

Unit 6

Vectors

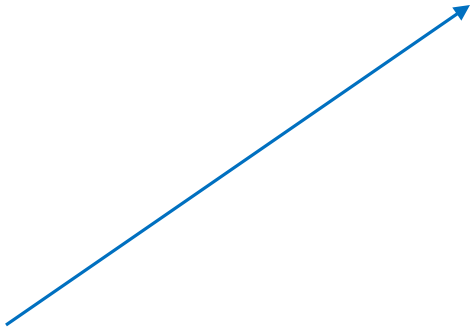
Outlines

- ❑ 6.1 Vector Space
- ❑ 6.2 Inner Product, Norm, and Distance
- ❑ 6.3 Cauchy-Schwarz Inequality
- ❑ 6.4 Statistical Measures

Unit 6.1

Vector Space

What is a Vector?



An arrow

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

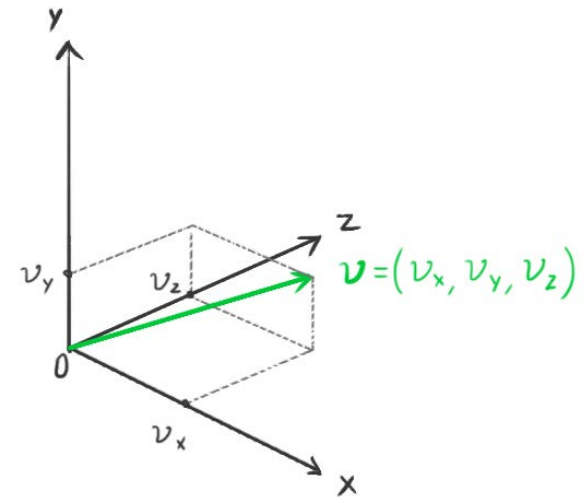
n numbers putting in
a certain order

An n -vector can be used to represent n quantities or values in an application.

The Use of Vectors: Examples

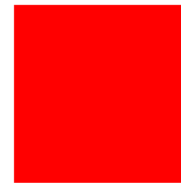
○ Motion

- A 3-vector to represent the velocity of an aircraft.

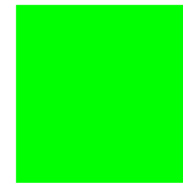


○ Color

- A 3-vector to represent the Red, Green, and Blue (RGB) intensity of values (often between 0 and 1).



(1, 0, 0)



(0, 1, 0)



(0, 0, 1)



(1, 1, 0)



(1, 0.5, 0.5)



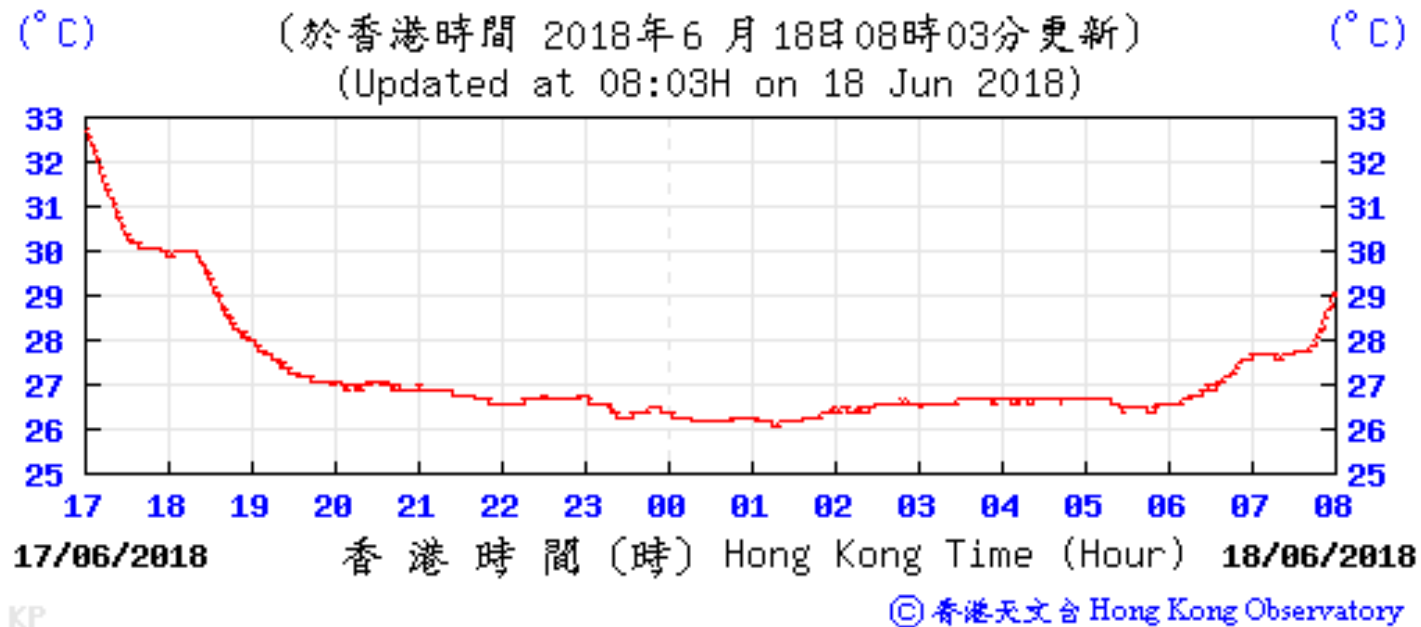
(0.5, 0.5, 0.5)

