LECTURE NOTES ON DVORETZKY'S THEOREM

STEVEN HEILMAN

Abstract. We present the first half of the paper [S]. In particular, the results below, unless otherwise stated, should be attributed to G. Schechtman. Below, P and E denote probability and expectation, respectively. We try to make explicit the measure space in question by adding an appropriate subscript to P, E whenever necessary.

1. Introduction

Our goal is to prove the following theorem (Theorem 3 from [S]), which proves Dvoretzky's Theorem (Thm. 1.7) with the best known ε dependence:

Theorem 2.3: $\exists\, c > 0$ such that $\forall\, n \in \mathbb{N}, \forall\, \varepsilon > 0$, every n-dimensional normed linear space X admits a subspace $Y \subseteq X$ such that $d(Y, \ell_2^k) \leq 1 + \varepsilon$, and $k > \frac{c\varepsilon}{(\log(1/\varepsilon))^2} \log n$.

Let us sketch the proof. If $M := \int_{S^{n-1}} ||x||\, d\mu(x)$ is large, then we can use (the proof and conclusion of) Milman's Dvoretzky theorem (Thm. 1.5) to get a sphere of large dimension. If M is small, then via our main Lemma (Lemma 2.1), the assumptions of the Alon-Milman theorem (Thm. 1.10) are satisfied. That is, we can find a cube of large dimension, which itself has a sphere of large dimension. Thus, in either case, we have our desired result.

Note that the new ingredient in this proof is the application of the Alon-Milman Theorem, which is enabled by the crucial Lemma 2.1. In the last class, Talagrand's proof [T] of this theorem was presented by Evan Chou and Lukas Koehler.

In Section 2, we prove Theorem 2.3 in more detail. Section 3 gives some preliminaries. Unfortunately, we will use a result (Theorem 1.5) which was not yet proven in class. In the following lecture, Sean Li will prove this result, which is known as Milman's extension of Dvoretzky's theorem, with best known constants.

Before we begin, we need to cite some relevant theorems and definitions.

Remark 1.1. Below, $g_1, g_2, \ldots$ will always designate standard normal real valued random variables on a probability space Ω. ($\mathbb{E}g_i = 0, \mathbb{E}g_i^2 = 1$, $P(g_i < t) = \int_{-\infty}^{t} e^{-x^2/2} dx/\sqrt{2\pi}$). Also, $r_j(t) := \operatorname{sign}\sin(2^j t)$, $t \in [0,1]$ denote the Rademacher functions. Moreover, c will be a constant that is allowed to change from line to line.

Remark 1.2. In the proofs of Lemma 2.1 and Theorem 2.3, we are going to use the triangle inequality in the form $\frac{||a+b||+||a-b||}{2} \geq \left|\left|\frac{1}{2}(a+b+a-b)\right|\right| = ||b||$. Let α and β be independent random variables, with symmetric distributions, so $P(\alpha < -t) = P(\alpha > t)$ for all $t > 0$. Integrating separately for $\beta < 0$ and $\beta > 0$ and taking expected values shows that $\mathbb{E}\,||\alpha a + \beta b|| \geq \mathbb{E}\,||\beta b||$. That is, throwing out vectors

Date: December 7th, 2010.

only makes the expectation smaller. One can also see this result as a consequence of the Contraction Principle.

Remark 1.3. In the proof of Theorem 2.3, we need: ℓ_∞^m contains a subspace of dimension at least $k = \frac{c}{\log(1/\varepsilon)} \log m$ which is $(1+\varepsilon)$-isomorphic to Euclidean space. This was almost proven in class (actually it was assigned as an exercise).

Recall: we consider an embedding $T\colon \ell_2^k \to \ell_\infty^m$ defined by $Tx := \{\langle x, x_i\rangle\}_{i=1}^m$, $\{x_i\}_{i=1}^m$ is an ε-net of S^{k-1} of size $m = \lfloor \left(\frac{3}{\varepsilon}\right)^k \rfloor$. That is, we "flatten" the sphere into an approximating polytope. Finally, such an ε-net exists by volume considerations (which we proved in class), and taking logs shows $\log m = k\log(3/\varepsilon)$, as desired.

Definition 1.4. Let X be a normed space. Define $E(X)$ by

$$E(X) := \sup\left\{\mathbb{E}_\Omega\left(\left|\left|\sum_{i=1}^n g_i u(e_i)\right|\right|_X : n\in\mathbb{N}, u\colon \ell_2^n \to X, ||u|| = 1\right)\right\}.$$

Theorem 1.5. ***(Milman's Dvoretzky Theorem)*** *$\exists$ a function $c(\varepsilon) > 0$ such that, for all $k \leq c(\varepsilon)E(X)^2$, $\ell_2^k \overset{(1+\varepsilon)}{\hookrightarrow} X$. Actually, one may take $k \leq c\varepsilon^2 E(X)^2$.*

That is, $\exists$ a linear invertible $T\colon \ell_2^k \to Y \subseteq X$ with $||T||\,||T^{-1}|| \leq (1+\varepsilon)$.

