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

Andre Milzarek · Fall Semester 2020/21

MDS 6106 – Introduction to Optimization

Exercise Sheet Nr.: 02

Name: 康昊宇 Kang Haoyu Student ID: 220041025

In the creation of this solution sheet, I worked together with:

Name: ______ Student ID: ______

Name: ______ Student ID: ______

Name: ______ Student ID: ______

For correction:

Exercise						Σ
Grading						

A2.1

a) If $f(x)$ is coercive. we should show that $\lim_{\|x\|\to\infty} f(x) = +\infty$

as x equal to $(-\infty, 0)$

$\lim_{x_1\to-\infty} f(x) = \frac{1}{3}x_1^3 - \frac{3}{2}x_1 = x_1\left(\frac{1}{3}x_1^2 - \frac{3}{2}\right) \to -\infty$ thus, f is not coercive

b) According to function f

$$\nabla f(x) = \begin{pmatrix} x_1^2 - \frac{3}{2} - x_2^2 \\ -2x_1x_2 + 4x_2^3 \end{pmatrix} \qquad \nabla^2 f(x) = \begin{pmatrix} 2x_1 & -2x_2 \\ -2x_2 & -2x_1 + 12x_2^2 \end{pmatrix}$$

X^* are stationary points, when $\nabla f(x^*) = 0$

$$\begin{pmatrix} x_1^2 - \frac{3}{2} - x_2^2 \\ -2x_1x_2 + 4x_2^3 \end{pmatrix} = 0 \Rightarrow x_1^* = \left(\frac{\sqrt{6}}{2}, 0\right) x_2^* = \left(-\frac{\sqrt{6}}{2}, 0\right) x_3^* = \left(\frac{3}{2}, \frac{\sqrt{3}}{2}\right) x_4^* = \left(\frac{3}{2}, -\frac{\sqrt{3}}{2}\right)$$

c) 1° put the x_1^* into the $\nabla^2 f(x) \Rightarrow \begin{pmatrix} \sqrt{6} & 0 \\ 0 & -\sqrt{6} \end{pmatrix} \Rightarrow \lambda_1 = \sqrt{6}, \lambda_2 = -\sqrt{6}$

so, $\nabla^2 f(x)$ is indefinite, $x_1^* = \left(\frac{\sqrt{6}}{2}, 0\right)$ is saddle point

2° put the $x_2^* = (-\frac{\sqrt{6}}{2}, 0)$ into the $\nabla^2 f(x_2^*) = \begin{pmatrix} -\sqrt{6} & 0 \\ 0 & \sqrt{6} \end{pmatrix} \Rightarrow \lambda_1 = \sqrt{6}, \lambda_2 = \sqrt{6}$

thus, $\nabla^2 f(x)$ is indefinite that means x_2^* is saddle point

3° put the $x_3^* = (\frac{3}{2}, \frac{\sqrt{3}}{2})$ into the $\nabla^2 f(x_3^*) = \begin{pmatrix} 3 & -\sqrt{3} \\ -\sqrt{3} & 6 \end{pmatrix} \Rightarrow \lambda_1\lambda_2 = 15 \quad \lambda_1 + \lambda_2 = 9$

so $\lambda_1, \lambda_2 > 0$, $\nabla^2 f(x)$ is positive definite which means x_3^* is local minimizer

4° put the $x_4^* = (\frac{3}{2}, -\frac{\sqrt{3}}{2})$ into the $\nabla^2 f(x_4^*) = \begin{pmatrix} 3 & \sqrt{3} \\ \sqrt{3} & 6 \end{pmatrix} \Rightarrow \lambda_1\lambda_2 = 15 \quad \lambda_1 + \lambda_2 = 9$

So, $\lambda_1, \lambda_2 > 0$ and $\nabla^2 f(x)$ is positive definite which means x_4^* is local minimizer

d)

According to the results from c), $\nabla^2 f(x_3^*)$ and $\nabla^2 f(x_4^*)$ both are positive definite

Thus, x_3^* and x_4^* must be strict local minimizer.

In addition, we know that f is not coercive since x equal to $(-\infty, 0)$

$f \to -\infty$, which means that f doesn't have global minimizer

So, f just possess strict local minimizer.

