Dvoretzky's Theorem and Concentration of Measure

David Miyamoto

November 13, 2016

Version 5

Contents

1 Introduction

Dvoretzky's Theorem is a result in convex geometry first proved in 1961 by Aryeh Dvoretzky. In informal terms, the theorem states that every compact, symmetric, convex (CSC) subset of $\mathbb{R}^N$ has a relatively high-dimensional slice that resembles an ellipsoid. For some intuition on the geometry involved, one could imagine 2-dimensional slices of a 3-dimensional cube. While the squares and rectangles are rather far from a 2-dimensional ellipsoid (a disk), the hexagon is somewhat closer. One could imagine that in higher dimensional cubes, we could do even better than a hexagon. Note that the 1-dimensional slice is a line segment, which is exactly an ellipsoid in 1 dimension. As this is the case for any 1-dimensional slice of any CSC body, we see that the "relatively high-dimensional" part of Dvoretzky's theorem is what keeps it from being a triviality.

The body of these notes relies heavily on Gilles Pisier's book *The Volume of Convex*

Bodies and Banach Space Geometry as well as on Yehoram Gordon's unpublished manuscript "Applications of the Gaussian Min-Max theorem".

1.1 Convex Geometry Basics

We will now formalize the language used above in order to introduce the two versions of Dvoretzky's theorem we will prove. Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be two N-dimensional Banach spaces (i.e. complete normed vector spaces). Since any two vectors spaces of the same dimension are isomorphic (in the sense that there exists an invertible linear map between them), we can define the Banach-Mazur distance between F and E as:

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

Note that $\|\cdot\|$ denotes the appropriate operator norm. For instance,

$$\|T\| = \sup\{\|T\alpha\|_F \mid \alpha \in E, \quad \|\alpha\|_E = 1\}. \tag{2}$$

Now $d(E,F)$ is at least 1 since $1 = \|TT^{-1}\| \leq \|T\|\|T^{-1}\|$. So $d(E,F) = 1$ when E and F are isometric. Then we may say E and F are close, or $(1+\epsilon)$-isomorphic, if $d(E,F) \leq 1+\epsilon$ for some small ϵ. An invertible map $T : E \to F$ such that $\|T\|\|T^{-1}\| \leq 1+\epsilon$ may be viewed as almost an isometry.

It will later be useful to develop an equivalent characterization of the Banach-Mazur distance. Suppose that $x_1, \ldots, x_N$ is a basis of E and $y_1, \ldots, y_N$ is a basis of F. Let T denote the isomorphism $x_k \mapsto y_k$. Then $d(E,F) \leq 1+\epsilon$ if for all $\alpha = \sum \alpha_i x_i \in E$:

$$\frac{1}{\sqrt{1+\epsilon}} \cdot \|\alpha\|_E \le \|T\alpha\|_F = \|\sum \alpha_i y_i\|_F \le \sqrt{1+\epsilon} \cdot \|\alpha\|_E. \tag{3}$$

To verify this, first take any $\|\alpha\|_E = 1$. Then $\|T\alpha\|_F \le \sqrt{1+\epsilon}$, so $\|T\| \le \sqrt{1+\epsilon}$. Now take any $\|\beta\|_F = 1$, then $(\sqrt{1+\epsilon})^{-1} \cdot \|T^{-1}\beta\|_E \le \|TT^{-1}\beta\|_F = 1$, so $\|T^{-1}\beta\|_E \le \sqrt{1+\epsilon}$. Thus $\|T^{-1}\| \le \sqrt{1+\epsilon}$. Hence $\|T\|\|T^{-1}\| \le 1+\epsilon$, and E is indeed $(1+\epsilon)$-isomorphic to F. This is an especially useful characterization of the Banach-Mazur distance, since (3) is much easier to check than our first definition.

As we now have a notion of distance between Banach spaces, it makes sense to draw a parallel between Banach spaces and convex bodies. More precisely, we will describe the bijection - up to isomorphism - between Banach spaces and CSC bodies in $\mathbb{R}^N$. Let $(E, \|\cdot\|_E)$ be an N-dimensional Banach space. Then its unit ball $\{v \in E \mid \|v\|_E \le 1\}$ can be identified with a CSC body B_E in $\mathbb{R}^N$ via any linear isomorphism $u : E \to \mathbb{R}^N$. Conversely, take any CSC body K in $\mathbb{R}^N$. Define the gauge $\|\cdot\|_K : \mathbb{R}^N \to [0,\infty)$ of K by

$$\|x\|_K = \inf\{r > 0 \mid x \in rK\}. \tag{4}$$

It is shown in section (A.4) that the gauge of K is a norm on $\mathbb{R}^N$, making $(\mathbb{R}^N, \|\cdot\|_K)$ a Banach space (setting aside the question of completeness). Thus, every Banach space E admits a CSC body B_E in $\mathbb{R}^N$, and vice versa. Therefore, any statement about CSC bodies in $\mathbb{R}^N$ can be interpreted as a statement about Banach spaces. With this correspondence, we present and interpret a version of Dvoretzky's theorem proved by Vitali Milman in 1971.

1.2 Motivation and the Gaussian Reformulation

Theorem (Dvoretzky's Theorem)**.** *For each $\epsilon > 0$ there is a number $\eta(\epsilon) > 0$ with the following property. Let $(E, \|\cdot\|)$ be an N-dimensional Banach space. Then E contains a* *subspace F of dimension $n = [\eta(\epsilon)\log N]$ such that $d(F, \ell_2^n) \le 1+\epsilon$.*[1]

The spaces E and F correspond to a CSC body B_E in $\mathbb{R}^N$ and a slice of this body through the origin, respectively. The slice F has dimension of order $\log N$, and $d(F,\ell_2^n) \le 1+\epsilon$ means the slice F is close to an ellipsoid. The parameter ϵ therefore dictates how close F is to an ellipsoid. What follows is an outline of the proof of this theorem using the concentration of measure phenomenon. To build motivation, we work backwards.

To prove Dvoretzky's theorem, we will show F can be found by choosing random subspaces of dimension n. To generate these random subspaces, take a linearly independent collection of $1 \le m \le N$ vectors $\{z_k\}_{k=1}^m$ in E. The magnitude of m determines an upper bound for the dimension of the subspace F which we will ultimately create from the z_k. Although the work immediately following is valid for any choice of z_k and m, to prove Dvoretzky's theorem in full we will later need to choose these parameters in a special way. For instance, we will require m to be of order $\sqrt{N}$. But for now, the z_k and m remain arbitrary.

Fix some $1 \le n \le m$. Let $\{g_{i,k}\}_{i=1}^{n}{}_{k=1}^{m}$ be independant standard Gaussian random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define $X_i = \sum_{k=1}^m g_{i,k} z_k$, giving a random vector in E. Notice that $\mathbb{E}\|X_i\| = \mathbb{E}\|X_j\|$ for all i and j, so we will denote X_1 by X and $M = \mathbb{E}\|X\|$.[2] Since choosing the basis vectors of F in the unit ball of E makes the analysis cleaner, let $Y_i = \frac{X_i}{M}$; this gives $\mathbb{E}\|Y_i\| = 1$, so we expect Y_i to have unit length. Take $F = \operatorname{span}\{Y_1, \dots, Y_n\}$, which is by construction a random n-dimensional subspace of E whose basis vectors are expected to be in the unit ball of E.

Let $\epsilon > 0$. Our ultimate goal is to show there exists some number $\eta(\epsilon)$ such that for any $n \le [\eta(\epsilon)\log N]$,

[1] ℓ_2^n denotes the Banach space $(\mathbb{R}^n, \|\cdot\|_2)$, which is $\mathbb{R}^n$ equipped with the usual Euclidean norm. Square brackets denote the floor function.

[2] Here $\mathbb{E}$ denotes the expectation.

$$\exists \omega \in \Omega \text{ s.t. } d(F(\omega), \ell_2^n) \leq 1 + \epsilon \tag{$\star$}$$

This result is a bit beyond our reach at this moment. Instead, we will determine weaker conditions on n that gaurentee ($\star$) holds. Recalling (3) from earlier in this introduction, suppose n is such that for all $\alpha \in \ell_2^n$,

$$\frac{1}{\sqrt{1+\epsilon}} \cdot \|\alpha\|_2 \leq \|\sum \alpha_i Y_i(\omega)\| \leq \sqrt{1+\epsilon} \cdot \|\alpha\|_2. \tag{5}$$

Then by the discussion surrounding (3), it follows that ($\star$) holds. Now, verifying (5) directly is challenging. However, we have a result, named Lemma 2 and proven in (2.2), that allows us to discretize the situation. Lemma 2 states that there exists some $\delta(\epsilon) = \delta > 0$ with the following property. If for some δ-net[3] A of the unit sphere in ℓ_2^n (denoted S^{n-1}), we have

$$\forall \alpha \in A, \quad \Big| \|\sum \alpha_i Y_i(\omega)\| - 1 \Big| \leq \delta, \tag{6}$$

then (5) holds, as desired. Take this δ. Our goal has now become to show that (14) holds for some δ-net A of S^{n-1} and $\omega \in \Omega$. But for any δ-net A, an ω satisfying (6) exists if

$$\mathbb{P}\left(\forall \alpha \in A, \quad \Big| \|\sum \alpha_i Y_i\| - 1 \Big| \leq \delta\right) > 0 \qquad \text{or equivalently} \tag{7}$$

$$\mathbb{P}\left(\exists \alpha \in A \text{ s.t. } \Big| \|\sum \alpha_i Y_i\| - 1 \Big| > \delta\right) < 1. \tag{8}$$

Therefore we want to choose n so that a δ-net A of S^{n-1} satisfying (8) exists. Take a

[3] Definition of δ-net: given a set $X \subseteq \ell_2^n$ and $\delta > 0$, a subset $A \subseteq X$ is a δ-net of X provided that for all points $x \in X$, there exists some $a \in A$ such that $\|x - a\|_2 \leq \delta$.

moment here to convince oneself that finding such an n will ensure that we have satisfied our ultimate goal, namely shown that $(\star)$ holds. Now, by using subadditivity and the fact $Y_i = \frac{X_i}{\mathbb{E}\,\|X\|}$, we see that the probability in (8) is bounded above by

$$|A| \cdot \mathbb{P}\left(\left|\|\textstyle\sum \alpha_i X_i\| - \mathbb{E}\,\|X\|\right| > \delta\,\mathbb{E}\,\|X\|\right), \qquad \text{for any } \alpha \in A. \tag{9}$$

We want to show this quantity can be forced below 1 by the proper choice of n. It is this desire to bound the probability in (9) that brings us to the concentration of measure phenomenon. Recall that since $X = \sum_{k=1}^{m} g_{1,k} z_k$ and the z_k are linearly independant, the third example of the concentration of measure phenomenon in the online notes gives that

$$\mathbb{P}\Big(\left|\|X\| - \mathbb{E}\,\|X\|\right| \geq \delta\,\mathbb{E}\,\|X\|\Big) \leq 2\exp\left(-\frac{\delta^2 d(X)}{4}\right).^4 \tag{10}$$

The probabilities in (10) and (9) appear quite similar; in fact, it turns out they are equal. Therefore we can use the bound in (10) to force the probability in (9) to be small. First, however, we will prove that these probabilities are equal; this is done by showing X and $\sum \alpha_i X_i$ have the same distribution. By definition of the X_i, we have

$$\sum \alpha_i X_i = \sum_{i=1}^{n} \alpha_i \left(\sum_{k=1}^{m} g_{i,k} z_k\right) = \sum_{k=1}^{m} \left(\sum_{i=1}^{n} \alpha_i g_{i,k}\right) z_k. \tag{11}$$

Now, $\sum_{i=1}^{n} \alpha_i g_{i,k}$ is a random variable distributed normally with mean 0 and variance $\sum_{i=1}^{n} \alpha_i^2$. But since $\alpha \in A$, and A is a δ-net of S^{n-1}, we know $\|\alpha\|_2 = 1$. Therefore $\sum_{i=1}^{n} \alpha_i g_{i,k} \sim N(0,1)$. So we see that X and $\sum \alpha_i X_i$ are sums of the form $\sum_{k=1}^{m} g_k z_k$, where the $g_k \sim N(0,1)$ and are independant. We may then conclude X and $\sum \alpha_i X_i$ the same distribution.

