$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}_{m\times n}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}}_{x} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}_{m\times 1}}_{b}$$

CX1104: **Linear Algebra for Computing**

Chap. No : **7.1.1**

Lecture : **Least Squares**

Topic : **Introduction**

Concept : **Consistency in a System of Equations**

Instructor: **A/P Chng Eng Siong**

TAs: **Zhang Su, Vishal Choudhari**

Rev: 30th June 2020

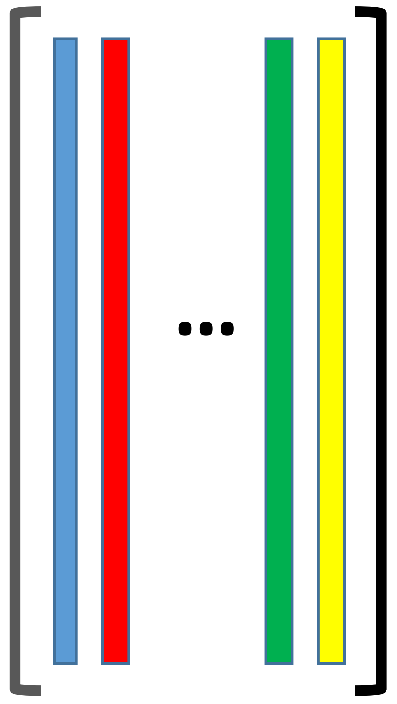
Consistency in a System of Equations

Consider solving the system of equations: $Ax = b$

Note:

- Matrix $A \in R^{M \times N}$, where
  - M denotes no. of rows/equations
  - N denotes no. of columns/unknowns
- $x \in R^N$
- $b \in R^M$
- The above system of equations can either be
  1. consistent (or)
  2. inconsistent.

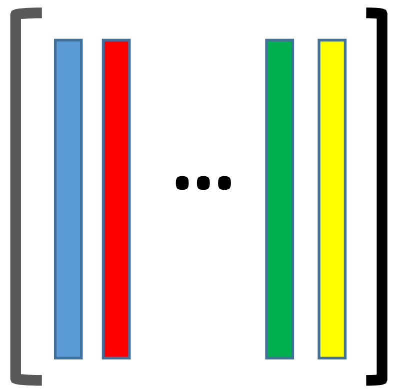
Based on M & N, there exist three cases:



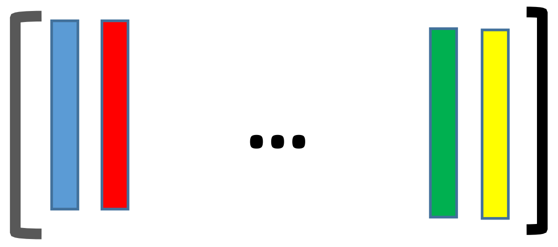
$M \gg N$

More equations, less unknowns.

Hence, **over-determined**! **Typically** this will result in inconsistent system of equations



$M \approx N$



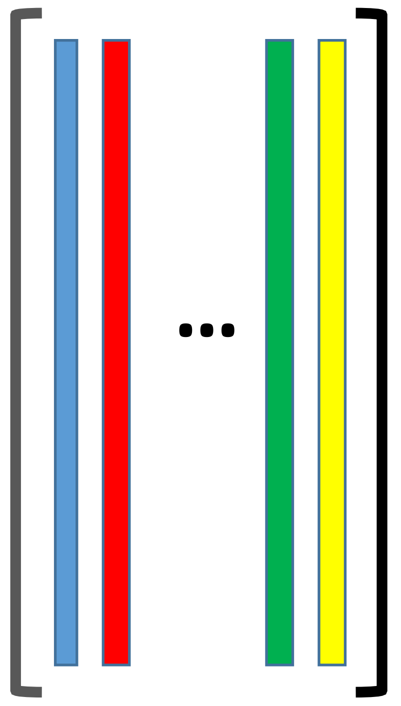
$M \ll N$

Less equations, more unknowns.

Hence, **under-determined**! **Typically,** this will result in infinite solutions

Consistency in a System of Equations

$$Ax = b$$



$M \gg N$

Consistent (solutions exist)

- b is in column space of A, i.e,
  b is formed by linear combinations of A's columns.
- Rank (A) = Rank ($A|b$), i.e,
  rank of A is same as that of the augmented matrix.

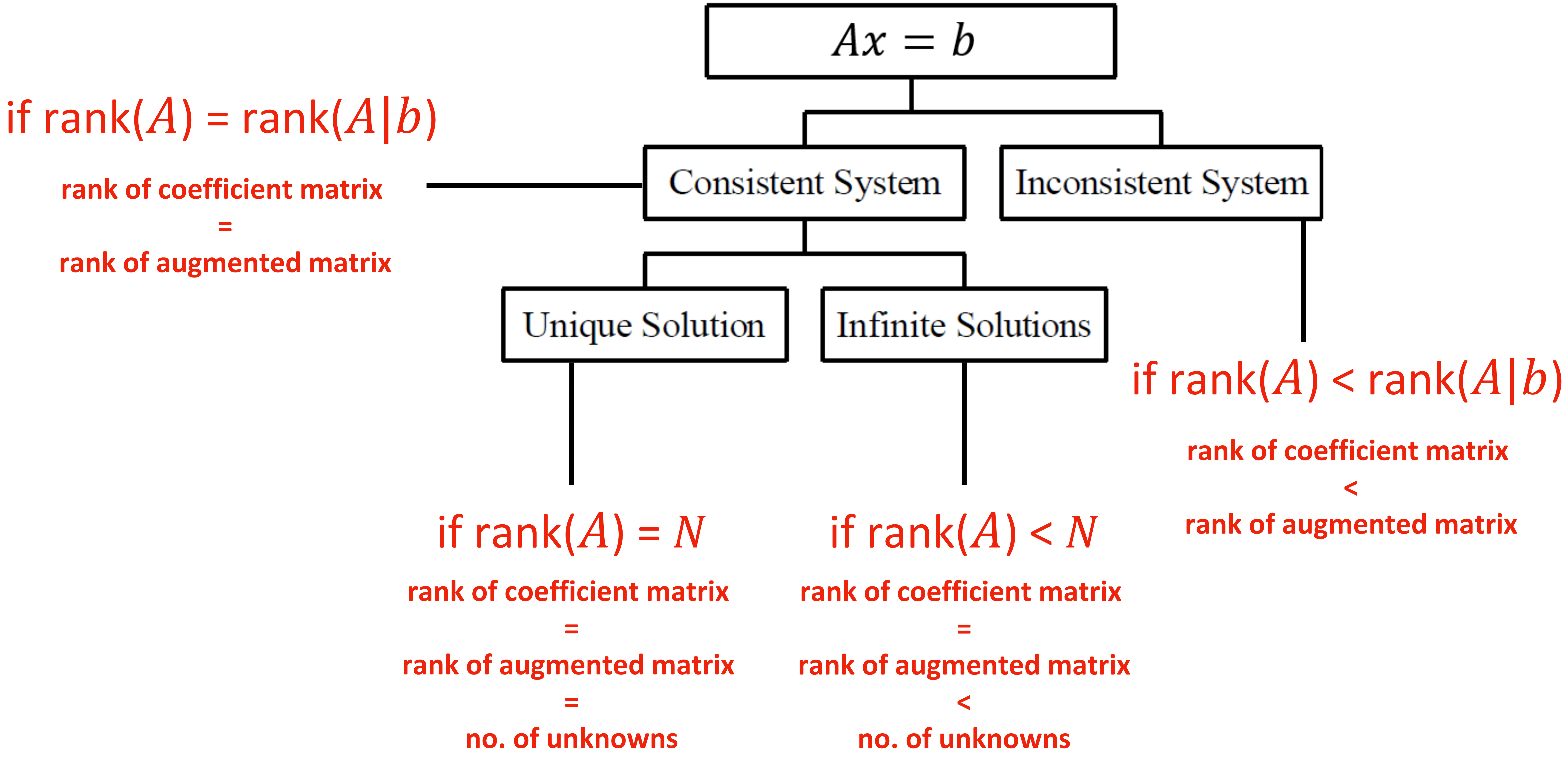
Inconsistent (solutions don't exist)

