ORIE 6300 Mathematical Programming I	November 6, 2014

Lecture 21

Lecturer: David P. Williamson *Scribe: Sijia Ma*

1 Overview of the Interior Point Methods

The ellipsoid method is not practically efficient for large scale problems, though very important theoretically, especially in combinatorial optimization. It can be viewed as an existence proof for an efficient algorithm. It inspired the search for a practically efficient and theoretically polynomial time algorithm.

Interior-point methods were initially devised by Karmakar (1984) (although they are closely related to barrier methods used for linear and nonlinear programming since the 1950s). Since then, great development has led to more sophisticated interior-point methods that are competitive with (and are sometimes faster than) the simplex method.

Interior-point methods represent a significant development in the theory and practice of linear programming. They combine the advantages of the simplex method and of the ellipsoid algorithm. From a theoretical point of view, they lead to efficient (polynomial time) algorithms and use interesting geometric ideas; from a practical point of view, they allow the solution to large scale problems that arise in many applications.

Consider the standard form LP and its dual:

$$\begin{array}{rlcrlcl}
\textbf{Primal} & & \qquad & \textbf{Dual} & & & \\
\min & c^T x & & \max & b^T y & & \\
\text{s.t.} & Ax = b & & \text{s.t.} & A^T y \leq c & \Leftrightarrow & A^T y + s = c \\
 & x \geq 0 & & & & & s \geq 0
\end{array}$$

We assume A is an $m \times n$ matrix with rank m. We define the feasible regions as follows:

$$\mathcal{F}(P) = \{x \in \mathbf{R}^n : Ax = b, x \geq 0\}$$

$$\mathcal{F}(D) = \{(y, s) \in \mathbf{R}^m \times \mathbf{R}^n : A^T y + s = c, s \geq 0\}$$

(interior)

$$\mathcal{F}^\circ(P) = \{x \in \mathbf{R}^n : Ax = b, x > 0\}$$

$$\mathcal{F}^\circ(D) = \{(y, s) \in \mathbf{R}^m \times \mathbf{R}^n : A^T y + s = c, s > 0\}$$

Interior-point methods generate a sequence of points in $\mathcal{F}^\circ(P)$ or in $\mathcal{F}^\circ(P) \times \mathcal{F}^\circ(D)$ converging to an optimal solution. In practice, we get with in 10^{-8} of optimal after 10-50 iterations. These iterations are more expensive either than a simplex pivot or an ellipsoid iteration. However, in $O(n \ln \frac{1}{\varepsilon})$ iterations, they come within a $(1 + \varepsilon)$ factor of the optimal value. Some interior-point methods only need $O(\sqrt{n} \ln \frac{1}{\varepsilon})$ iterations, but usually these algorithms work worse in practice.

2 Improving Direction

One idea to generate this sequence: given a feasible point $\bar{x} \in \mathcal{F}^\circ(P)$, we want to "improve" it, using the "steepest descent" approach to computing the next iteration, such that we keep $Ax = b$. We can ignore $x \geq 0$, since we are always in interior of the feasible region and hence $x > 0$.

We want

$$\begin{aligned}
x = \bar{x} + \alpha\bar{d} \quad \text{s.t. } Ax = b \Rightarrow \quad & A(\bar{x} + \alpha\bar{d}) = b \\
\Rightarrow \quad & \alpha A\bar{d} = 0 \quad (\text{Since } A\bar{x} = b) \\
\Rightarrow \quad & A\bar{d} = 0
\end{aligned}$$

To make sure that α indeed controls the step length, we require that $\|d\| \leq 1$ (so then $\|x - \bar{x}\| \leq \alpha$). We want $\bar{d}$ to be the "steepest descent" direction, so $\bar{d}$ should be the solution of the following, where we have $u = c$:

$$\begin{array}{ll} \min & u^T d \\ \text{s.t.} & Ad = 0 \qquad \text{for } u = c \\ & ||d|| \leq 1 \end{array}$$