Continuing from earlier, we can now use (10) to say that

[4] Notation $d(X) = \frac{\mathbb{E}\,\|X\|}{\sigma(X)}$, where $\sigma(X) = \sup\{(\zeta(z_1)^2 + \ldots + \zeta(z_m)^2)^{\frac{1}{2}} \mid \zeta \in E^*, \quad \|\zeta\| \leq 1\}$

$$\mathbb{P}\left(\left|\|\sum \alpha_i X_i\| - \mathbb{E}\|X\|\right| > \delta\,\mathbb{E}\|X\|\right) \le 2\exp\left(-\frac{\delta^2 d(X)}{4}\right). \tag{12}$$

This means we finally have an upper bound for the quantity in (9), namely

$$2\exp\left(\log|A| - \frac{\delta^2 d(X)}{4}\right). \tag{13}$$

To recap, we have that our ultimate goal ($\star$) holds if there exists n such that we can find a δ-net A of S^{n-1} which makes (13) strictly less than 1. As an aside, notice here that (13) can also be made small by controlling $d(X)$. This is achieved through manipulating the choice of the z_k. In fact, we will perform such manipulations later, for instance in section (1.3). For now, we will restrict ourselves to controlling n.

Here we introduce a result, named Lemma 1 and proven in section (2.1), which states that given n and $\delta > 0$, there is a δ-net A of S^{n-1} of size $f(\delta, n) = \left(1 + \frac{2}{\delta}\right)^n$. Therefore, if we can find some n (remember δ is fixed by ϵ by Lemma 2) such that

$$2\exp\left(\log f(\delta, n) - \frac{\delta^2 d(X)}{4}\right) < 1. \tag{14}$$

then the δ-net A from Lemma 1 will make (13) less than 1. In particular, notice that if $n \le \frac{\delta^2}{8\log\left(1+\frac{2}{\delta}\right)} d(X)$, then the right side of (14) is less than $2\exp\left(-\frac{\delta^2 d(X)}{8}\right)$. It can be shown this quantity is less 1 in the proper circumstances. Therefore, if we set $\eta_1(\epsilon) = \frac{\delta^2}{8\log\left(1+\frac{2}{\delta}\right)}$, then any $n \le [\eta_1(\epsilon) d(X)]$ satisfies (14). Following our chain of reasoning, this shows ($\star$) holds, and we can conclude there exists a subspace F of E that is $(1+\epsilon)$-isomorphic to ℓ_2^n. We call this intermediate result the Gaussian reformulation of Dvoretzky's theorem. It is similar to Dvoretzky's theorem, but gives a different upper bound for n that depends on both ϵ and

$d(X)$. To see the detailed proof of this theorem, see section (3).

1.3 The Full Dvoretzky Theorem

We are now in a position to establish Dvoretzky's theorem in its entirety. As stated above, our Gaussian reformulation result gives n as a function of ϵ and $d(X)$, whereas Dvoretzky's theorem gives n as a function of ϵ and $\log N$. It is therefore natural to assume Dvoretzky's theorem holds as a consequence of its Gaussian reformulation because we can show that for some universal constant C,

$$C \log N \leq d(X). \tag{15}$$

Unfortunately, the existance of C depends on the construction X. As alluded to earlier, this is the point where we need to control the z_k (and therefore m). Our goal is to ensure $d(X)$ remains large, and then to conclude (15) must hold. It turns out we can extract the necessary z_k and m from an extension of the Dvoretzky-Rogers Lemma, which is stated below.

Lemma (Extended Dvoretzky-Rogers Lemma)**.** *There are* $\bar{N} = \left[\frac{\sqrt{\left[\frac{N}{2}\right]}}{16}\right]$ *linearly independent elements* $\{z_k\}$ *of* E *such that for all* $\alpha \in \mathbb{R}^{\bar{N}}$*,*

$$\frac{1}{2} \sup |\alpha_k| \leq \left\| \sum \alpha_k z_k \right\| \leq 2 \left(\sum |\alpha_k|^2 \right)^{\frac{1}{2}} \tag{16}$$

Proving this lemma is challenging, and requires its own motivation and preliminary results. Its proof is also very geometric; the concentration of measure phenomenon has already exhausted its use. To see the proof and discussion of this lemma, please refer to section (4). For now, we will show how this lemma completes the proof of Dvoretzky's theorem.

Choose $m = \bar{N}$ and the z_k from the Extended Dvoretzky-Rogers Lemma. We show $d(X) = \frac{\mathbb{E}\|X\|}{\sigma(X)}$ is relatively large. To this end, first we prove $\sigma(X) \leq c$ for some constant c.

Then all that will remain is to ensure $\mathbb{E}\,\|X\|$ is large. By our choice of z_k and m, we get $c = 2$ in the following way:

$$\sigma(X) = \sup\left\{\left(\zeta(z_1)^2 + \cdots + \zeta(z_m)^2\right)^{\frac{1}{2}} \mid \zeta \in E^*, \|\zeta\| \le 1\right\} \tag{17}$$

$$= \sup\left\{\left\|(\zeta(z_1), \ldots, \zeta(z_m))\right\|_2 \mid \|\zeta\| \le 1\right\} \tag{18}$$

$$= \sup\left\{\sup\left\{\sum_{k=1}^{m} \alpha_k \zeta(z_k) \mid \|\alpha\|_2 \le 1\right\} \mid \|\zeta\| \le 1\right\} \tag{19}$$

$$= \sup\left\{\sum_{k=1}^{m} \alpha_k \zeta(z_k) \mid \|\alpha\|_2 \le 1, \|\zeta\| \le 1\right\}. \tag{20}$$

Now $\sum \alpha_k \zeta(z_k) = \zeta\left(\sum \alpha_k z_k\right)$, which is bounded above by $\|\zeta\| \cdot \|\sum \alpha_k z_k\|$. As $\|\zeta\| \le 1$ by necessity, and $\|\sum \alpha_k z_k\| \le 2\left(\sum |\alpha_k|^2\right)^{\frac{1}{2}} \le 2$ by the choice of z_k and since $\|\alpha\|_2 \le 1$ by necessity, we can conclude $\sigma(X) \le 2$. With this bound completed, we turn our attention to ensuring $\mathbb{E}\,\|X\|$ is sufficiently large.

Since $\sigma(X) \le 2$, the inequality $C \log N \le d(X)$ follows from proving there exists some constant c' such that $c'\sqrt{\log N} \le \mathbb{E}\,\|X\|$. To get bound, first notice that by the choice of the z_k,

$$\frac{1}{2}\,\mathbb{E} \sup |g_{1,k}| \le \mathbb{E}\,\left\|\sum g_{1,k} z_k\right\|. \tag{21}$$

Then, by a technical result named Lemma 4, proven in section (5.1), we have that there exists some constant c such that $c\sqrt{\log N} \le \mathbb{E} \sup |g_{1,k}|$. Taking $c' = \frac{c}{2}$ and $C = (c')^2$, we have then shown that $C \log N \le d(X)$. By earlier remarks, we conclude we have proven Dvoretzky's theorem.

The second version of Dvoretzky's theorem we will prove was established by Yehoram Gordan in 1985. This version is identical to the first, except it gives $\eta \sim \epsilon^2$ and $d(F, \ell_2^n) \le \frac{1+\epsilon}{1-\epsilon}$. The proof uses the Gaussian Minmax theorem and an intermediate result, as well as the

Extended Dvoretzky-Rogers Lemma. See section (6) for the details of this proof.

2 Discretization Lemmas

The introduction has how we will prove Milman's version of Dvoretzky's theorem. Herein we supply the details. We begin with Lemma 1 and Lemma 2, the two discretization lemmas. Note that although we use the Euclidean norm $\|\cdot\|_2$ throughout this section, this is unnecessarily restrictive; an arbitrary norm on $\mathbb{R}^n$ would also suffice.

2.1 Lemma 1

Lemma 1. *Denote the unit ball and sphere of ℓ_2^n by B^n and S^{n-1}, respectively. For any $\delta > 0$, there is a δ-net A of S^{n-1} such that*

$$|A| \leq \left(1 + \frac{2}{\delta}\right)^n. \tag{22}$$

Proof. Let $A = \{y_i\}_{i=1}^m$ be a maximal subset of S^{n-1} such that $\|y_i - y_j\|_2 \geq \delta$ for all $i \neq j$.[5] By maximality, any other $p \in S^{n-1} \setminus A$ is a distance less than δ from some y_i. Therefore A is a δ-net of S^{n-1}. We show A satisfies the theorem.

The balls B_i of radius $\frac{\delta}{2}$ centred at y_i are pairwise disjoint and contained in the ball B_0 of radius $1 + \frac{\delta}{2}$ centred at the origin. Further, we have $\text{vol}(B_i) = \left(\frac{\delta}{2}\right)^n \text{vol}(B^n)$ and $\text{vol}(B_0) = \left(1 + \frac{\delta}{2}\right)^n \text{vol}(B^n)$. Therefore

$$m \cdot \left(\frac{\delta}{2}\right)^n \text{vol}(B^n) = \sum_{i=1}^m \text{vol}(B_i) \leq \text{vol}(B_0) = \left(1 + \frac{\delta}{2}\right)^n \text{vol}(B^n). \tag{23}$$

Dividing by $\left(\frac{\delta}{2}\right)^n \text{vol}(B^n)$ and recalling that $|A| = m$ completes the proof. $\square$

[5] In this case, maximal means no set strictly containing A satisfies the condition that its points are all seperated by a distance at least δ.

