Interior Point Method I

Department of Statistics

November 9, 2023

University of Seoul

Interior Point Methods

Interior Point Methods

Consider a convex optimization problem

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \le 0 \quad i = 1, \dots, m \\
& Ax = b,
\end{aligned} \tag{1}
$$

where $f_0, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$ are convex and twice continuously differentiable, $A \in \mathbb{R}^{p \times n}$ with $\text{rank}(A) = p < n$.

For simplicity, assume that f_0 is quadratic and f_i for $i = 1, \cdots, m$ are linear.

If Slater's constraint qualification holds, x^* and $(\lambda^*, \nu^*) \in \mathbb{R}^m \times \mathbb{R}^p$ are primal and dual solution that the follow KKT conditions

$$\begin{aligned} Ax^* = b,\ f_i(x^*) &\leq 0,\ i = 1, \ldots, m \\ \lambda^* &\geq 0 \\ \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + A^\top \nu^* &= 0 \\ \lambda_i^* f_i(x^*) &= 0,\ i = 1, \ldots, m \end{aligned} \tag{2}$$

Without inequality constraints, the above KKT conditions derive a linear system that can be easily solved. However, the inequality constraints lead to a different story about the optimization problem.

Logarithmic barrier function

Consider an equivalent optimization problem to the original one:

$$\begin{array}{ll} \text{minimize} & f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) \\ \text{subject to} & Ax = b, \end{array} \tag{3}$$

where

$$I_-(u) = \begin{cases} 0 & u \le 0 \\ \infty & u > 0. \end{cases}$$

Note that I_- is not differentiable, which makes the problem computationally prohibitive. Instead, we use a differentiable function (see Figure 1 on the next page),

$$\hat{I}_-(u) = -(1/t)\log(-u) \text{ for } t > 0.$$

The original optimization problem is relaxed by

$$\begin{aligned} \text{minimize} \quad & f_0(x) + \sum_{i=1}^{m} -(1/t)\log(-f_i(x)) \\ \text{subject to} \quad & Ax = b. \end{aligned} \tag{4}$$

For a large $t > 0$ the relaxed objective function is well approximated to (3). Moreover, the second-order approximated algorithm, such as Newton's method, can be applied to solve (4) for fixed t. (see Example 15 in the slide of the dual problem)

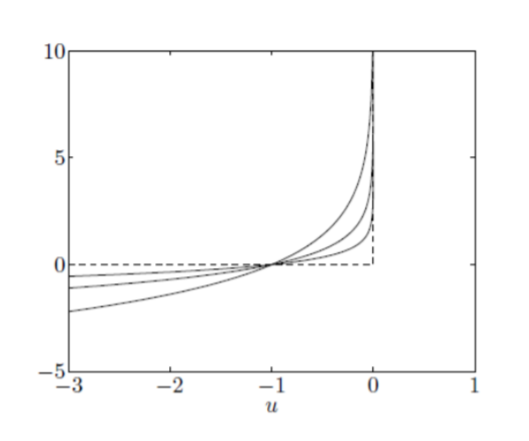

Figure 1: The dotted line displays $I_{-}(u)$, and the solid curves display $\hat{I}_{(}u)$. As t increases, $\hat{I}_{-}(u)$ becomes tighter above $\hat{I}_{(}u)$

Recall the Newton's method applied to (4).

- Set an initial $x^{(k)}$ for $k = 0$ where $Ax^{(k)} = b$ and $f_i(x^{(k)}) < 0$.
- Obtain the quadratic function $g(x|x^{(k)})$ by the second order approximation of the objective function (4) at $x^{(k)}$.
- Let $x = x^{(k)} + \nu$ and solve the following problem:

$$\begin{aligned} \min_{\nu} \quad & g(x^{(k)} + \nu | x^{(k)}) \\ \text{subject} \quad & A\nu = 0. \end{aligned}$$

  (The problem can be easily solved.)
- Update and repeat.

Let $\phi(x) = -\sum_{i=1}^{m} \log(-f_i(x))$ which is called the logarithmic barrier function for problem (1). Note that

$$\mathrm{dom}(\phi) = \{x \in \mathbb{R}^n \mid f_i < 0, i = 1, \ldots, m\}.$$

Then we consider the equivalent optimization problem to (4) as

$$\begin{aligned} \text{minimize} \quad & t f_0(x) + \phi(x) \\ \text{subject to} \quad & Ax = b, \end{aligned} \tag{5}$$

and denote the solution by $x^*(t)$. Next we investigate the properties of $x^*(t)$.

