

Problem Set 4

Convex Sets and Convex Functions

Keshav Choudhary ¹

Problem 6.6

To find the critical points, we calculate f_x, f_y and set these equal to zero.

$$\begin{aligned}f(x, y) &= 3x^2y + 4xy^2 + xy \\f_x(x, y) &= 4y^2 + 6xy + y \\f_y(x, y) &= 3x^2 + 8xy + x\end{aligned}$$

On setting the partial derivatives to zero, we get the saddle points as-

$$\begin{aligned}x=0, y=0 \\x=0, y=-\frac{1}{4} \\x=-\frac{1}{3}, y=0 \\x=-\frac{1}{9}, y=-\frac{1}{12}\end{aligned}$$

Checking for the maxima, minima and saddle points requires calculating the Hessian of the given equation. The Hessian of the equation is given by

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

For a two variable functions' minima to exist, the trace and determinant of the Hessian have to be positive. Whereas for the maxima, the trace has to be negative and the determinant of the Hessian has to be positive. If any of these conditions are not met then maxima or minima will not exist and the critical points may be saddle points.

In the case of $(-1/9, -1/12)$, $f_{xx} + f_{yy} < 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0$. Only $x = -\frac{1}{9}, y = -\frac{1}{12}$ satisfy the condition for the maxima. The other three points are the saddle points for the given function.

Problem 6.7

1. $Q^T = (A^T + A)^T = (A^T)^T + A^T = A^T + A = Q$. Hence, Q is a symmetric matrix.

We thus have

$$\begin{aligned}x^T Q x &= x^T A^T x + x^T A x = 2x^T A x \\&\Rightarrow x^T A x = 1/2 x^T Q x \\&\Rightarrow f(x) = \frac{1}{2} x^T Q x - b^T x + c\end{aligned}$$

¹I have worked with Navneeraj Sharma and Shekhar Kumar on this problem set

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

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

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 $Q^T x^* = b$

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

Consider $x_1, x_2 \in \text{conv}(S)$. Then, $\lambda_1 x_1 + (1 - \lambda_1)x_2$ is a finite sum of the form $\delta_1 x_1 + \delta_2 x_2$ where $\delta_1 + \delta_2 = 1$. By definition, this element $\in \text{conv}(S)$. Thus $\text{conv}(S)$ is convex.

Problem 7.2

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

$$\langle a, x \rangle = b$$

$$\langle a, y \rangle = b$$

Now, let $\lambda_1 + \lambda_2 = 1$ and $w = \lambda_1 x + \lambda_2 y$

$$\begin{aligned} \text{We have } \langle a, w \rangle &= \langle a, \lambda_1 x + \lambda_2 y \rangle \\ &= \langle a, \lambda_1 x \rangle + \langle a, \lambda_2 y \rangle \\ &= \lambda_1 \langle a, x \rangle + \lambda_2 \langle a, y \rangle \\ &= \lambda_1 b + \lambda_2 b \\ &= b \end{aligned}$$

Thus $\langle a, \lambda x + (1 - \lambda)y \rangle = b \Rightarrow \lambda x + (1 - \lambda)y \in P$. Hence a hyperplane is convex.

2. Let $H = \{x \in V | \langle a, x \rangle \leq b\}$, be a hyperplane. The proof follows analogously as above, except that in the last line we have $\langle a, x \rangle \leq b, \langle a, y \rangle \leq b$. Hence, $\langle a, \lambda x + (1 - \lambda)y \rangle \leq b$ which $\in H$. Hence Half Spaces are convex.

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, coupled with the fact that if $p \neq y, \|p - y\|^2 > 0$, 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 $\langle x - p, p - y \rangle \geq 0$

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

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

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

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

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

1. We are given that the function $g(t) : [0, 1] \rightarrow \mathbb{R}$ 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

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

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

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.

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

$$\begin{aligned}
f(tS^H S + (1-t)B) &= f(tS^H I S + (1-t)(S^H)^{-1}(S^H)BS^{-1}S) \\
&= f(S^H(tI + (1-t)(S^H)^{-1}BS^{-1})S) \\
\text{Let } X &= S^H, Y = tI + (1-t)(S^H)^{-1}BS^{-1}, Z = S \\
\text{The RHS can be written as } g(t) &= \\
&= -\log(\det(XYZ)) \\
&= -\log(\det(X)\det(Y)\det(Z)) \\
&= -\log(\det(XZ)\det(Y)) \\
&= -\log(\det(A)\det(Y)) \\
&= -\log(\det(A)) - \log(\det(Y)) \\
&= -\log(\det(A)) - \log(\det(tI + (1-t)(S^H)^{-1}BS^{-1}))
\end{aligned}$$

Hence part 2 is proved.

3. We now prove the next part

Let the Eigenvalues of $(S^H)^{-1}BS^{-1} = \lambda_1, \lambda_2 \dots \lambda_n$

Then the eigenvalues of Y are $t + (1-t)\lambda_1, t + (1-t)\lambda_2 \dots t + (1-t)\lambda_n$

Since the determinant is the product of the eigenvalues and since
log of product gives the sum of the logs, we have the result.

4. 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

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

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

$$\frac{f(x_1) - f(x_2)}{\lambda} \leq \frac{f(x_1) - (1 - \lambda)f(x_2)}{\lambda} \quad (2)$$

Which implies

$$\frac{f(x_1) - f(x_2)}{\lambda} \leq M \quad (3)$$

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, \mathbf{c} is $\mathbf{0}$ 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 $\phi \circ f(x)$ 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 $\phi \circ f(x^*) \leq \phi \circ f(x)$ for ϵ radius around $f(x)$. Again using the property that ϕ function is an increasing function we get $f(x^*) \leq f(x)$. Therefore x^* is a local minimizer of $f(x)$.