Matrix algebra

Zhixiang ZHANG

University of Macau

zhixzhang@um.edu.mo

January 9, 2025

Contents

Contents

Contents

Table of Contents

checklist

- vectors: column vector, row vector, vector addition and subtraction, scalar multiplication, dot product, length (norm), unit vector, angle, linear independence
- matrix: matrix addition and subtraction, scalar multiplication, matrix multiplication (product), matrix transpose, determinant, matrix rank, invertible (non-singular) squared matrix, symmetric must be squared matrix, orthogonal squared matrix, eigendecomposition, SVD, positive-(semi) definite matrix, quadratic forms, trace
- row space, column space, basis, kernel of a linear transformation

. . .

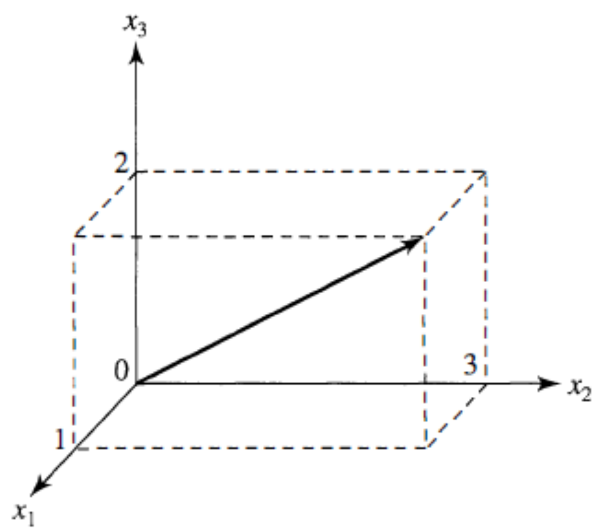
Table of Contents

Vector

An array $\mathbf{x}$ of n real numbers $x_1, \cdots x_n$ is called a vector, and it can be written as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{or} \quad \mathbf{x}^\top = [x_1, x_2, \ldots, x_n]$$

Geometrically, for $n = 3$:



Length, inner product and angle

- The *length* of a vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]^\top$, with n components, is defined by
$$\|\mathbf{x}\|_2 := \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$
This is the Euclidean norm.
- For an arbitrary number of dimensions n, we define the *inner product* of $\mathbf{x}$ and $\mathbf{y}$ as
$$\mathbf{x}^\top \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$
The inner product is denoted by either $\mathbf{x}^\top \mathbf{y}$ or $\mathbf{y}^\top \mathbf{x}$.
- The *angle* of two vectors $\mathbf{x}$ and $\mathbf{y}$ satisfies
$$\cos(\theta) = \frac{\mathbf{x}^\top \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} = \frac{\mathbf{x}^\top \mathbf{y}}{\sqrt{\mathbf{x}^\top \mathbf{x}} \sqrt{\mathbf{y}^\top \mathbf{y}}}.$$
We say $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal* (or perpendicular) if $\mathbf{x}^\top \mathbf{y}$ is zero.

Example

Given the vectors $\mathbf{x}^\top = [1, 3, 2]$ and $\mathbf{y}^\top = [-2, 1, -1]$, find $3\mathbf{x}$ and $\mathbf{x} + \mathbf{y}$. Next, determine the length of $\mathbf{x}$, the length of $\mathbf{y}$, and the angle between $\mathbf{x}$ and $\mathbf{y}$.

First,

$$3\mathbf{x} = 3\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \\ 6 \end{bmatrix}$$

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} + \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1-2 \\ 3+1 \\ 2-1 \end{bmatrix} = \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}$$

Next, $\mathbf{x}^\top\mathbf{x} = 1^2 + 3^2 + 2^2 = 14, \mathbf{y}^\top\mathbf{y} = (-2)^2 + 1^2 + (-1)^2 = 6$, and $\mathbf{x}^\top\mathbf{y} = 1(-2) + 3(1) + 2(-1) = -1$. Therefore,

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^\top\mathbf{x}} = \sqrt{14} = 3.742 \quad \|\mathbf{y}\|_2 = \sqrt{\mathbf{y}^\top\mathbf{y}} = \sqrt{6} = 2.449$$

and

$$\cos(\theta) = \frac{\mathbf{x}^\top\mathbf{y}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2} = \frac{-1}{3.742 \times 2.449} = -.109$$

so $\theta = 96.3°$.

Linear independence

Linear independence

A set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ is said to be *linearly dependent* if there exist constants $c_1, c_2, \ldots, c_k$, not all zero, such that

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k = \mathbf{0}$$

Linear dependence implies that at least one vector in the set can be written as a linear combination of the other vectors. Vectors of the same dimension that are not linearly dependent are said to be *linearly independent*.

Example

Consider the set of vectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

Setting

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 = \mathbf{0}$$

implies that

$$\begin{aligned} c_1 + c_2 + c_3 &= 0 \\ 2c_1 - 2c_3 &= 0 \\ c_1 - c_2 + c_3 &= 0 \end{aligned}$$

with the unique solution $c_1 = c_2 = c_3 = 0$. As we cannot find three constants c_1, c_2, and c_3, not all zero, such that $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 = \mathbf{0}$, the vectors $\mathbf{x}_1, \mathbf{x}_2$, and $\mathbf{x}_3$ are linearly independent.

