Mathematical Programming	**Lecture 25**
OR 630 Fall 2005	**November 29, 2005**
Notes by Arijit Chakrabarty	

Interior-point methods

The ellipsoid method is not practically efficient for large scale problems, though very important theoretically, especially in combinatorial optimisation. 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 (IPMs) 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 IPMs that are competitive with (and are sometimes faster than) the simplex method. We will see just an introduction to IPMs.

In theory, due to the work of Nesterov and Nemirovski, these can be applied to *any* convex programming problem but they are implementable for some important classes of problem including semi-definite and second-order cone programming.

Consider the standard form LP and its dual:

$$
\begin{array}{lrcl}
 & \min_x c^T x & & \\
(P) & Ax & = & b \\
 & x & \geq & 0
\end{array}
\qquad
\begin{array}{lrcccl}
 & \max_{y,s} b^T y & & & & \\
(D) & A^T y & + & s & = & c \\
 & & & s & \geq & 0
\end{array}
$$

where we assume $A \in \Re^{m \times n}$ has rank m.
Define:

$$
\begin{aligned}
\mathcal{F}(P) &:= \{x \in \Re^n : Ax = b, x \geq 0\}, \\
\mathcal{F}^\circ(P) &:= \{x \in \Re^n : Ax = b, x > 0\}, \\
\mathcal{F}(D) &:= \{(y,s) \in \Re^m \times \Re^n : A^T y + s = c, s \geq 0\}, \\
\mathcal{F}^\circ(D) &:= \{(y,s) \in \Re^m \times \Re^n : A^T y + s = c, s > 0\}.
\end{aligned}
$$

IPMs 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. They never (except when all feasible solutions are optimal) generate an optimal solution and hence are infinite iterative algorithms. However, in practice, we get solutions which are within a distance 10^{-8} of the optimal value in 10 to 50 iterations. However these iterations are more expensive than those of the simplex or ellipsoid methods. The number of iterations needed perhaps grows logarithmically in n. Theoretically we can show that such IPMs can generate ε-optimal solutions in $O(n \ln \frac{1}{\varepsilon})$ or $O(\sqrt{n} \ln \frac{1}{\varepsilon})$ iterations.

Suppose we are given $\bar{x} \in \mathcal{F}^\circ(P)$, and we want to "improve" it. Since $\bar{x} > 0$, we ignore non-negativities and try to decrease $c^T x$ staying in $\{x : Ax = b\}$.

N.B.: Norms in this lecture will always be considered in the $\mathcal{L}^2$ sense.
Look at the steepest-descent idea:

$$\min\{c^T x : Ax = b, \|x - \bar{x}\| \leq \alpha\} \quad (\alpha > 0).$$

The solution is :

$$x = \bar{x} + \alpha \bar{d}$$

where $\bar{d}$ is the solution of

$$\begin{array}{rcl} \min u^T d & & \\ Ad & = & 0 \\ \|d\| & \leq & 1 \end{array}$$

for $u = c$.

Proposition 1 *If u is not in the range of A^T, then the solution to the above problem is:*

$$\begin{array}{rcl} \bar{d} & = & \dfrac{P_A u}{\|P_A u\|}, \\ \textit{where } P_A & = & I - A^T(AA^T)^{-1}A. \end{array}$$

Also $u^T \bar{d} = -\|P_A u\| < 0$.

Proof: Note that, since A has full row rank, AA^T is positive definite and hence non-singular. So P_A is well defined. Also, $P_A u = 0$ implies that u lies in the range of A^T, so $P_A u \neq 0$ and $\bar{d}$ is well defined.
For any d with $Ad = 0$, we have

$$\begin{array}{rcl} (P_A u)^T d & = & u^T d - u^T A^T (AA^T)^{-1}(Ad) \\ & = & u^T d. \end{array}$$

So, we can minimise $(P_A u)^T d$ instead. If we minimise this over $\|d\| \leq 1$, we get that $\bar{d}$ is optimal by Cauchy-Schwarz. But since $AP_A = 0$, so $\bar{d}$ satisfies the $Ad = 0$ constraints too. This shows that $\bar{d}$ is the solution to the preceding problem. Finally,

$$\begin{array}{rcl} u^T \bar{d} & = & -\dfrac{u^T P_A u}{\|P_A u\|} \\ & = & -\|P_A u\| \\ & < & 0. \end{array}$$

(See the homework.)

