Empirical Processes and random projections

B. Klartag[1][*], S. Mendelson[2][†]

[1] School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA.
[2] Institute of Advanced Studies, The Australian National University, Canberra, ACT 0200, Australia.

Abstract

In this note we establish some bounds on the supremum of certain empirical processes indexed by sets of functions with the same L_2 norm. We present several geometric applications of this result, the most important of which is a sharpening of the Johnson-Lindenstrauss embedding Lemma. Our results apply to a large class of random matrices, as we only require that the matrix entries have a subgaussian tail.

1   Introduction

The goal of this article is to present bounds on the supremum of some empirical processes and to use these bounds in certain geometric problems. The original motivation for this study was the following problem: let $X_1, ..., X_k$ be independent random vectors in $\mathbb{R}^n$ and let $\Gamma : \mathbb{R}^n \to \mathbb{R}^k$ be the random operator $\Gamma v = \sum_{i=1}^k \langle X_i, v \rangle e_i$, where $\{e_1, ..., e_k\}$ is the standard orthonormal basis in $\mathbb{R}^k$. Assume for simplicity that $\mathbb{E}\langle X_i, t \rangle^2 = 1$ for any $1 \le i \le k$ and any $t \in S^{n-1}$, where S^{n-1} denotes the Euclidean sphere in $\mathbb{R}^n$. Given a subset $T \subset S^{n-1}$, we ask whether the (random) operator $\frac{1}{\sqrt{k}}\Gamma : \ell_2^n \to \ell_2^k$ almost preserves the norm of all elements of T. To ensure that Γ is almost norm preserving on T it suffices to analyze the supremum of the random variables $Z_t^k = \frac{1}{k}\|\Gamma t\|_{\ell_2^k}^2 - 1$, and to show that with positive probability, $\sup_{t \in T} |Z_t^k|$ is sufficiently small.

An important example is when the random vector $X_i = (\xi_1^i, ..., \xi_n^i)$, and $(\xi_j^i)_{i,j=1}^n$ are independent random variables with expectation 0, variance 1 and have a subgaussian tail, and thus, Γ is a random $k \times n$ subgaussian matrix. A standard concentration argument applied to each Z_t^k individually shows that if $|T| = n$ and if $k \ge c\frac{\log n}{\varepsilon^2}$, then with high probability, for every $t \in T$, $1 - \varepsilon \le \frac{1}{\sqrt{k}}\|\Gamma t\|_{\ell_2^k} \le 1 + \varepsilon$, and thus, Γ almost preserves the norm of all the elements in T. This simple fact is the basis of the celebrated Johnson-Lindenstrauss Lemma [7] which analyzes the ability to almost isometrically embed subsets of ℓ_2 in a

---

Hilbert space of a much smaller dimension; that is, for $A \subset \ell_2$, to find a mapping $f : T \to \ell_2^k$ such that for any $s, t \in A$

$$(1 - \varepsilon)\|s - t\|_{\ell_2} \leq \|f(s) - f(t)\|_{\ell_2^k} \leq (1 + \varepsilon)\|s - t\|_{\ell_2}.$$

We will call such a mapping an ε-isometry of A. In [7], Johnson and Lindenstrauss proved the following:

Theorem 1.1 *There exists an absolute constant c for which the following holds. If $A \subset \ell_2$, $|A| = n$ and $k = c\frac{\log n}{\varepsilon^2}$, there exists an ε-isometry $f : \ell_2 \to \ell_2^k$.*

The actual proof yields a little more than this statement, from which the connection to the problem we are interested in should become clearer. By considering the set of normalized differences $T = \left\{ \frac{x_i - x_j}{\|x_i - x_j\|} : i \neq j \right\}$, if one can find a linear mapping which almost preserves the norms of elements in T, this mapping would be the desired ε-isometry; and indeed, a random $k \times n$ subgaussian matrix is such a mapping.

Of course, the "complexity parameter" for T in the Johnson-Lindenstrauss Lemma is rather unsatisfying - the logarithm of the cardinality of the set T. One might ask if there is another, sharper complexity parameter that could be used in this case.

Let us consider a more general situation. Let (T, d) be a metric space, let $F_T = \{f_t : t \in T\} \subset L_2(\mu)$ be a set of functions on a probability space (Ω, μ) and denote $F_T^2 = \{f^2 : f \in F_T\}$. Given $(X_i)_{i=1}^k$, independent random variables distributed according to μ, let $\mu_k = \frac{1}{k}\sum_{i=1}^k \delta_{X_i}$ be the empirical uniform probability measure supported on $X_1, ..., X_k$ (i.e. μ_k is a random discrete measure). For any class of functions $F \subset L_1(\mu)$, set

$$\|\mu_k - \mu\|_F = \sup_{f \in F}\left|\int f d\mu_k - \int f d\mu\right| = \sup_{f \in F}\left|\frac{1}{k}\sum_{i=1}^k f(X_i) - \mathbb{E}f\right|,$$

and note that $\|\mu_k - \mu\|_F$ is a random variable. Our goal is to find a bound for $\|\mu_k - \mu\|_{F_T^2}$ that holds with a positive probability, under the assumption that $\int f^2 d\mu = 1$ for any $f \in F_T$. This bound should depend on the geometry of F_T (and, under some additional mild assumptions, reflect the metric structure of (T, d)). One possible way of bounding $\|\mu_k - \mu\|_F$ which is almost sharp, is based on a symmetrization argument due to Giné and Zinn [5], namely, that for every set F

$$\mathbb{E}\|\mu_k - \mu\|_F \leq C\mathbb{E}_X\mathbb{E}_g\left\|\frac{1}{k}\sum_{i=1}^k g_i\delta_{X_i}\right\|_F,$$

where $(g_i)_{i=1}^k$ are standard independent gaussian random variables and $C > 0$ is an absolute constant. Let

$$P_{X_1, .., X_k}(F) = \{(f(X_1), ..., f(X_k)) : f \in F\} \subset \mathbb{R}^k$$

2

be a random coordinate projection of F. Fixing such a random projection, one needs to bound the expectation of the gaussian process indexed by that projection. To that end, let us remind the reader of the definition of the γ-functionals (see [18]):

Definition 1.2 For a metric space (T, d) define

$$\gamma_\alpha(T, d) = \inf \sup_{t \in T} \sum_{s=0}^\infty 2^{s/\alpha} d(t, T_s),$$

where the infimum is taken with respect to all subsets $T_s \subset T$ with cardinality $|T_s| \leq 2^{2^s}$ and $|T_0| = 1$. If the metric is clear we denote the γ functionals by $\gamma_\alpha(T)$.

