MAI0130
Interior Point Methods for LP
Chapter 7.4-7.8

Johan Hagenbjörk

Produktionsekonomi - IEI

Linköpings universitet

December 15, 2016

- Central Path Algorithms.
- Primal Dual Methods.
- Solving the Linear System.

The material it a mix of [Murty, 2009] and [Nocedal and Wright, 2006]

Let Γ be a convext polytope

$$\Gamma = \{x : v = Ax - b \geq 0\}, \tag{1}$$

where v is the vector of slack variables associated with the inequality constraints.

Analytic Center

The analytic center Γ is defined as the point in Γ which maximizes the product of the slack variables assosiated with the inequality constraints, and A is $m \times n$.

This is equivalent of

$$\begin{aligned} \max \quad & \sum_{i=1}^{m} \log v_i \\ \text{subject to} \quad & v = Ax - b \geq 0 \end{aligned}$$

Finding a sequence of interior feasible solutions along a path called the central path, converging to the analytic center of the optimum face of the LP is the building block of some algorithms.
Consider

$$\begin{aligned} \min \quad & z(x) = c^T x \\ \text{subject to} \quad & Ax = b, x \geq 0 \end{aligned} \tag{2}$$

Let $\mu > 0$ and consider

$$\begin{aligned} \min \quad & z(x) = c^T x - \mu \left(\sum_{j=1}^{n} \log x_j \right) \\ \text{subject to} \quad & Ax = b, (x \geq 0) \end{aligned} \tag{3}$$

This problem (3) is known as the logarithmic barrier problem. By penalizing variables entering the negative region we can relax non-negativity.

youtu.be/MsgpSl5JRbI

It can be shown that it has a unique optimum for each $\mu > 0$. Letting y denote the vector of dual variables of Lagrange multipliers associated with the equality constraints of (3), the optimality conditions for a feasible solution $x(\mu)$ is that there exists a $y(\mu)$ satisfying the KKT conditions for (3).

$$\begin{aligned} Ax &= b \\ -A^T y - s &= -c^T \\ Xs &= \mu e \\ x, s &> 0 \end{aligned} \tag{4}$$

where $X = diag(x_1, ..., x_n)$.

It can be shown that the trajectory traced by $(x(\mu), y(\mu), s(\mu))$ exists and is unique for all μ if the original problem (2) and its dual have interior feasible solutions.

This trajectory is called the primal-dual central path. Every point $x(\mu)$ of the central path is the analytic center of the intersection of K with the objective plane through that point.

Consider the standard problem (2) and let x^r be an interior feasible solution. The method creates an ellipsoid in $\mathbb{R}^n$ around x^r by replacing $x \geq 0$ into

$$x \in E_r = \left\{ x : \left(\sum_{i=1}^{n} \frac{x_i - x_i^r}{x_i^2} \right)^2 \leq 1 \right\} \tag{5}$$

We then obtain the problem

$$\begin{aligned} \min \quad & z(x) = c^T x \\ \text{subject to} \quad & Ax = b \\ & \left(\sum_{i=1}^{n} \frac{x_i - x_i^r}{x_i^2} \right)^2 \leq 1 \end{aligned} \tag{6}$$

The intersection of the ellipsoid with the constraints $Ax = b$ is an ellipsoid $\bar{E}_r$ with center x_r.

The optimum of this problem can be computed analytically.

