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

Solution:

- Step1: $\boldsymbol{d}_k^s$ should lies in the negative direction of gradient $\nabla f(\boldsymbol{x}_k)$ to reach the minimum. Plus, consider the bound of its length, we get:
$$\boldsymbol{d}_k^s = -\Delta_k \frac{\nabla f(\boldsymbol{x}_k)}{\|\nabla f(\boldsymbol{x}_k)\|}.$$
- Step2: To obtain τ_k explicitly, we consider the cases of $\nabla f(\boldsymbol{x}_k)^T H_k \nabla f(\boldsymbol{x}_k) \le 0$ and $\nabla f(\boldsymbol{x}_k)^T H_k \nabla f(\boldsymbol{x}_k) > 0$ separately.

- When $\nabla f(\boldsymbol{x}_k)^T H_k \nabla f(\boldsymbol{x}_k) \leq 0$,

$$m_k(\tau \boldsymbol{d}_k^s) = f(\boldsymbol{x}_k) - \tau_k \Delta_k \|\nabla f(\boldsymbol{x}_k)\| + \tau_k^2 \frac{\Delta_k^2}{2\|\nabla f(\boldsymbol{x}_k)\|^2} \nabla f(\boldsymbol{x}_k)^T H_k \nabla f(\boldsymbol{x}_k),$$

  where the subproblem decreases monotonically with τ whenever $\nabla f(\boldsymbol{x}_k) \neq 0$. Thus $\tau_k = 1$ in this case.

- When $\nabla f(\boldsymbol{x}_k)^T H_k \nabla f(\boldsymbol{x}_k) \leq 0$, the optimal τ_k should be:

$$\tau_k = \min\{\frac{\Delta_k \|\nabla f(\boldsymbol{x}_k)\|}{\frac{\Delta_k^2}{\|\nabla f(\boldsymbol{x}_k)\|^2} \nabla f(\boldsymbol{x}_k)^T H_k \nabla f(\boldsymbol{x}_k)}, 1\}.$$

  where we use the fact that $\tau^\star = -\frac{b}{2a}$ if no constraints are given.

- In conclusion, we have

$$\boldsymbol{d}_k^c = -\tau_k \Delta_k \frac{\nabla f(\boldsymbol{x}_k)}{\|\nabla f(\boldsymbol{x}_k)\|},$$

  where

$$\tau_k = \begin{cases} 1 & \text{if } \nabla f(\boldsymbol{x}_k)^T H_k \nabla f(\boldsymbol{x}_k) \leq 0 \\ \min\{\frac{\|\nabla f(\boldsymbol{x}_k)\|^3}{\Delta_k \nabla f(\boldsymbol{x}_k)^T H_k \nabla f(\boldsymbol{x}_k)}, 1\} & \text{otherwise.} \end{cases}$$

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.)

Solution:

$$\begin{aligned} \rho_k &= \frac{f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + \boldsymbol{d}_k)}{m_k(\boldsymbol{0}) - m_k(\boldsymbol{d}_k)} \\ &= \frac{-\nabla f(\boldsymbol{x}_k)\boldsymbol{d}_k - \frac{1}{2}\boldsymbol{d}_k^T H_k' \boldsymbol{d}_k}{-\nabla f(\boldsymbol{x}_k)\boldsymbol{d}_k - \frac{1}{2}\boldsymbol{x}_k^T H_k \boldsymbol{x}_k}, \end{aligned}$$

where H_k' is the Hessian matrix at some point $\boldsymbol{x'} \in (\boldsymbol{x}_k, \boldsymbol{x}_k + \boldsymbol{d}_k)$. Since $\boldsymbol{d}_k \leq \Delta_k \to 0$, the second order terms $\boldsymbol{d}_k^T H_k' \boldsymbol{d}_k$ and $\boldsymbol{d}_k^T H_k \boldsymbol{d}_k$ can be omitted comparing to first order term $\nabla f(\boldsymbol{x}_k)$. Thus we get:

$$\begin{aligned} \lim_{\Delta_k \to 0} \rho_k &= \lim_{\Delta_k \to 0} \frac{-\nabla f(\boldsymbol{x}_k)\boldsymbol{d}_k - \frac{1}{2}\boldsymbol{d}_k^T H_k' \boldsymbol{d}_k}{-\nabla f(\boldsymbol{x}_k)\boldsymbol{d}_k - \frac{1}{2}\boldsymbol{x}_k^T H_k \boldsymbol{x}_k} \\ &= \lim_{\Delta_k \to 0} \frac{-\nabla f(\boldsymbol{x}_k)\boldsymbol{d}_k}{-\nabla f(\boldsymbol{x}_k)\boldsymbol{d}_k} \\ &= 1, \end{aligned}$$

when $\Delta_k \to 0$.

