CS 798 - Convexity and Optimization, Winter 2017, Waterloo

Lecture 18: Introduction to convex geometry

We will give an overview of the topics that we will cover, including Brunn-Minkowski and isoperimetric inequalities. We will also go through the basics in computing volumes, that will underlie the phenomenon of measure concentration.

Overview

In Vempala's notes, there are four basic structural facts about convex bodies:

① <u>Separation</u>: For a convex body K and any point $x \notin K$, there is a hyperplane separating K and x.

We have seen this in L03, and used it to derive duality theorems in L04-05, which in turn have applications everywhere (although we haven't seen many applications in this course).

We used duality to derive John's ellipsoid theorem in L06, and also in deriving the approximate Caratheodory's theorem in L11 and also in the algorithm for polytope intersection in L16, but there are much more.

② <u>Sandwiching / rounding</u>: Any convex body can be approximated by a simple convex body.

Given two convex bodies K_1 and K_2, we say K_1 is an α-rounding of K_2 if there exists a (center) point x in K_1 such that $K_1 \subset K_2 \subset x + \alpha(K_1 - x)$.

From John's theorem in L06, for any convex body $K \subseteq \mathbb{R}^n$ there is an ellipsoid which is a n-rounding of it.

This is a basic result in functional analysis as it allows us to approximate any norm by a quadratic norm.

Also, in L06, we see that the Dikin ellipsoid (Hessian of a convex function at a point) is a good approximation of the domain, and this is important in the analysis of interior point algorithms and also the fastest cutting plane algorithms.

③ <u>Brunn-Minkowski inequality</u>: This is a fundamental result about the volume distribution of a convex body.

We will explain the statement later after computing some volumes.

From Brunn-Minkowski, many interesting results can be derived including the Grünbaum's theorem that we used in the center of gravity method in L15, and also isoperimetric inequalities that are crucial in deriving the measure concentration results in probability theory, that have interesting applications everywhere.

④ <u>Concentration / isoperimetric / expansion</u>: It is a phenomenon about how mass is distributed in a convex body.

For example, we will soon see that most mass of a sphere is close to the equator (although a sphere is not a convex set), and this has important implications in probability theory.

One very important algorithmic problem is the efficient sampling of a random point in a convex body, whose analysis is based on the isoperimetric/expansion properties of a convex body.

If time permits, we will also see the approximate center of gravity cutting plane method, and the inertia ellipsoid that is a good rounding of a convex body.

Plan: We have already seen ① and ②, and our plan is to see ③ and ④.

The focus will shift towards the probabilistic implications, more than the optimization implications that we have seen in much of the course so far.

We will see the proofs of Brunn-Minkowski, Grünbaum, and some isoperimetric inequalities.

We should also see some algorithms for random sampling in convex bodies.

We may see other applications such as approximate center of gravity method, inertia ellipsoids, and an interesting result in (combinatorial) discrepancy minimization using convex geometry.

Gaussian distributions [Ross]

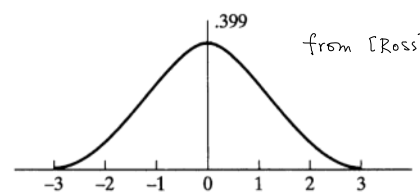
As this part of the course will be more about probability, it is good to recall some important facts.

We say that X is a normal random variable, or normally distributed, with parameters μ and σ^2 if the density of X is given by $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$ for $-\infty < x < \infty$.

We denote that X is normally distributed by $X \sim N(\mu, \sigma^2)$.

The density function is a bell-shaped curve that is symmetric about μ.

The picture shows the density function $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ for $N(0,1)$.



from [Ross]

Some basic facts about normal random variables:

- $\int_{-\infty}^{\infty} f(x)\,dx = 1$, i.e. it is a probability distribution, e.g. $\int_{-\infty}^{\infty} e^{-x^2/2}\,dx = \sqrt{2\pi}$.
- $E[X] = \mu$ and $Var[X] = \sigma^2$.
- If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ and X, Y are independent, then $X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

It is a good exercise to work it out. These facts are also implicit in the calculations for volumes later.

