Gaussian concentration & Dvoretsky's theorem

In today's lecture we prove a concentration result for multidimensional Gaussian (normal) random variables, and show how this can be applied to prove quite deep results in geometric functional analysis. In particular, we prove the Johnson–Lindenstrauss lemma and Dvoretsky's theorem, both by means of Gaussian concentration.

In Section 1 we state and prove a concentration result for high-dimensional standard Gaussians. A first application is given in Section 2, where we give rough estimates on the expected ℓ_2-norm of a standard $N(0, I_n)$ Gaussian. Section 3 then deals with the Johnson–Lindenstrauss lemma, and finally in Section 4 we prove Dvoretsky's theorem.

This lecture draws highly from earlier lectures. It should be pointed out that there are more direct ways to prove the big theorems such as Johnson–Lindenstrauss, but the given proof does a nice job at intertwining the results and ideas developed so far throughout this course.

Furthermore, the author would like to point out that in filling in the details of the proof given at lecture, it was necessary to change a few things here and there, mostly to make to proofs simpler. The proofs are still similar in spirit, but the constants and the arguments deviate from time to time.

1 Gaussian concentration

High-dimensional geometry features a number of interesting concentration phenomena. One of this is concentration of Lipschitz functions around their expected value. We will prove this in Theorem 1.6 below.

We establish the following result as a prerequisite to Theorem 1.6.

Theorem 1.1 *Let $A \subseteq \mathbb{R}^n$ be non-empty and Lebesgue measurable. Then for all $\varepsilon > 0$ one has*

$$\gamma_n\big(\mathbb{R}^n \setminus (A + \varepsilon B_2^n)\big) \;\leq\; \frac{e^{-\frac{1}{4}\varepsilon^2}}{\gamma_n(A)}.$$

To prove this theorem, we use the following lemma.

Lemma 1.2 *If $A \subseteq \mathbb{R}^n$ is non-empty and Lebesgue measurable, then one has*

$$\int_A d\gamma_n(x) \int_{\mathbb{R}^n \setminus A} e^{\frac{d(y,A)^2}{4}}\, d\gamma_n(y) \;\leq\; 1,$$

where $d(y, A) := \inf_{a \in A} \|y - a\|_2$ is the (Euclidean) distance from y to A.

Note: since $\int_{\mathbb{R}^n} d\gamma_n(x) = 1$ (the Gaussian measure is a probability measure), clearly one has

$$\int_A d\gamma_n(x) \int_{\mathbb{R}^n \setminus A} d\gamma_n(y) \;\leq\; 1.$$

Lemma 1.2 can be interpreted as adding weights to make the result non-trivial.

Proof of Lemma 1.2: We shall use the Prékopa–Leindler inequality. Define $f, g, m : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ by

$$\begin{aligned} f(x) &:= \frac{d\gamma_n}{d\lambda_n}(x) \cdot \mathbb{1}[x \in A]; \\ g(y) &:= \frac{d\gamma_n}{d\lambda_n}(y) \cdot \mathbb{1}[y \notin A] \cdot e^{\frac{d(y,A)^2}{4}}; \\ m(z) &:= \frac{d\gamma_n}{d\lambda_n}(z). \end{aligned}$$

(Here $\frac{d\gamma_n}{d\lambda_n}(x) = \left(\frac{1}{\sqrt{2n}}\right)^n \cdot e^{-\frac{\|x\|_2^2}{2}}$ denotes the Radon–Nikodym derivative of the Gaussian measure γ_n with respect to the Lebesgue measure λ_n.) We show that for all $x, y \in \mathbb{R}^n$ one has

$$f(x)^{\frac{1}{2}} g(y)^{\frac{1}{2}} \leq m(\tfrac{1}{2}x + \tfrac{1}{2}y). \tag{1.3}$$

If $x \notin A$ or $y \in A$, then $f(x)^{\frac{1}{2}} g(y)^{\frac{1}{2}} = 0$, so (1.3) is trivially true. So assume $x \in A$ and $y \notin A$. Then one has $\|x - y\|_2 \geq d(y, A)$, hence by the parallelogram identity:

$$\|x\|_2^2 + \|y\|_2^2 = 2\left(\left\|\tfrac{x+y}{2}\right\|_2^2 + \left\|\tfrac{x-y}{2}\right\|_2^2\right) \geq 2\left(\left\|\tfrac{x+y}{2}\right\|_2^2 + \tfrac{d(y,A)^2}{4}\right) = 2\left\|\tfrac{1}{2}x + \tfrac{1}{2}y\right\|_2^2 + \tfrac{d(y,A)^2}{2}.$$

In this setting we have

$$\begin{aligned} f(x)^{\frac{1}{2}} g(y)^{\frac{1}{2}} &= \left(\frac{1}{\sqrt{2n}}\right)^{\frac{n}{2}} e^{-\frac{\|x\|_2^2}{4}} \cdot \left(\frac{1}{\sqrt{2n}}\right)^{\frac{n}{2}} e^{-\frac{\|y\|_2^2}{4} + \frac{d(y,A)^2}{8}} \\ &= \left(\frac{1}{\sqrt{2n}}\right)^{n} \cdot e^{-\frac{\|x\|_2^2}{4} - \frac{\|y\|_2^2}{4} + \frac{d(y,A)^2}{8}} \\ &\leq \left(\frac{1}{\sqrt{2n}}\right)^{n} \cdot e^{-\frac{\|\frac{1}{2}x + \frac{1}{2}y\|_2^2}{2}} \\ &= m(\tfrac{1}{2}x + \tfrac{1}{2}y). \end{aligned}$$

This shows that (1.3) holds for all $x, y \in \mathbb{R}^n$. Consequently, by the Prékopa–Leindler inequality, we have

$$\left(\int_{\mathbb{R}^n} f(x)\, dx\right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^n} g(y)\, dy\right)^{\frac{1}{2}} \leq \int_{\mathbb{R}^n} m(z)\, dz, \tag{1.4}$$

where

$$\begin{aligned} \int_{\mathbb{R}^n} f(x)\, dx &= \int_A d\gamma_n(x); \\ \int_{\mathbb{R}^n} g(y)\, dy &= \int_{\mathbb{R}^n \setminus A} e^{\frac{d(y,A)^2}{4}}\, d\gamma_n(y); \\ \int_{\mathbb{R}^n} m(z)\, dz &= \int_{\mathbb{R}^n} d\gamma_n(z) = 1. \end{aligned}$$

Taking squares in (1.4), we find

$$\int_A d\gamma_n(x) \int_{\mathbb{R}^n \setminus A} e^{\frac{d(y,A)^2}{4}}\, d\gamma_n(y) \leq 1. \qquad \square$$

