Large-Scale Optimization

L22. LINEAR PROGRAM 2
Today

Linear Program

• Interior Point Method
Recall: Standard Form

\[ \begin{align*}
\min_{x \in \mathbb{R}^n} & \quad z = c^T x \\
A x & \geq b, \\
x & \geq 0.
\end{align*} \]

Data:
\[ c \in \mathbb{R}^n, \ A \in \mathbb{R}^{m \times n}, \ b \in \mathbb{R}^m \]

Inequalities can be converted to equalities by introducing slack variables:

\[ \begin{align*}
\min_{x_N \in \mathbb{R}^n, x_B \in \mathbb{R}^m} & \quad z = c^T x_N + 0^T x_B \\
x_B & = A x - b, \\
x_B, x_N & \geq 0.
\end{align*} \]

This is called the standard form (or the canonical form)
**Tableau Form**

\[
\begin{align*}
\min_{x_N \in \mathbb{R}^n, x_B \in \mathbb{R}^m} \quad & z = c^T x_N + 0^T x_B \\
\text{subject to} \quad & x_B = Ax - b, \\
\quad & x_B, x_N \geq 0.
\end{align*}
\]

This can be represented as a tableau:

| \( x_N \) | 1 \\
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_B ) =</td>
<td></td>
</tr>
<tr>
<td>( A )</td>
<td>(-b)</td>
</tr>
<tr>
<td>( c^T )</td>
<td>0</td>
</tr>
</tbody>
</table>
max \ z = 5x + 4y \\
\text{s.t.} \ x \leq 6, \\
\quad .25x + y \leq 6, \\
\quad 3x + 2y \leq 22, \\
\quad x, y \geq 0.
Ex.

\[
\begin{align*}
\min_{x,y} & \quad z = -5x - 4y \\
\text{s.t.} & \quad x \leq 6, \\
& \quad .25x + y \leq 6, \\
& \quad 3x + 2y \leq 22, \\
& \quad x, y \geq 0.
\end{align*}
\]

\[
\begin{array}{|ccc|}
\hline
x & y & 1 \\
\hline
u = & -1 & 0 & 6 \\
v = & -0.25 & -1 & 6 \\
w = & -3 & -2 & 22 \\
z = & -5 & -4 & 0 \\
\hline
\end{array}
\]
Simplex Algorithm

Repeat until optimality:

• Pricing
  – Choose a column (nonbasic var) to exchange

• Ratio test
  – Choose a row (basic var) to exchange

• Pivoting
  – Perform Jordan exchange of the chosen row/column
### Jordan Exchange

**u =**

\[
\begin{bmatrix}
  x & y & 1 \\
  -1 & 0 & 6 \\
  -0.25 & -1 & 6 \\
  -3 & -2 & 22 \\
  -5 & -4 & 0 \\
\end{bmatrix}
\]

**v =**

\[
\begin{bmatrix}
  x & y & u & v & w & 1 \\
  -1 & 0 & -1 & 0 & 0 & 6 \\
  -0.25 & -1 & 0 & -1 & 0 & 6 \\
  -3 & -2 & 0 & 0 & -1 & 22 \\
  -5 & -4 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

Eliminate x from rows except for the 1st

**x =**

\[
\begin{bmatrix}
  u & y & 1 \\
  -1 & 0 & 6 \\
  -0.25 & 1 & -18/4 \\
  3 & -2 & 4 \\
  5 & -4 & -30 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
  x & y & u & v & w & 1 \\
  -1 & 0 & -1 & 0 & 0 & 6 \\
  0 & 4 & -1 & 4 & 0 & -18 \\
  0 & -2 & 3 & 0 & -1 & 4 \\
  0 & -4 & 5 & 0 & 0 & -30 \\
\end{bmatrix}
\]

