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. $T(\mathbf{v}_1) +_\mathcal{W} T(\mathbf{v}_2) = T(\mathbf{v}_1 +_\mathcal{V} \mathbf{v}_2)$
  
  2. $T(c \cdot_\mathcal{V} \mathbf{v}) = c \cdot_\mathcal{W} T(\mathbf{v})$

  where "+$_\mathcal{W}$" and "+$_\mathcal{V}$" and ".$_\mathcal{W}$" and ".$_\mathcal{V}$" denote some general "additions" and "multiplications" defined over $\mathcal{W}$ and $\mathcal{V}$, respectively.

$\mathcal{V}$ 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.
  
  \[ T(a_1 \mathbf{v}_1 + (1 - a_1) \mathbf{v}_2) = a_1 T(\mathbf{v}_1) + (1 - a_1)T(\mathbf{v}_2) = a_1 \mathbf{w}_1 + (1 - a_1)\mathbf{w}_2. \]

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

• 0 in \( V \) will be transformed to 0 in \( W \).
  
  \[ T(\mathbf{0}) = T(0 \cdot \mathbf{v}) = 0 \cdot T(\mathbf{v}) = \mathbf{0}. \]

Note again that the 0 in \( V \) and the 0 in \( W \) may be different. For example,

\[ T \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \]
7.1 Kernel

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

$$T(v) = 0.$$ 

- 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

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

is a constant, independent of $w$.

**Proof:**

- Suppose distinct $v_1, v_2, v_3, \ldots, v_k$ satisfying $T(v_i) = w$ for every $1 \leq i \leq k$, where $k = |\mathcal{K}(w)|$ is the size of the set $\mathcal{K}(w)$. Then, either $|\mathcal{K}(\tilde{w})| \geq k$ or $|\mathcal{K}(\tilde{w})| = 0$ because for a given $\tilde{v}$ satisfying $T(\tilde{v}) = \tilde{w}$, we have

$$T(\tilde{v} + v_i - v_1) = T(\tilde{v}) + T(v_i) - T(v_1) = \tilde{w} \text{ for } 2 \leq i \leq k.$$  

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

- Note that since \( T(0) = 0 \) for a linear transformation, \( \mathcal{K}(0) \) cannot be empty. So

\[
\frac{|\mathbb{V}|}{|\mathcal{K}(0)|}
\]

will give the number of elements \( w \) in \( \mathbb{W} \) such that \( T(v) = w \) for some \( v \).

**Definition (Range):** The **range** of a transformation \( T \) is the set of all \( w \) such that

\[
T(v) = w \text{ for some } v.
\]

I.e.,

\[
\{ w \in \mathbb{W} : \exists v \text{ such that } T(v) = w \}.
\]

**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

\[
Av = w
\]

by properly selecting the matrix \( A \).

- So,

\[
\begin{cases} 
   \text{Kernel} = \text{Null space of } A \\
   \text{Range} = \text{Column space of } A
\end{cases}
\]
(Problem 16, Section 7.1) Suppose $T$ transposes every matrix $M$. Try to find a matrix $A$ which gives $AM = M^T$ 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.**

\[
\begin{align*}
T(M_1 + M_2) &= (M_1 + M_2)^T = M_1^T + M_2^T = T(M_1) + T(M_2) \\
T(c \cdot M) &= (c \cdot M)^T = cM^T = c \cdot T(M)
\end{align*}
\]

hence, it is a linear transformation.
7.1 Problem discussion

- There does not exist any matrix $A_{2\times2}$ satisfying $AM = M^T$.
- But there does exist a matrix $A_{4\times4}$ 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 $Av_{4\times1} = w_{4\times1}$ (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

\[ A_{m \times n}v_{n \times 1} = w_{m \times 1}? \]

Answer:

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

  \[ e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad e_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \]

- Then, \( T(e_i) = Ae_i \) gives

  \[ \begin{bmatrix} T(e_1) & T(e_2) & \cdots & T(e_n) \end{bmatrix} = A \begin{bmatrix} e_1 & e_2 & \cdots & e_n \end{bmatrix} = A. \]
7.2 The matrix of a linear transformation

Example. \( \mathbf{v}(x) = \) a polynomial of \( x \) of order 3, and \( T(\mathbf{v}) = \frac{\partial \mathbf{v}(x)}{\partial x} \).

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

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

Or in matrix form, \( A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \).
7.2 The matrix of a linear transformation

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

\[
\frac{\partial \mathbf{v}(x)}{\partial x} = A \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} = 2 + 3x^2.
\]
7.2 The matrix of a linear transformation

For a linear transformation

\[ T : \mathbb{V} \rightarrow \mathbb{W}, \]

how to find its equivalent matrix representation

\[ A_{m \times n} v_{n \times 1} = w_{m \times 1} \]

(by the bases other than \( e_1, e_2, \ldots, e_n \))?

