חלק תיאורטי

1.1.1 - Linear Algebra

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

$$A^T A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix}$$

נמצא את הערכים העצמיים של A^TA: ע"י דטרמיננטה והשוואה ל-0:

$$|A^TA - I\lambda| = \left| \begin{bmatrix} 2-\lambda & 0 & 2 \\ 0 & 2-\lambda & -2 \\ 2 & -2 & 4-\lambda \end{bmatrix} \right| = 0$$

פיתוח לפי עמודה ראשונה :)

$$2-\lambda\left((2-\lambda)\cdot(4-\lambda) - 2\cdot 2\right) - 0\cdot(\dots) + 2\left((0\cdot -2) - (2\cdot(2-\lambda))\right) = 0$$

$$= 2-\lambda\left(8 - 2\lambda - 4\lambda + \lambda^2 - 4\right) + 2(-4+\lambda) = 0$$

$$= \underbrace{(2-\lambda)}(\lambda^2 - 6\lambda + 4) - 4\underbrace{(2-\lambda)} = 0$$

$$= (\lambda^2 - 6\lambda)(2-\lambda) = 0$$

:)

$\boxed{\lambda_1 = 2}$

$\lambda^2 - 6\lambda = 0$

$\lambda(\lambda - 6) = 0$

$\boxed{\lambda_3 = 0}$ $\boxed{\lambda_2 = 6}$

הערכים הסינגולריים הם שורשי הע"ע:

$\Rightarrow$

$\sigma_1 = \sqrt{2}$

$\sigma_2 = \sqrt{6}$

$\sigma_3 = 0$

כעת נמצא את הוקטורים העצמיים:

$\lambda_1=2$

$$\begin{bmatrix} 2-2 & 0 & 2 \\ 0 & 2-2 & -2 \\ 2 & -2 & 4-2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & -2 \\ 2 & -2 & 2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \Rightarrow \boxed{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}} \cdot \sqrt{2}$$

נורמה

מתקיים

שהוקטור הנ"ל בכפל מימין עם המטריצה שווה לוקטור ה-0.

$\lambda_2=6$

$$\begin{bmatrix} 2-6 & 0 & 2 \\ 0 & 2-6 & -2 \\ 2 & -2 & 4-6 \end{bmatrix} = \begin{bmatrix} -4 & 0 & 2 \\ 0 & -4 & -2 \\ 2 & -2 & -2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \Rightarrow \boxed{\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}} \cdot \sqrt{6}$$

נורמה

$$-4x_1+2x_3=0 \Rightarrow x_3=2x_1$$
$$-4x_2-2x_3=0 \Rightarrow x_3=-2x_2$$
$$2x_1-2x_2-2x_3=0 \Rightarrow$$

$x_3=0$

$$\begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 2 & -2 & 4 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \Rightarrow \boxed{\begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}} \cdot \sqrt{3} \leftarrow$$ נורמה

$$2x_1+2x_3=0 \Rightarrow x_3=-x_1$$
$$2x_2-2x_3=0 \Rightarrow x_3=x_2$$

V- מטריצת וקטורים עצמיים מנורמלת ולכן:

$$\boxed{V=\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \end{bmatrix}}$$

$$\boxed{\Sigma = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{6} & 0 \\ 0 & 0 & 0 \end{bmatrix}}$$

מטריצה אלכסונית של הערכים הסינגולריים

נשתמש בנוסחה $AV=U\Sigma$ כדי למצוא את U:

נכפול

Σ^{-1} / $\Rightarrow$ $AV\Sigma^{-1}=U$

מימין

$A^{2\times3} \Rightarrow V^{3\times3},\ \Sigma^{2\times3},\ U^{2\times2}$

$$\Downarrow$$

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt2} & \frac{1}{\sqrt6} & \frac{1}{\sqrt3} \\ \frac{1}{\sqrt2} & -\frac{1}{\sqrt6} & -\frac{1}{\sqrt3} \\ 0 & \frac{2}{\sqrt6} & -\frac{1}{\sqrt3} \end{bmatrix}\begin{bmatrix} \sqrt2 & 0 & 0 \\ 0 & \sqrt6 & 0 \\ 0 & 0 & 0 \end{bmatrix} = U$$

$$\begin{bmatrix} \frac{1}{\sqrt2}+\frac{1}{\sqrt2} & \frac{1}{\sqrt6}-\frac{1}{\sqrt6} & \frac{1}{\sqrt3}-\frac{1}{\sqrt3} \\ \frac{1}{\sqrt2}-\frac{1}{\sqrt2} & \frac{1}{\sqrt6}+\frac{1}{\sqrt6}+\frac{4}{\sqrt6} & \frac{1}{\sqrt3}+\frac{1}{\sqrt3}-\frac{2}{\sqrt3} \\ 0 & 0 & 0 \end{bmatrix}\cdot\begin{bmatrix} \sqrt2 & 0 & 0 \\ 0 & \sqrt6 & 0 \\ 0 & 0 & 0 \end{bmatrix} = U$$

$$\Downarrow$$

$$= \begin{bmatrix} \sqrt2 & 0 & 0 \\ 0 & \sqrt6 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt2} & 0 & 0 \\ 0 & \frac{1}{\sqrt6} & 0 \\ 0 & 0 & 0 \end{bmatrix} = \boxed{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = U}$$

2) $v\in\mathbb{R}^n,\ u\in\mathbb{R}^m$

$$\Downarrow$$

מהגדרת מכפלה חיצונית:

$$A=\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}\cdot(u_1,\dots,u_m) = \begin{bmatrix} v_1u_1 & \cdots & v_1u_m \\ \vdots & & \vdots \\ v_nu_1 & \cdots & v_nu_m \end{bmatrix} \Rightarrow [v\otimes u]_{i,j} = v_i\cdot u_j$$

נסתכל על עמודות המטריצה:

$$\begin{pmatrix} v_1u_1 \\ \vdots \\ v_nu_1 \end{pmatrix}\quad\begin{pmatrix} v_1u_2 \\ \vdots \\ v_nu_2 \end{pmatrix}\quad\dots$$

נכתוב את הוקטורים כמכפלה (של המקדמים) סקלרית:

$$u_1\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \quad u_2\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \quad u_3\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

כלומר עמודות המטריצה הן כפל בסקלר שונה כל פעם של אותו וקטור:

$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$, כלומר וקטורי העמודות תלויים לינארית (ראה הגדרה של תלות לינארית)

ולכן $\text{rank}(A)=1$ כנדרש.

3) $x=\sum_{i=1}^{n} a_i u_i \in \mathbb{R}_n$ $\qquad a_i\in\mathbb{R}$, $(u_1,\dots,u_n)\in\mathbb{R}_n$

מלינאריות מכפלה פנימית

$$\langle x,u_j\rangle = \left\langle \sum_{i=1}^{n} a_i u_i , u_j \right\rangle = \sum_{i=1}^{n} a_i \langle u_i,u_j\rangle$$

כעת, נשים לב שהבסיס הוא אורתונורמלי ולכן מתקיים: $i\neq j$ $\langle u_i,u_j\rangle=0$

לכל $j,i\in n$ $\quad i=j$ $\langle u_i,u_j\rangle=1$

ולכן הסכום $\sum_{i=1}^{n} a_i\langle u_i,u_j\rangle$ הוא לא 0 רק כאשר $i=j$ ולכן:

$\boxed{a_j=\langle x,u_j\rangle}$ כלומר $\langle x,u_j\rangle = a_j \underbrace{\langle u_j,u_j\rangle}_{=1} = a_j$

כנדרש ☒

4) a) $x=\begin{pmatrix} 3 \\ -4 \\ 1 \\ -2 \end{pmatrix}$,

- $\|x\|_1 = 3+4+1+2=10$
- $\|x\|_2=\sqrt{3^2+4^2+1^2+2^2}=\sqrt{9+16+1+4}=\sqrt{30}$

