Chapter 7

Linear Transformations

Po-Ning Chen, Professor

Department of Electrical Engineering

National Chiao Tung University

Hsin Chu, Taiwan 30050, R.O.C.

7.1 The idea of a linear transformation

- A **transformation** T is simply a mapping from $\mathcal{V}$ to $\mathcal{W}$.
- It is sometimes denoted as $T : \mathcal{V} \mapsto \mathcal{W}$.
- A transformation is linear if
  1. $T(\boldsymbol{v}_1) +_{\mathcal{W}} T(\boldsymbol{v}_2) = T(\boldsymbol{v}_1 +_{\mathcal{V}} \boldsymbol{v}_2)$
  2. $T(c \cdot_{\mathcal{V}} \boldsymbol{v}) = c \cdot_{\mathcal{W}} T(\boldsymbol{v})$.

  Here, we of course need to define the "vector addition" and "scalar-to-vector multiplication" over $\mathcal{V}$ and $\mathcal{W}$.

  Hence, $\mathcal{V}$ and $\mathcal{W}$ are usually vector spaces $\mathbb{V}$ and $\mathbb{W}$.

  For simplicity, we will drop the subscripts in "+" and "·" (in case there is no ambiguity in these operations).

7.1 The idea of a linear transformation

Important notes on linear transformation

- A line segment will be transformed to a line segment.

$$T(a_1\boldsymbol{v}_1 + (1-a_1)\boldsymbol{v}_2) = a_1T(\boldsymbol{v}_1) + (1-a_1)T(\boldsymbol{v}_2) = a_1\boldsymbol{w}_1 + (1-a_1)\boldsymbol{w}_2.$$

- Hence, a triangle will be transformed into a triangle.

- $\mathbf{0}$ in $\mathbb{V}$ will be transformed to $\mathbf{0}$ in $\mathbb{W}$.

$$T(\mathbf{0}) = T(0\cdot\boldsymbol{v}) = 0\cdot T(\boldsymbol{v}) = \mathbf{0}.$$

  Note again that the above $\mathbf{0}$ may be different. For example,

$$T\left(\begin{bmatrix}0\\0\\0\end{bmatrix}\right) = \begin{bmatrix}0\\0\end{bmatrix}.$$

Definition (Kernel): The **kernel** of a transformation T is the set of all $\boldsymbol{v}$ such that

$$T(\boldsymbol{v}) = \mathbf{0}.$$

- The concept of "kernel" becomes more evidently important when the transformation T is linear.

- For a linear transformation, the number of elements in the set

$$\mathcal{K}(\boldsymbol{w}) \triangleq \{\boldsymbol{v} \ : \ T(\boldsymbol{v}) = \boldsymbol{w}\}$$

  is a constant, independent of $\boldsymbol{w}$.

  Proof:

  – Suppose distinct $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \ldots, \boldsymbol{v}_k$ satisfying $T(\boldsymbol{v}_i) = \boldsymbol{w}$ for every $1 \leq i \leq k$, where $k = |\mathcal{K}(\boldsymbol{w})|$ is the size of the set $|\mathcal{K}(\boldsymbol{w})|$. Then, either $|\mathcal{K}(\tilde{\boldsymbol{w}})| \geq k$ or $|\mathcal{K}(\tilde{\boldsymbol{w}})| = 0$ because for a give $\tilde{\boldsymbol{v}}$ satisfying $T(\tilde{\boldsymbol{v}}) = \tilde{\boldsymbol{w}}$, we have

$$T(\tilde{\boldsymbol{v}} + \boldsymbol{v}_i - \boldsymbol{v}_1) = T(\tilde{\boldsymbol{v}}) + T(\boldsymbol{v}_i) - T(\boldsymbol{v}_1) = \tilde{\boldsymbol{w}} \text{ for } 2 \leq i \leq k.$$

  – Since we can interchange the role of $\boldsymbol{w}$ and $\tilde{\boldsymbol{w}}$, we conclude that $|\mathcal{K}(\boldsymbol{w})| = |\mathcal{K}(\tilde{\boldsymbol{w}})|$ if they are positive. $\square$

7.1 Kernel

- Note that since $T(\mathbf{0}) = \mathbf{0}$ for a linear transformation, $\mathcal{K}(\mathbf{0})$ cannot be empty. So
$$\frac{|\mathbb{V}|}{|\mathcal{K}(\mathbf{0})|}$$
will give the number of elements $\boldsymbol{w}$ in $\mathbb{W}$ such that $T(\boldsymbol{v}) = \boldsymbol{w}$ for some $\boldsymbol{v}$.