2.2 Lemma 2

Lemma 2. *For each* $\epsilon > 0$ *there exists* $0 < \delta(\epsilon) = \delta < 1$ *with the following property. Let* n *be any natural number. Let* A *be a* δ*-net of* S^{n-1} *and let* $x_1, \ldots, x_n$ *be elements of a Banach space* $(B, \|\cdot\|)$*. If for all* $\alpha \in A$ *we have*

$$1 - \delta \leq \left\| \sum_{i=1}^{n} \alpha_i x_i \right\| \leq 1 + \delta, \tag{24}$$

then for all $\alpha \in \mathbb{R}^n$*,*

$$\frac{1}{\sqrt{1+\epsilon}} \cdot \|\alpha\|_2 \leq \left\| \sum_{i=1}^{n} \alpha_i x_i \right\| \leq \sqrt{1+\epsilon} \cdot \|\alpha\|_2. \tag{25}$$

In particular, assuming the x_i *are linearly independant, the subspace generated by* $x_1, \ldots, x_n$ *is* $(1+\epsilon)$*-isomorphic to* ℓ_2^n*.*

To develop some intuition, suppose the x_i are a basis of B. Then $T : \mathbb{R}^n \to B$ defined by $T(\alpha) = \sum \alpha_i x_i$ is a linear isomorphism. The lemma states that if T almost preserves the norm of elements of A, then T is almost an isometry. Thus, if we want to T to be almost an isometry, all we need is to ensure is that T almost preserves the lengths of elements of some δ-net. This will prove especially useful if A is finite, such as the net from Lemma 1, since then we can show two spaces are $(1+\epsilon)$ isomorphic by considering a finite number of points.

To prove the lemma, we first show that for any $0 < \delta < 1$ and α in S^{n-1}, we can bound $\|\sum \alpha_i x_i\|$ above and below by functions of δ. Taking δ small enough will then give $\left(\sqrt{1+\epsilon}\right)^{-1} \leq \|\sum \alpha_i x_i\| \leq \sqrt{1+\epsilon}$. Homogeniety of the norm $\|\cdot\|$ will allow us to capture the result for all points in $\mathbb{R}^n$.

Proof. Fix $0 < \delta < 1$ and assume (24) holds for some A and $x_1, \ldots, x_n \in B$. Take $\alpha \in S^{n-1}$. Since A is a δ-net of S^{n-1}, there exists $y^0 \in A$ such that $\|\alpha - y^0\|_2 = \lambda_1 \leq \delta$. Hence we can write $\alpha = y^0 + \lambda_1 \alpha'$, where $\lambda_1 \leq \delta$ and $\alpha' = \frac{\alpha - y^0}{\lambda_1} \in S^{n-1}$. As $\alpha' \in S^{n-1}$, we can similarly write $\alpha' = y^1 + \lambda_2 \alpha''$ where $y^1 \in A$, $\lambda_2 \leq \delta$, and $\alpha'' \in S$. Continuing, we get

$$\alpha = y^0 + \lambda_1(y^1 + \lambda_2(y^2 + \lambda_3(\ldots))) \tag{26}$$

where each $y^j \in A$ and $\lambda_j \le \delta$. Relabelling $\lambda_1 \cdots \lambda_j$ by λ_j, we get $\alpha = \sum_{j=0}^{\infty} \lambda_j y^j$ with $y^j \in A$ and $\lambda_j \le \delta^j$ (we take $\lambda_0 = 1$). We compute

$$\begin{aligned}
\left\| \sum \alpha_i x_i \right\| &= \left\| \sum_{i=1}^{n} \left(\sum_{j=0}^{\infty} \lambda_j y_i^j \right) x_i \right\| && (27)\\
&= \left\| \sum_{j=0}^{\infty} \left(\lambda_j \sum_{i=1}^{n} y_i^j x_i \right) \right\| && (28)\\
&\le \sum_{j=0}^{\infty} \lambda_j \left\| \sum_{i=1}^{n} y_i^j x_i \right\| && \text{by triangle inequality} \quad (29)\\
&\le \sum_{j=0}^{\infty} \delta^j \left\| \sum_{i=1}^{n} y_i^j x_i \right\| && \text{since } \lambda_j \le \delta^j. \quad (30)
\end{aligned}$$

Now $\| \sum_{i=1}^{n} y_i^j x_i \| \le 1 + \delta$ for each j, since $y^j \in A$. Therefore summing the geometric series yields

$$\left\| \sum \alpha_i x_i \right\| \le \sum_{j=0}^{\infty} \delta^j \left\| \sum_{i=1}^{n} y_i^j x_i \right\| \le \sum_{j=0}^{\infty} \delta^j (1 + \delta) = \frac{1 + \delta}{1 - \delta}. \tag{31}$$

This gives the upper bound for $\| \sum \alpha_i x_i \|$. For the lower bound, we pick up from (28), only this time we split the series

$$\left\|\sum \alpha_i x_i\right\| = \left\|\sum_{j=0}^{\infty}\left(\lambda_j \sum_{i=1}^{n} y_i^j x_i\right)\right\| \qquad \text{from (28)} \tag{32}$$

$$= \left\|\sum_{i=1}^{n} y_i^0 x_i - (-1)\sum_{j=1}^{\infty}\left(\lambda_j \sum_{i=1}^{n} y_i^j x_i\right)\right\| \qquad \text{recall } \lambda_0 = 1 \tag{33}$$

$$\geq \left\|\sum_{i=1}^{n} y_i^0 x_i\right\| - \left\|\sum_{j=1}^{\infty}\left(\lambda_j \sum_{i=1}^{n} y_i^j x_i\right)\right\| \qquad \text{reverse triangle inequality} \tag{34}$$

$$\geq \left\|\sum_{i=1}^{n} y_i^0 x_i\right\| - \sum_{j=1}^{\infty}\delta^j \left\|\sum_{i=1}^{n} y_i^j x_i\right\| \qquad \text{since } \lambda_j \leq \delta^j \text{ and triangle inequality.} \tag{35}$$

Since $y^0 \in A$, having (24) implies $1 - \delta \leq \|\sum y_i^0 x_i\|$. Further, by the same argument used (31), we can write

$$\sum_{j=1}^{\infty}\delta^j \left\|\sum_{i=1}^{n} y_i^j x_i\right\| \leq \sum_{j=1}^{\infty}\delta^j(1+\delta) = \frac{\delta(1+\delta)}{1-\delta}. \tag{36}$$

Putting everything together, we get the upper bound for $\|\sum \alpha_i x_i\|$ to be

$$\left\|\sum \alpha_i x_i\right\| \geq 1 - \delta - \frac{\delta(1+\delta)}{1-\delta} = \frac{1-3\delta}{1-\delta}. \tag{37}$$

Combining the bounds in (31) and (37), and recalling $\alpha \in S^{n-1}$ was arbitrary, we have for all $\alpha \in S^{n-1}$,

$$\frac{1-3\delta}{1-\delta} \leq \left\|\sum \alpha_i x_i\right\| \leq \frac{1+\delta}{1-\delta}. \tag{38}$$

Now assume $\epsilon > 0$. Take δ small so that $(\sqrt{1+\epsilon})^{-1} \leq \frac{1-3\delta}{1-\delta}$ and $\frac{1+\delta}{1-\delta} \leq \sqrt{1+\epsilon}$. Then (38) becomes

$$\frac{1}{\sqrt{1+\epsilon}} \leq \left\|\sum \alpha_i x_i\right\| \leq \sqrt{1+\epsilon}. \tag{39}$$

Hence we have shown (39) holds for all $\alpha \in S^{n-1}$. This completes the first part of the proof. Now take any $\alpha \in \mathbb{R}^n$. Homogeniety of $\|\cdot\|$ implies

$$\left\|\sum \alpha_i x_i\right\| = \|\alpha\|_2 \cdot \left\|\sum \frac{\alpha_i}{\|\alpha\|_2} x_i\right\|. \tag{40}$$

Since $\left(\frac{\alpha}{\|\alpha\|_2}\right) \in \mathbb{R}^n$ is a unit vector, the result in (39) gives

$$\frac{1}{\sqrt{1+\epsilon}} \cdot \|\alpha\|_2 \leq \left\|\sum \alpha_i x_i\right\| \leq \sqrt{1+\epsilon} \cdot \|\alpha\|_2. \tag{41}$$

This completes the proof.

□

3 The Gaussian Reformulation

We now state and prove the Gaussian reformulation of Dvoretzky's theorem. An extensive introduction of this proof was given in the section (1.2). This part is therefore devoted to rigorously stating and proving the Gaussian reformulation, foregoing lengthy motivational discussion.

Definition. *Consider a collection of m independent standard Gaussian variables g_k on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a collection $\{z_k\}_{k=1}^m$ from Banach space $(E, \|\cdot\|)$. Let $X = \sum_{k=1}^m g_k z_k$. We define*

$$\sigma(X) = \sup\left\{\left(\sum \zeta(z_k)^2\right)^{\frac{1}{2}} \mid \zeta \in E^*, \quad \|\zeta\| \leq 1\right\} \quad \textit{and} \quad d(X) = \left(\frac{\mathbb{E}\|X\|}{\sigma(X)}\right)^2. \tag{42}$$

Also recall the concentration of measure phenomenon. In the setting of the definition, we have for all $\delta > 0$,

$$\mathbb{P}\Big(\big|\|X\| - \mathbb{E}\,\|X\|\big| \geq \delta\,\mathbb{E}\,\|X\|\Big) \leq 2\exp\left(-\frac{\delta^2 d(X)}{4}\right). \tag{43}$$

Theorem (Gaussian reformulation)**.** *For every $\epsilon > 0$, there exists a number $\eta_1(\epsilon)$ with the following property. Let $m \leq N$, take m independant standard Gaussian random vairables g_k, and let $\{z_k\}_{k=1}^m$ be a collection of elements of an N-dimensional Banach space $(E, \|\cdot\|)$. Let $X = \sum g_k z_k$. Then for each $n \leq [\eta_1(\epsilon)d(X)]$, the space E contains a subspace F of dimension n which is $(1+\epsilon)$-isomorphic to ℓ_2^n.*

Proof. Fix $\epsilon > 0$. Take $\delta > 0$ to be the smaller of 1 and $\delta(\epsilon)$ obtained from Lemma 2. We claim $\eta_1(\epsilon) = \frac{\delta^2}{8\log\left(1+\frac{2}{\delta}\right)}$. Take any $n \leq [\eta_1(\epsilon)d(X)]$.

