Problem Set Week 4

Introduction to Continuous Optimization and Convex Sets

Shekhar Kumar

Exercise 6.6

Since, $f(x, y) = 3x^2y + 4xy^2 + xy$
We know that the critical points are achieved by taking first differential ,i.e.,

$$Df_x(x, y) = 0, i.e., 6xy + 4y^2 + y = 0, \tag{1}$$

and

$$Df_y(x, y) = 0, i.e, 3x^2 + 8xy + x = 0 \tag{2}$$

So these give us four following possibilities when we rewrite the equations above as, $y(6x + 4y + 1) = 0$ and $x(3x + 8y + 1) = 0$:

$$x = 0 \quad and \quad y = 0 \tag{3}$$

$$6x + 4y + 1 = 0 \quad and \quad 3x + 8y + 1 = 0 \tag{4}$$

$$6x + 4y + 1 = 0 \quad and \quad x = 0 \tag{5}$$

$$3x + 8y + 1 = 0 \quad and \quad y = 0 \tag{6}$$

Solving for the above equations (3) to (6), we get 4 sets of points as

$(0, 0), (0, -1/4), (-1/3, 0), (-1/9, -1/2)$.

Other than these points the critical points do not exist as the solution for the first differential equal to zero are these only.

Solving for the second order differentiation of $Df_x(x, y)$ and $Df_y(x, y)$, we get a $2 * 2$ matrix as follows:

Solving for terms within the matrix , we get the above matrix as :

$$\begin{bmatrix} 6y & 6x + 8y + 1 \\ 6x + 8y + 1 & 8x \end{bmatrix}$$

After getting the value of the matrix at all of the above four sets of points we see whether the condition of $D^2 f(x) > 0$ is satisfied or not for local minima, i.e., whether $f_{xx} + f_{yy} > 0$ and $f_{xx} f_{yy} - f_{xy}^2 > 0$

We see that none of the points satisfy the conditions above . Hence none of them is a minima. But then we have the condition for a local maxima as $f_{xx} + f_{yy} < 0$ and $f_{xx} f_{yy} - f_{xy}^2 > 0$.

We see that this condition is satisfied by (-1/9,-1/12) and hence it is the local maxima. The other four points are the saddle points.

Exercise 6.7

1. $f(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} - \mathbf{b}^T \mathbf{x} + c$ Let A be the square matrix $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and X be the matrix $\begin{bmatrix} x1 \\ x2 \end{bmatrix}$ so the $\mathbf{x}^T A \mathbf{x}$ is giving the result as $a_{11}x_1^2 + a_{21}x_2x_1 + a_{12}x_1x_2 + a_{22}x_2^2$. When we do $\mathbf{x}^T \mathbf{A}^T \mathbf{x}$, we get the same result. Meaning that both $\mathbf{x}^T \mathbf{A}^T \mathbf{x}$ and $\mathbf{x}^T A \mathbf{x}$ are the same and hence $\mathbf{x}^T Q \mathbf{x} = 2\mathbf{x}^T A \mathbf{x}$.

   Hence the above f($\mathbf{x}$) can be rewritten as $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q \mathbf{x} - \mathbf{b}^T \mathbf{x} + c$.

2. If minima of the quadratic exists then the first derivative of the quadratic at the minima will be zero. Differentiating $f(x)$ gives

$$\frac{1}{2}(x^T Q + x^T Q^T) = b^T \tag{6.7.1}$$

   Since from 1 above we have $Q^T = Q$. This shows that

$$x^T Q = b^T$$

   Taking transpose of both sides, we get our result $\mathbf{Q^T x^*} = \mathbf{b}$

3. As shown in the 2 above if the derivative of the quadratic is equated to zero then we get the first order necessary condition as $Q^T x^* = b$. The necessary and sufficient condition for a minima to exist is that the second derivative of the equation has to be positive definite.

   Differentiating (6.7.1) again we get its' derivative as Q. Therefore, if Q is positive definite then the quadratic will have a minima.

Problem 6.11

The Newton's method has the following iterative algorithm to identify the minimizer:

$$x_{t+1} = x_t - \frac{f^{'}(x_k)}{f^{"}(x_k)}$$

Inserting the value of $f^{'}(x_k)$ and $f^{"}(x_k)$ in the above equation gives

$$x_1 = x_0 - \left(\frac{2ax_0 + b}{2a}\right)$$

Therefore, the process converges in 1 iteration and x^* = -$\frac{b}{2a}$

Problem 6.15

See the Python code and answer in the included Jupyter notebook

Problem 7.1

Let $x, y \in conv(S)$. Further x,y can be written as $x = \sum_{i=1}^{k} \lambda_i a_i$, and $y = \sum_{i=1}^{k} \mu_i a_i$, where $\sum_{i=1}^{k} \lambda_i = 1$ and $\sum_{i=1}^{k} \mu_i = 1$.