An ellipsoid in $\mathbb{R}^n$ is the set of

$$E = \left\{ x : (x - x^0)^T D (x - x^0) \leq \rho^2 \right\}, \tag{7}$$

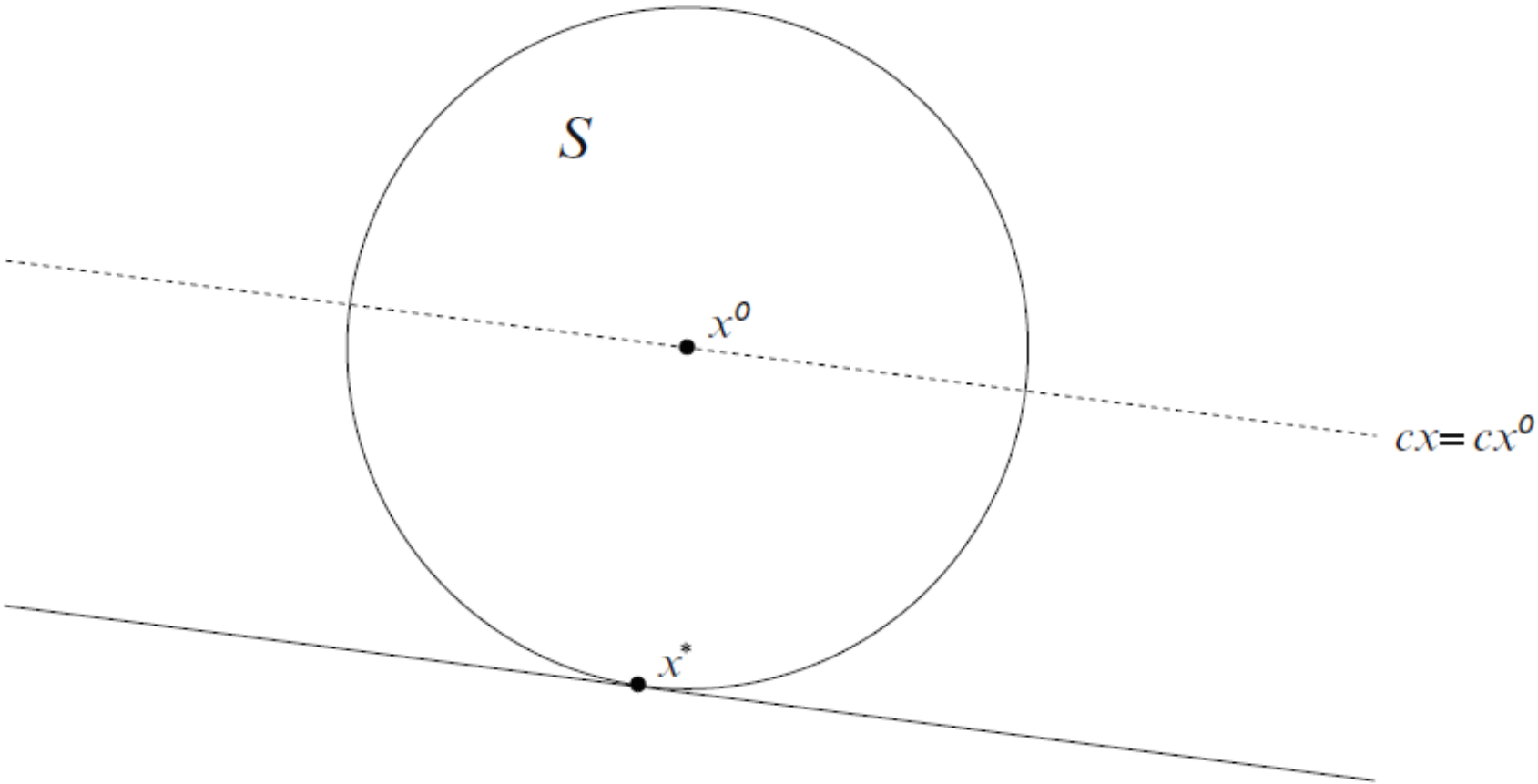
where D is a positive finite matrix.

