Homework 1

1) Proove:

f is convex $\Leftrightarrow$ $x_1, \dots, x_k \in D$, $\alpha_1 \dots \alpha_k \in [0,1]$ $\sum^k \alpha_i = 1$: $f(\sum^k \alpha_i x_i) \leq \sum^k \alpha_i f(x_i)$

Induction for proving $\Rightarrow$

1) $n=2$ $\quad f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$ (from definition of convexity)

2) Assume it holds for $n=k$:

$$\boxed{f(\sum^k \alpha_i x_i) \leq \sum_1^k \alpha_i f(x_i)}$$

Proove it holds for $n=k+1$:

$$f(t \cdot \sum_1^k \alpha_i x_i + (1-t)x_{k+1}) \underset{\substack{\text{because} \\ f \text{ is complex}}}{\leq} t f(\sum_1^k \alpha_i x_i) + (1-t) f(x_{k+1}) \underset{\text{assumption}}{\leq}$$

$$\leq t \cdot \sum_1^k \alpha_i f(x_i) + (1-t) f(x_{k+1}) \leq \sum_1^k \alpha_i t\, f(x_i) + (1-t) f(x_{k+1})$$

Define coefficients: $\beta_1 = \alpha_1 t \dots \beta_k = \alpha_k t$, $\beta_{k+1} = 1-t$

Check the condition: $\sum_1^k \alpha_i t + (1-t) = t + 1 - t = 1$ ✓

Finally: $\boxed{f(\sum^{k+1} \beta_i x_i) \leq \sum_1^{k+1} \beta_i f(x_i)}$

$\Leftarrow$ if $f(\sum \alpha_i x_i) \leq \sum \alpha_i f(x_i)$ we can set all $a_3, a_4, \dots, a_k = 0$

and $f(\alpha_1 x_1 + \alpha_2 x_2) \leq \alpha_1 f(x_1) + \alpha_2 f(x_2)$, since that is the convexity definition it follows that f is convex function.

2) $f(x,y) = x^2 + e^x + y^2 - xy \qquad f: [-2,2] \times [-2,2] \to R$

Finding L:

$\nabla f(x,y) = \begin{bmatrix} 2x + e^x - y \\ 2y - x \end{bmatrix}$; by definition f is L-lipschitz iff $\|\nabla f\| \leq L$

$\|\nabla f(x,y)\| = \sqrt{(2x + e^x - y)^2 + (2y-x)^2}$ because $\sqrt{x}$ is monotone increasing finding the max of $\sqrt{x}$ and x is the same problem.

$\frac{\|\nabla f\|}{\partial x} = 2(2x + e^x - y)(2 + e^x) - 2(2y - x) = 0$

$4x + 2e^x - 2y + 2xe^x + e^{2x} - ye^x - 2y + x = 0$

$5x + 2e^x + 2xe^x + e^{2x} - \frac{4}{5}(4x + e^x) - e^x \cdot \frac{1}{5}(4x + e^x) = 0$

$\frac{\|\nabla f\|}{\partial y} = -2(2x + e^x - y) + 2(2y - x) \cdot 2 = 0$

$-2x - e^x + y + 4y - 2x = 0$

$5y = 4x + e^x \qquad y = \frac{1}{5}(4x + e^x)$

$\Rightarrow$ no solution in R $\Rightarrow$ no max $\|\nabla f\|$

Lets check on the boundaries:

$$\|\nabla f\|^2 = (2x + e^x - y)^2 + (2y - x)^2$$

$x = 2$ $\quad \|\nabla f\|^2 = (4 + e^2 - y)^2 + (2y - 2)^2$

In this case we observe $\|\nabla f\|^2$ is larger as y is larger (or smaller), so it is checked in $y = 2 \,|\, -2$

$x = -2$ $\quad \|\nabla f\|^2 = (-4 + \frac{1}{e^2} - y)^2 + (2y - 2)^2$

Again when $y \to -\infty, +\infty$ $\|\nabla f\|^2$ is the biggest so check for $y = -2 \,|\, 2$

$\|\nabla f(2,2)\|^2 = (2 + e^2)^2 + 4 = /$ $\quad \|\nabla f(-2,2)\|^2 = (-6 + e^{-2})^2 + 4 = 38,39$

$\|\nabla f(2,-2)\|^2 = (6 + e^2)^2 + 36 = \boxed{215,26}$ $\quad \|\nabla f(-2,-2)\|^2 = (-2 + e^{-2})^2 + 36 = 39,47$

