An isomorphic Dvoretzky-Milman Theorem using general random ensembles

Shahar Mendelson *

October 28, 2020

Abstract

We construct rather general random ensembles that yield the optimal (isomorphic) estimate in the Dvoretzky-Milman Theorem. This is the first construction of non-gaussian/spherical ensembles that exhibit the optimal behaviour. The ensembles constructed here need not satisfy any rotation invariance and can be rather heavy-tailed.

1 Introduction

The Dvoretzky-Milman Theorem is one of the cornerstones of Asymptotic Geometric Analysis. The original result, due to Dvoretzky ([5]) is that $\ell_2$ is finitely represented in any infinite dimensional normed space. That means that given an infinite dimensional normed space $(E, \| \cdot \|)$, for any $\varepsilon > 0$ and integer $k$ there is a subspace $E_k \subset E$ of dimension $k$ such that $d(\ell_2^k, E_k) \leq 1 + \varepsilon$. Here $\ell_2^k = (\mathbb{R}^k, \| \cdot \|_2)$ and the distance is the Banach-Mazur distance, defined by

$$d(E, F) = \inf \{ \| T \|\| T^{-1} \| : T : E \to F \text{ is invertible} \}.$$ 

Dvoretzky’s result was actually quantitative, but with a suboptimal estimate. Milman, in his seminal work [10] (see also [2] for an extensive exposition), used the idea of concentration of measure, and specifically, the concentration of Lipschitz functions on the Euclidean sphere around their medians, to obtain the optimal finite dimensional estimate. Milman showed that each convex body $K \subset \mathbb{R}^n$ has a critical dimension $d^*(K)$ and for every $\varepsilon > 0$, a typical (with respect to the Haar measure on the grassmannian) subspace of $\mathbb{R}^n$ of dimension $s = c(\varepsilon)d^*(K)$ satisfies that $d(\ell_2^s, K \cap E) \leq 1 + \varepsilon$. Moreover, he showed that there is a linear image $K'$ of $K$ for which $d^*(K') \geq c\log n$ for an absolute constant $c$.

The following is the equivalent gaussian formulation of Milman’s result, due to Pisier ([12], see also [13]).

Let $K \subset \mathbb{R}^n$ be a convex body and denote by $\| \cdot \|_K$ the norm on $\mathbb{R}^n$ whose unit ball is $K$. Let $K^o$ be the dual body, i.e., for every $x \in \mathbb{R}^n$, $\| x \|_K = \sup_{t \in K^o} \langle x, t \rangle$. Let $(g_i)_{i=1}^n$ be independent standard gaussian random variables and set

$$\ell(K) = \mathbb{E} \left\| \sum_{i=1}^n g_i e_i \right\|_K = \mathbb{E} \sup_{t \in K^o} \sum_{i=1}^n g_i t_i.$$ 

*Mathematical Sciences Institute, The Australian National University. Email: shahar.mendelson@anu.edu.au

1A convex body in $\mathbb{R}^n$ is a convex, centrally symmetric set with a nonempty interior. There is an obvious equivalence between a convex body $K$ and the norm $\| \cdot \|_K$ whose unit ball is $K$. 

Definition 1.1. The critical dimension of $K$ is

$$d^*(K) = \left( \frac{\ell(K)}{\sup_{t \in K^\circ} \|t\|_2} \right)^2.$$ 

Here and in what follows, we denote by $B^n_2$ the unit ball of $\ell^n_2$, and $S^{n-1}$ is the Euclidean sphere in $\mathbb{R}^n$.

The gaussian version of Milman’s result is as follows:

Theorem 1.2. For every $\varepsilon > 0$ there exist constants $c_1$ and $c_2$ that depend on $\varepsilon$ such that the following holds. Let $K \subset \mathbb{R}^n$ be a convex body and set $d = c_1(\varepsilon)d^*(K)$. Let $T_g : \mathbb{R}^d \to \mathbb{R}^n$ be the random operator whose entries are independent, standard gaussian random variables and set $E = T_g \mathbb{R}^d$. Then with probability at least $1 - 2 \exp(-c_2(\varepsilon)d^*(K))$,

$$(1 - \varepsilon)\ell(K)(K \cap E) \subset T_g B^d_2 \subset (1 + \varepsilon)\ell(K)(K \cap E).$$

Remark 1.3. For reasons that will become clear immediately, the dependence on $\varepsilon$ is not of importance in the context of this article. However, it is a question that has been studied extensively over the years, and we refer the reader to the book [2] for a detailed discussion on that topic.

Note that Theorem 1.2 gives much more than the existence of an almost Euclidean subspace of $(\mathbb{R}^n, \|\|_K)$. Rather, it shows that a typical subspace (of the right dimension) with respect to the gaussian ensemble is almost Euclidean; Milman’s original proof shows that the same phenomenon is true with respect to the spherical ensemble.

To explain the role of the gaussian/spherical ensemble in Theorem 1.2 and highlight the difficulty in obtaining non-gaussian/non-spherical versions of that result, let us present an outline of the proof of Theorem 1.2.

The proof of Theorem 1.2 is remarkably simple: an operator $T : \ell^d_2 \to (\mathbb{R}^n, \|\|_K)$ is an almost isometric embedding of $\ell^d_2$ in $(\mathbb{R}^n, \|\|_K)$ if there is some $\alpha > 0$ such that for every $x \in S^{d-1}$,

$$(1 - \varepsilon)\alpha \leq \left\| \sum_{i=1}^d x_i T e_i \right\|_K \leq (1 + \varepsilon)\alpha; \quad (1.1)$$

it is an isomorphic embedding if $1 \pm \varepsilon$ is replaced by absolute constants $c$ and $C$. When $T$ is a gaussian operator, each $T e_i$ is distributed as $G$, the standard gaussian vector in $\mathbb{R}^n$, and the $d$ random vectors $\{T e_i, 1 \leq i \leq d\}$ are independent. Thanks to rotation invariance, for each $x \in S^{d-1}$, the random vector $\sum_{i=1}^d x_i T e_i$ has the same distribution as $\|x\|_2 G = G$, and by the gaussian concentration theorem (see, e.g. [13]), there is an absolute constant $c$ such that for any $u > 0$,

$$Pr \left( \|G\|_K - \ell(K) \geq u \right) \leq 2 \exp \left( -c \frac{u^2}{\sup_{t \in K^\circ} \|t\|_2^2} \right). \quad (1.2)$$

Setting $u = \varepsilon \ell(K)$, it follows from (1.2) that for any $x \in S^{d-1}$, (1.1) holds with probability at least $1 - 2 \exp(-\varepsilon^2 d^*(K))$ for $\alpha = \ell(K)$; the high probability estimate allows one to have uniform control on an $\varepsilon$-net in $S^{d-1}$; and thanks to standard convexity and approximation arguments the uniform estimate on $S^{d-1}$ can be ensured. We refer the reader to [13] for the complete proof.
While this proof is undeniably beautiful, it is also rather restrictive. It is based on two crucial features of the Gaussian vector: that for every \( x \in S^{d-1} \) the random variable \( \left\| \sum_{i=1}^{d} x_i T e_i \right\|_K \) concentrates around its mean, and that, by rotation invariance, the means \( \mathbb{E} \left\| \sum_{i=1}^{d} x_i T e_i \right\|_K \) are the same (or, for an isomorphic estimate, are equivalent). These features are shared by the spherical ensemble.

In what follows, we refer (somewhat inaccurately) to a random ensemble \( T \) as an optimal Dvoretzky-Milman ensemble if it exhibits the fact that an arbitrary convex body \( K \) has a subspace of dimension proportional to \( d^* (K) \) that is isomorphic (with absolute constants) to a Euclidean space. More accurately, an optimal Dvoretzky-Milman ensemble \( T \) satisfies that for every convex body \( \mathbb{K} \subset \mathbb{R}^n \), with probability at least \( 1/2 \), for \( d = c_0 d^* (K) \) and every \( x \in \mathbb{R}^d \),

\[
 c \ell (K) \| x \|_2 \leq \| T x \|_K \leq C \ell (K) \| x \|_2. \tag{1.3}
\]

Here, \( c_0, c \) and \( C \) are constants that might depend on properties of the ensemble but are independent of \( \mathbb{K}, n \) or \( d^* \).

As surprising as it may seem, it is hard to obtain a behaviour like (1.3) when leaving the Gaussian/spherical realm, even if the ensemble is generated by independent copies of a very well behaved random variable—as the next example shows.

**Example 1.4.** Let \( \mathbb{K} = B_{\infty}^n \), the unit cube in \( \mathbb{R}^n \); its dual body is \( B_1^n \), the unit ball in \( \ell_1^n = (\mathbb{R}^n, \| \cdot \|_1) \). It is standard to verify that

\[
 \ell (B_{\infty}^n) = \mathbb{E} \max_{1 \leq i \leq n} | g_i | \sim \sqrt{\log n},
\]

and clearly, \( \sup_{t \in B_1^n} \| t \|_2 = 1 \). Therefore, the critical dimension of the unit cube is \( d^* (B_{\infty}^n) \sim \log n \). By Theorem 1.2, there are absolute constants \( \delta > 0, c_0, c \) and \( C \), such that for \( d = c_0 \log n \), with probability at least \( 1 - 2/n^\delta \),

\[
 c \sqrt{\log n} \leq \inf_{x \in S^{d-1}} \| T g x \|_{\infty} \leq \sup_{x \in S^{d-1}} \| T g x \|_{\infty} \leq C \sqrt{\log n}.
\]

Thus, on that event, the random function \( \psi (x) = \| T g x \|_K \) is equivalent to \( \sqrt{\log n} \) on \( S^{d-1} \).

On the other hand, let \( \zeta \) be uniformly distributed in \([-1, 1]\). Set

\[
 T_\zeta = \{ (\zeta_{ij}) : 1 \leq i \leq n, 1 \leq j \leq d \}
\]

to be the random operator whose entries are independent copies of \( \zeta \).

Note that here we no longer have rotation invariance, and it turns out that with high probability, the random function

\[
 \phi (x) = \| T_\zeta x \|_\infty = \left\| \sum_{j=1}^{d} x_j T_\zeta e_j \right\|_{\infty}
\]
is not equivalent to one value on $S^{d-1}$. Indeed, for $y_1 = (1, 0, \ldots, 0)$, $\phi(y) \leq 1$ almost surely. At the same time, for $\eta = (\eta_j) \in \{-1, 1\}^d$ let $y_\eta = (\eta_j / \sqrt{d})_{j=1}^d \in S^{d-1}$, and

$$\phi(y_\eta) = \frac{1}{\sqrt{d}} \max_{i=1, \ldots, n} \left| \sum_{j=1}^d \eta_j \zeta_{ij} \right|.$$  

Denote by $|A|$ the cardinality of a set $A$ and observe that

$$\Pr \left( \{\exists i : |\{j : |\zeta_{ij}| \geq 1/2\}| \geq d/10\} \right) \geq 1 - 2 \exp(-c_1 dn)$$  

(1.4)

for an absolute constant $c_1$. For every realization of the event from (1.4), let $i^*$ be the corresponding index and set $\eta_j = \text{sgn}(\zeta_{i^*j})$ for $1 \leq j \leq d$. Clearly,

$$\phi(y_\eta) = \frac{1}{\sqrt{d}} \sum_{j=1}^d |\zeta_{i^*j}| \geq c_2 \sqrt{d}$$

for an absolute constant $c_2$, and recalling that $d = c_0 \log n$, we have that with probability at least $1 - 2 \exp(-c_0 n \log n)$

$$\inf_{x \in S^{d-1}} \left\| \sum_{i=1}^d x_i \zeta e_i \right\|_\infty \leq 1 \quad \text{and} \quad \sup_{x \in S^{d-1}} \left\| \sum_{i=1}^d x_i \zeta e_i \right\|_\infty \geq \sqrt{c_2 \log n}.$$  

This example illustrates a rather surprising fact: although 50 years have passed since Milman’s proof, there are no examples of random ensembles that are optimal Dvoretzky-Milman ensembles—other than the gaussian ensemble or the spherical one. The few known “non gaussian” versions of an isomorphic Dvoretzky-Milman Theorem hold only for convex bodies $K$ that have special structure, like cotype 2 (see, e.g. [9]).

Here we present a rather general construction that yields optimal Dvoretzky-Milman ensembles. The random operators need not satisfy rotation invariance of any kind; nor do they exhibit a gaussian-like concentration; in fact, they can be rather heavy-tailed.

Before presenting a more accurate description of the construction, let us present two examples. A third example is presented in Section 2.3.

