Lecture: Algorithms for LP, SOCP and SDP

Zaiwen Wen

Beijing International Center For Mathematical Research
Peking University

`http://bicmr.pku.edu.cn/~wenzw/bigdata2018.html`
wenzw@pku.edu.cn

Acknowledgement: this slides is based on chapters 13 and 14 of “Numerical Optimization”, Jorge Nocedal and Stephen Wright, Springer
some parts are based on Prof. Farid Alizadeh lecture notes

Outline

1. Properties of LP
2. Primal Simplex method
3. Dual Simplex method
4. Interior Point method

Standard form LP

$$
\textbf{(P)} \quad \begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}
\qquad\qquad
\textbf{(D)} \quad \begin{aligned} \max \quad & b^\top y \\ \text{s.t.} \quad & A^\top y + s = c \\ & s \geq 0 \end{aligned}
$$

- KKT condition

$$
\begin{aligned}
Ax &= b, \quad x \geq 0 \\
A^\top y + s &= c, \quad s \geq 0 \\
x_i s_i &= 0 \quad \text{for } i = 1, \ldots, n
\end{aligned}
$$

- **Strong duality**: If a LP has an optimal solution, so does its dual, and their objective fun. are equal.

dual \ primal	finite	unbounded	infeasible
finite	$\surd$	$\times$	$\times$
unbounded	$\times$	$\times$	$\surd$
infeasible	$\times$	$\surd$	$\surd$

Geometry of the feasible set

- Assume that $A \in \mathbb{R}^{m\times n}$ has **full row rank.** Let A_i be the ith column of A:

$$A = \begin{pmatrix} A_1 & A_2 & \dots & A_n \end{pmatrix}$$

- A vector x is a **basic feasible solution (BFS)** if x is feasible and there exists a subset $\mathcal{B} \subset \{1, 2, \dots, n\}$ such that
  - $\mathcal{B}$ contains exactly m indices
  - $i \notin \mathcal{B} \implies x_i = 0$
  - The $m \times m$ submatrix $B = [A_i]_{i\in\mathcal{B}}$ is nonsingular

  $\mathcal{B}$ is called a basis and B is called the basis matrix

Properties:

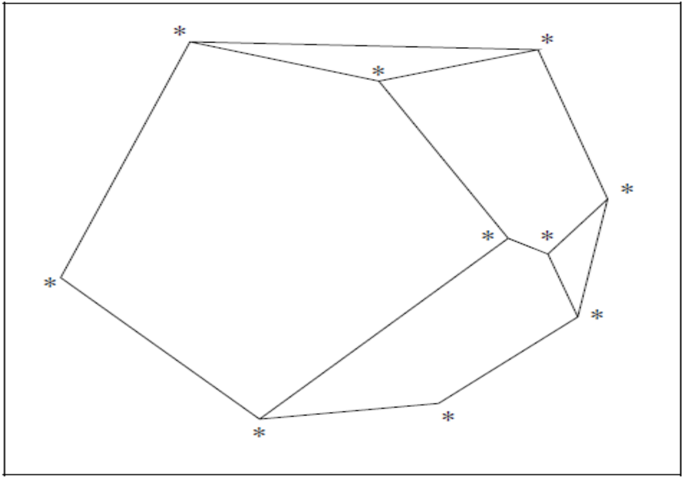
- If (P) has a nonempty feasible region, then there is at least one basic feasible point;
- If (P) has solutions, then at least one such solution is a basic optimal point.
- If (P) is feasible and bounded, then it has an optimal solution.

If (P) has a nonempty feasible region, then there is at least one BFS;

- Choose a feasible x with the minimal number (p) of nonzero x_i: $\sum_{i=1}^{p} A_i x_i = b$
- Suppose that $A_1, \ldots, A_p$ are linearly dependent $A_p = \sum_{i=1}^{p-1} z_i A_i$. Let $x(\epsilon) = x + \epsilon(z_1, \ldots, z_{p-1}, -1, 0, \ldots, 0)^\top = x + \epsilon z$. Then $Ax(\epsilon) = b$, $x_i(\epsilon) > 0$, $i = 1, \ldots, p$, for ϵ sufficiently small. There exists $\bar{\epsilon}$ such that $x_i(\bar{\epsilon}) = 0$ for some $i = 1, \ldots, p$. Contradiction to the choice of x.
- If $p = m$, done. Otherwise, choose $m - p$ columns from among $A_{p+1}, \ldots, A_n$ to build up a set set of m linearly independent vectors.

Polyhedra, extreme points, vertex, BFS

- A **Polyhedra** is a set that can be described in the form $\{x \in \mathbb{R}^n \mid Ax \geq b\}$
- Let P be a polyhedra. A vector $x \in P$ is an **extreme point** if we cannot find two vectors $y, z \in P$ (both different from x) such that $x = \lambda y + (1 - \lambda)z$ for $\lambda \in [0, 1]$
- Let P be a polyhedra. A vector $x \in P$ is a **vertex** if there exists some c such that $c^\top x < c^\top y$ for all $y \in P$ and $y \neq x$
- Let P be a nonempty polyhedra. Let $x \in P$. The following statements are equivalent: (i) x is vertex; (ii) x is an extreme point; (iii) x is a BFS
- A basis $\mathcal{B}$ is said to be **degenerate** if $x_i = 0$ for some $i \in \mathcal{B}$, where x is the BFS corresponding to $\mathcal{B}$. A linear program (P) is said to be degenerate if it has at least one degenerate basis.



