314092826 046197 יקטרינה גרימנוביץ' HW3

תרגיל 1

אם f קמורה ו-L-חלקה, האלגוריתם גרדיאנט בירידה עם גודל צעד $\eta \le \frac{1}{L}$ מתכנס למינימום גלובלי של f בקצב $O(\frac{1}{t})$. בשאלה זו ננתח תוצאות התכנסות של אלגוריתם גרדיאנט עבור f שהיא לא בהכרח קמורה.

תהיה f פונקציה L-חלקה ונסמן מינימום ע"י f^*. עבור גודל צעד קבוע: $t \ge 1$ $x_t = x_{t-1} - \eta \nabla f(x_{t-1})$

נניח כי $\eta \le 1/L$

(1) נוכיח כי

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

לפי הגדרה של f חלקות מתקיים:

f, כאשר היא $f: R^d \to R$ תהיה L-חלקה אם היא קמורה ומגזירה $\forall x,y \in R^d$:

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

לפי הנתון f היא L-חלקה ולכן נתייחס למקרה פרטי, בו

$x = x_{t-1}$

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

נציב ונקבל:

$$\begin{aligned} f(x_t) &\le f(x_{t-1}) + \nabla f(x_{t-1})^T(x_t - x_{t-1}) + \frac{L}{2}\|x_t - x_{t-1}\|_2^2 = \\ &= f(x_{t-1}) + \nabla f(x_{t-1})^T(x_{t-1} - \eta\nabla f(x_{t-1}) - x_{t-1}) + \frac{L}{2}\|x_{t-1} + \eta\nabla f(x_{t-1}) - x_{t-1}\|_2^2 \\ &= f(x_{t-1}) + \nabla f(x_{t-1})^T(+\eta\nabla f(x_{t-1})) + \frac{L}{2}\|+\eta\nabla f(x_{t-1})\|_2^2 \\ &= f(x_{t-1}) - \eta\nabla f(x_{t-1})^T\nabla f(x_{t-1}) + \frac{L}{2}|\eta|^2\|\nabla f(x_{t-1})\|_2^2 = \\ &= f(x_{t-1}) - \eta\|\nabla f(x-1)\|_2^2 + \frac{L}{2}\eta^2\|\nabla f(x_{t-1})\|_2^2 = \\ &= f(x_{t-1}) - \eta\left(1 - \frac{\eta L}{2}\right)\|\nabla f(x_{t-1})\|_2^2 \end{aligned}$$

$\nabla f^T(x_{t-1})\nabla f(x_{t-1}) = \|\nabla f(x_{t-1})\|_2^2$

$\eta \ge 0 \Rightarrow |\eta|^2 = \eta^2$

(2) יש להיעזר ב $\eta \le \frac{1}{L}$ ולהראות כי $\|\nabla f(x_{t-1})\|_2^2 \le \frac{2}{\eta}(f(x_{t-1}) - f(x_t))$

לפי הביטוי שהוכחנו בסעיף (1) נסדר אגפים ונקבל:

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

נחלק $:\eta$

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

$\times 2$

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

לפי ביטוי $\eta \le \frac{1}{L} \Leftarrow$

$$\begin{aligned} L\eta &\le 1 \\ -L\eta &\ge -1 \\ 2 - L\eta &\ge 2 - 1 \\ 2 - L\eta &\ge 1 \end{aligned}$$

לכן:

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

ולכן הביטוי מתקיים לכל $t \ge 1$

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

(3) יש להוכיח כי:

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

אם נחבר את כל ה $t = 1 \dots T+1$ של הביטוי שהוכחנו בסעיף הקודם, נקבל כי:

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

בנוסף גם מתקיים כי

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

$\sum_{t=0}^{T+1} f(x_{t-1}) = f_{x_0} + \sum_{t=0}^{T} f(x_t)$

$\sum_{t=0}^{T+1} f(x_t) = \sum_{t=0}^{T} f(x_t) + f(x_{T+1})$

ועבור אגף הימני:

$$\sum_{t=0}^{T+1}\frac{2}{\eta}(f(x_{t-1}) - f(x_t)) = \sum_{t=0}^{T+1}\frac{2}{\eta}f(x_{t-1}) - \sum_{t=0}^{T+1}\frac{2}{\eta}f(x_t) = \frac{2}{\eta}\sum_{t=0}^{T+1}f(x_{t-1}) - \frac{2}{\eta}\sum_{t=0}^{T+1}f(x_t) =$$

$$= \frac{2}{\eta}\left(f(x_0) + \sum_{t=0}^{T}f(x_t)\right) - \frac{2}{\eta}\left(\sum_{t=0}^{T}f(x_t) + f(x_{T+1})\right) = \frac{2}{\eta}(f(x_0) - f(x_{T+1}))$$

נשים לב כי לפי הנתון f^* הוא החסם מלמטה של f, לכן $f^* \le f(x_{T+1})$ ולכן:

$f(x_0) - f^* \ge f(x_0) - f(x_{T+1})$

ולכן

$$\sum_{t=0}^{T}\|\nabla f(x_t)\|_2^2 \le \sum_{t=0}^{T+1}\|\nabla f(x_{t-1})\|_2^2 \le \frac{2}{\eta}(f(x_0) - f(x_{T+1})) \le \frac{2}{\eta}(f(x_0) - f^*)$$

314092826 046197 יקטרינה גרמוביץ' HW3

תרגיל 1 המשך.

