ריבועים פחותים LS:

מערכות של משוואות וריבועים פחותים:

יש לנו m משוואות $f_i(x)$ $i\in\{1,\dots,m\}$ כך שבכל $f_i(x)$ יש n משתנים.

המטרה שלנו למצוא את x שיתן לנו:

$$\begin{bmatrix} f_1(\underline{x})=0 \\ f_2(\underline{x})=0 \\ \vdots \\ f_m(\underline{x})=0 \end{bmatrix}_{m\times 1}$$

זה בעצם וקטור של סקלרים

נעבור לייצוג אחר של הבעיה: נגדיר פונקציה חדשה

$$p(\underline{x}) = [f_1(\underline{x})]^2 + [f_2(\underline{x})]^2 + \dots [f_m(\underline{x})]^2 = \|f(\underline{x})\|_2^2$$

כך שנפתור את הדבר הבא:

$$\min_{\underline{x}} p(x) = \min_{\underline{x}} \sum_{k=1}^{m} [f_k(\underline{x})]^2 \leftarrow \text{שקול ל} \rightarrow f(x)=\underline{0} \rightarrow \min_{\underline{x}} \|f(x)\|_2^2$$

אם למערכת יש פתרון.

תזכורת להגדרה של נורמה:

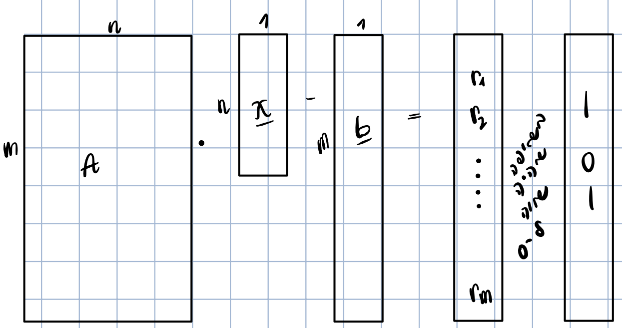
$$l^2 = \|x\|_2 = \left(\sum_{k=1}^{n} |x_k|^2\right)^{\frac{1}{2}}$$

$\forall k>0$,

- אם אין למערכת פתרון, הייצוג החדש עשיר יותר

כעת נסתכל על משוואות לינאריות:

$A\underline{x} = \underline{b}$ נעביר אגפים כדי שיהיה מערכת משוואות הומוגנית כך ש $A\underline{x} - \underline{b} = 0$. כלומר:



המערכת לאו בהכרח ריבועית!

פונקציית המחיר המשולבת תהיה בע"ת ה-LS:

$$p(x) = \sum_{k=1}^{m} (a_k \underline{x} - b_k)^2 = (A\underline{x}-b)^t \cdot (A\underline{x}-b)$$

$$\|\underline{r}\|^2 = \underline{r}^t \cdot r = \|A\underline{x}-b\|^2$$

מסמן בריבוע

$$= \|A\underline{x}-b\|_2^2 = \|\underline{r}\|^2$$

נורמה 2

השארית $Ax-b=r$

כלומר לסיכום מתקיים:

$$\left\{\begin{array}{l} a_1 x - b_1 = 0 \\ a_2 x - b_2 = 0 \\ \vdots \\ a_m x - b_m = 0 \end{array}\right\} \rightarrow \min_{\underline{x}} \sum_{k=1}^{m} (a_k \underline{x} - b_k)^2 \leftarrow \text{שקול ל} \rightarrow A\underline{x} = b \Rightarrow \min_{\underline{x}} \|A\underline{x}-b\|_2^2$$

אם למערכת משוואות יש פתרון

דרך נוספת להסתכל על אותה בעיה:

נניח כי בידינו סדרה בת $n<m$ וקטורים כלומר: $\{v_1, v_2, v_3, \dots, v_n\} \in R^m$

הוקטורים הנ"ל פורשים את תת-מרחב U במימד עד n - לכן b וקטור בתת מרחב, וכל להירקב בדרך הבאה:

$$\sum_{k=1}^{n} x_k v_k = \underbrace{\begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}_{m\times n}}_{A} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

נסמן את הקשר הנ"ל על ידי: הגדרת מט' A כשמורכבת שעמודותיה הן v_k ק שזו מטריצה $m \times n$

כעת נתון לנו וקטור $\underline{b}$ שרירותי באורך m ונרצה למצוא את הוקטור הקרוב ביותר ל-b מתוך וקטורי A כלומר את הטלתו על V. הדרך לעשות זאת:

$$\min_x \left\| \sum_{k=1}^{n} x_k v_k - b \right\|_2^2 = \min_x \| Ax - b \|$$

A מט' שעמודותיה הן וקטורי V

__מסקנה:__ כל בע"ת LS ניתנות כהתייחסות כבע"ת הטלה למרחב העמודות של מט' A.

__מינימיזציה של תבניות ריבועיות:__

$$p(x) = \|Ax-b\|_2^2 = (Ax-b)^t(Ax-b) = (x^tA^t - b^t)(Ax-b) =$$

$$= x^tA^tAx - x^tA^tb - Axb^t + b^tb$$

$$= x^tKx^t - 2x^tf + c$$

התוצאה של כל הביטוי הוא סקלר

כל אחד מהביטויים נותן בסופו של דבר תוצאה של סקלר.

