Lecture 2: Convex Optimization I

Lecturer: Santosh Vempala *Scribe: Steve Weis*

Applications of Random Walks

Optimization is the problem of minimizing some function over some domain: $\min f(x), x \in S$. We will consider convex optimization, the special case when the function and set are both convex.

Definition 1. *A* ***convex set*** *K is a subset of $\mathbb{R}^n$ such that for any two points in K, the line connecting the two points is also in K: $\forall x, y \in K$, $\forall \lambda, 0 \le \lambda \le 1$, $\lambda \cdot x + (1-\lambda) \cdot y \in K$.*

Definition 2. *A* ***convex function*** *is a function that satisfies the inequality $f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(x)$ for all $\lambda : 0 \le \lambda \le 1$.*

A **linear program** in $\mathbb{R}^n$ is an optimization problem of the form $\min(c^T x)$, $Ax \le b$, where A is a m x n matrix and both c and b are vectors. This is clearly a special case in $\mathbb{R}^n$ of convex optimization - K is a polyhedron and f is linear.

Convex optimization generalizes many discrete problems as well. An example is **Weighted Perfect Matching** in a graph: Given a graph $G = (V, E)$ and a set of edge weights w_{ij}, find the minimum weight perfect matching. We can frame this as a convex optimization problem by using a vector, $x \in \{0,1\}^{|E|}$ to indicate which edges are used in the matching M. Defining K to be the convex hull of all such perfect matchings of G, we can state the problem as: $\min(w^T x), x \in K$.

There is a problem, because getting a usable description of this convex hull K is not trivial. However, a result of Edmonds is that the following set of constraints defines the same convex set:

$$\forall i \textstyle\sum_j x_{ij} = 1$$

$$\forall i, j \in V, 0 \le x_{ij} \le 1$$

$$\forall S \subseteq V, |S| \text{ is odd}$$

$$\forall i \in S, j \notin S, \textstyle\sum x_{ij} \ge 1$$

Separation Oracles

Returning to the problem of solving convex optimizations brings up the question of how to get a description of $f(x)$ and K. For a function we can use an explicit description but what about how to describe K? Consider the the example of **Semi-Definite Programming**:

$\min(C \circ X), A_i \cdot X \leq b, X \succcurlyeq 0$

This signifies that the matrix X is positive semi-definite:

$\forall y \in \mathbb{R}^n, y^T X y \geq 0, \quad \text{or} \quad X(yy^T) \geq 0$

Fact 3. *For any convex set K, every point y is either in K or there exists some plane which separates y from K:*

$\forall y \in \mathbb{R}^n$ *either* $y \in K$ *or* $\exists a^T x = b$ *such that* $a^T y \leq b, \forall z \in K, a^T z > b$

A separation oracle can produce either $y \in K$ or $y \notin K$ and $a^T x = b$. To build such a separation oracle for positive semi-definite matrices, we simply need to check that all eigenvalues are greater than zero. This gives us an efficient $\Theta(n^3)$ time separation oracle. This is really a *weak* separation oracle because the eigenvalues can be approximated to any desired accuracy, but cannot be calculated exactly since they might be irrational.

It should be noted that it is hopeless to solve every convex optimization problem exactly since the optimum might be irrational. Instead we aim for getting ϵ-close to the optimum, in time proportion to $\log \frac{1}{\epsilon}$.

1 An algorithm for feasibility

We can reduce convex optimization to the following feasibility problem by using a binary search on the function value.

Definition 4. *The input to the* ***Feasibility Problem*** *is convex set K, a separation oracle and real values R and r and the output is either a point $x \in K$ or "K is empty".*

Theorem 5. *Suppose K contains an n-dimensional cube with side r and is itself contained in a n-dimensional cube with side R. The $n \log \frac{R}{r}$ is a lower bound on the number of calls to a separation oracle to solve the Feasibility problem*

This problem can be solved using the Ellipsoid method which runs in polynomial-time and makes $\Theta(n^2 \log \frac{R}{r})$ calls to the oracle.

Here is a different algorithm:

1. $P :=$ a cube of size R, $z := 0$
2. Check if $z \in K$. If so, we are done. Otherwise find a separating plane $a^T x \leq b$ containing K.
3. $P := P \cap \{a^T x \leq a^T z\}$
4. $Z := \frac{1}{N} \sum_{i=1}^{N} y_i$ where $y_1, \ldots, y_N$ are random points in P. Loop to step 2.

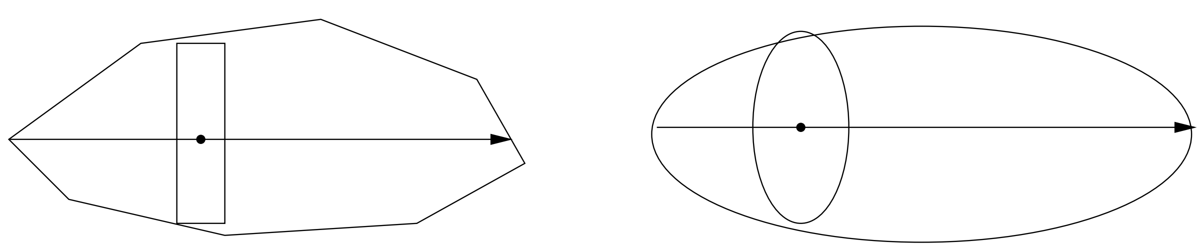
The idea of selecting the random points y is to approximate the centroid and cut the current polyhedron through that approximation. Cutting through the exact centroid does not always cut the polyhedron's n-dimensional volume in half, as can be demonstrated on a triangle, however it does cut the volume by a constant factor.