Definition (Range): The **range** of a transformation T is the set of all $\boldsymbol{w}$ such that
$$T(\boldsymbol{v}) = \boldsymbol{w} \text{ for some } \boldsymbol{v}.$$
I.e.,
$$\{\boldsymbol{w} \in \mathbb{W} \ : \ \exists\ \boldsymbol{v} \text{ such that } T(\boldsymbol{v}) = \boldsymbol{w}\}\ .$$

Important fact about linear transformation

- A linear transformation from a vector space $\mathbb{V}$ to a vector space $\mathbb{W}$ can always be represented as
$$A\boldsymbol{v} = \boldsymbol{w}$$
by properly selecting the matrix A.
- So, $\begin{cases} \textit{Kernel} = \textit{Null space} \text{ of } A \\ \textit{Range} = \textit{Column space} \text{ of } A \end{cases}$

(Problem 16, Section 7.1) Suppose T transposes every matrix M. Try to find a matrix A which gives $AM = M^{\mathsf{T}}$ for every M. Show that no matrix A will do it. *To professors:* Is this a linear transformation that doesn't come from a matrix.

Thinking over Problem 16: Define a transformation that maps a matrix $M_{2\times 2}$ to its transpose M^{T}. Is this a linear transformation?

Solution.

- $\begin{cases} T(M_1 + M_2) = (M_1 + M_2)^{\mathsf{T}} = M_1^{\mathsf{T}} + M_2^{\mathsf{T}} = T(M_1) + T(M_2) \\ T(c \cdot M) = (c \cdot M)^{\mathsf{T}} = cM^{\mathsf{T}} = c \cdot T(M) \end{cases}$

  hence, it is a linear transformation. □

7.1 Problem discussion

- There does not exist any matrix $A_{2\times 2}$ satisfying $AM = M^{\mathsf{T}}$.
- But there does exist a matrix $A_{4\times 4}$ satisfying

$$A \begin{bmatrix} m_{1,1} \\ m_{1,2} \\ m_{2,1} \\ m_{2,2} \end{bmatrix} = \begin{bmatrix} m_{1,1} \\ m_{2,1} \\ m_{1,2} \\ m_{2,2} \end{bmatrix}.$$

  So a linear transformation can always be represented as $A\boldsymbol{v}_{4\times 1} = \boldsymbol{w}_{4\times 1}$ (since the dimension of M is **four**).

7.2 The matrix of a linear transformation

For a linear transformation

$$T : \mathbb{V} \mapsto \mathbb{W},$$

how to find its equivalent matrix representation

$$A_{m\times n}\boldsymbol{v}_{n\times 1} = \boldsymbol{w}_{m\times 1}?$$

Answer:

- Denote the standard basis for vector space $\mathbb{V}$ by

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \boldsymbol{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

- Then, $T(\boldsymbol{e}_i) = A\boldsymbol{e}_i$ gives

$$\begin{bmatrix} T(\boldsymbol{e}_1) & T(\boldsymbol{e}_2) & \cdots & T(\boldsymbol{e}_n) \end{bmatrix} = A \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_n \end{bmatrix} = A.$$

□

7.2 The matrix of a linear transformation

Example. $\boldsymbol{v}(x) =$ a polynomial of x of order 3, and $T(\boldsymbol{v}) = \dfrac{\partial \boldsymbol{v}(x)}{\partial x}$.

- The bases for $\boldsymbol{v}(x)$ are 1, x, x^2, x^3. Or in vector forms, $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$.

- So, $A = \begin{bmatrix} T(1) & T(x) & \cdots & T(x^n) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2x & 3x^2 \end{bmatrix}$.