Vertices of a three-dimensional polyhedron (indicated by $*$)

Outline

The Simplex Method For LP

Basic Principle

Move from a BFS to its adjacent BFS unitil convergence (either optimal or unbounded)

- Let x be a BFS and $\mathcal{B}$ be the corresponding basis
- Let $\mathcal{N} = \{1, 2, \ldots, n\} \backslash \mathcal{B}$, $N = [A_i]_{i \in \mathcal{N}}$, $x_B = [x_i]_{i \in \mathcal{B}}$ and $x_N = [x_i]_{i \in \mathcal{N}}$
- Since x is a BFS, then $x_N = 0$ and $Ax = Bx_B + Nx_N = b$:

$$x_B = B^{-1} b$$

- Find exactly one $q \in \mathcal{N}$ and exactly one $p \in \mathcal{B}$ such that

$$\mathcal{B}^+ = \{q\} \cup (\mathcal{B} \backslash \{p\})$$

Finding $q \in \mathcal{N}$ to enter the basis

Let x^+ be the new BFS:

$$x^+ = \begin{pmatrix} x^+_{\mathcal{B}} \\ x^+_{\mathcal{N}} \end{pmatrix}, \quad Ax^+ = b \Longrightarrow x^+_{\mathcal{B}} = B^{-1}b - B^{-1}Nx^+_{\mathcal{N}}$$

The cost at x^+ is

$$\begin{aligned} c^\top x^+ &= c_B^\top x^+_{\mathcal{B}} + c_N^\top x^+_{\mathcal{N}} \\ &= c_B^\top B^{-1}b - c_B^\top B^{-1}Nx^+_{\mathcal{N}} + c_N^\top x^+_{\mathcal{N}} \\ &= c^\top x + (c_N^\top - c_B^\top B^{-1}N)x^+_{\mathcal{N}} \\ &= c^\top x + \sum_{j \in \mathcal{N}} \underbrace{(c_j - c_B^\top B^{-1}A_j)}_{s_j} x^+_j \end{aligned}$$

- s_j is also called **reduced cost.** It is actually the dual slackness
- If $s_j \geq 0$, $\forall j \in \mathcal{N}$, then x is optimal as $c^\top x^+ \geq c^\top x$
- Otherwise, find q such that $s_q < 0$. Then $c^\top x^+ = c^\top x + s_q x^+_q \leq c^\top x$

Finding $p \in \mathcal{B}$ to exit the basis

What is x^+: select $q \in \mathcal{N}$ and $p \in \mathcal{B}$ such that

$$x_{\mathcal{B}}^+ = B^{-1}b - B^{-1}A_q x_q^+, \quad x_q^+ \geq 0, x_p^+ = 0, x_j^+ = 0, j \in \mathcal{N} \setminus \{q\}$$

Let $u = B^{-1}A_q$. Then $x_{\mathcal{B}}^+ = x_{\mathcal{B}} - u x_q^+$

- If $u \leq 0$, then $c^\top x^+ = c^\top x + s_q x_q^+ \to -\infty$ as $x_q^+ \to +\infty$ and x^+ is feasible. **(P) is unbounded**

- If $\exists u_k > 0$, then find x_q^+ and p such that

  $$x_{\mathcal{B}}^+ = x_{\mathcal{B}} - u x_q^+ \geq 0, \quad x_p^+ = 0$$

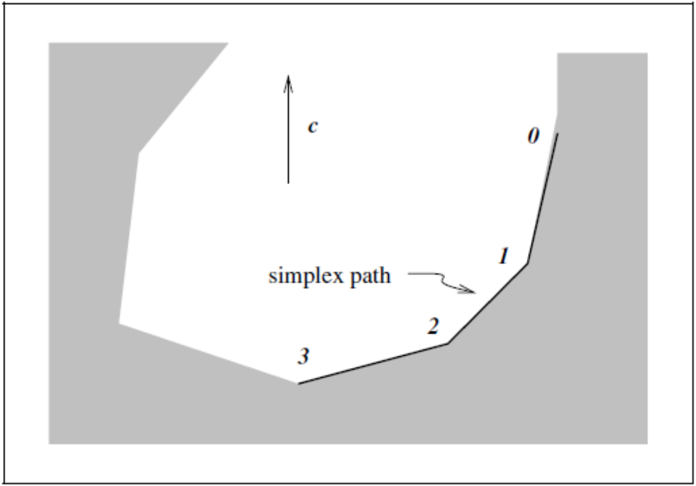
  Let p be the index corresponding to

  $$x_q^+ = \min_{i=1,\ldots,m | u_i > 0} \frac{x_{\mathcal{B}(i)}}{u_i}$$

An iteration of the simplex method

Typically, we start from a BFS x and its associate basis $\mathcal{B}$ such that $x_B = B^{-1}b$ and $x_N = 0$.

- Solve $y^\top = c_B^\top B^{-1}$ and then the reduced costs $s_N = c_N - N^\top y$
- If $s_N \geq 0$, x is optimal and stop; Else, choose $q \in \mathcal{N}$ with $s_q < 0$.
- Compute $u = B^{-1}A_q$. If $u \leq 0$, then (P) is unbounded and stop.
- If $\exists u_k > 0$, then find $x_q^+ = \min_{i=1,\dots,m | u_i > 0} \frac{x_{\mathcal{B}(i)}}{u_i}$ and use p to denote the minimizing i. Set $x_{\mathcal{B}}^+ = x_{\mathcal{B}} - u x_q^+$.
- Change $\mathcal{B}$ by adding q and removing the basic variable corresponding to column p of B.



Simplex iterates for a two-dimensional problem

Finite Termination of the simplex method

Theorem

Suppose that the LP (P) is nondegenerate and bounded, the simplex method terminates at a basic optimal point.