Remark 1.6. Sean Li will prove this theorem independently of our work. We will therefore use this result without further comment

Theorem 1.7. ***(Classical Dvoretzky)*** *Let X be a normed space of dimension n. There exists a function $c(\varepsilon) > 0$ such that, for all $k \leq c(\varepsilon)\log n$, $\ell_2^k \overset{(1+\varepsilon)}{\hookrightarrow} X$.*

For an expression for $c(\varepsilon)$, see Theorem 2.3. Now, recall three results from class.

Lemma 1.8. ***(Johnson-Lindenstrauss Lemma)*** *Let H be a Hilbert space. $\forall$ $\varepsilon \in (0,1)$, $\exists$ $c(\varepsilon) > 0$ such that, $\forall$ $n \in \mathbb{N}$, if $x_1, \ldots, x_n \in H$, then $\exists$ $k \leq c(\varepsilon)\log n$ and $y_1, \ldots, y_n \in \ell_2^k$ such that $\forall\, i, j$*

$$||x_i - x_j||_H \leq ||y_i - y_j||_2 \leq (1+\varepsilon)\,||x_i - x_j||_H\,.$$

Proof: (Rough Sketch) Using the probabilistic method on the orthogonal group, find a random projection that does what we want. ■

Lemma 1.9. ***(Dvoretzky-Rogers Lemma)*** *Let $||\cdot||$ be a norm on $\mathbb{R}^N$ with unit ball K. Let $\mathcal{E} \subseteq K$ be the John ellipsoid (i.e. the ellipsoid of maximal volume in K). Then $\exists$ $x_1, \ldots, x_N \in \mathbb{R}^N$ which are orthonormal with respect to $\langle\cdot,\cdot\rangle_{\mathcal{E}}$, and such that $||x_i|| \geq 2^{-N/(N-i+1)}$, $i = 1, \ldots, N-1$.*

Proof: (Rough Sketch) Inductively find new vectors of maximal norm, and pay attention to the volume of the ellipsoids.

Theorem 1.10. ***(Alon-Milman, Talagrand)*** *With the assumptions of Lemma 2.1, for any $0 < \varepsilon < 1$, $\exists$ a subspace of $\{\mathrm{span}\{e_i\}_{i=1}^n\, ||\cdot||\}$ of dimension $k \geq cn^{c\varepsilon/\log L}$ that is $(1+\varepsilon)$-isomorphic to ℓ_∞^k. Here $c > 0$ is a universal constant.*

Remark 1.11. In class, we proved $k \geq cn^{\frac{c\log(1+\varepsilon)}{\log M_n}}$ (see (2)). Using the assumptions and conclusion of Lemma 2.1, we have $M_n \leq 200L$, and then using $\log(1+\varepsilon) \approx \varepsilon$ gives Theorem 1.10.

2. Proof of Main Theorem

With the preliminary results of the Appendix (Section 3) and the following crucial Lemma 2.1, the proof of Theorem 2.3 will follow quickly. In the following Lemma, we use the notation of [T] (in particular, the constants M and M_n).

Lemma 2.1. ***(Main Lemma),*** [S] *Let $||\cdot||$ be a norm on $\mathbb{R}^N$ containing the unit Euclidean ball (so that $||\cdot|| \leq ||\cdot||_2$). Let $\{e_i\}_{i=1}^n$ be an orthonormal sequence in $\mathbb{R}^N$ (with respect to $||\cdot||_2$) satisfying $||e_i|| \geq 1/4$ for all i, and*

$$\sqrt{n}M = \mathbb{E}_\Omega \left|\left| \sum_{i=1}^n g_i e_i \right|\right| \leq L\sqrt{\log n} \quad \textbf{\textit{(small roundness)}} \tag{1}$$

(Recall $M := \int_{S^{N-1}} ||a||\, d\mu(a)$, with μ normalized Haar measure). Then, for all disjoint subsets $\sigma_1, \ldots, \sigma_{\lfloor\sqrt{n}\rfloor} \subseteq \{1, \ldots, n\}$ with $|\sigma_j| = \lfloor\sqrt{n}\rfloor$ for all j, $\exists$ a further subset $J \subseteq \{1, \ldots, \lfloor\sqrt{n}\rfloor\}$ of cardinality at least $\sqrt{n}/2$, and there are $\{x_j\}_{j\in J}$ with $x_j = \sum_{i\in\sigma_j} \lambda_i e_i$, $||x_j|| = 1$, and

$$M_n := \mathbb{E}_{t\in[0,1]} \left|\left| \sum_{j\in J} r_j(t) x_j \right|\right| \leq 200L \quad \textbf{\textit{(large cube-ness)}} \tag{2}$$

Remark 2.2. To summarize, we can rearrange and add together our n orthonormal vectors to get $\sqrt{n}/2$ vectors with small expected length.