(4) חסמו בעזרת את הסכום מסעיף הקודם על מנת לקבל $\min_{\tau=0\dots t}\|\nabla f(x_\tau)\|_2 \le \sqrt{\frac{2}{\eta(t+1)}(f(x_0)-f^*)}$
שים לב כי הראינו התכנסות בקצב של $O(\frac{1}{\sqrt{t}})$ או $\Theta(\frac{1}{\varepsilon^2})$
לנקודה שהיא ε-תת-סטציונרית, כלומר, נקודה x המקיימת $\|\nabla f(x)\|_2 \le \varepsilon$

נתבסס על סעיף הקודם, כאשר $\tau = t$, $T=t$ (נחליף ונקבל):

$$\sum_{\tau=0}^{t}\|\nabla f(x_\tau)\|_2^2 \le \frac{2}{\eta}(f(x_0)-f^*) \qquad (1)$$

על פי הגדרת המינימום מתקיים כי:
מתוך: $\|\nabla f(x_\tau)\|_2^2 \ge \min_{\tau=0..t}\|\nabla f(x_\tau)\|_2^2$ נקבל

$$\sum_{\tau=0}^{t}\|\nabla f(x_\tau)\|_2^2 \ge (t+1)\cdot \min\|\nabla f(x_\tau)\|_2^2 = (t+1)\left(\min_{\tau=0\dots t}\|\nabla f(x_\tau)\|_2\right)^2$$

נציב בחזרה ב-(1):

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

$t+1>0$

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

ניתן להוציא שורש הואיל ושני צדי האי שוויון חיוביים

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

תרגיל 2

תהיה $f:\mathbb{R}^n\to\mathbb{R}$ פ' קמורה וחלקה. מריצים אלגוריתם הגרדיאנט עם סדרת גדלי צעד $\{\eta_t\}_{t\ge0}$
כלומר, העדכון יהיה: לכל $t\ge0$ $x_{t+1} = x_t - \eta_t\nabla f(x_t)$
נסמן ממזער של f ע"י x^*,
נסמן $h_t \triangleq f(x_t)-f(x^*)$ $d_t \triangleq \|x_t - x^*\|$

(1) יש להוכיח כי לכל $t\ge0$ מתקיים: $d_{t+1}^2 \le d_t^2 - 2\eta_t h_t + \eta_t^2\|\nabla f(x_t)\|_2^2$

ניתן להיעזר בקמירות של f על מנת לחסום מלמעלה את $\nabla f(x_t)^T(x_t-x^*)$

מתוך הגדרה כי f - פונקציה קמורה עבור $\forall x,y\in\mathbb{R}^n$ מתקיים: $f(y)-f(x)\le\nabla f(y)^T(y-x)$ (1)
נבחר כי $x=x_t$, $y=x^*$ לכן מתוך הביטוי נקבל:

נציב ב(1)

$$f(y)-f(x) = f(x^*)-f(x_t) = -h_t$$
$$(y-x) = (x^*-x_t)$$
$$-h_t \le \nabla f(x^*)^T(x^*-x_t) \iff -\nabla f(x^*)^T(x^*-x_t)\le h_t$$
$$\nabla f(x^*)^T(x_t-x^*)\le h_t \Rightarrow \nabla f(x^*)^T d_t \le h_t$$

$$d_{t+1}^2 = \|x_{t+1}-x^*\|_2^2 = \|(x_{t+1}-x_t)+(x_t-x^*)\|_2^2 = \|x_{t+1}-x_t\|_2^2 + 2(x_{t+1}-x_t)^T\overbrace{(x_t-x^*)}^{d_t} + \overbrace{\|x_t-x^*\|_2^2}^{d_t^2} =$$

בנוסף: $x_{t+1} - x_t - \eta_t\nabla f(x_t) \Leftarrow x_{t+1}-x_t = -\eta_t\nabla f(x_t)$

$$= \|-\eta_t\nabla f(x_t)\|_2^2 - 2\eta\nabla f(x_t)^T d_t + d_t^2 = ②$$

נבחר כי $x=x^*$, $y=x_t$ לכן מתוך הביטוי של תכונת הקמירות (1) נקבל:

$$\left.\begin{array}{l} f(y)-f(x) = f(x_t)-f(x^*) = h_t \\ (y-x) = (x_t-x^*)\end{array}\right\} \to h_t \le \nabla f(x_t)^T\cdot(x_t-x^*) \Rightarrow -\nabla f(x_t)^T(x_t-x^*)\le -h_t$$

