(20942) מבוא ללמידה חישובית | סיכום הרצאה 3

מנחה: שי מימון
סמסטר: 2022א'
נכתב על ידי: מתן כהן

1 חזרה מתמטית - Matrix Calculus

1.1 סיווגי מטריצות

הגדרה: אם M מטריצה ממשית סימטרית מסדר $n \times n$ אז:

- M תקרא **חיובית מוגדרת** (positive-definite) אם"ם לכל וקטור $\mathbf{x} \in \mathbb{R}^n$ כך ש-$\mathbf{x} \neq \mathbf{0}$ מתקיים:

$$\mathbf{x}^T M \mathbf{x} > 0$$

- M תקרא **אי-שלילית מוגדרת** (positive-semidefinite) אם"ם לכל וקטור $\mathbf{x} \in \mathbb{R}^n$ מתקיים:

$$\mathbf{x}^T M \mathbf{x} \geq 0$$

- בצורה מקבילה מגדירים מטריצה **שלילית לחלוטין** ומטריצה **אי-חיובית**
- אם לא ניתן להגדיר סימן למטריצה היא תיקרא indefinite

דוגמה: מטריצה $A = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$

נבדוק:

$$\begin{aligned} \mathbf{x}^T A \mathbf{x} &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= 2x_1^2 - 2x_1x_2 + x_2^2 \\ &= x_1^2 + (x_1^2 - 2x_1x_2 + x_2^2) \\ &= \underbrace{x_1^2}_{\geq 0} + \underbrace{(x_1 - x_2)^2}_{\geq 0} \end{aligned}$$

$$\forall \mathbf{x} \neq 0 > 0$$

ולכן A חיובית מוגדרת

דוגמה: מטריצה $B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$

נבדוק ונמצא שלא ניתן לדעת האם חיובית או שלילית ונקבע כי היא היא indefinite:

$$\begin{aligned} \mathbf{x}^T B \mathbf{x} &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^2 \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= x_1^2 + 4x_1x_2 + x_2^2 \\ &= (x_1 + x_2)^2 + 2x_1x_2 \end{aligned}$$

למה: תהא A מטריצה חיובית מוגדרת (או אי-שלילית מוגדרת). אזי כל אברי האלכסון במטריצה הם חיוביים (או אי-שליליים).

משפט: אפיון מטריצה על פי ערכים עצמיים

תהא A מטריצה סימטרית מסדר $n \times n$, אז:

- A היא חיובית מוגדרת $\iff$ כל הערכים העצמיים שלה הם חיוביים
- A היא אי-שלילית מוגדרת $\iff$ כל הערכים העצמיים שלה הם אי-שליליים
- A היא שלילית מוגדרת $\iff$ כל הערכים העצמיים שלה הם שליליים
- A היא אי-חיובית מוגדרת $\iff$ כל הערכים העצמיים שלה הם אי-חיוביים
- A היא עם סימן לא מוגדר $\iff$ יש ל A לפחות ע"ע אחד חיובי וע"ע אחד שלילי

דוגמה: נסווג את המטריצה $A = \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix}$ על פי הע"ע שלה

נמצא ערכים עצמיים:

$$det(A) = \begin{vmatrix} 4 & 1 \\ 1 & 3 \end{vmatrix} = 11 = \lambda_1 \cdot \lambda_2$$
$$tr(A) = 4 + 3 = 7 = \lambda_1 + \lambda_2$$

נציב:

$$\lambda_2 = 7 - \lambda_1$$

ואז:

$$\lambda_1 \cdot (7 - \lambda_1) = 11$$

ולכן:

$$\Rightarrow \lambda_1^2 - 7\lambda_1 + 11 = 0$$
$$\Rightarrow \lambda_1 = \frac{7 \pm \sqrt{5}}{2} > 0$$

לכן A היא חיובית מוגדרת

1.2 גרדיאנט

הגדרה: יהי $\mathbf{x} \in \mathbb{R}^n$ ותהא $f : \mathbb{R}^n \to \mathbb{R}^m$ פונקציה דיפרנציאבילית (גזירה) ב-$\mathbf{x}$ אזי הגרדיאנט של f בנקודה $\mathbf{x}$ (וקטור) מוגדרת כוקטור קואורדינטות בצורה הבאה:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \\ \frac{\partial f}{\partial x_2}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

הגדרה: ההסיין ($Hessian$) של f אשר ניתנת לגזירה פעמיים בנקודה $\mathbf{x} \in U$ הוא המטריצה בגודל $n \times n$ שבאלכסון הראשי גוזרים פעמיים לפי אותו משתנה ובכל תא i, j גוזרים לפי משתנה x_i ומשתנה x_j:

$$H = \nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & & \vdots \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1} & \frac{\partial^2 f}{\partial x_n x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

עבור הסיין עם נגזרת שנייה רציפה - נקבל מטריצה סימטרית.

