Introduction to Optimization Homework (2)

Dimitri Tabatadze · Saturday 16-03-2024

Homework Assignment Not graded

Problem 2.1:

(a) First calculate the gradient

$$f(x_1, x_2) = (4x_1^2 - x_2)^2 = 16x_1^4 - 8x_1^2 x_2 + x_2^2 \Longrightarrow \nabla f(x_1, x_2) = \langle 64x_1^3 - 16x_1x_2, -8x_1^2 + 2x_2 \rangle.$$

At stationary points, $\nabla f(x) = 0$

$$\nabla f(x_1, x_2) = \langle 64x_1^3 - 16x_1x_2, -8x_1^2 + 2x_2 \rangle = 0 \Longrightarrow \begin{cases} 64x_1^3 - 16x_1x_2 = 0 \\ -8x_1^2 + 2x_2 = 0 \end{cases} \Rightarrow 4x_1^2 = x_2.$$

At such points, $f(u, 4u^2) = 0$ This means that all points $\{(u, 4u^2) : u \in \mathbb{R}\}$ are stationary points.

$$H_{f(x_1,x_2)} = \begin{pmatrix} 192x_1^2 - 16x_2 & -16x_1 \\ -16x_1 & 2 \end{pmatrix} \Longrightarrow H_{f(u,4u^2)} = 2\begin{pmatrix} 64u^2 & -8u \\ -8u & 1 \end{pmatrix}, \det\left(H_{f(u,4u^2)}\right) = 0$$

so second derivative test is not suitable for this example. If we let $d \in \mathbb{R}^2$ be an arbitrary non-zero vector we can write

$$\begin{aligned} f(x + d) &= \left(4(x_1 + d_1)^2 - (x_2 + d_2)\right)^2 \\ &= (4x_1^2 - x_2 + 8x_1d_1 + 4d_1^2 - d_2)^2 \\ &= (4x_1^2 - x_2)^2 + 2(4x_1^2 - x_2)(8x_1d_1 + 4d_1^2 - 2x_2d_2 - d_2^2) + (8x_1d_1 + 4d_1^2 - 2x_2d_2 - d_2^2)^2 \end{aligned}$$

and in the case where $4x_1^2 = x_2$ we get that

$$\begin{aligned} f(x + d) &= (4x_1^2 - x_2)^2 + 2(4x_1^2 - x_2)(8x_1d_1 + 4d_1^2 - 2x_2d_2 - d_2^2) + (8x_1d_1 + 4d_1^2 - 2x_2d_2 - d_2^2)^2 \\ &= 0 + 0 + (8x_1d_1 + 4d_1^2 - 2x_2d_2 - d_2^2)^2 \\ &\geq 0. \end{aligned}$$

Therefore the minimum of f is 0 and is attained at points $\{(u, 4u^2) : u \in \mathbb{R}\}$.

Problem 2.2:

Proof: Let $S = \{x \in \mathbb{R}^n : f(x) \leq f(y)\}$ be a sub-level set for some y and $a, b \in S$ such that $f(a) = f(b)$. Since f is convex, i.e. $(1 - \lambda)f(a) + \lambda f(b) \geq f((1 - \lambda)a + \lambda b)$ if some c is on the line between a and b, $f(c) \leq f(a) = f(b)$ meaning that if S contains a and b it will also contain c. □

Problem 2.3:

(a) *Proof*:

$$\begin{aligned} f(g((1 - \lambda)a + \lambda b)) &\leq f((1 - \lambda)g(a) + \lambda g(b)) && \text{(by convexity and monotonousity)} \\ &\leq (1 - \lambda)f(g(a)) + \lambda f(g(b)) \end{aligned}$$

□

(b) *Proof*: Let $f(x) = e^{-x}, g(x) = x^2, f(g(x)) = e^{-x^2}$. At points -1 and 1 value of the function is $f(g(-1)) = f(g(1)) = e^{-1}$ which is evidently less than the value at 0 which is $f(g(0)) = e^0$ therby contradicting the convexity. □

Problem 2.4:

Let $\bar{x}$ and $\bar{y}$ be the optimizers of the problem and $\lambda \in (0, 1)$. Then, by convexity of $f(x)$, we know that any point $c = (1 - \lambda)x + \lambda y$ gives us

$$f(c) \le (1 - \lambda)f(x) + \lambda f(y) = f(x) = f(y)$$

and since $f(x) \le f(a) \forall a \in K$ we get that $f(c) = f(x) = f(y) \Longrightarrow c \in \{x \in K : f(x) \le f(y) \forall y \in K\}$.

