Assignment 2

Optimization

Student Number : 40178580

CAT I Problems

1.1

We know that the gradient of a vector is the partial derivatives of each component of x:

$$\nabla f(x) = \frac{\delta}{\delta x_1}(c^T x_1)$$

Since c is a constant:

$$\nabla f(x) = \frac{\delta}{\delta x_1}(c^T x_1) = c^T \frac{\delta}{\delta x_1}(x_1)$$

The derivative of x_1 with respect to itself is simply 1, and so we get:

$$\frac{\delta}{\delta x_1}(c^T x_1) = c^T$$

Now, when we repeat this for all elements of X (x1, x2, x3,…xn) we will get:

$$\nabla f(x) = [c_1, c_2, c_3, \dots\ c_n]^T$$

The transpose will simple be C, thus:

$$\nabla f(x) = c$$

1.2

The gradient will simply be :

$$\frac{\partial}{\partial xi} = 1$$

As we are differentiating each component of X with respect to itself. The resulting gradient of the sum will be a vector of length n consisting of 1's. Ie:

$$\nabla f = [1, 1, 1, \dots 1]$$

We will get the same array for h, as again we are differentiating each component of x with respect to itself, however this time the vector will be 2n in length.

$$\nabla h = [1, 1, 1, \dots 1]$$

1.3

$$f1(x) = 0.5x^T Q x + q^T x + c$$

We need to calculate the 2nd partial derivitves of f1 and orginise into the following matrix:

$$H = \begin{matrix} \frac{\partial 2f1}{\partial {x^2}_1} & \frac{\partial 2f1}{\partial x_1 \partial x_2} & \frac{\partial 2f1}{\partial x_1 \partial x_n} \\ \frac{\partial 2f1}{\partial x_2 \partial x_1} & \frac{\partial 2f1}{\partial {x^2}_2} & \frac{\partial 2f1}{\partial x_2 \partial x_n} \\ \frac{\partial 2f1}{\partial x_n \partial x_1} & \frac{\partial 2y}{\partial x_n \partial x_2} & \frac{\partial 2f1}{\partial {x^2}_n} \end{matrix}$$

1. For the diagonal elements ($\frac{\partial 2f1}{\partial {x^2}_1}$)

$$\frac{\partial 2f1}{\partial {x^2}_1} = \frac{\partial}{\partial xi}(0.5x^T Qx + \ q^T x + c)$$

$$\frac{\partial 2f1}{\partial {x^2}_1} = \ Q_{ii}$$

2. For the non-diagonal elements ($\frac{\partial 2f1}{\partial x_i \partial x_j}$)

$$\frac{\partial 2f1}{\partial x_i \partial x_j} = \ \frac{\partial}{\partial xi}(0.5x^T Qx + \ q^T x + c)$$

Since Q is symmetric

$$\frac{\partial 2f1}{\partial x_i \partial x_j} = \ \frac{\partial 2f1}{\partial x_j \partial x_i}$$

And

$$\frac{\partial 2f1}{\partial {x^2}_1} = \ Q_{ii}$$

We can assemble the matrix :

$$H = \begin{matrix} Q_{11} & Q_{12} & Q_{1n} \\ Q_{21} & Q_{22} & Q_{2n} \\ Q_{n1} & Q_{n2} & Q_{nn} \end{matrix}$$

This the Hessian on f1 is simply H = Q.

We can rewrite the 2nd equation as:

$$f2(x) = 0.5(Ax - b)^T Q(Ax - b))$$

Now lets calculate 2nd partial derivatives, starting with diagonals:

$$\frac{\partial 2f2}{\partial {x^2}_1} = \ \frac{\partial}{\partial xi}(0.5(x^T A^T - b^T)\big(Q(Ax - b)\big) \rightarrow First\ derivative$$

Now we use the chain rule to find the solution, we get:

$$\frac{\partial}{\partial xi}(0.5(x^T A^T - b^T)\big(Q(Ax - b)\big) \\ = 0.5\left(\frac{\partial}{\partial xi}(x^T A^T - b^T)\right)Q(Ax - b) + 0.5(x^T A^T - b^T)Q(\frac{\partial}{\partial xi}(Ax - b))$$