Proof: Ideas: Gaussian concentration, the probabilistic method, and symmetry. Corollary 3.6 (applied to $m = \sqrt{n}$ and using $||e_i|| \geq 1/4$) shows that

$$\mathbb{E} \left|\left| \sum_{i\in\sigma_j} g_i e_i \right|\right| \geq \frac{\sqrt{\log n}}{30\sqrt{2}} \geq \frac{\sqrt{\log n}}{50} \tag{3}$$

Define $T\colon (\mathbb{R}^{\sqrt{N}}, ||\cdot||_2) \to (\mathbb{R}, |\cdot|)$ by $T(\{\lambda_i\}_{i\in\sigma_j}) = ||\sum_{i\in\sigma_j} \lambda_i e_i||$. Using the orthonormality of the e_i (and that $||\cdot|| \leq ||\cdot||_2$), the map T is 1-Lipschitz. So, we can apply (3) and Lemma 3.8 to get

$$P\left(||\sum_{i\in\sigma_j} g_i e_i|| \leq \frac{1}{100}\sqrt{\log n} \right) \leq P\left(\left| ||\sum_{i\in\sigma_j} g_i e_i|| - \mathbb{E}||\sum_{i\in\sigma_j} g_i e_i|| \right| > \frac{1}{100}\sqrt{\log n} \right)$$
$$\leq 2e^{-2\frac{1}{10{,}000\pi^2}\log n}.$$

So, for $n \geq 4^{5{,}000\pi^2}$, this probability is $\leq 1/2$, for every j. Let A_j denote the event $\{||\sum_{i\in\sigma_j} g_i e_i|| > \frac{1}{100}\sqrt{\log n}\}$. Let A denote the following event: there exists at least one subset $J \subseteq \{1, \ldots, \lfloor\sqrt{n}\rfloor\}$ with $|J| \geq \lfloor\sqrt{n}\rfloor/2$ and A_j occurs for all $j \in J$. Then $P(A) \geq 1/2$, from Proposition 3.9. Now, extend the probability space Ω to include independent Rademacher functions, apply our assumptions, and observe

$$
\begin{aligned}
L\sqrt{\log n} &\geq \mathbb{E}_g \left|\left| \sum_{j=1}^{\lfloor\sqrt{n}\rfloor} \sum_{i\in\sigma_j} g_i e_i \right|\right| \quad \text{, by (1)} \\
&= \mathbb{E}_r \mathbb{E}_g \left|\left| \sum_{j=1}^{\lfloor\sqrt{n}\rfloor} r_j \sum_{i\in\sigma_j} g_i e_i \right|\right| \quad \text{, by symmetry of the } g_i \\
&\geq \mathbb{E}_r \mathbb{E}_g \left(\left|\left| \sum_{j=1}^{\lfloor\sqrt{n}\rfloor} r_j \sum_{i\in\sigma_j} g_i e_i \right|\right| 1_A \right) \geq \frac{1}{2} \mathbb{E}_g \left(\left(\mathbb{E}_r \left|\left| \sum_{j=1}^{\lfloor\sqrt{n}\rfloor} r_j \sum_{i\in\sigma_j} g_i e_i \right|\right| \right) \middle| 1_A \right).
\end{aligned}
$$

Here we used the definition of conditional expectation, and that $P(A) \geq 1/2$. In conclusion, by the definition of A, if we choose $\omega \in A$, then $\exists\ J \subseteq \{1, \ldots, \lfloor\sqrt{n}\rfloor\}$ with $|J| \geq \lfloor\sqrt{n}/2\rfloor$ as above, so that A_j holds for all $j \in J$. That is, $\tilde{x}_j := \sum_{i\in\sigma_j} g_i(\omega) e_i$ satisfies $||\tilde{x}_j|| > \frac{1}{100}\sqrt{\log n}$, and by the inequality above (and Remark 1.2), there exists $\omega \in A$ such that

$$\mathbb{E}_r \left|\left| \sum_{j\in J} r_j \tilde{x}_j \right|\right| \leq 2L\sqrt{\log n}.$$

So, taking $x_j := \tilde{x}_j / ||\tilde{x}_j||$ completes the Lemma, since then

$$\mathbb{E}_r \left|\left| \sum_{j\in J} r_j x_j \right|\right| \leq 2 \cdot 100L \frac{\sqrt{\log n}}{\sqrt{\log n}},$$

which is (2), as desired. ■

Theorem 2.3. ***(Dvoretzky, Best Constants),*** [S] $\exists\ c, d > 0$ *such that for all* $n \in \mathbb{N}$ *and all* $0 < \varepsilon < d < 1$*, every* n*-dimensional normed space admits a subspace* Y *with* $d(Y, \ell_2^k) \leq (1+\varepsilon)$*, and* $k > \frac{c\varepsilon}{(\log(1/\varepsilon))^2}\log n$.

Equivalently, every symmetric convex body in $\mathbb{R}^n$ *admits a* k*-dimensional section that contains an Euclidean ball and is contained in* $1+\varepsilon$ *times that ball, where* $k > \frac{c\varepsilon}{(\log(1/\varepsilon))^2}\log n$.