Projection of a vector on another vector

The *projection* of a vector $\mathbf{x}$ on a vector $\mathbf{y}$ is

$$\text{Projection of } \mathbf{x} \text{ on } \mathbf{y} = \frac{(\mathbf{x}^\top \mathbf{y})}{\mathbf{y}^\top \mathbf{y}} \mathbf{y} = \frac{(\mathbf{x}^\top \mathbf{y})}{\|\mathbf{y}\|_2} \frac{1}{\|\mathbf{y}\|_2} \mathbf{y}$$

where the vector $\|\mathbf{y}\|_2^{-1} \mathbf{y}$ has unit length. The length of the projection is

$$\text{Length of projection } = \frac{|\mathbf{x}^\top \mathbf{y}|}{\|\mathbf{y}\|_2} = \|\mathbf{x}\|_2 \left| \frac{\mathbf{x}^\top \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \right| = \|\mathbf{x}\|_2 |\cos(\theta)|$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$.

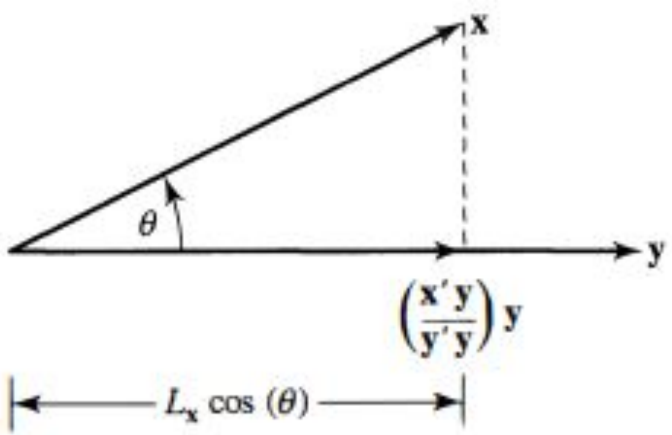


Table of Contents

Matrix multiplication (product)

Matrix multiplication

The *matrix multiplication* $\mathbf{AB}$ of an $m \times n$ matrix $\mathbf{A} = \{a_{ij}\}$ and an $n \times k$ matrix $\mathbf{B} = \{b_{ij}\}$ is the $m \times k$ matrix $\mathbf{C}$ whose elements are

$$c_{ij} = \sum_{\ell=1}^{n} a_{i\ell} b_{\ell j} \quad i = 1, 2, \ldots, m \quad j = 1, 2, \ldots, k$$

Note that for the product $\mathbf{AB}$ to be defined, the column dimension of $\mathbf{A}$ must equal the row dimension of $\mathbf{B}$.

Let $\mathbf{I}$ denote the square matrix with ones on the diagonal and zeros elsewhere, then $\mathbf{AI} = \mathbf{IA} = \mathbf{A}$ The matrix $\mathbf{I}$ acts like 1 in ordinary multiplication ($1 \cdot a = a \cdot 1 = a$), so it is called the *identity matrix*.

Matrix transpose

- The *transpose* operation $\mathbf{A}^\top$ of a matrix changes the columns into rows, so that the first column of $\mathbf{A}$ becomes the first row of $\mathbf{A}^\top$, the second column becomes the second row, and so forth. In other words, $(\mathbf{A}^\top)_{ji} = (\mathbf{A})_{ij}$.

  reflect ~~rotate~~ in 90° degree

- Example: If

$$\underset{(2\times 3)}{\mathbf{A}} = \begin{bmatrix} 3 & -1 & 2 \\ 1 & 5 & 4 \end{bmatrix}$$

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

$A = (\cdots) \in \mathbb{R}^{m\times n}$

then

$$\underset{(3\times 2)}{\mathbf{A}^\top} = \begin{bmatrix} 3 & 1 \\ -1 & 5 \\ 2 & 4 \end{bmatrix}$$

$A^T = (\quad) \in \mathbb{R}^{n\times m}$

$A_{31} = (A^T)_{13}$

Properties of matrix multiplication

implemented.

$x \in (\quad) \in \mathbb{R}^{n \times p}$ image data. $(\quad)$ stored in a matrix. pcla tools.

e.g. $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times k}$

For all matrices $\mathbf{A}, \mathbf{B}$. and $\mathbf{C}$ (of dimensions such that the indicated products are defined) and a scalar c, verify: $(AB)^T = B^T A^T$: prove $(AB)^T = B^T A^T$.

ij entry is the same for varification.