Now since A and b are constants, the derivatives are 0.

$$\frac{\partial}{\partial xi}(Ax - b) = Ai$$

$$So\ now\ we\ have:$$
$$0.5(A_i^T - 0)Q(Ax - b)$$

Which is

$$0.5(A_i^T Q(Ax - b)$$

The 2nd term in the main equation also has $\frac{\partial}{\partial xi}(Ax - b) = Ai$ which hwe know is Ai and we get:

$$0.5(x^T A^T - b^T)QAi$$

Putting it all together:

$$\frac{\partial 2f2}{\partial x^2{}_1} = 0.5(A_i^T Q(Ax - b) + 0.5(x^T A^T - b^T)QAi$$

Now, moving onto the non-diagonal elements ($\frac{\partial 2f1}{\partial x_i \partial x_j}$) and again using chain rule.

$$\frac{\partial 2f1}{\partial x_i \partial x_j} = 0.5(\frac{\partial}{\partial xi}(x^T A^T - b^T)A_i^T Q(Ax - b) + 0.5(x^T A^T - b^T)Q(\frac{\partial}{\partial xj}(Ax - b)$$

A and B are constants so

$$\frac{\partial 2f1}{\partial x_i \partial x_j} = 0$$

So the overall Hessian matrix would be a matrix that has H = $0.5(A_i^T Q(Ax - b) + 0.5(x^T A^T - b^T)QAi$ on the diagonal elements. The non-diagonal would all be 0.

1.4

$$f(x) = \|x\| + c^T + 8088$$

The constraint is $\sum_{i=1}^{n} xi < a$

1) Looking at each individual component of the function, first we are dealing with some norm of x, which is convex by nature. The 2nd term cTx is a linear operator and thus is also convex. The 3rd term is a constant and will have no effect on the convexity of the overall function. This means the function itself is convex since there are no non-convex elements.
2) Looking at the constraint, which is a linear inequality which is convex in nature and thus does not effect the confexity established in (1)

CAT II

2.1

As before, we are dealing with the square of the Euclidean norm which is a convex function. cTx is still a linear function. And the constant does not effect convexity.

Now the constraint is different, with the norm being constrained to a value of 5 representing a sphere of radius 5. The constraint is NOT convex because the set of points that satisfy this constraint would not be convex. Since the constraint is not convex, the overall function is not convex.

2.2

A function is convex if and only if, for any two points x and y in its domain and for any value λ in the interval [0 1] the following inequality holds true:

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

Now we can prove convexity of g(x) = f(Ax+b)

1) Take 2 arbitrary points x and y in the domain of g
2) Consider a new point z which falls in the interval defined by λ ([0 1]) where z = λx + (1-λ)y
3) Now we can look at g(x) , g(y) and g(z)

$$g(z) = f(Az+b)$$
$$g(x) = f(Ax+b)$$
$$g(y) = f(Ay+b)$$

G(z) comes from the definition of g(x), g(x) is one point and g(y) is the new point.

Using convexity properties of f(x) and simplifying , we get

$$f(\lambda(\mathrm{Ax}+\mathrm{b}) + (1-\lambda)(\mathrm{Ay}+\mathrm{b})) = \mathrm{f}(\mathrm{Ax}+\mathrm{b}) + (1-\lambda)(\mathrm{Ay} + (1-\lambda)\mathrm{b})$$

Now using matrix multiplication and vector addition we can rewrite as :

$$f(\mathrm{A}\,(\lambda\mathrm{x} + (1-\lambda)\mathrm{y}) + b(\lambda + (1-\lambda)$$

Using the properites of convexity again we get:

$$f(\mathrm{A}\,(\lambda\mathrm{x} + (1-\lambda)\mathrm{y}) + b(\lambda + (1-\lambda)$$
$$\leq \mathrm{f}(\mathrm{Ax}+\mathrm{b}) + (1-\lambda)\mathrm{f}(\mathrm{Ay}+b)$$

Subbing back into inequalities:

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

Since the inequality holds for random x, y and λ. G(x) is a convex function.