**Theorem 1.5.** There exist absolute constants $c_0, c_1, c_2, c$ and $C$ such that the following holds: Let $K \subset \mathbb{R}^n$ and set $d^*(K)$ to be its critical dimension. Let $d = c_0 d^*(K)$ and $m = c_1 n$, and set

$$\Gamma_1 = m^{-1/2}(\zeta_{ij}) : \mathbb{R}^n \to \mathbb{R}^m, \quad \Gamma_2 = (\zeta'_{ij}) : \mathbb{R}^m \to \mathbb{R}^d,$$

where $\{(\zeta_{ij}) : 1 \leq i \leq m, \ 1 \leq j \leq n\}$ and $\{(\zeta'_{ij}) : 1 \leq i \leq d, \ 1 \leq j \leq m\}$ are independent random variables that are uniformly distributed in $[-1, 1]$. Define

$$\Gamma : \Gamma_1 \Gamma_2^* : \mathbb{R}^d \to \mathbb{R}^n.$$

Then, with probability at least $1 - 2 \exp(-c_2 d^*(K))$, for every $x \in S^{d-1}$,

$$c\ell(K) \leq \|\Gamma x\|_K \leq C\ell(K).$$
Theorem 1.5 recovers the isomorphic Dvoretzky-Milman Theorem using the product of two matrices of the type $T_\zeta$ of appropriate dimensions. However, as Example 1.4 shows, a single matrix of the same type is not a suitable choice.

Before we formulate the second example, let us recall some standard definitions.

**Definition 1.6.** Let $Y$ be a centred random vector in $\mathbb{R}^\ell$.

- $Y$ is isotropic if its covariance is the identity; that is, if for every $t \in \mathbb{R}^\ell$
  \[ \mathbb{E} \langle Y, t \rangle^2 = \| t \|_2^2. \]

- $Y$ is log-concave if has a density that is a log-concave function.

- $Y$ is $L$-subgaussian if for any $t \in \mathbb{R}^\ell$ and every $p \geq 2$,
  \[ \| \langle Y, t \rangle \|_{L_p} \leq L \sqrt{p} \| \langle Y, t \rangle \|_{L_2}. \]

- Let $q > 2$. $Y$ satisfies an $L_q - L_2$ norm equivalence with constant $L$ if for every $t \in \mathbb{R}^\ell$,
  \[ \| \langle Y, t \rangle \|_{L_q} \leq L \| \langle Y, t \rangle \|_{L_2}. \]

Observe that if, in addition to being $L$-subgaussian, $Y$ is isotropic, then for every $t \in \mathbb{R}^\ell$ and $p \geq 1$,
\[ \| \langle Y, t \rangle \|_{L_p} \leq L \sqrt{p} \| t \|_2. \] Clearly, the analogous observation is true if $Y$ satisfies $L_q - L_2$ norm equivalence with constant $L$.

**Remark 1.7.** To put these definitions in some context, the random vector $Y = (\zeta_i)_{i=1}^\ell$ is isotropic; $L$-subgaussian for an absolute constant $L$; and log-concave.

In what follows, all the random vectors we consider are symmetric and isotropic. Given integers $n$ and $d$, we consider random vectors $Z$ in $\mathbb{R}^n$ and random vectors $X$ in $\mathbb{R}^d$. For $m \geq n$ let $Z_1, ..., Z_m$ be independent copies of $Z$ and set $X_1, ..., X_m$ to be independent copies of $X$. Define $\Gamma_1$ to be the random matrix whose rows are $Z_i/\sqrt{m}, 1 \leq i \leq m$, and set $\Gamma_2$ to be the matrix whose columns are $X_1, ..., X_m$. Finally, let
\[ \Gamma = \Gamma_1^* \Gamma_2^*: \mathbb{R}^d \to \mathbb{R}^n. \] (1.5)

As we explain, if the rows of $\Gamma_1$ and the columns of $\Gamma_2$ satisfy an appropriate mixture of tail and small-ball properties then $\Gamma = \Gamma_1^* \Gamma_2^*$ is an optimal Dvoretzky-Milman ensemble.

The second outcome of our main result is as follows:

**Theorem 1.8.** For every $L, L' \geq 1$ there are constants $c_0, ..., c_2, c$ and $C$ that depend on $L$ and $L'$ such that the following holds. Consider a convex body $K \subset \mathbb{R}^n$, set $d = c_0 d^*(K)$ and put $m = c_1 n$. Let $Z$ be an isotropic, $L$-subgaussian random vector in $\mathbb{R}^n$ and let $X$ be an isotropic, log-concave random vector in $\mathbb{R}^d$. Assume that the densities of all of the one-dimensional marginals $\langle Z, t \rangle$ and $\langle X, v \rangle$ are bounded by $L'$.

If $\Gamma$ is as in (1.5), then with probability at least
\[ 1 - 2 \exp \left( -c_2 \min \{ (nd^*(K))^{1/4}, d^*(K) \} \right), \]
for every $x \in S^{d-1}$,
\[ c\ell(K) \leq \| \Gamma x \|_K \leq C\ell(K). \]
Remark 1.9. Although $Z$ and $X$ need not have iid coordinates, they still share some of the features of $\zeta_i$. In particular, the one-dimensional marginals of $X$ exhibit a fast tail decay. We shall present another example in Section 2.3 which shows that $X$ can be heavy-tailed.

Let us turn to the formulation of the main result of this note, beginning with the required features of $Z$ and $X$.

1.1 The Assumptions

The first assumption we require is that both $X$ and $Z$ satisfy a rather weak small-ball property:

**Assumption 1.1.** There is a constant $\kappa_0$ such that for every $t \in \mathbb{R}^n$ and every $v \in \mathbb{R}^d$, 

$$
\Pr(|\langle Z, t \rangle| \leq \kappa_0 \|t\|_2) \leq \frac{1}{1000}
$$

and

$$
\Pr(|\langle X, v \rangle| \leq \kappa_0 \|v\|_2) \leq \frac{1}{1000}.
$$

**Remark 1.10.** The choice of $1/1000$ is completely arbitrary; in fact, any uniform probability estimate strictly smaller than $1/2$ suffices for our purposes.

The second assumption deals with the tail behaviour of one-dimensional marginals of $X$ and $Z$.

**Assumption 1.2.** Assume that

1. $Z$ is $L$-subgaussian.

2. $X$ satisfies an $L_q - L_2$ norm equivalence for some $q > 2$ with constant $L$. In particular, since $X$ is isotropic, it follows that for every $v \in \mathbb{R}^d$, $\| \langle X, v \rangle \|_{L_q} \leq L \|v\|_2$.

In itself, Assumption 1.2 is not enough and additional information on $\Gamma_2$ is required. To formulate this final assumption, denote by $\|a\|_0$ the cardinality of the support of the vector $a$, i.e., $\|a\|_0 = |\{i : a_i \neq 0\}|$.

**Assumption 1.3.** For $\kappa_1 \geq 1$ and well chosen $0 < \delta, \theta < 1/4$, there is a nontrivial event $A$ such that for any $(X_i)_{i=1}^m \in A$,

$$
\sup_{v \in B^d_2} \|\Gamma_2 v\|_2 \leq \kappa_1 \sqrt{m},
$$

(1.6)

and

$$
\sup_{\{a \in \mathbb{R}^m : \|a\|_2 \leq 1, \|a\|_0 \leq \theta m\}} \left\| \sum_{i=1}^m a_i X_i \right\|_2 \leq \delta \sqrt{m}.
$$

(1.7)

**Remark 1.11.** Assumption 1.3 has strong ties to the study of the extremal singular values of $\Gamma_2$. Thanks to these connections, there are many scenarios in which (1.6) and (1.7) are known to be true on a large event. Indeed, (1.6) implies that the largest singular value of $\Gamma_2/\sqrt{m}$ satisfies $\lambda_{\text{max}}(\Gamma_2/\sqrt{m}) \leq \kappa_1$. And in fact, in all the examples we present here, $\kappa_1 = 2$ suffices: under rather mild assumptions on $X$, for any $m \geq d$, with high probability

$$
\lambda_{\text{max}}(\Gamma_2/\sqrt{m}) \leq 1 + c \sqrt{d/m}.
$$

The second condition in the definition of $A$, Equation (1.7), is closely related to Bai-Yin type estimates on $\lambda_{\text{max}}(\Gamma_2/\sqrt{m})$ and $\lambda_{\text{min}}(\Gamma_2/\sqrt{m})$ (see, e.g., [1, 8, 15]).
Although the event $A$ depends only on $(X_i)^m_{i=1}$, we view it with respect to the product $(X_i, Z_i)^m_{i=1}$. Note that the event actually depends on five parameters: $m$, $d$ (because $X$ is a random vector in $\mathbb{R}^d$), $\kappa_1$, $\delta$ and $\theta$. While $\kappa_1$ can be arbitrary (and, as noted previously, in most interesting cases, $\kappa_1 = 2$ suffices), a suitable choice of $\delta$ and $\theta$ requires more care: for the embedding result to hold, these parameters need to satisfy several constraints that depend on the values of $\kappa_0$, $\kappa_1$, $L$ and $q$. As a result, the event $A$ should be understood as corresponding to a fixed choice of $m$, $d$, $\kappa_1$, $\delta$ and $\theta$; the proof will dictate the constraints on $m$, $d$, $\delta$ and $\theta$, and one has to show that for a set of parameters that satisfy those constraints the event $A$ has nontrivial probability. To be more accurate:

- The proof of our main result is based on a net argument. From here on, fix $0 < \rho < 1/4$ and denote by $V$ a maximal $\rho$-separated subset of $S^{d-1}$ with respect to $\ell_2$ norm. In what follows, we shall call such a set a $\rho$-net. By a volumetric estimate

$$|V| \leq \exp(d \log(5/\rho)),$$

a fact we use frequently.

The right value of $\rho$ turns out to be a constant that depends on $\kappa_0$, $\kappa_1$, $L$ and $q$ (see the proof of Theorem 1.12 below).

- Given a suitable choice of $\rho$, $m$ has to satisfy that

$$m \geq c_0 \max \left\{ \frac{d^*(K)}{\rho}, n \right\}$$

for a constant $c_0$ that depends on $\kappa_0$, $\kappa_1$, $L$ and $q$.

- The parameters $\delta$ and $\theta$ must satisfy that for suitable constants $c_1$ and $c_2$ that depend on $\kappa_0$, $\kappa_1$, $L$ and $q$,

$$\delta \leq \frac{c_1}{\log(5/\rho)}, \quad \text{and} \quad \theta^{\frac{q-2}{2(q+2)} \sqrt{\log(e/\theta)}} \leq \frac{c_2}{\log(5/\rho)}.$$  \hspace{1cm} (1.9)

- At the same time, for values of $\delta$ and $\theta$ that satisfy (1.9), the dimension $d$ of the subspace that is close to Euclidean will be

$$d = c_3(\kappa_0, L)(\theta^{4/(2+q)} \log^5(5/\rho))d^*(K).$$  \hspace{1cm} (1.10)

1.2 The main result

We show that

**Theorem 1.12.** There are constants $c$, $c'$ and $C$ that depend on $\kappa_0$, $\kappa_1$, $L$ and $q$ such that the following holds. Set $m$ as in (1.8), $\theta$ and $\delta$ as in (1.9) and $d$ as in (1.10). Then with probability at least $\Pr(A) - 2\exp(-c'd^*(K))$, for every $x \in S^{d-1}$

$$c\ell(K) \leq \|\Gamma x\| \leq C\ell(K).$$
Writing $\langle \Gamma v, t \rangle = \langle \Gamma_2^* v, \Gamma_1 t \rangle$, Theorem 1.12 follows once we show that there is an event $B$ of probability at most $2 \exp(-c'd^*(K))$ such that on $A \setminus B$,

$$\sup_{v \in B_2^d} \sup_{t \in K^\circ} \langle \Gamma_2^* v, \Gamma_1 t \rangle \leq C\ell(K),$$

and

$$\inf_{v \in S^{d-1}} \sup_{t \in K^\circ} \langle \Gamma_2^* v, \Gamma_1 t \rangle \geq c\ell(K).$$

To that end, let $0 < \rho < 1/4$ be specified later. By convexity (for the standard argument, see, e.g., [2]), it is enough to establish the upper estimate on $U$ that is a $1/5$-net in $B_2^d$; and by a volumetric estimate, there is a $1/5$-net in $B_2^d$ whose cardinality is at most $\exp(10d)$.

Thus, for the upper bound it suffices to show that:

**Theorem 1.13.** There are constants $c, c', C_0$ and $C$ that depend only on $\kappa_1, L$ and $q$ for which the following holds. Let $d \leq n$ and $m \geq C_0 n$. There is an event $B_1$ of probability at most $2 \exp(-c'd)$ such that on $A \setminus B_1$, for every $v \in U$,

$$\sup_{t \in K^\circ} \langle \Gamma_2^* v, \Gamma_1 t \rangle \leq C\ell(K).$$

The reverse inequality is far more delicate. The key step is to obtain a uniform lower bound on the $\rho$-net $V \subset S^{d-1}$.

**Theorem 1.14.** There are constants $c_0, \ldots, c_4$ that depend on $\kappa_0, \kappa_1, L$ and $q$ such that the following holds. Fix $0 < \delta, \theta < 1/4$, let

$$m \geq c_0 \max\{d^*(K)/\rho, n\} \quad \text{and} \quad d = c_1 (\theta^{4/(2+q)} / \log^5(5/\rho)) d^*(K).$$

There is an event $B_2$ such that $\Pr(B_2) \leq 2 \exp(-c_2 d \log(5/\rho))$ and on $A \setminus B_2$, for every $v \in V$,

$$\sup_{t \in K^\circ} \langle \Gamma_2^* v, \Gamma_1 t \rangle \geq \left(c_3 - c_4 \log(5/\rho) \left(\delta + \theta^{(q-2)/(2+q)} \sqrt{\log(e/\theta)}\right)\right) \ell(K).$$

In particular, if

$$\delta \leq \frac{c_3}{8c_4 \log(5/\rho)}, \quad \text{and} \quad \theta^{(q-2)/(2+q)} \sqrt{\log(e/\theta)} \leq \frac{c_3}{8c_4 \log(5/\rho)},$$

then on the same event, for every $v \in V$,

$$\sup_{t \in K^\circ} \langle \Gamma_2^* v, \Gamma_1 t \rangle \geq \frac{c_3}{2} \ell(K).$$

The combination of Theorem 1.13 and Theorem 1.14 leads to the proof of Theorem 1.12.

