Numerical Optimization

Linear Programming - Interior Point Methods

Shirish Shevade

Computer Science and Automation
Indian Institute of Science
Bangalore 560 012, India.

NPTEL Course on Numerical Optimization

Computational Complexity of Simplex Algorithm

Example[1]

$$\begin{aligned}
\min \quad & -2^{n-1}x_1 - 2^{n-2}x_2 - \ldots - 2x_{n-1} - x_n \\
\text{s.t.} \quad & x_1 \leq 5 \\
& 4x_1 + x_2 \leq 25 \\
& 8x_1 + 4x_2 + x_3 \leq 125 \\
& \vdots \\
& 2^n x_1 + 2^{n-1}x_2 + \ldots + 4x_{n-1} + x_n \leq 5^n \\
& x_j \geq 0 \; \forall\, j = 1, \ldots, n
\end{aligned}$$

Simplex method, starting at $\boldsymbol{x} = \boldsymbol{0}$, would visit all 2^n extreme points before reaching the optimal solution.

[1] V. Klee and G.J. Minty, *How good is the simplex algorithm?*. In O. Shisha, editor, *Inequalities*, II, pp. 159-175, Academic Press, 1971

$$\begin{aligned} \min \quad & -4x_1 - 2x_2 - x_3 \\ \text{s.t.} \quad & x_1 \le 5 \\ & 4x_1 + x_2 \le 25 \\ & 8x_1 + 4x_2 + x_3 \le 125 \\ & x_1, x_2, x_3 \ge 0 \end{aligned}$$

Iteration	Basic Vectors	Objective function
1	x_4, x_5, x_6	0
2	x_1, x_5, x_6	-20
3	x_1, x_2, x_6	-30
4	x_4, x_2, x_6	-50
5	x_4, x_2, x_3	-75
6	x_1, x_2, x_3	-95
7	x_1, x_5, x_3	-105
8	x_4, x_5, x_3	-125

Simplex Algorithm is not a polynomial time algorithm (number of computational steps grows as an exponential function of the number of variables, rather than as a polynomial function)

Interior Point Methods for Linear Programming

- Points generated are in the “interior" of the feasible region
- Based on nonlinear programming techniques
- Some interior points methods:
  - Affine Scaling
  - Karmarkar’s Method

We consider the linear program,

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

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\text{rank}(\boldsymbol{A}) = m$.

Affine Scaling

Idea:

(1) Use *projected steepest descent direction* at every iteration

Given a feasible interior point $\boldsymbol{x}^k$ at current iteration k.

$$\boldsymbol{A}\boldsymbol{x}^k = \boldsymbol{b}, \boldsymbol{x}^k \geq \boldsymbol{0}$$

Let $\boldsymbol{d}$ denote a direction such that $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k \boldsymbol{d}$, $\alpha^k > 0$.
Therefore,

$$\boldsymbol{A}\boldsymbol{x}^{k+1} = \boldsymbol{b} \;\Rightarrow\; \boldsymbol{A}\boldsymbol{d} = \boldsymbol{0}$$

Projected Steepest Descent Direction

Project the steepest descent direction, $-\boldsymbol{c}$, on the *null space of* $\boldsymbol{A}$

Let $\hat{\boldsymbol{c}} = -\boldsymbol{c} = \boldsymbol{p_c} + \boldsymbol{q}$ where

- $\boldsymbol{p_c} \in$ Null Space$(\boldsymbol{A})$. $\therefore \boldsymbol{A}\boldsymbol{p_c} = \boldsymbol{0}$.
- $\boldsymbol{q} \in$ Row Space$(\boldsymbol{A})$. $\therefore \boldsymbol{q} = \boldsymbol{A}^T\boldsymbol{\lambda}$.

$$\boldsymbol{A}\hat{\boldsymbol{c}} = \boldsymbol{A}\boldsymbol{A}^T\boldsymbol{\lambda} \;\Rightarrow\; \boldsymbol{\lambda} = (\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A}\hat{\boldsymbol{c}}$$

$$\begin{aligned}
\boldsymbol{p_c} &= \hat{\boldsymbol{c}} - \boldsymbol{q} \\
&= \hat{\boldsymbol{c}} - \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A}\hat{\boldsymbol{c}} \\
&= (\boldsymbol{I} - \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A})\hat{\boldsymbol{c}} \\
&= -(\boldsymbol{I} - \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A})\boldsymbol{c} \\
&= -\boldsymbol{P}\boldsymbol{c}
\end{aligned}$$

where $\boldsymbol{P}$ denotes the *projection matrix.*

Idea:

