Interior-point methods
Path-following method

Slides by Nicolai Truong, SDU

Table of Contents

Table of Contents

Linear Programming

Consider the primal form in linear programming:

$$\begin{aligned} \text{maximize} \quad & c^T x \\ \text{subject to} \quad & Ax \le b \\ & x \ge 0 \end{aligned}$$

And the corresponding dual problem:

$$\begin{aligned} \text{minimize} \quad & b^T y \\ \text{subject to} \quad & A^T y \le c \\ & y \ge 0 \end{aligned}$$

Linear Programming

Both problems can be converted into equality form:

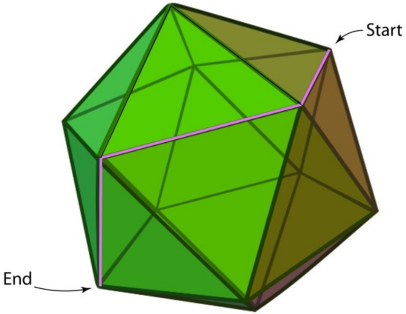
$$\begin{aligned} \text{maximize} \quad & c^T x \\ \text{subject to} \quad & Ax + w = b \\ & x, w \geq 0 \end{aligned}$$

And:

$$\begin{aligned} \text{minimize} \quad & b^T y \\ \text{subject to} \quad & A^T y + z = c \\ & y, z \geq 0 \end{aligned}$$

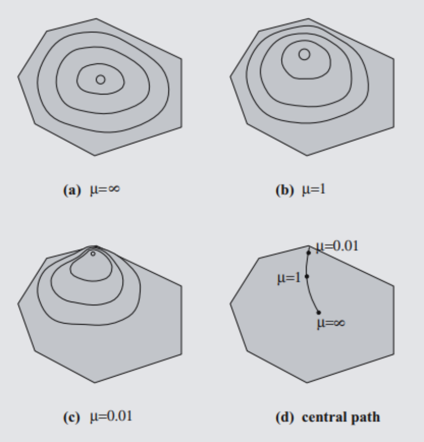
Simplex algorithm

Simplex Method finds the optimal solution by traversing along the edges from one vertex to another.



Interior Point method

The Interior Point method starts inside the polytope and iteratively converges to the optimal solution



Non-standard notation

Non-standard notation ahead! Given a lower-case letter denoting a vector quantity, we also have an upper-case letter denoting a diagonal matrix whose entries corresponds to elements the vector quantity.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \implies X = \begin{bmatrix} x_1 & & & \\ & x_2 & & \\ & & \ddots & \\ & & & x_n \end{bmatrix}$$

Barrier function

- How do we avoid converging outside the feasibility region?
- Barrier problem:

$$BP(\mu): \begin{array}{rl} \text{maximize} & c^T x + \mu \sum_j \log x_j + \mu \sum_i \log w_i \\ \text{subject to} & Ax + w = b \end{array}$$

- Nonlinear objective function: logarithmic barrier function
- Family of problems indexed by parameter $\mu > 0$

Lagrange Multipliers and Barrier Problem

We Lagrange relax to the barrier function, we get the following problem:

$$L(x, w, y) = c^T x + \mu \sum_j \log x_j + \mu \sum_i \log w_i + y^T (b - Ax - w)$$

We get the first-order optimality conditions when we take the derivatives and set them to zero.

$$\begin{aligned}
\frac{\partial L}{\partial x_j} &= c_j + \mu \frac{1}{x_j} - \sum_i y_i a_{ij} && = 0, \quad j = 1, 2, ..., n. \\
\frac{\partial L}{\partial w_i} &= \mu \frac{1}{w_j} - y_1 && = 0, \quad i = 1, 2, ..., m. \\
\frac{\partial L}{\partial y_i} &= b_i - \sum_j a_{ij} x_j - w_i && = 0, \quad i = 1, 2, ..., m.
\end{aligned}$$

Lagrange Multipliers and Barrier Problem

This can be written in matrix form:

$$
\begin{aligned}
A^T y - \mu X^{-1} e &= c \\
y &= \mu W^{-1} e \\
Ax + w &= b
\end{aligned}
$$

Note: X and W are the diagonal matrices containing diagonal entries.
Vector e is the vector of all ones.

