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

DDA6010/CIE6010 · Assignment 1

Due: 23:59, September 20

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 Coerciveness and Optimality (30pts).

(a) Determine whether the following functions are coercive and find their global minimizers, if any.

(1) $f(x,y,z) = x^2 - 2xy + y^2 + z^2$

(2) $f(x,y) = x^4 + y^4 - 4xy$

(b) Calculate all stationary points of the function $f(x,y) = x^4 - xy^2 - \frac{1}{2}x^2 + y^2$ and investigate whether the stationary points are local maximizer, local minimizer, or saddle points.

Solution:

(a) (1) $f(x,y,z) = (x,y,z)\,\boldsymbol{A}\begin{pmatrix} x \\ y \\ z \end{pmatrix}$, where $\boldsymbol{A} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

The eigenvalues of $\boldsymbol{A}$ are $\lambda_1 = 0, \lambda_2 = 1, \lambda_3 = 2$, which implies that $\boldsymbol{A} \succeq 0$. Thus, $f(x,y,z)$ is not coercive.

Solving $\nabla f(x^*,y^*,z^*) = \mathbf{0}$, which is $2\boldsymbol{A}\begin{pmatrix} x^* \\ y^* \\ z^* \end{pmatrix} = \mathbf{0}$, we have $z^* = 0, x^* = y^* \in \mathbb{R}$ and $f(x^*,y^*,z^*) = 0$ which is the global minimal value, considering the positive semi-definite matrix $\boldsymbol{A}$.

Thus, the global minimizer is $(x^*,y^*,z^*) = (a,a,0), \quad \forall a \in \mathbb{R}$.

(2) Note that $x^4+y^4 \geq \frac{1}{2}(x^2+y^2)^2$ and $xy \leq \frac{1}{2}(x^2+y^2)$. It implies that

$$f(x,y) = x^4+y^4-4xy \geq \frac{1}{2}(x^2+y^2)^2 - 2(x^2+y^2) = \frac{1}{2}(x^2+y^2-2)^2 - 2$$

When $\|(x,y)\| \to \infty$, $x^2+y^2 \to \infty$. Then, $(x^2+y^2-2)^2 \to \infty$.
Thus, $\lim_{\|(x,y)\|\to\infty} f(x,y) = \infty$, which means that $f(x,y)$ is coercive.

$$\nabla f(x,y) = \begin{pmatrix} 4x^3-4y \\ 4y^3-4x \end{pmatrix}, \quad \nabla^2 f(x,y) = \begin{pmatrix} 12x^2 & -4 \\ -4 & 12y^2 \end{pmatrix}$$

Solving $\nabla f(x,y) = \mathbf{0}$, we have $(x,y) = (0,0)$, $(1,1)$ or $(-1,-1)$. Besides,

$$\nabla^2 f(0,0) = \begin{pmatrix} 0 & -4 \\ -4 & 0 \end{pmatrix} \text{ is indefinite }, \ \nabla^2 f(1,1) = \begin{pmatrix} 12 & -4 \\ -4 & 12 \end{pmatrix} \succ 0,$$

and

$$\nabla^2 f(1,1) = \begin{pmatrix} 12 & -4 \\ -4 & 12 \end{pmatrix} \succ 0.$$

Considering that $f(1,1) = f(-1,-1) = -2$, we have the global minimizers $(1,1)$ and $(-1,-1)$.

(b) The gradient and Hessian of f are given by

$$\nabla f(x,y) = \begin{pmatrix} 4x^3-y^2-x \\ -2xy+2y \end{pmatrix}, \ \nabla^2 f(x,y) = \begin{pmatrix} 12x^2-1 & -2y \\ -2y & -2x+2 \end{pmatrix}$$

Hence, all the stationary points $\mathbf{x} := (x,y)^\intercal$ are

$$\mathbf{x}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ \mathbf{x}_2 = \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix}, \ \mathbf{x}_3 = \begin{pmatrix} -\frac{1}{2} \\ 0 \end{pmatrix} \mathbf{x}_4 = \begin{pmatrix} 1 \\ \sqrt{3} \end{pmatrix}, \ \mathbf{x}_5 = \begin{pmatrix} 1 \\ -\sqrt{3} \end{pmatrix}.$$