Consider the problem

$$\begin{aligned} \min \quad & z(x) = c^T x \\ \text{subject to} \quad & Ax = b \\ & (x - x^0)^T D(x - x^0) \leq \rho^2 \end{aligned} \tag{8}$$

For the special case where the ellipsoid is a sphere, $D = I$, we get the solution

$$x^\star = x^0 + \rho(-c^T)/\|c\| \tag{9}$$


S
x^o
$cx = cx^o$
x^*

Let x^0 be a given point and consider the problem

$$\begin{aligned} \min \quad & z(x) = c^T x \\ \text{subject to} \quad & Ax = b \\ & (x - x^0)^T (X^0)^{-2} (x - x^0) \leq \rho^2 \end{aligned} \tag{10}$$

which can be transformed into the sphere problem (8) by

$$y^T = e^T + (x - x^0)^T (X^0)^{-1} \tag{11}$$

which gives us the expression for the point x as

$$x^T = (y - e)^T X^0 (x^0)^T = y^T X^0 \tag{12}$$

This gives us

$$\begin{aligned} \min \quad & z(x) = c^T x \\ \text{subject to} \quad & (y-e)^T(y-e) \leq \rho^2 \end{aligned} \tag{13}$$

with the optimal solution

$$x^\star = x^0 - \frac{\rho (X^0)^2 c^T}{\|X^0 c^T\|}$$

Now consider the problem

$$\begin{aligned} \min \quad & z(x) = c^T x \\ \text{subject to} \quad & Ax = b \\ & (x - x^0)^T (x - x^0) \leq \rho^2 \end{aligned} \tag{14}$$

where A is a matrix of order $m \times n$ and full row rank m. And x^0 is a point in the space $H = \{x : Ax = b\}$. Denoting the ellipsoid B we know that the center x^0 of B is in H and that $G \cap B$ is another ball which has center x^0 and radius ρ and is totally contained in H.

Since A is of full rank, the orthogonal projection of c^T into the subspace $\{x : Ax = 0\}$ is Pc^T where $P = I - A^T(AA^T)^{-1}A$ is the projection matrix.

Now consider the problem

$$\begin{aligned} \min \quad & z(x) = c^T x \\ \text{subject to} \quad & Ax = b \\ & (x - x^0)^T (X^0)^{-2} (x - x^0) \leq \rho^2 \end{aligned} \tag{15}$$

where A is a matrix of order $m \times n$ and full row rank m. And x^0 is a point in the space $H = \{x : Ax = b\}$.
Solving (15) is equivalent to minimizing cPx on $H \cap B$ and using the solution to the sphere problem (9) we get

$$x^0 - \rho \frac{Pc^T}{\|Pc^T\|} \tag{16}$$

Using the transformation of variables previously defined in(11), the problem becomes

$$\begin{aligned} \min \quad & z(x) = cX^0 y + \text{constant} \\ \text{subject to} \quad & AX^0 y = b \\ & (y-e)^T(y-e) \leq \rho^2 \end{aligned} \tag{17}$$

We know that the optimal y is given by

$$\bar{y} = e - \rho \frac{P_0 X^0 c^T}{\|P_0 X^0 c^T\|}$$

where P_0 is the projection matrix

$$P_0 = I - X^0 A^T (A(X^0)^2 A^T)^{-1} A X^0$$

is the projection matrix.

Using (12) we get the optimal point x

$$\bar{x} = X^0 \bar{y} = X^0 \left(e - \rho \frac{P_0 X^0 c^T}{\| P_0 X^0 c^T \|} \right) =$$
$$x^0 - \rho \left(e - \rho \frac{P_0 X^0 c^T}{\| P_0 X^0 c^T \|} \right)$$

If $\bar{x}_j^r = 0$ for atleast one $j = 1, ..., n$, then $\bar{x}_j^r$ is an optimum solution of (2).

If the tentative dual slack vector $s^r = c^T - A^T y^r \leq 0$ the objective function is unbounded in the problem (2).

If either of the above is satisfied the algorithm terminates. Else a step is taken in the direction given by $d^r = \bar{x}^r - x^r$.

