HW2

S22001011 冯靖轩

June 19, 2023

1 Introduction

Implementation of the Primal-dual interior-point method

2 Primal-dual interior-point method

2.1 Problem

The convex optimization problems that include inequality constraints:

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

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

We also assume that the problem is strictly feasible, i.e., $\exists x \in \mathcal{D}$ satisfying $Ax = b$ and $f_i(x) < 0$ for $i = 1, \ldots, m$. This means that Slater's constraint qualification holds, and therefore strong duality holds, so there exists dual optimal $\lambda^* \in R^m, \nu^* \in R^p$, which together with x^* satisfy the KKT conditions:

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

2.2 Method

Rewrite the problem (77) and make the inequality constraints implicit in the objective:

$$\begin{array}{ll} \min & f_0(x) + \sum_{i=1}^{m} I_-\left(f_i(x)\right) \\ \text{s.t.} & Ax = b, \end{array}$$

where

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

The basic idea of the barrier method is to approximate the indicator function I_- by the function

$$\hat{I}_-(u) = -(1/t)\log(-u), \quad \operatorname{dom}\hat{I}_- = -R_{++}$$

where t is a parameter that sets the accuracy of the approximation. Obviously, $\hat{I}_-$ is convex, nondecreasing and differentiable.

Substituting $\hat{I}_-$ for I_- in (79) gives the approximation

$$\begin{aligned} &\min f_0(x) + \sum_{i=1}^{m} -(1/t)\log\left(-f_i(x)\right) \\ &\text{s.t. } Ax = b. \end{aligned}$$

The function

$$\phi(x) = -\sum_{i=1}^{m} \log\left(-f_i(x)\right),$$

is called the logarithmic barrier for the problem (77). Its domain is the set of points that satisfy the inequality constraints of (77) strictly:

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

We multiply the objective of (80) by t, and consider the equivalent problem

$$\begin{aligned} \min \quad & tf_0(x) + \phi(x) \\ \text{s.t.} \quad & Ax = b. \end{aligned}$$

We assume problem (82) can be solved via Newton's method, and, that it has a unique solution for each $t > 0$.

For $t > 0$ we define $x^*(t) = \arg\min_x \{tf_0(x) + \phi(x) \text{ s.t. } Ax = b\}$ as the solution of (82).

The central path associated with problem (77) is defined as the set of points $\{x^*(t) \mid t > 0\}$, which we call the central points.

Points on the central path are characterized by the following necessary and sufficient conditions: $x^*(t)$ is strictly feasible, i.e., satisfies

$$Ax^*(t) = b, \quad f_i\left(x^*(t)\right) < 0, i = 1, \ldots, m$$

and $\exists \hat{\nu} \in R^p$ such that

$$\begin{aligned} 0 &= t\nabla f_0\left(x^*(t)\right) + \nabla\phi\left(x^*(t)\right) + A^\top\hat{\nu} \\ &= t\nabla f_0\left(x^*(t)\right) + \sum_{i=1}^{m} \frac{1}{-f_i\left(x^*(t)\right)}\nabla f_i\left(x^*(t)\right) + A^\top\hat{\nu} \end{aligned}$$

holds.

Every central point yields a dual feasible point. Define

$$\lambda_i^*(t) = -\frac{1}{tf_i\left(x^*(t)\right)}, \quad i = 1, \ldots, m, \quad \nu^*(t) = \frac{\hat{\nu}}{t}.$$

Because $f_i\left(x^*(t)\right) < 0, i = 1, \ldots, m$, it's clear that $\lambda^*(t) > 0$.

Since we have assumed that $x^*(t)$ is the unique solution to problem (82) for each $t > 0$, a point is equal to $x^*(t)$ if and only if $\exists \lambda, \nu$ such that

$$Ax = b, \quad f_i(x) \leq 0, \quad i = 1, \ldots, m$$
$$\lambda \geq 0$$
$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + A^\top \nu = 0$$
$$-\lambda_i f_i(x) = 1/t, \quad i = 1, \ldots, m.$$

The only difference between (85) and the KKT condition (78) is that the complementarity condition $-\lambda_i f_i(x) = 0$ is replaced by the condition $-\lambda_i f_i(x) = 1/t$ In particular, for large $t, x^*(t)$ and $\lambda^*(t), \nu^*(t)$ 'almost' satisfy the KKT optimality conditions for the problem (77).