Let A be the δ-net of S^{n-1} of cardinality at most $\left(1+\frac{2}{\delta}\right)^n$, as given by Lemma 1. Set $\{g_{i,k}\}_{i=1}^{n}{}_{k=1}^{m}$ to be a collection of independant standard Gaussians, and defeine $X_i = \sum_{k=1}^m g_{i,k} z_k$. Then, for any $\alpha \in A$, both $\sum_{i=1}^n \alpha_i X_i$ and X have the same distribution, as proven in section (1.2). Therefore, denoting $\mathbb{E}\,\|X\| = M$, our concentration inequality allows

$$\mathbb{P}\Big(\big|\|\sum \alpha_i X_i\| - M\big| \geq \delta M\Big) \leq 2\exp\left(-\frac{\delta^2 d(X)}{4}\right) \quad \text{and therefore} \tag{44}$$

$$\mathbb{P}\Big(\exists \alpha \in A \text{ s.t. } \big|\|\sum \alpha_i X_i\| M^{-1} - 1\big| \geq \delta\Big) \leq 2\exp\left(n\log\left(1+\frac{2}{\delta}\right) - \frac{\delta^2 d(X)}{4}\right). \tag{45}$$

The second inequality follows from the bound on $|A|$ and finite subadditivity of $\mathbb{P}$. Since the exponential is an increasing function, we can replace n in the right side of (45) with the larger $\epsilon_1(\epsilon)d(X)$. Then we can recall the definition of $\epsilon_1(\epsilon)$ to conclude the right side of (45) is bounded above by

$$2\exp\left(\eta_1(\epsilon)d(X)\log\left(1+\frac{2}{\delta}\right) - \frac{\delta^2 d(X)}{4}\right) = 2\exp\left(-\frac{\delta^2 d(X)}{8}\right). \tag{46}$$

We now consider two cases.

- If the quantity in (46) is larger than or equal to 1, we must have $d(X) \leq -\frac{8 \log \frac{1}{2}}{\delta^2}$. Rearranging and then using $\delta \leq 1$, we see

$$\eta_1(\epsilon) d(X) \leq -\frac{\log \frac{1}{2}}{\log\left(1 + \frac{2}{\delta}\right)} < 1. \tag{47}$$

  This means that n can only taken to be 0. Since any 0 dimensional subspace F of E is $(1+\epsilon)$-isomorphic to ℓ_2^0, the theorem holds.

- Otherwise, if the quantity in (46) is less than 1, we can use the fact this quantity bounds the probability in (45) to see

$$\mathbb{P}\Big(\exists \alpha \in A \text{ s.t. } \big| \| \sum \alpha_i X_i \| M^{-1} - 1 \big| \geq \delta \Big) < 1. \tag{48}$$

  In other words, there exists some $\omega \in \Omega$ such that for all $\alpha \in A$ (denoting $Y_i = \frac{X_i}{M}$),

$$1 - \delta \leq \| \sum \alpha_i Y_i(\omega) \| \leq 1 + \delta. \tag{49}$$

  But then the vectors $Y_i(\omega)$, the δ-net A, and the choice of δ satisfy the hypotheses of Lemma 2. So by the lemma, $F = \operatorname{span}\{Y_1(\omega), \ldots, Y_n(\omega)\}$ is a subspace of E that is $(1+\epsilon)$-isomorphic to ℓ_2^n. Thus the theorem holds.

Therefore, we have shown that taking $\eta_1(\epsilon) = \frac{\delta^2}{8 \log(1+\frac{2}{\delta})}$, where $\delta = \delta(\epsilon)$ is taken from Lemma 2, satisfies the conclusion of Dvoretzky's theorem. This completes the proof.

□

In appendix (A.3), we show that for $\epsilon \leq \frac{1}{9}$, we can take $\delta(\epsilon) = \frac{\epsilon}{9}$. Hence for sufficiently small ϵ, the above result gives $\eta_1(\epsilon) = \frac{\epsilon^2}{648\log(1+\frac{18}{\epsilon})}$, which is of the order $\frac{\epsilon^2}{\log(1+\frac{1}{\epsilon})}$.

4 The Extended Dvoretzky-Rogers Lemma

Using the notation from section (3), our goal, as outlined in (1.3), is to find X such that $d(X)$ is bounded below by $C \log N$ for some universal constant C. Once this bound is established, Dvoretzky's theorem follows from its Gaussian reformulation. Recall X is constructed from our choice of $\{z_k\}_{k=1}^m$ from E. The appropriate z_k and m will come from an extension of a classical result known as the Dvoretzky-Rogers Lemma. The proof of the lemma relies on Lewis' theorem and Lemma 3, which we prove first. Lewis' theorem in particular is of singular interest. The extension of the lemma is presented after, and relies on a proposition due to M. Talagrand.

4.1 Lewis' Theorem

Here we state and prove Lewis' theorem, a result shown by D. Lewis in 1979. It is a generalization of a theorem of F. John concerning ellipsoids of maximal volume in CSC bodies. First, note Lewis' theorem deals with an arbitrary norm α on $\mathcal{L}(\mathbb{R}^n, E)$, the space of linear maps $\mathbb{R}^n \to E$. For our later purposes we will take α to be the usual operator norm. Regardless, we will reproduce Lewis' result in full generality.

Definition. *Let U and V be vector spaces, and α be a norm on $\mathcal{L}(U, V)$. Then the dual norm α^* is a norm on $\mathcal{L}(V, U)$ defined by*

$$\alpha^*(v) = \sup\{tr(vT) \mid T : U \to V, \alpha(T) \leq 1\}. \tag{50}$$

Theorem (Lewis' Theorem)**.** *Let E be an n-dimensional vector space and let α be a norm on $\mathcal{L}(\mathbb{R}^n, E)$. There exists an isomorphism $u : \mathbb{R}^n \to E$ such that*

$$\alpha(u) = 1 \ \textit{and}\ \alpha^*(u^{-1}) = n. \tag{51}$$

We interpret this theorem. An ellipsoid in a Banach space $(E, \|\cdot\|)$ is the image of the unit ball B^n of ℓ_2^n under an isomorphism $u : \mathbb{R}^n \to E$; thus we identify ellipsoids with isomorphisms. Note this identification is not necessarily unique, but its existance is sufficient for our purposes. If this isomorphism u has operator norm 1, then the ellipsoid $u(B^n)$ is contained inside B_E, the unit ball of E. If we give E some basis, then we can further say $\mathrm{vol}(u(B^n)) = c|\det u|$ for some constant c. Hence, maximizing det over the unit ball in $\mathcal{L}(\mathbb{R}^n, E)$ (equipped with the operator norm) gives an ellipsoid in B_E of maximum volume. F. John proved in 1948 that this ellipsoid is unique up to an orthogonal transformation. Hence, ellipsoids of this type are called John ellipsoids.

Lewis' theorem is an enumeration of some properties of the John ellipsoid of B_E, should we take α to be the operator norm. In particular, the u in Lewis' theorem is the John ellipsoid, and (51) helps prove the uniqueness of u should we desire to do so. The proof of Lewis' theorem is an exercise in linear algebra, and begins by maximizing det over the unit ball of $\mathcal{L}(\mathbb{R}^n, E)$ equipped with an arbitrary norm.

Proof. Let $K = \{T : \mathbb{R}^n \to E \mid \alpha(T) \le 1\}$, the unit ball in $\mathcal{L}(\mathbb{R}^n, E)$ equipped with the norm α. Since $\mathcal{L}(\mathbb{R}^n, E)$ is finite dimensional, K is compact. Consider the determinant map $T \mapsto \det(T)$ for some given basis of E. Since det is continuous and K is compact, det attains its maximum over K. Say this occurs at $u \in K$. We must have $\det(u) \neq 0$, so u is invertible. We show u satisfies (51) by bounding $tr(u^{-1}T)$ for any $T : \mathbb{R}^n \to E$.

Take T, and note $\frac{u+T}{\alpha(u+T)}$ is in K. Hence

$$\left|\det\left(\frac{u+T}{\alpha(u+T)}\right)\right| \leq |\det(u)| \qquad \text{by choice of } u \tag{52}$$

$$|\det(u+T)| \leq \alpha(u+T)^n |\det(u)| \qquad \text{by homogeniety} \tag{53}$$

$$\left|\frac{\det(u+T)}{\det(u)}\right| \leq \alpha(u+T)^n \qquad \text{since } \det(u) \neq 0 \tag{54}$$

$$|\det(1+u^{-1}T)| \leq \alpha(u+T)^n \qquad \text{by multiplicativity.} \tag{55}$$

The triangle inequality for α implies $\alpha(u+T)^n \leq (\alpha(u)+\alpha(T))^n$, which is at most $(1+\alpha(T))^n$ since u is in K. So inequality (55) gives $|\det(1+\epsilon u^{-1}T)| \leq (1+\epsilon\alpha(T))^n$, where $\epsilon > 0$ can be added since T is arbitrary. For small ϵ, we have $\det(1+\epsilon u^{-1}T) = 1 + \epsilon tr(u^{-1}T) + o(\epsilon)$, and thus for all $T : \mathbb{R}^n \to E$

$$1 + \epsilon tr(u^{-1}T) + o(\epsilon) \leq (1+\epsilon\alpha(T))^n \tag{56}$$

$$tr(u^{-1}T) + \frac{o(\epsilon)}{\epsilon} \leq \frac{(1+\epsilon\alpha(T))^n - 1}{\epsilon} \tag{57}$$

$$tr(u^{-1}T) \leq n\alpha(T) \qquad \text{letting } \epsilon \to 0. \tag{58}$$

Inequality (58) holds in particular for $T = u$, giving $n = tr(u^{-1}u) \leq n\alpha(u)$ and thus $1 \leq \alpha(u)$. As $u \in K$ we get $1 = \alpha(u)$. Now (58) further gives $tr(u^{-1}T) \leq n$ for all T, so $\alpha^*(u^{-1}) \leq n$. But $n = tr(u^{-1}u) \leq \alpha^*(u^{-1})$ is clear, and we conclude $\alpha^*(u^{-1}) = n$. Thus u satisfies (51), and the proof is complete.

□

4.2 Lemma 3

This lemma is a technical exercise in linear algebra, and of marginal interest in its own right. Although there is a version of the lemma that holds in infinite dimensional spaces, given in appendix (A.2), we only use the version for finite dimensional spaces.

Definition. *Let $T : U \to V$ be a linear map between Banach spaces. Define $K_n = \{\|T - S\| \mid S : U \to V, rank(S) < n\}$, where $\|\cdot\|$ is the usual operator norm. Then define $a_n(T) = \inf K_n$. Note here that $K_1 = \{\|T\|\}$, and thus $a_1(T) = \|T\|$, since the only $S : U \to V$ of rank less than 1 is the zero map.*

Lemma 3. *Let $T : H \to X$ be a linear operator from a Hilbert space $(H, |\cdot|)$ of dimension N to a Banach space $(X, \|\cdot\|)$. There exists an orthonormal sequence $\{f_n\}_{n=1}^N$ in H such that $\|Tf_n\| \geq a_n(T) - \epsilon$ for any n.*

Proof. We construct the sequence $\{f_n\}_{n=1}^N$ recursively. By definition $a_1(T) = \|T\| = \sup\{\|Tv\| \mid |v| = 1\}$. Now, $\|\cdot\|$ is continuous, and since H is finite dimensional we have that T is continuous and the set $\{v \in H \mid |v| = 1\}$ is compact. Therefore, $\{\|Tv\| \mid |v| = 1\}$ is compact, and hence there exists some $f_1 \in H$ with $|f_1| = 1$ such that $a_1(T) = \|Tf_1\|$. Take f_1 to be the first vector in our sequence.

