מודלים לא לינאריים בחקר ביצועים - תרגיל בית 6

3 במאי 2021

הנחיות להגשה

- הגשת התרגיל היא עד יום ראשון, 23.5.21, בשעה 12:00 בצהריים.
- יש להגיש שני קבצים עם שם קובץ HW##_ID1_ID2:
  - PDF המכיל את תשובותיכם כולל קודים ופלטים.
  - ZIP המכיל את כל הקודים.

שאלה 1

תהי $\mathbf{A} \in \mathbb{R}^{m\times n}$ ששורותיה הם הווקטורים $\mathbf{a}_1^T, \ldots, \mathbf{a}_m^T$. נניח ש-$\mathbf{a}_i$ אינו וקטור האפס לכל $1 \leq i \leq m$. יהי וקטור $\mathbf{b} \in \mathbb{R}^m$ ונגדיר את הפוליגון $\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \leq \mathbf{b}\}$. המרכז האנליטי של $\mathcal{P}$ מוגדר להיות פתרון בעיית האופטימיזציה

$$\min_{\mathbf{x}\in \text{interior}(\mathcal{P})} \left\{ f(\mathbf{x}) \equiv -\sum_{i=1}^{m} \ln\left(\mathbf{b}_i - \mathbf{a}_i^T \mathbf{x}\right) \right\}$$

הערה: המשמעות הגאומטרית של המרכז האנליטי היא שזו הנקודה שמכפלת מרחקיה מכל אחד מגבולות הפוליגון היא הגדולה ביותר. לנקודה זו יש משמעות בתחומים שונים בפיזיקה.

א. הוכיחו כי $\nabla^2 f(\mathbf{x}) \succeq 0$ לכל $\mathbf{x} \in \text{interior}(\mathcal{P})$.

הכוונה: שימו לב שלפונקציה זו יש חלקים לינאריים.

ב. האם בהכרח מתקבל לבעיה פתרון מיטבי לכל $\mathbf{A} \in \mathbb{R}^{m\times n}$ ולכל $\mathbf{b} \in \mathbb{R}^m$?

ג. בהרצאה ראינו את משפט ההתכנסות של שיטת ניוטון הטהורה עבור f המוגדרת בכל $\mathbb{R}^n$. ניתן לנסח משפט דומה עבור f המוגדרת בתחום הגדרה פתוח כשכל האיטרציות $\left\{\mathbf{x}^k\right\}_{k\geq 0}$ נמצאות בתוך התחום.

הראו שעבור

$$\mathbf{A} = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

אז לבעיה למציאת המרכז האנליטי קיים פתרון מיטבי בתחום $\text{interior}(\mathcal{P})$, וגם שלא קיים $m > 0$ כך ש-$\nabla^2 f(\mathbf{x}) \succeq m\mathbf{I}_2$ לכל $\mathbf{x}$ בתחום ההגדרה של f.

הערה: המשמעות של כך היא שמצאתם פונקציה f בעלת פתרון מיטבי שמשפט ההתכנסות של שיטת ניוטון הטהורה לא מבטיח עבורה התכנסות.

ד. ממשו פונקציית python המופעלת באופן הבא:

```
xs, fs = analytic_center(A, b, x0)
```

הפונקציה מחשבת את המרכז האנליטי של הפוליגון $\mathcal{P}$ באמצעות שיטת ניוטון עם גודל צעד הנבחר ע"י עקיבה לאחור עם פרמטרים $s = 2$, $\alpha = \frac{1}{4}$ ו־$\beta = \frac{1}{2}$ ונקודת התחלה $\mathbf{x}^0$. השתמשו בתנאי העצירה $\|\nabla f(\mathbf{x})\| \leq 10^{-6}$.

הפלט xs הוא מטריצה שבה העמודה k היא האיטרציה $\mathbf{x}^k$, כולל האיטרציה 0. הפלט fs הוא וקטור באורך כמות האיטרציות, כולל האיטרציה 0, כך ש־$\texttt{fs(k)} = f\left(\mathbf{x}^k\right)$.