- nondegenerate: $x_{\mathcal{B}} > 0$ and $c^\top x$ is bounded
- A strict reduction of $c^\top x$ at each iteration
- There are only a finite number of BFS since the number of possible bases $\mathcal{B}$ is finite (there are only a finite number of ways to choose a subset of m indices from $\{1, 2, \ldots, n\}$), and since each basis defines a single basic feasible point

Finite termination does not mean a polynomial-time algorithm

Linear algebra in the simplex method

- Given B^{-1}, we need to compute $\bar{B}^{-1}$, where

$$B = [A_1, \ldots, A_m], \quad \bar{B} := B^+ = [A_1, \ldots, A_{p-1}, A_q, A_{p+1}, \ldots, A_m]$$

- the cost of inversion $\bar{B}^{-1}$ from scratch is $O(m^3)$
- Since $BB^{-1} = I$, we have

$$\begin{aligned} B^{-1}\bar{B} &= [e_1, \ldots e_{p-1}, u, e_{p+1}, \ldots, e_m] \\ &= \begin{pmatrix} 1 & & u_1 & & \\ & \ddots & \vdots & & \\ & & u_p & & \\ & & \vdots & \ddots & \\ & & u_m & & 1 \end{pmatrix}, \end{aligned}$$

where e_i is the ith column of I and $u = B^{-1}A_q$

Linear algebra in the simplex method

- Apply a sequence of "elementary row operation"
  - For each $j \neq p$, we add the p-th row times $-\frac{u_j}{u_p}$ to the jth row. This replaces u_j by zero.
  - We divide the pth row by u_p. This replaces u_p by one.

$$Q_{ip} = I + D_{ip}, \quad (D_{ip})_{jl} = \begin{cases} -\frac{u_j}{u_p}, & (j,l) = (i,p) \\ 0, & \text{otherwise} \end{cases}, \text{ for } i \neq p$$

- Find Q such that $QB^{-1}\bar{B} = I$. Computing $\bar{B}^{-1}$ needs only $O(m^2)$
- What if B^{-1} is computed by the LU factorization, i.e., $B = LU$?
  L is is unit lower triangular, U is upper triangular.
  Read section 13.4 in "Numerical Optimization", Jorge Nocedal and Stephen Wright,

An iteration of the revised simplex method

Typically, we start from a BFS x and its associate basis $\mathcal{B}$ such that $x_B = B^{-1}b$ and $x_N = 0$.

- Solve $y^\top = c_B^\top B^{-1}$ and then the reduced costs $s_N = c_N - N^\top y$
- If $s_N \geq 0$, x is optimal and stop; Else, choose $q \in \mathcal{N}$ with $s_q < 0$.
- Compute $u = B^{-1}A_q$. If $u \leq 0$, then (P) is unbounded and stop.
- If $\exists u_k > 0$, then find $x_q^+ = \min\limits_{i=1,\dots,m|u_i>0} \frac{x_{\mathcal{B}(i)}}{u_i}$ and use p to denote the minimizing i. Set $x_{\mathcal{B}}^+ = x_{\mathcal{B}} - u x_q^+$.
- Form the $m \times (m+1)$ matrix $[B^{-1} \mid u]$. Add to each one of its rows a multiple of the pth row to make the last column equal to the unit vector e_p. The first m columns of the result is the matrix $\bar{B}^{-1}$.

Selection of the entering index (pivoting rule)

Reduced costs $s_N = c_N - N^\top y$, $c^\top x^+ = c^\top x + s_q x_q^+$

- Dantzig: chooses $q \in \mathcal{N}$ such that s_q is the most negative component
- Bland's rule: choose the smallest $j \in \mathcal{N}$ such that $s_j < 0$; out of all variables x_i that are tied in the test for choosing an exiting variable, select the one with with the smallest value i.
- Steepest-edge: choose $q \in \mathcal{N}$ such that $\frac{c^\top \eta_q}{\|\eta_q\|}$ is minimized, where

$$x^+ = \begin{pmatrix} x_B^+ \\ x_N^+ \end{pmatrix} = \begin{pmatrix} x_B \\ x_N \end{pmatrix} + \begin{pmatrix} -B^{-1}A_q \\ e_q \end{pmatrix} x_q = x + \eta_q x_q^+$$

  efficient computation of this rule is available

Degenerate steps and cycling

Let q be the entering variable:

$$x_{\mathcal{B}}^{+} = B^{-1}b - B^{-1}A_q x_q^{+} = x_B - x_q^{+}u, \text{ where } u = B^{-1}A_q$$

- Degenerate step: there exists $i \in \mathcal{B}$ such that $x_i = 0$ and $u_i > 0$. Then $x_i^{+} < 0$ if $x_q^{+} > 0$. Hence, $x_q^{+} = 0$ and do the pivoting
- Degenerate step may still be useful because they change the basis $\mathcal{B}$, and the updated $\mathcal{B}$ may be closer to the optimal basis.
- cycling: after a number of successive degenerate steps, we may return to an earlier basis $\mathcal{B}$
- Cycling has been observed frequently in the large LPs that arise as relaxations of integer programming problems
- Avoid cycling: Bland's rule and Lexicographically pivoting rule

Finding an initial BFS

The two-phase simplex method

$$\textbf{(P)} \quad \begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \qquad \textbf{(P0)} \quad \begin{aligned} \tilde{f} = \min \quad & z_1 + z_2 + \ldots + z_m \\ \text{s.t.} \quad & Ax + z = b \\ & x \geq 0, \quad z \geq 0 \end{aligned}$$

- A BFS to (P0): $x = 0$ and $z = b$
- If x is feasible to (P), then $(x, 0)$ is feasible to (P0)
- If the optimal cost $\tilde{f}$ of (P0) is nonzero, then (P) is infeasible
- If $\tilde{f} = 0$, then its optimal solution must satisfies: $z = 0$ and x is feasible to (P)
- An optimal basis $\mathcal{B}$ to (P0) may contain some components of z

Finding an initial BFS

(x, z) is optimal to (P0) with some components of z in the basis

- Assume $A_1, \ldots, A_k$ are in the basis matrix with $k < m$. Then

$$B = [A_1, \ldots, A_k \mid \text{ some columns of } I]$$

$$B^{-1}A = [e_1, \ldots, e_k, B^{-1}A_{k+1}, \ldots, B^{-1}A_n]$$

- Suppose that ℓth basic variable is an artificial variable

- If the ℓth row of $B^{-1}A$ is zero, then $g^\top A = 0^\top$, where $g^\top$ is the ℓth row of B^{-1}. If $g^\top b \neq 0$, (P) is infeasible. Otherwise, A has linearly dependent rows. Remove the ℓth row.