  Or in matrix form, $A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$.

- Hence, if $\boldsymbol{v}(x) = 1 + 2x + x^3 = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$, then

$$\frac{\partial \boldsymbol{v}(x)}{\partial x} = A \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} = 2 + 3x^2.$$

$\square$

7.2 The matrix of a linear transformation

For a linear transformation

$$T : \mathbb{V} \mapsto \mathbb{W},$$

how to find its equivalent matrix representation

$$A_{m\times n}\boldsymbol{v}_{n\times 1} = \boldsymbol{w}_{m\times 1}$$

(by the bases other than $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$)?

Answer:

- Denote a basis for vector space $\mathbb{V}$ by $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\ldots$, $\boldsymbol{v}_n$.

- Denote a basis for vector space $\mathbb{W}$ by $\boldsymbol{w}_1$, $\boldsymbol{w}_2$, $\ldots$, $\boldsymbol{w}_m$.

- Suppose that
$$T(\boldsymbol{v}_i) = b_{1,i}\boldsymbol{w}_1 + b_{2,i}\boldsymbol{w}_2 + \cdots + b_{m,i}\boldsymbol{w}_m.$$

  Then, $T(\boldsymbol{v}_i) = A\boldsymbol{v}_i$ gives

$$A\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n\end{bmatrix} = \begin{bmatrix}T(\boldsymbol{v}_1) & T(\boldsymbol{v}_2) & \cdots & T(\boldsymbol{v}_n)\end{bmatrix} = \begin{bmatrix}\boldsymbol{w}_1 & \cdots & \boldsymbol{w}_m\end{bmatrix}\begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,m} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,m} \end{bmatrix}.$$

Hence,

$$\begin{aligned} A &= \begin{bmatrix} T(\boldsymbol{v}_1) & T(\boldsymbol{v}_2) & \cdots & T(\boldsymbol{v}_n) \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_m \end{bmatrix} \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,m} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,m} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}^{-1} \end{aligned}$$

□

Example (Example 6 in the textbook): T projects a vector in $\Re^2$ onto the line passing via $(0, 0)$ and $(1, 1)$. Find the projection matrix A.

Solution 1:

- $A = \left[T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \; T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \right] = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$. □

Solution 2:

- Choose $\boldsymbol{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ and $\boldsymbol{v}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$.
- Then $T(\boldsymbol{v}_1) = \boldsymbol{v}_1$ and $T(\boldsymbol{v}_2) = \boldsymbol{0}$.
- Hence, $A = \left[T(\boldsymbol{v}_1) \;\; T(\boldsymbol{v}_2)\right] \left[\boldsymbol{v}_1 \;\; \boldsymbol{v}_2\right]^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$. $\square$

Solution 3:

- From Chapter 4, we know that the projection matrix onto a line is given by

$$A = \boldsymbol{a}\left(\boldsymbol{a}^{\mathsf{T}}\boldsymbol{a}\right)^{-1}\boldsymbol{a}^{\mathsf{T}} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

  where $\boldsymbol{a} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. $\square$

7.2 The matrix of a linear transformation

Change of basis is also a linear transformation.

Example (Example 9 in the textbook): A linear transformation T transforms

$$\text{input } \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix}, \text{ where } \boldsymbol{v} = s_1\boldsymbol{v}_1 + \cdots + s_n\boldsymbol{v}_n,$$

to

$$\text{output } \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}, \text{ where } \boldsymbol{v} = t_1\boldsymbol{w}_1 + \cdots + t_n\boldsymbol{w}_n.$$

Find the matrix $A_{n\times n}$ such that $T(\boldsymbol{s}) = A\boldsymbol{s} = \boldsymbol{t}$.

Answer:

- $\boldsymbol{v} = \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \end{bmatrix} \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_n \end{bmatrix} \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}$

  $\Longrightarrow \begin{bmatrix} \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_n \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \end{bmatrix} \boldsymbol{s} = \boldsymbol{t}.$

- Hence, $A = \begin{bmatrix} \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_n \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \end{bmatrix}$. □

When $\begin{bmatrix} \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_n \end{bmatrix} = I$ as the textbook does, $A = \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \end{bmatrix}$.