If we are able to show that $\alpha x + (1 - \alpha)y \in conv(S)$ for $\alpha \in [0, 1]$ then we would have proven that $conv(S)$ is convex.

$$\alpha x + (1-\alpha)y = \alpha \sum_{i=1}^{k} \lambda_i a_i + (1-\alpha) \sum_{i=1}^{k} \mu_i a_i$$

RHS can be written as :

$$\sum_{i=1}^{k} (\alpha\lambda_i + (1-\alpha)\mu_i) a_i$$

It can be seen that $\sum_{i=1}^{k} (\alpha\lambda_i + (1-\alpha)\mu_i) = 1$. Thus, $\alpha x + (1-\alpha)y \in conv(S)$.

Therefore, $conv(S)$ is convex.

Problem 7.2

1. Let $P = \{x \in V | \langle a, x\rangle = b\}$, be a hyperplane. For $x, y \in P$, if

   $$\langle a, \lambda x + (1-\lambda)y\rangle = \lambda\langle a, x\rangle + (1-\lambda)\langle a, y\rangle = b \ \forall \lambda \in [0, 1]$$

   Using property of inner product spaces for real numbers it is easy to show that the hyperplane **P** will be a convex set.

   Thus, $\lambda x + (1-\lambda)y \in P$.

2. Let $H = \{x \in V | \langle a, x\rangle \leq b\}$, be a hyperplane. For $x, y \in H$, if

   $$\langle a, \lambda x + (1-\lambda)y\rangle = \lambda\langle a, x\rangle + (1-\lambda)\langle a, y\rangle \leq b \ \forall \lambda \in [0, 1]$$

   Using property of inner product spaces for real numbers it is easy to show that the hyperplane **H** will be a convex set.

   Thus, $\lambda x + (1-\lambda)y \in H$.

Problem 7.4

1. We can write the LHS of the given equation as

$$\|x-y\|^2 = \|(x-p)+(p-y)\|^2 = \langle (x-p)+(p-y),(x-p)+(p-y)\rangle$$

   Using the rules for inner product spaces for real numbers we can write RHS of above equations as

$$\langle (x-p)+(p-y),(x-p)+(p-y)\rangle = \|x-p\|^2 + \|p-y\|^2 + 2\langle x-p, p-y\rangle$$

2. We are given that $\langle x-p, p-y\rangle \geq 0 \ \forall y \in C$

   Using the result shown in 1 above. It is clear that $\|x-y\| > \|x-p\|$

3. The given norm can be written as shown below

$$\|x-z\|^2 = \|x-\lambda y-(1-\lambda)p\|^2$$

$$= \|x-p\|^2 + 2\lambda\langle x-p,(p-y)\rangle + \lambda^2\|(y-p)\|^2$$

4. Using (7.15), and setting $z = y$. Then, using (7.15) and result in 2 above, we get

$$0 \leq \|x-y\|^2 - \|x-p\|^2 = 2\lambda\langle x-p, p-y\rangle + \lambda^2\|y-p\|^2$$

   Dividing the whole equation by λ gives us our result for the equation given then

$$0 \leq 2\langle x-p, p-y\rangle + \lambda\|y-p\|^2$$

   Since the above equation holds for every λ and $y \in C$, setting $\lambda = 0$ gives us our result $\langle x-p, p-y\rangle \geq 0$

subsection*Problem 7.8 We are given the function $g(x) = f(Ax+b)$

$$g(\lambda x + (1-\lambda)y) = f(A(\lambda x + (1-\lambda)y) + b)$$

RHS can be further written as:

$$f(A(\lambda x + (1-\lambda)y + b) = f(\lambda(Ax+b) + (1-\lambda)(Ay+b))$$

Since it is given that f is convex, we can use its' property. Which gives us

$$f(\lambda(Ax+b) + (1-\lambda)(Ay+b)) \leq \lambda f(Ax+b) + (1-\lambda)f(Ay+b)$$

which can be further written as

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

Hence we get our result.

Problem 7.12

1. Let $A, B \in PD_n(\mathbb{R})$. A matrix,M is positive definite if for any non-zero vector x, $x^T Mx > 0$

   $$x^T(\lambda A + (1-\lambda)B)x = \lambda x^T Ax + (1-\lambda)x^T Bx$$

   Since $x^T Ax > 0, x^T Bx > 0, \lambda x^T Ax + (1-\lambda)x^T Bx > 0$. Thus the combination also $\in PD_n(\mathbb{R})$

2. We prove the following one by one to get the main result.

(a) We are given that the function $g(t) : [0, 1] \to \mathbb{R}$ given by $g(t) = f(tA + (1-t)B)$ is convex. From the chapter, we have the result that if f is a real-valued function on a convex set $C \subset \mathbb{R}^n$, then

$$f \text{ is convex } \Leftrightarrow$$

$$\forall x_1, x_2 \subset C, \text{ the map } g : [0,1] \to \mathbb{R} \text{ given by } g(t) = f(tx_1 + (1-t)x_2 \text{ is convex.}$$

Since from the previous result we have $PD_n(\mathbb{R})$ is convex, substituting C by $PD_n(\mathbb{R})$ and x_1, x_2 by $A, B \in PD_n(\mathbb{R})$, we have the result.

