University of Chicago Professional Education

MSCA 37016

Advanced Linear Algebra for Machine Learning

Session 3

Shaddy Abado Ph.D.

Agenda: Session #3

- Projections
- Least Squares Approximation
- QR Decomposition
- Linear transformations
- Visualization of linear transformations

BASIC CONCEPTS NEEDED FOR THIS SESSION

A_{mxn}

Definition: Column Space - $C(A)$

Given a matrix A with columns in $\mathcal{R}^m$, these columns and all their linear combinations <u>form a subspace</u> of $\mathcal{R}^m$.

$$C(A) = span\,\{v_1, \cdots, v_n\}$$

For example,

For $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 3 \\ 0 & 4 \end{bmatrix}$

The **column space** of $\boldsymbol{A}$ is the <u>plane</u> through the origin in $\mathcal{R}^3$ containing $\begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$

$\boldsymbol{A}_{mxn} x_{nx1} = b_{mx1}$

Column Space - $C(\boldsymbol{A})$

Question:

Given a matrix $\boldsymbol{A}$, for what vectors $\boldsymbol{b}$ does $\boldsymbol{Ax} = \boldsymbol{b}$ have a solution $\boldsymbol{x}$?

Answer:

The system $\boldsymbol{A}x = \boldsymbol{b}$ is solvable <u>if and only if</u> $\boldsymbol{b}$ is in the column space of A (i.e., $C(\boldsymbol{A})$)

Core Idea:

The Column Space $C(\boldsymbol{A})$ describes all the attainable $\vec{b}$'s

$A_{mxn}x_{nx1} = 0_{mx1}$

Definition: Nullspace

The *nullspace* $N(\boldsymbol{A})$ of a matrix $\boldsymbol{A}$ is the collection of all solutions $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ to the equation $\boldsymbol{A}x = \vec{0}$.

Definition - Special solutions:

The nullspace of $\boldsymbol{A}$ consists of all the combinations of the special solutions to $\boldsymbol{A}x = \vec{0}$.

Again:

Only singular matrices have a nullspace that contains more than just the zero vector.

Definition: Basis

Motivation:

Find <u>enough</u> Independent vectors to span the space (and not more).

A *basis* for a vector space is a sequence of vectors that are

(1) linearly independent and

(2) span the space.

The basis of a space tells us everything we need to know about that space.

Definition: Orthogonal Vectors

Recall:

v and u are said to be orthogonal to each other if $v \cdot u = 0$ or $\boldsymbol{v}^T \boldsymbol{u} = 0$

Vector Notation

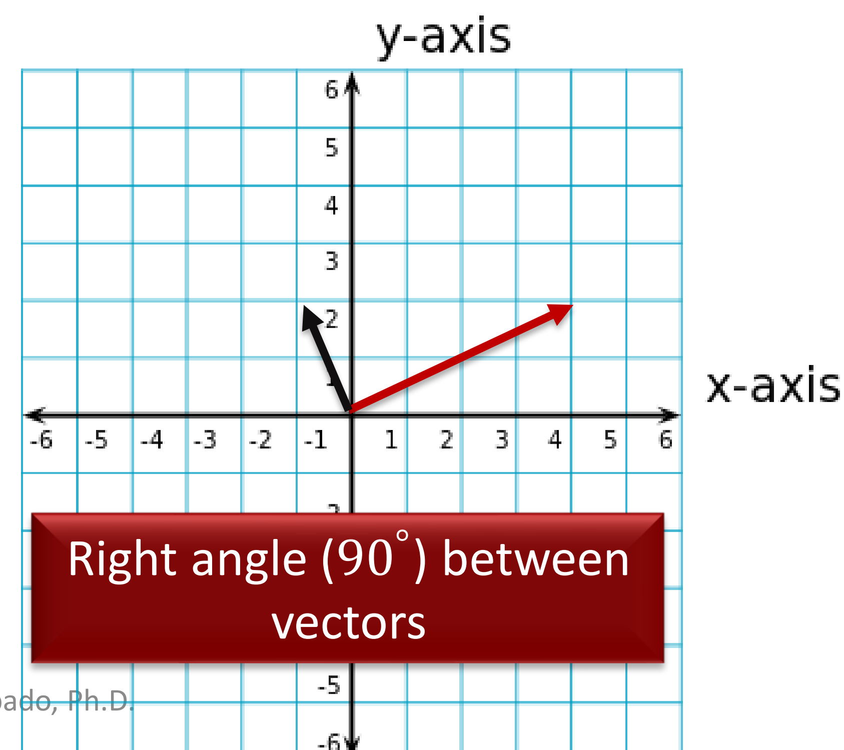
$$v \cdot u = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix} =$$

$$4 \cdot -1 + 2 \cdot 2 = 0$$

Matrix Notation

$$v^T u = \begin{bmatrix} 4 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} =$$

$$4 \cdot -1 + 2 \cdot 2 = 0$$



Right angle (90°) between vectors

Definition: Orthogonal Matrix

The vectors $v_1, \cdots, v_n$ are <u>orthonormal</u> if

$$v_i{}^T v_j = \begin{cases} 0 & when\ i \neq j & Orthogonal\ vectors \\ 1 & when\ i = j & unit\ vectors:\ \|v_i\| = 1 \end{cases}$$

A matrix with orthonormal columns is assigned the special letter $\boldsymbol{Q}$

Examples

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad Q = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -2/\sqrt{2} & 1/\sqrt{3} \end{bmatrix}$$

Orthonormal Matrix - Properties

1. Q is not required to be square
2. $Q^TQ = I$
3. When Q is square, $Q^TQ = QQ^T = I$ means that $Q^T = Q^{-1}$

Example 1

$$Q = \frac{1}{3}\begin{bmatrix} 2 & -2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & -2 \end{bmatrix} \qquad Q^T = Q^{-1} = \frac{1}{3}\begin{bmatrix} 2 & 1 & 2 \\ -2 & 2 & 1 \\ 1 & 2 & -2 \end{bmatrix} \qquad Q^TQ = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Example 2

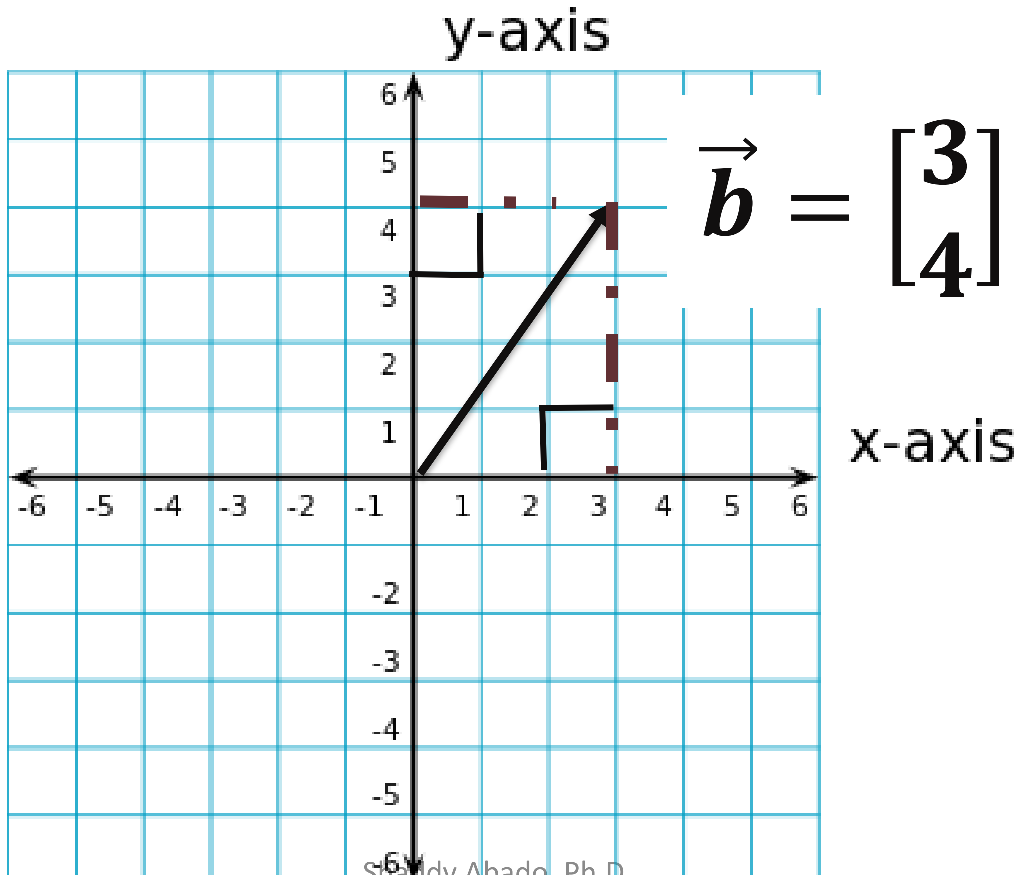
$$Q = \frac{1}{3}\begin{bmatrix} 1 & -2 \\ 2 & -1 \\ 2 & 2 \end{bmatrix} \qquad Q^TQ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad QQ^T \neq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Check

PROJECTIONS

An Orthogonal Line Projection

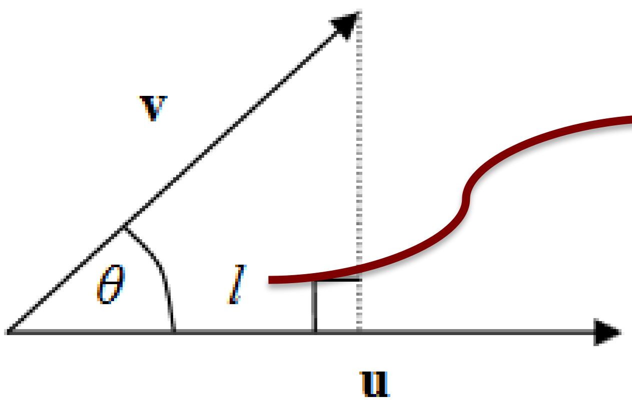
Given vector b, what is the orthogonal projections onto x- and y-axes?



Scalar Projection - Cosine Formula

$$\frac{\boldsymbol{v} \cdot \boldsymbol{u}}{\|\boldsymbol{v}\|\,\|\boldsymbol{u}\|} = \cos \boldsymbol{\theta}$$

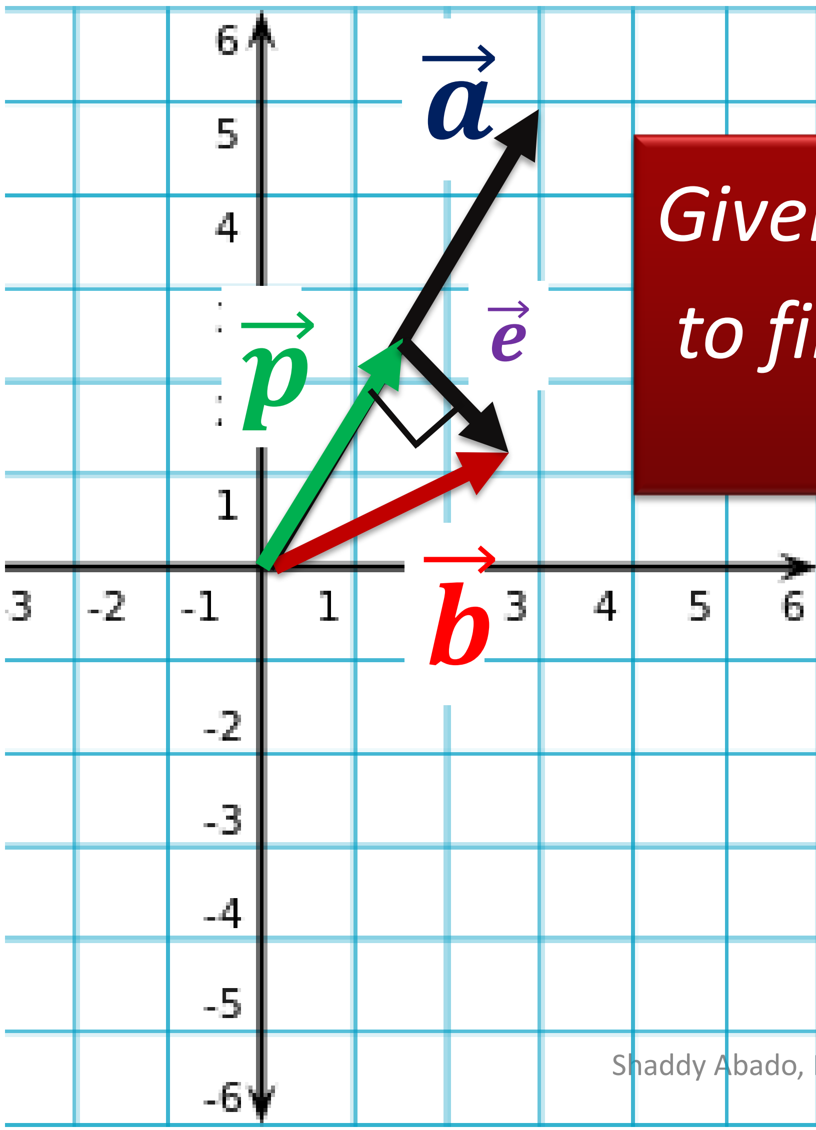
This relation can be used to provide a simple way of calculating the **Orthogonal <u>scalar</u> projection** of one vector in the direction of another



$$\boldsymbol{l} = \|\boldsymbol{v}\| \cos \boldsymbol{\theta} = \frac{\boldsymbol{v} \cdot \boldsymbol{u}}{\|\boldsymbol{u}\|}$$

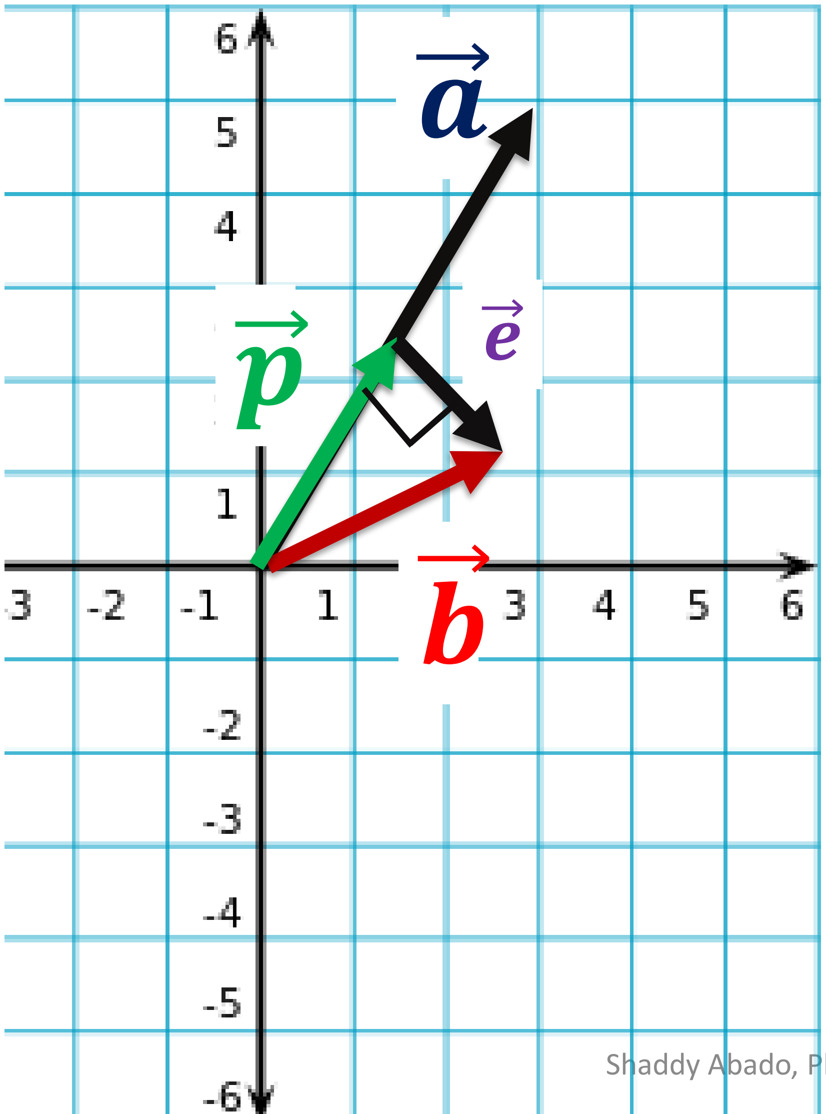
"How much" of a vector v is in a given direction (i.e., **components**)?

Vector Projection – Problem Statement



Given vectors $\vec{a}$ and $\vec{b}$, how to find the projection of $\vec{b}$ onto $\vec{a}$?

An Orthogonal Line Projection



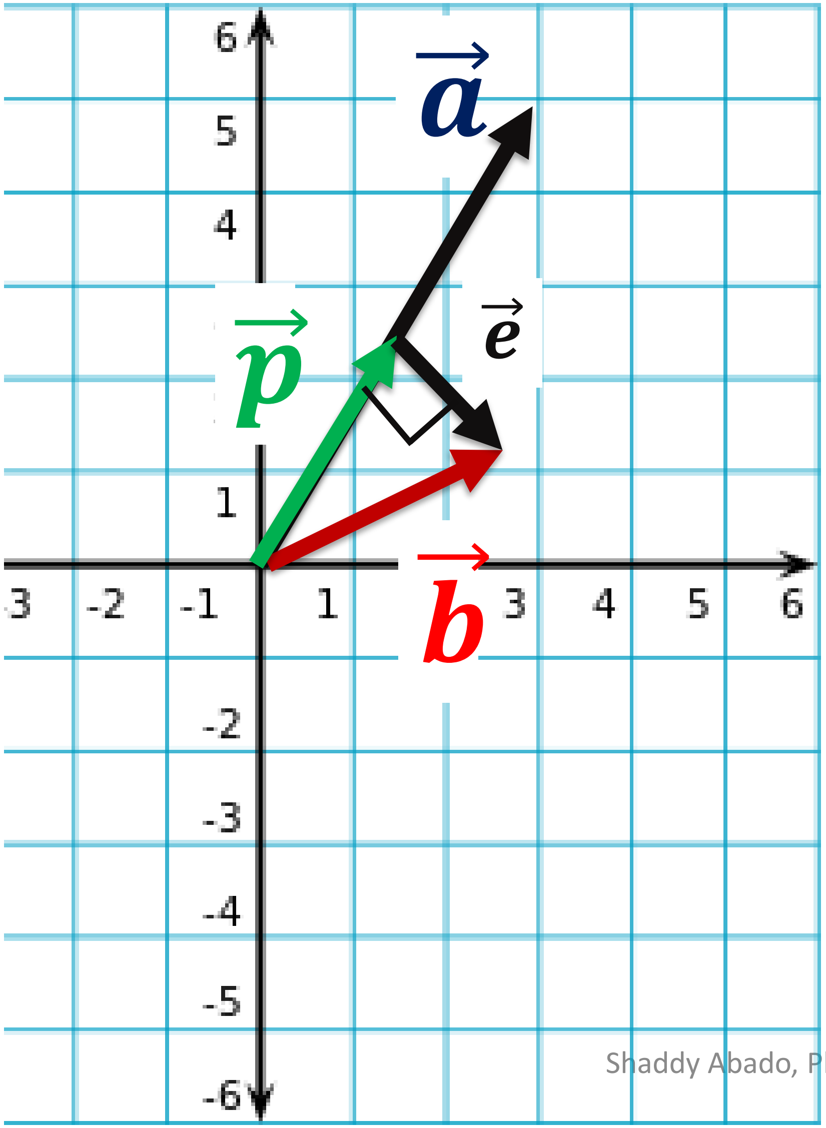
Given vectors $\vec{a}$ and $\vec{b}$, how to decompose it into a:

1. "projection" piece $\vec{p}$ which is multiple of $\vec{a}$
2. And an error/residual piece $\vec{e}$ which is orthogonal to $\vec{a}$

$$\vec{b} = \vec{e} + \vec{p}$$

$$\vec{b} = \vec{e} + \underbrace{\hat{x}}_{\text{Scalar}}\vec{a}$$

An Orthogonal Line Projection



Scalar

$$\vec{b} = \vec{e} + \hat{x}\vec{a}$$

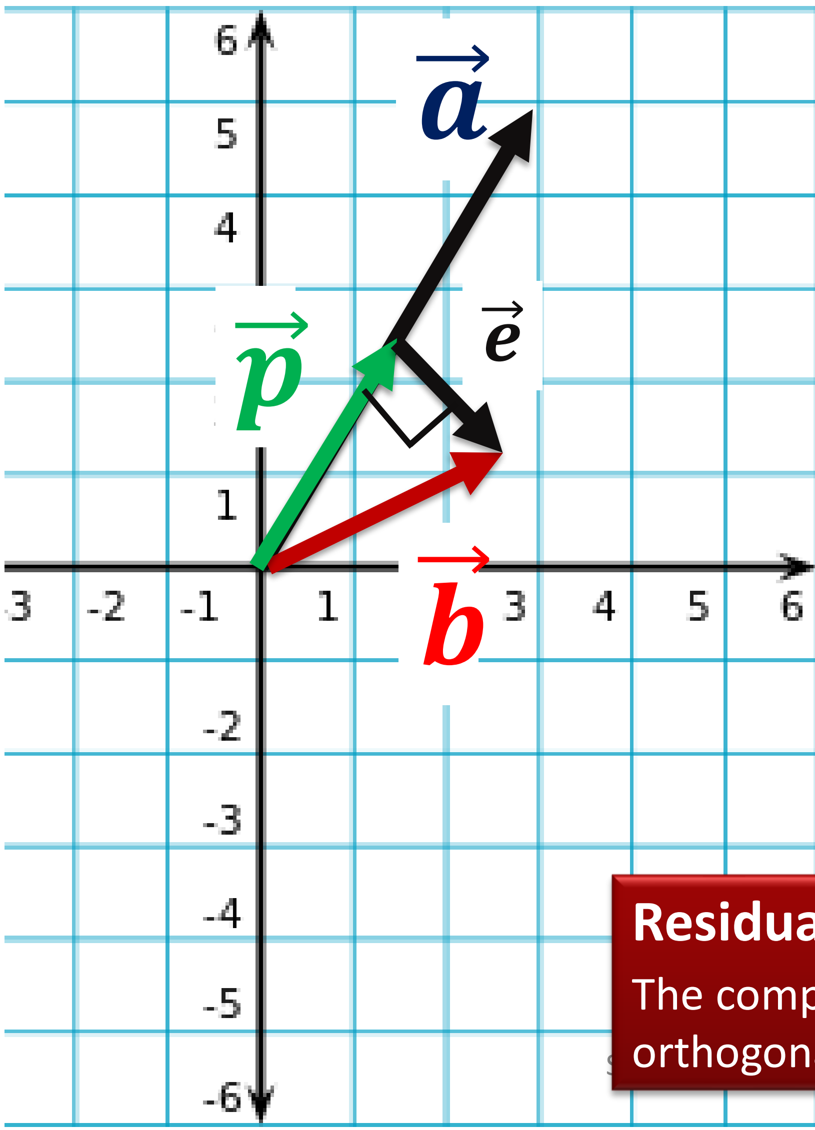
$$\hat{x} = \frac{\vec{a} \cdot \vec{b}}{\vec{a} \cdot \vec{a}}$$

or

$$\hat{x} = \frac{a^T b}{a^T a}$$

See Extra Slides for derivation

An Orthogonal Line Projection



$$\hat{x} = \frac{a^T b}{a^T a}$$

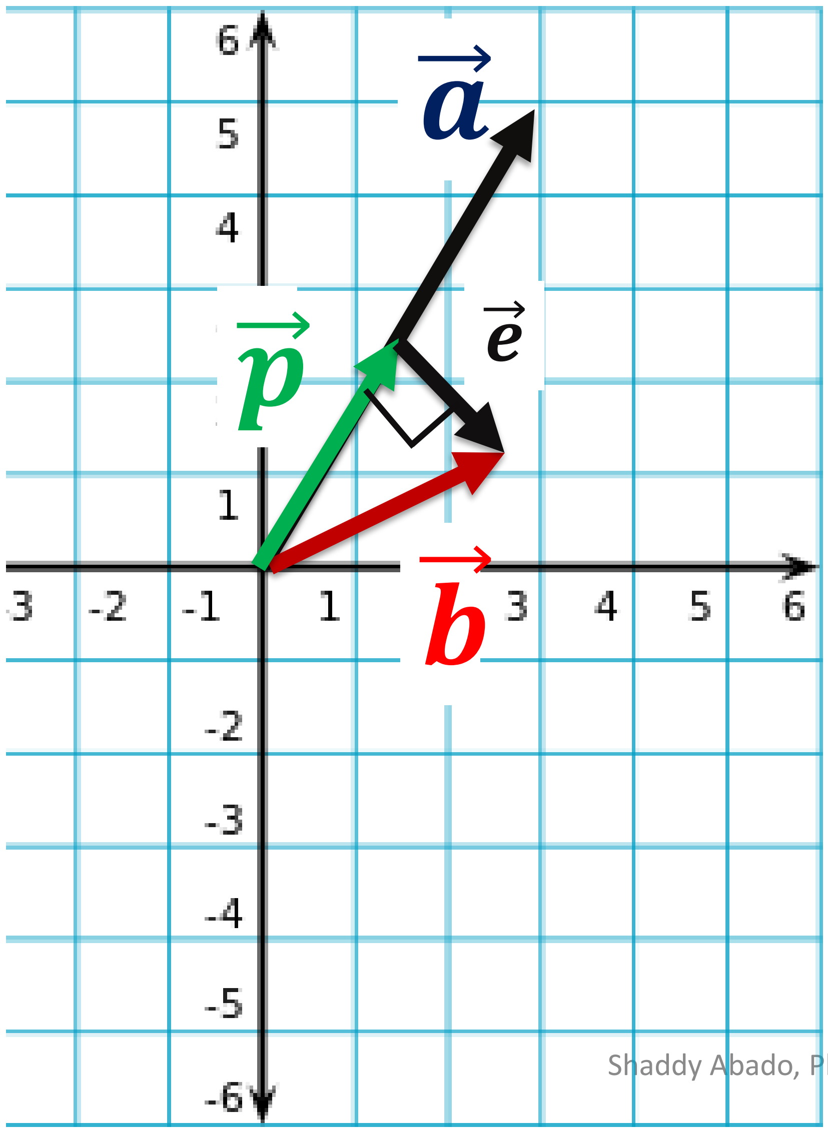
Scalar

$$\vec{b} = \vec{e} + \vec{p}$$
$$= \vec{e} + \frac{a^T b}{a^T a}\vec{a}$$

Residual:
The component of $\vec{b}$ orthogonal to $\vec{a}$

Projection:
Orthogonal projection of $\vec{b}$ onto $\vec{a}$

An Orthogonal Line Projection - Check



$$\vec{b} = \vec{e} + \vec{p}$$
$$= \vec{e} + \frac{a^T b}{a^T a}\vec{a}$$

If $\vec{b} = \vec{a}$,

then $\frac{a^T b}{a^T a} = 1$

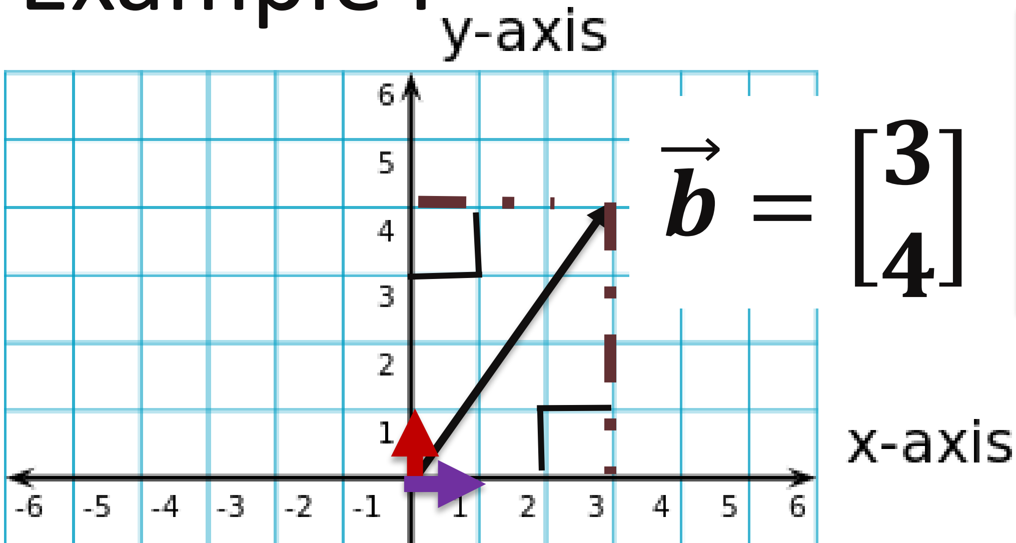
$\vec{p} = \vec{a}$ and $\vec{e} = 0$

If $\vec{b}$ and $\vec{a}$ are orthogonal

then $\frac{a^T b}{a^T a} = 0$

$\vec{p} = 0$ and $\vec{e} = \vec{b}$

An Orthogonal Line Projection – Example I



$$\vec{b} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

$$\vec{b} = \vec{e} + \vec{p} = \vec{e} + \frac{a^T b}{a^T a}\vec{a}$$

$$\|b\|^2 = \|p\|^2 + \|e\|^2$$

$$\vec{a} = e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\hat{x} = \frac{a^T b}{a^T a} = \frac{3}{1} = 3$$ **Weight**

Projection $$\vec{p} = \frac{a^T b}{a^T a}\vec{a} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

$$\vec{e} = \vec{b} - \vec{p} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} - \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$$

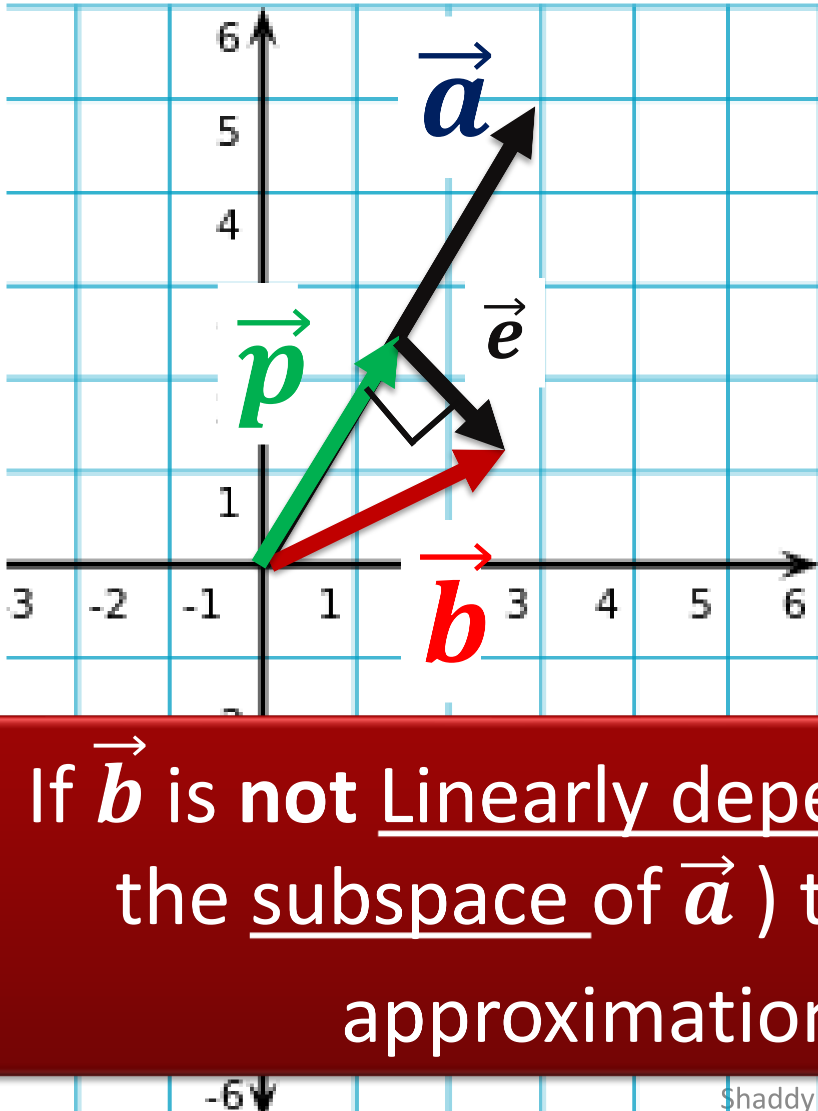
$$\vec{a} = e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\hat{x} = \frac{a^T b}{a^T a} = \frac{4}{1} = 4$$ **Weight**

Projection $$\vec{p} = \frac{a^T b}{a^T a}\vec{a} = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$$

$$\vec{e} = \vec{b} - \vec{p} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} - \begin{bmatrix} 0 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

Again: What are we trying to do?



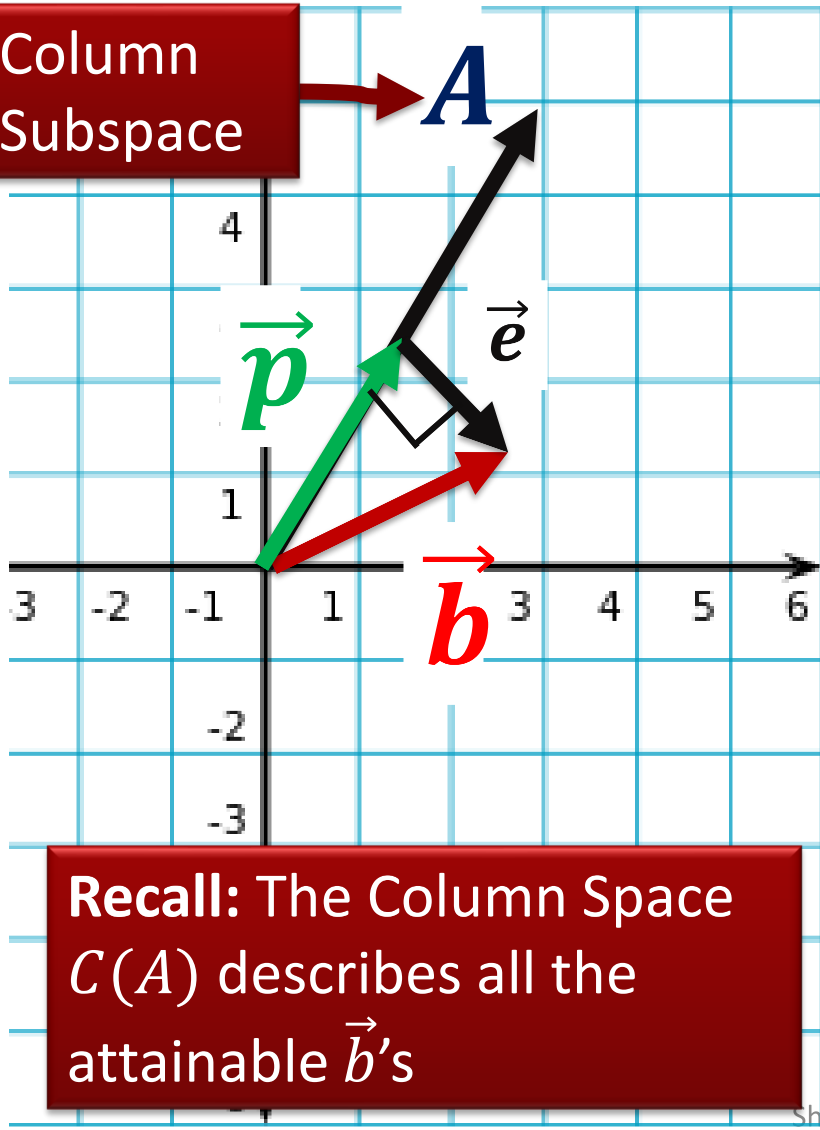
$$\vec{b} = \vec{e} + \vec{p}$$
$$= \vec{e} + \frac{a^T b}{a^T a}\vec{a}$$

If $\vec{b}$ is **not** Linearly dependent on $\vec{a}$ (i.e., **not** in the subspace of $\vec{a}$) then we find the best approximation ($\vec{p}$) to $\vec{b}$ in $\vec{a}$.

$$A_{mxn}x_{nx1} = b_{mx1}$$

Orthogonal Projection onto Subspace

Column Subspace



Recall: The Column Space $C(A)$ describes all the attainable $\vec{b}$'s

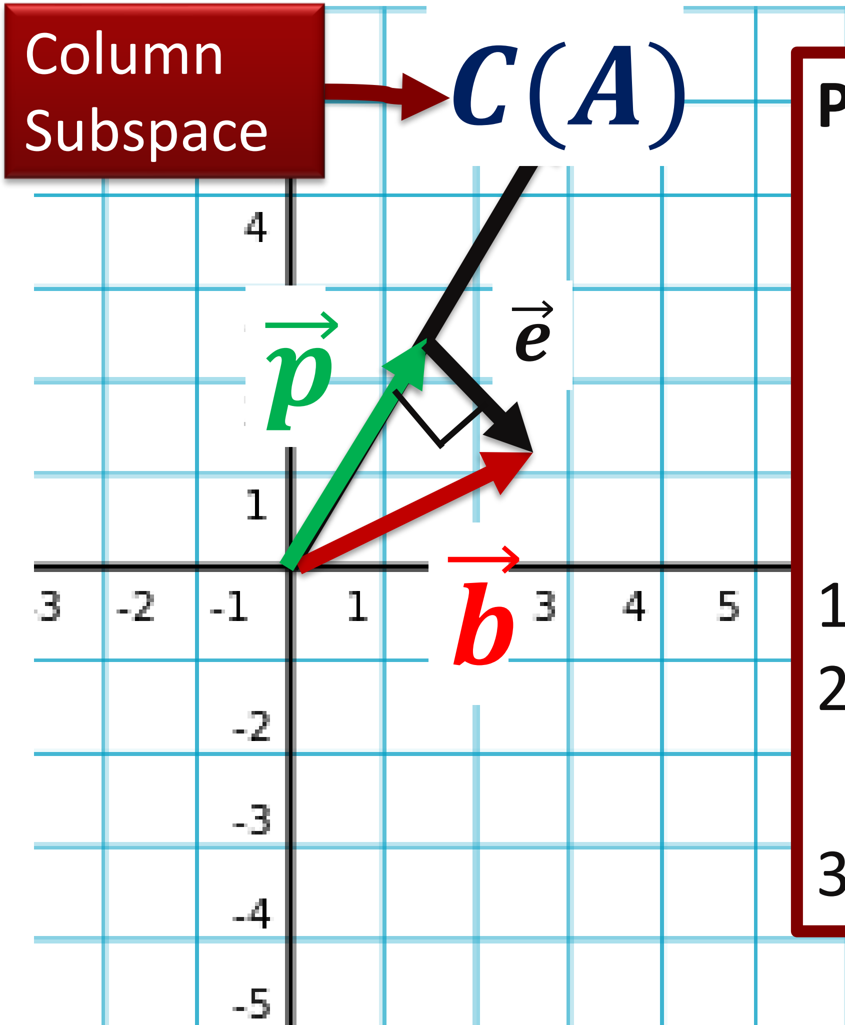
Recall:

Knowing that there is no unique solution, we are trying to find the <u>best approximate solution</u> for $A_{mxn}x_{nx1} = b_{mx1}$

Goal:

Given vector $\vec{\boldsymbol{b}}$ and assuming that matrix $\boldsymbol{A}$ is full ranks (i.e., all column vectors are linearly independent); then find the closest $\vec{\boldsymbol{p}}$ in $\boldsymbol{A}$ (where $\vec{\boldsymbol{p}}$ will be linear combination of matrix $\boldsymbol{A}$ columns)

Orthogonal Projection onto Subspace



Plan:

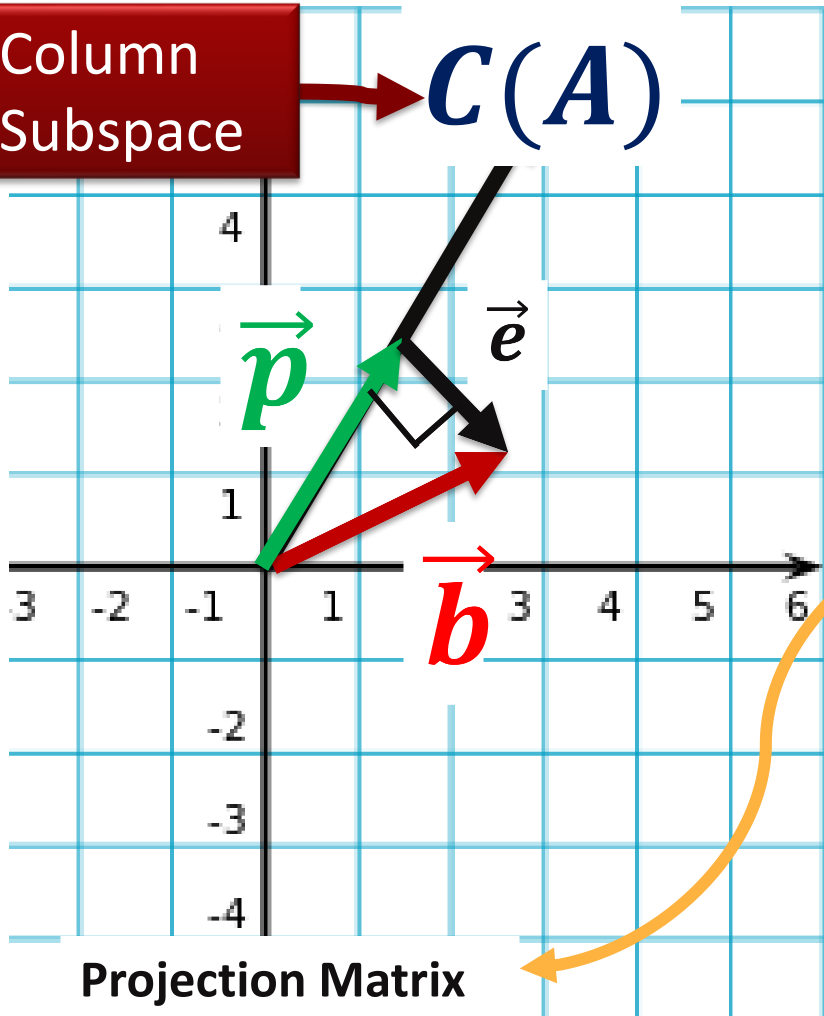
Similar to orthogonal line projection, we will compute the projection onto n-dimensional subspace by:

1. Finding 'weight' $\hat{x}$ (Vector)
2. Finding the <u>projection</u> ($\vec{p} = \boldsymbol{A}\hat{x}$) and <u>error</u> ($\vec{e}$) <u>vectors</u>
3. Find the projection matrix $\boldsymbol{P}$

Recall: The Column Space $C(A)$ describes all the attainable $\vec{b}$'s

Orthogonal Projection onto Subspace

Column Subspace



$$\vec{b} = \vec{e} + \vec{p}$$

$$\vec{b} = \vec{e} + A\hat{x}$$

'weights' Vector

Projection of b onto subspace A

$$\vec{p} = A\hat{x} = = \boxed{A\left(A^T A\right)^{-1} A^T} b$$

Residual

$$\vec{e} = \vec{b} - A\hat{x}$$

Projection Matrix

$$P = A\left(A^T A\right)^{-1} A^T$$

The projection vector $\vec{p}$ is the best approximation of $\vec{b}$ in $C(A)$

The error vector $\vec{e}$ is perpendicular to the subspace A

Orthogonal Projection onto Subspace

$(n = 1)$

$A_{mxn} x_{nx1} = b_{mx1}$

Orthogonal Line Projection

$n = 1$

Vector

Projection Matrix

$$\boldsymbol{P} = \frac{aa^T}{a^T a}$$

Projection of b onto vector a

$$\vec{p} = \boldsymbol{P}b = \vec{a}\hat{x} = \frac{a^T b}{a^T a}\vec{a}$$

$\hat{x}$ - Scalar

Orthogonal Projection onto a subspace

$n > 1$

Matrix

Projection Matrix

$$\boldsymbol{P} = \boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T$$

Projection of b onto subspace A

$$\vec{p} = \boldsymbol{P}b = \boldsymbol{A}\hat{x} = \boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T b$$

$\hat{x}$ - Vector

The linear independence of matrix A Columns guarantee that the inverse matrix exist

Example: Orthogonal Projection onto Subspace

Projection matrix: $P = A(A^TA)^{-1}A^T$

Projection vector: $\vec{p} = Pb = A\hat{x} = A(A^TA)^{-1}A^Tb$

Find the Orthogonal Projection of $\vec{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$ onto $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$

Symmetric

$$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$$

$$A^Tb = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

Weights vector

$$\hat{x} = (A^TA)^{-1}A^Tb = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}^{-1}\begin{bmatrix} 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$$

Projection matrix

$$P = A(A^TA)^{-1}A^T = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}^{-1}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}$$

Symmetric

Projection vector

$$\vec{p} = A(A^TA)^{-1}A^Tb = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 5 \\ -3 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$$

Error vector

$$\vec{e} = \vec{b} - \vec{p} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

$$P = A\left(A^T A\right)^{-1} A^T$$

A few more notes …

$A^T A$ is invertible if and only if A has linearly independent columns (Full column ranked)

Recall:

Matrix A_{mxn} is rectangular ➔ There is no inverse matrix

$A^T A$ ➔ A square nxn matrix

However, when A has independent columns then matrix $A^T A$ is

- A square nxn matrix
- Symmetric
- Invertible

Definition:

Normal Equation

Knowing that there is no unique solution, we are trying to find the best approximate solution $\hat{x}$ for

$$Ax = b \rightarrow A\hat{x} = p$$

$$\hat{x} = \left(A^T A\right)^{-1} A^T b$$ **Normal Equation**

$\hat{x}$ - **Vector**

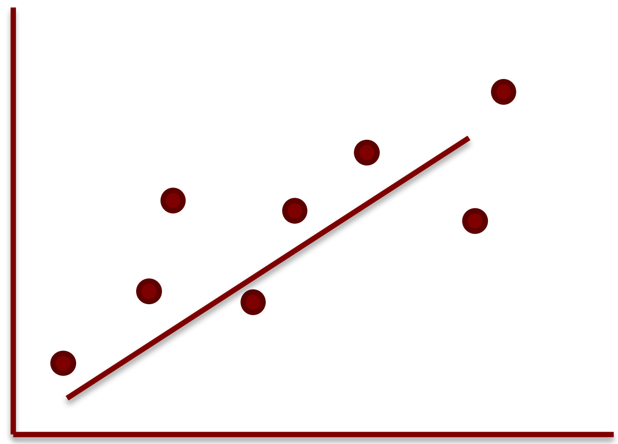
LEAST SQUARES APPROXIMATION

THE UNIVERSITY OF CHICAGO

Motivation:

Least Squares Estimate

For real life applications, data is contaminated with many error samples, and thus is not useful as a predictive tool.

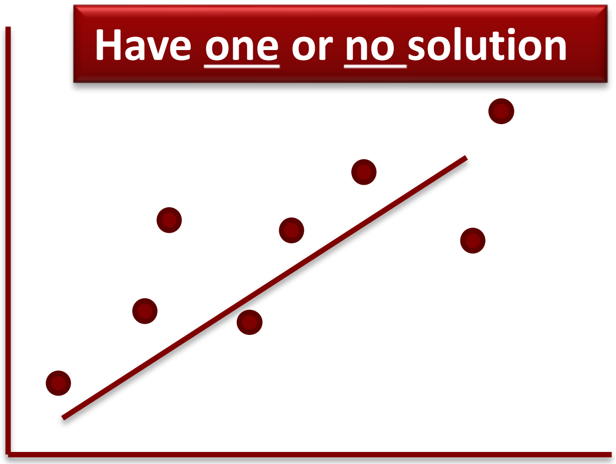


Over-determined
System of Equations

- In such cases, we needed to separate the most important information from the less important information (noise).
- It may be more useful to choose a lower order curve which does not exactly pass through all available points, but which does minimize the residual.

Over-Determined System of Equations

Over-determined System of Equations $m > n$
(# of Equations > # of Unknowns)



$$\boxed{A}\, x = b$$

$$\boldsymbol{A}_{mxn} x_{nx1} = b_{mx1}$$

Goal:

Is there a line which fits the data in some optimal sense?

We want the “best" possible solution

Least Squares Approximation

Over-determined System of Equations with no solution

$$\boldsymbol{A}x = b$$

In linear algebra terms, b is not in the column space $C(\boldsymbol{A})$ of $\boldsymbol{A}$. But we can "project" b onto $C(\boldsymbol{A})$ and get the vector $\vec{p} = \boldsymbol{A}\hat{x}$ in $C(\boldsymbol{A})$ that most closely resembles b.

$$\hat{x} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{b}$$

Normal Equation

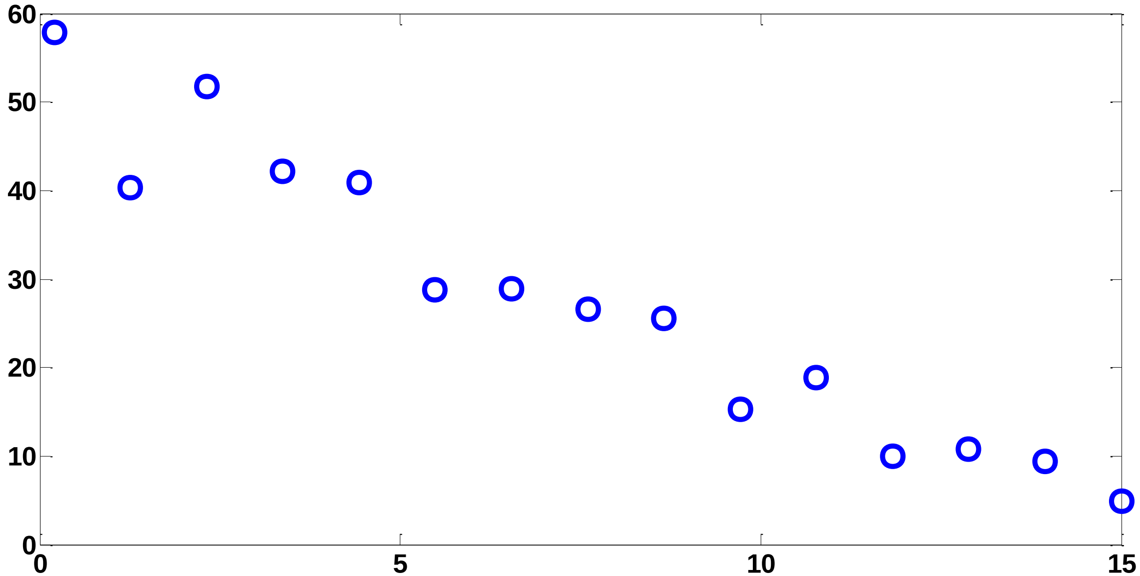
Recall

$$\vec{p} = \boldsymbol{P}b = \boldsymbol{A}\hat{x} = \boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T b$$

Example I - LSE

x_i	y_i
0.20	57.87
1.26	40.35
2.31	51.78
3.37	42.16
4.43	40.96
5.49	28.78
6.54	28.95
7.60	26.59
8.66	25.52
9.71	15.23
10.77	18.92
11.83	9.98
12.89	10.81
13.94	9.37
15.00	4.91

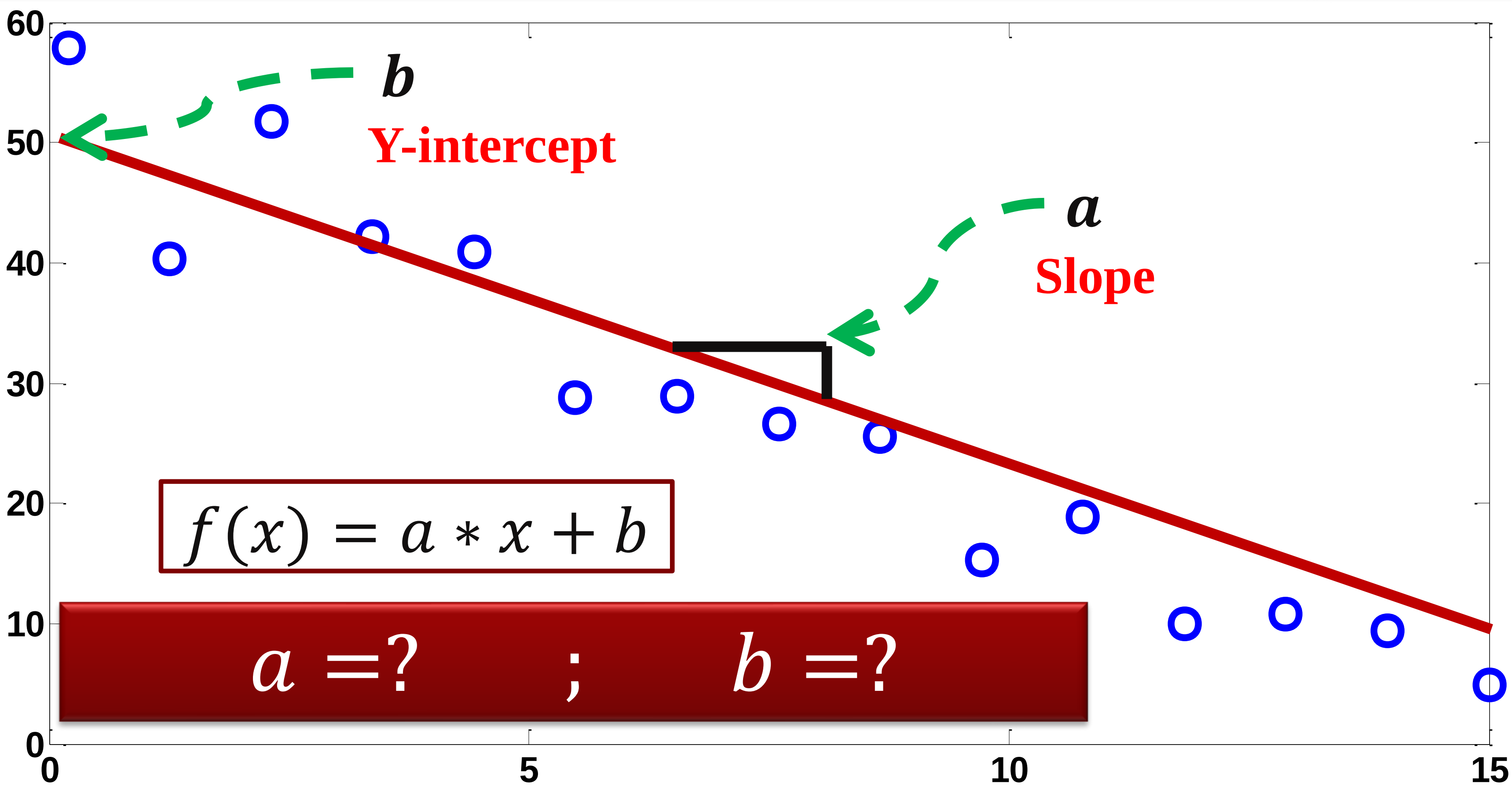
Step 1) Examines the data



We can notice a strong linear relationship between x and y.

Example I - LSE

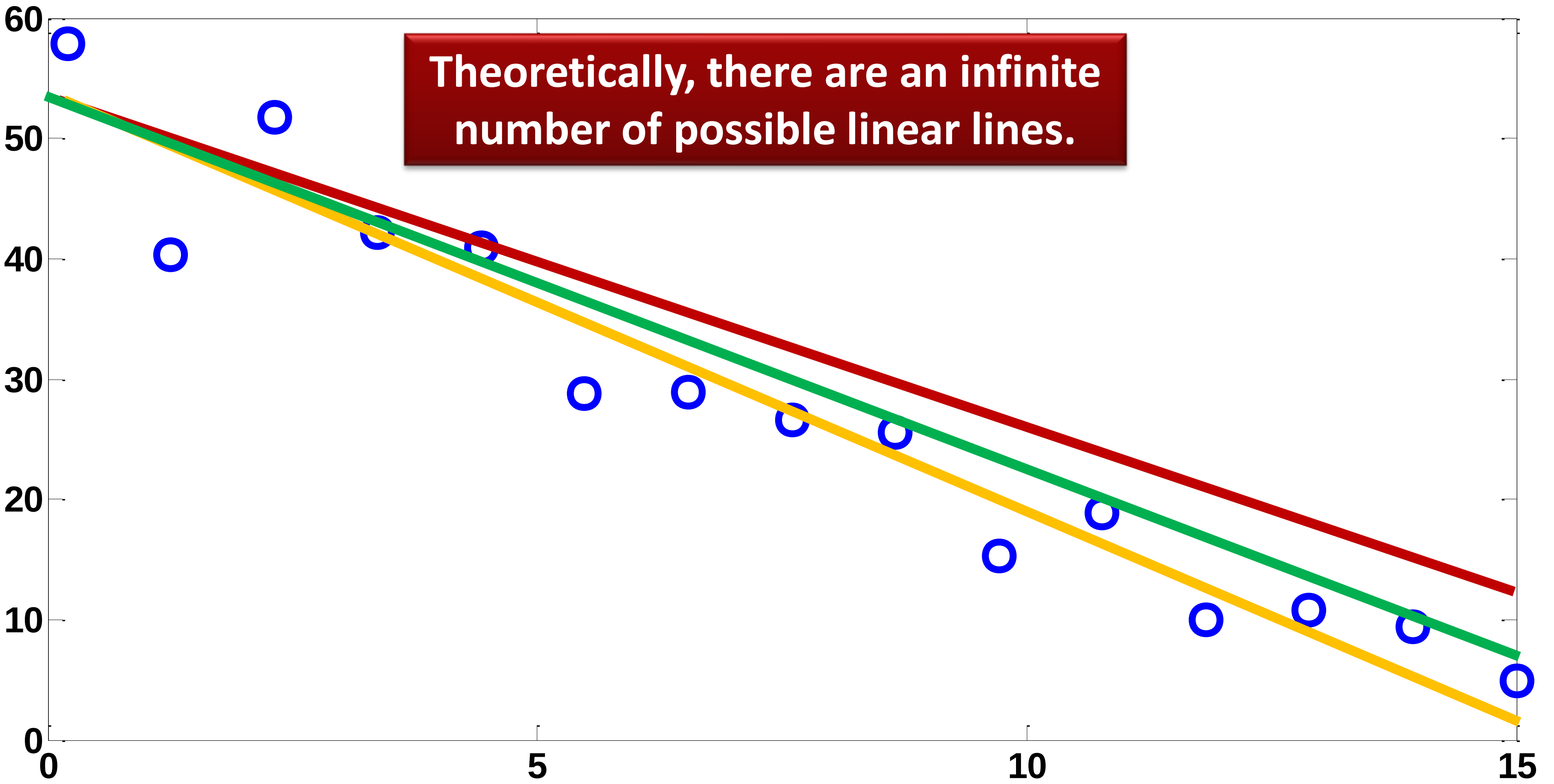
Step 2) Makes a non-unique judgment of what the functional form might be



Example I – LSE (Linear equations in two variables)

$$f(x) = a * x + \boxed{b}$$

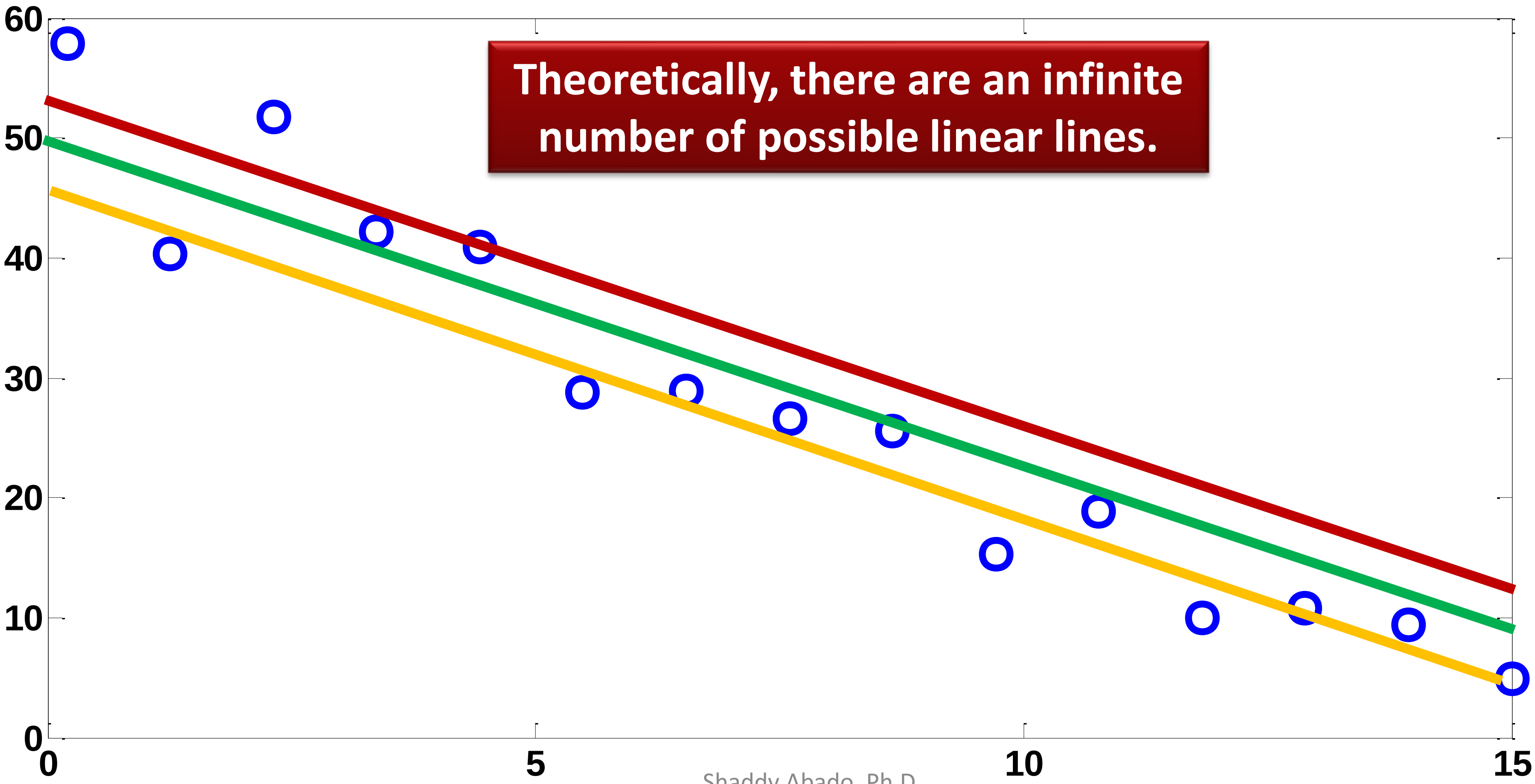
Fixed



Example I – LSE
(Linear equations in two variables)