① $(AB)^T_{ij}$

② $B^T A^T_{ij}$

For any $1 \le i \le k$, $1 \le j \le m$

② $(B^T A^T)_{ij} \overset{\text{def of multi}}{=} \sum_{b=1}^{n} (B^T)_{ib} (A^T)_{bj} \overset{\text{def of T}}{=} \sum_{b=1}^{n} B_{bi} A_{jb} = \sum_{b=1}^{n} A_{jb} B_{bi} \overset{\text{def of multi}}{=} (AB)_{ji}$

① $(AB)^T_{ij} \overset{\text{def of T}}{=} (AB)_{ji}$

$\therefore (B^T A^T)_{ij} = (AB)^T_{ij}$

$\therefore B^T A^T = (AB)^T$

$(ABC)^T = C^T B^T A^T$

- $c(\mathbf{AB}) = (c\mathbf{A})\mathbf{B}$
- $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$
- $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
- $(\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA}$
- $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$

There are several important differences between the algebra of matrices and the algebra of real numbers. Two of these differences are as follows:

Matrix multiplication

product of two matrices $(X^TX)^{-1}X^Ty$

$m\times n \cdot n\times k$

$R^{m\times n}$ jth $R^{n\times k}$

ith A B

$C=AB$ C_{ij}-entry

$C_{ij} = \sum A_{ik}B_{kj}$, $k=1,2,\cdots,n$.

independent of i and j

only k is changing

1. Matrix multiplication is, in general, not commutative. That is, in general, $\mathbf{AB} \neq \mathbf{BA}$. For example

$$\begin{bmatrix} 3 & -1 \\ 4 & 7 \end{bmatrix}\begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 \\ 14 \end{bmatrix}$$

$a\cdot I = I\cdot a = a$

a is a number

$AI = IA = A$.

$I = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$

but

$2\times1 \cdot 2\times2$ not defined

$$\begin{bmatrix} 0 \\ 2 \end{bmatrix}\begin{bmatrix} 3 & -1 \\ 4 & 7 \end{bmatrix}$$

is not defined.

Matrix multiplication

2. Let $\mathbf{0}$ denote the zero matrix, that is, the matrix with zero for every element. In the algebra of real numbers, if the product of two numbers, ab, is zero, then $a = 0$ or $b = 0$. In matrix algebra, however, the product of two nonzero matrices may be the zero matrix. Hence,

$$\underset{(m\times n)(n\times k)}{\mathbf{AB}} = \underset{(m\times k)}{\mathbf{0}}$$

does not imply that $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$. For example,

each entry is 0

$$\begin{bmatrix} 3 & 1 & 3 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \\ -5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

It is true, however, that if either $\underset{(m\times n)}{\mathbf{A}} = \underset{(m\times n)}{\mathbf{0}}$ or $\underset{(n\times k)}{\mathbf{B}} = \underset{(n\times k)}{\mathbf{0}}$, then $\underset{(m\times n)(n\times k)}{\mathbf{AB}} = \underset{(m\times k)}{\mathbf{0}}$.

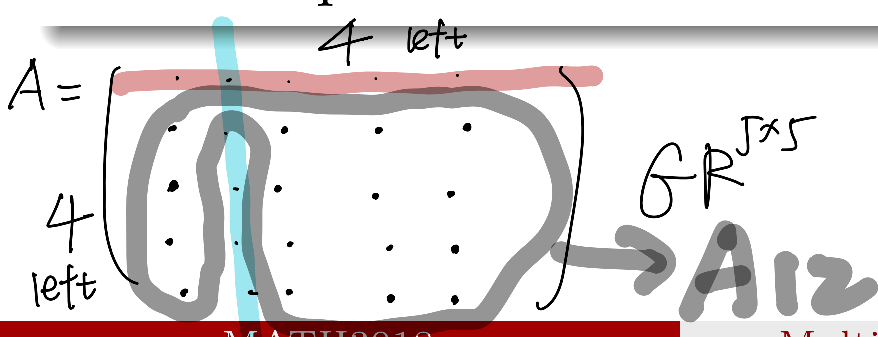
Determinant in matrix inverse

Determinant recursively

The *determinant* of the square $k \times k$ matrix $\mathbf{A} = \{a_{ij}\}$, denoted by $|\mathbf{A}|$, is the scalar

$$
\begin{aligned}
|\mathbf{A}| &= a_{11} && \text{if } k = 1 \\
|\mathbf{A}| &= \textstyle\sum_{j=1}^{k} a_{1j}\,|\mathbf{A}_{1j}|\,(-1)^{1+j} && \text{if } k > 1
\end{aligned}
$$

where $\mathbf{A}_{1j}$ is the $(k-1) \times (k-1)$ matrix obtained by deleting the first row and j th column of $\mathbf{A}$. Also, $|\mathbf{A}| = \sum_{j=1}^{k} a_{ij}\,|\mathbf{A}_{ij}|\,(-1)^{i+j}$, with the i th row in place of the first row.