נציב את הביטוי הזה ב-②

$$= ② \le \|-\eta_t\nabla f(x_t)\|_2^2 - 2\eta_t h_t + d_t^2 = |\eta_t|^2\|\nabla f(x_t)\|_2^2 - 2\eta_t h_t + d_t^2 = d_t^2 - 2\eta_t h_t + \eta_t^2\|\nabla f(x_t)\|_2^2$$

314092826 046197 יקטרינה גרימוביץ' HW3

תרגיל 2 המשך

(2) הוכיחו כי לכל $t\ge 0$ מתקיים $d_{t+1}^2 - d_t^2 \le -\frac{h_t}{2L}$

נבחר גודל הצעד בזמן t להיות: $\eta_t = \frac{h_t}{\|\nabla f(x_t)\|_2^2}$

נעזר בלמה 1: תהי $f:\mathbb{R}^n\to\mathbb{R}$ פונקציה L-חלקה בעלת מינימום גלובלי בנק' x^*.
אז לכל $x\in\mathbb{R}^n$ מתקיים: $\|\nabla f(x)\|_2^2 \le 2L(f(x)-f(x^*))$

לפי סעיף 1 ראינו כי $d_{t+1}^2 \le d_t^2 - 2\eta_t h_t + \eta_t^2 \|\nabla f(x_t)\|_2^2$ לכל $t\ge 0$, לכן:

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

נציב בביטוי הזה את גודל הצעד הנבחר:

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

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

כעת מתוך למה 1 עבור $x=x_t$ נקבל: $\|\nabla f(x_t)\|_2^2 \le 2L(f(x_t)-f(x^*))$

$$h_t = f(x_t)-f(x^*) \Rightarrow \|\nabla f(x_t)\|_2^2 \le 2L\cdot h_t$$

$$\frac{1}{2L} \le \frac{h_t}{\|\nabla f(x_t)\|_2^2} \to -\frac{h_t}{\|\nabla f(x_t)\|_2^2} \le -\frac{1}{2L}$$

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

(3) הוכיחו כי לכל $T\ge 1$ מתקיים $\sum_{t=0}^{T-1} h_t \le 2Ld_0^2$

והסיקו כי $f(\bar{x}_T) - f(x^*) \le \frac{2Ld_0^2}{T} = \frac{2L\|x_0-x^*\|_2^2}{T}$ כאשר $\bar{x}_T = \frac{1}{T}\sum_{t=0}^{T-1} x_t$

נוכיח קודם את הביטוי $\sum_{t=0}^{T-1} h_t \le 2Ld_0^2$. מתוך הסעיף (2) ראינו כי לכל $t\ge0$ מתקיים
$d_{t+1}^2 - d_t^2 \le -\frac{h_t}{2L}$

אם נבודד את h_t נקבל: $h_t \le 2L(d_t^2 - d_{t+1}^2)$

עבור $t=0,\dots,T-1$ ניתן לרשום סכום: בדומה לתרגיל 1:

סכום טלסקופי

$$\sum_{t=0}^{T-1} h_t \le \sum_{t=0}^{T-1} 2L(d_t^2 - d_{t+1}^2) = \sum_{t=0}^{T-1} 2L d_t^2 - \sum_{t=0}^{T-1} 2L d_{t+1}^2 =$$

$$= \left(\sum_{t=1}^{T-1} 2Ld_t^2 + 2Ld_0^2\right) - \sum_{t=1}^{T} 2Ld_t^2 = 2L\left[\left(\sum_{t=1}^{T-1} d_t^2 + d_0^2\right) - \left(\sum_{t=1}^{T-1} d_t^2 + d_T^2\right)\right]$$

$$= 2L(d_0^2 - d_T^2) = 2Ld_0^2 \underbrace{-2Ld_T^2}_{\text{הביטוי הזה לא חיובי}} \le 2Ld_0^2$$

לכן $\sum_{t=0}^{T-1} h_t \le 2Ld_0^2$

נציב בביטוי את h_t, d_t ונקבל:

$$\sum_{t=0}^{T-1}(f(x_t)-f(x^*)) \le 2L\|x_0-x^*\|_2^2 \quad :T$$

$$\frac{1}{T}\left(\sum_{t=0}^{T-1}(f(x_t)-f(x^*))\right) \le \frac{2L\|x_0-x^*\|_2^2}{T} \quad (1)$$

מכיוון ש-f היא פונקציה קמורה ולכן מתקיים א"ש Jensen: יהיו $x_1,\dots,x_N\in C$, $\lambda_1,\dots,\lambda_N\ge 0$ כך ש- $\sum_{i=1}^N \lambda_i = 1$: $f\left(\sum_{i=1}^N \lambda_i x_i\right) \le \sum_{i=1}^N \lambda_i f(x_i)$