Proof of Theorem 1.1: If $y \in \mathbb{R}^n \setminus (A + \varepsilon B_2^n)$ is given, then for all $x \in A$ we have $|x - y| > \varepsilon$, so we find $d(y, A) \geq \varepsilon$. Therefore it follows from Lemma 1.2 that

$$\begin{aligned} e^{\frac{\varepsilon^2}{4}} \cdot \gamma_n\big(\mathbb{R}^n \setminus (A + \varepsilon B_2^n)\big) &= \int_{\mathbb{R}^n \setminus (A+\varepsilon B_2^n)} e^{\frac{\varepsilon^2}{4}} \, d\gamma_n(y) \\ &\leq \int_{\mathbb{R}^n \setminus (A+\varepsilon B_2^n)} e^{\frac{d(y,A)^2}{4}} \, d\gamma_n(y) \\ &\leq \int_{\mathbb{R}^n \setminus A} e^{\frac{d(y,A)^2}{4}} \, d\gamma_n(y) \\ &\leq \frac{1}{\gamma_n(A)}. \end{aligned}$$

□

Definition 1.5 *For $L \in \mathbb{R}_{>0}$, we say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is L-*Lipschitz *if for all $x, y \in \mathbb{R}^n$ one has $|f(x) - f(y)| \leq L \cdot \|x - y\|_2$.*

Theorem 1.6 (Gaussian Concentration for Lipschitz functions) *There is an absolute constant $c > 0$ with the following property: for any $n \in \mathbb{N}_1$ and $L > 0$, if $f : \mathbb{R}^n \to \mathbb{R}$ is L-Lipschitz, then for all $\varepsilon > 0$ one has*

$$\Pr_{z \sim N(0, I_n)} \big[\, |f(z) - \mu| \geq \varepsilon L \,\big] \leq 4 \cdot e^{-c\varepsilon^2}, \tag{1.7}$$

where

$$\mu := \mathop{\mathbb{E}}_{z \sim N(0, I_n)} f(z).$$

Note that one has to check that f has an expected value, i.e. that $\int_{\mathbb{R}^n} |f(x)| \, d\gamma_n(x)$ is finite. We claim that this is always the case if f is Lipschitz; this is left as an exercise to the reader.

Proof of Theorem 1.6: Let $m \in \mathbb{R}$ be a *median* of f, that is, a real number m satisfying

$$\Pr_{z \sim N(0, I_n)} \big[f(z) \geq m\big] \geq \frac{1}{2}, \qquad \text{and} \qquad \Pr_{z \sim N(0, I_n)} \big[f(z) \leq m\big] \geq \frac{1}{2}.$$

(Again, one has to check that such an m exists. We leave it as an exercise to show that any real-valued random variable has a median.) Write $A := \{x \in \mathbb{R}^n : f(x) \leq m\}$, then one has

$$\begin{aligned} \Pr_{z \sim N(0, I_n)} \big[f(z) - m \geq \varepsilon L\big] &\leq \Pr_{z \sim N(0, I_n)} \big[z \in \mathbb{R}^n \setminus (A + \varepsilon B_2^n)\big] && \text{(since } f \text{ is } L\text{-Lipschitz)} \\ &\leq \frac{e^{-\frac{\varepsilon^2}{4}}}{\gamma_n(A)} && \text{(by Theorem 1.1)} \\ &\leq 2 \cdot e^{-\frac{\varepsilon^2}{4}}. && \text{(since } \gamma_n(A) \geq \tfrac{1}{2}\text{)} \end{aligned}$$

One similarly shows that $\Pr_{z \sim N(0, I_n)} \big[f(z) - m \leq -\varepsilon L\big] \leq 2 \cdot e^{-\frac{\varepsilon^2}{4}}$, so we have

$$\Pr_{z \sim N(0, I_n)} \big[\, |f(z) - m| \geq \varepsilon L \,\big] \leq 4 \cdot e^{-\frac{\varepsilon^2}{4}}.$$

We will now show that the mean μ and the median m are, in a sense, not too far apart. Indeed, we have

$$\begin{aligned}
|\mu - m| \;=\; \left| \mathop{\mathbb{E}}_{z\sim N(0,I_n)} \big[f(z) - m \big] \right| \;&\leq\; \mathop{\mathbb{E}}_{z\sim N(0,I_n)} \big[\, |f(z) - m| \,\big] \\
&= \int_0^\infty \mathop{\Pr}_{z\sim N(0,I_n)} \big[\, |f(z) - m| \geq t \,\big]\, dt \\
&\leq \int_0^\infty 4 \cdot e^{-(\frac{t}{L})^2/4}\, dt \\
&= 4L\sqrt{2} \int_0^\infty e^{-s^2/2}\, ds \qquad\qquad \text{(substituted } s = \tfrac{t}{L\sqrt{2}}\text{)} \\
&= 4L\sqrt{2} \cdot \sqrt{2\pi} \cdot \gamma_1(\mathbb{R}_{\geq 0}) \\
&= 4L\sqrt{\pi}. \qquad\qquad \text{(since } \gamma_1(\mathbb{R}_{\geq 0}) = \tfrac{1}{2}\text{)}
\end{aligned}$$

Write $K := 4\sqrt{\pi}$, so we have $|\mu - m| \leq KL$. Now we claim that the constant $c := \frac{1}{4} \cdot \frac{1}{(K+1)^2}$ suffices to prove (1.7). We distinguish two cases.

- For $\varepsilon \leq K+1$, we note that $\frac{1}{4} < 1 < \ln(4)$ holds, so we have

  $$4 \cdot e^{-c\varepsilon^2} \;=\; 4 \cdot e^{-\frac{1}{4}\cdot(\frac{\varepsilon}{K+1})^2} \;>\; 4 \cdot e^{-\ln(4)\cdot 1^2} \;=\; 1.$$

  Therefore (1.7) is trivially satisfied in this case.

- If $\varepsilon > K+1$, then we have

  $$0 \;<\; \frac{1}{2(K+1)} \;=\; \frac{1}{2} - \frac{K}{2(K+1)} \;<\; \frac{1}{2} - \frac{K}{2\varepsilon},$$

  and therefore

  $$\frac{\frac{(\varepsilon-K)^2}{4}}{\varepsilon^2} \;=\; \left(\frac{1}{2} - \frac{K}{2\varepsilon}\right)^2 \;>\; \left(\frac{1}{2(K+1)}\right)^2 \;=\; c.$$

  It follows that $-\frac{(\varepsilon-K)^2}{4} < -c\varepsilon^2$. Using the (reverse) triangle inequality, we find

  $$|f(z) - m| \;\geq\; |f(z) - \mu| - |\mu - m| \;\geq\; |f(z) - \mu| - KL,$$

  so it follows from the above that

  $$\mathop{\Pr}_{z\sim N(0,I_n)} \big[\, |f(z) - \mu| \geq \varepsilon L \,\big] \;\leq\; \mathop{\Pr}_{z\sim N(0,I_n)} \big[\, |f(z) - m| \geq (\varepsilon - K)L \,\big] \;\leq\; 4 \cdot e^{-\frac{(\varepsilon-K)^2}{4}} \;<\; 4 \cdot e^{-\frac{\varepsilon^2}{4}}.$$

In either case, we find that (1.7) holds, where c is an absolute constant. □

Remark 1.8 We give a rough estimate of the constant c from Theorem 1.6. In the proof, we showed that $c := \frac{1}{4} \cdot \frac{1}{(K+1)^2}$ suffices, where $K = 4\sqrt{\pi}$. A crude estimate gives $K + 1 < 10$, so we may take $c > \frac{1}{400}$.