To construct the next vector f_2, let $S_1 = (\text{span}\{f_1\})^\perp$. Then define $S : H \to X$ by

$$Sv = \begin{cases} Tv & \text{if } v \in \text{span}\{f_1\} \\ 0 & \text{if } v \in S_1. \end{cases} \tag{59}$$

Clearly S is linear. We also have $\text{rank}(S) < 2$ since S is potentially nonzero only on a subspace of dimension 1. Finally, $(T - S)v = T|_{S_1}v$ for all $v \in S_1$, and is 0 otherwise. It follows that $\|T|_{S_1}\| = \|T - S\|$, and hence $\|T|_{S_1}\|$ is in K_2. By definition $a_2(T)$ we therefore get

$$\|T|_{S_1}\| \geq a_2(T). \tag{60}$$

But, by similar continuity and compactness arguments used above, there must exists some $f_2 \in S_1$ such that $\|T|_{S_1}f_2\| = \|T|_{S_1}\|$ and $|f_2| = 1$. Hence by inequality (60) and the definition of restriction of a map we conclude $\|Tf_1\| \geq a_2(T)$, as desired. Take f_2 to be the second vector in our sequence. Repeating this procedure generates an orthonormal list of vectors $\{f_n\}_{n=1}^N$ in H with the desired properties, proving the lemma. $\square$

4.3 The Dvoretzky-Rogers Lemma

We have now have the tools to prove the Dvoretzky-Rogers Lemma. We begin, however, with motivation for the proof, as well as a description regarding the contributions of Lewis' theorem and Lemma 3.

Lemma (Dvoretzky-Rogers Lemma)**.** *Let $(E, \|\cdot\|)$ be an N-dimensional Banach space. Let $\bar{N} = \left[\frac{N}{2}\right]$. Then there are $\bar{N}$ linearly independent elements z_k in E such that for all $\alpha \in \mathbb{R}^{\bar{N}}$*

$$\left\| \sum_{K=1}^{\bar{N}} \alpha_k z_k \right\| \leq \left(\sum_{k=1}^{\bar{N}} |\alpha_k|^2 \right)^{\frac{1}{2}}, \tag{61}$$

and $\|z_k\| \geq \frac{1}{2}$ for all k.

First we discuss the geometry of the lemma. Consider the z_k from the lemma. Denote $F = \text{span}\{z_1, \ldots, z_{\bar{N}}\}$ and let T be the usual isomorphism $\mathbb{R}^{\bar{N}} \to F$ defined by $e_k \mapsto z_k$. Then, the first part of the lemma states that $\|T\alpha\| \leq \|\alpha\|_2$. In other words, $T(B^{\bar{N}}) \subseteq B_F$. The second part simply means that T does not shrink the basis vectors e_k too much. Therefore the Dvoretzky-Rogers Lemma states there exists an $\bar{N}$-dimensional subspace F of E such that B_F contains the image of $B^{\bar{N}}$ in such a way so that the images of the basis vectors e_k remain relatively large. Of course, if we can find such an F (in lieu of the z_k), we would expect some basis of F to satisfy the lemma.

Now we motivate the proof. We need to find an $\bar{N}$-dimensional subspace F of E where the usual isomorphism $T : \mathbb{R}^{\bar{N}} \to F$ shrinks the unit ball $B^{\bar{N}}$ but still keeps the basis vectors e_k as big as possible. To find F, we can instead find a map $T' : \mathbb{R}^N \to E$ that, when restricted to an isomorphism from $\mathbb{R}^{\bar{N}}$, behaves like the desired T. We may then take $F = \text{span}\{T'e_1, \ldots, T'e_{\bar{N}}\}$. A good candidate for T' is some John ellipsoid v of B_E, since v shrinks B^N into B_E in a way that maximizes its volume, thereby making it likely v keeps the e_k large. So we search for John ellipsoids v of B_E such that we can take $F = \text{span}\{ve_1, \dots, ve_{\bar{N}}\}$. In other words, we want v where taking $z_k = v_k e_k$ satisfies both (61) and $\|z_k\| \geq \frac{1}{2}$ for $1 \leq k \leq \bar{N}$.

The desired John ellipsoid v is in fact be given by $e_k \mapsto uf_k = z_k$, where $u : \ell_2^N \to E$ is the John ellipsoid in Lewis' theorem (where α is the operator norm) and the f_k are from Lemma 3. Below, we verify that this choice of z_k indeed satisfies the lemma.

Proof. Take α to be the operator norm on $\mathcal{L}(\ell_2^N, E)$. By Lewis' theorem, there exists a John ellipsoid $u : \ell_2^N \to E$ such that $\alpha(u) = 1$ and $\alpha^*(u^{-1}) = N$. Then by Lemma 3, noting ℓ_2^N is finite dimensional, we can find an orthonormal collection of vectors $\{f_k\}_{k=1}^N$ in ℓ_2^N such that $\|uf_k\| \geq a_k(u)$ for all $k \leq N$. We now show that $a_k(u) \geq 1 - \frac{k}{N}$.

Fix k and let P be an orthogonal projection on ℓ_2^N with rank less than k. Then

$$
\begin{aligned}
N - k \leq \text{rank}(I - P) &= tr(I - P) && \text{since } P \text{ is idempotent} && (62)\\
&= tr(u^{-1}u(I - P)) && && (63)\\
&\leq \alpha^*(u^{-1})\|u(I - P)\| && \text{by definition of } \alpha^* && (64)\\
&= N\|u - uP\| && \text{since } \alpha^*(u^{-1}) = N. && (65)
\end{aligned}
$$

Rearranging gives $\|u - uP\| \geq 1 - \frac{k}{N}$. As $\text{rank}\, uP < k$ by choice of P, we then have by definition of $a_k(u)$ that $a_k(u) \geq \|u - uP\| \geq 1 - \frac{k}{N}$, as we set out to show. So then, by choice of f_k, we have $\|uf_k\| \geq a_k(u) \geq 1 - \frac{k}{N}$, and in particular $\|uf_k\| \geq \frac{1}{2}$ for $k \leq \bar{N}$.

Thus, taking $z_k = uf_k$ satisfies the lemma, and the proof is complete. □

4.4 Talagrand's Proposition

Comparing the Dvoretzky-Rogers Lemma we just proved with its extension used in the introduction, we observe some similarities. Namely, for α in Euclidean space, both give $\|\alpha\|_2$

as an upper bound to $\|\alpha_k z_k\|$. However, whereas the Dvoretzky-Rogers Lemma gives $\left[\frac{N}{2}\right]$ vectors z_k of norm larger than $\frac{1}{2}$, its extension gives $\left[\frac{\sqrt{\left[\frac{N}{2}\right]}}{16}\right]$ vectors z_k such that half the supremum norm of α was less than $\|\sum \alpha_k z_k\|$. With this comparison in mind, we will show that the vectors in the extension are simply a subset of the $\left[\frac{N}{2}\right]$ vectors z_k from the lemma. That such a subset exists is due to a result of M. Talagrand, named Proposition 2 in his 1995 paper "Embedding of ℓ_2^∞ and a Theorem of Alon and Milman". We adapt his proof here.

Lemma (Talagrand's Proposition)**.** *Suppose $(E, \|\cdot\|)$ is a Banach space and $\{z_i\}_{i=1}^m$ is a collection of vectors in E such that $\|z_i\| = 1$ for $i = 1, \dots, m$. Let*

$$Z_m = \sup\{\sum_{i=1}^m |\zeta(z_i)| \mid \zeta \in E^*, \quad \|\zeta\| \le 1\}. \tag{66}$$

There is a subset A of $\{1, \dots, m\}$ with $|A| \ge \frac{m}{8Z_m}$ such that for all collections of real numbers $\{\alpha_i\}_{i\in A}$, we have

$$\|\sum_{i\in A} \alpha_i z_i\| \ge \frac{1}{2} \sup_{i\in A} |\alpha_i|. \tag{67}$$

Proof. Let $\delta = \frac{1}{4Z_m} < 1$. Let $\{\delta_i\}_{i=1}^m$ be a collection of independent Bernouilli random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\delta_i \sim \text{Bernouilli}(\delta)$. For any $1 \le i \le m$, fix $\zeta_i \in E^*$ such that $\zeta_i(z_i) = \|\zeta_i\| = 1$. For convenience, denote $\zeta_i(z_j)$ by z_{ij}. By choice of ζ_i, we see

$$\sum_{j=1}^m |z_{ij}| = \sum_{j=1}^m |\zeta_i(z_j)| \in \{\sum_{j=1}^m |\zeta(x_j)| \mid \zeta \in E^*, \|\zeta\| \le 1\}. \tag{68}$$

Therefore $\sum_{j=1}^m |z_{ij}| \le Z_m$. Let $K = \{(i,j) \mid 1 \le i \ne j \le m\}$. Now we can compute

$$\mathbb{E}\left(\sum_{(i,j)\in K}\delta_i\delta_j|z_{ij}|\right) = \sum_{(i,j)\in K}(\mathbb{E}\,\delta_i\,\mathbb{E}\,\delta_j)\,|z_{ij}| \quad \text{as } \mathbb{E} \text{ is linear and } \delta_i, \delta_j \text{ are independant} \tag{69}$$

$$= \delta^2 \sum_{(i,j)\in K}|z_{ij}| \quad \text{since } \mathbb{E}\,\delta_i = \delta \tag{70}$$

$$\leq \delta^2 m Z_m \quad \text{since each } \sum_{\substack{j\leq m\\ i\neq j}}|z_{ij}| \leq \sum_{j=1}^{m}|z_{ij}| \leq Z_m \tag{71}$$

$$= \frac{m\delta}{4} \quad \text{by definition of } \delta. \tag{72}$$

Using the linearity of the expectation, followd by the above inequality, then allows us to write

$$\mathbb{E}\left(\sum_{i=1}^{m}\delta_i - 2\sum_{(i,j)\in K}\delta_i\delta_j|z_{ij}|\right) \geq m\delta - 2\cdot\frac{m\delta}{4} = \frac{m\delta}{2}. \tag{73}$$

Therefore there exists $\omega \in \Omega$ such that

$$\sum_{i=1}^{m}\delta_i(\omega) - 2\sum_{(i,j)\in K}\delta_i(\omega)\delta_j(\omega)|x_{ij}| \geq \frac{m\delta}{2}. \tag{74}$$