במקרה שלנו נתון כי: $\bar{x}_T = \frac{1}{T}\sum_{t=0}^{T-1} x_t$:

$$f(\bar{x}_T) = f\left(\frac{1}{T}\sum_{t=0}^{T-1} x_t\right) = f\left(\sum_{t=0}^{T-1}\frac{1}{T}x_t\right) \le \sum_{t=0}^{T-1}\frac{1}{T}f(x_t)$$

$$f(\bar{x}_T) \le \frac{1}{T}\sum_{t=0}^{T-1} f(x_t)$$

$$\cdot T \quad T(f(\bar{x}_T) - f(x^*)) \le T\left(\frac{1}{T}\sum_{t=0}^{T-1} f(x_t) - f(x^*)\right) = \frac{T}{T}\sum_{t=0}^{T-1} f(x_t) - Tf(x^*) = \sum_{t=0}^{T-1} f(x_t) - \sum_{t=0}^{T-1} f(x^*) = \sum_{t=0}^{T-1}(f(x_t)-f(x^*)) \quad (2)$$

מתוך (1)+(2) נקבל:

$$f(\bar{x}_T) - f(x^*) \le \frac{1}{T}\left(\sum_{t=0}^{T-1}(f(x_t)-f(x^*))\right) \le \frac{2L\|x_0-x^*\|_2^2}{T}$$

HW3 גרימוביץ' יקטרינה 046197 314092826

תרגיל 3

(1) נתונה בעיית אופטימיזציה 1D : $\max_{0 \le x \le S} \left\{ 2\sqrt{S-x} + \frac{x}{\sqrt{S}} \right\}$

הוכיחו כי פתרון האופטימלי של הבעיה הוא $2\sqrt{S}$ והסיקו כי לכל $0 \le x \le S$ מתקיים $2\sqrt{S-x} + \frac{x}{\sqrt{S}} \le 2\sqrt{S}$

הדרכה: הראו כי נגזרת של פ' מטרה היא שלילית בתחום. הסיקו כי פתרון מתקבל על השפה.

נגדיר $f(x) = 2\sqrt{S-x} + \frac{x}{\sqrt{S}}$ היא פ' מטרה. על מנת למצוא את הפתרון האופטימלי נמצא את $f(x)$. פ' זאת גזירה בתחום הנתון, בפרט:

$$f'(x) = \frac{d}{dx}\left(2\cdot(S-x)^{\frac{1}{2}} + \frac{1}{\sqrt{S}}x\right) = \frac{1}{\sqrt{S}} - \frac{1}{\sqrt{S-x}}$$

נראה כי נגזרת של פ' מטרה הוא שלילית.

נניח כי $f'(x) < 0$:

$$\frac{1}{\sqrt{S}} - \frac{1}{\sqrt{S-x}} < 0$$

$$\frac{\sqrt{S-x} - \sqrt{S}}{\sqrt{S(S-x)}} < 0 \quad \rightarrow \quad \text{מכנה הוא חיובי}$$

עבור $x \in (0,S)$, בנוסף $\sqrt{S-x} < \sqrt{S}$ בתחום חיובי, לכן המונה מתקבל תמיד שלילי ולכן $f'(x) < 0$ עבור $x \in (0,S)$ כלומר בתחום הפנימי של תחום ההגדרה של הבעיה ולכן הפונקציה הוא יורדת לאורך כל התחום הנתון והמקסימום מתקבל בקצה השמאלי שלו כלומר ב $x=0$: $f^*(x=0) = 2\sqrt{S}$

לכן מאמת $f^* = 2\sqrt{S}$ הוא ערך המקסימלי שהפונקציה יכולה לקבל בתחום הנתון:
כלומר: לכל $0 \le x \le S$ מתקיים: $f(x) \le f^*$ ולכן ניתן להסיק כי

$$f(x) = \underline{2\sqrt{S-x} + \frac{x}{\sqrt{S}}} \le \max_{0 \le x \le S}\left\{2\sqrt{S-x} + \frac{x}{\sqrt{S}}\right\} \underline{= 2\sqrt{S}} = f^*$$

$$2\sqrt{S-x} + \frac{x}{\sqrt{S}} \le 2\sqrt{S} \quad \forall x: 0 \le x \le S.$$

(2) הוכיחו את למה 5 מופיעה בתרגיל 3 מתרגול 6.
לכל סדרת מספרים אי שליליים $a_1 \dots a_n$ מתקיים:

$$(*) \quad \sum_{i=1}^{n} \frac{a_i}{\sqrt{\sum_{j=1}^{i} a_j}} \le \underbrace{2\sqrt{\sum_{i=1}^{n} a_i}}_{(2)}$$

מתוך סעיף הקודם: $\forall x: 0 \le x \le S$ מתקיים: $2\sqrt{S-x} + \frac{x}{\sqrt{S}} \le 2\sqrt{S}$. (1) $\quad S>0$

נוכיח באינדוקציה את הביטוי.