3 Exact line search

[20pts] Consider minimizing the following quadratic function

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

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

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

Solution:

First we compute the first and second derivatives with respect to α:

$$\begin{aligned}
f(\boldsymbol{x}_k + \alpha\boldsymbol{d}_k) &= \frac{1}{2}(\boldsymbol{x}_k + \alpha\boldsymbol{d}_k)^T Q(\boldsymbol{x}_k + \alpha\boldsymbol{d}_k) - \boldsymbol{b}^T(\boldsymbol{x}_k + \alpha\boldsymbol{d}_k),\\
\frac{\partial f}{\partial \alpha} &= \boldsymbol{x}_k^T Q\boldsymbol{d}_k + \alpha\boldsymbol{d}_k^T Q\boldsymbol{d}_k - \boldsymbol{b}^T\boldsymbol{d}_k,\\
\frac{\partial^2 f}{\partial \alpha^2} &= \boldsymbol{d}_k^T Q\boldsymbol{d}_k \geq 0.
\end{aligned}$$

where the last inequality comes from the positive definiteness of Q. Thus we just need to find the α that satisfies the first order necessary condition which gives:

$$\frac{\partial f}{\partial \alpha} = 0 \quad \Rightarrow \quad \alpha^\star = \frac{\boldsymbol{b}^T\boldsymbol{d}_k - \boldsymbol{x}_k^T Q\boldsymbol{d}_k}{\boldsymbol{d}_k^T Q\boldsymbol{d}_k}.$$

4 The conjugate gradient algorithm

[20pts] Let $A \in \mathbb{R}^{n\times n}$ be a positive definite matrix. Show that if the directions $\boldsymbol{d}_0, \ldots, \boldsymbol{d}_k \in \mathbb{R}^n$, $k \leq n-1$, 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 $\boldsymbol{d}_i^T A\boldsymbol{d}_i = 0, \;\; \forall i, j, \;\; i \neq j$.)

Solution:

Assume that $\boldsymbol{d}_0, \ldots, \boldsymbol{d}_k \in \mathbb{R}^n$ are not linearly independent. Accordingly, we can express $\boldsymbol{d}_m = \sum_{i\neq m} \alpha_i \boldsymbol{d}_i$. Then we choose $\boldsymbol{d}_j$ such that $j \neq m$ and $\alpha_j \neq 0$ (this always works because $\boldsymbol{d}_m = \vec{0}$ if not). Then we have:

$$\begin{aligned}
\boldsymbol{d}_m^T A\boldsymbol{d}_j &= (\sum_{i\neq m} \alpha_i \boldsymbol{d}_i)^T A\boldsymbol{d}_j\\
&= \alpha_j \boldsymbol{d}_j^T A\boldsymbol{d}_j\\
&\neq 0,
\end{aligned}$$

which violates the fact that $\boldsymbol{d}_0, \ldots, \boldsymbol{d}_k$ are A-conjugate. The assumption is unreasonable and $\boldsymbol{d}_0, \ldots, \boldsymbol{d}_k$ are linearly independent in this case.

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 $\theta_k \to 0$.

Solution:
$\boldsymbol{d}_k$ is a global solution of the TR subproblem if and only if for Some $\lambda \geq 0$ we have

$$\begin{aligned}
(H_k + \lambda I)\boldsymbol{d}_k &= -\nabla f(\boldsymbol{x}_k), \\
(H_k + \lambda I) &\geq 0, \\
\lambda &\geq 0, \\
\lambda(\Delta_k - \|\boldsymbol{d}_k\|) &= 0.
\end{aligned}$$

(i) For sufficiently large Δ_k, $\boldsymbol{x}_k + \boldsymbol{d}_k$ will locate within the trust region so that the constraint on the step can be ignored. Accordingly, $\lambda = 0$ and the subproblem is equivalent to finding the Newton direction:

$$(H_k + 0I)\boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k) \quad \Rightarrow \quad \boldsymbol{d}_k = -H_k^{-1}\nabla f(\boldsymbol{x}_k),$$

which is basically solved by the Newton step.

(ii) When $\Delta_k \to 0$, then $\boldsymbol{d}_k \to 0$ and $\lambda \to +\infty$. This conclusion can be seen from meeting the condition $(H_k + \lambda I)\boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k)$. In such case, $(H_k + \lambda I)\boldsymbol{d}_k \to \lambda I$, $\lambda \to +\infty$. Thus we have:

$$\begin{aligned}
\boldsymbol{d}_k &= -(H_k + \lambda I)^{-1}\nabla f(\boldsymbol{x}_k) \\
&\approx -\frac{1}{\lambda}\nabla f(\boldsymbol{x}_k).
\end{aligned}$$

Accordingly, the angle $\theta_k = \arccos \frac{-\nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k}{\|\nabla f(\boldsymbol{x}_k)\| \|\boldsymbol{d}_k\|} \to \arccos \frac{\frac{1}{\lambda}\|\nabla f(\boldsymbol{x}_k)\|^2}{\frac{1}{\lambda}\|\nabla f(\boldsymbol{x}_k)\|^2} = \arccos 1 = 0$. So $\theta_k \to 0$.