לכן: $x^tA^tb = Axb^t = x^t \cdot f$

וגם נשים לב כי $A^tA = K$

וגם הפונקציה (כ-ו) רב ממדית n $\quad p(x^*)=0$

וזו מט' גראם K שהוא סימטרית וחיובית מוגדרת תמיד (PSD)

הערה

$$\boxed{x^* \text{ הוא השארית האופטימלית (0 במקרה הטוב)}}$$

הפתרון המקורב הטוב ביותר

__משפט:__ בהינתן התבנית הריבועית: $p(x) = x^tKx - 2x^tf + c$ $\quad$ PD

המיוצגת על ידי $\{K, f, c\}$ אם K חיובית מוגדרת (כלומר $rkA = n$).

$$\boxed{\text{אז } p(x) \text{ יש נקודת מינימום יחידה גלובלית הנתונה על ידי } x^* = K^{-1}f}$$

ערך הפונקציה בנקודה הזו הוא:

$$p(x^*) = c - (x^*)^t K (x^*)$$

לסיכום עבור הבעיה המקורית שלנו מחפשים כי אם עמודות של A בת"ל:

$$p(x) = \|Ax-b\|_2^2 = (Ax-b)^t(Ax-b)$$

הפתרון יהיה:

$$x^* = K^{-1}f = (A^tA)^{-1}A^tb$$

נשים לב לדבר הבא: אם נכפול A^tA משמאל נקבל:

$$(A^tA)x^* = A^tb$$

$\Downarrow$ נוציא A^t החוצה ונעביר אגפים

$$A^t(Ax^*-b)=0$$

משוואה זה נקראת המשוואה הנורמלית

הסיבה לכך שבפתרון האופטימלי מתקבל כי השאריות אנכיות לעמודות המטריצה A.

כי כאשר וקטור אנך לוקטור זה אומר ש:

$$\boxed{\;v_1\;}\;\boxed{v_2}=0$$

בהסתכלות על מט' ריבועית והפיכה:

$$\underset{n\times n}{\boxed{\underset{\text{הפיכה}}{A}}}\;\underset{n\times 1}{\boxed{x}} = \underset{n\times 1}{\boxed{b}}$$

איך נמצא את ה-x האופטימלי? נסתכל על שני דרכים (ונראה שהן שקולות.

דרך 1

$x^*=A^{-1}b$ פתרון מדויק

בעזרת LS

$$\min_{\underline{x}} \|Ax-b\|_2^2 \to x^* = \underset{(K)}{(A^tA)}^{-1}A^tb$$

$$\to x^* = A^{-1}\cdot \underbrace{A^{-t}\cdot A^t}_{\text{מצטמצם}} b = A^{-1}b$$

A הפיכה אז:

$A^t=A$

בעצם כאשר A אינה ריבועית והפיכה נשתמש ב LS ונשים לב כי בעצם LS הוא הכללה להיפוך של מט' שנקראת: pseudo-inverse

סימון

$$\{A\} = A^+ = (A^tA)^{-1}A^t$$

$$\hookrightarrow A^+\cdot A = (A^tA)^{-1}A^tA = A^{-1}A^{-t}A^tA = I$$

איך LS מתחבר עם פירוק LU ו-Cholesky:

$$A = L\cdot D\cdot L^t = M\cdot M^t$$

$$(M=L\cdot D^{\frac{1}{2}})$$

- מתחילים עם בעיית LS שהוגדרה:

$$p(x)=\|Ax-b\|_2^2=(Ax-b)^t(Ax-b)$$

- כאמור הפתרון הנדרש הוא למעשה פתרון מ"מ:

$$(A^tA)x^*=A^tb$$

- נעשה פירוק cholesky עבור מט' A^tA

משפט: בהינתן הבעיה הריבועית

$$p(x) = x^t K x - 2x^t f + c$$

המאופיינת ע"י $\{K, f, c\}$ אם K חיובית חצי מוגדרת, אזי:

x^* לא יחיד
אבל כן גלובלי.