עבור $\underline{\underline{n=1}}$:

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

עבור n=1 (2) מתקבל $\sqrt{\sum_{i=1}^{1} a_i} = \sqrt{a_1}$

נניח כעת כי הטענה תקפה עבור $\underline{n}$, נוכיח כי הטענה מתקיימת גם עבור $\underline{\underline{n+1}}$ צעדים:
כלומר נרצה להראות כי

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

נסמן (לצורך עבודה נוחה עם (1)) כי $S = \sum_{i=1}^{n+1} a_i$ כאשר $x = a_{n+1}$

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

מתוך הנחה כי הביטוי (*) מתקיים עבור n צעדים

נרשום את הביטוי: $\sqrt{\sum_{i=1}^{n} a_i} = \sqrt{\sum_{i=1}^{n+1} a_i - a_{n+1}}$ ונציב אותו וגם את S, x בביטוי:

$$= \frac{x}{\sqrt{S}} + 2\sqrt{S-x} \underset{\text{נובע מ(1)}}{\le} 2\sqrt{S} \underset{\text{נציב בחזרה}}{=} 2\sqrt{\sum_{i=1}^{n+1} a_i}$$

לכן הביטוי מתקיים עבור n+1, לכן הלמה תקפה לכל צעד וסדרת מספרים אי שליליים $a_1 \dots a_n$

314092826 046197 יקטרינה גרומוביץ' HW3

תרגיל 4

בעיה $\min_{x\in C} f(x)$ כאשר f קמורה ו-C קמורה, קומפקטית ובעלת קוטר D.

אלגוריתם גרדיאנט המוטל עם בחירת גודל צעד AdaGrad: $\eta = \frac{D}{\sqrt{2\sum_{t=1}^{t}\|g_t\|_2^2}}$

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

עבור f פי' ליפשיץ רציפה עם קבוע G הדבר שקול לקצב התכנסות של $O\left(\frac{GD}{\sqrt{T}}\right)$

נראה בשאלה כי אם f היא L-חלקה אז מתקבל קצב התכנסות $O\left(\frac{LD^2}{T}\right)$

(1) יש להוכיח: תהיה $f:\mathbb{R}^n\to\mathbb{R}$ פונקציה L חלקה בעלת מינימום גלובלי מתקבל ב x^*

אז לכל $x\in\mathbb{R}^n$ מתקיים: $\|\nabla f(x)\|_2^2 \le 2L\left(f(x) - f(x^*)\right)$

לפי הרצאה, משפט sufficient decrease:

תהיה $f:\mathbb{R}^d\to\mathbb{R}$ פונקציה גזירה וחלקה עם פרמטר L. עם $\eta := \frac{1}{L}$, GD מקיים כי:

$$f(x_{t+1}) \le f(x_t) - \frac{1}{2L}\|\nabla f(x_t)\|^2 \qquad t \ge 0$$

מתוך משפט הזה וגם מתוך העברת אגפים וסידור מחדש נקבל:

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

$$2L\left(f(x_t) - f(x_{t+1})\right) \ge \|\nabla f(x_t)\|_2^2 \qquad \text{או} \qquad -\|\nabla f(x_t)\|_2^2 \ge \left(f(x_{t+1}) - f(x_t)\right)2L$$

נרצה כעת לסכום את כל t מ-0 עד $t=T-1$

(נשים לב כי כאן יש סכום טלסקופי, כמו שראינו בשאלות נוספות):

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

$$= 2L\left[f(x_0) + \sum_{t=1}^{T-1}f(x_t) - \sum_{t=1}^{T}f(x_t)\right] = 2L\left[f(x_0) + \sum_{t=1}^{T-1}f(x_t) - \sum_{t=1}^{T-1}f(x_t) - f(x_T)\right] =$$

$$= 2L\left[f(x_0) - f(x_T)\right] \qquad (1)$$

בנוסף ניתן לחסום את Σ של נורמה ע' נורמה של אחד האיברים, מכיון שכל הרכיבים של Σ הם אי שליליים, לכן: $\|\nabla f(x_0)\|_2^2 \le \sum_{t=0}^{T-1}\|\nabla f(x_t)\|_2^2$ (2)

בנוסף, מכיון שמדובר בבעיית $\min f(x)$ עם עבור x^* נקבל מינימום גלובלי נחייב להתקיים כי לכל $x\in\mathbb{R}^n$: $f(x^*) \le f(x)$ לכן יתקיים כי:

$$f(x_0) - f(x^*) \ge f(x_0) - f(x)$$

לכן מתוך 1 והביטוי שהתקבל $\sum_{t=0}^{T-1}\|\nabla f(t)\|_2^2 \le 2L\left[f(x_0) - f(x_T)\right] \le 2L\left[f(x_0) - f(x^*)\right]$

