Unit 8

Linearity

Albert Sung

Outlines

- 8.1 Linearity
- 8.2 Matrix Multiplication
- 8.3 Geometric Transformations
- 8.4 Regression Model

Unit 8.1

Linearity

Mapping of a Vector

- In this unit, we consider f: $\mathbb{R}^n \rightarrow \mathbb{R}^m$.

$$y = f(x),$$

  where x is an n-vector and y is an m-vector.

- In general, f is a *vector-valued* function.
  - i.e., the output y is a vector.
- In the special case when $m = 1$, f is a scalar-valued function.
  - i.e., the output y is a scalar.

Linearity: An Illustration

mori soba = 800 yen

sushi = 200 yen

- Exchange rate: 100 yen = HK$ 7.00
- How much are 2 plates of mori soba and 4 plates of sushi in HK dollars?

Linear Functions

- A function f is linear if it satisfies the following two conditions:

1. Additivity
$$f(x+y) = f(x) + f(y)$$

2. Scaling
$$f(cx) = cf(x)$$

- Putting the two properties together,

Superposition
$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

- A function f is linear if it satisfies the superposition property.

In our example, f is the function to convert from yen to HK dollars.

Zero-in Zero-out Property

- A linear function f must have
$$f(\mathbf{0}) = \mathbf{0}.$$
- Why?
  - Hint: Use scaling property.
- Zero-in zero-out is a necessary condition for f to be linear.
- Is it sufficient? Why?

Class Exercises

- Is each of the following scalar-valued functions linear?

a) $f(x) = \mathbf{avg}(x) = \frac{x_1 + x_2 \ldots + x_n}{n}$

b) $f(x) = \mathbf{min}(x) = \min\{x_1, x_2, \ldots, x_n\}$

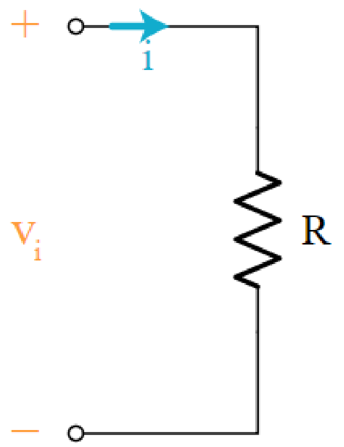
c) $f(x) = x_1 - x_2 + 4$

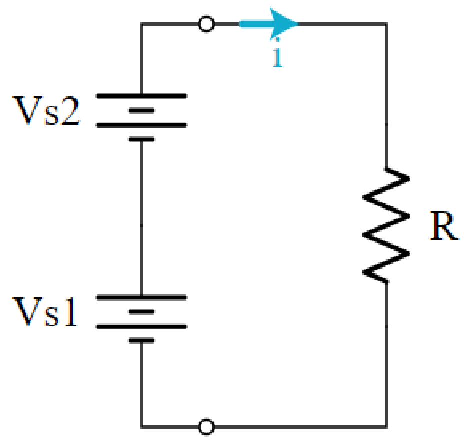
Application: Ohm's Law

- The current is given by Ohm's Law:

$$i = f(v_i) = \frac{v_i}{R}.$$

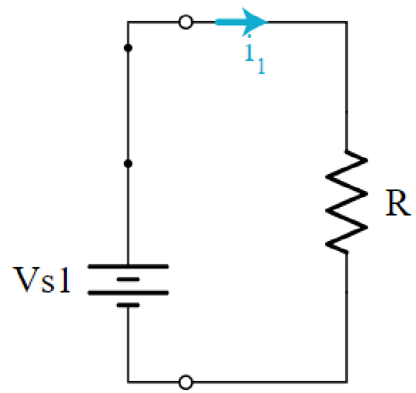
  - The resistor can be viewed as a function that takes in a voltage and outputs a current.
- It is a *linear* function.
- The figure on the right is a toy problem to demonstrate the use of *superposition principle*.

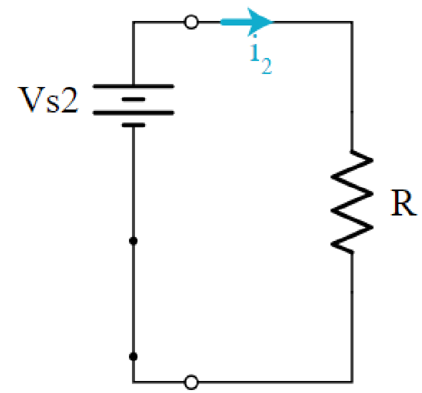




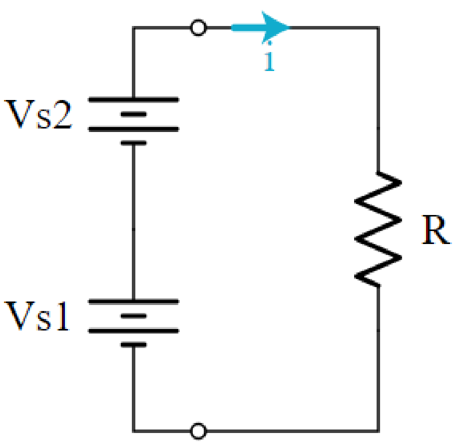
Using Superposition

- $f(V_{s1}) = \dfrac{V_{s1}}{R}$
- $f(V_{s2}) = \dfrac{V_{s2}}{R}$





- $$\begin{aligned} i &= f(V_{s1} + V_{s2}) \\ &= f(V_{s1}) + f(V_{s2}) \\ &= \frac{V_{s1}}{R} + \frac{V_{s2}}{R} \end{aligned}$$



Although $f(V_{s1} + V_{s2})$ can be computed directly in this example, we demonstrate the use of superposition.

Unit 8.2

Matrix Multiplication

Matrix-Vector Multiplication

- Let A be an $m \times n$ matrix and x be an n-vector.
- Then

$$Ax = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{bmatrix},$$

where a_i^T is the i-th row of A.

Alternative Representation

- $$Ax = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

Matrix Multiplication is Linear

- Consider a function
$$f(x) = Ax,$$
where A is a matrix and x is a vector.
- It is easy to check that superposition holds:
$$\begin{aligned} f(\alpha x + \beta y) &= A(\alpha x + \beta y) \\ &= \alpha Ax + \beta Ay \qquad \text{(matrix algebra)} \\ &= \alpha f(x) + \beta f(y). \end{aligned}$$
- Hence, any matrix leads to a linear transformation.

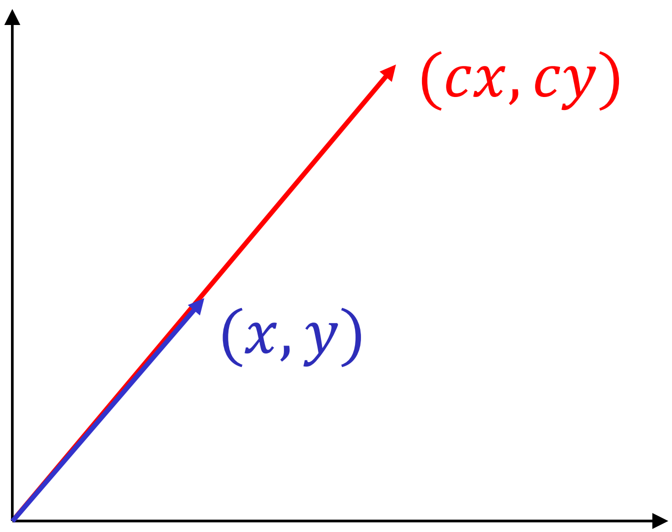
Is the converse true?

Does every linear transformation lead to a matrix?