(b) To prove $f(X) = -\log(\det(X))$ is convex. Let $A = S^H S$. Hence,

$$f(tS^H S + (1-t)B)$$

$$= f(tS^H IS + (1-t)(S^H)^{-1}(S^H)BS^{-1}S)$$

$$= f(S^H(tI + (1-t)(S^H)^{-1}BS^{-1})S)$$

Let $X = S^H, Y = tI + (1-t)(S^H)^{-1}BS^{-1}, Z = S$

The RHS can be written as g(t) =

$$= -\log(\det(XYZ))$$

$$= -\log(\det(X)\det(Y)\det(Z))$$

$$= -\log(\det(XZ)\det(Y)$$

$$= -\log(\det(A)\det(Y))$$

$$= -\log(\det(A)) - \log(\det(Y))$$

$$= -\log(\det(A)) - \log(\det(tI + (1-t)(S^H)^{-1}BS^{-1}))$$

Hence part (b) is proved

(c) We now prove the next part. Let $\lambda_1, \lambda_2 ... \lambda_n$ be the Eigenvalues of $(S^H)^{-1}BS^{-1}$. Then the eigenvalues of Y are $t+(1-t)\lambda_1, t+(1-t)\lambda_2 ... t+(1-t)\lambda_n$.Since the determinant is the product of the eigenvalues and since log of product gives the sum of the logs, we have the result.

(d) To prove this we differentiate $g(t)$ twice and obtain

$$g''(t) = \sum_{i=1}^{n} \frac{(1-\lambda_i)^2}{(t+(1-t)\lambda_i)^2} \geq 0 \forall t \in [0,1]$$

The proof of the main result follows from the theorem in the chapter since $g''(t) \geq 0 \Rightarrow g$ is convex which in turn implies that $f(X)$ is convex.

Problem 7.13

Suppose function f is not a constant and $\exists\ x' \in \mathbb{R}^n$ such that $f(x') \neq f(x)$. Let the value of the bound be M. Finally, suppose $f(x) \leq M$ for all x. We are given that f is convex, therefore

$$\begin{aligned} f(x_1) &= f\left(\lambda \frac{x_1 - (1-\lambda)x_2}{\lambda} + (1-\lambda)x_2\right) \\ &\leq \lambda f\left(\frac{x_1 - (1-\lambda)x_2}{\lambda}\right) + (1-\lambda)f(x_2) \end{aligned}$$

Rearranging the terms in above inequality, we can rewrite it as

$$\frac{f(x_1) - (1-\lambda)f(x_2)}{\lambda} \leq f\left(\frac{x_1 - (1-\lambda)x_2}{\lambda}\right) \leq M \tag{7}$$

We are given that f is bounded from above. Therefore,

$$\frac{f(x_1) - f(x_2)}{\lambda} \leq \frac{f(x_1) - (1-\lambda)f(x_2)}{\lambda} \tag{8}$$

Which implies

$$\frac{f(x_1) - f(x_2)}{\lambda} \leq M \tag{9}$$

As λ go to 0, the term on the LHS grows without bound, which contradicts the fact that f is bounded, therefore f has to be a constant function.

Problem 7.20

We are given that f and $-f$ both are convex. Hence, we know that

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

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

Combining the two equations gives us

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

The above result shows that f is a linear transformation in $\mathbb{R}^n$ as shown in the chapter on inner product spaces. Therefore, $f(x)$ can be written as a linear transformation where as per Definition 7.4.1, **c** is **0** and this implies f is an affine function.

Problem 7.21

Let's prove the if part first. We are given that x^* is a local minimizer of $f(x)$. This implies that in a small δ radius neighborhood around x^*, $f(x^*) \leq f(x)$. we are given that function $\phi \ o \ f(x)$ is an increasing function. Therefore, it will preserve the inequality relationship for $f(x)$. Hence we get the if side of the result.

Let's prove the only if part of the result. We are given that $\phi \ o \ f(x^*) \leq \phi \ o \ f(x)$ for ϵ radius around $f(x)$. Aain using the property that ϕ function is an increasing function we get $f(x^*) \leq f(x)$. Therefore x^* is a local minimizer of $f(x)$.