Home Assignment 3

<u>שאלה 1:</u>

<u>סעיף a:</u>

i. $\underline{f(x) = e^{ax}}$

תחום ההגדרה של הפונקציה הינו כל R, שהוא תחום קמור.
נראה שבתחום זה הנגזרת השנייה של הפונקציה חיובית:
$f``(x) = a^2e^{ax}$
$e^{ax} > 0 \wedge a^2 \geq 0$
סה"כ
$f``(x) \geq 0$
מכאן הפונקציה קמורה

ii. $\underline{f(x) = -\log(x)}$

תחום ההגדרה של הפונקציה הינו R^+ . נראה כי תחום זה הוא תחום קמור:
יהיו $x, y \in R^+$ נראה שלכל $\alpha \in [0,1]$ מתקיים: $\alpha x + (1-\alpha)y \in R^+$.
כאשר $\alpha = 0$ $\alpha x + (1-\alpha)y = y \in R^+$
כאשר $\alpha = 1$ $\alpha x + (1-\alpha)y = x \in R^+$
כאשר $\alpha \in (0,1)$ אז $\alpha, (1-\alpha), x, y \in R^+$ ולכן $\alpha x + (1-\alpha)y \in R^+$
- כעת נבדוק את הנגזרת השנייה של הפונקציה:

$$f`(x) = -\frac{1}{xlna} \rightarrow f``(x) = \frac{1}{x^2lna}$$

נסמן: (a הבסיס הלוגריתמי)
ישנם 3 מקרים:
1. $a < 1$

$$\Rightarrow lna < 0 \Rightarrow f``(x) \leq 0$$

ואז הפונקציה קעורה
2. $a > 1$

$$\Rightarrow lna > 0 \Rightarrow f``(x) \geq 0$$

ואז הפונקציה קמורה
3. $a = 1$

$\Rightarrow lna = 0 \Rightarrow$נגזרת שנייה אינה מוגדרת

iii. $\underline{f(x) = \log(x)}$

בדומה לb , תחום ההגדרה הינו R^+ וזה תחום קמור.
- נבדוק את הנגזרת השנייה של הפונקציה:

$$f`(x) = \frac{1}{xlna} \rightarrow f``(x) = -\frac{1}{x^2lna}$$

נחלק למקרים:
1. $a < 1$

אז נקבל:

$$\Rightarrow lna < 0 \Rightarrow f``(x) > 0$$

אז הפונקציה קמורה

2. $a > 1$

אז נקבל:

$$\Rightarrow lna < 0 \Rightarrow f``(x) < 0$$

אז הפונקציה קעורה

3. $a = 1$

$\Rightarrow lna = 0 \Rightarrow$נגזרת שנייה אינה מוגדרת

iv. $\underline{f(x) = |x|^a, a \geq 1}$

ניתן לרשום את הפונקציה כך:

תחום ההגדרה של הפונקציה הינו כל R, שהוא תחום קמור. נראה כי הפונקציה קמורה בתחום זה.
נחלק למקרים:
כאשר $a = 1$: אז $f(x) = |x|$ ונראה כי ההגדרה של קמירות אכן מתקיימת לכל $x, y \in R$ ו- $\alpha \in [0,1]$.

$$\mathrm{f}(ax + (1-a)y) = |\mathrm{ax} + (1-\mathrm{a})\mathrm{y}| \leq |\mathrm{ax}| + |(1-\mathrm{a})\mathrm{y}| = \mathrm{a}|\mathrm{x}| + (1-\mathrm{a})|\mathrm{y}|$$
$$= \mathrm{af}(x) + (1-\mathrm{a})\mathrm{f}(\mathrm{y})$$