Definition 6. *A **centroid** of a convex body K is defined as:* $\frac{1}{Vol(K)} \int_{x \in K} x dx$.

Theorem 7. *For any convex body K, any cut through its centroid has at least $\frac{1}{e}$ of the volume on each side.*

Proof of Theorem 7:

To show this, consider the process of **symmetrizing** the polyhedron. This is accomplished by constructing an axis perpendicular to the separating cut through the centroid. At each point along the axis, replace the cross-section of the polyhedron with a (n-1)-dimensional ball of equal volume, illustrated as follows:



Our claim is that the newly constructed region is also convex. Consider two cross sections $A := \{x | a^T x = t_1\}$ and $B := \{x | a^T x = t_2\}$. Also consider the region $C := \{x | x = \lambda y + (1-\lambda) z, y \in A, z \in B\}$, i.e. $C = \lambda A + (1-\lambda) B$ for some $\lambda : 0 \leq \lambda \leq 1$. Since we defined each cross section as a n-dimensional ball, $Vol(A) = f(n) r^{n-1}$. To prove this, we will use the following lemma, proved by Brunn-Minkowski:

Lemma 8. $\forall A, B$ *convex* $\subseteq \mathbb{R}^n$, $Vol(A+B)^{\frac{1}{n}} \geq Vol(A)^{\frac{1}{n}} + Vol(B)^{\frac{1}{n}}$

Proof of Lemma 8:

Let A and B be defined as n-dimensional cuboids with sides $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ respectively. Then $A+B$ will have sides of length at most $(a_i + b_i)$. From this, we can see the following:

$Vol(A+B)^{\frac{1}{n}} = (\Pi(a_i + b_i))^{\frac{1}{n}} \geq (\Pi a_i)^{\frac{1}{n}} + (\Pi b_i)^{\frac{1}{n}} = Vol(A)^{\frac{1}{n}} + Vol(B)^{\frac{1}{n}}$

The proof can be extended to unions of cuboids and then to the general case when we approximate a convex set by a union of cuboids.

Returning to the proof of Theorem 7, define $A' := \lambda A$ and $B' := (1-\lambda)B$. Using the above lemma:

$Vol(A'+B')^{\frac{1}{n-1}} \geq Vol(A')^{\frac{1}{n-1}} + Vol(B')^{\frac{1}{n-1}}$

$Vol(\lambda A + (1-\lambda)B)^{\frac{1}{n-1}} \geq Vol(\lambda A)^{\frac{1}{n-1}} + Vol((1-\lambda)B)^{\frac{1}{n-1}}$

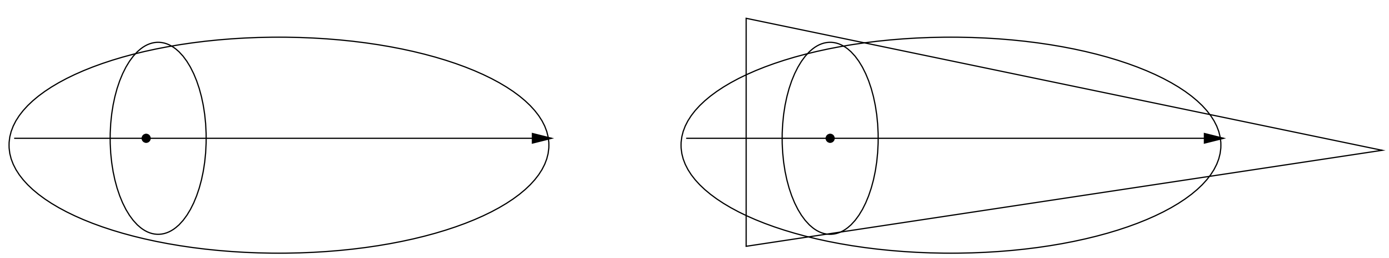
$Vol(C)^{\frac{1}{n-1}} \geq Vol(\lambda A)^{\frac{1}{n-1}} + Vol((1-\lambda)B)^{\frac{1}{n-1}}$

Since $Vol(\lambda A)^{\frac{1}{n-1}} = (f(n)\lambda^{n-1}r^{n-1})^{\frac{1}{n-1}} = \lambda(f(n)r^{n-1})^{\frac{1}{n-1}} = \lambda Vol(A)^{\frac{1}{n-1}}$, we arrive at:

$Vol(C)^{\frac{1}{n-1}} \geq \lambda Vol(A)^{\frac{1}{n-1}} + (1-\lambda)Vol(B)^{\frac{1}{n-1}}$

This means that the process of symmetrization maintains convexity.

To lower bound the volume on one of the sides of the cut, we replace a half of K with a n-dimensional cone and the other half with a truncated cone equal to the original volume. In class, this was accomplished in two parts: the "right" half is a cone tapering down from the original cut radius and the "left" half is a truncated cone. This is illustrated as follows:



This process only moves the center of gravity to the right and so the new right "half" is at most the old right "half" in volume. Thus we can focus on the new convex body, name a cone.

First, we will find the centroid z of the cone of base radius R and height h:

$$z = \frac{1}{Vol(K)} \int_0^h t f(n) r(t)^{n-1} t dt = \frac{n f(n)}{f(n) R^{n-1} h} \int_0^h t \left(\frac{tR}{h}\right)^{n-1} dt = \frac{nh}{n+1}$$

Using the centroid z, we can lower bound the volume of one of the halves and can show that Theorem 7 holds:

Vol(half K) $= \int_0^{\frac{nh}{n+1}} f(n)(\frac{tR}{h})^{n-1} dt = \frac{Ah}{n}(\frac{n}{n+1})^n \geq$ Vol(K) $(1 - \frac{1}{n+1})^n \geq \frac{1}{e} \cdot Vol(K)$