Fixed

$$f(x) = \boxed{a} * x + b$$



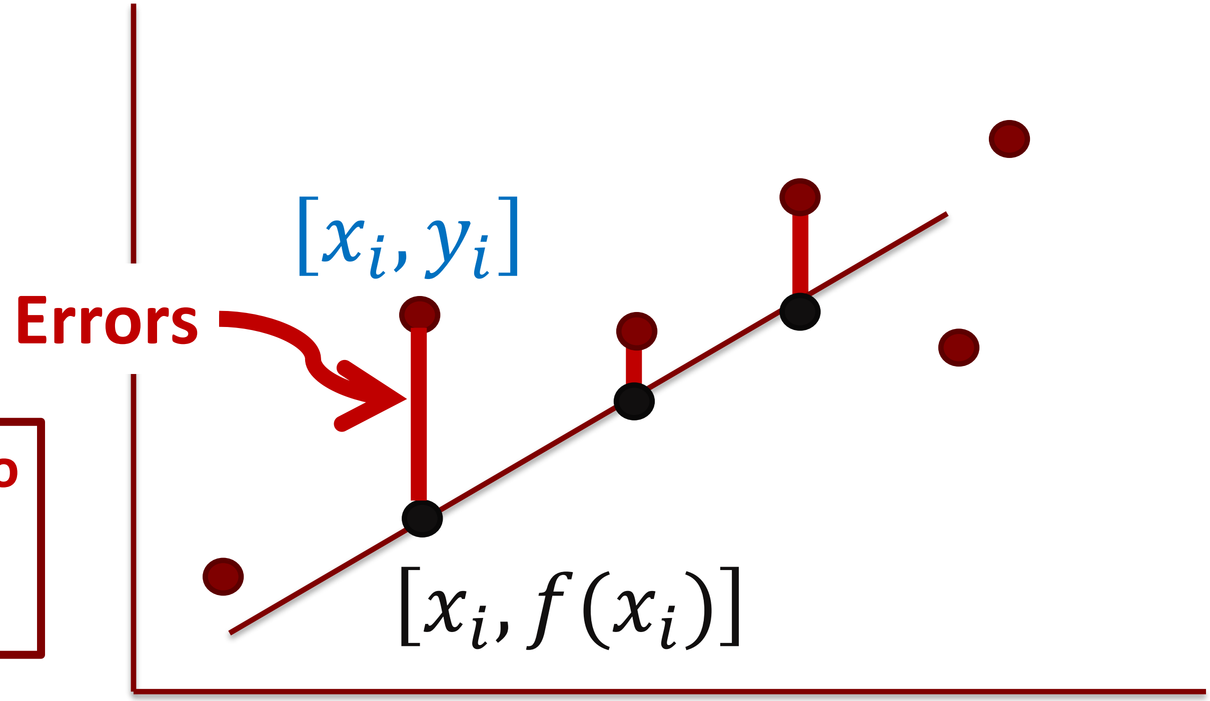
Example I – LSE (Linear equations in two variables)

Target Function

$$f(x_i) = y_i = a * x_i + b$$

Vertical distance (Error) to the line:

$$[(a * x_i + b) - y_i]$$



Goal:

Minimize the vertical distance to the target function

How?

Minimize the “projected” distance of each measured point to the target function (line).

How to measure the “best” approximation?

Minimizing the sum of the squares of the vertical distances (errors) from the points to the subspace

Sum of Squared Errors

$$SSE = \frac{1}{2}\sum_{i=1}^{N} Error^2 = \frac{1}{2}\sum_{i=1}^{N}[f(x_i) - y_i]^2$$
$$= \frac{1}{2}\sum_{i=1}^{N}[(a * x_i + b\ ) - y_i]^2 \rightarrow Small$$

This is the least squares approximation to the points.
The error (i.e., SSE) is as small as possible.

$$\vec{e} = \vec{b} - A\hat{x}$$

Normal Equation

Goal:

Minimizing the squared length of the vector $A\hat{x} - b$

$$\frac{1}{2}\|A\hat{x} - b\|^2 \to 0$$

Normal Equation

$$A^T A\hat{x} = A^T b$$

$$\hat{x} = \left(A^T A\right)^{-1} A^T b$$

Example I - LSE

(Over-determined System of Equations)

Step 3) Substitutes each data point into the assumed form to form an over-determined system of linear equations

x_i	Y_i
0.20	57.87
1.26	40.35
2.31	51.78
3.37	42.16
4.43	40.96
5.49	28.78
6.54	28.95
7.60	26.59
...	...
15.00	4.91

$$y_i = a_0 + a_1 * x_i$$

$$\boxed{\begin{gathered} 57.87 = a_0 + a_1 * 0.2 \\ 40.35 = a_0 + a_1 * 1.26 \\ \vdots \\ 4.91 = a_0 + a_1 * 15 \end{gathered}}$$

- Solve for a_0 and a_1
- This is an overdetermined system of equations (exact solution does not exist)

Example I - LSE (Matrix Notation)

$$57.87 = a_0 + a_1 * 0.2$$
$$40.35 = a_0 + a_1 * 1.26$$
$$\vdots$$
$$4.91 = a_0 + a_1 * 15$$

$$\boldsymbol{A\hat{x} = b}$$

- The x-values of the data points become the entries in the matrix A (**<u>Coefficient matrix</u>**)
- The y-values of the data points become vector b
- The coefficients a_0 and a_1 become the approximation $\hat{x}$ (**Unknowns**)

$$\underbrace{\begin{bmatrix} 1 & 0.2 \\ 1 & 1.26 \\ \vdots & \vdots \\ 1 & 15 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a_0 \\ a_1 \end{bmatrix}}_{\hat{x}} = \underbrace{\begin{bmatrix} 57.87 \\ 40.35 \\ \vdots \\ 4.91 \end{bmatrix}}_{b}$$

Example I - LSE

(Over-determined System of Equations)

x_i	Y_i
0.20	57.87
1.26	40.35
2.31	51.78
3.37	42.16
4.43	40.96
5.49	28.78
...	...
15.00	4.91

$$y_i = a_0 + a_1 * x_i$$

$$\begin{bmatrix} 1 & 0.2 \\ 1 & 1.26 \\ \vdots & \vdots \\ 1 & 15 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 57.87 \\ 40.35 \\ \vdots \\ 4.91 \end{bmatrix}$$

$$A \quad x \quad = \quad b$$

Step 4) Uses the normal equation to solve for the coefficients which best represent the given data.

$$\boldsymbol{A^T A \hat{x} = A^T b}$$

$$\boldsymbol{\hat{x} = \left(A^T A\right)^{-1} A^T b}$$

$$A^T A\hat{x} = A^T b$$
$$\hat{x} = \left(A^T A\right)^{-1} A^T b$$

Example I - LSE

$$\begin{array}{ccc} A & x = & b \end{array}$$

$$\begin{bmatrix} 0.2 & 1 \\ 1.26 & 1 \\ \vdots & \vdots \\ 15 & 1 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 57.87 \\ 40.35 \\ \vdots \\ 4.91 \end{bmatrix}$$

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \left(A^T A\right)^{-1} A^T b$$

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 52.81 \\ -3.33 \end{bmatrix}$$

Y-intercept

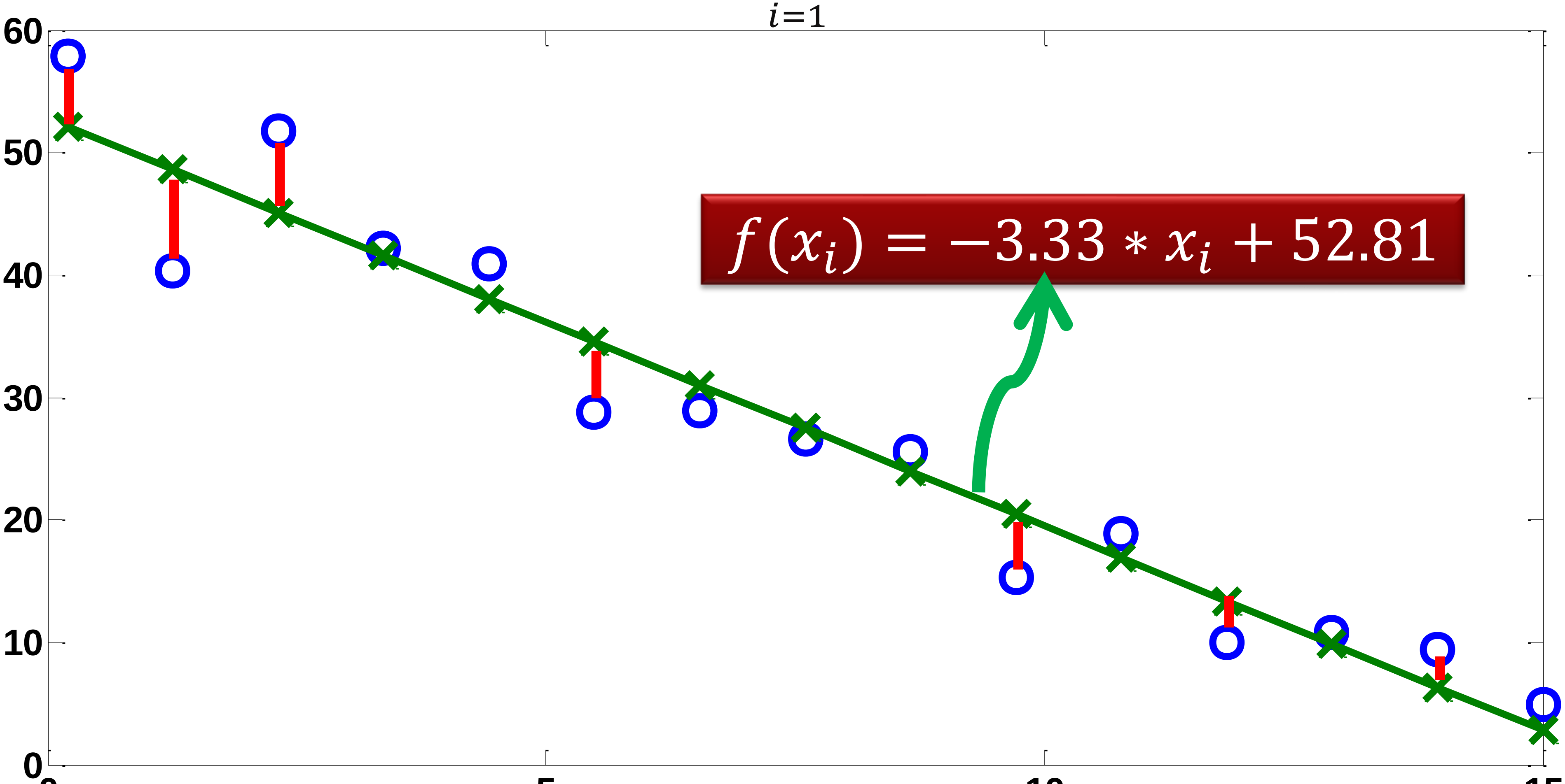
Slope

R^2 (Coefficient of Determination)

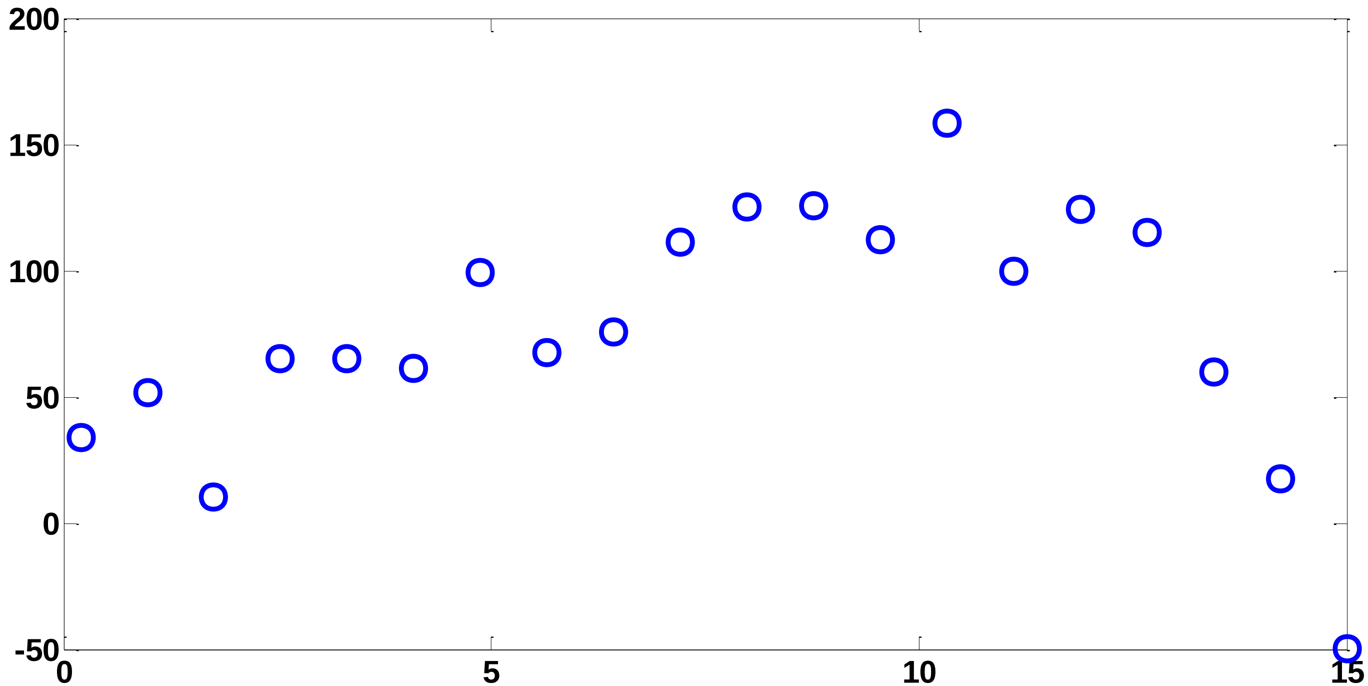
$$R^2 = \frac{\sum_{i=1}^{N}(f(x_i) - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} = \frac{3478.1}{3730.3} = 0.932$$

Sum of Squared Errors

$$SSE = \sum_{i=1}^{15}[f(x_i) - y_i]^2 = 252.1$$



Example II - LSE



x_i	y_i
0.20	33.86
0.98	51.84
1.76	10.53
2.54	65.16
3.32	65.35
4.09	61.12
4.87	99.39
5.65	67.46
6.43	75.49
7.21	111.41
7.99	125.07
8.77	125.66
9.55	112.28
10.33	158.37
11.11	99.59
11.88	124.40
12.66	115.21
13.44	60.08
14.22	17.62
15.00	-49.74

Step 1) Examines the data
Step 2) Makes a non-unique judgment of what the functional form might be

Find the 1st, 2nd and 3rd order polynomial fit

Example II - LSE

$$A^T A\hat{x} = A^T b$$
$$\hat{x} = \left(A^T A\right)^{-1} A^T b$$

$$f(x_i) = a_1 * x_i + a_0$$

$$\boldsymbol{A} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix} = \begin{pmatrix} 1 & 0.2 \\ 1 & 0.98 \\ \vdots & \vdots \\ 1 & 15 \end{pmatrix} \quad \boldsymbol{b} = \begin{pmatrix} 33.86 \\ 51.84 \\ \vdots \\ -49.74 \end{pmatrix}$$

$$\boxed{\hat{x} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 69.0 \\ 0.98 \end{pmatrix}}$$

$$f(x_i) = a_2 * x_i^2 + a_1 * x_i + a_0$$

$$\boldsymbol{A} = \begin{pmatrix} 1 & x_1 & (x_1)^2 \\ 1 & x_2 & (x_2)^2 \\ \vdots & \vdots & \vdots \\ 1 & x_N & (x_N)^2 \end{pmatrix} = \begin{pmatrix} 1 & 0.2 & (0.2)^2 \\ 1 & 0.98 & (0.98)^2 \\ \vdots & \vdots & \vdots \\ 1 & 15 & (15)^2 \end{pmatrix} \quad \boldsymbol{b} = \begin{pmatrix} 33.86 \\ 51.84 \\ \vdots \\ -49.74 \end{pmatrix}$$

$$\boxed{\hat{x} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} -8.3 \\ 32.2 \\ -2.1 \end{pmatrix}}$$

Example II - LSE

$$A^T A\hat{x} = A^T b$$
$$\hat{x} = \left(A^T A\right)^{-1} A^T b$$

$$f(x_i) = a_3 * x_i^3 + a_2 * x_i^2 + a_1 * x_i + a_0$$

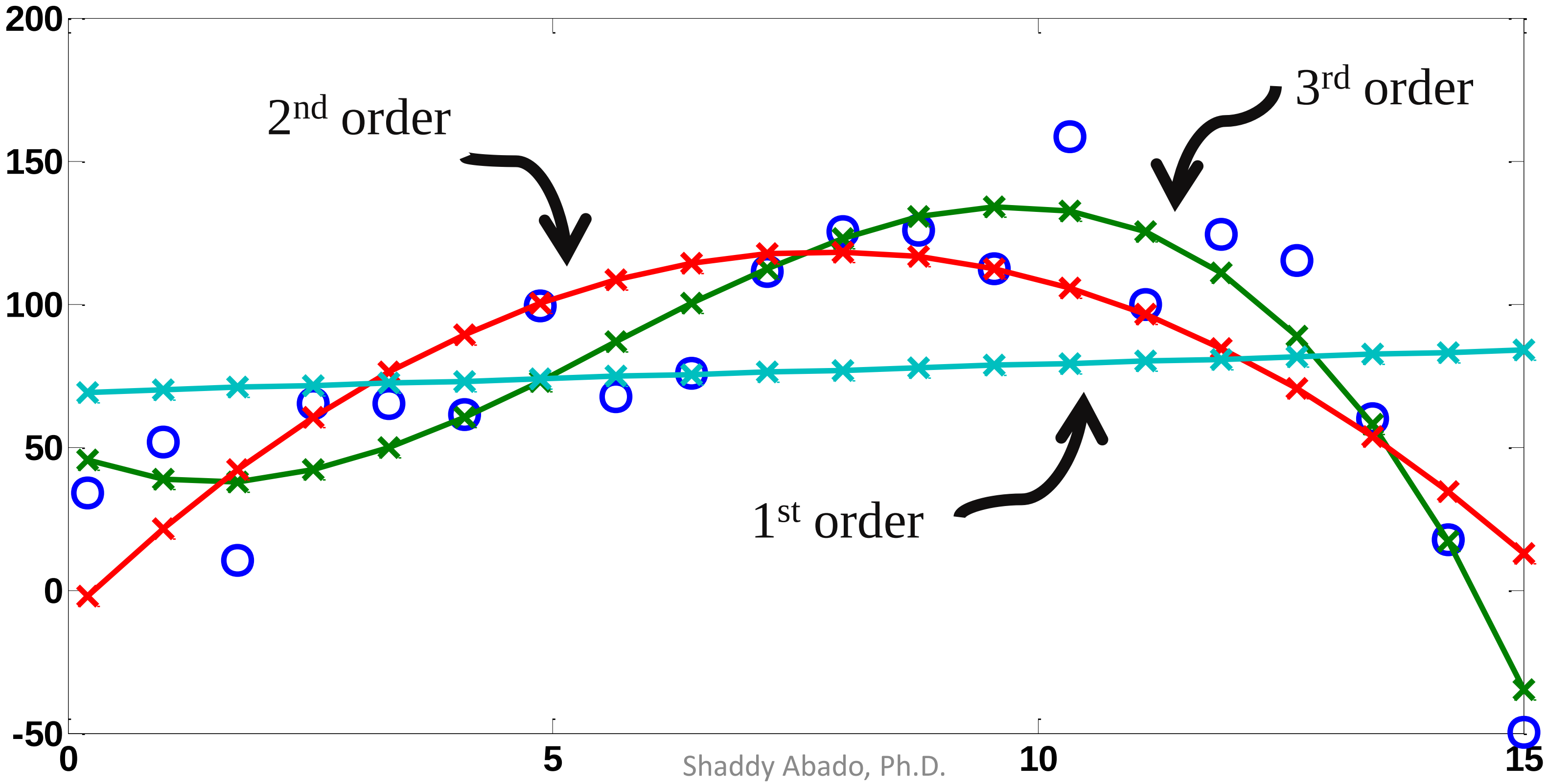
$$A = \begin{pmatrix} 1 & x_1 & (x_1)^2 & (x_1)^3 \\ 1 & x_2 & (x_2)^2 & (x_2)^3 \\ 1 & \vdots & \vdots & \vdots \\ 1 & x_N & (x_N)^2 & (x_N)^3 \end{pmatrix} = \begin{pmatrix} 1 & 0.2 & (0.2)^2 & (0.2)^3 \\ 1 & 0.98 & (0.98)^2 & (0.98)^3 \\ 1 & \vdots & \vdots & \vdots \\ 1 & 15 & (15)^2 & (15)^3 \end{pmatrix}$$

$$\boldsymbol{b} = \begin{pmatrix} 33.86 \\ 51.84 \\ \vdots \\ -49.74 \end{pmatrix}$$

$$\boxed{\hat{x} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 48.1 \\ -14.9 \\ 5.7 \\ -0.3 \end{pmatrix}}$$