Lagrange Multipliers and Barrier Problem

When we introduce an extra vector $z = \mu X^{-1} e$, we can rewrite our first-order optimality conditions like this:

$$\begin{aligned} Ax + w &= b \\ A^T y - z &= c \\ z &= \mu X^{-1} e \\ y &= \mu W^{-1} e \end{aligned}$$

Lagrange Multipliers and Barrier Problem

When we multiply X and W on respectively the third and fourth equation, we get the following equations:

$$\begin{aligned} Ax + w &= b \\ A^T y - z &= c \\ XZe &= \mu e \\ YWe &= \mu e \end{aligned}$$

Lagrange Multipliers and Barrier Problem

When we multiply X and W on respectively the third and fourth equation, we get the following equations:

$$\begin{aligned} Ax + w &= b \\ A^T y - z &= c \\ XZe &= \mu e \\ YWe &= \mu e \end{aligned}$$

Componentwise, the third and fourth equation can be written like this:

$$\begin{aligned} x_j z_j &= \mu & \qquad j &= 1, 2, ..., n \\ y_i w_i &= \mu & \qquad i &= 1, 2, ..., m \end{aligned}$$

Lagrange Multipliers and Barrier Problem

When we multiply X and W on respectively the third and fourth equation, we get the following equations:

$$\begin{aligned} Ax + w &= b \\ A^T y - z &= c \\ XZe &= \mu e \\ YWe &= \mu e \end{aligned}$$

Componentwise, the third and fourth equation can be written like this:

$$\begin{aligned} x_j z_j &= \mu & \qquad j &= 1, 2, ..., n \\ y_i w_i &= \mu & \qquad i &= 1, 2, ..., m \end{aligned}$$

μ-complementarity conditions: $2n + 2m$ equations in $2n + 2m$ unknowns.

Does a solution exist and if so, is it unique?

Lagrange Multipliers and Barrier Problem

Theorem 1

There exists a solution to the barrier problem if and only if both the primal and the dual feasible regions have nonempty interior.

Corollary 2

If a primal feasible set (or, for that matter, its dual) has a nonempty interior and is bounded, then for each $\mu > 0$ there exists a unique solution

$$(x_\mu, w_\mu, y_\mu, z_\mu)$$

Table of Contents

Path-Following Method

1. Estimate appropriate value for μ.
2. Compute step directions $(\Delta x, \Delta w, \Delta y, \Delta z)$ pointing at the point $(x_\mu, w_\mu, y_\mu, z_\mu)$ on the central path.
3. Compute a new step length parameter θ such that the new point:

$$\begin{aligned} \tilde{x} &= x + \theta \Delta x, & \tilde{y} &= y + \theta \Delta y \\ \tilde{w} &= w + \theta \Delta w, & \tilde{z} &= z + \theta \Delta z \end{aligned}$$

4. Replace (x, w, y, z) with the new solution $(\tilde{x}, \tilde{y}, \tilde{w}, \tilde{z})$

Path-Following Method
1. Estimating μ

- We must find an appropriate value for μ.
- Too high, we might converge to the analytic center of the feasible set.
- Too low, we will converge to the edge of the feasible set that might be suboptimal.

$$\mu = \delta \frac{z^T x + y^T w}{n + m}$$

Path-Following Method
2. Compute step directions $(\Delta x, \Delta w, \Delta y, \Delta z)$

Recall the given equations defining the point $(x_\mu, w_\mu, y_\mu, z_\mu)$ on the central path:

$$\begin{aligned} Ax + w &= b \\ A^T y - z &= c \\ XZe &= \mu e \\ YWe &= \mu e \end{aligned}$$

Path-Following Method
2. Compute step directions $(\Delta x, \Delta w, \Delta y, \Delta z)$

Given a new point $(x_\mu + \Delta x, w_\mu + \Delta w, y_\mu + \Delta y, z_\mu + \Delta z)$, we get the following equations:

$$
\begin{aligned}
A(x + \Delta x) + (w + \Delta w) &= b \\
A^T(y + \Delta y) - (z + \Delta z) &= c \\
(X + \Delta X)(Z + \Delta Z)e &= \mu e \\
(Y + \Delta Y)(W + \Delta W)e &= \mu e
\end{aligned}
$$

Path-Following Method
2. Compute step directions $(\Delta x, \Delta w, \Delta y, \Delta z)$