$\sqrt{215,26} = 14,67 = L$

Finding β and α

- Eigenvalues of $\nabla^2 f$ β-smooth funct. lie on $[0, \beta]$
- Eigenvalues of $\nabla^2 f$, of α-strongly convex funct. are all $\geq \alpha$

$$\nabla^2 f = \begin{bmatrix} 2+e^x & -1 \\ -1 & 2 \end{bmatrix} \qquad \det(\nabla^2 f - \lambda I_2) = \begin{bmatrix} 2+e^x-\lambda & -1 \\ -1 & 2-\lambda \end{bmatrix} = 4 - 2\lambda + 2e^x - 2\lambda - \lambda e^x + \lambda^2 - 1 = 0$$

$$= \lambda^2 - \lambda(4 + e^x) + 2e^x + 3 = 0$$

$$\lambda_{1|2} = \frac{4 + e^x \pm \sqrt{16 + \cancel{8e^x} + e^{2x} - \cancel{8e^x} - 12}}{2} = \frac{4 + e^x \pm \sqrt{4 + e^{2x}}}{2}$$

The biggest λ is at $x = 2$ $\quad \beta = \frac{4 + e^2 + \sqrt{4 + e^4}}{2} = \boxed{9,522 = \beta}$

The smalles λ is at $x = -2$ $\quad \alpha = \frac{4 + e^{-2} - \sqrt{4 + e^{-4}}}{2} = \boxed{1,07 = \alpha}$

Prooving it is convex:

$$\begin{bmatrix} 2+e^x & -1 \\ -1 & 2 \end{bmatrix}$$

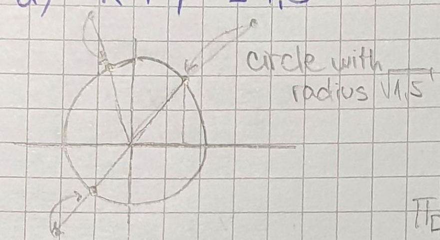
Sylvester thm:

- check $\det(|2 + e^x|) \geq 0$ ✓
- check $\det\left(\begin{vmatrix} 2+e^x & -1 \\ -1 & 2 \end{vmatrix}\right) = 4 + 2e^x - 1 = 3 + 2e^x \geq 0$ ✓

$\boxed{\text{Therefore } \nabla^2 f \text{ is PSD.}}$ $\quad e^x \geq 0$

3) Find $\Pi_d : \mathbb{R}^2 \to K$ for the following convex sets:

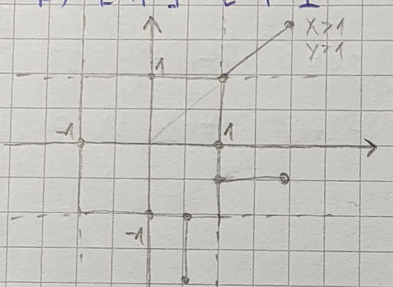
a) $x^2+y^2 \leq 1{,}5$



We can see that the newly projected coordinates x,y and the old should be lin. dependant!

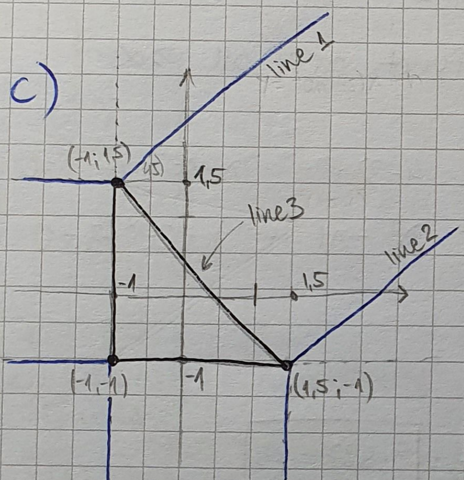
$$\Pi_D(\vec{x}) = \begin{cases} \vec{x} & \text{if } \vec{x} \in K \\ \dfrac{\vec{x}}{\|\vec{x}\|} \cdot \sqrt{1{,}5} & \text{else} \end{cases}$$

b) $[-1,1] \times [-1,1]$



$$\Pi_D(x,y) = \begin{cases} (x,y) & \text{if } (x,y) \in K \\ (1,y) & x>1 \quad |y| \leq 1 \\ (-1,y) & x<-1 \quad |y| \leq 1 \\ (x,1) & y>1 \quad |x| \leq 1 \\ (x,-1) & y<-1 \quad |x| \leq 1 \\ (1,1) & x>1 \quad y>1 \\ (-1,1) & x<1,\ y>1 \\ (-1,-1) & x<1,\ y<1 \\ (1,-1) & x>1,\ y<1 \end{cases}$$

c)