Vectors

More Examples

○ Time Series (or Discrete-Time Signals)

- An n -vector to represent the value of some quantity at different **times**.



Grass temperature at King's Park

More Examples

○ Data Science

- An n -vector to represent the **features** or **attributes** of an object.
- Examples:
 - a 6-vector could give the age, gender, height, weight, blood pressure and temperature of a patient admitted to a hospital.
 - a 2-vector could give the coursework mark and exam mark of a class of students.

Special Vectors

□ Zero Vector:

- $\mathbf{0}_n$: An n -vector with all entries equal to 0.
 - Sometimes simply written as $\mathbf{0}$.

□ Ones Vector:

- $\mathbf{1}_n$: An n -vector with all entries equal to 1.
 - Sometimes simply written as $\mathbf{1}$.

□ Unit Vectors:

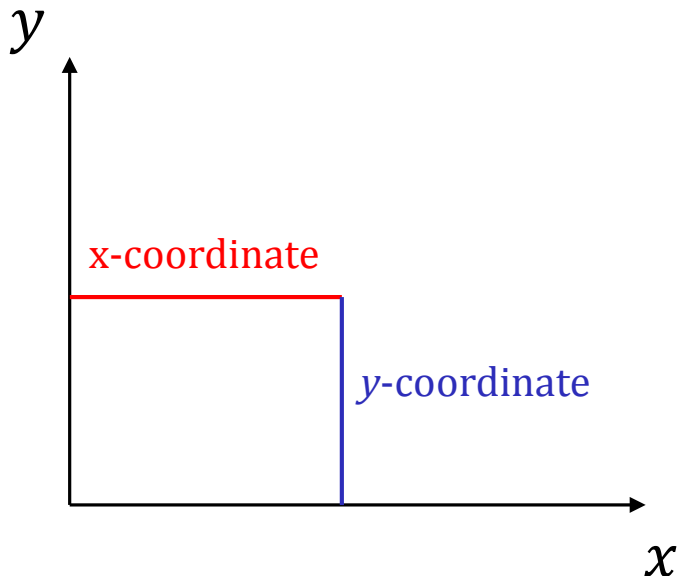
- e_i : An n -vector with all entries equal to 0 except entry i equal to 1.
- Example: In 3-dimensional space,

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

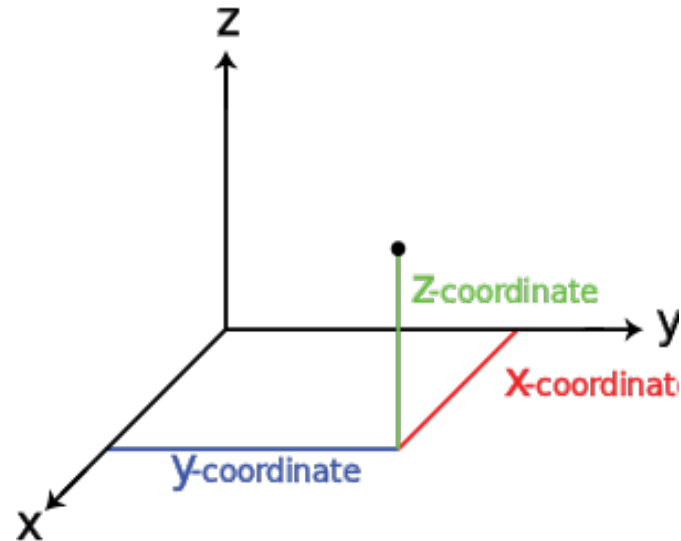
Vector Spaces: Examples

A vector is an element of a **vector space**.

a set with some special properties



2-dimensional
Euclidean space, $\mathbb{R}^2$



3-dimensional
Euclidean space, $\mathbb{R}^3$

What is a Vector Space?

