הטכניון, מכון טכנולוגי לישראל
הפקולטה להנדסת חשמל

שיטות חישוביות באופטימיזציה

046197

סמסטר אביב תשפ"א 2021

מגיש:

עידו יחזקאל	204397368

תרגיל 1-

תהי f פונקציה L-חלקה וחסומה מלמטה על ידי f^*

כלל עדכון הגרדיאנט: $X_t = X_{t-1} - \eta \nabla f(x_{t-1})$

נניח כי $\eta \le 1/L$

1. צ"ל $f(x_t) \le f(x_{t-1}) - \eta\left(1-\frac{\eta L}{2}\right)\|\nabla f(x_{t-1})\|_2^2$

<u>הוכחה-</u> לפי הגדרת L-smooth:

$$f(x_t) \le f(x_{t-1}) + \nabla f(x_{t-1})^T (x_t - x_{t-1}) + \frac{L}{2}\|x_t - x_{t-1}\|_2^2$$

$$x_t - x_{t-1} = -\eta \nabla f(x_{t-1})$$

$$\Rightarrow f(x_t) \le f(x_{t-1}) + \nabla f(x_{t-1})^T\left(-\eta \nabla f(x_{t-1})\right) + \frac{L}{2}\eta^2 \|\nabla f(x_{t-1})\|_2^2$$

$$f(x_t) \le f(x_{t-1}) - \eta \|\nabla f(x_{t-1})\|_2^2 + \frac{L}{2}\eta^2 \|\nabla f(x_{t-1})\|_2^2$$

$$f(x_t) \le f(x_{t-1}) - \eta\left(1 - \frac{L}{2}\eta\right)\|\nabla f(x_{t-1})\|_2^2 \quad \blacksquare$$

2. צ"ל $\frac{2}{\eta}\left(f(x_{t-1}) - f(x_t)\right) \ge \|\nabla f(x_{t-1})\|_2^2$

<u>מסעיף 1</u> $\quad \eta\left(1-\frac{L}{2}\eta\right)\|\nabla f(x_{t-1})\|_2^2 \le f(x_{t-1}) - f(x_t)$

$$\eta\left(1 - \frac{L}{2}\cdot\frac{1}{L}\right) \le \eta\left(1 - \frac{L}{2}\eta\right) \quad \Rightarrow \frac{1}{2}\eta \|\nabla f(x_{t-1})\|_2^2 \le f(x_{t-1}) - f(x_t)$$

$$\Leftrightarrow \frac{1}{2}\eta \le \eta\left(1 - \frac{L}{2}\eta\right) \qquad \Rightarrow \|\nabla f(x_{t-1})\|_2^2 \le \frac{2}{\eta}\left(f(x_{t-1}) - f(x_t)\right)$$

$$\Leftrightarrow \frac{1}{2} \le 1 - \frac{L}{2}\eta$$

$$\Leftrightarrow \eta \le 1/L \qquad \blacksquare$$

3. צ"ל
$$\sum_{t=0}^{T} \|\nabla f(x_t)\|_2^2 \le \frac{2}{\eta}\left(f(x_0) - f^*\right)$$

מסעיף 2:
$$\|\nabla f(x_t)\|_2^2 \le \frac{2}{\eta}\left(f(x_t) - f(x_{t+1})\right)$$

$$\sum_{t=0}^{T} \|\nabla f(x_t)\|_2^2 \le \sum_{t=0}^{T} \frac{2}{\eta}\left(f(x_t) - f(x_{t+1})\right)$$

$$= \frac{2}{\eta}\left[f(x_0) - \cancel{f(x_1)} + \cancel{f(x_1)} - \cancel{f(x_2)} + \dots + \cancel{f(x_T)} - f(x_{T+1})\right]$$

טור טלסקופי

$$= \frac{2}{\eta}\left[f(x_0) - f(x_{T+1})\right] \underset{f^* \le f(x_{T+1})}{\le} \frac{2}{\eta}\left(f(x_0) - f^*\right)$$

$$\sum_{t=0}^{T} \|\nabla f(x_t)\|_2^2 \le \frac{2}{\eta}\left(f(x_0) - f^*\right) \quad \blacksquare$$