A 2.2

a) If $f(x)$ is coercive. we should show that $\lim\limits_{\|x\|\to\infty} f(x) = +\infty$

$$\lim_{\|x\|\to\infty} f(x) = x_1^4 - 2x_1^2 + x_2^2 + 2x_2x_3 + 2x_3^2$$

$$= x_1^2(x_1^2-3) + x_1^2 + \frac{1}{3}x_2^2 + \left(\frac{\sqrt{6}}{3}x_2 + \frac{\sqrt{6}}{2}x_3\right)^2 + \frac{1}{2}x_3^2$$

$$\geq -\frac{9}{4} + \frac{1}{3}\|x\|^2 + \underbrace{\frac{1}{6}x_3^2 + \frac{2}{3}x_1^2}_{\geq 0} + \underbrace{\left(\frac{\sqrt{6}}{3}x_2 + \frac{\sqrt{6}}{2}x_3\right)^2}_{\geq 0} \geq \lim_{\|x\|\to\infty} \frac{1}{3}\|x\|^2 - \frac{9}{4} \longrightarrow +\infty$$

thus $f(x)$ is coercive

b) If we want to verify if it is convex, we can compute whether $\nabla^2 f(x)$ is positive semidefinite

$$\nabla f(x) = \begin{pmatrix} 4x_1^3 - 4x_1 \\ 2x_2 + 2x_3 \\ 2x_2 + 4x_3 \end{pmatrix} \qquad \nabla^2 f(x) = \begin{pmatrix} 12x_1^2 - 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 4 \end{pmatrix}$$

$$|\nabla^2 f(x) - \lambda I| = \begin{vmatrix} 12x_1^2-4-\lambda & 0 & 0 \\ 0 & 2-\lambda & 2 \\ 0 & 2 & 4-\lambda \end{vmatrix} = (12x_1^2-4-\lambda)[(2-\lambda)(4-\lambda)-4]$$

$$= (12x_1^2-4-\lambda)(8-6\lambda+\lambda^2-4)$$

$$= (12x_1^2-4-\lambda)(\lambda^2-6\lambda+4)$$

thus $\lambda_1 = 3+\sqrt{5}$ $\lambda_2 = 3-\sqrt{5}$ $\lambda_3 = 12x_1^2-4$ $\quad = (12x_1^2-4-\lambda)[(\lambda-3)^2-5] = 0$

if $x_1 \in \left(-\frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3}\right)$ $\lambda_3 < 0$ $\Rightarrow \nabla^2 f(x)$ is not positive semidefinite on $\mathbb{R}^3$ for any x

$\Rightarrow$ function f is not convex on $\mathbb{R}^3$

c) according to (b) $\nabla f(x) = \begin{pmatrix} 4x_1^3 - 4x_1 \\ 2x_2 + 2x_3 \\ 2x_2 + 4x_3 \end{pmatrix}$ If we want to find stationary points $\nabla f(x)$ should be equal to zero $\Rightarrow$ $\nabla f(x) = \begin{pmatrix} 4x_1^3 - 4x_1 \\ 2x_2 + 2x_3 \\ 2x_2 + 4x_3 \end{pmatrix} = 0 \Rightarrow x_1^* = (0, 0, 0)$

$x_2^* = (1, 0, 0)$

$x_3^* = (-1, 0, 0)$

Thus, x_1^*, x_2^*, x_3^* are stationary points

1° put x_1^* into the $\nabla^2 f(x) \Rightarrow \nabla^2 f(x_1^*) = \begin{pmatrix} -4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 4 \end{pmatrix} \Rightarrow \lambda_1 = 3+\sqrt{5}$ $\lambda_2 = 3-\sqrt{5}$ $\lambda_3 = -4$

so $\nabla^2 f(x_1^*)$ is indefinite $\Rightarrow x_1^* = (0, 0, 0)$ is paddle point

2° put x_2^* into the $\nabla^2 f(x) \Rightarrow \nabla^2 f(x_2^*) = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 4 \end{pmatrix} \Rightarrow \lambda_1 = 3+\sqrt{5}$ $\lambda_2 = 3-\sqrt{5}$ $\lambda_3 = 8$

so $\nabla^2 f(x_2^*)$ is positive definite $\Rightarrow x_2^* = (1, 0, 0)$ is strict local minimum point

3°. put x_3^* into the $\nabla^2 f(x) \Rightarrow \nabla^2 f(x_3^*) = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 4 \end{pmatrix} \Rightarrow \lambda_1 = 3+\sqrt{5}$ $\lambda_2 = 3-\sqrt{5}$, $\lambda_3 = 8$

so $\nabla^2 f(x_3^*)$ is positive definite $\Rightarrow x_3^* = (-1, 0, 0)$ is strict local minimum point

plus, $f(x_2^*) = -1$ $f(x_3^*) = -1$, and f is coercive that means global minimum points are generated from local minimum points and $f(x_2^*) = f(x_3^*)$

Thus, $x_2^* = (1, 0, 0)$ and $x_1^* = (-1, 0, 0)$ both are non-strict global minimum point

A2.3

a) X_1 is not convex set.