- There exists j such that the ℓth entry of $B^{-1}A_j$ is nonzero. Then A_j is linearly independent to $A_1, \ldots, A_k$. Perform elementary row operation to replace $B^{-1}A_j$ to be the ℓth unit vector. Driving one of z out of the basis

The primal simplex method for LP

$$\textbf{(P)}\quad \begin{array}{ll} \min & c^\top x \\ \text{s.t.} & Ax = b \\ & x \ge 0 \end{array} \qquad \textbf{(D)}\quad \begin{array}{ll} \max & b^\top y \\ \text{s.t.} & A^\top y + s = c \\ & s \ge 0 \end{array}$$

- KKT condition

$$\begin{array}{rcll} Ax & = & b, & x \ge 0 \\ A^\top y + s & = & c, & s \ge 0 \\ x_i s_i & = & 0 & \text{for } i = 1, \ldots, n \end{array}$$

- The primal simplex method generates

$$\begin{array}{rcl} x_B & = & B^{-1} b \ge 0, \quad x_N = 0, \\ y & = & B^{-T} c_B, \\ s_B & = & c_B - B^\top y = 0, s_N = c_N - N^\top y ? 0 \end{array}$$

Outline

The dual simplex method for LP

- The dual simplex method generates

$$\begin{aligned} x_B &= B^{-1}b ? 0, \quad x_N = 0, \\ y &= B^{-T}c_B, \\ s_B &= c_B - B^\top y = 0, s_N = c_N - N^\top y \geq 0 \end{aligned}$$

- If $x_B \geq 0$, then (x, y, s) is optimal
- Otherwise, select $q \in \mathcal{B}$ such that $x_q < 0$ to exit the basis, select $r \in \mathcal{N}$ to enter the basis, i.e., $s_r^+ = 0$
- The update is of the form

$$\begin{aligned} s_B^+ &= s_B + \alpha e_q \quad \text{obvious} \\ y^+ &= y + \alpha v \quad \text{requirement} \end{aligned}$$

The dual simplex method for LP

- What is v? Since $A^\top y^+ + s^+ = c$, it holds

$$s_B^+ = c_B - B^\top y^+$$
$$\Longrightarrow \quad s_B + \alpha e_q = c_B - B^\top (y + \alpha v) \Longrightarrow e_q = -B^\top v$$

- The update of the dual objective function

$$\begin{aligned} b^\top y^+ &= b^\top y + \alpha b^\top v \\ &= b^\top y - \alpha b^\top B^{-T} e_q \\ &= b^\top y - \alpha x_B^\top e_q \\ &= b^\top y - \alpha x_q \end{aligned}$$

- Since $x_q < 0$ and we maximize $b^\top y^+$, we choose α as large as possible, but require $s_N^+ \geq 0$

The dual simplex method for LP

- Let $w = N^\top v = -N^\top B^{-T} e_q$. Since $Ay + s = c$ and $A^\top y^+ + s^+ = c$, it holds

$$s_N^+ = c_N - N^\top y^+ = s_N - \alpha N^\top v = s_N - \alpha w \geq 0$$

- The largest α is

$$\alpha = \min_{j \in \mathcal{N}, w_j > 0} \frac{s_j}{w_j}.$$

Let r be the index at which the minimum is achieved.

$$s_r^+ = 0, \quad w_r = A_r^\top v > 0$$

- (D) is unbounded if $w \leq 0$

The dual simplex method for LP: update of x^+

We have: $Bx_B = b$, $x_q^+ = 0$, $x_r^+ = \gamma$ and $Ax^+ = b$, i.e.,

$$Bx_{\mathcal{B}}^+ + \gamma A_r = b \Longrightarrow x_{\mathcal{B}}^+ = B^{-1}b - \gamma B^{-1}A_r,$$

where $Bd = A_r$. Then $Ax^+ = b$ gives

$$B(x_{\mathcal{B}} - \gamma d) + \gamma A_r = b \text{ for any } \gamma.$$

Since it is required $x_q^+ = 0$, we set

$$\gamma = \frac{x_q}{d_q}, \text{ where } d_q = d^\top e_q = A_r^\top B^{-T} e_q = -A_r^\top v = -w_r < 0.$$

Therefore

$$x_i^+ = \begin{cases} x_i - \gamma d_i, & \text{for } i \in \mathcal{B} \text{ with } i \neq q, \\ 0, & i = q, \\ 0, & i \in \mathcal{N} \text{ with } i \neq r, \\ \gamma, & i = r \end{cases}$$

An iteration of the dual simplex method

Typically, we start from a dual feasible (y, s) and its associate basis $\mathcal{B}$ such that $x_B = B^{-1}b$ and $x_N = 0$.

- If $x_B \geq 0$, then x is optimal and stop. Else, choose q such that $x_q < 0$.
- Compute $v = -B^{-T}e_q$ and $w = N^\top v$. If $w \leq 0$, then (D) is unbounded and stop.
- If $\exists w_k > 0$, then find $\alpha = \min_{j \in \mathcal{N}, w_j > 0} \frac{s_j}{w_j}$ and use r to denote the minimizing j. Set $s_B^+ = s_B + \alpha e_q$, $s_N^+ = s_N - \alpha w$ and $y^+ = y + \alpha v$.
- Change $\mathcal{B}$ by adding r and removing the basic variable corresponding to column q of B.