**Proof of Theorem 1.12.** Since the conditions on $d$ and $m$ imposed in Theorem 1.14 are more restrictive than those in Theorem 1.13 we choose $d$ and $m$ as in the latter. All that remains is to prove the uniform lower bound and determine a suitable choice of $\rho$.

Let $B_1$ and $B_2$ be the events from Theorem 1.13 and Theorem 1.14, respectively.
For every $x \in S^{d-1}$ let $v \in V$ satisfy that $\|x - v\|_2 \leq \rho$. On $A \setminus (B_1 \cup B_2)$, we have, using the same notation as in the two theorems,

$$
\sup_{t \in K^o} \langle \Gamma_1^* x, t \rangle \geq \sup_{t \in K^o} \langle \Gamma_2^* v, \Gamma_1 t \rangle - \sup_{t \in K^o} \langle \Gamma_2^* (x - v), \Gamma_1 t \rangle \geq \left( \frac{c_3}{2} - C \rho \right) \ell(K) \geq \frac{c_3}{4} \ell(K),
$$

provided that $\rho \leq c_3/4C$. Since $c_3$ and $C$ depend on $\kappa_0, \kappa_1, L$ and $q$, so does $\rho$; and as a result so do $\theta, \delta, m$ and $d$ with the choices specified in (1.9), (1.8) and (1.10).

In Section 4 we prove Theorem 1.13 and Theorem 1.14 is proved in Section 5. The proofs are based on Talagrand’s generic chaining mechanism and properties of Bernoulli processes that are outlined in Section 3. A key component of the proof of Theorem 1.14 is that the supremum of certain Bernoulli processes, indexed by random subsets of $\mathbb{R}^m$, actually dominate the expected supremum of their gaussian counterparts, and do so with high probability. This type of domination goes in the opposite direction of what one usually expects. It holds here only because of the special structure of the (random) indexing sets, and the role of $\Gamma_1$ and $\Gamma_2$ is to generate that special structure with sufficiently high probability.

Section 2 is devoted to the proofs of Theorem 1.5, Theorem 1.8 and to another example, presented in Section 2.3. The latter shows that $X$ can be heavy-tailed, but still $\Gamma$ is an optimal Dvoretzky-Milman ensemble.

We end this introduction with a word about notation. Throughout, absolute constants are denoted by $c, C, c', \ldots$. Their value may change from line to line, though they remain unchanged within each proof/statement. Unless mentioned otherwise, all constants are absolute—meaning that they are just positive numbers that do not depend on any of the parameters of the problem. If a constant does depend on some parameter, that will be made explicit by writing $c(L), c(L, q), \ldots$. We write $a \lesssim b$ if there is an absolute constant $c$ such that $a \leq cb$, and $a \sim b$ means that $a \lesssim b$ and $b \lesssim a$.

2 Examples

Let us return to the two examples mentioned previously. The key is to show that in both cases $\mathcal{A}$ is a large event.

2.1 Proof of Theorem 1.5

Recall that $\zeta$ is distributed uniformly in $[-1, 1]$ and that $X = (\zeta_i)_{i=1}^d$ and $Z = (\zeta_j')_{i=1}^n$ have independent coordinates distributed as $\zeta$. Observe that $X$ and $Z$ are isotropic random vectors that are $L$-subgaussian for an absolute constant $c$. Also, it is standard to verify that they satisfy the small ball property of Assumption 1.1. Therefore, both $\kappa_0$ and $L$ are absolute constants in this case.

All that remains is to show that $\mathcal{A}$ has high probability when $X$ is an arbitrary $L$-subgaussian random vector. As a result, the assertion of Theorem 1.5 actually holds when $X$ and $Z$ are subgaussian random vectors that satisfy the small-ball property. In such a case, for $c = c(\kappa_0, L)$, with probability at least $1 - 2 \exp(-cd^*(K))$, $\Gamma$ is an optimal Dvoretzky-Milman ensemble.

Before turning to the proof, let us describe certain features of subgaussian random variables. The proofs of all these facts are standard and can be found, for example, in [7, 16, 4].
Definition 2.1. Let $\xi$ be a centred random variable. For $1 \leq \alpha \leq 2$, the $\psi_\alpha$ norm of $\xi$ is defined by

$$\inf \{ c > 0 : \mathbb{E} \exp(|\xi/c|^\alpha) \leq 2 \}.$$ 

There is a well-known equivalence between the $\psi_\alpha$ norm of a random variable, the growth of its moment and its tail behaviour:

Theorem 2.2. Let $1 \leq \alpha \leq 2$ and consider a centred random variable $\xi$. The following are equivalent:

1. $\|\xi\|_{\psi_\alpha} \leq L$;
2. For every $p \geq 2$, $\|\xi\|_{L^p} \leq L'^{1/\alpha}$;
3. For every $u \geq 1$, $\Pr(|\xi| \geq L''u) \leq 2 \exp(-u^\alpha)$.

Moreover, the constants $L$, $L'$ and $L''$ are equivalent in the sense that $L \sim L' \sim L''$.

Next, let us turn to a useful feature of independent random variables with bounded $\psi_2$ norms:

Lemma 2.3. There exists an absolute constant $c$ such that the following holds. Let $\xi_1, \ldots, \xi_\ell$ be independent, centred random variables. Then for any $a \in \mathbb{R}^\ell$,

$$\| \sum_{i=1}^\ell a_i \xi_i \|_{\psi_2} \leq c \left( \sum_{i=1}^\ell a_i^2 \| \xi_i \|_{\psi_2}^2 \right)^{1/2}.$$ 

Clearly, a centred random vector $Y$ in $\mathbb{R}^\ell$ is $L$-subgaussian if, for every $t \in \mathbb{R}^\ell$, $\| \langle Y, t \rangle \|_{\psi_2} \lesssim L \| Y \|_{L_2}$.

Lemma 2.4. For every $L \geq 1$ there is a constant $C$ that depends on $L$ such that the following holds. Let $X_1, \ldots, X_m$ be independent copies of an isotropic, $L$-subgaussian random vector $X$ in $\mathbb{R}^d$. Then for $1 \leq k \leq m$, with probability at least $1 - 2 \exp(-c_0 m)$,

$$\sup_{\{a \in \mathbb{R}^m : \|a\|_2 \leq 1, \|a\|_0 \leq k\}} \left\| \sum_{i=1}^m a_i X_i \right\|_2 \leq C(L) \left( \sqrt{d+k \log(em/k)} + t \right).$$
Consider $k = \theta m$ such that $k \log(em/k) \geq d$. Setting $t = \sqrt{k \log(em/k)}$ it follows that
\[
\sup_{\{a \in \mathbb{R}^m: \|a\|_2 \leq 1, \|a\|_0 \leq \theta m\}} \left\| \sum_{i=1}^{m} a_i X_i \right\|_2 \leq c_2(L) \sqrt{\theta \log(e/\theta) \sqrt{m}}
\]
with probability at least $1 - 2 \exp(-\theta m \log(e/\theta))$. Hence, in the context of [L7], for any $0 < \delta \leq 1/4$ one may select any $\theta \leq \delta^2/c^2_2 \log(e/\delta)$.

Returning to Theorem [L5] note that $L$ is an absolute constant and that $\kappa_1 = 2$. Moreover, $\delta$ and $\theta$ are constants that depend on $L$, and $m = cn$, $d = c'd'(K)$ for constants $c$ and $c'$ that also depend on $L$—which means they are all absolute constants. As a result, $\Pr(A) \geq 1 - 2 \exp(-c''d''(K))$ and $\Gamma$ is an optimal Dvoretzky-Milman ensemble with probability at least $1 - 2 \exp(-c'd''(K))$, as claimed.

**Proof of Lemma [2.4]:** Set $U_k = \{a \in \mathbb{R}^m : \|a\|_2 \leq 1, \|a\|_0 \leq k\}$ and note that
\[
\sup_{a \in U_k} \left\| \sum_{i=1}^{m} a_i X_i \right\|_2 = \sup_{v \in B_2^d} \sup_{a \in U_k} \sum_{i=1}^{m} a_i \langle X_i, v \rangle.
\]
By convexity, it suffices to control the supremum over two $1/10$-nets: $\tilde{V} \subset B_2^d$ and $\tilde{U} \subset U_k$. Given $I \subset \{1, \ldots, m\}$, let $B_{2}^{I}$ be the set of points in $B_2^m$ that are supported on $I$ and observe that $U_k = \bigcup_{|I|=k} B_{2}^{I}$. Therefore, by a volumetric estimate, $|\tilde{V}| \leq \exp(c_0 d)$ and $|\tilde{U}| \leq \exp(c_0 k \log(em/k))$ for an absolute constant $c_0$. Fix $v \in \tilde{V}$ and $u \in \tilde{U}$ and by (2.1), the random variable $y = \sum_{i=1}^{m} u_i \langle X_i, v \rangle$ satisfies that $\|y\|_{L_2} \leq cL \|v\|_2 \|u\|_3 \leq cL$. Using the equivalent formulation of the $\varphi_2$ norm from Lemma [2.2] we have that $Pr(\|y\| \geq c'Lx) \leq 2 \exp(-x^2)$. The proof is completed by setting $x = c_1 \max\{\sqrt{d}, \sqrt{k \log(em/k)}\} + t$ followed by the union bound.

### 2.2 Proof of Theorem [1.8]

Let $X$ be an isotropic, log-concave random vector in $\mathbb{R}^d$. By Borell’s Lemma (see, e.g., [3]), $X$ satisfies $\psi_1 - L_2$ norm equivalence with an absolute constant $C$. In particular, for any $q > 2$, $X$ satisfies an $L_q - L_q$ norm equivalence with constant $Cq$. Set, for example, $q = 6$ and the required norm equivalence in Assumption [1.2] is satisfied for an absolute constant.

The assumption that all the one-dimensional marginals $\langle X, v \rangle$ and $\langle Z, t \rangle$ have densities that are uniformly bounded by $L'$ implies that Assumption [1.1] holds as well. Following the notation used in the two assumptions we denote the constants by $L$ and $\kappa_0$ respectively (and clearly $\kappa_0$ depends only on $L'$).

As in the proof of Theorem [1.5], the difficulty is in verifying that $A$ has large probability. Thankfully, that has been addressed as part of the study of the extremal singular values of random matrices with iid log-concave rows (columns). The wanted estimate was established in [4], with a minor but important caveat from our perspective: an upper bound on $m$ as a function of $d$. That caveat was removed in [14], (see Thm 11.4.1 and Prop. 11.4.4):

**Theorem 2.5.** There exist absolute constants $c$ and $C$ such that the following holds. Let $Y$ be an isotropic, log concave random vector in $\mathbb{R}^d$, and let $Y_1, \ldots, Y_m$ be independent copies of $Y$. Then for $1 \leq k \leq m$ and $u \geq 1$, with probability at least $1 - c \exp(-u)$,
\[
\sup_{\{a \in \mathbb{R}^m: \|a\|_2 \leq 1, \|a\|_0 \leq k\}} \left\| \sum_{i=1}^{m} a_i Y_i \right\|_2 \leq C \left( u + \sqrt{k \log(em/k)} + \max_{1 \leq i \leq m} \|Y_i\|_2 \right).
\]
Moreover, if \( T \) is the random matrix whose rows are \( (Y_i/\sqrt{m})_{i=1}^m \), then conditioned on the event \( \max_{1 \leq i \leq m} \|Y_i\|_2 \leq (m\ell)^{1/4} \), with probability at least \( 1 - c\exp(-\ell) - \exp(-(m\ell)^{1/4}) \), we have

\[
1 - c' \sqrt{\frac{\ell}{m}} \leq \lambda_{\min}(T) \leq \lambda_{\max}(T) \leq 1 + c' \sqrt{\frac{\ell}{m}}.
\]

The second result we require here is due to Paouris [11]. The formulation is from [3]:

**Theorem 2.6.** There exists an absolute constant \( C_1 \) such that, for any integer \( \ell \), any isotropic log-concave random vector \( Y \) on \( \mathbb{R}^{\ell} \) and any \( u \geq 1 \),

\[
Pr(\|Y\|_2 \geq C_1 u\sqrt{\ell}) \leq 2 \exp(-u\sqrt{\ell}).
\]

The combination of Theorem 2.5 and Theorem 2.6 clearly implies the following:

**Corollary 2.7.** There exist absolute constants \( c \) and \( C \) such that, with probability at least \( 1 - 2\exp(-c\min\{(\ell m)^{1/4}, \ell\}) \), for every \( 1 \leq k \leq m \),

\[
\sup_{\{a \in \mathbb{R}^m : \|a\|_2 \leq 1, \|a\|_\infty \leq k\}} \left\| \sum_{i=1}^m a_i Y_i \right\|_2 \leq C \left( (\ell m)^{1/4} + \sqrt{k \log(em/k)} \right),
\]

and

\[
1 - c' \sqrt{\frac{\ell}{m}} \leq \lambda_{\min}(T) \leq \lambda_{\max}(T) \leq 1 + c' \sqrt{\frac{\ell}{m}}.
\]

**Proof.** Fix \( 1 \leq k \leq m \) and set \( u = (m/\ell)^{1/4} \geq 1 \). By Theorem 2.6, with probability at least \( 1 - 2m \exp(-(m\ell)^{1/4}) \geq 1 - 2 \exp(-c(m\ell)^{1/4}) \), \( \max_{1 \leq i \leq m} \|Y_i\|_2 \leq (m\ell)^{1/4} \). Therefore, applying Theorem 2.5 for any \( 1 \leq k \leq m \), it follows that with probability at least \( 1 - 2 \exp(-c_1(m\ell)^{1/4}) \),

\[
\sup_{\{a \in \mathbb{R}^m : \|a\|_2 \leq 1, \|a\|_\infty \leq k\}} \left\| \sum_{i=1}^m a_i Y_i \right\|_2 \leq C \left( \sqrt{k \log(em/k)} + (m\ell)^{1/4} \right), \tag{2.2}
\]

The proof of the first claim is concluded by the union bound with respect to \( 1 \leq k \leq m \). The proof of the second claim is immediate. \( \blacksquare \)

Finally, set \( \ell = d \) and let \( m \geq n \) to be specified in what follows. Consider the event from Corollary 2.7. By the second claim from that corollary we may set \( \kappa_1 = 2 \).