Central path

Let $x^*(t)$ be a solution of (5). The central path is a collection of $x^*(t)$ for $t > 0$. By KKT conditions (convex objective function with equality constraints),

- (Stationary Condition) There exists a $\hat{\nu} \in \mathbb{R}^p$ such that

$$\mathbf{0} \quad = \quad t\nabla f_0(x^*(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(x^*(t))} \nabla f_i(x^*(t)) + A^\top \hat{\nu} \tag{6}$$

- (Feasibility condition) $x^*(t)$ is strictly feasible

$$Ax^*(t) = b, \qquad f_i(x^*(t)) < 0,\ i = 1, \ldots, m;$$

Let

$$\lambda_i^* = -\frac{1}{t f_i(x^*(t))}, \; i = 1, \ldots, m, \qquad \nu^* = \hat{\nu}/t,$$

then we will claim that the pair λ^*, ν^* dual feasible. (Note that λ^* and ν^* depend on t.)

- $\lambda^* > \mathbf{0}$ because $f_i(x^*(t)) < 0, \; i = 1, \ldots, m.$
- $\nabla f_0(x^*(t)) + \sum_{i=1}^m \lambda^* \nabla f_i(x^*(t)) + A^\top \nu^* = 0$ implies that x^* minimizes the Lagrangian,

$$L(x, \lambda^*, \nu^*) = f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + (\nu^*)^\top (Ax - b).$$

Hence, we know that the dual function associated with (1) is

$$\begin{aligned} g(\lambda^*, \nu^*) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x)) + (\nu^*)^\top (Ax - b) \right) \\ &= f_0(x^*(t)) + \sum_{i=1}^m \lambda_i^* f_i(x^*(t)) + (\nu^*)^\top (Ax^*(t) - b) \\ &= f_0(x^*(t)) - m/t. \end{aligned}$$

The first equality holds by the definition of dual function, and the second equality holds by the optimality of $L(x, \lambda^*, \nu^*)$, and the last equality holds because

$$\begin{aligned} \lambda_i^* f_i(x^*(t)) = -\frac{1}{t f_i(x^*(t))} f_i(x^*(t)) &= -1/t, \\ (Ax^*(t) - b) &= 0. \end{aligned}$$

Theorem 1 (suboptimality)

Let p^* be the optimal value of (1), then

$$f_0(x^*(t)) - p^* \leq m/t.$$

(proof) Let p^* be the optimal value of (1), then $g(\lambda^*, \nu^*) \leq p^*$ by weak duality. Thus,

$$f_0(x^*(t)) - m/t = g(\lambda^*, \mu^*) \leq p^*.$$

Finally, we obtain $f_0(x^*(t)) - p^* \leq m/t$. ($x^*(t)$ is no more than m/t-suboptimal)

The centrality condition (perturbed KKT conditions)

The KKT conditions for a relaxed problem can be understood as a modified KKT condition:

- $x = x^*(t)$ is a point on the central path if and only if there exists $\lambda = \lambda^*(t)$ and $\nu = \nu^*(t)$ such that

$$
\begin{aligned}
Ax = b, \quad f_i(x) &\leq 0, \qquad i = 1, \ldots, m \\
\lambda &> 0 \\
\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i x) + A^T \nu &= 0 \\
\lambda_i f_i(x) &= -1/t, \qquad i = 1, \ldots, m
\end{aligned}
\tag{7}
$$

- The complementary slackness condition $-\lambda_i f_i(x) = 0$ is replaced by $\lambda_i f_i(x) = -1/t$.

interpretation of the central path

- The inequality conditions are relaxed by a barrier function.
- The associated KKT conditions are relaxed. The complementary slackness condition is relaxed as

$$\lambda_i f_i(x) = -1/t$$

- The modified problem can be solved by the Newton-Raphson algorithm with equality constraints.
- The solution to the problem is suboptimal.

The Barrier method

If we set an upper bound of error as $\epsilon > 0$, we only let $t = m/\epsilon$ and solve the modified problem. However, we encounter the problem between numerical stability and accuracy since the algorithm becomes unstable for a large t.

(HW) Discuss the above problem related to the convergence of Newton's algorithm.

