Numerical Optimization, 2020 Fall

Homework 8

Due 14:59 (CST), Dec. 10, 2020

(NOTE: Homework will not be accepted after this due for any reason.)

Throughout this assignment, we focus on the following trust region subproblem, which reads

$$\begin{aligned} \min_{\boldsymbol{d}\in\mathbb{R}^n} \quad & m_k(\boldsymbol{d}) := f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k + \tfrac{1}{2}\boldsymbol{d}_k^T H_k \boldsymbol{d}_k \\ \text{s.t.} \quad & \|\boldsymbol{d}\| \le \Delta_k, \end{aligned} \tag{1}$$

where $\Delta_k > 0$ is the trust-region radius.

Note: Throughout this assignment, the notion of positive definiteness applies exclusively to symmetric matrices. Thus whenever we say that a matrix is positive (semi)definite, we implicitly assume that the matrix is symmetric.

1 Cauchy point calculation

[20pts] Please write down a closed-form expression of the Cauchy point. (Make sure you provided detailed proof; otherwise you won't earn marks.)

Specifically, first solve the a linear version of (1) to obtain vector $\boldsymbol{d}_k^s$, that is,

$$\boldsymbol{d}_k^s = \arg\min_{\boldsymbol{d}\in\mathbb{R}^n} f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k \qquad \text{s.t.} \quad \|\boldsymbol{d}\| \le \Delta_k. \tag{2}$$

Then, calculate the scalar $\tau_k > 0$ that minimizes $m_k(\tau \boldsymbol{d}_k^s)$ subject to the trust region bound, that is

$$\tau_k = \arg\min_{\tau\ge 0} \ m_k(\tau \boldsymbol{d}_k^s) \qquad \text{s.t.} \quad \|\tau \boldsymbol{d}_k^s\| \le \Delta_k. \tag{3}$$

Set $\boldsymbol{d}_k^c = \tau_k \boldsymbol{d}_k^s$.

(1) $d_k^s = \underset{d \in \mathbb{R}^n}{\arg\min}\ f(x_k) + \nabla f(x_k)^T d_k$

s.t. $\|d\| \le \Delta_k$

$\min\ f(x_k) + \nabla f(x_k)^T d_k = f(x_k) + \nabla f(x_k)^T \left[-\frac{\Delta_k}{\|\nabla f(x_k)\|} \nabla f(x_k) \right]$ Since $\forall d \in \mathbb{R}^n, d = \cos\theta \nabla f(x_k), \theta \in [0, 2\pi), \|d\| = \Delta_k$

$d_k^s = -\frac{\Delta_k}{\|\nabla f(x_k)\|} \nabla f(x_k)$ $\quad \nabla f(x_k)^T d = \nabla f(x_k)^T \nabla f(x_k) \cos\theta \ge -\nabla f(x_k)^T \nabla f(x_k)$ when $\theta = -\pi$

(2) $\tau_k = \underset{\tau > 0}{\arg\min}\ m_k(\tau d_k^s)$

s.t. $\|\tau d_k^s\| \le \Delta_k$

$m_k(\tau d_k^s) = f(x_k) + \nabla f(x_k)^T d_k^s \tau + \frac{1}{2} d_k^{sT} H_k d_k^s \tau^2$

$\|\tau\| \le 1$

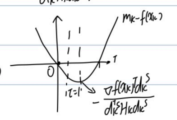
$m_k = \frac{1}{2} d_k^{sT} H_k d_k^s \left[\tau^2 + \frac{2\nabla f(x_k)^T d_k^s}{d_k^{sT} H_k d_k^s} \tau \right] + f(x_k)$ $\quad \nabla f(x_k)^T d_k^s < 0$

$= \frac{1}{2} d_k^{sT} H_k d_k^s \tau \left(\tau + \frac{2\nabla f(x_k)^T d_k^s}{d_k^{sT} H_k d_k^s} \right) + f(x_k)$