Proof: We may assume (by taking an appropriate linear transformation) that $X = (\mathbb{R}^n, ||\cdot||)$ and S^{n-1} is the ellipsoid of maximal volume in $K := B_{X,||\cdot||}$. By Dvoretzky-Rogers (Lemma 1.9), $\exists\ \{e_1, \ldots, e_{\lfloor n/2\rfloor}\} \subseteq \mathbb{R}^n$ orthonormal, with $||e_i|| \geq 1/4$, $i = 1, \ldots, \lfloor n/2\rfloor$. Since $S^{n-1} \subseteq K$, $||\cdot|| \leq ||\cdot||_2$. Define $E := \mathbb{E}\,||\sum_{i=1}^n g_i e_i||$. Applying Theorem 1.5 (recall the definition of $E(X)$ and E), $\exists\ Y \subseteq X$ with $d(Y, \ell_2^k) \leq (1+\varepsilon)$, and $k > c\varepsilon^2 E^2$. So, we have two cases:

Case 1 (large ball): $\varepsilon^2 E^2 \geq \frac{\varepsilon}{(\log(1/\varepsilon))^2}\log n$.

Case 2 (large cube): $\varepsilon^2 E^2 < \frac{\varepsilon}{(\log(1/\varepsilon))^2}\log n$.

For Case 1, we have $k > \frac{c\varepsilon}{(\log(1/\varepsilon))^2}\log n$ (by assumption), so the theorem is proven. For Case 2, we apply Remark 1.2, use the definition of E, and rewrite the assumption for this case to get **small roundness:**

$$\mathbb{E}_\Omega || \sum_{i=1}^{\lfloor n/2\rfloor} g_i e_i || \leq E \leq \frac{1}{\sqrt{\varepsilon}\log\frac{1}{\varepsilon}}\sqrt{\log n} \tag{4}$$

We therefore apply Lemma 2.1 with $L = \frac{1}{\sqrt{\varepsilon}\log(1/\varepsilon)}$, giving **large cube-ness**: $M_{n/2} \leq 200L$.

By Alon-Milman (Theorem 1.10), we can take $m \geq c(n/2)^{\frac{c\varepsilon}{\log(\varepsilon^{-1/2}(\log\frac{1}{\varepsilon})^{-1})}}$ and $\exists$ $Y \subseteq X$ of dimension m with $d(Y, \ell_\infty^m) \leq (1+\varepsilon)$. Then by Remark 1.3, Y contains a subspace Z of dimension k where $Z \subseteq Y \subseteq X$ and

$$k \geq \frac{c}{\log(1/\varepsilon)} \log m \geq \frac{c}{\log\frac{1}{\varepsilon}} \left(\frac{c\varepsilon}{\log(\varepsilon^{-1/2}(\log\frac{1}{\varepsilon})^{-1})} \log n \right) \geq \frac{c\varepsilon}{(\log\frac{1}{\varepsilon})^2} \log n,$$

for ε small, where Z is $(1+\varepsilon)^2$-isomorphic to ℓ_2^k (and $(1+\varepsilon)^2 \leq 1+3\varepsilon$ for $0 < \varepsilon < 1$).

The final statement of the theorem follows since an n-dimensional ellipsoid has a spherical section of dimension at least $n/16$ (we proved this in class). ■

Remark 2.4. Concerning the computation of ε near the end of the proof, we have

$$\begin{aligned}
&\frac{\tilde{c}}{\log\frac{1}{\varepsilon}} \left(\frac{\varepsilon}{\log(\varepsilon^{-1/2}(\log\frac{1}{\varepsilon})^{-1})} \log n \right) \geq \frac{c\varepsilon}{(\log\frac{1}{\varepsilon})^2} \log n \\
&\quad \Leftrightarrow \frac{1}{\log(\varepsilon^{-1/2}(\log\frac{1}{\varepsilon})^{-1})} \geq \frac{C}{\log\frac{1}{\varepsilon}} \\
&\quad \Leftrightarrow \log\frac{1}{\varepsilon} \geq C \log(\varepsilon^{-1/2}(\log\frac{1}{\varepsilon})^{-1}) \\
&\quad \Leftrightarrow \frac{1}{\varepsilon} \geq \left(\varepsilon^{-1/2}(\log\frac{1}{\varepsilon})^{-1} \right)^C \\
&\quad \Leftrightarrow \varepsilon^{-1/C+1/2} \geq (\log\frac{1}{\varepsilon})^{-1} \\
&\quad \Leftrightarrow \varepsilon^{1/C-1/2} \leq -\log\varepsilon
\end{aligned}$$

For small $C = c/\tilde{c} > 0$, the last inequality is only true for ε small.

Acknowledgements: Thanks to Evan Chou and Assaf Naor for fixing an error in the proof of Lemma 2.1, and thanks to Sean Li for helpful discussions.

3. Appendix: Preliminary Things

Proposition 3.1. Let g be a standard normal real valued gaussian random variable. We have the following estimates on the distribution function of g:

(a) $\frac{1}{\lambda}e^{-\lambda^2/2} \geq \int_\lambda^\infty e^{-x^2/2}dx$, for $\lambda > 0$.
(b) $\frac{1}{2\lambda}e^{-\lambda^2/2} \leq \int_\lambda^\infty e^{-x^2/2}dx$, for $\lambda > \sqrt{2}$.
(c) $\int_\lambda^\infty e^{-x^2/2}dx \sim \frac{1}{\lambda}e^{-\lambda^2/2}$, as $\lambda \to \infty$.