- b is NOT in column space of A, i.e,
  b is NOT formed by linear combinations of A's columns.
- **Typically** occurs when $M \gg N$ (**over-determined**), i.e,
  there exist more equations than unknowns.
- The rows of A are dependent but,
  their corresponding b values are not consistent.
- Rank (A) < Rank ($A|b$), i.e,
  rank of A is less than that of the augmented matrix.

Consistency in a System of Equations

A system of equations can be consistent or inconsistent. What does that mean?

A system of equations $Ax = b$ is consistent if there is a solution, and it is inconsistent if there is no solution. However, consistent system of equations does not mean a unique solution, that is, a consistent system of equation may have a unique solution or infinite solutions.



NOTE: Rank (A) is the maximum number of independent rows or columns of A.

You can find number of independent row or columns by:
1. row reduction process
2. rank(A) in MATLAB

Ref: https://en.wikipedia.org/wiki/Augmented_matrix
Ref: https://www.mathsisfun.com/algebra/matrix-rank.html

Note: rank (A) > rank ($A|b$) is never possible. Why?

Examples

$\text{rank}(A) = \text{rank}(A|b) = N$

Consistent and Unique Solution

a) The system of equations

$$\begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

is a consistent system of equations as it has a unique solution, that is,

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

```
>> A_b = [ 2 4 6; 1 3 4]

A_b =

     2     4     6
     1     3     4

>> rank(A_b)

ans =

     2
```

Inconsistent and No solutions Exist

c) The system of equations

$$\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

$\text{rank}(A) < \text{rank}(A|b)$

```
>> A = [2 4; 1 2]

A =

     2     4
     1     2

>> rank(A)

ans =

     1
```

```
>> A_b = [ 2 4 6; 1 2 4]

A_b =

     2     4     6
     1     2     4

>> rank(A_b)

ans =

     2
```

Consistent and Having Infinite Solutions

b) The system of equations

$$\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$$

```
>> A_b = [ 2 4 6; 1 2 3]

A_b =

     2     4     6
     1     2     3

>> rank(A_b)

ans =

     1
```

is also a consistent system of equations but it has infinite solutions as given as follows.

Expanding the above set of equations,

$$2x + 4y = 6$$
$$x + 2y = 3$$

you can see that they are the same equation. Hence any combination of (x, y) that satisfies

$$2x + 4y = 6$$

is a solution. For example $(x, y) = (1,1)$ is a solution and other solutions include $(x, y) = (0.5,1.25)$, $(x, y) = (0,\ 1.5)$ and so on.

$\text{rank}(A) = \text{rank}(A|b) < N$

Ref: Introduction to Matrix Algebra, Autar Kaw (pg 68)

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}_{m\times n}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}}_{x} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}_{m\times 1}}_{b}$$

CX1104: **Linear Algebra for Computing**

Chap. No : **7.1.2**

Lecture : **Least Squares**

Topic : **Introduction**

Concept : **The Least Squares Problem**

Instructor: **A/P Chng Eng Siong**

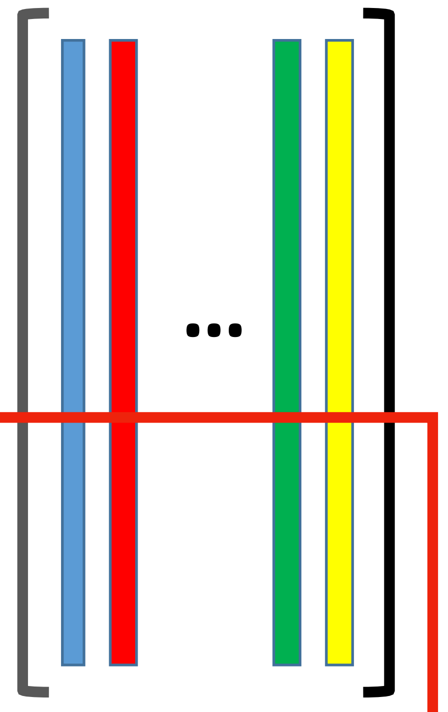
TAs: **Zhang Su, Vishal Choudhari**

Rev: 2nd July 2020

Consistency in a System of Equations

$$Ax = b$$

Consider



$M \gg N$

Consistent (solutions exist)

- b is in column space of A, i.e,
  b is formed by linear combinations of A's columns.
- Rank (A) = Rank $(A|b)$, i.e,
  rank of A is same as that of the augmented matrix.

Inconsistent (solutions don't exist)

- b is NOT in column space of A, i.e,
  b is NOT formed by linear combinations of A's columns.
- **Typically** occurs when $M \gg N$ (**over-determined**), i.e,
  there exist more equations than unknowns.
- The rows of A are dependent but,
  their corresponding b values are not consistent.
- Rank (A) < Rank $(A|b)$, i.e,
  rank of A is less than that of the augmented matrix.

Least Squares Solution for Inconsistent Equations

Consider solving the system of equations: $Ax = b$

Note:

- Matrix $A \in R^{M \times N}$, where
  - M denotes no. of rows/equations
  - N denotes no. of columns/unknowns
- $x \in R^N$
- $b \in R^M$
- When $M \gg N$,
  - the system is over-determined
  - the equations may be inconsistent
  - there may be no solution

Best we can do?

Find x such that Ax is as close to b as possible!

If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

for all $\mathbf{x}$ in $\mathbb{R}^n$.

Think of $A\mathbf{x}$ as an *approximation* to $\mathbf{b}$. The smaller the distance between $\mathbf{b}$ and $A\mathbf{x}$, given by $\|\mathbf{b} - A\mathbf{x}\|$, the better the approximation. The **general least-squares problem** is to find an $\mathbf{x}$ that makes $\|\mathbf{b} - A\mathbf{x}\|$ as small as possible. The adjective "least-squares" arises from the fact that $\|\mathbf{b} - A\mathbf{x}\|$ is the square root of a sum of squares.