דוגמה: הראו: $\nabla_w(\mathbf{w}^T\mathbf{b}) = \mathbf{b}$

מתקיים:

$$\mathbf{w}^T\mathbf{b} = \sum_{i=1}^{n} w_i b_i$$

$$\Rightarrow \nabla_{\mathbf{w}} \left(\sum_{i=1}^{n} w_i b_i \right) = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \mathbf{b}$$

טענה: מתקיים: $\nabla_{\mathbf{w}} \left(\mathbf{w}^T A \mathbf{w} \right) = \left(A + A^T \right) \mathbf{w}$

1.3 נקודות מינימום ומקסימום

הגדרה: מינימום ומקסימום גלובליים

תהא $f : S \to \mathbb{R}$ מוגדרת על קבוצה $S \subseteq \mathbb{R}^n$. אז:

(1) אם $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ לכל $\mathbf{x} \in S$ אז $\mathbf{x}^* \in S$ היא **נקודת מינימום גלובלית** של f מעל S

(2) אם $f(\mathbf{x}) > f(\mathbf{x}^*)$ לכל $\mathbf{x} \neq \mathbf{x}^* \in S$ אז $\mathbf{x}^* \in S$ היא **נקודת מינימום (במובן החזק) גלובלית** של f מעל S

(3) אם $f(\mathbf{x}) \leq f(\mathbf{x}^*)$ לכל $\mathbf{x} \in S$ אז $\mathbf{x}^* \in S$ היא **נקודת מקסימום גלובלית** של f מעל S

(4) אם $f(\mathbf{x}) < f(\mathbf{x}^*)$ לכל $\mathbf{x} \in S$ אז $\mathbf{x}^* \in S$ $\mathbf{x}^* \neq$ היא **נקודת מקסימום (במובן החזק) גלובלית** של f מעל S

הגדרה: מינימום ומקסימום מקומיים

נגדיר את $open - ball$ (סביבת כדור) עם מרכז $\mathbf{c} \in \mathbb{R}^n$ ורדיוס r ב:

$$B(\mathbf{c}, r) = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{c}\| < r\}$$

תהא $f : S \to \mathbb{R}$ מוגדרת על הקבוצה $S \subseteq \mathbb{R}^n$ אז:

(1) אם קיים $r > 0$ עבורו $f(\mathbf{x}^*) \leq f(\mathbf{x})$ לכל $\mathbf{x} \in S \cap B(\mathbf{x}^*, r)$ אז $\mathbf{x}^*$ היא נקודת **מינימום מקומית** של f מעל S

(2) אם קיים $r > 0$ עבורו $f(\mathbf{x}^*) < f(\mathbf{x})$ לכל $\mathbf{x} \neq \mathbf{x}^* \in S \cap B(\mathbf{x}^*, r)$ אז $\mathbf{x}^*$ היא נקודת **מינימום (במובן החזק) מקומית** של f מעל S

(3) אם קיים $r > 0$ עבורו $f(\mathbf{x}^*) \geq f(\mathbf{x})$ לכל $\mathbf{x} \in S \cap B(\mathbf{x}^*, r)$ אז $\mathbf{x}^*$ היא נקודת **מקסימום מקומית** של f מעל S

(4) אם קיים $r > 0$ עבורו $f(\mathbf{x}^*) > f(\mathbf{x})$ לכל $\mathbf{x} \neq \mathbf{x}^* \in S \cap B(\mathbf{x}^*, r)$ אז $\mathbf{x}^*$ היא נקודת **מקסימום (במובן החזק) מקומית** של f מעל S

משפט: תהא $f : U \to \mathbb{R}$ מעל $U \subseteq \mathbb{R}^n$. אם קיימת נקודה $\mathbf{x}^*$ שהיא אופטימום מקומי (מקסימום או מינימום) וגם כל הנגזרות החלקיות של f קיימות בנקודה $\mathbf{x}^*$ אז $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

