1 Theoretical problems

1. Prove equivalence 1. 4. of Proposition 2.3 in notes.

   Proposition is:

$$f(\sum_{i=1}^{k}\alpha_i x_i) \leq \sum_{i=1}^{k}\alpha_i f(x_i) \mid \sum_{i=1}^{k}\alpha_i = 1$$

   which needs to be proven that it is same as (attribute of convex function):

$$f(tx+(1-t)y) \leq tf(x)+(1-t)f(y) \mid t \in [0,1]$$

   If we follow principles of induction:
   For (n=k=1) we have: $f(\alpha_1 x_1) \leq \alpha_1 f(x_1) \rightarrow \alpha_1 = t$
   For(k=n) we have:

$$f(\sum_{i=1}^{n}\alpha_i x_i) \leq \sum_{i=1}^{n}\alpha_i f(x_i)$$

$$f(\alpha_1 x_1 + ... + \alpha_n x_n) \leq \alpha_1 f(x_1) + ... + \alpha_n f(x_n)$$

$$f(\alpha_n x_n + \frac{(1-\alpha_n)}{(1-\alpha_n)}\sum_{i=1}^{n-1}\alpha_i x_i) \leq \alpha_n f(x_n) + (1-\alpha_n)\sum_{i=1}^{n-1}\frac{1}{1-\alpha_n}\alpha_i x_i$$

$$f(\alpha_n x_n + (1-\alpha_n)\sum_{i=1}^{n-1}\frac{1}{(1-\alpha_n)}\alpha_i x_i) \leq \alpha_n f(x_n) + (1-\alpha_n)\sum_{i=1}^{n-1}\frac{1}{1-\alpha_n}\alpha_i x_i$$

   note: we can see that this shape reminds us on function condition of convexity. With we can prove that this condition holds by induction as:

$$f(\sum_{i=1}^{n}\alpha_i x_i) \leq \sum_{i=1}^{n}\alpha_i f(x_i)$$

   Almost identical procedure is done for $k = n + 1$:

$$f(\sum_{i=1}^{n+1}\alpha_i x_i) \leq \sum_{i=1}^{n+1}\alpha_i f(x_i)$$

$$f(\alpha_1 x_1 + ... + \alpha_{n+1} x_{n+1}) \leq \alpha_1 f(x_1) + ... + \alpha_{n+1} f(x_{n+1})$$

$$f(\alpha_{n+1} x_{n+1} + (1-\alpha_{n+1})\sum_{i=1}^{n}\frac{1}{(1-\alpha_{n+1})}\alpha_i x_i) \leq \alpha_{n+1} f(x_n) + (1-\alpha_{n+1})\sum_{i=1}^{n}\frac{1}{1-\alpha_{n+1}}\alpha_i x_i$$

2. Let $f(x,y) = x^2 + e^x + y^2 - xy$. Function f restricted to $K = [-2,2] \times [-2,2]$ is Lipschitz, smooth and strongly convex. Find some corresponding (preferably optimal) constants L, α and β on K. Furthermore, prove f is convex.

$$\nabla f = \begin{bmatrix} 2x + e^x - y \\ 2y - x \end{bmatrix}$$

$\frac{\partial^2 f}{\partial x^2} = 2 + e^x$
$\frac{\partial^2 f}{\partial y^2} = 2$
$\frac{\partial^2 f}{\partial x \partial y} = -1$

Lipschitz constant:

$|f(x,y_1) - f(x,y_2)| = |y_1^2 - y_2^2 + x(y_1 - y_2)|$
$|(y_1 - y_2)(y_1 + y_2) + x(y_1 - y_2)| = |(y_1 - y_2)(x + y_1 + y_2)|$
$\leq |y_1 - y_2||x + y_1 + y_2| \leq |y_1 - y_2|(|x| + |y_1| + |y_2|)$

If you take in consideration K which is square on the graph we can say that K is essentially constraint in form of $|x| \leq 1$, $|y| \leq 1$. Thus:

$|f(x,y_1) - f(x,y_2)| = 6|y_1 - y_2|$ With this we have $L = 6$

$\beta - smooth$ convex function is defined as :
$\|\nabla f(x) - \nabla f(y)\| \leq \beta\|x - y\| \rightarrow \|\nabla^2 f(x,y)\| \leq \beta$

Having in mind that:

$$\nabla^2 f = \begin{bmatrix} 2 + e^x \\ 2 \end{bmatrix}$$

we can write down:

$\sqrt{(2+e^x)^2 + 2^2} \leq \beta \,|\, x \in [-2,2]$ eigenvalues for $\nabla^2 f \rightarrow \lambda_1 = 2\lambda_2 \approx 2.13$ or high value with $x = 2$ is $\lambda_2 \approx 9.38$.