$\cdot \|x\|_\infty = y$

נורמת המקסימום החזירה את הערך הנמוך ביותר, ונצאת מאחר וכל ערכי x
גדולים שווים מ-1 ולכן שתי הנורמות האחרות בוודאות יחזירו ערך שגדול
מ-y בסוף נורמת המקסימום מסומה בגרפים בין 1 ל-y.

b) $\underline{\ell_1}$:

$\|x\|_1 = 1$

$\Downarrow$

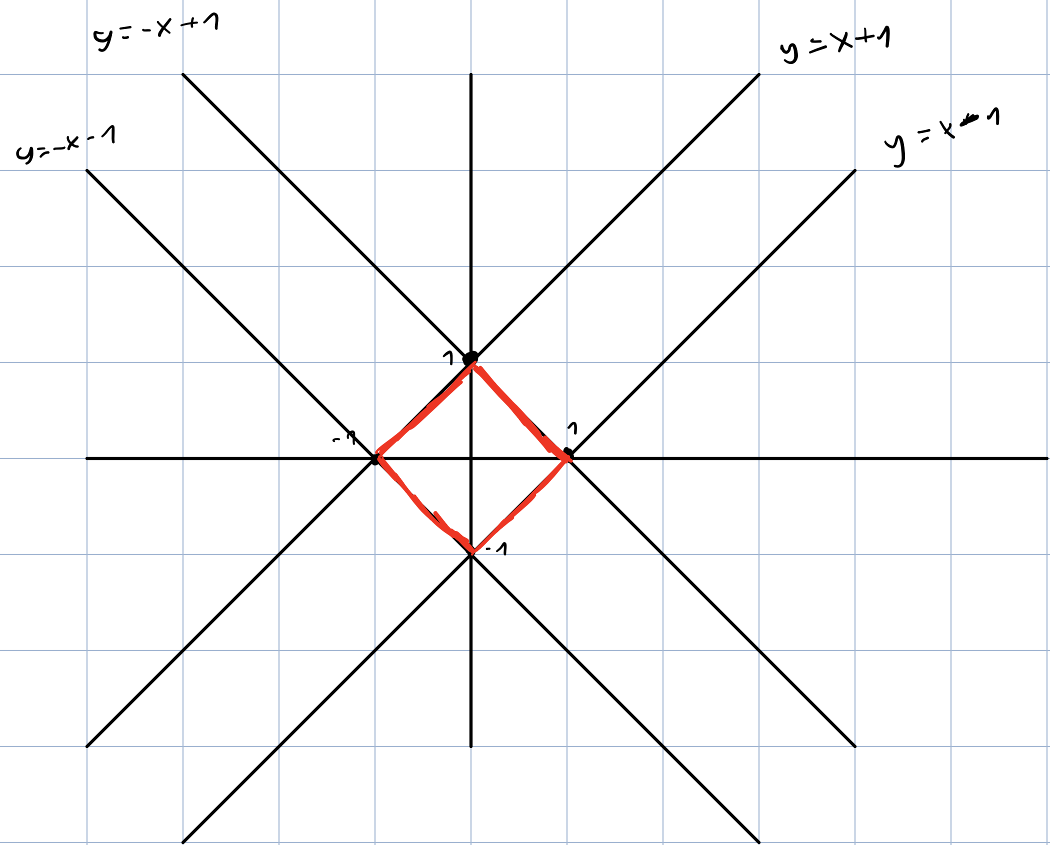
$x = (x,y) \Rightarrow |x| + |y| = 1$

$x + y = 1 \Rightarrow y = 1 - x$

