חישוב מורד קירוב:

נתון: $h:\mathbb{R}^2\to\mathbb{R}; h(x_1,x_2)=|x_1-1|+c$

חשב את $Prox_{2h}\left(\begin{bmatrix}2\\3\end{bmatrix}\right)$ ואת $Prox_{5h}\left(\begin{bmatrix}-3\\7\end{bmatrix}\right)$ כלומר רוצים לדעת כמה יהיה שווה הערך של הוקטור $\vec{u}$ באופרטור הפרוקסימלי בוקטורים הנתונים

תחילה נחשב את התת דיפרנציאל של $h(x_1,x_2)=|x_1-1|+c$ ונקבל: $\partial h(x_1,x_2)=\begin{cases}\{(1,0)^T\}; x_1>1\\ \{(a,0)^T\}\mid -1\le a\le 1; x_1=1\\ \{(-1,0)^T\}; x_1<1\end{cases}$ נחשב אותו דבר עם α ונקבל

$\alpha\partial h(u_1,u_2)=\begin{cases}\{(\alpha,0)^T\}; u_1>1\\ \{(\alpha a,0)^T\}\mid -1\le a\le 1; u_1=1\\ \{(-\alpha,0)^T\}; u_1<1\end{cases}$ החלפה נתונה לנו בכל אחד מן חישובי הקרבות שלנו וכעט כל מה שנותר לעשות זה להציב את אז זה לפי הנוסחא של חישוב קירוב

הגרדיאנט $(\vec{x}-\vec{u})\in\alpha\partial h(\vec{u}) \iff \vec{u}=\mathrm{Prox}_{\alpha h}(\vec{x})$ ולכן נגדיר $\vec{v}\in\partial h(u_1,u_2)$ נציב ונקבל $\vec{v}=\begin{cases}\{(\alpha,0)^T\}; v_1>1\\ \{(\alpha a,0)^T\}\mid -1\le a\le 1; v_1=1\\ \{(-\alpha,0)^T\}; v_1<1\end{cases}$ נרצה למצוא את $\vec{u}$ כך ש

$(\vec{x}-\vec{u})\in\alpha\partial h(u_1,u_2)$ ונקבל $\begin{matrix}\vec{x}-\vec{u}=v\\ \vec{x}-\vec{v}=\vec{u}\end{matrix}$ נציב בחזרה במשוואה שלנו ונקבל כי $\vec{u}=Prox_{\alpha h}(\vec{x})=\vec{x}-\vec{v}=\begin{cases}\vec{x}^T-(\alpha,0)^T; u_1>1\\ \vec{x}^T-(\alpha a,0)^T; u_1=1\\ \vec{x}^T-(-\alpha,0)^T; u_1<1\end{cases}$ נקח כל מקרה לגופו ונראה:

אם $u_1>1$ אז $u_1=x_1-\alpha$ לפי הנוסחא ולכן $x_1-1>\alpha$

אם $u_1=1$ אז $u_1=x_1-\alpha$ לפי הנוסחא ולכן $|x_1-1|=\alpha$

אם $u_1<1$ אז $u_1=x_1+\alpha<1$ לפי הנוסחא ולכן $x_1-1<-\alpha$

סה"כ נקבל: $\vec{u}=\begin{cases}\vec{x}-(\alpha,0)^T; x_1-1>\alpha\\ \vec{x}-(\alpha a,o)^T; |x_1-1|=\alpha a\\ \vec{x}-(-\alpha,0)^T; x_1-1<\alpha\end{cases}$ כעט כל שנותר הוא להציב את הוקטורים ולראות מה נקבל: $\vec{u}=Prox_{\alpha h}=\begin{cases}\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}\alpha\\0\end{bmatrix}; x_1-1>\alpha\\ \begin{bmatrix}1\\x_2\end{bmatrix}; |x_1-1|\le\alpha\\ \begin{bmatrix}x_1\\x_2-\end{bmatrix}\begin{bmatrix}-\alpha\\0\end{bmatrix}; x_1-1<-\alpha\end{cases}$ נציב לפי מה שביקשו

$$Prox_{2h}\left(\begin{bmatrix}2\\3\end{bmatrix}\right)\Rightarrow \begin{matrix}\alpha=2\\ x_1-1=2-1=1\\ |x_1-1|=1<2=\alpha\Rightarrow\vec{u}=\begin{bmatrix}1\\x_2\end{bmatrix}=\begin{bmatrix}1\\3\end{bmatrix}\end{matrix}$$

$$Prox_{5h}\left(\begin{bmatrix}-3\\7\end{bmatrix}\right)\Rightarrow \begin{matrix}\alpha=5\\ x_1-1=-3-1=-4\\ |x_1-1|=4<5=\alpha\Rightarrow\vec{u}=\begin{bmatrix}1\\x_2\end{bmatrix}=\begin{bmatrix}1\\7\end{bmatrix}\end{matrix}$$

שיטת נסטרוב:

נתונים:

$\vec{b}=\begin{bmatrix}8\\8\\16\end{bmatrix}$, $f(\vec{x})=\frac{1}{2}\vec{x}^TA\vec{x}-\vec{b}\vec{x}, A=\begin{pmatrix}3&1&4\\1&3&4\\4&4&8\end{pmatrix}$, $\vec{x}_k=\begin{bmatrix}-1/2\\7/6\\13/6\end{bmatrix}$; $L=12;\rho_{k-1}=\sqrt{\frac{89}{110}};\vec{x}_{k-1}=\begin{bmatrix}0\\2\\3\end{bmatrix}$ **צריך למצוא את $\vec{x}_{k+1}$ בשיטת נסטרוב לפונקציה קמורה לא במובן**

החזק

נתחיל להשתמש בפסאודו קוד לחישוב שיטת נסטרוב

קודם נחשב את הρ_k ונראה: $1+\rho_k(\rho_{k-1}^2-1)-\rho_k^2=0\to 1+\rho_k(\frac{89}{110}-1)-\rho_k^2=0\to\rho_k^2+\frac{21}{110}\rho_k-1=10\to\rho_k=\frac{10}{11},\frac{-11}{10}\to\rho_k\in[0,1],\rho_k=\frac{10}{11}$