Because we can verify it by list a counter-example

we can assume that X_1 is in one-dimension

Set $\alpha=1$, $\beta=4$ $a=1$ $\Rightarrow$ $X_1=\{x\in R: 1\le x^2\le 4\}$

Let $x=-1\in X_1$ $y=1\in X_1$, $\lambda\in[0,1]$

So $[\lambda x+(1-\lambda)y]^2=(1-2\lambda)^2$ $\lambda\in[0,1]$ if $\lambda=\frac{1}{2}$ $[\lambda x+(1-\lambda)y]^2=0\notin X_1$

Thus X_1 is not a convex set

2°. $X_2=\{x\in\mathbb{R}^n: \|x-a\|_2\le\|x-b\|_2\}$, $a,b\in R^n$, $a\neq b$

$$\Rightarrow \|x-a\|_2^2-\|x-b\|_2^2\le 0 \Rightarrow (x-a)^T(x-a)-(x-b)^T(x-b)\le 0$$

$$\Rightarrow 2x^T(b-a)+a^Ta-b^Tb\le 0 \Rightarrow X_2=\{x\in\mathbb{R}^n: 2x^T(b-a)+a^Ta-b^Tb\le 0\}$$

Thus set $x\in X_2$ $y\in X_2$ $\Rightarrow$ $2[\lambda x+(1-\lambda)y]^T(b-a)+a^Ta-b^Tb$

$$= \underbrace{2\lambda x^T(b-a)}_{\le\lambda(b^Tb-a^Ta)}+\underbrace{2(1-\lambda)y^T(b-a)}_{\le(1-\lambda)(b^Tb-a^Ta)}+a^Ta-b^Tb$$

Thus X_2 is a convex set ≤ 0

(b) According to question. Set $x \in X$ and $y \in X$, thus x and y are positive semidefinite and symmetric $n\times n$ matrices.

Therefore $\begin{cases} d^T x \cdot d \geq 0 \\ \sum_{i=1}^{n} x_{ii} = 1 \end{cases}$ $\begin{cases} d^T y \cdot d \geq 0 \\ \sum_{i=1}^{n} y_{ii} = 1 \end{cases}$

Set $\lambda \in [0,1]$, $b = \lambda x + (1-\lambda) y$

$$d^T b d = d^T \cdot [\lambda x + (1-\lambda) y] \cdot d = \underbrace{\lambda d^T x d}_{\geq 0} + \underbrace{(1-\lambda) d^T y \cdot d}_{\geq 0} \geq 0$$

$$\sum_{i=1}^{n} b_{ii} = \lambda \sum_{i=1}^{n} x_{ii} + (1-\lambda) \sum_{i=1}^{n} y_{ii} = \lambda + (1-\lambda) = 1 \Rightarrow tr(b) = 1$$

Thus $b = \lambda x + (1-\lambda) y \in$ set $X \Rightarrow X$ is a convex subset of $R^{n\times n}$

(c) 1° False. Set $Y = X_1 \cup X_2$ if $x \in Y$ and $y \in Y \Rightarrow \begin{cases} x \in X_1 \text{ or } x \in X_2 \\ y \in X_1 \text{ or } y \in X_2 \end{cases}$

when $\lambda x + (1-\lambda) y \in Y$, it means that $\lambda x + (1-\lambda) y \in X_1$ $(x \in X_1, y \in X_1)$

or $\lambda x + (1-\lambda) y \in X_2$ $(x \in X_2, y \in X_2)$

So, if $X_1 \subseteq X_2$ or $X_2 \subseteq X_1$, the union of two convex is also a convex set

2° True.

Since $f: R^n \to R$ be a concave, we can get that

$$\lambda f(x) + (1-\lambda) f(y) \leq f(\lambda x + (1-\lambda) y) \qquad \lambda \in [0,1]$$

if we limit $f(x) \geq 0$ and $f(y) \geq 0$

$$f(\lambda x + (1-\lambda) y) \geq \lambda f(x) + (1-\lambda) f(y) \geq 0$$

$$\Rightarrow f(\lambda x + (1-\lambda) y) \geq 0$$

Thus set $X := \{x \in R^n : f(x) \geq 0\}$ is convex

A 2.4

a) 1° we can compute the $\nabla^2 f(x)$ to verify

Since $x \in R$, $f'(x) = \frac{1}{2} \cdot \frac{-2x^{-3}}{\sqrt{1+x^{-2}}}$ $f''(x) = \frac{1}{2} \cdot \frac{6x^{-4} \cdot \sqrt{1+x^{-2}} - \frac{1}{2} \cdot \frac{4x^{-6}}{\sqrt{1+x^{-2}}}}{1+x^{-2}} = \frac{\frac{6x^{-4}+6x^{-6}-2x^{-6}}{\sqrt{1+x^{-2}}}}{1+x^{-2}}$