By the celebrated "majorizing measures" Theorem (see the discussion before Theorem 2.1 for more details), if $\{X_t : t \in T\}$ is a gaussian process, and $d_2^2(s, t) = \mathbb{E}|X_s - X_t|^2$ is its covariance structure, then

$$c_1 \gamma_2(T, d_2) \leq \mathbb{E} \sup_{t \in T} X_t \leq c_2 \gamma_2(T, d_2), \tag{1}$$

where $c_1, c_2 > 0$ are absolute constants. Observe that for every $X_1, ..., X_k$ the process $\left\{ \frac{1}{\sqrt{k}} \sum_{i=1}^k g_i f(X_i) : f \in F \right\}$ is gaussian and its covariance structure satisfies

$$\mathbb{E}_g \left| \frac{1}{\sqrt{k}} \sum_{i=1}^k g_i f_1(X_i) - \frac{1}{\sqrt{k}} \sum_{i=1}^k g_i f_2(X_i) \right|^2 = \|f_1 - f_2\|_{L_2(\mu_k)}^2.$$

Thus,

$$\mathbb{E}\|\mu_k - \mu\|_F \leq \frac{C}{\sqrt{k}} \mathbb{E}_X \gamma_2(P_{X_1,..,X_k}(F), L_2(\mu_k)),$$

and in particular this general bound holds for the set F_T^2. Unfortunately, although this bound is almost sharp and is given solely in terms of γ_2, it is less than satisfactory for our purposes as it involves averaging γ_2 of random coordinate projections of F with respect to the random $L_2(\mu_k)$ structure, and thus could be rather hard to estimate.

On the other hand, if one wishes to estimate the expectation of the supremum of a general empirical process $\mathbb{E}\|\mu_k - \mu\|_F$ with a more accessible geometric parameter of the set F which does not depend on the random coordinate structure of F, the best that one can hope for in general is a bound which is a combination of γ_1 and γ_2. This too follows from Talagrand's generic chaining approach [18] (see Theorem 2.1 and Lemma 3.2 here).

The main result we present here is that if a process is indexed by a set of functions with the same L_2 norm, then with non-zero probability the γ_1 part of the bound can be removed. In some sense, the process behaves as if it were a purely subgaussian process, rather than a combination of a subgaussian and sub-exponential.

To formulate the exact result, we require the notion of the Orlicz norms ψ_p. Let X be a random variable and define the ψ_p norm of X as $\|X\|_{\psi_p} = \inf_{C>0} \mathbb{E} \exp\left(\frac{|X|^p}{C^p}\right) \leq 2$. A standard argument shows that if X has a bounded ψ_p norm then the tail of $|X|$ decays faster than $2\exp(-u^p/\|X\|_{\psi_p}^p)$ [19]. In particular, for $p = 2$, it means that X has a subgaussian tail.

Theorem 1.3 *Let (Ω, μ) be a probability space and let $X, X_1, X_2, ..., X_k$ be independent random variables distributed according to μ. Set T to be a collection of functions, such that for every $f \in T$, $\mathbb{E}f^2(X) = \|f\|_{L_2}^2 = 1$ and $\|f\|_{\psi_2} \leq \beta$. Define the random variable*

$$Z_f^k = \frac{1}{k}\sum_{i=1}^{k} f^2(X_i) - \|f\|_{L_2}^2.$$

Then for any $e^{-c'\gamma_2^2(T,\|\cdot\|_{\psi_2})} < \delta < 1$, with probability larger than $1 - \delta$,

$$\sup_{f \in T} |Z_f^k| \leq c(\delta, \beta)\frac{\gamma_2(T, \|\cdot\|_{\psi_2})}{\sqrt{k}} \tag{2}$$

where $c' > 0$ is an absolute constant and $c(\delta, \beta)$ depends solely on δ, β.

The three applications we present have a geometric flavor. The first, which motivated our study, is concerned with random projections and follows almost immediately from Theorem 1.3.

Theorem 1.4 *For every $\beta > 0$ there exists a constant $c(\beta)$ for which the following holds. Let $T \subset S^{n-1}$ be a set and let $\Gamma : \mathbb{R}^n \to \mathbb{R}^k$ be a random matrix whose entries are independent, identically distributed random variables, that satisfy $\mathbb{E}\Gamma_{i,j} = 0$, $\mathbb{E}\Gamma_{i,j}^2 = 1$ and $\|\Gamma_{i,j}\|_{\psi_2} \leq \beta$. Then, with probability larger than $1/2$, for any $x \in T$ and $\varepsilon \geq c\frac{\gamma_2(T)}{\sqrt{k}}$,*

$$1 - \varepsilon \leq \frac{1}{\sqrt{k}}\|\Gamma x\|_{\ell_2^k} < 1 + \varepsilon.$$

In the case of a Gaussian projection, our result was proved by Gordon (Corollary 1.3 in [6]). However, that proof uses the special structure of Gaussian variables and does not seem to generalize to arbitrary subgaussian random variables.

Note that for a set $T \subset S^N$ of cardinality n, $\gamma_2(T, \|\cdot\|_2) \leq c\sqrt{\log n}$. In particular, Theorem 1.4 improves the Johnson-Lindenstrauss embedding result for any random operator with i.i.d ψ_2 entries which are centered and have variance 1. From here on we shall refer to such an operator as a ψ_2 *operator*. The novelty in Theorem 1.4 is that the *logn* term can be improved to $\gamma_2^2(T)$ in the general case, for any ψ_2 operator.