We rewrite the equations so the unknowns are on the left and the data on the right:

$$\begin{aligned} A\Delta x + \Delta w &= b - Ax - w \\ A^T\Delta y - \Delta z &= c - A^T y + z \\ Z\Delta x + X\Delta z + \Delta X\Delta Z e &= \mu e - XZe \\ W\Delta y + Y\Delta w + \Delta Y\Delta W e &= \mu e - YWe \end{aligned}$$

Path-Following Method

2. Compute step directions $(\Delta x, \Delta w, \Delta y, \Delta z)$

Then we transform the equations into a linear system by dropping the nonlinear terms:

$$\begin{aligned} A\Delta x + \Delta w &= b - Ax - w \\ A^T \Delta y - \Delta z &= c - A^T y + z \\ Z\Delta x + X\Delta z &= \mu e - XZe \\ W\Delta y + Y\Delta w &= \mu e - YWe \end{aligned}$$

Path-Following Method
3. Compute a new step length parameter θ

- Whenever we find the step direction, we need to determine the **step length** θ.
- Recall that we want to replace (x, w, y, z) with the new solution $(\tilde{x}, \tilde{y}, \tilde{w}, \tilde{z})$ by:

$$\begin{aligned} \tilde{x} &= x + \theta\Delta x, & \tilde{y} &= y + \theta\Delta y \\ \tilde{w} &= w + \theta\Delta w, & \tilde{z} &= z + \theta\Delta z \end{aligned}$$

The solution to this system of linear equation corresponds to the aplication of the Newton method on the primal-dual equations and μ-complementary equations.

Path-Following Method

3. Compute a new step length parameter θ

- We need to guarantee that:

  $$x_j + \theta \Delta x > 0, \quad j = 1, 2, \ldots, n.$$

Path-Following Method

3. Compute a new step length parameter θ

- We need to guarantee that:

$$x_j + \theta \Delta x > 0, \quad j = 1, 2, \ldots, n.$$

- Similarly for w, y and z.

Path-Following Method

3. Compute a new step length parameter θ

- We need to guarantee that:

$$x_j + \theta \Delta x > 0, \quad j = 1, 2, \ldots, n.$$

- Similarly for w, y and z.

$$\frac{1}{\theta} = \max_{ij} \left\{ -\frac{\Delta x_j}{x_j}, -\frac{\Delta w_i}{w_i}, -\frac{\Delta y_i}{y_i}, -\frac{\Delta z_j}{z_j} \right\}$$

Path-Following Method

3. Compute a new step length parameter θ

- We need to guarantee that:

$$x_j + \theta \Delta x > 0, \quad j = 1, 2, \ldots, n.$$

- Similarly for w, y and z.

$$\frac{1}{\theta} = \max_{ij} \left\{ -\frac{\Delta x_j}{x_j}, -\frac{\Delta w_i}{w_i}, -\frac{\Delta y_i}{y_i}, -\frac{\Delta z_j}{z_j} \right\}$$

- We introduce a parameter r that is close to but strictly less than one.

$$\theta = r \left(\max_{ij} \left\{ -\frac{\Delta x_j}{x_j}, -\frac{\Delta w_i}{w_i}, -\frac{\Delta y_i}{y_i}, -\frac{\Delta z_j}{z_j} \right\} \right)^{-1} \wedge 1 \qquad a \wedge b \equiv \min\{a, b\}$$

Path-Following Method
4. Replace (x, w, y, z) with $(\tilde{x}, \tilde{y}, \tilde{w}, \tilde{z})$

Now we can replace (x, w, y, z) with $(\tilde{x}, \tilde{y}, \tilde{w}, \tilde{z})$:

$$\begin{aligned} \tilde{x} &= x + \theta \Delta x, & \tilde{y} &= y + \theta \Delta y \\ \tilde{w} &= w + \theta \Delta w, & \tilde{z} &= z + \theta \Delta z \end{aligned}$$