$-x + y = 1 \Rightarrow y = 1 + x$

$x - y = 1 \Rightarrow y = x - 1$

$-x - y = 1 \Rightarrow y = -x - 1$



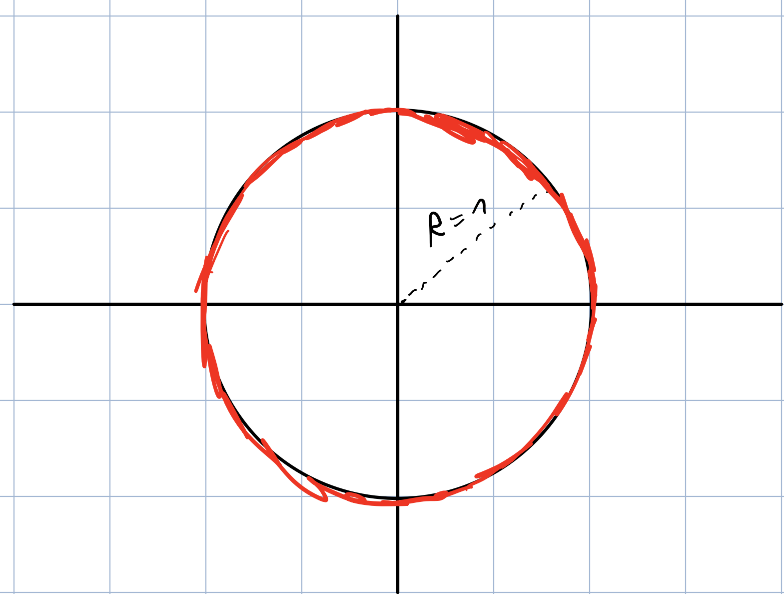
$\underline{\ell_2}$:

$\|x\|_2 = 1$

$\Downarrow$

$\sqrt{x^2 + y^2} = 1$

נוסחת מרחק
מרווסית
הקירים
ולכן נקבל
מעגל



ℓ_∞:

$\|x\|_\infty = 1$

$\Updownarrow$

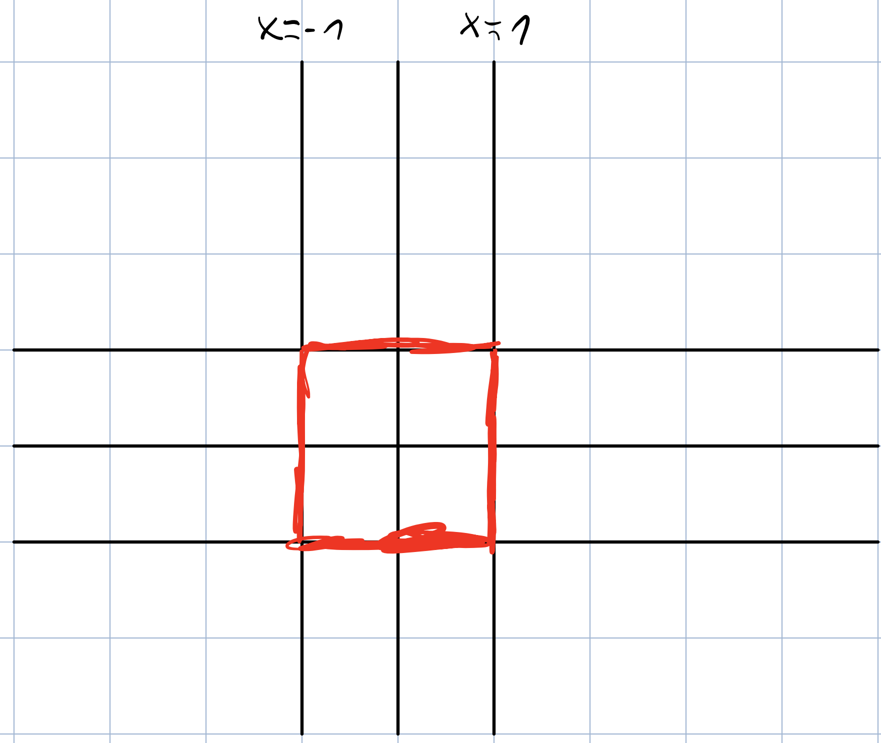
$\mathrm{Max}\{x, y\} = 1$

כלומר:

$x = \pm 1 \,,\ -1 < y < 1$

או $\Rightarrow$

$y = \pm 1 \,,\ -1 < x < 1$



השרטוטים בכלל משקפים איך כל נורמה מודדת "אורך" בצורה שונה:

ℓ_1: מעוין: משנה משמעות גבולה יותר לגבולה בשני הצירים במקביל:

כלומר בצורת אלכסון → כלומר האורך מושפע הכי הרבה משינוי בכל הצירים במקביל

ℓ_2: עיגול: בכל גידול האורך משתנה באותו אופן ולכן גם מייצג מרחק מוחלט.

ℓ_∞: ריבוע: האורך מושפע הכי הרבה משינוי בציר (הקואורדינטה) הגדולה ביותר.

1.1.2 - Multivariate Calculus

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

חוק השרשרת:

$$\nabla h(\sigma) = J_f(\sigma)^T \cdot \nabla h(f(\sigma))$$

$$\Downarrow$$

ראינו בתרגול ש: $\nabla_x \|g(x)\|^2 = 2g(x)^T \cdot \nabla_x g(x)$

ולכן:

$$\nabla h(\sigma) = J_f(\sigma)^T \cdot \nabla h(f(\sigma)) = J_f(\sigma)^T \cdot 2 \cdot \frac{1}{2} \cdot (f(\sigma) - y)$$

$$= \boxed{J_f(\sigma)^T \cdot (f(\sigma) - y)}$$

c) $S(x)_j = \dfrac{e^{x_j}}{\sum_{l=1}^{K} e^{x_l}}$ $\qquad S : \mathbb{R}^K \to [0,1]^K$:הגדרה

$$J_S \overset{\text{הגדרה}}{=} \begin{bmatrix} \frac{\partial s_1(x)}{\partial x_1} & \cdots & \frac{\partial s_1(x)}{\partial x_K} \\ \vdots & & \vdots \\ \frac{\partial s_K(x)}{\partial x_1} & \cdots & \frac{\partial s_K(x)}{\partial x_K} \end{bmatrix}$$

בתרגול ראינו שהנגזרת של כל קואורדינטה שבה $j=i$, כלומר על האלכסון היא:

$$\frac{\partial s_i}{\partial x_j} = \frac{d}{dx_j}\left(\frac{e^{x_j}}{\sum_k e^{x_k}}\right) \overset{\text{כלל המנה}}{=} \frac{e^{x_i}\sum e^{x_k} - e^{x_i} \cdot e^{x_j}}{\left(\sum_k e^{x_k}\right)^2} = \frac{e^{x_i}}{\sum e^{x_k}} \cdot \frac{\sum e^{x_k} - e^{x_j}}{\sum e^{x_k}} = S_i(1 - S_i)$$

נסתכל על המקרים בהם $i \neq j$

כלל המנה $\qquad$ גוזרים לפי משתנה אחר

$$\frac{\partial s_i}{\partial x_j} = \frac{d}{dx_j}\left(\frac{e^{x_i}}{\sum_k e^{x_k}}\right) \overset{\downarrow}{=} \frac{0 \cdot \sum_k e^{x_k} - e^{x_j} \cdot e^{x_i}}{\left(\sum_k e^{x_k}\right)^2} = \frac{-e^{x_j} \cdot e^{x_i}}{\left(\sum_k e^{x_k}\right)^2}$$

$$= \frac{-e^{x_j}}{\sum_k e^{x_k}} \cdot \frac{e^{x_i}}{\sum_k e^{x_k}} = -s_j \cdot s_i$$