The second application that follows from the theorem is an estimate on the so called *Gelfand numbers* of a convex, centrally symmetric body, that is, on the ability to find sections of the body of a "small" Euclidean diameter (a question which is also known as the "low M^*-estimate"). Indeed, with a similar line of reasoning to the one used in [11, 6, 9] (where such an estimate on the Gelfand numbers was obtained using a random orthogonal projection or a gaussian operator), one can establish the next corollary.

Corollary 1.5 *There exists an absolute constant c for which the following holds. Let $V \subset \mathbb{R}^n$ be a convex, centrally symmetric body, set $1 \leq k \leq n$ and let Γ be a random, $k \times n$ ψ_2 matrix. Then, with probability at least $1/2$, $\operatorname{diam}(V \cap \ker(\Gamma)) \leq c\gamma_2(V)/\sqrt{k}$.*

Let us point out that this corollary does not use the full strength of our result, as its proof only requires that for $\rho \sim \gamma_2(V)/\sqrt{k}$, $0 \notin \Gamma T$, where $T = V/\rho \cap S^{n-1}$. Thus, Corollary 1.5 is implied by a one-sided isomorphic condition, that $0 < \inf_{t \in T} \|\Gamma t\|$, rather than the two-sided quantitative estimate we actually have in Theorem 1.4.

The third application is related to a Dvorezky-type theorem. As $T \subset \mathbb{R}^n$, the parameter $\gamma_2(T)$ has a natural geometric interpretation: it is approximately proportional to the mean width of T. Indeed, if $T \subset S^{n-1} \subset \mathbb{R}^n$ then $\gamma_2(T)$ is equivalent to $\sqrt{n}w(T)$, where $w(T)$ is the mean-width of $T \subset \mathbb{R}^n$ (see, for example, [10], chapter 9). Note that exactly the same estimate as in Theorem 1.4, which is $k \geq c\left(\frac{\gamma_2(T)}{\varepsilon}\right)^2$, appears in a result related to Dvorezky's Theorem due to Milman (e.g. [9], Section 2.3.5). There, random k-dimensional projections of a convex body are analyzed for k much larger than the critical value - that is, k is larger than the dimension in which a random projection is almost Euclidean. It turns out that even though a typical projection is far from being Euclidean, a regular behavior may be observed for the diameter of the projected body. In our setting, this result could be described as follows: if $T \subset S^{n-1}$ is a centrally-symmetric set and Γ is a random orthogonal projection, then for $k \geq C\frac{\gamma_2^2(T)}{\varepsilon^2}$ with high probability

$$1 - \varepsilon \leq \max_{t \in T} \frac{1}{\sqrt{k}}\|\Gamma t\|_{\ell_2^k} \leq 1 + \varepsilon. \tag{3}$$

This consequence of a Dvorezky-type theorem also follows from our arguments, even for an arbitrary ψ_2-matrix Γ. Let us emphasize that for $k \sim \gamma_2^2(T)$ the convex hull of ΓT is approximately a Euclidean ball. Thus, for $k << \gamma_2^2(T)$ it is impossible to obtain an estimate in the spirit of (3), even for $\varepsilon = \frac{1}{2}$, as the Euclidean ball cannot be well embedded in a lower-dimensional space.

We end this introduction with a notational convention. Throughout, all absolute constants are positive and are denoted by c or C. Their values may change from line to line or even within the same line. $C(\varphi)$ and C_φ denote a constant which depends only on the parameter φ (which is usually a real number), and $a \sim_\varphi b$ means that $c_\varphi b \leq a \leq C_\varphi b$. If the constants are absolute we use the notation $a \sim b$.

2 Generic chaining

Our analysis is based on Talagrand's *generic chaining* approach [17, 18], which is a way of controlling the expectation of the supremum of centered processes indexed by a metric space (compare also with [1] for another application of a similar flavor). The generic chaining was introduced by Talagrand [18] as a way of simplifying the *majorizing measures* approach used in the analysis of some stochastic processes (in particular of gaussian processes), and in numerous other applications (see [18, 16, 3] and references therein). We will not discuss this beautiful approach in detail, but rather mention the results we require.

The majorizing measures Theorem (in its modern version) states, as shown in (1), that for any $T \subset \ell_2$, $\gamma_2(T, \| \ \|_2)$ is equivalent to the expectation of the supremum of the gaussian process $\{X_t : t \in T\}$, where $X_t = \sum_{i=1}^{\infty} g_i t_i$. The upper bound in its original form is due to Fernique (building on earlier ideas of Preston) [4, 12, 13], while the lower was established by Talagrand in [15]. Let us reformulate the corresponding result for subgaussian processes as well as for processes whose tails are a mixture of a subgaussian part and a sub-exponential part.

Theorem 2.1 *[18] Let d_1 and d_2 be metrics on T space and let $\{Z_t : t \in T\}$ be a centered stochastic process.*

1. *If there is a constant α such that for every $s, t \in T$ and every $u > 0$,*
$$Pr\left(\{|Z_s - Z_t| > u\}\right) \leq \alpha \exp\left(-\frac{u^2}{d_2^2(s,t)}\right), \text{ then for any } t_0 \in T,$$

$$\mathbb{E}\sup_{t \in T}|Z_t - Z_{t_0}| \leq C(\alpha)\gamma_2(T, d_2).$$

2. *If there is a constant α such that for every $s, t \in T$ and every $u > 0$,*
$$Pr\left(\{|Z_s - Z_t| > u\}\right) \leq \alpha \exp\left(-\min\left(\frac{u^2}{d_2^2(s,t)}, \frac{u}{d_1(s,t)}\right)\right), \text{ then for any } t_0 \in T,$$
$$\mathbb{E}\sup_{t \in T}|Z_t - Z_{t_0}| \leq C(\alpha)\left(\gamma_2(T, d_2) + \gamma_1(T, d_1)\right).$$

The main result of this article, formulated in Theorem 1.3, follows as a particular case of the next proposition and the rest of this section is devoted to its proof.