① $d_k^{sT} H_k d_k^s > 0$

$\Updownarrow$

$\nabla f(x_k)^T H_k \nabla f(x_k) > 0$

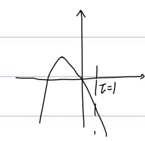


② $d_k^{sT} H_k d_k^s \le 0$

$\Updownarrow$

$\nabla f(x_k)^T H_k \nabla f(x_k) \le 0$

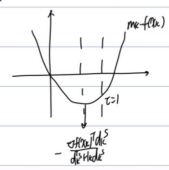
$\tau_k = 1$



1° $-\frac{\nabla f(x_k)^T d_k^s}{d_k^{sT} H_k d_k^s} \ge 1$

$\tau_k = 1$

2° $-\frac{\nabla f(x_k)^T d_k^s}{d_k^{sT} H_k d_k^s} < 1$



$\tau_k = \frac{-\nabla f(x_k)^T d_k^s}{d_k^{sT} H_k d_k^s}$

$= \frac{\frac{\Delta_k}{\|\nabla f(x_k)\|} \cdot \nabla f(x_k)^T \nabla f(x_k)}{\left(\frac{\Delta_k}{\|\nabla f(x_k)\|} \right)^2 \nabla f(x_k)^T H_k \nabla f(x_k)}$

$= \frac{\|\nabla f(x_k)\|^3}{\Delta_k \nabla f(x_k)^T H_k \nabla f(x_k)}$

$\therefore$ ①② $\Rightarrow d_k^c = \begin{cases} -\frac{\Delta_k}{\|\nabla f(x_k)\|} \nabla f(x_k) & \nabla f(x_k)^T H_k \nabla f(x_k) \le 0 \text{ or } \nabla f(x_k)^T H_k \nabla f(x_k) > 0 \text{ and } \frac{\|\nabla f(x_k)\|^3}{\Delta_k \nabla f(x_k)^T H_k \nabla f(x_k)} \ge 1 \\ -\frac{\|\nabla f(x_k)\|^2 \nabla f(x_k)}{\nabla f(x_k)^T H_k \nabla f(x_k)} & \nabla f(x_k)^T H_k \nabla f(x_k) > 0 \text{ and } \frac{\|\nabla f(x_k)\|^3}{\Delta_k \nabla f(x_k)^T H_k \nabla f(x_k)} < 1 \end{cases}$

2 Local convergence for trust region methods

[20pts] Given a step $\boldsymbol{d}_k$, consider the ratio (with positive denominator):

$$\rho_k := \frac{f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + \boldsymbol{d}_k)}{m_k(\boldsymbol{0}) - m_k(\boldsymbol{d}_k)}. \tag{4}$$

Show that if $\Delta_k \to 0$, then $\rho_k \to 1$. (This proves that for Δ_k sufficiently small, $m_k(\boldsymbol{d})$ approximates $f(\boldsymbol{x}_k + \boldsymbol{d}_k)$ well.)

2. As the algorithm will stop when $\nabla f(x^k) = 0$, so we consider $\nabla f(x^k) \ne 0$, and we choose d_k such that $\nabla f(x^k)^T d_k < 0$

$f(x_k) - f(x_k + d_k) = -\nabla f(x_k)^T d_k - \frac{1}{2} d_k^T \nabla^2 f(x_k) d_k - o(\|d_k\|^2)$ $\quad \|d_k\| \le \Delta_k,\ \Delta_k \to 0 \Rightarrow \|d_k\| \to 0 \Rightarrow d_k \to 0$

$m(0) - m(d_k) = -\nabla f(x_k)^T d_k - \frac{1}{2} d_k^T \nabla^2 f(x_k) d_k > 0$

$\lim_{\Delta_k \to 0} \rho_k = \lim_{\|d_k\| \to 0} \frac{f(x_k) - f(x_k + d_k)}{m(0) - m(d_k)}$ $\quad \because \nabla f(x_k)^T d_k + \frac{1}{2} d_k^T \nabla^2 f(x_k) d_k = [\nabla f(x_k)^T + \frac{1}{2} d_k^T \nabla^2 f(x_k)] d_k < 0 \quad d_k \to 0$

