Unit 7

Vectors

Albert Sung

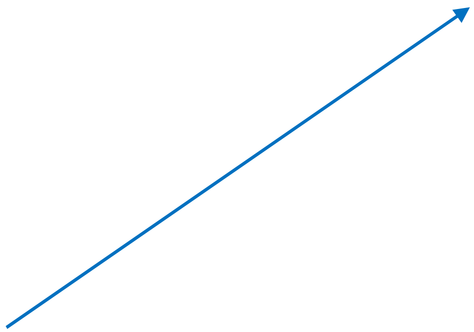
Outlines

- 7.1 Vector Space
- 7.2 Inner Product, Norm, and Distance
- 7.3 Cauchy-Schwarz Inequality
- 7.4 Statistical Measures

Unit 7.1

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.

Special Vectors

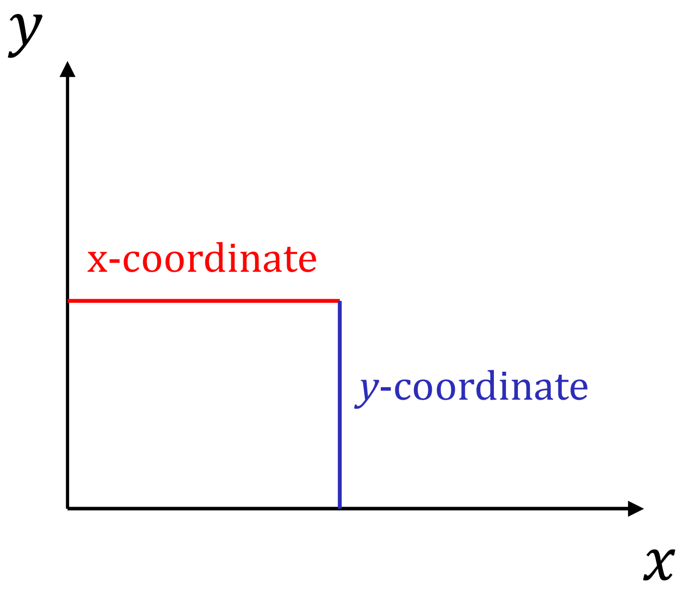
- Zero Vector:
  - $\mathbf{0}_n$: An n-vector with all entries equal to 0.
    - Sometimes simply written as $\mathbf{0}$.
- Ones Vector:
  - $\mathbf{1}_n$: An n-vector with all entries equal to 1.
    - Sometimes simply written as $\mathbf{1}$.
- Unit Vectors:
  - e_i: An n-vector with all entries equal to 0 except entry i equal to 1.
  - Example: In 3-dimensional space,

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

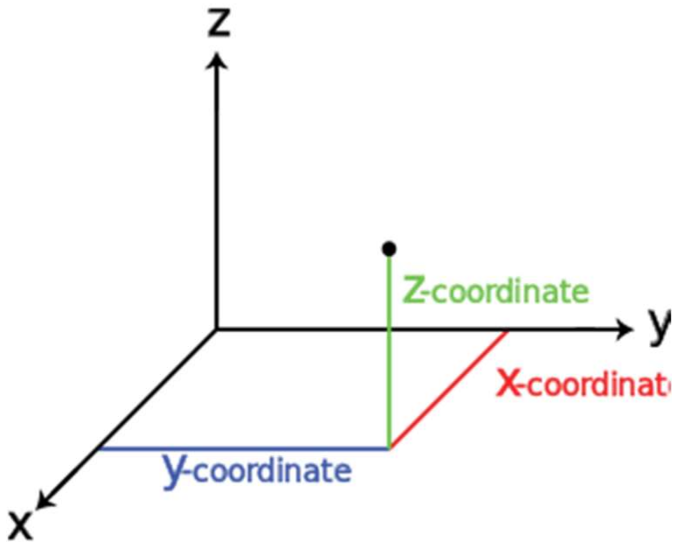
Vector Spaces: Examples

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 within the space.

