Winter School
Combinatorial and algorithmic aspects of convexity
Paris, Institut Henri Poincaré, January 19-23, 2015.

Algorithmic aspects of convexity.
Santosh Vempala's lecture I.

Scribes: Joe Anderson and Alperen Ergur

1 Introduction

Convex geometry is a rich and classical field of Mathematics. On the other hand, computational complexity and the theory of algorithms are only about half a century old. When one studies *complexity*, the objective is to study the amount of resources (space, time, randomness, communication) needed to solve a particular problem, as a function of the size of the input. For *geometric* problems, the input size is generally a function of the dimension (e.g. the input is a set of points in $\mathbb{R}^n$). We want to address the following type of question: "What is the complexity of a given problem as a function of dimension?" We have typically two kinds of behavior: the ones which are polynomial (P) and the ones which are exponential or worse-NP (decision problems with polysize proofs of Yes answer). The question "P=NP?" has a great importance and the convexity is a frontier of polytime!

See [GLS88] for more information of many aspects of these problems.

We will focus on the following problems, each of which is to be solved by an algorithm taking input of a particular type and format and returning some output with particular guarantees:

1. Optimization

   Input: a function $f : \mathbb{R}^n \to \mathbb{R}$ and a number $\epsilon > 0$.

   Output: $x \in \mathbb{R}^n$ such that $f(x) \geq \max_{y\in\mathbb{R}^n} f(y) - \epsilon$.

2. Integration

   Input: A function $f : \mathbb{R}^n \to \mathbb{R}$ and a number $\epsilon > 0$.

   Output: $V \in \mathbb{R}$ such that $(1-\epsilon)\int f \leq V \leq (1+\epsilon)\int f$.

3. Sampling

   Input: Function $f : \mathbb{R}^n \to \mathbb{R}$ and $\epsilon > 0$.

   Output: A sample X drawn from a distribution π such that $d(\pi, \pi_f) \leq \epsilon$, where π_f is the distribution whose density is proportional to f.

4. Learning

   Input: Set of data $\{(x_i, f(x_i)\}$ where each x_i is drawn according to an unknown distribution $\mathcal{D}$, a number $\epsilon > 0$.

   Output: A function g such that $P_{\mathcal{D}}(g(x) \neq f(x)) \leq \epsilon$.

5. Rounding

   Input: Function $f : \mathbb{R}^n \to \mathbb{R}_+$ such that $\int f < \infty$, and $c_1, c_2 \in \mathbb{R}$

   Output: An affine transformation T such that

   (a) $E_{Tf}(\|x\|^2) = n$

   (b) $P_{Tf}(\|x\| \leq c_1) \geq c_2$

In the above examples, one must consider how a function f can be provided as an input argument; to do so, one can imagine passing a function pointer to a function which is a black box to compute f at a given point. Then the complexity under this model is the number of calls to such a black box needed by the oracle to give an acceptable output.

Problem: without more hypotheses, the problems 1 to 5 are all intractable. No efficient algorithm is possible.
But if we add some **convexity** hypotheses on f, then it become possible. For example in problem 1 and 2 we moreover assume:

Example 1 (Convex Optimization)**.** Minimize a function f over K where f is convex and $K \subseteq \mathbb{R}^n$ is convex. This can be solved with accuracy ϵ and complexity $\text{poly}(n, \log 1/\epsilon)$ by the Ellipsoid Algorithm (to be discussed in detail in the next talk).

Example 2. Let $f = \chi_K$ for a convex body $K \subseteq \mathbb{R}^n$. The task is to compute $\text{vol}(K) = \int f$ and it can be solved with complexity $\text{poly}(n, 1/\epsilon)$ (see [DFK91]).

2 Preliminaries

Definition 1. Let $S \subset \mathbb{R}^n$. We call S convex if $a, b \in S \implies [a, b] \in S$, where $[a, b] = \{\lambda a + (1 - \lambda)b : \lambda \in [0, 1]\}$.

Definition 2. Let $f : \mathbb{R}^n \to \mathbb{R}$. We call f concave if $f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y)$ for $\lambda \in [0, 1]$ and all $x, y \in \mathbb{R}^n$.