Lay, Linear Algebra and its Applications (4th Edition)

360 CHAPTER 6 Orthogonality and Least Squares

Definitions

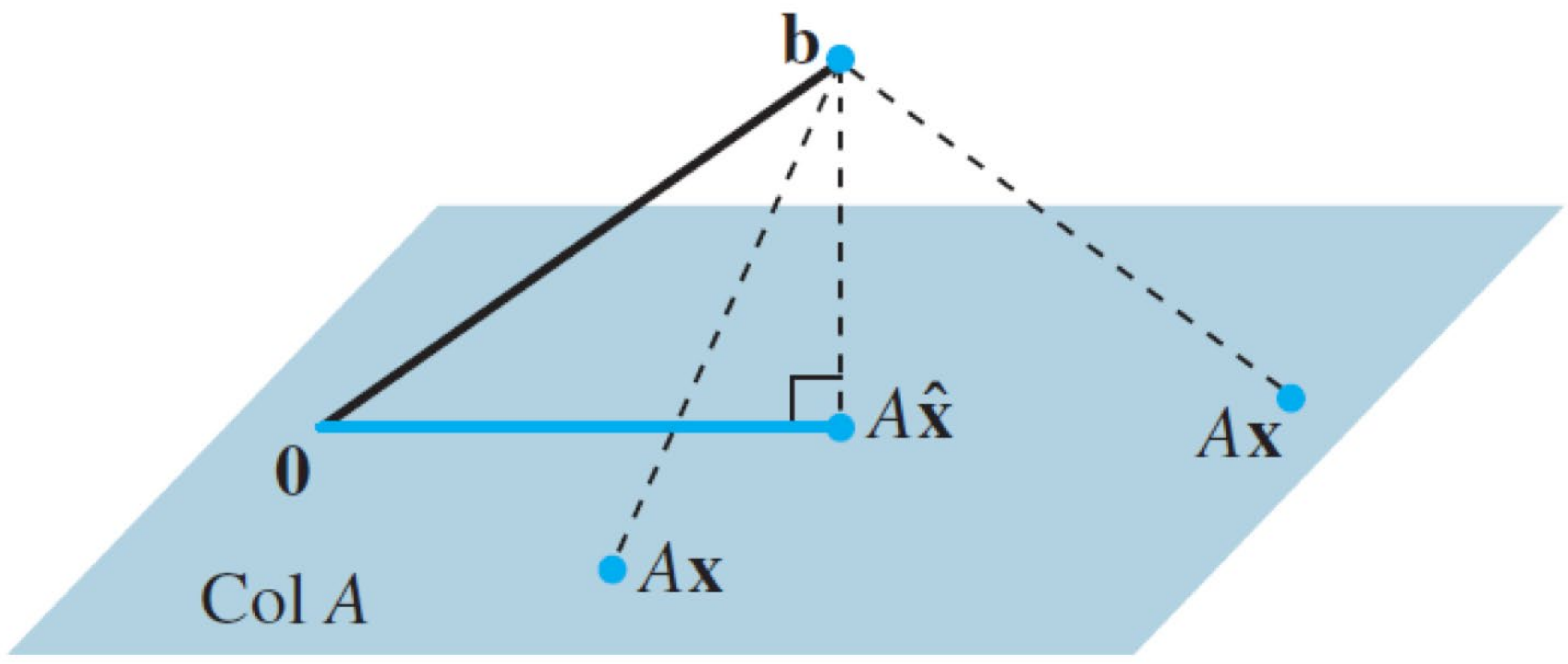


FIGURE 1 The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.

The most important aspect of the least-squares problem is that no matter what $\mathbf{x}$ we select, the vector $A\mathbf{x}$ will necessarily be in the column space, Col A. So we seek an $\mathbf{x}$ that makes $A\mathbf{x}$ the closest point in Col A to $\mathbf{b}$. See Fig. 1. (Of course, if $\mathbf{b}$ happens to be in Col A, then $\mathbf{b}$ *is* $A\mathbf{x}$ for some $\mathbf{x}$, and such an $\mathbf{x}$ is a "least-squares solution.")

If a linear system is consistent, then its exact solutions are the same as its least squares solutions, in which case the least squares error is zero.

NOTE:
When the linear system $Ax = b$ is inconsistent, b does not lie in the column space of A.

Lay, Linear Algebra and its Applications (4th Edition)

360 CHAPTER 6 Orthogonality and Least Squares

To explain the terminology in this problem, suppose that the column form of $\mathbf{b} - A\mathbf{x}$ is

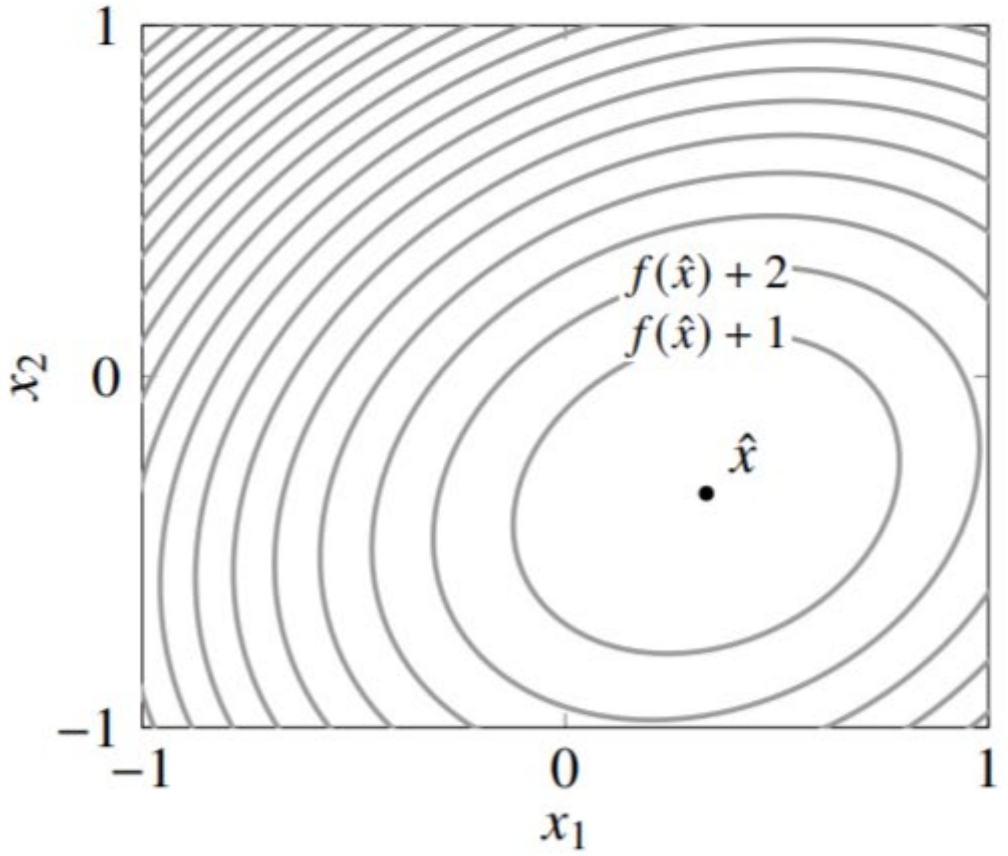
$$\mathbf{b} - A\mathbf{x} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}$$

The term "least squares solution" results from the fact that minimizing $\|\mathbf{b} - A\mathbf{x}\|$ also has the effect of minimizing $\|\mathbf{b} - A\mathbf{x}\|^2 = e_1^2 + e_2^2 + \cdots + e_m^2$.