Barrier method algorithm

- Given strictly feasible x, $t := t^{(0)} > 0$, $\mu > 1$, tolerance $\epsilon > 0$
- Repeat:
  - Centering step : Compute $x^*(t)$ by minimizing $tf_0 + \phi$, subject to $Ax = b$, starting at x
  - Update $x := x^*(t)$
  - Stopping criterion : quit if $m/t < \epsilon$
  - Increase t: $t := \mu t$

- Execution of step 1 as a centering step or an outer iteration
- Newton iterations or steps executed during the centering step as inner iterations
- At each inner step, we have a primal feasible point
- At each outer step, we have a dual feasible point

Example 2 (Linear programming)

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & \mathbf{c}^T x \\
\text{subject to} \quad & Ax = \mathbf{b} \\
& x \succeq 0
\end{aligned}
\tag{8}
$$

where A is $p \times n(p < n)$ full rank matrix, $\mathbf{c} \in \mathbb{R}^p$, $\mathbf{b} \in \mathbb{R}^n$.

The objective function with barrier function is given by

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & B(x,t) = t\mathbf{c}^T x - \sum_{i=1}^{n} \log(x_i) \\ \text{subject to} \quad & Ax = b \end{aligned} \tag{9}$$

where real $t > 0$

The gradient and hessian of (9) is given by

$$\begin{aligned} \text{Gradient} &: \nabla_x B(x,t) = t\mathbf{c} - X^{-1}e \\ \text{Hessian} &: \nabla_x^2 B(x,t) = X^{-2} \end{aligned}$$

where $X = \text{diag}(x_1, \ldots, x_n)$, $e = (1, \ldots, 1)^T$. The quadratic function is obtained by second-order approximation near x of the logarithmic barrier function:

$$\begin{aligned} \min_{\Delta \in \mathbb{R}^p} \quad & t\mathbf{c}^T x - \sum_{i=1}^{n} \log(x_i) + (t\mathbf{c} - X^{-1}e)^T \Delta + \frac{1}{2}\Delta^T X^{-2} \Delta \\ \text{subject to} \quad & A(x + \Delta) = \mathbf{b} \end{aligned}$$

Note that $Ax = \mathbf{b}$.

The corresponding Lagrangian function is given by

$$L(v,\nu) = t\mathbf{c}^T x - \sum_{i=1}^{n} \log(x_i) + (t\mathbf{c} - X^{-1}e)^T \Delta + \frac{1}{2}\Delta^T X^{-2} \Delta + \nu^T A\Delta$$

The KKT conditions are

$$\begin{aligned} t\mathbf{c} - X^{-1}e + X^{-2}\Delta + A^T\nu &= 0 \text{ (stationarity)} \\ A\Delta &= 0 \text{ (primal feasibility)} \end{aligned}$$

and the linear equations corresponding to the conditions are written by

$$\begin{bmatrix} X^{-2} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta \\ \nu \end{bmatrix} = \begin{bmatrix} -t\mathbf{c} + X^{-1} \\ 0 \end{bmatrix}.$$

Since A is full rank, the linear system has the unique solution (primal and dual solution).

The solution is given by

$$\begin{bmatrix} \Delta \\ \nu \end{bmatrix} = \begin{bmatrix} X^{-2} & A^T \\ A & 0 \end{bmatrix}^{-1} \begin{bmatrix} -t\mathbf{c} + X^{-1} \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} X^2 - X^2 A^T \Theta^{-1} A X^2 & X^2 A^T \Theta^{-1} \\ \Theta^{-1} A X^2 & -\Theta^{-1} \end{bmatrix} \begin{bmatrix} -t\mathbf{c} + X^{-1} \\ 0 \end{bmatrix}$$

where $\Theta = -AX^2A^T$. (See the inversion of block matrix for the second equation)

How to find an initial x satisfying $Ax = b$.

- $x = (A^\top A)^- A^\top b$ where G^- is the generalized inverse matrix of G.
- Solve $\min_x \ \|Ax - b\|^2$ by the gradient method.

Example 3 (Isotonic regression)

$$\begin{aligned} \min_{\beta_0,\cdots,\beta_p} \quad & \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2 \\ \text{subject to} \quad & \beta_1 \le \beta_2 \le \cdots \le \beta_p \end{aligned}$$

(hint) Let $\delta_j = \beta_j - \beta_{j-1}$ for $j \geq 2$ then $\delta_j \geq 0$.