TU Dortmund, Dr. Sangkyun Lee
\[
\begin{align*}
\begin{array}{ccc}
x & y & 1 \\
u = & -1 & 0 & 6 \\
v = & -0.25 & -1 & 6 \\
w = & -3 & -2 & 22 \\
z = & -5 & -4 & 0 \\
x = & -1 & 0 & 6 \\
v = & -0.25 & 1 & -18/4 \\
w = & 3 & -2 & 4 \\
z = & 5 & -4 & -30 \\
x = & -1 & 0 & 6 \\
v = & -1.25 & 0.5 & 2.5 \\
y = & 1.5 & -0.5 & 2 \\
z = & -1 & 2 & -38 \\
x = & 0.8 & -0.4 & 4 \\
u = & -0.8 & 0.4 & 2 \\
y = & -1.2 & 0.1 & 5 \\
z = & 0.8 & 1.6 & -40
\end{array}
\end{align*}
\]

This is optimal: no further improvement possible!
Complexity of LP

When a solution exists, it must be a vertex of the convex polyhedron defined by constraints

However, in the worst case the number of vertices in $\mathbb{R}^n$ defined with $\ell$ inequality constraints and bounds can be

$$\binom{\ell}{n} \quad n = 10, \ell = 20, \binom{\ell}{n} \approx 185,000$$

Therefore a brute force approach to find a solution can very slow
Interior Point Method: Motivation

\[ \min_{x \in \mathbb{R}^n} \quad c^T x \quad \quad A \in \mathbb{R}^{m \times n} \]

\[ A x = b, \quad \lambda \in \mathbb{R}^m \]

\[ x \geq 0. \quad \xi \in \mathbb{R}_+^n \]

KKT conditions:

\[ \mathcal{L}(x; \lambda, \xi) = c^T x - \lambda^T (A x - b) - \xi^T x \]

\[ \frac{\partial \mathcal{L}}{\partial x} = c - A^T \lambda - \xi = 0 \]

\[ x \geq 0 \quad \xi \geq 0 \quad \lambda_i (A_i x - b_i) = 0, \quad i = 1, \ldots, m, \]

\[ A x = b \]

\[ \xi_i x_i = 0, \quad i = 1, \ldots, n. \]
KKT Conditions

Solving LP = finding $x, \lambda, \xi$ satisfying the KKT conditions:

$$Ax = b$$

$$A^T \lambda + \xi = c$$

$$x \geq 0, \quad \xi \geq 0$$

$$x_i^T \xi_i = 0, \quad i = 1, \ldots, n$$

$$XSe = 0$$
Dual Problem

$$\max_{\lambda, \xi} \inf_x L(x; \lambda, \xi)$$

$$L(x; \lambda, \xi) = c^T x - \lambda^T (A x - b) - \xi^T x$$

$$= (c - \xi)^T x = (A^T \lambda)^T x = \lambda^T A x = b^T \lambda$$

This is convex in $x$, therefore KKT conditions is necessary and sufficient for optimality in $x$:

$$A x = b \quad A^T \lambda + \xi = c \quad x^T \xi = 0 \quad x \geq 0, \ \xi \geq 0$$

$$\max_{\lambda, \xi} \ b^T \lambda$$

s.t. $A^T \lambda + \xi = c$, $\xi \geq 0$
**Duality**

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad c^T x \\
\text{s.t.} & \quad Ax = b, \\
& \quad x \geq 0.
\end{align*}
\]

\[
\begin{align*}
\max_{\lambda, \xi} & \quad b^T \lambda \\
\text{s.t.} & \quad A^T \lambda + \xi = c, \\
& \quad \xi \geq 0
\end{align*}
\]

Weak duality: \( b^T \lambda^* \leq c^T x^* \)

Strong duality: \( b^T \lambda^* = c^T x^* \)  

(Check lectures 7 & 8)

Duality gap: \( c^T x - b^T \lambda = \xi^T x = n \left( \frac{1}{n} \sum_{i=1}^{n} \xi_i x_i \right) \)

Duality measure, \( \mu \)
Primal-Dual System of Equations

\[ Ax = b \]
\[ A^T \lambda + \xi = c \]
\[ XSe = 0 \]
\[ x \geq 0, \ \xi \geq 0 \]

\[ F(x, \lambda, \xi) := \begin{bmatrix} Ax - b \\ A^T \lambda + \xi - c \\ XSe \end{bmatrix} \]

2n+m vars

LP solution: \((x^*, \lambda^*, \xi^*)\) satisfying

\[ F(x^*, \lambda^*, \xi^*) = 0, \ x^* \geq 0, \ \xi^* \geq 0. \]
Newton-Rapson Method

\[ F : \mathbb{R}^N \rightarrow \mathbb{R}^M \]