- ❑ A vector space is
 - a set of elements, and
 - two operations are possible within the space.

- ❑ The two operations are
 - i. (Vector Addition) Adding two vectors, and
 - ii. (Scalar Multiplication) Multiplying a vector by a scalar.
- These operations need to satisfy **eight properties** to be defined in the next two slides.

Vector Addition

1. (Commutative) $x + y = y + x$.
2. (Associative) $x + (y + z) = (x + y) + z$.
3. (Zero) There exists an element $\mathbf{0}$, called zero vector, such that $x + \mathbf{0} = x$ for all x .
4. (Inverse) For each x , there exists a unique vector $-x$ such that $x + (-x) = \mathbf{0}$.

Scalar Multiplication

- 5. (Associative) $(c_1 c_2)x = c_1(c_2 x).$
- 6. (Unitarity) $1x = x.$
- 7. (Distributive I) $c(x + y) = cx + cy.$
- 8. (Distributive II) $(c_1 + c_2)x = c_1 x + c_2 x.$

You are not required to memorize them but you should understand what they mean.

Vector Subspace

- A subspace of a vector space is a nonempty subset that satisfies two requirements:
 - i. If x and y are in the subspace, then $x + y$ is in the subspace.
 - ii. If x is in the subspace, then cx is still in the subspace for any scalar c .
- In other words, a subspace is a subset which is “closed” under addition and scalar multiplication.

Linear Combination

- ❑ Let $a_1, a_2, \dots, a_m$ be vectors.
- ❑ The vector $y = \beta_1 a_1 + \beta_2 a_2 + \dots + \beta_m a_m$ is said to be a linear combination of $a_1, a_2, \dots, a_m$.
- ❑ Example
 - There are n students enrolled in EE2302.
 - a_1 : the course work marks of the n students
 - a_2 : the examination marks of the n students
 - $y = 0.4a_1 + 0.6a_2$: the final mark of the n students
 - It's called weighted average when $\beta_1 + \beta_2 + \dots + \beta_m = 1$, and all $\beta_i \geq 0$

Standard Basis

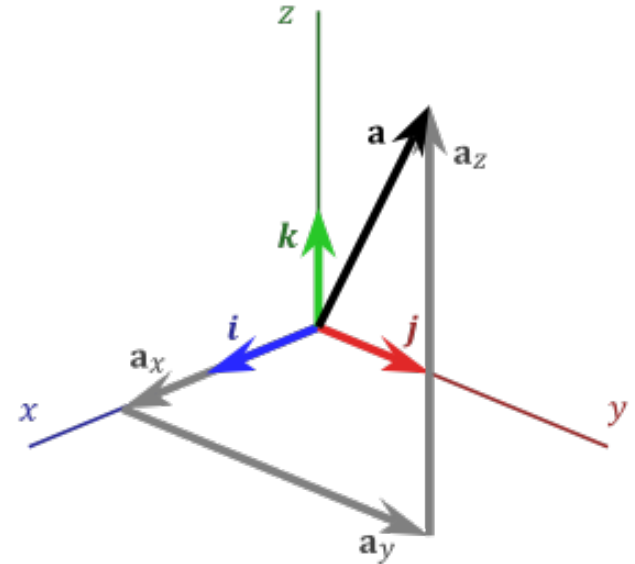
- In the n -dimensional Euclidean space, $\mathbb{R}^n$, $e_1, e_2, \dots, e_n$ are called standard basis.
- Any vector $x \in \mathbb{R}^n$ can be written as a linear combination of them.

- In 3-dimensional space,

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

- Any vector a can be expressed as

$$a = \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} = a_x e_1 + a_y e_2 + a_z e_3.$$



Unit 6.2

Inner Product, Norm, and Distance

Inner Product

- The inner product (or dot product) of two vectors $u = (u_1, u_2, \dots, u_n)$ and $v = (v_1, v_2, \dots, v_n)$ is defined as the scalar

$$\begin{aligned} u^T v &= [u_1 \quad u_2 \quad \dots \quad u_n] \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \\ &= u_1 v_1 + u_2 v_2 + \dots + u_n v_n \end{aligned}$$

Properties of Inner Product

1. Commutativity

$$a^T b = b^T a$$

2. Linearity (in the first argument)

$$(\gamma a)^T b = \gamma(a^T b)$$

$$(a + c)^T b = a^T b + c^T b$$

3. Positive-Definiteness

$$a^T a \geq 0$$

with equality holds if and only if $a = \mathbf{0}$.

Applications

□ Price-quantity

- p : a vector of prices of n goods
- q : a vector of quantities of n goods
- $p^T q$: total cost of the goods given by q .