$= \frac{6x^{-4}+4x^{-6}}{(1+x^{-2})\sqrt{1+x^{-2}}}$

For any $x \in R_{++}$, $f''(x) > 0$, thus $f(x) = \sqrt{1+x^{-2}}$ is convex

2° we can use definition of general composition to solve

Let $h(x) = Ax-b$ $g(x) = \|x\|^2$, Since $h(x)$ is linear function and $g(x)$ is convex function

So $f_1(x) = g(h(x))$ is convex and $f_2(x) = u\|x\|^2$ is also a convex function

$f(x) = \frac{1}{2} f_1(x) + u f_2(x)$ should be a convex function through sum rule

3° Let $f_1(x) = \|x\|^2$ $h_1(x) = 0$ $h_2(x) = 1$ $h_3(x,y) = -b_i(a_i^T x + y)$

According to definition of Taking Maximum, $h_1(x)$ $h_2(x)$ and $h_3(x,y)$ are convex function, so $h(x) = \max\{0, 1, -b_i(a_i^T x + y)\}$ is convex function

Let $f_2(x) = \sum_{i=1}^{m} h(x,y)$, due to definition of sum rule, $h(x)$ is convex, so $f_2(x)$ is convex. $f(x,y) = \frac{\lambda}{2} f_1(x) + f_2(x,y)$, because $f_1(x)$ and $f_2(x)$ all convex, $f(x,y)$ is convex function

(b) Function g is not convex, accoring to definitor of convex function

$$\lambda g(x) + (1-\lambda) g(y) = \lambda (f(x))^2 + (1-\lambda)(f(y))^2$$

$$g(\lambda x + (1-\lambda) y) = (f[\lambda x + (1-\lambda) y])^2$$

$\Rightarrow$ Since f is convex

$$\Rightarrow g(\lambda x + (1-\lambda) y) = (f\underbrace{[\lambda x + (1-\lambda) y]}_{\leq \lambda f(x) + (1-\lambda) f(y)})^2 \begin{cases} \leq (\lambda f(x) + (1-\lambda) f(y))^2, & (|\lambda x + (1-\lambda) y| < |\lambda f(x) + (1-\lambda) f(y)|) \\ \geq (\lambda f(x) + (1-\lambda) f(y))^2, & (|f[\lambda x + (1-\lambda) y]| > |\lambda f(x) + (1-\lambda) f(y)|) \end{cases}$$

$$\Rightarrow (f[\lambda x + (1-\lambda) y])^2 \geq \lambda^2 (f(x))^2 + (1+\lambda^2 - 2\lambda)(f(y))^2 + 2\lambda(1-\lambda) f(x) f(y)$$

$$= \lambda^2 (f(x))^2 + (\lambda - \lambda^2)(f(x))^2 + (1+\lambda^2 - 2\lambda)(f(y))^2 + (\lambda - \lambda^2)(f(y))^2 - \lambda(1-\lambda)[f(x) - f(y)]^2$$

$$= \lambda (f(x))^2 + (1-\lambda)[f(y)]^2 - \underbrace{\lambda(1-\lambda)[f(x) - f(y)]^2}_{\text{when } \leq 0}$$

$$\geq \lambda g(x) + (1-\lambda) g(y)$$

Thus, it exists the situation that $g[\lambda x + (1-\lambda) y] \geq \lambda g(x) + (1-\lambda) g(y)$

As for $x \longmapsto \frac{1}{2}(\|x\|^2 - 1)^2$ Let $f(x) = \|x\|^2 - 1$ and it's a convex function $\mathbb{R}^d \to \mathbb{R}$

$h(x) = x^2$ $\quad g(x) = \frac{1}{2} h(f(x)) = \frac{1}{2}(f(x))^2$, and we've verified that $g(x) = (f(x))^2$ above is not convex, So $x \mapsto \frac{1}{2}(\|x\|^2 - 1)^2$ is not convex yet.