4. $$\min_{\tau=0,\dots,t} \|\nabla f(x_\tau\|_2 \le \sqrt{\frac{2}{\eta(t+1)}\left(f(x_0) - f^*\right)}$$

מסעיף 3:
$$\frac{1}{t+1} \cdot \sum_{\tau=0}^{t} \|\nabla f(x_\tau)\|_2^2 \le \frac{2}{\eta(t+1)}\left(f(x_0 - f^*\right)$$

$$\min_{\tau=0,\dots,t} \|\nabla f(x_\tau)\|_2^2 \le \mathbb{E}\left[\|\nabla f(x_\tau)\|_2^2\right] \le \frac{2}{\eta(t+1)}\left(f(x_0) - f^*\right)$$

נשים לב כי מכיוון ש-$0 \le \|\nabla f(x_\tau)\|$ הרי שאותו מינימום מתקבל על ידי הוצאת שורש, שהיא מונוטונית עבור ערכים אי שליליים.

$$\min_{\tau=0,\dots,t} \|\nabla f(x_\tau)\|_2 \le \sqrt{\frac{2}{\eta(t+1)}\left(f(x_0) - f^*\right)}$$

תרגיל 2-

יהי $f:\mathbb{R}^n \to \mathbb{R}$ פונקציה קמורה - convex ו- L-smooth

מריצים אלגוריתם GD עם עדכון מהצורה:

$$\{\eta_t\}_{t\ge 0} : \quad x_{t+1} = x_t - \eta_t \nabla f(x_t)$$

נסמן:

$$h_t \triangleq f(x_t) - f(x^*) ; \quad d_t \triangleq \|x_t - x^*\|$$

1. צ"ל: לכל $t \ge 0$

$$d_{t+1}^2 \le d_t^2 - 2\eta_t h_t + \eta_t^2 \|\nabla f(x_t)\|_2^2$$

הוכחה: מקמירות של f:

$$f(x^*) \ge f(x_t) + \nabla f(x_t)^T (x^* - x_t)$$

$$\Rightarrow \nabla f(x_t)^T (x_t - x^*) \ge f(x_t) - f(x^*) = h_t$$

$$\|x_{t+1} - x^*\|_2^2 = \|x_t - \eta_t \nabla f(x_t) - x^*\|_2^2 =$$

$$\|x_t - x^*\|_2^2 - 2\eta_t \nabla f(x_t)^T (x_t - x^*) + \eta_t^2 \|\nabla f(x_t)\|_2^2$$

$$\Rightarrow d_{t+1}^2 = d_t^2 - 2\eta_t \nabla f(x_t)^T (x_t - x^*) + \eta_t^2 \|\nabla f(x_t)\|_2^2$$

$$\Rightarrow d_{t+1}^2 \le d_t^2 - 2\eta_t h_t + \eta_t^2 \|\nabla f(x_t)\|_2^2 \quad \blacksquare$$

ונבחר גודל צעד בזמן t:

$$\eta_t = \frac{h_t}{\|\nabla f(x_t)\|_2^2}$$

צ"ל ②

$$d_{t+1}^2 - d_t^2 \le \frac{-h_t}{2L}$$

$$d_{t+1}^2 - d_t^2 \le -2\eta_t h_t + \eta_t^2 \|\nabla f(x_t)\|_2^2$$

$$= \frac{-2h_t^2}{\|\nabla f(x_t)\|_2^2} + \frac{h_t^2}{\|\nabla f(x_t)\|_2^4}\|\nabla f(x_t)\|_2^2$$

$$= \frac{-h_t^2}{\|\nabla f(x_t)\|_2^2}$$

<u>מהלמה:</u>

$$* \ \|\nabla f(x_t)\|_2^2 \le 2L\left(f(x_t) - f(x^*)\right) \le 2L\, h_t$$

$$d_{t+1}^2 - d_t^2 \le \frac{-h_t^2}{\|\nabla f(x_t)\|_2^2} \underset{\uparrow}{\le} \frac{-h_t^2}{2L h_t} = -\frac{h_t}{2L}$$

שבר שלילי ולכן הקטנתי ע"י הגדלת מכנה
מגדילה את הביטוי.