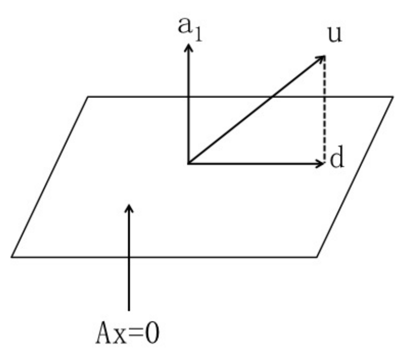

Figure 1: d is actually the unit projection vector of u on hyperplane $Ax = 0$

Lemma 1 *If $u = c$, and $\nexists y$ s.t. $A^T y = u$, then the solution to the above problem is*

$$\bar{d} = -\frac{P_A u}{||P_A u||},$$

where $P_A = I - A^T(AA^T)^{-1}A$.

Note that, suppose $\exists y$ s.t. $A^T y = c$, then $s = 0$, and $(y, s) \notin \mathcal{F}^\circ(D)$. We claim that $\bar{x}$ is optimal by complementary slackness. Actually, since $\bar{x}_i > 0$ and $(A^T y) = c_i$, $\forall i$, we have $\bar{x}$ is optimal for the primal and $(y, 0)$ is optimal for the dual.

Proof: First, we check if everything in the lemma is well-defined. Since we assumed that A has full rank, AA^T is positive definite and therefore $(AA^T)^{-1}$ exists.

We start by showing that $P_A u \neq 0$. Suppose, for a contradiction, that

$$P_A u = 0$$

$$\Rightarrow (I - A^T (AA^T)^{-1} A) u = 0$$

$$\Rightarrow u = A^T (AA^T)^{-1} Au$$

Since u is now $A^T z$ for vector $z = (AA^T)^{-1}Au$, we see that u^T is a linear combination of the rows of A, which contradicts our assumption, and proves the claim.

Consider any d such that $Ad = 0$. Then since $P_A^T = I - A^T(AA^T)^{-1}A = P_A$, we have

$$(P_A u)^T d = u^T {P_A}^T d = u^T (I - A^T (AA^T)^{-1} A) d = u^T d$$

since $Ad = 0$.

So, rather than considering our original linear objective function, it is equivalent to solve the optimization problem

$$\begin{array}{ll} \min & (P_A u)^T d \\ s.t. & Ad = 0 \\ & \| d \| \leq 1. \end{array}$$

Suppose that we ignore the constraint $Ad = 0$ for the moment. Now, we are simply optimizing over the unit ball. By the Cauchy-Schwarz theorem, we know that

$$-\|x\|\|y\| \leq x^T y \leq \|x\|\|y\|.$$

Cauchy-Schwarz gives us a lower bound on the objective function:

$$\begin{aligned}(P_A u)^T d &\geq - \|P_A u\| \, \|d\| \\ &\geq - \|P_A u\| \, .\end{aligned}$$

So $-\|P_A u\|$ is the smallest objective function we can hope to get.

If we set

$$\bar{d} = -\frac{P_A u}{\|P_A u\|}$$

then

$$(P_A u)^T \bar{d} = -\frac{\|P_A u\|^2}{\|P_A u\|} = -\|P_A u\|,$$

so that this gives us the best possible objective function.

Also,

$$\begin{aligned}A\bar{d} &= -\frac{AP_A u}{\|P_A u\|} \\ &= -\frac{A(I - A^T(AA^T)^{-1}A)u}{\|P_A u\|} \\ &= -\frac{(A - AA^T(AA^T)^{-1}A)u}{\|P_A u\|} \\ &= -\frac{(A - A)u}{\|P_A u\|} \\ &= 0\end{aligned}$$

It follows that $\bar{d}$ optimizes the descent direction optimization problem. Furthermore, since $u^T \bar{d} = (P_A u)^T \bar{d} = -\|P_A u\| < 0$, it follows that $u^T(x+d) < u^T x$, so that taking a step in the direction of $\bar{d}$ improves the objective function value. □

3 Affine-scaling Direction

We notice that d does not depend on x! It is good if we are not close to the boundary, but potentially bad if we are.

One solution: we rescale the problem to look like we are at $e = (1 \; 1 \; \dots \; 1)^T$. Let our current iterate at this step be x. Transform this to $\hat{x} = e$ by re-scaling as defined below. By the steepest descent step, get $\bar{d}$ in this transformed space. We then perform the inverse of our transform to map the new point back to our original space (as described below).