Definition 3. We call f log-concave if $f(\lambda x + (1 - \lambda)y) \geq f(x)^\lambda f(y)^{1-\lambda}$ for $\lambda \in [0, 1]$ and all $x, y \in \mathbb{R}$.

Definition 4. We call f harmonic-concave if

$$f(\lambda x + (1 - \lambda)y) \geq \frac{1}{\frac{\lambda}{f(x)} + \frac{1-\lambda}{f(y)}}$$

for $\lambda \in [0, 1]$ and all $x, y \in \mathbb{R}$.

Definition 5. We call f quasi-concave if $f(\lambda x + (1-\lambda)y) \geq \min\{f(x), f(y)\}$ for $\lambda \in [0, 1]$ and all $x, y \in \mathbb{R}$.

Definition 6. We call f star-shaped if there is some $x_0 \in S$ such that for all $y \in S$, $f(\lambda x_0 + (1 - \lambda)y) \geq \lambda f(x_0) + (1 - \lambda)f(y)$.

Example 3. Exercise in relation to the problem 4 (sampling). Consider the following algorithmic task: given a point $x \in \mathbb{R}^n$, produce a "label" for x, i.e., compute $\ell(x) : \mathbb{R}^n \to \{-1, 1\}$. The points will be given sequentially, each after the previous has been assigned a label. Moreover, after each label is produced – and before the next point is supplied – the algorithm receives feedback on its choice of label, either "correct" or "incorrect". If the algorithm has chosen the label incorrectly, it is charged one; otherwise, no charge. The ultimate goal of the algorithm designer is to limit the number of mistakes in *any* setting. However, since the labels may be chosen by an adversary, one cannot hope to have a good strategy in general, since the adversary can always claim you are wrong, not having to have assigned a ground truth labeling *a priori*.

One possibility to have a "good" algorithm, however, is to restrict the problem setting to one in which there is some halfspace H defined by its normal vector $a \in \mathbb{R}^n$; the ground truth is then

$$\ell(x) = \begin{cases} 1 & \langle x, a \rangle \geq 0 \\ -1 & \langle x, a \rangle < 0. \end{cases}$$

Using this knowledge, after n points are given, the algorithm has a set of constraints $\langle a, x_i \rangle < 0$ or $\langle a, x_i \rangle \geq 0$, where a is the unknown. To guess the label of the new point x, compute $\langle x, a \rangle$ over all a satisfying the previous constraints and return the majority answer, (supposing this can be done efficiently).

Note that to compute ℓ (*i.e.* a), one needs $n(b+1)$ bits, where b is the bit length of a. Then using the majority decision, when a mistake is made, one eliminates at least half of the valid choices of a. Thus, after $2^{n(b+1)}$ rounds of this game, the number of mistakes will be at most $n(b+1)$.

Instead of using the majority, which takes exponentially many rounds to become reliable, one can simply choose an a from the remaining feasible region uniformly at random. This will give the number of expected mistakes at most $2n(b+1)$, since the algorithm has at least 50% chance to agree with the majority on every round.

Exercise: Use sampling to make the argument rigorous.

3 Structure of convex bodies

We now study some algorithmic problems and see how one can use convexity to construct good solutions. In this section we'll note basic structural facts about convex bodies. These facts will be useful for later discussions.

3.1 Structure I: Separation

Convex sets have separation oracles. That is, there is an algorithm which, given a point $x \in \mathbb{R}^n$, either says that $x \in K$ or returns a halfspace H such that $K \subset H$ and $x \notin H$. This hyperplane is guaranteed by using convexity of K, considering the point y in K closest to $x \notin K$. Then one takes the hyperplane perpendicular to $y - x$ passing through y. Furthermore, the closest point in K to $x \notin K$ is unique, because if there were two, say y_1 and y_2, then one could take the point $(y_1 + y_2)/2$ which would be strictly closer to x.

3.2 Structure II: Brunn-Minkowski

Let $A, B \subset \mathbb{R}^n$ be compact. Note that $\mathrm{vol}(\lambda A) = \lambda^n \mathrm{vol}(A)$ for $\lambda > 0$.