Proposition 2.2 *Let (T, d) be a metric space and let $\{Z_x\}_{x \in T}$ be a stochastic process. Let $k > 0$, $\varphi : [0, \infty) \to \mathbb{R}$ and set $W_x = \varphi(|Z_x|)$ and $\varepsilon = \frac{\gamma_2(T,d)}{\sqrt{k}}$. Assume that for some $\eta > 0$ and $e^{-c_1(\eta)k} < \delta < \frac{1}{4}$ the following holds:*

(1) For any $x, y \in T$ and $t < \delta_0 = \frac{4}{\eta} \log \frac{1}{\delta}$,

$$Pr\left(|Z_x - Z_y| > td(x,y)\right) < exp\left(-\frac{\eta}{\delta_0}kt^2\right).$$

(2) For any $x, y \in T$ and $t > 1$,

$$Pr\left(|W_x - W_y| > t d(x, y)\right) < exp(-\eta k t^2).$$

(3) For any $x \in T$, with probability larger than $1 - \delta$, $|Z_x| < \varepsilon$.

(4) φ is increasing, differentiable at zero and $\varphi'(0) > 0$.

Then, with probability larger than $1 - 2\delta$,

$$\sup_{x \in T} |Z_x| < C\varepsilon$$

where $C = C(\varphi, \delta, \eta) > 0$ depends solely on its arguments.

Note the difference between requirements (1) and (2) in Proposition 2.2. In (1) we are interested in the domain $t < C$, while in (2) the domain of interest is $t > 1$. The proof is composed of two steps. The first reduces the problem of estimating $\sup_{t \in T} |Z_t|$ to the easier problem of estimating $\sup_{t \in T'} Z_t$, where $T' \subset T$ is an ε-cover of T (that is, a set T' such that for any $t \in T$ there is some $s \in T'$ for which $d(t, s) \leq \varepsilon$).

Lemma 2.3 *Let $\{W_t\}_{t \in T}$ be a stochastic process that satisfies requirement (2) in Proposition 2.2. Let also ε, k, η be as in Proposition 2.2. Then there exist constants c_1 and c_2 which depend solely on η from Proposition 2.2 and a set $T' \subset T$ with $|T'| \leq 4^k$, such that with probability larger than $1 - e^{-c_1 k}$,*

$$\sup_{t \in T} W_t < \sup_{t \in T'} W_t + c_2 \varepsilon.$$

Proof. Let $\{T_s : s \geq 0\}$ be a collection of subsets of T such that $|T_s| = 2^{2^s}$, $|T_0| = 1$ and for which $\gamma_2(T, d)$ is "almost" attained. For any $x \in T$, let $\pi_s(x)$ be a nearest point to x in T_s with respect to the metric d. We may assume that $\gamma_2(T, d) < \infty$, and thus the sequence $(\pi_s(x))$ converges to x for any $x \in T$. Clearly, by the definition of γ_2 and the triangle inequality,

$$\sum_{i=0}^{\infty} 2^{s/2} d(\pi_s(x), \pi_{s+1}(x)) \leq C \gamma_2(T, d). \tag{4}$$

Let s_0 be the minimal integer such that $2^{s_0} > k$, put $T' = T_{s_0}$ and note that $|T'| = 2^{2^{s_0}} < 2^{2k} = 4^k$. It remains to prove that with high probability, for any $x \in T$

$$W_x - W_{\pi_{s_0}(x)} = \sum_{s=s_0}^{\infty} \left(W_{\pi_{s+1}(x)} - W_{\pi_s(x)}\right) < C\varepsilon, \tag{5}$$

as this implies that $\sup_{t \in T} W_t \leq \sup_{t \in T'} W_t + C\varepsilon$. Fix $s > s_0$ and $x \in T_{s+1}, y \in T_s$. Since $\frac{\log |T_s||T_{s+1}|}{k} > 1$ then by the increment assumption on W, the probability that

$$|W_x - W_y| < c'_\eta \sqrt{\frac{\log |T_s||T_{s+1}|}{k}} d(x, y) \tag{6}$$

7

is larger than $1 - e^{c''_\eta \log |T_s||T_{s+1}|} > 1 - \frac{1}{|T_s|^2|T_{s+1}|}$ for the appropriate choice of the constants $c'_\eta, c''_\eta > 0$ which depend only on η. Hence, with probability larger than $1 - \frac{1}{|T_s|}$, condition (6) holds for all $x \in T_{s+1}$ and $y \in T_s$. It follows that with probability larger than $1 - \exp(-\tilde{c}_\eta k)$, for all $x \in T$ and any $s \geq s_0$,

$$\left|W_{\pi_{s+1}(x)} - W_{\pi_s(x)}\right| < C_\eta 2^{s/2} \frac{1}{\sqrt{k}} d(\pi_{s+1}(x), \pi_s(x)),$$

since $\log |T_s| = 2^s$. Applying (4), with probability at least $1 - \exp(-c'_\eta k)$, for any $x \in T$,

$$W_x - W_{\pi_{s_0}(x)} \leq \sum_{s=s_0}^{\infty} |W_{\pi_{s+1}(x)} - W_{\pi_s(x)}| < \frac{C_\eta}{\sqrt{k}} \sum_{s=s_0}^{\infty} 2^{s/2} d(\pi_{s+1}(x), \pi_s(x))$$

$$< C_\eta \frac{\gamma_2(T)}{\sqrt{k}}$$

and thus (5) holds, as $\frac{\gamma_2(T)}{\sqrt{k}} < \varepsilon$. ∎

The next step is to bound the maximum of a stochastic process, when indexed by a small set.

Lemma 2.4 *Let (T', d) be a metric space, assume that $|T'| \leq 4^k$ and let $\{Z_t\}_{t \in T'}$ be a stochastic process that satisfies requirements (1) and (3) in Proposition 2.2. If $\eta, \varepsilon, \delta, k$ are as in Proposition 2.2 then with probability larger than $1 - \frac{3}{2}\delta$,*

$$\sup_{t \in T'} |Z_t| \leq C(\eta, \delta)\varepsilon$$

where $C(\eta, \delta)$ depends solely on δ and η.