Concentration of measure

The Gaussian distribution plays a central role in probability theory because of the central limit theorem, which says that if $X_1, \ldots, X_n$ are independent random variables with mean μ and variance σ^2, then the distribution of their sum $\sum_{i=1}^{n} X_i$ converges to $N(n\mu, n\sigma^2)$ as $n \to \infty$.

This holds for any identically distributed random variables X_i, or even non-identical random variables under some assumptions.

This gives us a good idea that the sum of a large number of random variables is concentrated around its mean, as the Gaussian random variable is concentrated around its mean.

Let $X \sim N(0,1)$ and $\Pr(X \geq t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-x^2/2}\,dx$.

A useful estimate is that $\left(\frac{1}{t} - \frac{1}{t^3}\right) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \leq \Pr(X \geq t) \leq \frac{1}{t} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$.

So, when t is large, $\Pr(X \geq t) \approx \frac{1}{t} e^{-t^2/2}$, and so the probability that the sum is t standard deviation far away is exponentially decreasing with t^2.

We should keep this bound in mind when we discuss about measure concentration.

The proof of the estimate is using the trick that $\left(1 - \frac{1}{3x^4}\right) \Pr(X=x) \leq \Pr(X=x) \leq \left(1 + \frac{1}{x^2}\right) \Pr(X=x)$ but the LHS and RHS can be computed using integration by parts (good exercise).

Random unit vectors

Here is a simple way to generate a random unit vector.

Let $r = (X_1, \ldots, X_n) \in \mathbb{R}^n$ be a vector where each entry is an independent $N(0,1)$ variable.

The density function for r is $\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-X_i^2/2} = \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot e^{-\sum_{i=1}^{n} X_i^2/2} = \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot e^{-\|r\|_2^2/2}$.

Note that the density function only depends on the distance of the point to the origin.

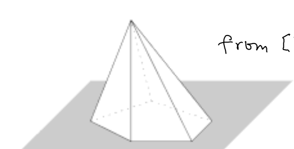
Therefore, the distribution of r is spherically symmetric. So, $r/\|r\|$ is a random unit vector.

Volume [Ball, lecture 1]

Having an understanding of the volumes of simple objects will give us good intuitions about probability.

Let's start with computing the volumes of a unit ball $B_2^n := \{ x \in \mathbb{R}^n \mid \|x\|_2^2 \leq 1 \}$ and a unit sphere $S^{n-1} = \{ x \in \mathbb{R}^n \mid \|x\|_2^2 = 1 \}$ in their respective dimension.

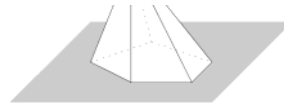
Cone The two volumes are closely related through the concept of a cone.



from [Ball]

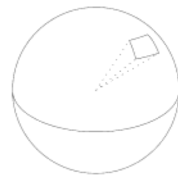
Given a convex body B of dimension $n-1$, a cone of base B is simply the convex hull of a single point and B.

Given a convex body B of dimension $n-1$, a cone of base B is simply the convex hull of a single point and B.



If the point is of height one above the base, then $\mathrm{vol}(\text{cone}) = \int_0^1 (t)^{n-1}\mathrm{vol}(B)\,dt = \frac{1}{n}\mathrm{vol}(B)$.

If the point is of height h, then $\mathrm{vol}(\text{cone}) = \frac{h}{n}\mathrm{vol}(B)$.



from [Ball]

Now, from the picture, we see that $\mathrm{vol}(B_2^n) = \frac{1}{n}\mathrm{vol}(S^{n-1})$, by dividing the ball into cones.

Ball Let's focus on the volume of the unit ball now, denoted by v_n.

We consider integration using polar coordinates.

In general, given a function $f: \mathbb{R}^n \to \mathbb{R}$, $\int_{\mathbb{R}^n} f = \int_{r=0}^{\infty}\int_{S^{n-1}} f(r\theta)\cdot r^{n-1}\,d\theta\,dr$, where $d\theta$ is the volume measure of the sphere, and the factor r^{n-1} appears as the sphere of radius r has volume r^{n-1} times that of S^{n-1}.