Note that if $c \in \mathcal{R}(\mathcal{A}^T)$, then all feasible solutions are optimal. So assume henceforth $c \notin \mathcal{R}(\mathcal{A}^T)$.
We cannot keep going in direction $\bar{d}$ from any $\bar{x} \in \mathcal{F}^\circ(P)$, since then we would converge to

a non-optimal point. But $\bar{d}$ looks good if $\bar{x}$ is far from all non-negative constraints, *e.g.,* if $\bar{x} = e := \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \in \Re^n$.

Given any arbitrary $\bar{x} \in \mathcal{F}^\circ(P)$, we can rescale the problem so that $\bar{x}$ looks like e.

$$\text{Let } \bar{X} := \mathrm{Diag}(\bar{x}) := \begin{bmatrix} \bar{x}_1 & & \\ & \ddots & \\ & & \bar{x}_n \end{bmatrix}.$$

Consider the linear transformation $x \longmapsto \hat{x} = \bar{X}^{-1}x$. This transforms (P) to

$$\begin{array}{lrcl} & \min(\bar{X}c)^T\hat{x} & & \\ (\bar{P}) & (A\bar{X})\hat{x} & = & b \\ & \hat{x} & \geq & 0, \end{array}$$

and $\bar{x}$ is transformed to $e \in \mathcal{F}^\circ(\bar{P})$. In this scaled problem, we can consider moving in direction $\hat{d} := \frac{-P_{A\bar{X}}(\bar{X}c)}{\|P_{A\bar{X}}(\bar{X}c)\|}$. This corresponds to moving in the original space in direction

$$\bar{d} := \frac{-\bar{X}P_{A\bar{X}}(\bar{X}c)}{\|P_{A\bar{X}}\bar{X}c\|}.$$

Note that $\hat{d}$ solves :

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

So $\bar{d}$ solves

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

This direction $\bar{d}$ is called the (primal) <u>affine-scaling direction</u> and one can get algorithms based on it suggested by I.I.Dikin (1967), and rediscovered by several people after Karmakar's paper. This method is reasonably efficient, but is not thought to be polynomial. Note that

$$P_{A\bar{X}} = I - \bar{X}A^T(A\bar{X}^2A^T)^{-1}\bar{X},$$

so that the matrix that needs to be inverted (or more accurately, for which we have to solve a linear system) *changes* at each iteration. Hence each iteration involves $O(m^3)$ work. Note that $\bar{d}$ is required to satisfy $\|\bar{X}^{-1}d\| \leq 1$, so this is steepest descent *with respect to some norm that*

changes with $\bar{x}$.
Also $-\bar{X}P_{A\bar{X}}\bar{X}c$ solves

$$\begin{array}{lcll} \min & c^T d & + & \frac{1}{2} d^T \bar{X}^{-2} d \\ & Ad & & = 0, \end{array}$$

which can be viewed (perhaps) as minimizing the quadratic approximation to some nonlinear function while staying in the feasible set. This raises the question: Is there a function F with $\nabla^2 F(x) = X^{-2}$? Yes!

$$\begin{aligned} F(x) &:= -\ln(x) \\ &:= -\sum_j \ln(x_j) \end{aligned}$$

satisfies it. This is called the *logarithmic barrier function*. Introducing F answers another question for IPMs: how can we compare two points $\bar{x}$ and $\hat{x}$ in $\mathcal{F}^\circ(P)$? Now we have two criteria: $c^T x$ compares their objective function values while $F(x)$ compares how close they are to the "center" of the feasible region.