Continuous Optimization: Assignment 1

Due on April 30, 2024

Honglu Ma **Hiroyasu Akada** **Mathivathana Ayyappan**

Exercise 1

(a)

Claim: $lim_{k\to\infty} x^{(k)} = \frac{1}{30}$
Proof:
Let $\epsilon > 0$ be given. Choose $N = \max\{\frac{1}{90\epsilon}, 8\}$. Assume $n > N$. We have

$$n > N \Rightarrow n > 9 > \sqrt[3]{600} \Rightarrow n^3 > 600 \Rightarrow 5n^3 > 3000 \Rightarrow 10n^3 - 5n^3 > 3000 \Rightarrow 3000 + 5n^3 < 10n^3$$

and obviously

$$900n^4 > 150n^3 + 900n^4$$

To check the validity of the limit we need to show $\left|x^{(n)} - x^\star\right| < \epsilon$ where $x^\star = \frac{1}{30}$.

$$\begin{aligned}
\left|\frac{n^4 - 100}{5n^3 + 30k^4} - \frac{1}{30}\right| &= \left|\frac{30n^4 - 3000 - 5n^3 - 30k^4}{150n^3 + 900n^4}\right| \\
&= \left|\frac{-3000 - 5n^3}{150n^3 + 900n^4}\right| \\
&= \frac{3000 + 5n^3}{150n^3 + 900n^4} \\
&< \frac{10n^3}{900n^4} = \frac{1}{90n} && \text{(by the inequalities above)} \\
&< \frac{1}{90N} \\
&< \frac{1}{\frac{90}{90\epsilon}} = \epsilon
\end{aligned}$$

(b)

We have the following accumulation points:

$$x^{(8n+1)} = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}, \quad x^{(8n+2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad x^{(8n+3)} = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}, \quad x^{(8n+4)} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

$$x^{(8n+5)} = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix}, \quad x^{(8n+6)} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad x^{(8n+7)} = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix}, \quad x^{(8n)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad n \in \mathbb{N}$$

We can prove by showing that $x^{(k)}$ is not a cauchy sequence thus does not converges (Proposition A.5, Lecture Script) i.e. $\exists \epsilon > 0$ such that $\forall N \in \mathbb{N}$, $\exists n, m > N$ such that $\left\|x^{(n)} - x^{(m)}\right\| \geq \epsilon$.
Proof:
Let $\epsilon = 1$, for all $N \in \mathbb{N}$, choose $n = 8N$ and $m = 8N + 4$. We have

$$\left\|x^{(8N)} - x^{(8N+4)}\right\| = \left\|\begin{pmatrix} 1 \\ 0 \end{pmatrix} - \begin{pmatrix} -1 \\ 0 \end{pmatrix}\right\| = 2 \geq 1 = \epsilon$$

Exercise 2

(a)

The interior of a set C is defined as the union of all open sets contained in C. The closure of a set C is the set C together with all of its limit points.

(i) Let $C = \mathbb{R}$. We have $\text{int}(C) = \mathbb{R}$ and $\text{cl}(C) = \mathbb{R}$. It is obvious that $\text{int}(\text{cl}(C)) = \text{int}(C) = \mathbb{R}$.

(ii) Let $C = \mathbb{Q}$. We have $\text{int}(C) = \emptyset$ and $\text{cl}(C) = \mathbb{R}$. This is because every open interval in $\mathbb{R}$ contains irrational numbers. Thus $\text{int}(\text{cl}(C)) = \text{int}(\mathbb{R}) = \mathbb{R} \neq \text{int}(C) = \emptyset$.

(b)

Consider function $f(x) = \sqrt{||x||_1}$, $x \in \mathbb{R}^2$ has sublevel sets $\{x \in \mathbb{R}^2 \mid f(x) \leq \alpha\}$. The function can also be written as $f(x) = \sqrt{|x_1| + |x_2|}$, $x = (x_1, x_2)$.
Pick two elements x, y from the sublevel set of α such that $f(x) \leq \alpha$ and $f(y) \leq \alpha$. We have

$$f(x) = \sqrt{|x_1| + |x_2|} \leq \alpha \Rightarrow |x_1| + |x_2| \leq \alpha^2 \text{ and similarly } |y_1| + |y_2| \leq \alpha^2$$

