Spring 2023

SWCON253: Machine Learning

Lecture 02
Review of Linear Algebra

Jinwoo Choi
Assistant Professor
CSE, Kyung Hee University

Disclaimer: Many slides have been borrowed from Hui Yong Kim, who may have borrowed some of them from others. Any time a slide does not show any explicit credit, I credit the slide to Hui Yong Kim.

Contents

References

- *Linear Algebra Review* by J. Zico Kolter (http://www.cs.cmu.edu/~zkolter/course/linalg/)
- *Review of Linear Algebra* by Xinkun Nie & Fereshte Khani (http://cs229.stanford.edu/syllabus-fall2020.html)
- *Linear Algebra for Machine Learning* by Sargur N. Srihari (https://cedar.buffalo.edu/~srihari/CSE574/)

1. Notation & Basic Operations

1. Vector & Matrix
2. Special Matrices
3. Additions & Multiplications
4. Transpose
5. Trace

Vector & Matrix

- By $x \in \mathbb{R}^n$, we denote a vector with n entries.

1-D array of numbers

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

↳ 열벡터

$$x^T = [x_1 \ x_2 \ \cdots \ x_n].$$

↳ 행벡터

- By $A \in \mathbb{R}^{m \times n}$ we denote a matrix with m rows and n columns, where the entries of A are real numbers.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix}.$$

2-D array of numbers

a set of column vectors

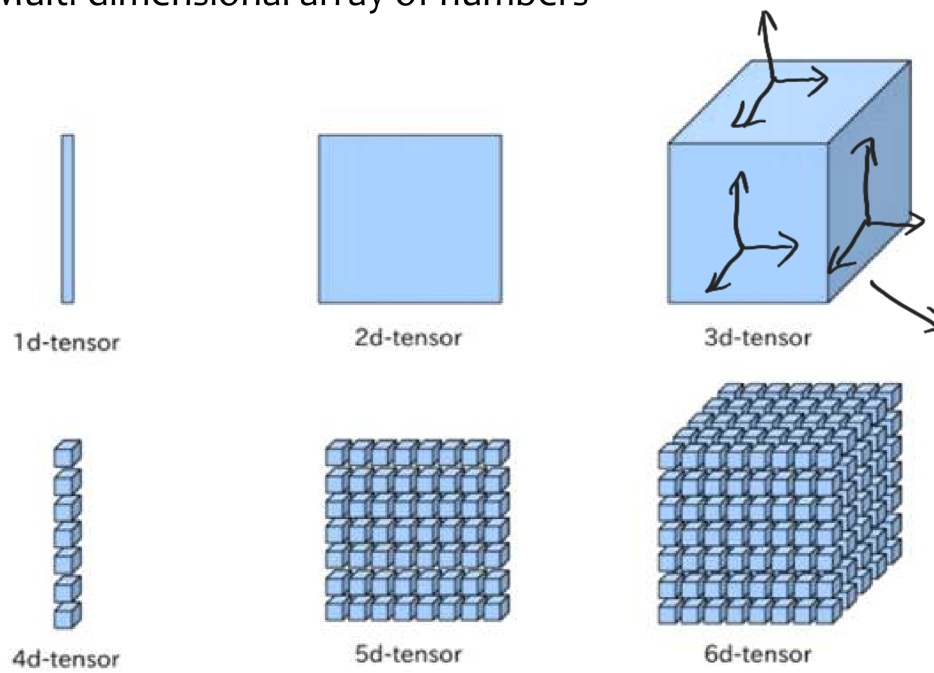
a set of row vectors

$a^{①}$ → 숫자가 위에 = 열 벡터

$a_{①}$ → 숫자가 밑에 = 행 벡터

Tensor → 동시에 2개 이상의 방향을 표시해야할 때 필요

Multi-dimensional array of numbers



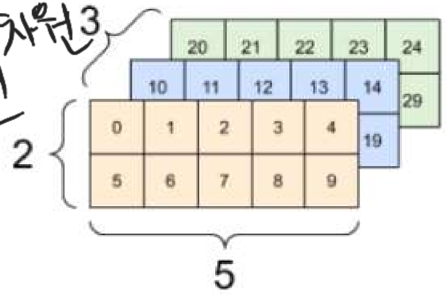

```
tf.Tensor(
[[[ 0  1  2  3  4]
  [ 5  6  7  8  9]]

 [[10 11 12 13 14]
  [15 16 17 18 19]]

 [[20 21 22 23 24]
  [25 26 27 28 29]]], shape=(3, 2, 5)
```

→ 면마다 새로운 직교 좌표계를 표현

→ 상위 차원3 표현



Product of Vectors

"row × column"

inner product or *dot product*

Inner: row X column

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i . \quad = y^T x$$

outer product

outer: column x row

$$xy^T \in \mathbb{R}^{m \times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_n \end{bmatrix} . \quad = (yx^T)^T$$

Special Matrices – *Zero*

The Zero Matrix

$$0 \in \mathbb{R}^{m \times n} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Special Matrices – *Identity*

The Identity Matrix

$$I \in \mathbb{R}^{n \times n} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

~ *collection of* ***unit vectors*** *in* R^n

Has the property that for any $A \in \mathbb{R}^{m \times n}$

$$AI = A = IA$$

Special Matrices – *Diagonal*

Diagonal Matrices

For $d \in \mathbb{R}^n$

$$\operatorname{diag}(d) \in \mathbb{R}^{n \times n} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}$$

- For example, the identity is given by $I = \operatorname{diag}(1)$

$$1 \in \mathbb{R}^n = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

Scalar Operations: *aA+b*

A scalar can be added to a matrix.

A matrix can be multiplied by a scalar.

$$B = aA + b \quad \Rightarrow \quad B_{i,j} = aA_{i,j} + b$$

Matrix Addition: *A+B*

- For two matrices *of the same size and type*, $A, B \in \mathbb{R}^{m \times n}$ addition is just sum of corresponding elements

$$A + B = C \in \mathbb{R}^{m \times n} \quad \Longleftrightarrow \quad C_{ij} = A_{ij} + B_{ij}$$

Addition is *undefined* for matrices of different sizes $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{p \times q}$

(Note: Numpy's *broadcasting* allows addition of different sized arrays)

Matrix-Vector Multiplication: *Ax*

- If we write A by rows, then we can express Ax as,

$$y = Ax = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{bmatrix}.$$

- If we write A by columns, then we have:

$$y = Ax = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a^1 \end{bmatrix} x_1 + \begin{bmatrix} a^2 \end{bmatrix} x_2 + \ldots + \begin{bmatrix} a^n \end{bmatrix} x_n .$$

y is a *linear combination* of the *columns* of A.