בדיקת הקלט היחידה שעליכם לבצע היא שנקודת ההתחלה $\mathbf{x}^0$ נמצאת בתחום ההגדרה $\text{interior}\left(\mathcal{P}\right)$ של f (תוכלו להשתמש בפונקציה all). ניתן להניח שלבעיה הנתונה יש פתרון מיטבי.

ה. כיתבו סקריפט בשם test_analytic_center שמוצא את המרכז האנליטי של המערכת

$$\begin{cases} 2x + 10y \leq 1 \\ x \leq 0 \\ -x + 3y \leq 2 \\ -x - y \leq 2 \end{cases}$$

באמצעות הפונקציה שכתבתם בסעיף הקודם. נקודת ההתחלה היא $x = -1.99$ ו־$y = 0$. הסקריפט יוצר שני תרשימים:

- בתרשים הראשון רואים את קווי הגובה $\left\{f\left(\mathbf{x}^k\right)\right\}_{k\geq 0}$ יחד עם הנקודות $\left\{\mathbf{x}^k\right\}_{k\geq 0}$ בתחום

$$\mathcal{Q} = \left\{x \in [-2, 0], y \in \left[-2, \frac{1}{2}\right]\right\} \supset \mathcal{P}$$

השתמשו בפונקציה contour תחת plt.subplots(). שימו לב ש־f לא מוגדרת בכל $\mathcal{Q}$ עבור גבהים אלה, לכן בשביל לחשב את קווי הגובה רק בתחום $\mathcal{P}$ כיתבו את שורות הקוד הבאות:

```
z[(~np.isreal(z))] = 10**10
z = z.real
z = z.reshape(x.T.shape).T
fs.reverse()
```

כאשר z מייצג את ערכי הפונקציה f בתחום $\mathcal{Q}$. באותו תרשים שרטטו גם אם הפוליגון $\mathcal{P}$. השתמשו לצורך כך בפונקציות Polygon ו־PatchCollection.

- בתרשים השני רואים את $f\left(\mathbf{x}^k\right) - f\left(\mathbf{x}^N\right)$ לכל $1 \leq k \leq N-1$ כאשר N היא האיטרציה האחרונה. שרטטו את ציר ה־y בסולם לוגריתמי.

שאלה 2

א. כיתבו פונקציית python המממשת את שיטת ניוטון ההיברידית עבור שיטת חיפוש קווי גנרית. הפונקציה מופעלת באופן הבא:

```
x, fs, gs, ts, newton = hybrid_newton(f, gf, hf, lsearch, x0, eps)
```

כאשר הקלטים הם:

- $\texttt{f}, \texttt{gf}, \texttt{hf}$ – הפונקציה f, הגרדיאנט ∇f וההסיאן $\nabla^2 f$ בהתאמה.
- x0 – נקודת ההתחלה של השיטה.
- eps – רמת דיוק. השיטה תעצור לפי תנאי העצירה $\left\|\nabla f\left(x^k\right)\right\| \leq \varepsilon$.
- lsearch – פונקציית חיפוש קווי המופעלת באופן הבא:

```
tk = lsearch(xk, gk, direction)
```

כאשר tk הוא גודל הצעד המוחזר, xk היא האיטרציה הנוכחית, gk הוא הגרדיאנט באיטרציה הנוכחית ו־direction הוא מערך עם שני רכיבים, כך ש־direction[0] הוא הכיוון שהאלגוריתם בחר ו־direction[1] הוא string עם הערכים 'newton' ו־'grad' בהתאם.