.3 צ״ל לכל $T \geq 1$ מתקיים

$$\sum_{t=0}^{T-1} h_t \leq 2L d_0^2$$

<u>מסעיף 2:</u>

$$h_t \leq 2L\left(d_t^2 - d_{t+1}^2\right)$$

$$\sum_{t=0}^{T-1} h_t = 2L \sum_{t=0}^{T-1}\left(d_t^2 - d_{t+1}^2\right) = 2L\left[d_0^2 - d_1^2 + d_1^2 - d_2^2 + \dots + d_{T-1}^2 - d_T^2\right]$$

<u>טור טלסקופי:</u>

$$\leq 2L\left(d_0^2 - d_T^2\right) \leq 2L d_0^2.$$

$$\sum_{t=0}^{T-1} f(x_t) - f(x^*) \leq 2L d_0^2 \quad /:T$$

$$\sum_{t=0}^{T-1} \frac{1}{T}\left(f(x_t) - f(x^*)\right) \leq \frac{2L d_0^2}{T}$$

$$f(\bar{x}_t) - f(x^*) \underset{\Uparrow}{\leq} \sum_{t=0}^{T-1} \frac{1}{T} f(x_t) - f(x^*) \leq \frac{2L d_0^2}{T}$$

אי שוויון ינסן.

תרגיל 3:

נתונה בעיית אופטימיזציה חד מימדית הבאה:

$$\max_{0\le x\le 5}\left\{2\sqrt{5-x}+\frac{x}{\sqrt{5}}\right\}$$

צ"ל פתרון אופטימלי הוא $2\sqrt{5}$.

הוכחה:

$$f(x)=2\sqrt{5-x}+\frac{x}{\sqrt{5}}\ ;\quad 0\le x\le 5$$

$$f'(x)=2\cdot\frac{1}{2\sqrt{5-x}}\cdot(-1)+\frac{1}{\sqrt{5}}$$

$$=\frac{1}{\sqrt{5}}-\frac{1}{\sqrt{5-x}}$$

נראה כי נגזרת שלילית לכל $0\le x$:

$$\frac{1}{\sqrt{5}}-\frac{1}{\sqrt{5-x}}\le 0$$

$$\Leftrightarrow \frac{1}{\sqrt{5}}\le\frac{1}{\sqrt{5-x}}$$

$$\Leftrightarrow \sqrt{5-x}\le\sqrt{5}$$

$$\Leftrightarrow 5-x\le 5$$

$$x\ge 0$$

ולכן בתחום הנגזרת היא שלילית $\Rightarrow$ הפונקציה יורדת

ולכן מקסימום מתקבל על השפה $x=0$

$$\max_{0\le x\le 5} f(x)=f(0)=2\sqrt{5-0}+\frac{0}{\sqrt{5}}=2\sqrt{5}$$

ולכן $2\sqrt{5-x}+\frac{x}{\sqrt{5}}\le 2\sqrt{5}\ \ \forall\ 0\le x\le 5.$

2. ג׳)

$$\sum_{i=1}^{n} \frac{a_i}{\sqrt{\sum_{j=1}^{i} a_j}} \le 2\sqrt{\sum_{i=1}^{n} a_i}$$

לכל סדרת מספרים $a_1,\dots,a_n$ אי שליליים.

נוכיח באינדוקציה:

בסיס $i=1$:

$$\sum_{i=1}^{1} \frac{a_i}{\sqrt{\sum_{j=1}^{i} a_j}} = \frac{a_1}{\sqrt{a_1}} = \sqrt{a_1} \le 2\sqrt{a_1}.$$

נניח כי הטענה נכונה עבור $n=k$.