כאשר $a > 1$ נוכיח לפי ההגדרה השנייה לקמירות(נשים לב שהפונקציה גזירה ב- $x \neq 0$):

$$f'(x) = \{$$

נראה כי $f(x) \geq f(y) + f'(y)(x-y)$
נסמן: $g(x) = f(x) - f(y) - f'(y)(x-y)$
נראה כי $g(x) \geq 0$
נשים לב כי:

$$g'(x) = f'(x) - f'(y)$$

כאשר $x, y \geq 0$:

$$g'(x) = f'(x) - f'(y) = ax^{a-1} - ay^{a-1} = a(x^{a-1} - y^{a-1})$$

אז $g'(x) < 0$ אם $y > x$

כאשר $x, y \leq 0$:

$$g'(x) = f'(x) - f'(y) = -a(-x)^{a-1} + a(-y)^{a-1} = a((-y)^{a-1} - (-x)^{a-1})$$

אז $g'(x) < 0$ אם $y > x$

כאשר $x \leq 0 \leq y$:

$$g'(x) = f'(x) - f'(y) = -a(-x)^{a-1} - a(y)^{a-1} = -a((-x)^{a-1} + (y)^{a-1})$$

מתקיים תמיד $g'(x) \leq 0$
כמו כן $g(0) = f(0) - f(y) - f'(y)(-y) = -y^a + ay^{a-1}y = -y^a + ay^a = y^a(a-1) \geq 0$

מאחר והפונקציה תמיד יורדת ובנק 0 היא חיובית אז מתקיים ש $g(x) \geq 0$

כאשר $y \leq 0 \leq x$:

$$g'(x) = f'(x) - f'(y) = a(x)^{a-1} + a(-y)^{a-1} = a((-y)^{a-1} + (x)^{a-1})$$

מתקיים תמיד $g'(x) \geq 0$

כמו כן $g(0) = f(0) - f(y) - f'(y)(-y) = -(-y)^a + a(-y)^{a-1}(-y) = (-y)^a + a(-y)^a =$
$(-y)^a(a+1) \geq 0$

ומכאן שגם במקרה הזה מתקיים $g(x) \geq 0$

סה"כ הראנו שבכל המקרים מתקיים $g(x) \geq 0$ ולכן תמיד מתקיים $f(x) \geq f(y) + f'(y)(x-y)$ שזו הגדרת קמירות.

v. $\underline{f(x) = x^3}$

תחום ההגדרה של הפונקציה הינו כל R, שהוא תחום קמור. נתבונן על הנגזרת השנייה:

$$f``(x) = 6x$$

נשים לב כי הנגזרת חיובית כאשר x חיובי ושלילית כאשר x שלילי, ולכן אינה קעורה בתחום הגדרתה.

<u>סעיף b:</u>

$$f(x) = x^T Ax + b^T x + c$$

$$f`(x) = (A + A^T)x + b^T$$
$$f``(x) = A + A^T$$

בהנחה ש$A \geq 0$, מתקיים

$$A \geq 0 \Leftrightarrow A^T \geq 0$$

מהגדרת מטריצה חיובית:

$$xAx^T + xA^T x^T \geq 0$$

ולכן מתקיים

$$x(A + A^T)x^T \geq 0$$

כלומר $A + A^T \geq 0$

מכאן שלכל $SPD\ A$,

$$f``(x) \geq 0$$

<u>סעיף c:</u>

עלינו להוכיח כי :

$$f\ is\ convex \leftrightarrow f(y) \geq f(x) + \nabla f(x)^T(y - x)$$

כיוון ראשון:

$$f\ is\ convex \rightarrow f(y) \geq f(x) + \nabla f(x)^T(y - x)\ \ \forall x, y \in \Omega$$

מהגדרת קמירות, אנו יודעים שלכל x, y מתקיים $f(\theta y + (1-\theta)x) \leq \theta f(y) + (1-\theta) f(x)$, כלומר נקבל: $f\big(x + (\theta)(y - x)\big) \leq \theta\big(f(y) - f(x)\big) + f(x)$
נשים לב כאשר $\theta \to 0$ נקבל לפי טור טיילור:

$$f\big(x + \theta(y - x)\big) = f(x) + \theta(y - x) f'(x) = f(x) + \theta \nabla f(x)^T (y - x)$$

ולכן נקבל:

$$f(x) + \theta \nabla f(x)^T (y - x) \leq \theta\big(f(y) - f(x)\big) + f(x)$$
$$\to \nabla f(x)^T (y - x) \leq f(y) - f(x)$$
$$\to f(y) \geq f(x) + \to \nabla f(x)^T (y - x)$$

כנדרש.