□ Probability and expected value

- v : a vector whose i -th entry represents the value of the i -th outcome.
- p : a vector whose i -th entry represents the probability of the i -th outcome.
- Example: Throw of a fair die
 - $v = (1, 2, 3, 4, 5, 6)$
 - $p = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$
 - Expected value = $p^T v = 3.5$



Applications

□ Sentiment Analysis (of movie reviews)

- x : an n -vector to represent the histogram of word occurrences in a document, from a dictionary of n words.
- w : an n -vector to represent the sentiment categories of the n words.
 - $w_i = 1$ if word i is positive (e.g. good, nice) 
 - $w_i = -1$ if word i is negative (e.g. bad, terrible) 
 - $w_i = 0$ if word i is neutral 
- $w^T x$: a (crude) measure of the sentiment of the movie.

Norm

- The norm of a vector can be defined by the inner product:

$$\|x\| \triangleq \sqrt{x^T x},$$

which can be interpreted as the **length** of x .

- The **Euclidean norm** then follows:

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

- Any real-valued function of a vector that satisfies the three properties in the next slide is called a (general) norm. In this unit, we'll only consider the Euclidean norm.

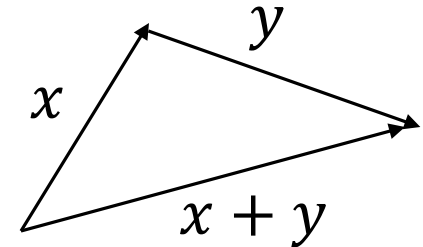
Properties of Norm

- Absolute Scalability (or Absolute Homogeneity)

$$\|\beta x\| = |\beta| \|x\|$$

- Triangle Equality

$$\|x + y\| \leq \|x\| + \|y\|$$



(Unobvious algebraically;
to be proved later)

- Positive-definiteness

$$\|x\| \geq 0$$

with equality holds only if $x = \mathbf{0}$.

Quadratic Formula for Vector Norm

□ A useful formula:

$$\|x + y\|^2 = \|x\|^2 + 2x^T y + \|y\|^2$$

□ How to prove it?

○ Hint: $\|x + y\|^2 = (x + y)^T (x + y)$

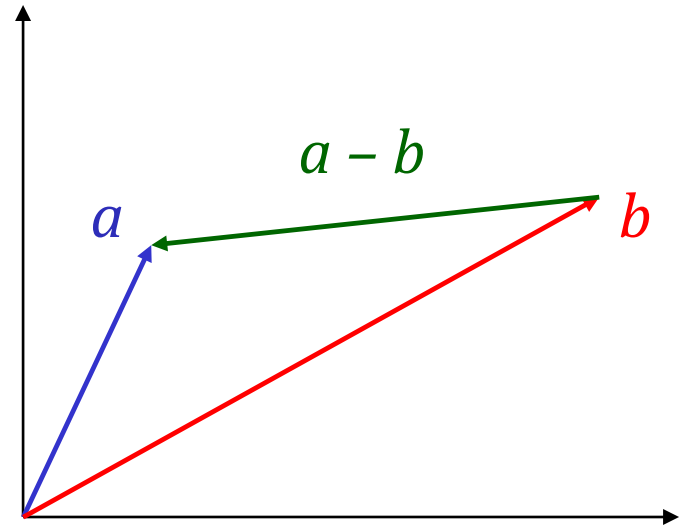
Euclidean Distance

- The distance between two vectors a and b can be defined by the norm:

$$d(a, b) \triangleq \|a - b\|.$$

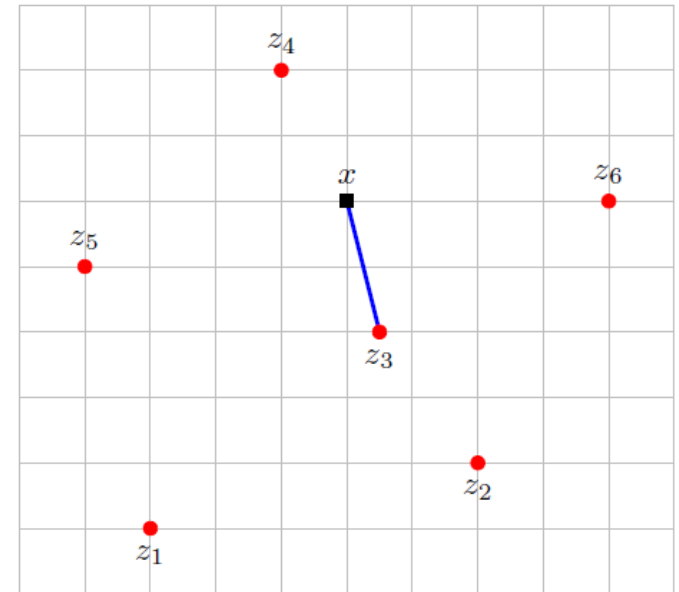
- Symmetry:

$$d(a, b) = d(b, a)$$



Example: Nearest Neighbor

- $z_1, z_2, \dots, z_m$ are all n -vectors.
- x is another n -vector.
- We say that z_j is the **nearest neighbour** of x if
$$\|x - z_j\| \leq \|x - z_i\|, \forall i.$$
- This concept is used in many applications:
 - e.g., demodulation, decoding, data clustering, ...