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

Eight Properties

No need to memorize them.

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 $x + (-x) = \mathbf{0}$.
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) $(c_1 + c_2)x = c_1 x + c_2 x$.

Example 1: Matrices

- Consider the set of all 2× 2 matrices with real entries, denoted by $\mathbb{R}^{2\times 2}$.
- The two operations are defined as follows:
  - Addition: $\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} w & x \\ y & z \end{bmatrix} = \begin{bmatrix} a+w & b+x \\ c+y & d+z \end{bmatrix}$.
  - Scalar Multiplication: $\alpha \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \alpha a & \alpha b \\ \alpha c & \alpha d \end{bmatrix}$.
- Do these two operations satisfy the eight conditions?
  - Yes. (The checking is tedious, thus omitted).
  - Details can be found here: https://www.youtube.com/watch?v=ug3FpapN8Ng (start from 7:38)

Example 2: Real Functions

- Consider the set of all functions $f: \mathbb{R} \to \mathbb{R}$.
- The two operations are defined as follows:
  - Addition: $(f + g)(x) = f(x) + g(x)$.
  - Scalar Multiplication: $(\alpha f)(x) = \alpha f(x)$.
- Next, check the eight conditions.
- Yes, they are satisfied.
- Zero element is the constant function $\mathbf{0}(x) = 0$.

Example 3: Polynomials

- A real polynomial p is of the form
$$p = a_0 + a_1 x + \cdots + a_n x^n,$$
where the coefficients are real numbers.

- The two operations are defined in the usual way.
- It can be checked that the set of all polynomials is a vector space.

Subspace

- A vector space is a set with two operations that satisfy eight conditions.
- Its subset is called a subspace if the subset is also a vector space.
- We only need to check:

Is it closed under addition and scalar multiplication?

"Closed" means that the result remains in the subset.

- The eight conditions will automatically be satisfied, since it is a subset of a vector space.

Example 4: The x-y plane in $\mathbb{R}^3$

- Is the x-y plane a subspace of $\mathbb{R}^3$?
- The x-y plane consists of all vectors in the form of $(x, y, 0)$.
- Closed under addition:
  - $(x_1, y_1, 0) + (x_2, y_2, 0) = (x_1 + x_2, y_1 + y_2, 0)$ is still in the x-y plane.
- Closed under scalar multiplication:
  - $c(x, y, 0) = (cx, cy, 0)$ is still in the x-y plane.
- Therefore, it is a subspace.

Example 5: A non-example

- What if we life the x-y plane by 1 unit along the z-axis? Is it a subspace of $\mathbb{R}^3$?
- No.
  - Not closed under addition:
    $$(x_1, y_1, 1) + (x_2, y_2, 1) = (x_1 + x_2, y_1 + y_2, 2)$$
  - Not closed under scalar multiplication if $c \neq 1$:
    $$c(x, y, 1) = (cx, cy, c)$$
- Recall that condition 3 says that there must be a zero element. In this case, the zero vector is not in the lifted x-y plane, so it must not be a vector space.

Linear Combination

- Let $a_1, a_2, \dots, a_m$ belong to a vector space.
- The vector $y = \beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_m a_m$ is said to be a linear combination of $a_1, a_2, \dots, a_m$.
  - y also belongs to the vector space.
- 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
  - $y = 0.55a_1 + 0.45a_2$: the final mark of the n students
    - It's called weighted average when $\beta_1 + \beta_2 + \cdots + \beta_m = 1$, and all $\beta_i \geq 0$