Proof: Let $\lambda > 0$, and observe

$$\begin{aligned}
\int_{\lambda}^{\infty} e^{-x^2/2} &= \int_0^{\infty} e^{-(y+\lambda)^2/2} dy \quad \text{, c.o.v.} \\
&= \int_0^{\infty} e^{-y^2/2} e^{-2\lambda y/2} e^{-\lambda^2/2} dy \\
&\leq e^{-\lambda^2/2} \int_0^{\infty} e^{-\lambda y} dy \quad , e^{-y^2/2} \leq 1 \\
&= e^{-\lambda^2/2} \left[\frac{-1}{\lambda} e^{-\lambda y} \right]_{y=0}^{y=\infty} \quad \text{, using } \lambda > 0 \\
&= \frac{1}{\lambda} e^{-\lambda^2/2},
\end{aligned}$$

which proves (a). Now, let $\lambda > \sqrt{2}$ and observe

$$\begin{aligned}
&\frac{d}{d\lambda} \left[(\lambda^{-1} - \lambda^{-3}) e^{-\lambda^2/2} \right] \\
&\quad = -e^{-\lambda^2/2} + \lambda^{-2} e^{-\lambda^2/2} - \lambda^{-2} e^{-\lambda^2/2} + 3\lambda^{-4} e^{-\lambda^2/2} = -(1 - 3\lambda^{-4}) e^{-\lambda^2/2}
\end{aligned}$$

Therefore, the Fundamental Theorem of Calculus gives

$$\begin{aligned}
\int_{\lambda}^{\infty} e^{-x^2/2} dx &\geq \int_{\lambda}^{\infty} (1 - 3x^{-4}) e^{-x^2/2} dx \quad \text{, by monotonicity} \\
&= (\lambda^{-1} - \lambda^{-3}) e^{-\lambda^2/2} \quad \text{, by differentiating the integral,} \quad (*) \\
&\geq \frac{1}{2\lambda} e^{-\lambda^2/2} \quad \text{, since } \lambda > \sqrt{2} \text{, so } (1/2)\lambda^{-1} - \lambda^{-3} > 0
\end{aligned}$$

Actually, from the above sequence of inequalities ending at $(*)$, we see that we have proved (c) as well (using (a) too). ■.

Proposition 3.2. Let $\{g_1, \ldots, g_n\}$ be n standard real valued gaussians. Then

$$P\left(\max_{i=1,\ldots,n} |g_i| < c\sqrt{\log n} \right) \to \begin{cases} 0, & c < \sqrt{2}, c > 0 \\ 1, & c > \sqrt{2} \end{cases} .$$

Proof:

$$\begin{aligned}
P\left(\max_{i=1,\ldots,n} |g_i| < \lambda \right) &= P\left(|g_1| < \lambda \right)^n \quad \text{, by definition of max} \\
&= \left(\int_{-\lambda}^{\lambda} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right)^n \quad \text{, by definition of the } g_i \\
&= \left(1 - 2 \int_{\lambda}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right)^n \\
&\leq \left(1 - \frac{1}{\sqrt{2\pi}} \frac{1}{\lambda} e^{-\lambda^2/2} \right)^n \quad \text{, by Proposition 3.1(b).}
\end{aligned}$$

So, setting $\lambda = c\sqrt{\log n}$ we get

$$P\left(\max_{i=1,\dots,n}|g_i|<\lambda\right)\le\left(1-\frac{1}{c\sqrt{2\pi}\sqrt{\log n}}(e^{-\log n})^{c^2/2}\right)^n$$
$$=\left(1-\frac{1}{c\sqrt{2\pi}\sqrt{\log n}}\frac{n^{-(c^2/2)+1}}{n}\right)^n.$$

Then, using the power series expansion of log (recall, $-\log(1-x)=x+x^2/2+x^3/3+\cdots$ for $|x|<1$) shows that (for $c_n=o(n)$)

$$\log(1-c_n/n)^n=n\log(1-c_n/n)\approx -c_n+O(c_n/n).$$

So, for $c\le\sqrt{2}$, $\log P\left(\max_{1,\dots,n}|g_i|<c\sqrt{\log n}\right)\to-\infty$. And for $c>\sqrt{2}$, $\log P\left(\max_{1,\dots,n}|g_i|<c\sqrt{\log n}\right)\to 0$, using Proposition 3.1(a). ■

Proposition 3.3. Let $x_1,\dots,x_n$ be unit norm vectors in a normed space X. Let ε_i be independent, $P(\varepsilon_i=1)=P(\varepsilon_i=-1)=1/2$. Then for all $a_1,\dots,a_n\in\mathbb{R}$,

$$P_{\varepsilon_i}\left(\left|\left|\sum_{i=1}^n\varepsilon_i a_i x_i\right|\right|<\max_{1\le i\le n}|a_i|\right)\le 1/2.$$