$$\Pi_D(x,y) = \begin{cases} (x,y) & \text{if } (x,y) \in K \\ (-1,y) & -1 \leq y \leq 1{,}5 \quad x<-1 \\ (x,-1) & -1 \leq x \leq 1{,}5 \quad y<-1 \\ (-1,-1) & x<-1 \quad y<-1 \\ (-1;1{,}5) & y>1{,}5 \ \wedge \ y>x+2{,}5 \\ (1{,}5;-1) & x>1{,}5 \ \wedge \ y<x-2{,}5 \\ \left(\frac{1}{2}(x-y+\frac{1}{2});\ \frac{1}{2}(\frac{1}{2}+y-x)\right) & \text{otherwise} \end{cases}$$

line 1: $1{,}5 = 1 \cdot (-1) + n$

$n = 2{,}5$ $\qquad \underline{y = x + 2{,}5}$

line 2: $k=1$ $\qquad -1 = 1{,}5 + n$ $\qquad n = -2{,}5$

$\underline{y = x - 2{,}5}$

line 3: $k = \dfrac{-1-1{,}5}{1{,}5+1} = \dfrac{-2{,}5}{2{,}5} = -1$ $\qquad y+1 = (-1)(x-1{,}5)$

$\underline{y = -x + 0{,}5}$

Slope of perpendicular will be: $k=1$

Say we have point (x_1, y_1) $\qquad \{ y_1 = x_1 + n \quad n = y_1 - x_1$

intersection of two lines

$$\begin{cases} y = x + (y_1 - x_1) \\ y = -x + 0{,}5 \end{cases}$$

$2y = 0{,}5 + (y_1 - x_1)$ $\qquad \underline{y = \frac{1}{2}(0{,}5 + y_1 - x_1)}$

$x = 0{,}5 - y$

$x = \frac{1}{2} - \frac{1}{4} - \frac{1}{2}y_1 + \frac{1}{2}x_1 = +\frac{1}{4} - \frac{1}{2}y_1 + \frac{1}{2}x_1 = \underline{\frac{1}{2}(x_1 - y_1 + \frac{1}{2})}$

4) Let $f(x,y) = x^2 + 2y^2$. $x_1 = (1,1)$

(a) What is the minimal value that can be achieved with one step of the GD i.e. find min of $f(x_2)$

$$x_2 = x_1 - \gamma \nabla f(x_1) = x_1 - \gamma \cdot \begin{bmatrix} 2x \\ 4y \end{bmatrix} = \begin{bmatrix} 1-\gamma 2x \\ 1-\gamma 4y \end{bmatrix} = \begin{bmatrix} 1-2\gamma \\ 1-4\gamma \end{bmatrix}$$

$$f(x_2) = (1-2\gamma x)^2 + 2(1-4\gamma x)^2$$

$$\frac{\nabla f(x_2)}{\partial \gamma} = 2(1-2\gamma)\cdot(-2) + 4(1-4\gamma)\cdot(-4) = -\cancel{4}(1-2\gamma) - \cancel{16}^{4}(1-4\gamma) = 0$$

$$-1+2\gamma = 4-16\gamma \qquad 18\gamma = +5 \qquad \gamma = \frac{5}{18}$$

$$f(x_2)\Big|_{\gamma=\frac{5}{18}} = \frac{2}{9}$$

b) How close can GD come to x^* in one step?

$$\nabla f = \begin{bmatrix} 2x \\ 4y \end{bmatrix} \qquad \begin{matrix} 2x=0 \\ 4y=0 \end{matrix} \qquad (x,y) = (0,0) = x^*$$

$$x_2 = \begin{bmatrix} 1-2\cdot\gamma \\ 1-4\cdot\gamma \end{bmatrix}$$

$$\|x_2 - x^*\| = \sqrt{(1-2\gamma)^2 + (1-4\gamma)^2}$$

Again, minimizing $\|x_2 - x^*\|$ is equal to minimizing $\|x_2 - x^*\|^2$

$$\|x_2 - x^*\|^2 = (1-2\gamma)^2 + (1-4\gamma)^2$$

$$\frac{\partial \|x_2 - x^*\|^2}{\partial \gamma} = 2(1-2\gamma)\cdot(-2) + 2(1-4\gamma)\cdot(-4) = -4 + 8\gamma - 8 + 32\gamma = 0$$

$$40\gamma - 12 = 0 \qquad 40\gamma = 12 \qquad \gamma = \frac{12}{40} = \frac{3}{10}$$

(given $\|x_2 - x^*\|^2$ is quadratic, and leading coef. is >0 then $\cup$ and $\gamma = \frac{3}{10}$ is min)

$$\|x_2 - x^*\|^2 = \sqrt{\left(1-\frac{6}{10}\right)^2 + \left(1-\frac{12}{10}\right)^2} =$$

$$= \sqrt{\frac{16}{100} + \frac{4}{100}} = \sqrt{\frac{20}{100}} = \frac{2\sqrt{5}}{10} = \boxed{\frac{\sqrt{5}}{5}}$$