כיוון שני:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \ \ \forall x, y \ \to \ f \ is \ convex$$

יהיו $x, y, z \in \Omega$ אז מתקיים:

$$f(x) \geq f(z) + \nabla f(z)^T (x - z)$$
$$f(y) \geq f(z) + \nabla f(z)^T (y - z)$$

נכפול את הא"ש הראשון בα ואת הא"ש השני ב$1-\alpha$:

$$\alpha f(x) \geq \alpha f(z) + \alpha \nabla f(z)^T (x - z)$$
$$(1-\alpha) f(y) \geq (1-\alpha) f(z) + (1-\alpha) \nabla f(z)^T (y - z)$$

נסכום את הא"ש ונקבל:

$$\alpha f(x) + (1-\alpha) f(y) \geq \alpha f(z) + \alpha \nabla f(z)^T (x - z) + (1-\alpha) f(z) + (1-\alpha) \nabla f(z)^T (y - z)$$
$$\alpha f(x) + (1-\alpha) f(y) \geq f(z) + \nabla f(z)^T (\alpha x - \alpha z + y - \alpha y + \alpha z - z)$$
$$\alpha f(x) + (1-\alpha) f(y) \geq f(z) + \nabla f(z)^T (\alpha x - z + y - \alpha y)$$

נציב $z = \alpha x + (1-\alpha) y$

$$\alpha f(x) + (1-\alpha) f(y) \\ \geq f(\alpha x + (1-\alpha) y) + \nabla f(\alpha x + (1-\alpha) y)^T (\alpha x - (\alpha x + (1-\alpha) y) + y - \alpha y)$$

$$\alpha f(x) + (1-\alpha) f(y) \geq f(\alpha x + (1-\alpha) y) + \nabla f(\alpha x + (1-\alpha) y)^T (0)$$

$$\alpha f(x) + (1-\alpha) f(y) \geq f(\alpha x + (1-\alpha) y)$$

ולפי הגדרת קמירות נקבל כי f פונקציה קמורה כנדרש.

<u>שאלה 2:</u> מצורף קוד
כאשר השתמשנו בחישובים הבאים:

$$argmin_x = (x-y)^T(x-y) + \frac{\lambda}{2}(x^TG^TGx) = x^Tx - x^Ty - y^Tx + y^Ty + \frac{\lambda}{2}(x^TG^TGx)$$
$$f`(x) = 2x - y - y + \lambda G^TGx = 2x - 2y + \lambda G^TGx$$
$$f`(x) = 2x - 2y + \lambda G^TGx = 0$$
$$(2 + \lambda G^TG)x = 2y$$
$$x = (2 * I + \lambda G^TG)^{-1}2y$$

$$argmin \left|\left|x-y\right|\right|_2^2 + \lambda\left|\left|Gx\right|\right|_W^2 = (x-y)^T(x-y) + (Gx)^TW(Gx)$$
$$= x^Tx - xy - y^Tx - y^Ty + x^TGWGx$$