Vector-Matrix Multiplication: $x^T A$

- If we write A by columns, then we can express $x^\top A$ as,

$$y^T = x^T A = x^T \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} x^T a^1 & x^T a^2 & \cdots & x^T a^n \end{bmatrix}$$

- expressing A in terms of rows we have:

$$y^T = x^T A \;=\; \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix}$$

$$= \; x_1 \begin{bmatrix} — & a_1^T & — \end{bmatrix} + x_2 \begin{bmatrix} — & a_2^T & — \end{bmatrix} + \ldots + x_m \begin{bmatrix} — & a_m^T & — \end{bmatrix}$$

y^T is a linear combination of the *rows* of A.

Matrix Multiplication: *AB*

- For two matrices $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$, their product is

$$AB = C \in \mathbb{R}^{m\times p} \iff C_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}$$

Multiplication is undefined when number of columns in A doesn't equal number or rows in B (one exception: cA for $c \in \mathbb{R}$ taken to mean scaling A by c)

Matrix Multiplication – *Different Views*

1. As a set of vector-vector products / inner

$$C = AB = \underset{\text{row}}{\begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix}} \underset{\text{column}}{\begin{bmatrix} | & | & & | \\ b^1 & b^2 & \cdots & b^p \\ | & | & & | \end{bmatrix}} = \begin{bmatrix} a_1^T b^1 & a_1^T b^2 & \cdots & a_1^T b^p \\ a_2^T b^1 & a_2^T b^2 & \cdots & a_2^T b^p \\ \vdots & \vdots & \ddots & \vdots \\ a_m^T b^1 & a_m^T b^2 & \cdots & a_m^T b^p \end{bmatrix}.$$

2. As a sum of outer products

$$C = AB = \underset{\text{column}}{\begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix}} \underset{\text{row}}{\begin{bmatrix} — & b_1^T & — \\ — & b_2^T & — \\ & \vdots & \\ — & b_n^T & — \end{bmatrix}} = \sum_{i=1}^{n} a^i b_i^T .$$

Matrix Multiplication – *Different Views (cont'd)*

3. As a set of matrix-vector products.

$$C = AB = A \begin{bmatrix} | & | & & | \\ b^1 & b^2 & \cdots & b^p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ Ab^1 & Ab^2 & \cdots & Ab^p \\ | & | & & | \end{bmatrix}. \qquad (2)$$

Here the ith column of C is given by the matrix-vector product with the vector on the right, $c_i = Ab_i$. These matrix-vector products can in turn be interpreted using both viewpoints given in the previous subsection.

4. As a set of vector-matrix products.

$$C = AB = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} B = \begin{bmatrix} — & a_1^T B & — \\ — & a_2^T B & — \\ & \vdots & \\ — & a_m^T B & — \end{bmatrix}.$$

Matrix Multiplication – *Diagonal Matrix*

- Multiplying $A \in \mathbb{R}^{m\times n}$ by a diagonal matrix $D \in \mathbb{R}^{n\times n}$ on the right scales the *columns* of A

$$A\boxed{D} = \begin{bmatrix} | & | & & | \\ d_1 a_1 & d_2 a_2 & \cdots & d_n a_n \\ | & | & & | \end{bmatrix}$$

→ column scaler

(diagonal on the right)

- Multiplying by a diagonal matrix $D \in \mathbb{R}^{m\times m}$ on the left scales the *rows* of A

$$\boxed{D}A = \begin{bmatrix} — & d_1 a_1^T & — \\ — & d_2 a_2^T & — \\ & \vdots & \\ — & d_m a_m^T & — \end{bmatrix}$$

→ row scaler

(diagonal on the left)

Matrix Multiplication – *Important Properties*

- Associative: $(A \in \mathbb{R}^{n\times n},\ B \in \mathbb{R}^{n\times p},\ C \in \mathbb{R}^{p\times q})$

$$A(BC) = (AB)C$$

- Distributive: $(A \in \mathbb{R}^{m\times n},\ B, C \in \mathbb{R}^{n\times p})$

$$A(B + C) = AB + AC$$

- *NOT* commutative: (the dimensions might not even make sense, but this doesn't hold even when the dimensions are correct)

$$\boxed{AB \neq BA}$$

Transpose ~ *flipping*

The *transpose* of a matrix results from "flipping" the rows and columns. Given a matrix $A \in \mathbb{R}^{m \times n}$, its transpose, written $A^T \in \mathbb{R}^{n \times m}$, is the $n \times m$ matrix whose entries are given by

$$(A^T)_{ij} = A_{ji}.$$

– The mirror image across a diagonal line

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} \\ A_{2,1} & A_{2,2} & A_{2,3} \\ A_{3,1} & A_{3,2} & A_{3,3} \end{bmatrix} \Rightarrow A^T = \begin{bmatrix} A_{1,1} & A_{2,1} & A_{3,1} \\ A_{1,2} & A_{2,2} & A_{3,2} \\ A_{1,3} & A_{2,3} & A_{3,3} \end{bmatrix}$$

The following properties of transposes are easily verified:

- $(A^T)^T = A$
- $\boxed{(AB)^T = B^T A^T}$
- $(A + B)^T = A^T + B^T$

Special Matrices – *Symmetric*

Symmetric Matrices

- Symmetric matrix: $A \in \mathbb{R}^{n \times n}$ with $A = A^T$ 대칭행렬
- Arise naturally in many settings

  Gram matrix

  – For $A \in \mathbb{R}^{m \times n}$, $\boxed{A^T A \in \mathbb{R}^{n \times n} \text{ is symmetric}}$ A^TA는 무조건 대칭

$$\underset{(m\times n)}{A} = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} \rightarrow \underset{(n\times m)}{A^T} = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_n^T & - \end{bmatrix} \rightarrow \underset{(n\times n)}{A^TA} = \begin{bmatrix} a_1^Ta_1 & a_1^Ta_2 & \cdots & a_1^Ta_n \\ a_2^Ta_1 & a_2^Ta_2 & \cdots & a_2^Ta_n \\ \vdots & & & \vdots \\ a_n^Ta_1 & a_n^Ta_2 & \cdots & a_n^Ta_n \end{bmatrix}$$

$(A^TA)^T = A^TA$ $\therefore$ symmetric

$(AA^T)^T = AA^T$ $\therefore$ symmetric

Symmetric matrix examples
- distance matrix
- covariance matrix

Trace ~ *sum of diagonals*

The *trace* of a square matrix $A \in \mathbb{R}^{n \times n}$, denoted $\mathrm{tr}A$, is the sum of diagonal elements in the matrix:

$$\mathrm{tr}A = \sum_{i=1}^{n} A_{ii}.$$

Trace: 모든 대각 성분들의 합

The trace has the following properties:

- For $A \in \mathbb{R}^{n \times n}$, $\mathrm{tr}A = \mathrm{tr}A^T$.
- For $A, B \in \mathbb{R}^{n \times n}$, $\mathrm{tr}(A + B) = \mathrm{tr}A + \mathrm{tr}B$.
- For $A \in \mathbb{R}^{n \times n}$, $t \in \mathbb{R}$, $\mathrm{tr}(tA) = t\,\mathrm{tr}A$.
- For A, B such that AB is square, $\mathrm{tr}AB = \mathrm{tr}BA$.
- For A, B, C such that ABC is square, $\mathrm{tr}ABC = \mathrm{tr}BCA = \mathrm{tr}CAB$, and so on for the product of more matrices.