7.2 Combinations of linear transformation

- Sometimes, we need to determine the linear transformation of a linear transformation.

$$T : \mathbb{V} \mapsto \mathbb{W} \quad \text{and} \quad S : \mathbb{U} \mapsto \mathbb{V}.$$

Then, what is

$$TS : \mathbb{U} \mapsto \mathbb{W}?$$

I.e., $T(S(\boldsymbol{u}))$.

Answer:

– If $S(\boldsymbol{u}) = B\boldsymbol{u}$ and $T(\boldsymbol{v}) = A\boldsymbol{v}$, then $\boxed{TS(\boldsymbol{u}) = T(B\boldsymbol{u}) = AB\boldsymbol{u}}$.

$\square$

7.2 Combinations of linear transformation

We can prove the trigonometry formula using composition of linear transformation.

Example (Example 8 in the textbook): S rotates by θ and T rotates by $-\theta$.
So $TS(\boldsymbol{u}) = \boldsymbol{u}$. This proves $\cos^2(\theta) + \sin^2(\theta) = 1$ as

$$\begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} = \begin{bmatrix} \cos^2(\theta) + \sin^2(\theta) & 0 \\ 0 & \cos^2(\theta) + \sin^2(\theta) \end{bmatrix} = I.$$

$\square$

7.2 Wavelets transforms

- What are wavelets?

  Answer: "Wavelets" are "little waves," which have different lengths and are localized at different places.

  Example. Haar basis.

$$\boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{w}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \quad \boldsymbol{w}_3 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{w}_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$

  Note that the first one has no "waveform" but just a flat vector.

- The four vectors above are orthogonal, and can form a basis. I.e., any vector $\boldsymbol{v}$ can be written as the form:

$$\boldsymbol{v} = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + c_3\boldsymbol{w}_3 + c_4\boldsymbol{w}_4 = \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \boldsymbol{w}_3 & \boldsymbol{w}_4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix}.$$

In interpretation of these coefficients:

c_1	the average of components of $\boldsymbol{v}$
c_2	the difference between the first half and the second half
c_3	the detail of the first half
c_4	the detail of the second half

- The wavelet transforms are especially useful in data compression.

  Example. Continue from the previous example.

  If we do not need the detail of the second half, we can ignore c_4 and compress the data.

7.2 Discrete Fourier transform

- A very useful transform is the discrete Fourier transform.

- It has the shape of

$$F = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \kappa & \kappa^2 & \cdots & \kappa^{n-1} \\ 1 & \kappa^2 & \kappa^4 & \cdots & \kappa^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \kappa^{n-1} & \kappa^{2(n-1)} & \cdots & \kappa^{(n-1)^2} \end{bmatrix}$$

  where

$$\kappa = e^{\imath 2\pi/n}.$$

- Since $\kappa^n = 1$, the jth column of F is **approximately** a wave of cycle period $n/(j-1)$.

  Example. Suppose $n = 10$. Then, the 3rd column consists of

$$1, \kappa^2, \kappa^4, \kappa^6, \kappa^8, \kappa^{10}, \kappa^{12}, \kappa^{14}, \kappa^{16}, \kappa^{18}$$

  which is equivalent to

$$\underbrace{1, \kappa^2, \kappa^4, \kappa^6, \kappa^8}_{\text{cycle 1}}, \underbrace{1, \kappa^2, \kappa^4, \kappa^6, \kappa^8}_{\text{cycle 2}}$$

7.2 Discrete Fourier transform

- So the Fourier transform decomposes the signal/vectors into waves of different (cycle) frequencies.

7.3 Polar decomposition

- Any complex number $x + \imath y$ can be equivalently represented as

$$x + \imath\, y = re^{\imath\theta},$$

where $r = \sqrt{x^2 + y^2}$ and $\theta = \tan^{-1}(y/x)$.