Example: Document Dissimilarity

- ❑ Suppose x and y represent the histograms of word occurrences for two documents.
- ❑ Then $d(x, y)$ measures the dissimilarity of the two documents.

Unit 6.3

Cauchy-Schwarz Inequality

Cauchy-Schwarz Inequality

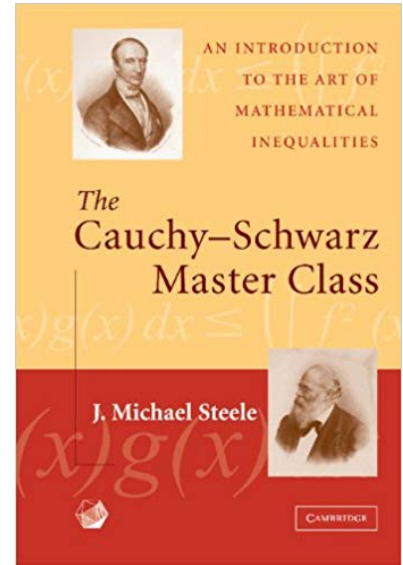
- An important inequality:

$$|a^T b| \leq \|a\| \|b\|$$

This can be written out as

$$|a_1 b_1 + \cdots + a_n b_n| \leq (a_1^2 + \cdots + a_n^2)^{1/2} (b_1^2 + \cdots + b_n^2)^{1/2}$$

- This inequality is clearly true if either a or b is a zero vector. In the proof shown in next page, we assume that both of them are non-zero.



Proof: $(a \neq 0, b \neq 0)$

□ Define $\alpha \triangleq \|a\|$ and $\beta \triangleq \|b\|$.

$$\begin{aligned}\square \quad 0 &\leq \|\beta a - \alpha b\|^2 \\ &= \|\beta a\|^2 - 2(\beta a)^T(\alpha b) + \|\alpha b\|^2 \quad (\text{quadratic formula}) \\ &= \beta^2 \|a\|^2 - 2\beta\alpha(a^T b) + \alpha^2 \|b\|^2 \\ &= 2\|a\|^2 \|b\|^2 - 2\|a\| \|b\| (a^T b) \\ &= 2\|a\| \|b\| (\|a\| \|b\| - a^T b)\end{aligned}$$

□ Therefore, $a^T b \leq \|a\| \|b\|$. (because $a \neq 0, b \neq 0$)

□ Apply this inequality to $-a$ and b , $-a^T b \leq \|a\| \|b\|$.

□ Combining them, $|a^T b| \leq \|a\| \|b\|$. *Q.E.D.*

Proof of Triangle Inequality

$$\begin{aligned}\square \|a + b\|^2 &= \|a\|^2 + 2a^T b + \|b\|^2 && \text{(quadratic formula)} \\ &\leq \|a\|^2 + 2\|a\|\|b\| + \|b\|^2 && \text{(Cauchy-Schwarz)} \\ &= (\|a\| + \|b\|)^2\end{aligned}$$

$$\begin{aligned}\square \text{ Taking the square root,} \\ \|a + b\| &\leq \|a\| + \|b\|\end{aligned}$$

Q.E.D.

Angle

- The **angle** between two vectors a and b is defined as

$$\theta \triangleq \cos^{-1} \left(\frac{a^T b}{\|a\| \|b\|} \right)$$

By Cauchy-Schwarz inequality, this ratio is always between -1 and $+1$.



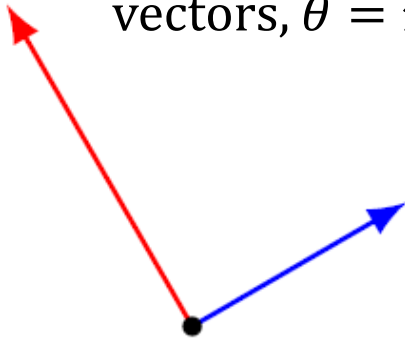
- In other words, θ is the unique number between 0 and π (measured in radians) that satisfies

$$a^T b = \|a\| \|b\| \cos \theta$$

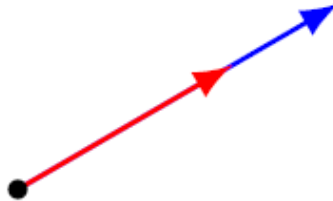
- a and b are said to be **orthogonal** if $\theta = \frac{\pi}{2}$.
 - Or equivalently, $a^T b = 0$.