Path-Following Method
Pseudo-code of the path-following method

initialize $(x, w, y, z) > 0$
while (not optimal) {

$$\rho = b - Ax - w$$
$$\sigma = c - A^T y + z$$
$$\gamma = z^T x + y^T w$$
$$\mu = \delta \frac{\gamma}{n+m}$$

solve:

$$\begin{aligned} A\Delta x + \Delta w &= \rho \\ A^T \Delta y - \Delta z &= \sigma \\ Z\Delta x + X\Delta z &= \mu e - XZe \\ W\Delta y + Y\Delta w &= \mu e - YWe \end{aligned}$$

$$\theta = r \left(\max_{ij} \left\{ -\frac{\Delta x_j}{x_j}, -\frac{\Delta w_i}{w_i}, -\frac{\Delta y_i}{y_i}, -\frac{\Delta z_j}{z_j} \right\} \right)^{-1} \wedge 1$$

$$x \leftarrow x + \theta \Delta x, \qquad w \leftarrow w + \theta \Delta w$$
$$y \leftarrow y + \theta \Delta y, \qquad z \leftarrow z + \theta \Delta z$$

}

Optimality

- We have converged to a solution, but how do we know if it is optimal?

Optimality

- We have converged to a solution, but how do we know if it is optimal?
- Recall from the duality theory that we need to meet the following criteria for the solution to be optimal:
  Primal feasibility:

  $$\|\rho\|_1 = \|b - Ax - w\|_1$$

  Dual feasibility:

  $$\|\sigma\|_1 = \|c - A^T y + z\|_1$$

  Complementarity:

  $$\gamma = z^T x + y^T w$$

Optimality

- When do we stop?
- Let $\epsilon > 0$ be a small tolerance and $M < \infty$ be a large finite tolerance
- $\|x\|_\infty > M$ then the primal problem is unbounded.
- $\|y\|_\infty > M$ then the dual problem is unbounded.
- If $\|\rho\|_1 < \epsilon$, $\|\sigma\|_1 < \epsilon$, and $\gamma < \epsilon$ then we found our optimal solution!

Table of Contents

Implementation example

Consider the following LP problem:

Primal problem:

$$\begin{aligned} \max \quad & 5x_1 + 4x_2 + 3x_3 \\ \text{s.t.} \quad & 2x_1 + 3x_2 + x_3 \leq 5 \\ & 4x_1 + x_2 + 2x_3 \leq 11 \\ & 3x_1 + 4x_2 + 2x_3 \leq 8 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

Dual problem:

$$\begin{aligned} \max \quad & 5y_1 + 11y_2 + 8y_3 \\ \text{s.t.} \quad & 2y_1 + 4y_2 + 3_y3 \leq 5 \\ & 3y_1 + y_2 + 4y_3 \leq 4 \\ & y_1 + 2y_2 + 2y_3 \leq 3 \\ & y_1, y_2, y_3 \geq 0 \end{aligned}$$

Implementation example

Both are converted into equality form:

Primal problem:

$$
\begin{aligned}
\max \quad & 5x_1 + 4x_2 + 3x_3 \\
\text{s.t.} \quad & 2x_1 + 3x_2 + x_3 + w_1 = 5 \\
& 4x_1 + x_2 + 2x_3 + w_2 = 11 \\
& 3x_1 + 4x_2 + 2x_3 + w_3 = 8 \\
& x_1, x_2, x_3, w_1, w_2, w_3 \geq 0
\end{aligned}
$$

Dual problem:

$$
\begin{aligned}
\max \quad & 5y_1 + 11y_2 + 8y_3 \\
\text{s.t.} \quad & 2y_1 + 4y_2 + 3_{y_3} + z_1 = 5 \\
& 3y_1 + y_2 + 4y_3 + z_2 = 4 \\
& y_1 + 2y_2 + 2y_3 + z_3 = 3 \\
& y_1, y_2, y_3, z_1, z_2, z_3 \geq 0
\end{aligned}
$$

Implementation example

Recall that we are following the pseudo-code:

initialize $(x, w, y, z) > 0$

while (not optimal) {

$$\rho = b - Ax - w$$

$$\sigma = c - A^T y + z$$

$$\gamma = z^T x + y^T w$$

$$\mu = \delta \frac{\gamma}{n+m}$$

solve:

$$\begin{aligned} A\Delta x + \Delta w &= \rho \\ A^T \Delta y - \Delta z &= \sigma \\ Z\Delta x + X\Delta z &= \mu e - XZe \\ W\Delta y + Y\Delta w &= \mu e - YWe \end{aligned}$$

$$\theta = r \left(\max_{ij} \left\{ -\frac{\Delta x_j}{x_j}, -\frac{\Delta w_i}{w_i}, -\frac{\Delta y_i}{y_i}, -\frac{\Delta z_j}{z_j} \right\} \right)^{-1} \wedge 1$$

$$x \leftarrow x + \theta \Delta x, \qquad w \leftarrow w + \theta \Delta w$$

$$y \leftarrow y + \theta \Delta y, \qquad z \leftarrow z + \theta \Delta z$$

}

Implementation example

Initialize $(x, w, y, z) > 0$ with arbitrary values:

$$x = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix}, w = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix}, y = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix}, z = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix}$$