Now take a point $z := \lambda x + (1-\lambda)y$, $\lambda \in [0, 1]$. Also $z_1 = \lambda x_1 + (1-\lambda)y_1$, $z_2 = \lambda x_2 + (1-\lambda)y_2$

$$\begin{aligned}
|z_1| + |z_2| &= |\lambda x_1 + (1-\lambda)y_1| + |\lambda x_2 + (1-\lambda)y_2| \\
&\leq \lambda|x_1| + (1-\lambda)|y_1| + \lambda|x_2| + (1-\lambda)|y_2| \\
&\leq \lambda(|x_1| + |x_2|) + (1-\lambda)(|y_1| + |y_2|) \\
&\leq \lambda\alpha^2 + (1-\lambda)\alpha^2 = \alpha^2
\end{aligned}$$

thus $\{x \in \mathbb{R}^2 \mid f(x) \leq \alpha\}$ is a convex set.
Now we need to show that f is not convex. Let $x = (0, 0)$, $y = (1, 0)$ and $\lambda = \frac{1}{2}$. We have $f(x) = 0$, $f(y) = 1$ and

$$f(\lambda x + (1-\lambda)y) = \sqrt{\frac{1}{2}} > \lambda f(x) + (1 - \frac{1}{2})f(y) = \frac{1}{2}$$

Check here for visualization of the sublevel sets.

(c)

Let v_{N_A} be v's projection onto the null space and $\lambda \in \mathbb{R}^m$. We know that v can be decomposite into

$$\begin{aligned}
v = A^\top\lambda + v_{N_A} &\Rightarrow Av = A(A^\top\lambda + v_{N_A}) \\
&\Rightarrow Av = AA^\top\lambda + 0 \\
&\Rightarrow Av = AA^\top\lambda \\
&\Rightarrow (AA^\top)^{-1}Av = \lambda
\end{aligned}$$

Replace λ in the original equation we get

$$\begin{aligned}
v = A^\top(AA^\top)^{-1}Av + v_{N_A} &\Rightarrow v - A^\top(AA^\top)^{-1}Av = v_{N_A} \\
&\Rightarrow (I - A^\top(AA^\top)^{-1}A)v = v_{N_A}
\end{aligned}$$

(d)

To show equivalence of two sets A and B, we need to show both $A \subseteq B$ and $B \subseteq A$.

(i) Show $\{x \in E \mid f(x) \geq c\} \subseteq \bigcap_{k=1}^{\infty}\{x \in E \mid f(x) > c - \frac{1}{k}\}$
Take $v \in \{x \in E \mid f(x) \geq c\}$ we have

$$f(v) \geq c > c - \frac{1}{k} \text{ for all } k \geq 1$$

This shows that $v \in \{x \in E \mid f(x) > c - \frac{1}{k}\}$ for all $k \geq 1$, i.e. $v \in \bigcap_{k=1}^{\infty}\{x \in E \mid f(x) > c - \frac{1}{k}\}$.
Show $\bigcap_{k=1}^{\infty}\{x \in E \mid f(x) > c - \frac{1}{k}\} \supseteq \{x \in E \mid f(x) \geq c\}$
Proof by contradiction:
Take $v \in \bigcap_{k=1}^{\infty}\{x \in E \mid f(x) > c - \frac{1}{k}\}$. Assume that $v \notin \{x \in E \mid f(x) \geq c\}$ we have

$$\begin{aligned} f(v) < c &\Rightarrow c - f(v) > 0 \\ &\Rightarrow \exists k \geq 1 \text{ such that } c - f(v) - \frac{1}{k} \geq 0 \\ &\Rightarrow c - \frac{1}{k} \geq f(v) \end{aligned} \qquad \text{Contradiction}$$

(ii) Show $\{x \in E \mid f(x) > c\} \subseteq \bigcup_{k=1}^{\infty}\{x \in E \mid f(x) \geq c + \frac{1}{k}\}$
Take $v \in \{x \in E \mid f(x) > c\}$. Let $f(v) = c + \frac{1}{k}$.
We can always pick a k' such that $k' \geq k$ so that $f(v) = c + \frac{1}{k} \geq c + \frac{1}{k'}$.
This shows that $\exists k' > 1$ such that $v \in \{x \in E \mid f(x) \geq c + \frac{1}{k'}\}$.
Show $\{x \in E \mid f(x) > c\} \supseteq \bigcup_{k=1}^{\infty}\{x \in E \mid f(x) \geq c + \frac{1}{k}\}$
Take $v \in \bigcup_{k=1}^{\infty}\{x \in E \mid f(x) \geq c + \frac{1}{k}\}$.

$$f(v) \geq c + \frac{1}{k} > c$$

Exercise 3

(i)

The assumption of h is a vector with same entry is incorrect which leads to a wrong result. Please see the correct answer below.
Suppose each the entry of h is t, we have $||h|| = \sqrt{nt^2}$. To make the derivation cleaner, we represent