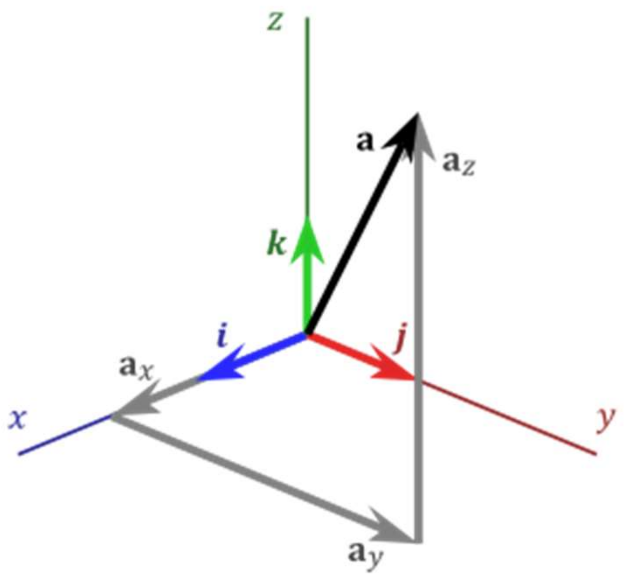
Standard Basis

- In the n-dimensional Euclidean space, $\mathbb{R}^n$, the set $\{e_1, e_2, \dots, e_n\}$ is called standard basis.
- Any vector $x \in \mathbb{R}^n$ can be written as a linear combination of the standard basis vectors.
  - In 3-dimensional space,

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

  - Any vector a can be expressed as

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



Linear Combination of Audio Signals

- (6.5 min) https://www.youtube.com/watch?v=sNigRX9-z1A

Unit 7.2

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

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

$$= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

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 $a = \mathbf{0}$.

Norm

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

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

which can be interpreted as the length of x.

  - 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 focus mainly on the Euclidean norm.

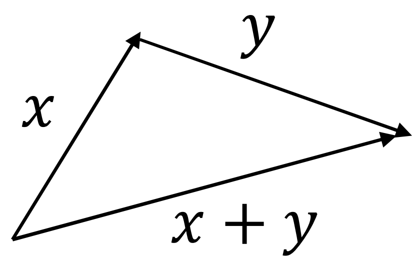
Properties of Norm

1) Absolute Scalability (or Absolute Homogenity)

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

2) Triangle Inequality

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



(Unobvious algebraically; to be proved later)

3) Positive-definiteness

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

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

Quadratic Formula for Vector Norm

- A useful formula:

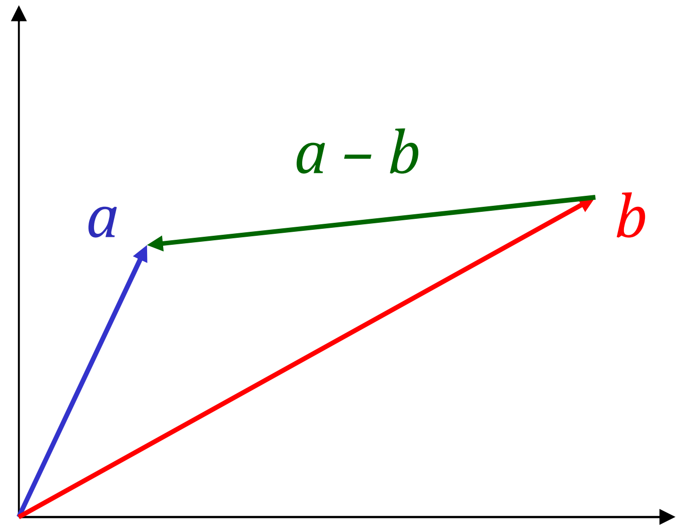
$$\boxed{\|x + y\|^2 = \|x\|^2 + 2x^T y + \|y\|^2}$$

- 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:

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



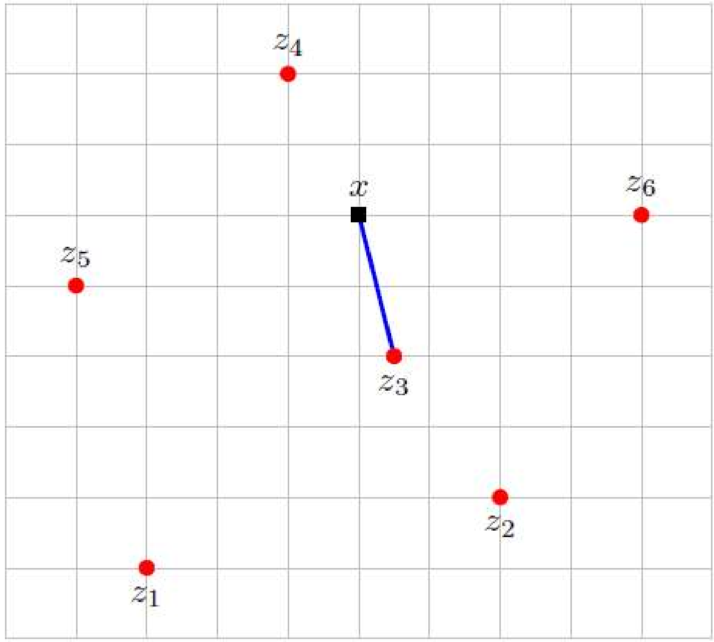
- Symmetry:

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

Nearest Neighbor

- $z_1, z_2, \dots, z_k$ 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, clustering, …

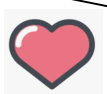
Example: Movie Preference

- Alice, Bob, and Charlie rank the four movies using a 5-point scale:

$a = (2, 1, 5, 5)$

$\|b - a\| = \sqrt{38}$

$b = (5, 3, 2, 1)$



Who has a similar taste as me?

$\|c - a\| = \sqrt{6}$

$c = (1, 3, 4, 5)$

Unit 7.3

Cauchy-Schwarz Inequality

Cauchy-Schwarz Inequality

- An important inequality:

$$\boxed{|a^T b| \le \|a\| \|b\|}$$

This can be written out as

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

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

Triangle Inequality (proof)

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

- Taking the square root,

$$\boxed{\|a+b\| \le \|a\| + \|b\|}$$

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 $+1$.

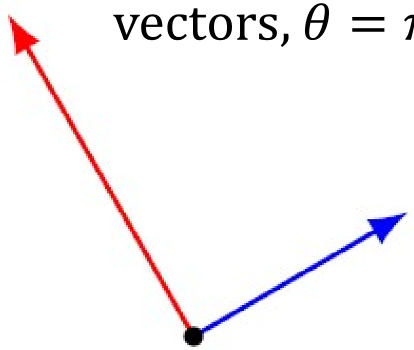
- In other words, θ is the unique number between 0 and π (in radians) that satisfies

$$\boxed{a^T b = \|a\|\|b\| \cos\theta}$$

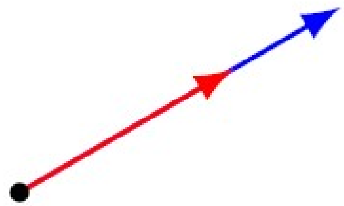
- a and b are said to be orthogonal if $\theta = \frac{\pi}{2}$.
  - Or equivalently, $a^T b = 0$.

Examples

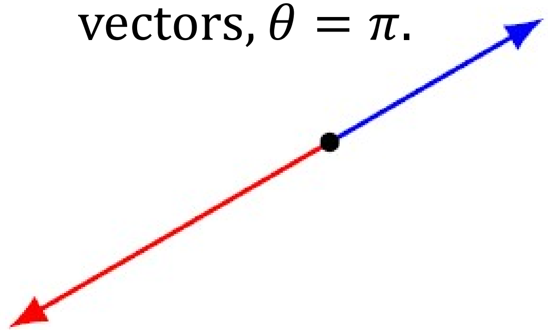
Orthogonal vectors, $\theta = \pi/2$.



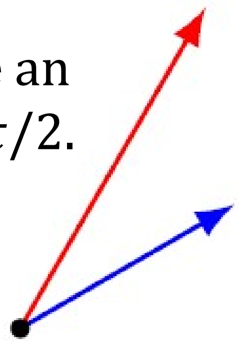
Aligned vectors, $\theta = 0$.