Example

fast in R for det()

- If $\mathbf{I}$ is the $k \times k$ identity matrix, $|\mathbf{I}| = 1$.
-

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22}(-1)^2 + a_{12}a_{21}(-1)^3 = a_{11}a_{22} - a_{12}a_{21}$$

-

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} (-1)^2 + a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} (-1)^3 + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} (-1)^4$$

$$= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{21}a_{32}a_{13} - a_{31}a_{22}a_{13} - a_{21}a_{12}a_{33} - a_{32}a_{23}a_{11}$$

Rank important, close to PCA Interception, Factor Analysis

- The *row rank* of a matrix is the maximum number of linearly independent rows, considered as vectors (that is, row vectors). The *column rank* of a matrix is the rank of its set of columns, considered as $\boxed{\text{vectors.}}$
- The row rank and the column rank of a matrix are equal. Thus, the rank of a matrix is either the row rank or the column rank.
- For example, for the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 5 & -1 \\ 0 & 1 & -1 \end{bmatrix} \xrightarrow[\text{rank}]{\substack{(-2)\text{row}_1 + \text{row}_2 \\ \text{row reduction} \\ \text{Solve}}} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 3 & -3 \\ 0 & 1 & -1 \end{bmatrix}$$

the rank is two. $\Rightarrow$ row rank $= 2$ $\Rightarrow$ rank $= 2$

$$\xrightarrow{-3\text{row}_3 + \text{row}_2} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 3 & -3 \\ 0 & 0 & 0 \end{bmatrix}$$

Matrix inverse

Matrix inverse

If there exists a matrix $\mathbf{B}$ such that

$$\underbrace{\underset{(k\times k)}{\mathbf{B}}\underset{(k\times k)}{\mathbf{A}} = \underset{(k\times k)}{\mathbf{A}}\underset{(k\times k)}{\mathbf{B}} = \underset{(k\times k)}{\mathbf{I}}}_{},$$

then $\underline{\mathbf{B}}$ is called the *inverse* of $\mathbf{A}$ and is denoted by $\mathbf{A}^{-1}$.

might have MP-inverse / pseudo inverse not discussed. no squared matrix

For a $k \times k$ matrix $\mathbf{A}$:

the existence of $\mathbf{A}^{-1}$

rank(A) = k = rank_row(A) = rank_column(A)

$\overset{\text{or}}{\iff}$ the columns (rows) of $\mathbf{A}$ are linear independent

$\overset{\text{or}}{\iff}$ $\mathbf{A}$ has full column (row) rank

$\overset{\text{or}}{\iff}$ $|\mathbf{A}| \neq 0$

Matrix inverse

calculated

- In general, $\mathbf{A}^{-1}$ has (i,j)-th entry $\left[(-1)^{j+i}\left|\mathbf{A}_{ji}\right|/|\mathbf{A}|\right]$, where $\mathbf{A}_{ji}$ is the matrix obtained from $\mathbf{A}$ by deleting the j th row and i th column.
- Specifically, the inverse of any 2×2 matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

is given by

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

△ same as questions

Some properties of matrix inverse and determinant

Let $\mathbf{A}$ and $\mathbf{B}$ be square matrices of the same dimension, and let the indicated inverses exist. Then the following hold:

- $(\mathbf{A}^{-1})^\top = (\mathbf{A}^\top)^{-1}$
- $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

Let $\mathbf{A}$ an $\mathbf{B}$ be $k \times k$ square matrices.

- $|\mathbf{A}| = |\mathbf{A}^\top|$
- If each element of a row (column) of $\mathbf{A}$ is zero, then $|\mathbf{A}| = 0$
- If any two rows (columns) of $\mathbf{A}$ are identical, then $|\mathbf{A}| = 0$
- If $\mathbf{A}$ is nonsingular, then $|\mathbf{A}| = 1/|\mathbf{A}^{-1}|$; that is, $|\mathbf{A}||\mathbf{A}^{-1}| = 1$.
- $|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}|$
- $|c\mathbf{A}| = c^k|\mathbf{A}|$, where c is a scalar.

$\det(cA) = c^k \det(A) \quad A \in R^{k\times k}$

$cA = \begin{bmatrix} cA_{11} & cA_{12} & \cdots & cA_{1k} \\ cA_{21} & cA_{22} & & \\ \vdots & & \ddots & \\ cA_{k1} & & & cA_{kk} \end{bmatrix} \sim \sum A_{i_1j_1}' A_{i_2j_2}' \cdots A_{i_kj_k}' \sim \sum c^k \cdot A_{i_1j_1} A_{i_2j_2} \cdots A_{i_kj_k}$

Diagonal matrix

- The calculations of matrix inverse are usually proceeded by a computer, especially when the dimension is greater than three.
- Diagonal matrices have inverses that are easy to compute. For example,

$\det = \sum_{i=1}^{k} a_{ii}$

$$\begin{bmatrix} a_{11} & 0 & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 & 0 \\ 0 & 0 & a_{33} & 0 & 0 \\ 0 & 0 & 0 & a_{44} & 0 \\ 0 & 0 & 0 & 0 & a_{55} \end{bmatrix} \text{ has inverse } \begin{bmatrix} \frac{1}{a_{11}} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{a_{22}} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{a_{33}} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{a_{44}} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{a_{55}} \end{bmatrix}$$

if all the $a_{ii} \neq 0$.

similarly $A^T A$.