2 An application: expected norm of a normal random vector

As a simple application of the results from Section 1, we show how to use Gaussian concentration to give an estimate of the expected ℓ_2-norm of a random variable $z \sim N(0, I_n)$. This is given in the "lazy man's approximation" in Proposition 2.2 below. (We call it so because there are better estimates available if one is willing to do some more work. However, the result gives a nice illustration of the power of concentration, giving us results with very little additional effort.)

The approximation given in Proposition 2.2 is not just an interesting application, we shall also use it in our proof of the Johnson–Lindenstrauss lemma in Section 3 below.

Before stating and proving Proposition 2.2, we need the following auxiliary result.

Proposition 2.1 *Let $c > 0$ be the constant from Theorem 1.6. Then for any $n \in \mathbb{N}_1$ and any L-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ one has*

$$\operatorname*{Var}_{z\sim N(0,I_n)} f(z) \;\leq\; \frac{4L^2}{c}.$$

Proof: We use both the result and the method of proof from Theorem 1.6. Write $\mu := \mathop{\mathbb{E}}_{z\sim N(0,I_n)} f(z)$, then we have

$$\begin{aligned}
\operatorname*{Var}_{z\sim N(0,I_n)} f(z) \;=\; \mathop{\mathbb{E}}_{z\sim N(0,I_n)} \left[\left(f(z)-\mu\right)^2 \right] \;&=\; \int_0^\infty \Pr\left[(f(z)-\mu)^2 \geq t \right] dt \\
&=\; \int_0^\infty \Pr\left[\, |f(z)-\mu| \geq \sqrt{t}\,\right] dt \\
&\leq\; \int_0^\infty 4\cdot e^{-c\left(\frac{\sqrt{t}}{L}\right)^2} dt \qquad \text{(by Theorem 1.6)} \\
&=\; \int_0^\infty 4\cdot e^{-\frac{c}{L^2}t}\, dt \\
&=\; \left[-\frac{4L^2}{c}\cdot e^{-\frac{c}{L^2}t} \right]_{t=0}^{\infty} \\
&=\; 0 - \left(-\frac{4L^2}{c}\cdot 1 \right) \\
&=\; \frac{4L^2}{c}.
\end{aligned}$$

□

Proposition 2.2 (Lazy man's approximation) *There is an absolute constant c' such that for all $n \in \mathbb{N}_1$ one has*

$$\sqrt{n} - \frac{c'}{\sqrt{n}} \;\leq\; \mathop{\mathbb{E}}_{z\sim N(0,I_n)} \|z\|_2 \;\leq\; \sqrt{n}.$$

Proof: Write $\mu := \mathop{\mathbb{E}}_{z\sim N(0,I_n)} \|z\|_2$. By linearity of expectation, we have

$$\mathop{\mathbb{E}}_{z\sim N(0,I_n)} \|z\|_2^2 \;=\; \mathop{\mathbb{E}}_{z\sim N(0,I_n)} \left[\sum_{i=1}^n z_i^2 \right] \;=\; \sum_{i=1}^n \left(\mathop{\mathbb{E}}_{z\sim N(0,I_n)} z_i^2 \right) \;=\; \sum_{i=1}^n 1 \;=\; n,$$

so it follows from the Cauchy–Schwarz inequality that

$$\begin{aligned}
\mu \;=\; \mathop{\mathbb{E}}_{z\sim N(0,I_n)} \|z\|_2 \;&=\; \int_{\mathbb{R}^n} \|z\|_2 \; d\gamma_n(z) \\
&=\; \left| \int_{\mathbb{R}^n} \|z\|_2 \cdot 1 \; d\gamma_n(z) \right| \\
&\leq\; \left(\int_{\mathbb{R}^n} \|z\|_2^2 \; d\gamma_n(z) \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^n} 1^2 \; d\gamma_n(z) \right)^{\frac{1}{2}} \\
&=\; \sqrt{\mathop{\mathbb{E}}_{z\sim N(0,I_n)} \|z\|_2^2} \cdot \sqrt{1} \\
&=\; \sqrt{n}.
\end{aligned}$$

Let $c > 0$ be the constant from Theorem 1.6. Since $\|\cdot\|_2$ is 1-Lipschitz (clearly), it follows from Proposition 2.1 that

$$n - \mu^2 \;=\; \mathop{\mathbb{E}}_{z\sim N(0,I_n)} \|z\|_2^2 - \left(\mathop{\mathbb{E}}_{z\sim N(0,I_n)} \|z\|_2 \right)^2 \;=\; \mathop{\mathrm{Var}}_{z\sim N(0,I_n)} \|z\|_2 \;\leq\; \frac{4}{c}.$$

Now set $c' := \frac{4}{c}$. We claim that this constant suffices. To that end, we distinguish two cases.

- If $0 < n < c'$, then we have $\sqrt{n} - \frac{c'}{\sqrt{n}} = \frac{n-c'}{\sqrt{n}} < 0 \leq \mu$, and we are done.
- If $n \geq c'$, then by the above we have $\sqrt{n-c'} \leq \mu$, hence

$$\mu \;\geq\; \sqrt{n-c'} \;>\; \sqrt{n-c'} \cdot \frac{\sqrt{n-c'}}{\sqrt{n}} \;=\; \frac{n-c'}{\sqrt{n}} \;=\; \sqrt{n} - \frac{c'}{\sqrt{n}}. \qquad \square$$

Corollary 2.3 *If $\lambda \in \mathbb{R}_{>0}$ and $z \sim N(0, \lambda I_n)$, then one has*

$$\sqrt{\lambda n} - c'\sqrt{\frac{\lambda}{n}} \;\leq\; \mathbb{E}\,\|z\|_2 \;\leq\; \sqrt{\lambda n},$$

where c' is the constant from Proposition 2.2.

Proof: If $y \sim N(0, I_n)$, then z and $\sqrt{\lambda} \cdot y$ are identically distributed, so the result follows from Proposition 2.2. $\square$

Remark 2.4 In Remark 1.8, we gave a rude estimate of $c > \frac{1}{400}$ on the constant of Theorem 1.6. Consequently, in the proof of Proposition 2.2 we set $c' := \frac{4}{c}$, so our crude estimate becomes $c' < 1600$. (Even though not much care was taken to find optimal constants, we nevertheless see that the constants are not astronomical.)