עכשיו אפשר לחשב את β_k ידוע כי $\beta_k=\rho_k\rho_{k-1}^2$ ולכן $\beta_k=\frac{10}{11}\cdot\frac{89}{110}=\frac{89}{121}$ מכאן ניתן לחשב את $\vec{y}_k$ בעזרת β_k שמצאנו $\vec{y}_k=\vec{x}_k+\beta_k(\vec{x}_k-\vec{x}_{k-1})$ נציב ונקבל:

$\vec{y}_k=\begin{bmatrix}-0.5\\7/6\\13/6\end{bmatrix}+\frac{89}{121}\cdot\left(\begin{bmatrix}-1/2\\7/6\\13/6\end{bmatrix}-\begin{bmatrix}0\\2\\3\end{bmatrix}\right)=\frac{1}{121}\cdot\begin{bmatrix}-105\\67\\188\end{bmatrix}$ **כעט כל שנותר לנו זה לחשב בחזרה את $\vec{x}_{k+1}$ ונראה כי $\nabla f\left(\vec{y}_k\right)=A^T\cdot\vec{x}-\vec{b}$ ונחשב**

$\begin{pmatrix}3&1&4\\1&3&4\\4&4&8\end{pmatrix}\cdot\frac{1}{121}\cdot\begin{bmatrix}-105\\67\\188\end{bmatrix}-\begin{bmatrix}8\\8\\16\end{bmatrix}=-\frac{1}{121}\begin{bmatrix}464\\120\\584\end{bmatrix}$ **ומכאן ניתן לחשב את x_{k+1} עם הנוסחא: $x_{k+1}=\vec{y}_k-\frac{1}{L}\nabla f(\vec{y}_k)$ ונקבל**

$$x_{k+1}=\frac{1}{121}\cdot\begin{bmatrix}-105\\67\\188\end{bmatrix}-\frac{1}{12}\cdot\frac{1}{121}\begin{bmatrix}464\\120\\584\end{bmatrix}=\begin{bmatrix}\frac{199}{363}\\ \frac{7}{11}\\ \frac{710}{363}\end{bmatrix}$$

משפטים נכונים לשאלות:

* סכום של שתי פונקציות קמורות הוא פונקציה קמורה
* מקסימום של פונקציות קמורות הוא פונקציה קמורה
* הקבוצה $\Omega=\{x\in R^n \mid ||\vec{x}||\le r\}$ כאשר $r>0$ היא קבוצה קמורה
* תהי $A\in R^{m\times n}$ ו $\vec{b}\in R^m$ אזי הקבוצה $\Omega=\{x\in R^n\mid A\vec{x}=\vec{b}\}$ היא קבוצה קמורה
* הקבוצה $\Omega=\{x\in R^n\mid x_i\ge 0, i=1,2,\ldots,n\}$ כאשר הקבוצה קמורה
* אם $f:\Omega\subset\mathbb{R}^n\to\mathbb{R}$ גזירה פעמיים וקמורה במובן החזק כל ש: $\nabla f\left(\vec{x}^*\right)=\vec{0}$ אז $\vec{x}^*$ היא נקודת מינימום גלובלי יחידה.
* משפט: אם f קמורה וגזירה בנקודה $\vec{x}$ אז $\partial f=[\nabla^T f(\vec{x})]$
* נגדיר $f:\mathbb{R}^n\to\mathbb{R}$ נקודה פנימית בתחום ההגדרה של f אז קיים תת גרדיאנט של f בנקודה $\vec{x}$
* צירוף לינארי של פונקציות קמורות עם מקדמים אי שליליים הוא פונקציה קמורה
* פונקציה קמורה פחות פונקציה לינארית היא פונקציה קמורה
* אם קיימת פונקציה מסוג $f:\mathbb{R}\to\mathbb{R}$ ונתון כי $\vec{b}\in\mathbb{R}^n$ ונגדיר פונקציה חדשה כך: גם היא תהיה קמורה $g(x)=f\left(A\vec{x}^1+\vec{b}\right)$
* אם נקח את כל הפונקציות הקמורות מכל המימדים שיש לנו בפונקציה מסוג $f_i:\mathbb{R}\to\mathbb{R}$ ונגדיר פונקציה חדשה שלוקחת את המקסימום של אותן פונקציות בצורה הבאה: $f\left(\vec{x}\right)=\max_{i=1,\ldots,n} f_i(\vec{x})\cap_i f_i$ גם היא תהיה קמורה
* $f(\vec{x})=\frac{1}{2}\vec{x}^TA\vec{x}-\vec{b}\vec{x}\to\nabla f(\vec{x})=A^T\cdot\vec{x}-\vec{b}$

שיטת מורד

נתונים:

$$f\left(\vec{x}\right)=\vec{x}^TA\vec{x}-\vec{b}^T\vec{x}, A=\begin{pmatrix}2&-1&1\\-1&4&-1\\1&-1&2\end{pmatrix},\vec{b}=\begin{pmatrix}0\\4\\-2\end{pmatrix},\vec{x}_0=\begin{pmatrix}1\\1\\1\end{pmatrix}$$

תחילה נמצא את הנגזרת של f <— $\nabla f\left(\vec{x}\right)=A^T\cdot\vec{x}-\vec{b}$ ועכשיו נרצה למצוא את פונקציית

ההסיאן של f ואנחנו יודעים כי היא מטריצת A לכן נמצא את הערכים העצמאיים ונראה:

$\begin{vmatrix}2-\lambda&-7&1\\-1&4-\lambda&-1\\1&-1&2-\lambda\end{vmatrix}=1,2,5$, אנחנו יודעים כי בכדי למצוא את L חלקה נרצה את הערך העצמי

המקסימלי ביותר לכן $L\ge 5$ ואנחנו יודעים שבשביל שהפונקציה תהיה קמורה במובן החזק נרצה לקחת

את הm המינימלי ביותר ולכן $0<m\le 1$, נראה כי מדובר במקרה של פונקציה שהיא L חלקה וf קמורה

במובן החזק מודולו m ולכן $T=\Omega(\frac{L}{m}\cdot\ln(\frac{1}{\epsilon}))$, לפי הנוסחא הידוע לנו נחשב:

$f\left(\vec{x}_k\right)-f^*\le\left(1-\frac{m}{L}\right)^k\cdot(f(\vec{x}_0)-f^*)$ בשביל לפתור את האי שוויון הזה נצטרך למצוא את