$= 1 + \lim_{\|d_k\| \to 0} \frac{o(\|d_k\|^2)}{\nabla f(x_k)^T d_k + \frac{1}{2} d_k^T \nabla^2 f(x_k) d_k}$ $\quad \therefore [\nabla f(x_k)^T + \frac{1}{2} d_k^T \nabla^2 f(x_k) + d_k^T] d_k < 0 \quad d_k \to 0 \Rightarrow \nabla f(x_k)^T d_k + \frac{1}{2} d_k^T \nabla^2 f(x_k) d_k \le -\|d_k\|^2 \quad d_k \to 0$

$\left| \lim_{\|d_k\| \to 0} \frac{o(\|d_k\|^2)}{\nabla f(x_k)^T d_k + \frac{1}{2} d_k^T \nabla^2 f(x_k) d_k} \right| \le \left| \lim_{\|d_k\| \to 0} \frac{o(\|d_k\|^2)}{\|d_k\|^2} \right|$

$= \left| \lim_{\|d_k\| \to 0} o(1) \right|$

$= 0$

$\therefore \lim_{\Delta_k \to 0} \rho_k = 1$

3 Exact line search

[20pts] Consider minimizing the following quadratic function

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \quad f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T Q\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}, \tag{5}$$

where $Q \in \mathbb{R}^{n\times n}$ is positive definite and $\boldsymbol{b} \in \mathbb{R}^n$.

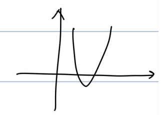
Let $\boldsymbol{d}_k$ be a descent direction at the kth iterate. Suppose that we search along this direction from $\boldsymbol{x}^k$ for a new iterate, and the line search are exact. Please find the stepsize α. This can be achieved exactly solving the following one-dimensional minimization problem

$$\min_{\alpha>0} \quad f(\boldsymbol{x}_k + \alpha\boldsymbol{d}_k). \tag{6}$$

3. $f(x_k+\alpha d_k) = \frac{1}{2}(x_k+\alpha d_k)^T Q(x_k+\alpha d_k) - b^T(x_k+\alpha d_k)$

$= \frac{1}{2}x_k^T Q x_k + x_k^T Q d_k\alpha + \frac{1}{2}d_k^T Q d_k\alpha^2 - b^T x_k - b^T d_k\alpha$

$= \frac{1}{2}d_k^T Q d_k\alpha^2 + (x_k^T Q - b^T)d_k\alpha + \frac{1}{2}x_k^T Q x_k - b^T x_k$



$\because$ Q is positive definite

$\therefore$ $d_k^T Q d_k > 0$

$= \frac{1}{2}d_k^T Q d_k\left[\alpha^2 - \frac{2(x_k^T Q - b^T)d_k}{d_k^T Q d_k}\alpha\right] + \frac{1}{2}x_k^T Q x_k - b^T x_k$

$\alpha = \frac{(x_k^T Q - b^T)d_k}{d_k^T Q d_k}$

4 The conjugate gradient algorithm

[20pts] Let $A \in \mathbb{R}^{n\times n}$ be a positive definite matrix. Show that if the directions $\boldsymbol{d}_0, \ldots, \boldsymbol{d}_k \in \mathbb{R}^n$, $k \le n-1$, are A-conjugate, then they are linearly independent. (Hint: We say that a set of nonzero vectors $\boldsymbol{d}_1, \ldots, \boldsymbol{d}_m \in \mathbb{R}^n$ are A-conjugate if $\boldsymbol{d}_i^T A\boldsymbol{d}_i = 0, \quad \forall i, j, \quad i \ne j$.)