Examples

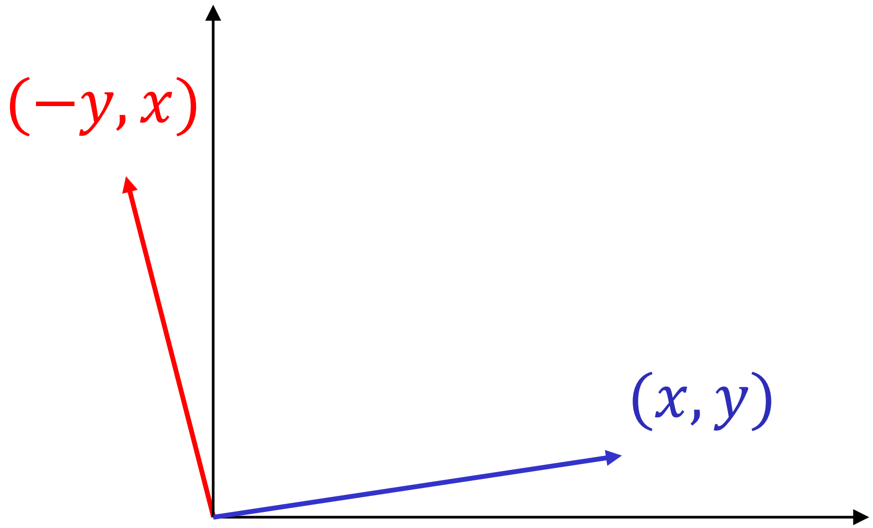
Stretching by the factor *c*

- $A = \begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}$ stretches every vector x.



Rotation by 90^o

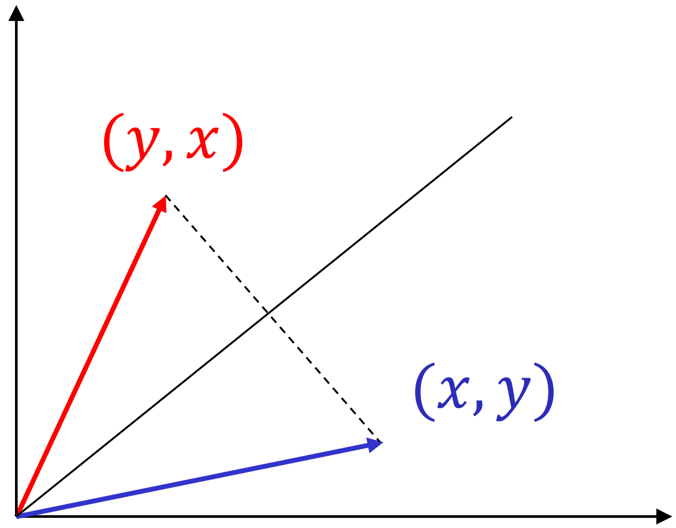
- $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ turns every point (x, y) to $(-y, x)$.



Examples

Reflection across 45^o line

- $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ flips every vector x to its mirror image.



Projection onto x-axis

- $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ projects every point (x, y) to the x-axis.

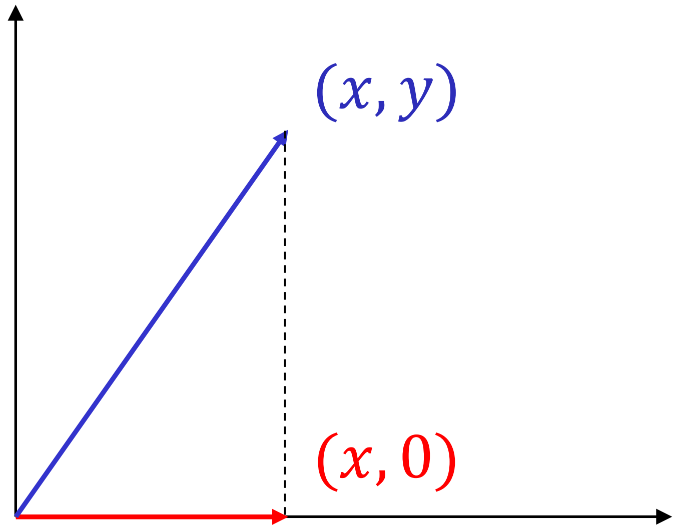
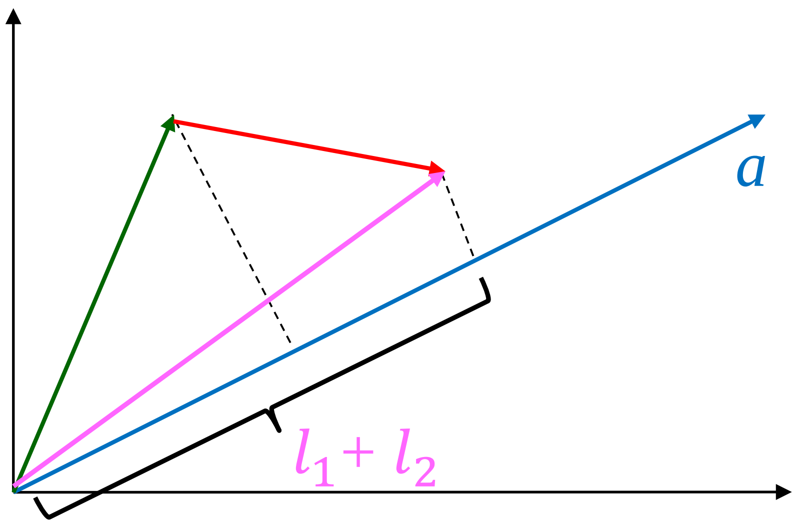
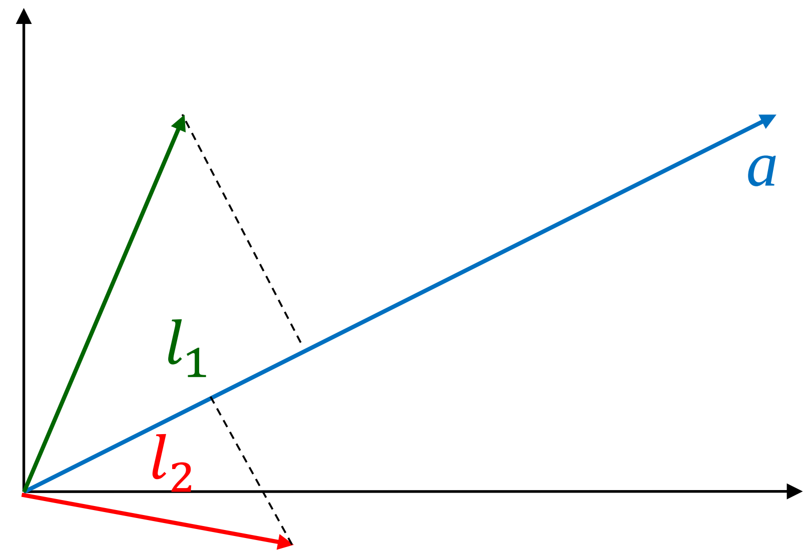


Illustration: Projection is Linear





Add two vectors and then project onto a.

$\equiv$

Project two vectors onto a and then add them.

Representation of Linear Functions

If $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear function, then it can be represented as

$$f(x) = Ax,$$

where A is an $m \times n$ matrix, and x is an n-vector.

- To see this, we write x as $x_1 e_1 + x_2 e_2 + \cdots + x_n e_n$.
- By superposition,

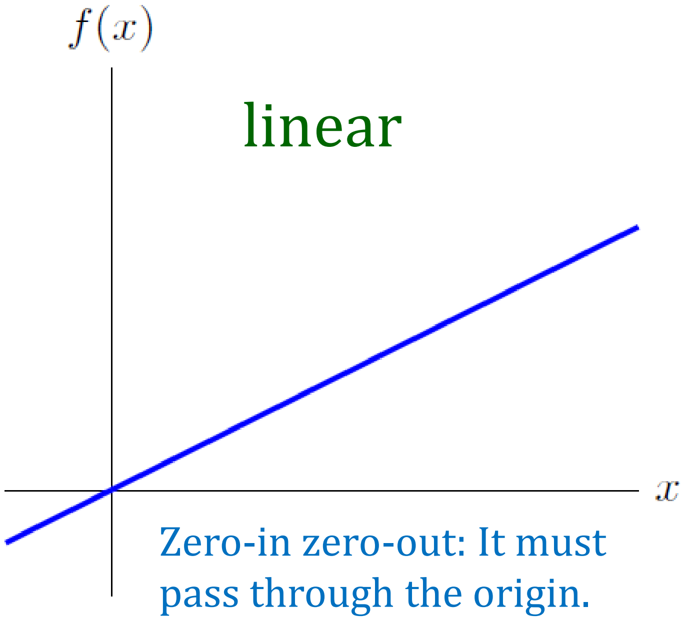
$$f(x) = x_1 f(e_1) + x_2 f(e_2) + \cdots + x_n f(e_n)$$

