

# Solution to Homework 1

Total 40 points

**Problem 1.** (10 points) Judge the properties of the following sets (openness, closeness, boundedness, compactness) and give their interiors, closures, and boundaries:

- (a)  $\mathcal{C}_1 = \emptyset$ .
- (b)  $\mathcal{C}_2 = \mathbb{R}^n$ .
- (c)  $\mathcal{C}_3 = \{(x, y)^\top | x \geq 0, y > 0\}$ .
- (d)  $\mathcal{C}_4 = \{k | k \in \mathbb{Z}\}$ .
- (e)  $\mathcal{C}_5 = \{k^{-1} | k \in \mathbb{Z}\}$ .

**Solution.**

- (a)  $\mathcal{C}_1$  is open, closed, bounded and compact.  $\mathcal{C}_1^\circ = \bar{\mathcal{C}}_1 = \partial\mathcal{C}_1 = \emptyset$ .
- (b)  $\mathcal{C}_2$  is open and closed.  $\mathcal{C}_2^\circ = \bar{\mathcal{C}}_2 = \mathbb{R}^n$  and  $\partial\mathcal{C}_2 = \emptyset$ .
- (c)  $\mathcal{C}_3^\circ = \{(x, y)^\top | x > 0, y > 0\}$ ,  $\bar{\mathcal{C}}_3 = \{(x, y)^\top | x \geq 0, y \geq 0\}$  and  $\partial\mathcal{C}_3 = \{(x, y)^\top | x = 0, y \geq 0\} \cup \{(x, y)^\top | x \geq 0, y = 0\}$ .
- (d)  $\mathcal{C}_4$  is closed.  $\mathcal{C}_4^\circ = \emptyset$  and  $\bar{\mathcal{C}}_4 = \partial\mathcal{C}_4 = \{k | k \in \mathbb{Z}\}$ .
- (e)  $\mathcal{C}_5$  is bounded.  $\mathcal{C}_5^\circ = \emptyset$  and  $\bar{\mathcal{C}}_5 = \partial\mathcal{C}_5 = \{k | k^{-1} \in \mathbb{Z}\} \cup \{0\}$ .

**Problem 2.** (12 points) For each of the following sequence, determine the rate of convergence and the rate constant:

- (a)  $x_k = 1 + 5 \times 10^{-2k}$ .
- (b)  $x_k = 2^{-2^k}$ .
- (c)  $x_k = 3^{-k^2}$ .
- (d)  $x_{k+1} = x_k/2 + 2/x_k$ ,  $x_1 = 4$ .

**Solution.**

- (a) As  $\lim_{k \rightarrow \infty} x_k = 1$ , and  $\lim_{k \rightarrow \infty} \frac{5 \times 10^{-2(k+1)}}{5 \times 10^{-2k}} = 0.01$ , thus  $r = 1$ ,  $C = 0.01$ .
- (b) As  $\lim_{k \rightarrow \infty} x_k = 0$ , and  $\lim_{k \rightarrow \infty} \frac{2^{-2^{k+1}}}{(2^{-2^k})^2} = 1$ , thus  $r = 2$ ,  $C = 1$ .
- (c) As  $\lim_{k \rightarrow \infty} x_k = 0$ , and  $\lim_{k \rightarrow \infty} \frac{3^{-(k+1)^2}}{3^{-k^2}} = 0$ , thus  $r = 1$ ,  $C = 0$ .
- (d)  $\lim_{k \rightarrow \infty} x_k = 2$ , and  $\lim_{k \rightarrow \infty} \frac{x_{k+1}-2}{(x_k-2)^2} = \lim_{k \rightarrow \infty} \frac{x_k/2+2/x_k-2}{(x_k-2)^2} = \lim_{k \rightarrow \infty} \frac{1}{2x_k} = \frac{1}{4}$ , thus  $r = 2$ ,  $C = \frac{1}{4}$ .