- We can re-state the fact as:

Every complex number has the **polar form** as $e^{\imath\theta} r$, where r is **non-negative** and $e^{\imath\theta}$ is the rotation with respect to the x-axis.

- Analogously (but not exactly):

Every real square matrix A has the **polar decomposition form** as QH, where H is a **non-negative definite** matrix and Q is an orthogonal matrix.

If A is invertible, then H is **positive definite**.

Proof:

– (Reduced) SVD gives $A = U\Sigma V^{\mathsf{T}}$ with the diagonals of Σ are all chosen non-negative and U and V are both orthogonal matrix.

– Then, $A = U\Sigma V^{\mathsf{T}} = \underbrace{UV^{\mathsf{T}}}_{=Q}\underbrace{V\Sigma V^{\mathsf{T}}}_{=H}$. $\square$

7.3 Polar decomposition

- We can similarly prove that:

  Every real square matrix A has the **polar decomposition form** as KQ, where K is a **non-negative definite** matrix and Q is an orthogonal matrix.

  If A is invertible, then K is **positive definite**.

  Proof:

  – (Reduced) SVD gives $A = U\Sigma V^{\mathsf{T}}$ with the diagonals of Σ are all chosen non-negative and U and V are both orthogonal matrix.

  – Then, $A = U\Sigma V^{\mathsf{T}} = \underbrace{U\Sigma U^{\mathsf{T}}}_{=K}\underbrace{UV^{\mathsf{T}}}_{=Q}$. $\square$

7.3 Pseudoinverse or Moore-Penrose pseudoinverse

- A non-square matrix A does not have "inverse" (but may have left-inverse and right-inverse).

- But in terms of SVD, we can define its **pseudoinverse**.

Definition (Pseudoinverse): The psdueoinverse of a matrix A is

$$A^+_{n\times m} = V_{n\times n}\Sigma^+_{n\times m}U^{\mathsf{T}}_{m\times m},$$

where

$$\Sigma^+ = \begin{bmatrix} \sigma_1^{-1} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_2^{-1} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_r^{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}_{n\times m}$$

and

$$A_{m\times n} = U_{m\times m}\Sigma_{m\times n}V^{\mathsf{T}}_{n\times n}$$

is the SVD of A.

7.3 Pseudoinverse or Moore-Penrose pseudoinverse

– If A is invertible, then $A^{-1} = A^{+}$.

– $\begin{cases} A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i \text{ and } A^{+}\boldsymbol{u}_i = \dfrac{1}{\sigma_i}\boldsymbol{v}_i & \text{for } 1 \le i \le r. \\ A\boldsymbol{v}_i = \boldsymbol{0} \text{ and } A^{+}\boldsymbol{u}_i = \boldsymbol{0} & \text{for } i > r. \end{cases}$

– $\begin{cases} A_{m\times n}A^{+}_{n\times m}A_{m\times n} = U\Sigma \underbrace{V^{\mathrm{T}}V}_{=I} \Sigma^{+} \underbrace{U^{\mathrm{T}}U}_{=I} \Sigma V^{\mathrm{T}} = A_{m\times n} \\ A^{+}_{n\times m}A_{m\times n}A^{+}_{n\times m} = V\Sigma^{+}U^{\mathrm{T}}\, U\Sigma V^{\mathrm{T}}\, V\Sigma^{+}U^{\mathrm{T}} = A^{+}_{n\times m} \end{cases}$

– $\begin{cases} \boldsymbol{C}(A) = \boldsymbol{R}(A^{+}) = \boldsymbol{R}(A^{\mathrm{T}}) \\ \boldsymbol{R}(A) = \boldsymbol{C}(A^{+}) = \boldsymbol{C}(A^{\mathrm{T}}) \end{cases}$ So $A\boldsymbol{x} \in \boldsymbol{C}(A)$ and $A^{+}\boldsymbol{x} \in \boldsymbol{C}(A^{+}) = \boldsymbol{R}(A)$.

Note both $A^{\mathrm{T}}\boldsymbol{x}$ and $A^{+}\boldsymbol{x}$ are in $\boldsymbol{R}(A)$, but the mapping results could be different. See the below example $A^{\mathrm{T}} = \sigma\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}}$ and $A^{+} = \frac{1}{\sigma}\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}}$.

