| **10-725: Optimization** | **Fall 2012** |
|---|---|

## Lecture 15: October 15

*Lecturer: Barnabas Poczos* — *Scribes: Christian Kroer, Fanyi Xiao*

**Note**: *LaTeX template courtesy of UC Berkeley EECS dept.*

**Disclaimer**: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

This lecture's notes illustrate some uses of various LaTeX macros. Take a look at this and imitate.

## 15.1 Outline

- Linear programs
- Simplex algorithm
- Running time of Simplex
- Cutting planes and ellipsoid methods for linear programming
- Cutting planes and ellipsoid method for unconstrained minimization

## 15.2 Linear Programs

The inequality form of LPs:

$$\begin{aligned} \min\ & c^T x \\ s.t.\ & Ax \leq b \\ & l \leq x \leq u \end{aligned}$$

The standard form of LPs:

$$\begin{aligned} \min\ & c^T x \\ s.t.\ & Ax \leq b \\ & x \geq 0, b \geq 0 \end{aligned}$$

Any LP can be written into the standard form.

## 15.3 Simplex Method

The Simplex algorithm is a combinatorial search algorithm for solving linear programs. Conceptually, it jumps from one intersection of half-planes to another and increase/decrease the objective along this process

until convergence.
The running complexity of the simplex method has long been believed to be polynomial in the size of the problem $(m, n)$, where $m$ is the number of constraints whereas $n$ is the number of variables. The worst case is to numerate all possible configurations of basic variables which results in $C_n^m = \frac{n!}{(n-m)!m!}$ steps. However, the Klee-Minty example demonstrates that the running time is not polynomial [1]

$$\max_{x_1,\ldots,x_n} \sum_{j=1}^{n} 10^{n-j} x_j$$

$$s.t.\ 2\sum_{j=1}^{i} 10^{i-j} x_j + x_i \le 100^{i-1}, i = 1, 2, ..., n$$

$$x_j \ge 0$$

This example requires $2^n - 1$ pivot steps to get the optimal solution. After shown to be not a polynomial algorithm, people start looking for new algorithms other than simplex. Khachiyan's ellipsoid method is the first to be proved running at polynomial complexity for LPs. However, it is usually slower than simplex in real problems. Karmarkar's method is the first algorithm works reasonably well in practice while has polynomial complexity in theory. It falls into the category of interior point method.

## 15.4 The ellipsoid method for linear programs

### 15.4.1 The feasibility problem

Given some polyhedral set

$$\Omega = \{y \in \mathbb{R}^m : y^T a_j \le c_j, j = 1, \ldots, n\}$$

The feasibility problem is to find some $x \in \Omega$. It is possible to prove that solving the feasibility problem is equivalent to solving general LPs.

### 15.4.2 The ellipsoid method

We now start to describe how the ellipsoid method works. For this we make the following assumptions:

- (A1) $\Omega$ can be covered with a finite ball of radius $R$
  - That is: $\Omega \subseteq \{y \in \mathbb{R}^m : |y - y_0| \le R\} = S(y_0, R)$
  - We assume that $R, y_0$ are both known to us
- (A2) There exists a ball with radius $r$ that fits inside of $\Omega$
  - That is, there exists $r, y^*$ such that $S(y^*, r) \subset \Omega$
  - r is known, $y^*$ is unknown

We will iteratively compute smaller and smaller $y_0, R$ until $y_0$ is inside $\Omega$.

For this we will need ellipsoids.

[1] For a concrete example, please see `http://www.math.ubc.ca/~israel/m340/kleemin3.pdf`

**Definition 15.1 (Ellipsoid)** *An ellipsoid is a set*

$$\mathcal{E} = \{y \in \mathbb{R}^m : (y - z)^T Q(y - z) \leq 1\}$$

*Where $z \in \mathbb{R}^m$ is the center, and $Q \in \mathbb{R}^{n \times m}$ is positive definite.*

Some properties of ellipsoids are:

- Axes of the ellipsoid are the eigenvectors of $Q$
- Lengths of the axes are $\lambda_1^{-\frac{1}{2}}, \dots, \lambda_m^{-\frac{1}{2}}$
- Volume of the ellipsoid is equal to the volume of the unit sphere times the determinant of $Q^{-\frac{1}{2}}$:

$$\begin{aligned}\text{VOL}(\mathcal{E}) =& \text{VOL}(S(0,1))\text{DET}(Q^{-\frac{1}{2}})\\ =& \text{VOL}(S(0,1)) \prod_i \lambda_i^{-\frac{1}{2}}\end{aligned}$$

As mentioned, the ellipsoid method works by construction a series of ellipsoids, we let $\mathcal{E}_k$ denote the $k$-th ellipsoid. The center of the $k$-th ellipsoid is $y_k$, and the parameter matrix is $Q = B_k^{-1}$, where $B_k \succ 0$.