$f^*, f(\vec{x}_0)$ לכן נחשב: $f(\vec{x}_0)=f\begin{pmatrix}1\\1\\1\end{pmatrix}=0.5\cdot(111)\cdot A\cdot\begin{pmatrix}1\\1\\1\end{pmatrix}-\vec{b}^T\cdot\begin{pmatrix}1\\1\\1\end{pmatrix}=1$

$f^*=\min\text{ point of }f(\vec{x})$ אנחנו יודעים שמכיוון שf קמורה במובן החזק יש נקודת מינימום גלובלית

אחת יחידה לכן נרצה לדעת מתי $\nabla f\left(\vec{x}\right)=0$ אנחנו כבר יודעים כי $\nabla f\left(\vec{x}\right)=A^T\cdot\vec{x}-\vec{b}$ ולכן

נרצה לחשב את $A^T\cdot\vec{x}-\vec{b}=0\Rightarrow A^T\cdot\vec{x}=\vec{b}$ נראה כי:

$\left(\begin{array}{ccc|c}2&-1&1&0\\-1&4&-1&4\\1&-1&2&-2\end{array}\right)=\left(\begin{array}{ccc|c}1&0&0&1\\0&1&0&1\\0&0&1&-1\end{array}\right)$ כעט נוכל לדעת מהי f^* על ידי החישוב

$$f^*=f\begin{pmatrix}1\\1\\-1\end{pmatrix}=0.5\cdot(11-1)\cdot A\cdot\begin{pmatrix}1\\1\\-1\end{pmatrix}-\vec{b}^T\cdot\begin{pmatrix}1\\1\\-1\end{pmatrix}=0.5\cdot(11-1)\cdot\begin{pmatrix}2&-1&1\\-1&4&-1\\1&-1&2\end{pmatrix}\cdot\begin{pmatrix}1\\1\\-1\end{pmatrix}-(042)\cdot\begin{pmatrix}1\\1\\-1\end{pmatrix}=-3$$

כעט נוכל לחשב את $f\left(\vec{x}_0\right)-f^*$ ונקבל כי זה שווה ל4 $f\left(\vec{x}_k\right)-f^*=1-(-3)=4$ לכן ניתן

להגיד כי $f\left(\vec{x}_k\right)-f^*\le\left(1-\frac{m}{L}\right)^k\cdot(f(\vec{x}_0)-f^*)$ נציב ונראה כי:

$$f\left(\vec{x}_k\right)-f^*\le\left(1-\frac{1}{5}\right)^k\cdot 4$$

היטל הגראדינט:

יש למצוא מינימום לפונקציה $f:\mathbb{R}^3\to\mathbb{R}; o.5\vec{x}Q\vec{x}-\vec{c}^T\vec{x}+d; d\in\mathbb{R}$ **כאשר ידוע כי** $Q>0$

בכפוף לאילוץ $\vec{x}\in\Omega;\Omega=\{\vec{x}\mid A\vec{x}=\vec{b}\}$

$$Q=\begin{pmatrix}2&-1&0\\-1&2&-1\\0&-1&2\end{pmatrix}; A=\begin{pmatrix}0&1&0\\-1&0&-1\end{pmatrix};\vec{c}=\begin{bmatrix}1\\-2\\-1\end{bmatrix};\vec{b}=\begin{bmatrix}1\\-2\end{bmatrix};\vec{x}_k=\begin{bmatrix}1\\1\\1\end{bmatrix}$$

תחילה נמצא את מטריצת ההטלה ועל מנת לעשות זאת נחשב

$$(AA^T)^{-1}=\left(\begin{pmatrix}0&1&0\\-1&0&-1\end{pmatrix}\begin{pmatrix}0&-1\\1&0\\0&-1\end{pmatrix}\right)^{-}1=\begin{pmatrix}1&0\\0&0.5\end{pmatrix}$$

$$I-A^T(AA^T)^{-1}A=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}-\begin{pmatrix}0&-1\\1&0\\0&-1\end{pmatrix}\begin{pmatrix}1&0\\0&0.5\end{pmatrix}\begin{pmatrix}0&1&0\\-1&0&-1\end{pmatrix}=\begin{pmatrix}0.5&0&-0.5\\0&0&0\\-0.5&0&0.5\end{pmatrix}$$

כעט נבדוק מהו $\vec{g}_k$ ונחשב זאת כך: $\vec{g}_k=\nabla f(\vec{x})=Q\vec{x}-\vec{c}$

$$\begin{pmatrix}2&-1&0\\-1&2&-1\\0&-1&2\end{pmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix}-\begin{bmatrix}1\\-2\\-1\end{bmatrix}=\begin{bmatrix}0\\2\\2\end{bmatrix}$$

כעט נחשב את $\vec{\alpha}_k, \vec{p}_k$; $P\vec{g}_k=\vec{p}_k$; $\alpha_k=\frac{\vec{g}_k^T\vec{p}_k}{\vec{p}_k^TQ\vec{p}_k}$ $\vec{p}_k=\begin{pmatrix}0.5&0&-0.5\\0&0&0\\-0.5&0&0.5\end{pmatrix}\begin{bmatrix}0\\2\\2\end{bmatrix}=\begin{bmatrix}-1\\0\\1\end{bmatrix}$

נחשב $\vec{g}_k^T\vec{p}_k=(0\ \ 2\ \ 2)\begin{bmatrix}-1\\0\\1\end{bmatrix}=2$, $\vec{p}_k^TQ\vec{p}_k=[-1\ \ 0\ \ 1]Q\begin{bmatrix}-1\\0\\1\end{bmatrix}=4$

וכעט ניתן לחשב $\vec{x}_{k+1}=\vec{x}_k-\alpha_kP_{N(A)}\left(\nabla^Tf(\vec{x})\right)$

$$\vec{x}_{k+1}=\begin{bmatrix}1\\1\\1\end{bmatrix}-0.5\begin{bmatrix}-1\\0\\1\end{bmatrix}=\begin{bmatrix}3/2\\1\\1/2\end{bmatrix}$$

חיפוש קווי:

נתונה פונקציה: $f(\vec{x})=3x_1^2+2x_1x_2+2x_x^2$, עם הנקודה והכיוון

$\vec{x}_k=(-1,2)^T,\vec{d}_k=(1,-2)^T,\bar{\alpha}=4,\beta=0.25$ ותנאי וולף $c_1=10^{-3},c_2=0.8$ הפעילו