Let $I = \{1 \leq i \leq m \mid \delta_i(\omega) = 1\}$. Of course, for all $i \notin I$, we necessarily have $\delta_i(\omega) = 0$. Then we have

$$\sum_{i=1}^{m}\delta_i(\omega) = |I| \text{ and} \tag{75}$$

$$2\sum_{(i,j)\in K}\delta_i(\omega)\delta_j(\omega)|z_{ij}| = 2\sum_{\substack{i,j\in I\\ i\neq j}}|z_{ij}|. \tag{76}$$

This means that (74) becomes

$$|I| - 2\sum_{\substack{i,j\in I\\ i\neq j}} |z_{ij}| \geq \frac{m\delta}{2} \quad \text{i.e.} \quad \sum_{i\in I}\left(1 - 2\sum_{j\in I\setminus\{i\}} |z_{ij}|\right) \geq \frac{m\delta}{2}. \tag{77}$$

Now for each $i \in I$, we have $1 - 2\sum_{j\in I\setminus\{i\}} |z_{ij}| \leq 1$. Since the sum of these $|I|$ terms is at least $\frac{m\delta}{2}$, it then follows that at least $\frac{m\delta}{2}$ of these terms are positive. In other words, if we take A to be the collection of these positive terms, namely

$$A = \{i \in I \mid 0 \leq 1 - 2\sum_{j\in I\setminus\{i\}} |x_{ij}| \quad \text{i.e.} \quad \sum_{j\in I\setminus\{i\}} |x_{ij}| \leq \frac{1}{2}\}, \tag{78}$$

then $|A| \geq \frac{m\delta}{2}$. We show A is the subset of $\{1, \ldots, m\}$ that satisfies the theorem. Fix a collection of real numbers $\{\alpha_i\}_{i\in A}$ and take $l \in A$ such that $|\alpha_l| = \sup_{i\in A} |\alpha_i|$. Then

$$\begin{aligned}
\|\sum_{i\in A} \alpha_i z_i\| &\geq \left|\zeta_l\left(\sum_{i\in A} \alpha_i z_i\right)\right| && \text{since } \|\zeta_l\| \leq 1 && (79)\\
&= \left|\alpha_l \zeta_l(z_l) + \sum_{i\in A\setminus\{l\}} \alpha_i \zeta_l(z_i)\right| && && (80)\\
&\geq |\alpha_l \zeta_l(z_l)| - \sum_{i\in A\setminus\{l\}} |\alpha_i||\zeta_l(z_i)| && \text{by the reverse and regular triangle inequality} && (81)\\
&\geq |\alpha_l|\left(|\zeta_l(z_l)| - \sum_{i\in A\setminus\{l\}} |\zeta_l(z_i)|\right) && \text{replacing each } |\alpha_i| \text{ with the larger } |\alpha_l| && (82)\\
&= |\alpha_l|\left(1 - \sum_{i\in A\setminus\{l\}} |z_{li}|\right) && \text{since } |\zeta_l(z_l)| = 1 && (83)\\
&\geq |\alpha_l|\frac{1}{2} && \text{since } \sum_{i\in A\setminus\{l\}} |z_{li}| \leq \sum_{i\in I\setminus\{l\}} |z_{li}| \leq \frac{1}{2} \text{ as } l \in A && (84)\\
&= \frac{1}{2}\sup_{i\in A} |\alpha_i| && \text{by choice of } \alpha_l. && (85)
\end{aligned}$$

Since the collection $\{\alpha_i\}_{i\in A}$ was arbitrary, A indeed satisfies the theorem. $\square$

4.5 The Extended Dvoretzky-Rogers Lemma

With the Dvoretzky-Rogers Lemma and Talagrand's proposition at our disposal, we may now prove the Extended Dvoretzky-Rogers Lemma as used in the introduction.

Lemma (The Extended Dvoretzky-Rogers Lemma). *Suppose $(E,\|\cdot\|)$ is an N-dimensional Banach space. Then there exists a collection $\{z_k\}$ of $\bar{N} = \left[\frac{\sqrt{[\frac{N}{2}]}}{16}\right]$ vectors in E such that for all $\alpha \in \mathbb{R}^{\bar{N}}$,*

$$\frac{1}{2}\sup|\alpha_k| \le \left\|\sum \alpha_k z_k\right\| \le 2\left(\sum|\alpha_k|^2\right)^{\frac{1}{2}}. \tag{86}$$

We will get a preliminary collection of $\left[\frac{N}{2}\right]$ vectors from the Dvoretzky-Rogers Lemma, then take a subset of this collection using Talagrand's Proposition, and then discard a few more vectors so the size of the remaining collection is as desired.

Proof. First, suppose we had a collection $\{z_k\}$ that satisfied either inequality in (86). Then any subset must satisfy the same inequality (with the number of terms approppariately adjusted). To see this, simply apply the inequality in question to $\{z_k\}$ in the case where $\alpha_k = 0$ whenever z_k is not in the subset in question.

Now take the $m = \left[\frac{N}{2}\right]$ vectors $\{z'_k\}$ given by the Dvoretzky-Rogers Lemma. Let $z_k = 2z'_k$. Then $\|\sum \alpha_k z_k\| \le 2\|\alpha\|_2$ for all $\alpha \in \mathbb{R}^m$, and each $\|z_k\| \ge 1$. Then Talagrand's proposition applied to $\{z_k\}$ gives a collection of $l \geq= \frac{m}{8Z_m}$ vectors $\{z_{k_j}\}$ from $\{z_k\}$ such that $\frac{1}{2}\sup|\alpha_{k_j}| \le \left\|\sum \alpha_{k_j} z_{k_j}\right\|$ for all $\alpha \in \mathbb{R}^l$.

Recall $Z_m = \sup\{\sum_{k=1}^m |\zeta(z_k)| \mid \zeta \in E^*, \|\zeta\| \le 1\}$. Then, for each $\zeta \in E^*$, an elementary property of norms on Euclidean spaces gives

$$\sum|\zeta(z_k)| \le \sqrt{m}\left(\sum|\zeta(z_k)|^2\right)^{\frac{1}{2}}. \tag{87}$$

Therefore $Z_m \leq \sqrt{m} \cdot \sup\{(\sum |\zeta(z_k)|^2)^{\frac{1}{2}} \mid \zeta \in E^*, \|\zeta\| \leq 1\}$. But since $\| \sum \alpha_{k_j} z_{k_j} \| \leq 2\|\alpha\|_2$ for all $\alpha \in \mathbb{R}^l$ (remember $\{z_{k_j}\}$ is a subset of $\{z_k\}$, which satisfies the right side of (86)), we can conclude by the calculation carried out in section (1.3) around (17) that $Z_m \leq \sqrt{m} \cdot 2$. Now we have

$$l \geq \left[\frac{m}{8Z_m}\right] \geq \left[\frac{m}{16\sqrt{m}}\right] = \left[\frac{\sqrt{m}}{16}\right] = \bar{N}. \tag{88}$$

Consider the collection $\{z_{k_j}\}_{j=1}^{\bar{N}}$. Since it is a subset of $\{z_{k_j}\}$, which satisfies (86), we conclude that $\{z_{k_j}\}_{j=1}^{\bar{N}}$ also satisfies (86). Noting that $\bar{N} = \left[\frac{\sqrt{[\frac{N}{2}]}}{16}\right]$ completes the proof. □

5 Dvoretzky's Theorem

With both the Gaussian reformulation of Dvoretzky's theorem and the Extended Dvoretzky-Rogers Lemma proven, we can now tackle the version of Dvoretzky's theorem stated in the introduction. However, recalling the end of section (1.3), there is one more technical lemma we need before we can proceed. This is Lemma 4, which is proven first.

5.1 Lemma 4

Lemma 4. *Let $\{g_k\}_{k=1}^{N}$ be independant standard Gaussian variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\bar{N} = \left[\frac{\sqrt{[\frac{N}{2}]}}{16}\right]$. Then there exists a constant c such that*

$$c\sqrt{\log N} \leq \mathbb{E} \sup_{k \leq \bar{N}} |g_k|. \tag{89}$$

Proof. We first establish a preliminary result. Fix $n > 1$. Then we claim

$$\mathbb{P}\left(|g_1| > \sqrt{\log n}\right) \geq \frac{1}{n}. \tag{90}$$

By definition of a standard Gaussian variable, for $\lambda > 1$ we have $\mathbb{P}(|g| > \lambda) = 2\int_\lambda^\infty e^{-\frac{t^2}{2}}dt$. Then we may calculate

$$
\begin{aligned}
2\int_\lambda^\infty e^{-\frac{t^2}{2}}dt &\geq \int_\lambda^\infty e^{-\frac{t^2}{2}}(1+t^{-2})dt && \text{since } t^{-2} < 1 && (91)\\
&= -e^{-\frac{t^2}{2}}t^{-1}\Big|_\lambda^\infty && && (92)\\
&= \frac{e^{-\frac{\lambda^2}{2}}}{\lambda}. && && (93)
\end{aligned}
$$

Letting $\lambda = \sqrt{\log n}$ we can conclude by substitution that

$$
\mathbb{P}\left(|g| > \sqrt{\log n}\right) \geq \frac{e^{-\frac{\log n}{2}}}{\sqrt{\log n}} = \left(\frac{1}{n\log n}\right)^{\frac{1}{2}} \geq \left(\frac{1}{n^2}\right)^{\frac{1}{2}} = \frac{1}{n} \tag{94}
$$

The last inequality used the fact $n \geq \log n$. Therefore $\mathbb{P}\left(|g| > \sqrt{\log n}\right) \geq \frac{1}{n}$ for $n > 1$. Direct computation gives the result for $n = 1$, thus establishing this preliminary result.

Now then, letting $n = \sqrt{N}$, we have, by the above result

$$
\mathbb{P}\left(|g_1| > \sqrt{\log\sqrt{N}}\right) \geq \frac{1}{\sqrt{N}}. \tag{95}
$$

Then we can compute

$$
\begin{aligned}
\mathbb{P}\left(\sup_{k\leq \bar{N}}|g_k| \leq \sqrt{\log\sqrt{N}}\right) &= \mathbb{P}\left(|g_1| \leq \sqrt{\log\sqrt{N}}\right)^{\bar{N}} && && (96)\\
&\leq \left(1 - \frac{1}{\sqrt{N}}\right)^{\bar{N}} && \text{by (95)} && (97)\\
&\leq e^{-\frac{1}{16\sqrt{2}}} && && (98)
\end{aligned}
$$

Rearranging gives $1 - e^{-\frac{1}{16\sqrt{2}}} \le \mathbb{P}\left(\sup_{k \le \bar{N}} |g_k| \ge \sqrt{\log \sqrt{N}}\right)$. Then by Markov's inequality,

$$1 - e^{-\frac{1}{16\sqrt{2}}} \le \frac{\mathbb{E}\sup_{k \le \bar{N}} |g_k|}{\sqrt{\log \sqrt{N}}} \tag{99}$$

$$\left(1 - e^{-\frac{1}{16\sqrt{2}}}\right) \cdot \frac{1}{\sqrt{2}} \sqrt{\log N} \le \mathbb{E} \sup_{k \le \bar{N}} |g_k|. \tag{100}$$

We have therefore proven the lemma, with $c = \left(1 - e^{-\frac{1}{16\sqrt{2}}}\right) \cdot \frac{1}{\sqrt{2}}$. $\square$

5.2 Dvoretzky's Theorem

Theorem. *For each $\epsilon > 0$ there is a number $\eta(\epsilon) > 0$ with the following property. Let $(E, \|\cdot\|)$ be an N-dimensional Banach space. Then E contains a subspace F of dimension $n = [\eta(\epsilon) \log N]$ such that $d(F, \ell_2^n) \le 1 + \epsilon$.*

Proof. Apply the Extended Dvoretzky-Rogers Lemma to E to get $\bar{N} = \left[\frac{\sqrt{[\frac{N}{2}]}}{16}\right]$ vectors $\{z_k\}$ in E such that for all $\alpha \in \mathbb{R}^{\bar{N}}$,

$$\frac{1}{2} \sup |\alpha_k| \le \| \sum \alpha_k z_k \| \le 2\|\alpha\|_2. \tag{101}$$

Let $\{g_k\}_{k=1}^{\bar{N}}$ be independent standard Gaussian variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Set $X = \sum g_k z_k$. We claim $C \log N \le d(X)$ for some constant C.

