The Chinese University of Hong Kong, Shenzhen
School of Data Science

DDA 6010/CIE 6010 · Assignment 2
Due: 23:59, October 7

Instructions:

- Assignment problems must be carefully and clearly answered to receive full credit. Complete sentences that establish a clear logical progression are highly recommended.
- You must submit your assignment in Blackboard. Please upload a pdf file with codes. The file name should be in the format **last name-first name-student ID-hw1**, e.g. **Zhang-San-123456789-hw1**.
- Please make your solutions legible and write your solutions in English. You are strongly encouraged to type your solutions in LaTeX/Markdown or others.
- Late submission will **not** be graded.
- Each student **must not copy** assignment solutions from another student or from any other source.
- For those questions that ask you to write MATLAB/Python/other codes to solve the problem. Please attach your code in the **pdf file**. You also need to clearly state (write or type) the optimal solution and the optimal value you obtained. However, you do not need to attach the outputs in the command window of MATLAB/Python/others.

Problem 1 Condition Number (10 pts).
Consider the minimization problem

$$\min \left\{ x^T Q x : x \in \mathbb{R}^2 \right\},$$

where Q is a symmetric positive definite 2×2 matrix. Suppose we use the diagonal scaling matrix

$$D = \begin{pmatrix} Q_{11}^{-1} & 0 \\ 0 & Q_{22}^{-1} \end{pmatrix},$$

where $Q_{ij} \in \mathbb{R}$ denotes the (i,j)-element of Q. Show that the above scaling matrix improves the condition number of Q in the sense that

$$\kappa\left(D^{1/2} Q D^{1/2}\right) \le \kappa(Q).$$

Problem 2 Smooth Function (20 pts).

(a) Let $g(x) = x^T A x + 2b^T x + c$, where A is a symmetric $n \times n$ matrix, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Show that the smallest Lipschitz constant of ∇f is $2\|A\|$.

(b) Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is an L-smooth function. Prove that

$$\|\nabla f(x)\|^2 \leq 2L\left(f(x) - \bar{f}\right),$$

where $\bar{f}$ satisfies $f(x) \geq \bar{f}$ for any $x \in \mathbb{R}^d$.

Problem 3 Gradient Descent Method with Error (35 pts).

Consider the objective function $f \in C_{1,1}^L(\mathbb{R}^n)$ and the sequence $\{x_k\}_k$ generated by

$$x_{k+1} = x_k - \alpha_k g_k,$$

where $g_k = \nabla f(x_k) + e_k$, the term $e_k \in \mathbb{R}^n$ is the error from the computation of gradient, and the sequence of stepsizes $\{\alpha_k\}$ satisfies

$$0 < \alpha_k \leq 1/L, \quad \sum_{k=0}^{\infty} \alpha_k = \infty, \quad \text{and} \quad \sum_{k=0}^{\infty} \alpha_k^2 < \infty.$$

Assume that $\inf_{x \in \mathbb{R}^n} f(x) > -\infty$ and the error terms satisfy $\langle e_k, \nabla f(x_k)\rangle = 0$ and

$$\|e_k\|^2 \leq c_1[f(x_k) - f^*] + c_2, \ \forall k \in \mathbb{N},$$

where $c_1 \geq 0$ and $c_2 \geq 0$ are some constants and $f^* := \inf_x f(x)$. Prove:

(a) The approximate descent property of $f(x_k)$:

$$f(x_{k+1}) - f^* \leq \left(1 + \frac{\alpha_k^2 L c_1}{2}\right)[f(x_k) - f^*] - \frac{\alpha_k}{2}\|\nabla f(x_k)\|^2 + \frac{\alpha_k^2 L c_2}{2}, \ \forall k \in \mathbb{N}.$$

(b) The sequence $\{f(x_k)\}_{k\geq 0}$ converges to some finite value and it holds $\sum_{k=0}^{\infty} \alpha_k \|\nabla f(x_k)\|^2 < \infty$.

(c) The term $\|x_{k+1} - x_k\|^2$ satisfies

$$\|x_{k+1} - x_k\|^2 \leq \alpha_k^2 \|\nabla f(x_k)\|^2 + \alpha_k^2 c_1 [f(x_k) - f^*] + \alpha_k^2 c_2.$$

(d) Read Section 2 and Section 3.1 in [1] and show $\lim_{k\to\infty}\|\nabla f(x_k)\| = 0$. (Hint: You can refer to the proof of [1, Theorem 2.1] (Part I in Appendix A).)

Problem 4 Second-Order Method (20pts).
Let f be a twice continuously differentiable function satisfying $LI \succeq \nabla^2 f(x) \succeq mI$ for some $L > m > 0$ and let x^* be the unique minimizer of f over $\mathbb{R}^n$.

(a) Show that

$$f(x) - f(x^*) \geq \frac{m}{2}\|x - x^*\|^2$$

for any $x \in \mathbb{R}^n$.

(b) Let $\{x_k\}_{k\geq 0}$ be the sequence generated by damped Newton's method with constant stepsize $t_k = \frac{m}{L}$. Show that

$$f(x_k) - f(x_{k+1}) \geq \frac{m}{2L}\nabla f(x_k)^T \left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k),$$

and $x_k \to x^*$ as $k \to \infty$.

Problem 5 Implementation of Gradient Method (15 pts).
Consider the quadratic minimization problem

$$\min \left\{ x^T A x : x \in \mathbb{R}^5 \right\},$$

where A is the 5×5 Hilbert matrix defined by

$$A_{i,j} = \frac{1}{i+j-1}, \quad i, j = 1, 2, 3, 4, 5.$$

The matrix can be constructed via the MATLAB command *A=hilb(5)*. Run the following methods and compare the number of iterations required by each of the methods when the initial vector is $x_0 = (1, 2, 3, 4, 5)^T$ to obtain a solution x with $\|\nabla f(x)\| \leq 10^{-4}$:

(a) gradient method with backtracking stepsize rule and parameters $\alpha = 0.5, \beta = 0.5$, and $s = 1$;

(b) gradient method with backtracking stepsize rule and parameters $\alpha = 0.1, \beta = 0.5$, and $s = 1$;

(c) gradient method with exact line search;

(d) diagonally scaled gradient method with diagonal elements $D_{ii} = \frac{1}{A_{ii}}, i = 1, 2, 3, 4, 5$ and exact line search;

(e) diagonally scaled gradient method with diagonal elements $D_{ii} = \frac{1}{A_{ii}}, i = 1, 2, 3, 4, 5$ and backtracking line search with parameters $\alpha = 0.1, \beta = 0.5$, $s = 1$.

References

[1] Xiao Li and Andre Milzarek. A unified convergence theorem for stochastic optimization methods. *Advances in Neural Information Processing Systems*, 35:33107–33119, 2022.