Outline

Primal-Dual Methods for LP

$$
\textbf{(P)} \quad \begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \qquad\qquad \textbf{(D)} \quad \begin{aligned} \max \quad & b^\top y \\ \text{s.t.} \quad & A^\top y + s = c \\ & s \geq 0 \end{aligned}
$$

- KKT condition

$$
\begin{aligned} Ax &= b, \quad x \geq 0 \\ A^\top y + s &= c, \quad s \geq 0 \\ x_i s_i &= 0 \quad \text{for } i = 1, \ldots, n \end{aligned}
$$

- Perturbed system

$$
\begin{aligned} Ax &= b, \quad x \geq 0 \\ A^\top y + s &= c, \quad s \geq 0 \\ x_i s_i &= \sigma \mu \quad \text{for } i = 1, \ldots, n \end{aligned}
$$

Newton's method

- Let (x, y, s) be the current estimate with $(x, s) > 0$
- Let $(\Delta x, \Delta y, \Delta s)$ be the search direction
- Let $\mu = \frac{1}{n} x^\top s$ and $\sigma \in (0, 1)$. Hope to find

$$
\begin{aligned}
A(x + \Delta x) &= b \\
A^\top (y + \Delta y) + s + \Delta s &= c \\
(x_i + \Delta x_i)(s_i + \Delta s_i) &= \sigma \mu
\end{aligned}
$$

- dropping the nonlinaer term $\Delta x_i \Delta s_i$ gives

$$
\begin{aligned}
A \Delta x &= r_p := b - Ax \\
A^\top \Delta y + \Delta s &= r_d := c - A^\top y - s \\
x_i \Delta s_i + \Delta x_i s_i &= (r_c)_i := \sigma \mu - x_i s_i
\end{aligned}
$$

Newton's method

- Let $\mathsf{L}_x = \mathrm{Diag}(x)$ and $\mathsf{L}_s = \mathrm{Diag}(s)$. The matrix form is:

$$\begin{pmatrix} A & 0 & 0 \\ 0 & A^\top & I \\ \mathsf{L}_s & 0 & \mathsf{L}_x \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta s \end{pmatrix} = \begin{pmatrix} r_p \\ r_d \\ r_c \end{pmatrix}$$

- Solving this system we get

$$\begin{aligned} \Delta y &= (A\mathsf{L}_s^{-1}\mathsf{L}_x A^\top)^{-1}(r_p + A\mathsf{L}_s^{-1}(\mathsf{L}_x r_d - r_c)) \\ \Delta s &= r_d - A^\top \Delta y \\ \Delta x &= -\mathsf{L}_s^{-1}(\mathsf{L}_x \Delta s - r_c) \end{aligned}$$

- The matrix $A\mathsf{L}_s^{-1}\mathsf{L}_x A^\top$ is symmetric and positive definite if A is full rank

The Primal-Dual Path-following Method

Given (x^0, y^0, s^0) with $(x^0, s^0) \geq 0$. A typical iteration is

- Choose $\mu = (x^k)^\top s^k/n$, $\sigma \in (0,1)$ and solve

$$\begin{pmatrix} A & 0 & 0 \\ 0 & A^\top & I \\ \mathsf{L}_{s^k} & 0 & \mathsf{L}_{x^k} \end{pmatrix} \begin{pmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{pmatrix} = \begin{pmatrix} r_p^k \\ r_d^k \\ r_c^k \end{pmatrix}$$

- Set

$$(x^{k+1}, y^{k+1}, s^{k+1}) = (x^k, y^k, s^k) + \alpha_k(\Delta x^k, \Delta y^k, \Delta s^k),$$

  choosing α_k such that $(x^{k+1}, s^{k+1}) > 0$

The choices of centering parameter σ and step length α_k are crucial to the performance of the method.

The Central Path

- The primal-dual feasible and strictly feasible sets:

$$\begin{aligned} \mathcal{F} &= \{(x,y,s) \mid Ax = b, A^\top y + s = c, (x,s) \geq 0\} \\ \mathcal{F}^o &= \{(x,y,s) \mid Ax = b, A^\top y + s = c, (x,s) > 0\} \end{aligned}$$

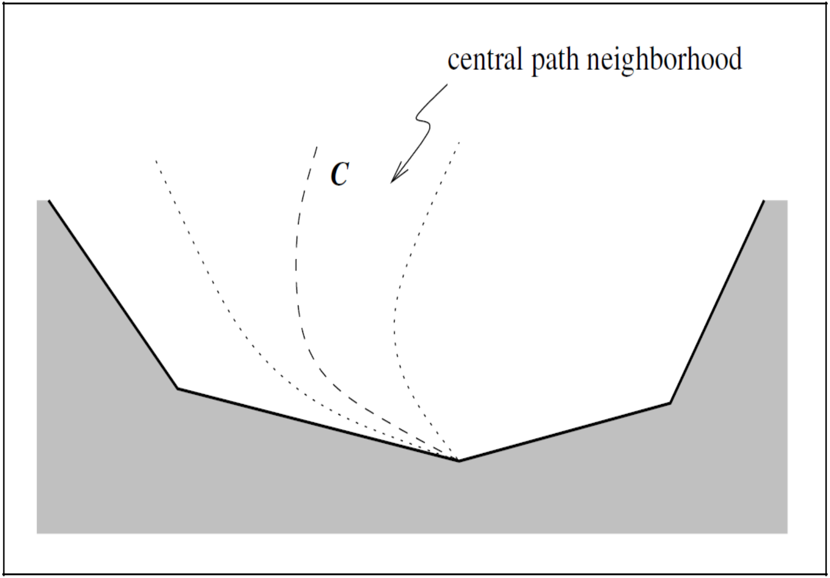
- The central path is $\mathcal{C} = \{(x_\tau, y_\tau, s_\tau) \mid \tau > 0\}$, where

$$\begin{aligned} Ax_\tau &= b, \quad x_\tau > 0 \\ A^\top y_\tau + s_\tau &= c, \quad s_\tau > 0 \\ (x_\tau)_i (s_\tau)_i &= \tau \quad \text{for } i = 1, \ldots, n \end{aligned}$$