הנקודה $\mathbf{x}^*$ תיקרא גם **נקודה סטציונרית (נקודה חשודה לקיצון מקומי)**

משפט: הכללה של הגדרת נקודת קיצון על פי הנגזרת השנייה של פונקציה רבת משתנים.
תהא $f : U \to \mathbb{R}$ מוגדרת על קבוצה פתוחה $U \subseteq \mathbb{R}^n$. אם f גזירה ברציפות פעמיים מעל U וגם $\mathbf{x}^*$ היא נקודה חשודה לקיצון אז:

(1) אם $\nabla^2 f(\mathbf{x}^*) \succ 0$ (המטריצה המתקבלת מהגזירה השנייה היא חיובית מוגדרת), אז $\mathbf{x}^*$ היא נקודת מינימום (**במובן החזק**) של f מעל U

(2) אם $\nabla^2 f(\mathbf{x}^*) \prec 0$ (המטריצה המתקבלת מהגזירה השנייה היא שלילית מוגדרת), אז $\mathbf{x}^*$ היא נקודת מקסימום (**במובן החזק**) של f מעל U

הערה: התנאים תקפים גם עבור $\preceq$ ו- $\succeq$ ומטריצות אי-שליליות או אי-חיוביות בצורה דומה

משפט: תהא f גזירה פעמיים ברציפות מעל $\mathbb{R}^n$. אם $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ לכל $\mathbf{x} \in \mathbb{R}^\mathbf{n}$ ו-$\mathbf{x}^* \in \mathbb{R}^n$ נקודה חשודה לקיצון של f אז $\mathbf{x}^*$ היא מינימום גלובלי של f.

הערה: יכולות להיות מספר נקודות מינימום גלובליות (שוות)

1.4 פונקציות קונבקסיות (קמורות)

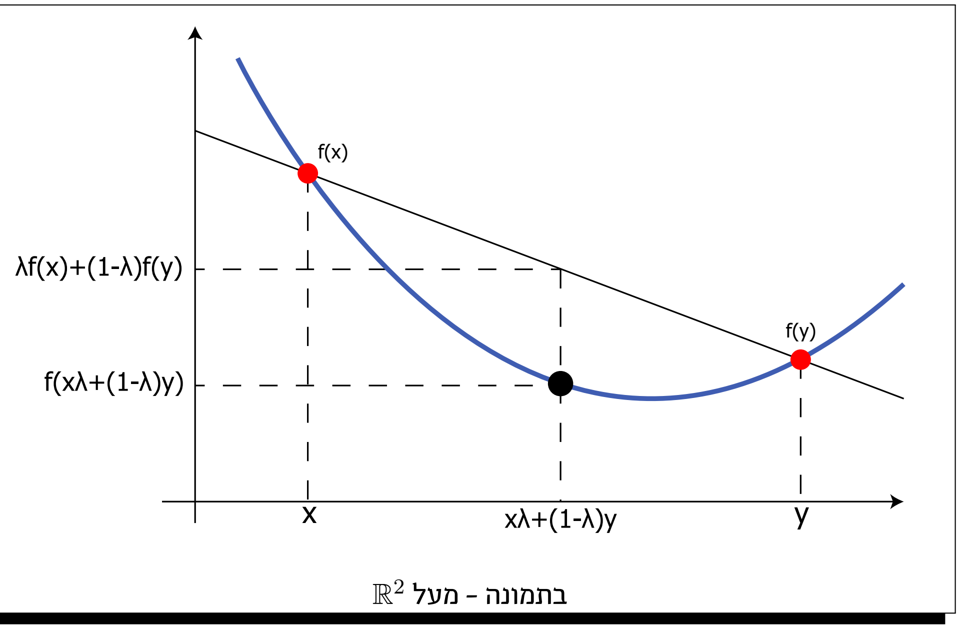
הגדרה: **קבוצה קונבקסית (קמורה)** - קבוצת נקודות במרחב וקטורי היא **קמורה** אם לכל שתי נקודות בתורה, גם הקטע המחבר את שתי הנקודות נמצא כולו בתוכה.