Answer:

- Denote a basis for vector space \( \mathbb{V} \) by \( v_1, v_2, \ldots, v_n \).
- Denote a basis for vector space \( \mathbb{W} \) by \( w_1, w_2, \ldots, w_m \).
- Suppose that

\[ T(v_i) = b_{1,i}w_1 + b_{2,i}w_2 + \cdots + b_{m,i}w_m. \]

Then, \( T(v_i) = Av_i \) gives

\[
A \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} T(v_1) & T(v_2) & \cdots & T(v_n) \end{bmatrix} = \begin{bmatrix} w_1 & \cdots & w_m \end{bmatrix} \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,m} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,m} \end{bmatrix}.
\]
7.2 The matrix of a linear transformation

Hence,

\[ A = \left[ T(\mathbf{v}_1) \ T(\mathbf{v}_2) \ \cdots \ T(\mathbf{v}_n) \right] \left[ \mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n \right]^{-1} \]

\[ = \left[ \mathbf{w}_1 \ \cdots \ \mathbf{w}_m \right] \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,m} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,m} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \cdots \\ \mathbf{v}_n \end{bmatrix}^{-1} \]

\[ = \left[ \mathbf{w}_1 \ \cdots \ \mathbf{w}_m \right] \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,m} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,m} \end{bmatrix} \]

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

Solution 1:

\( A = \left[ T \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \ T \left( \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \right] = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \).
7.2 The matrix of a linear transformation

Solution 2:

• Choose \( \mathbf{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \) and \( \mathbf{v}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \).

• Then \( T(\mathbf{v}_1) = \mathbf{v}_1 \) and \( T(\mathbf{v}_2) = \mathbf{0} \).

• Hence, \( A = \begin{bmatrix} T(\mathbf{v}_1) & T(\mathbf{v}_2) \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \).

Solution 3:

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

\[
A = \mathbf{a} (\mathbf{a}^\top \mathbf{a})^{-1} \mathbf{a}^\top = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}
\]

where \( \mathbf{a} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \).


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

\[
\begin{bmatrix}
\vdots \\
s_1 \\
\vdots \\
s_n
\end{bmatrix}, \text{ where } \mathbf{v} = s_1 \mathbf{v}_1 + \cdots + s_n \mathbf{v}_n,
\]

to

\[
\begin{bmatrix}
\vdots \\
t_1 \\
\vdots \\
t_n
\end{bmatrix}, \text{ where } \mathbf{v} = t_1 \mathbf{w}_1 + \cdots + t_n \mathbf{w}_n.
\]

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

*Answer:*

- \( \mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \cdots \mathbf{v}_n \end{bmatrix} \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 \cdots \mathbf{w}_n \end{bmatrix} \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix} \)

- \( \Rightarrow \begin{bmatrix} \mathbf{w}_1 \cdots \mathbf{w}_n \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{v}_1 \cdots \mathbf{v}_n \end{bmatrix} \mathbf{s} = \mathbf{t} \).

- Hence, \( A = \begin{bmatrix} \mathbf{w}_1 \cdots \mathbf{w}_n \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{v}_1 \cdots \mathbf{v}_n \end{bmatrix} \).

When \( \begin{bmatrix} \mathbf{w}_1 \cdots \mathbf{w}_n \end{bmatrix} = I \) as the textbook does, \( A = \begin{bmatrix} \mathbf{v}_1 \cdots \mathbf{v}_n \end{bmatrix} \).
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., \( T(S(u)) \).

**Answer:**

- If \( S(u) = Bu \) and \( T(v) = Av \), then \( TS(u) = T(Bu) = ABu \).  \( \square \)
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 $\theta$ and $T$ rotates by $-\theta$. So $TS(u) = u$. This proves $\cos^2(\theta) + \sin^2(\theta) = 1$ as

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

\[
\begin{align*}
\mathbf{w}_1 &= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \\
\mathbf{w}_2 &= \begin{bmatrix} 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \\
\mathbf{w}_3 &= \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \\
\mathbf{w}_4 &= \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}
\end{align*}
\]

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 \( \mathbf{v} \) can be written as the form:

\[
\mathbf{v} = c_1 \mathbf{w}_1 + c_2 \mathbf{w}_2 + c_3 \mathbf{w}_3 + c_4 \mathbf{w}_4 = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \mathbf{w}_3 & \mathbf{w}_4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix}
\]
7.2 Wavelets transforms

In interpretation of these coefficients:

- $c_1$ the average of components of $\mathbf{v}$
- $c_2$ 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

\[
\kappa = e^{i2\pi/n}.
\]

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

\textit{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 $x + iy$ can be equivalently represented as

$$x + iy = re^{i\theta},$$

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 $e^{i\theta}r$, where $r$ is **non-negative** and $e^{i\theta}$ 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 $A = U\Sigma V^T$ with the diagonals of $\Sigma$ are all chosen non-negative and $U$ and $V$ are both orthogonal matrix.