- Central path neighborhoods, for $\theta, \gamma \in [0, 1)$:

$$\begin{aligned} \mathcal{N}_2(\theta) &= \{(x,y,s) \in \mathcal{F}^o \mid \|\mathsf{L}_x \mathsf{L}_s e - \mu e\|_2 \leq \theta\mu\} \\ \mathcal{N}_{-\infty}(\gamma) &= \{(x,y,s) \in \mathcal{F}^o \mid x_i s_i \geq \gamma\mu\} \end{aligned}$$

Tyically, $\theta = 0.5$ and $\gamma = 10^{-3}$



Central path, projected into space of primal variables x, showing a typical neighborhood $\mathcal{N}$

The Long-Step Path-following Method

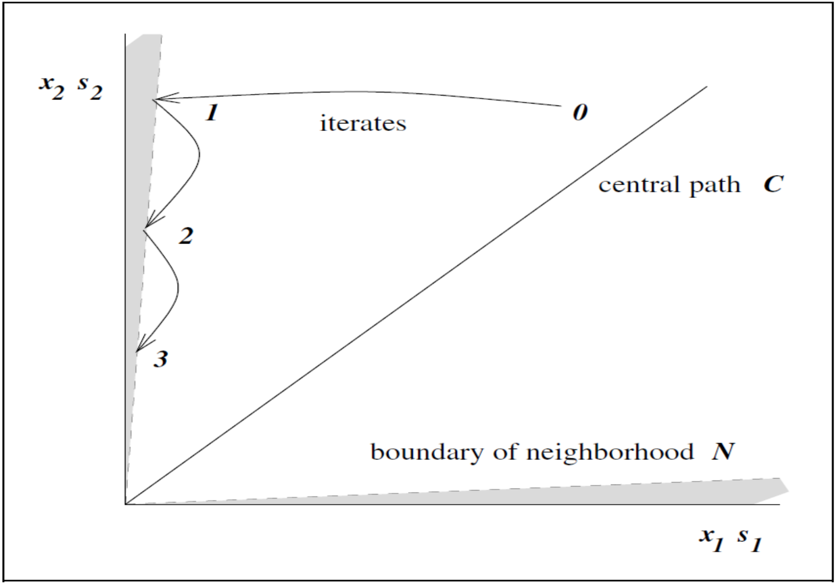
Given $(x^0, y^0, s^0) \in \mathcal{N}_{-\infty}(\gamma)$. A typical iteration is

- Choose $\mu = (x^k)^\top s^k/n$, $\sigma \in (0,1)$ and solve

$$\begin{pmatrix} A & 0 & 0 \\ 0 & A^\top & I \\ \mathsf{L}_{s^k} & 0 & \mathsf{L}_{x^k} \end{pmatrix} \begin{pmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{pmatrix} = \begin{pmatrix} r_p^k \\ r_d^k \\ r_c^k \end{pmatrix}$$

- Set α_k be the largest value of $\alpha \in [0,1]$ such that $(x^{k+1}, y^{k+1}, s^{k+1}) \in \mathcal{N}_{-\infty}(\gamma)$ where

$$(x^{k+1}, y^{k+1}, s^{k+1}) = (x^k, y^k, s^k) + \alpha_k(\Delta x^k, \Delta y^k, \Delta s^k),$$


$x_2\ s_2$
1
iterates
0
2
3
central path C
boundary of neighborhood N
$x_1\ s_1$

Analysis of Primal-Dual Path-Following

1. If $(x, y, s) \in \mathcal{N}_{-\infty}(\gamma)$, then $\|\Delta x \circ \Delta s\| \leq 2^{-3/2}(1 + 1/\gamma)n\mu$
2. The long-step path-following method yields
$$\mu_{k+1} \leq \left(1 - \frac{\delta}{n}\right)\mu_k,$$
where $\delta = 2^{3/2}\gamma\frac{1-\gamma}{1+\gamma}\sigma(1-\sigma)$
3. Given $\epsilon, \gamma \in (0, 1)$, suppose that the starting point $(x^0, y^0, s^0) \in \mathcal{N}_{-\infty}(\gamma)$. Then there exists $K = O(nlog(1/\epsilon))$ such that
$$\mu_k \leq \epsilon\mu_0, \quad \text{for all } k \geq K$$

Proof of 3:

$$\begin{aligned} \log(\mu_{k+1}) &\leq \log\left(1 - \frac{\delta}{n}\right) + \log(\mu_k) \\ \log(1+\beta) &\leq \beta, \quad \forall\beta > -1 \end{aligned}$$

Barrier Methods

A general strategy for solving convex optimization problem:

$$
\begin{array}{lll}
(P) & \min & c^\top x \\
& \text{s.t.} & x \in C,
\end{array}
$$

where C is convex. Find a barrier function $b(x)$: Int $C \to \mathbb{R}$

- $b(x)$ is convex on Int C
- for any sequence of points $\{x_i\}$ approaching boundary bd(C), $b(x_i) \to \infty$
- We can replace the problem

$$
(II) \quad \min \quad c^\top x + \mu b(x)
$$

- If x_μ is the optimum of (II) and x^* of (I) then
  - $x_\mu \in$ Int$C$
  - As $\mu \to 0$, $x_\mu \to x^*$

- For the positive orthant $\{x \mid x \geq 0\}$, a barrier is

$$b(x) = -\sum_i \ln(x_i)$$

- For the semidefinite cone $\{X \mid X \succeq 0\}$, a barrier is

$$b(x) = -\ln\det(X)$$

- We will discuss the second order cone shortly

Barriers for LP and SDP

Thus LP can be replaced by

Primal $(\mathrm{P})_\mu$

$$\begin{aligned} \min \quad & c^\top x - \mu \sum_i \ln x_i \\ \text{s.t.} \quad & Ax = b \\ & x > 0 \end{aligned}$$