(2) Position the current point close to the centre of the feasible region

For example, one possible choice is the point:
$\mathbf{1} = (1, 1, \ldots, 1)^T$
Given a point $\boldsymbol{x}^k$ in the interior of the feasible region, define $X^k = diag(x^k)$.
Define the transformation, $\boldsymbol{y} = T(\boldsymbol{x}) = X^{k^{-1}}\boldsymbol{x}$.

$$\therefore \boldsymbol{y}^k = X^{k^{-1}}\boldsymbol{x}^k = \mathbf{1} \text{ or } X^k\boldsymbol{y}^k = \boldsymbol{x}^k$$

Affine Scaling Algorithm:

- Start with any interior point $\boldsymbol{x}^0$
- while (stopping condition is not satisfied at the current point)
  - Transform the current problem into an equivalent problem in $\boldsymbol{y}-$space so that the current point is close to the centre of the feasible region
  - Use projected steepest descent direction to take a step in the $\boldsymbol{y}-$space without crossing the feasible set boundary
  - Map the new point back to the corresponding point in the $\boldsymbol{x}-$space

  endwhile

Stopping Condition for an Affine Scaling Algorithm

Consider the following primal and dual problems:

Primal Problem (P)

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

Dual Problem (D)

$$\begin{aligned} \max \quad & \boldsymbol{b}^T\boldsymbol{\mu} \\ \text{s.t.} \quad & \boldsymbol{A}^T\boldsymbol{\mu} \leq \boldsymbol{c} \end{aligned}$$

For any primal and dual feasible $\boldsymbol{x}$ and $\boldsymbol{\mu}$,

$$\boldsymbol{c}^T\boldsymbol{x} \geq \boldsymbol{b}^T\boldsymbol{\mu} \qquad \text{(Weak Duality)}$$

At optimality,

$$\boldsymbol{c}^T\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{\mu} = 0 \qquad \text{(Strong Duality)}$$

Idea: Use the duality gap, $\boldsymbol{c}^T\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{\mu}$, to check optimality

Q. How to get $\boldsymbol{\mu}$?

Consider the primal problem:

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

Define the Lagrangian function,

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{\mu}^T(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) - \boldsymbol{\lambda}^T\boldsymbol{x}$$

Assumption: $\boldsymbol{x}$ is primal feasible and $\boldsymbol{\lambda} \geq \boldsymbol{0}$
KKT conditions at optimality:

$$\begin{aligned} \nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{0} \;\Rightarrow\; & \boldsymbol{A}^T\boldsymbol{\mu} + \boldsymbol{\lambda} = \boldsymbol{c} \\ & \lambda_i x_i = 0 \;\forall\; i = 1, \ldots, n \end{aligned}$$

Defining $X = diag(x)$, the KKT conditions are

$$X(\boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{\mu}) = \boldsymbol{0}.$$

Solve the following problem to get $\boldsymbol{\mu}$:

$$\min_{\boldsymbol{\mu}} \quad \|X\boldsymbol{c} - X\boldsymbol{A}^T\boldsymbol{\mu}\|^2$$

$$\therefore \boldsymbol{\mu} = (\boldsymbol{A}X^2\boldsymbol{A}^T)^{-1}\boldsymbol{A}X^2\boldsymbol{c}$$

Thus, at a given point $\boldsymbol{x}^k$,

$$\text{Duality Gap} = \boldsymbol{c}^T\boldsymbol{x}^k - \boldsymbol{b}^T\boldsymbol{\mu}^k$$

where $\boldsymbol{\mu}^k = (\boldsymbol{A}X^{k^2}\boldsymbol{A}^T)^{-1}\boldsymbol{A}X^{k^2}\boldsymbol{c}$ and $X^k = diag(\boldsymbol{x}^k)$.