$(AA^T)^T = (A^T)^T A^T = AA^T \Rightarrow AA^T$ symmetric.

Symmetric matrix and Orthogonal matrix

$(A+B+C)^T = A^T + B^T + C^T$

$X \in \mathbb{R}^{n \times p}$ $S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T$

$= \frac{1}{n-1} X^T (I - \frac{1}{n} 1 1^T) X$

$(AB)^T = B^T A^T$

$S^T = S$: verified: $S^T = \left(\frac{1}{n-1}\sum (X_i - \bar{X})(X_i - \bar{X})^T\right)^T = \frac{1}{n-1}\sum (X_i - \bar{X})(X_i - \bar{X})^T = S$

- A square matrix is said to be *symmetric* if $\mathbf{A} = \mathbf{A}^\top$ or $a_{ij} = a_{ji}$ for all i and j.
- Another special class of square matrices are the orthogonal matrices, characterized by

eigenvectors

why Q is invertible for orthogonal matrix

$QQ^T = I$ $Q^{-1} Q Q^T = Q^{-1} I \Rightarrow Q^T = Q^{-1}$

$\det(AB) = \det(A)\det(B)$

$\det(QQ^T) = \det(I) = 1$

$$\mathbf{Q}\mathbf{Q}^\top = \mathbf{Q}^\top\mathbf{Q} = \mathbf{I} \text{ or } \mathbf{Q}^\top = \mathbf{Q}^{-1}$$

$= \det(Q)\det(Q^T)$

$= (\det(Q))^2$

The name derives from the property that if $\mathbf{Q}$ has i th row $\mathbf{q}_i^\top$, then $\mathbf{Q}\mathbf{Q}^\top = \mathbf{I}$ implies that $\mathbf{q}_i^\top \mathbf{q}_i = 1$ and $\mathbf{q}_i^\top \mathbf{q}_j = 0$ for $i \neq j$, so the rows have unit length and are mutually perpendicular (orthogonal). According to the condition $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$, the columns have the same property.

If $Q = \begin{pmatrix} k_1^T \\ k_2^T \\ \vdots \\ k_k^T \end{pmatrix}$ $QQ^T = \begin{pmatrix} k_1^T \\ k_2^T \\ \vdots \\ k_k^T \end{pmatrix} \begin{pmatrix} k_1 & k_2 & \cdots & k_k \end{pmatrix}$

Eigenvalues and eigenvectors

$$= \begin{bmatrix} k_1^T k_1 & \cdots & k_1^T k_k \\ & k_2^T k_2 & \\ k_k^T k_1 & & \ddots \end{bmatrix}$$

A square matrix **A** is said to have an *eigenvalue* λ, with corresponding *eigenvector* $\mathbf{x} \neq \mathbf{0}$, if

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}.$$

$A(2x) = \lambda(2x)$ ✓

$Ax = (2\lambda)x$ ✗

$$= \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

- Let **A** be a $k \times k$ square matrix and **I** be the $k \times k$ identity matrix. Then the scalars $\lambda_1, \lambda_2, \ldots, \lambda_k$ satisfying the polynomial equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$ (determinant function.) are the eigenvalues (or characteristic roots) of a matrix **A**. The equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$ (as a function of λ) is called the characteristic equation. $Ax = \lambda x$

$Ax_i = \lambda_i x_i : \det(A - \lambda I) = 0. = \det\begin{pmatrix} a_{11}-\lambda, a_{12}, \cdots, a_{1k} \\ a_{21} \quad a_{22}-\lambda \\ \vdots \quad \ddots \\ a_{k1} \quad a_{kk}-\lambda \end{pmatrix} = C_k\lambda^k + C_{k-1}\lambda^{k-1} + \cdots + C_1\lambda + C_0 = 0$

C_i dependents on entries of A.

- The eigenvector associated with λ_i is the solution **x** to $(A - \lambda_i\mathbf{I})\mathbf{x} = \mathbf{0}$. It is usual practice to determine an eigenvector so that it has length unity.

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{k-1} \geq \lambda_k$

Solve $(A - \lambda_i I_k)X = 0$ to find X_i

$\lambda^2 - 4\lambda + 4 = 0$

$\lambda_1 = \lambda_2 = 2$ multiple eigenvalue

if $\det = 0$

$\det \neq 0 \Rightarrow$ inverse $\Rightarrow X = (A - \lambda_i I_k)^{-1} \cdot 0$

$Ax = \lambda x \quad (x \neq 0)$

why eigenvalues solve $\det(A-\lambda I) = 0$.