Proof. Let $\{T_s : s \in \mathbb{N}\}$ be a collection of subsets of T' such that $|T_s| \leq 2^{2^s}$, $|T_0| = 1$ and for which $\gamma_2(T')$ is "almost" attained. Again, for any $x \in E$, let $\pi_s(x)$ be a nearest point to x in T_s, and because $|T'| < 4^k$, we may assume that $T' = T_s$ for $s > s_0 = \lfloor \log_2 k \rfloor + 1$. Thus, it is evident that

$$Z_x - Z_{\pi_0(x)} = \sum_{s=1}^{\infty} \left(Z_{\pi_s(x)} - Z_{\pi_{s-1}(x)}\right) = \sum_{s=1}^{\lfloor \log_2 k \rfloor + 2} \left(Z_{\pi_s(x)} - Z_{\pi_{s-1}(x)}\right).$$

Fix some $u \leq \frac{\delta_0}{2} = 2\frac{1}{\eta} \log \frac{1}{\delta}$ and $s \leq s_0$. Selecting $t = u2^{(s+1)/2}/\sqrt{k} \leq \delta_0$ in condition (1) in Proposition 2.2, it follows that for a fixed $x \in T'$,

$$Pr\left(\left|Z_{\pi_s(x)} - Z_{\pi_{s-1}(x)}\right| > u2^{(s+1)/2} \frac{d(\pi_s(x), \pi_{s-1}(x))}{\sqrt{k}}\right) \leq e^{-\frac{\eta}{\delta_0}u^2 2^{s+1}}.$$

Hence, with probability larger than $1 - \sum_{s=1}^{\infty} 2^{2^s} 2^{2^{s-1}} e^{-\frac{\eta}{\delta_0}u^2 2^{s+1}}$, for every $x \in$

T'

$$|Z_x - Z_{\pi_0(x)}| = \left| \sum_{s=1}^{s_0+1} \left(Z_{\pi_s(x)} - Z_{\pi_{s-1}(x)} \right) \right| \le \sqrt{2}u \sum_{s \ge 1} 2^{s/2} \frac{d(\pi_s(x), \pi_{s-1}(x))}{\sqrt{k}}$$

$$\le cu \frac{\gamma_2(T')}{\sqrt{k}}.$$

By condition (3) in Proposition 2.2, with probability larger than $1 - \delta$ we have $|Z_{\pi_0(x)}| < \varepsilon$. Hence, with probability larger than $1 - \sum_{s=1}^{\infty} 2^{2^{s+1}} e^{-\frac{\eta}{\delta_0} u^2 2^s} - \delta$, for every $x \in T'$,

$$|Z_x| < |Z_{\pi_0(x)}| + cu \frac{\gamma_2(T')}{\sqrt{k}} < |Z_{\pi_0(x)}| + cu\varepsilon < c(u+1)\varepsilon.$$

It remains to estimate the probability, and to that end, put $u = \frac{2}{\eta} \log \frac{1}{\delta}$ and observe that

$$1 - \sum_{s=1}^{\infty} 2^{2^{s+1}} e^{-\frac{\eta}{\delta_0} u^2 2^{s+1}} - \delta \ge 1 - \delta - \sum_{s=2}^{\infty} (2\delta)^s > 1 - \frac{3}{2}\delta,$$

as claimed. ∎

Proof of Proposition 2.2. By condition (4),

$$\tilde{C} = \tilde{C}(\varphi) = \sup_{0 < t < 1} \frac{\varphi^{-1}(\varphi(t) + t)}{t} < \infty$$

and hence for any $0 < \varepsilon < 1$,

$$\varphi(\varepsilon) + \varepsilon < \varphi(\tilde{C}\varepsilon). \tag{7}$$

Let $T' \subset T$ be the set from Lemma 2.3. Applying Lemma 2.3 and since φ is increasing,

$$\sup_{t \in T} |Z_t| = \varphi^{-1} \left(\sup_{t \in T} W_t \right) \le \varphi^{-1} \left(\sup_{t \in T'} W_t + c_2 \varepsilon \right)$$

$$= \varphi^{-1} \left(\varphi \left(\sup_{t \in T'} |Z_t| \right) + c_2 \varepsilon \right)$$

with probability larger than $1 - e^{-c_1 k} > 1 - \frac{\delta}{2}$. Since $|T'| < 4^k$ then by Lemma 2.4, with probability larger than $1 - \frac{3}{2}\delta$,

$$\sup_{t \in T'} |Z_t| < C(\eta, \delta)\varepsilon.$$

and hence, by (7), with probability larger than $1 - 2\delta$

$$\sup_{t \in T} |Z_t| < \varphi^{-1} \left[\varphi \left(C(\eta, \delta)\varepsilon \right) + c_2 \varepsilon \right] < \varphi^{-1} \varphi \left[\tilde{C}(\varphi) \max\{C(\eta, \delta), c_2\}\varepsilon \right]$$

$$= c(\eta, \delta, \varphi)\varepsilon.$$

∎

3 Empirical Processes

Next we show how Theorem 1.3 follows from Proposition 2.2. A central ingredient in the proof is the well known Bernstein's inequality.

Theorem 3.1 *[19, 2] Let $X_1, ..., X_k$ be independent random variables with zero mean such that for every i and every $m \geq 2$, $\mathbb{E}|X_i|^m \leq m! M^{m-2} v_i/2$. Then, for any $v \geq \sum_{i=1}^{k} v_i$ and any $u > 0$,*

$$Pr\left(\left\{\left|\sum_{i=i}^{k} X_i\right| > u\right\}\right) \leq 2\exp\left(-\frac{u^2}{2(v+uM)}\right).$$

It is easy to see that if $\mathbb{E}\exp(|X|/b) \leq 2$, i.e., if $\|X\|_{\psi_1} \leq b$ then $\sum_{m=1}^{\infty} \frac{\mathbb{E}|X|^m}{b^m m!} \leq 2$, and the assumptions of Theorem 3.1 are satisfied for $M = \|X\|_{\psi_1}$ and $v = 4k\|X\|_{\psi_1}^2$. Hence,

$$Pr\left(\left\{\left|\frac{1}{k}\sum_{i=i}^{k} X_i\right| > u\right\}\right) \leq 2\exp\left(-ck\min\left\{\frac{u^2}{\|X\|_{\psi_1}^2}, \frac{u}{\|X\|_{\psi_1}}\right\}\right) \qquad (8)$$

Lemma 3.2 *Using the notation of Theorem 1.3, for any $f, g \in T$ and $u > 0$,*

$$Pr\left(\left|Z_f^k - Z_g^k\right| > u\|f-g\|_{\psi_2}\right) \leq \exp\left(-ck\min\{u, u^2\}\right)$$

and

$$Pr\left(\left|Z_f^k\right| > u\right) \leq \exp\left(-c'(\beta)k\min\{u, u^2\}\right)$$

where c is a universal constant and $c'(\beta)$ depends solely on β from Theorem 1.3.