$f' = f(x+h)$ and $f = f(x)$. We can rewrite the equation in the form of vector entry terms:

$$\begin{aligned}
&\lim_{t\to 0} \frac{\sqrt{\sum_i^n (f_i' - f_i - t\sum_j^m A_{ij})^2}}{\sqrt{nt^2}} = 0\\
&\Rightarrow \lim_{t\to 0} \sqrt{\frac{\sum_i^n (f_i' - f_i - t\sum_j^m A_{ij})^2}{nt^2}} = 0\\
&\Rightarrow \lim_{t\to 0} \frac{\sum_i^n (f_i' - f_i - t\sum_j^m A_{ij})^2}{nt^2} = 0\\
&\Rightarrow \lim_{t\to 0} \left(\sum_i^n \frac{(f_i' - f_i - t\sum_j^m A_{ij})^2}{nt^2}\right) = 0\\
&\Rightarrow \lim_{t\to 0} \left(\sum_i^n \frac{(f_i' - f_i)^2 - 2t(f_i' - f_i)\sum_j^m A_{ij} + t^2(\sum_j^m A_{ij})^2}{nt^2}\right) = 0\\
&\Rightarrow \lim_{t\to 0} \left(\sum_i^n \frac{(f_i' - f_i)^2}{nt^2} - \frac{2(f_i' - f_i)\sum_j^m A_{ij}}{nt} + \frac{(\sum_j^m A_{ij})^2}{n}\right) = 0\\
&\Rightarrow \lim_{t\to 0} \left(\sum_i^n \frac{(f_i' - f_i)^2}{nt^2}\right) - \lim_{t\to 0}\left(\sum_i^n \frac{2(f_i' - f_i)\sum_j^m A_{ij}}{nt}\right) + \lim_{t\to 0}\left(\sum_i^n \frac{(\sum_j^m A_{ij})^2}{n}\right) = 0\\
&\Rightarrow \lim_{t\to 0} \left(\frac{1}{nt^2}\sum_i^n (f_i' - f_i)^2\right) - \lim_{t\to 0}\left(\frac{2}{nt}\sum_i^n (f_i' - f_i)\sum_j^m A_{ij}\right) + \frac{1}{n}\sum_i^n(\sum_j^m A_{ij})^2 = 0
\end{aligned}$$

Proof: Assume that $\exists A_1, A_2 : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h\to 0} \frac{||f(x+h) - f(x) - A_1 h||}{||h||} = 0 \text{ and } \lim_{h\to 0} \frac{||f(x+h) - f(x) - A_2 h||}{||h||} = 0$$

With the previous derivation, we can write the two equations as such:

$$\lim_{t\to 0} \left(\frac{1}{nt^2}\sum_i^n (f_i' - f_i)^2\right) - \lim_{t\to 0}\left(\frac{2}{nt}\sum_i^n (f_i' - f_i)\sum_j^m A_{1ij}\right) + \frac{1}{n}\sum_i^n(\sum_j^m A_{1ij})^2 = 0$$

and

$$\lim_{t\to 0} \left(\frac{1}{nt^2}\sum_i^n (f_i' - f_i)^2\right) - \lim_{t\to 0}\left(\frac{2}{nt}\sum_i^n (f_i' - f_i)\sum_j^m A_{2ij}\right) + \frac{1}{n}\sum_i^n(\sum_j^m A_{2ij})^2 = 0$$

Exercise 3 continued on next page...

Subtracting the two equations:

$$\lim_{t\to 0}\left(\frac{2}{nt}\sum_i^n (f_i'-f_i)(\sum_j^m A_{1ij}-\sum_j^m A_{2ij})\right)+\frac{1}{n}\sum_i^n(\sum_j^m A_{1ij})^2-\frac{1}{n}\sum_i^n(\sum_j^m A_{2ij})^2=0$$
$$\Rightarrow \lim_{t\to 0}\left(\sum_i^n \frac{2(f_i'-f_i)}{nt}(\sum_j^m A_{1ij}-\sum_j^m A_{2ij})\right)+\frac{1}{n}\sum_i^n(\sum_j^m A_{1ij})^2-\frac{1}{n}\sum_i^n(\sum_j^m A_{2ij})^2=0$$
$$\Rightarrow \sum_i^n \lim_{t\to 0}\left(\frac{2(f_i'-f_i)}{nt}(\sum_j^m A_{1ij}-\sum_j^m A_{2ij})\right)+\frac{1}{n}\sum_i^n(\sum_j^m A_{1ij})^2-\frac{1}{n}\sum_i^n(\sum_j^m A_{2ij})^2=0$$
$$\Rightarrow \sum_i^n \left((\sum_j^m A_{1ij}-\sum_j^m A_{2ij})\lim_{t\to 0}\frac{2(f_i'-f_i)}{nt}\right)+\frac{1}{n}\sum_i^n(\sum_j^m A_{1ij})^2-\frac{1}{n}\sum_i^n(\sum_j^m A_{2ij})^2=0$$