Dual $(\mathrm{D})_\mu$

$$\begin{aligned} \max \quad & b^\top y + \mu \sum_i \ln s_i \\ \text{s.t.} \quad & A^\top y + s = c \\ & s > 0 \end{aligned}$$

Thus SDP can be replaced by

Primal $(\mathrm{P})_\mu$

$$\begin{aligned} \min \quad & \langle C, X \rangle - \mu \ln \det(X) \\ \text{s.t.} \quad & \langle A_i, X \rangle = b_i \\ & X \succ 0 \end{aligned}$$

Dual $(\mathrm{D})_\mu$

$$\begin{aligned} \max \quad & b^\top y + \mu \ln \det(S) \\ \text{s.t.} \quad & \sum_i y_i A_i + S = C \\ & S \succ 0 \end{aligned}$$

Applying standard optimality condition we get

- LP: $\mathcal{L}(x,y) = c^\top x - \mu \sum_i \ln x_i - y^\top (b - Ax)$
- SDP: $\mathcal{L}(x,y) = \langle C, X \rangle - \mu \ln \det(X) - \sum_i y_i (b_i - \langle A_i, X \rangle)$

The Karush-Kuhn-Tucker condition requires that at the optimum

$$\nabla_X \mathcal{L} = 0$$

which translates into

(LP)

$$\nabla_y \mathcal{L} = b - Ax = 0$$

$$\frac{\partial \mathcal{L}}{\partial x_i} = c_i - \frac{\mu}{x_i} - (y^\top A)_i = 0$$

(SDP)

$$\nabla_y \mathcal{L} = (b_i - \langle A_i, X \rangle) = 0$$

$$\nabla_X = C - \mu X^{-1} - \sum_i y_i A_i = 0$$

- In LP: define $s_i = \frac{\mu}{x_i}$, then s is dual feasible
- In SDP: define $S = \mu S^{-1}$, then S is dual feasible

The optimality conditions result in the square system

(LP)	(SDP)
$Ax = b$	$\langle A_i, X\rangle = b_i$
$A^\top y + s = c$	$\sum_i y_i A_i + S = C$
$x_i = \frac{\mu}{s_i}$	$X = \mu S^{-1}$

- In LP: if we write $x_i s_i = \mu$, we get relaxed complementarity
- In SDP: if we write $XS = \mu I$, we get relaxed complementarity

Newton's method for SDP

- Let X, y, S be initial estimates Then
- If we use $XS = \mu I$, ΔX is not symmetric
- Since $X, S \succ 0$ then $XS = \mu I$ iff $X \circ S = \frac{XS+SX}{2} = \mu I$
- Now applying Newton, we get

$$
\begin{aligned}
\langle A_i, X + \Delta X \rangle &= b_i \\
\sum_i (y_i + \Delta y_i) A_i + S + \Delta S &= C \\
(X + \Delta X) \circ (S + \Delta S) &= \mu I
\end{aligned}
$$

Newton's method

Expanding and throwing out nonlinear terms

$$\begin{aligned} \langle A_i, \Delta X\rangle &= (r_p)_i \\ \sum_i \Delta y_i A_i + \Delta S &= R_d \\ S \circ \Delta X + \Delta S \circ X &= R_c \end{aligned}$$

where

$$\begin{aligned} (r_p)_i &= b_i - \langle A_i, X\rangle \\ R_d &= C - \sum_i y_i A_i - S \\ R_c &= \mu I - X \circ S \end{aligned}$$

- In matrix form

$$\begin{pmatrix} \mathcal{A} & 0 & 0 \\ 0 & \mathcal{A}^\top & I \\ \mathcal{L}_S & 0 & \mathcal{L}_X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta s \end{pmatrix} = \begin{pmatrix} r_p \\ r_d \\ r_c \end{pmatrix}$$

- $vec(A)$ is a vector made by stacking columns of a matrix A
- $\mathcal{A}$ is a matrix whose rows are $vec(A_i)$
- $x = vec(X)$, $s = vec(S)$ ...
- $\mathcal{L}_X$ (and $\mathcal{L}_S$) are matrix representations of L_X (and L_S) operators
- $\mathcal{L}_X = X \otimes I + I \otimes X$ and $\mathcal{L}_S = S \otimes I + I \otimes S$
- Kronecker product: $A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix}$

  http://en.wikipedia.org/wiki/Kronecker_product

- Solving this system we get

$$\begin{aligned}
\Delta y &= (\mathcal{A}\mathcal{L}_S^{-1}\mathcal{L}_X\mathcal{A}^\top)^{-1}(r_p + \mathcal{A}\mathcal{L}_S^{-1}(\mathcal{L}_X r_d - r_c)) \\
\Delta s &= r_d - \mathcal{A}^\top \Delta y \\
\Delta x &= -\mathcal{L}_S^{-1}(\mathcal{L}_X \Delta s - r_c)
\end{aligned}$$

- The matrix $\mathcal{A}\mathcal{L}_s^{-1}\mathcal{L}_x\mathcal{A}^\top$ is not symmetric because $\mathcal{L}_S$ and $\mathcal{L}_X$ do not commute!
- In LP, it is quite easy to compute $A\mathsf{L}_s^{-1}\mathsf{L}_x A^\top$
- Most computational work in LP involves solving the system

$$(A\mathsf{L}_s^{-1}\mathsf{L}_x A^\top)v = u$$

- in SDP even computing $\mathcal{A}\mathcal{L}_s^{-1}\mathcal{L}_x\mathcal{A}^\top$ is fairly expensive (in this form requires solving Lyapunov equations)