PSD

כל וקטור המקיים את המשוואה: $Kx^* = f \to x^* = K^{-1} f$

יהיה פתרון אופטימלי של הבעיה, עם ערך המינימום:

הוא לכן

$$p(x^*) = c - (x^*)^t K (x^*)$$

עמודותיה של K תלויות לינארית!
(במצב זה יש אינסוף פתרונות
כאשר $\ker K \neq \{0\}$ יש כמה וקטורים ב $\ker$

גזירת ביטויים מטריציים-וקטוריים:

הנתונה ע"י $p(x): \mathbb{R}^n \to \mathbb{R}^1$

$$p(x) = \|Ax - b\|_2^2 = (Ax - b)^t (Ax - b)$$

הוקטור הגרדיאנט של הפונקציה מוגדר ∇:

וקטור שמוגדר ∇ של שורה בו
זה נגזרת הפונקציה לפי x_i
$i \in \{1, \dots, n\}$

$$\nabla p(x) = \begin{bmatrix} \frac{dp(x)}{dx_1} \\ \vdots \\ \frac{dp(x)}{dx_2} \end{bmatrix}$$

∇ של הנגזרות (שווה לאפס)

נרצה כעת להכליל:

איך מתרגמים ביטוי וקטורי-מטריצי לוקטור גרדיאנט זה, $\nabla p(x)$:

<u>עבור m משוואות</u>:

$$p(x) = \sum_{k=1}^{m} (a_k x - b_k)^2 = \left\| \begin{bmatrix} - a_1 - \\ \vdots \\ - a_m - \end{bmatrix} \cdot x - \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \right\|_2^2 = \|Ax - b\|_2^2$$

$$\boxed{\nabla p(x)} = 2 \sum_{k=1}^{m} a_k^t (a_k x - b_k) = 2 \cdot \begin{bmatrix} | & & | \\ a_1^t & \dots & a_m^t \\ | & & | \end{bmatrix} \cdot \begin{bmatrix} a_1 x - b_1 \\ \vdots \\ a_m x - b_m \end{bmatrix} = \boxed{2A^t (Ax - b)}$$

- מט' A מגודל $m \times n$ (m שורות, n עמודות)
- וקטור x באורך n איברים
- וקטור b באורך m איברים
- נוסחת וקטור הגרדיאנט צריכה להיות חוקית ∇ שהוא באורך n נגזרות.

סיכום:

$$p(x) = \|Ax-b\|_2^2 \longrightarrow \nabla p(x) = 2A^t(Ax-b)$$

כי זה כמו $(Ax-b)^2$:

כאשר זוכרים $A'=A^t$!

כאשר משווים ל-0 כדי להגיע למינימום ע״י למשוואה הנורמלית

$$A^tAx = A^tb$$

כדי לבדוק אם מינימום נבדוק את הנגזרת השנייה

A^tA מוגדר חיובי

עבור פונקציה במבנה אחר:

טוב גם עבור ממדים גדולים יותר

$$p(x) = x^tKx - 2x^tf + c$$

$$\nabla p(x) = 2(Kx-f)$$

אם נשווה ל-0 כדי למצוא מינימום

$$2(Kx-f)=0 \rightarrow Kx=f$$

נשים לב שהגענו לנגזרות:

כלומר, לא נצטרך לגזור את $p(x)$ לפי כמה משתנים.

אלא נחשב את $K=A^tA$ ואת $f=A^tb$

ואז המשוואה יוצאת $x^*=K^{-1}f$

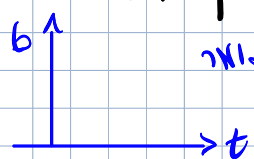
כך נמצא את x^* שהוא הפתרון האופטימלי הקטן ביותר.

כאופטימלי יהיה 0.

התאמת עקומות לנתונים בעזרת ריבועים פחותים:

- מהינתן סט של נתונים $\{t_i, b_i\}_{i=1}^m$ ניתן להתאים לנתונים אלו קירוב.