Proof: We may assume symmetry and rearrangement that $a_1=\max_{1\le i\le n}|a_i|$. Suppose $||a_1x_1+\sum_{i=2}^n\varepsilon_i a_i x_i||<a_1$. Then

$$\left|\left|a_1x_1-\sum_{i=2}^n\varepsilon_i a_i x_i\right|\right|=\left|\left|a_1x_1+a_1x_1-a_1x_1-\sum_{i=2}^n\varepsilon_i a_i x_i\right|\right|$$
$$\ge||2a_1x_1||-\left|\left|a_1x_1+\sum_{i=2}^n\varepsilon_i a_i x_i\right|\right|$$
$$>a_1\quad\text{, using }||x_i||=1\text{ and our assumption.}$$

Let D be the event $\{||\sum_{i=1}^n\varepsilon_i a_i x_i||<a_1\}$, and let C be the event defined by $\{||\sum_{i=1}^n\varepsilon_i a_i x_i||>a_1\}$. For $\{\varepsilon_i\}_{i=1}^n\subseteq D$, we have shown: there exists a unique $\{\varepsilon_i'\}_{i=1}^n=\{\varepsilon_1,-\varepsilon_2,\dots,-\varepsilon_n\}$ associated to $\{\varepsilon_i\}_{i=1}^n$ with $\{\varepsilon_i'\}_{i=1}^n\subseteq C$. So $P(C)\ge P(D)$, i.e.

$$P\left(\left|\left|\sum_{i=1}^n\varepsilon_i a_i x_i\right|\right|>a_1\right)\ge P\left(\left|\left|\sum_{i=1}^n\varepsilon_i a_i x_i\right|\right|<a_1\right).$$

Thus

$$1\ge P\left(\left|\left|\sum_{i=1}^n\varepsilon_i a_i x_i\right|\right|\ne\max_{1\le i\le n}|a_i|\right)$$
$$=P\left(\left|\left|\sum_{i=1}^n\varepsilon_i a_i x_i\right|\right|<a_1\right)+P\left(\left|\left|\sum_{i=1}^n\varepsilon_i a_i x_i\right|\right|>a_1\right)\quad\text{, by definition of }a_1$$
$$\ge 2P\left(\left|\left|\sum_{i=1}^n\varepsilon_i a_i x_i\right|\right|<a_1\right)$$

■

Remark 3.4. Letting $x_1=x_2$, $a_1=a_2=1$, and $a_i=0$ otherwise, we see that the Proposition is sharp.

Proposition 3.5. Let $x_1, \ldots, x_n$ be vectors in a normed space with $||x_i|| \geq 1/4$ for all i. Let $g_1, \ldots, g_n$ be standard normal gaussian random variables on a probability space Ω. Then

$$P_\Omega\left(\left|\left|\sum_{i=1}^n g_i x_i\right|\right| < \frac{\sqrt{\log n}}{10}\right) \leq 2/3.$$

Proof:

$$\begin{aligned}
&P_\Omega\left(\left|\left|\sum_{i=1}^n g_i x_i\right|\right| < \frac{\sqrt{\log n}}{10}\right)\\
&\leq P_\Omega\left(\left|\left|\sum_{i=1}^n g_i x_i\right|\right| < \frac{\sqrt{\log n}}{10} \;\wedge\; \frac{\sqrt{\log n}}{10} < \max_{1\leq i\leq n} |g_i|\,||x_i||\right)\\
&\quad + P_\Omega\left(\max_{1\leq i\leq n} |g_i|\,||x_i|| \leq \frac{\sqrt{\log n}}{10}\right)\\
&\leq P_{\Omega,\varepsilon_i}\left(\left|\left|\sum_{i=1}^n (\varepsilon_i\,|g_i|\,||x_i||)\frac{x_i}{||x_i||}\right|\right| < \max_{1\leq i\leq n} |g_i|\,||x_i||\right)\\
&\quad + P_\Omega\left(\max_{1\leq i\leq n} |g_i| \leq \frac{2\sqrt{\log n}}{5}\right) \quad \text{, since } \frac{1}{||x_i||} \leq 4 \text{ and } g_i \text{ is symmetric}\\
&\leq 1/2 + o(1) \leq 2/3 \quad \text{, from Proposition 3.3 and Proposition 3.2.}
\end{aligned}$$

In the penultimate equality, we first condition on g_i, then integrate in Ω. ■

Corollary 3.6. *With the assumptions of Proposition 3.5,* $\mathbb{E}_\Omega\left(||\sum_{i=1}^m g_i x_i||\right) \geq \sqrt{\log m}/30$

Definition 3.7. Let Y, Z be two Banach spaces (finite or infinite dimensional). Suppose there exists a linear isomorphism $T\colon Y \to Z$. Define the **Banach-Mazur distance** $d(Y,Z)$ as

$$d(Y,Z) = \inf\{||T||\,||T^{-1}||\colon T\colon Y \twoheadrightarrow Z, \text{ a linear isomorphism}\}.$$

Note that dilating T has no effect on $||T||\,||T^{-1}||$.

Lemma 3.8. ***(Gaussian Concentration)*** *Let* $F\colon (\mathbb{R}^n, ||\cdot||_2) \to \mathbb{R}$ *be a Lipschitz function with constant* σ*. Then*

$$P(|F(g_1,\ldots,g_n) - \mathbb{E}F(g_1,\ldots,g_n)| > C) \leq 2e^{-2C^2/\pi^2\sigma^2}.$$

Proof: Idea: Gaussian integration by parts.