$\|\nabla^2 f(x,y)\| \leq \beta \rightarrow \beta \approx 9.599$

For α - smoothness (strong convexity): $\nabla^2 f \geq \alpha I \rightarrow \begin{bmatrix} 2 + e^x \\ 2 \end{bmatrix} \geq \begin{bmatrix} \alpha = 9.389 \\ \alpha = 2 \end{bmatrix}$

3. Find formulas for projections $R^2 \to K$ to the closest point for the following convex sets: $x^2+y^2 \leq 1.5$ $[-1,1]\times[-1,1]$; and the triangle with vertices $(-1,-1), (1.5,-1), (-1,1.5)$.

$a)(x^2+y^2 \leq 1.5:$

We can represent this problem with projection onto circle as intersection approach. If we create line from $(x,y) \in \mathbb{R}^2$ to center of circle $(0,0)$ we can find projection of this point as intersection of circle and this line. Thus:

$$Slope \to k = \frac{y_2 - y_1}{x_2 - x_1} \to k = 1.5$$
$$y - y_1 = k(x - x_1) \to y = 1.5x$$

Example: $(x,y) = (2,3)$ From circle equation we get: $x_{1,2} = \pm 0.774$; $y_{1,2} = \pm 1.16$

With this we are considering solution (projection point) closest to considered point:

$argmin(z - x) \mid z \in C, x \in \mathbb{R}^2 \mid C \subset \mathbb{R}^2$.

In this case it is $proj(x,y) = (0.774, 1.16)$.

$b)ic):$

We can look these problem as intersection problems between line segments. If we take into consideration that we have areas of xy plane defined by line segments we can project our (x,y) point to closest line segment as intersection of line segment with a line defined passing through two points; our (x,y) point and a point which is part of line segment. Forming this line we will have a line which is normal to line segment and thus representing shortest distance to that line. We can define our two lines to begin with as being connected between points P_1,P_2 for the set line and P,P_L (through the our point and normal point) for the second. The equations of these lines are defined simply as:

$$P_1 + t(P_2 - P_1)$$
$$P + t(P_L - P)$$
$$x_1 + t_a(x_2 - x_1) = x + t_b(x_L - x)$$
$$y_1 + t_a(y_2 - y_1) = y + t_b(y_L - y)$$

$$t = -\frac{(x_1 - x)(x_2 - x_1) + (y_1 - y)(y_2 - y_1)}{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

If $t \in [0,1]$ then perpendicular line is intersecting one of our set lines and we can calculate minimum distance as:

$$d = \frac{|(x_2 - x_1)(y_1 - y) - (y_2 - y_1)(x_1 - x)|}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$

We are going to repeat this procedure for all of our lines which are defining our set (i.e. square and triangle in our case). With this approach we are going to have the closest projected point on our set.

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

   (a) What is the minimal function value that can be achieved with one step of the gradient descend, i.e., find the minimum of $f(x_2)$.

   (b) How close to the actual minumum x^* of function f can we get with one step of the gradient descend, i.e., find the minimum of the distance from x^* to x_2.

$$\nabla f = \begin{bmatrix} 2x \\ 4y \end{bmatrix}$$

$$H_f = \begin{bmatrix} 2 & 0 \\ 4 & y \end{bmatrix}$$

$a)$:

We can look at this problem as finding minimum of gradient descent formula as :

$$\phi(\gamma) \to x_{k+1} = x_k - \gamma \nabla f(x, y)$$

$$\phi(\gamma) = f(x_k - \gamma \nabla f(x, y))$$

$$\phi'(\gamma) = \nabla f(x_k - \gamma \nabla f(x, y)) \mid x_1 = (1, 1)$$

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

$$\nabla f(1 - 2\gamma, 1 - 4\gamma) = \begin{bmatrix} 2 - 4\gamma \\ 4 - 16\gamma \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

$$\gamma = 0.416 \to \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0.168 \\ -0.664 \end{bmatrix}$$

$b)$:

We can look at this problem as:

$$d(x^*, x_2) = \sqrt{(y^* - y_2)^2 + (x^* - x_2)^2} = \sqrt{(y^*)^2 - 2y^* y_2} = 0.6849 \mid (x^*, y^*) = (0, 0)$$

$$2dd' = 2y_k y_k' + 2x_k x_k'$$

$$\begin{bmatrix} x_k \\ yk \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \gamma \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

$$x_k = 1 - 2\gamma \to x_k' = -2$$

$$y_k = 1 - 4\gamma \to y_k' = -4$$

$$d' = \frac{40\gamma - 12}{2d} = 0 \rightarrow \gamma = 0.33$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0.4 \\ -0.2 \end{bmatrix}$$