Example

Example

$$A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$



- the least squares solution $\hat{x}$ minimizes

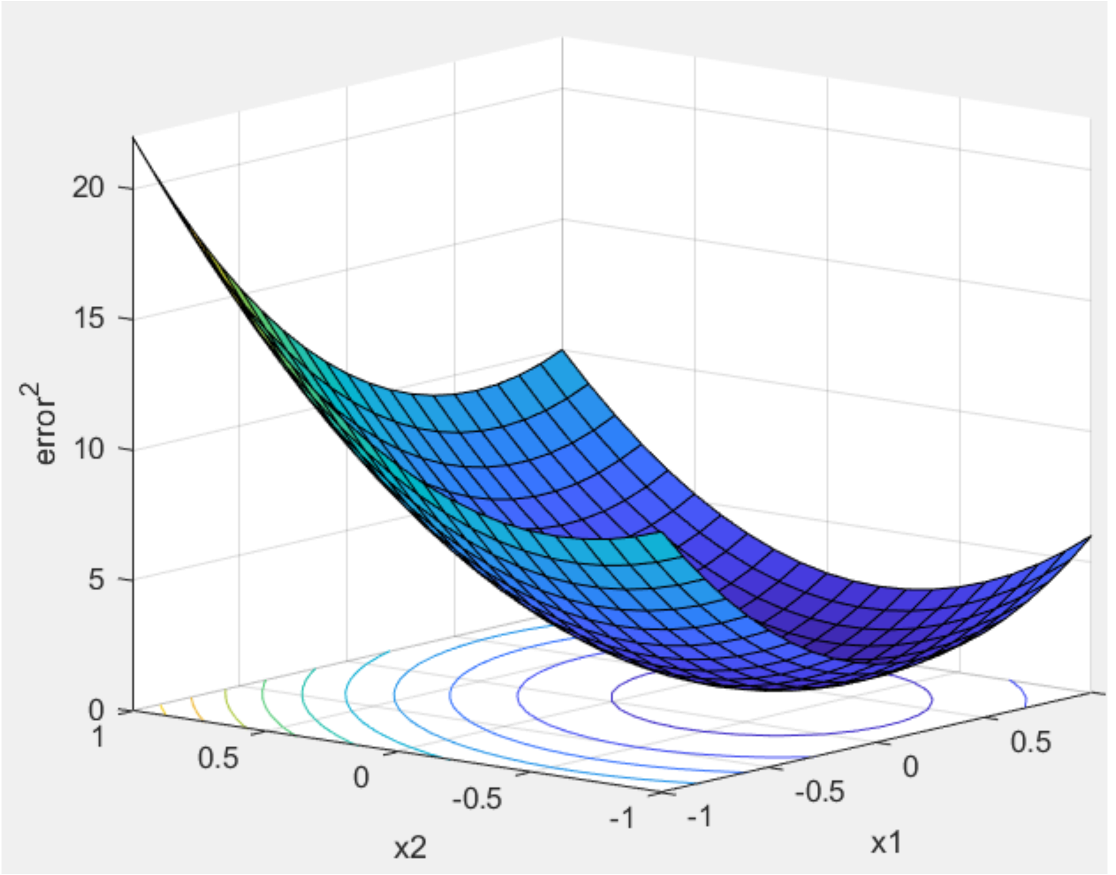
$$f(x) = \|Ax - b\|^2 = (2x_1 - 1)^2 + (-x_1 + x_2)^2 + (2x_2 + 1)^2$$

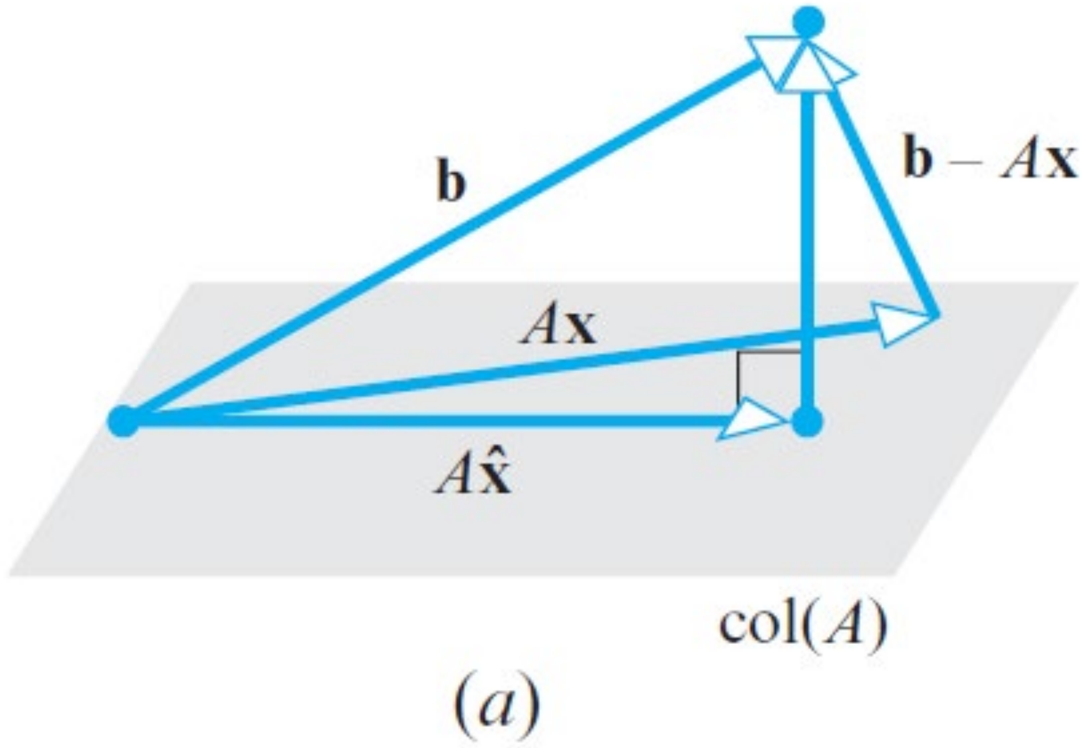
- to find $\hat{x}$, set derivatives with respect to x_1 and x_2 equal to zero:

$$10x_1 - 2x_2 - 4 = 0, \qquad -2x_1 + 10x_2 + 4 = 0$$

solution is $(\hat{x}_1, \hat{x}_2) = (1/3, -1/3)$

Least squares 8.3






```
opStr =

    'x1=0.30,   x2=-0.30,   err^2=0.680 '
```

```
%ch6_4_Ex1.m
%Chng Eng Siong, plotting the error wrt x
close all; clear all;
A = [2 0; -1 1; 0 2];
b = [1 0 -1]';
[x1,x2] = meshgrid(-1:0.1:1, -1:0.1:1);
[m,n] = size(x1);
z = zeros(m,n);
for i=1:m
    for j=1:n
        z(i,j) = norm(b - (x1(i,j)*A(:,1)+x2(i,j)*A(:,2))).^2;
    end
end
surfc(x1,x2,z)
xlabel('x1'); ylabel('x2'); zlabel('error^2');

% Lets print the min value and the x vector
minIdx = find(z == min(z(:)));
x1(minIdx), x2(minIdx), z(minIdx)
opStr = sprintf('x1=%0.2f,  x2=%0.2f,  err^2=%.3f ',x1(minIdx),x2(minIdx),z(minIdx))
```

Boyd – Lecture 6

Ref: https://www.youtube.com/watch?v=ZKhZclqfR5E&t=1850s

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}_{m\times n}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}}_{x} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}_{m\times 1}}_{b}$$