Now, given any \( 0 < \delta < 1/4 \), let \( m \geq (2C/\delta)^4 d^* (K) \) (and recall that one has the freedom to choose \( m \) that is large enough). Set \( \theta \) to satisfy that \( (Ch \log(e/\theta)) \leq \delta/2 \) and therefore,

\[
\sup_{\{a \in \mathbb{R}^m : \|a\|_2 \leq 1, \|a\|_\infty \leq \theta m\}} \left\| \sum_{i=1}^m a_i X_i \right\|_2 \leq \delta.
\]

Hence, it is evident that \( Pr(\mathcal{A}) \geq 1 - 2 \exp(-c_2 \min\{(nd)^{1/4}, d\}) \), as claimed. \( \blacksquare \)
2.3 A heavy-tailed example

In the final example we present, the random vector $X$ can be heavy-tailed.

**Theorem 2.8.** For $\kappa_0$, $L$ and $q > 8$ there exist constants $c_0$ and $c_1$ that depend on $\kappa_0$, $L$ and $q$ such that the following holds. Let $K \subset \mathbb{R}^n$ be a convex body. Set $d = c_0 d(K^*)$ and $m = c_1 n$. Assume that

1. $Z$ is an isotropic, $L$-subgaussian random vector in $\mathbb{R}^n$.
2. $X$ is an isotropic random vector in $\mathbb{R}^d$ that satisfies $L_q$-$L_2$ norm equivalence with constant $L$ for some $q > 8$.
3. $\|X\|_2 \leq 100\sqrt{d}$ almost surely.
4. $Z$ and $X$ satisfy the small-ball property with constant $\kappa_0$.

Then for $\beta \geq 1$, with probability at least $1 - c_2/m$, we have that

$$\sup_{v \in B_2^d} \|\Gamma_2\| \leq 2\sqrt{m},$$

and

$$\sup_{\{\|a\|_2 \leq 1, \|a\|_0 \leq \theta m\}} \left\| \sum_{i=1}^m a_i X_i \right\|_2 \leq c_3 \left( \sqrt{\beta \sqrt{d}} + \theta^{1/4} \sqrt{m} \right)$$

for absolute constant $c_2$ and $c_3$.

In particular, with probability $1 - c_2/m - 2 \exp(-cd)$, $\Gamma$ is an optimal Dvoretzky-Milman ensemble.

**Remark 2.9.** Clearly, the constant 100 in (3) is arbitrary. Moreover, the assumption that $X$ has a bounded diameter almost surely can be relaxed considerably: all that is needed is that the event \{max_{1 \leq i \leq m} \|X_i\|_2 \leq (md)^{1/4}\} has nontrivial probability. For the sake of simplicity we shall not pursue this further and keep the bounded diameter assumption.

The proof of Theorem 2.8 has appeared implicitly in [8], where it was shown that random matrix $\Gamma_2$ satisfies a Bai-Yin estimate: with probability at least $1 - c/m^\beta$,

$$\sup_{v \in S^{d-1}} \left| \frac{1}{m} \sum_{i=1}^m (X_i, v)^2 - 1 \right| \leq c' \sqrt{\frac{d}{m}}.$$

Clearly, on that event $\sup_{v \in B_2^d} \|\Gamma_2 v\|_2 \leq 2\sqrt{m}$.

Moreover, the crucial part in the proof of the Bai-Yin estimate from [8] is an upper estimate on

$$\sup_{\{\|a\|_2 \leq 1, \|a\|_0 \leq k\}} \left\| \sum_{i=1}^m a_i X_i \right\|_2$$

for a specific choice of $k$. A careful examination of the proof (see Lemma 2.3 and Theorem 2.4 in [8]) reveals that the choice of $k$ is actually arbitrary.
Lemma 2.10. Following the notation of Theorem 2.8, for \(1 \leq k \leq m\) and \(\beta \geq 1\), with probability at least \(1 - \frac{c}{m^\beta}\),

\[
\sup_{\{||a||_2 \leq 1, \ ||a||_0 \leq k\}} \left\| \sum_{i=1}^{m} a_i X_i \right\|_2 \leq c' \left( \sqrt{\beta} \max_{1 \leq i \leq m} ||X_i||_2 + (mk)^{1/4} \right);
\]

here \(c\) is an absolute constant and \(c'\) depends on \(L\) and \(q\).

Since the proof of Lemma 2.10 is almost identical to the combination of Lemma 2.3 and Theorem 2.4 from [8], we shall not reproduce it here. The combination of the two facts imply that \(Pr(A) \geq 1 - \frac{c}{m^\beta}\), and the assertion of Theorem 1.12 holds by invoking Theorem 2.8.

3 Generic Chaining and Bernoulli processes

The proofs of Theorem 1.13 and Theorem 1.14 are based on Talagrand’s generic chaining method, which was introduced as a way of controlling the suprema of random processes. We refer to Talagrand’s invaluable manuscript [14] for an extensive survey on generic chaining and its applications. Here, we present the basic notions we require, and only in the setup that is needed in what follows.

Let \(V \subset (\mathbb{R}^\ell, \ |\ |)\). A collection of subsets of \(V\), \((V_s)_{s \geq 0}\), is an admissible sequence of \(V\) if \(|V_0| = 1\) and for every \(s \geq 1\), \(|V_s| \leq 2^{2^s}\). The idea is that the sets \(V_s\) are increasingly fine approximations of \(V\), and that the \(\gamma_2\) functional, defined below, captures some “aggregate complexity” of \(V\).

To define the \(\gamma_2\) functional, let \(\pi_s v\) be the nearest point to \(v\) in \(V_s\) with respect to the norm \(|\ |\), and set \(\Delta_s v = \pi_{s+1} v - \pi_s v\).

Definition 3.1. Let

\[
\gamma_2(V, |\ |) = \inf_{v \in V} \sup_{s \geq 0} \left( \sum_{s \geq 0} 2^{s/2} ||\Delta_s v|| + ||\pi_0 v|| \right),
\]

where the infimum is taken with respect to all admissible sequences of \(V\).

The \(\gamma_2\) functional is a natural object when trying to obtain upper bounds on random processes that exhibit a subgaussian behaviour with respect to the norm \(|\ |\). For example, let \(Y\) be a symmetric random vector in \(\mathbb{R}^\ell\), and consider \(\sup_{v \in V} \langle Y, v \rangle\), the supremum of the random process \(v \rightarrow \langle Y, v \rangle\). By “subgaussian behaviour of the random process” we mean that for every \(p \geq 2\) and every \(v \in \mathbb{R}^\ell\), \(||\langle Y, v \rangle||_p \leq L \sqrt{p} ||v||\). Writing the telescopic sum

\[
\langle Y, v \rangle = \sum_{s \geq 0} \langle Y, \Delta_s v \rangle + \langle Y, \pi_0 v \rangle
\]

it follows that for “s-links” in all possible chains, it suffices to control the random variables \(\{\langle Y, \Delta_s v \rangle : v \in V\}\). The latter is a collection consisting of at most \(2^{2^s} 2^{s+1} \leq 2^{2s+2}\) random variables. The uniform control over the s-links is easily obtained because \(Pr(|\langle Y, \Delta_s v \rangle| \geq \sqrt{\beta} \max_{1 \leq i \leq m} ||X_i||_2 + (mk)^{1/4})\).
\( c \text{Lu}^{2s/2} \| \Delta_s v \| \leq 2 \exp(-u^{2s}), \) and the uniform control over all chains is obtained by taking the union bound over \( s. \)

This remarkably simple argument reveals why the \( \gamma_2 \) functional is a natural upper bound on the supremum of a subgaussian process. Talagrand discovered that, in fact, the reverse inequality was true for gaussian processes. In particular, in the context of the standard gaussian random vector in \( \mathbb{R}^\ell, \) he showed that:

**Theorem 3.2.** There exist absolute constants \( c \) and \( C \) such that, for every integer \( \ell \) and every \( V \subset \mathbb{R}^\ell, \)

\[
c\gamma_2(V, \| \|_2) \leq E \sup_{v \in V} \sum_{i=1}^{\ell} g_i v_i \leq C \gamma_2(V, \| \|_2).
\]

The upper bound is (in some equivalent formulation) due to Fernique, though this version was established by Talagrand. The lower bound is Talagrand’s celebrated *Majorizing Measures Theorem*. The proof of both parts can be found in [14].

Theorem 3.2 implies that behaviour of the supremum of a gaussian process is determined by a metric invariant relative to a single metric. Unfortunately, obtaining a lower bound on the supremum of the corresponding Bernoulli process is considerably harder. In what follows we denote by \( (\varepsilon_i)_{i=1}^\ell \) independent, symmetric \( \{-1,1\} \)-valued random variables and the Bernoulli process indexed by \( V \subset \mathbb{R}^\ell \) is

\[
v \rightarrow \sum_{i=1}^{\ell} \varepsilon_i v_i.
\]

It should be stressed that \( E \sup_{v \in V} \sum_{i=1}^{\ell} \varepsilon_i v_i \) and \( E \sup_{v \in V} \sum_{i=1}^{\ell} g_i v_i \) need not be equivalent, as the example of \( V = B_1^\ell \) shows. At the same time, it is straightforward to verify that there are absolute constant \( c \) and \( c' \) such that for any \( V \subset \mathbb{R}^\ell, \)

\[
\frac{c}{\sqrt{\log \ell}} E \sup_{v \in V} \sum_{i=1}^{\ell} g_i v_i \leq E \sup_{v \in V} \sum_{i=1}^{\ell} \varepsilon_i v_i \leq c' E \sup_{v \in V} \sum_{i=1}^{\ell} g_i v_i,
\]

which is the best that one can hope for.

Despite the logarithmic gap, we show in what follows that for the indexing sets \( V \) we are interested in, the expected supremum of the Bernoulli process is actually equivalent to the expected supremum of the gaussian process. To that end we require three features of the Bernoulli vector. Firstly, by [6], we have that for \( (a_i)_{i=1}^\ell \in \mathbb{R}^\ell \)

\[
\| \sum_{i=1}^{\ell} \varepsilon_i a_i \|_{L_p} \sim \sum_{i \leq p} a_i^* + \sqrt{p} \left( \sum_{i > p} (a_i^*)^2 \right)^{1/2},
\]

where \( (a_i^*)_{i=1}^\ell \) denotes the nonincreasing rearrangement of \( (|a_i|)_{i=1}^\ell \).

Secondly, Bernoulli processes satisfy a version of Sudakov’s minoration, a fact that was established by Talagrand (see, e.g., [7]). Here we use an equivalent formulation, noticed independently by R. Latała and the author.
Theorem 3.3. There exist absolute constants $c_1$ and $c_2$ for which the following holds. Let $U \subset \mathbb{R}^\ell$. For every $\delta \leq \sup_{v \in U} \|v\|_2$ and $2 \leq p \leq \ell$, if $|U| \geq \exp(p)$ and the set $\{\sum_{i=1}^\ell \varepsilon_i v_i : v \in U\}$ is $c_1 \delta \sqrt{p}$-separated in $L_p$, then

$$
\mathbb{E} \sup_{v \in U} \sum_{i=1}^\ell \varepsilon_i v_i \geq c_2 \delta \sqrt{p}.
$$

Finally, the supremum of a Bernoulli process exhibits a gaussian-like concentration, a fact that is also due to Talagrand (see, e.g. [7]):

Theorem 3.4. There exists an absolute constant $c$ such that, for $x > 0$,

$$
Pr\left( \left| \sup_{v \in V} \sum_{i=1}^\ell \varepsilon_i v_i - \mathbb{E} \sup_{v \in V} \sum_{i=1}^\ell \varepsilon_i v_i \right| > x \right) \leq 2 \exp\left(-c \frac{x^2}{\sup_{v \in V} \|v\|_2^2}\right).
$$

4 Proof of Theorem 1.13

The proof of Theorem 1.13 is based on a chaining argument for a Bernoulli process indexed by a random set.