נגזור לפי x ונקבל: $2x - 2y + 2G^TWGx = 0$
$$x = (I + G^TWG)^{-1} * y$$

<u>שאלה 3:</u>

<u>סעיף *a*:</u>

נתון כי
$$F(\theta) = \frac{1}{2}\left|\left|f(\theta) - y^{obs}\right|\right|_2^2$$
נסמן: $y^{obs} = c$
אז נקבל כי:
$$F(\theta) = \frac{1}{2}(f(\theta) - y^{obs})^T(f(\theta) - y^{obs}) = \frac{1}{2}(f(\theta)^Tf(\theta) - f(\theta)^Tc - c^Tf(\theta) + c^Tc)$$
ולכן:
$$\nabla F(\theta) = \frac{1}{2}(2f(\theta) * f`(\theta) - f`(\theta)c - f`(\theta)c) = f(\theta)f`(\theta) - f`(\theta)c = f`(\theta)(f(\theta) - c)$$
$$= J^T(f\theta - c) = J^T(f(\theta) - y^{obs})$$

<u>סעיף *b*:</u>

ממשוואה (3) : $d_{GN}^k = arg\min \frac{1}{2}\left|\left|f(\theta^K) + J(\theta^K)d - y^{obs}\right|\right|_2^2$

נגזור את הפונקציה ונשווה ל0 על מנת למצוא את הפתרון:

נסמן: $g(d) = \frac{1}{2}\left|\left|f(\theta^K) + J(\theta^K)d - y^{obs}\right|\right|_2^2$
נשתמש בהוכחה מסעיף קודם כי: $\nabla F(\theta) = J^T(f(\theta) - y^{obs})$
אז
$$g'(x) = J^T(\theta^k)(f(\theta^k) + J(\theta^k)d - y^{obs}) =$$
$$(J^T(\theta^k)f(\theta^k) + J^T(\theta^k)J(\theta^k)d - J^T(\theta^k)y^{obs}) = 0$$

$$\rightarrow J^T(\theta^k)J(\theta^k)d = J^T(\theta^k)y^{obs} - J^T(\theta^k)f(\theta^k) = J^T(\theta^k)(y^{obs} - f(\theta^k))$$

מסעיף א': $J^T(y^{obs} - f) = -\nabla F(\theta)$

ולכן נקבל: $J^TJd = J^T(y^{obs} - f) = -\nabla F(\theta)$ כנדרש.

<u>סעיף *c*:</u>

נפתור את המשוואה: $d_{LM}^k = arg\min \frac{1}{2}\left|\left|f(\theta^K) + J(\theta^K)d - y^{obs}\right|\right|_2^2 + \frac{\mu}{2}\left|\left|d\right|\right|_2^2$

נסמן: $g(d) = \frac{1}{2}\left|\left|f(\theta^K) + J(\theta^K)d - y^{obs}\right|\right|_2^2 + \frac{\mu}{2}\left|\left|d\right|\right|_2^2$

נגזור ונשווה ל-0:

$$g'(d) = J^T(\theta^k)(f(\theta^k) + J(\theta^k)d - y^{obs}) + \mu d = 0$$
$$d(J^T(\theta^k)J(\theta^k) + \mu) = J^T(\theta^k)y^{obs} - J^T(\theta^k)f(\theta^k)$$
$$\rightarrow d = (J^TJ + \mu I)^{-1}(J^Ty^{obs} - J^Tf)$$

נשים לב שהמטריצה $J^TJ + \mu I$ היא הפיכה שכן J^TJ היא סימטרית ו $\mu I > 0$ ולכן המטריצה כולה הפיכה.

<u>סעיף *d*:</u>

נשים לב שקיבלנו כי $d_{LM} = -(J^TJ + \mu I)^{-1} * \nabla F$

המטריצה J^TJ היא סימטרית ולפחות *SemiPositiveDefinite* אך עם הוספת μI נקבל כי $J^TJ + \mu I$ הינה *PD*. ולכן גם המטריצה ההופכית שלה $(J^TJ + \mu I)$. על כן, d_{LM} המתקבל הינו בכיוון המנוגד לגרדיאנט ולכן כיוון ירידה.

<u>סעיף *e*:</u>

קוד מצורף

<u>**שאלה 4:**</u> קוד מצורף