- Then, $A = U\Sigma V^T = U\underbrace{V^T}_=QV\underbrace{\Sigma V^T}_=H$  \[\square\]
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 $A = U\Sigma V^T$ with the diagonals of $\Sigma$ are all chosen non-negative and $U$ and $V$ are both orthogonal matrix.
- Then, $A = U\Sigma V^T = U\Sigma U^T U V^T = KQ$. $\square$
7.3 Pseudoinverse or Moore-Penrose pseudoinverse

- A non-square matrix $A$ does not have “inverse” (but may have left-inverse or right-inverse).

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

**Definition (Pseudoinverse):** The pseudoinverse of a matrix $A$ is

$$A_{n \times m}^+ = V_{n \times n} \Sigma_{n \times m}^+ U_{m \times m}^T,$$

where

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

and

$$A_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T$$

is the SVD of $A$. 
7.3 Pseudoinverse or Moore-Penrose pseudoinverse

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

\[
\begin{align*}
A_v = \sigma_i u_i & \quad \text{and} \quad A^+ u_i = \frac{1}{\sigma_i} v_i \quad \text{for } 1 \leq i \leq r. \\
A_v = 0 & \quad \text{and} \quad A^+ u_i = 0 \quad \text{for } i > r.
\end{align*}
\]

\[
\begin{align*}
A_{m \times n} A^+_{n \times m} A_{m \times n} &= U \Sigma V^T \Sigma^+ U^T \Sigma V^T = A_{m \times n} \\
A^+_{n \times m} A_{m \times n} A^+_{n \times m} &= V \Sigma^+ U^T U \Sigma V^T V \Sigma^+ U^T = A^+_{n \times m}
\end{align*}
\]

\[
\begin{align*}
C(A) &= R(A^+) = R(A^T) \\
R(A) &= C(A^+) = C(A^T)
\end{align*}
\]

So $A x \in C(A)$ and $A^+ x \in C(A^+) = R(A)$.

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

Example. Find $A^+$ of $A = \sigma u v^T$.

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

So, $A^+_{n \times m} = \begin{bmatrix} v & V_{n \times (n-1)} \end{bmatrix} \begin{bmatrix} \sigma^{-1} & 0 \\ 0 & 0_{(n-1) \times (m-1)} \end{bmatrix} \begin{bmatrix} u & U_{m \times (m-1)} \end{bmatrix}^T = \frac{1}{\sigma} v u^T$. \qed
7.3 Pseudoinverse and least square approximation

What is the relation between pseudoinverse & least square approximation?

- If \(Ax = b\) has no solution, then we turn to find \(\hat{x}\) such that \(\|A\hat{x} - b\|^2\) is minimized.

- In such case, the solution \(\hat{x}\) will satisfy the normal equations: \(A^TA\hat{x} = A^Tb\).

- Can we say one of the solutions is given by \(x^+ = A^+b\)?

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

- Let us check this convenient solution.

\[
A^T A(x^+) = A^T A A^+ b = V \Sigma U^T U \Sigma V^T \Sigma^+ U^T b = V \Sigma U^T b = A^T b.
\]

- Based on the above derivation, any \((x^+ + x^{(n)})\), where \(x^{(n)} \in N(A)\), is also a solution. In fact, these give all the solutions of \(A^T A x = A^T b\).
7.3 Pseudoinverse and least square approximation

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

\[ \|x^+ + x^{(n)}\| \geq \|x^+\|. \]

Hence, \( x^+ \) 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 $AA^+ (Ab) = (Ab)$; hence, $AA^+$ maps any vector in $\mathbf{C}(A)$ to itself. Hence,

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

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

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

- Similarly, $A^+ A$ is the projection matrix onto $\mathbf{R}(A)$. 
7.3 Linear transform for basis changing (revisited)

SVD (i.e., $A_{n\times n} = U\Sigma V^T$) can be regarded as changing from input basis $v$’s to output basis $u$’s.