II)

Following the same process as before, considering random x,y and λ values. We get :

$$g(z) = max(Af1(z), f2(z), \dots fp(z))$$
$$g(x) = max(Af1(x), f2(x), \dots fp(x))$$
$$g(y) = max(Af1(y), f2(y), \dots fp(y))$$

Using the convexity property for each f:

$$fi(\lambda \ + (1 - \lambda)\mathrm{y}) \leq \ \lambda \mathrm{fi(x)} + (1 - \ \lambda)\mathrm{fi(y)}$$

Now we consider the maximum values of fi$(\mathrm{fi}(\lambda \mathrm{x} + (1 - \lambda)\mathrm{fi(y)} \mid 1 \ \leq i \leq p$

Since g(z) is the max of value fi(z), we can rewrite it as:

$$g(\lambda \mathrm{x} \ + (1 - \lambda)\mathrm{y}) \leq \max(\lambda fi(x) + (1 - \ \lambda)fi(y)| \ 1 \leq i \leq p$$

Now, using the following property of the max:

$$\max(a1, a2, ap) \leq max(b1, b2, bp)$$

For all ai <= bi and 1 < I < p

Therfor we can say further simplify the inequality above to:

$$g(\lambda \mathrm{x} \ + (1 - \lambda)\mathrm{y}) \leq \max(\lambda fi(x) + (1 - \ \lambda)fi(y). \ \leq \ \lambda \max(\mathrm{fi(x)} + (1 - \lambda)\max(\mathrm{fi(y)})|$$

Since this inequality holds, g(x) is convex.

iii)

Both f1*f2 and f1-f2 are not guaranteed to be convex.

The product can be convex if both f1(x) and f2(x) are non-negative and increasing functions over the entire domain.

The product can be non-convex if either f1 or f2 change sign and is not strictly increasing.

Eg f1(x) = x and f2(x) = -x, f1*f2 = x^2

The difference can be convex if f1(x) is convex and f2 is concave.

Convesrly, if both f1 and f2 are convex, then the resulting function will not be convex

Eg f1 = x^2 and f2 = x   f1 – f2 = x^2 -x

2.3

We use a function with oscillatory behaviour. Consider the function

$$f(x) = \sin(x) + x$$

This function has infinitely many stationary points, where the deriviative is equal to 0. To find these points we can set the deriviative of f(x) equal to 0 and solve for x

$$\frac{df}{dx} = \cos(x) + 1 = 0$$

Here x is pi + 2 (pi) k, where k is an integer. This is where the stationary points occur.

Looking at the stationary points, the derivative will always be positive for all x because the range of cosine is [-1, 1] and when we add one we get [0, 2]

Since the derivative is always positive, it means f(x) is always increasing, therfor none of the stationary points at 2(pi)k is a local and global minimum because the values keep increasing.

So we have a continuously differentiable function with infinite stationary points . None of whuch are a minimizer.

2.4

We need a function that oscillates between -1 and 1 without converging to a maximum value. Consider f(x) = sin(x)

This doesn't have a minimum value in the sense that it doesn't converge.

2.5

3.1

We need to show that that for any two points x and y in the lvel set, the line segment joining connecting the points are contained within the level set itself.

Let x and y be two arbitrary points in the level set lev<α f, meaning that f(x) < α and f(y) < α. Now consider λ in the interval [0 1]. We want to show that the point z = $\lambda\text{x} + (1-\lambda)\text{y}$ is also in the lvel set f(z) < α

Because f is convex, it satisfies:

$$f(\lambda\text{x} + (1-\lambda)\text{y}) < \lambda\text{f}(\text{x}) + (1-\lambda)\text{f}(\text{y})$$

We can use this property to show that:

$z = \lambda x + (1-\lambda)y$ is in fact in the level set.

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

f(x) < α and f(y) < α , we can replace them in the inequality:

$$f(z) = \lambda\,\alpha + (1-\lambda)\alpha = \alpha$$

Now we have shown that f(z) is less than or equal to α, meaning z = $\lambda x + (1-\lambda)y$ is in the level set.