3 The Johnson–Lindenstrauss lemma

We now come to one of the main theorems of this lecture (and indeed this course): the Johnson–Lindenstrauss lemma, which states that we can represent a finite set of vectors in high-dimensional Euclidean space by vectors in a relatively low-dimensional (Euclidean) space via a linear map which approximately preserves the distances between these points.

The proof given here relies only on Gaussian concentration, though this is certainly not the only (or even the "standard") way to prove it.

THEOREM 3.1 (THE JOHNSON–LINDENSTRAUSS LEMMA) *For $n, m \in \mathbb{N}_1$, let $v_1, \ldots, v_n \in \mathbb{R}^m$ and $\varepsilon \in (0,1)$ be fixed. Then there exist $d \in \mathbb{N}_1$ and a linear map $T : \mathbb{R}^m \to \mathbb{R}^d$ such that*

(i) For all $i, j \in [n]$ one has

$$(1-\varepsilon) \cdot \|v_i - v_j\|_2 \;\leq\; \|Tv_i - Tv_j\|_2 \;\leq\; (1+\varepsilon) \cdot \|v_i - v_j\|_2.$$

(ii) $d = O\big(\frac{\log(n)}{\varepsilon^2}\big)$.

PROOF: Let $c \in \mathbb{R}_{>0}$ be the constant obtained from Theorem 1.6. Set $d := \lceil \frac{12\ln(2n)}{c\varepsilon^2} \rceil$, and note that we have $d > 0$ and therefore $d \geq 1$. Let T be a random $d \times m$ matrix whose entries are i.i.d. $N(0, \frac{1}{d})$ random variables. We show that T satisfies (i) with positive probability.

Note that for arbitrary $w \in \mathbb{R}^m$ we have $(1-\varepsilon)\|w\|_2 \leq \|Tw\|_2 \leq (1+\varepsilon)\|w\|_2$ if and only if $\big|\, \|Tw\|_2 - \|w\|_2 \,\big| \leq \varepsilon\|w\|_2$. We claim that for fixed $w \in \mathbb{R}^m$ one has

$$\Pr_T\Big[\,\big|\, \|Tw\|_2 - \|w\|_2 \,\big| > \varepsilon\|w\|_2\Big] \;\leq\; 4e^{-\frac{c\varepsilon^2 d}{4}}, \tag{3.2}$$

For $w = 0$ this is trivially true, for we have $\Pr[0 > 0] = 0$, so we assume henceforth that $w \neq 0$ holds. By homogeneity, we may assume without loss of generality that $\|w\|_2 = \sqrt{d}$ holds. Note that the i-th coordinate of Tw is given by

$$(Tw)_i \;=\; T_{i1}w_1 + \cdots + T_{im}w_m,$$

so it is a linear combination of i.i.d. $N(0, \frac{1}{d})$ random variables $T_{i1}, \ldots, T_{im}$. As such, $(Tw)_i$ is again normally distributed, with mean 0 and variance $\frac{w_1^2}{d} + \cdots + \frac{w_m^2}{d} = \frac{\|w\|_2^2}{d} = \frac{d}{d} = 1$. Since the rows of T are independent, we conclude that Tw is distributed as $N(0, I_d)$.

In the proof of Proposition 2.2, we showed that

$$\mu \;:=\; \mathop{\mathbb{E}}_T \|Tw\|_2 \;=\; \mathop{\mathbb{E}}_{z \sim N(0, I_d)} \|z\|_2 \;\in\; \left[\sqrt{d} - \frac{4}{c\sqrt{d}}\,,\, \sqrt{d}\right].$$

(The point here is that we showed that $c' := \frac{4}{c}$ suffices to prove Proposition 2.2. Using the constant $\frac{4}{c}$ instead of a second constant c' makes the analysis slightly easier here.)

As a consequence, we have

$$\begin{aligned} 0 \;\leq\; \sqrt{d} - \mu \;&\leq\; \frac{4}{c\sqrt{d}} \\ &=\; \frac{4}{cd} \cdot \sqrt{d} \end{aligned}$$

$$
\begin{aligned}
&\leq \frac{4}{c} \cdot \frac{c\varepsilon^2}{12\ln(2n)} \cdot \sqrt{d} \\
&< \frac{\varepsilon^2}{2} \cdot \sqrt{d} && \text{(since } 12\ln(2n) > 8 \text{ whenever } n \geq 1\text{)} \\
&< \frac{\varepsilon}{2} \cdot \sqrt{d}. && \text{(since } \varepsilon < 1\text{)}
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\Pr_T \left[\left| \, \|Tw\|_2 - \|w\|_2 \, \right| > \varepsilon \|w\|_2 \right] &= \Pr_T \left[\left| \, \|Tw\|_2 - \sqrt{d} \, \right| > \varepsilon\sqrt{d} \right] \\
&\leq \Pr_T \left[\left| \, \|Tw\|_2 - \mu \, \right| > \varepsilon\sqrt{d} - \left(\sqrt{d} - \mu\right) \right] \\
&\leq \Pr_T \left[\left| \, \|Tw\|_2 - \mu \, \right| \geq \frac{\varepsilon}{2} \cdot \sqrt{d} \right] \\
&\leq 4 \cdot e^{-\frac{c\varepsilon^2 d}{4}},
\end{aligned}
$$

where in the last step we used that $Tw \sim N(0, I_d)$ and that $\|\cdot\|_2$ is 1-Lipschitz, so that we may apply Theorem 1.6. This proves our claim (3.2).

Now, to prove the theorem, we take a union bound over the n^2 events

$$
\left\{ \left\{ \, \left| \, \|Tv_i - Tv_j\|_2 - \|v_i - v_j\|_2 \, \right| > \varepsilon \|v_i - v_j\|_2 \right\} \, : \, i, j \in [n] \right\},
$$

from which it follows that

$$
\begin{aligned}
\Pr_T \left[\, T \text{ does not satisfy (i)} \, \right] &\leq n^2 \cdot 4 \cdot e^{-\frac{c\varepsilon^2 d}{4}} \\
&\leq n^2 \cdot 4 \cdot e^{-\frac{c\varepsilon^2}{4} \cdot \frac{12\ln(2n)}{c\varepsilon^2}} \\
&= (2n)^2 \cdot e^{-3\ln(2n)} \\
&= \frac{(2n)^2}{(2n)^3} \\
&= \frac{1}{2n} \\
&< 1.
\end{aligned}
$$

It follows that T satisfies (i) with positive probability, so in particular there exists an instance of T which satisfies (i). Clearly d is $O(\frac{\log(n)}{\varepsilon^2})$, so (ii) is also met. □

Remark 3.3 Note that one can always find a linear map $T : \mathbb{R}^m \to \mathbb{R}^d$ satisfying property (i) of Theorem 3.1 if $d \geq n$: simply choose an isometric embedding of $\text{span}(v_1, \ldots, v_n)$ into $\mathbb{R}^d$. It follows that the result of Theorem 3.1 is trivial whenever ε is very small, say $\varepsilon \leq \frac{1}{\sqrt{n}}$, for then we have $\frac{\log(n)}{\varepsilon^2} \geq n\log(n)$. The most interesting case, therefore, is if ε is in the intermediate range between $\frac{1}{\sqrt{n}}$ and 1.

