# IML - Ex 1 - Theoretical Part

1. נתונה מטריצה $A=\begin{pmatrix}1 & 1 & 0\\ 1 & -1 & 2\end{pmatrix}$, נחשב את ה-SVD שלה.

ראשית, נחשב את $A^TA$:

$$A^TA=\begin{pmatrix}1 & 1\\ 1 & -1\\ 0 & 2\end{pmatrix}\begin{pmatrix}1 & 1 & 0\\ 1 & -1 & 2\end{pmatrix}=\begin{pmatrix}2 & 0 & 2\\ 0 & 2 & -2\\ 2 & -2 & 4\end{pmatrix}$$

נמצא את הערכים העצמיים שלה ע"י פתירת $0=\mathrm{Det}(A^TA-\lambda I)$:

$$A^TA-\lambda I=\begin{pmatrix}2-\lambda & 0 & 2\\ 0 & 2-\lambda & -2\\ 2 & -2 & 4-\lambda\end{pmatrix}$$

לפי הנוסחה לדטר' 3x3 שלמדנו בלינארית:

$$\mathrm{Det}(A^TA-\lambda I)=$$

$$(2-\lambda)\cdot\mathrm{Det}\begin{pmatrix}2-\lambda & -2\\ -2 & 4-\lambda\end{pmatrix}+2\cdot\det\begin{pmatrix}0 & 2\\ 2-\lambda & -2\end{pmatrix}=$$

$$(2-\lambda)\cdot\left((2-\lambda)(4-\lambda)-((-2)\cdot(-2))\right)+2\left(-2(2-\lambda)\right)=$$

$$(2-\lambda)(\lambda^2-6\lambda+4)+4\lambda-8=2\lambda^2-12\lambda+8-\lambda^3+6\lambda^2-4\lambda+4\lambda-8$$

$$=-\lambda^3+8\lambda^2-12\lambda=-\lambda(\lambda^2-8\lambda+12)=-\lambda(\lambda-2)(\lambda-6)$$

בהשוואה ל-0 נקבל: ערכים עצמיים. $\lambda=6,2,0$.

נחפש וקטורים עצמיים: עבור $\lambda=0$: $(A-0I)v=0$:

$$\begin{pmatrix}2 & 0 & 2\\ 0 & 2 & -2\\ 2 & -2 & 4\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}=0 \Rightarrow \begin{matrix}2x+2z=0\\ 2y-2z=0\\ 2x-2y+4z=0\end{matrix}$$

$x=-z$ 3-נ,נ3)

$\Rightarrow y=z \Rightarrow -2z-2z+4z=0$ ✓

לכן $v_1=\begin{pmatrix}-2\\2\\2\end{pmatrix}$ נבחר $z=1$ ונרמל:

$$\|v_1\|=\sqrt{(-1)^2+1^2+1^2}=\sqrt{3}$$

ולכן $v_1=\frac{1}{\sqrt3}\begin{pmatrix}-1\\1\\1\end{pmatrix}$.

עבור $\lambda=2$:

$$(A-2I)v_2=0 \Rightarrow \begin{pmatrix}0&0&2\\0&0&-2\\2&-2&2\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=0$$

$$\begin{aligned}2z&=0\\-2z&=0\\2x-2y+2z&=0\end{aligned} \Rightarrow \begin{aligned}z&=0\\x&=y\end{aligned}$$

לכן $v_2=\begin{pmatrix}y\\y\\0\end{pmatrix}$

נבחר $y=1$ ונרמל:

$$\|v_2\|=\sqrt{(1)^2+1^2}=\sqrt{2}$$

סה"כ $v_2=\sqrt{2}\begin{pmatrix}1\\1\\0\end{pmatrix}$

עבור $\lambda=6$:

$$(A-6I)(v_3)=0 \Rightarrow \begin{pmatrix}-4&0&2\\0&-4&-2\\2&-2&-2\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=0$$

$$\begin{aligned}-4x+2z&=0\\-4y-2z&=0\\2x-2y-2z&=0\end{aligned} \Rightarrow \begin{aligned}z&=2x\\z&=-2y\\y&=-x\end{aligned}$$