Let $\Gamma = \Gamma_1^* \Gamma_2^*$ and set $U$ to be a $1/2$ net in $B_d^d$ whose cardinality is at most $\exp(10d)$. The goal is to show that with the wanted probability,

$$
\max_{v \in U} \sup_{t \in K^o} \langle \Gamma u, t \rangle \leq C(\ell(K)
$$

for a constant $C$ that depends on $\kappa_1$, $L$ and $q$. To that end, for $v \in S^{d-1}$ and $t \in K^o$ set

$$
W_{v, t} = \langle \Gamma_1^* \Gamma_2^* v, t \rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^m \langle X_i, v \rangle \langle Z_i, t \rangle.
$$

Let $(\varepsilon_i)_{i=1}^m$ be independent, symmetric $\{-1, 1\}$-valued random variables that are also independent of $(X_i, Z_i)_{i=1}^m$. Set

$$
W_{\varepsilon, t}^o = \frac{1}{\sqrt{m}} \sum_{i=1}^m \varepsilon_i \langle X_i, v \rangle \langle Z_i, t \rangle
$$

and by the symmetry of $Z$ and $X$ and the independence of $(\varepsilon_i)_{i=1}^m$, $(X_i)_{i=1}^m$ and $(Z_i)_{i=1}^m$ the process

$$
\{(v, t) \to W_{\varepsilon, t}, v \in U, t \in K^o\}
$$

has the same distribution as the process

$$
\{(v, t) \to W_{\varepsilon, t}^o, v \in U, t \in K^o\}.
$$

Let $s_0$ be the smallest integer such that $2^{s_0} \geq 20d$ and for an integer $s$, let $k_s$ be the smallest integer such that $\gamma 2^{s+3} \leq k \log(e m / k)$ for a constant $\gamma$ to be specified in what follows. Let $(T_s)_{s \geq 0}$ be an optimal admissible sequence of $K^o$ with respect to the $\ell_2$ norm. As always, denote by $\pi_s t$ the nearest point to $t$ in $T_s$, set $\Delta_s t = \pi_{s+1} t - \pi_s t$ and observe that

$$
\gamma_2(K^o, \|\|_2) = \sum_{s \geq 0} 2^{s/2} \|\Delta_s t\|_2 + \|\pi_0 t\|_2.
$$
Clearly,
\[
\sum_{i=1}^{m} \varepsilon_i \langle X_i, v \rangle \langle Z_i, t \rangle = \sum_{s \geq s_0}^{m} \sum_{i=1}^{m} \varepsilon_i \langle X_i, v \rangle \langle Z_i, \Delta_{st} \rangle + \sum_{i=1}^{m} \varepsilon_i \langle X_i, v \rangle \langle Z_i, \pi_{st} \rangle.
\]

Conditioned on \((X_i, Z_i)_{i=1}^{m}\), for any \(v \in \mathbb{R}^d, t \in \mathbb{R}^n\) and \(I \subset \{1, \ldots, m\}\),
\[
(\mathbb{E}|W_{v,t}^{\circ}|^p)^{1/p} = \|W_{v,t}^{\circ}\|_{L_p(\varepsilon)} \leq \frac{1}{\sqrt{m}} \sum_{i \in I} |\langle X_i, v \rangle \langle Z_i, t \rangle| + \sqrt{p} \left( \frac{1}{m} \sum_{i \in I} |\langle X_i, v \rangle \langle Z_i, t \rangle|^2 \right)^{1/2},
\]
for an absolute constant \(c\). In particular, let \(I = I_{v,t} \) be the union of the sets of indices corresponding to the \(k\) largest values of \(|\langle X_i, v \rangle|_{i=1}^{m}\) and the \(k\) largest values of \(|\langle Z_i, t \rangle|_{i=1}^{m}\).

Here and throughout we abuse notation and write \(\langle Z_i, t \rangle^*\) (resp. \(\langle X_i, v \rangle^*\)) for the \(i\)-th largest element in the non-increasing rearrangement of \(|\langle Z_j, t \rangle|_{j=1}^{m}\) (resp. \(|\langle X_j, v \rangle|_{j=1}^{m}\)). Therefore,
\[
\|W_{v,t}^{\circ}\|_{L_p(\varepsilon)} \leq \|(X_i, v)\|_2 \left( \frac{1}{m} \sum_{i \leq 2k} (\langle X_i, v \rangle^* \langle Z_i, t \rangle^* \rangle^2 \right)^{1/2} + \sqrt{p} \left( \frac{1}{m} \sum_{i \geq k} (\langle X_i, v \rangle^* \langle Z_i, t \rangle^* \rangle^2 \right)^{1/2}.
\]

(4.1)

The key observation is the following:

**Theorem 4.1.** There is an event \(B\) of probability at most \(2 \exp(-d)\) such that on \(A \setminus B\) the following holds:

1. For every \(v \in U\), \(\|(X_i, v)\|_2 \leq \kappa_1 \sqrt{m}\).
2. For every \(t \in K^\circ, v \in U\) and every \(s \geq s_0\),
\[
\left( \frac{1}{m} \sum_{i \geq k_s} (\langle X_i, v \rangle^* \langle Z_i, \Delta_{st} \rangle^* \rangle^2 \right)^{1/2} \leq c_1(q, L) \|\Delta_{st}\|_2.
\]
3. For every \(t \in K^\circ\) and every \(s \geq s_0\),
\[
\left( \frac{1}{m} \sum_{i \leq 2k_s} (\langle Z_i, \Delta_{st} \rangle^* \rangle^2 \right)^{1/2} \leq c_2(L) \frac{2^{s/2}}{\sqrt{m}} \|\Delta_{st}\|_2.
\]
4. For every \(t \in K^\circ\) and \(v \in U\),
\[
\left( \frac{1}{m} \sum_{i \leq 2k_{s_0}} (\langle Z_i, \pi_{st} \rangle^* \rangle^2 \right)^{1/2} \leq c_2(L) \frac{2^{s_0/2}}{\sqrt{m}} \sup_{t \in K^\circ} \|t\|_2,
\]

and
\[
\left( \frac{1}{m} \sum_{i \geq k_{s_0}} (\langle X_i, v \rangle^* \langle Z_i, \pi_{st} \rangle^* \rangle^2 \right)^{1/2} \leq c_1(q, L) \sup_{t \in K^\circ} \|t\|_2.
\]
The proof of Theorem 4.1 is based on several standard facts that are outlined in the next lemma.

**Lemma 4.2.** Let \((\xi_i)_{i=1}^m\) be independent copies of a centred random variable \(\xi\). Set \(1 \leq k \leq m\) and \(u \geq 2\).

(a) If \(\xi \in L_q\) and \(1 \leq r < q\), then with probability at least \(1 - 2u^{-qk} \exp(-c_0 k \log(em/k))\),

\[
\xi_k^* \leq eu\|\xi\|_{L_q} \left(\frac{m}{k}\right)^{1/r},
\]

where \(c_0 \sim (q/r - 1)\). In particular, if \(\|\xi\|_{L_q} \leq L\|\xi\|_{L_2}\) then on that event,

\[
\xi_k^* \leq eLu\|\xi\|_{L_2} \left(\frac{m}{k}\right)^{1/r}.
\]

(b) If \(\xi\) is \(L\)-subgaussian then with probability at least \(1 - 2 \exp(-u^2 k \log(em/k))\),

\[
\xi_k^* \leq c_1 Lu\|\xi\|_{L_2} \sqrt{k \log(em/k)}
\]

and

\[
\left(\sum_{i \leq k} (\xi_i^*)^2\right)^{1/2} \leq c_1 Lu\|\xi\|_{L_2} \sqrt{k \log(em/k)}
\]

for an absolute constant \(c_1\).

**Proof.** The proof of the first claim is an immediate outcome of a binomial estimate. Indeed,

\[
Pr \left( \exists I, |I| = k : \forall i \in I \ |\xi_i| \geq eu\|\xi\|_{L_q} \left(\frac{m}{k}\right)^{1/r} \right) \leq \binom{m}{k} Pr \left( |\xi| \geq eu\|\xi\|_{L_q} \left(\frac{m}{k}\right)^{1/r} \right)
\]

\[
\leq \left(\frac{em}{k}\right)^k \cdot u^{-qk} \left(\frac{k}{em}\right)^{kq/r} = u^{-qk} \exp \left(- \left(\frac{q}{r} - 1\right) k \log \left(\frac{em}{k}\right)\right).
\]

The first part of (b) follows from the second part because

\[
\xi_k^* \leq \left(\frac{1}{k} \sum_{i \leq k} (\xi_i^*)^2\right)^{1/2}.
\]

To prove the second part, let

\[
\|x\|_{[k]} = \max_{|I|=k} \left(\sum_{i \in I} x_i^2\right)^{1/2}
\]

which is a norm on \(\mathbb{R}^m\), and set \(B\) to be the unit ball of its dual norm. Let \(x = (\xi_i)_{i=1}^m\) and denote by \(G\) the standard gaussian random vector in \(\mathbb{R}^m\). Since \((\xi_i)_{i=1}^m\) is a \(c_0 L\)-subgaussian random vector, it follows from gaussian dominance (e.g., via the chaining mechanism, see [14]) that for every norm \(\|\|\) on \(\mathbb{R}^m\) and any \(1 \leq p < \infty\),

\[
(\mathbb{E}\|x\|^p)^{1/p} \lesssim L (\mathbb{E}\|G\|^p)^{1/p}.
\]
Hence, by the strong-weak inequality for a gaussian random vector (which follows from the gaussian concentration theorem), there is an absolute constant $c_1$ such that

$$ \left( \mathbb{E}\|x\|_k^p \right)^{1/p} \lesssim L \left( \mathbb{E}\|G\|_k^p \right)^{1/p} \leq c_1 L \left( \mathbb{E}\|G\|_k + \sqrt{p} \sup_{z \in B} \|z\|_2 \right). $$

It is straightforward to verify that $\mathbb{E}\|G\|_k \lesssim \sqrt{k \log(em/k)}$ and that $\sup_{z \in B} \|z\|_2 = 1$. The claim follows by setting $p = u^2 k \log(em/k)$ and invoking Chebyshev’s inequality. 

**Proof of Theorem 4.1.** Property (1) holds on the event $A$; therefore, it is enough to verify that Properties (2) – (4) hold with high probability.

Recall that for every $t \in \mathbb{R}^n$, $(Z, t)$ is $L$-subgaussian and for every $v \in \mathbb{R}^d$, $\|\langle X, v \rangle\|_{L_q} \leq L \|v\|_2 = L$. Observe that

$$ |\{ \Delta_s t : t \in T \}| \leq 2^s \cdot 2^{s+1} \leq 2^{2s+2}, \text{ and } |U| \leq \exp(10d). $$

Fix $s \geq s_0$ and consider the two parts of Lemma 4.2 for $k \geq k_s$: Part (a) for every $\xi = \langle X, v \rangle$ for $v \in V$ and $r = 1 + q/2$ and Part (b) for every $\xi = \langle X, \Delta_s t \rangle$, $t \in K^\circ$. The probability with which the two parts hold for a pair $(v, t)$ is $1 - 2 \exp(-c_0 k \log(em/k))$

Hence, the two parts hold uniformly for every $v \in U$ and $\Delta_s t$ with probability at least

$$ 1 - 2^{2s+2} \exp(10d) \cdot 2 \exp(-c_0 k \log(em/k)). $$

Taking the union bound for $k \geq k_s$, both parts hold uniformly with probability at least

$$ 1 - 2^{2s+2} \exp(10d) \exp(-c_1 k_s \log(em/k_s)) \geq 1 - 2^{2s+2} \exp(10d) \exp(-c_1 \cdot (\gamma/2)2^{s+3}) \geq 1 - \exp(2^{s+2}), $$

for a suitable choice of the absolute constant $\gamma$.

On that event, and by the isotropicity of $Z$ and $X$, there is an absolute constant $c$ such that

$$ \left( \frac{1}{m} \sum_{i \geq k_s} \langle X_i, v \rangle^* \langle Z_i, \Delta_s t \rangle^* \right)^{1/2} \leq c u L \|v\|_2 \|\Delta_s t\|_2 \cdot \left( \frac{1}{m} \sum_{j \geq k_s} \left( \frac{m}{j} \right)^{2/r} \log(em/j) \right)^{1/2} \leq c_2(q) L \|\Delta_s t\|_2, $$

as required. Property (2) now follows from the union bound on $s \geq s_0$ and recalling that $2^{s_0} \geq 20d$.

Property (3) follows in a similar fashion, thanks to the second component in Part (b) of Lemma 4.2 and the union bound for $s \geq s_0$.