Examples

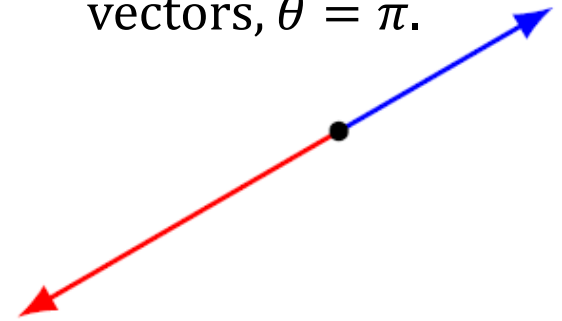
Orthogonal
vectors, $\theta = \pi/2$.



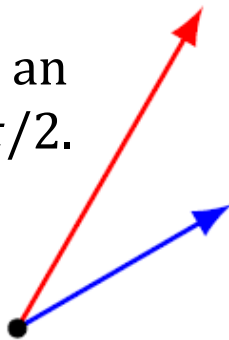
Aligned vectors,
 $\theta = 0$.



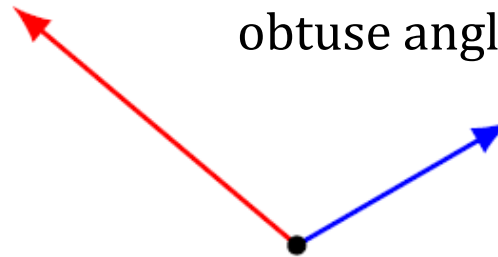
Anti-aligned
vectors, $\theta = \pi$.



Vectors that make an
acute angle, $\theta < \pi/2$.



Vectors that make an
obtuse angle, $\theta > \pi/2$.



Example: Document Similarity via Angles

- If x and y represent the word count distributions for two documents, their angle can also be used as a measure of document similarity.

Unit 6.4

Statistical Measures

Mean

- Given an n -vector $x = (x_1, x_2, \dots, x_n)$.
- The mean (or average) of the n elements is given by

$$\mathbf{avg}(x) \triangleq \frac{1}{n} \mathbf{1}^T x = \frac{x_1 + x_2 + \dots + x_n}{n}$$

- Traditionally, the mean is denoted by μ .

De-Meaned Vector

- Given any vector x , the associated de-meaned vector is obtained by subtracting $\mathbf{avg}(x)$ from each entry of x :

$$\tilde{x} = x - \mu \mathbf{1}$$

- It is useful for understanding how the entries of x deviate from their mean, μ .
- Q: What is $\tilde{x}$ if all entries of x are equal?

Root-Mean-Square Value

- The root-mean-square (RMS) of an n -vector x is defined as “**norm normalized by root- n** ”,

$$\mathbf{rms}(x) \triangleq \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}} = \frac{\|x\|}{\sqrt{n}}$$

- It is useful in comparing norms of vectors with different dimensions.
- It is also called quadratic mean; it tells us what a “typical” value of $|x_i|$ is.
- Q: What is the RMS of $(1, 1, 1)$? $(1, 1, 2)$?

Example: Estimation Error

- ❑ Suppose the n -vector y represents a time series of temperature at some location.
- ❑ Suppose the n -vector $\hat{y}$ represents an estimation of y based on other information.
- ❑ The RMS estimation error,

$$\mathbf{rms}(y - \hat{y})$$

is often used to measure how accurate the estimation is.

- ❑ It is a **normalized** distance, so that the value is independent of n .
- ❑ This concept is used in many signal processing applications.

Standard Deviation

- The standard deviation of x is defined as the **RMS of the de-meaned vector** $x - \mu\mathbf{1}$, i.e.,

$$\mathbf{std}(x) \triangleq \sqrt{\frac{(x_1 - \mu)^2 + \cdots + (x_n - \mu)^2}{n}}.$$

- We interpret this number as a “typical” value by which the entries differ from the mean of the entries.
- It can also be expressed as

$$\mathbf{std}(x) = \frac{\|x - \mu\mathbf{1}\|}{\sqrt{n}} = \frac{\|x - (\mathbf{1}^T x / n)\mathbf{1}\|}{\sqrt{n}}.$$

- Traditionally, the standard deviation is denoted by σ .