לכן $v_3=\begin{pmatrix}x\\-x\\2x\end{pmatrix}$ נבחר $x=1$ ונרמל:

$$\|v_3\|=\sqrt{1^2+(-1)^2+2^2}=\sqrt{6}$$

בסה"כ $v_3=\sqrt{6}\begin{pmatrix}1\\-1\\2\end{pmatrix}$. לכן לפי מה שטענו בתרגול:

$$V=\begin{pmatrix}\frac{1}{\sqrt6}&\frac{1}{\sqrt2}&-\frac{1}{\sqrt3}\\-\frac{1}{\sqrt6}&\frac{1}{\sqrt2}&\frac{1}{\sqrt3}\\\frac{2}{\sqrt6}&0&\frac{1}{\sqrt3}\end{pmatrix},\quad \Sigma=\begin{pmatrix}\sqrt6&0&0\\0&\sqrt2&0\end{pmatrix}$$

ועבור U:

$$u_1 = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{pmatrix}\left(\frac{1}{\sqrt{6}}\right)\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$u_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{pmatrix}\left(\frac{1}{\sqrt{2}}\right)\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

ולכן $U = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

2) $v$ הינו וקטור עמודה ב-$\mathbb{R}^n$ ו-$u$ הינו וקטור שורה ב-$\mathbb{R}^m$
לכן המכפלה שהינו $v \cdot u^T$ הינו בעצם הכפלה של שני וקטורי
עמודה (הראשון באורך $n$ והשני באורך $m$) ואפשר להסתכל על זה
כמכפלת מטריצות (הראשונה בגודל $n \times 1$ והשנייה בגודל $1 \times m$)
שתוצאתה $A$ מטריצה בגודל $n \times m$.

$$A = v \cdot u^T = \begin{pmatrix} v_1u_1 & \cdots & v_1u_m \\ \vdots & \ddots & \vdots \\ v_nu_1 & \cdots & v_nu_m \end{pmatrix}$$

העמודה ה-$j$ של $A$ היא $v \cdot u_j = A_j = \begin{pmatrix} v_1u_j \\ \vdots \\ v_nu_j \end{pmatrix}$ כלומר כל
עמודה היא הכפלה $v$ בסקלר $u_j$, ובגלל שכל העמודות הן אותו הוקטור
$v$ מוכפל ע"י סקלרים שונים נובע שהעמודות הן ת"ל לפי הגדרה.
הדבר זהה גם בשורות (הכפלה של $u$ בסקלר $v_j$) לכן ניתן
לראות מתוך שתי העובדות האלו שהמטריצה $A$ דרגתה 1.

3. יהי $(u_1,\dots,u_n)$ בסיס אורתונורמלי אזי לכל $i\neq j$ $\langle u_i,u_j\rangle=0$, לכל $i$ $\langle u_i,u_i\rangle=1$. יהי הוקטור $x\in\mathbb{R}^n$

ידוע לפיכך $\exists$ $a$ וקטורי הבסיס $B$ לומר $x=\sum_{i=1}^{n} a_i\cdot u_i$.

יהי אינדקס $1\le k\le n$: $\langle x,u_k\rangle=\langle\sum_{i=1}^{n} a_iu_i,u_k\rangle \overset{\text{לינ}}{=}$

$$\sum_{i=1}^{n}\langle a_iu_i,u_k\rangle \overset{\text{לינ}}{=} \sum_{i=1}^{n} a_i\langle u_i,u_k\rangle \overset{\text{לפי שוויון}}{=} a_k$$

לכן סה"כ לכל $i\in[1,n]$: $a_i=\langle x,u_i\rangle$, כנדרש.

4. נתון הוקטור $x=\begin{pmatrix}3\\-4\\1\\-2\end{pmatrix}$. נחשב:

$$\|x\|_1=\sum_i|x_i|=3+1+4+2=10$$

$$\|x\|_2=\sqrt{\sum_i x_i^2}=\sqrt{9+1+16+4}=\sqrt{30}$$

$$\|x\|_\infty=\max_i|x_i|=4$$

הנורמה שנותנה את הערך הכי נמוך היא $\|x\|_\infty=4$, מכיוון שנורמה זו לוקחת ערך יחיד הערך מוחלט שבוקטור של $x$ ויצא שהערך היחיד בקואורדינטה הוא 4 שזה ערך נמוך יחסית לשאר הנורמות ~~האחרות~~ יצאו הערכים גבוהים יותר כי הם סוכמים מס' ערכים, ולא ערך בודד.