הגדרה: פונקציה $f : C \to \mathbb{R}$ מוגדרת על קבוצה קונבקסית $C \subseteq \mathbb{R}^n$ נקראת **קונבקסית (קמורה)** אם לכל $\mathbf{x}, \mathbf{y} \in C$ ו-$\lambda \in [0,1]$ מתקיים:

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

אינטואיציה: אם $\lambda \in [0,1]$ אז :

- נתבונן בביטוי השמאלי
  - בהכרח $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ היא נקודה שחייבת להמצא בין $\mathbf{x}$ לבין $\mathbf{y}$! (ממוצע משוקלל).
- נתבונן בביטוי הימני
  - בהכרח עבור $\lambda = 0$ נקבל: $0 \cdot f(\mathbf{x}) + (1-0) f(\mathbf{y}) = f(\mathbf{y})$
  - בהכרח עבור $\lambda = 1$ נקבל: $1 \cdot f(\mathbf{x}) + (1-1\cdot) f(\mathbf{y}) = f(\mathbf{x})$
  - $\star$ לכן כל נקודה $\lambda \in [0,1]$ תניב לנו ערך על הישר שעובר בין 2 הנקודות $\mathbf{x}$ ו-$\mathbf{y}$
- נסיק כי אי השווין מתאר שלכל זוג נקודות $\mathbf{x}, \mathbf{y} \in C$, ולכל $\lambda \in [0,1]$, הביטוי שמתאר את ערך הפונקציה בכלל הנקודות בין $\mathbf{x}$ ל-$\mathbf{y}$ תמיד קטן או שווה לישר שעובר בין $f(\mathbf{x})$ ל- $f(\mathbf{y})$.



בתמונה - מעל $\mathbb{R}^2$

הגדרה: פונקציה $f:C\to\mathbb{R}$ מוגדרת על על קבוצה קונבקסית $C\subseteq\mathbb{R}^n$ נקראת **קונבקסית (קמורה) ממש** אם לכל $\mathbf{x}\neq\mathbf{y}\in C$ ו- $\lambda\in[0,1]$ מתקיים:

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

באופן סימטרי מגדירים קבוצה **קעורה** וקעורה ממש עם $\geq$ ו- $>$ בהתאמה.

משפט: תהא f פונקציה גזירה פעמיים על קבוצה קמורה $C\subseteq\mathbb{R}^n$.
נגיד כי f קמורה מעל C **אם ורק אם** $\nabla^2 f(\mathbf{x})\succeq\mathbf{0}$ (ההסיין של f בנקודה $\mathbf{x}$ היא מטריצה אי שלילית) לכל $\mathbf{x}\in C$

משפט: תהא f פונקציה גזירה וקמורה מעל קבוצה קמורה $C\subseteq\mathbb{R}^n$. אם $\nabla f(\mathbf{x}^*)=\mathbf{0}$ עבור $\mathbf{x}^*\in C$ אז $\mathbf{x}^*$ היא נקודת מינימום של f מעל C.

1.5 קירוב לינארי וקוואדרטי של פונקציות רבות משתנים

הרחבה למקרה הסקלרי של טור טיילור

משפט: קירוב לינארי: תהא $f:U\to\mathbb{R}$ מוגדרת על קבוצה $U\subseteq\mathbb{R}^n$ וגזירה פעמיים ברציפות מעל U. יהא $\mathbf{x}\in U$ ו $r>0$ אשר מקיימים $B(\mathbf{x},r)\subseteq U$. אז עבור כל $\mathbf{y}\in B(\mathbf{x},r)$ קיים $\xi\in[\mathbf{x},\mathbf{y}]$ כך ש

$$f(\mathbf{y})=f(\mathbf{x})+\nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x})+\frac{1}{2}(\mathbf{y}-\mathbf{x})^T\nabla^2 f(\xi)(\mathbf{y}-\mathbf{x})$$

משפט: קירוב קוואדרטי: תהא $f:U\to\mathbb{R}$ מוגדרת על קבוצה $U\subseteq\mathbb{R}^n$ וגזירה פעמיים ברציפות מעל U. יהא $\mathbf{x}\in U$ ו $r>0$ אשר מקיימים $B(\mathbf{x},r)\subseteq U$. אז עבור כל $\mathbf{y}\in B(\mathbf{x},r)$:

$$f(\mathbf{y})=f(\mathbf{x})+\nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x})+\frac{1}{2}(\mathbf{y}-\mathbf{x})^T\nabla^2 f(\mathbf{x})(\mathbf{y}-\mathbf{x})+o(\left(\|\mathbf{y}-\mathbf{x}\|^2\right)$$