The proof of Property (4) is identical to the proofs of Properties (2) and (3), by considering the set of the $L$-subgaussian random variables $\{\langle Z, \pi_{s_0} t \rangle : t \in T \}$ whose cardinality is at most $2^{2s_0}$. We omit the standard details. 

**Proof of Theorem 4.13.** Let $B$ be the event from Theorem 4.1. Conditioned on $A \setminus B$ and invoking (4.1), there is an absolute constant $c_0$ such that the following holds. Let $s \geq s_0,$
\( w \geq 2 \), and set \( p = w^{2s+3} \). Hence, for every \( v \in U \) and \( t \in K^o \),
\[
\|W_{v, \Delta t}^o\|_{L_p(\varepsilon)}
\leq \|(X_i, v)\|_2 \left( \frac{1}{m} \sum_{i \leq 2k_s} ((Z_i, \Delta t)^*)^2 \right)^{1/2} + c_0 w^{2(s+3)/2} \left( \frac{1}{m} \sum_{i \geq k_s} ((X_i, v)^* (Z_i, \Delta t)^*)^2 \right)^{1/2}
\leq c_1(q, L, \kappa_1) \left( \sqrt{\frac{2s^2}{m}} \|\Delta s t\|_2 + w^{2s/2} \|\Delta s t\|_2 \right) \leq 2c_1 w^{2s/2} \|\Delta s t\|_2,
\]
By Chebyshev’s inequality,
\[
Pr_\varepsilon \left( \|W_{v, \Delta t}^o\| \geq 2c_1 w^{2s/2} \|\Delta s t\|_2 \right) \leq \exp(-w^{2s+3}). \tag{4.2}
\]
Recall that \( \{\{\Delta s t : t \in T\} : |U| \leq 2^{2s+2} \), that \( |U| \leq \exp(10d) \), and the choice of \( w \). By the union bound, (4.2) holds uniformly for every \( t \in K^o \) and \( v \in U \) as long as \( s \geq s_0 \). And, by the union bound for \( s \geq s_0 \), there is an absolute constant \( c_2 \) such that with probability at least \( 1 - 2 \exp(-c_2 w^2d) \) with respect to \( (\varepsilon_i)_{i=1}^m \),
\[
\left| \sum_{s \geq s_0} W_{v, \Delta t}^o \right| \leq c_3(q, L, \kappa_1) w \sum_{s \geq s_0} 2^{s/2} \|\Delta s t\|_2 \leq c_3w \ell(K),
\]
where we have used the fact that
\[
\sum_{s \geq s_0} 2^{s/2} \|\Delta s t\|_2 \leq \gamma_2(K^o, \| \|_2) \lesssim \ell(K).
\]
Also, setting \( p = w^{2s_0} \), it follows that for every \( t \in K^o \) and every \( v \in U \),
\[
\|W_{v, \pi_{s_0} t}^o\|_{L_p(\varepsilon)}
\leq \|(X_i, v)\|_2 \left( \frac{1}{m} \sum_{i \leq 2k_{s_0}} ((Z_i, t)^*)^2 \right)^{1/2} + c_0 w^{2s_0/2} \left( \frac{1}{m} \sum_{i \geq k_{s_0}} ((X_i, v)^* (Z_i, t)^* )^2 \right)^{1/2}
\leq c_4(q, L, \kappa_1) w^{2s_0/2} \|\pi_{s_0} t\|_2 \leq c_4 w \sqrt{d} \sup_{t \in K^o} \|t\|_2 \leq c_4 w \ell(K),
\]
because \( d \leq d^*(K) \). Hence, by the union bound, with probability at least \( 1 - 2 \exp(-c_5 w^2d) \) with respect to \( (\varepsilon_i)_{i=1}^m \), for every \( t \in K^o \) and \( v \in U \),
\[
\|W_{v, \pi_{s_0} t}^o\| \leq c_6 w \ell(K).
\]
To summarize, conditioned on \( A \setminus B \), with probability at least \( 1 - 2 \exp(-c w^2d) \) with respect to \( (\varepsilon_i)_{i=1}^m \), for every \( v \in U \) and \( t \in K^o \),
\[
\|W_{v, t}\| \leq \|W_{v, \pi_{s_0} t}^o\| + \left| \sum_{s \geq s_0} W_{v, \Delta t}^o \right| \leq c'(\kappa_1, q, L) w \ell(K).
\]
Now the claim follows using a suitable choice of \( w = c''(\kappa_1, q, L) \) and a Fubini argument.
5 Proof of Theorem 1.14

Recall that $V$ is a $\rho$ net in $S^{d-1}$ for some fixed $0 < \rho \leq 1/4$. The idea behind the lower bound is that, mainly thanks to the small-ball property, the set $\Gamma_1 K^o$ consists of vectors in a "good position". As a result, conditioned on a large subset of $\mathcal{A}$ and for every $v \in V$, the supremum of the Bernoulli process $W_{v,t}$ dominates $\sim \ell(K)$ with sufficiently high probability. That high probability leads to a uniform estimate in $V$.

5.1 Preliminary estimates

The starting point is the following lemma, which was first noticed by A. Pajor.

**Lemma 5.1.** There exists an absolute constant $\kappa_2$ for which the following holds. For any $T \subset \mathbb{R}^n$ there is a subset $T' \subset T$ of cardinality at most $\exp(\kappa_2 n)$ such that

$$\mathbb{E} \sup_{t \in T'} \sum_{i=1}^{n} g_i t_i \geq \frac{1}{2} \mathbb{E} \sup_{t \in T} \sum_{i=1}^{n} g_i t_i.$$ 

**Proof.** Let $T'$ be a maximal $\varepsilon$-separated subset of $T$ whose cardinality is $\exp(k)$. By Sudakov’s inequality (see, e.g., [7, 13]), there is an absolute constant $c$ such that

$$\varepsilon \leq \frac{c}{\sqrt{k}} \mathbb{E} \sup_{t \in T} \sum_{i=1}^{n} g_i t_i.$$ 

Since $T \subset T' + \varepsilon B_2^n$ it follows that

$$\mathbb{E} \sup_{t \in T} \sum_{i=1}^{n} g_i t_i \leq \mathbb{E} \sup_{t \in T'} \sum_{i=1}^{n} g_i t_i + \mathbb{E} \sup_{v \in B_2^n} \sum_{i=1}^{n} g_i v_i$$

$$\leq \mathbb{E} \sup_{t \in T'} \sum_{i=1}^{n} g_i t_i + \varepsilon \sqrt{n} \leq \mathbb{E} \sup_{t \in T'} \sum_{i=1}^{n} g_i t_i + \sqrt{n} \mathbb{E} \sup_{t \in T} \sum_{i=1}^{n} g_i t_i$$

$$\leq \mathbb{E} \sup_{t \in T'} \sum_{i=1}^{n} g_i t_i + \frac{1}{2} \mathbb{E} \sup_{t \in T} \sum_{i=1}^{n} g_i t_i$$

provided that $k \geq 4c^2 n$. 

With Lemma 5.1 in mind, let $T' \subset K^o$ satisfy that

$$|T'| \leq \exp(\kappa_2 n) \quad \text{and} \quad \mathbb{E} \sup_{t \in T'} \sum_{i=1}^{n} g_i t_i \geq \frac{1}{2} \ell(K) \quad (5.1)$$

for a suitable absolute constant $\kappa_2$.

Next, let us describe the properties of $\Gamma_1^o V$ and of $\Gamma_1 T'$ that are required in the proof. Let $s_0$ be the smallest integer that satisfies $2^s \geq 20d$ and set $k_8$ to be the smallest integer that satisfies $2^{s+3} \leq k \log(em/k)$.

**Definition 5.2.** For $u \geq 2$ let $\Omega_u$ be the event on which:
For every $t \in \mathbb{R}^m$, 
\[
\frac{1}{m} \sum_{i=1}^{m} \langle Z_i, t \rangle^2 \leq 2 \|t\|_2^2.
\]

(2) Let $(T_s)_{s \geq 0}$ be an optimal admissible sequence of $T'$. Then for every $s \geq s_0$ and every $t \in T'$,
\[
\left( \sum_{i \leq 2k_s} ((Z_i, \Delta_s t)^*)^2 \right)^{1/2} \leq Lu 2^{s/2} \|\Delta_s t\|_2;
\]
\[
\langle Z_j, \Delta_s t \rangle^* \leq Lu \|\Delta_s t\|_2 \sqrt{\log(em/j)} \text{ for every } j \geq k_s;
\]
and
\[
\left( \sum_{i \leq 2k_{s_0}} ((Z_i, \pi_{s_0} t)^*)^2 \right)^{1/2} \leq Lu 2^{s_{0}/2} \|\pi_{s_0} t\|_2;
\]
\[
\langle Z_j, \pi_{s_0} t \rangle^* \leq Lu \|\pi_{s_0} t\|_2 \sqrt{\log(em/j)} \text{ for every } j \geq k_{s_0}.
\]

(3) \[ \sup_{\|a\|_2 \leq 1, \|a\|_0 \leq \theta m} \|\sum_{i=1}^{m} a_i X_i\|_2 \leq \delta \sqrt{m}. \]

(4) For every $v \in V$ and every $j \geq d$, \[ (\langle X_i, v \rangle)^*_j \leq Lu \left( \frac{em}{J} \right)^{q/q+2}. \]

(5) For every $t_1, t_2 \in T'$ and $v \in V$,
\[
|\{i : |\langle X_i, v \rangle| \geq \kappa_0 \text{ and } |\langle Z_i, t_1 - t_2 \rangle| \geq \kappa_0 \|t_1 - t_2\|_2\}| \geq 0.98m. \tag{5.2}
\]

We begin by showing that $\Omega_u$ is a nontrivial event.

**Theorem 5.3.** There exist constants $c_0$ and $c_1$ that depend on $\kappa_0, \kappa_1, q$ and $L$ such that the following holds. Let
\[
m \geq c_0 \max \left\{ \frac{d^*(K)}{\rho}, n \right\}, \tag{5.3}
\]
and set $u \geq 4 \sqrt{\log(5/\rho)}$. There exists an event $B_u$, such that $\Omega_u \supset A \setminus B_u$ and
\[
Pr(B_u) \leq 2 \exp(-c_1 d \log u).
\]

The proof of Theorem 5.3 is similar to the proof of Theorem 4.1. We only outline the standard argument.

**Proof.** Part (1) holds with probability at least $1 - 2 \exp(-c(L)m)$. Indeed, it suffices to consider a net in $\mathbb{S}^{n-1}$ whose cardinality is at most $\exp(10n)$. By the $\psi_1$ version of Bernstein’s inequality (see, e.g. [4]) we have that
\[
Pr \left( \left| \frac{1}{m} \sum_{i=1}^{m} \langle Z_i, t \rangle^2 - \|t\|_2^2 \right| \geq \frac{\|t\|_2^2}{4} \right) \leq \exp(-c_0(L)m),
\]
and the wanted estimate follows from the union bound and the choice of $m$. 


The proof of Part (3) is identical to the proof of Theorem 5.1. Following the latter, it is evident that the claim holds with probability at least $1 - 2\exp(-c_2 u^2 d)$ provided that $u \geq c_3 \sqrt{\log(5/\rho)}$ for constants $c_2$ and $c_3$ that depend on $L$ and $q$.

Part (3) holds on the event $A$.

To prove Part (4), let $r = 1 + q/2$ and recall that $\|\langle X, v \rangle\|_{L_q} \leq L\|v\|_2$. By Lemma 4.2, with probability at least $1 - u^{-q_j} \exp(-c_3(q_j \log(em/j))$

$$\langle (X_j, v) \rangle^* \leq Lu \left( \frac{em}{j} \right)^{1/r}.$$

(5.4)

Since $|V| \leq \exp(d \log(5/\rho))$ and $m \geq c_4d/\rho$, it follows from the union bound that with probability at least $1 - 2\exp(-c_5(q)d(\log u + \log(5/\rho)))$, (5.4) holds for every $v \in V$ and every $j \geq 1$. Indeed, note that

$$2^{s_0} \log(em/2^{s_0}) \geq 20d \log(em/10d) \geq 20d \log(c_4/10\rho) \geq 20d \log(5/\rho)$$

provided that $c_4$ is a sufficiently large absolute constant.

Finally, Part (5) is evident by the small-ball assumption on each $\langle Z, t \rangle$ and $\langle X, v \rangle$. Indeed, for every $t \in \mathbb{R}^n$, $Pr(|\langle Z, t \rangle| \leq \kappa_0 \|t\|_2) \leq 1/1000$. By a binomial estimate there is an absolute constant $c_5$ such that with probability at least $1 - 2\exp(-c_5m)$, $|\{i : |\langle Z_i, t \rangle| \geq \kappa_0\}| \geq 0.99m$. A similar argument used for $|\langle X, v \rangle|$ and the fact that $|T'| \leq \exp(c_2m)$ and $|V| \leq \exp(d \log(5/\rho))$ completes the proof—by invoking the union bound and recalling the choice of $m$.

\[ \square \]

5.2 A lower bound on Bernoulli processes

A crucial component in the proof of the lower bound is an equivalence result between the expected supremum of a Bernoulli process and of a gaussian one indexed by the same set—under certain structural assumptions on the indexing set.