- How about SOCP?
- What is an appropriate barrier for the convex cone

$$\mathcal{Q} = \{x \mid x_0 \geq \|\bar{x}\|\}?$$

- By analogy we expect relaxed complementary conditions turn out to be $x \circ s = \mu e$

Algebra Associated with SOCP

In SDP

- The barrier $\ln\det(X) = \sum_i \ln\lambda_i(X)$
- For each symmetric $n \times n$ matrix X, there is a characteristic polynomial, such that
  - $p(t) = p_0 + p_1 t + \ldots + p_{n-1}t^{n-1} + t^n$
  - roots of $p(t)$ are eigenvalues of X
  - $\mathrm{Tr}(X) = p_{n-1}$, $\det(X) = p_0$
  - roots of $p(t)$ are real numbers
  - $p(X) = 0$ by Cayley-Hamilton Theorem
  - There is orthogonal matrix Q: $X = Q\Lambda Q^\top = \lambda_1 q_1 q_1^\top + \ldots + \lambda_n q_n q_n^\top$

SOCP

- Remember

$$x \circ s = \begin{pmatrix} x^\top s \\ x_0 \bar{s} + s_0 \bar{x} \end{pmatrix} \quad \mathsf{L}_x = \mathrm{Arw}(x) = \begin{pmatrix} x_0 & \bar{x}^\top \\ \bar{x} & x_0 I \end{pmatrix} \quad e = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

- It is easy to verify

$$x \circ x - 2x_0 x + (x_0^2 - \|\bar{x}\|^2)e = 0$$

- Define the characteristic polynomial

$$p(t) = t^2 - 2x_0 t + (x_0^2 - \|\bar{x}\|^2) = (t - (x_0 + \|\bar{x}\|)(t - (x_0 - \|\bar{x}\|)$$

- Define eigenvalues of x roots of $p(t)$: $\lambda_{1,2} = x_0 \pm \|\bar{x}\|$
- Define $\mathrm{Tr}(x) = 2x_0$ and $\det(x) = x_0^2 - \|\bar{x}\|^2$

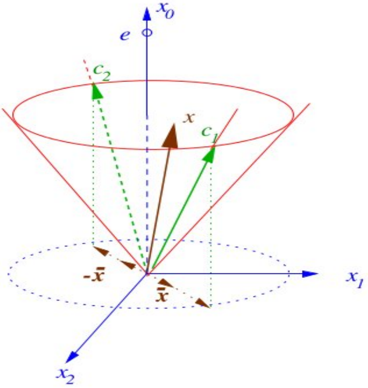
- For each x define

$$c_1 = \frac{1}{2}\begin{pmatrix} 1 \\ \frac{\bar{x}}{\|\bar{x}\|} \end{pmatrix} \quad c_2 = \frac{1}{2}\begin{pmatrix} 1 \\ -\frac{\bar{x}}{\|\bar{x}\|} \end{pmatrix}$$

- We can verify

$$x = \lambda_1 c_1 + \lambda_2 c_2$$

- This relation is the spectral decomposition of the vectors in SOCP Algebra



to get c_1 and c_2, (i) project x to $x_1, \ldots, x_n$ plane, (ii) normalize $\bar{x}$ and $-\bar{x}$ (iii) lift the normalized vectors up to touch the boundary of the cone

- Define for any real number t, $x^t = \lambda_1^t c_1 + \lambda_2^t c_2$ whenever λ_i^t is defined

$$x^{-1} = \frac{1}{\lambda_1} c_1 + \frac{1}{\lambda_2} c_2 = \frac{1}{\det(x)} \begin{pmatrix} x_0 \\ -\bar{x} \end{pmatrix}$$

- Now we can define an appropriate barrier for $\mathcal{Q}$

$$-\ln\det(x) = -\ln(x_0^2 - \|\bar{x}\|^2)$$

$$\nabla_x(-\ln\det x) = \frac{2}{\det(x)} \begin{pmatrix} x_0 \\ -\bar{x} \end{pmatrix} = 2x^{-1}$$

- we can replace SOCP problem with

$$\begin{array}{ll} \max & c^\top x - \mu \ln \det x \\ \text{s.t.} & Ax = b \\ & x \succ_{\mathcal{Q}} 0 \end{array}$$

- The Lagrangian

$$\mathcal{L}(x, y) = c^\top x - \mu \ln \det x - y^\top (b - Ax)$$

- Applying KKT

$$\begin{aligned} b - Ax &= 0 \\ c - \mu x^{-1} - A^\top y &= 0 \end{aligned}$$

- Setting $s = \mu x^{-1}$ we can see that s is dual feasible

Newton's method

Thus we have to solve the following system

$$\begin{aligned} Ax &= b \\ A^\top y + s &= c \\ x \circ s &= 2\mu e \end{aligned}$$

Using Newton's method, we get

$$\begin{aligned} A(x + \Delta x) &= b \\ A^\top (y + \Delta y) + s + \Delta s &= c \\ (x_i + \Delta x_i) \circ (s_i + \Delta s_i) &= 2\mu e \end{aligned}$$

Newton's method

Now expanding and dropping nonlinear terms

$$\begin{aligned} A\Delta x &= b - Ax \\ A^\top \Delta y + \Delta s &= c - A^\top y - s \\ x \circ \Delta s + \Delta x \circ s &= 2\mu e - x \circ s \quad \text{nonlinear term } \Delta x \circ \Delta s \text{ was dropped} \end{aligned}$$

- In matrix form

$$\begin{pmatrix} A & 0 & 0 \\ 0 & A^\top & I \\ \mathsf{L}_s & 0 & \mathsf{L}_x \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta s \end{pmatrix} = \begin{pmatrix} r_p \\ r_d \\ r_c \end{pmatrix} \quad \text{where} \quad \begin{aligned} r_p &= b - Ax \\ r_d &= c - A^\top y - s \\ r_c &= 2\mu e - x \circ s \end{aligned}$$