A 2.f

a) $f_\beta(x) = \frac{1}{2}(x-b)^T(x-b) + \frac{\beta}{2}\left[\begin{pmatrix}\vdots\end{pmatrix}^T x\right]^2.$ $\quad \left[\begin{pmatrix}\vdots\end{pmatrix}^T h\right]^T \begin{pmatrix}\vdots\end{pmatrix}^T h.$

$\nabla f_\beta(x) = x-b+\beta\cdot\begin{pmatrix}\vdots\end{pmatrix}\begin{pmatrix}\vdots\end{pmatrix}^T x$

$\nabla^2 f_\beta(x) = I + \beta\begin{pmatrix}\vdots\end{pmatrix}\begin{pmatrix}\vdots\end{pmatrix}^T$

b) If f_β is strongly convex. $h^T\nabla^2 f_\beta(x)h \geq \mu\|h\|^2$, $\mu>0$

$\Rightarrow h^T\nabla^2 f_\beta(x)h = \left(h^T+\beta h^T\begin{pmatrix}\vdots\end{pmatrix}\begin{pmatrix}\vdots\end{pmatrix}^T\right)h = h^Th + \beta h^T\begin{pmatrix}\vdots\end{pmatrix}\begin{pmatrix}\vdots\end{pmatrix}^T h$

$= \|h\|^2 + \beta\left\|\begin{pmatrix}\vdots\end{pmatrix}^T h\right\|^2$

$= (1+\alpha\beta)\|h\|^2$

so there exists $\mu>0$, $(1+\alpha\beta)\|h\|^2 \geq \mu\|h\|^2$

so f_β is strongly convex for all $\beta \geq 0$

(c) According to (b), we know that f_β is strongly convex, that is to say $\nabla^2 f_\beta(x) \succ 0$,
So $\nabla^2 f(x)$ is a positive definite which means $f_\beta(x)$ only has one stationary point
Through the result of (a), we know $\nabla f(x) = x - b + \beta \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}^T x$
If we want to find stationary point $\nabla f(x) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

So $\nabla f(x) = x - b + \beta \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}^T x = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

$$\Rightarrow \left(I + \beta \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}^T \right) x = b$$

$$x + \beta \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \sum_{i=1}^{n} x_i = b \Rightarrow \begin{pmatrix} x_1 + \beta \sum_{i=1}^{n} x_i \\ x_2 + \beta \sum_{i=1}^{n} x_i \\ \vdots \\ x_n + \beta \sum_{i=1}^{n} x_i \end{pmatrix} = b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

$$\Rightarrow (1 + n\beta) \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} b_i$$

$$\sum_{i=1}^{n} x_i = \frac{\sum_{i=1}^{n} b_i}{1 + n\beta} \Rightarrow x_i = b_i - \frac{\beta \sum_{i=1}^{n} b_i}{n\beta}$$

$$\Rightarrow x_\beta^* = \begin{pmatrix} b_1 - \beta \sum_{i=1}^{n} b_i / 1+n\beta \\ \vdots \\ b_n - \beta \sum_{i=1}^{n} b_n / 1+n\beta \end{pmatrix} = b - \frac{1}{1+n\beta} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\begin{pmatrix} \beta \\ \vdots \\ \beta \end{pmatrix}^T b$$ and $f_\beta(x)$ is a strict convex function.

So, x_β^* is a strict local minimizer

(d) According to the result of x_β^*

$$x_\beta^* = \begin{pmatrix} b_1 - \beta\sum_{i=1}^{n} b_i / 1+n\beta \\ \vdots \\ b_n - \beta\sum_{i=1}^{n} b_i / 1+n\beta \end{pmatrix} \qquad x_\beta^{*(1)} = b_1 - \beta\sum_{i=1}^{n} b_i / 1+n\beta$$

$$\Rightarrow \lim x_1^* = \lim_{\beta\to\infty} x_\beta^{*(1)} = \lim_{\beta\to\infty} b_1 - \frac{\sum_{i=1}^{n} b_i}{\frac{1}{\beta}+n} = b_1 - \frac{\sum_{i=1}^{n} b_i}{n}$$

Thus $x^* = \begin{pmatrix} b_1 - \sum_{i=1}^{n} b_i / n \\ \vdots \\ b_n - \sum_{i=1}^{n} b_i / n \end{pmatrix} = b - \frac{1}{n}\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}^T b$

Put the x^* into the constraint.

1° $\mathbb{1}^T x^* = \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}^T \left(b - \frac{1}{n}\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}^T b\right) = \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}^T b - \frac{1}{n}\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}^T\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}^T b$

$= \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}^T b - \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}^T b = 0$

2° $\mathbb{1}^T x^* = \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}^T \begin{pmatrix} b_1 - \sum_{i=1}^{n} b_i / n \\ \vdots \\ b_n - \sum_{i=1}^{n} b_i / n \end{pmatrix} = \sum_{i=1}^{n} x_i^* = \sum_{i=1}^{n} b_i - n \cdot \frac{1}{n}\sum_{i=1}^{n} b_i = 0$