CX1104: **Linear Algebra for Computing**

Chap. No : **7.1.3**

Lecture : **Least Squares**

Topic : **Solving the Least Squares Problem**

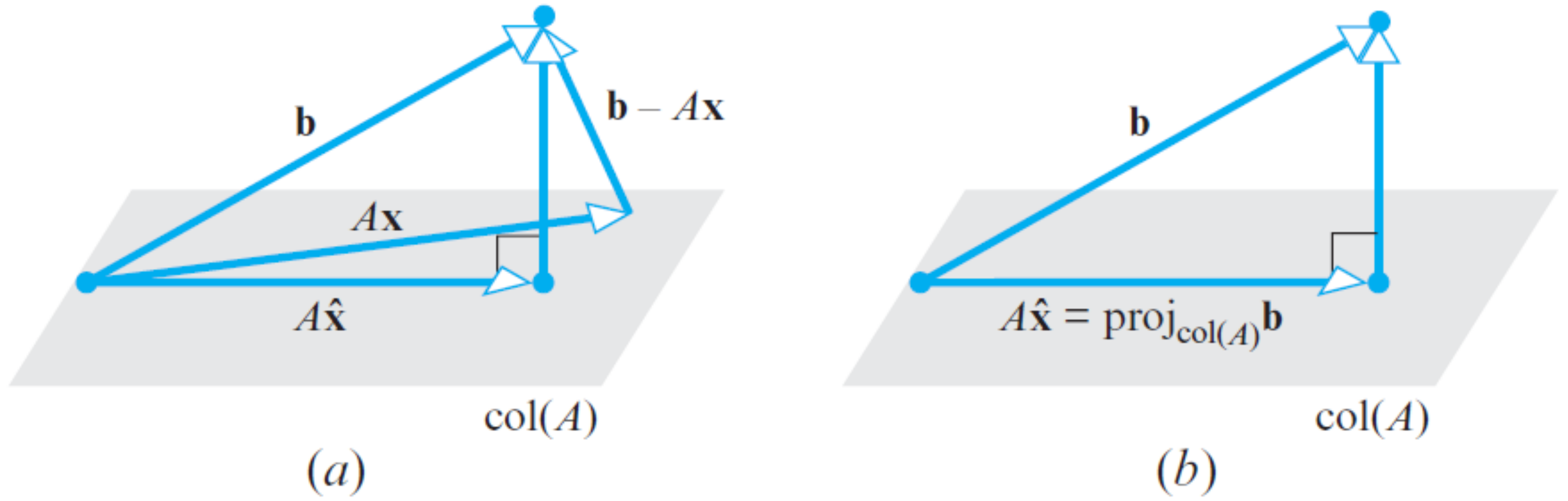
Concept : **Best Approx. Theorem and Normal Equation**

Instructor: **A/P Chng Eng Siong**

TAs: **Zhang Su, Vishal Choudhari**

Rev: 2nd July 2020

Best Approximation Theorem



THEOREM 6.4.1 Best Approximation Theorem

If W is a finite-dimensional subspace of an inner product space V, and if $\mathbf{b}$ is a vector in V, then $\text{proj}_W \mathbf{b}$ is the ***best approximation*** *to $\mathbf{b}$ from W in the sense that*

$$\|\mathbf{b} - \text{proj}_W \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\|$$

for every vector $\mathbf{w}$ in W that is different from $\text{proj}_W \mathbf{b}$.

Anton:

6.4 Best Approximation; Least Squares 379

Proof For every vector $\mathbf{w}$ in W, we can write

$$\mathbf{b} - \mathbf{w} = (\mathbf{b} - \text{proj}_W \mathbf{b}) + (\text{proj}_W \mathbf{b} - \mathbf{w})$$

But $\text{proj}_W \mathbf{b} - \mathbf{w}$, being a difference of vectors in W, is itself in W; and since $\mathbf{b} - \text{proj}_W \mathbf{b}$ is orthogonal to W, the two terms on the right side of (1) are orthogonal. Thus, it follows from the Theorem of Pythagoras (Theorem 6.2.3) that

$$\|\mathbf{b} - \mathbf{w}\|^2 = \|\mathbf{b} - \text{proj}_W \mathbf{b}\|^2 + \|\text{proj}_W \mathbf{b} - \mathbf{w}\|^2$$

If $\mathbf{w} \neq \text{proj}_W \mathbf{b}$, it follows that the second term in this sum is positive, and hence that

$$\|\mathbf{b} - \text{proj}_W \mathbf{b}\|^2 < \|\mathbf{b} - \mathbf{w}\|^2$$

Since norms are nonnegative, it follows (from a property of inequalities) that

$$\|\mathbf{b} - \text{proj}_W \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\| \quad ◀$$

The Normal Equation

$$Ax = b$$

Multiplying both sides by A^T

$$A^TA\mathbf{x} = A^T\mathbf{b}$$

Normal Equation!

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$.

Proved in the next slide!

Solution of the General Least-Squares Problem

Given A and $\mathbf{b}$ as above, apply the Best Approximation Theorem in Section 6.3 to the subspace Col A. Let

$$\hat{\mathbf{b}} = \text{proj}_{\text{Col}\,A}\,\mathbf{b}$$

Because $\hat{\mathbf{b}}$ is in the column space of A, the equation $A\mathbf{x} = \hat{\mathbf{b}}$ *is* consistent, and there is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}} \tag{1}$$

Since $\hat{\mathbf{b}}$ is the closest point in Col A to $\mathbf{b}$, a vector $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $\hat{\mathbf{x}}$ satisfies (1). Such an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ is a list of weights that will build $\hat{\mathbf{b}}$ out of the columns of A. See Fig. 2. [There are many solutions of (1) if the equation has free variables.]