$Ax = \lambda x \Leftrightarrow (A - \lambda I)x = 0$

$\overset{\text{this means}}{\Longleftrightarrow} x_1 \cdot 1_{col} \text{ of } A-\lambda I + x_2 \cdot 2_{col} \text{ of } A-\lambda I + \dots + x_k \cdot k_{col} \text{ of } A-\lambda I$

$= 0. \Longrightarrow$ linear dependent

$(x \neq 0)$

$\Longleftrightarrow \text{rank}(A-\lambda I) < k$

$\Longleftrightarrow \det(A-\lambda I) = 0.$

Eigenvalue decomposition

$$A^{\in\mathbb{R}^{k\times k}} = U\Lambda U^T = \overset{\text{orthogonal}}{(u_1, u_2, \cdots, u_k)} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_k \end{pmatrix} \begin{pmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_k^T \end{pmatrix}$$

$$= (\lambda_1 u_1, \lambda_2 u_2, \cdots, \lambda_k u_k) \begin{pmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_k^T \end{pmatrix}$$

$$= \sum_{i=1}^{k} \lambda_i u_i u_i^T$$

usually $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$

Theorem [eigenvalue decomposition of symmetric matrix]

might $\lambda_i = 0$, might $\lambda_i = \lambda_j$ where $i < j$

Let $\mathbf{A}$ be a $k \times k$ square symmetric matrix. Then $\mathbf{A}$ can be expressed in terms of its k eigenvalue-eigenvector pairs $(\lambda_i, \mathbf{u}_i)$ as

If $\lambda_i \neq \lambda_j$, then $u_i^T u_j = 0$.

$$\begin{cases} Au_i = \lambda_i u_i \rightarrow u_j^T A u_i = u_j^T \lambda_i u_i \\ Au_j = \lambda_j u_j \rightarrow u_i^T A u_j = u_i^T \lambda_j u_j \end{cases}$$

$$\mathbf{A} = \sum_{i=1}^{k} \lambda_i \mathbf{u}_i \mathbf{u}_i^\top .$$

In a matrix form, we can write $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$ where $\mathbf{U}$ is an $k \times k$ orthogonal matrix whose ith column is the eigenvector $\mathbf{u}_i$ of $\mathbf{A}$, and $\boldsymbol{\Lambda}$ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues, and usually sorted in descending order.

why we have such decomposition:

For each distinct eigenvalue find associated eigenvector

For multiple eigenvalue, Gram-schmidt to get orthogonal basis

$AU = (Au_1, Au_2, \cdots, Au_k) = (\lambda_1 u_1, \lambda_2 u_2, \cdots, \lambda_k u_k)$

$= (u_1, u_2, \dots, u_k) \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & \vdots & \lambda_k \end{bmatrix}$

$= U\Lambda$

Example

$A = AI = AUU^T = U\Lambda U^T$

$\begin{pmatrix} 1 & \\ & 1 \end{pmatrix} = U \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} U^T$

where $U = \begin{bmatrix} \sin\theta & \cos\theta \\ -\cos\theta & \sin\theta \end{bmatrix}$ orthogonal transform in 2-dimensional case

u_1? $(A - \lambda_1 I_2)x = 0$

Solve x

If x_1 solves $(A - \lambda_1 I_2) x_1 = 0$

So $2x_1$ solves $(A - \lambda_1 I_2) x_1 = 0$

So make x_1 has unit length

For example, let

$$\mathbf{A} = \begin{bmatrix} 2.2 & .4 \\ .4 & 2.8 \end{bmatrix}$$

Then

$u_1 = \frac{x_1}{\|x_1\|_2}$

$$|\mathbf{A} - \lambda \mathbf{I}| = \lambda^2 - 5\lambda + 6.16 - .16 = (\lambda - 3)(\lambda - 2)$$

so $\mathbf{A}$ has eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 2$. The corresponding eigenvectors are $\mathbf{u}_1^\top = [1/\sqrt{5}, 2/\sqrt{5}]$ and $\mathbf{u}_2^\top = [2/\sqrt{5}, -1/\sqrt{5}]$, respectively. Consequently,

not unique as $-u_1^T$ solve the equation.

check that $u_1^T u_2 = 0$

$u_2^T u_1 = 0$

$$\mathbf{A} = \begin{bmatrix} 2.2 & .4 \\ .4 & 2.8 \end{bmatrix} = 3 \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} + 2 \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{bmatrix}.$$

Why eigenvalue decomposition important

EigenValue Decomposition EVD illustrated

Singular Value Decomposition SVD : $A \in R^{n \times d} = \underset{n\times n}{U} \underset{n\times d}{S} \underset{d\times d}{V^T}$ (U: orthogonal; V: orthogonal)

$d \le n$ case

$= (u_1, \cdots, u_d, \cancel{u_{d+1}, \cdots, u_n}) \begin{pmatrix} s_1 & & \\ & s_2 & 0 \\ & & s_d \\ & 0 & \end{pmatrix}_{n\times d} \begin{pmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_d^T \end{pmatrix}_{d\times d}$

singular vectors $n\times n$; singular value matrix; not important

$= (u_1 \cdots u_d) \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_d \end{pmatrix} \begin{pmatrix} v_1^T \\ \vdots \\ v_d^T \end{pmatrix}$

$R^{n\times d}$ $R^{d\times d}$ $R^{d\times d}$

PCA or Factor Analysis

- The spectral decomposition is an important analytical tool. With the eigen-decomposition of sample covariance matrices, we are very easily able to demonstrate certain statistical results (later in this course).
- Another advantage is certain matrix operation becomes easier, e.g., calculating A^{-1} or A^{100}. We can also find the square root of a positive-definite matrix using eigenvalue decomposition.