Example II - LSE

Order	***SSE***
1^{st}	~45000
2^{nd}	~18000
3^{rd}	~6400



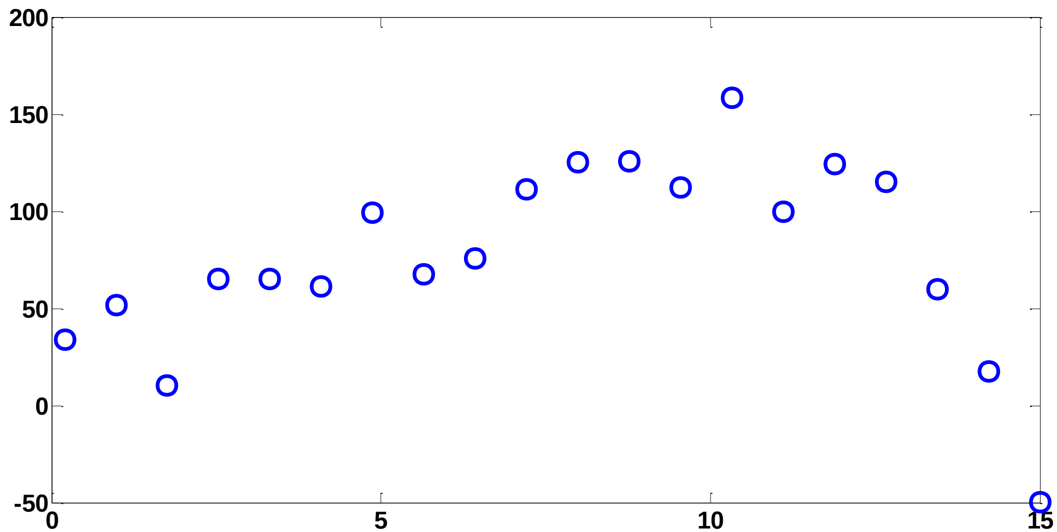
BREAK

THE UNIVERSITY OF CHICAGO

Weighted Least Squares

Motivation:

Suppose the measurements that produced the entries in b are not equally reliable.



Idea:

More importance is given to more reliable measurements: $\boldsymbol{A\vec{x} = b}$ ➔ $\boldsymbol{WA\vec{x} = Wb}$

Weighted Least Squares

Sum of Squared Errors

$$SSE = \frac{1}{2}\sum_{i=1}^{N} Error_i^2 = \frac{1}{2}\sum_{i=1}^{N}[f(x_i) - y_i]^2 \rightarrow Small$$

$$\|\boldsymbol{A}\hat{x} - \boldsymbol{b}\|^2 \rightarrow \boldsymbol{0} \quad \Rightarrow \quad \hat{x} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{b}$$

Normal equation for ordinary least squares

Weighted Sum of Squared Errors

$$WSSE = \frac{1}{2}\sum_{i=1}^{N} w_i * Error_i^2 = \frac{1}{2}\sum_{i=1}^{N} w_i[f(x_i) - y_i]^2 \rightarrow Small$$

$$\|\boldsymbol{W}(\boldsymbol{A}\hat{x} - \boldsymbol{b})\|^2 \rightarrow \boldsymbol{0} \quad \Rightarrow \quad \hat{x} = \left(\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{b}$$

$\boldsymbol{W}$ – Diagonal matrix (when errors are independent) with positive values

Normal equation for weighted least squares

Weighted Least Squares

$$\boldsymbol{W} = \begin{pmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{pmatrix}$$

OLSE $\hat{x} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{b}$

WLSE $\hat{x} = \left(\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{b}$

Note I

Ordinary least squares is a special case where all the weights $w_i = 1$.

If $\boldsymbol{W} = \boldsymbol{I}$ then $\hat{x} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{b}$

Note II

To apply weighted least squares, we need to know the weights $w_1 \ldots w_n$. In many real-life situations, the weights are not known apriori.

Python

scipy.linalg.lstsq

$$b = \begin{pmatrix} 33.86 \\ 51.84 \\ \vdots \\ -49.74 \end{pmatrix} \quad A = \begin{pmatrix} 1 & 0.2 & (0.2)^2 & (0.2)^3 \\ 1 & 0.98 & (0.98)^2 & (0.98)^3 \\ 1 & \vdots & \vdots & \vdots \\ 1 & 15 & (15)^2 & (15)^3 \end{pmatrix}$$

```
In [9]: # scipy.linalg.lstsq
        # https://docs.scipy.org/doc/scipy-0.18.1/reference/generated/scipy.linalg.lstsq.html#scipy.linalg.lstsq

        order=3
        b=Y
        A = CreateA(X,order)

        w,Res, rank, sv = LA.lstsq(A,b)

        print('3rd Order fit (using lstsq)')
        print("Weights = ",w)
        print("Residuals = ",Res)
        print("Rank = ",rank,'\n')

        3rd Order fit (using lstsq)
        Weights =  [ 48.09748761 -14.91235192   5.78729892  -0.34405083]
        Residuals =  6440.54203236
        Rank =  4
```

Normal Equation

$$\hat{x} = \left(A^T A\right)^{-1} A^T b$$

```
order=3
b=Y
A = CreateA(X,order)

w=LA.inv(np.dot(A.T,A)).dot(A.T).dot(b)
print('3rd Order fit (using matrix multiplication)')
print("Weights = ",w)

3rd Order fit (using matrix multiplication)
Weights =  [ 48.09748761 -14.91235192   5.78729892  -0.34405083]
```

QR Decomposition

A real square matrix A may be decomposed as

$$\boldsymbol{A} = \boldsymbol{QR}$$

Orthogonal Matrix
$\boldsymbol{mxn}$

Upper triangular w/ positive diagonals
$(\boldsymbol{R} = \boldsymbol{Q}^T\boldsymbol{A})$
$\boldsymbol{nxn}$

Notes:

- Any mxn matrix $\boldsymbol{A}$ with <u>independent columns</u> can be decomposed into $\boldsymbol{A} = \boldsymbol{QR}$
- If A is invertible, then the factorization is unique if we require the diagonal elements of R to be positive

QR Decomposition - Example

$$\boldsymbol{A} \quad = \quad \boldsymbol{Q} \quad \boldsymbol{R}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & -3 \\ 0 & -2 & 3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{18} \\ 0 & \sqrt{6} & -\sqrt{6} \\ 0 & 0 & \sqrt{3} \end{bmatrix}$$

```python
import numpy as np
import numpy.linalg as la


A = np.array([[1,2,3],
              [-1,0,-3],
              [0,-2,3]]
            )

Q,R = la.qr(A)
```

QR Decomposition & LSE

Recall

$$\boldsymbol{A^T A\hat{x} = A^T b}$$

$$\boldsymbol{A = QR}$$

$$\boldsymbol{A^T A = (QR)^T QR} \rightarrow$$

$$\boldsymbol{= R^T Q^T QR} \rightarrow$$

$$\boldsymbol{= R^T R} \rightarrow$$

Q *Orthogonal matrix*

R *Upper triangular Matrix*

$$\boldsymbol{\cancel{R^T} R\hat{x} = \cancel{R^T} Q^T b} \rightarrow$$

$$\boldsymbol{LSE\!:\ \ R\hat{x} = Q^T b}$$

$$\boldsymbol{(Solved\ using\ back\ substitution)}$$

LINEAR FUNCTIONS AND LINEAR TRANSFORMATIONS

What is Linear Algebra?

Linear algebra is the study of vectors and linear functions

- What are vectors?
- What are linear functions?

In broad terms, vectors are things you can
(1) add and (2) scalar multiply

Linear functions are functions of vectors that respect these properties $T(c\boldsymbol{v} + d\boldsymbol{w}) = cT(\boldsymbol{v}) + dT(\boldsymbol{w})$

Linear and Nonlinear Relationships

Linear relationships are the main objective of study in this course.

Linear

$$f(x, y, z \ldots.) = w_0 + xw_1 + yw_1 + zw_3 + \cdots$$

$$f(x, y, z \ldots.) = w_0 + w_1 Sin(x) + Sin(y)w_2 + cos(y)w_3 + \cdots$$

$$f(x, y, z \ldots.) = w_0 + w_1 log(x) + exp(z)w_2 + yw_3 + \cdots$$

Linear in the unknown parameters

Nonlinear

$$f(x, y, z \ldots.) = w_0 + x^{w_1} + exp(w_2 z) + \cdots$$

$$f(x, y, z \ldots.) = w_0 + \frac{w_3 z}{w_1 x + w_2 y} + \cdots$$

Motivation:

Linear Transformation

Function

$$\boldsymbol{f}$$

Input

$$x$$

Output

$$\boldsymbol{f}(x)$$

Transformation

$$\boldsymbol{T}$$

Input

$$v$$

Output

$$\boldsymbol{T}(v)$$

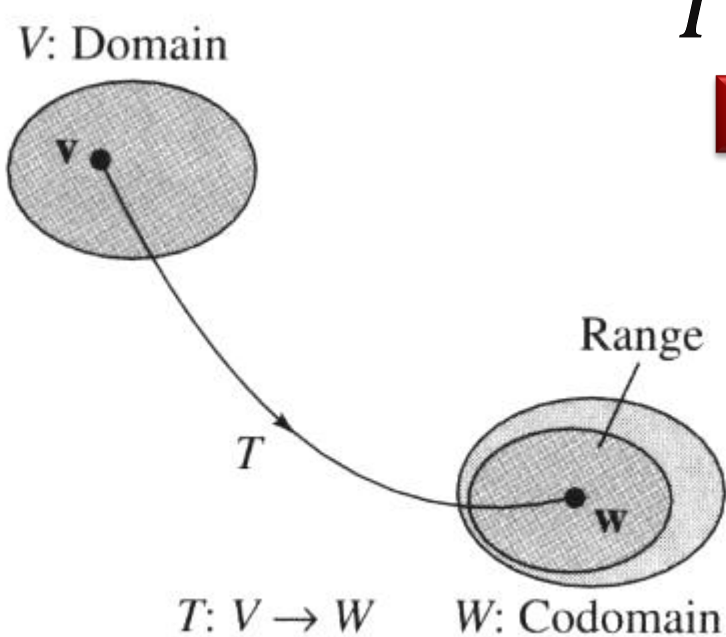
We can also refer to linear transformation as

- Linear functions
- Linear map
- Linear operator

Motivation:

Linear Transformation

Linear transformations is a way of moving from one vector space to another



$$T : V \overset{mapping}{\rightarrow} W$$

Domain **Codomain**

$V, W\ Vector\ Spaces$

Definition: Linear Transformation

From a linear algebra point of view, the most important transformations are those which preserve linear combinations.

Definition (Two axioms of linear transformations)

A transformation T is *linear* if:

$$T(\boldsymbol{v} + \boldsymbol{w}) = T(\boldsymbol{v}) + T(\boldsymbol{w})$$

Additivity

and

$$T(c\boldsymbol{v}) = cT(\boldsymbol{v})$$

Homogeneity

for all vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ and for all scalars c.

Equivalently, we can combine the previous two rules into one:

$$T(c\boldsymbol{v} + d\boldsymbol{w}) = cT(\boldsymbol{v}) + dT(\boldsymbol{w})$$

for all vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ and scalars c and d.

Linearity:

Additivity

+

Homogeneity

Definition: Linear Transformation

$$T(c\boldsymbol{v} + d\boldsymbol{w}) = cT(\boldsymbol{v}) + dT(\boldsymbol{w})$$

Note I – Null vector: Linear transformation maps $\mathbf{0}$ to $\mathbf{0}$: $T(\mathbf{0}) = \mathbf{0}$

If the input is $v = \mathbf{0}$

It is impossible to move the origin

Then the output must be $T(v) = T(\mathbf{0}) = \mathbf{0}$

Additivity

because if not, it couldn't be true that $T(c\mathbf{0}) = cT(\mathbf{0})$.

Homogeneity

Note II - Linearity:

In Engineering and physics: Superposition principle

If

$$u = v_1 * c_1 + \ldots + v_n c_n$$

Then

$$T(u) = c_1 * T(v_1) + c_2 * T(v_2) + \ldots + c_n * T(v_n)$$

The Superposition Principle – The most important idea in linear algebra

$$T(\boldsymbol{v} + \boldsymbol{w}) = T(\boldsymbol{v}) + T(\boldsymbol{w})$$

Linear transformations map any linear combination of inputs to the same linear combination of outputs.

If you know the outputs of T for the inputs v and w, you can deduce the output T for any linear combination of the vectors v and w by computing the appropriate linear combination of the outputs $T(v)$ and $T(w)$.

Source: "No bullshit guide to linear algebra“ by Ivan Savov

Definitions:

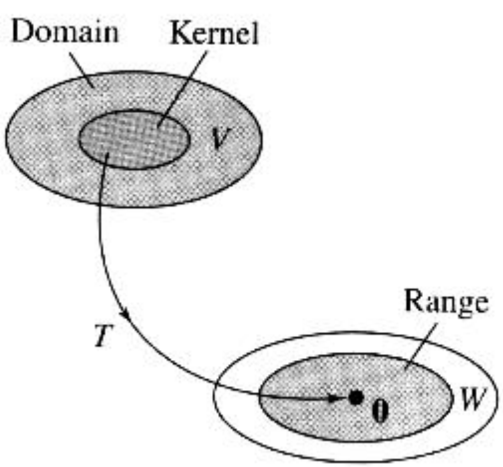
Range and Kernel of T Transformation

Range of T:

Set of all outputs $T(v)$

Kernel of T:

Set of all inputs for which $T(\boldsymbol{v}) = \mathbf{0}$



More about this later

Example I - Linear Transformation

Identity transformation

Input:

$$\boldsymbol{v}$$

Transformation:

$$T(\boldsymbol{v}) = \boldsymbol{v}$$

Linear

Why?

Note that: $T(c\boldsymbol{u}) = c\boldsymbol{u} = cT(\boldsymbol{u})$; $T(d\boldsymbol{w}) = d\boldsymbol{w} = dT(\boldsymbol{w})$

Then:

$$T(c\boldsymbol{u} + d\boldsymbol{w}) = c\boldsymbol{u} + d\boldsymbol{w} = cT(\boldsymbol{u}) + dT(\boldsymbol{w})$$

$$T(c\boldsymbol{u} + d\boldsymbol{w}) = cT(\boldsymbol{u}) + dT(\boldsymbol{w})$$

$$T(\boldsymbol{0}) = \boldsymbol{0}$$

Example II - Linear Transformation

Dot Product: $a \cdot v$

Input:

$$\boldsymbol{v}$$

Transformation:

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

Linear

Why?

Note that:

$$T(c\boldsymbol{u}) = \boldsymbol{a} \cdot c\boldsymbol{u} = c(\boldsymbol{a} \cdot \boldsymbol{u}) = cT(\boldsymbol{u}); \quad T(d\boldsymbol{w}) = dT(\boldsymbol{w});$$

Then:

$$T(c\boldsymbol{u} + d\boldsymbol{w}) = \mathbf{a} \cdot (c\boldsymbol{u} + d\boldsymbol{w}) =$$
$$c(\boldsymbol{a} \cdot \boldsymbol{u}) + d(\boldsymbol{a} \cdot \boldsymbol{w}) = cT(\boldsymbol{u}) + dT(\boldsymbol{w})$$

$$T(c\boldsymbol{u} + d\boldsymbol{w}) = cT(\boldsymbol{u}) + dT(\boldsymbol{w})$$
$$T(\mathbf{0}) = \mathbf{0}$$

Example III - Linear Transformation

Find the Kernel and Range of the Linear transformation $T\colon\ R^3 \to R^2$ defined by

$$T(x, y, z) = (z, x)$$

Kernel of $\boldsymbol{T}$:

Set of all inputs for which $T(\boldsymbol{v}) = \mathbf{0}$

Range of $\boldsymbol{T}$:

Set of <u>all outputs</u> $T(v)$

Kernel of $\boldsymbol{T}$:

Any vector that has only a y $-$component will be sent to the zero vector

$$Ker(T)\ =\ span\,\{(0,1,0)\}$$

Range of $\boldsymbol{T}$:

$$Range(T)\ =\ R^2$$

Example I - Nonlinear Transformation

Project every 3-dimentional vector onto horizontal plan $z = 1$

Input:

$$\boldsymbol{v} = (x, y, z)$$

Transformation:

$$T(\boldsymbol{v}) = (x, y, 1)$$

Not Linear

Why?

Doesn't transform $v = 0$ into $T(v) = 0$

If:

$$\boldsymbol{v} = \boldsymbol{0} = (0,0,0)$$

then:

$$T(\boldsymbol{0}) = (0,0,1) \neq \boldsymbol{0}$$

$$T(c\boldsymbol{v} + d\boldsymbol{w}) = cT(\boldsymbol{v}) + dT(\boldsymbol{w})$$

$$T(\boldsymbol{0}) = \boldsymbol{0}$$

MATRICES AS LINEAR MAPPINGS

Motivation:

Matrices as Linear Transformations

The purpose of this section is to make the connection between <u>matrix theory</u> and <u>linear transformation.</u>

Frequently, the best way to understand a linear transformation is to find the matrix that lies behind the transformation.

Goal:

We want to show that every linear transformation leads to a matrix.

$\boldsymbol{T}(\vec{v}) = \boldsymbol{A}\vec{v}$

Observation:

Matrix multiplication satisfies the rules of linearity

"Proof":

Given a matrix $\boldsymbol{A}$,

$$\boldsymbol{T}(c\vec{v} + d\vec{u}) = \boldsymbol{A}(c\vec{v} + d\vec{u}) = c(\boldsymbol{A}\vec{v}) + d(\boldsymbol{A}\vec{u})$$
$$= c\boldsymbol{T}(\vec{v}) + d\boldsymbol{T}(\vec{u})$$

Therefore, we can define the linear transformation

$$\boldsymbol{T}(\vec{v}) = \boldsymbol{A}\vec{v}$$

Conclusion:

Any matrix leads immediately to a linear transformation.

Mapping Vector Spaces

$$u_{mx1} = \boldsymbol{A}_{mxn} \cdot v_{nx1}$$

We can think of $\boldsymbol{A}$ as a linear transformation taking a vector v into m-dimensional column vector.

$$u_{1xn} = v_{1xm} \cdot \boldsymbol{A}_{mxn}$$

We can think of $\boldsymbol{A}$ as a linear transformation taking a vector v into n-dimensional row vector.

Conclusion:

We can view any $m\ by\ n$ matrix $\boldsymbol{A}$ as a function that maps one vector space onto another.