4 Dvoretsky's theorem

The final topic of this lecture is Dvoretsky's theorem, which is also the deepest result of the lecture. Recall that the Banach–Mazur distance measures how closely two symmetric convex bodies resemble one another. It was shown that $d(\ell_2^n, \ell_\infty^n) = \sqrt{n}$, so in this sense the hypersphere and the hypercube are quite far apart. Dvoretsky's theorem, rather remarkably, shows that *any* symmetric convex $K \subseteq \mathbb{R}^n$ body is close to being an ellipsoid on some subspace $L \subseteq \mathbb{R}^n$ of relatively large dimension. The precise statement is as follows.

Theorem 4.1 (Dvoretsky's theorem, geometric version) *For every $r \in \mathbb{R}_{>0}$, there is a constant $C_r > 0$ with the following property. For any $n \in \mathbb{N}_1$, any symmetric convex body $K \subseteq \mathbb{R}^n$, and any $\varepsilon \in (0, r]$, there exists a linear subspace $W \subseteq \mathbb{R}^n$ and a full-dimensional ellipsoid $E \subseteq W$ such that $\dim(W) \geq C_r \cdot \frac{\varepsilon^2 \log(n)}{\log(1+\frac{1}{\varepsilon})}$ and $E \subseteq K \cap W \subseteq (1+\varepsilon)E$.*

The theorem can be reformulated in the language of functional analysis. To that end, we introduce the following notation: if X, Z are normed spaces, then we write $Z \xhookrightarrow{1+\varepsilon} X$ to denote that there exist a subspace $Y \subseteq X$ and an invertible linear map $T : Z \to Y$ with $\|T\| \cdot \|T^{-1}\| \leq 1 + \varepsilon$. (Note that T is only injective when viewed as a map $Z \to X$.)

Theorem 4.2 (Dvoretsky's theorem, functional analytic version) *For every $r \in \mathbb{R}_{>0}$, there is a constant $C_r > 0$ with the following property. For any $n \in \mathbb{N}_1$, any n-dimensional normed space X, and any $\varepsilon \in (0, r]$, there is some $d \geq C_r \cdot \frac{\varepsilon^2 \log(n)}{\log(1+\frac{1}{\varepsilon})}$ such that $\ell_2^d \xhookrightarrow{1+\varepsilon} X$.*

The proof of Dvoretsky's theorem comes in two parts. First, we prove a special version for the case where the norm of X is 1-Lipschitz (the "Dvoretsky criterion", see Subsection 4.1). After that, an assortment of geometric and combinatorial arguments is used to solve the general case (see Subsection 4.2). The lecture is concluded with a few closing remarks in Subsection 4.3.

Before stating and proving the Dvoretsky criterion, we give a simple characterisation of norms which are 1-Lipschitz.

Proposition 4.3 *Let $K \subseteq \mathbb{R}^n$ be a symmetric convex body. Then $\|\cdot\|_K$ is 1-Lipschitz if and only if $B_2^n \subseteq K$.*

Proof: Note that we have $B_2^n \subseteq K$ if and only if $\|x\|_K \leq \|x\|_2$ for all $x \in \mathbb{R}^n$. If this is the case, then we have $|\|x\|_K - \|y\|_K| \leq \|x - y\|_K \leq \|x - y\|_2$ for all $x, y \in \mathbb{R}^n$, so $\|\cdot\|_K$ is 1-Lipschitz. Conversely, suppose that $\|\cdot\|_K$ is 1-Lipschitz. Then in particular for all $x \in \mathbb{R}^n$ we have $|\|x\|_K - \|0\|_K| \leq \|x - 0\|_2$, hence $\|x\|_K \leq \|x\|_2$. □

4.1 The Dvoretsky criterion

We are now ready to state and prove the Dvoretsky criterion.

Lemma 4.4 (Dvoretsky criterion) *For every $r \in \mathbb{R}_{>0}$, there is a constant $C'_r > 0$ with the following property. Let $X = (\mathbb{R}^n, \|\cdot\|_X)$ be a normed space such that $\|\cdot\|_K$ is 1-Lipschitz, and let $\varepsilon \in (0, r]$ be given. Write $\mu := \mathop{\mathbb{E}}_{z \sim N(0, I_n)} \|z\|_X$. Then there exists some $d \geq C'_r \cdot \frac{\varepsilon^2 \mu^2}{\log(1+\frac{1}{\varepsilon})}$ such that $\ell_2^d \xhookrightarrow{1+\varepsilon} X$.*

Before we proceed with the proof, let us recall the following lemmas from lecture 7:

LEMMA 4.5 *Let X and Y be finite-dimensional normed spaces, and let $T \subseteq \partial B_X$ be an ε-net for ∂B_X with the additional property that $B_X \subseteq \frac{1}{1-\varepsilon}\operatorname{conv}(\pm T)$. Then for any linear map $M : X \to Y$ one has*

$$\|M\|_{X\to Y} \leq \frac{1}{1-\varepsilon} \sup_{t\in T} \|Mt\|_Y.$$

If additionally $\dim(X) = \dim(Y)$ *and* $-\varepsilon\|M\|_{X\to Y} + \inf_{t\in T}\|Mt\|_Y > 0$, *then M is invertible and*

$$\|M^{-1}\|_{Y\to X} \leq \frac{1}{-\varepsilon\|M\|_{X\to Y} + \inf_{t\in T}\|Mt\|_Y}.$$

(Of course, if T is finite, then the supremum/infimum can be replaced by a maximum/minimum.)

LEMMA 4.6 *Let $B \subseteq \mathbb{R}^n$ be a symmetric convex body. Then there exists a finite set $T \subseteq \partial B$ with $|T| \leq N^\circ(B, \frac{\varepsilon}{2}B) \leq (1+\frac{4}{\varepsilon})^n$ such that T is an ε-net for ∂B and* $\operatorname{conv}(\pm T) \subseteq B \subseteq \frac{1}{1-\frac{\varepsilon}{2}}\operatorname{conv}(\pm T)$.

We are now ready to prove the Dvoretsky criterion.