2. Linear Independence & Inverse

1. Linear Independence
2. Span & Linear Transform
3. Rank & Inverse
4. Solving Linear Equations

Linear Independence

선형 종속

A set of vectors $\{x_1, x_2, \ldots x_n\} \subset \mathbb{R}^m$ is said to be *(linearly) dependent* if one vector belonging to the set *can* be represented as a linear combination of the remaining vectors; that is, if

$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i$$

선형 독립

for some scalar values $\alpha_1, \ldots, \alpha_{n-1} \in \mathbb{R}$; otherwise, the vectors are *(linearly) independent*.

$(x_1, x_2, \cdots, x_i)$ 중에 어떤 원소 x_k가 다른 원소들의 합으로 표현될 수 있으면 선형 종속

Example:

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad x_2 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix} \quad x_3 = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

are linearly dependent because $x_3 = -2x_1 + x_2$.

Span

The *span* of a set of vectors $\{x_1, x_2, \ldots x_n\}$ is the set of all vectors that can be expressed as a linear combination of $\{x_1, \ldots, x_n\}$. That is,

$$\underbrace{\mathrm{span}}_{\text{공간}}(\{x_1, \ldots x_n\}) = \left\{ v : v = \sum_{i=1}^{n} \alpha_i x_i, \ \ \alpha_i \in \mathbb{R} \right\}.$$

→ 얘네들의 선형 결합으로 표현할수 있는 벡터들의 집합

- Note that the ***span*** of n ***linearly independent*** vectors is $\mathbb{R}^n$.

Linear Transform: *v = Ax*

◆ *v = Ax*

- An *m* x *n* matrix *A* maps an *n*-dimensional vector *x* into an *m*-dimensional vector *v*.

$$x \in \mathbb{R}^n \xrightarrow[A \in \mathbb{R}^{m \times n}]{v = Ax} v \in \mathbb{R}^m$$

- [illegible] ***linear***, it is called a *linear transform*.

For $\forall \alpha, \beta \in \mathbb{R}$ and $\forall x_1, x_2 \in \mathbb{R}^n$,

$$A(\alpha x_1 + \beta x_2) = \alpha(Ax_1) + \beta(Ax_2)$$

Range & Null Space: $\mathcal{R}(A)$ & $\mathcal{N}(A)$

The *range* or the columnspace of a matrix $A \in \mathbb{R}^{m \times n}$, denoted $\mathcal{R}(A)$, is the the span of the columns of A.

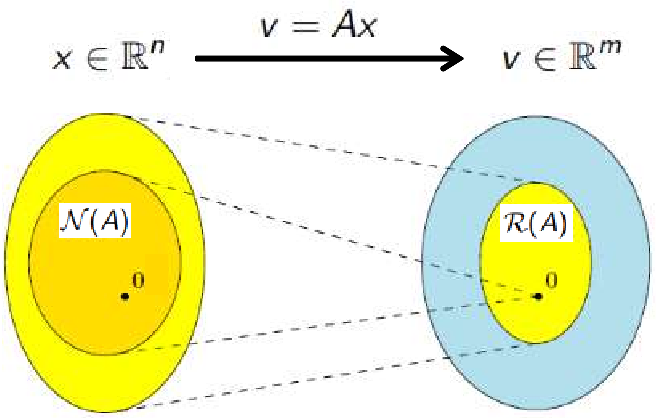
$$\mathcal{R}(A) = \{v \in \mathbb{R}^m : v = Ax, x \in \mathbb{R}^n\}.$$ (image)

공역

열공간 — A의 열벡터들이 만드는 공간

The *nullspace* of a matrix $A \in \mathbb{R}^{m \times n}$, denoted $\mathcal{N}(A)$ is the set of all vectors that equal 0 when multiplied by A, i.e.,

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\}.$$ (kernel)



Rank ~ # of lin. ind. columns (rows)

The *column rank* of a matrix $A \in \mathbb{R}^{m \times n}$ is the size of the largest subset of columns of A that constitute a linearly independent set.

column rank = row rank = matrix rank

The *row rank* is the largest number of rows of A that constitute a linearly independent set.

For any matrix $A \in \mathbb{R}^{m \times n}$, it turns out that the column rank of A is equal to the row rank of A (prove it yourself!), and so both quantities are referred to collectively as the *rank* of A, denoted as $\mathrm{rank}(A)$.

rank = 열 또는 행의 부분공간 중 가장 큰 선형 독립 집합의 원소 수.

Properties of the Rank

- For $A \in \mathbb{R}^{m \times n}$, $\mathrm{rank}(A) \leq \min(m, n)$. If $\mathrm{rank}(A) = \min(m, n)$, then A is said to be *full rank*.

크기와 rank가 같으면 full rank

- For $A \in \mathbb{R}^{m \times n}$, $\mathrm{rank}(A) = \mathrm{rank}(A^T)$.
- For $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, $\mathrm{rank}(AB) \leq \min(\mathrm{rank}(A), \mathrm{rank}(B))$.
- For $A, B \in \mathbb{R}^{m \times n}$, $\mathrm{rank}(A + B) \leq \mathrm{rank}(A) + \mathrm{rank}(B)$.

Inverse

The *inverse* of a square matrix $A \in \mathbb{R}^{n \times n}$ is denoted A^{-1}, and is the unique matrix such that

$$A^{-1}A = I = AA^{-1}.$$

A^{-1} exists $\iff Ax \neq 0$ for all $x \neq 0$

- We say that A is *invertible* or *non-singular* if A^{-1} exists and *non-invertible* or *singular* otherwise.
- In order for a square matrix A to have an inverse A^{-1}, then A must be full rank.