Implementation example

Initialize $(x, w, y, z) > 0$ with arbitrary values:

$$x = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix}, w = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix}, y = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix}, z = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix}$$

Then initialize b, c, A and A^T

$$b = \begin{bmatrix} 5 \\ 11 \\ 8 \end{bmatrix}, c = \begin{bmatrix} 5 \\ 4 \\ 3 \end{bmatrix}, A = \begin{bmatrix} 2, 3, 1 \\ 4, 1, 2 \\ 3, 4, 2 \end{bmatrix}, A^T = \begin{bmatrix} 2, 4, 3 \\ 3, 1, 4 \\ 1, 2, 2 \end{bmatrix}$$

Implementation example

$$\rho = b - Ax - w = \begin{bmatrix} 5 \\ 11 \\ 8 \end{bmatrix} - \begin{bmatrix} 2,3,1 \\ 4,1,2 \\ 3,4,2 \end{bmatrix} \cdot \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} - \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 4.3 \\ 10.2 \\ 7 \end{bmatrix}$$

Implementation example

$$\rho = b - Ax - w = \begin{bmatrix} 5 \\ 11 \\ 8 \end{bmatrix} - \begin{bmatrix} 2,3,1 \\ 4,1,2 \\ 3,4,2 \end{bmatrix} \cdot \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} - \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 4.3 \\ 10.2 \\ 7 \end{bmatrix}$$

$$\sigma = c - A^T y + z = \begin{bmatrix} 5 \\ 4 \\ 3 \end{bmatrix} - \begin{bmatrix} 2,4,3 \\ 3,1,4 \\ 1,2,2 \end{bmatrix} \cdot \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 4.2 \\ 3.3 \\ 2.6 \end{bmatrix}$$

Implementation example

$$\rho = b - Ax - w = \begin{bmatrix} 5 \\ 11 \\ 8 \end{bmatrix} - \begin{bmatrix} 2,3,1 \\ 4,1,2 \\ 3,4,2 \end{bmatrix} \cdot \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} - \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 4.3 \\ 10.2 \\ 7 \end{bmatrix}$$

$$\sigma = c - A^T y + z = \begin{bmatrix} 5 \\ 4 \\ 3 \end{bmatrix} - \begin{bmatrix} 2,4,3 \\ 3,1,4 \\ 1,2,2 \end{bmatrix} \cdot \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 4.2 \\ 3.3 \\ 2.6 \end{bmatrix}$$

$$\gamma = z^T x + y^T w = \begin{bmatrix} 0.1, 0.1, 0.1 \end{bmatrix} \cdot \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} + \begin{bmatrix} 0.1, 0.1, 0.1 \end{bmatrix} \cdot \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} = 0.06$$

Implementation example

$$\rho = b - Ax - w = \begin{bmatrix} 5 \\ 11 \\ 8 \end{bmatrix} - \begin{bmatrix} 2,3,1 \\ 4,1,2 \\ 3,4,2 \end{bmatrix} \cdot \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} - \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 4.3 \\ 10.2 \\ 7 \end{bmatrix}$$

$$\sigma = c - A^T y + z = \begin{bmatrix} 5 \\ 4 \\ 3 \end{bmatrix} - \begin{bmatrix} 2,4,3 \\ 3,1,4 \\ 1,2,2 \end{bmatrix} \cdot \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 4.2 \\ 3.3 \\ 2.6 \end{bmatrix}$$

$$\gamma = z^T x + y^T w = \begin{bmatrix} 0.1, 0.1, 0.1 \end{bmatrix} \cdot \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} + \begin{bmatrix} 0.1, 0.1, 0.1 \end{bmatrix} \cdot \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} = 0.06$$

$$\mu = \delta \frac{\gamma}{n+m} = 0.1 \frac{0.06}{3+3} = 0.001$$

Where $\delta = 0.1$, $n = 3$ and $m = 3$.