Example. Find A^{+} of $A = \sigma\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$.

Answer. $A_{m\times n} = \begin{bmatrix}\boldsymbol{u} & U_{m\times(m-1)}\end{bmatrix} \begin{bmatrix}\sigma & 0 \\ 0 & 0_{(m-1)\times(n-1)}\end{bmatrix} \begin{bmatrix}\boldsymbol{v} & V_{n\times(n-1)}\end{bmatrix}^{\mathrm{T}}$.

So, $A^{+}_{n\times m} = \begin{bmatrix}\boldsymbol{v} & V_{n\times(n-1)}\end{bmatrix} \begin{bmatrix}\sigma^{-1} & 0 \\ 0 & 0_{(n-1)\times(m-1)}\end{bmatrix} \begin{bmatrix}\boldsymbol{u} & U_{m\times(m-1)}\end{bmatrix}^{\mathrm{T}} = \dfrac{1}{\sigma}\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}}$. □

7.3 Pseudoinverse and least square approximation

What's the relation between pseudoinverse & least square approximation?

- if $A\boldsymbol{x} = \boldsymbol{b}$ has no solution, then we turn to find $\hat{\boldsymbol{x}}$ such that $\|A\hat{\boldsymbol{x}} - \boldsymbol{b}\|^2$ is minimized.

- In such case, the solution $\hat{\boldsymbol{x}}$ will satisfy the normal equations: $A^{\mathsf{T}}A\hat{\boldsymbol{x}} = A^{\mathsf{T}}\boldsymbol{b}$.

- Can we say one of the solutions is given by $\boldsymbol{x}^{+} = A^{+}\boldsymbol{b}$?

  Answer: Yes, but it only gives us one convenient solution, not the complete solution as given by the normal equations.

  – Let's check this **convenient solution**.

  $$A^{\mathsf{T}}A(\boldsymbol{x}^{+}) = A^{\mathsf{T}}AA^{+}\boldsymbol{b} = V\Sigma \cancel{U^{\mathsf{T}}U}\,\Sigma \cancel{V^{\mathsf{T}}V}\,\Sigma^{+}U^{\mathsf{T}}\boldsymbol{b} = V\Sigma U^{\mathsf{T}}\boldsymbol{b} = A^{\mathsf{T}}\boldsymbol{b}.$$

  – Based on the above derivation, any $(\boldsymbol{x}^{+} + \boldsymbol{x}^{(n)})$, where $\boldsymbol{x}^{(n)} \in \boldsymbol{N}(A)$, is also a solution. In fact, these give all the solutions of $A^{\mathsf{T}}A\boldsymbol{x} = A^{\mathsf{T}}\boldsymbol{b}$.

7.3 Pseudoinverse and least square approximation

– $\boldsymbol{x}^+ = A^+\boldsymbol{b} \in \boldsymbol{C}(A^+) = \boldsymbol{R}(A)$. So $\boldsymbol{x}^{(n)} \perp \boldsymbol{x}^+$. As a result,

$$\|\boldsymbol{x}^+ + \boldsymbol{x}^{(n)}\| \geq \|\boldsymbol{x}^+\|.$$

Hence, $\boldsymbol{x}^+$ is exactly the solution with the minimum length (among all solutions).

Note that Figure 7.4 in the textbook is wrong in that $A^+A \neq \begin{bmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{bmatrix}$.

So do not use this relationship.

7.3 Pseudoinverse and projections

- $AA^+A = A$ implies $AA^+(A\boldsymbol{b}) = (A\boldsymbol{b})$; hence, AA^+ maps any vector in $\boldsymbol{C}(A)$ to itself. Hence,

$$\left\{\boldsymbol{b} : AA^+\boldsymbol{x} = \boldsymbol{b} \text{ for some } \boldsymbol{x}\right\} \subset \boldsymbol{C}(A).$$

  Since AA^+ and A has the same rank, and since both of the above two sets form vector spaces,

$$\left\{\boldsymbol{b} : AA^+\boldsymbol{x} = \boldsymbol{b} \text{ for some } \boldsymbol{x}\right\} = \boldsymbol{C}(A).$$

- Similarly, A^+A is the projection matrix onto $\boldsymbol{R}(A)$.