Properties (Assuming $A, B \in \mathbb{R}^{n \times n}$ are non-singular):

- $(A^{-1})^{-1} = A$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^{-1})^T = (A^T)^{-1}$. For this reason this matrix is often denoted A^{-T}.

$$(A+B)^{-1} \neq A^{-1} + B^{-1}$$

Solving Linear Equations: *Using Inverse*

- If exists, ***the inverse can be used*** to solve a system of linear equations
  - • Two linear equations

$$\begin{aligned} 4x_1 \quad - 5x_2 &= -13 \\ -2x_1 \quad + 3x_2 &= 9 \end{aligned}$$

• In vector form, $Ax = b$, with

$$A = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad b = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$$

• Solution using inverse

$$\begin{aligned} Ax &= b \\ A^{-1}Ax &= A^{-1}b \\ x &= A^{-1}b \end{aligned}$$

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \iff \boldsymbol{A}^{\top}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\top}\boldsymbol{b} \iff \boldsymbol{x} = (\boldsymbol{A}^{\top}\boldsymbol{A})^{-1}\boldsymbol{A}^{\top}\boldsymbol{b}$$

- Note that the inverse may not exists
- Even exists, ***the inverse is not used*** in practice
  - It cannot be represented with sufficient precision
- Gaussian elimination also has disadvantages
  - It has numerical instability (division by small no.)
  - $O(n^3)$ for n x n matrix
- Iterative algorithms are used instead
  - Many different decompositions can be used depending on the characteristics of A, x, and b

Solving Linear Equations: *Using Gaussian Elimination*

1. Find a particular solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$
2. Find all solutions to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$
3. Combine the solutions from 1. and 2. to the general solution.

A(*m*x*n*): *m* equations with *n* unknowns

- Under-determined*: m < n*
- Over-determined*: m > n*

$$\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix}$$

Reduced Row Echelon Form

Remark (Gaussian Elimination). *Gaussian elimination* is an algorithm that performs elementary transformations to bring a system of linear equations into reduced row echelon form. ◇

1. a solution is $[42, 8, 0, 0]^\top$

2.
$$\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \left(\lambda_1 \begin{bmatrix} 8 \\ 2 \\ -1 \\ 0 \end{bmatrix} \right) = \lambda_1(8\boldsymbol{c}_1 + 2\boldsymbol{c}_2 - \boldsymbol{c}_3) = \boldsymbol{0}$$

$$\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \left(\lambda_2 \begin{bmatrix} -4 \\ 12 \\ 0 \\ -1 \end{bmatrix} \right) = \lambda_2(-4\boldsymbol{c}_1 + 12\boldsymbol{c}_2 - \boldsymbol{c}_4) = \boldsymbol{0}$$

3.
$$\left\{ \boldsymbol{x} \in \mathbb{R}^4 : \boldsymbol{x} = \begin{bmatrix} 42 \\ 8 \\ 0 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} 8 \\ 2 \\ -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} -4 \\ 12 \\ 0 \\ -1 \end{bmatrix}, \lambda_1, \lambda_2 \in \mathbb{R} \right\}$$

non-pivot columns are expressed as a combination of the *pivot* columns

3. Norm, Orthogonality & Projections

1. Norm
2. Angle
3. Orthogonality
4. Projection & Least Squares

Norm ~ *length of a vector*

A ***norm*** of a vector $\|x\|$ is informally a measure of the "length" of the vector.

More formally, a norm is any function $f : \mathbb{R}^n \to \mathbb{R}$ that satisfies 4 properties:

1. For all $x \in \mathbb{R}^n$, $f(x) \geq 0$ (non-negativity).
2. $f(x) = 0$ if and only if $x = 0$ (definiteness).
3. For all $x \in \mathbb{R}^n$, $t \in \mathbb{R}$, $f(tx) = |t| f(x)$ (homogeneity).
4. For all $x, y \in \mathbb{R}^n$, $f(x + y) \leq f(x) + f(y)$ (triangle inequality).

- Note: $\|x - y\|$ is a measure of the "***distance***" of the two vectors.

p-Norms

- ℓ_2 norm (Euclidean norm)

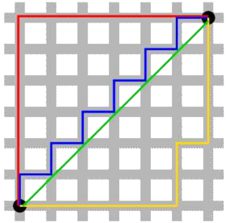
$$\|x\|_2 = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

- ℓ_1 norm (Manhattan norm)

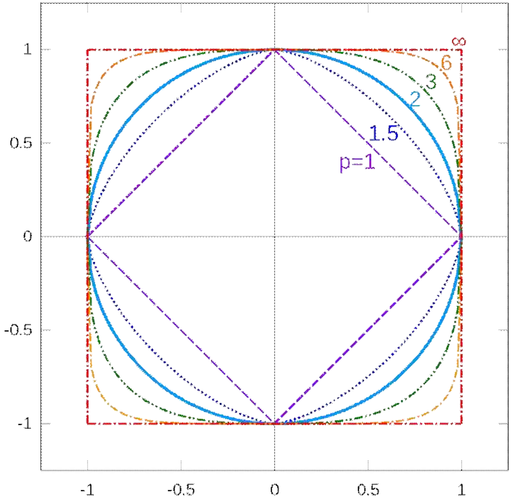
$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

- ℓ_∞ norm (Max norm)

$$\|x\|_\infty = \max_{i=1,\dots,n} |x_i|$$



Illustrations of unit circles in R^2
(length-one vectors from the origin)



In fact, all three norms presented so far are examples of the family of ℓ_p norms, which are parameterized by a real number $p \geq 1$, and defined as

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

Above figure is from: https://commons.wikimedia.org/w/index.php?curid=17428655

Frobenius Norm ~ *a norm for matrices*

$$\mathbb{R}^{m\times n} \rightarrow \mathbb{R}$$

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2} = \sqrt{\operatorname{tr}(A^TA)}.$$

Cf. Frobenius norm is used for
reg

$$J_R = J + \lambda \sum_{i=1}^{L} \left\| W^{(i)} \right\|_F$$

$$\underset{(m\times n)}{A} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ & \vdots & & \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}$$

$$\begin{aligned} a_1^Ta_1 &= A_{11}^2 + A_{21}^2 + \cdots + A_{m1}^2 \\ a_2^Ta_2 &= A_{12}^2 + A_{22}^2 + \cdots + A_{m2}^2 \\ &\vdots \\ a_n^Ta_n &= A_{1n}^2 + A_{2n}^2 + \cdots + A_{mn}^2 \end{aligned} \rightarrow A^TA = \begin{bmatrix} a_1^Ta_1 & a_1^Ta_2 & \cdots & a_1^Ta_n \\ a_2^Ta_1 & a_2^Ta_2 & \cdots & a_2^Ta_n \\ \vdots & & \ddots & \vdots \\ a_n^Ta_1 & a_n^Ta_2 & \cdots & a_n^Ta_n \end{bmatrix} \rightarrow \operatorname{tr}(A^TA) = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2$$

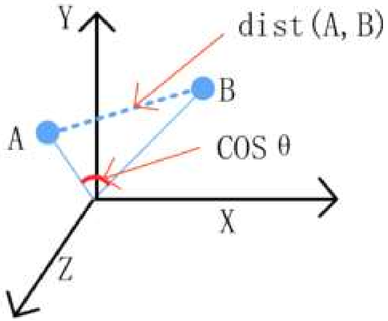
Angle *between Vectors*

Dot product of two vectors can be written in terms of their L^2 norms and angle θ between them

$$x^T y \Rightarrow ||x||_2 ||y||_2 \cos\theta$$

Cosine between two vectors is a measure of their (*orientation*) ***similarity***

$$\cos\theta = \frac{x^T y}{||x||_2 ||y||_2}$$



Orthogonal *Vectors*

- Two vectors $x, y \in \mathbb{R}^n$ are *orthogonal* if

$$x^T y = 0$$

$(\cos(\theta) = 0 \; or \; \theta = \pi/2)$ $x \perp y$

- They are *orthonormal* if, in addition,

$$\|x\|_2 = \|y\|_2 = 1$$

A vector $x \in \mathbb{R}^n$ is ***normalized*** if $\|x\|_2 = 1$.