Theorem 5.4. For every $0 < \lambda < 1$ there exists a constant $c_0 = c_0(\lambda)$ for which the following holds. Let $U \subset \mathbb{R}^m$ and assume that there is $\eta > 0$ such that for every $v, w \in U$ and every $1 \leq p \leq \lambda m$,

$$\| \sum_{i=1}^{m} \varepsilon_i (v - w)_i \|_{L_p} \geq \eta \| \sum_{i=1}^{m} g_i (v - w)_i \|_{L_p}.$$ 

Then

$$E \sup_{v \in U} \sum_{i=1}^{m} \varepsilon_i v_i \geq c_0 \eta E \sup_{v \in U} \sum_{i=1}^{m} g_i v_i.$$ 

The proof of Theorem 5.4 is almost identical to the proof of the majorizing measures theorem. It is based on Talagrand’s construction of an admissible sequence, by showing that the functional $\phi(U) = E \sup_{v \in U} \sum_{i=1}^{m} \varepsilon_i v_i$ satisfies the following growth condition:

Definition 5.5. Let $(U,d)$ be a metric space. The functional $\phi$ satisfies the growth condition if there are $r \geq 4$ and $c_0$ such that for every integer $s$, every $k = 2^s$ and every $a > 0$ the following holds. If $v \in U$, $v_1, \ldots, v_k \in B(v, ra)$ are a-separated and $H_i \subset U \cap B(v_i, a/r)$ then

$$\phi\left( \bigcup_{\ell \leq m} H_\ell \right) \geq c_0 a^{2^{s/2}} + \min_{\ell \leq m} \phi(H_\ell).$$ 

23
The growth condition was used by Talagrand to construct an admissible sequence for the metric space \((U,d)\).

**Theorem 5.6.** [14] If \((U,d)\) satisfies the growth condition in Definition 5.5 with parameters \(r\) and \(c_0\) then

\[
\gamma_2(U,d) \leq c_1(\phi(U) + \text{diam}(U,d)),
\]

where \(c_1 = c_1(r,c_0)\).

**Proof of Theorem 5.4.** By the majorizing measures theorem, \(\gamma_2(U,\|\cdot\|_2) \sim E\sup_{v \in U} \sum_{i=1}^{m} g_i v_i\), and therefore, by Lemma 5.1 it suffices to consider \(U' \subset U\) of cardinality at most \(\exp(\kappa_2 m)\).

We will use Theorem 5.6 for \(U' \subset \mathbb{R}^m\) and upper bound \(\gamma_2(U',\|\cdot\|_2)\).

Let us show that \(U'\) satisfies the growth condition for our choice of functional \(\phi(H) = E\sup_{v \in H} \sum_{i=1}^{m} \varepsilon_i v_i\).

Fix \(r > 4\) to be named later. Let \(k \leq \exp(\kappa_2 m)\), assume that \(\{v_{\ell}, \ 1 \leq \ell \leq k\}\) is \(a\)-separated in \(E\), set \(H_\ell \subset B(v_{\ell}, \sigma)\) for \(\sigma = a/r\) and put \(H = \bigcup_{\ell \leq k} H_\ell\).

For every \(v \in U'\), let \(X_v = \sum_{i=1}^{m} \varepsilon_i v_i\), \(Y_v = (\sup_{v \in H} X_v) - X_v\) and \(W = \max_{1 \leq k} |Y_v - EY_v|\).

Applying the concentration inequality for Bernoulli processes, it follows that for every \(u > 0\),

\[
Pr(|Y_v - EY_v| \geq u) \leq 2 \exp(-c_1 u^2/\sigma^2)
\]

for a suitable absolute constant \(c_1\). Therefore, \(Pr(W \geq u) \leq 2k \exp(-c_1 u^2/\sigma^2)\), and thus \(EW \leq c_2 \sigma \sqrt{\log k}\).

Note that

\[
\sup_{v \in H} X_v = \max_{\ell \leq k} X_{v_{\ell}} + \min_{\ell \leq k} EY_{\ell} - W
\]

and that

\[
E\sup_{v \in H} X_v = \phi\left(\bigcup_{\ell \leq k} H_\ell\right) \quad \text{and} \quad EY_{\ell} = E\sup_{v \in H_\ell} X_v = \phi(H_\ell).
\]

Therefore, integrating (5.5), there is an absolute constant \(c_3\) such that

\[
\phi\left(\bigcup_{\ell \leq k} H_\ell\right) \geq E\max_{\ell \leq k} X_{v_{\ell}} + \min_{\ell \leq k} \phi(H_\ell) - c_3 \sigma \sqrt{\log k},
\]

and the growth condition follows once we show that

\[
E\max_{1 \leq k} X_{v_{\ell}} \geq 2c_3 \cdot (a/r) \sqrt{\log k}
\]

for an appropriate choice of \(r\).

For \(1 \leq \ell \leq k\), let \(u_{\ell} = v_{\ell} - v_1\), and clearly

\[
E\sup_{1 \leq k} X_{v_{\ell}} = E\sup_{1 \leq k} X_{u_{\ell}}
\]

because \(E X_{v_1} = 0\). Also, \(u_i - u_j = v_i - v_j\), and thus \(\{u_{\ell} : 1 < \ell \leq k\}\) is \(a\)-separated in \(E\).

Let \(G\) be the standard gaussian vector in \(\mathbb{R}^m\). By the assumed equivalence between \(\|X_{s-t}\|_{L_p}\) and \(\|G, s-t\|_{L_p}\) it is evident that for every \(i \neq j\), \(i, j > 0\) and every \(2 \leq p \leq \lambda m\),

\[
\|X_{u_i - u_j}\|_{L_p} \geq \eta \|G, u_i - u_j\|_{L_p} \geq c_4 \eta \sqrt{p} u_i - u_j \|_2 \geq c_4 \eta \sqrt{pa}.
\] (5.6)
Applying Sudakov’s minoration for Bernoulli processes (Theorem 3.3) for \( p \sim \log k \), we have that
\[
\mathbb{E} \sup_{\ell \leq k} X_{u_\ell} \geq c_5 \eta a \sqrt{\log k}.
\]
This concludes the proof of the growth condition for \( k \leq \exp(\lambda m) \) by making a suitable choice of \( r \) that is large enough.

Finally, since \( k \leq \exp(\kappa_2 m) \), it remains to consider \( \lambda m \leq \log k \leq \kappa_2 m \), and show that a similar norm equivalence to (5.6) is true in this range. Indeed, observe that for \( \lambda m \leq p \leq m \),
\[
\|X_{u_i - u_j}\|_{L_p} \geq \|X_{u_i - u_j}\|_{L_{\lambda m}} \geq \eta \|\langle G, u_i - u_j \rangle\|_{L_{\lambda m}} \geq c_6 \eta \sqrt{\lambda m} \|\langle G, u_i - u_j \rangle\|_{L_p}.
\]
Hence, the same estimate as in (5.6) holds with a constant that depends only on \( \lambda \), concluding the proof of the growth condition.

5.3 Returning to the lower bound
With all the ingredients set in place, let \( T' \) be as in (5.1), fix \( v \in V \) and let us obtain a high probability lower bound on
\[
\sup_{t \in T'} \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \varepsilon_i \langle X_i, v \rangle \langle Z_i, t \rangle,
\]
conditioned on the event \( \Omega_u \). Note that to be of any use, the lower bound should hold uniformly for every \( v \in V \), implying that the required individual probability estimate has to “defeat” \( |V| \leq \exp(d \log(5/\rho)) \).

For \( v \in V \), let \( J_v \) be the set of indices of the \( \theta m \) largest coordinates of \( (|\langle X_i, v \rangle|)_{i=1}^{m} \). Consider the random variables
\[
U_{v,t}^o = \frac{1}{\sqrt{m}} \sum_{i \in J_v^c} \varepsilon_i \langle X_i, v \rangle \langle Z_i, t \rangle,
\]
and the first order of business is to obtain a high probability lower bound on \( \sup_{t \in T'} U_{v,t}^o \), conditioned on \( \Omega_u \).

Fix \( v \in V \), set
\[
A_v = \left\{ \left( \frac{1}{\sqrt{m}} \langle X_i, v \rangle \langle Z_i, t \rangle \mathbf{1}_{J_v^c}(i) \right)_{i=1}^{m} : t \in T' \right\} \subset \mathbb{R}^m
\]
and observe that on \( \Omega_u \),
\[
\sup_{a \in A_v} \|a\|_2 \leq \max_{t \in T'} \left( \frac{1}{m} \sum_{i \in J_v^c} \langle X_i, v \rangle^2 \langle Z_i, t \rangle^2 \right)^{1/2} \leq (\langle X_{\theta m}, v \rangle)^* \max_{t \in T'} \left( \frac{1}{m} \sum_{i=1}^{m} \langle Z_i, t \rangle^2 \right)^{1/2} \leq 2(\langle X_{\theta m}, v \rangle)^* \sup_{t \in K^o} \|t\|_2 \leq 2Lu^2 \left( \frac{e}{\theta} \right)^{1/r} \sup_{t \in K^o} \|t\|_2.
\]
By the concentration theorem for Bernoulli processes, we have that

\[
Pr_{\varepsilon} \left( \left| \sup_{a \in A_v} \sum_{i=1}^{m} \varepsilon_i a_i - \mathbb{E}_\varepsilon \sup_{a \in A_v} \sum_{i=1}^{m} \varepsilon_i a_i \right| \geq x \right) \leq 2 \exp \left( -c_0 \frac{x^2}{\sup_{a \in A_v} \|a\|_2^2} \right)
\]

\[
\leq 2 \exp \left( -c_1(L) \left( \frac{\theta^{2/r}}{u^4} \right) \frac{x^2}{\sup_{t \in K^c} \|t\|_2^2} \right).
\]

and setting \( x = \frac{1}{2} \mathbb{E}_\varepsilon \sup_{a \in A_v} \sum_{i=1}^{m} \varepsilon_i a_i \), it is evident that

\[
\sup_{a \in A_v} \sum_{i=1}^{m} \varepsilon_i a_i \geq \frac{1}{2} \mathbb{E}_\varepsilon \sup_{a \in A_v} \sum_{i=1}^{m} \varepsilon_i a_i
\]

with probability at least

\[
1 - 2 \exp \left( -c_3(L) \left( \frac{\theta^{2/r}}{u^4} \right) \left( \frac{\mathbb{E}_\varepsilon \sup_{a \in A_v} \sum_{i=1}^{m} \varepsilon_i a_i}{\sup_{t \in K^c} \|t\|_2^2} \right)^2 \right).
\]

Therefore, if one can show that for a suitable constant \( c_4 \),

\[
\mathbb{E}_\varepsilon \sup_{a \in A_v} \sum_{i=1}^{m} \varepsilon_i a_i \geq c_4 \ell(K)
\]  

(5.7)

and if

\[
c_3 c_4^2 \frac{\theta^{2/r}}{u^4} \mathbb{E} \sup_{a \in A_v} \sum_{i=1}^{m} \varepsilon_i a_i \geq 4d \log(5/\rho),
\]

then with probability at least

\[
1 - 2 \exp \left( -\frac{c_3}{2} \frac{c_4^2 \theta^{2/r}}{u^4} \mathbb{E} \sup_{a \in A_v} \sum_{i=1}^{m} \varepsilon_i a_i \right)
\]  

(5.9)

with respect to \( (\varepsilon_i)_{i=1}^{m} \), for every \( v \in V \)

\[
\sup_{t \in T'} \frac{1}{\sqrt{m}} \sum_{i \in J_v} \varepsilon_i \langle X_i, v \rangle \langle Z_i, t \rangle \geq \frac{c_4}{2} \ell(K).
\]  

(5.10)

It should be noted that, following the restriction on \( u \) from Theorem 5.3, a suitable choice of \( u \) will be proportional to \( \log(5/\rho) \).

For the proof of (5.7) we show that for every \( v \in V \), the Bernoulli process restricted to the coordinates \( J_v^c \) is likely to be large. That is based on Property (5) satisfied by \( \Omega_u \) — most of the coordinates of \( \langle X_i, v \rangle \) and \( \langle Z_i, t \rangle \) are nontrivial.