Implementation example

Crucial part, we create our linear system of equations.

$$A\Delta x + \Delta w = \rho \tag{1}$$
$$A^T\Delta y - \Delta z = \sigma \tag{2}$$
$$Z\Delta x + X\Delta z = \mu e - XZe \tag{3}$$
$$W\Delta y + Y\Delta w = \mu e - YWe \tag{4}$$

We define right-hand side first:

$$rhs1 = \begin{bmatrix} 4.3 \\ 10.2 \\ 7 \end{bmatrix}, rhs2 = \begin{bmatrix} 4.2 \\ 3.3 \\ 2.6 \end{bmatrix}, rhs3 = \begin{bmatrix} -0.009 \\ -0.009 \\ -0.009 \end{bmatrix} rhs4 = \begin{bmatrix} -0.009 \\ -0.009 \\ -0.009 \end{bmatrix}$$

Recall that $X = \begin{bmatrix} 0.1, 0, 0 \\ 0, 0.1, 0 \\ 0, 0, 0.1 \end{bmatrix}$. Similarly for W, Y, and Z.

Implementation example

Left-hand side, we define M1, M2, M3 and M4 that corresponds to the right-hand side.

$$M1 = \begin{bmatrix} 2,3,1,1,0,0,0,0,0,0,0,0 \\ 4,1,2,0,1,0,0,0,0,0,0,0 \\ 3,4,2,0,0,1,0,0,0,0,0,0 \end{bmatrix}$$

$$M2 = \begin{bmatrix} 0,0,0,0,0,0,2,4,3,-1,-0,-0 \\ 0,0,0,0,0,0,3,1,4,-0,-1,-0 \\ 0,0,0,0,0,0,1,2,2,-0,-0,-1 \end{bmatrix}$$

$$M3 = \begin{bmatrix} 0.1,0,0,0,0,0,0,0,0,0.1,0,0 \\ 0,0.1,0,0,0,0,0,0,0,0,0.1,0 \\ 0,0,0.1,0,0,0,0,0,0,0,0,0.1 \end{bmatrix}$$

$$M4 = \begin{bmatrix} 0,0,0,0.1,0,0,0.1,0,0,0,0,0 \\ 0,0,0,0,0.1,0,0,0.1,0,0,0,0 \\ 0,0,0,0,0,0.1,0,0,0.1,0,0,0 \end{bmatrix}$$

Implementation example

Using Python (3.10) and numpy (1.26.4), we solve the system using `np.linalg.solve(M, rhs)` to retrieve the delta values:

$$\Delta x = \begin{bmatrix} 2.011 \\ -0.153 \\ 1.147 \end{bmatrix}, \Delta w = \begin{bmatrix} -0.411 \\ 0.015 \\ -0.716 \end{bmatrix}, \Delta y = \begin{bmatrix} 0.321 \\ -0.105 \\ 0.626 \end{bmatrix}, \Delta z = \begin{bmatrix} -2.101 \\ 0.063 \\ -1.237 \end{bmatrix}$$

Then we retrieve θ:

$$\begin{aligned} \theta = r\left(\max_{ij}\left\{-\frac{\Delta x_j}{x_j}, -\frac{\Delta w_i}{w_i}, -\frac{\Delta y_i}{y_i}, -\frac{\Delta z_j}{z_j}\right\}\right)^{-1} \wedge 1 \\ = 0.043 \end{aligned}$$

Implementation example

New solution:

$$x = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} + 0.011 \cdot \begin{bmatrix} 2.011 \\ -0.153 \\ 1.147 \end{bmatrix} = \begin{bmatrix} 0.186 \\ 0.093 \\ 0.149 \end{bmatrix}$$

$$w = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} + 0.011 \cdot \begin{bmatrix} -0.411 \\ 0.015 \\ -0.716 \end{bmatrix} = \begin{bmatrix} 0.082 \\ 0.100 \\ 0.069 \end{bmatrix}$$

$$y = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} + 0.011 \cdot \begin{bmatrix} 0.321 \\ -0.105 \\ 0.626 \end{bmatrix} = \begin{bmatrix} 0.114 \\ 0.095 \\ 0.127 \end{bmatrix}$$

$$z = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.1 \end{bmatrix} + 0.011 \cdot \begin{bmatrix} -2.101 \\ 0.063 \\ -1.237 \end{bmatrix} = \begin{bmatrix} 0.01 \\ 0.103 \\ 0.047 \end{bmatrix}$$

Implementation example

Is this the optimal solution?
Primal feasibility:

$$\|\rho\|_1 = 20.58$$

Dual feasibility:

$$\|\sigma\|_1 = 9.67$$

Complementarity:

$$\gamma = 0.068$$

As $\|\rho\|_1 \geq \epsilon$, $\|\sigma\|_1 \geq \epsilon$ and $\gamma \geq \epsilon$, we have not found an optimal solution yet. Therefore, we continue on our second iteration with the new x, w, y and z values.