~~i unit ball in $\mathbb{R}^2$ for $l_1$ norm: $\|x\|_1=|x_1|+|x|_2=1$~~

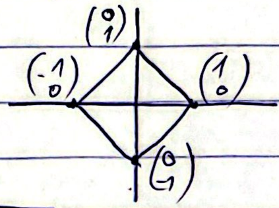
~~ברביע הראשון כלומר כש- $x_1,x_2\ge 0$ יש לנו $x=\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix}$~~

~~ברביע השני כלומר כש $x_1\le 0$, $x_2\ge 0$ יש $x=\begin{pmatrix}-1\\0\end{pmatrix}$~~

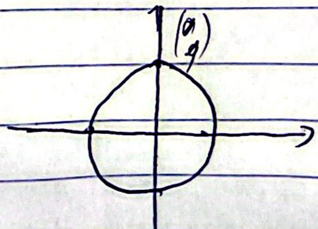
~~ברביעי יש את $\begin{pmatrix}0\\-1\end{pmatrix}$~~

i. unit balls in $\mathbb{R}^2$

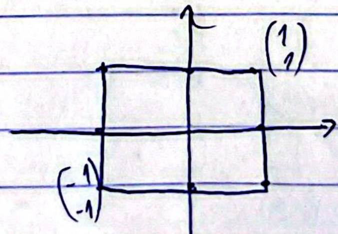
- עבור $l_1$: לפי הגדרה $\|x\|_1 = |x_1| + |x_2| = 1$
לכן הנורמה עבור x הם $\binom{-1}{0}, \binom{0}{-1}, \binom{0}{1}, \binom{1}{0}$. שרטוט:




- עבור $l_2$: לפי הגדרה $\|x_2\| = \sqrt{x_1^2+x_2^2} = 1$ כלומר
כל הערכים שראינו ב-$l_1$ בנוסף לכל הוקטורים שנמצאים במרחק
1 מהראשית, למשל: $\binom{\frac{\sqrt{2}}{2}}{\frac{\sqrt{2}}{2}}$ שרטוט: עיגול היחידה סביב הראשית:




- עבור $l_\infty$ לפי הגדרה $\|x\|_\infty = \max(|x_1|,|x_2|) = 1$. כל הנקודות
בהן $x_2 \vee x_1 = \pm 1$. שרטוט:




הנורמות האלו מדגימות את האופן שבו כל נורמה מגדירה "אורך".
$l_1$ מחברת קואורדינטות ולכן הגבולות שלה הם ליניאריים,
$l_2$ מדגימה מרחק אוקלידי: מה שיוצר מעגל יחידה סביב הראשית
ו-$l_\infty$ משתמשת במקסימום מה שיוצר 4 קווים שכל ערך מוגבל
בכל קואורדינטה $(1,-1)$.
כל אחת מהנורמות יוצרת גבולות גאומטריים ברורים שמסמנים את
אורך הוקטור לפי הגדרתם.

1.1.2

(b) נתונה $h(\sigma)=\frac{1}{2}\|f(\sigma)-y\|^2$, $f:\mathbb{R}^d\to\mathbb{R}^d$, $a\in\mathbb{R}^d$. נחשב את $\nabla h(\sigma)$:

נרשום מחדש:

$$\frac{1}{2}\|f(\sigma)-y\|^2=\frac{1}{2}(f(\sigma)-y)^T(f(\sigma)-y)=$$

$$\frac{1}{2}(f(\sigma)^Tf(\sigma)-2f(\sigma)^Ty+y^Ty)=$$

$$\frac{1}{2}\|f(\sigma)\|^2-f(\sigma)^Ty+\frac{1}{2}\|y\|^2$$

לפי כלל השרשרת:

$$\frac{\partial h}{\partial \sigma_i}=\frac{\partial h}{\partial f}\frac{\partial f}{\partial \sigma_i}=\frac{\partial}{\partial f}\left(\frac{1}{2}\|f(\sigma)\|^2-f(\sigma)^Ty+\frac{1}{2}\|y\|^2\right)\cdot\frac{\partial f}{\partial \sigma_i}=$$

$$(f(\sigma)-y)^T\cdot\frac{\partial f}{\partial \sigma_i}$$

ומאחר ו- $\frac{\partial f}{\partial \sigma_i}=J_\sigma(f)$ סה"כ נקבל $\nabla h(\sigma)=(f(\sigma)-y)^TJ_\sigma(f)$.