**Theorem 5.7.** There is a constant \( \kappa_3 \) that depends only on \( \kappa_0 \) such that, conditioned on the event \( \Omega_u \), for every \( v \in V \),

\[
\mathbb{E}_\varepsilon \sup_{a \in A_v} \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \varepsilon_i a_i = \mathbb{E}_\varepsilon \sup_{t \in T'} \frac{1}{\sqrt{m}} \sum_{i \in J_v^c} \varepsilon_i \langle X_i, v \rangle \langle Z_i, t \rangle \geq \kappa_3 \ell(K).
\]
Proof. On \( \Omega_u \) we have that for every \( v \in V \) and \( t_1, t_2 \in T' \), the cardinality of 
\[
I_{v,t_1,t_2} = \{ i : |\langle X_i, v \rangle| \geq \kappa_0 \text{ and } |\langle Z_i, t_1 - t_2 \rangle| \geq \kappa_0 \|t_1 - t_2\|_2 \}
\]
is at least \( 0.98m \). Recall that \( \theta < 1/4 \), implying that \( |J_v| = \theta m \leq m/4 \) and that
\[
|I_{v,t_1,t_2} \cap J_v| \geq \frac{m}{2}.
\]
Set
\[
I_v = \{ i : |\langle X_i, v \rangle| \geq \kappa_0 \} \supset I_{v,t_1,t_2}.
\]
By the contraction inequality for Bernoulli processes (see, e.g. [7]) used twice,
\[
\mathbb{E}_\varepsilon \sup_{t \in T'} \left| \frac{1}{\sqrt{m}} \sum_{i \in J_v} \varepsilon_i \langle X_i, v \rangle \langle Z_i, t \rangle \right| \geq \mathbb{E}_\varepsilon \sup_{t \in T'} \left| \frac{1}{\sqrt{m}} \sum_{i \in J_v \cap I_v} \varepsilon_i \langle X_i, v \rangle \langle Z_i, t \rangle \right| \geq \kappa_0 \mathbb{E}_\varepsilon \sup_{t \in T'} \left| \frac{1}{\sqrt{m}} \sum_{i \in J_v \cap I_v} \varepsilon_i \langle Z_i, t \rangle \right|.
\]
Denote by \( (a^*_i) \) the nonincreasing rearrangement of \( \{a_i\}_{i \in J_v \cap I_v} \) and fix \( p \leq m/4 \). Invoking the characterization of the \( L_p \) norm of a linear form of the Bernoulli vector, we have that, for any \( t_1, t_2 \in T' \),
\[
\left\| \frac{1}{\sqrt{m}} \sum_{i \in J_v \cap I_v} \varepsilon_i \langle Z_i, t_1 - t_2 \rangle \right\|_{L_p(\varepsilon)} \geq \sum_{\ell \leq p} \langle Z_i, t_1 - t_2 \rangle^* + \sqrt{p} \left( \sum_{i > p} \langle Z_i, t_1 - t_2 \rangle^* \right)^{1/2} \geq \sqrt{p} \left( \sum_{i > p} \langle Z_i, t_1 - t_2 \rangle^* \right)^{1/2} \geq \sqrt{p} \kappa_0 \|t_1 - t_2\|_2 \geq c_1 \kappa_0 \|G, t_1 - t_2\|_{L_p}.
\]
Hence, by Theorem 5.4 applied to the set
\[
F_v = \left\{ m^{-1/2} \langle Z_i, t \rangle, \ i \in J_v \cap I_v, \ t \in T' \right\}
\]
there is an absolute constant \( c_2 \) such that
\[
\gamma_2(F_v, \| \|_2) \leq c_2 \kappa_0^{-1} \mathbb{E}_\varepsilon \sup_{t \in T'} \frac{1}{\sqrt{m}} \sum_{i \in J_v \cap I_v} \varepsilon_i \langle Z_i, t \rangle.
\]
At the same time, and considering only coordinates in \( J_v^c \cap I_v \), it is evident that for every
\( x = m^{-1/2} \langle Z_i, t_1 \rangle \) and \( y = m^{-1/2} \langle Z_i, t_2 \rangle \), \( \|x - y\|_2 \geq \kappa_0 \|t_1 - t_2\|_2 \). Indeed, \( |J_v^c \cap I_{v,t_1,t_2}| \geq m/2 \) and on \( \Omega_u \), for every \( i \in J_v^c \cap I_{v,t_1,t_2}, \ |\langle Z_i, t_1 - t_2 \rangle| \geq \kappa_0 \|t_1 - t_2\|_2 \). Thus,
\[
\gamma_2(F_v, \| \|_2) \geq c_2 \kappa_0 \gamma_2(T', \| \|_2),
\]
and by the majorizing measures theorem, Lemma 5.1 and the definition of \( T' \)
\[
\mathbb{E}_\varepsilon \sup_{t \in T'} \frac{1}{\sqrt{m}} \sum_{i \in J_v} \varepsilon_i \langle Z_i, v \rangle \geq c_1(\kappa_0) \mathbb{E}_\varepsilon \sup_{t \in T'} \sum_{i=1}^m g_i t_i \geq c_1 \frac{1}{2} \ell(K),
\]
as required.
Finally, to complete the proof of a high probability lower bound on \( \sup_{t \in T} W_v^m \), we show that the contribution of each set of indices \( J_v \) to the supremum of the Bernoulli process is not that big.

**Lemma 5.8.** Let \( u = 4 \sqrt{\log(5/p)} \) and set \((X_i, Z_i)_{i=1}^m \in \Omega_u \). Then with probability at least \( 1 - \exp(-u^2d/2) \) with respect to \((\varepsilon_i)_{i=1}^m\), for every \( v \in V \)

\[
\sup_{t \in K^o} \left| \frac{1}{\sqrt{m}} \sum_{i \in J_v} \varepsilon_i \langle X_i, v \rangle \langle Z_i, t \rangle \right| \leq c(L)u \left( \delta + u \theta^{(q-2)/2(q+2)} \sqrt{\log(e/\theta)} \right) \ell(K).
\]

**Proof.** Recall that \( s_0 \) is the smallest integer such that \( 2^s \geq 20d \) and \( k_s \) is the smallest integer such that \( 2^{s+3} \leq k \log(em/k) \). Let \((T_s)_{s \geq 0} \) be an optimal admissible sequence of \( K^o \) and fix \( v \in V \). Denote by \( I \) the union of the set of indices of the \( k_s \) largest coordinates of \((\langle X_i, v \rangle)_{i \in J_v} \) and of \((\langle Z_i, t \rangle)_{i \in J_v} \). By the definition of \( \Omega_u \),

\[
\exp\left(\sum_{i \in I} \sum_{k_s \leq 2^m} \left( \langle X_i, v \rangle, \langle Z_i, t \rangle \right) \right) \leq 1 + \frac{1}{\sqrt{m}} \sum_{i \in J_v} \left| \langle X_i, v \rangle \langle Z_i, t \rangle \right| + \sqrt{p} \left( \frac{1}{m} \sum_{i \in I} \langle X_i, v \rangle \langle Z_i, t \rangle \right)^{1/2}
\]

\[
\leq \left( \sum_{i \in J_v} \langle X_i, v \rangle \right)^{1/2} \left( \frac{1}{m} \sum_{i \leq 2^m} \left( \langle X_i, \Delta_s t \rangle \right)^2 \right) \leq \sqrt{p} \left( \frac{1}{m} \sum_{i \leq 2^m} \left( \langle X_i, v \rangle \langle Z_i, \Delta_s t \rangle \right)^2 \right)^{1/2}
\]

\[
\leq c(L)u \|\Delta_s t\|_2 \left( \delta \sqrt{m} \cdot \frac{1}{m} \sqrt{k_s \log(em/k_s)} + \sqrt{p} \left( \frac{1}{m} \sum_{i=k_s+1}^{\theta_m} \left( \frac{m}{i} \right)^r \log(em/i) \right) \right)^{1/2} = (*),
\]

where \( r = 1 + q/2 \). Setting \( p = u^2 2^{s+3} \), we have that

\[
(*) \leq c_1(L)u \left( \delta + u \theta^{1/2-1/r} \sqrt{\log(e/\theta)} \right) 2^{s/2} \|\Delta_s t\|_2,
\]

(5.11)

and by Chebyshev’s inequality, followed by the union bound over all \( \{\Delta_s t : t \in K^o\} \), it is evident that with probability at least \( 1 - 2^{2s+3} \exp(-u^2 2^{s+3}) \geq 1 - \exp(-u^2 2^{s+3}) \) with respect to \((\varepsilon_i)_{i=1}^m\), \ref{5.11} holds for every \( t \in K^o \). Next, by the union bound for \( s \geq s_0 \), \ref{5.11} holds with probability at least \( 1 - 2^s \exp(-u^2 2^{s_0+1}) \) for every \( t \in K^o \) and every \( s \geq s_0 \). On that event,

\[
\left| \frac{1}{\sqrt{m}} \sum_{i \in J_v} \varepsilon_i \langle X_i, v \rangle \langle Z_i, t - \pi_{s_0} t \rangle \right| \leq c_1(L)u \left( \delta + u \theta^{1/2-1/r} \sqrt{\log(e/\theta)} \right) \ell(K).
\]

Finally, by an identical argument, only this time for the random variables

\[
\frac{1}{\sqrt{m}} \sum_{i \in J_v} \varepsilon_i \langle X_i, v \rangle \langle Z_i, \pi_{s_0} t \rangle,
\]
for \( p = u^2 2^{s_0 + 3} \),

\[
\left\| \frac{1}{\sqrt{m}} \sum_{i \in J_v} \varepsilon_i \langle X_i, v \rangle \langle Z_i, \pi_{s_0} t \rangle \right\|_{L_p(\varepsilon)} \leq \left( \sum_{i \in J_v} \langle X_i, v \rangle^2 \right)^{1/2} \left( \frac{1}{m} \sum_{i \leq 2k_{s_0}} (\langle Z_i, \pi_{s_0} t \rangle^* \rangle)^2 \right)^{1/2} + \sqrt{p} \left( \frac{1}{m} \sum_{k_{s_0} \leq i \leq \theta m} (\langle X_i, v \rangle^* \langle Z_i, \pi_{s_0} t \rangle^* \rangle)^2 \right)^{1/2} \leq c_2(L) u \left( \delta + u \theta^{1/2-1/r} \log^{1/2} (e/\theta) \right) 2^{s_0/2} \| \pi_{s_0} t \|_2,
\]

and again, \( 2^{s_0/2} \| \pi_{s_0} t \|_2 \leq \sqrt{d} \sup_{t \in K^c} \| t \|_2 \leq \ell(K) \).

Chebyshev’s inequality, the union bound over \( \pi_{s_0} t \in T_{s_0} \), followed by the union bound over \( v \in V \) completes the proof, recalling that \( 2^{s_0} \geq 20d \) and that \( u = 4 \sqrt{\log(5/\rho)} \).

The proof of Theorem 1.14 is now clear: one combines (5.10) (a lower bound that holds uniformly w.r.t \( v \in V \) on the Bernoulli process for the indices in each \( J_v^c \)) and Lemma 5.8 (an upper bound that holds uniformly w.r.t \( v \in V \) on the Bernoulli process for the indices in each \( J_v \)).

\section*{References}

[1] Radoslaw Adamczak, Alexander E. Litvak, Alain Pajor, and Nicole Tomczak-Jaegermann. Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. \textit{J. Amer. Math. Soc.}, 23(2):535–561, 2010.

[2] Shiri Artstein-Avidan, Apostolos Giannopoulos, and Vitali D. Milman. \textit{Asymptotic geometric analysis. Part I}, volume 202 of \textit{Mathematical Surveys and Monographs}. American Mathematical Society, Providence, RI, 2015.

[3] Silouanos Brazitikos, Apostolos Giannopoulos, Petros Valettas, and Beatrice-Helen Vritsiou. \textit{Geometry of isotropic convex bodies}, volume 196 of \textit{Mathematical Surveys and Monographs}. American Mathematical Society, Providence, RI, 2014.

[4] Djalil Chafai, Olivier Guédon, Guillaume Lecué, and Alain Pajor. \textit{Interactions between compressed sensing random matrices and high dimensional geometry}, volume 37 of \textit{Panoramas et Synthèses}. Société Mathématique de France, Paris, 2012.

[5] Aryeh Dvoretzky. Some results on convex bodies and Banach spaces. In \textit{Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960)}, pages 123–160. Jerusalem Academic Press, Jerusalem; Pergamon, Oxford, 1961.

[6] Pawel Hitczenko. Domination inequality for martingale transforms of a Rademacher sequence. \textit{Israel J. Math.}, 84(1-2):161–178, 1993.

[7] Michel Ledoux and Michel Talagrand. \textit{Probability in Banach spaces}, volume 23 of \textit{Ergebnisse der Mathematik und ihrer Grenzgebiete (3)}. Springer-Verlag, Berlin, 1991.
[8] Shahar Mendelson and Grigoris Paouris. On the singular values of random matrices. *J. Eur. Math. Soc. (JEMS)*, 16(4):823–834, 2014.

[9] Shahar Mendelson and Nicole Tomczak-Jaegermann. A subgaussian embedding theorem. *Israel J. Math.*, 164:349–364, 2008.

[10] V. D. Milman. A new proof of A. Dvoretzky’s theorem on cross-sections of convex bodies. *Funkcional. Anal. i Priložen.*, 5(4):28–37, 1971.

[11] G. Paouris. Concentration of mass on convex bodies. *Geom. Funct. Anal.*, 16(5):1021–1049, 2006.

[12] Gilles Pisier. Probabilistic methods in the geometry of Banach spaces. In *Probability and analysis (Varenna, 1985)*, volume 1206 of *Lecture Notes in Math.*, pages 167–241. Springer, Berlin, 1986.

[13] Gilles Pisier. *The volume of convex bodies and Banach space geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989.

[14] Michel Talagrand. *Upper and lower bounds for stochastic processes*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge*. Springer, Heidelberg, 2014.

[15] Konstantin Tikhomirov. Sample covariance matrices of heavy-tailed distributions. *Int. Math. Res. Not. IMRN*, (20):6254–6289, 2018.

[16] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996.

[17] Roman Vershynin. *High-dimensional probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018.