Jacobian matrix: \[ J \in \mathbb{R}^{M \times N} \]

\[ [J(z)]_{ij} = \frac{\partial F_i(z)}{\partial z_j}, \quad i = 1, \ldots, M, \quad j = 1, \ldots, N. \]

When \( F \) is smooth, from Taylor approximation we have:

\[ F(z + \Delta z) \approx F(z) + J(z)\Delta z \quad \text{for small } \Delta z \]
Newton-Rapson Method

Find $\Delta z$ such that the linear approximation becomes the zero:

$$F(z^k) + J(z^k)\Delta z^k = 0$$

If $J(z^k)$ is square and nonsingular,

$$\Delta z^k = -J(z^k)^{-1}F(z^k)$$

Then update the approx. root of $F(z) = 0$:

$$z^{k+1} = z^k + \Delta z^k$$

The sequence converges quadratically when

- When $F$ is Lipschitz continuously differentiable, and
- $z^0$ is not far from the solution
  - Otherwise, a stepsize $\alpha^k$ need to be chosen, and

$$z^{k+1} = z^k + \alpha^k \Delta z^k$$
Primal-Dual Interior Point Method

Find a root of

$$F(x, \lambda, \xi) = \begin{bmatrix} Ax - b \\ A^T \lambda + \xi - c \\ XSe \end{bmatrix} = 0$$

Jacobian matrix:

$$J(x, \lambda, \xi) = \begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{bmatrix} \quad N \times N, \quad N = 2n+m$$

- Start from an interior point: \((x^0, \lambda^0, \xi^0)\) s.t. \(x^0 > 0\) and \(\xi^0 > 0\).
- Apply a modified Newton-Rapson to ensure \(x^k > 0\) and \(\xi^k > 0\).
  - When \(A\) has full row rank, this ensures \(J\) to be nonsingular!
Exercise: J nonsingular

\[ J(x, \lambda, \xi) = \begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{bmatrix} \]

Hint: use block inversion lemma that \( \begin{bmatrix} A & B \\ C & D \end{bmatrix} \) is invertible if \( A \) and \( D - CA^{-1}B \) are invertible.
Path-Following Approach

Motivation: too many Newton iteration might be necessary to satisfy the complementarity conditions

Let \((x(\tau), \lambda(\tau), \xi(\tau))\) be a point satisfying

\[
Ax = b \\
A^T \lambda + \xi = c \\
XSe = \tau e \\
x > 0, \; \xi > 0
\]

This defines a perturbed KKT system. It can be shown that this system has a unique solution for any \(\tau > 0\).
Central Path

\[ C := \{ z(\tau) = (x(\tau), \lambda(\tau), \xi(\tau) : \tau > 0 \} \]

An illustration of a central path, with \( z(\tau) \rightarrow z^* \) as \( \tau \downarrow 0 \)
Central Path

Note that

\[ XSe = \tau e \quad \longrightarrow \quad \tau = \frac{1}{n} x^T \xi = \mu \text{ (duality measure)} \]

In practice, # of required Newton steps could be large, and therefore we stay in a neighborhood of \( C \):

\[ N(\gamma) := \{ (x, \lambda, \xi) : Ax = b, A^T \lambda + \xi = c, (x, \xi) \geq 0, x_i \xi_i \geq \gamma \mu, \ i = 1, \ldots, n \} \]

\[ \gamma \in (0, 1] \]

That is, each \( x_i \xi_i \) is at least the \( \gamma \)-fraction of their average
Path-Following Algorithm

Choose $\sigma_{\text{min}}$ and $\sigma_{\text{max}}$ s.t. $0 < \sigma_{\text{min}} < \sigma_{\text{max}} < 1$;

Choose an initial point $(x^0, \lambda^0, \xi^0)$ with $x^0 > 0$ and $\xi^0 > 0$.