Observe term $\sum_i^n\left((\sum_j^m A_{1ij}-\sum_j^m A_{2ij})\lim_{t\to 0}\frac{2(f_i'-f_i)}{nt}\right)$. As $t\to 0$, $f_i'-f_i\to 0$, we cannot determine $\lim_{t\to 0}\frac{2(f_i'-f_i)}{nt}$. To make each term of the sum equals to 0, we need $\sum_j^m A_{1ij}=\sum_j^m A_{2ij}$ for each row i.

Correct Answer

No need to expand any term. Just treat $||f(x+h)-f(x)-Ah||$ like $f(x+h)-f(x)-Ah$ without the norm.

(ii)

If $A = Df(x)$ is admitted as the derivative of f at x then we can write

$$\begin{aligned} f(x+h)=f(x)+Ah \\ &\Rightarrow \lim_{h\to 0} f(x+h)=\lim_{h\to 0}(f(x)+Ah)\\ &\Rightarrow \lim_{h\to 0} f(x+h)=f(x)\\ &\Rightarrow \lim_{h\to 0} f(x+h)=f(x)\end{aligned}$$

which implies continuity of f at x.

Correct Answer

Let $h=x-a$.

$$\begin{aligned} f(x+h)=f(x)+Df(x)h+o(||h||) &\Rightarrow f(x)=f(a)+Df(a)\cdot(x-a)+o(||x-a||)\\ &\Rightarrow \lim_{x\to a} f(x)=\lim_{x\to a} f(a)+\lim_{x\to a}(Df(a)\cdot(x-a))+\lim_{x\to a} o(||x-a||)\\ &\Rightarrow \lim_{x\to a} f(x)=f(a)+Df(a)\cdot\lim_{x\to a}(x-a)+\lim_{x\to a} o(||x-a||)\\ &\Rightarrow \lim_{x\to a} f(x)=f(a)+Df(a)\cdot 0+\lim_{x\to a} o(||x-a||)\\ &\Rightarrow \lim_{x\to a} f(x)=f(a)+\lim_{x\to a} o(||x-a||)\\ &\Rightarrow \lim_{x\to a} f(x)=f(a)+\lim_{x\to a}\left(\frac{||x-a||o(||x-a||)}{||x-a||}\right)\\ &\Rightarrow \lim_{x\to a} f(x)=f(a)\end{aligned}$$

(iii)

We can continue to derive from part (i):

$$\lim_{t\to 0}\left(\frac{1}{nt^2}\sum_i^n (f_i'-f_i)^2\right)-\lim_{t\to 0}\left(\frac{2}{nt}\sum_i^n (f_i'-f_i)\sum_j^m A_{ij}\right)+\frac{1}{n}\sum_i^n(\sum_j^m A_{ij})^2=0$$

$$\Rightarrow \frac{1}{n}\left(\sum_i^n \lim_{t\to 0}\frac{(f_i'-f_i)^2}{t^2}-\sum_i^n 2\sum_j^m A_{ij}\lim_{t\to 0}\frac{f_i'-f_i}{t}+\sum_i^n(\sum_j^m A_{ij})^2\right)=0$$

$$\Rightarrow \frac{1}{n}\sum_i^n\left(\lim_{t\to 0}\frac{(f_i'-f_i)^2}{t^2}-2\sum_j^m A_{ij}\lim_{t\to 0}\frac{f_i'-f_i}{t}+(\sum_j^m A_{ij})^2\right)=0$$

$$\Rightarrow \frac{1}{n}\sum_i^n\left(\lim_{t\to 0}\frac{f_i'-f_i}{t}-\sum_j^m A_{ij}\right)^2=0$$

$$\Rightarrow \lim_{t\to 0}\frac{f_i'-f_i}{t}=\sum_j^m A_{ij}$$

which implies that the sum of each row i of A is how sensitive f_i is to a small change of t. A is the Jacobian matrix of f and is express as:

$$A=\begin{pmatrix}\frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n}\end{pmatrix}$$

Correct Answer

Let $h=te_j$ where e_j is the jth canonical basis vector.