Since this holds, we can conclude that the level sets of f are convex sets for all alpha.

ii)

We can consider the function f(x) = - x^4

This function is non-convex as its 2nd deriviative is negative (-12x^2). . If we look at the level sets for this function, they are convex sets for all alpha.

X^4 >= - α

Taking the 4th root, x>= α^$(\frac{1}{4})$ which means the level set for any α is a symmetric interval around the origin.

III)

To show that f is convex if and only if its epigragh is a convex set, we need to prove both directions.

1. If f is convex, then the epi f is a convex set.

Let (x1, $\alpha 1$) and (x2, $\alpha 2$) be two points in epi f. This means that f(x1) <= $\alpha 1$ and f(x2) <= $\alpha 2$

Now we repeat and consider λ in the interval [0, 1]. We want to show that a new point (z, b) is in epi f.

By the definition of convexity, for any x valie we have:

$$f(z) = f(\lambda x1 + (1-\lambda)x2) < \lambda f(x1) + (1-\lambda)f(x2)$$

Since f(x1) < $\alpha 1$ and f(x2) < $\alpha 2$:

$$b = f(\lambda x1 + (1-\lambda)x2) < \lambda f(\alpha 1) + (1-\lambda)f(\alpha 2)$$

Which implies that (z, b) is in epi f

2. If epi f is convex, then f is convex

Suppose epi f is a convex set

Repeat process

$$f(\lambda x1 + (1-\lambda)x2) < \lambda f(x1) + (1-\lambda)f(x2)$$

For any $\alpha 1$ = f(x1) and $\alpha 2$ = f(x2), the points (x1, $\alpha 1$) and (x2, $\alpha 2$) are in epi f.

Because epi f is convex, the point

$$\lambda x1 + (1-\lambda)x2 \text{ and } \lambda\alpha 1 + (1-\lambda)\alpha 2$$

Must also be in epi f. This gives:

$$f(\lambda x1 + (1-\lambda)x2) < \lambda f(\alpha 1) + (1-\lambda)f(\alpha 2) =< \lambda f(x1) + (1-\lambda)f(x2)$$

Therefor, we have shown that if epi f is in a convex set, then f is a convex function.

3.2

KKT system can be written as

$$Qx* + A^T\lambda = -q$$
$$Ax* = b$$

1. If K is non-singular, then ker(Q) $\cap$ ker(A) = 0. Lets assume K is non-singular, this means that the KKT system has a unique solution for any given Q, A, q and b values.

To prove that ker(Q) $\cap$ ker(A) = 0, we consider the following:

Suppose there is a non-zero vector 'v' in both ker(Q) and ker(A), meaning Qv = 0 and Av = 0.

If we consider the system:

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{matrix} x* \\ \lambda \end{matrix} = \begin{matrix} -q \\ b \end{matrix}$$

Multiply the first block by v

$$Qvx* + A^T v\lambda = -qv$$
$$\text{Since Qv} = 0 \text{ and Av} = 0$$
$$-qv = 0$$

This implies that qv = 0 but we assumed that v is non-zero, so this cannot be true. We can conclude that there cannot be a non-zero vector in both ker(A) and ker(Q). Hence ker(Q) $\cap$ ker(A) = 0

Now, if ker(Q) $\cap$ ker(A) = 0, then K is non-singular. Lets assume the KKT system has a non-zeor solution (x, λ), then we have :

$$Qx* + A^T\lambda = -q$$

$$Ax = b$$

We can say that Ax = 0. If we multiply 1st equation by $x *^T$ we get

$$x *^T (Qx * + q) + x *^T A^T \lambda = 0$$

Rearranging:

$$x *^T Qx * + x *^T q + x *^T A^T \lambda = 0$$

We know Ax = 0, so $A^T \lambda = 0$

$$x *^T Qx * + x *^T q = 0$$

Now consider the vector v [x, 0]

Apply it to the system:

$$v^T K v = \begin{bmatrix} x *^T Qx * + x *^T q & 0 \\ 0 & 0 \end{bmatrix}$$

Thus $v^T K v$ is zero for all non-zero solutions v. Therfor the KKT matrix is singular.

So, the KKT matric is non-singular if and only if ker(Q) $\cap$ ker(A) = 0