Properties of Standard Deviation

- Adding a constant:

$$\mathbf{std}(x + c\mathbf{1}) = \mathbf{std}(x)$$

- σ is the RMS of the demeaned vector of x . The constant will be removed after “demeaning”.

- Multiplying by a scalar:

$$\mathbf{std}(ax) = |a|\mathbf{std}(x)$$

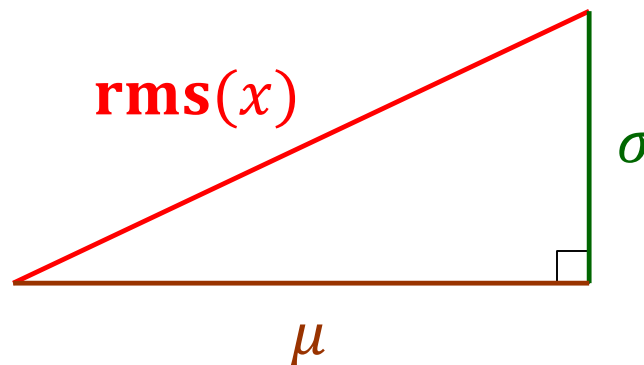
- σ is defined by RMS, which is a normalized norm. Multiplying the vector by a (no matter it is positive or negative) will “lengthen” the vector by $|a|$.

Useful Identity

$$\mathbf{rms}(x)^2 = \mu^2 + \sigma^2$$

(Proof omitted.)

- Its form is the same as the Pythagoras' Theorem:



Chebyshev's Inequality for Data Set

The proportion of entries of x that satisfy

$$|x_i - \mu| \geq m\sigma$$

is less than or equal to $\frac{1}{m^2}$.

Proof

- Let k of its entries satisfy

$$|x_i - \mu| \geq m\sigma$$

- The other $n - k$ entries must satisfy

$$|x_i - \mu| \geq 0 \quad (\text{because of the absolute sign})$$

- By definition,

$$\sigma = \mathbf{std}(x) = \sqrt{\frac{\sum (x_i - \mu)^2}{n}} \geq \sqrt{\frac{k(m\sigma)^2}{n}} = \sqrt{\frac{k}{n}} m\sigma$$

- Cancelling out σ and rearranging the terms,

$$\frac{k}{n} \leq \frac{1}{m^2}$$

Q.E.D.

Example

- ❑ Computers from a particular company are found to last on average for 3 years without any hardware malfunction, with standard deviation of 2 months. At least what proportion of the computers last between 31 months and 41 months?

- ❑ Answer:

- $\sigma = 2, m\sigma = 5$. Therefore, $m = 2.5$.

- At least $1 - \frac{1}{m^2} = 1 - 16\% = 84\%$.