Proof. Clearly,

$$Z_f^k - Z_g^k = \frac{1}{k}\sum_{i=1}^{k}(f-g)(X_i)(f+g)(X_i).$$

Let $Y_i = (f-g)(X_i) \cdot (f+g)(X_i) = f^2(X_i) - g^2(X_i)$. Then for any $u > 0$,

$Pr\left(|Y_i| > 4u\|f-g\|_{\psi_2}\|f+g\|_{\psi_2}\right) \leq$
$Pr\left(|(f-g)(X_i)| > 2\sqrt{u}\|f-g\|_{\psi_2}\right) + Pr\left(\left\{|(f+g)(X_i)| > 2\sqrt{u}\|f+g\|_{\psi_2}\right\}\right)$
$\qquad \leq 2e^{-u},$

which implies that $\|Y_i\|_{\psi_1} \leq c_1\|f-g\|_{\psi_2}\|f+g\|_{\psi_2} \leq c_2\|f-g\|_{L_2}$, as $\|f\|_{L_2} = \|g\|_{L_2} = 1$. In particular, since $Y_1, ..., Y_k$ are independent and $\mathbb{E}Y_i = 0$, then by (8)

$$Pr\left(\left|Z_f^k - Z_g^k\right| > u\right) \leq 2\exp\left(-ck\min\left\{\frac{u^2}{\|f-g\|_{L_2}^2}, \frac{u}{\|f-g\|_{L_2}}\right\}\right).$$

The estimate for $Pr\left(\left|Z_f^k\right| > u\right)$ follows the same path, where we define $Y_i = f^2(X_i) - 1$, use the fact that $\|f(X)\|_{\psi_2} \leq \beta$ and apply (8). ∎

Proof of Theorem 1.3. We will show that the conditions (1)-(4) of Proposition 2.2 are satisfied, for the choice of $\varphi(t) = \sqrt{1+t}$ and $d(f,g) = 2\|f-g\|_{\psi_2}$.

(1) Fix $\eta \leq c$, the constant from Lemma 3.2. Assume that $t < \delta_0 = 4\frac{1}{\eta}\log\frac{1}{\delta}$. By Lemma 3.2,

$$Pr\left(|Z_f - Z_g| > t\|f-g\|_{\psi_2}\right) < exp(-\eta k \min\{t, t^2\}) < \exp\left(-\eta k \frac{t^2}{\delta_0}\right).$$

(2) By the triangle inequality,

$$W_f - W_g = \left(\frac{1}{k}\sum_{i=1}^{k} f^2(X_i)\right)^{1/2} - \left(\frac{1}{k}\sum_{i=1}^{k} g^2(X_i)\right)^{1/2} \leq \left(\frac{1}{k}\sum_{i=1}^{k}(f-g)^2(X_i)\right)^{1/2}$$

Applying (8) for $t > 1$,

$$Pr\left(|W_f - W_g| > t\|f-g\|_{\psi_2}\right) \leq Pr\left(\frac{1}{k}\sum_{i=1}^{k}(f-g)^2(X_i) > t^2\|f-g\|_{\psi_2}^2\right)$$

$$= Pr\left(\frac{1}{k}\sum_{i=1}^{k}(f-g)^2(X_i) > t^2\|(f-g)^2\|_{\psi_1}\right) < \exp(-ckt^2)$$

$$\leq \exp(-\eta k t^2)$$

since we assumed that η is smaller than the constant in (8).

(3) For any $x \in T$, by Lemma 3.2,

$$Pr\left(|Z_x| > \varepsilon\right) < \exp(-\eta k \varepsilon^2) < \delta.$$

(4) $\varphi'(0) = \frac{1}{2} > 0$.

∎

Remark: We may formulate a variant of Theorem 1.3, as follows

Theorem 3.3 *Let (Ω, μ) be a probability space and let $X, X_1, X_2, ..., X_k$ be independent random variables distributed according to μ. Set T to be a collection of functions such that for every $f \in T$, $\mathbb{E}f(X) = 0$ and $\|f(X)\|_{\psi_1} \leq \beta$. Define a metric on T by $d(f,g) = \max\{\|(f-g)(X)\|_{\psi_1}, \sqrt{\|(f-g)(X)\|_{\psi_1}}\}$ and consider the random variable*

$$Z_f^k = \frac{1}{k}\sum_{i=1}^{k} f(X_i).$$

Then for any $e^{-c'\gamma_2^2(T,d)} < \delta < 1$, with probability larger than $1 - \delta$,

$$\sup_{f \in T} |Z_f^k| \leq c(\beta, \delta)\frac{\gamma_2(T,d)}{\sqrt{k}} \tag{9}$$

where $c(\beta, \delta)$ depends solely on β and δ and c' is an absolute constant.

The proof of Theorem 3.3 is almost identical to that of Theorem 1.3, with the only difference being the fact that requirement (2) in Proposition 2.2 is satisfied. In this case,

$$
W_f - W_g = \sqrt{1 + \left| \frac{1}{k} \sum_{i=1}^{k} f(X_i) \right|} - \sqrt{1 + \left| \frac{1}{k} \sum_{i=1}^{k} g(X_i) \right|}
$$

$$
\leq \sqrt{\left| \frac{1}{k} \sum_{i=1}^{k} (f - g)(X_i) \right|}
$$

and for $t > 1$,

$$
Pr\left(|W_f - W_g| > t\sqrt{\|f - g\|_{\psi_1}} \right) \leq Pr\left(\left| \frac{1}{k} \sum_{i=1}^{k} (f - g)(X_i) \right| > t^2 \|f - g\|_{\psi_1} \right)
$$

$$
< \exp(-ckt^2)
$$

by (8).