The maximum step length θ_r is the maximum value that keeps

$$x_j^r + \theta d_j^r \geq 0, \quad \forall j.$$

It can be verified that that is

$$\theta_r = \min \left\{ j, \frac{\|X^r s^r\|}{x_j^r s_j^r} : s_j^r > 0 \right\}$$

$$x^{r+1} = x^r + \alpha \theta_r d^r, 0 < \alpha < 1.$$

Typically $\alpha = 0.95$

Newton's method is a method for solving non-linear equations.

$$x^{r+1} = x^r - (\nabla f(x^r))^{-1} f(x^r)$$

I the Jacobian is non-singular, the method takes a step of length 1 in this direction.

The central path-following primal-dual methods are some of the most popular methods.

$$\begin{aligned} \min \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{18}$$

And its dual

$$\begin{aligned} \max \quad & b^T y \\ \text{subject to} \quad & A^T y + s = c \\ & x, s \geq 0 \end{aligned} \tag{19}$$

The complementary slackness conditions are $x_j s_j = 0$.

Solving the LP is equivalent to finding a solution $(x, s) \geq 0$ to the following system of $(2n + m)$ equations with $(2n + m)$ unknown, if the constraint maxtrix A is $m \times n$.

$$F(x, y, s) = \begin{bmatrix} A^T y + s - c \\ Ax - b \\ XSe \end{bmatrix} = 0 \tag{20}$$

which is a non-linear system because of the last equation.

The central path is parametrized using $\mu > 0$. For each $\mu > 0$ the point $(x^\mu, y^\mu, s^\mu) \in \mathcal{C}$ satisfies $(x^\mu, s^\mu) > 0$ and

$$\begin{aligned} A^T y^\mu + s^\mu &= c^T \\ Ax^\mu &= b \\ x_j^\mu s_j^\mu &= \mu, \forall j = 1, ..., n \end{aligned}$$

If $\mu = 0$ this defines the optimality conditions.

Following the central path while decreasing μ to 0 implies that all complementary products $x_j s_j$ are equal. Thus μ is a measure of closeness to optimality.

However, two difficulties arise

- Finding a starting point on $\mathcal{C}$ with all $x_j s_j$ equal.
- $\mathcal{C}$ is a non-linear curve.

LINKÖPINGS UNIVERSITET

We introduce the measure of deviation from the central path $\mathcal{C}$ as

$$\frac{\|XSe - \mu e\|}{\mu}.$$

Usually the 2-norm is used with a parameter θ

$$\|XSe - \mu e\| \leq \mu,$$

where $0 < \theta < 1$, where 0.5 is a common value. By keeping all iterates within this kind of neighbourhood, path-following methods reduce all $x_j s_j$ to 0 at about the same rate.

LINKÖPINGS UNIVERSITET

Moving from the current point, we may only take a small step. Thus the non-negativity constraints $(x, s) \geq 0$ can be ignored when computing this step. To find the direction, we solve

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -XSe + \sigma\mu e \end{pmatrix} \tag{21}$$

where $\sigma = 1$ is the centering direction, which will take us towards the point $(x^\mu, y^\mu, s^\mu) \in \mathcal{C}$, but may produce a small improvement in objective value. Using $\sigma = 1$ is a pure Newton step. Algorithms usually choose a trade off in the interval $(0, 1)$.

$$\begin{aligned} \min \quad & \frac{1}{2}x^T x \\ \text{subject to} \quad & Ax = b \end{aligned} \tag{22}$$

$$\begin{aligned} \min \quad & \frac{1}{2}s^T s \\ \text{subject to} \quad & A^T y + s = c \end{aligned} \tag{23}$$

The solution of these problems can be written as

$$\begin{aligned} \bar{x} &= A^T(AA^T)^{-1}b \\ \bar{y} &= (AA^T)^{-1}Ac \\ \bar{s} &= c - A^T\bar{y} \end{aligned}$$

We must modify the solutions to avoid non-negative values of $\bar{x}, \bar{s}$.

$$\hat{x} = \bar{x} + \max(-3/2 \min \bar{x}_i, 0)e+$$
$$\hat{s} = \bar{s} + \max(-3/2 \min \bar{s}_i, 0)e$$

We then scalars to ensure the components are not too close to zero

$$x^0 = \hat{x} + \frac{1}{2}\frac{\hat{x}^T\hat{s}}{e^T\hat{x}}e$$
$$s^0 = \hat{s} + \frac{1}{2}\frac{\hat{s}^T\hat{x}}{e^T\hat{x}}e$$

The computational cost of finding this starting point is about the same as one step of the primal-dual method.