If the function f is invariant under rotation, then $\int_{\mathbb{R}^n} f = n v_n \int_{r=0}^{\infty} f(r)\cdot r^{n-1}\,dr$.

We will use the function $f(x) = \exp(\sum_{i=1}^n x_i^2)$, which is invariant under rotation and is also easy to integrate directly as the variables are independent.

On one hand, $\int_{\mathbb{R}^n} f = \int_{\mathbb{R}^n} \prod_{i=1}^n e^{-x_i^2}\,dx = \prod_{i=1}^n \int_{-\infty}^{\infty} e^{-x_i^2}\,dx_i = \sqrt{\pi}^n$.

On the other hand, $\int_{\mathbb{R}^n} f = n v_n \int_{r=0}^{\infty} e^{-r^2} r^{n-1}\,dr$

$= \frac{1}{2} n v_n \int_{t=0}^{\infty} e^{-t} t^{\frac{n}{2}-1}\,dt$ (by letting $t = r^2$)

Using integration by parts, $\int_0^\infty e^{-t}t^k\,dt = k\int_0^\infty e^{-t}t^{k-1}\,dt$.

The base cases are $\int_0^\infty e^{-t}\,dt = 1$ and $\int_0^\infty e^{-t}t^{\frac{1}{2}}\,dt = \frac{\sqrt{\pi}}{2}$.

Let $\Gamma(k) = (k-1)\Gamma(k-1)$ be the gamma function with $\Gamma(1) = 1$ and $\Gamma(\frac{1}{2}) = \frac{\sqrt{\pi}}{2}$.

Then $\sqrt{\pi}^n = \int_{\mathbb{R}^n} f = \frac{1}{2} n v_n \Gamma(\frac{n}{2}) = v_n \Gamma(\frac{n}{2}+1)$

Therefore, $v_n = \dfrac{(\pi)^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}$.

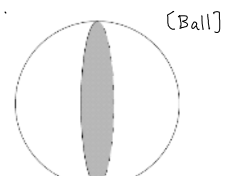
Using Stirling's approximation that $n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$, we have $v_n \approx \left(\frac{2\pi e}{n}\right)^{\frac{n}{2}}$.

This is exponentially smaller than a cube of $[0,1]^n$.

In particular, for a ball to have volume 1, its radius needs to be $\approx \sqrt{\frac{n}{2\pi e}}$.

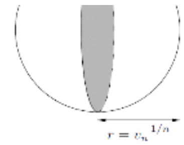
Slice Let's compute the volume of a slice near the equator of a ball of volume 1.

[Ball]



The central slice is an $(n-1)$-dimensional ball of radius $r = v_n^{-\frac{1}{n}}$, since $r^n v_n = 1$.

The volume of the central slice is $v_{n-1} r^{n-1} = v_{n-1} v_n^{-\frac{(n-1)}{n}}$



The volume of the central slice is $v_{n-1} r^{n-1} = v_{n-1} v_n^{-\frac{(n-1)}{n}}$

$$\approx \left(\frac{2\pi e}{n-1}\right)^{\frac{n-1}{2}} \left(\frac{n}{2\pi e}\right)^{\frac{n}{2}\left(\frac{n-1}{n}\right)}$$

$$= \left(\frac{n}{n-1}\right)^{\frac{n-1}{2}} \approx \sqrt{e} \quad \text{when } n \to \infty \quad \text{as} \quad \lim_{n\to\infty} \left(1+\frac{1}{n}\right)^n = e$$

For a slice which is at distance x from the center, it's radius is $\sqrt{r^2-x^2}$ and its volume is

$$v_{n-1}\left(\sqrt{r^2-x^2}\right)^{n-1} = v_{n-1} r^{n-1} \left(\frac{\sqrt{r^2-x^2}}{r}\right)^{n-1} \approx \sqrt{e}\left(1-\frac{x^2}{r^2}\right)^{\frac{n-1}{2}} \approx \sqrt{e}\left(1-\frac{2\pi e x^2}{n}\right)^{\frac{n-1}{2}} \quad \text{using } r \approx \sqrt{\frac{n}{2\pi e}}$$

$$\approx \sqrt{e}\exp\left(-\pi e x^2\right) \quad \text{using } 1-x \approx e^{-x} \text{ for small } x$$

Thus, if we project the mass distribution of the ball of volume 1 onto a single direction, we get a distribution that is approximately Gaussian with variance $\sigma^2 = \frac{1}{2\pi e}$.