הפונקציה מחזירה את הפלטים הבאים:

- x – הנקודה אליה התכנסה השיטה.
- $\texttt{fs}, \texttt{gs}, \texttt{ts}$ – מערכים שגודלם ככמות האיטרציות כפי שהוגדרו בתרגיל בית 4.
- newton – מערך שגודלו ככמות האיטרציות כך ש־$\texttt{newton}[\texttt{k}] = 1$ אם באיטרציה ה־k האלגוריתם בחר בכיוון $\mathbf{d}^k$ לפי שיטת ניוטון, ו־0 אם לפי שיטת הגרדיאנט.

את הבדיקה האם $\nabla^2 f(x) \succ 0$ יש לבצע באמצעות פירוק שולסקי.

ב. כיתבו פונקציית python המופעלת באופן הבא:

```
lsearch = hybrid_back(f, alpha, beta, s)
```

המחזירה פונקציית חיפוש קווי כפי שהוגדרה בסעיף הקודם, המממשת את שיטת עקיבה לאחור עבור כיוון שנבחר לפי שיטת הגרדיאנט, וגודל צעד קבוע בגודל 1 עבור כיוון שנבחר לפי שיטת ניוטון (שיטת ניוטון הטהורה). הקלטים של הפונקציה הם פונקציית המטרה f, פרמטרים $s > 0$ ו־$\alpha, \beta \in (0,1)$. אין צורך לבדוק את תקינות הקלט.

ג. הפעילו את השיטה שכתבתם בסעיף א' עם החיפוש הקווי שכתבתם בסעיף ב' עם פרמטרים $s = 1$, $\alpha = \frac{1}{4}$ ו־$\beta = \frac{1}{2}$ על הפונקציה $f(x,y) = x^4 + y^4 - 36xy$ עם נקודת ההתחלה $(200,0)^T$.

צרפו תרשים אחד הכולל:

- המערך $\texttt{fs} + 162$ בסולם לוגריתמי בכל איטרציה, יחד עם סוג הכיוון (ניוטון/גרדיאנט) אותו האלגוריתם בחר. השתמשו ב־$\varepsilon = 10^{-6}$.
- המערך $\texttt{fs} + 162$ לפי שיטת הגרדיאנט עם עקיבה לאחור עם פרמטרים $s = 1$, $\alpha = \frac{1}{4}$ ו־$\beta = \frac{1}{2}$ ואותם נקודת התחלה ותנאי עצירה. תוכלו להשתמש בפונקציות אותן כתבתם בתרגיל בית 4 (רק זיכרו להתאים את תנאי העצירה).

בהסתמך על התרשים, האם האלגוריתם ההיברידי כפי שהוא מתואר בשאלה הוא אלגוריתם ירידה? הציעו שינוי שניתן ליישם בו בשביל להפוך אותו לאלגוריתם ירידה (אין צורך לספק הוכחה מתמטית).

שאלה 3

בעיית פרמה־וובר מוגדרת כפתרון בעיית האופטימיזציה

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \equiv \sum_{i=1}^{m} \omega_i \|\mathbf{x} - \mathbf{a}_i\|,$$

עבור $\mathcal{A} \equiv \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\} \subset \mathbb{R}^n$ ו־$\omega_1, \omega_2, \dots, \omega_m > 0$ קבועים. נגדיר $h\colon \mathbb{R}^n \times (\mathbb{R}^n \backslash \mathcal{A}) \to \mathbb{R}$ ו־$L\colon \mathbb{R}^n \backslash \mathcal{A} \to \mathbb{R}$ על־ידי

$$h(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{m} \omega_i \frac{\|\mathbf{x} - \mathbf{a}_i\|^2}{\|\mathbf{y} - \mathbf{a}_i\|}$$

$$L(\mathbf{x}) = \sum_{i=1}^{m} \frac{\omega_i}{\|\mathbf{x} - \mathbf{a}_i\|}$$