ולכן היעקוביאן נראה כך:

$$J_s(x) = \begin{bmatrix} s_1(1-s_1) & -s_1 \cdot s_2 & \cdots & -s_1 s_k \\ -s_2 s_1 & s_2(1-s_2) & & -s_2 s_k \\ \vdots & & \ddots & \vdots \\ -s_k s_1 & \cdots & \cdots & s_k(1-s_k) \end{bmatrix}$$

1.2: Linear Regression:

$X \in R^{m \times d}$, $y \in R^m$

מטריצת הדגימות עם הפיצ'רים

וקטור התוצאות

5) a) $\Rightarrow$

נכפיל משמאל ב X^T

$$v \in \ker(X) \Rightarrow X \cdot v = 0 \Rightarrow X^T \cdot X \cdot v = X^T \cdot 0 = 0 \Rightarrow v \in \ker(X^T X)$$

$\Leftarrow$

נכפיל משמאל ב v^T

$$v \in \ker(X^T X) \Rightarrow X^T X \cdot v = 0 \Rightarrow v^T X^T X v = v^T \cdot 0 \Rightarrow \|Xv\|^2 = 0$$

תכונת נורמה

$$\Rightarrow Xv = 0 \Rightarrow v \in \ker(X)$$

$\blacksquare$

b) $A \in \mathbb{R}^{m \times m}$

קיים $w \in \mathbb{R}^m$

$$v \in Im(A^T) \Rightarrow A^T w = v$$

$$ker(A)^{\perp} = \{ v \mid v^T w = 0, \forall w \in ker(A) \}, \quad Aw = 0 \Downarrow w^T A^T = 0$$

כלומר, כל וקטור ב $Ker(A)$ אורתוגונלי לכל עמודה ב A^T

כלומר מימדי A^T ו $Ker(A)^{\perp}$ שווים או במילים אחרות $Im(A^T) = ker(A)^{\perp}$

c) X עמודות אך בלתי הפיכה כלומר קיים $v \neq 0$ כך ש:

$$Xv = 0$$

$$Xw = y \Leftrightarrow y \in Im(X^T) \underset{\text{סעיף ב'}}{\Leftrightarrow} y \in ker(X)^{\perp} \Leftrightarrow y \perp ker(X^T)$$

תנאי כב מקיים פתרון יחיד רק אם X הפיכה ומאחר והיא לא
ישנם אינסוף פתרונות.

d) $X^T X w = X^T y$

לפי סעיף a: $ker(X^T X) = ker(X)$, נניח ש $X^T X$ הפיכה אז
מתקיים $ker(X) = \{0\}$ כלומר ישנו פתרון יחיד ל X.

נניח ש $X^T X$ לא הפיכה אז מימד הגרעין של X גדול מר ואם מסעיף C כלומר והיו אינסוף
פתרונות שיפתרו את המשוואה הנורמלית.

1.2.2:

1) $x^{+}y \overset{\text{הגדרה}}{=} V\Sigma^{+}U^{T}y$ $(*)$

2) $x^{T}x = (U\Sigma V^{T})^{T}(U\Sigma V^{T}) = V\Sigma^{T}U^{T}U\Sigma V^{T}$

$= V\Sigma^{2}V^{T} \Rightarrow [x^{T}x]^{-1} = V(\Sigma^{2})^{-1}V^{T}$

U אורתוגונלית $(\Leftarrow)$ $U^{T}U = I$

Σ אלכסונית $(\Leftarrow)$ $\Sigma^{T} = \Sigma$

V אורתוגונלית

$\Downarrow$ אז

$[x^{T}x]^{-1}x^{T}y = V(\Sigma^{2})^{-1}V^{T}x^{T}y = V(\Sigma^{2})^{-1}V^{T}V\Sigma U^{T}y$

$= V(\Sigma^{2})^{-1}\Sigma U^{T}y \overset{\text{הגדרה } +}{=} V\Sigma^{+}U^{T}y \overset{(*)}{=} x^{+}y$

V אורתוגונלית $(\Leftarrow)$ $V^{+}V = I$

$\blacksquare$