(c) נתונה פונקציית softmax: $S(x)_j=\frac{e^{x_j}}{\sum_{l=1}^{k}e^{x_l}}$, $S:\mathbb{R}^k\to[0,1]^k$.

נחשב $J_S$: מתקיים לכל $1\le j\le k$ המטריצה $J_S$ בגודל $k\times k$

כשהקואורדינטה ה-$ij$ שלה, $(J_S)_{ij}=\frac{\partial S_i(x)}{\partial x_j}$.

נגדיר עבור המכנה $D(x)=\sum_{l=1}^{k}e^{x_l}$ והמונה $g(x)=e^{x_i}$

אז לפי חוק המנה:

$$\frac{\partial}{\partial x_j}\left(\frac{g}{D}\right)=\frac{\frac{\partial g}{\partial x_j}D-g\frac{\partial D}{\partial x_j}}{D^2}$$

עבור המונה - $\frac{\partial g}{\partial x_j}=\frac{\partial e^{x_i}}{\partial x_j}$

אם $i=j$: $\frac{\partial e^{x_j}}{\partial x_j}=e^{x_j}$

אם $i\neq j$: $\frac{\partial e^{x_i}}{\partial x_j}=0$.

עבור המכנה:

$$\frac{\partial D}{\partial x_j} = \frac{\partial}{\partial x_j}\sum_{l=1}^{k} e^{x_l} = \sum_{l=1}^{k}\frac{\partial}{\partial x_j} e^{x_l}$$

בנוסף, אם $j=l$ $\frac{\partial}{\partial x_l} e^{x_l} = e^{x_l}$ אחרת $\frac{\partial}{\partial x_j} e^{x_l} = 0$

לכן

$$\frac{\partial S_i}{\partial x_j} = \frac{\frac{\partial}{\partial x_j}(e^{x_i})\cdot D(x) - e^{x_i}\cdot \frac{\partial}{\partial x_j} D(x)}{D(x)^2} =$$

נשתמש בהגדרה $\delta_{ij} = \begin{cases} 1 & i=j \\ 0 & i\neq j \end{cases}$ אז $\frac{\partial}{\partial x_j} e^{x_i} = \delta_{ij} e^{x_i}$

$$\frac{\delta_{ij} e^{x_i} D(x) - e^{x_i} e^{x_j}}{D(x)^2} = \frac{e^{x_i}}{D(x)} \cdot \frac{\delta_{ij} D(x) - e^{x_j}}{D(x)} =$$

נציין ש- $\frac{e^{x_i}}{D(x)} = S_i(x)$ ו- $\frac{e^{x_j}}{D(x)} = S_j(x)$ מתקיים:

$$S_i(x)\cdot\left(\delta_{ij} - \frac{e^{x_j}}{D(x)}\right) = S_i(x)(\delta_{ij} - S_j(x))$$ נפריד בין מקרים:

אם $i=j \Leftarrow \delta_{ij}=1$ $\frac{\partial S_i}{\partial x_i} = S_i(x)(1-S_i(x))$

אחרת $\delta_{ij}=0 \Leftarrow i\neq j$ $\frac{\partial S_i}{\partial x_j} = -S_i(x)S_j(x)$

לכן ניתן לכתוב את זה בצורה הבאה:

$$J_x(S) = \mathrm{diag}(S) - SS^T.$$

$\ker(X)=\ker(X^TX)$ צ"ל 3

1.2.1

S. (a) נראה הכלה דו כיוונית: כיוון ראשון $\subseteq$:

יהי $v\in\ker(X)$ אזי $Xv=0$, מתקיים!

$$X^TX(v)=X^T(Xv)=X^T0=0$$

לכן $v\in\ker(X^TX)$. כיוון שני $\supseteq$:

יהי $v\in\ker(X^TX)$ אזי $X^TXv=0$. מתקיים כי:

$$\|Xv\|^2=(Xv)^T(Xv)=v^TX^TXv\overset{\text{הנחה}}{=}v^T0=0$$

ו- $0=\|Xv\|^2 \iff Xv=0$ לכן $v\in\ker(X)$ כנדרש.