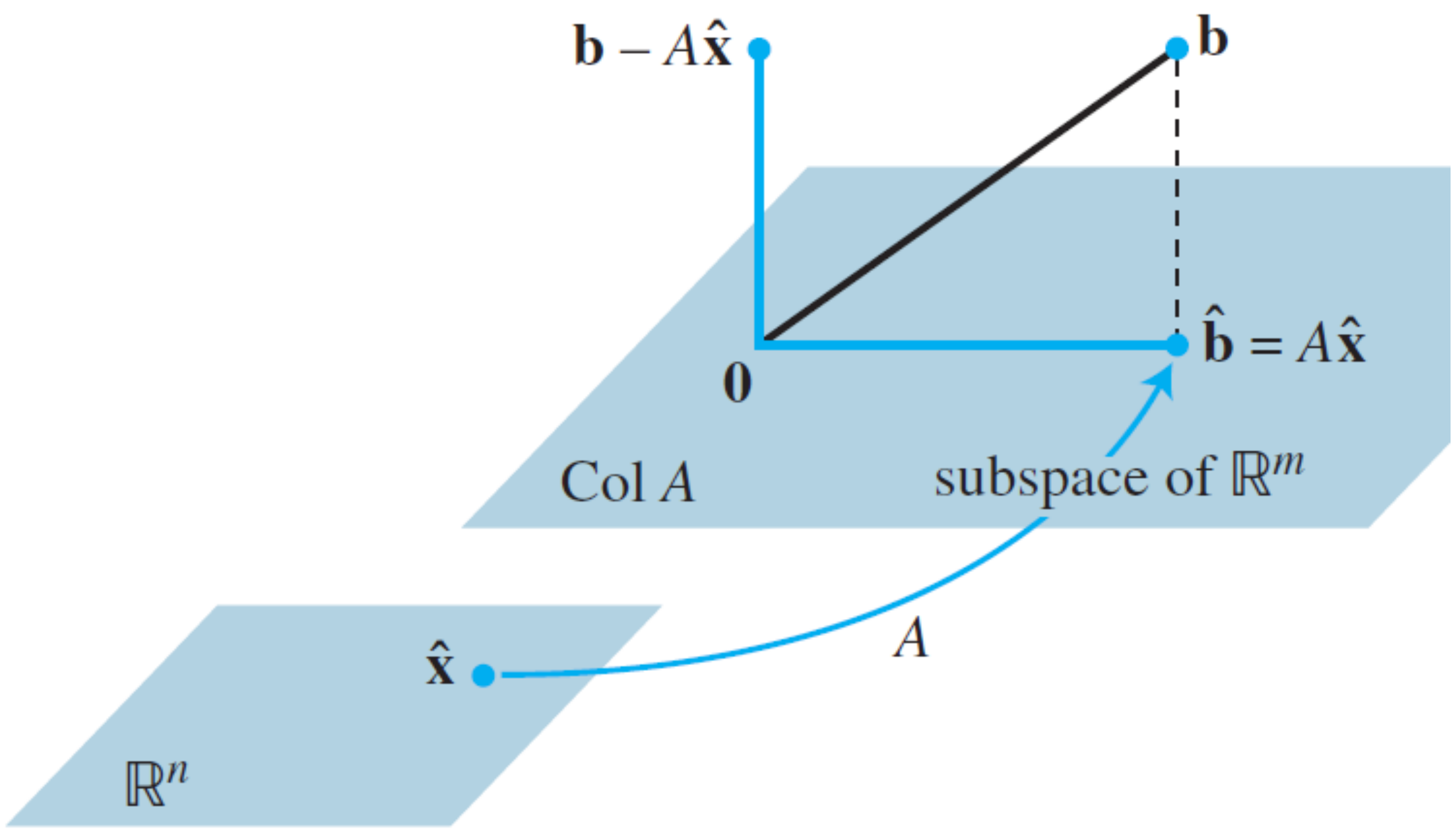


FIGURE 2 The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

Lay, Linear Algebra and its Applications (4th Edition)

6.5 Least-Squares Problems **361**

Why called "Normal"?

Ref: https://mathworld.wolfram.com/NormalEquation.html

The Normal Equation Proof

Suppose $\hat{\mathbf{x}}$ satisfies $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. By the Orthogonal Decomposition Theorem in Section 6.3, the projection $\hat{\mathbf{b}}$ has the property that $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to Col A, so $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to each column of A. If $\mathbf{a}_j$ is any column of A, then $\mathbf{a}_j \cdot (\mathbf{b} - A\hat{\mathbf{x}}) = 0$, and $\mathbf{a}_j^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$. Since each $\mathbf{a}_j^T$ is a row of A^T,

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0} \tag{2}$$

(This equation also follows from Theorem 3 in Section 6.1.) Thus

$$\begin{aligned} A^T\mathbf{b} - A^TA\hat{\mathbf{x}} &= \mathbf{0} \\ A^TA\hat{\mathbf{x}} &= A^T\mathbf{b} \end{aligned}$$

These calculations show that each least-squares solution of $A\mathbf{x} = \mathbf{b}$ satisfies the equation

$$A^TA\mathbf{x} = A^T\mathbf{b} \tag{3}$$

The matrix equation (3) represents a system of equations called the **normal equations** for $A\mathbf{x} = \mathbf{b}$. A solution of (3) is often denoted by $\hat{\mathbf{x}}$.

Lay, Linear Algebra and its Applications (4th Edition)

6.5 Least-Squares Problems **361**

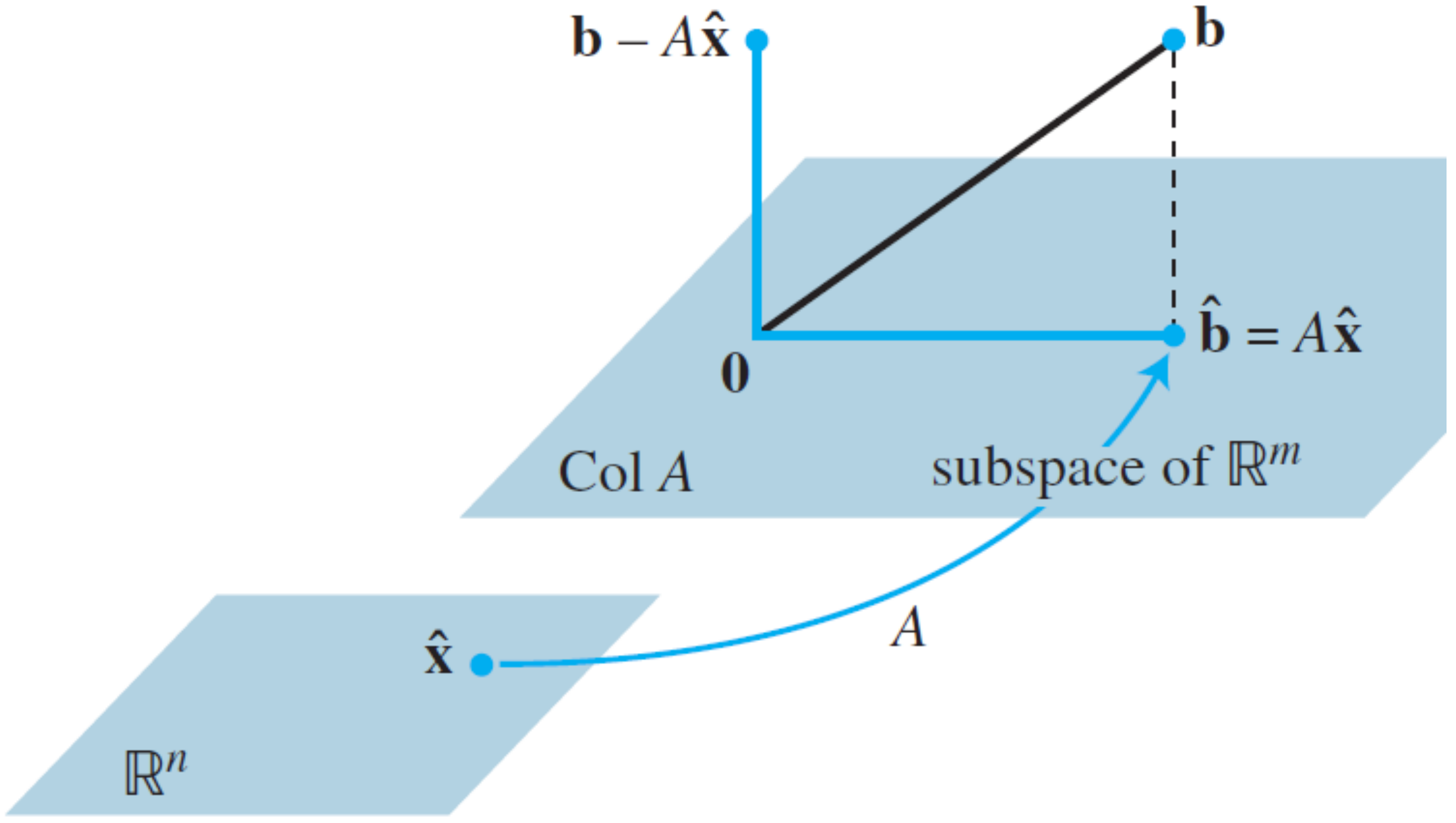


FIGURE 2 The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

THEOREM 14

Let A be an $m \times n$ matrix. The following statements are logically equivalent:

a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

b. The columns of A are linearly indpendent.

c. The matrix A^TA is invertible.