4. $d_0, \cdots d_{n-1} \in \mathbb{R}^n$ are A-conjugate

Suppose $d_0, \cdots d_{n-1}$ are linear dependent

$\exists \alpha_0, \cdots \alpha_{n-1}$ are not all 0 such that $\alpha_0 d_0 + \cdots + \alpha_{n-1} d_{n-1} = 0$

We define $\alpha_i \neq 0$

$$d_i^T A [\alpha_0 d_0 + \cdots + \alpha_{n-1} d_{n-1}] = 0$$

$$\alpha_0 d_i^T A d_0 + \cdots + \alpha_i d_i^T A d_i + \cdots + \alpha_{n-1} d_i^T A d_{n-1} = 0$$

$$\alpha_i d_i^T A d_i = 0$$

$\because$ A is positive definite

$\therefore d_i^T A d_i > 0$

$\therefore \alpha_i = 0$ contradiction

$\therefore \alpha_0 = \alpha_1 = \cdots = \alpha_{n-1} = 0$

$\therefore d_1, \cdots d_n$ are linear independent

5 Trust region subproblems

Consider the trust region subproblem (1), and H_k is positive definite. Let θ_k denote the angle between $\boldsymbol{d}_k$ and $-\nabla f(\boldsymbol{x}_k)$, defined by

$$\cos\theta_k = \frac{-\nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k}{\|\nabla f(\boldsymbol{x}_k)\| \|\boldsymbol{d}_k\|}.$$

Show that

(i) [10pts] For sufficiently large Δ_k, the trust region subproblem (1) will be solved by the Newton step.

(ii) [10pts] When Δ_k approaches 0, the angle $\theta_k \to 0$.

5. $\min_{d \in \mathbb{R}^n} m_k(d) = f(x_k) + \nabla f(x_k)^T d_k + \frac{1}{2} d_k^T H_k d_k$

s.t. $\|d\| \leq \Delta_k$

(i) As H_k is positive definite

$\therefore \min_{d \in \mathbb{R}^n} m_k(d) \Leftrightarrow \nabla m_k(d) = 0$

$\nabla m_k(d) = \nabla f(x_k) + H_k d_k = 0$

$d_k = -H_k^{-1} \nabla f(x_k)$

When we add the constraint condition $\|d\| \leq \Delta_k$

then $\forall \Delta_k \geq \|H_k^{-1} \nabla f(x_k)\|$

$\|d_k\| = \|H_k^{-1} \nabla f(x_k)\| \leq \Delta_k$

$\therefore \min_{d \in \mathbb{R}^n} m_k(d) \Leftrightarrow \min_{d \in \mathbb{R}^n} m_k(d)$

s.t. $\|d\| \leq \Delta_k$

$\Delta_k \geq \|H_k^{-1} \nabla f(x_k)\|$ (sufficiently large Δ_k)

where $d_k = -H_k^{-1} \nabla f(x_k)$

(ii) $\min_{d \in \mathbb{R}^n} m_k(d) \Leftrightarrow \min_{d \in \mathbb{R}^n} \frac{\nabla f(x_k)^T d_k + \frac{1}{2} d_k^T H_k d_k}{\|\nabla f(x_k)\| \|d_k\|}$ we consider d_k such that $\nabla f(x)^T d_k < 0$

$\lim_{\Delta_k \to 0} \frac{\nabla f(x_k)^T d_k + \frac{1}{2} d_k^T H_k d_k}{\|\nabla f(x_k)\| \|d_k\|}$ When $\Delta_k \to 0$, in order to minimize

$= \lim_{\|d_k\| \to 0} \frac{-\|\nabla f(x_k)^T d_k\|}{\|\nabla f(x_k)\| \|d_k\|}$ $m_k(d)$, we consider $\|d_k\| = \Delta_k$

$\geq \lim_{\|d_k\| \to 0} \frac{-\|\nabla f(x_k)\|^T \|d_k\|}{\|\nabla f(x_k)\| \|d_k\|}$ ("=" when $d_k = -\frac{\Delta_k}{\|\nabla f(x_k)\|} \nabla f(x_k)$)

$= -1$

$\therefore d_k \to -\frac{\Delta_k}{\|\nabla f(x_k)\|} \nabla f(x_k)$ $\Delta_k \to 0$

$\cos\theta_k = \frac{-\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\| \|d_k\|} \to 1$ $\Delta_k \to 0$

$\because \theta_k \in [0, 2\pi)$

$\therefore \theta_k \to 0$ $\Delta_k \to 0$