The modified KKT conditions (87) can be expressed as $r_t(x, \lambda, \nu) = 0$, where $t > 0$ and

$$r_t(x, \lambda, \nu) = \begin{bmatrix} \nabla f_0(x) + J[f(x)]^\top \lambda + A^\top \nu \\ -\operatorname{diag}(\lambda) f(x) - (1/t)\mathbf{1} \\ Ax - b \end{bmatrix}.$$

Here $f : R^n \to R^m$ and $J[f]$ are given by

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix}, \quad J[f(x)] = \begin{bmatrix} \nabla f_1(x)^\top \\ \vdots \\ \nabla f_m(x)^\top \end{bmatrix}$$

If x, λ, ν satisfy $r_t(x, \lambda, \nu) = 0$ (and $f_i(x) < 0$), then $x = x^*(t), \lambda = \lambda^*(t)$ and $\nu = \nu^*(t)$ - The first block component of r_t,

$$r_{\text{dual}} = \nabla f_0(x) + J[f(x)]^\top \lambda + A^\top \nu$$

is called the dual residual. - The last block component, $r_{\text{pri}} = Ax - b$, is called the primal residual. - The middle block

$$r_{\text{cent}} = -\operatorname{diag}(\lambda) f(x) - (1/t)\mathbf{1}$$

is the centrality residual, i.e., the residual for the modified complementarity condition.

Let $y = (x, \lambda, \nu)$ denote the current point and $\delta_y = (\delta_x, \delta_\lambda, \delta_\nu)$ denote the Newton step for solving the equation $r_t(x, \lambda, \nu) = 0$, for fixed t where $f(x) < 0, \lambda > 0$ The Newton step is characterized by

$$r_t(y + \delta_y) \approx r_t(y) + J[r_t(y)]\, \delta_y = 0.$$

In terms of x, λ, ν, we have

$$\begin{bmatrix} \nabla^2 f_0(x) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(x) & J[f(x)]^\top & A^\top \\ -\operatorname{diag}(\lambda) J[f(x)] & -\operatorname{diag}(f(x)) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \delta_x \\ \delta_\lambda \\ \delta_\nu \end{bmatrix} = - \begin{bmatrix} r_{\text{dual}} \\ r_{\text{cent}} \\ r_{\text{pri}} \end{bmatrix}$$

The primal-dual search direction $\delta_{y_{\text{pd}}} = \left(\delta_{x_{\text{pd}}}, \delta_{\lambda_{\text{pd}}}, \delta_{\nu_{\text{pd}}}\right)$ is defined as the solution of (92).

We define the surrogate duality gap, for any x that satisfies $f(x) < 0$ and $\lambda \geq 0$, as

$$\hat{\eta}(x, \lambda) = -f(x)^\top \lambda.$$

2.3 Algorithm

Primal-dual interior-point method

Input: x that satisfies

$$f_1(x) < 0, \ldots, f_m(x) < 0, \lambda > 0, \gamma > 1, \epsilon_{\text{feas}} > 0, \epsilon > 0.$$

1: **procedure** MAIN

2: **while** $\|r_{\text{pri}}\|_2 \leq \epsilon_{\text{feas}}$, $\|r_{\text{dual}}\|_2 \leq \epsilon_{\text{feas}}$, and $\hat{\eta} \leq \epsilon$ **do**

(1) Determine t. Set $t := \gamma(m/\hat{\eta})$.

(2) Compute primal-dual search direction $\delta_{y_{\text{pd}}}$.

(3) Line search and update. Determine step length $\alpha > 0$ and set $y := y + \alpha\delta_{y_{\text{pd}}}$.

The line search in step 3 is a standard backtracking line search. For a step size α, let

$$y^+ = \begin{bmatrix} x^+ \\ \lambda^+ \\ \nu^+ \end{bmatrix} = \begin{bmatrix} x \\ \lambda \\ \nu \end{bmatrix} + \alpha \begin{bmatrix} \delta_{x_{\text{pd}}} \\ \delta_{\lambda_{\text{pd}}} \\ \delta_{\nu_{\text{pd}}} \end{bmatrix}$$

Let

$$\alpha^{\max} = \sup\{\alpha \in [0,1] \mid \lambda + \alpha\delta_\lambda \geq 0\} = \min\left\{1, \min\left\{\frac{-\lambda_i}{\delta_{\lambda_i}} \mid \delta_{\lambda_i} < 0\right\}\right\}$$

to be the largest positive step length that gives $\lambda^+ \geq 0$.

We start backtracking with $\alpha = 0.99\alpha^{\max}$, and multiply α by $\beta \in (0,1)$ until we have $f(x^+) < 0$. We continue multiplying α by β until we have

$$\left\|r_t\left(x^+, \lambda^+, \nu^+\right)\right\|_2 \leq (1 - \tau\alpha)\|r_t(x, \lambda, \nu)\|_2 .$$

Here τ is typically chosen in the range $[0.01, 0.1]$.

3 Results

See "results.pdf"

4 Result Analysis

We use two examples to show that our implementation is able to solve common convex optimization problems. We experimentally show that the interior point method works well with different initial point.

Increasing the value of γ doesn't lead to failure perhaps because the examples we select are easy to solve.