When these statements are true, the least-squares solution $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} \tag{4}$$

Examples

EXAMPLE 1 Find a least-squares solution of the inconsistent system $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

SOLUTION To use normal equations (3), compute:

$$A^TA = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

Then the equation $A^TA\mathbf{x} = A^T\mathbf{b}$ becomes

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

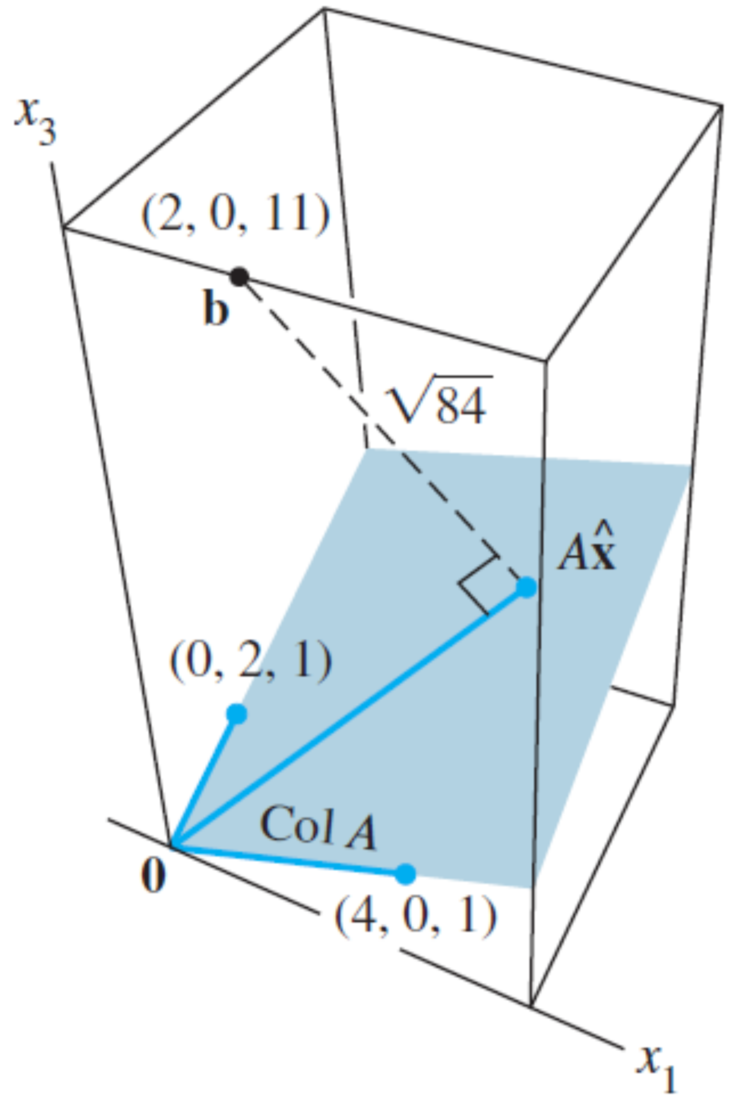


FIGURE 3

Lay, Linear Algebra and its Applications (4th Edition)

Row operations can be used to solve this system, but since A^TA is invertible and 2×2, it is probably faster to compute

$$(A^TA)^{-1} = \frac{1}{84}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$$

and then to solve $A^TA\mathbf{x} = A^T\mathbf{b}$ as

$$\begin{aligned} \hat{\mathbf{x}} &= (A^TA)^{-1}A^T\mathbf{b} \\ &= \frac{1}{84}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}\begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84}\begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \end{aligned}$$

In many calculations, A^TA is invertible, but this is not always the case. The next

Examples

EXAMPLE 3 Given A and $\mathbf{b}$ as in Example 1, determine the least-squares error in the least-squares solution of $A\mathbf{x} = \mathbf{b}$.