את שיטת החיפוש הקווי הנסוג ומצאו את l הטבעי הקטן ביותר כך שעבור $\alpha_{k,l}=\bar{\alpha}\beta^l$ בשלב העדכון

מתקיימים שני וולף החלשים ביחד וקבעו את שלב העדכון:

נחשב את גרדיאנט תחילה: $\nabla f(\vec{x})=\begin{pmatrix}6x_1+2x_2\\2x_1+4x_2\end{pmatrix}$ ונחשב בנקודה: $\vec{x}_k=(-1,2)^T$

$$\begin{pmatrix}6x_1+2x_2\\2x_1+4x_2\end{pmatrix}=\begin{pmatrix}-6+4\\-2+8\end{pmatrix}=\begin{pmatrix}-2\\6\end{pmatrix}$$

נזכר בתנאי וולף:

תנאי הירידה המספקת - $f\left(\vec{x}_k+\alpha\vec{d}_k\right)\le f\left(\vec{x}_k\right)+c_1\alpha\nabla f\left(\vec{x}_k\right)\vec{d}_k$

תנאי גרדיאנט - $f\left(\vec{x}_k+\alpha\vec{d}_k\right)\cdot\vec{d}_k\ge c_2\cdot\nabla f\left(\vec{x}_k\right)\cdot\vec{d}_k$

נתחיל עם $\bar{\alpha}$ הנתונה ונראה האם אחד משני תנאי וולף מתקיימים או לא, אם הם לא מתקיימים ננסה למצוא a אחר ונעלה את הl שלנו, אם הם מתקיימים בפעם הראשונה אז ה$l=1$ וכך ניתן לפתור את $\vec{x}_{k+1}$.

נציב $\alpha=\bar{\alpha}=4$, $\vec{x}_{k+1}=\begin{pmatrix}-1\\2\end{pmatrix}+\alpha\cdot\begin{pmatrix}-1\\2\end{pmatrix}=\begin{pmatrix}-1+\alpha\\2-2\alpha\end{pmatrix}=\vec{x}_k+\alpha\vec{d}_k$ ונקבל

$\vec{x}_{k+1}=\begin{pmatrix}3\\-6\end{pmatrix}$, נבדוק האם התנאי הראשון מתקיים :

אגף שמאל: $f\begin{pmatrix}3\\-6\end{pmatrix}=3\cdot3^2+2\cdot3\cdot(-6)+2\cdot(-6)^2=63$

אגף ימין: $f(\vec{x}_k)=3\cdot(-1)+2\cdot(-1)\cdot2+2\cdot2^2=7$

$c_1\alpha\nabla f\left(\vec{x}_k\right)\vec{d}_k=0.8\cdot4\cdot(-2\ \ 6)\cdot\begin{pmatrix}1\\-2\end{pmatrix}=-0.056$, סה"כ: $7-0.057=6.944$,

ניתן לראות שהתנאי הראשון לא מתקיים: $63\not\le 6.944$

ולכן צריך לעדכן את הα באמצעות הנוסחא הנתונה: $\alpha_{k,l}=\bar{\alpha}\beta^l$ ולכן $\alpha_{k,1}=4\cdot0.25=1$ נציב

בשנית ב$\vec{x}_{k+1}$ ונקבל $\vec{x}_{k+1}=\begin{pmatrix}0\\0\end{pmatrix}$ ונבדוק שוב את התנאי הראשון בשנית כאשר אנחנו כבר יודעים כי

אגף ימין: $f(\vec{x}_k+\alpha\vec{d}_k)=\begin{pmatrix}-1+\alpha\\2-2\alpha\end{pmatrix}$ נציב ונקבל 0

אגף שמאל: $c_1\alpha\nabla f\left(\vec{x}_k\right)\vec{d}_k=10^{-3}\cdot1\cdot(-2\ \ 6)\cdot\begin{pmatrix}1\\-2\end{pmatrix}=7-0.014=6.984$

ניתן לראות כי התנאי הראשון מתקיים $0\le 6.984$

נבדוק את התנאי השני **(למרות שבמבחן לפי איצקוביץ אין צורך)**

$$f\left(\vec{x}_k+\alpha\vec{d}_k\right)\cdot\vec{d}_k\ge c_2\cdot\nabla f\left(\vec{x}_k\right)\cdot\vec{d}_k$$

אגף שמאל:

$$\nabla f\left(\vec{x}_k+\alpha\vec{d}_k\right)\cdot\vec{d}_k=0$$

אגף ימין:

$$c_2\cdot\nabla f\left(\vec{x}_k\right)\cdot\vec{d}_k=0.8\cdot(-2\ \ 6)\cdot\begin{pmatrix}-1\\2\end{pmatrix}=-11.2$$

נראה כי גם הוא מתקיים ולכן שני תנאי וולף מתקיימים עבור $l=1$ ו- $\vec{x}_{k+1}=\begin{pmatrix}0\\0\end{pmatrix}$

חישוב בשיטת תת גראדיאנט:

נתונים: $\vec{x}_k=\begin{bmatrix}0\\1\end{bmatrix}$, $f(x_1,x_2)=2|x_1|+2|x_2|$ **יש למצוא את $\vec{x}_{k+1}$ בשיטת תת גרדיאנט**

כאשר $\alpha_k=\frac{1}{k^{3/4}}$ תחילה נזכר באלגוריתם: $\vec{x}_{k+1}=\vec{x}_k-\alpha_k\vec{g}_k$ ונבדוק האם ה$\vec{g}_k$ שייך לתת

דיפרנציאל אנחנו כבר מכירים את $f(x_1,x_2)$ מהתרגיל הקודם רק שבמקרה הזה $A=B=2$

ולכן $\vec{x}=[0//1]\Rightarrow\vec{g}_k\in(a,2\,\mathrm{Bsign}\,x_2)^T\mid a\in[-2,2]$ וכעט כל שנותר לנו הוא לבדוק

איפה $\vec{g}_k=\begin{bmatrix}a\\2\end{bmatrix};-2\le a\le 2$ כעט נציב ונקבל $\vec{x}_{k+1}=\begin{bmatrix}0\\1\end{bmatrix}-\frac{1}{k^{3/4}}\begin{bmatrix}a\\2\end{bmatrix}$ במקרה שלנו

יהיו אופציות לתשובות ונראה מי מהן נכון ע"י הצבה.