At each iteration $k$, we have $\Omega \subset \mathcal{E}_k$. We can then check whether $y_k \in \Omega$. If it is, then we have found an element of $\Omega$, and we're done. If not, there is at least one constraint that $y_k$ violates.

Suppose it is the $j$th constraint, we have: $a_j^T y_k > c_j$. It follows from this:

$$\Omega \subset \mathcal{E}_k \cap \{y : a_j^T y \leq a_j^T y_k\}$$

By the definition of the ellipsoid, we have that for all points $y$ in the ellipsoid, $a_j^T y \leq c_j$. Since the $j$th constraint is violated, we have $c_j < a_j^T y_k$, which gives the intersection above.

We have that $\{y : a_j^T y \leq a_j^T y_k\}$ is a halfspace, and that it cuts the ellipsoid in half. Let

$$\frac{1}{2}\mathcal{E}_k = \mathcal{E}_k \cap \{y : a_j^T y \leq a_j^T y_k\}$$

Now, the next step is to fit a new ellipsoid around this smaller set. Formally, the successor ellipsoid $\mathcal{E}_{k+1}$ is defined to be the minimal volume ellipsoid containing $\frac{1}{2}\mathcal{E}_k$.

It is constructed as follows:

First, we define:

$$\tau = \frac{1}{m+1}, \quad \delta = \frac{m^2}{m^2 - 1}, \quad \sigma = 2\tau$$

$$y_{k+1} = y_k - \frac{\tau}{(a_j^T B_k a_j)^{\frac{1}{2}}} b_k a_j$$

$$B_{k+1} = \delta(B_k - \sigma \frac{B_k a_j a_j^2 B_k}{a_j^T B_k a_j}$$

We will not prove that this is indeed the minimizer here.

Now, we compare the ratios of the two ellipsoids. We know that

$$\mathcal{E}_{k+1} = \mathcal{E}(y_{k+1}, B_{k+1}^{-1}) \supset \frac{1}{2}\mathcal{E}_k$$

**Theorem 15.2**

$$\frac{VOL(\mathcal{E}_{k+1})}{VOL(\mathcal{E}_k)} = (\frac{m^2}{m^2-1})^{(m-1)/2} < \exp(\frac{-1}{2(m+1)}) < 1$$

Again, we don't prove this.

**Convergence** For the initial step,

$$\Omega \subset \{y \in \mathbb{R}^m : |y - y_0| \leq R\}$$

Where $y_0, R$ are known. We start the ellipsoid method from the $S(y_0, R)$ ellipsoid.

Now, consider the ratio

$$\frac{\text{VOL}(\mathcal{E}_{2m})}{\text{VOL}(\mathcal{E}_0)} = \frac{\text{VOL}(\mathcal{E}_{2m})}{\text{VOL}(\mathcal{E}_{2m-1})} \cdot \ldots \cdot \frac{\text{VOL}(\mathcal{E}_1)}{\text{VOL}(\mathcal{E}_0)}$$

We can upper bound this using Theorem 15.2. We get

$$\frac{\text{VOL}(\mathcal{E}_{2m})}{\text{VOL}(\mathcal{E}_{2m-1})} \cdot \ldots \cdot \frac{\text{VOL}(\mathcal{E}_1)}{\text{VOL}(\mathcal{E}_0)} \leq \exp(\frac{-2m}{2(m+1}) < \frac{1}{2}$$

Intuitively, this means that if we run the ellipsoid method for $2m$ steps, then the volume of the ellipsoid will be halved. Thus, in $O(m)$ iterations, the ellipsoid method can reduce the volume of the ellipsoid to half of its initial volume.

**How many iterations do we need to get into $S(y^*, r)$?**

We start from the sphere $S(y_0, R)$. Its volume is $\text{VOL}(\mathcal{E}_0) = cR^m$. After $k$ steps, we want an ellipsoid with volume $\text{VOL}(\mathcal{E}_k) \leq c\tau^m$.