כלומר



- נעשה את הקירוב בעזרת משפחה של $\{\phi_k(t)\}_{k=1}^p$ פונקציות f כך

אשר מורכבות לינארית, כך ש $x_1, x_2, \dots, x_p$ הם המקדמים, (p מתאר כמות הפונקציות)

$$f(t) = \sum_{k=1}^p x_k \phi_k(t)$$

לכן מתקיים הביטוי הבא:

$$\{\hat{x}_i\}_{i=1}^p = \underset{\{x_i\}_{i=1}^p}{\mathrm{Arg\,min}} \sum_{i=1}^m \left(\sum_{k=1}^p x_k \phi_k(t_i) - b_i \right)^2$$

מעבר על m הנקודות

מעבר על p הפונקציות עבור כל נקודה

לכן בבעיה של מינ׳ - וקטורית מתקבל:

$$\left\| \begin{bmatrix} \phi_1(t_1) & \phi_2(t_1) & \cdots & \phi_p(t_1) \\ \phi_1(t_2) & \phi_2(t_2) & \cdots & \phi_p(t_2) \\ \vdots & & & \\ \phi_1(t_m) & \phi_2(t_m) & \cdots & \phi_p(t_m) \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \right\| = \|Ax-b\|_2^2 \xrightarrow[x]{} \min$$

המקדמים האופטימליים מתקבלים ע״י:

$$A^t(Ax-b)=0 \implies \hat{x} = (A^tA)^{-1}A^tb$$

לכן כאשר מוצאים את $\hat{x}$ האופטימלים נציב אותם מחדש כדי לקבל את התאמה הטובה לעקומה:

$$\hat{f}(t) = \sum_{k=1}^{p} \hat{x}_k \phi_k(t) \quad \text{אופטימלי}$$

WLS ~~ריבועים פחותים משוקללים:~~

[illegible] סט נתונים [illegible]: $\{t_i, b_i, w_i\}_{i=1}^m$ כש- w_i הוא מדד לאמינות הנקודה. ככל ש w_i גדול יותר כך הנקודה אמינה יותר ונרצה להשתמש בה.

נרצה להתאים לנתונים קירוב מהצורה:

$$f(t) = \sum_{k=1}^{p} x_k \phi_k(t)$$

כשלוקחים את w_i בחשבון.

נעשה את הדבר הבא:

$$\{\hat{x}_i\}_{i=1}^p = \underset{\{x_i\}_{i=1}^p}{\operatorname{Arg\,min}} \sum_{i=1}^m w_i \underbrace{\left(\sum_{k=1}^p x_k \phi_k(t_i) - b_i\right)^2}$$

$$\sum_{i=1}^m w_i r_i^2(\underline{x})$$

מינימום משוקלל של השגיאות

כעת [illegible] מט׳ אלכסונית W שמכילה את המשקלות באלכסונה הראשי.

$$(Ax-b)^t W (Ax-b) = \|Ax-b\|_W^2 \xrightarrow[x]{} \min$$

נפתח סוגריים:

$$x^t \underbrace{A^tWA}_{K} x - 2x^t \underbrace{A^tWb}_{f} + \underbrace{b^tWb}_{c}$$

ומתקיים:

$$A^tWAx^* = A^tWb$$

• אם A בלת״ל / האלכסון W חיובים ממש אז הפתרון יהיה יחיד כאשר A תהיה הפיכה!

A PD

ב- W אין ערכי אפס אז מינימום!

רגולריזציה של LS:

אין פתרון למערכת $Ax=b$

b לא נפרש מעמודות A

עמודות A בת"ל → פתרון יחיד לבעיית LS

עמודות A ת"ל → אינסוף פתרונות לבעיית ה-LS

ברגע שאין פתרון למשוואה $Ax-b=0$

<u>הרעיון הבסיסי:</u> פתרונה של בעיית ה-LS נתון ע"י המשוואה הנורמלית

$$(A^t\cdot A)x^* = A^t b$$

ריבוי פתרונות קורה כאשר המטריצה A^tA אינה הפיכה, כלומר סינגולרית

<u>משפט:</u> לבעיה הריבועית הבאה עם $\lambda>0$

$$\min \|Ax-b\|_2^2+\lambda\|x\|_2^2$$

יש פתרון יחיד הנתון ע"י:

$$x^*=(A^tA+\lambda I)^{-1}A^tb$$

הביטוי שהתווסף הוא הרגולריזציה והופך את הבעיה לרגולרית (מיוצבת)

<u>נשים לב:</u> המטריצה $A^tA+\lambda I$ עבור $\lambda>0$ היא מטריצה PD

$$\forall x\neq 0,\ x^t(A^tA+\lambda I)x = \|Ax\|_2^2+\lambda\|x\|_2^2>0$$

לכן, ניתן להגדיר בע"ית LS כך שהמשוואה הנורמלית החדשה תהיה: $(A^tA+\lambda I)x^*=A^tb$

בעצם! הרגולריזציה כאן מתעדפת פתרונות קטנים ב-L_2 בשביל הקנס המושת על הפתרון הארוך

<u>בצורה הכללית:</u>

$$\min_x \|Ax-b\|_2^2+\lambda\|\overset{\text{מט'}}{C}x\|_2^2$$

אם C^tC חיובית מוגדרת, אז הפתרון לבעיה יחיד.