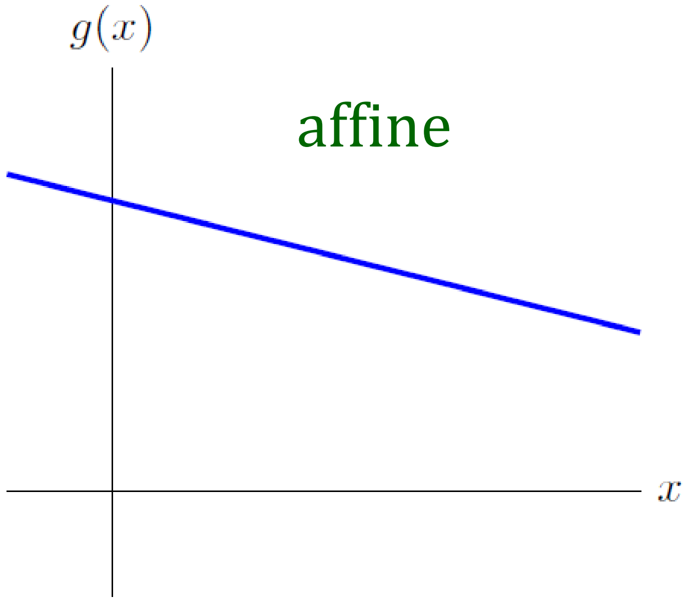
$$= \begin{bmatrix} | & | & \cdots & | \\ f(e_1) & f(e_2) & \ddots & f(e_n) \\ | & | & \cdots & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Affine Functions

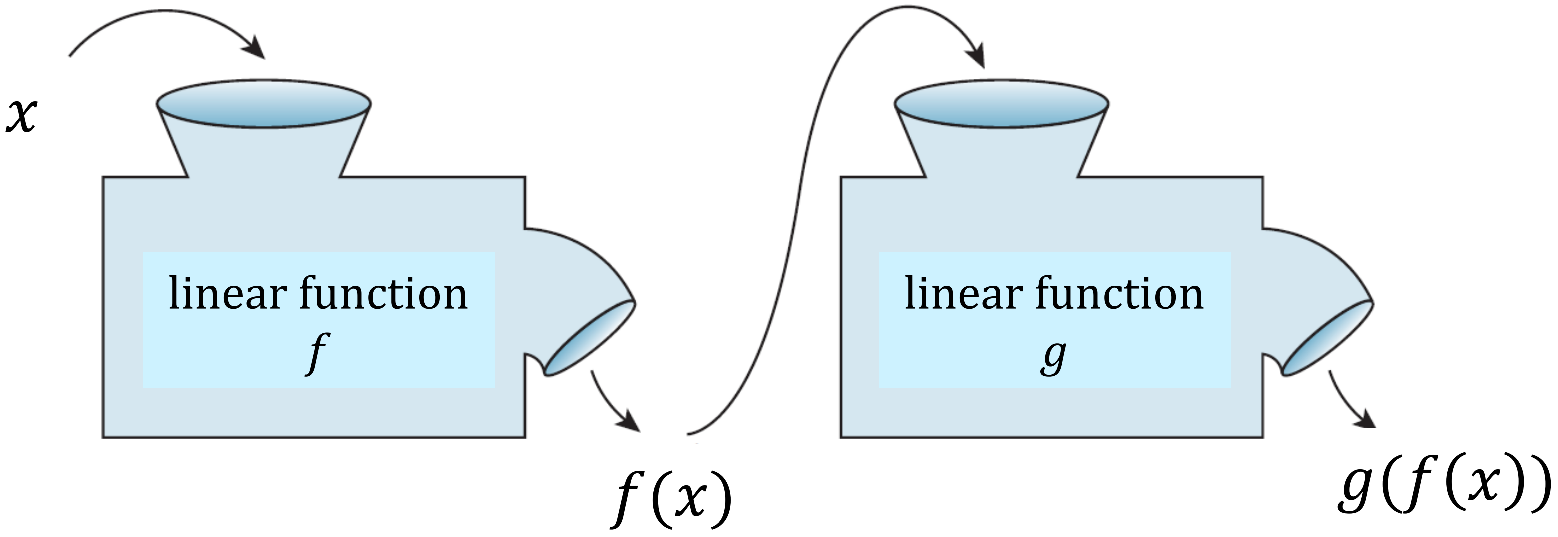
- A function g is affine if it is the sum of a linear function and a constant, i.e.,

$$g(x) = Ax + c.$$





Composition of Linear Functions



- Is the composite function $g \circ f$ linear?

Linearity Preserves under Composition

- Since f and g are linear, they can be expressed as

$$f(x) = Ax \quad \text{and} \quad g(y) = By.$$

- Consider the composition:

$$g\big(f(x)\big) = B(Ax) = BAx = Cx,$$

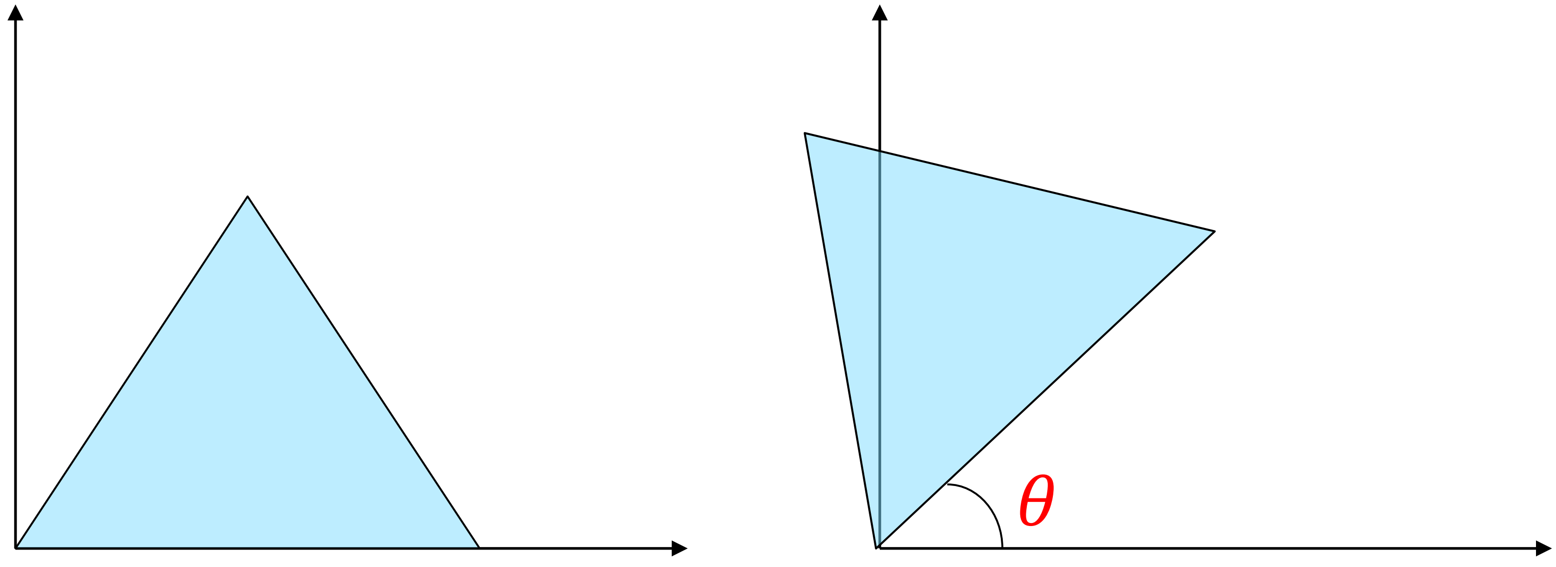
where $C = BA$ is a matrix.