The Hessians of f at $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are

$$\nabla^2 f(\mathbf{x}_1) = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}, \ \nabla^2 f(\mathbf{x}_2) = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \ \nabla^2 f(\mathbf{x}_3) = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}.$$

The Hessian $\nabla^2 f(\mathbf{x}_1)$ is indefinite and thus, $\mathbf{x}_1$ is a saddle point. The Hessian $\nabla^2 f(\mathbf{x}_2)$ and $\nabla^2 f(\mathbf{x}_3)$ are positive definite, and hence, $\mathbf{x}_2$ and $\mathbf{x}_3$ are strict local minimizer. The Hessians of f at $\mathbf{x}_4$ and $\mathbf{x}_5$ are given by

$$\nabla^2 f(\mathbf{x}_4) = \begin{pmatrix} 11 & -2\sqrt{3} \\ -2\sqrt{3} & 0 \end{pmatrix}, \ \nabla^2 f(\mathbf{x}_5) = \begin{pmatrix} 11 & 2\sqrt{3} \\ 2\sqrt{3} & 0 \end{pmatrix}.$$

It is easy to see that $\det(\nabla^2 f(\mathbf{x}_4))$ and $\det(\nabla^2 f(\mathbf{x}_5))$ are both negative while $\mathrm{tr}(\nabla^2 f(\mathbf{x}_4))$ and $\mathrm{tr}(\nabla^2 f(\mathbf{x}_5))$ are positive. Consequently, the Hessians $\nabla^2 f(\mathbf{x}_4)$ and $\nabla^2 f(\mathbf{x}_5)$ must be indefinite and $\mathbf{x}_4$ and $\mathbf{x}_5$ are saddle points.

Problem 2 Quadratic Optimization Problems (20pts).

(a) Consider the following regularized quadratic problem:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n, y\in\mathbb{R}} g(\boldsymbol{x}, y) := \frac{\lambda}{2}\|\boldsymbol{x}\|_2^2 + \sum_{i=1}^{m} \max\left\{0, 1 - b_i\left(\boldsymbol{a}_i^\intercal \boldsymbol{x} + y\right)\right\}, \tag{1}$$

where $\lambda > 0$, $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m) \in \mathbb{R}^{n\times m}$, and $b_i \in \{-1, 1\}$ for any i. Verify that Problem (1) always possesses a global optimal solution.

(b) Consider the unconstrained quadratic optimization problem:

$$\min_{\alpha\in\mathbb{R}} f(\boldsymbol{x} - \alpha\nabla f(\boldsymbol{x})),\ f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\intercal \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\intercal \boldsymbol{x} + c,$$

where $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is a symmetric and positive definite matrix, $\boldsymbol{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Derive the optimal solution for $\alpha \in \mathbb{R}$.

Solution:

(a) Firstly, suppose all the b_i's have the same value. We have the following observation:

$$g(\boldsymbol{x}, y) \geq 0,\ g(\boldsymbol{x}, y) = 0 \ \Leftrightarrow \frac{\lambda}{2}\|\boldsymbol{x}\|_2^2 = 0, \text{ and } 1 - b_i\left(\boldsymbol{a}_i^\intercal \boldsymbol{x} + y\right) \leq 0,\ \forall i. \tag{2}$$

Condition (2) implies $\boldsymbol{x} = \boldsymbol{0}$ and

$$\begin{cases} 1 + y \leq 0, & b_i = -1,\ \forall i \\ 1 - y \leq 0, & b_i = 1,\ \forall i \end{cases}.$$

In this case, Problem (1) has global solutions given the set

$$\left\{(\boldsymbol{x}, y) : \{\boldsymbol{0}\} \times (-\infty, -1] \text{ or } \{\boldsymbol{0}\} \times [1, \infty)\right\}.$$

Secondly, suppose not all b_i's have the same value. We can verify that $g(\boldsymbol{x}, y)$ is a coercive function in this case. Then, Problem (1) possesses at least one global optimal solution.