PROOF OF LEMMA 4.4: Define $\alpha_r \in \mathbb{R}_{>0}$ by $\alpha_r := \frac{1}{r+2}$. One easily shows that the following inequalities hold for all $\varepsilon \in (0, r]$:

$$1 - \tfrac{\alpha_r}{4}\varepsilon > 1 - \alpha_r\varepsilon > 0;$$

$$\frac{1+\alpha_r\varepsilon}{1-\alpha_r\varepsilon} \leq 1+\varepsilon; \tag{4.7}$$

$$\frac{1}{1-\frac{\alpha_r}{4}\varepsilon} < 1 + \tfrac{\alpha_r}{2}\varepsilon; \tag{4.8}$$

$$(1+\tfrac{\alpha_r}{2}\varepsilon)(1+\tfrac{\alpha_r}{4}\varepsilon) < 1+\alpha_r\varepsilon; \tag{4.9}$$

$$-\tfrac{\alpha_r}{4}\varepsilon(1+\alpha_r\varepsilon) + 1 - \tfrac{\alpha_r}{4}\varepsilon > 1-\alpha_r\varepsilon > 0. \tag{4.10}$$

To illustrate, we prove (4.7). Note that we have

$$\begin{aligned}
(1+\varepsilon)(1-\alpha_r\varepsilon) &= 1+\varepsilon-\alpha_r\varepsilon-\alpha_r\varepsilon^2 \\
&\geq 1+\varepsilon-\alpha_r\varepsilon-\alpha_r r\varepsilon && \text{(since } \varepsilon \leq r\text{)} \\
&= 1+\big(\alpha_r(r+2)-\alpha_r(r+1)\big)\varepsilon && \text{(since } \alpha_r = \tfrac{1}{r+2}\text{)} \\
&= 1+\alpha_r\varepsilon.
\end{aligned}$$

(The inequalities (4.8), (4.9) and (4.10) are even simpler – expand the expression and get rid of the terms involving ε^2 by using the estimate $\alpha_r\varepsilon^2 \leq \alpha_r r\varepsilon = \frac{r}{r+2}\varepsilon < \varepsilon$.)

Choose $d := \max(1, \lceil \frac{c\alpha_r^2\varepsilon^2\mu^2 - 16\ln(4)}{16\ln(1+\frac{16}{\alpha_r\varepsilon})} \rceil - 1)$, where c is the constant from Theorem 1.6. It is routine (but nevertheless quite challenging) to show that there exists a constant $C_r' > 0$ depending only on r such that $d \geq C_r' \cdot \frac{\varepsilon^2\mu^2}{\ln(1+\frac{1}{\varepsilon})}$ for all $\mu > 0$ and all $\varepsilon \in (0, r]$. We leave this as an exercise to the reader.[1]

[1]Several pages of estimates in my scratch pad show that we may take $C_r' = \frac{c\ln(1+\frac{1}{r})}{32(r+2)^2\ln(2+32\frac{r+2}{r})}$, where c is the constant from Theorem 1.6. For $r = 1$, this yields $C_1' > \frac{c}{2000} > \frac{1}{800\,000}$.

We show that $\ell_2^d \overset{1+\varepsilon}{\hookrightarrow} X$. If $d = 1$, then for any injective linear map $f : \mathbb{R} \to X$ there is some $\lambda \in \mathbb{R}$ such that λf is an isometry, so clearly $\ell_2^1 \overset{1}{\hookrightarrow} X$. Assume henceforth that we are in the situation $d > 1$. Then we have $d < \frac{c\alpha_r^2\varepsilon^2\mu^2 - 16\ln(4)}{16\ln(1+\frac{16}{\alpha_r\varepsilon})}$, so $d\ln(1+\frac{16}{\alpha_r\varepsilon}) + \ln(4) - c(\frac{\alpha_r}{4}\varepsilon\mu)^2 < 0$, and it follows that

$$\left(1+\tfrac{16}{\alpha_r\varepsilon}\right)^d \cdot 4 \cdot e^{-c(\frac{\alpha_r}{4}\varepsilon\mu)^2} \;<\; 1. \tag{4.11}$$

Let T be a random $n \times d$ matrix whose entries are i.i.d. $N(0,1)$ standard normal random variables. We show that T is a map $\ell_2^d \overset{1+\varepsilon}{\hookrightarrow} X$ with high probability.

For fixed $y \in S^{d-1}$, we note that Ty is distributed as $N(0, I_n)$, analogously to what we had in the proof of the Johnson–Lindenstrauss theorem (Theorem 3.1). As such, for every $y \in S^{d-1}$ we have

$$\Pr_T\left[\,\big|\|Ty\|_X - \mu\big| > \delta\,\right] \;=\; \Pr_{z\sim N(0,I_n)}\left[\,\big|\|z\|_X - \mu\big| > \delta\,\right] \;\le\; 4\cdot e^{-c\delta^2}, \tag{4.12}$$

by Theorem 1.6 (Gaussian concentration).

By Lemma 4.6 we may choose an $(\frac{\alpha_r}{4}\varepsilon)$-net $N_{r,\varepsilon}$ of S^{d-1} such that $|N_{r,\varepsilon}| \le (1+\frac{16}{\alpha_r\varepsilon})^d$ and $B_2^d \subseteq \frac{1}{1-\frac{\alpha_r}{4}\varepsilon}\operatorname{conv}(\pm N_{r,\varepsilon})$. We shall say that T is *good* if for all $x \in N_{r,\varepsilon}$ one has $\big|\|Tx\|_X - \mu\big| \le \frac{\alpha_r}{4}\varepsilon\mu$. By taking a union bound, it follows from (4.11) and (4.12) that

$$\Pr\left[T \text{ is not good}\right] \;\le\; \left(1+\tfrac{16}{\alpha_r\varepsilon}\right)^d \cdot 4 \cdot e^{-c(\frac{\alpha_r}{4}\varepsilon\mu)^2} \;<\; 1.$$

In other words, T is good with positive probability. Choose such a T, and consider it as an operator $\ell_2^d \to Y := \operatorname{im}(T) \subseteq X$. Then, by Lemma 4.5, we have

$$\begin{aligned}
\|T\|_{\ell_2^d\to Y} \;&\le\; \frac{1}{1-\frac{\alpha_r}{4}\varepsilon}\cdot\max_{x\in N_{r,\varepsilon}}\|Tx\|_X \\
&\le\; \frac{1}{1-\frac{\alpha_r}{4}\varepsilon}\cdot(1+\tfrac{\alpha_r}{4}\varepsilon)\mu \\
&\le\; (1+\tfrac{\alpha_r}{2}\varepsilon)(1+\tfrac{\alpha_r}{4}\varepsilon)\mu && \text{(by (4.8))} \\
&<\; (1+\alpha_r\varepsilon)\mu. && \text{(by (4.9))}
\end{aligned}$$

Furthermore, by the above and (4.10) we have

$$-\tfrac{\alpha_r}{4}\varepsilon\|T\|_{\ell_2^d\to Y} + \min_{x\in N_{r,\varepsilon}}\|Tx\|_X \;>\; -\tfrac{\alpha_r}{4}\varepsilon(1+\alpha_r\varepsilon)\mu + (1-\tfrac{\alpha_r}{4}\varepsilon)\mu \;>\; (1-\alpha_r\varepsilon)\mu \;>\; 0,$$

so it follows from the second part of Lemma 4.5 that T is invertible with

$$\|T^{-1}\|_{Y\to\ell_2^d} \;\le\; \frac{1}{-\frac{\alpha_r}{4}\varepsilon\|T\|_{\ell_2^d\to Y} + \min_{x\in N_{r,\varepsilon}}\|Tx\|_X} \;<\; \frac{1}{(1-\alpha_r\varepsilon)\mu}.$$