נוכיח כי הטענה נכונה עבור $n=k+1$.

$$\sum_{i=1}^{k+1} \frac{a_i}{\sqrt{\sum_{j=1}^{i} a_j}} = \sum_{i=1}^{k} \frac{a_i}{\sqrt{\sum_{j=1}^{i} a_j}} + \frac{a_{k+1}}{\sqrt{\sum_{j=1}^{k+1} a_j}}$$

$$\underset{\substack{\uparrow \\ \text{הנחת אינדוקציה}}}{\le} 2\sqrt{\sum_{i=1}^{k} a_i} + \frac{a_{k+1}}{\sqrt{\sum_{j=1}^{k+1} a_j}}$$

$$= 2\sqrt{\sum_{i=1}^{k+1} a_i - a_{k+1}} + \frac{a_{k+1}}{\sqrt{\sum_{j=1}^{k+1} a_j}}$$

נסמן $S=\sum_{j=1}^{k+1} a_j$, $X=a_{k+1}$ ונשים לב כי $0 \le a_{k+1} \le \sum_{i=1}^{k+1} a_i$

$$= 2\sqrt{S-X} + \frac{X}{\sqrt{S}} \underset{\substack{\uparrow \\ \text{מטענה 1}}}{\le} 2\sqrt{S} = 2\sqrt{\sum_{i=1}^{k+1} a_i}$$

ולכן הוכחתי באינדוקציה כי הטענה נכונה לכל סדרת מספרים $a_1,\dots,a_n$ אי שליליים.

$\min_{x \in C} f(x)$

תרגיל 4 - נתעניין בבעיה הבאה:

1. צ"ל למה 1:

תהי $f:\mathbb{R}^n \to \mathbb{R}$ פונקציה L חלקה בעלת מינימום גלובלי מתקבל בנקודה x^* אזי לכל $x \in \mathbb{R}^n$

$$\|\nabla f(x)\|_2^2 \le 2L(f(x) - f(x^*))$$

הוכחה-

בתרגיל 1 ראינו כי עבור f L חלקה (ללא הנחת קמירות) ואלגוריתם GD עם גודל צעד קבוע מתקיים:

$$\|\nabla f(x_{t-1})\|_2^2 \le 2L(f(x_{t-1}) - f(x_t)) \quad \forall t \ge 1$$

עבור $t=1$:

$$\|\nabla f(x_0)\|_2^2 \le 2L(f(x_0) - f(x_1)) \le 2L(f(x_0) - f(x^*))$$

נשים לב שבאלגוריתם ה-GD אין הנחות על נקודת ההתחלה x_0 ולכן האי שוויון נכון לכל $x \in \mathbb{R}^n$

$$\|\nabla f(x)\|_2^2 \le 2L(f(x) - f(x^*)) \quad \blacksquare$$

2. צ"ל:

$$\sum_{t=1}^{T}\left(f(x_t) - f(x^*)\right) \le 4LD^2$$

$$\sum_{t=1}^{T}\left(f(x_t) - f(x^*)\right) \le \sqrt{2D^2 \sum_{t=1}^{T} \|g_t\|_2^2}$$

$$\|g_t\|^2 \underset{\uparrow}{=} \|\nabla f(x_t)\|_2^2 \le 2L\left(f(x_t) - f(x^*)\right)$$

f חלקה בתחום.

$$\sum_{t=1}^{T}\left(f(x_t) - f(x^*)\right) \le \sqrt{2D^2 \cdot 2L \sum_{t=1}^{T} f(x_t) - f(x^*)}$$

$$\frac{\sum_{t=1}^{T}\left(f(x_t) - f(x^*)\right)}{\sqrt{\sum_{t=1}^{T} f(x_t) - f(x^*)}} \le \sqrt{4LD^2}$$

$$\sqrt{\sum_{t=1}^{T}\left(f(x_t) - f(x^*)\right)} \le \sqrt{4LD^2} \quad /^2$$

שני צדדים חיוביים.