Definition 7. The Minkowski sum of A and B is

$$A + B = \{x + y : x \in A, y \in B\}.$$

Theorem 8 (Brunn-Minkowski)**.** *Let $A, B \subset \mathbb{R}^n$ be compact and $\lambda \in [0,1]$. Then*

$$\operatorname{vol}(\lambda A + (1-\lambda)B)^{1/n} \geq \lambda \operatorname{vol}(A)^{1/n} + (1-\lambda)\operatorname{vol}(B)^{1/n}.$$

Consequence: the volume of the sections of a convex body by parallel hyperplanes is a $1/(n-1)$-concave function. Indeed, let K be a convex body and u be a unit vector which determines a fixed direction. For $a \in K$, let $H_a = \{x; \langle x - a, u\rangle = 0\}$. We apply the Brunn-Minkowski inequality in $\mathbb{R}^{n-1}$ to the sets $A = K \cap H_a$ and $B = K \cap H_b$. From the convexity of K one has $\frac{A+B}{2} \subset K \cap H_{\frac{a+b}{2}}$. Thus

$$\operatorname{vol}(K \cap H_{\frac{a+b}{2}})^{\frac{1}{n-1}} \geq \operatorname{vol}\left(\frac{A+B}{2}\right)^{\frac{1}{n-1}} \geq \frac{\operatorname{vol}(K\cap H_a)^{\frac{1}{n-1}} + \operatorname{vol}(K \cap H_b)^{\frac{1}{n-1}}}{2}.$$

Proof of Brunn-Minkowski's theorem:
We first prove the case where the two sets A and B in $\mathbb{R}^n$ are axis-aligned parallelopipeds of side length $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$, respectively. This gives us $\operatorname{vol}(A) = \prod a_i$ and $\operatorname{vol}(B) = \prod b_i$, and $\operatorname{vol}(A+B) = \prod(a_i + b_i)$. Then, using the arithmetico-geometric inequality;

$$\begin{aligned}\frac{\operatorname{vol}(A)^{1/n} + \operatorname{vol}(B)^{1/n}}{\operatorname{vol}(A+B)^{1/n}} &= \left(\prod \frac{a_i}{a_i + b_i}\right)^{1/n} + \left(\prod \frac{b_i}{a_i+b_i}\right)^{1/n} \\ &\leq \frac{1}{n}\sum \frac{a_i}{a_i+b_i} + \frac{1}{n}\sum\frac{b_i}{a_i+b_i} = 1.\end{aligned}$$

Next, we take A and B to be finite sets of disjoint parallelepipeds and proceed by induction on the total number of parallelepipeds. Note that one can find an axis hyperplane H with at least one full member of A (or B, but we assume that this is achieved for A slog) on each side. Then we decompose A and B into A^+, A^-, B^+, and B^- depending on which side of H they lie.

We can then translate B so that $\frac{\operatorname{vol}(A)}{\operatorname{vol}(B)} = \frac{\operatorname{vol}(A^+)}{\operatorname{vol}(B^+)}$ and we apply the inductive hypothesis:

$$\begin{aligned}
\mathrm{vol}(A+B) &\geq \mathrm{vol}(A^+ + B^+) + \mathrm{vol}(A^- + B^-) \\
&\geq (\mathrm{vol}(A^+)^{1/n} + \mathrm{vol}(B^+)^{1/n})^n + (\mathrm{vol}(A^-)^{1/n} + \mathrm{vol}(B^-)^{1/n})^n \\
&= \mathrm{vol}(A^+)\left(1 + \left(\frac{\mathrm{vol}(B^+)}{\mathrm{vol}(A^+)}\right)^{1/n}\right)^n + \mathrm{vol}(A^-)\left(1 + \left(\frac{\mathrm{vol}(B^-)}{\mathrm{vol}(A^-)}\right)^{1/n}\right)^n \\
&= \mathrm{vol}(A^+)\left(1 + \left(\frac{\mathrm{vol}(B)}{\mathrm{vol}(A)}\right)^{1/n}\right)^n + \mathrm{vol}(A^-)\left(1 + \left(\frac{\mathrm{vol}(B)}{\mathrm{vol}(A)}\right)^{1/n}\right)^n \\
&= \mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n}.
\end{aligned}$$

For the last step, we approximate any compact sets A and B by finite unions of parallelepipeds and take the limit.