7.3 Linear transform for basis changing (revisited)

SVD ($A = U\Sigma V^{\mathsf{T}}$) can be regarded as changing from **input basis $\boldsymbol{v}$'s** to **outout basis $\boldsymbol{u}$'s**.

Example (Example 9 in the textbook): A linear transformation T transforms

$$\text{input } \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix}, \text{ where } \boldsymbol{w} = s_1 \boldsymbol{v}_1 + \cdots + s_n \boldsymbol{v}_n,$$

to

$$\text{output } \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}, \text{ where } \boldsymbol{w} = t_1 \boldsymbol{u}_1 + \cdots + t_n \boldsymbol{u}_n.$$

Find the matrix $B_{n\times n}$ such that $T(\boldsymbol{s}) = B\boldsymbol{s} = \boldsymbol{t}$.

Answer:

- Previously in slide 7-13, we obtain

$$B = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \end{bmatrix} = U^{-1} V.$$

7.3 Linear transform for basis changing (revisited)

- An alternative choice of B is $B = U^{\mathsf{T}} \underbrace{(U\Sigma V^{\mathsf{T}})}_{A} V = U^{\mathsf{T}} A V$ for some Σ with non-zero diagonals.

  –

$$
\begin{aligned}
A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i \implies B\boldsymbol{s} &= U^{\mathsf{T}} A V \boldsymbol{s} \\
&= U^{\mathsf{T}} A (s_1 \boldsymbol{v}_1 + \cdots + s_n \boldsymbol{v}_n) \\
&= U^{\mathsf{T}} \left(\underbrace{\sigma_1 s_1}_{t_1} \boldsymbol{u}_1 + \cdots + \underbrace{\sigma_n s_n}_{t_n} \boldsymbol{u}_n \right) \\
&= U^{\mathsf{T}} U \boldsymbol{t} = \boldsymbol{t}.
\end{aligned}
$$

□

7.3 Linear transform for basis changing (revisited)

Further generalization (for $m \neq n$):

$$\boxed{U^{\mathsf{T}} A(s_1 \boldsymbol{v}_1 + \cdots + s_n \boldsymbol{v}_n) = \underbrace{\sigma_1 s_1}_{t_1} \boldsymbol{u}_1 + \cdots + \underbrace{\sigma_n s_n}_{t_n} \boldsymbol{u}_n}$$

- $U^{\mathsf{T}} A = U^{\mathsf{T}} U \Sigma V^{\mathsf{T}}$ maps a vector in the form of $(s_1 \boldsymbol{v}_1 + \cdots + s_n \boldsymbol{v}_n)$ (space spanned by $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$) to a vector in the form of $(t_1 \boldsymbol{u}_1 + \cdots + t_m \boldsymbol{u}_m)$ (space spanned by $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m$).

- If the rank of A is r, then $t_{r+1} = \cdots = t_m = 0$. Then, $U^{\mathsf{T}} A$ maps a vector in the vector space $\boldsymbol{C}(V)$ to a vector in the subspace $\boldsymbol{C}\left(\begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r \end{bmatrix}\right)$.

7.3 Pseudoinverse, left-inverse, right-inverse

It is natural to infer that:

1. Some matrices only have **left inverse** but have no **right inverse**.
2. Some matrices only have **right inverse** but have no **left inverse**.
3. Some matrices have neither **left inverse** nor **right inverse**.
4. Some matrices have both **left inverse** and **right inverse**.
   In such case, the inverse exists, and is equal to the **left inverse** and also the **right inverse**.

Question: When do each of the above cases happen?

1. If A has full column rank, i.e., $r = n$
2. If A has full row rank, i.e., $r = m$.
3. If A has no full column and row rank, i.e., $r < n$ and $r < m$.
4. If A has full column and row rank, i.e., $r = n = m$.