(b) Denote $\boldsymbol{d} = -\nabla f(\boldsymbol{x})$. Then,

$$\begin{aligned}
\phi(\alpha) := f(\boldsymbol{x} - \alpha\nabla f(\boldsymbol{x})) &= f\left(\boldsymbol{x} + \alpha\boldsymbol{d}\right) \\
&= \frac{1}{2}\left(\boldsymbol{x} + \alpha\boldsymbol{d}\right)^\intercal \boldsymbol{A}\left(\boldsymbol{x} + \alpha\boldsymbol{d}\right) + \boldsymbol{b}^\intercal\left(\boldsymbol{x} + \alpha\boldsymbol{d}\right) + c \\
&= \frac{1}{2}\alpha^2\boldsymbol{d}^\intercal\boldsymbol{A}\boldsymbol{d} + \alpha\boldsymbol{x}^\intercal\boldsymbol{A}\boldsymbol{d} + \alpha\boldsymbol{b}^\intercal\boldsymbol{d} + \frac{1}{2}\boldsymbol{x}^\intercal\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\intercal\boldsymbol{x} + c \\
&= \frac{1}{2}\alpha^2\boldsymbol{d}^\intercal\boldsymbol{A}\boldsymbol{d} + \alpha\nabla f(\boldsymbol{x})^\intercal\boldsymbol{d} + f(x).
\end{aligned}$$

Since $\phi'(\alpha) = \alpha\boldsymbol{d}^\intercal\boldsymbol{A}\boldsymbol{d} + \nabla f(\boldsymbol{x})^\intercal\boldsymbol{d}$, setting $\phi'(\alpha) = 0$ leads to

$$\alpha = -\frac{\nabla f(\boldsymbol{x})^\intercal\boldsymbol{d}}{\boldsymbol{d}^\intercal\boldsymbol{A}\boldsymbol{d}} = \frac{\|\nabla f(\boldsymbol{x})\|^2}{\nabla f(\boldsymbol{x})^\intercal\boldsymbol{A}\nabla f(\boldsymbol{x})}.$$

Problem 3 Optimality Condition (30pts).
Consider the function

$$f(x_1, x_2) = \exp\left(x_1^2 + x_2^2(1 - x_1)^3\right)$$

Prove that

(a) The point $z = (0, 0)$ is the unique stationary point of the function f.

(b) The point z is also a local minimal of f.

(c) The function f is bounded from below, but z is not the global minimal of f and f does not have any global minimal.

Solution:

(a) Note that

$$\nabla f(x_1, x_2) = \begin{pmatrix} \left(2x_1 - 3x_2^2(1-x_1)^2\right) \exp\left(x_1^2 + x_2^2(1-x_1)^3\right) \\ 2x_2(1-x_1)^3 \exp\left(x_1^2 + x_2^2(1-x_1)^3\right) \end{pmatrix}.$$

Solving $\nabla f(x_1^*, x_2^*) = 0$, we have $2x_2^*(1-x_1^*)^3 = 0$ and $2x_1^* - 3(x_2^*)^2(1-x_1^*)^2 = 0$. If $x_2^* \neq 0$, we have $x_1^* = 1$ by the first equation, which conflicts with the second equation. Hence, the point $z = (0, 0)$ is the unique stationary point of the function f.

(b) Note that $\nabla^2 f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \succ 0$ By the second order optimality condition, $z = (0, 0)$ is the local minimal of f.

(c) It is obvious that $f(x_1, x_2) > 0$ as the exponential function. Thus, f is bounded from below. With the fact that $f(4, 1) = \exp(-11) < 1 = f(0, 0)$, z is not the global minimal of f. Besides, since z is the only one satisfied the first-order optimality condition for f defined on $\mathbb{R}^2$, we get our conclusion that f does not have any global minimal.