$$\sum_{t=1}^{T} f(x_t) - f(x^*) \le 4LD^2$$

$$f(\bar{x}_T) - f(x^*) \underset{\uparrow}{\le} \frac{1}{T}\sum_{t=1}^{T} f(x_t) - f(x^*) \le \frac{4LD^2}{T}$$

א. שוויון ינסן

■

תרגיל 5-

נתונה הבעיה:

$$\min_x F(x) = \frac{1}{m}\sum_{i=1}^{m} f_i(x)$$

$\{f_i\}_{i=1}^{m}$ הן L-smooth.

$$g = \nabla f_i(x) \qquad i \sim \text{Uniform}(\{1,\dots,m\})$$

$$x_{t+1} = x_t - \eta g_t$$

סימון:

$$h_t = F(x_t) - F(x^*) \quad ; \quad \eta = \frac{\mu}{L^2} = \frac{\gamma}{L} \quad ; \quad \gamma = \frac{\mu}{L}$$

1. צ"ל $F(x)$ L חלקה:

מכך ש- $f_i(x)$ חלקות:

$$f_i(y) \le f_i(x) + \nabla f_i(x)^T (y-x) + \frac{L}{2}\|y-x\|_2^2$$

$$\sum_{i=1}^{m} f_i(y) \le \sum_{i=1}^{m} f_i(x) + \sum_{i=1}^{m} \nabla f_i(x)^T (y-x) + m\frac{L}{2}\|y-x\|_2^2 \quad / \cdot \frac{1}{m}$$

$$\frac{1}{m}\sum_{i=1}^{m} f_i(y) \le \frac{1}{m}\sum_{i=1}^{m} f_i(x) + \frac{1}{m}\sum_{i=1}^{m} \nabla f_i(x)^T (y-x) + \frac{L}{2}\|y-x\|_2^2$$

$$F(y) \le F(x) + \nabla F(x)^T (y-x) + \frac{L}{2}\|y-x\|_2^2$$

ולכן F היא L-smooth לפי הגדרה.

$$h_{t+1} - h_t \le -\eta \nabla F(x_t)^T g_t + \frac{L}{2}\eta^2 \|g_t\|_2^2$$ ס' 3 .2

הוכחה-

מהגדרת חלקות:

$$F(x_{t+1}) - F(x_t) \le \nabla F(x_t)^T (x_{t+1} - x_t) + \frac{L}{2}\|x_{t+1} - x_t\|_2^2$$

מכלל העדכון:

$$x_{t+1} - x_t = -\eta g_t$$

$$F(x_{t+1}) - F(x^*) - (F(x_t) - F(x^*)) \le \nabla F(x_t)^T(-\eta g_t) + \frac{L}{2}\eta^2 \|g_t\|_2^2$$

$$h_{t+1} - h_t \le -\eta \nabla F(x_t)^T g_t + \frac{L}{2}\eta^2 \|g_t\|_2^2 \quad \blacksquare$$

$$\mathbb{E}[h_{t+1} - h_t] \le -\mu^2 \mathbb{E}[h_t]$$ ס' 3 .3

$$\mathbb{E}[h_{t+1} - h_t] \le \mathbb{E}[-\eta \nabla F(x_t)^T g_t] + \frac{L}{2}\eta^2 \mathbb{E}[\|g_t\|_2^2]$$

נחשב את הביטויים:

$$\mathbb{E}[\|g_t\|_2^2] = \mathbb{E}[\mathbb{E}[\|g_t\|_2^2 | x_t]] \le \mathbb{E}[2L(F(x_t) - F(x^*))]$$
$$\le 2L\,\mathbb{E}[h_t]$$

$$\mathbb{E}[-\eta \nabla F(x_t)^T g_t] = \mathbb{E}[\mathbb{E}[-\eta \nabla F(x_t)^T g_t | x_t]]$$
$$= \mathbb{E}[-\eta \nabla F(x_t)^T \mathbb{E}[g_t | x_t]] = \mathbb{E}[-\eta \nabla F(x_t)^T \nabla F(x_t)]$$