4 Applications to random Embeddings

Theorem 4.1 *For every $\beta > 0$ there exists a constant $c(\beta)$ for which the following holds. Let $T \subset S^{n-1}$ be a set, and let $\Gamma : \mathbb{R}^n \to \mathbb{R}^k$ be a random operator whose rows are independent vectors $\Gamma_1, .., \Gamma_k \in \mathbb{R}^n$. Assume that for any $x \in \mathbb{R}^n$ and $1 \leq i \leq k$, $\mathbb{E}\langle \Gamma_i, x \rangle^2 = \frac{1}{k} \|x\|^2$ and $\|\langle \Gamma_i, x \rangle\|_{\psi_2} \leq \beta \|\langle \Gamma_i, x \rangle\|_{L_2}$. Then, with probability larger than $1/2$, for any $x \in T$ and any $\varepsilon \geq c(\beta)\frac{\gamma_2(T)}{\sqrt{k}}$,*

$$
1 - \varepsilon \leq \|\Gamma x\|_{\ell_2^k} < 1 + \varepsilon.
$$

Proof. For $x \in T$ define

$$
Z_x^k = \|\Gamma x\|_{\ell_2^n}^2 - 1 = \frac{1}{k} \sum_{i=1}^{k} \langle \sqrt{k}\Gamma_i, x \rangle^2 - 1.
$$

Then $\mathbb{E}Z_x^k = 0$ and $\|\sqrt{k}\langle \Gamma_i, x - y \rangle\|_{\psi_2} \leq \beta \|x - y\|_{\ell_2^n}$. The assumptions of Theorem 1.3 are satisfied, and hence for $\delta = \frac{1}{2}$, with probability at least $1/2$,

$$
\sup_{x \in T} \left| \|\Gamma x\|_{\ell_2^k}^2 - 1 \right| < \varepsilon,
$$

as claimed. ∎

An important example is when the elements of the matrix Γ are independent random variables $\Gamma_{i,j}$, such that $\mathbb{E}\Gamma_{i,j} = 0$, $Var(\Gamma_{i,j}) = \frac{1}{k}$ and $\|\Gamma_{i,j}\|_{\psi_2} \leq \beta\|\Gamma_{i,j}\|_{L_2}$, since in that case, Theorem 1.4 holds. Indeed, one just needs to verify that $\|\sum_{i=1}^n \Gamma_{i,j}x_i\|_{\psi_2} \leq c(\beta)\|x\|_{\ell_2^n}$. To that end, recall that for any random variable X,

$$\|X\|_{\psi_2} < c \quad \Rightarrow \quad \forall u \; \mathbb{E}\exp(uX) < \exp\left(3c^2u^2\right) \quad \Rightarrow \quad \|X\|_{\psi_2} < 20c^2.$$

Hence, the fact that $\|\sum_{i=1}^n \Gamma_{i,j}x_i\|_{\psi_2} \leq c(\beta)\|x\|_{\ell_2^n}$ follows from

$$\mathbb{E}e^{u\sum_{i=1}^n \Gamma_{i,j}x_i} = \prod_{i=1}^n \mathbb{E}e^{ux_i\Gamma_{i,j}} \leq \prod_{i=1}^n e^{c(\beta)u^2x_i^2} = e^{c'(\beta)u^2\|x\|_{\ell_2^n}}.$$

As was mentioned in the introduction, in the special case where $\Gamma_{i,j}$ are independent, standard gaussian variables, shorter proofs of Theorem 1.4 exist. The first proof in that case is due to Gordon [6] and is based on a comparison Theorem for Gaussian variables. Another simple proof which uses Talagrand's generic chaining approach may be described as follows. Let Γ be an $k \times n$ gaussian matrix. Since the gaussian measure is rotation invariant, it follows that for any $t \in S^{n-1}$, $\frac{1}{\sqrt{k}}\mathbb{E}\|\Gamma t\|_{\ell_2^k}$ is independent of t, and we denote it by A. It is standard to verify that A is equivalent to an absolute constant.

Consider the process $Z_t = \frac{1}{A\sqrt{k}}\|\Gamma t\|_{\ell_2^k} - 1$ indexed by $T \subset S^{n-1}$ and note that this process is centered. In [14], G. Schechtman proved that there is some absolute constant c such that for any $s, t \in S^{n-1}$,

$$Pr\left(\left\{\left|\|\Gamma t\|_{\ell_2^k} - \|\Gamma s\|_{\ell_2^k}\right| \geq u\right\}\right) \leq 2\exp\left(-c\frac{u^2}{\|s-t\|_{\ell_2^n}^2}\right). \tag{10}$$

Hence, the process Z_t is subgaussian with respect to the Euclidean metric, implying that for any $t_0 \in T$

$$\mathbb{E}\sup_{t\in T}|Z_t| \leq \mathbb{E}\sup_{t\in T}|Z_t - Z_{t_0}| + \mathbb{E}\left|\frac{1}{A\sqrt{k}}\|\Gamma t_0\|_{\ell_2^k} - 1\right|$$
$$\leq C_1\frac{\gamma_2(T)}{A\sqrt{k}} + \frac{C_2}{A\sqrt{k}} \leq \varepsilon,$$

where the last inequality follows from Theorem 2.1, combined with a standard estimate on $\mathbb{E}\left|\|\Gamma t\|_{\ell_2^k} - \mathbb{E}\|\Gamma t\|_{\ell_2^k}\right|$ and the fact that $k \geq C\gamma_2^2(T)/\varepsilon^2$.