ולכן מתוך (2) נקבל: $\|\nabla f(x_0)\|_2^2 \le 2L\left[f(x_0) - f(x^*)\right]$

מכיון שבבחירת x_0 אנחנו הגדרנו אותו כאיבר מתוך הסכום,

ניתן להסיק כי הביטוי שהתקבל יהיה תקף לכל x מתוך ספרת הx ב-$\mathbb{R}^n$ ולכן $\|\nabla f(x)\|_2^2 \le 2L\left[f(x) - f(x^*)\right]$.

תרגיל 4 המשך

נניח כי מינימום הגלובלי של f מתקבל בתוך תחום C ולכן ניתן להשתמש בלמה מהסעיף הקודם לכל $x \in C$

(2) הוכיחו כי עבור f שהיא L-חלקה, אלגוריתם הגרדיאנט המותאם עם בחירת גודל הצעד כמו מקיים $\sum_{t=1}^{T}\left(f(x_t)-f(x^*)\right) \le 4LD^2$ והסיקו כי $f(\bar{x}_T)-f(x^*) \le \frac{4LD^2}{T}$

לפי נתוני השאלה עבור אלגוריתם גרדיאנט המותאם עם בחירת גודל הצעד AdaGrad נקבל כי:

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

על מנת להגיע לביטוי המבוקש,

נרצה לחסום את הביטוי הימני ובמדויק את הגבול של הסכום של $\|g_t\|_2^2$, מתוך הגדרה $g_t=\nabla f(x_t)$

לכן לפי למה מסעיף 1 נקבל $\|g_t\|_2^2=\|\nabla f(x)\|_2^2 \le 2L\left(f(x)-f(x^*)\right)$ לכן אם נבצע סכום: $\sum_{t=1}^{T}\|g_t\|_2^2 \le 2L\sum_{t=1}^{T}\left(f(x)-f(x^*)\right)$

אם נציב את זה בביטוי (*) נקבל:

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

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

נשים לב כי הביטוי $\sum_{t=1}^{T}\left(f(x_t)-f(x^*)\right)>0$ (נמצא בתוך התחום C) של הביטוי הוא חיובי כי x^* הוא מינימום הגלובלי שמתקבל בתחום
לכן ניתן להסיק כי אם $\sqrt{\sum_{t=1}^{T}\left(f(x_t)-f(x^*)\right)}>0$ נחלק בו:

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

בנוסף אם ניקח $\bar{x}_T=\frac{1}{T}\sum_{t=1}^{T}x_t$ ראינו כי

$$f(\bar{x}_T)-f(x^*) \le \frac{1}{T}\left(\sum_{t=1}^{T}\left(f(x_t)-f(x^*)\right)\right)$$

לכן:

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

314092826 046197 גרימוביץ' יקטרינה HW3

תרגיל 5

נתון מינימיזציה: $\min_x F(x) := \frac{1}{m}\sum_{i=1}^{m} f_i(x)$ כאשר $\{f_i\}_{i=1}^{m}$ הן L חלקות.

מתרגיל 4, תרגיל 2, תחת הנחת reliability, קצב התכנסות של אלגוריתם גרדיאנט הסטוכסטי הינו $O(\frac{1}{\epsilon})$ (לעומת $O(\frac{1}{\epsilon^2})$ עבור מקרה כללי).

נראה כי אם F קמורה במובן חזק עם קבוע μ, אז קצב התכנסות אלגוריתם $O(e^{-\gamma^2 t})$ (לעומת $\Theta(\frac{1}{\epsilon})$ עבור מקרה כללי) $\gamma = \mu/L$

נניח כי משערך הגרדיאנט: $i \sim Uniform(\{1..m\})$, $g = \nabla f_i(x)$, וכלל עדכון: $x_{t+1} = x_t - \eta g_t$

וגודל הצעד: $\eta = \frac{\mu}{L^2} = \frac{\gamma}{L}$

נסמן $h_t = F(x_t) - F(x^*)$

SGD

(1) נוכיח כי $F_{(x)}$ L-חלקה

לפי הגדרה של פ' חלקות, $f_i: \mathbb{R}^d \to \mathbb{R}$ לכל $i=1...m$ היא L-חלקה אם היא ~~גזירה~~ וגזירה

לכל $x,y \in \mathbb{R}^d$ ומתקיים: $\|\nabla f(x) - \nabla f(y)\| \le L\|x-y\|$ (1)

בנוסף, לפי הנתון $F(x) = \frac{1}{m}\sum_{i=1}^{m} f_i(x)$, לפי המשפט אשר מתאים כי $F(x)$ היא פ' סכום של פונקציות קמורות וגזירות ברציפות ולכן $F(x)$ גם גזירה ברציפות. לכן, ניתן לרשום כי:

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

לכן מכאן לכל $x,y \in \mathbb{R}^d$ נרשום:

$$\|\nabla F(x) - \nabla F(y)\| = \left\|\frac{1}{m}\sum_{i=1}^{m}\nabla f_i(x) - \frac{1}{m}\sum_{i=1}^{m}\nabla f_i(y)\right\| = \frac{1}{m}\left\|\sum_{i=1}^{m}\nabla f_i(x) - \sum_{i=1}^{m}\nabla f_i(y)\right\| =$$

$$= \frac{1}{m}\left\|\sum_{i=1}^{m}(\nabla f_i(x) - \nabla f_i(y))\right\| \underset{\text{מאי שוויון המשולש}}{\le} \frac{1}{m}\sum_{i=1}^{m}\|\nabla f_i(x) - \nabla f_i(y)\| \underset{\text{מתוך (1)}}{\le} \frac{1}{m}\sum_{i=1}^{m} L\|x-y\| =$$

$$= \frac{1}{m}\cdot mL\|x-y\| = L\|x-y\|$$

לכן מתקבל: $\|\nabla F(x) - \nabla F(y)\| \le L\|x-y\|$

לכן F היא L-חלקה

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

מתוך ההרצאה, לפי הגדרה מס' 2 מתקיים כי אם F חלקה עם פרמטר L אז:

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

(ניתן להשתמש בביטוי זה כי אמרנו כי F היא קמורה במובן החזק כמבוא לשאלה)

נבחר כי $x = x_t$, $y = x_{t+1}$ ונציב:

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

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

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

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

היות ו $\eta \ge 0$:

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

נשים לב כעת כי הוגדר: $h_t = F(x_t) - F(x^*)$ לכן $h_{t+1} = F(x_{t+1}) - F(x^*)$

$$F(x_{t+1}) - F(x_t) = F(x_{t+1}) - F(x^*) - (F(x_t) - F(x^*)) = h_{t+1} - h_t$$

לכן:

$$h_{t+1} - h_t \le -\eta\nabla F(x_t)^T g_t + \frac{L}{2}\eta^2\|g_t\|_2^2 \quad (3)$$

(3) הוכיחו כי $E[h_{t+1} - h_t] \le -\gamma^2 E[h_t]$

הדרכה: היעזרו בא"י שוויון שהוכחו בכיתה $E[\|g_t\|_2^2 \mid x_t] \le 2L(F(x_t) - F(x^*))$ (1)

וכן בלמה הבאה: $\|\nabla F(x)\|_2^2 \ge 2\mu(F(x) - F(x^*)) \quad \forall x \in \mathbb{R}^n$ (2)

מתוך הביטוי מסעיף 2 (3):

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

$$= \underbrace{-\eta E[\nabla F(x_t)^T g_t]}_{(4)} + \frac{L}{2}\eta^2 \underbrace{E[\|g_t\|_2^2]}_{(5)}$$

נרדה כעת למבוא חסם עליון עבור כל אחד מבין הביטויים הללו

$$(4)\ -\eta E[\nabla F(x_t)^T g_t] = -\eta E\left[E[\nabla F(x_t)^T g_t \mid x_t]\right] =$$

$$= -\eta E\left[\nabla F(x_t)^T E[g_t \mid x_t]\right] = (*)$$

היעזר במשפט התוחלת (תרגול 7)

יהיה Y RV- בנוסף X RV-

(6) $E_x[X] = E_y[E_x[X|Y]]$

פונקציה של Y

7

314092826 D4G197 ארימוביץ' יקטרינה HW3

תרגיל 5 המשך

המשך סעיף (3)

נתבונן כעת על הביטוי $E[g_t|x_t]$.

ראינו בהרצאה כי g_t (משתנה אקראי) לפי הגדרה הוא unbiased estimate של $\nabla f(x_t)$ כך ש:

$$E[g_t|x_t=x]=\frac{1}{n}\sum_{i=1}^{n}\nabla f_i(x)=\nabla f(x)$$

לכן במקרה שלנו עבור $i\sim uniform(\{1...m\})$ ניתן לרשום כי

$$E[g_t|x_t]=\frac{1}{m}\sum_{i=1}^{m}\nabla f_i(x_t)\underset{\text{לפי הגדרה של } F(x)}{=}\nabla F(x_t)$$

נציב בחזרה ל (*) ונקבל:

$$(*)=-\eta E[\nabla F(x_t)^T\nabla F(x_t)]\le$$ לפי (2)