- Hence, $g\big(f(x)\big)$ is a linear function of x.

Unit 8.3

Geometric Transformations

Rotation



- Useful operation in computer graphics.
- Is it a linear transformation?

Linear Mapping via Standard Basis

- Every vector $x \in \mathbb{R}^n$ can be written as

$$x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n.$$

- By superposition, if we know $v_i = Ae_i$ for all i, then we can determine

$$\begin{aligned} Ax &= x_1 A e_1 + x_2 A e_2 + \cdots + x_n A e_n \\ &= x_1 v_1 + x_2 v_2 + \cdots + x_n v_n \\ &= \begin{bmatrix} | & | & \cdots & | \\ v_1 & v_2 & \ddots & v_n \\ | & | & \cdots & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \end{aligned}$$

Example

- Let e_1 and e_2 be the standard basis of $\mathbb{R}^2$.
- Consider a linear function $f: \mathbb{R}^2 \to \mathbb{R}^3$.
- Suppose we know that

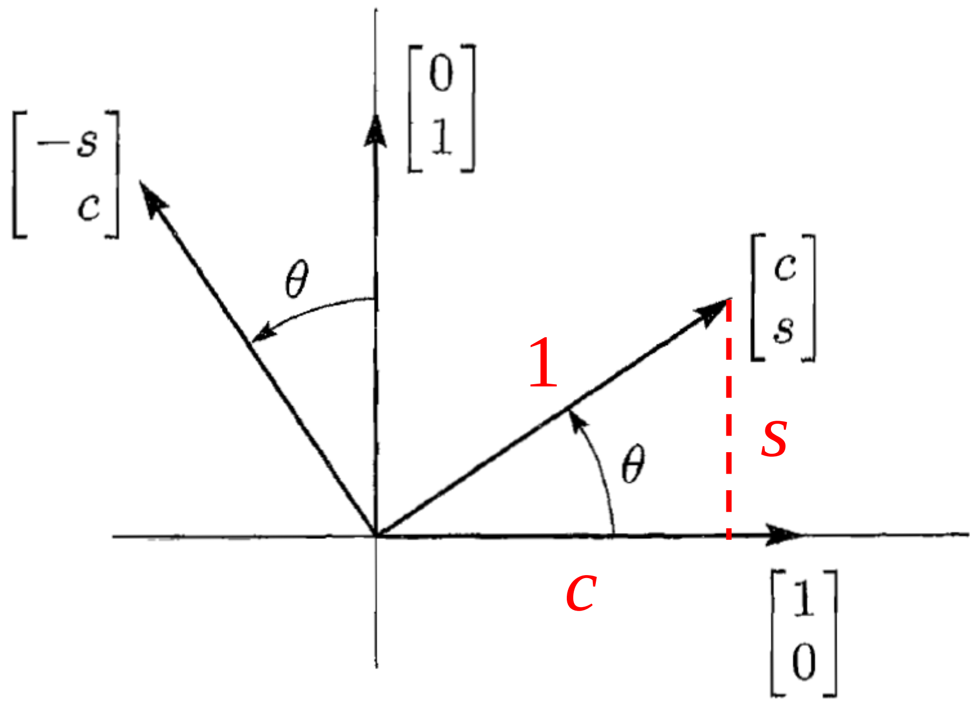
$$f(e_1) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \text{ and } f(e_2) = \begin{bmatrix} 2 \\ 5 \\ 4 \end{bmatrix}.$$

- A vector $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ can be written as $x = x_1 e_1 + x_2 e_2$.
- Hence, $f(x) = x_1 f(e_1) + x_2 f(e_2) = \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = Ax$,

  where $A = \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 3 & 4 \end{bmatrix}$.

Rotation through θ

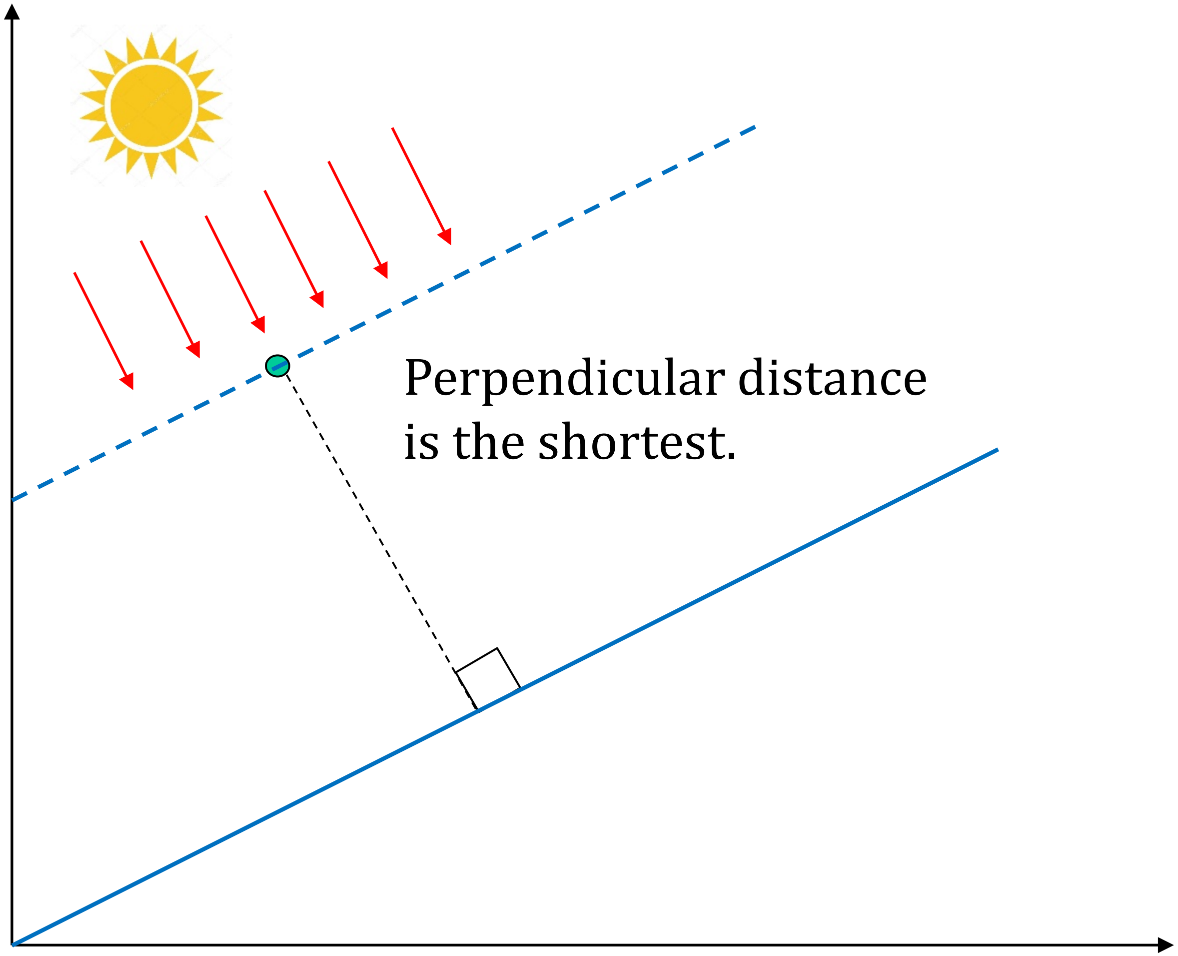
- Let $c = \cos\theta$ and $s = \sin\theta$.