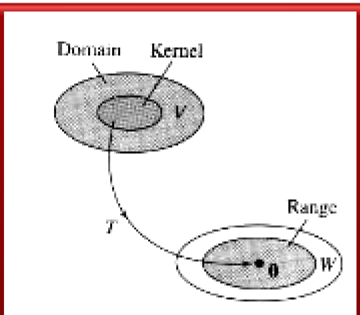
The Consequence of Linearity and Basis

Suppose that basis consists of the n vectors $v_1, \cdots, v_n$.
Then every other vector v is a combination of those particular vectors (i.e., they span the space).
Therefore, linearity determines Av

$$T(v) = Av$$

If

$$v = v_1 * c_1 + \ldots + v_n c_n$$



Then

$$Av = c_1(Av_1) + \ldots + c_n(Av_n)$$

If we know Av_i for every vector in a basis, then we know Av for each vector in the entire space.

Example I:

Transformations Represented by Matrices

Find the matrix representation of the Linear transformation $T\colon\ R^3 \to R^2$ defined by

$$T(x, y, z) = (x - 2y, 2x + y)$$

Input Basis (Standard Basis)

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Transformation

$$T(e_1) = T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$T(e_2) = T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

$$T(e_3) = T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Example I:

Transformations Represented by Matrices

$$A = \left[T(e_1) \mid T(e_2) \mid T(e_3)\right]$$

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

Standard Matrix

Check

$T:\ R^3 \to R^2$

$T(x, y, z) = (x-2y, 2x+y)$

$$A\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & -2 & 0 \\ 2 & 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x-2y \\ 2x+y \end{bmatrix}$$

Note

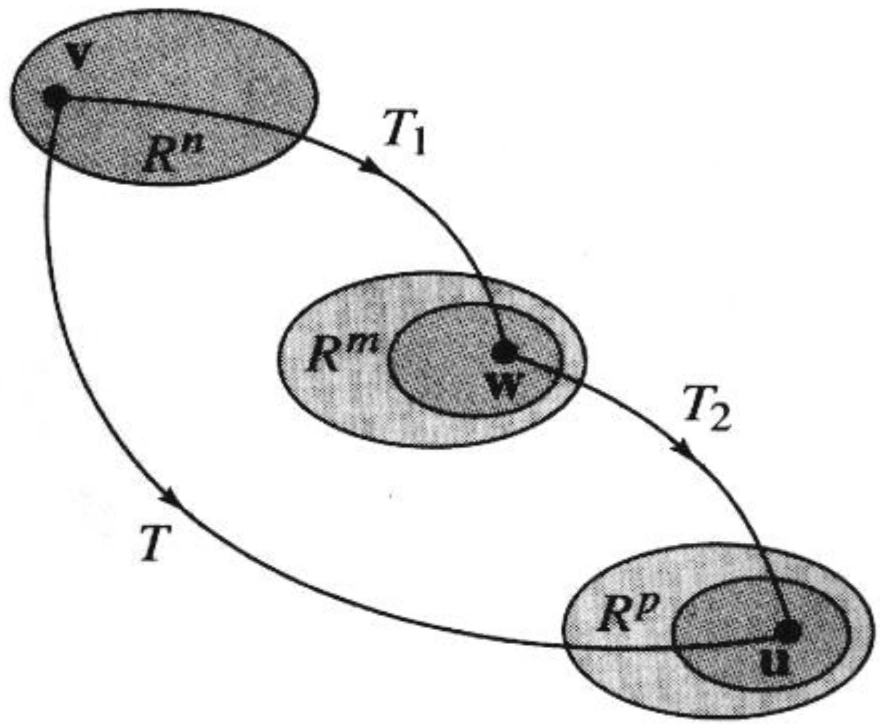
$$A = \begin{bmatrix} 1 & -2 & 0 \\ 2 & 1 & 0 \end{bmatrix} \begin{array}{l} \leftarrow \ 1x-2y+0z \\ \leftarrow \ 2x+1y+0z \end{array}$$

Definition:

Composition of Linear Transformations

Let $T_1 : R^n \to R^m$ and $T_2 : R^m \to R^p$ be L.T.

Find $T : R^n \to R^p$



Composition of Transformations

Definition:

Composition of Linear Transformations

Let $T_1 : R^n \rightarrow R^m$ and $T_2 : R^m \rightarrow R^p$ be L.T. with A_1 and A_2 , then

1) The composition $T\colon\ R^n \rightarrow R^p, defined\ by$

$$\boldsymbol{T}(v) = \boldsymbol{T}_2(\boldsymbol{T}_1(v))$$ is Linear Transformation

2) Matrix $\boldsymbol{A}$ for $\boldsymbol{T}$ is given by the matrix product

$$\boldsymbol{A} = \boldsymbol{A}_2\boldsymbol{A}_1$$

Matrix multiplication is just the operation of composing two linear transformations.

Recall

$$\boldsymbol{A}_2\boldsymbol{A}_1 \neq \boldsymbol{A}_1\boldsymbol{A}_2$$

Definition:

Composition of Linear Transformations

Let $T_1 : R^n \to R^m$ and $T_2 : R^m \to R^p$ be L.T. with A_1 and A_2 , then

$$T\colon\ R^n \to R^p$$

$$\boldsymbol{A} = \boldsymbol{A}_2\boldsymbol{A}_1$$

$$pxn \ = \ (pxm)(mxn)$$

Looking at matrix as a linear transformation gives us a natural explanation for the definition of matrix multiplication.

Definition:

Inverse of Linear Transformations

If $\boldsymbol{T_1 : R^n \to R^n}$ and $\boldsymbol{T_2 : R^n \to R^n}$ are L.T. such that for every $\mathbf{v}$ **in** $\boldsymbol{R^n}$

$$T_2(T_1(\mathbf{v})) = \mathbf{v} \quad \text{and} \quad T_1(T_2(\mathbf{v})) = \mathbf{v}$$

Then T_2 is called the inverse of T_1 and T_2 is said to be invertible

Note I - Uniqueness

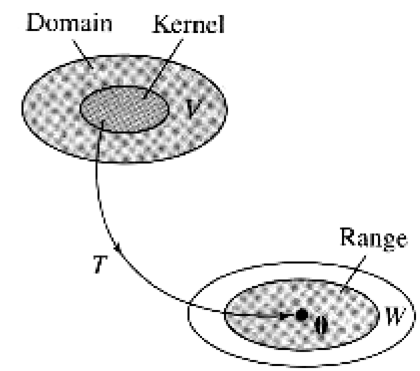
If the transformation $\boldsymbol{T}$ is invertible, then the inverse is unique and denoted by $\boldsymbol{T}^{-1}$.

Note II – Matrix inverse

If $\boldsymbol{T}$ is invertible with standard matrix $\boldsymbol{A}$, then the standard matrix for $\boldsymbol{T}^{-1}$ is $\boldsymbol{A}^{-1}$.

Recall:

Range and Kernel of T Transformation



Range of T:

Set of <u>all outputs</u> $T(v)$

Range corresponds to the column space $C(A)$

Kernel of T:

Set of all inputs for which $T(v) = \mathbf{0}$

Kernel corresponds to the nullspace $N(A)$

Example: Linear Transformation from R^2 into R^2

$$M = \begin{pmatrix} 1 & -1 \\ -2 & 2 \end{pmatrix}$$

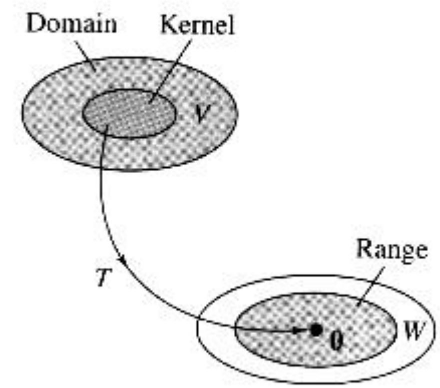
Linear transformation from R^2 into R^2

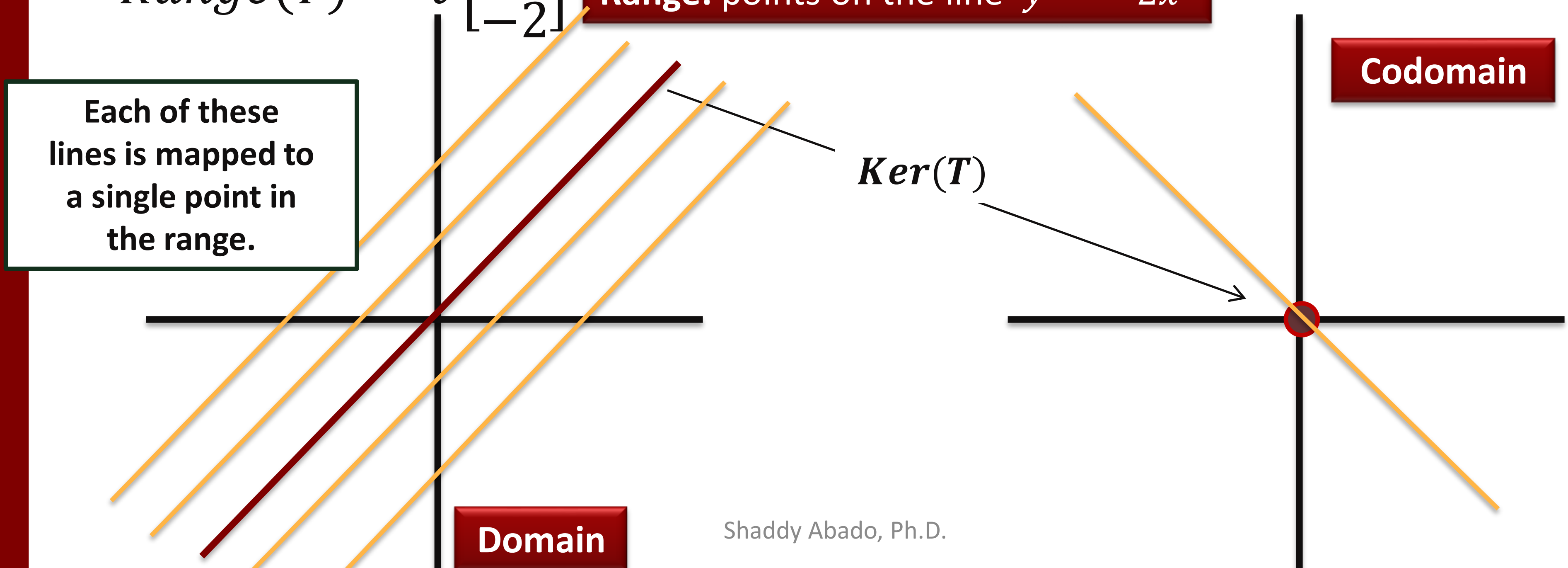
$$Ker(T) = t\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Nullspace: points on the line $y = x$

$$Range(T) = t\begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

Range: points on the line $y = -2x$





Summary

- The central idea of linear algebra is to exploit the hidden simplicity of linear functions.

- Every matrix transformation is linear transformation.

There is always a matrix $\boldsymbol{A}$ hiding behind transformation $\boldsymbol{T}$.

Summary

- A Matrix carries all the essential information on linear transformation. If the basis is known, and the matrix is known, then the transformation of every vector is known.

- For any linear transformation $\boldsymbol{T}$ we can find a matrix $\boldsymbol{A}$ so that $T(\boldsymbol{v}) = A\boldsymbol{v}$. If the transformation is invertible, the inverse transformation has the matrix $\boldsymbol{A}^{-1}$.

- The product of two transformations $T_1\!: \boldsymbol{v} \rightarrow \boldsymbol{A_1 v}$ and $T_2\!: \boldsymbol{w} \rightarrow \boldsymbol{A_2 w}$ corresponds to the product $\boldsymbol{A_2 A_1}$ of their matrices. This is where matrix multiplication came from!

BREAK

THE UNIVERSITY OF CHICAGO

VISUALIZATION OF LINEAR TRANSFORMATIONS

THE UNIVERSITY OF CHICAGO

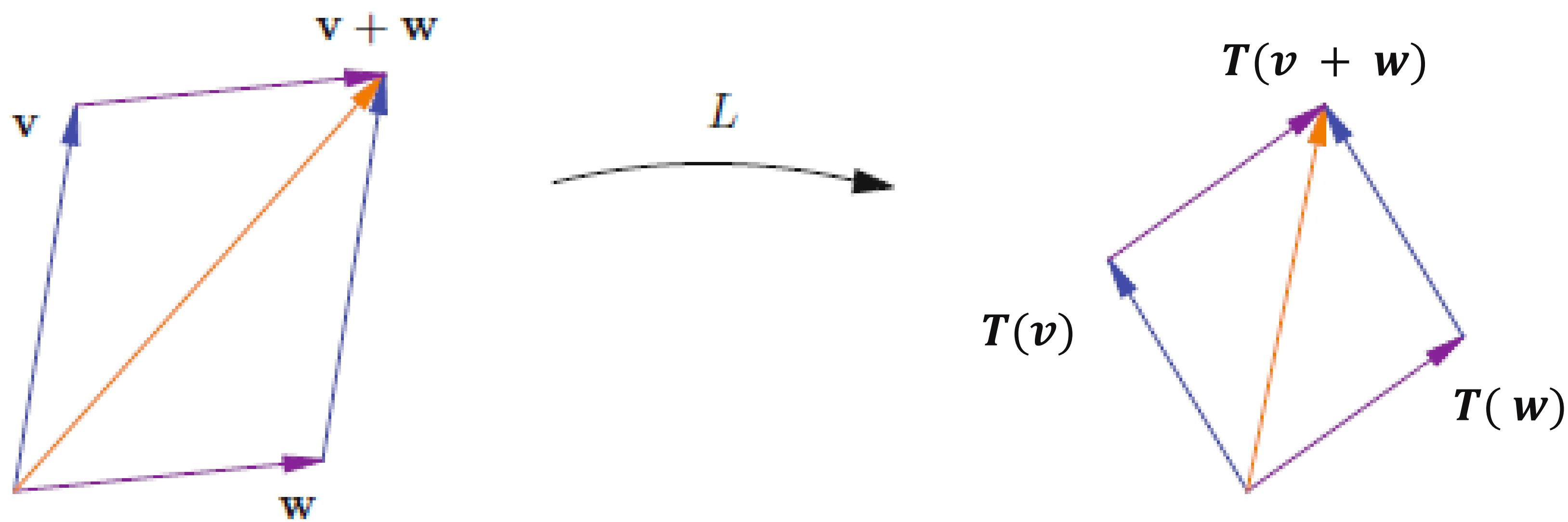
Analytical Geometry

Analytical Geometry is concerned with defining and representing geometrical shapes in a numerical way and extracting numerical information from shapes' numerical definitions and representations.

Linear transformations includes most of the useful transformations of analytical geometry: stretchings, projections, reflections, rotations, and combinations of these.

Linear Function on Euclidean Space

Rhombus **Square**



$$T(v + w) = T(v) + T(w)$$

Computer Graphics

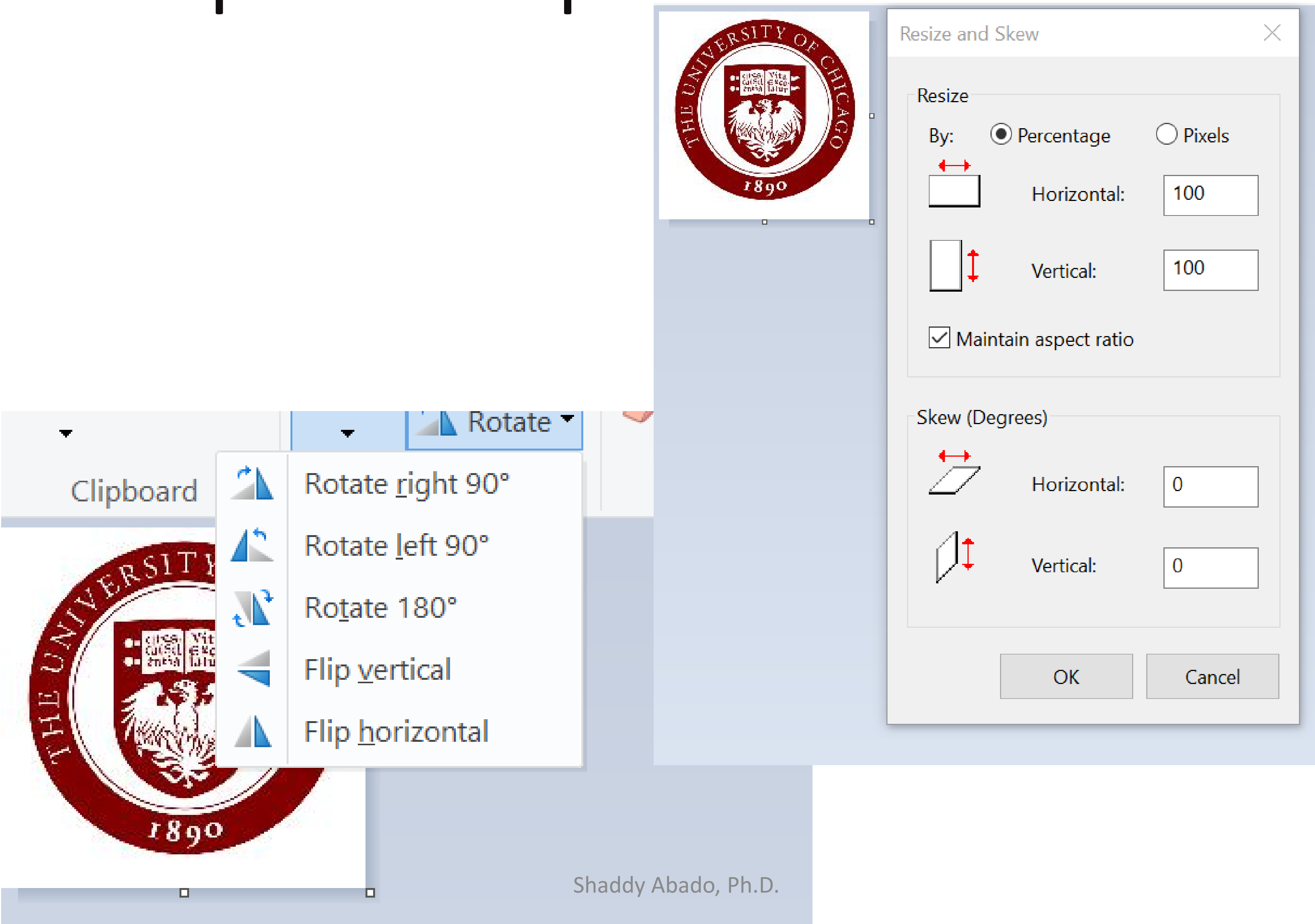
Computer graphic deals with manipulating images digitally. For example, images can:

- Change scale
- Rotate
- Projected into lower dimension
- Etc.

Linear transformations can be used to change an image's:

- Size (scaling): by m in all or directions or by different factors m_1, m_2 in different directions.
- Orientation (rotation): e.g., Around an axis through the origin.
- Projection: onto a plane through the origin.