Step 1: Equivalent problem formulation to get considerable improvement in the objective function
Given $\boldsymbol{x}^k$, define $X^k = diag(\boldsymbol{x}^k)$.
Define a transformation T as,

$$\boldsymbol{y} = T(\boldsymbol{x}) = X^{k^{-1}}\boldsymbol{x}$$

Therefore,

$$X^k\boldsymbol{y}^k = \boldsymbol{x}^k \text{ and } \boldsymbol{y}^k = \mathbf{1}.$$

Using this transformation,

$$\left.\begin{array}{ll} \min & \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} & A\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geq \mathbf{0} \end{array}\right\} \equiv \begin{array}{ll} \min & \boldsymbol{c}^T X^k \boldsymbol{y} \\ \text{s.t.} & AX^k\boldsymbol{y} = \boldsymbol{b},\ \boldsymbol{y} \geq \mathbf{0} \end{array}$$

which can written in standard form as

$$\begin{array}{ll} \min & \bar{\boldsymbol{c}}^T\boldsymbol{y} \\ \text{s.t.} & \bar{A}\boldsymbol{y} = \boldsymbol{b},\ \boldsymbol{y} \geq \mathbf{0} \end{array}$$

where $\bar{\boldsymbol{c}} = X^k\boldsymbol{c}$ and $\bar{A} = AX^k$.

Step 2: Find the projected steepest direction and step length at $\boldsymbol{y}^k$ for the problem,

$$\begin{array}{ll} \min & \bar{\boldsymbol{c}}^T \boldsymbol{y} \\ \text{s.t.} & \bar{\boldsymbol{A}}\boldsymbol{y} = \boldsymbol{b},\ \boldsymbol{y} \geq \boldsymbol{0} \end{array}$$

Given $\boldsymbol{x}^k$, $\boldsymbol{y}^k = X^{k^{-1}} \boldsymbol{x}^k = \boldsymbol{1}$.
The projected direction of $-\bar{\boldsymbol{c}}$ on the null space of $\bar{\boldsymbol{A}}$ is,

$$\therefore \boldsymbol{d}^k = -(\boldsymbol{I} - X^k \boldsymbol{A}^T (\boldsymbol{A} X^{k^2} \boldsymbol{A}^T)^{-1} \boldsymbol{A} X^k) X^k \boldsymbol{c}.$$

Let $\alpha^k (> 0)$ denotes the step length.

$$\begin{aligned} \boldsymbol{y}^{k+1} &= \boldsymbol{y}^k + \alpha^k \boldsymbol{d}^k \\ &= \boldsymbol{1} + \alpha^k \boldsymbol{d}^k \geq \boldsymbol{0} \end{aligned}$$

$$\text{Let } \alpha_{max} = \min_{j: \boldsymbol{d}_j^k < 0} -\frac{1}{\boldsymbol{d}_j^k} \text{ and set } \alpha^k = .9 * \alpha_{max}.$$

Step 3: $\boldsymbol{x}^{k+1} = X^k \boldsymbol{y}^{k+1}$

Affine Scaling Algorithm (to solve an LP in Standard Form)

(1) Input: $\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{x}^0, \epsilon$

(2) Set $k := 0$.

(3) $X^k = diag(\boldsymbol{x}^k)$

(4) $\boldsymbol{\mu}^k = (\boldsymbol{A}{X^k}^2\boldsymbol{A}^T)^{-1}\boldsymbol{A}{X^k}^2\boldsymbol{c}$

(5) **while** $(\boldsymbol{c}^T\boldsymbol{x}^k - \boldsymbol{b}^T\boldsymbol{\mu}^k) > \epsilon$

  (a) $\boldsymbol{d}^k = -(\boldsymbol{I} - X^k\boldsymbol{A}^T(\boldsymbol{A}{X^k}^2\boldsymbol{A}^T)^{-1}\boldsymbol{A}X^k)X^k\boldsymbol{c}$

  (b) $\alpha^k = .9 * \min_{j:\boldsymbol{d}_j^k<0} -\frac{1}{\boldsymbol{d}_j^k}$

  (c) $\boldsymbol{x}^{k+1} = X^k(1 + \alpha^k\boldsymbol{d}^k)$

  (d) $X^{k+1} = diag(\boldsymbol{x}^{k+1})$

  (e) $\boldsymbol{\mu}^{k+1} = (\boldsymbol{A}{X^{k+1}}^2\boldsymbol{A}^T)^{-1}\boldsymbol{A}{X^{k+1}}^2\boldsymbol{c}$