By a mollifying argument, we may assume $F \in C^1$. Let $H = (h_1, \ldots, h_n)$ be equal in distribution to $G = (g_1, \ldots, g_n)$. Define $G_\theta := G\sin\theta + H\cos\theta$, $\theta \in (0, \pi/2)$. Recall that the gaussian distribution is invariant under rotation ($\sqrt{1-t}g_i + \sqrt{t}h_i \overset{d}{=} (1-t+t)^{1/2}g_i = g_i$). So, $G_\theta \overset{d}{=} G$, $\frac{d}{d\theta}G_\theta = G\cos\theta - H\sin\theta \overset{d}{=} H$, and G_θ, $\frac{d}{d\theta}G_\theta$ are independent, since they are jointly normal with zero covariance. That is $(G_\theta, \frac{d}{d\theta}G_\theta) \overset{d}{=} (G,H)$. Let $\mathbb{E}_G, \mathbb{E}_H, \mathbb{E}$ denote expectation with respect to G,

H, and both G and H, respectively. For a convex function ϕ

$$\begin{aligned}
\mathbb{E}_G\phi(F(G) - \mathbb{E}_H F(H)) &\leq \mathbb{E}_G\mathbb{E}_H\phi(F(G) - F(H)) \quad \text{, by Jensen's} \\
&= \mathbb{E}\phi\left(\int_0^{\pi/2} \frac{d}{d\theta}F(G_\theta)d\theta\right) = \mathbb{E}\phi\left(\int_0^{\pi/2}\left\langle \nabla_{\mathbb{R}^n}F(G_\theta), \frac{d}{d\theta}G_\theta\right\rangle d\theta\right) \\
&= \mathbb{E}\phi\left(\int_0^{\pi/2} \frac{2}{\pi}\frac{\pi}{2}\left\langle \nabla_{\mathbb{R}^n}F(G_\theta), \frac{d}{d\theta}G_\theta\right\rangle d\theta\right) \\
&\leq \frac{2}{\pi}\int_0^{\pi/2} \mathbb{E}\phi\left(\frac{\pi}{2}\left\langle \nabla_{\mathbb{R}^n}F(G_\theta), \frac{d}{d\theta}G_\theta\right\rangle\right) d\theta \quad \text{, Jensen, Fubini} \\
&= \mathbb{E}\phi\left(\frac{\pi}{2}\langle \nabla_{\mathbb{R}^n}F(G), H\rangle\right) \quad \text{, by equality of joint distributions.}
\end{aligned}$$

Let $\lambda \in \mathbb{R}$ and set $\phi(t) = e^{\lambda t}$. From basic properties of gaussians ($ag_1 + bg_2 \overset{d}{=} (a^2+b^2)^{1/2}g_1$, and $\int_{\mathbb{R}} e^{tx}d\gamma(x) = \int_{\mathbb{R}} e^{-tx}d\gamma(x) = e^{t^2/2}$) we have

$$\begin{aligned}
&\mathbb{E}_G\exp(\lambda(F(G) - \mathbb{E}_H F(H))) \leq \mathbb{E}\exp\left(\lambda\frac{\pi}{2}\sum_{i=1}^n \frac{\partial F}{\partial x_i}(G)h_i\right) \\
&\quad = \mathbb{E}\exp\left(\lambda\frac{\pi}{2}\left(\sum_{i=1}^n\left(\frac{\partial F}{\partial x_i}(G)\right)^2\right)^{1/2} h_1\right) = \mathbb{E}_G\exp\left(\lambda^2\frac{\pi^2}{8}\sum_{i=1}^n\left(\frac{\partial F}{\partial x_i}(G)\right)^2\right) \\
&\quad \leq \exp(\lambda^2\pi^2\sigma^2/8) \quad \text{, since } ||\nabla F||_2 \leq \sigma.
\end{aligned}$$

This is our desired moment estimate, which gives our result via Markov's inequality. Let $\lambda = 4C/\pi^2\sigma^2 > 0$ and observe

$$\begin{aligned}
&P(|F(G) - \mathbb{E}F(H)| > C) = P(e^{\lambda|F(G)-\mathbb{E}F(H)|} > e^{\lambda C}) \\
&\quad \leq e^{-\lambda C}\mathbb{E}e^{\lambda|F(G)-\mathbb{E}F(H)|} \leq 2e^{-\lambda C}\mathbb{E}e^{\lambda(F(G)-\mathbb{E}F(H))} \quad \text{, using } e^{\lambda|F|} \leq e^{\lambda F} + e^{-\lambda F} \\
&\quad \leq 2e^{-\lambda C}e^{\lambda^2\pi^2\sigma^2/8} = 2e^{-4C^2/\pi^2\sigma^2}e^{2C^2/\pi^2\sigma^2} = 2e^{-2C^2/\pi^2\sigma^2}.
\end{aligned}$$

■

Proposition 3.9. Let $\{A_j\}_{j=1}^k$ be independent events on a probability space Ω with $P(A_j) \geq 1/2$ and k even. Define A as the event

$$A := \left\{\sum_{j=1}^k 1_{A_j} \geq k/2\right\} = \left\{(\#A_j) \geq \frac{k}{2}\right\}.$$

Then $P(A) \geq 1/2$.