א. הראו שהפונקציה $h(\mathbf{x},\mathbf{y})$ $\mathbf{x} \to$ היא ריבועית עם מטריצה מובילה $L(\mathbf{y})\mathbf{I}_n$ והסיקו ממשפט הקירוב הריבועי מסדר שני כי

$$h(\mathbf{x},\mathbf{y}) = h(\mathbf{y},\mathbf{y}) + \nabla_{\mathbf{x}} h(\mathbf{y},\mathbf{y})^T (\mathbf{x}-\mathbf{y}) + L(\mathbf{y}) \|\mathbf{x}-\mathbf{y}\|^2, \ \forall \mathbf{x},\mathbf{y} \in \mathbb{R}^n.$$

ב. הוכיחו כי

$$\begin{aligned} h(\mathbf{y},\mathbf{y}) &= f(\mathbf{y}), \ \forall \mathbf{y} \in \mathbb{R}^n \\ \nabla_{\mathbf{x}} h(\mathbf{y},\mathbf{y}) &= 2\nabla f(\mathbf{y}), \ \forall \mathbf{y} \in \mathbb{R}^n \\ h(\mathbf{x},\mathbf{y}) &\geq 2f(\mathbf{x}) - f(\mathbf{y}), \ \forall \mathbf{x},\mathbf{y} \in \mathbb{R}^n \end{aligned}$$

והסיקו כי מתקיימת למת הירידה הבאה:

$$f(\mathbf{x}) \leq f(\mathbf{y}) + \nabla f(\mathbf{y})^T (\mathbf{x}-\mathbf{y}) + \frac{L(\mathbf{y})}{2} \|\mathbf{x}-\mathbf{y}\|^2, \ \forall \mathbf{x},\mathbf{y} \in \mathbb{R}^n.$$

שאלה 4

עבור m פונקציות $f_i\colon \mathbb{R}^n \to \mathbb{R}$ גזירות ברציפות ו־m קבועים $c_i \in \mathbb{R}$ נסתכל על בעיית המינימיזציה

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left\{ g(\mathbf{x}) \equiv \sum_{i=1}^m (f_i(\mathbf{x}) - c_i)^2 \right\}$$

נגדיר את הפונקציה $F\colon \mathbb{R}^n \to \mathbb{R}^m$ ע"י

$$F(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) - c_1 \\ f_2(\mathbf{x}) - c_2 \\ \vdots \\ f_m(\mathbf{x}) - c_m \end{pmatrix}$$

ונסמן את מטריצת היעקוביאן של F בנקודה $\mathbf{x} \in \mathbb{R}^n$ ב־$\mathbf{J}(\mathbf{x})$.

א. הוכיחו כי $\nabla g(\mathbf{x}) = 2\mathbf{J}(\mathbf{x})^T F(\mathbf{x})$.

שיטת גאוס־ניוטון היא שיטה למציאת ממזער של g ע"י קירוב לינארי. האיטרציות מוגדרות באופן הבא

$$\mathbf{x}^{k+1} = \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^m \left(f_i(\mathbf{x}^k) + \nabla f_i(\mathbf{x}^k)^T (\mathbf{x} - \mathbf{x}^k) - c_i \right)^2$$

ב. הראו ששיטת גאוס־ניוטון פותרת בכל איטרציה k בעיית ריבועים פחותים מהצורה

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left\| \mathbf{J}(\mathbf{x}^k) \mathbf{x} - \mathbf{b}^k \right\|^2$$

כאשר $\mathbf{b}^k \in \mathbb{R}^m$ וקטור התלוי ב־F ובמטריצת היעקוביאן שלה. בתשובתכם כיתבו את $\mathbf{b}^k$ באופן מפורש ע"י $\mathbf{J}(\mathbf{x}^k)$ ו־$F(\mathbf{x}^k)$.

ג. הניחו ש־$\mathbf{J}(\mathbf{x}^k)$ מדרגת עמודות מלאה לכל k. הראו ששיטת גאוס־ניוטון היא למעשה שיטת גרדיאנט מדורגת המופעלת על g. כיתבו במפורש את המטריצה המדרגת $\mathbf{D}^k$ והסבירו מדוע היא מוגדרת חיובית לכל איטרציה k.