Orthogonal *Matrices*

- A matrix $U \in \mathbb{R}^{n \times n}$ is orthogonal if all it's columns are orthonormal, i.e.,

$$U^T U = I = U U^T \quad \Longleftrightarrow \quad \text{all its columns are orthogonal to each other}$$

$U^{-1} = U^T$

(linearly independent)

Transformations with orthogonal matrices preserve Euclidean ***distances*** and ***angles***!

$$\|\boldsymbol{A}\boldsymbol{x}\|^2 = (\boldsymbol{A}\boldsymbol{x})^\top(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^\top\boldsymbol{I}\boldsymbol{x} = \boldsymbol{x}^\top\boldsymbol{x} = \|\boldsymbol{x}\|^2$$

$$\cos\omega = \frac{(\boldsymbol{A}\boldsymbol{x})^\top(\boldsymbol{A}\boldsymbol{y})}{\|\boldsymbol{A}\boldsymbol{x}\|\,\|\boldsymbol{A}\boldsymbol{y}\|} = \frac{\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{y}}{\sqrt{\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x}\boldsymbol{y}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{y}}} = \frac{\boldsymbol{x}^\top\boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|}$$

Linear Independence vs. *Orthogonality*

Consider a set of vectors **S** = $\{v_1, \dots v_n\}$

Clearly No!

S = $\{v_1, \dots v_n\}$ is **linearly independent**

S = $\{v_1, \dots v_n\}$ is **orthogonal**

Yes if $v_k \neq 0, \forall k$

$\underline{0}^T \underline{v} = 0$, $\underline{0} = 0 \cdot \underline{v}$

$\underline{0}$ is orthogonal to $\forall \underline{v}$.

$\underline{0}$ is linearly dependent to $\forall \underline{v}$.

Projection

- **Definition:** *Idempotence*
  - A ***projection matrix*** P is a linear transformation from a vector space to itself such that $P^2 = P$.
  - Such mapping is called a ***projection***.

$$\begin{tikzcd} x \in \mathbb{R}^n \arrow[r, "v = Px"] & v \in \mathbb{R}^n \arrow[loop right, "P"] \end{tikzcd}$$

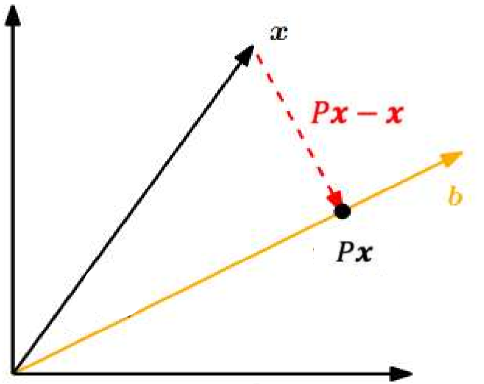
- **Examples**

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$P = \begin{bmatrix} 0 & 0 \\ \alpha & 1 \end{bmatrix}$$

Projection onto 1-D Subspaces

Assume we are given a line (1-dimensional subspace) through the origin with basis vector $\boldsymbol{b} \in \mathbb{R}^n$. The line is a one-dimensional subspace $U \subseteq \mathbb{R}^n$ spanned by $\boldsymbol{b}$. When we project $\boldsymbol{x} \in \mathbb{R}^n$ onto U, we seek the vector $\boldsymbol{v} = P\boldsymbol{x} \in U$ that is closest to $\boldsymbol{x}$.



(a) Projection of $\boldsymbol{x} \in \mathbb{R}^2$ onto a subspace U with basis vector $\boldsymbol{b}$.

The projection $\boldsymbol{v} = P\boldsymbol{x}$ is closest to $\boldsymbol{x}$.

$\Longrightarrow$ $\|P\boldsymbol{x} - \boldsymbol{x}\|$ is minimal.

$\Longrightarrow$ $(P\boldsymbol{x} - \boldsymbol{x})$ is orthogonal to $\boldsymbol{b}$.

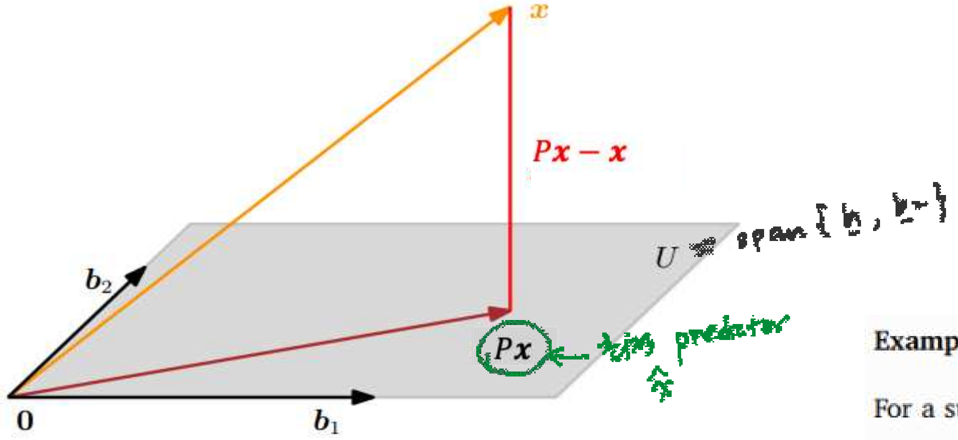
$$\Longrightarrow \quad P = \frac{\boldsymbol{b}\boldsymbol{b}^T}{\|\boldsymbol{b}\|^2}$$

Projection onto General Subspaces

$\boldsymbol{x} \in \mathbb{R}^n$

lower-dimensional subspaces $U \subseteq \mathbb{R}^n$ with $\dim(U) = m$

Assume that $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m)$ is an ordered basis of U.