הראינו הכלה דו כיוונית ולכן השוויון מתקיים, כנדרש.

(b) עבור A מטריצה ריבועית צ"ל $Im(A^T)=\ker(A)^\perp$.

$\ker(A)^\perp=\{z\in\mathbb{R}^n \mid \forall x\in\ker(A)\ z^Tx=0\}$. נראה הכלה דו כיוונית:

כיוון ראשון $\subseteq$: יהי $z\in Im(A^T)$ אזי קיים $y\in\mathbb{R}^n$ כך ש- $z=A^Ty$

יהי $x\in\ker(A)$ אזי $Ax=0$. $z^Tx=(A^Ty)^Tx=$

$$y^TAx=y^T0=0$$

לכן z מאונך ל-x וכיוון ש-x שרירותי, z מאונך $\forall x\in\ker(A)$

כלומר $z\in\ker(A)^\perp$

כיוון שני: יהי $z\in\ker(A)^\perp$ אזי $\forall x\in\ker(A)\ z^Tx=0$

הניסוח $y\perp\ker(A)$ לכן יש פתרון y ל- $z=A^Ty$

כלומר $z\in Im(A^T)$ ומכיוון ש z שרירותי נקבל

ש- $Im(A^T)\supseteq\ker(A)^T$

ומסה"נ - $Im(A^T)=\ker(A)^\perp$.

c) $Xw=y$ מערכת משוואות לינאריות לא הומוגנית

$X$ ריבועית לא הפיכה.

צ"ל שלמערכת יש $\infty$ פתרונות אמ"מ $y \perp \ker(X^T)$.

נניח שלמערכת יש $\infty$ פתרונות. נסמן ש- $\forall v\in\ker(X)\ Xv=0$

הפתרונות של המערכת הן מהצורה $\forall v\in\ker(X)\ w=w_p+v$

כש- $w_p$ הוא פתרון ספציפי.

לכן במערכת: $\forall v\in\ker(X)\ y=X(w_p+v)$

לכן, מכיוון ש-$X$ ריבועית ולא הפיכה $\ker(X)\neq\{0\}$, מתקיים

ש- $y\in Im(X)$.

ואנחנו יודעים מלינארית ש- $(\ker(X^T))^{\perp} = Im(X)$

לכן מכאן נובע $y\perp\ker(X^T) \Leftarrow y\in\ker(X^T)^{\perp}$ כנדרש.

כעת נניח כי $y\perp\ker(X^T)$ כלומר $y\in Im(X)$ משמע

שקיים $w_p\in\mathbb{R}^n$ כך ש- $Xw_p=y$ לכן יש לפחות פתרון אחד

למערכת משוואות ומכיוון ש-$X$ ריבועית לא הפיכה

$\ker(X)\neq\{0\}$ ולכן מרחב הפתרונות $\{w_p+v \mid v\in\ker(X)\}$ הוא

אינסופי. כנדרש.

d) ~~נניח שבמקרה זה $X^TX$ הפיכה, אזי $\ker(X^TX)=\{0\}$ לכן~~

~~הפתרון למערכת הנורמלית הוא יחיד $w=(X^TX)^{-1}X^Ty$~~

~~אם $X^TX$ לא הפיכה אזי הפתרון של המערכת $y=Xw$ יש~~

~~אינסוף פתרונות מהצורה $w\in\{w_p+v\mid v\in\ker(X)\}$.~~

d) נחלק למקרים, אם $X^TX$ הפיכה, ההופכי שלה $(X^TX)^{-1}$ קיימת ולכן במשוואה $X^TXw=X^Ty$ ניתן להכפיל בהופכי ונקבל $w=(X^TX)^{-1}X^Ty$ לכן $w$ הוא פתרון יחיד.

אם $X^TX$ לא הפיכה, אז הגרעין שלה לא טריויאלי כלומר $\ker(X^TX)\neq\{0\}$ ומכיוון שהראנו ש- $\ker(X)=\ker(X^TX)$ מתקיים ש- $\ker(X)=\{0\}$ ולכן גם $X$ לא הפיכה.

כדי שלמערכת תהיה פתרון צריך להתקיים $X^Ty\in Im(X^TX)$ מלינארית יודעים ש- $\ker(X^TX)^{\perp}=Im(X^TX)$

ומכיוון ש- $\ker(X^TX)=\ker(X)$ צריך בסה"כ להתקיים $X^Ty\perp\ker(X)$.