Hence, the rotation matrix is given by

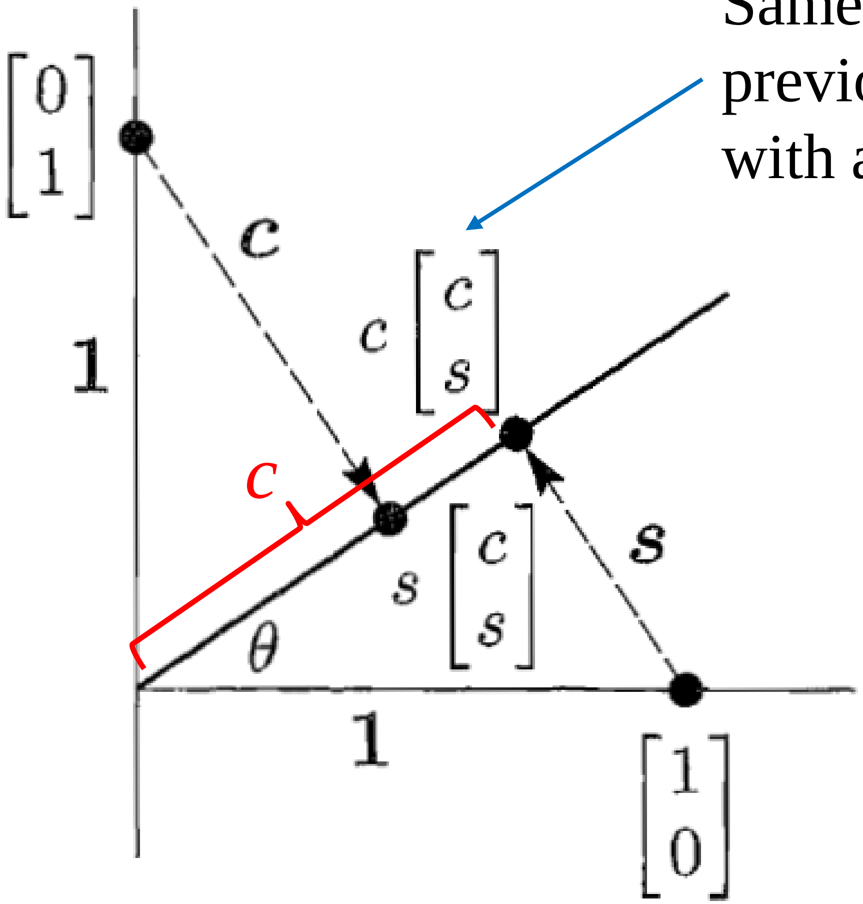
$$R = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}$$

What's the shortest distance between a point and a line?



Perpendicular distance is the shortest.

Projection onto the θ-line



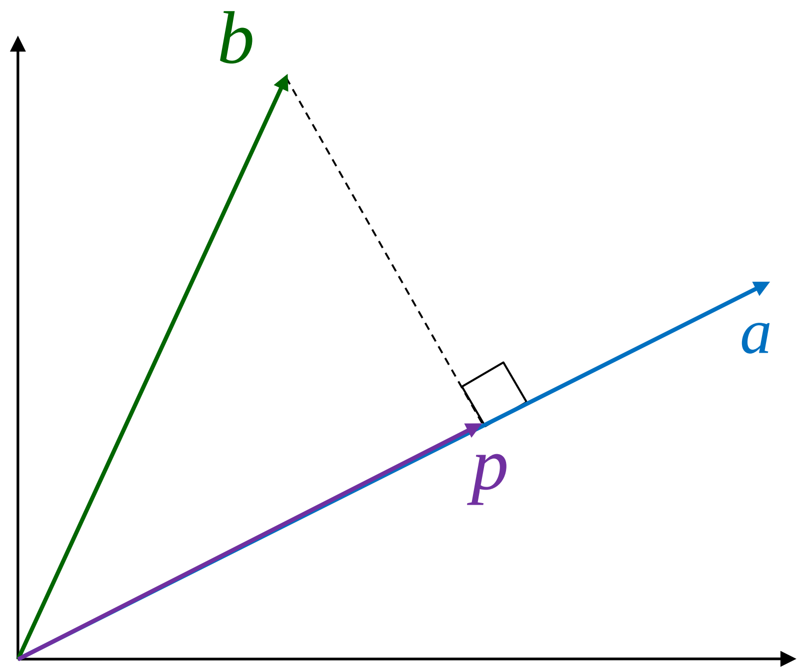
Same as in the previous slide, except with a factor of c.

Hence, the projection matrix is given by

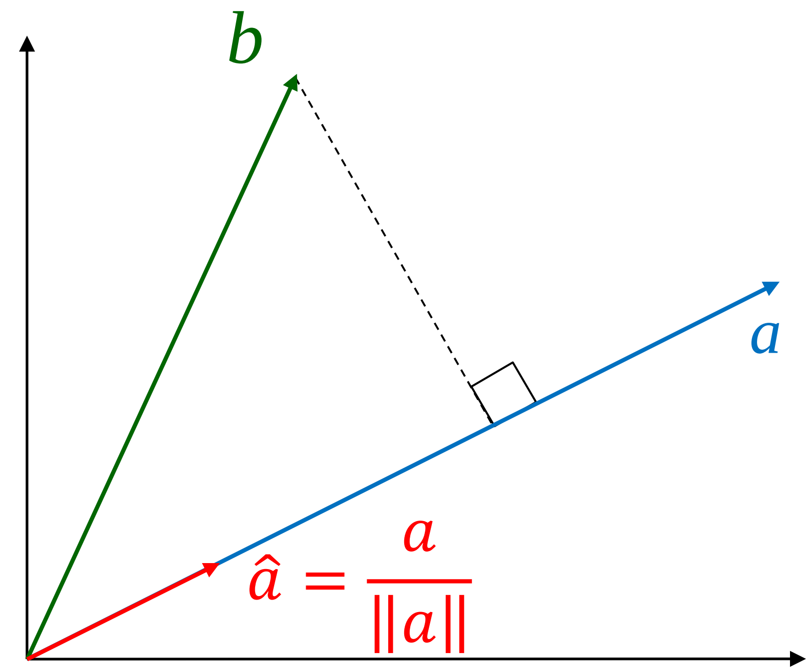
$$P = \begin{bmatrix} c^2 & cs \\ cs & s^2 \end{bmatrix}$$

How about projection in the n-dimensional space?

Projection of b onto a

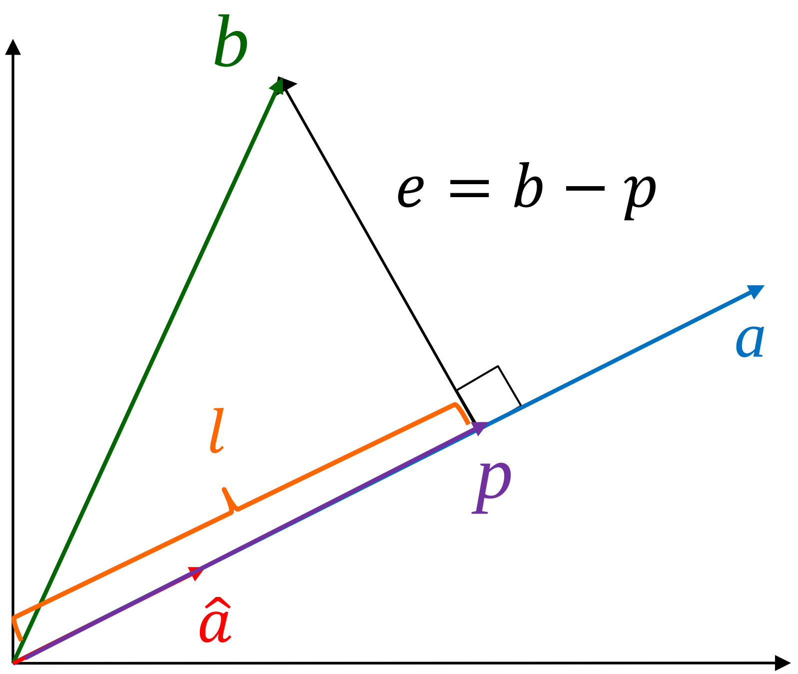


We want to find p.



We normalize a to obtain the unit vector $\hat{a}$, which points to the same direction as a but of unit length.