\begin{align*}
\text{for } & k = 0, 1, 2, \ldots \\
& \text{Choose } \sigma_k \in [\sigma_{\text{min}}, \sigma_{\text{max}}]; \\
& \text{Solve for } (\Delta x^k, \Delta \lambda^k, \Delta \xi^k):
\end{align*}

\[\begin{bmatrix}
A & 0 & 0 \\
0 & A^T & I \\
S^k & 0 & X^k
\end{bmatrix}
\begin{bmatrix}
\Delta x^k \\
\Delta \lambda^k \\
\Delta \xi^k
\end{bmatrix}
= -\begin{bmatrix}
Ax^k - b \\
A^T \lambda^k + \xi^k - c \\
X^k S^k e - \sigma_k \mu_k e
\end{bmatrix},
\]

where $\mu_k = (x^k)^T \xi^k / n$.

Choose $\alpha_{\text{max}}$ to be the largest positive value such that

$$(x^k, \xi^k) + \alpha(\Delta x^k, \Delta \xi^k) \geq 0;$$

Set $\alpha_k = \min(1, (1 - \gamma_k) \cdot \alpha_{\text{max}})$ for some $\gamma_k \in (0, 1)$.

Set $(x^{k+1}, \lambda^{k+1}, \xi^{k+1}) = (x^k, \lambda^k, \xi^k) + \alpha_k(\Delta x^k, \Delta \lambda^k, \Delta \xi^k)$;

end (for).
Predictor-Corrector IP Method

Mehrotra (1992): a foundation of modern IP methods

- (Predictor) Compute a Newton step with $\sigma_k = 0$.
  - If $\alpha_k \approx 1$ can be taken, set $\sigma_k \approx 0$
  - If $\alpha_k$ short, then choose larger $\sigma_k$

- (Corrector) Choose the step by solving

$$
\begin{bmatrix}
A & 0 & 0 \\
0 & A^T & I \\
S^k & 0 & X^k
\end{bmatrix}
\begin{bmatrix}
\Delta x^k \\
\Delta y^k \\
\Delta s^k
\end{bmatrix} = -
\begin{bmatrix}
A x^k - b \\
A^T \lambda^k + \xi^k - c \\
X^k S^k e - \sigma_k \mu_k e + \Delta X^k_{pre} \Delta S^k_{pre} e
\end{bmatrix},
$$

- This formulation is from a second-order Newton’s method

- + Heuristics to choose the initial point, gamma$_k$, and the largest stepsize per iteration (which have good impact on practical performance)
Solving the Newton System

\[
\begin{bmatrix}
A & 0 & 0 \\
0 & A^T & I \\
S & 0 & X
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta \lambda \\
\Delta \xi
\end{bmatrix}
=
\begin{bmatrix}
r_b \\
r_c \\
r_{xs}
\end{bmatrix}
\]

\[S\Delta x + X\Delta \xi = r_{xs} \quad \Rightarrow \quad \Delta \xi = -X^{-1}S\Delta x + X^{-1}r_{xs}\]

\[
\begin{bmatrix}
-X^{-1}S & A^T \\
A & 0
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta \lambda
\end{bmatrix}
=
\begin{bmatrix}
r_c - X^{-1}r_{xs} \\
r_b
\end{bmatrix}
\]

\[\Delta x = -S^{-1}X \left[ r_p - X^{-1}r_{xs} - A^T \Delta \lambda \right]\]

\[A(S^{-1}X)A^T \Delta \lambda = r_b + AS^{-1}X \left[ r_c - X^{-1}r_{xs} \right]\]

Symmetric positive definite when A has full row rank: use Cholesky factorization
Interior-Point vs. Simplex

Interior-Point

• Much less # of iterations than simplex
• Each iteration cost more (solving 1 or 2 Newton systems)
  • When A is sparse, then can be done faster
• Issues in obtaining a vertex solution:

$$\min_{x \in \mathbb{R}^n} c^T x$$

\[ A x \geq b, \quad x \geq 0. \]

$$\min_{x, s} c^T x$$

\[ A x - s = b, \quad x \geq 0, \quad s \geq 0. \]

IP solutions:
\[ x > 0, \quad s > 0 \]

\[ A x > b, \quad x > 0 \]
Reference

Ferris, Mangasarian, & Wright, Linear Programming with Matlab, MPS-SIAM Series on Optimization, 2007