By the preceding inequalities and (4.7), we have $\|T\|_{\ell_2^d\to Y}\cdot\|T^{-1}\|_{Y\to\ell_2^d} < 1+\varepsilon$. □

Remark 4.13 At the end of the preceding proof, we note *a posteriori* that $d \le n$ must hold, since $T : \ell_2^d \to Y \subseteq X$ is invertible. This is not at all clear from the definition of d, especially since it depends on the constant c from Gaussian concentration. We claim that, in fact, this proof can be used to give an upper bound on the (optimal) constant c for which Theorem 1.6 holds, but we leave the details to the (interested) reader.

4.2 Completing the proof of Dvoretsky's theorem

We proceed to prove the functional analytic version of Dvoretsky's theorem (Theorem 4.2). Note that this theorem is stated for an abstract normed space, whereas the Dvoretsky criterion (Lemma 4.4) is formulated for a "concrete" normed space $X = (\mathbb{R}^n, \|\cdot\|_X)$. Thus, in order to prove Dvoretsky's theorem, we will assume that X is an abstract normed space, and our goal will be to find an isomorphism $M : X \to \mathbb{R}^n$ such that:

(i) $\|\cdot\|_{MB_X}$ is 1-Lipschitz (i.e. $B_2^n \subseteq MB_X$);

(ii) $\mu := \mathop{\mathbb{E}}_{z\sim N(0,I_n)} \left[\|z\|_{MB_X} \right]$ is $\Omega(\sqrt{\log(n)})$.

We shall do so in the following way.

Proposition 4.14 *There exists an invertible linear map $M : X \to \mathbb{R}^n$ such that B_2^n is the John ellipsoid of MB_X.*

Proof: Let $E \subseteq X$ be the John ellipsoid of B_X, and choose M such that $M^{-1}B_2^n = E$. Then it is easy to see that $B_2^n = ME$ is the John ellipsoid of B_X. □

This simple proposition already gets us halfway towards our goal: if we choose M like this, then we get (i) for free. We will show that this construction also gives us (ii). We use the following combinatorial proposition.

Proposition 4.15 *Let $y_1, \ldots, y_N \in S^{n-1}$ and $\lambda_1, \ldots, \lambda_N \in \mathbb{R}_{\geq 0}$ be such that $I_n = \sum_{j=1}^n \lambda_j y_j y_j^\top$. Write $Q := \mathrm{conv}(\pm y_1, \ldots, \pm y_N)$. Then $N(Q, \frac{1}{4}B_2^n) \geq P(Q, \frac{1}{4}B_2^n) > \frac{n}{2}$.*

Proof: The first inequality follows from general packing bounds (where we use that $\frac{1}{4}B_2^n$ is symmetric). We prove $P(Q, \frac{1}{4}B_2^n) \geq \frac{n}{2}$. Note that each of the following statements implies the previous and the next one:

$$\begin{aligned}
(y_i + \tfrac{1}{4}B_2^n) \cap (y_j + \tfrac{1}{4}B_2^n) &\neq \varnothing; \\
\|y_i - y_j\|_2 &\leq \tfrac{1}{2}; \\
\|y_i - y_j\|_2^2 &\leq \tfrac{1}{4}; \\
2 - 2\langle y_i, y_j\rangle &\leq \tfrac{1}{4}; \\
\langle y_i, y_j\rangle &\geq 1 - \tfrac{1}{8}.
\end{aligned}$$

As such, we see that $(y_i + \frac{1}{4}B_2^n) \cap (y_j + \frac{1}{4}B_2^n) \neq \varnothing$ if and only if $\langle y_i, y_j\rangle \geq 1 - \frac{1}{8}$. For all $i \in [N]$, define $N_i \subseteq [N]$ by

$$N_i := \{j \in [N] : \langle y_i, y_j\rangle \geq 1 - \tfrac{1}{8}\}.$$

Furthermore, we define a measure μ on $\mathcal{P}([N])$ by setting $\mu(S) := \sum_{j\in S} \lambda_j$ for all $S \subseteq [N]$. Then we have $\mu([N]) = \sum_{j=1}^N \lambda_j = \sum_{j=1}^N \lambda_j \,\mathrm{tr}(y_j y_j^\top) = \mathrm{tr}(I_n) = n$, where we used that for all $j \in [N]$ one has $\mathrm{tr}(y_j y_j^\top) = \mathrm{tr}(y_j^\top y_j) = \langle y_j, y_j\rangle = \|y_j\|^2 = 1$. Moreover, for all $i \in [N]$ we find

$$1 = \|y_i\|_2^2 = y_i^\top I_n y_I = y_i^\top \left(\sum_{j=1}^N \lambda_j y_j y_j^\top \right) y_i = \sum_{j=1}^N \lambda_j \langle y_i, y_j\rangle^2 \geq \sum_{j\in N_i} \lambda_j (1 - \tfrac{1}{8})^2 = (1 - \tfrac{1}{8})^2 \mu(N_i),$$

hence $\mu(N_i) \leq \frac{1}{(1-\frac{1}{8})^2} < 2$.

Now let $T \subseteq \{y_1, \ldots, y_N\}$ be the centres of an optimal packing (among all packings using only the vectors $y_1, \ldots, y_N$ as centres), and write $S := \{i \in [N] : y_i \in T\}$. Note that every $j \in [N]$ must belong to some N_i with $i \in S$, for otherwise the packing could be extended by adding y_j. It follows that $\bigcup_{i \in S} N_i = [N]$, so we have

$$n = \mu([N]) \leq \sum_{i \in S} \mu(N_i) < 2|S|.$$

It follows that $|S| > \frac{n}{2}$. □

To complete the proof of Dvoretsky's theorem, we just need the following lemma.

Lemma 4.16 *There is an absolute constant $\gamma > 0$ with the following property: let $n \in \mathbb{N}_1$ be given, and let $K \subseteq \mathbb{R}^n$ be a symmetric convex body such that the John ellipsoid of K is B_2^n. Then*

$$\mathop{\mathbb{E}}_{z \sim N(0, I_n)} \left[\|z\|_K \right] \geq \gamma \sqrt{\log(n)}.$$

Proof: By the Sudakov inequality, we may choose an absolute constant $\gamma' > 0$ such that for every $n \in \mathbb{N}_1$, every symmetric convex body $P \subseteq \mathbb{R}^n$ and every $t \in \mathbb{R}_{>0}$ one has

$$N(P^\circ, tB_2^n) \leq e^{\gamma' (\frac{\ell}{t})^2}, \qquad \text{where } \ell := \mathop{\mathbb{E}}_{z \sim N(0, I_n)} \|z\|_P. \tag{4.17}$$

We shall use this to prove that

$$\mathop{\mathbb{E}}_{z \sim N(0, I_n)} \|z\|_K \geq \tfrac{1}{4} \cdot \sqrt{\frac{\ln(\frac{n}{2})}{\gamma'}},$$

which is easily seen to be of the desired form.