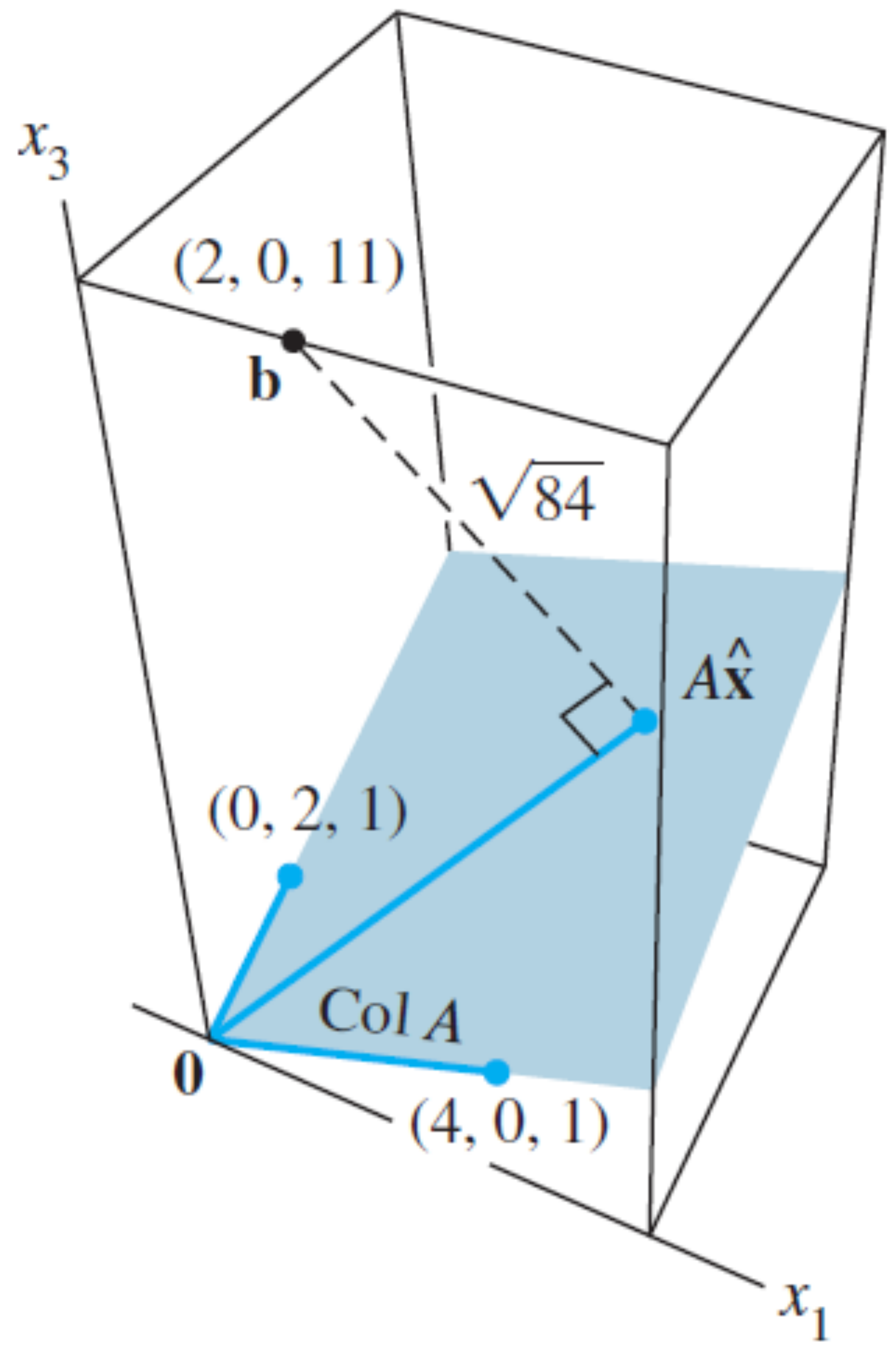


FIGURE 3

SOLUTION From Example 1,

$$\mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} \quad \text{and} \quad A\hat{\mathbf{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$$

Hence

$$\mathbf{b} - A\hat{\mathbf{x}} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -4 \\ 8 \end{bmatrix}$$

and

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| = \sqrt{(-2)^2 + (-4)^2 + 8^2} = \sqrt{84}$$

The least-squares error is $\sqrt{84}$. For any $\mathbf{x}$ in $\mathbb{R}^2$, the distance between $\mathbf{b}$ and the vector $A\mathbf{x}$ is at least $\sqrt{84}$. See Fig. 3. Note that the least-squares solution $\hat{\mathbf{x}}$ itself does not appear in the figure. ■

Lay, Linear Algebra and its Applications (4th Edition)

Examples

EXAMPLE 2 Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}$$

Note the linear dependency in the rows and columns of A:

- Column 1 = Column 2 + Column 3 + Column 4
- Rows 1 & 2 are same, but their corresponding b values are different (inconsistent)
- Rows 3 & 4 are same, but their corresponding b values are different (inconsistent)
- Rows 5 & 6 are same, but their corresponding b values are different (inconsistent)

SOLUTION Compute

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}}_{A^TA}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}$$

Note that A^TA is always a square matrix.

Lay, Linear Algebra and its Applications (4th Edition)

The augmented matrix for $A^TA\mathbf{x} = A^T\mathbf{b}$ is

Reduced to

$$\begin{bmatrix} \underbrace{\begin{matrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{matrix}}_{A^TA} & \underbrace{\begin{matrix} 4 \\ -4 \\ 2 \\ 6 \end{matrix}}_{A^Tb} \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The general solution is $x_1 = 3 - x_4$, $x_2 = -5 + x_4$, $x_3 = -2 + x_4$, and x_4 is free. So the general least-squares solution of $A\mathbf{x} = \mathbf{b}$ has the form

$$\hat{\mathbf{x}} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$ ■

Note: Here, there are infinitely many solutions with the same least square error.

Note: Here, A^TA is not invertible (its determinant is 0).

A^TA may not be invertible if:

- some columns are linearly dependent (i.e. we have redundant features) (**as in this example**)
  - solution: remove the linear dependency
- too many features ($m < n$)
  - solution: delete some features, there are too many features for the amount of data we have

Ref: http://mlwiki.org/index.php/Normal_Equation

Ref: Andrew Ng discussing this phenomenon-
https://www.coursera.org/lecture/machine-learning/normal-equation-noninvertibility-zSiE6

Examples

```
% pg 362 Lay's book, Example 2 - Least Squares, when A'A is singular
close all; clear all;
A = [1 1 0 0; 1 1 0 0; 1 0 1 0; 1 0 1 0; 1 0 0 1; 1 0 0 1];
b = [-3 -1 0 2 5 1]';

AtA = A'*A
rank_Ata = rank(AtA)   % A'*A is singular, we check its rank

% Ax = b;
x1 = pinv(A)*b
x2 = inv(A'*A)*A'*b   % This is what we think we should do
% compare inv(A'*A) bs pinv(A'*A)
disp("using normal inverser (A'*A):");
inv(A'*A)
disp("using pinverser (A'*A):");
pinv(A'*A)

x3 = pinv(A'*A)*A'*b   % This is what Andy Ng suggest to d
```

```
AtA =

     6     2     2     2
     2     2     0     0
     2     0     2     0
     2     0     0     2
```

```
rank_Ata =

     3
```

```
x1 =

    0.5000
   -2.5000
    0.5000
    2.5000
```

```
x2 =

     1
    -3
     0
     2
```

$A^T A$

```
Warning: Matrix is close to singular or badly scaled.
> In Lay_example2_pg362 (line 8)
```

```
ans =

   1.0e+15 *

    1.5012   -1.5012   -1.5012   -1.5012
   -1.5012    1.5012    1.5012    1.5012
   -1.5012    1.5012    1.5012    1.5012
   -1.5012    1.5012    1.5012    1.5012
```

$A^T A$ is non-invertible. Hence MATLAB computes its inverse as a very large value $\Longrightarrow \infty$

```
ans =

    0.0938    0.0312    0.0313    0.0313
    0.0313    0.3437   -0.1562   -0.1563
    0.0312   -0.1562    0.3438   -0.1562
    0.0313   -0.1562   -0.1563    0.3438
```

```
x3 =

    0.5000
   -2.5000
    0.5000
    2.5000
```

NOTE: Pseudo-inverse (pinv) will be introduced later.

Reference

Some Useful readings:

1) Fogel: Learning Goals: find the best solution (by one measure, anyway) of inconsistent equation. Learn to apply the algebra, geometry, and calculus of projections to this problem.
   http://staff.imsa.edu/~fogel/LinAlg/PDF/33%20Least%20Squares.pdf

2) Why normal equation always have a solution: unique or infinite even if A has dependent column
   a) https://math.stackexchange.com/questions/2920398/how-do-the-normal-equations-always-have-a-solution
   b) https://math.stackexchange.com/questions/72222/existence-of-least-squares-solution-to-ax-b
   c) https://stats.stackexchange.com/questions/63143/question-about-a-normal-equation-proof

3) Prof Walker, Wocester Polytechnic Institute: https://users.wpi.edu/~walker/MA3257/HANDOUTS/least-squares_handout.pdf

Reference

Some Useful readings:

1) Fogel: Learning Goals: find the best solution (by one measure, anyway) of inconsistent equation. Learn to apply the algebra, geometry, and calculus of projections to this problem.
http://staff.imsa.edu/~fogel/LinAlg/PDF/33%20Least%20Squares.pdf

2) Why normal equation always have a solution: unique or infinite even if A has dependent column
a) https://math.stackexchange.com/questions/2920398/how-do-the-normal-equations-always-have-a-solution
b) https://math.stackexchange.com/questions/72222/existence-of-least-squares-solution-to-ax-b
c) https://stats.stackexchange.com/questions/63143/question-about-a-normal-equation-proof

3) Prof Walker, Wocester Polytechnic Institute: https://users.wpi.edu/~walker/MA3257/HANDOUTS/least-squares_handout.pdf