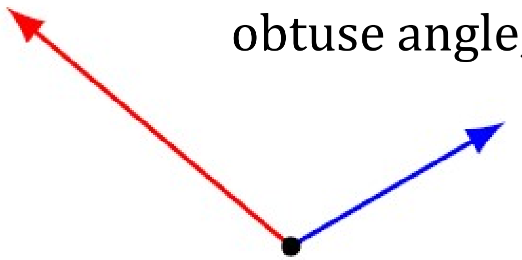
Anti-aligned vectors, $\theta = \pi$.



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



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



Unit 7.4

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

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

- Traditionally, the mean is denoted by μ.
- It is also called arithmetic mean.

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?

Geometric Mean

- The geometric mean is defined as $\sqrt[n]{x_1 x_2 \dots x_n}$.
- Comparison with arithmetic mean.
  - multiplication instead of addition
  - n-th root instead of division by n.

- Given any n non-negative numbers,

$$\boxed{\text{arithmetic mean} \geq \text{geometric mean}}$$

A.M. ≥ G.M. (for $n = 2$)

Prove $\frac{x_1+x_2}{2} \geq \sqrt{x_1 x_2}$ using Cauchy-Schwarz inequality.

Proof:

$$\|a\|\|b\| \geq |a^T b|$$

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 also called quadratic mean, which roughly reflects the typical value of $|x_i|$.
- Q: What is the RMS of $(1, 1, 1)$? $(1, 1, 2)$?

Application: 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 the accuracy of an estimation.
- 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 c to each component:

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

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

- Multiplying by a scalar:

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

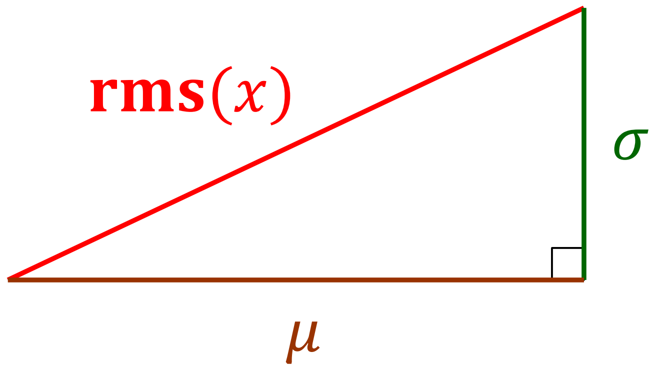
  - Standard deviation is a "length" (normalized by the dimension).
  - Multiplying the vector by a (no matter it is positive or negative) will "lengthen" the vector by $|a|$.

Useful Identity

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

(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
$$|x_i - \mu| \geq 0 \qquad \text{(because of the absolute sign)}$$
- 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 certain 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 $1 - \frac{1}{m^2} = 1 - 16\% = 84\%$.