Projection Length as Inner Project



We want to find the length l, and

$$p = l\hat{a}.$$

- Geometry:
  - $e = (b - p)$ is perpendicular to $\hat{a}$.
- Therefore,

$$(b - l\hat{a})^T\hat{a} = 0,$$

- Expanding,

$$b^T\hat{a} - l = 0.$$

- Rearranging the terms,

$$l = b^T\hat{a}.$$

- Projection vector

$$p = l\hat{a} = (b^T\hat{a})\hat{a}$$

Projection Matrix

- Expressed in terms of a,

$$p = (b^T\hat{a})\hat{a} = \left(\frac{b^T a}{\|a\|}\right)\frac{a}{\|a\|} = \frac{a^T b}{a^T a}a.$$

- Moving a to the front, we obtain

$$p = a\frac{a^T b}{a^T a} = \boxed{\frac{aa^T}{a^T a}}b.$$

The numerator aa^T is a column times a row, which gives a square matrix.

- The projection matrix P is given by

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

The denominator $a^T a$ is a scalar.

Unit 8.4

Regression Model

Regression Model

- In many applications, we want to estimate the relationship between a dependent variable vector y and an independent variable vector, $x = (x_1, x_2, .., x_n)$.
  - This is called regression.
- A common regression model is to use an affine function to approximate the relationship:

$$\hat{y} = \beta^T x + v$$

  - $\hat{y}$: predicted value
  - x: n-vector (called feature vector or regressor vector)
  - β: weight vector or coefficient vector
  - v: offset or intercept

We consider only the case where $\hat{y}$ is a scalar.

Example: House Price Regression

Measurement data

House	x_1 (area)	x_2 (beds)	y (price)	$\hat{y}$ (prediction)
1	0.846	1	115.00	161.37
2	1.324	2	234.50	213.61
3	1.150	3	198.00	168.88
4	3.037	4	528.00	430.67
5	3.984	5	572.50	552.66

Table 2.3 Five houses with associated feature vectors shown in the second and third columns. The fourth and fifth column give the actual price, and the price predicted by the regression model.

- Model: $\hat{y} = \beta^T x + v = \beta_1 x_1 + \beta_2 x_2 + v$
  - y: selling price of a house in some area
  - x_1: house area (in 1000 square feet)
  - x_2: number of bedrooms

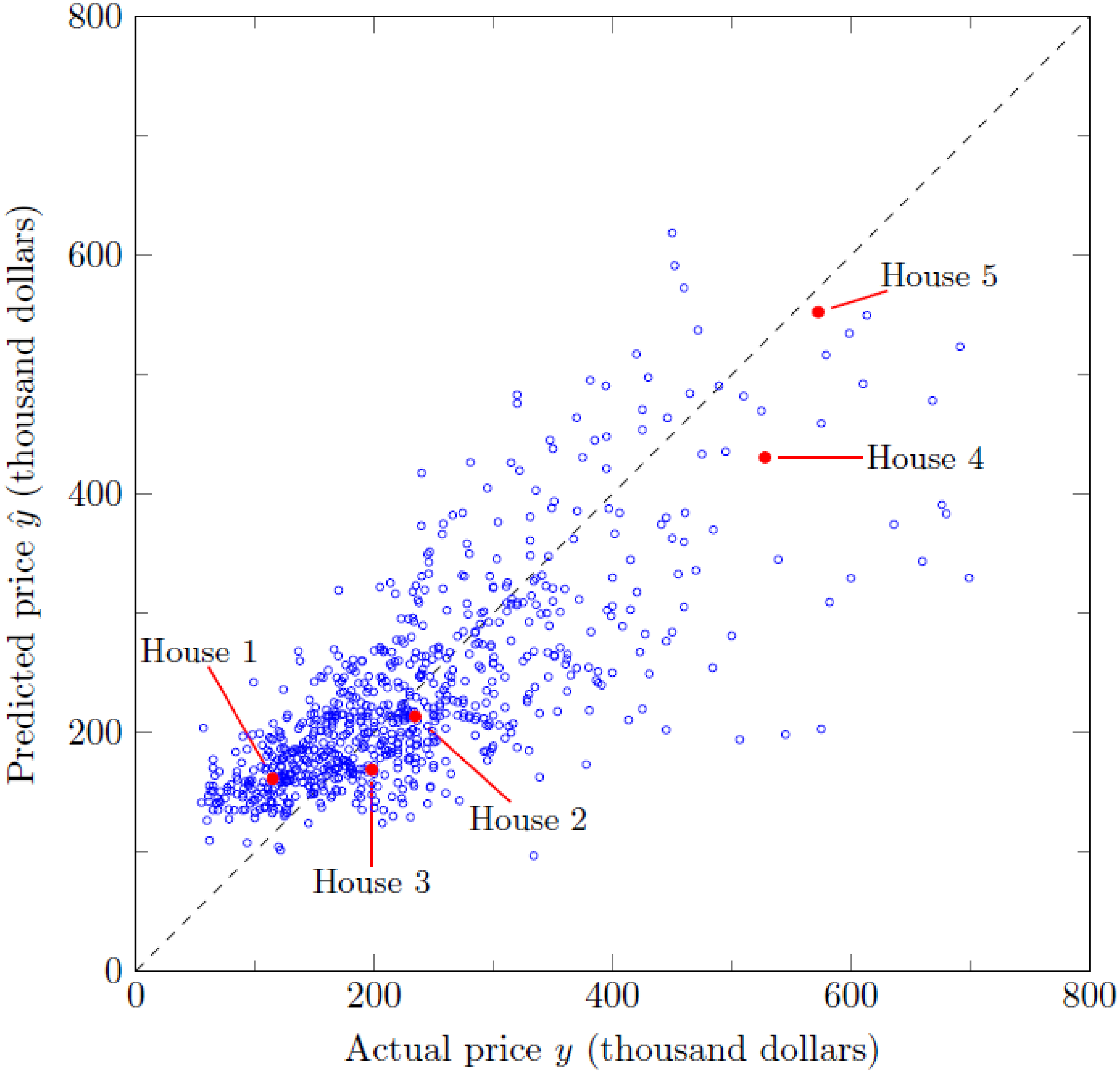


Figure 2.4 Scatter plot of actual and predicted sale prices for 774 houses sold in Sacramento during a five-day period.

Scatter Plot

Model parameters:

$\beta = (148.73, -18.85)$
$v = 54.40$

Special Case 1: No Feature Vector

- Special Case: No feature vector, only offset

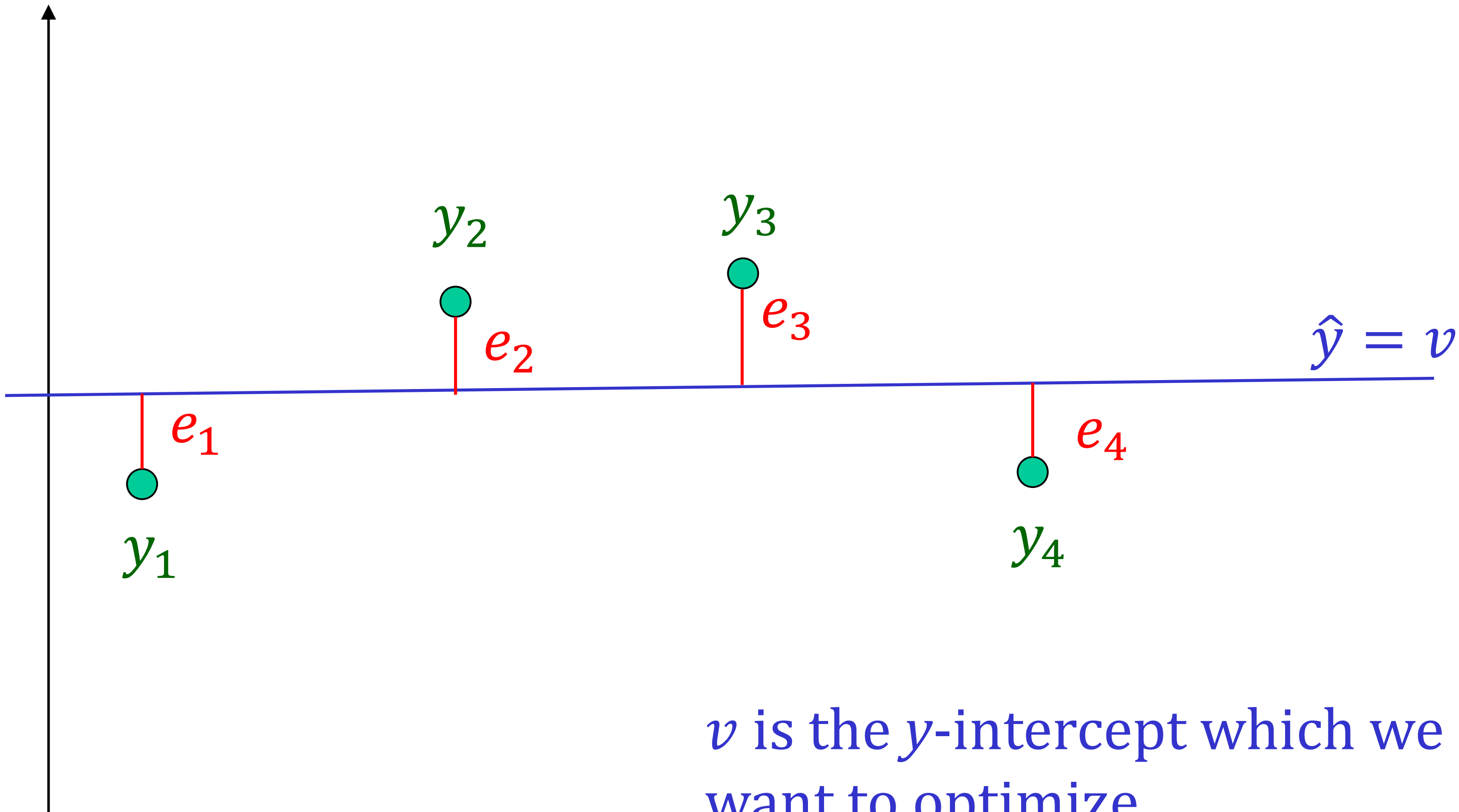
$$\hat{y} = v$$

- Objective: Given n data values $y_1, y_2, \dots, y_n$, we want to find an estimate v which minimizes the total squared error:

$$f(v) = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (y_i - v)^2$$

  - Note: It is the same as minimizing the RMS error (why?)

Horizontal-Line Fitting



v is the y-intercept which we want to optimize.

Least-Square Solution

- Taking the derivative,

$$f(v) = \sum_{i=1}^{n} (y_i - v)^2, \qquad \frac{df}{dv} = 2\sum_{i=1}^{n} (v - y_i)$$

- Setting the derivative to zero,

$$\sum_{i=1}^{n} (v - y_i) = 0, \qquad nv - \sum_{i=1}^{n} y_i = 0$$

- Rearranging the terms,

$$v = \frac{\sum_{i=1}^{n} y_i}{n} = \mathbf{avg}(y)$$

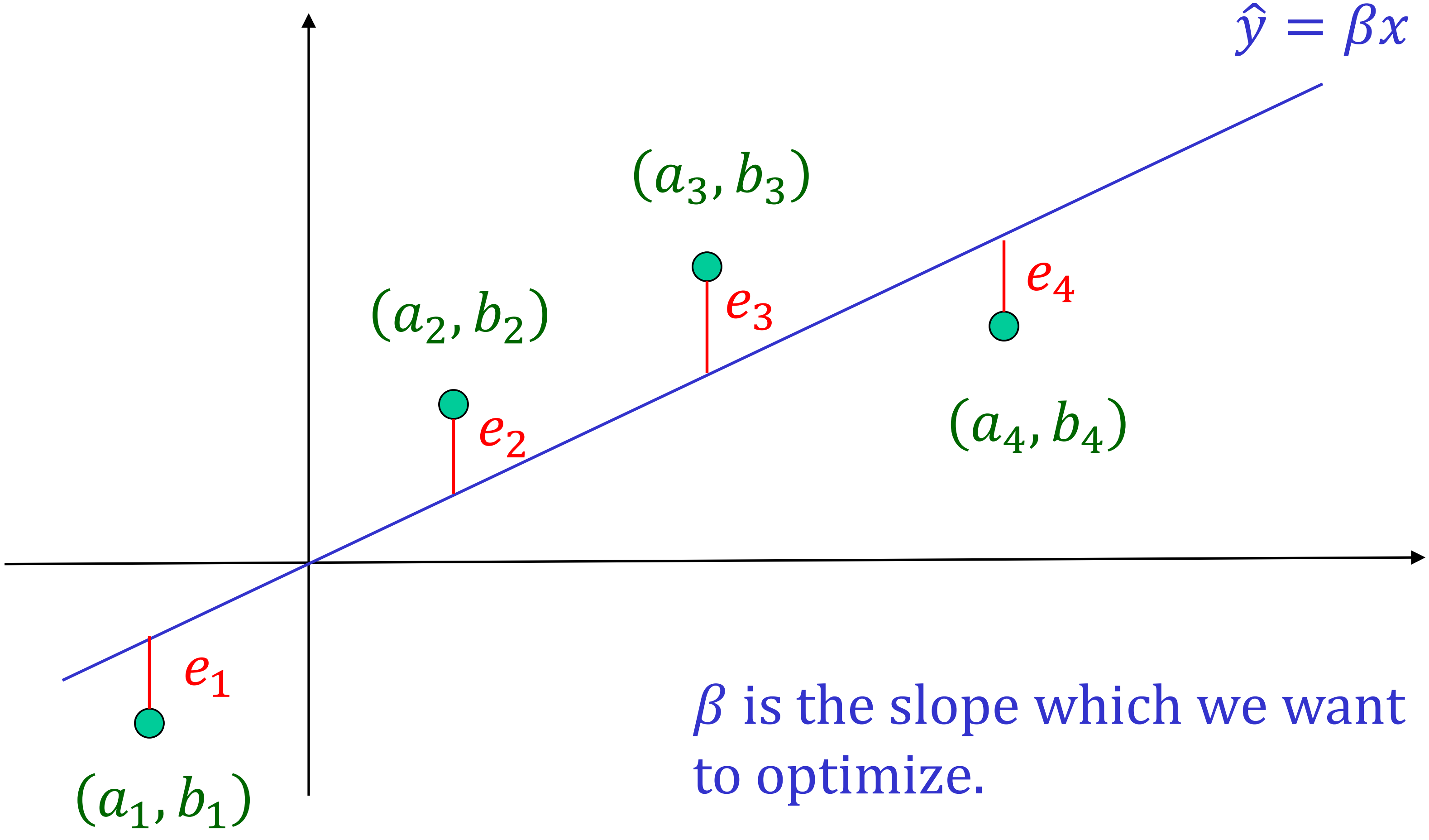
Special Case 2: One Feature, Zero Offset

- Special Case: Only one feature (i.e., x is a scalar) with zero offset (i.e., $v = 0$)

$$\hat{y} = \beta x$$

- Geometric interpretation:
  - Given n data points, $(a_1, b_1), \dots, (a_n, b_n)$, find a straight line which passes through the origin to fit the data points.

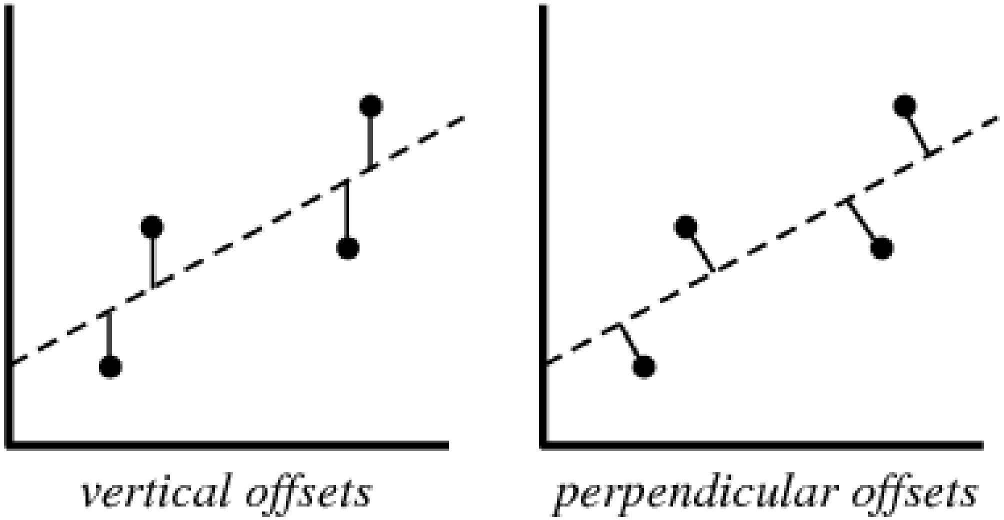
Line Fitting



β is the slope which we want to optimize.