$$\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m] \in \mathbb{R}^{n \times m},$$

$$\boldsymbol{P} = B(B^\top B)^{-1}B^\top .$$

Example 3.11 (Projection onto a Two-dimensional Subspace)

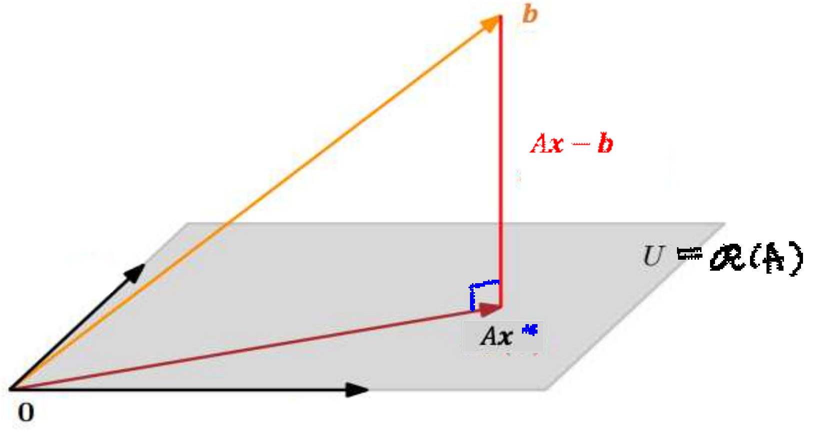
For a subspace $U = \text{span}[\underbrace{\begin{bmatrix}1\\1\\1\end{bmatrix}}_{b_1}, \underbrace{\begin{bmatrix}0\\1\\2\end{bmatrix}}_{b_2}] \subseteq \mathbb{R}^3$ and $\boldsymbol{x} = \begin{bmatrix}6\\0\\0\end{bmatrix} \in \mathbb{R}^3$

$$\boldsymbol{B} = \begin{bmatrix}1 & 0\\1 & 1\\1 & 2\end{bmatrix} \qquad \hat{x} = Px = \begin{bmatrix}5\\2\\-1\end{bmatrix} .$$

Projection & Least Squares

Using projections, we can find approximate solutions to linear equations A***x*** = ***b***.

- Suppose ***b*** does not lie in the span of A. Given that the linear equation cannot be solved exactly.
- We can find an approximate solution by computing the ***orthogonal projection*** of ***b*** onto the span of A.



$Ax^* \perp b$

This problem arises often in practice, and the solution is called the ***least-squares solution*** of an over-determined system.

$A x = b$

Squared Error

$\|Ax-b\|^2 = (Ax-b)^T(Ax-b)$

$= x^TA^TAx - b^TAx - x^TA^Tb + b^Tb$

$\frac{\partial}{\partial x}\|Ax-b\|^2 = 2A^TAx - 2A^Tb = 0$

Orthogonality: $(A\boldsymbol{x})^T(A\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{0}$

$$\Longrightarrow \boxed{\boldsymbol{x}^* = (A^TA)^{-1}A^T\boldsymbol{b}}$$

$$\Longrightarrow \boxed{\boldsymbol{x} = (A^TA)^{-1}A^T\boldsymbol{b}}$$

Least Square Solution (미분=0)

4. Determinant, Decomposition, Quadratic Forms

1. Determinant
2. Eigenvector & Eigenvalue
3. Eigendecomposition
4. Quadratic Forms
5. Positive Definite
6. Singular Value Decomposition (SVD)

Determinant ~ *volume*

The ***determinant*** of a square matrix $A \in \mathbb{R}^{n \times n}$,

is a function $\det : \mathbb{R}^{n \times n} \to \mathbb{R}$, and is denoted $|A|$ or $\det A$.

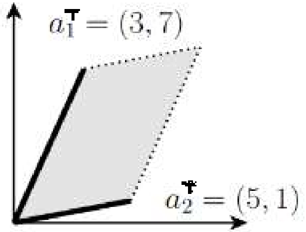
Measures how much multiplication by the matrix expands or contracts space

Given a matrix $\begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_n^T & — \end{bmatrix}$, consider the set of points $S \subset \mathbb{R}^n$ as follows:

$$S = \{v \in \mathbb{R}^n : v = \sum_{i=1}^{n} \alpha_i a_i \text{ where } 0 \le \alpha_i \le 1, i = 1, \ldots, n\}.$$

The absolute value of the determinant of A,

is a measure of the "volume" of the set S



$$A = \begin{bmatrix} 3 & 7 \\ 5 & 1 \end{bmatrix}$$

$|\det A|$ is the area of the parallelogram

Determinant: *Definition*

- Can be formally defined by three properties

1. Determinant of identity is one: $\det I = 1$

2. Multiplying a row by scalar $t \in \mathbb{R}$ scales determinant:

$$\det \begin{bmatrix} — & t a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_n^T & — \end{bmatrix} = t \det A$$

3. Swapping rows negates determinant:

$$\det \begin{bmatrix} — & a_2^T & — \\ — & a_1^T & — \\ & \vdots & \\ — & a_n^T & — \end{bmatrix} = -\det A$$

In case you are wondering, it is not immediately obvious that a function satisfying the above three properties exists. In fact, though, such a function does exist, and is unique (which we will not prove here).

Determinant: *Properties*

- Important properties

For $A, B \in \mathbb{R}^{n\times n}$,

- $\det A = \det A^T$
- $\det AB = \det A \det B$
- $\det A = 0 \Leftrightarrow$ A singular (non-invertible)
- $\det A^{-1} = 1/\det A$

Determinant: *Formula*

Let $A \in \mathbb{R}^{n\times n}$, $A_{\setminus i,\setminus j} \in \mathbb{R}^{(n-1)\times(n-1)}$ be the *matrix* that results from deleting the ith row and jth column from A.

The general (recursive) formula for the determinant is

$$\begin{aligned} |A| &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{\setminus i,\setminus j}| \quad (\text{for any } j \in 1,\ldots,n) \\ &= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |A_{\setminus i,\setminus j}| \quad (\text{for any } i \in 1,\ldots,n) \end{aligned}$$

with the initial case that $|A| = a_{11}$ for $A \in \mathbb{R}^{1\times 1}$. If we were to expand this formula completely for $A \in \mathbb{R}^{n\times n}$, there would be a total of $n!$ (n factorial) different terms. For this reason, we hardly ever explicitly write the complete equation of the determinant for matrices bigger than 3×3.