Graded Homework Assignment

Problem 2.1:

$f(x_1, x_2) = x_1^2 - 5x_1x_2^2 + 5x_2^4$

(a) To determine all stationary points of f, first we find

$$\nabla f(x_1, x_2) = \langle 2x_1 - 5x_2^2, -10x_1x_2 + 20x_2^3 \rangle$$

then we find such (x_1, x_2) that $\nabla f(x_1, x_2) = 0$

$$\begin{cases} 2x_1 - 5x_2^2 = 0 \\ -10x_1x_2 + 20x_2^3 = 0 \end{cases} \Longrightarrow \begin{cases} 2x_1 - 5x_2^2 = 0 \\ -x_1 + 2x_2^2 = 0 \end{cases} \Longrightarrow \begin{cases} x_1 = 0 \\ x_2 = 0 \end{cases}$$

(b)
- $f(x_1, 0) = x_1^2$. $\bar{x}_1 = 0$ is a global minimizer since $0^2 \le x_1^2$ for all x_1.
- $f(0, x_2) = 5x_2^4$. $\bar{x}_2 = 0$ is a global minimizer since $5 \cdot 0^4 \le 5 \cdot x_2^4$ for all x_2.

(c) We know that $\bar{x} = 0$ is a stationary point. We need to find the Hessian of f

$$H_{f(x_1,x_2)} = \begin{pmatrix} 2 & -10x_2 \\ -10x_2 & -10x_1 + 50x_2^2 \end{pmatrix}$$

Now we just plug in $\bar{x} = 0$ and obtain

$$H_{f(0,0)} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

the we find the eigenvalues

$$\det(H_f - \lambda I) = \lambda(2 - \lambda) = 0 \Longrightarrow \lambda = 0, 2.$$

This means that $H_{f(0,0)}$ is positive semi-definite and that doesn't tell us anthing about the class of the stationary point.

Consider the curves

- $A = \{(2u^2, u) : u \in \mathbb{R}\}$

$$f(2u^2, u) = 4u^4 - 10u^4 + 5u^4 = -u^4$$

- $B = \{(u, 0) : u \in \mathbb{R}\}$

$$f(B) = \{u^2 - 5u \cdot 0 + 5 \cdot 0^4 : u \in \mathbb{R}\} = \{u^2 : u \in \mathbb{R}\}$$

It's clear that $\bar{x}$ has values both greater and lower than $f(\bar{x}) = 0$. This means that the point $\bar{x}$ is a saddle point. Figure 1 also illustrates that quite well.

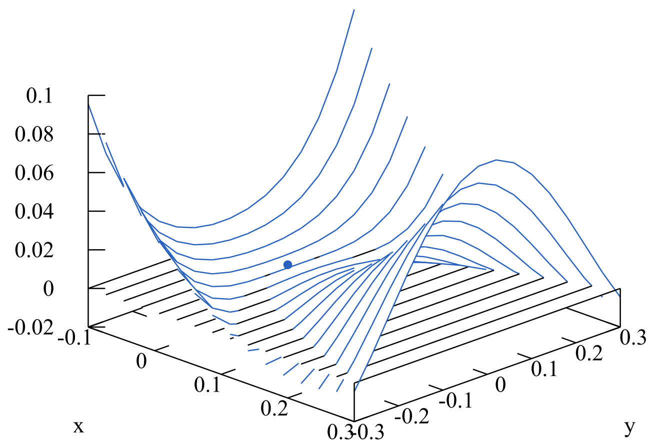

Figure 1: plot of f near $(0,0)$

Problem 2.2:

- f is bounded below $\Longleftarrow b \in \{Ay : y \in \mathbb{R}^n\}$

  Proof: let $b = Ay$

$$\begin{aligned} f(x) &= x^T A x + 2b^T x + c \\ &= x^T A x + 2y^T A x + c \\ &= \langle A^{\frac{1}{2}}(x+y), A^{\frac{1}{2}}(x+y)\rangle - y^T A y + c \\ &= \underbrace{(x+y)^T A (x+y)}_{\geq 0 \text{ (PTSD)}} - \underbrace{y^T A y + c}_{\text{constant}} \\ &\geq -y^T A y + c \end{aligned}$$

  □

- f is bounded below $\Longrightarrow b \in \{Ay : y \in \mathbb{R}^n\}$

  Proof: Since A is positive semi-definite and not positive definite, assume $\det(A) = 0$ and $\ker(A) \neq \{0\}$. If we let $x \in \ker(A)$ we get

$$\begin{aligned} f(x) &= x^T A x + 2b^T x + c \\ &= 0 + 2b^T x + c. \end{aligned}$$

  From here we observe that if b were to be of the form Ay we would have

$$\begin{aligned} f(x) &= 2b^T x + c \\ &= 2y^T \underbrace{Ax}_{=0} + c \\ &= c \end{aligned}$$

  but otherwise, $f(x)$ would be a linear function in respect to x which is not bounded from below. □

Problem 2.3:

1. Let $a = (1,0)$ and $b = (-1,0)$, $\|a\|^2 = \|b\|^2 = 1$ therefore $a, b \in A$. The midpoint is $c = (0,0)$ has $\|c\|^2 = 0 \neq 1 \Longrightarrow c \notin A$. A is not convex.
2. *Proof*: Let $a, b \in \mathbb{R}^n$ and $a_{\max}, b_{\max} \leq 1 \Longleftrightarrow a, b \in B$. Now take the midpoint c of a and b,

$$c = \left(\frac{a_1 + b_1}{2}, \frac{a_2 + b_2}{2}, \dots, \frac{a_n + b_n}{2}\right).$$

We can safely assume that $a_i = a_{\max} \forall i$ and $b_i = b_{\max} \forall i$ giving us

$$c = \left(\frac{a_{\max} + b_{\max}}{2}, \frac{a_{\max} + b_{\max}}{2}, \dots, \frac{a_{\max} + b_{\max}}{2}\right)$$

from which it's clear that $c_{\max} = (a_{\max} + b_{\max})2$. We now prove that $\frac{a_{\max}+b_{\max}}{2} \leq \max\{a_{\max}, b_{\max}\}$. Assume $a_{\max} \geq b_{\max}$, we get

$$\begin{aligned}
\frac{a_{\max} + b_{\max}}{2} &\leq a_{\max} \\
a_{\max} + b_{\max} &\leq 2a_{\max} \\
b_{\max} &\leq a_{\max}
\end{aligned}$$

□

3. Let $a = (10, 0)$ and $b = (0, 10)$, $\min a = \min b = 0 \leq 1$ therefore $a, b \in C$. The midpoint is $c = (5, 5)$ has $\min c = 5 \nleq 1 \Longrightarrow c \notin C$. C is not convex.

Problem 2.4:

Proof:

- f is convex $\Longrightarrow g_{x,d}$ is convex.

$$\begin{aligned}
g_{x,d}((1-\lambda)a + \lambda b) &= f(x + ((1-\lambda)a + \lambda b)d) \\
&= f((1-\lambda)(x + ad) + \lambda(x + bd)) \\
&\leq (1-\lambda)f(x + ad) + \lambda f(x + bd) \\
&= (1-\lambda)g_{x,d}(a) + \lambda g_{x,d}(b)
\end{aligned}$$

- f is convex $\Longleftarrow g_{x,d}$ is convex.

$$\begin{aligned}
f((1-\lambda)a + \lambda b) &= g_{a,b-a}((1-\lambda) \cdot 0 + \lambda \cdot 1) \\
&\leq (1-\lambda)g_{a,b-a}(0) + \lambda g_{a,b-a}(1) \\
&= (1-\lambda)f(a) + \lambda f(b)
\end{aligned}$$

□