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

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

to

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

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

Answer:

• Previously in Slide 7-13, we obtain

$$B = [\mathbf{u}_1 \cdots \mathbf{u}_n]^{-1} [\mathbf{v}_1 \cdots \mathbf{v}_n] = U^{-1}V.$$
7.3 Linear transform for basis changing (revisited)

Find a linear transformation $T$ or a mapping from the space $\text{span}(v_1, v_2, \ldots, v_n)$ to the space $\text{span}(u_1, u_2, \ldots, u_n)$.

- An alternative choice of $B$ is $B = U^T \left( U \Sigma V^T \right) V = U^T A V$ for some $\Sigma$ with non-zero diagonals.

$$A v_i = \sigma_i u_i \implies B s = U^T A V s$$

$$= U^T A (s_1 v_1 + \cdots + s_n v_n)$$

$$= U^T \left( \sigma_1 s_1 u_1 + \cdots + \sigma_n s_n u_n \right)$$

$$= U^T U t = t.$$
7.3 Linear transform for basis changing (revisited)

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

\[
A_{m \times n}(s_1 \mathbf{v}_1 + \cdots + s_n \mathbf{v}_n) = \sigma_1 s_1 \mathbf{u}_1 + \cdots + \sigma_n s_n \mathbf{u}_n
\]

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

- If the rank of \( A \) is \( r \), then \( t_{r+1} = \cdots = t_m = 0 \) (because \( \sigma_{r+1} = \cdots = \sigma_{m+1} = 0 \)). As a consequence, \( A \) maps a vector in the vector space \( \mathcal{C}(V) = \mathcal{C}([\mathbf{v}_1 \cdots \mathbf{v}_n]) \) to a vector in the subspace \( \mathcal{C}([\mathbf{u}_1 \cdots \mathbf{u}_r]) \).
It is natural to infer that:

1. Some matrices only have \textit{left inverse} but have no \textit{right inverse}.
2. Some matrices only have \textit{right inverse} but have no \textit{left inverse}.
3. Some matrices have neither \textit{left inverse} nor \textit{right inverse}.
4. Some matrices have both \textit{left inverse} and \textit{right inverse}.

In such case, the inverse exists, and is equal to the \textit{left inverse} and also the \textit{right inverse}.

**Question:** When do each of the above cases happen?

1. If $A$ has full column rank ($r = n$) but has no full row rank ($r < m$).
2. If $A$ has full row rank ($r = m$) but has no full column rank ($r < n$).
3. If $A$ has no full column and no full 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 $A_{m \times n} = U_{m \times r} \Sigma_{r \times r} V_{r \times n}^T$. Hence,

$$\begin{align*}
A^T A &= V_{n \times r} \Sigma_{r \times r} U_{r \times m}^T U_{m \times r} \Sigma_{r \times r} V_{r \times n}^T = V_{n \times r} \Sigma_r^2 V_{r \times n}^T \\
A A^T &= U_{m \times r} \Sigma_{r \times r} V_{r \times n}^T V_{n \times r} \Sigma_{r \times r} U_{r \times m}^T = U_{m \times r} \Sigma_r^2 U_{r \times m}^T
\end{align*}$$

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

   In such case, $A^+ = \text{left inverse} = (A^T A)^{-1} A^T$
   $$= V_{n \times r} \Sigma_r^{-2} V_{r \times n}^T V_{n \times r} \Sigma_r \Sigma_r V_{r \times n}^T = V_{n \times r} \Sigma_r^{-1} U_{r \times m}.$$  

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

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

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 $(A^T A)^{-1} A^T = A^T (A A^T)^{-1}$.

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

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

\[ A_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2\sqrt{2} \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ \end{bmatrix} \]

\[ A_2 = \begin{bmatrix} 2 & 2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 2\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \]

\[ A_3 = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \]
7.3 Pseudoinverse, left-inverse, right-inverse

Solution.

\[
A_1^+ = \begin{bmatrix} 1 \\ \Sigma^+ \end{bmatrix} \begin{bmatrix} 1/(2\sqrt{2}) & 0 \\ \sqrt{2} & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1/4 & 1/4 \end{bmatrix} \implies A_1^+A_1 = [1]
\]

\[
A_2^+ = \begin{bmatrix} 1 \\ \Sigma^+ \end{bmatrix} \begin{bmatrix} 1/(2\sqrt{2}) \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1/4 \\ 1/4 \end{bmatrix} \implies A_2A_2^+ = [1]
\]

\[
A_3^+ = \begin{bmatrix} 1 \\ \Sigma^+ \end{bmatrix} \begin{bmatrix} 1/4 & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1/8 & 1/8 \\ 1/8 & 1/8 \end{bmatrix} \implies A_3^+A_3 = A_3A_3^+ = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}
\]