$[\nabla F(x_t)^T\nabla F(x_t)]=\|\nabla F(x_t)\|^2$

$$\le -\eta E[2\mu(F(x_t)-F(x^*))]=$$

$$=-\eta E[2\mu h_t]=-\eta 2\mu E[h_t]=$$ לפי הגדרת $h_t=F(x_t)-F(x^*)$:

$$=-2\frac{\gamma}{L}\mu E[h_t]=\underline{-2\gamma^2E[h_t]}$$ נציב את $\eta=\frac{\gamma}{L}$: $\frac{\mu}{L}=\gamma$

כעת נמצא גם חסם עליון עבור ביטוי (5):

$$(5)=\frac{L}{2}\eta^2E[\|g_t\|_2^2]\overset{\text{לפי } 6}{=}\frac{L}{2}\eta^2E[E[\|g_t\|_2^2|x_t]]=$$

$$\le\frac{L}{2}\eta^2E[2L(F(x_t)-F(x^*))]=$$ לפי אי שוויון (1)

$$=\frac{L}{2}\eta^2E[2Lh_t]=$$ לפי h_t נציב:

$$=\frac{L}{\cancel{2}}\eta^2\cancel{2}LE[h_t]=$$ לפי $\eta L=\gamma$

$$=\underline{\gamma^2E[h_t]}$$

נאחד בחזרה את הביטויים ונקבל:

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

(4) הוכיחו כי $E[h_t]\le e^{-\gamma^2t}E[h_0]$

כלומר, $E[F(x_t)-F(x^*)]\le e^{-\gamma^2t}(F(x_0)-F(x^*))$

מתוך סעיף הקודם $E[h_{t+1}-h_t]\le-\gamma^2E[h_t]$ (8)

נשים לב, כי במקרה הזה, מתוך הגדרה של $\eta=\frac{\gamma}{L}$, L לא יכול להיות שווה ל 0, כי במקרה הזה גודל הצעד לא יהיה מוגדר, לכן בכל

(על פי הגדרת חלקות L: $\forall x,y\in\mathbb{R}^n \quad F(y)\le F(x)+\nabla F(x)^T(y-x)+\frac{L}{2}\|y-x\|^2$)

בנוסף הפונקציה לא תהיה שטוחה (עזרת לכך גם הגדרת קמירות ממש:

$\forall x,y\in\mathbb{R}^n \quad F(y)\ge F(x)+\nabla F^T(y-x)+\frac{\mu}{2}\|y-x\|^2$

$L\ge\mu \Rightarrow \gamma=\frac{\mu}{L}\le 1$

ומכיוון שלפי הגדרה סדר לכן [illegible] ממש (זה שם מוכיח מה שהוגדר בהגדרת קמירות ממש.

מתוך (8) מתוך לינאריות של התוחלת:

נציב את הביטוי הזה $E[h_{t+1}-h_t]=E[h_{t+1}]-E[h_t]$

בחזרה באי שוויון ונקבל $E[h_{t+1}]\le(1-\gamma^2)E[h_t]$

מכיון שזה מתקיים לכל t נמצא קשר בין $E[h_t]$ ל $E[h_0]$

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

כעת נטפל בביטוי $(1-\gamma^2)^t$ ו"להביא" אותו ל $e^{-\gamma^2t}$

ראינו מקודם כי $\gamma\in(0,1]$, נרצה להראות כי $(1-\gamma^2)\le e^{-\gamma^2}$

נפעיל ln: $\ln(1-\gamma^2)\le-\gamma^2$

נסמן: $g(\gamma)=-\gamma^2-\ln(1-\gamma^2)$ ונראה כי זו פונקציה:

$g'(\gamma)=-2\gamma-\frac{-2\gamma}{1-\gamma^2}=-2\gamma+\frac{2\gamma}{1-\gamma^2}=\frac{\gamma^3}{1-\gamma^2}>0$ ← מונוטונית עולה בתחום הזה.

היות וכרגע מונה חיובי, היות ו $\gamma<1$ אז מכנה גם חיובי

לכן הפונקציה הזו היא אכן מונוטונית עולה ב $\gamma\in(0,1)$

נשים לב כי נקודה ראשונית היא $g(0)=0$ אך מכיון שאנו מעוניינים להפוך את המקסימום, לכן עבור $\gamma=1$ ניתן לראות כי הביטוי שואף ל $+\infty$ כלומר בתחום העליון $g(\gamma)>0$ ולכן ניתן להסיק כי $(1-\gamma^2)\le e^{-\gamma^2}$

בנוסף מכיון ש $t\ge0$ אז נעלה את שני האגפים ב t ואי השוויון נשמר. לכן גם $(1-\gamma^2)^t\le e^{-t\gamma^2}$ ולכן:

$$E[h_t]\le\dots\le(1-\gamma^2)^tE[h_0]\le e^{-t\gamma^2}E[h_0]$$

נציב את $h_t=F(x_t)-F(x^*)$ בחזרה ונקבל:

$$E[F(x_t)-F(x^*)]\le e^{-t\gamma^2}E[F(x_0)-F(x^*)]$$