* אם f חלקה במובן ליפשיץ ו L חלקה כלומר הפונקציה קמורה באיזור שבין $\vec{x},\vec{y}$ נוכל למצוא פונקציה חדשה קמורה בנקודה שהיא תחסום את הפונקציה מלמעלה.
* אם f גזירה פעמיים והיא חלקה במובן ליפשיץ אז פונקציית ההסיאן שלה קטנה שווה לL כפול מטריצת היחידה: $\mathbf{H_f}\le L\mathbf{I}$
* אם f גזירה פעמיים כך ש: $\mathbf{H_f}\le L\mathbf{I}$ אז היא חלקה במובן ליפשיץ.
* אם f היא פונקציה קמורה ותת גראדיאנט $\vec{g}$ מקיים $||\vec{g}||\le G$ אז מספר האיטרציות הדרושות לדיוק $\epsilon>0$ בשיטת הsubgradient method עם הצעד הקבוע $\Omega\left(\frac{1}{\epsilon}^2\right)$
* אם f חלקה במובן ליפשיץ ו L חלקה וקמורה במובן החזק מודולו m אז בשטת נסטרוב מס האיטרציות לדיוק יהיה $\Omega\left(\sqrt{L/m}\ln(1/\epsilon)\right)$
* אם נקח את כל הפונקציות הלינאריות שמתחת לפונקציה קמורה f ונקרא להם $\mathscr{U}$ אז המקסימום של כל הפונקציות הללו היא גם פונקציה קמורה אך המינימום אינו בהכרח פונקציה קמורה.
* אם נתונה לנו פונקציה קמורה כאשר האפיגרף שלה קמור וסגור אז כל פתרון למציאת נקודת מינימום תהיה נקודת מינימום גלובלי בפונקציה
* אם הפונקציה $f:\Omega\to\mathbb{R}$ קמורה ונתון כי Ω קמורה וסגורה אז קבוצת הפתרונות של נקודות המינימום בפונקציה קמורה
* עקרון הדואליות: אם נקח את הפונקציה $f:\Omega\to\mathbb{R}$ כקמורה וגזירה ממעלה ראשונה ניתן להגיד כי מישור לפונקציה בנקודה מסוימת הוא הערכה טובה לפונקציה באותה נקודה בצורה הבאה: $f\left(\vec{y}\right)=f\left(\vec{x}\right)+\left(\nabla f\left(\vec{x}\right)\right)\left(\vec{y}-\vec{x}\right)$ כאשר ידוע כי $f\left(\vec{x}\right)+(\left(\nabla f\left(\vec{x}\right)\right)\left(\vec{y}-\vec{x}\right)$ הוא מישור בוקטור כלומר המישור עצמו נמצא מתחת לפונקציה.

שיטת מורד המראה:

יש לחשב את $\vec{x}_{k+1}$ בשיטת המראה עם נתונים:

$$f(\vec{x})=\frac{1}{2}\vec{x}^TA\vec{x}-\vec{b}\vec{x}, A=\begin{pmatrix}3&-1&0\\-1&2&-1\\0&-1&3\end{pmatrix},\vec{b}=\begin{pmatrix}\frac{4}{11}\\ \frac{4}{11}\\ \frac{4}{11}\end{pmatrix}, m=1, L=5,$$

$$h=\sum_{i=1}^n x_i\ln x_i:\Omega$$

$$\Omega=\left[\vec{x}^T=(x_1,x_2,\ldots,x_n)\mid x_i>0; i=1,2,\ldots,n;\sum_{i=1}^n x_i=1\right]$$

$$D_h(\vec{x}^*,\vec{x}_0)<R=0.125;\vec{x}_k=\begin{pmatrix}0.3\\0.3\\0.4\end{pmatrix}; T=3$$

תחילה נבין את הנתונים

אנחנו יודעים שהפונקציה היא קמורה במובן החזק מודולו $m=1$

אנחנו יודעים ש: $h=\sum_{i=1}^n x_i\ln x_i$ היא בעלת אנטרופיה שלילית ואנחנו יודעים שכאשר נפעיל על

הפונקציה הזו סטיית ברגמן נקבל כי הקבוע הוא $D_h(\vec{x}^*,\vec{x}_0)<R=0.125$ ובסה"כ מבצעים 3 איטרציות $T=3$

תחילה נמצא את גרדיאנט של הפונקציה של $h(\vec{x})$

נגדיר תחילה נגזרות חלקיות

$$\frac{\partial h}{\partial x_i}=\ln x_i+x_i\cdot\frac{1}{x_i}=\ln x_i+1$$

$\nabla^T h(\vec{x})=\begin{pmatrix}\ln x_1+1\\\ln x_2+1\\\ln x_3+1\end{pmatrix}$ נציב ב$\vec{x}_k$ ונקבל $\begin{pmatrix}\ln 0.3+1\\\ln 0.3+1\\\ln 0.4+1\end{pmatrix}$ עכשיו ניתן לחשב את α נציב את הנתונים

ונראה כי $\alpha=\frac{\sqrt{2mR}}{L}\cdot\frac{1}{\sqrt{T+1}}=\frac{\sqrt{2}\cdot1\cdot0.125}{5}\cdot\frac{1}{\sqrt{3+1}}=0.05$

כעט נכפול אותה בטרנספורט של גרדיאנט בפונקציה הנתונה ונראה כי

$$\nabla f(\vec{x})=A^T\cdot\vec{x}-\vec{b}\Rightarrow\nabla f(\vec{x}_k)=\begin{pmatrix}3&-1&0\\-1&2&-1\\0&-1&3\end{pmatrix}\cdot\begin{pmatrix}0.3\\0.3\\0.4\end{pmatrix}-\begin{pmatrix}\frac{4}{11}\\ \frac{4}{11}\\ \frac{4}{11}\end{pmatrix}=\begin{pmatrix}0.2363\\-0.4634\\-0.5363\end{pmatrix}$$

כעט נעבור למרחב הדואלי ונחשב את $\vec{y}_k$ ניזכר כי: $\nabla h(\vec{x}_{k+1})=\nabla h(\vec{x}_k)-\alpha\nabla f(\vec{x}_k)$