  (f) $k := k + 1$

endwhile

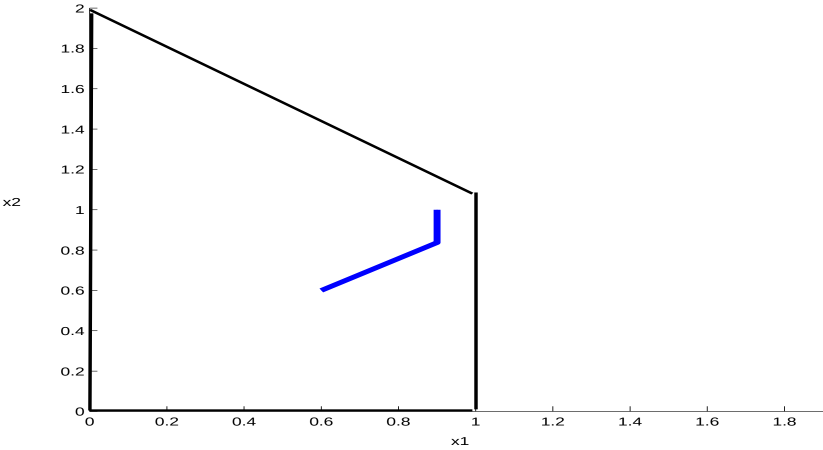
Output : $\boldsymbol{x}^* = \boldsymbol{x}^k$

Application of Affine Scaling Algorithm to the problem,

$$\begin{aligned} \min \quad & -3x_1 - x_2 \\ \text{s.t.} \quad & x_1 + x_2 \leq 2 \\ & x_1 \leq 1 \\ & x_1, x_2 \geq 0 \end{aligned}$$

- $\boldsymbol{x}^* = (1,1)^T$

Iteration	${\boldsymbol{x}^k}^T$	$\|\boldsymbol{x}^k - \boldsymbol{x}^*\|$
0	(.6, .6)	.57
1	(.87, .81)	.23
2	(.88, 1.00)	.12
3	(.90, .99)	.1026
4	(.90, 1.00)	.1001


x2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
x1

Karmarkar's Method

Assumptions:

- The problem is in *homogeneous* form:

$$\begin{aligned} \min \quad & \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \\ & \boldsymbol{1}^T\boldsymbol{x} = 1 \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

- Optimum objective function value is 0

Idea:

(1) Use projective transformation to move an interior point to the centre of the feasible region

(2) Move along projected steepest descent direction

Projective Transformation:

Consider the transformation, T, defined by,

$$\boldsymbol{y} = T(\boldsymbol{x}) = \frac{X^{k-1}\boldsymbol{x}}{\mathbf{1}^T X^{k-1}\boldsymbol{x}}, \quad \boldsymbol{x} \neq \mathbf{0}$$

Remarks:

- $\boldsymbol{x} \mapsto \boldsymbol{y} \qquad (\boldsymbol{y} = (\frac{1}{n}, \ldots, \frac{1}{n})^T, \mathbf{1}^T\boldsymbol{y} = 1)$
- Inverse transformation:
  $\boldsymbol{x} = (\mathbf{1}^T X^{k-1}\boldsymbol{x})(X^k\boldsymbol{y}) \;\Rightarrow\; \mathbf{1}^T\boldsymbol{x} = (\mathbf{1}^T X^{k-1}\boldsymbol{x})\mathbf{1}^T(X^k\boldsymbol{y})$.
  For every feasible $\boldsymbol{x}$, $\mathbf{1}^T\boldsymbol{x} = 1 \;\Rightarrow\; \mathbf{1}^T X^{k-1}\boldsymbol{x} = \dfrac{1}{\mathbf{1}^T X^k\boldsymbol{y}}$

$$\therefore T^{-1}(\boldsymbol{y}) = \boldsymbol{x} = \frac{X^k\boldsymbol{y}}{\mathbf{1}^T X^k\boldsymbol{y}}$$

Using the transformation,

$$\boldsymbol{x} = \frac{X^k \boldsymbol{y}}{\mathbf{1}^T X^k \boldsymbol{y}}$$

$$\begin{array}{ll} \min & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t.} & A\boldsymbol{x} = \mathbf{0} \\ & \mathbf{1}^T \boldsymbol{x} = 1 \\ & \boldsymbol{x} \geq \mathbf{0} \end{array} \equiv \begin{array}{ll} \min & \dfrac{\boldsymbol{c}^T X^k \boldsymbol{y}}{\mathbf{1}^T X^k \boldsymbol{y}} \\ \text{s.t.} & A X^k \boldsymbol{y} = \mathbf{0} \\ & \mathbf{1}^T \boldsymbol{y} = 1 \\ & \boldsymbol{y} \geq \mathbf{0} \end{array} \equiv \begin{array}{ll} \min & \boldsymbol{c}^T X^k \boldsymbol{y} \\ \text{s.t.} & A X^k \boldsymbol{y} = \mathbf{0} \\ & \mathbf{1}^T \boldsymbol{y} = 1 \\ & \boldsymbol{y} \geq \mathbf{0} \end{array}$$