About 95% of the mass is within two standard derivations, which lies in the slab $\{x \mid -\frac{1}{2} \le x_1 \le \frac{1}{2}\}$.

Note that the radius is roughly $\sqrt{\frac{n}{2\pi e}}$ and so really most of the mass is close to the equator.

The volume of the ball concentrates close to any subspace of dimension $n-1$, but that does not mean that the volume concentrates near the center, as it is easy to see that most mass is near the surface.

So, the mass distribution in high dimension is not consistent with our intuition in low dimension.

<u>Caps</u> Scaling back to a ball of radius one, the fraction of mass with $x_1 \ge \frac{c}{\sqrt{n}}$ is roughly equal to the probability that $\Pr(X \ge c)$ for $X \sim N(0,1)$, which is about $\frac{1}{c} e^{-\frac{c^2}{2}}$ as we estimated.

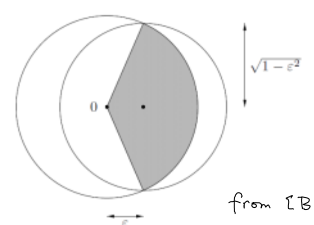
So, if we randomly pick a unit vector, then $\Pr(x_1 \ge \frac{c}{\sqrt{n}}) \le \frac{1}{c} e^{-\frac{c^2}{2}}$.

It also implies that two random vectors u, v are almost orthogonal: By rotational invariance, we can assume that $u = e_1$, and thus $\Pr(\langle u,v\rangle \ge \frac{c}{\sqrt{n}}) = \Pr(v_1 \ge \frac{c}{\sqrt{n}}) \le \frac{1}{c} e^{-\frac{c^2}{2}}$.

This is also an upper bound of a spherical cap: for a fixed unit vector v, a spherical cap $C(\varepsilon, v)$ is defined as $\{ u \in \mathbb{R}^n \mid \langle u,v\rangle \ge \varepsilon \}$.

By the above estimate, we have $\mathrm{vol}(C(\varepsilon,v))$ is at most a $e^{-\frac{\varepsilon^2 n}{2}}$ fraction of $\mathrm{vol}(S^{n-1})$.

There is also a nice way to see this more directly.

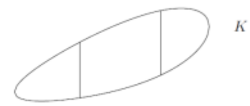


from [Ball]

From the picture, $\frac{\mathrm{vol}(C(\varepsilon,v))}{\mathrm{vol}(S^{n-1})} = \frac{\mathrm{vol}(\text{cone})}{v_n} \le \frac{\sqrt{1-\varepsilon^2}^{\,n} v_n}{v_n} \le e^{-\frac{\varepsilon^2 n}{2}}$.

This argument is similar to the argument used in geometric descent, and only holds when $\varepsilon \le \frac{1}{\sqrt{2}}$, so that the vertical side is not shorter than the horizontal side.

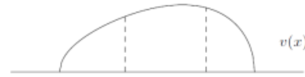
Brunn-Minkowski inequality and isoperimetric inequality [Ball, lecture 5]



[Ball]

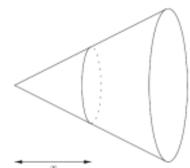
This is a preview of the topics in the next couple lectures.

Let $K \subset \mathbb{R}^2$, let $v(x)$ be the length of the line with the first coordinate being x, then it is easy to see that v is concave on its support.



This is not true in 3-dimension, as the area of a cone's section increases with x^2.

Brunn proved that this is the worst case (as it is the most "expanding" case).



[Ball]

Theorem (Brunn) Let K be a convex body in $\mathbb{R}^n$, let $u \in \mathbb{R}^n$, and for each $r \in \mathbb{R}$ let H_r be the hyperplane $\{x \in \mathbb{R}^n \mid \langle x, u \rangle = r\}$.