Proof: The idea here is to compare the case $P(A_j) \geq 1/2$ to the case $P(\tilde{A}_j) = 1/2$. To do this, we need to extend the probability space Ω. By extending Ω, let $\{y_j\}_{j=1}^k$ be independent random variables that are uniform on $[0,1]$. Define $B_j := \{y_j \leq P(A_j)\}$. By the uniformity of the y_j,

$$1_{B_j} = 1_{y_j \leq P(A_j)} = 1_{y_j^{-1}([0,P(A_j)])} \overset{d}{=} 1_{A_j}.$$

So, by independence [1] of the y_j and A_j (which implies independence of the $1_{y_j \leq P(A_j)}$ and the A_j),

$$\sum_j 1_{B_j} \stackrel{d}{=} \sum_j 1_{A_j}. \qquad (*)$$

We therefore conclude that

$$P((\#B_j) \geq k/2) = P((\#A_j) \geq k/2).$$

Now, define $C_j := \{y_j \leq 1/2\}$. Since $P(A_j) \geq 1/2$, $C_j \subseteq B_j$. So,

$$\{(\#C_j) \geq k/2\} \subseteq \{(\#B_j) \geq k/2\}. \qquad (**)$$

Since $P(C_j) = 1/2$ for all j, and the C_j are independent, $P\{(\#C_j) \geq k/2\} \geq 1/2$. One can see this, for example, by writing

$$\Omega = \bigcup_\ell (C_1^{\ell_1} \cap \cdots \cap C_k^{\ell_k}) =: \bigcup_\ell D_\ell,$$

where this disjoint union runs over all multi-indices ℓ with $\ell = (\ell_1, \ldots, \ell_k)$, and ℓ_i is either the complement operator ($C_i^{\ell_i} = C_i^c$), or the identity ($C_\ell^{\ell_i} = C_i$). Note that $P(D_\ell) = (1/2)^k$. Define $|\ell|$ as the number of ℓ_i which are identity operators.

Note that $D_\ell \subseteq \{(\#C_j) \geq k/2\}$ if and only if $|\ell| \geq k/2$. Now, by symmetry, for every $D_\ell = C_1^{\ell_1} \cap \cdots \cap C_k^{\ell_k}$ with $|\ell| > k/2$, there exists a unique ℓ' defined by $D_{\ell'} = C_1^{\ell_1 c} \cap \cdots \cap C_k^{\ell_k c}$, which satisfies $|\ell'| < k/2$. This association $\ell \leftrightarrow \ell'$ therefore partitions Ω (less sets with $|\ell| = k/2$) into two disjoint sets of equal measure: $\cup_{\ell\colon |\ell| > k/2} D_\ell$ and $\cup_{\ell\colon |\ell| < k/2} D_\ell$. Therefore,

$$\begin{aligned} P((\#C_j) \geq k/2) &= P(\cup_{\ell\colon |\ell| \geq k/2} D_\ell) \\ &= \frac{1}{2} P(\cup_{\ell\colon |\ell| \neq k/2} D_\ell) + P(\cup_{\ell\colon |\ell| = k/2} D_\ell) \geq 1/2. \end{aligned} \qquad (\dagger)$$

Combining $(*)$, $(**)$, and $(\dagger)$, we get $P(A) \geq 1/2$, as desired. ∎

References

[D] R. Durrett, Probability: theory and examples. Third edition. Duxbury Press, Belmont, CA, 2005.

[MS] V. Milman, and G. Schechtman, Asymptotic theory of finite-dimensional normed spaces. With an appendix by M. Gromov. Lecture Notes in Mathematics, 1200. Springer-Verlag, Berlin, 1986.

[S] G. Schechtman, *Two observations regarding embedding subsets of Euclidean spaces in normed spaces.* Adv. Math. **200** (2006), no. 1, 125135.

[T] M. Talagrand, *Embedding of ℓ_k^∞ and a theorem of Alon and Milman.* Geometric aspects of functional analysis (Israel, 19921994), 289293, Oper. Theory Adv. Appl., **77**, Birkhäuser, Basel, 1995.

[1] Suppose $a \stackrel{d}{=} c$, $b \stackrel{d}{=} d$, and a, b, c, d are independent. Then $a + b \stackrel{d}{=} c + d$.

$$\begin{aligned} P(a+b<\lambda) &= \cup_{\gamma\in\mathbb{Q}} P(a<\lambda-\gamma, b<\gamma) \\ &= \cup_{\gamma\in\mathbb{Q}} P(a<\lambda-\gamma)P(b<\gamma) \quad \text{, by independence} \\ &= \cup_{\gamma\in\mathbb{Q}} P(c<\lambda-\gamma)P(d<\gamma) \quad \text{, by equality of distributions} \\ &= \cup_{\gamma\in\mathbb{Q}} P(c<\lambda-\gamma, d<\gamma) \quad \text{, by independence} \\ &= P(c+d<\lambda) \end{aligned}$$