Least-Square Fitting

- Note that least-square fitting minimizes the vertical offsets, *not* the perpendicular offsets.



- This provides a much simpler analytical form, which can be easily generalized to best-fit polynomial curves.

Hubble's Law

$$\text{Recessional Velocity} = \underbrace{\text{Hubble's constant}}_{\text{estimated by regression}} \times \text{distance}$$

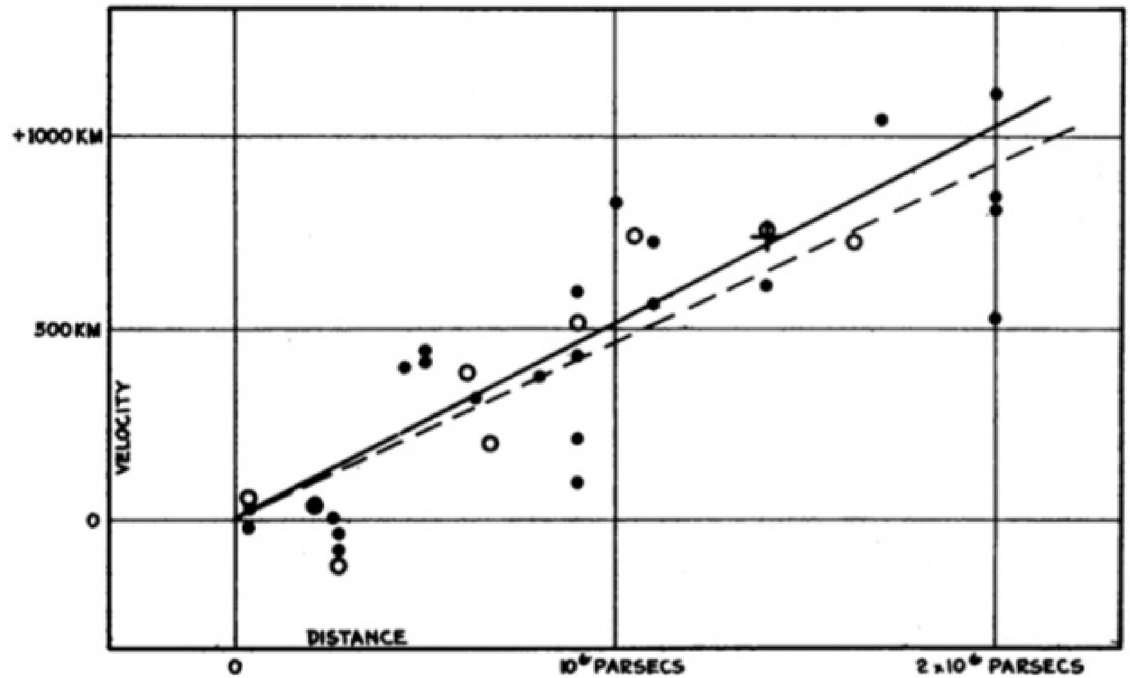


FIGURE 1

Velocity-Distance Relation among Extra-Galactic Nebulae.

Hubble's Diagram (1929):

The first observational evidence for the expansion of the universe.

(2 min) https://www.youtube.com/watch?v=Cd8fweRI8E8

Least-Square Solution

- Total Squared Error:

$$f(\beta) = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (b_i - \beta a_i)^2$$

- Taking its derivative,

$$\frac{df}{d\beta} = \sum_{i=1}^{n} 2(b_i - \beta a_i)(-a_i) = 2\beta \sum_{i=1}^{n} a_i^2 - 2\sum_{i=1}^{n} a_i b_i$$

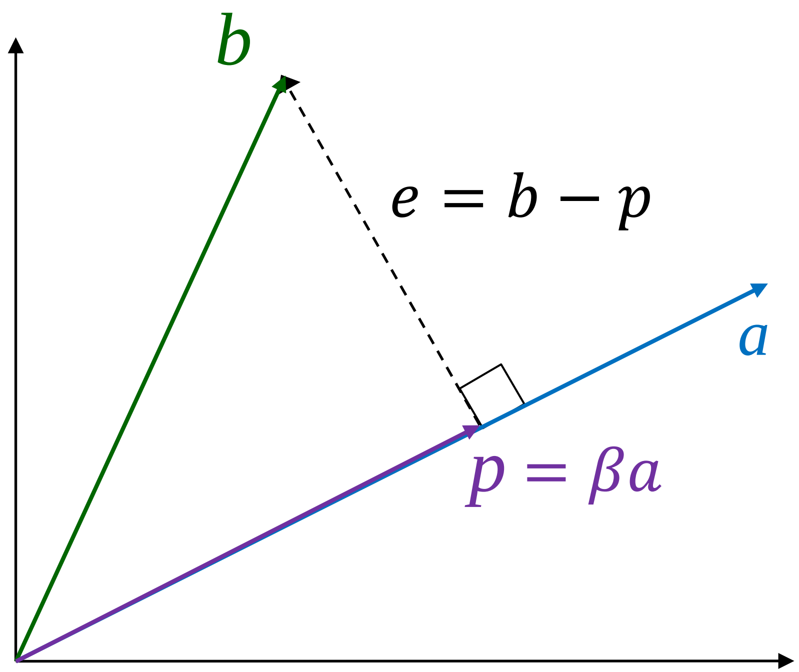
- Setting the derivative to zero,

$$\beta a^T a - a^T b = 0, \text{ or } \beta = \frac{a^T b}{a^T a}.$$

Geometric Interpretation

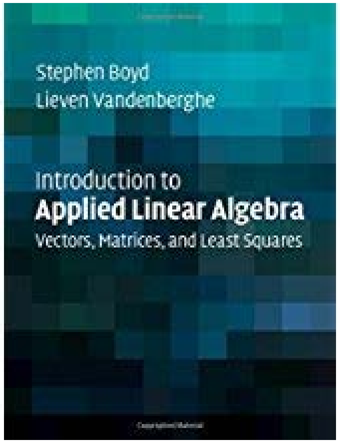
- Regression (one feature, zero offset):

  $$\hat{y} = \beta x, \text{ where } \beta = \frac{a^T b}{a^T a}.$$

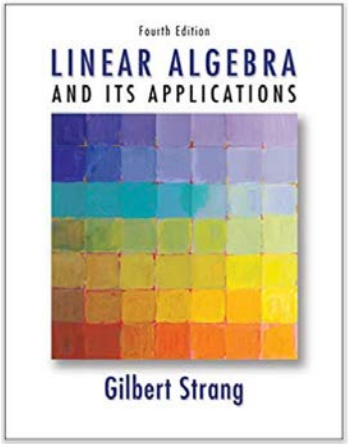


- Regression is the projection of b onto a in the n dimensional space.
  - b is the measured values (e.g. price of a flat) of n data points
  - a is the feature values (e.g. area of a flat) of the n points.
  - e is the error vector, which is perpendicular to a.
- β is the scaling factor of a to obtain the projection vector p.

Recommended Reading



- Chapter 2, S. Boyd and L. Vandenberghe, *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares*, Cambridge University Press, 2018.
  - Available on the web, http://web.stanford.edu/~boyd/vmls/



- Sections 2.6, 3.2, and 3.3, G. Strang, *Linear Algebra and its Applications*, 4th ed., Thomson Learning, 2006.
  - This book is more advanced.