**Problem 3.** (18 points) Compute the **gradient** and the **Hessian** of the following functions (write in vector or matrix form, rather than entrywise), give details:

- (a)  $f(\mathbf{x}) = (\mathbf{a}^\top \mathbf{x})(\mathbf{b}^\top \mathbf{x})$ .
- (b)  $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2$ .
- (c)  $f(\mathbf{x}) = \log \sum_{i=1}^m \exp(\mathbf{a}_i^\top \mathbf{x} + b_i)$ .

**Solution.**

(a)

$$\begin{aligned} \nabla f(\mathbf{x}) &= \frac{\partial(\mathbf{a}^\top \mathbf{x})}{\partial \mathbf{x}} (\mathbf{b}^\top \mathbf{x}) + (\mathbf{a}^\top \mathbf{x}) \frac{\partial(\mathbf{b}^\top \mathbf{x})}{\partial \mathbf{x}} = \mathbf{a}(\mathbf{b}^\top \mathbf{x}) + (\mathbf{a}^\top \mathbf{x})\mathbf{b} = (\mathbf{ab}^\top + \mathbf{ba}^\top)\mathbf{x}. \\ \nabla^2 f(\mathbf{x}) &= (\mathbf{ab}^\top + \mathbf{ba}^\top). \end{aligned}$$

(b) As  $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2 = \frac{1}{2} (\mathbf{Ax} - \mathbf{b})^\top (\mathbf{Ax} - \mathbf{b})$ , we know

$$\begin{aligned} \nabla f(\mathbf{x}) &= \mathbf{A}^\top (\mathbf{Ax} - \mathbf{b}). \\ \nabla^2 f(\mathbf{x}) &= \mathbf{A}^\top \mathbf{A}. \end{aligned}$$

(c) Let  $g(\mathbf{y}) = \log \sum_{i=1}^m \exp(y_i)$ , then  $f(\mathbf{x}) = g(\mathbf{Ax} + \mathbf{b})$  where  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_m]^\top$  and  $\mathbf{b} = [b_1, \dots, b_m]^\top$ . On the other hand, let  $h(\mathbf{y}) = \sum_{i=1}^m \exp(y_i)$ , then  $g(\mathbf{y}) = \log(h(\mathbf{y}))$ . According to the chain rule (page 38 and 41 of lecture 1), we can get

$$\begin{aligned} \nabla g(\mathbf{y}) &= \frac{1}{\sum_{i=1}^m \exp(y_i)} \begin{bmatrix} \exp(y_1) \\ \vdots \\ \exp(y_m) \end{bmatrix}. \\ \nabla^2 g(\mathbf{y}) &= \frac{1}{\sum_{i=1}^m \exp(y_i)} \begin{bmatrix} \exp(y_1) & & \\ & \ddots & \\ & & \exp(y_m) \end{bmatrix} - \frac{1}{(\sum_{i=1}^m \exp(y_i))^2} \begin{bmatrix} \exp(y_1) \\ \vdots \\ \exp(y_m) \end{bmatrix} [\exp(y_1), \dots, \exp(y_m)]^\top \end{aligned}$$

Thus we have

$$\begin{aligned} \nabla f(\mathbf{x}) &= \mathbf{A}^\top \nabla g(\mathbf{Ax} + \mathbf{b}) = \frac{1}{\mathbf{1}^\top \mathbf{z}} \mathbf{A}^\top \mathbf{z} \\ \nabla^2 f(\mathbf{x}) &= \mathbf{A}^\top \nabla^2 g(\mathbf{Ax} + \mathbf{b}) \mathbf{A} = \mathbf{A}^\top \left( \frac{1}{\mathbf{1}^\top \mathbf{z}} \text{diag}(\mathbf{z}) - \frac{1}{(\mathbf{1}^\top \mathbf{z})^2} \mathbf{z} \mathbf{z}^\top \right) \mathbf{A}. \end{aligned}$$

where  $z_i = \exp(\mathbf{a}_i^\top \mathbf{x} + b_i)$  and  $\text{diag}(\mathbf{x})$  denotes a square matrix which has  $\mathbf{x}$  on the diagonal and zero everywhere else.