Most computational time is taken up solving the linear system (21). THe matrix in these systems are usually large and sparse, and the structure allows us to reformulate them as systems with more compact symmetric coefficient matrices which are easier to factor than the original one.

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta s \end{pmatrix} = \begin{pmatrix} -r_c \\ -r_b \\ -r_{xs} \end{pmatrix} \tag{24}$$

Since x and s are strictly positive, the diagonal matrices X and S are non-singular. We can eliminate Δs by $X\Delta s = -r_{xs} - S\Delta x$ and by multiplying with $-X^{-1}$ obtain

$$\Delta s = -X^{-1}r_{xs} - X^{-1}S\Delta x$$

Consider the third equation

$$S\Delta x + X\Delta s = -r_{xs}$$

Add $-X^{-1}$ times the third equation to the first equation.

$$X^{-1}S\Delta x + \underbrace{X^{-1}X}_{=I}\Delta s = -X^{-1}r_{xs}$$

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ X^{-1}S & 0 & -I \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta s \end{pmatrix} = \begin{pmatrix} -r_c \\ -r_b \\ X^{-1}r_{xs} \end{pmatrix}$$

$$\begin{pmatrix} -D^{-2} & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} -r_c + X^{-1} r_{xs} \\ -r_b \end{pmatrix} \tag{25}$$

where $D = S^{-1/2} X^{1/2}$. This is known as the agumented system. We can go further by eliminating Δx and adding AD^2 times the first equation to the second in order to cancel out the term $A\Delta x$.

$$\begin{pmatrix} -AD^2 D^{-2} & AD^2 A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} -AD^2 r_c + AD^2 X^{-1} r_{xs} \\ -r_b \end{pmatrix}$$

using the definition $D = S^{-1/2} X^{1/2}$

$$\begin{pmatrix} -A & AD^2 A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} -AXS^{-1} r_c + AS^{-1} r_{xs} \\ -r_b \end{pmatrix}$$

LINKÖPINGS UNIVERSITET

Adding equation one to equation two yields

$$\begin{pmatrix} -A & AD^2A^T \\ 0 & AD^2A^T \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} -AXS^{-1}r_c + AS^{-1}r_{xs} \\ -r_b - AXS^{-1}r_c + AS^{-1}r_{xs} \end{pmatrix}$$

$$\begin{aligned} D^2A^T\Delta y &= -r_b - AXS^{-1}r_c + AS^{-1}r_{xs} \\ \Delta s &= -r_c + A^T\Delta y \\ \Delta x &= -S^{-1}r_{xs} - XS^{-1}\Delta s \end{aligned} \tag{26}$$

where the expressions for Δs and Δx are obtained from the original system (24).

Most implementations of primal dual interior point solvers are based on formulations like (26), which are called the normal equations.

General purpose Choleskey software can be applied to AD^2A^T but modifications are needed because the matrix may be ill-conditioned or singular.

All path-following methods have been shown to be polynomial time algorithms (where simplex is worst case exponential). Each step requires a full matrix inversion, a rather expensive task for a large scale problem. However the number of steps are smaller in interior point methods than in the simplex method.

Compared to the simplex method

+ Better worst case complexity.
+ Easier to program.
- Cannot take advantage of warm start information as good.

Bibliography I

Katta G Murty. *Optimization for decision making*. Springer, 2009.

Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

LINKÖPINGS UNIVERSITET