ונראה: $\begin{pmatrix}\ln 0.3+1\\\ln 0.3+1\\\ln 0.4+1\end{pmatrix}-\alpha\cdot\begin{pmatrix}0.2363\\-0.4634\\-0.5363\end{pmatrix}=\begin{pmatrix}-0.2157\\-0.1807\\0.0568\end{pmatrix}$

מכאן נחזור למרחב האמיתי שלנו של נזכר כי: $\vec{x}_{k+1}=\nabla h(\vec{x}_k)^{-1}-\alpha_k\nabla^Tf(\vec{x}_k)$

אנחנו יודעים שבמקרה של אנטרופיה שלילית גרדיאנט ההופכי הוא

$h(\vec{v})=\sum_{i=1}^n v_i\ln v_i\Rightarrow((\nabla h)^{-1}(\vec{v}))_j=\frac{e^{v_j}}{\sum_{i=1}^n e^{v_i}}$ ולכן ניתן לחשב בצורה הבאה:

וכעט ניתן לחשב: $\begin{pmatrix}-0.2157\to v_1\\-0.1807\to v_2\\0.0568\to v_3\end{pmatrix}$

$$\vec{x}_{k+1}=\frac{e^{v_j}}{\sum_{i=1}^n e^{v_i}}=\frac{1}{e^{-0.2157}+e^{-0.1807}+e^{0.0568}}\cdot\begin{pmatrix}-0.2157\\-0.1807\\0.0568\end{pmatrix}=\begin{pmatrix}0.299\\0.309\\0.392\end{pmatrix}$$

חישוב תת דיפרנציאל:

יש לחשב את תת הדיפרנציאל של כל אחת מן הפונקציות הבאות:

$h:\mathbb{R}^n\to\mathbb{R}^2;\quad h(\vec{x})=f(A\vec{x});\quad A\in\mathbb{R}^{2\times m};\quad\vec{x}\in\mathbb{R}^n$

$f:\mathbb{R}^2\to\mathbb{R};\quad f(\vec{z})=f(z_1,z_2)=|z_1|+|z_2|$ **תחילה נתייחס לפונקציה f ונפצל**

אותה ל2 פונקציות שונות כמו חוק 2 ונראה $\partial f_1(\vec{x})+\partial f_2(\vec{x})=\partial(f_1(\vec{x})+f_2(\vec{x}))(\vec{x})$ **נציב ונקבל**

$$\partial f_1(\vec{x})+\partial f_2(\vec{x})=\begin{cases}(\mathrm{Asign}\,z_1,\mathrm{Bsign}\,z_2); z_1,z_2\ne0\\(a,\mathrm{Bsign}\,z_2)\mid -1\le a\le1; z_1=0,z_2\ne0\\(\mathrm{Asign}\,z_1,b)\mid -1\le b\le1; z_1\ne0,z_2=0\\(a,b)\mid -1\le a\le1,-1\le b\le1; z_1,z_2=0\end{cases}$$

כעט נחשב את $h(\vec{x})$ אנחנו יודעים כי $A_{2\times x}\cdot\vec{x}_{n\times1}$ ומכפלה כזו תביא לנו וקטור מהסוג $\begin{bmatrix}(A\vec{x})_1\\(A\vec{x})_2\end{bmatrix}$

לכן נשתמש במשפט 3 $\partial h(\vec{x})=A^T\partial f(A\vec{x}+\vec{b})$ ומכן ניתן לחשב

$$A^T\partial f(A\vec{x})=\begin{cases}\left[\mathrm{Asign}(A\vec{x}_1),\mathrm{Bsign}(A\vec{x}_2)\right]; A\vec{x}_1,A\vec{x}_2\ne0\\\left[a,\mathrm{Bsign}(A\vec{x}_2)\right]\mid -1\le a\le1; A\vec{x}_1=0,A\vec{x}_2\ne0\\\left[\mathrm{Asign}(A\vec{x}_1),b\right]\mid -1\le b\le1; A\vec{x}_1\ne0,A\vec{x}_2=0\\\left[a,b\right]\mid -1\le b\le1,-1\le a\le1; A\vec{x}_1A\vec{x}_2=0\end{cases}$$

General case

$$f(x_1,x_2)=A|x_1|+B|x_2|; A,B>0\implies\partial f(\vec{x})=\begin{cases}\{(A\,\mathrm{sign}(x_1),B\,\mathrm{sign}(x_2))^T\} & x_1,x_2\ne0\\\{(a,B\,\mathrm{sign}(x_2))^T\mid a\in[-A,A]\} & x_1=0,x_2\ne0\\\{(A\,\mathrm{sign}(x_1),b)^T\mid b\in[-B,B]\} & x_1\ne0,x_2=0\\\{(a,b)^T\mid a\in[-A,A],b\in[-B,B]\} & \vec{x}=\vec{0}\end{cases}$$

חיפוש קווי מראה

סוג הפונקציה	קצב התכנסות	מספר איטרציות	חישוב הנקודה הבאה	נוסחאות נוספות
הפונקציה היא L חלקה וקמורה במובן החזק	$f(\bar{\vec{x}}_T)-f^*\le\frac{L\sqrt{2R}}{\sqrt{m}}\cdot\frac{1}{\sqrt{T+1}}$	$k=\Omega((\frac{L^2R}{m})\cdot\frac{1}{\epsilon^2})$	$\vec{x}_{k+1}=\nabla h(\vec{x}_k)^{-1}-\alpha_k\nabla^T f(\vec{x}_k)$	$\nabla h(\vec{x}_{k+1})=\nabla h(\vec{x}_k)-\alpha\nabla f(\vec{x}_k)$ $\alpha=\frac{\sqrt{2}mR}{L}\cdot\frac{1}{\sqrt{T+1}}$
אנטרופיה שלילית	$h(\vec{x})=\sum_{i=1}^{n}x_i\ln x_i$	$h(\vec{v})=\sum_{i=1}^{n}v_i\ln v_i\Rightarrow((\nabla h)^{-1}(\vec{v}))_j=\frac{e^{v_j}}{\sum_{i=1}^{n}e^{v_i}}$		$\bar{\vec{x}}_T=\frac{\sum_{j=0}^{k}\alpha_j\vec{x}_j}{\sum_{i=0}^{k}\alpha_i}$
אנטרופיה חיובית	$h(\vec{x})=\frac{1}{2}\Rightarrow\nabla(\vec{x})=\vec{z}^T\Rightarrow\frac{1}{2}\|\|\vec{x}-\vec{z}\|\|^2$	$h(\vec{v})=\frac{1}{2}\|\|\vec{v}\|\|\Rightarrow(\nabla h)^{-1}(\vec{v})=\vec{v}$		$f(\bar{\vec{x}}_T)-f^*\le\frac{1}{2T}\|\|\vec{x}_0-\vec{x}^*\|\|$