$$f(x+te_j)-f(x)=Df(x)te_j+o(||te_j||)$$

$$\Rightarrow \frac{f(x+te_j)-f(x)}{t}=\frac{Df(x)te_j}{t}+\frac{o(||te_j||)}{t}$$

$$\Rightarrow \frac{f(x+te_j)-f(x)}{t}=Df(x)e_j+\frac{o(||te_j||)}{t}$$

$$\Rightarrow \lim_{t\to 0}\frac{f(x+te_j)-f(x)}{t}=\lim_{t\to 0}Df(x)e_j+\lim_{t\to 0}\frac{o(||te_j||)}{t}$$

$$\Rightarrow \lim_{t\to 0}\frac{f(x+te_j)-f(x)}{t}=Df(x)e_j+\lim_{t\to 0}\frac{o(||te_j||)}{t}$$

$$\Rightarrow \lim_{t\to 0}\frac{f(x+te_j)-f(x)}{t}=Df(x)e_j+\lim_{t\to 0}\frac{|t|o(||e_j||)}{t}$$

$$\Rightarrow \lim_{t\to 0}\frac{f(x+te_j)-f(x)}{t}=Df(x)e_j$$

The left hand side denotes the directional derivative of f at x in the direction of e_j while the right hand side is the definition of the jth column of the matrix $Df(x)$. Thus each column of $Df(x)$ is the directional derivative of f at x along the direction of the canonical basis vectors i.e. the partial derivatives of f at x.

Exercise 4

(a)

The function can be rewritten as

$$f(u) = \frac{1}{2}||u - c||^2 + \frac{\mu}{2}||Au||^2$$

where $u \in \mathbb{R}^n$, $A \in \mathbb{R}^{N \times N}$ and each element at row $i < N$ and column j is defined as

$$A_{ij} := \begin{cases} -1 & \text{if } i = j \\ 1 & \text{if } i = j + 1 \\ 0 & \text{otherwise} \end{cases}$$

and its last row is defined as $A_{Nj} := 0$

(b)

For each entry of $\nabla f(u)$ we have

$$\frac{\partial f}{\partial u_i} = \begin{cases} u_i - c_i + \mu(u_i - u_{i+1}) & \text{if } i = 1 \\ u_i - c_i + \mu(-u_{i-1} + 2u_i - u_{i+1}) & \text{if } 1 < i < n \\ u_i - c_i + \mu(-u_{i-1} + u_i) & \text{if } i = n \end{cases}$$

(c)

$$\nabla f(u) = u - c + \mu A^\top A u$$

which can be verified since each element of $A^\top A$ can be expressed as

$$(A^\top A)_{1j} = \begin{cases} 1 & \text{if } j = 1 \\ -1 & \text{if } j = 2 \\ 0 & \text{otherwise} \end{cases}$$

$$(A^\top A)_{ij} = \begin{cases} -1 & \text{if } j = i - 1 \\ 2 & \text{if } j = i \\ -1 & \text{if } j = i + 1 \\ 0 & \text{otherwise} \end{cases}, 2 \leq i \leq N - 2$$

$$(A^\top A)_{Nj} = \begin{cases} -1 & \text{if } j = N - 1 \\ 1 & \text{if } j = N \\ 0 & \text{otherwise} \end{cases}$$

(d)

$$\begin{aligned} \nabla f(u) = 0 &\Rightarrow u - c + \mu A^\top A u = 0 \\ &\Rightarrow (\mu A^\top A + I)u = c \end{aligned}$$

$\mu A^\top A + I$ is symmetric. if $\det(\mu A^\top A + I) \neq 0$ then $\mu A^\top A + I$ is invertible and we can write

$$u = (\mu A^\top A + I)^{-}1c$$

Exercise 4 continued on next page...

(e)

We need to show that the matrix $\mu A^\top A + I$ is invertible, *i.e.*, $\mu A^\top A + I$ is positive definite.
Let λ be an eigenvalue of $\mu A^\top A + I$ and v be the corresponding eigenvector. Then we have

$$(\mu A^\top A + I)v = \lambda v.$$

By multiplying both sides by $v^\top$, we get

$$v^\top(\mu A^\top A + I)v = \lambda v^\top v.$$

Since $\mu A^\top A + I$ is symmetric, we have $v^\top(\mu A^\top A + I)v = \mu\|Av\|^2 + \|v\|^2$.
Since $\mu > 0$, we have $\mu\|Av\|^2 + \|v\|^2 > 0$ and thus, we get $\lambda > 0$.
This shows that all eigenvalues of $\mu A^\top A+I$ are positive, and hence it is positive definite and invertible.