Problem 4 Coding: Nonlinear Least Square (20pts).
Generate 50 points $\{(x_i, y_i), i = 1, 2, \cdots, 50\}$ through the following code. (Type them by yourself to avoid the error caused by indentation.)

```
randn('seed',314);
x=linspace(0,1,50);
y=2*x.^2-3*x+1+0.05*randn(size(x));
```

Or you could generate your own random sample through the same quadratic rule with your favorite language. Find the quadratic function $y_i = ax_i^2 + bx_i + c$ that best fits the points in the least squares sense.

(a) Formulate the above least square problem into the following form:

$$\min_{\boldsymbol{z} \in \mathbb{R}^3} \|\boldsymbol{X}\boldsymbol{z} - \boldsymbol{y}\|^2,$$

where $\boldsymbol{z} = (a, b, c)^\intercal \in \mathbb{R}^3$.

(b) What are analytical form of a, b, c in terms of $\boldsymbol{X}$ and $\boldsymbol{y}$ that solves the least squares solution?

(c) Plot the points along with the derived quadratic function. The resulting plot should look like the one in Figure 1. You can apply Matlab/Python or any other programs to generate the picture.

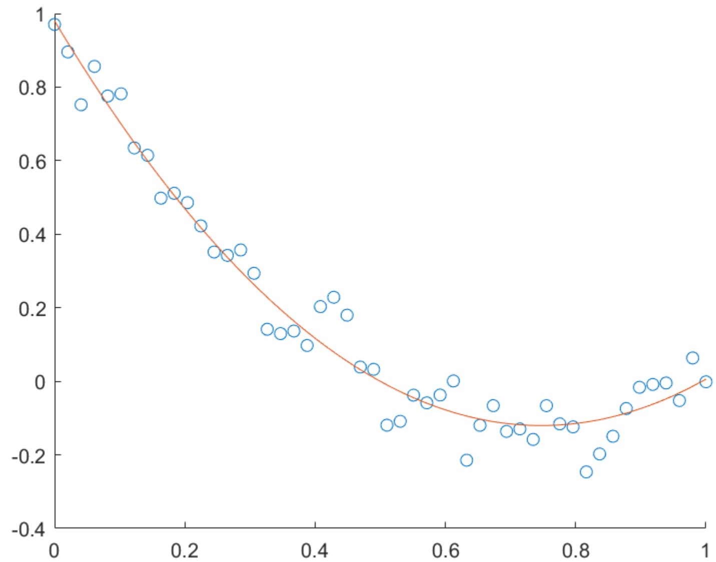

Figure 1: 50 points and their best quadratic least squares fit.

Solution:

(a) We only need to minimize $\sum_{i=1}^{50}\left(ax_i^2+bx_i+c-y_i\right)^2$, which is

$$\min f(\boldsymbol{z}) := \|\boldsymbol{X}\boldsymbol{z}-\boldsymbol{y}\|_2^2, \text{ where } \boldsymbol{X} = \begin{pmatrix} x_1^2 & x_1 & 1 \\ \vdots & \vdots & \dots \\ x_{50}^2 & x_{50} & 1 \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_{50} \end{pmatrix}, \quad \boldsymbol{z} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

(b) To solve the least square problem, we have

$$\nabla f(\boldsymbol{z}) = 2(\boldsymbol{X}^\intercal\boldsymbol{X}\boldsymbol{z}-\boldsymbol{X}^\intercal\boldsymbol{y}) \text{ and } \nabla^2 f(\boldsymbol{z}) = 2\boldsymbol{X}^\intercal\boldsymbol{X} \succ 0$$

Thus, solving $\nabla f(\boldsymbol{z}^*)=0$, it implies that $\begin{pmatrix} a \\ b \\ c \end{pmatrix} = (\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\boldsymbol{y}$.

(c) We get the solution through the following code.

```
A = [x.^2;x;ones(size(x))]';
coeff = inv(A'*A)*A'*y';
z = linspace(0,1,1000);
y1 = coeff(1)*z.^2 + coeff(2)*z + coeff(3);
scatter(x,y)
hold on
plot(z,y1)
```