We get

$$\frac{\text{VOL}(\mathcal{E}_k)}{\text{VOL}(\mathcal{E}_0)} \leq (\frac{\tau}{R})^m \leq (\frac{1}{2})^{\frac{k}{m}}$$

Where the last inequality is because we halve the size in every $m$ steps. Rewriting, we get

$$\begin{aligned} m \log \frac{\tau}{R} &\leq \frac{k}{m} \log \frac{1}{2} && \Leftrightarrow \\ m \log \frac{R}{\tau} &\geq \frac{k}{m} \log 2 && \Leftrightarrow \\ k &\leq O(m^2 \log \frac{R}{\tau}) \end{aligned}$$

Hence we can reduce the volume to less than that of a sphere of radius $r$ in $O(m^2 \log \frac{R}{\tau})$ iterations.

A single iteration of the ellipsoid method requires $O(m^2)$ operations. Thus, the entire process requires $O(m^4 \log \frac{R}{\tau})$ operations.

Thus, we can solve the feasibility problem.

## 15.5 Solving general LPs with the ellipsoid method

Consider a primal LP:

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \\ & x \geq 0 \end{aligned}$$

and its dual:

$$\begin{aligned} \min \quad & y^T b \\ \text{s.t.} \quad & y^T A \geq c^T \\ & y \geq 0 \end{aligned}$$

From duality theory we know that for any $x, y$ such that they are primal and dual feasible solutions, we have $c^T x \leq b^T y$. For optimal $x^*, y^*$, we know that $c^T x^* = b^T y^*$. Thus, we can write the feasibility program

$$\begin{aligned} -c^T x + b^T y &\leq 0 \\ Ax &\leq b \\ -A^T y &\leq c \\ x, y &\geq 0 \end{aligned}$$

Where any feasible $x, y$ for this problem must be primal and dual optimal solutions.

Thus, we can bound the number of operations needed for solving a linear program by:

$$O((m+n)^4 \log \frac{R}{\tau})$$

## 15.6 Cutting Plane and Ellipsoid Method for Unconstrained Convex Optimization

Initialize with an ellipsoid containing the optimal point:

$$x^* \in \epsilon_k = \{x \in R^n : (x - x_k)^T P_k^{-1} (x - x_k) \leq 1\}$$

Now if we know how to compute $\partial f$, the subgradient of the function $f(x)$, we can use the following inequality:

$$\partial f(x_k)(x^* - x_k) \leq f(x^*) - f(x_k) \leq 0$$

With this, we can draw a line perpendicular to $\partial f(x_k)$ to cut the plane into 2 half-planes and have $x^*$ in one of them:

$$x^* \in H(x_k) = \{x \in R^n : \partial f(x_k)^T x \leq \partial f(x_k)^T x_k\}$$

Therefore we have:

$$x^* \in H(x_k) \cap \epsilon_k = \{x \in R^n : \partial f(x_k)^T x \leq \partial f(x_k)^T x_k \text{ and } (x - x_k)^T P_k^{-1} (x - x_k) \leq 1\}$$

Thus, we can construct a new ellipsoid $\epsilon_{k+1}$ with the minimum volume containing the set $H(x_k) \cap \epsilon_k$:

$$H(x_k) \cap \epsilon_k \in \epsilon_{k+1}$$

We iterate the above procedure until we satisfy the stopping criteria $f(x_k) - f(x^*) \leq \epsilon$. The explicit form of constructing new ellipsoid is as follows:

$$\begin{aligned} x_{k+1} =& x_k - \frac{1}{n+1} P_k \tilde{g}_{k+1} \\ P_{k+1} =& \frac{n^2}{n^2-1}(P_k - \frac{2}{n+1} P_k \tilde{g}_{k+1} \tilde{g}_{k+1}^T P_k) \end{aligned}$$

Where $\tilde{g}_{k+1} = \frac{g_{k+1}}{\sqrt{g_{k+1}^T P_k g_{k+1}}}$.
The cutting plane method could also be used for constrained problem as long as it is easy to compute the intersection of the feasible region and the set $H(x_k) \cap \epsilon_k$. We can update $\epsilon_k + 1$ as the ellipsoid with the minimum volume containing $H(x_k) \cap \epsilon_k \cap C$, where $C$ is the feasible region of the problem.

# References

[LY] D.G. Luenberger and Y. Ye, "Linear and Nonlinear Programming," 2008

[BV] S. Boyd and L. Vandenberghe, "Convex Optimization," 2004