Then the function $r \to \mathrm{vol}(K \cap H_r)^{\frac{1}{n-1}}$ is concave on its support.

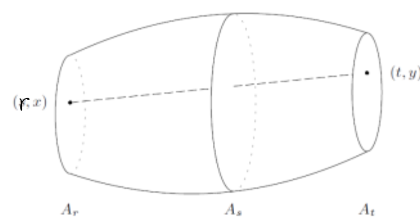
Minkowski came up with a very useful reformulation.

The Minkowski sum of two sets A, B is defined as $A+B := \{a+b \mid a \in A, b \in B\}$.

One way to think of it is to add a copy of B for each point of A.

For Brunn's theorem, if A_r and A_t are two slices of a convex body K, the only place we used convexity is to ensure that $(1-\lambda)A_r + \lambda A_t \subseteq K$.

So Minkowski reformulate Brunn's theorem as follows.



Theorem (Brunn-Minkowski inequality) If A and B are nonempty compact subsets of $\mathbb{R}^n$ then

$$\mathrm{vol}((1-\lambda)A + \lambda B)^{\frac{1}{n}} \geq (1-\lambda)\,\mathrm{vol}(A)^{\frac{1}{n}} + \lambda\,\mathrm{vol}(B)^{\frac{1}{n}}.$$

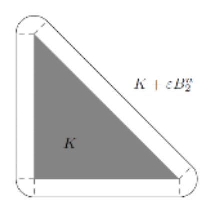
Let see an immediate application of Minkowski's formulation to prove a classical isoperimetric inequality.

Theorem (isoperimetric inequality) Among bodies of a given volume, Euclidean balls have least surface area.

Proof. Let C be a compact set in $\mathbb{R}^n$ whose volume is equal to that of B_2^n.

The surface area of C can be written as $\mathrm{vol}(\partial C) = \lim_{\varepsilon \to 0} \dfrac{\mathrm{vol}(C+\varepsilon B_2^n) - \mathrm{vol}(C)}{\varepsilon}$.

By Brunn-Minkowski inequality, $\mathrm{vol}(C+\varepsilon B_2^n)^{\frac{1}{n}} \geq (1-\varepsilon)\,\mathrm{vol}(\frac{C}{1-\varepsilon})^{\frac{1}{n}} + \varepsilon\,\mathrm{vol}(B_2^n)^{\frac{1}{n}} = \mathrm{vol}(C)^{\frac{1}{n}} + \varepsilon\,\mathrm{vol}(B_2^n)^{\frac{1}{n}}$



from [Ball]

$$\begin{aligned}
\text{So, } \mathrm{vol}(\partial C) &\geq \lim_{\varepsilon \to 0} \frac{\left(\mathrm{vol}(C)^{\frac{1}{n}} + \varepsilon\,\mathrm{vol}(B_2^n)^{\frac{1}{n}}\right)^n - \mathrm{vol}(C)}{\varepsilon} \\
&\geq \lim_{\varepsilon \to 0} \frac{\mathrm{vol}(C) + n\varepsilon\,\mathrm{vol}(C)^{\frac{n-1}{n}}\,\mathrm{vol}(B_2^n)^{\frac{1}{n}} - \mathrm{vol}(C)}{\varepsilon} \\
&= n\,\mathrm{vol}(C)^{\frac{n-1}{n}}\,\mathrm{vol}(B_2^n)^{\frac{1}{n}} = n\,\mathrm{vol}(B_2^n) \quad \text{as } \mathrm{vol}(C) = \mathrm{vol}(B_2^n).
\end{aligned}$$

$= \mathrm{vol}(\partial B_2^n)$ by the cone argument. $\square$

With the isopermetric inequalities, one can readily use them to derive concentration inequalities in probability. The Brunn-Minkowski theorem will be useful in those applications, which will be described later.

References - [Ball, lecture 1, lecture 5], highly recommended.

- Hopcroft, Kannan. Foundations of data science, chapter 2.
- Ross, A first course in probability.