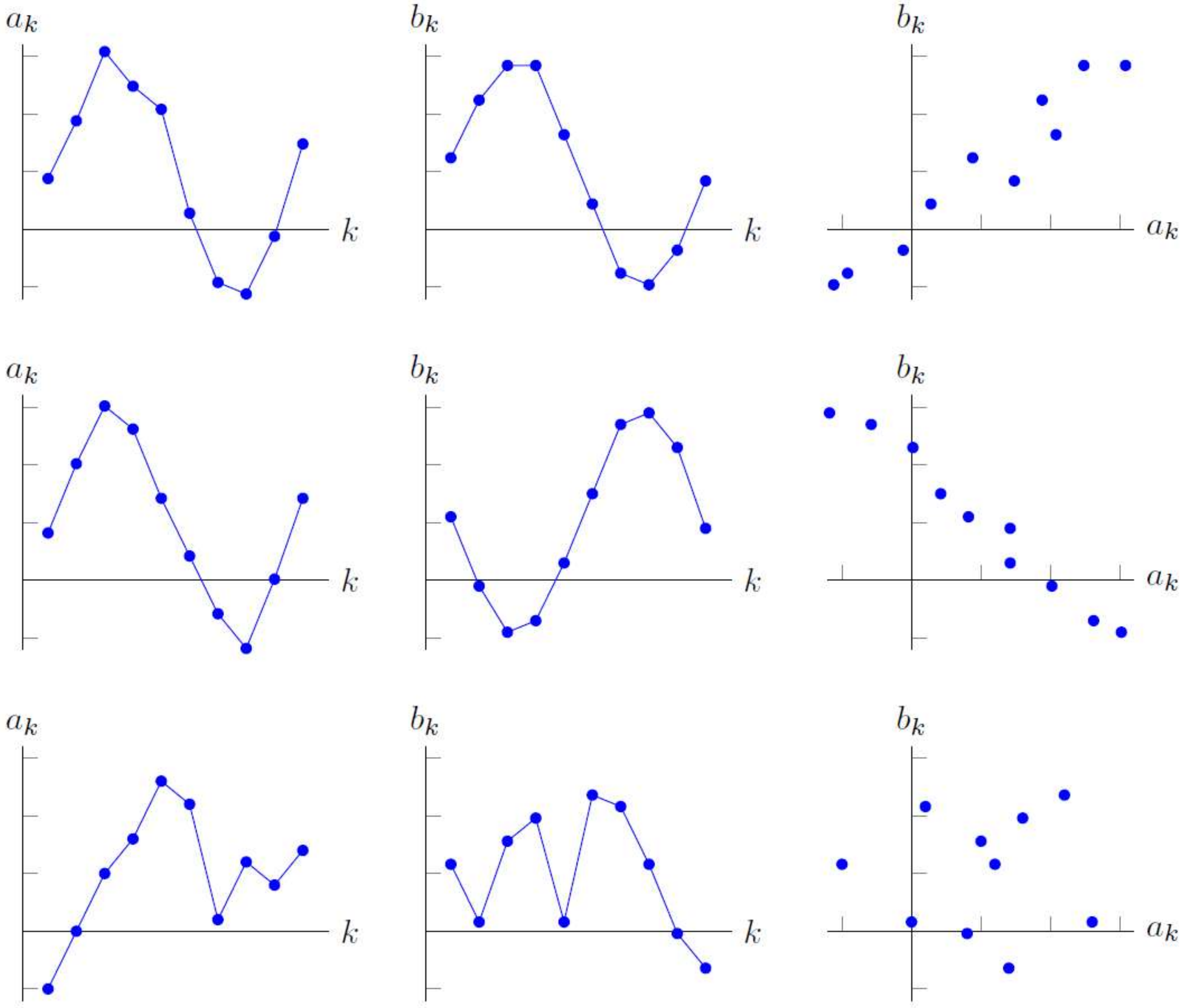
Correlation Coefficient

- 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



$\rho = 0.968$

$\rho = -0.988$

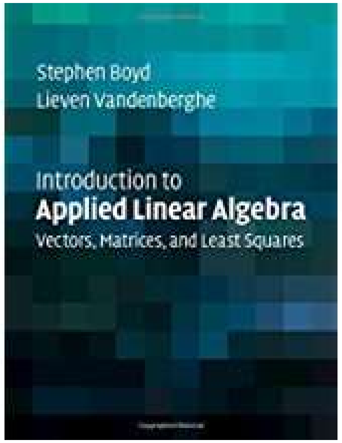
$\rho = 0.004$

Example: Coursework Score up to Test 1

- There are 101 students enrolled in this course.
- Consider the following three 101-vectors:
  - t: Scores of Test 1
  - a: Scores of Assignment 1 – 4
  - q: Scores of Quizzes 1 – 4

- Correlation between t and a: 0.39
- Correlation between t and q: 0.41
- Correlation between q and a: 0.55

Recommended Reading



- Chapters 1, 3 and 4, 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/
  - There are many examples and exercises in this book!