Note that the argument in end of section (1.3) and Lemma 4 together imply that for $c = \left(1 - e^{-\frac{1}{16\sqrt{2}}}\right) \cdot \frac{1}{\sqrt{2}}$,

$$\frac{c}{2}\sqrt{\log N} \le \frac{1}{2} \mathbb{E} \sup |g_k| \le \mathbb{E}\,\|X\|. \tag{102}$$

Then since $\sigma(X) \le 2$, as by section 1.3 around equation (17), we have

$$\frac{c^2}{4}\log N \cdot \frac{1}{4} \le \left(\frac{\mathbb{E}\|X\|}{2}\right)^2 \le d(X). \tag{103}$$

Letting $C = \frac{c^2}{16}$ gives the desired inequality.

Now take η_1 from the Gaussian reformulation of Dvoretzk's theorem. Denoting $\eta(\epsilon) = C\eta_1(\epsilon)$, we have $[\eta(\epsilon)\log N] \le [\eta_1(\epsilon)d(X)]$. By the Gaussian reformulation of the Dvoretzky theorem, it follows that there exists a subspace of E of dimension $n = [\eta(\epsilon)\log N]$ that is $(1+\epsilon)$-isomorphic to ℓ_2^n. This completes the proof. $\square$

Notice that since $\eta = C\eta_1$, the function η is of the order $\frac{\epsilon^2}{\log(\frac{1}{\epsilon})}$ for $\epsilon \le \frac{1}{9}$.

6 Alternate Proof With Min-Max Theorem

Now we study Gordon's version and proof of Dvoretzky's theorem. As mentioned in the introduction, Gordon's result gives $\eta \sim \epsilon^2$, which is considered an improvement on the η given in our first version of Dvoretzky's theorem. As many proofs are omitted here, we will state the various prerequisite results and then describe how they prove Dvoretzky's theorem without much motivation.

6.1 Gaussian Min-Max Theorem

The following theorem is proved in section 9 of the online notes, and is also stated in Gordon's paper. Here we forgo the proof.

Theorem (Min-Max Theorem)**.** *Let $\{X_{i,j}\}_{i=1}^{n}{}_{j=1}^{m}$ and $\{Y_{i,j}\}_{i=1}^{n}{}_{j=1}^{m}$ be two collections of Gaussian vectors. Suppose that both*

$$\mathbb{E}(X_{i,j} - X_{i,l})^2 \geq \mathbb{E}(Y_{i,j} - Y_{i,l})^2 \qquad \textit{for all } i,j,l \tag{104}$$

$$\mathbb{E}(X_{i,j} - X_{k,l})^2 \leq \mathbb{E}(Y_{i,j} - Y_{k,l})^2 \qquad \textit{for all } i,j,k,l \textit{ with } i \neq k. \tag{105}$$

Then $\mathbb{E}\min_i \max_j X_{i,j} \geq \mathbb{E}\min_i \max_j Y_{i,j}$.

6.2 Gordon's Theorem 2

The following theorem and its proof are found in Gordon's paper as a consequence of the Gaussian Min-Max theorem. First, we establish some notation - although some of the following objects are also used earlier in this paper, we remain closer to Gordon's notation in this section. Let m and N be natural numbers. Let T be a subset of S^{m-1} and take $\{x_i\}_{i=1}^N$ to be a collection of vectors in a normed space $(X, \|\cdot\|)$. Let $\{g_i\}_{i=1}^N$ and $\{h_j\}_{j=1}^m$ be independant standard Gaussian variables, and set

$$\mathbb{E}(\{x_i\}) = \mathbb{E}\left\|\sum_{i=1}^N g_i x_i\right\| \quad \mathbb{E}^*(T) = \mathbb{E}\max_{t\in T}\left\{\sum_{j=1}^m h_j t_j\right\}. \tag{106}$$

Theorem (Gordon's Theorem 2)**.** *Let* $0 < \epsilon < 1$*. Assume* $\left\|\sum_{i=1}^N \tau_i x_i\right\| \leq \|\tau\|_2$ *for all* $\tau \in \mathbb{R}^N$*. If* $\mathbb{E}^*(T) < \mathbb{E}(\{x_i\})$*, then there is a linear map* $A : \mathbb{R}^m \to X$ *such that*

$$\frac{\max_{t\in T}\|A(t)\|}{\min_{t\in T}\|A(t)\|} \leq \frac{\mathbb{E}(\{x_i\}) + \mathbb{E}^*(T)}{\mathbb{E}(\{x_i\}) - \mathbb{E}^*(T)}. \tag{107}$$

If we further have $\mathbb{E}^*(T) < \epsilon\,\mathbb{E}(\{x_i\})$*, then*

$$\frac{\max_{t\in T}\|A(t)\|}{\min_{t\in T}\|A(t)\|} \le \frac{1+\epsilon}{1-\epsilon}. \tag{108}$$

6.3 Dvoretzky's Theorem

Theorem. *Let $0 < \epsilon < 1$. Take n and m natural numbers with $1 \le m \le c\epsilon^2 \log n$, where $c = \frac{(c')^2}{4}$, and c' is the constant from Lemma 4. Given a normed space $X = (\mathbb{R}^n, \|\cdot\|)$, there exists an m-dimensional subspace Y of X such that $d(Y, \ell_2^m) \le \frac{1+\epsilon}{1-\epsilon}$.*

Proof. First, apply the Extended Dvoretzky-Rogers Lemma to get a sequence of $N = \left[\frac{\sqrt{[\frac{n}{2}]}}{16}\right]$ vectors $\{x_i\}_{i=1}^N$ in X such that for all $t \in \mathbb{R}^N$, we have

$$\frac{1}{2}\max_i |t_i| \le \left\|\sum_{i=1}^N t_i x_i\right\| \le 2\left(\sum_{i=1}^N t_i^2\right)^{\frac{1}{2}}. \tag{109}$$

Take $T = S^{m-1}$ and $\{g_i\}_{i=1}^N$, $\{h_j\}_{j=1}^m$ to be independant standard Gaussian variables. We will show the collection $\{x_i\}$ satisfies the conditions of Gordon's Theorem 2. The first condition is satisfied up to the constant 2, which is fine for our purposes. Now, we have

$$\mathbb{E}^*(T) = \mathbb{E}\max_{t\in S^{m-1}}\left\{\sum_{j=1}^m h_j t_j\right\} = \mathbb{E}\sup_{t\in S^{m-1}}\{h\cdot t\} = \mathbb{E}\|h\| \le \sqrt{m}. \tag{110}$$

The last inequality is not obvious, and a proof given in the appendix, section (A.1). Continuing on:

$$\begin{aligned}
\sqrt{m} &\leq \epsilon\sqrt{c}\sqrt{\log n} && \text{by choice of } m && (111)\\
&\leq \epsilon\frac{1}{2}\,\mathbb{E}\max_{i\leq N}|g_i| && \text{by choice of } c && (112)\\
&\leq \epsilon\,\mathbb{E}\left\|\sum_{i=1}^{N} g_i x_i\right\| && \text{by choice of } x_i && (113)\\
&= \epsilon\,\mathbb{E}(\{x_i\}). && && (114)
\end{aligned}$$

Therefore, $\mathbb{E}^*(T) \leq \epsilon\,\mathbb{E}(\{x_i\})$, and we can apply Gordon's Theorem 2 to get a map $A : \mathbb{R}^m \to X$ such that, letting $Y = A(\ell_2^m)$:

$$d(Y, \ell_2^m) \leq \|A\|\|A^{-1}|_Y\| = \frac{\max_{t\in T}\|A(t)\|}{\min_{t\in T}\|A(t)\|} \leq \frac{1+\epsilon}{1-\epsilon}. \tag{115}$$

This completes the proof. □

A Appendix

A.1 Estimating the Expectation of a Gaussian Vector

In section (6.3), we used a well-known by not obvious inequality. This inequality is proven here.

Theorem. *Suppose $h = (h_1, \ldots, h_n)$ is a standard Gaussian random vector in $\mathbb{R}^n$. Then $\frac{n}{\sqrt{n+1}}\,\mathbb{E}\|h\| \leq \sqrt{n}$.*

Proof. We will first directly integrate $\mathbb{E}\|h\|$. By definition

$$\mathbb{E}\|h\| = \frac{1}{(2\pi)^{\frac{n}{2}}}\int_{\mathbb{R}^n}\|x\|e^{-\frac{\|x\|}{2}}dx. \tag{116}$$

Now, moving to spherical coordinates, we see that

$$\mathbb{E}\|h\| = \frac{1}{(2\pi)^{\frac{n}{2}}}\int_0^\infty r^n e^{-\frac{r^2}{2}}dr\prod_{k=2}^{n-1}\int_0^\pi \sin^{k-1}(\phi_{n-k})d\phi_{n-k}\int_0^{2\pi}d\phi_{n-1}. \tag{117}$$

Now it is a property of the beta function that $B(x,y) = 2\int_0^{\frac{\pi}{2}}(\sin\theta)^{2x-1}(\cos\theta)^{2y-1}d\theta$ and that $B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$. Using these, we can confirm

$$\int_0^\pi \sin^{k-1}(\phi_{n-k})d\phi_{n-k} = B\left(\frac{k}{2},\frac{1}{2}\right) = \frac{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{k+1}{2}\right)} \tag{118}$$