However, the equations for determinants of matrices up to size 3×3 are fairly common, and it is good to know them:

$$\begin{aligned} |[a_{11}]| &= a_{11} \\ \left|\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right| &= a_{11}a_{22} - a_{12}a_{21} \end{aligned}$$

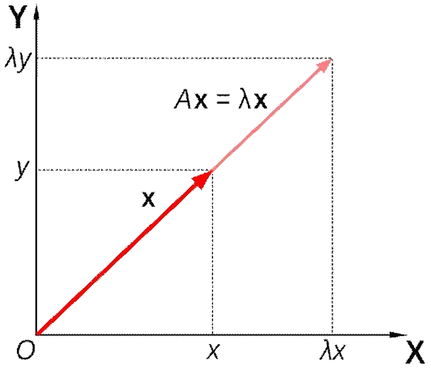
$$\left|\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right| = \begin{array}{l} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ \quad -a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \end{array}$$

Eigenvector & Eigenvalue

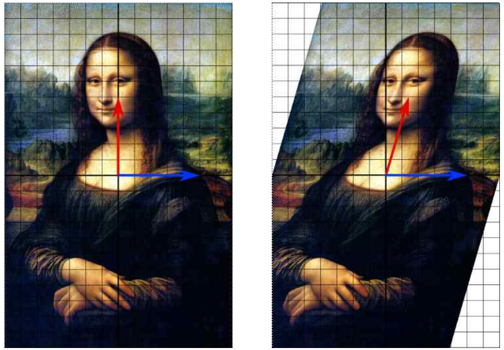
Given a square matrix $A \in \mathbb{R}^{n \times n}$, we say that $\lambda \in \mathbb{C}$ is an ***eigenvalue*** of A and $x \in \mathbb{C}^n$ is the corresponding ***eigenvector*** if

$$\boxed{Ax = \lambda x, \quad x \neq 0.}$$

Intuitively, this definition means that multiplying A by the vector x results in a new vector that points in the same direction as x, but scaled by a factor λ.



Matrix *A* acts by stretching the vector *x*, not changing its direction, so *x* is an eigenvector of *A*.



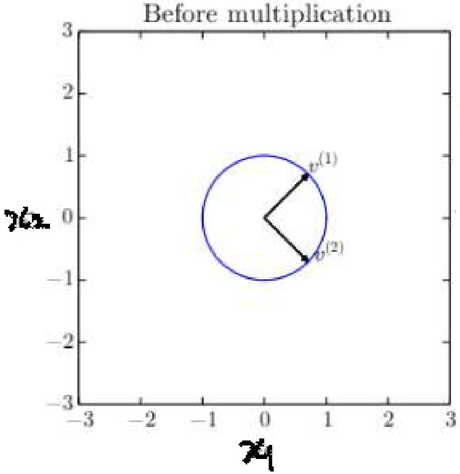
In this shear mapping, the red arrow changes direction, but the blue arrow does not. Thus the blue arrow is an eigenvector of this mapping.

Figures are from https://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors

Eigenvector & Eigenvalue (cont'd)

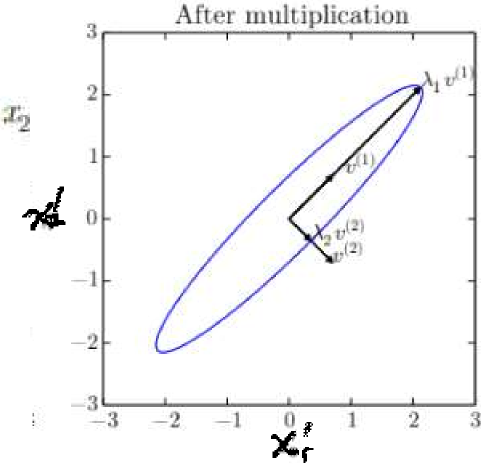
- Example of 2×2 matrix
- Matrix A with two orthonormal eigenvectors
  - $v^{(1)}$ with eigenvalue λ_1, $v^{(2)}$ with eigenvalue λ_2

Plot of unit vectors $u \in \mathbb{R}^2$ (circle)



with two variables x_1 and x_2

Plot of vectors Au (ellipse)



Characteristic Polynomial

We can rewrite the equation above to state that (λ, x) is an eigenvalue-eigenvector pair of A if,

$$(\lambda I - A)x = 0, \quad x \neq 0.$$

But $(\lambda I - A)x = 0$ has a non-zero solution to x if and only if $(\lambda I - A)$ has a non-empty nullspace, which is only the case if $(\lambda I - A)$ is singular, i.e.,

$$|(\lambda I - A)| = 0. \quad = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\ldots(\lambda_n - \lambda)$$

We can now use the previous definition of the determinant to expand this expression $|(\lambda I - A)|$ into a (very large) polynomial in λ, where λ will have degree n. It's often called the characteristic polynomial of the matrix A.

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \qquad |A - \lambda I| = \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} = 3 - 4\lambda + \lambda^2$$

$\lambda=1$,
$\lambda=3$

$$v_{\lambda=1} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, v_{\lambda=3} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Eigendecomposition

- Suppose that matrix A has n linearly independent eigenvectors $\{\boldsymbol{v}^{(1)},\dots,\boldsymbol{v}^{(n)}\}$ with eigenvalues $\{\lambda_1,\dots,\lambda_n\}$
- Concatenate eigenvectors to form matrix V
- Concatenate eigenvalues to form vector $\lambda=[\lambda_1,\dots,\lambda_n]$
- Eigendecomposition of A is given by

$$A = V\mathrm{diag}(\lambda)V^{-1}$$

$$\begin{cases} A\underline{v}_1 = \lambda_1 \underline{v}_1 \\ A\underline{v}_2 = \lambda_2 \underline{v}_2 \\ \vdots \\ A\underline{v}_n = \lambda_n \underline{v}_n \end{cases}$$

$$A\begin{bmatrix} | & & | \\ v^{(1)} & \cdots & v^{(n)} \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \lambda_1 v^{(1)} & \cdots & \lambda_n v^{(n)} \\ | & & | \end{bmatrix}$$

$$AV = V\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

column scalar (diagonal on the right)

$$\Rightarrow A = \boxed{V\Lambda V^{-1}}$$

Properties of Eigenvalues

- $\mathrm{tr}\, A = \sum_{i=1}^{n} \lambda_i$
- $\det A = \prod_{i=1}^{n} \lambda_i$
- $\mathrm{rank}(A) =$ number of non-zero eigenvalues

  Matrix is singular $\Longleftrightarrow$ any eigenvalue is zero

- Eigenvalues of A^{-1} are $1/\lambda_i, i = 1, \ldots, n,$
  eigenvectors are the same

$\because$ $tr(V \Lambda V^{-1}) = tr(\Lambda V^{-1} V) = tr(\Lambda)$

$\because$ $det(V \Lambda V^{-1}) = det V \cdot det \Lambda \cdot det V^{-1}$

$= det \Lambda$

$\left(det V^{-1} = \frac{1}{det V} \right)$

$\because$ $Ax = \lambda x$

$x = \lambda A^{-1} x$

$\frac{1}{\lambda} x = A^{-1} x$

Eigendecomposition of *Real Symmetric Matrices*

let's assume that A is a symmetric real matrix

- Every real symmetric matrix A can be decomposed into real-valued eigenvectors and eigenvalues

$$\boxed{A = Q\Lambda Q^{\mathrm{T}}}$$

where Q is an orthogonal matrix composed of eigenvectors of A: $\{\boldsymbol{v}^{(1)},\ldots,\boldsymbol{v}^{(n)}\}$

