Introduction to Optimization — Homework 2

Dimitri Tabatadze · 12:06, March 1, 2024

Homework Assignment Not graded

1. (a) First calculate the gradient

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

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

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

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

$$H_f(x_1,x_2) = \begin{pmatrix} 192x_1^2 - 16x_2 & -16x_1 \\ -16x_1 & 2 \end{pmatrix} \implies H_f(u,4u^2) = 2\begin{pmatrix} 64u^2 & -8u \\ -8u & 1 \end{pmatrix}, \ \det(H_f(u,4u^2)) = 0$$

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

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

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 $\{(u, 4u^2) : u \in \mathbb{R}\}$.

2. 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.

3. (a) *Proof.*

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

$\square$

(b) Let $f(x) = e^{-x}, g(x) = x^2, f(g(x)) = e^{-x^2}$. At points $-1, 1$ 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.

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

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

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

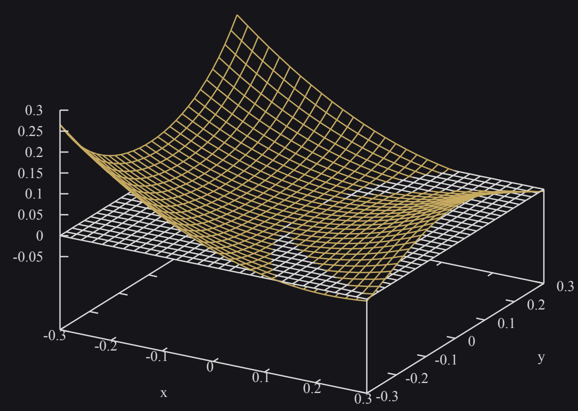

Figure 1: plot of f near $(0,0)$

Graded Homework Assignment

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

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

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

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

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

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

$$H_f(x_1, x_2) = \begin{pmatrix} 2 & -10x_2 \\ -10x_2 & -10x_1 + 60x_2^2 \end{pmatrix}$$

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

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

then we find the eighenvalues

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

This means that $H_f(0,0)$ is positive semi-definite and that doesn't tell us anything about the class of the stationary point. From figure 1 it's clear that the point is not a minimizer but a saddle point.

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

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

   - f is bounded below $\implies$ $b \in \{Ay : y \in \mathbb{R}^n\}$ Since A is positive semi-definite and not positive definite, $\det(A) = 0$ and $\ker(A) \neq \{0\}$. If we let $x \in \ker(A)$ we get

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

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

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

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

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 $\|c\|^2 = 0 \implies c \notin A$. A is not convex.
   2. In 2. I proved that level sets of a convex function are convex and all norms are convex, therefore the sub-level set $\{x \in \mathbb{R}^n : \max_{i=1,2,\ldots,n} x_i \leq 1\} = \{x \in \mathbb{R}^n : \|x\|_\infty \leq 1\}$ is convex.
   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 $\min c = 5 \not\leq 1 \implies c \notin C$. C is not convex.

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

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