היטל גראדיאנט קירוב

סוג הפונקציה	קצב התכנסות	מספר איטרציות	חישוב הנקודה הבאה	נוסחאות נוספות
הפונקציה מורכבת מ2 פונקציות אחת קמורה וחלקה בעוד האחרת רק חלקה	$f(\vec{x}_T)-f(\vec{x}^*)\le\frac{L\|\vec{x}_0-\vec{x}^*\|^2}{2T}$	$\Omega(1/\epsilon)$	$\vec{y}_{k+1}=\vec{x}_k-\alpha_k\nabla^T g(\vec{x}_k)$ $\vec{x}_{k+1}=\mathrm{Prox}_{\alpha_k h}(\vec{y}_{k+1})$	$\alpha=1/L$
הפונקציה מורכבת מ2 פונקציות אחת קמורה וחלקה בעוד האחרת רק חלקה במובן החזק	$\|\vec{x}_T-\vec{x}^*\|\le\left(\frac{\kappa-1}{\kappa+1}\right)^T\|\vec{x}_0-\vec{x}^*\|;\quad\kappa=\frac{L}{m}$	$\Omega\left(((L-m)/2m)\ln(1/\epsilon)\right)$	$\vec{x}_{k+1}=\mathrm{Prox}_{\alpha_k h}(\vec{x}_k-\alpha_k\nabla^T g(\vec{x}_k))=\vec{x}_k-\alpha_k\vec{G}_{\alpha_k}(\vec{x}_k)$	$\alpha=2/(L+m)$

גרדינט צמוד

סוג הפונקציה	קצב התכנסות	מספר איטרציות	חישוב הנקודה הבאה	נוסחאות נוספות
קיימת מטריצה Q סימטרית כך שיהיו לה וקטורים צמודים/אורטוגונלים/אורטוגונלים ביחס למטריצה	$\|\vec{x}_k-\vec{x}^*\|_A\le 2\|\vec{x}_0-\vec{x}^*\|_A\left(\frac{\sqrt{L/m}-1}{1+\sqrt{L/m}}\right)^k$	$k=\Omega\left(\sqrt{L/m}\cdot\ln(1/\epsilon)\right)$	$\vec{x}_{k+1}=\vec{x}_k+\alpha_k\cdot\vec{p}_k$	$\alpha_i=\frac{\vec{p}_i^TQ\vec{x}^*}{\vec{p}_i^TQ\vec{p}_i}$ $\gamma_k=\frac{\vec{r}_{k+1}^T\vec{r}_{k+1}}{\vec{r}_k^T\vec{r}_k}$
אם הפונקציה חלקה וקמורה במובן החזק מודולו m				$\vec{P}_k=-\vec{r}_k+\gamma_{k-1}\vec{p}_{k-1}$
משפט הכיוון הצמוד	אם נתון שהQ גדולה מ-0 וגם יש וקטורים צמודים לQ		$\vec{p}_{k+1}=-\vec{r}_{k+1}+\gamma_k\vec{p}_k$ $\vec{r}_{k+1}=\vec{r}_k+\alpha_kQ\vec{p}_k$	$\alpha_k=-\frac{\vec{p}^T\vec{r}_k}{\vec{p}_k^TQ\vec{p}_k}$ $\vec{r}_k=\nabla f(\vec{x}_k)$

תת גרדיאנט

סוג הפונקציה	קצב התכנסות	מספר איטרציות	חישוב הנקודה הבאה	נוסחאות נוספות
פונקציה קמורה וגם D0 = ||X1 - X*||	$f(\tilde{\vec{x}}_T)-f(\vec{x}^*)\le\frac{D_0^2+G^2\sum_{k=1}^{T}\alpha_j^2}{2\sum_{k=1}^{T}\alpha_k}$	$\Omega(1/\epsilon^2)$	$\vec{x}_{k+1}=\vec{x}_k-\alpha_k\vec{g}_k$	$\partial h(\vec{x})=A^T\partial f(A\vec{x}+\vec{b})$
במקרה של צעד קבוע בכל פעם	$f(\tilde{\vec{x}}_T)-f(\vec{x}^*)\le\frac{1}{2}(\theta+\theta^{-1})\frac{D_0G}{\sqrt{T}}$	$\alpha=\frac{\theta D_0}{G\sqrt{T}}$	General case $f(x_1,x_2)=A\|x_1\|+B\|x_2\|;A,B>0\implies\partial f(\vec{x})=\begin{cases}\{(A\,\mathrm{sign}(x_1),B\,\mathrm{sign}(x_2))^T\} & x_1,x_2\neq 0\\ \{(a,B\,\mathrm{sign}(x_2))^T\mid a\in[-A,A]\} & x_1=0,x_2\neq 0\\ \{(A\,\mathrm{sign}(x_1),b)^T\mid b\in[-B,B]\} & x_1\neq 0,x_2=0\\ \{(a,b)^T\mid a\in[-A,A],b\in[-B,B]\} & \vec{x}=\vec{0}\end{cases}$	
בנורמת צעד קבוע	$f(\tilde{\vec{x}}_T)-f(\vec{x}^*)\le\frac{D_0^2G+GT\alpha^2}{2T\alpha}=\frac{D_0^2G}{2T\alpha}+\frac{G\alpha}{2}$	$\alpha_k=\frac{\alpha}{\|\|\vec{g}_k\|\|}$		
באורך צעד יורד	$f(\tilde{\vec{x}}_T)-f(\vec{x}^*)\le\frac{D_0^2+G^2\theta^2(\ln T+1)}{2\theta\sqrt{T}}$	$\alpha=\frac{\theta}{\sqrt{k}}$		