Given $\bar{x} \in \mathcal{F}^o(P)$, let

$$\bar{X} = Diag(\bar{x}) = \begin{bmatrix} \bar{x}_1 & 0 & \cdots & 0 \\ 0 & \bar{x}_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & \bar{x}_n \end{bmatrix}$$

and consider the linear transformation,

$$x \rightarrow \hat{x} = \bar{X}^{-1} x.$$

This transforms $\bar{x}$ to e and our original optimization problem becomes

$$\begin{array}{ll} \min & c^T(\bar{X}\hat{x}) = (\bar{X}c)^T\hat{x} \\ s.t. & A(\bar{X}\hat{x}) = b \Rightarrow (A\bar{X})\hat{x} = b \\ & \hat{x} \geq 0. \end{array}$$

Therefore, in our transformed space, we compute the descent direction

$$\hat{d} = -\frac{P_{A\bar{X}}\bar{X}c}{\|P_{A\bar{X}}\bar{X}c\|}.$$

Note, $\hat{d}$ solves

$$\begin{array}{ll} \min & (\bar{X}c)^T d \\ s.t. & (A\bar{X})d = 0 \\ & \| d \| \leq 1. \end{array}$$

If we now map $\hat{d}$ from the transformed space back to our original space, we have derived our new descent direction:

$$\bar{d} = \bar{X}\hat{d} = -\frac{\bar{X}P_{A\bar{X}}\bar{X}c}{\|P_{A\bar{X}}\bar{X}c\|}.$$

Note, $\bar{d}$ solves

$$\begin{array}{ll} \min & c^T d \\ s.t. & Ad = 0 \\ & \|\bar{X}^{-1}d\| \leq 1. \end{array}$$

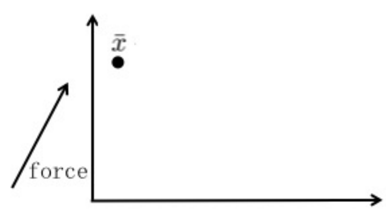

Figure 2: $\bar{x}$ is not to go beyond the boundary

Say $\bar{x}_i$ is very small, then $\bar{X}_{ii}^{-1}$ is very large. In order to make $||\bar{X}^{-1}d|| \leq 1$, it requires that $\bar{d}_i$ is very small. The direction $\bar{d}$ is called *affine-scaling direction* and was introduced by Dikin (1967). This direction gives a good algorithm, but it is not known if the algorithm terminates in polynomial time.

4 Barrier Function

We notice that $\bar{X}P_{A\bar{X}}\bar{X}c$ (the direction of $\bar{d}$) is also the solution to the problem

$$\begin{array}{ll} \min & c^T d + \frac{1}{2}d^T\bar{X}^{-2}d = c^T d + \frac{1}{2}||\bar{X}^{-1}d||^2 \\ \text{s.t.} & Ad = 0 \end{array}$$

It is a tradeoff between minimizing the objective function and step length. The form of this problem looks like minimizing some nonlinear F such that

$$\nabla^2 F = \bar{X}^{-2}$$

The question is that, does such an F exist?
The answer is yes. $F(x)$ is the logarithmic barrier function:

$$F(x) \equiv -\ln(x) = -\sum_j \ln(x_j)$$

So it is to optimize

$$\theta_\mu(x) = c^T x + \mu F(x).$$

Consider the minimization of $F(x)$,

$$\begin{aligned}
\min F(x) \quad &\Leftrightarrow \quad \min -\sum_j \ln(x_j) \\
&\Leftrightarrow \quad \min -\ln(\prod_j x_j) \\
&\Leftrightarrow \quad \max \prod_j x_j
\end{aligned}$$

for $x \in \mathcal{F}^\circ(P)$. Maximizing $\prod_j x_j$ for $x \in \mathcal{F}^\circ(P)$ gives the *analytic center* of $\mathcal{F}^\circ(P)$. In other words, it is a trade-off between minimizing objective function and staying in the center of $\mathcal{F}^\circ(P)$ parametrized by μ, which is called the *barrier parameter.*