7.3 Pseudoinverse, left-inverse, right-inverse

Conceptual proof: SVD tells us that $A_{m\times n} = U_{m\times r}\Sigma_{r\times r}V^{\mathsf{T}}_{r\times n}$. Hence,

$$\begin{cases} A^{\mathsf{T}}A = V_{n\times r}\Sigma_{r\times r}U^{\mathsf{T}}_{r\times m}U_{m\times r}\Sigma_{r\times r}V^{\mathsf{T}}_{r\times n} = V_{n\times r}\Sigma^2_{r\times r}V^{\mathsf{T}}_{r\times n} \\ AA^{\mathsf{T}} = U_{m\times r}\Sigma_{r\times r}V^{\mathsf{T}}_{r\times n}V_{n\times r}\Sigma_{r\times r}U^{\mathsf{T}}_{r\times m} = U_{m\times r}\Sigma^2_{r\times r}U^{\mathsf{T}}_{r\times m} \end{cases}$$

1. If $r = n$, then the left inverse is equal to $(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}$.

   $$\begin{aligned} \text{In such case, } A^+ &= \text{left inverse} = (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} \\ &= V_{n\times r}\Sigma^{-2}_{r\times r}V^{\mathsf{T}}_{r\times n}V_{n\times r}\Sigma_{r\times r}U^{\mathsf{T}}_{r\times n} = V_{n\times r}\Sigma^{-1}_{r\times r}U^{\mathsf{T}}_{r\times m}. \end{aligned}$$

2. If $r = m$, then the right inverse is equal to $A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}$.

   In such case, $A^+ =$ right inverse $= A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}$.

3. It is not possible to find $B_{n\times m}$ and $C_{n\times m}$ such that $B_{n\times m}A_{m\times n} = I_{n\times n}$ and $A_{m\times n}C_{n\times m} = I_{m\times m}$.

   In such case, A^+ still exists but it is neither left inverse nor right inverse.

4. The inverse is equal to $(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} = A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}$.

   In such case, $A^+ =$ inverse $= (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} = A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}$.

This is the reason why A^+ is named the **pseudoinverse**. It is the left or right inverse whenever they exist!

7.3 Pseudoinverse, left-inverse, right-inverse

Example. (Worked Example 7.3A) For the first three cases, let's examine

$$A_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 2\sqrt{2} \\ 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1 \end{bmatrix}}_{V^{\mathsf{T}}}$$

$$A_2 = \begin{bmatrix} 2 & 2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \end{bmatrix}}_{U^{\mathsf{T}}} \underbrace{\begin{bmatrix} 2\sqrt{2} & 0 \end{bmatrix}}_{\Sigma} \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{V^{\mathsf{T}}}$$

$$A_3 = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} = \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{V^{\mathsf{T}}}$$

Solution.

$$A_1^+ = \underbrace{\begin{bmatrix}1\end{bmatrix}}_{V}\underbrace{\begin{bmatrix}1/(2\sqrt{2}) & 0\end{bmatrix}}_{\Sigma^+}\underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}}_{U^{\mathsf{T}}} = \begin{bmatrix}1/4 & 1/4\end{bmatrix} \implies A_1^+A_1 = \begin{bmatrix}1\end{bmatrix}$$

$$A_2^+ = \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}}_{V}\underbrace{\begin{bmatrix}1/(2\sqrt{2})\\ 0\end{bmatrix}}_{\Sigma^+}\underbrace{\begin{bmatrix}1\end{bmatrix}}_{U^{\mathsf{T}}} = \begin{bmatrix}1/4\\ 1/4\end{bmatrix} \implies A_2A_2^+ = \begin{bmatrix}1\end{bmatrix}$$

$$A_3^+ = \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}}_{V}\underbrace{\begin{bmatrix}1/4 & 0\\ 0 & 0\end{bmatrix}}_{\Sigma^+}\underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}}_{U^{\mathsf{T}}} = \begin{bmatrix}1/8 & 1/8\\ 1/8 & 1/8\end{bmatrix} \implies A_3^+A_3 = A_3A_3^+ = \frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$$

$\square$