$A^{-1} = (U\Lambda U^T)^{-1} = (U^T)^{-1}\Lambda^{-1}U^{-1} = U\Lambda^{-1}U^T = (u_1, \cdots, u_k)\begin{bmatrix} \frac{1}{\lambda_1} & & \\ & \ddots & \\ & & \frac{1}{\lambda_k} \end{bmatrix}\begin{pmatrix} u_1^T \\ \vdots \\ u_k^T \end{pmatrix}$ $UU^T = I \Rightarrow U^T = U^{-1}$

$A^{100} = U\Lambda U^T U\Lambda U^T \cdots U\Lambda U^T$
$= U\Lambda^{100}U^T$
$= (u_1, \cdots, u_k)\begin{bmatrix} \lambda_1^{100} & & \\ & \ddots & \\ & & \lambda_k^{100} \end{bmatrix}\begin{bmatrix} u_1 \\ \vdots \\ u_k \end{bmatrix}$

Singular Value Decomposition

A *Singular Value Decomposition* (SVD) of an $n \times d$ matrix A expresses the matrix as the product of three "simple" matrices:

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$$

where:

- $\mathbf{U}$ is an $n \times n$ orthogonal matrix
- $\mathbf{V}$ is an $d \times d$ orthogonal matrix;
- $\mathbf{S}$ is an $n \times d$ matrix with non-zero entries on main-diagonal, sorted in descending order.

SVD

$d>n$

$A = \tilde{U}\,\tilde{S}\,V$

$d\times n \quad n\times n \quad n\times n$

$\sum_{i=1}^{\min(n,d)} s_i u_i v_i^T$

$n>d$

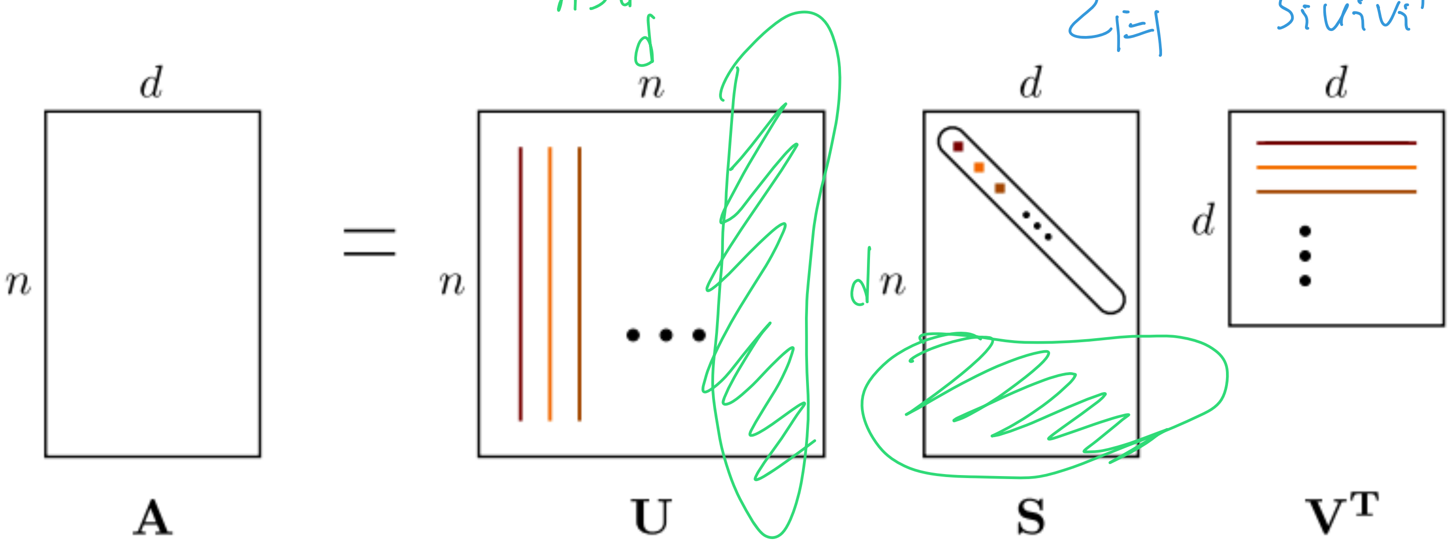


Figure: The singular value decomposition (SVD). Each singular value in **S** has an associated left singular vector in **U**, and right singular vector in **V**.

SVD

$A = USV^T ? \in \mathbb{R}^{n\times d}$ eigen decomposition

$AA^T = USV^TVS^TU$
$= USS^TU$
$= \sum_{i=1}^{n} s_i^2 u_i u_i^T$

$s_i = \sqrt{\lambda_i}$ where λ_i are non-zero eigenvalues of AA^T

U_i are eigenvectors.