כעת, אם $v\in\ker(X)$ לכן $Xv=0$ $y^TXv=y^T0=0$ ומכאן שלכל $y$ $X^Ty\in\ker(X)^{\perp}$. מכיוון שהגרעין של $X^TX$ לא טריויאלי, הפתרונות הן $w=w_p+v$ וכפי שהראנו, יש אינסוף פתרונות כאלה. כנדרש.

6. מכיוון ש- $X$ היא מטריצה $m\times d$, $X^TX$ היא מטריצה ריבועית מסדר $d\times d$. ובנוסף ש- $X^TX$ הפיכה, דרגתה חייבת להיות $d$ מה שאומר שהעמודות ב- $X$ הן בת"ל. דרגתה של $X$ היא לכל היותר המינ' מבין $m,d$ לכן:

אם $m<d$ אז $d>m\geq rk(X)$ ו- $d\geq rk(X^TX)$ מה שאומר שאינה הפיכה.

לכן בהכרח $m\geq d$ ו- $rk(X)=d$

מכאן נובע של- SVD יש $d$ ערכים סינגולריים חיוביים.

לכן $\Sigma$ היא מטריצה $m\times d$ ו-

$\forall i\in[d]\ \Sigma_{ii}=\sigma_i$

$\forall i\neq j\ \ \Sigma_{ij}=0$

ו- $\Sigma^+$ היא מטריצה $d\times m$

$\forall i\in[d]\ \ \Sigma^+_{ij}=\frac{1}{\sigma_i}$

else $\ 0$

כעת מתקיים ש- $X=U\Sigma V^T$ ו- $X^T=(U\Sigma V^T)^T=V\Sigma^TU^T$

לכן $X^TX=(V\Sigma^TU^T)(U\Sigma V^T)=$

מכיוון שU אורתונורמלית

$V\Sigma^T\Sigma V^T$

נתבונן ב- $\Sigma^T\Sigma$: מהגדרתם נובע שמטריצה $d\times d$.

$\forall i=j\ \ (\Sigma^T\Sigma)_{ij}=\sum_{k=1}^{m}(\Sigma^T)_{ik}\Sigma_{ki}=\sigma_i\sigma_i=\sigma_i^2$

$\forall i\neq j\ \ (\Sigma^T\Sigma)_{ij}=0$

לכן $\Sigma^T\Sigma=diag(\sigma_1^2,\dots,\sigma_d^2)$ ובסה"כ

$X^TX=V\,diag(\sigma_1^2,\dots,\sigma_d^2)\,V^T$.

$V$ אורתונורמלית ו- $\sigma_i>0$ לכל $i$ אז המטריצה הפיכה עם ההופכית:

$(X^TX)^{-1}=(V\,diag(\sigma_1^2,\dots,\sigma_d^2)V^T)^{-1}=$

$V\,diag(\frac{1}{\sigma_1^2},\dots,\frac{1}{\sigma_d^2})V^T$

ובסה״כ $(X^TX)^{-1}X^T = V\text{diag}(\frac{1}{\sigma_1^2},\dots\frac{1}{\sigma_d^2})V^T(V\Sigma^TU^T) =$

$$V\text{diag}(\frac{1}{\sigma_1^2},\dots\frac{1}{\sigma_d^2})\Sigma^TU^T$$

מתקיים - $\text{diag}(\frac{1}{\sigma_1^2},\dots\frac{1}{\sigma_d^2})\Sigma^T = \Sigma^+$ מכיוון ש-

$\forall\ i=j\le d\ \Sigma^T_{ij} = \sigma_i \Rightarrow \frac{1}{\sigma_i^2}\sigma_i = \frac{1}{\sigma_i}$

$\forall\ j\ne i\ \Sigma^T_{ij} = 0$

$\forall\ j>d\ \Sigma^T_{ij} = 0$

ולכן השיוויון מתקיים.

$$(X^TX)^{-1}X^T = V\Sigma^+U^T$$

לכן בסה״כ

ומכיוון שגם $X^+ = V\Sigma^+U^T$ יש שיוויון בין המטריצות אם נכפיל ב $y$:

$(X^TX)^{-1}X^Ty = X^+y$ כנדרש.