Then, using the fact that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ and $\int_0^{2\pi}d\phi_{n-1} = 2\pi$, we can write

$$\begin{aligned}
\mathbb{E}\|h\| &= \frac{2\pi\pi^{\frac{n-2}{2}}}{(2\pi)^{\frac{n}{2}}}\int_0^\infty r^n e^{-\frac{r^2}{2}}dr\prod_{k=2}^{n-1}\frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{k+1}{2}\right)} && && (119)\\
&= \frac{2^{\frac{2-n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\int_0^\infty r^n e^{-\frac{r^2}{2}}dr && \text{by cancelling} && (120)\\
&= \frac{2^{\frac{2-n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\int_0^\infty 2^{\frac{n}{2}}t^{\frac{n}{2}}e^{-t}\frac{dt}{(2t)^{\frac{1}{2}}} && \text{letting } \frac{r^2}{2} = t && (121)\\
&= \frac{\sqrt{2}}{\Gamma\left(\frac{n}{2}\right)}\int_0^\infty t^{\frac{n-1}{2}}e^-tdt && && (122)\\
&= \frac{\sqrt{2}\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} && \text{by definition of } \Gamma. && (123)
\end{aligned}$$

Now we will prove the theorem by induction. The base case has $n = 1$. By direct computation we can verify that

$$\frac{1}{\sqrt{2}} < \frac{\sqrt{2}\Gamma(1)}{\Gamma\left(\frac{1}{2}\right)} = \frac{\sqrt{2}}{\sqrt{\pi}} < 1. \tag{124}$$

Now, assume the inductive hypothesis, namely that for some natural $n = k$ we have

$$\frac{k}{\sqrt{k+1}} \leq \frac{\sqrt{2}\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \leq \sqrt{k}. \tag{125}$$

Therefore, taking reciprocals and then multiplying by k gives (noting $\frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$):

$$\sqrt{k} \leq \frac{k}{2} \frac{\sqrt{2}\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{k+1}{2}\right)} \leq \sqrt{k+1}. \tag{126}$$

Now since $x\Gamma(x) = \Gamma(x+1)$, the middle term becomes $\frac{\sqrt{2}\Gamma(\frac{k+2}{2})}{\Gamma(\frac{k+1}{2})}$. Noting that $\sqrt{k+1} \leq \sqrt{k+2}$ and $\frac{k+1}{\sqrt{k+2}} \leq \sqrt{k}$, we conclude

$$\frac{k+1}{\sqrt{k+2}} \leq \frac{\sqrt{2}\Gamma\left(\frac{k+2}{2}\right)}{\Gamma\left(\frac{k+1}{2}\right)} \leq \sqrt{k+1}. \tag{127}$$

This completes the induction and thus the proof. □

A.2 Lemma 3: The Infinite Dimensional Case

In section (4.2), we alluded to the fact that Lemma 3 had a variant that held in infinite dimensional spaces. We state and prove this variant here.

Definition. *Let $T : U \to V$ be a linear map between Banach spaces. Define $K_n = \{\|T - S\| \mid S : U \to V, rank(S) < n\}$, where $\|\cdot\|$ is the usual operator norm. Then define $a_n(T) = \inf K_n$. Note here that $K_1 = \{\|T\|\}$, and thus $a_1(T) = \|T\|$, since the only $S : U \to V$ of rank less than 1 is the zero map.*

Lemma (Infinite Dimensional Version of Lemma 3)**.** *Let $T : H \to X$ be a linear operator from a Hilbert space $(H, |\cdot|)$ to a Banach space $(X, |\cdot|)$. For all $\epsilon > 0$, there exists an orthonormal* *sequence* $\{f_n\}$ *in* H *such that* $\|Tf_n\| \geq a_n(T) - \epsilon$ *for any* n.

The proof proceeds in exactly the same way as in Lemma 3, only we cannot appeal to compactness of closed unit balls or to continuity of T. This being the case, details already covered in the proof of Lemma 3 are ommited here.

Proof. We again construct the sequence $\{f_n\}$ recursively. Since $\|T\| = \sup\{\|Tv\| \mid |v| = 1\}$, there exists $f_1 \in H$ such that $\|Tf_1\| \geq \|T\| - \epsilon$ and $|f_1| = 1$. We can take f_1 to be the first vector in our sequence.

Let the subspace S_1 and the map $S : X \to H$ be as defined in the proof of Lemma 3 (that is, section 4.2). Then as before we have

$$\|T|_{S_1}\| \geq a_2(T). \tag{128}$$

Now find $f_2 \in S_1$ such that $\|T|_{S_1} f_2\| \geq \|T|_{S_1}\| - \epsilon$ and $|f_2| = 1$. Then by inequality (128) we have $\|Tf_2\| = \|T|_{S_1} f_2\| \geq a_2(T) - \epsilon$. So we can take f_2 to be the second vector in our sequence. Repeating this procedure generates an orthonormal list of vectors $\{f_n\}$ in H with the desired properties. This completes the proof. □

A.3 Estimating $\delta(\epsilon)$

In section (2.2), we found that Lemma 2 holds by taking $0 < \delta(\epsilon) = \delta$ to be such that

$$\frac{1}{\sqrt{1+\epsilon}} \leq \frac{1-3\delta}{1-\delta} \text{ and } \frac{1+\delta}{1-\delta} \leq \sqrt{1+\epsilon}. \tag{129}$$

That δ exists can be easily seen. However, we will calculate here the size of δ compared to ϵ for small ϵ. First note that we can rewrite (129) as

$$\sqrt{1+\epsilon} \geq \max\left(\frac{1-\delta}{1-3\delta}, \frac{1+\delta}{1-\delta}\right). \tag{130}$$

Then, for $0 < \delta < \frac{1}{3}$ we have

$$\frac{1-\delta}{1-3\delta} - \frac{1+\delta}{1-\delta} = \frac{4\delta^2}{(1-3\delta)(1-\delta)} > 0. \tag{131}$$

Since we are interested primarily in small δ stipulating $\delta < \frac{1}{3}$ is reasonable. Then we can rewrite (130) as $\sqrt{1+\epsilon} \geq \frac{1-\delta}{1-3\delta}$. Rearranging this inequality leads to

$$(8+9\epsilon)\delta^2 - (4+6\epsilon)\delta + \epsilon \geq 0, \tag{132}$$

and by concavity of this quadratic we see the above is satisfied when $0 < \delta \leq \frac{2+3\epsilon-2\sqrt{1+\epsilon}}{8+9\epsilon}$ (remember we do not want large δ). Now, supposing $\epsilon \leq \frac{1}{9}$, we see

$$\begin{aligned}
\frac{2+3\epsilon-2\sqrt{1+\epsilon}}{8+9\epsilon} &\geq \frac{2+3\epsilon-2\sqrt{1+\epsilon}}{8+9\epsilon} && \text{since } \sqrt{1+\epsilon} \leq 1+\epsilon && (133)\\
&= \frac{\epsilon}{8+9\epsilon} && && (134)\\
&\geq \frac{\epsilon}{9} && \text{since } \epsilon \leq \frac{1}{9}. && (135)
\end{aligned}$$

Therefore, if $\epsilon \leq \frac{1}{9}$, we can set $\delta(\epsilon) = \min\left(\frac{\epsilon}{9}, \frac{1}{3}\right) = \frac{\epsilon}{9}$. This result is what allows us to calculate η_1 and η in the first version of Dvoretzky's theorem.

A.4 Proving the Gauge Gives a Norm

In section (1.1), we stated that given a compact, symmetric, convex body $K \subseteq \mathbb{R}^N$ containing 0, the gauge $\|\cdot\|_K$ defined a norm on $\mathbb{R}^N$. This is not obvious; we give a proof here. Denote $\|\cdot\|_K = \phi$. We need to show ϕ is homogeneous, satisfies the triangle inequality, and achieves zero only on the zero vector.

- First, we verify the triangle inequality. Before we begin, notice that if $x \in r_1 K$ and $y \in r_2 K$, then $x + y \in (r_1 + r_2)K$. To see this, let $z = \frac{x+y}{r_1+r_2}$. Then $z \in K$ since K is convex, and so $(r_1 + r_2)z = x + y$ is in $(r_1 + r_2)K$. Now take any $x, y \in \mathbb{R}^N$ and any $\epsilon > 0$. Then by definition of the gauge, we have both $x \in \left(\phi(x) + \frac{\epsilon}{2}\right) K$ and $y \in \left(\phi(y) + \frac{\epsilon}{2}\right) K$. This means $x + y \in (\phi(x) + \phi(y) + \epsilon)K$, and hence that

$$\phi(x + y) \leq \phi(x) + \phi(y) + \epsilon. \tag{136}$$

  Since (136) is true for all $\epsilon > 0$, we conclude the triangle inequality does indeed hold.

- Second, we check that $\phi(x) = 0$ if and only if $x = 0$. Since $0 \in 0K$, and ϕ is at least 0, we have $\phi(0) = 0$. Now suppose $\phi(x) = 0$. Then we have $x \in \frac{1}{n}K$ for all naturals n. This means that $x \in \bigcap_{n \in \mathbb{N}} \frac{1}{n}K$. Now, using the facts that K is compact, symmetric, and convex, we see the $\frac{1}{n}K$ form a nested sequence of compact intervals of decreasing diameter. By Cantor's intersection theorem, it follows that $\bigcap_{n \in \mathbb{N}} \frac{1}{n}K$ is a singleton set. This set obviously contains 0, and since it also contains x we see $x = 0$. This proves $\phi(x) = 0$ if and only if $x = 0$.

- Now we show ϕ is homogeneous. Suppose $x \in \mathbb{R}^N$ and $\alpha \in \mathbb{R}$. If $\alpha = 0$, then $\phi(\alpha x) = \phi(0) = 0 = 0 \cdot \phi(x)$, so homogeniety holds. So let $\alpha \neq 0$. Then take $\epsilon > 0$. We have

$$
\begin{aligned}
\phi(\alpha x) &= \inf\{r > 0 \mid \alpha x \in rK\} && && (137)\\
&= \inf\{r > 0 \mid |\alpha| x \in rK\} && \text{by symmetry of } K && (138)\\
&= \inf\{r > 0 \mid x \in \frac{r}{|\alpha|}K\} && \text{since } \alpha \neq 0 && (139)\\
&= \inf\{|\alpha|\frac{r}{|\alpha|} > 0 \mid x \in \frac{r}{|\alpha|}K\} && && (140)\\
&= |\alpha| \inf\{r' > 0 \mid x \in r'K\} && \text{letting } r' = \frac{r}{|\alpha|} && (141)\\
&= |\alpha|\phi(x). && && (142)
\end{aligned}
$$

This shows ϕ is homogeneous. As this was the final property of ϕ we had to verify, we conclude the gauge of K is indeed a norm.

References

[1] Gilles Pisier, *The Volume of Convex Bodies and Banach Space Geometry*, Cambridge University Press, Cambridge, 1st edition, 1989.

[2] Yehoram Gordon, *Applications of the Gaussian Min-Max Theorem.* Unpublished. Linked on the course website.

[3] Michel Talagrand *Embedding of ℓ_k^∞ and a theorem of Alon and Milman*, Geometric aspects of functional analysis (Israel), Operator Theorey Advances and Applications, 77, Birkhauser, Basel, 1995.