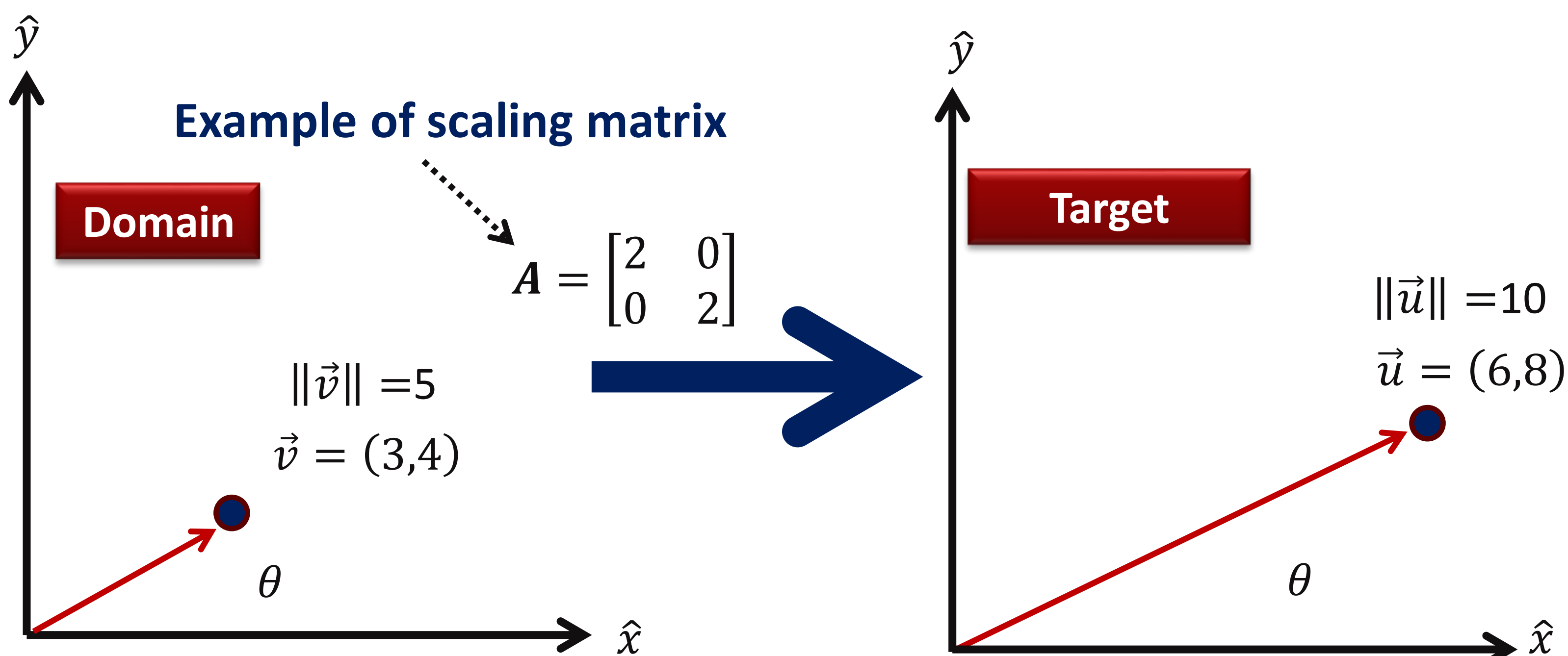
Computer Graphics



Scaling Matrix

Scaling Matrix –
Leaves the direction of the vector unchanged, but changes its length

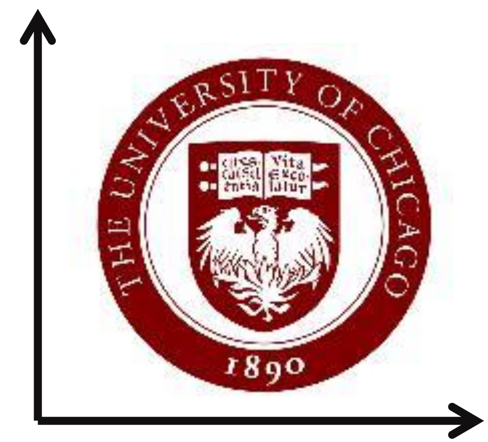
$$x = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \rightarrow A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \rightarrow Ax = \begin{bmatrix} 6 \\ 8 \end{bmatrix}$$

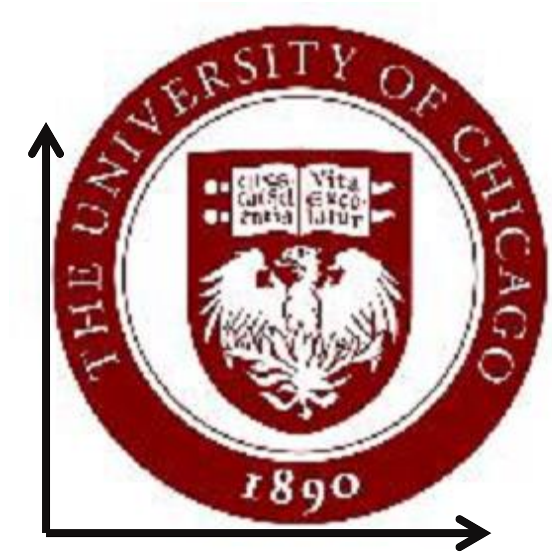


Scaling Matrix

$> 1 \rightarrow$ ***Stretch***

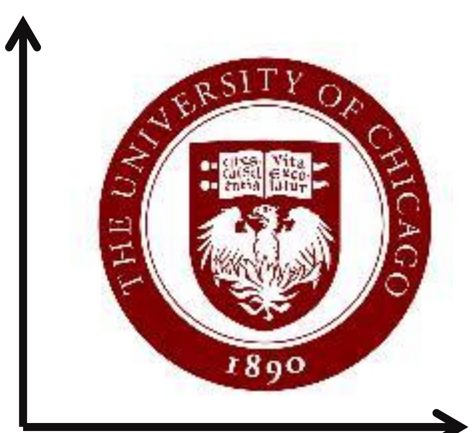
$< 1 \rightarrow$ ***Shrink***





Uniform scale

$$\begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}$$



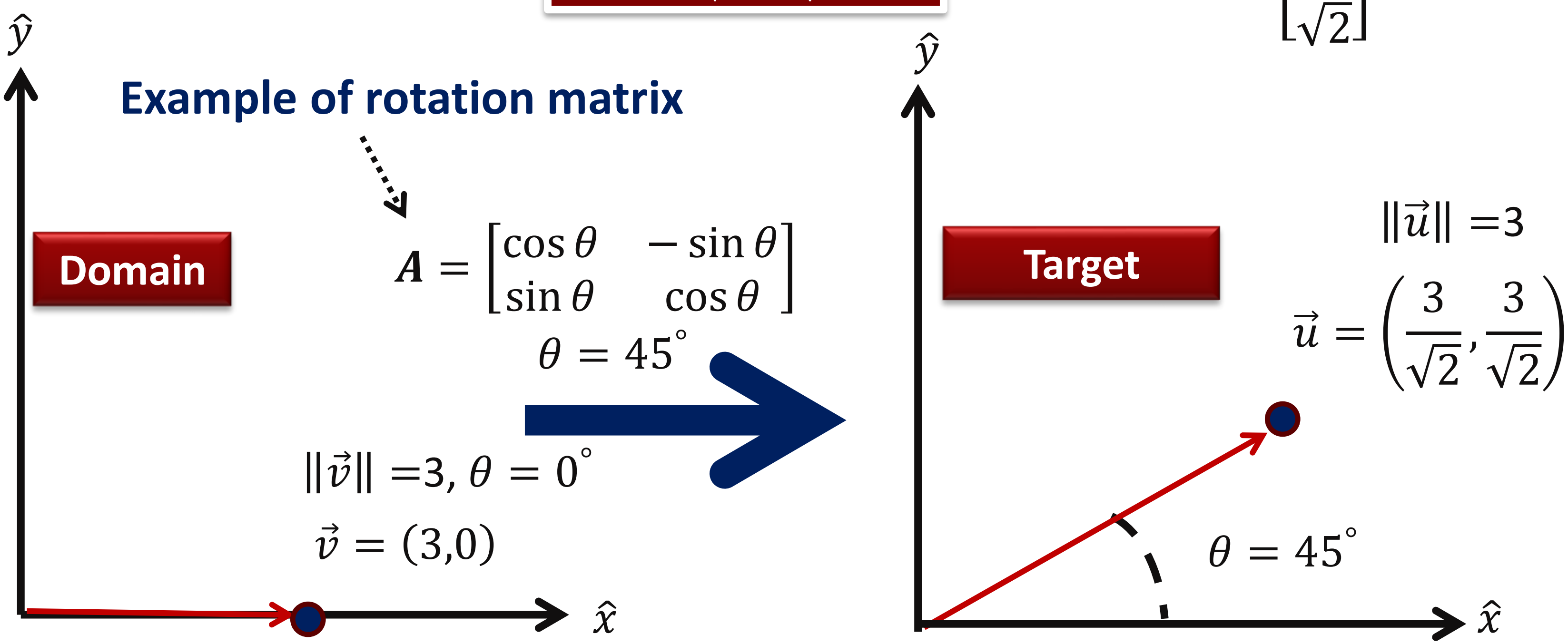
Nonuniform scale (Dilation)

$$\begin{bmatrix} 1.5 & 0 \\ 0 & 0.8 \end{bmatrix}$$

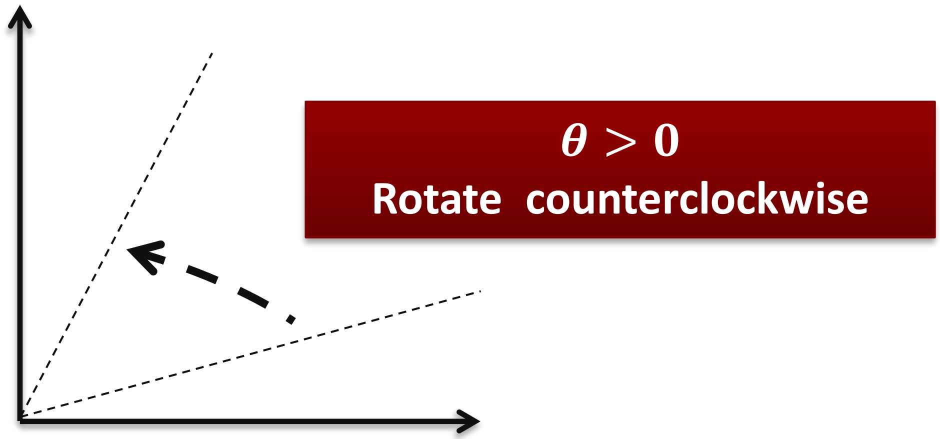
Rotation Matrix

Rotation Matrix –
Changes the direction of vector, but leaves its norm unchanged

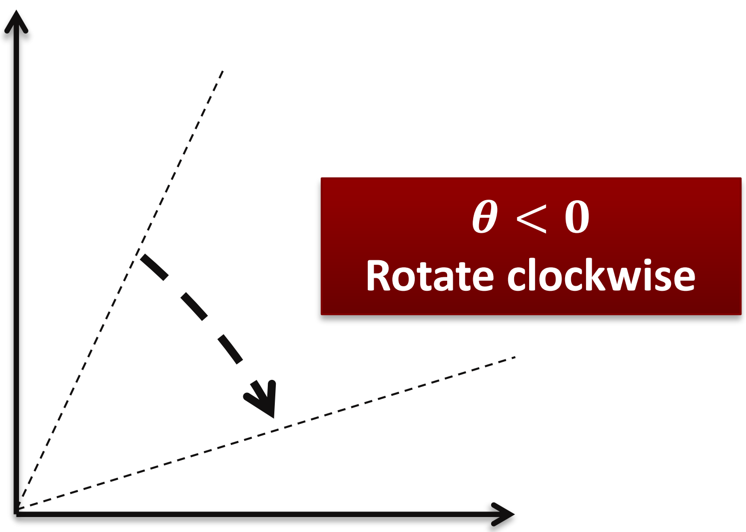
$$x = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \rightarrow A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \rightarrow Ax = \begin{bmatrix} \frac{3}{\sqrt{2}} \\ \frac{3}{\sqrt{2}} \end{bmatrix}$$



Rotation Matrix



$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

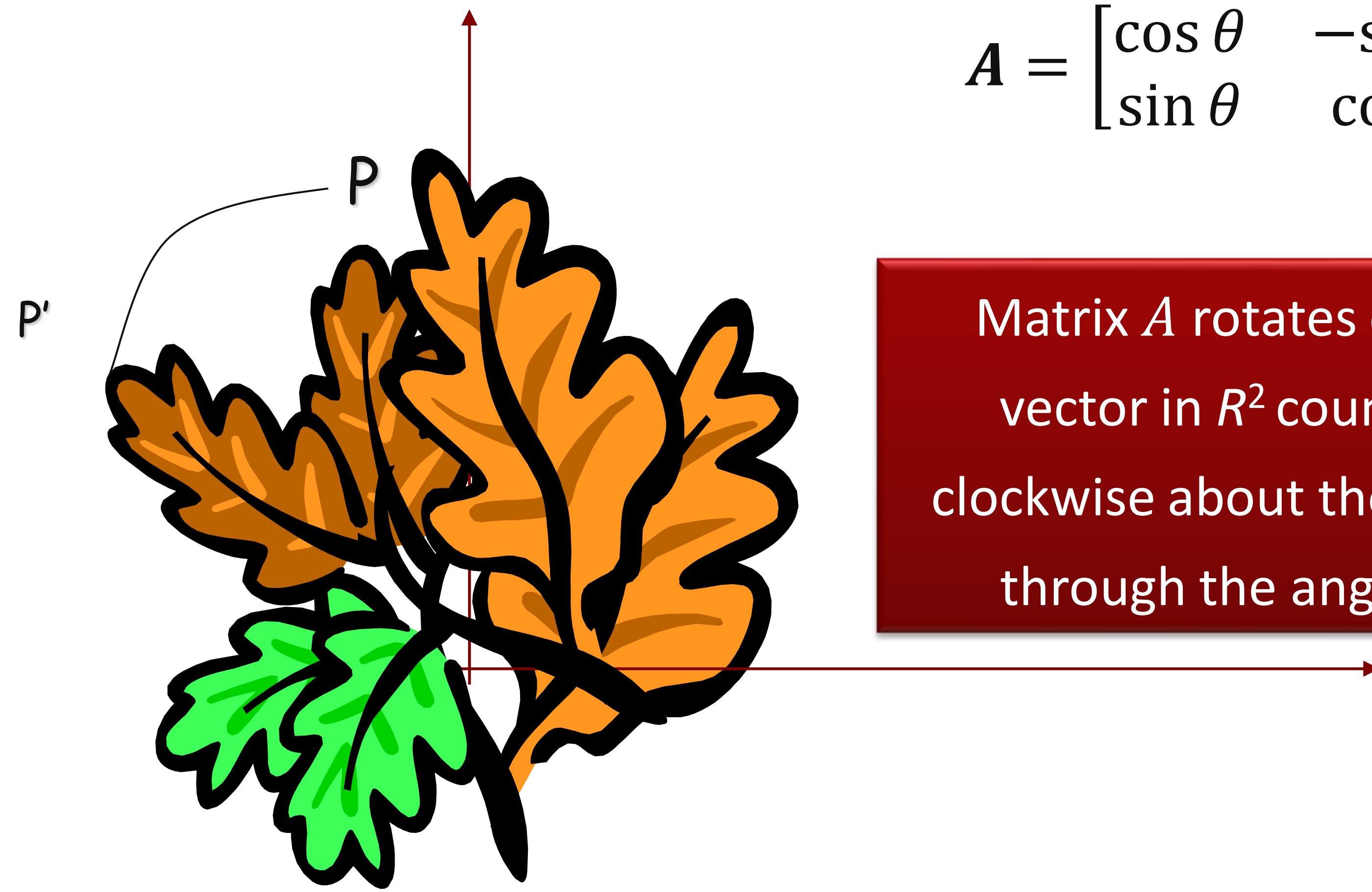


$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

The rotation matrix is:

1. orthogonal matrix
2. Determinant = +1

Rotation Matrix



$$T : R^2 \to R^2$$

$$\boldsymbol{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Matrix A rotates every vector in R^2 counter-clockwise about the origin through the angle θ.

Rotation Matrix - Three Dimensional

XY rotation matrix:

$$\begin{bmatrix} cos\theta & -sin\theta & 0 \\ sin\theta & cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

XZ rotation matrix:

$$\begin{bmatrix} cos\theta & 0 & sin\theta \\ 0 & 1 & 0 \\ -sin\theta & 0 & cos\theta \end{bmatrix}$$

YZ rotation matrix:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\theta & -sin\theta \\ 0 & sin\theta & cos\theta \end{bmatrix}$$

The rotation matrix is:

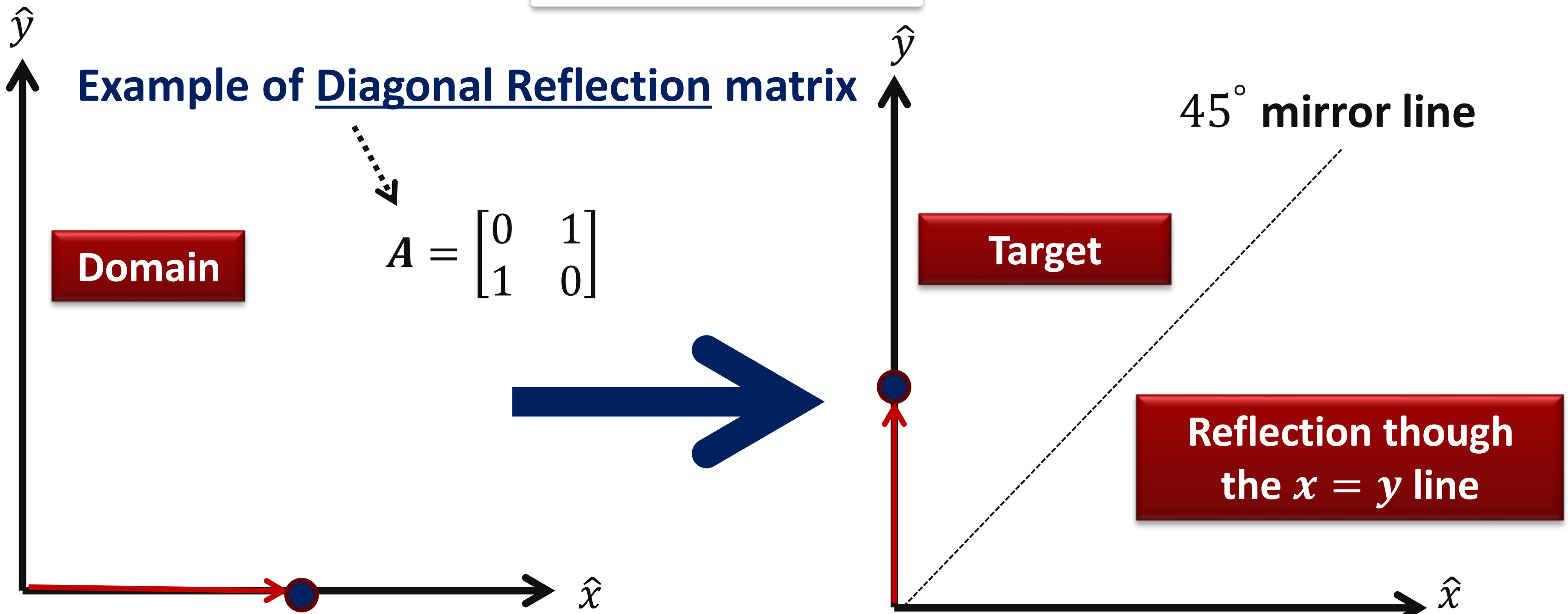
1. orthogonal matrix
2. Determinant = +1

Reflection Matrix

Reflection matrices:
orthogonal matrix
& Determinant = -1

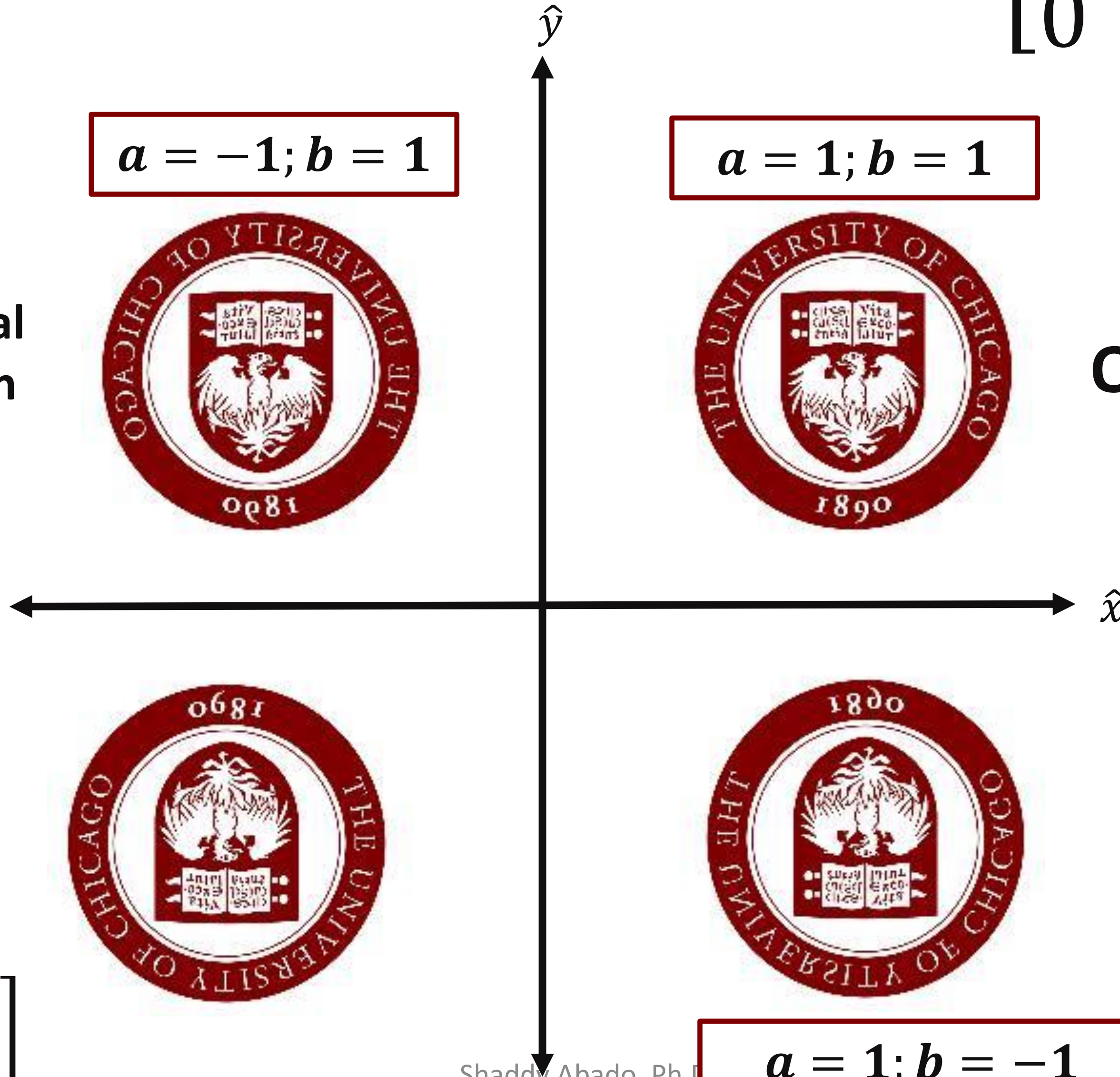
Reflection Matrix –
Reflects a vector across one or more coordinate axis

$$x = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \rightarrow \boxed{A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}} \rightarrow Ax = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$