By John's decomposition of the identity, we may choose vectors $y_1, \ldots, y_N \in \partial K \cap S^{n-1}$ and scalars $\lambda_1, \ldots, \lambda_N \geq 0$ such that $I_n = \sum_{j=1}^N \lambda_j y_j y_j^\top$. Recall from lecture 3 (the proof of John's theorem) that we automatically have $y_1, \ldots, y_N \in K^\circ$, or equivalently

$$K \subseteq \left\{ x \in \mathbb{R}^n : |\langle x, y_i \rangle| \leq 1 \text{ for all } i \in [N] \right\} =: P.$$

Note that we have $P = \{y_1, \ldots, y_N\}^\circ$, so it follows from the bipolar theorem that P° is the absolutely convex hull of $y_1, \ldots, y_N$, which is simply $\operatorname{conv}(\pm y_1, \ldots, \pm y_N)$. Setting $t = \frac{1}{4}$ and applying (4.17) and Proposition 4.15, we find

$$\mathop{\mathbb{E}}_{z \sim N(0, I_n)} \|z\|_K \geq \mathop{\mathbb{E}}_{z \sim N(0, I_n)} \|z\|_P \geq \tfrac{1}{4} \cdot \sqrt{\frac{\ln(N(P^\circ, \frac{1}{4} B_2^n))}{\gamma'}} \geq \tfrac{1}{4} \cdot \sqrt{\frac{\ln(\frac{n}{2})}{\gamma'}},$$

which proves our claim. □

Putting together the ideas and results from this subsection (Proposition 4.14, Lemma 4.16) and the Dvoretsky criterion (Lemma 4.4) completes the proof of Dvoretsky's theorem.

4.3 Closing remarks

We conclude this section with some thoughts on the bound obtained in Dvoretsky's theorem.

Remark 4.18 How to think of the term $\frac{1}{\log(1+\frac{1}{\varepsilon})}$? Perhaps it is instructive to think of it like this: set $x := \frac{1}{\varepsilon}$, then the lower bound from Dvoretsky's theorem becomes $d \geq \frac{\log(n)}{x^2 \log(1+x)}$. So the term $\frac{1}{\log(1+\frac{1}{\varepsilon})}$ acts as a negligible (but non-constant) extra factor.

Remark 4.19 (A rough estimate) Alternatively, we may consider the following rough estimate. One easily shows that for all $x \geq 0$ one has $0 \leq \ln(1+x) \leq x$. (Use that $\ln(1) = 0$ and $\frac{d}{dx}\ln(1+x) = \frac{1}{1+x}$, which is ≤ 1 on $\mathbb{R}_{\geq 0}$.) Therefore we have $0 < \ln(1+\frac{1}{\varepsilon}) \leq \frac{1}{\varepsilon}$, and thus

$$\frac{1}{\ln(1+\frac{1}{\varepsilon})} \geq \varepsilon, \qquad \text{for all } \varepsilon > 0.$$

This is only a very rough estimate, which gets progressively worse as ε gets smaller. (If $\varepsilon \to 0$, then $x := \frac{1}{\varepsilon} \to \infty$, and we note that $\ln(1+x)$ grows much slower than x.) Nevertheless, it is quite good if ε is very large, and we use this in Remark 4.20 below.

Remark 4.20 Why upper bound the allowed values of ε? Dvoretsky's theorem gives a trade-off between accuracy and dimension, and it would be interesting to know what happens if we choose in favour of dimension. However, this question remains unanswered in this form of Dvoretsky's theorem. We claim that the statement does not hold unconditionally, that is, there does not exist an absolute constant $C > 0$ which works for all r.

For the constants C'_r constructed in the proof of the Dvoretsky criterion (Lemma 4.4), we have $\lim_{r\to\infty} C'_r = 0$, but that could have been a result of poor choices. However, suppose that such a constant $C > 0$ does exist. Then if we consider a regime where $X = \ell_\infty^n$ and $\varepsilon = \sqrt[3]{n}$ as $n \to \infty$, we find that the rough estimate from Remark 4.19 gives us

$$d \geq C \cdot \varepsilon^3 \log(n) = C \cdot n \log(n).$$

So as n gets larger, we will eventually have $d > n$ (which is absurd), as well as $\ell_2^d \overset{1+\sqrt[3]{n}}{\hookrightarrow} \ell_\infty^n$, which is equally absurd (as we established in lecture 1 that $d(\ell_2^n, \ell_\infty^n) = \sqrt{n}$). This explains why we need to work with an upper bound on ε. The theorem doesn't tell us what happens if we choose in favour of dimension, only what happens if we choose in favour of accuracy.

Remark 4.21 In Remark 3.3 we saw that the Johnson–Lindenstrauss lemma is trivial if ε is too small compared to n (say, $\varepsilon \leq \frac{1}{\sqrt{n}}$), as $n \to \infty$. A similar situation occurs in Dvoretsky's theorem: if we have $\varepsilon \leq \frac{1}{\sqrt{\log(n)}}$, then we find

$$C_r \cdot \frac{\varepsilon^2 \log(n)}{\log(1+\frac{1}{\varepsilon})} = C_r \cdot \frac{\log(n)}{(\frac{1}{\varepsilon})^2 \log(1+\frac{1}{\varepsilon})} \leq C_r \cdot \frac{\log(n)}{\log(n)\log(1+\sqrt{\log(n)}\,)} = C_r \cdot \frac{1}{\log(1+\sqrt{\log(n)}\,)},$$

so for large enough n this only tells us that $d \geq 1$.

Remark 4.22 Finally, we point out that Dvoretsky's theorem occurs in various different incarnations in the literature. For instance, [Mat02] states and proves the following:

Theorem 4.23 (Dvoretzky's theorem, cf. [Mat02, Theorem 14.6.1]) *For any natural number k and any real $\varepsilon > 0$, there exists an integer $n = n(k, \varepsilon)$ with following property. For any n-dimensional centrally symmetric convex body $K \subseteq \mathbb{R}^n$, there exists a k-dimensional linear subspace $L \subseteq \mathbb{R}^n$ such that the section $K \cap L$ is $(1 + \varepsilon)$-almost spherical.*

The best known estimates give $n(k, \varepsilon) = e^{O(k/\varepsilon^2)}$.

References

[Mat02] Jiří Matoušek, *Lectures on Discrete Geometry* (2012), Graduate Texts in Mathematics 212, Springer.