מורד התלול ביותר חיפוש קווי נסוג

סוג הפונקציה	קצב התכנסות	מספר איטרציות	חישוב הנקודה הבאה	נוסחאות נוספות
הפונקציה היא L חלקה וחסומה במלרע	$\min_{0\le k\le T-1}\|\|\nabla f(\vec{x}_n)\|\|\le\sqrt{\frac{2L(f(\vec{x}_0)-\tilde{f})}{T}}$	$\Omega(L/\epsilon^2)$	$\vec{x}_{k+1}=\vec{x}_k+\alpha_k\beta^l\vec{d}_k$	$\alpha_{k,l}=\bar{a}\beta^l$
הפונקציה היא L חלקה וקמורה	$f(\vec{x}_T)-f^*\le\frac{L}{2T}\|\vec{x}_0-\vec{x}^*\|^2$	$\Omega(L/\epsilon)$		$f(\vec{x}_k+\alpha\vec{d}_k)\le f(\vec{x}_k)+c_1\alpha\nabla f(\vec{x}_k)\vec{d}_k$
הפונקציה היא L חלקה וקמורה במובן החזק	$f(\vec{x}_k)-f^*\le\left(1-\frac{m}{L}\right)^k\cdot(f(\vec{x}_0)-f^*)$	$\Omega\left((L/m)\ln(1/\epsilon)\right)$		$f(\vec{x}_k+\alpha\vec{d}_k)\cdot\vec{d}_k\ge c_2\cdot\nabla f(\vec{x}_k)\cdot\vec{d}_k$

שיטת נסטרוב

סוג הפונקציה	קצב התכנסות	מספר איטרציות	חישוב הנקודה הבאה	נוסחאות נוספות
הפונקציה היא L חלקה וקמורה במובן החזק	$f(\vec{x})=0.5\cdot\vec{x}^T\cdot Q\vec{x}-\vec{b}\cdot\vec{x}$	$\Omega\left(\sqrt{L/m}\ln(1/\epsilon)\right)$	$\vec{x}_{k+1}=\vec{x}_k-\alpha_k\nabla^T f(\vec{x}_k+\beta_k(\vec{x}_k-\vec{x}_{k-1}))+\beta_k(\vec{x}_k-\vec{x}_{k-1})$	$k(Q)=\frac{\lambda_{\max}}{\lambda_{\min}}=\frac{L}{m}$ $\beta_k\equiv\beta=\frac{\sqrt{L}-\sqrt{m}}{\sqrt{L}+\sqrt{m}}=\frac{\sqrt{k}-1}{\sqrt{k}+1};$ $\vec{y}_k=\vec{x}_k+\beta_k(\vec{x}_k-\vec{x}_{k-1})$ $\alpha_k\equiv\alpha=\frac{1}{L}$

הפונקציה היא L חלקה וקמורה	$f(\vec{x}_T)-f^*\le\frac{2L}{(T+1)^2}\|\vec{x}_0-\vec{x}^*\|^2$	$\Omega(\sqrt{L/\epsilon})$	$x_{k+1}=\vec{y}_k-\frac{1}{L}\nabla f(\vec{y}_k)$
			for $k=0,1,2,\cdots$ ***do***
			Set $\vec{y}_k:=\vec{x}_k+\beta_k(\vec{x}_k-\vec{x}_{k-1})$
			Set $\vec{x}_{k+1}:=\vec{y}_k-(1/L)\nabla f(\vec{y}_k)$
			Define ρ_{k+1} to be the root in [0,1] of the following equation :
			$1+\rho_{k+1}(\rho_k^2-1)-\rho_{k+1}^2=0$
			Set $\beta_{k+1}=\rho_{k+1}\rho_k^2$
			end for
				$\beta_k=\rho_k\rho_{k-1}^2$
				$1+\rho_k\rho_{k-1}^2-\rho_k^2=\rho_k$
				$1+\rho_k(\rho_{k-1}^2-1)-\rho_k^2=0$
				$\rho_k\in[0,1]$

מורד התלול ביותר היטל הגרדיאנט

סוג הפונקציה	קצב התכנסות	מספר איטרציות	חישוב הנקודה הבאה תקף לכל סוגי הפונקציה	נוסחאות נוספות תקף להכל מלבד הנתונים על a
הפונקציה היא L חלקה	$\min_{0\le k\le T-1}\|\vec{x}_{k+1}-\vec{x}_k\|\le\sqrt{\frac{2f(\vec{x}_0)-\tilde{f}}{LT}}$	$\Omega(L/\epsilon^2)$	$\vec{x}_{k+1}=\underset{\vec{x}\in\mathbb{R}^n}{\mathrm{Argmin}}\left\{f(\vec{x}_k)+\nabla f(\vec{x}_k)(\vec{x}-\vec{x}_k)+\frac{1}{2\alpha_k}\|\vec{x}-\vec{x}_k\|^2\right\}\implies\vec{x}_{k+1}=\vec{x}_k-\alpha_k\nabla f(\vec{x}_k)$	$\vec{p}_0=-\nabla f(\vec{x})$ $P_{N(A)}=I-A^T(AA^T)^{-1}A$ $\alpha=\frac{1}{L}$
הפונקציה היא l חלקה וקמורה	$f\left(\frac{1}{T+1}\sum_{k=0}^{T}\vec{x}_k\right)-f^*\le\frac{\|\vec{x}_0-\vec{x}^*\|_2^2}{2\alpha(T+1)}+\frac{\alpha}{2(T+1)}\sum_{k=0}^{T}\|\nabla f(\vec{x}_k)\|_2^2$	$\Omega(L/\epsilon)$	$\vec{x}_{k+1}=\vec{x}_k-\alpha_kP_{N(A)}\left(\nabla^T f(\vec{x})\right)$	$\alpha_k=a$ $\vec{g}_k=\nabla f(\vec{x})=Q\vec{x}-\vec{c}$
הפונקציה היא l חלקה וקמורה במובן החזק	$\left(\frac{1}{T+1}\sum_{k=0}^{T}\vec{x}_k\right)-f^*\le\frac{M\|\vec{x}_0-\vec{x}^*\|}{\sqrt{T+1}}$	$\Omega(1/\sqrt{\epsilon})$	$\vec{x}_{k+1}=\vec{x}_k-\alpha_kP_{N(A)}\left(\nabla^T f(\vec{x})\right)$	$\alpha=\frac{\|\|\vec{x}_0-\vec{x}^*\|\|}{M\sqrt{T+1}}$