Table of Contents

The KKT System

Given a system of equations:

$$A\Delta x + \Delta w = \rho \tag{5}$$

$$A^T \Delta y - \Delta z = \sigma \tag{6}$$

$$Z\Delta x + X\Delta z = \mu e - XZe \tag{7}$$

$$W\Delta y + Y\Delta w = \mu e - YWe \tag{8}$$

The KKT System

Given a system of equations:

$$A\Delta x + \Delta w = \rho \tag{5}$$

$$A^T\Delta y - \Delta z = \sigma \tag{6}$$

$$Z\Delta x + X\Delta z = \mu e - XZe \tag{7}$$

$$W\Delta y + Y\Delta w = \mu e - YWe \tag{8}$$

In the previous example, it might be possible to solve a linear system of equations in a small problem. But what if the problem is larger? We transform the system into a **symmetric** linear system in matrix form:

$$\left[\begin{array}{c|c} -XZ^{-1} & \begin{matrix} -I & \\ A & I \end{matrix} \\ \hline \begin{matrix} -I & A^T \\ & I \end{matrix} & YW^{-1} \end{array}\right] \left[\begin{array}{cc} \Delta z & \Delta y \\ \hline \Delta x & \Delta w \end{array}\right] = \left[\begin{array}{cc} -\mu Z^{-1}e + x & \rho \\ \hline \sigma & \mu W^{-1}e - y \end{array}\right]$$

This is the Karush-Kuhn-Tucker system (KKT system)

The Reduced KKT System

The KKT system can be reduced further. We solve for Δz and Δw in equations 7 and 8.

$$\begin{aligned}
\Delta z &= X^{-1}(\mu e - XZ\epsilon - Z\Delta x) \\
\Delta w &= Y^{-1}(\mu e - YW\epsilon - W\Delta y)
\end{aligned}$$

Then we substitute Δz and Δw into equations 5 and 6.

$$\begin{aligned}
A\Delta x - Y^{-1}W\Delta y &= \rho - \mu Y^{-1}e + w \\
A^T\Delta y - X^{-1}Z\Delta x &= \sigma + \mu X^{-1}e - z
\end{aligned}$$

This gives us the following reduced KKT System.

The Reduced KKT System

$$\begin{bmatrix} -Y^{-1}W & A \\ A^T & X^{-1}Z \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta x \end{bmatrix} = \begin{bmatrix} b - Ax - \mu Y^{-1}e \\ c - A^T y + \mu X^{-1}e \end{bmatrix}$$

The Reduced KKT System is still symmetric. However, we can keep reducing the system into normal equations.

Normal Equations

Given the reduced KKT System:

$$A\Delta x - Y^{-1}W\Delta y = \rho - \mu Y^{-1}e + w \tag{9}$$
$$A^T\Delta y - X^{-1}Z\Delta x = \sigma + \mu X^{-1}e - z \tag{10}$$

We solve for Δy in equation 9 and eliminate it from 10 OR
We solve for Δx in equation 10 and eliminate it from 9. We choose the latter.

$$\Delta x = -XZ^{-1}(c - A^T y + \mu X^{-1}e - A^T\Delta y)$$

Then we eliminate Δx:

$$\begin{aligned} -(Y^{-1}W + AXZ^{-1}A^T)\Delta y = b - Ax - \mu Y^{-1}e \\ - AXZ^{-1}(c - A^T y + \mu X^{-1}e) \end{aligned}$$

Normal Equations

- This gives us a system of normal equations in primal form.
- Similarly, if we choose the former option, we get a system of normal equations in dual form.
- Problem: If A has a dense column, then we end up with a dense matrix which is difficult to solve in primal form.
- Same problem if A has a dense row, which is also difficult to solve in dual form.
- Should we then use the reduced KKT matrix? Possible if the matrices are positive definite!