- Equivalence based on the assumption: Optimal objective function value is 0 and $\mathbf{1}^T X^k \boldsymbol{y} > 0$

Consider the problem,

$$\begin{aligned} \min \quad & \boldsymbol{c}^T X^k \boldsymbol{y} \\ \text{s.t.} \quad & A X^k \boldsymbol{y} = \boldsymbol{0} \\ & \boldsymbol{1}^T \boldsymbol{y} = 1 \\ & \boldsymbol{y} \geq \boldsymbol{0} \end{aligned}$$

- Step 1: Find a projected steepest descent direction in the $\boldsymbol{y}-$ space (Projection of $-X^k\boldsymbol{c}$ onto the subspace of $\{\boldsymbol{d} : AX^k\boldsymbol{d} = \boldsymbol{0}, \boldsymbol{1}^T\boldsymbol{d} = 0, \boldsymbol{d} \geq \boldsymbol{0}\}$)
  Find $\boldsymbol{d}$ by solving

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|X^k\boldsymbol{c} - \boldsymbol{d}\|^2 \\ \text{s.t.} \quad & AX^k\boldsymbol{d} = \boldsymbol{0}, \end{aligned}$$

  projecting it onto the null space of $\boldsymbol{1}^T$ and ensure $\boldsymbol{d} \geq \boldsymbol{0}$.

Consider the problem,

$$\begin{array}{ll} \min & \frac{1}{2}\|X^k \boldsymbol{c} - \boldsymbol{d}\|^2 \\ \text{s.t.} & AX^k \boldsymbol{d} = \boldsymbol{0} \end{array}$$

$$\mathcal{L}(\boldsymbol{d}, \boldsymbol{\mu}) = \frac{1}{2}\|X^k \boldsymbol{c} - \boldsymbol{d}\|^2 + \boldsymbol{\mu}^T \boldsymbol{A}\boldsymbol{x}^k \boldsymbol{d}$$

$$\begin{aligned} \nabla_d \, \mathcal{L}(\boldsymbol{d}, \boldsymbol{\mu}) = \boldsymbol{0} \; \Rightarrow \; & -(X^k \boldsymbol{c} - \boldsymbol{d}) + X^k A^T \boldsymbol{\mu} = 0 \\ \Rightarrow \; & \boldsymbol{d} = X^k \boldsymbol{c} - X^k A^T \boldsymbol{\mu} \end{aligned}$$

$$\boldsymbol{0} = AX^k \boldsymbol{d} = AX^{k^2} \boldsymbol{c} - AX^{k^2} A^T \boldsymbol{\mu} \; \Rightarrow \; AX^{k^2} \boldsymbol{c} = AX^{k^2} A^T \boldsymbol{\mu}$$

Projection of $-\boldsymbol{d}$ on the null space of $\boldsymbol{1}^T$:

$$\boldsymbol{d}^k = -(\boldsymbol{I} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^T)(X^k \boldsymbol{c} - X^k A^T \boldsymbol{\mu})$$

Karmarkar's Projective Scaling Algorithm (for homogeneous Linear Program)

(1) Input: Homogeneous LP, $\boldsymbol{A}, \boldsymbol{c}, \epsilon$

(2) Set $k := 0$, $\boldsymbol{x}^k = \frac{1}{n}\mathbf{1}$

(3) $X^k = diag(\boldsymbol{x}^k)$

(4) **while** $\boldsymbol{c}^T\boldsymbol{x}^k > \epsilon$

  (a) Find the projected steepest descent direction $\boldsymbol{d}^k$

  (b) $\boldsymbol{y}^{k+1} = \frac{1}{n}\mathbf{1} + \frac{\delta}{\sqrt{n(n-1)}} \frac{\boldsymbol{d}^k}{\|\boldsymbol{d}^k\|}$ $(\delta = \frac{1}{3})$

  (c) $\boldsymbol{x}^{k+1} = T^{-1}(\boldsymbol{y}^{k+1})$

  (d) $X^{k+1} = diag(\boldsymbol{x}^{k+1})$

  (e) $k := k + 1$

endwhile

Output : $\boldsymbol{x}^* = \boldsymbol{x}^k$