Λ is a diagonal matrix of eigenvalues $\{\lambda_1,\ldots,\lambda_n\}$

- By convention order entries of Λ in descending order:
- Decomposition is not unique when two eigenvalues are the same

Quadratic Forms

Given a square matrix $A \in \mathbb{R}^{n \times n}$ and a vector $x \in \mathbb{R}^n$, the scalar value $\boxed{x^T Ax}$ is called a ***quadratic form***. Written explicitly, we see that

$$x^T Ax = \sum_{i=1}^{n} x_i (Ax)_i = \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} A_{ij} x_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j \ .$$

$$\begin{aligned} x^T A x = \; & A_{11} x_1^2 + A_{12} x_1 x_2 + A_{13} x_1 x_3 + \cdots \\ & + A_{21} x_2 x_1 + A_{22} x_2^2 + A_{23} x_2 x_3 + \cdots \\ & + A_{31} x_3 x_1 + A_{32} x_3 x_2 + A_{33} x_3^2 + \cdots \\ & + \cdots \end{aligned}$$

a ***quadratic form*** is a polynomial with terms all of degree two

$$4x^2 + 2xy - 3y^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 4 & 1+a \\ 1-a & -3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$N(\mathrm{x} \mid \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} \underbrace{(\mathrm{x} - \boldsymbol{\mu})^T}_{x^T} \underbrace{\Sigma^{-1}}_{A} \underbrace{(\mathrm{x} - \boldsymbol{\mu})}_{x} \right\}$$

Positive Definite

A symmetric matrix $A \in \mathbb{S}^n$ is:

임의의 벡터x에 대해

- *positive definite* (PD), denoted $A \succ 0$ if for all non-zero vectors $x \in \mathbb{R}^n$, $x^T Ax > 0$.
- *positive semidefinite* (PSD), denoted $A \succeq 0$ if for all vectors $x^T Ax \geq 0$.
- *negative definite* (ND), denoted $A \prec 0$ if for all non-zero $x \in \mathbb{R}^n$, $x^T Ax < 0$.
- *negative semidefinite* (NSD), denoted $A \preceq 0$) if for all $x \in \mathbb{R}^n$, $x^T Ax \leq 0$.
- *indefinite*, if it is neither positive semidefinite nor negative semidefinite — i.e., if there exists $x_1, x_2 \in \mathbb{R}^n$ such that $x_1^T Ax_1 > 0$ and $x_2^T Ax_2 < 0$.

One important property of positive definite and negative definite matrices is that they are always full rank, and hence, invertible.

Given any matrix $A \in \mathbb{R}^{m \times n}$ (not necessarily symmetric or even square), the matrix $\boxed{G = A^T A}$ (sometimes called a ***Gram matrix***) is always positive semidefinite. Further, if $m \geq n$ and A is full rank, then $G = A^T A$ is positive definite.

증명

$\|Ax\|^2 \geq 0$

$= (Ax)^T Ax$

$= x^T \underline{A^T A} x$

Definiteness & Eigenvalue Signs

1. If all $\lambda_i > 0$, then the matrix A s positivedefinite
2. If all $\lambda_i \geq 0$, it is positive semidefinite
3. Likewise, if all $\lambda_i < 0$ or $\lambda_i \leq 0$, then A is negative definite or negative semidefinite respectively.
4. Finally, if A has both positive and negative eigenvalues, say $\lambda_i > 0$ and $\lambda_j < 0$, then it is indefinite.

Singular Value Decomposition (SVD) ★ 중요

- SVD is more general than eigendecomposition
  - If A is not square, eigendecomposition is undefined
  - Every real matrix has a SVD

$$\boxed{A = UDV^{\top}}$$

- U and V are orthogonal matrices
- D is a diagonal matrix not necessarily square
  - Elements of Diagonal of D are called *singular values of A*
  - Columns of U are called *left singular vectors*
  - Columns of V are called *right singular vectors*

- Left singular vectors of A are eigenvectors of AA^{T}
- Right singular vectors of A are eigenvectors of $A^{\mathrm{T}}A$
- Nonzero singular values of A are square roots of eigen values of $A^{\mathrm{T}}A$. Same is true of AA^{T}

밑에 설명

Use of SVD in ML

1. SVD is used in generalizing matrix inversion
   - Moore-Penrose inverse (discussed next)
2. Used in Recommendation systems
   - Collaborative filtering (CF)

cf) $A_{n\times n} = V\Lambda V^{-1}$

$A_{m\times n} = U\ \Sigma\ V^{T}$

$$m\underbrace{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}_{n} = m\underbrace{\begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix}}_{n}\underbrace{\begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & \sigma_n \end{bmatrix}}_{n} n\underbrace{\begin{bmatrix} - v_1^T - \\ - v_2^T - \\ \vdots \\ - v_3^T - \end{bmatrix}}_{n}$$

row space basis

column space basis

rank(A) = r

$\sigma_1, \sigma_2, \cdots, \sigma_r$: r non-zero eigen value (singular value)

$\sigma_{r+1}, \sigma_{r+2}, \cdots, \sigma_n$: 0

$u_1, u_2, \cdots, u_r$: orthonormal basis of row space

$v_1, v_2, \cdots, v_r$: 〃 column

Moore-Penrose Pseudoinverse 유사 역

- $A^+ = VD^+U^\top$

often we can represent data with smaller # of basis than the original # of basis

- Solution to y = Ax (using the pseudoinverse): $\boldsymbol{x} = \boldsymbol{A}^+\boldsymbol{y}$

If the equation has:

- Exactly one solution: this is the same as the inverse.
- No solution: this gives us the solution with the smallest error $||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}||_2$. least square solution
- Many solutions: this gives us the solution with the smallest norm of $\boldsymbol{x}$.

given $A_{m\times n}$ find $U\ V\ \Sigma$

$$
\begin{aligned}
A^TA &= A^TU\Sigma V^T \\
&= (U\Sigma V^T)^T U\Sigma V^T \\
&= V\Sigma^T U^T U\Sigma V^T \\
&= V\Sigma^T\Sigma V^T \\
&= V\Sigma^2 V^T
\end{aligned}
$$

with eigenvalue decomposition

get V, Σ

$$
\begin{aligned}
AA^T &= U\Sigma V^T A^T \\
&= U\Sigma U^T
\end{aligned}
$$

eigenvalue decomposition get U