A second, analogous case in which the preceding argument works, is that of projections onto random subspaces. Consider $G_{n,k}$, the Grassmanian of k-dimensional subspaces of $\mathbb{R}^n$ and recall that there exists a unique rotation invariant probability measure on $G_{n,k}$. Assume that P is an orthogonal projection onto a random k-dimensional subspace in $\mathbb{R}^n$ and let $\Gamma = \sqrt{n}P$. Note that for a suitable A, which is equivalent to an absolute constant, the process $Z_t = \frac{1}{A\sqrt{k}}\|\Gamma t\|_{\ell_2^k} - 1$, indexed by $T \subset S^{n-1}$, is centered. Next, we shall prove

an analog of (10) for such random operators. As in the gaussian case, Theorem 2.1 implies that all of our results also hold for random operators which are orthogonal projections onto random subspaces. The proof of the following lemma is similar to Schechtman's proof in [14].

Lemma 4.2 *There exists an absolute constant c for which the following holds. Fix $s, t \in S^{n-1}$ and let P be a random projection of rank k. Then,*

$$Pr\left(\left\{\left|\|Ps\|_{\ell_2^k} - \|Pt\|_{\ell_2^k}\right| > u\sqrt{\frac{k}{n}}\|s-t\|_{\ell_2^n}\right\}\right) \leq \exp\left(-cu^2 k\right).$$

Proof. Let $d = \|s-t\|_{\ell_2^n}$ and set $\Omega_d = \{(x,y) : x, y \in S^{n-1}, \|x-y\|_{\ell_2^n} = d\}$. There exists a unique rotation invariant probability measure μ_d on Ω_d (see, for example, the first pages of [10]). Rather than fixing s, t and randomly selecting P, one may equivalently fix an orthogonal projection P of rank k and prove that

$$\mu_d\left(\left\{\left|\|Ps\|_{\ell_2^k} - \|Pt\|_{\ell_2^k}\right| > u\sqrt{\frac{k}{n}}\|s-t\|_{\ell_2^n}\right\}\right) \leq \exp\left(-cu^2 k\right).$$

Note that if one conditions on $x = \frac{s+t}{2}$ then $y = \frac{s-t}{2}$ is distributed uniformly on the sphere $\frac{d}{2}S^{n-2}$. Thus it is enough to show that for any fixed x,

$$Pr\left(\left\{y \in S^{n-2} : \left|\left\|Px + \frac{d}{2}Py\right\|_{\ell_2^k} - \left\|Px - \frac{d}{2}Py\right\|_{\ell_2^k}\right| > du\sqrt{\frac{k}{n}}\right\}\right) \leq \exp\left(-cu^2 k\right).$$

Since the function $f(y) = \left\|Px + \frac{d}{2}Py\right\|_{\ell_2^k} - \left\|Px - \frac{d}{2}Py\right\|_{\ell_2^k}$ is Lipschitz on S^{n-2} with a constant d and since its expectation is zero, then by the standard concentration inequality on the sphere (see, e.g. [10]),

$$Pr\left(y \in S^{n-2} : |f(y)| > ud\right) \leq e^{-cu^2 n},$$

which completes the proof. ∎

Note that this argument relays heavily on the structure of the gaussian random variables (particularly, the rotation invariance), which is the reason why one can have a purely subgaussian behavior for the process Z_t. For a general ψ_2 operator, the best that one could hope for off-hand is a mixture of subgaussian and sub-exponential tails, and in order to obtain an upper bound solely in terms of γ_2 one requires the more general argument.

References

[1] S. Artstein, The change in the diameter of a convex body under a random sign-projection, preprint.

[2] G. Bennett, Probability inequalities for the sum of independent random variables, J. Am. Stat. Assoc. 57, 33-45, 1962.

[3] R.M. Dudley, *Uniform Central Limit Theorems*, Cambridge Studies in Advanced Mathematics 63, Cambridge University Press, 1999.

[4] X. Fernique, Régularité des trajectoires des fonctiones aléatoires gaussiennes, Ecole d'Eté de Probabilités de St-Flour 1974, Lecture notes in Mathematics 480, Springer-Verlag 1-96, 1975.

[5] E. Giné and J. Zinn: Some limit theorems for empirical processes, Ann. Probab. 12(4), 929–989, 1984.

[6] Y. Gordon, On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$. Geom. Aspects of Funct. Anal., Israel seminar, Lecture Notes in Math., 1317, Springer-Verlag, 84-106, 1988.

[7] W.B. Johnson, J. Lindenstrauss, Extensions of Lipschitz mappings into a Hilbert space, Contemp. math. 26, 189-206, 1984.

[8] V.D. Milman, Geometrical inequalities and mixed volumes in the local theory of Banach spaces, Astérisque 131, 373-400, 1985.

[9] V.D. Milman, Topics in Asymptotic Geometric Analysis, Geom. Funct. Anal., Special Volume - GAFA2000, Birkhäuser, 792-815, 2000.

[10] V.D. Milman, G. Schechtman, *Asymptotic theory of finite dimensional normed spaces*, Lecture Notes in Mathematics 1200, Springer, 1986.

[11] A. Pajor, N. Tomczak-Jaegermann, Subspaces of small codimension of finite-dimensional Banach spaces, Proc. AMS, 97 (4), 637-642, 1986.

[12] C. Preston, Banach spaces arising from some integral inequalities, Indiana Math. J. 20, 997-1015, 1971.

[13] C. Preston, Continuity properties of some Gaussian processes, Ann. Math. Statist. 43, 285-292, 1972.

[14] G. Schechtman, A remark concerning the dependence on ϵ in Dvoretzky's theorem. Geom. Aspects of Funct. Anal. (1987–88), Lecture Notes in Math., 1376, Springer, Berlin, 274–277, 1989.

[15] M. Talagrand, Regularity of Gaussian processes, Acta Math. 159, 99-149, 1987.

[16] M. Talagrand, Majorizing Measures, the generic chaining, Ann. Probab. 24, 1049-1103, 1996.

[17] M. Talagrand, Majorizing Measures without measures, Ann. Probab. 29, 411-417, 2001.

[18] M. Talagrand, *The generic chaining*, forthcoming.

[19] A. W. Van der Vaart, J. A. Wellner, *Weak convergence and Empirical Processes*, Springer-Verlag, 1996.