□

3.3 Structure III: Sandwiching

Theorem 9 (John's Theorem)**.** *Let $K \subseteq \mathbb{R}^n$ be convex. Then there exists an ellipsoid E such that $E \subseteq K \subseteq nE$.*

Proof idea. Use the ellipsoid contained in K of maximum volume. □

Remark 1:For K centrally symmetric, there exists and ellipsoid E with property $E \subseteq K \subseteq \sqrt{n}E$.
Remark 2: for any deterministic algorithm using a membership oracle, one cannot compute an approximation to such an ellipsoid with better than $n^{3/2}$ accuracy.
$\mathbb{E}_K$ denotes expected value with respect to the uniform measure over K.

Theorem 10 (Inertia Ellipsoid)**.** *Let K be a convex body with center of gravity $\mathbb{E}_K(x) = x_0$ and inertia matrix $\mathbb{E}_K[(x-x_0)(x-x_0)^T] = A$. The inertia ellipsoid is $E = \{y : (y-x_0)A^{-1}(y-x_0) \leq 1\}$. Then*

$$\sqrt{\frac{n+1}{n}}E \subseteq K \subseteq \sqrt{n(n+1)}E.$$

The inertia ellipsoid is easy to construct, a linear number of samples is enough. It gives the empirical covariance matrix.

Definition 11 (Milman's Ellipsoid)**.** Let $N(A, B)$ be the minimum number of translates of B needed to cover A. A Milman Ellipsoid of K is an ellipsoid E such that $N(E, K)N(K, E) = 2^{O(n)}$.

It can be computed by a probabilistic algorithm in polytime, but the best deterministic algorithm needs $2^{\Omega(n)}$ and this is optimal.

3.4 Structure IV: Concentration of mass

Recall that in $\mathbb{R}^n$, the unit hypercube has volume 1. However, the unit sphere has volume about c^n/n^n, approaching 0 with n.

Let B be a ball of radius r. Then $\operatorname{vol}(tB) = t^n \operatorname{vol}(B)$. Compare this with the volume of a ball of radius $(1-1/n)r$: $(1-1/n)^n \operatorname{vol}(B) \approx vol(B)/e$. This means that in high dimension, about a third of the mass is contained in a very thin shell (of width $\approx 1/n$) near the boundary.

In an opposite way, if one look at a slice H_t of the same ball at distance t from the origin, then

$$\frac{\operatorname{vol}_{n-1}(H_t)}{\operatorname{vol}_{n-1}(H_0)} = \left(\frac{\sqrt{r^2 - t^2}}{r}\right)^{n-1} = \left(1 - \frac{t^2}{r^2}\right)^{\frac{n-1}{2}} = e^{-\frac{t^2(n-1)}{2r^2}}.$$

Hence one gets a constant fraction of the volume in a slab of width $c/\sqrt{n}$.

The first case is related to the thin shell conjecture which asks if for any convex body of volume one and in isotropic position a large part of the mass is contained in a shell of constant radius.

4 Basics for tomorrow: convex optimization

Let K be a convex body given by a separation oracle and by two numbers $0 < r < R$ such that there exists x_0 unknown so that $B(x_0, r) \subset K \subset B(0, R)$. The goal will be to minimize a convex function f over K, *i.e.* to find a point $x \in \mathbb{R}^n$ such that $f(x) \geq \max_{y \in \mathbb{R}^n} f(y) - \epsilon$.

This problem reduces to the feasibility problem: if K is non-empty, find a single point in K. We first find a point in $\{x \in K; f(x) \leq t\}$ then ask if there is a point in $\{x \in K; f(x) \leq t/2\}$, Next, use binary search to find another point and repeat. This will produce the expected result in $O(\log(1/\varepsilon)$ steps. In general, results of this type will depend on parameters x_0, r, and R such that $x_0 + rB_2^n \subseteq K \subseteq RB_2^n$.

References

[DFK91] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM (JACM)*, 38(1):1–17, 1991.

[GLS88] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag, Berlin, 1988.