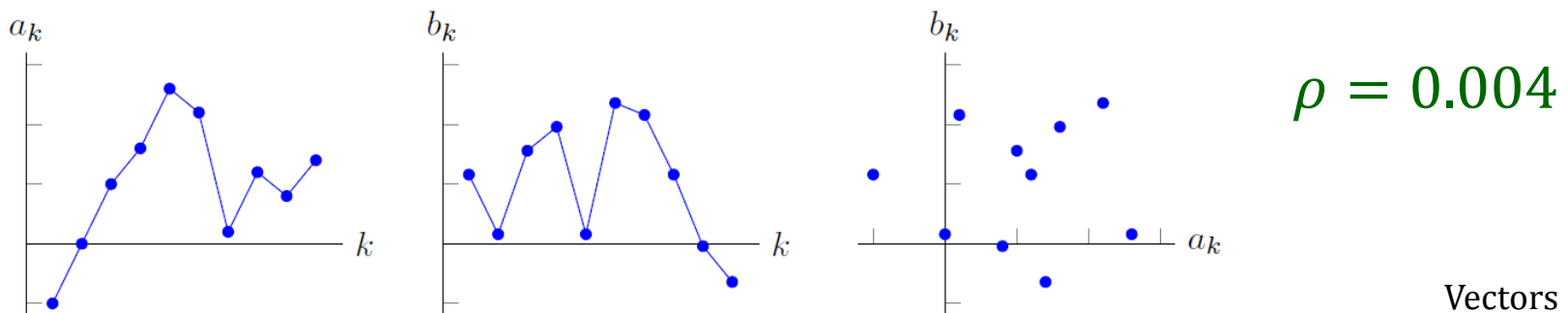
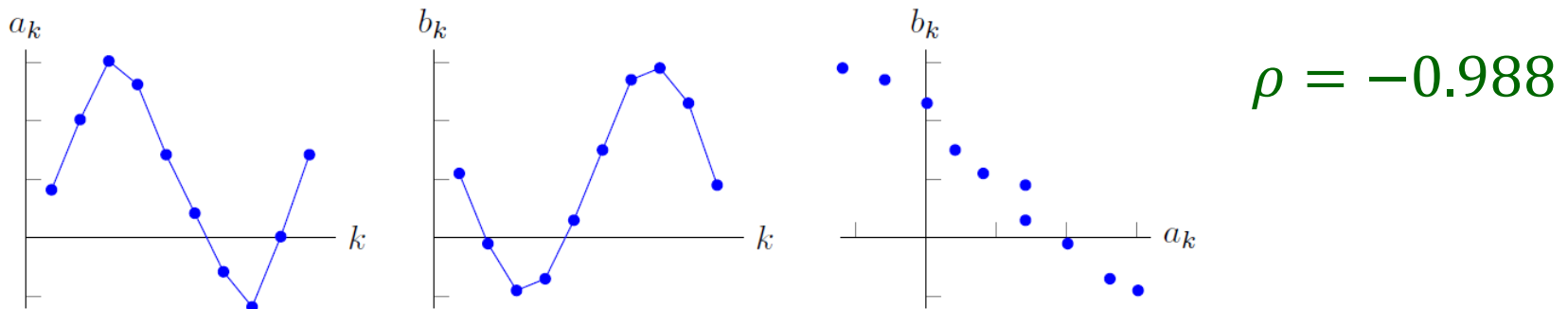
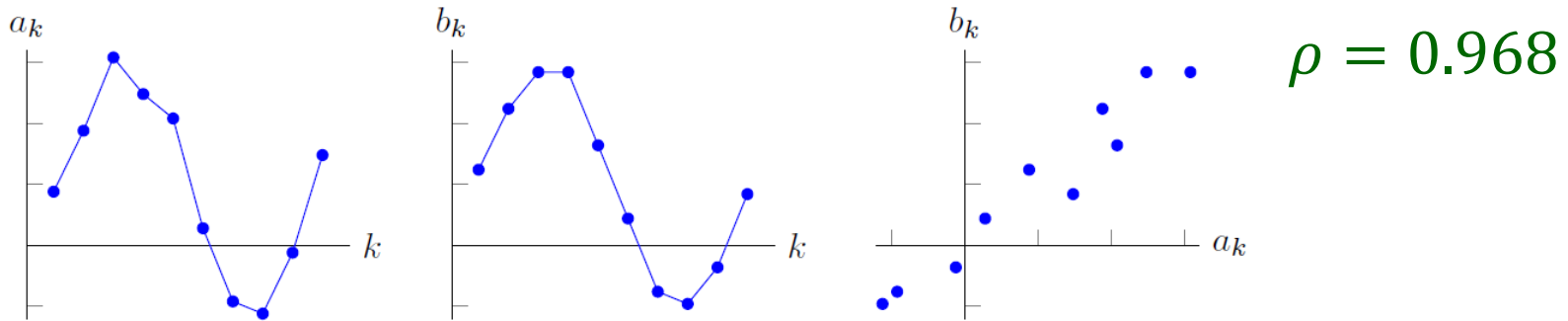
Correlation Coefficient

- ❑ Let $\tilde{a}$ and $\tilde{b}$ be the demeaned vectors of a and b , respectively.
- ❑ Assume that $\tilde{a}$ and $\tilde{b}$ are not zero.
 - i.e., not all entries of the original vector are equal.
- ❑ The **correlation coefficient** of a and b is defined as

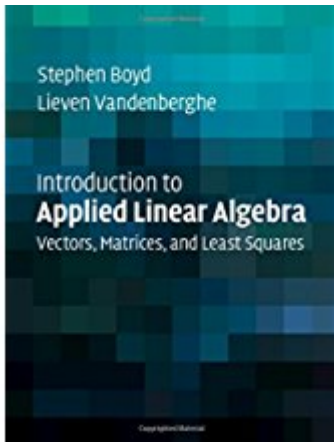
$$\rho \triangleq \frac{\tilde{a}^T \tilde{b}}{\|\tilde{a}\| \|\tilde{b}\|} = \cos \theta$$

- $|\rho|$ measures the **strength** of **linear** relationship.
- The sign of ρ measures the indicates the **direction** of the relationship.
- When $\rho = 0$, a and b are said to be **uncorrelated**.

Example: Three pairs of vectors with their scatter plots



Recommended Reading



- Chapters 1 and 3, S. Boyd and L. Vandenberghe, *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares*, Cambridge University Press, 2018.
 - The book is available on the web,
<http://web.stanford.edu/~boyd/vmls/>