מאומד בלתי מוטה.

$$= -\eta\, \mathbb{E}[\|\nabla F(x_t)\|_2^2] \le -\eta\, 2\mu\, \mathbb{E}[F(x_t) - F(x^*)] \le -2\mu\eta\, \mathbb{E}[h_t]$$

ונחבר את האי-שוויונות:

$$\mathbb{E}[h_{t+1} - h_t] \le -2\mu\eta\,\mathbb{E}[h_t] + \frac{L}{2}\eta^2 \cdot 2L\,\mathbb{E}[h_t]$$

$$= \left(-2\mu \cdot \frac{\mu}{L^2} + \cancel{L^2} \cdot \frac{\mu^2}{L^{\cancel{4}2}}\right)\mathbb{E}[h_t] = (-2\gamma^2 + \gamma^2)\,\mathbb{E}[h_t]$$

$$\Rightarrow \mathbb{E}[h_{t+1} - h_t] \le -\gamma^2\,\mathbb{E}[h_t]$$

$$\mathbb{E}[h_t] \le e^{-\gamma^2 t}\,\mathbb{E}[h_0]$$ ג .4

הוכחה:

מסעיף 3:

$$\mathbb{E}[h_{t+1}] - \mathbb{E}[h_t] \le -\gamma^2\,\mathbb{E}[h_t]$$

$$\mathbb{E}[h_{t+1}] \le (1-\gamma^2)\,\mathbb{E}[h_t] \qquad \forall t \ge 0$$

לכן:

$$\mathbb{E}[h_t] \le (1-\gamma^2)\,\mathbb{E}[h_{t-1}] \le (1-\gamma^2)^2\,\mathbb{E}[h_{t-2}] \le \dots \le (1-\gamma^2)^t\,\mathbb{E}[h_0]$$

ועוד לכן:

$\forall x, y \in \mathbb{R}^n$

$$F(y) \le F(x) + \nabla F(x)^T (y-x) + \frac{L}{2}\|y-x\|_2^2$$ הגדרת L-smooth

הגדרת Strongly convex

$$F(y) \ge F(x) + \nabla F(x)^T (y-x) + \frac{\mu}{2}\|y-x\|_2^2$$

$$\Rightarrow F(x) + \nabla F(x)^T (y-x) + \frac{\mu}{2}\|y-x\|_2^2 \le F(x) + \nabla F(x)^T (y-x) + \frac{L}{2}\|y-x\|_2^2$$

$$\Rightarrow \mu/2 \le L/2 \Rightarrow \mu/L \le 1 \Rightarrow \boxed{\gamma \le 1}$$

ונראה כי $e^{-\gamma^2} \geq (1-\gamma^2)$ עבור $0 < \gamma \leq 1$

$$\omega(\gamma) = e^{-\gamma^2} - (1-\gamma^2)$$

$$\omega'(\gamma) = -2\gamma e^{-\gamma^2} + 2\gamma$$

$$2\gamma(1 - e^{-\gamma^2}) \geq 0$$

$$\iff \gamma > 0 \qquad e^{-\gamma^2} \leq 1$$

$$\iff 1 \leq e^{\gamma^2}$$

$$\iff 0 < \gamma$$

ולכן עבור $\gamma > 0$ הפונקציה בעלת נגזרת אי שלילית ולכן לא יורדת בנוסף עבור

$$\omega(0) = 1 - (1-0) = 0$$

ולכן בתחום $0 < \gamma \leq 1$ הפונקציה $\omega(\gamma)$ לא יורדת

ובפרט $\omega(0) = 0$ ולכן $\omega(\gamma) = e^{-\gamma^2} - (1-\gamma^2) \geq 0$

$$\mathbb{E}[h_t] \leq (1-\gamma^2)^t \, \mathbb{E}[h_0] \leq e^{-\gamma^2 t} \, \mathbb{E}[h_0]$$

∎