Reflection Matrix

$$\boldsymbol{A} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$$



$\boldsymbol{a = -1; b = 1}$

Horizontal reflection

$\boldsymbol{a = 1; b = 1}$

Original

Vertical & Horizontal reflection

$$\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$$

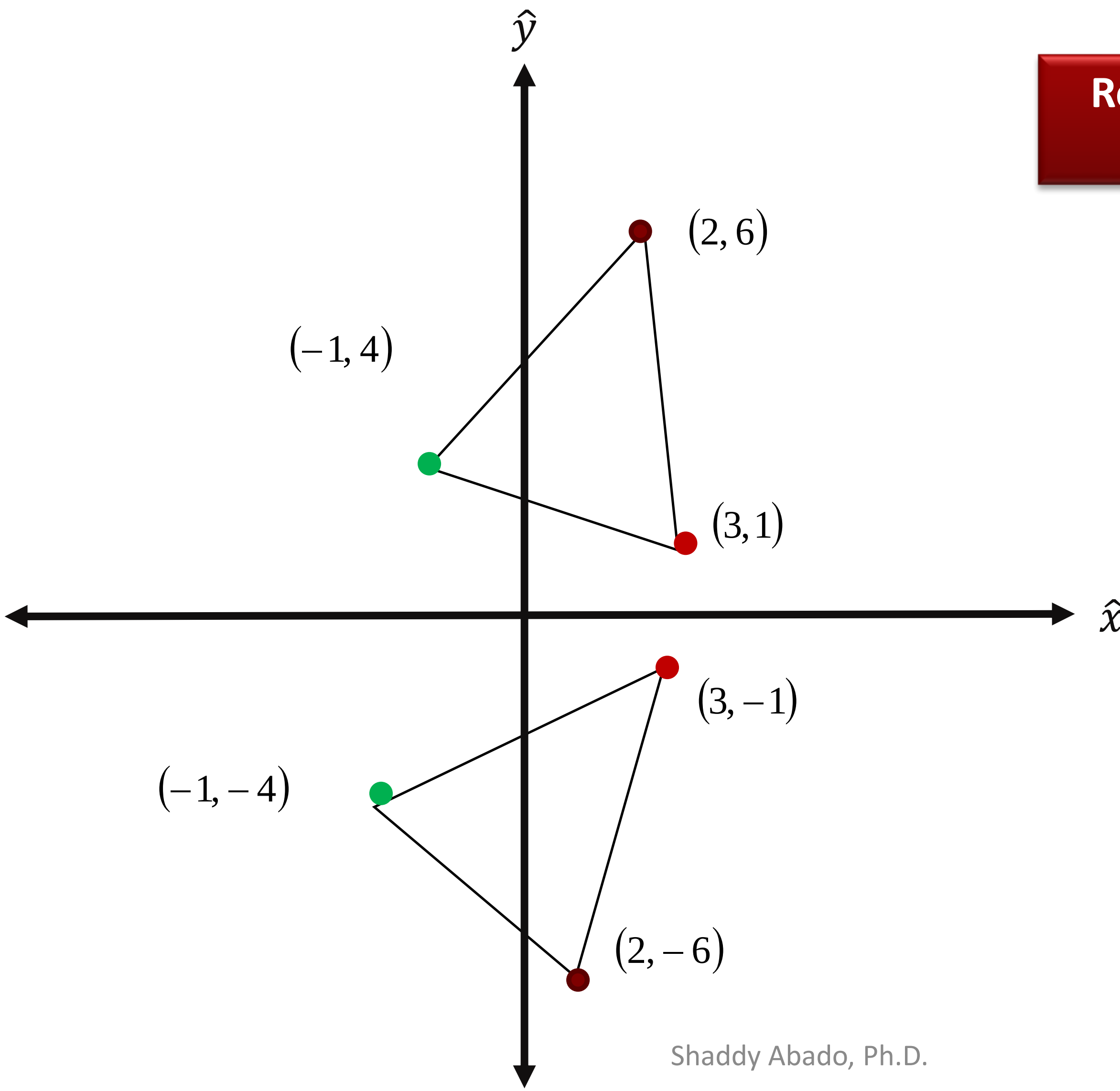
Vertical reflection

$\boldsymbol{a = 1; b = -1}$

Reflection Matrix

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

Reflection though the x axis



$$L\left(\begin{bmatrix} -1 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ -4 \end{bmatrix},$$

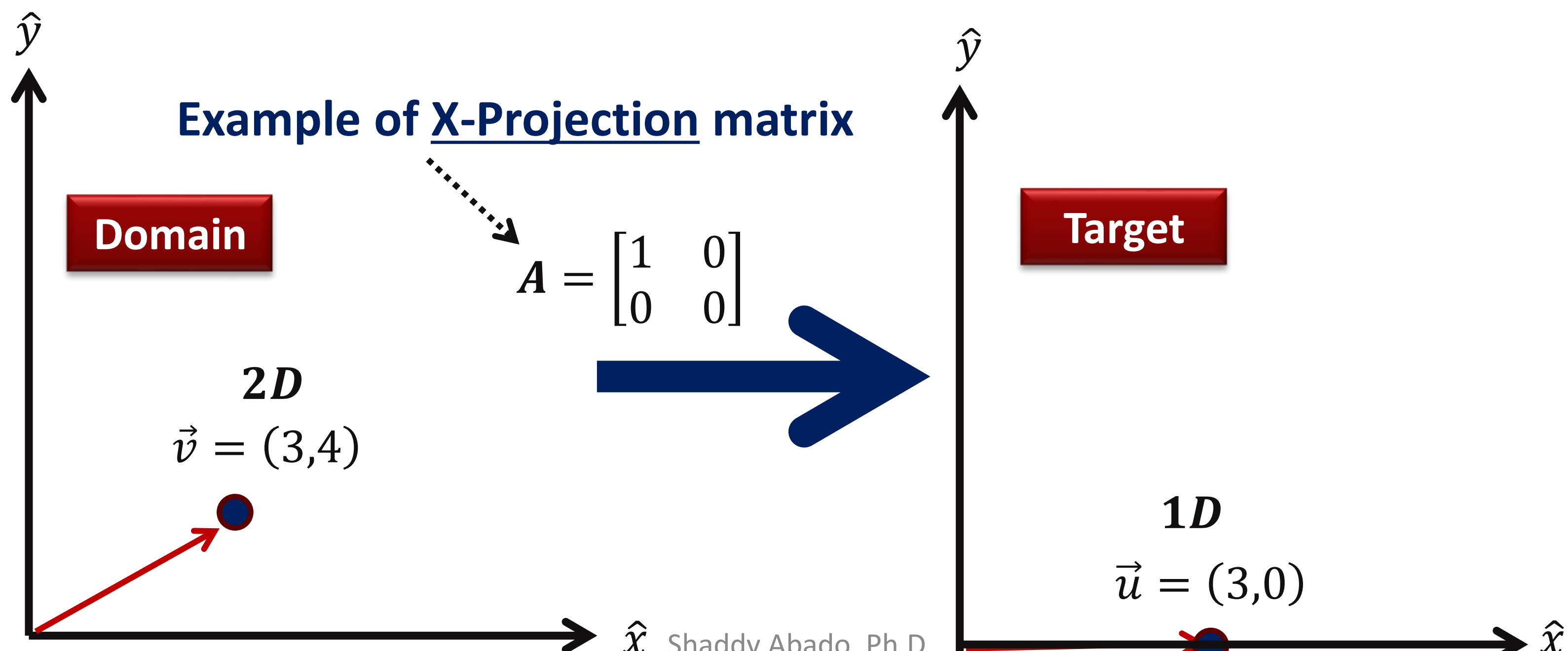
$$L\left(\begin{bmatrix} 3 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ -1 \end{bmatrix},$$

$$L\left(\begin{bmatrix} 2 \\ 6 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ -6 \end{bmatrix}$$

Projection Matrix

Projection Matrix –
Takes a vector into lower dimensional subspace (e.g., line, plane)

$$x = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \longrightarrow \boxed{A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}} \longrightarrow Ax = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$



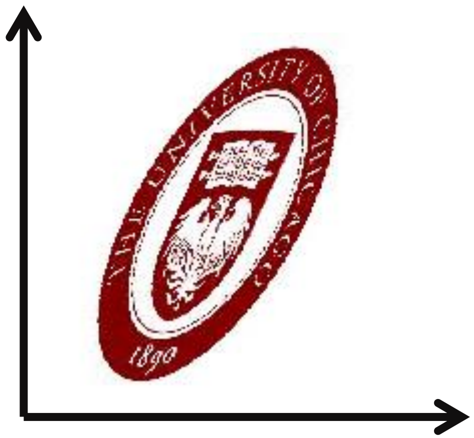
Shear Matrix

Horizontal shear (parallel to the x-axis) by a factor m

y coordinates are unaffected, but x coordinates are translated linearly with y

$$A = \begin{bmatrix} 1 & m \\ 0 & 1 \end{bmatrix}$$

$$x' = x + m * y$$
$$y' = y$$

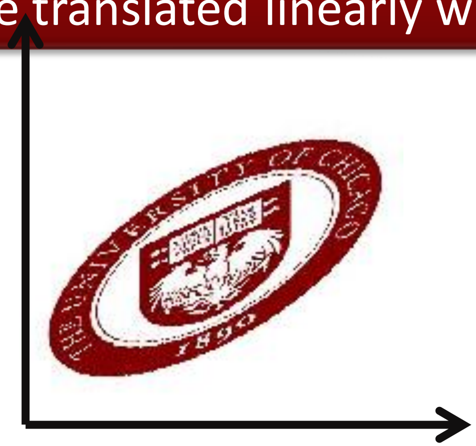


Vertical shear (parallel to the y-axis) by a factor m

x coordinates are unaffected, but y coordinates are translated linearly with x

$$A = \begin{bmatrix} 1 & 0 \\ m & 1 \end{bmatrix}$$

$$x' = x$$
$$y' = y + m * x$$



A few more notes …

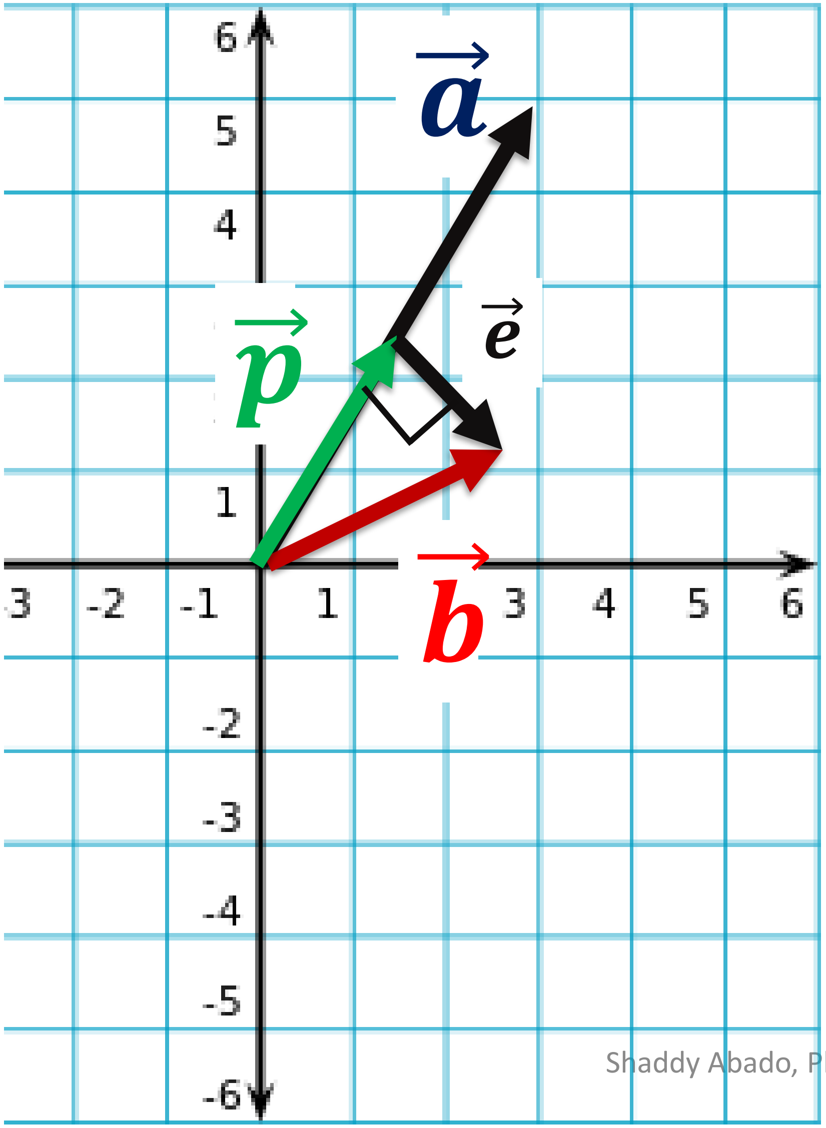
- Additional transformations are available. For example:
  - Image Downsampling and Upsampling
  - Image Cropping
  - Translation (position shifting through the origin to another point)
  - Permutation

- Different matrix transformations can be combined by applying them one after another. (i.e., Composition of Linear Transformations)
  - For example: First rotate through an angle of 45° counter-clockwise, then scale by a factor of 1/2 horizontally, and then rotates back through an angle of 45° clockwise.
  - The combined effect of all three transformations is to scale by a factor of 1/2 along a line that is inclined at angle of 45°.
  - All three matrices can be combined into a single matrix that embodies this overall transformation.

EXTRA SLIDES

THE UNIVERSITY OF CHICAGO

* An Orthogonal Line Projection



Scalar

$$\vec{b} = \vec{e} + \hat{x}\vec{a}$$

$\vec{e}$ is orthogonal to $\vec{a}$ if and only if:

$$0 = \vec{e} \cdot \vec{a}$$

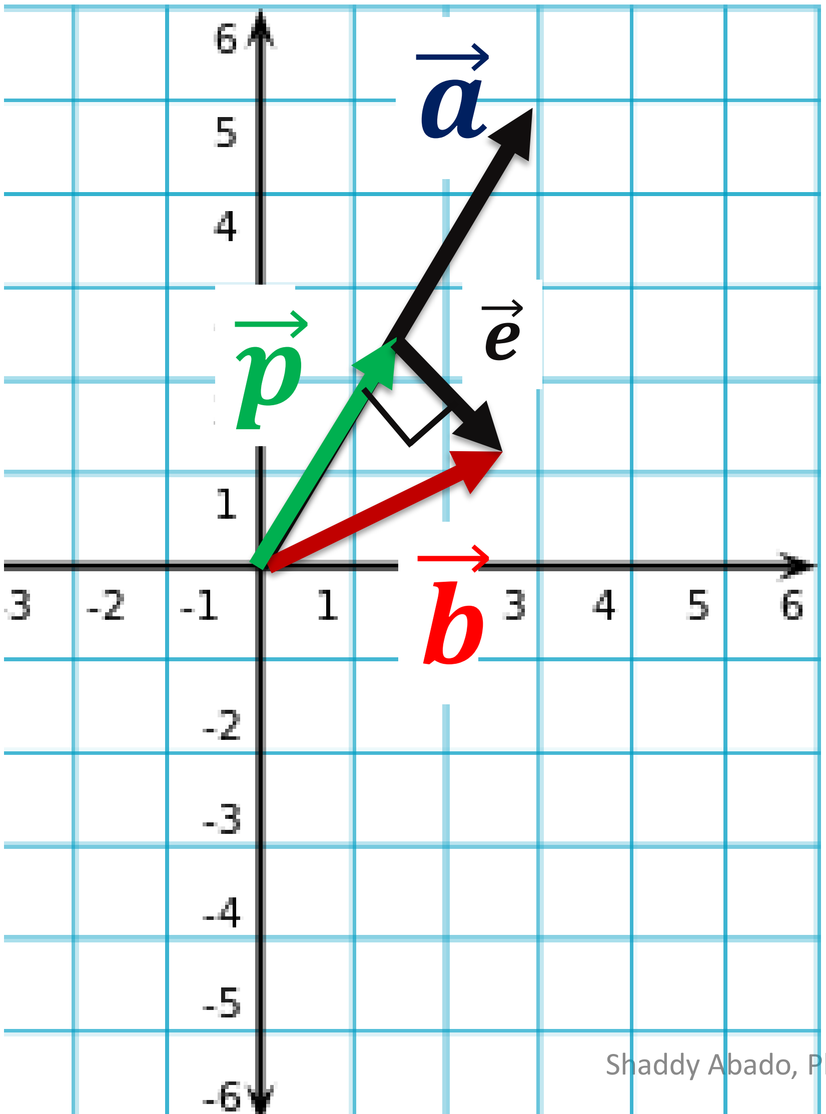
$$0 = \left(\vec{b} - \hat{x}\vec{a}\right) \cdot \vec{a}$$

$$0 = \vec{a} \cdot \vec{b} - \hat{x}\vec{a} \cdot \vec{a}$$

$$\hat{x} = \frac{\vec{a} \cdot \vec{b}}{\vec{a} \cdot \vec{a}}$$

$$\hat{x} = \frac{a^T b}{a^T a}$$

An Orthogonal Line Projection



$$\vec{b} = \vec{e} + \vec{p}$$
$$= \vec{e} + \frac{a^T b}{a^T a}\vec{a}$$

Right Triangle (Pythagoras theorem)

$$\|b\|^2 = \|p\|^2 + \|e\|^2$$

$$\vec{b} = \vec{e} + \vec{p}$$
$$= \vec{e} + \frac{a^T b}{a^T a}\vec{a}$$

* Projection Matrix

The projection matrix P_{mxm} multiply $\overrightarrow{b_{mx1}}$ to give $\overrightarrow{p_{mx1}}$ (i.e., $\vec{p} = P\vec{b}$)

$$P_{mxm} = \frac{aa^T}{a^T a}$$

Project $\vec{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ onto $\vec{a} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$

$$P_{mxm} = \frac{aa^T}{a^T a} = \frac{1}{9}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}\begin{bmatrix} 1 & 2 & 2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}$$

$$\vec{p} = P\vec{b} = \frac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 5 \\ 10 \\ 10 \end{bmatrix}$$

* Orthogonal Projection onto subspace

$$\vec{b} = \vec{e} + \vec{p}$$
$$\vec{b} = \vec{e} + A\hat{x}$$

Vector

is orthogonal to A if and only if:

b_{mx1}
e_{mx1}
$\hat{x}_{nx1}$
p_{mx1}
A_{mxn}
P_{mxm}

$$\boxed{\begin{aligned} 0 &= A^T e \\ 0 &= A^T(b - \hat{x}A) \\ 0 &= A^T b - \hat{x}A^T A \\ \hat{x} &= \left(A^T A\right)^{-1} A^T b \end{aligned}}$$

$$\vec{p} = A\left(A^T A\right)^{-1} A^T b$$

Projection of b onto subspace A

$$P = A\left(A^T A\right)^{-1} A^T$$

Projection Matrix

* Normal Equation for a Straight line

$y_i = a * x_i + b$

$$\boldsymbol{A^T A \hat{x} = A^T b}$$

$$\boldsymbol{\hat{x} = \left(A^T A\right)^{-1} A^T b}$$

$$\begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & N \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum x_i y_i \\ \sum y_i \end{bmatrix}$$

$\boldsymbol{A^T A}$ $\boldsymbol{\hat{x}}$ $\boldsymbol{A^T b}$