$X = \sum_{i=1}^{320} s_i u_i v_i^T$ $\quad s_1 \geq s_2 \geq \cdots \geq s_{320}$

$X_5 = \sum_{i=1}^{5} s_i u_i v_i^T$

$s_1 = 10$
$s_2 = 5$
$s_{10} = 1$
$s_u = 0.11$

$5\times(n+d+1) \leq n\times d$

- The columns of $\mathbf{U}$ and $\mathbf{V}$ are called the *left singular vectors* and *right singular vectors* of $\mathbf{A}$, respectively. The entries on the main-diagonal position are *singular values* of $\mathbf{A}$.
- Another way to write the SVD is $\mathbf{A} = \sum_{i=1}^{\min\{n,d\}} s_i \mathbf{u}_i \mathbf{v}_i^\top$, where $\mathbf{u}_i$, $\mathbf{v}_i$ are the i-th column of $\mathbf{U}$, $\mathbf{V}$, respectively.
- How to obtain SVD of $\mathbf{A}$? The columns of $\mathbf{V}$ are the eigenvectors of $\mathbf{A}^\top\mathbf{A}$, the singular values are the square roots of the eigenvalues of $\mathbf{A}^\top\mathbf{A}$, and the columns of $\mathbf{U}$ are the eigenvectors of $\mathbf{A}\mathbf{A}^\top$. Remark: When $n > d$, computing those eigenvectors associated with non-zero eigenvalues of $\mathbf{A}\mathbf{A}^\top$ is sufficient to obtain the SVD.

$A^TA = VS^TU^TUSV^T$
$= VS^TSV^T$

$\underbrace{\begin{pmatrix} \ddots & 0 \end{pmatrix}}_{S^T} \underbrace{\begin{pmatrix} \ddots \\ 0 \end{pmatrix}}_{S} \Rightarrow \begin{bmatrix} s_1^2 & & & 0 \\ & s_2^2 & & \\ 0 & & \ddots & \\ & & & s_{\min(n,d)}^2 \end{bmatrix}$

Positive definite matrix and quadratic form

Positive-definite matrix& quadratic form

When a $k \times k$ symmetric matrix $\mathbf{A}$ satisfies

$$\mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0$$

for all $\mathbf{x}^\top = [x_1, x_2, \ldots, x_k] \in \mathbb{R}^k$, the matrix $\mathbf{A}$ is said to be *positive semi-definite* (or *nonnegative definite*), denoted by $\mathbf{A} \succeq 0$. If equality holds only for the vector $\mathbf{x}^\top = [0, 0, \ldots, 0]$, then $\mathbf{A}$ is said to be *positive definite*, denoted by $\mathbf{A} \succ 0$.

Because $\mathbf{x}^\top \mathbf{A}\mathbf{x}$ has only squared terms x_i^2 and product terms $x_i x_k$, it is called a *quadratic form*. If $\mathbf{A}$ is positive definite, $\mathbf{x}^\top \mathbf{A}\mathbf{x}$ is called a positive definite quadratic form.

A $k \times k$ symmetric matrix $\mathbf{A}$ is a positive definite matrix if and only if every eigenvalue of $\mathbf{A}$ is positive!

Quadratic form and Distance

AA^T

For any matrix $A \in \mathbb{R}^{n \times n}$

If we use the distance formula introduced in Chapter 1, the distance from the origin satisfies the general formula

$AA^T \succeq 0$

$$(\text{ distance })^2 = a_{11}x_1^2 + a_{22}x_2^2 + \cdots + a_{pp}x_p^2$$
$$+ 2 (a_{12}x_1x_2 + a_{13}x_1x_3 + \cdots + a_{p-1,p}x_{p-1}x_p)$$

For any $x \in \mathbb{R}^n$

x^TAA^Tx

$(A^Tx)^T(A^Tx)$

provided that (distance) $)^2 > 0$ for all $[x_1, x_2, \ldots, x_p] \neq [0, 0, \ldots, 0]$.
Setting $a_{ij} = a_{ji}$, $i \neq j$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, p$, we have

$= y^Ty = \|y\|_2 \geq 0.$

$E((X_i - E(X_i))(X_i - E(X_i))^T)$

$$0 < (\text{ distance })^2 = [x_1, x_2, \ldots, x_p] \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

or

$$0 < (\text{ distance })^2 = \mathbf{x}^\top \mathbf{A} \mathbf{x} \quad \text{for } \mathbf{x} \neq \mathbf{0}.$$

Geometrical understanding

- Expressing distance as the square root of a positive definite quadratic form allows us to give a geometrical interpretation based on the eigenvalues and eigenvectors of the matrix $\mathbf{A}$.
- For example, suppose $p = 2$. Then the points $\mathbf{x}^\top = [x_1, x_2]$ of constant distance c from the origin satisfy

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} = a_{11} x_1^2 + a_{22} x_2^2 + 2 a_{12} x_1 x_2 = c^2$$

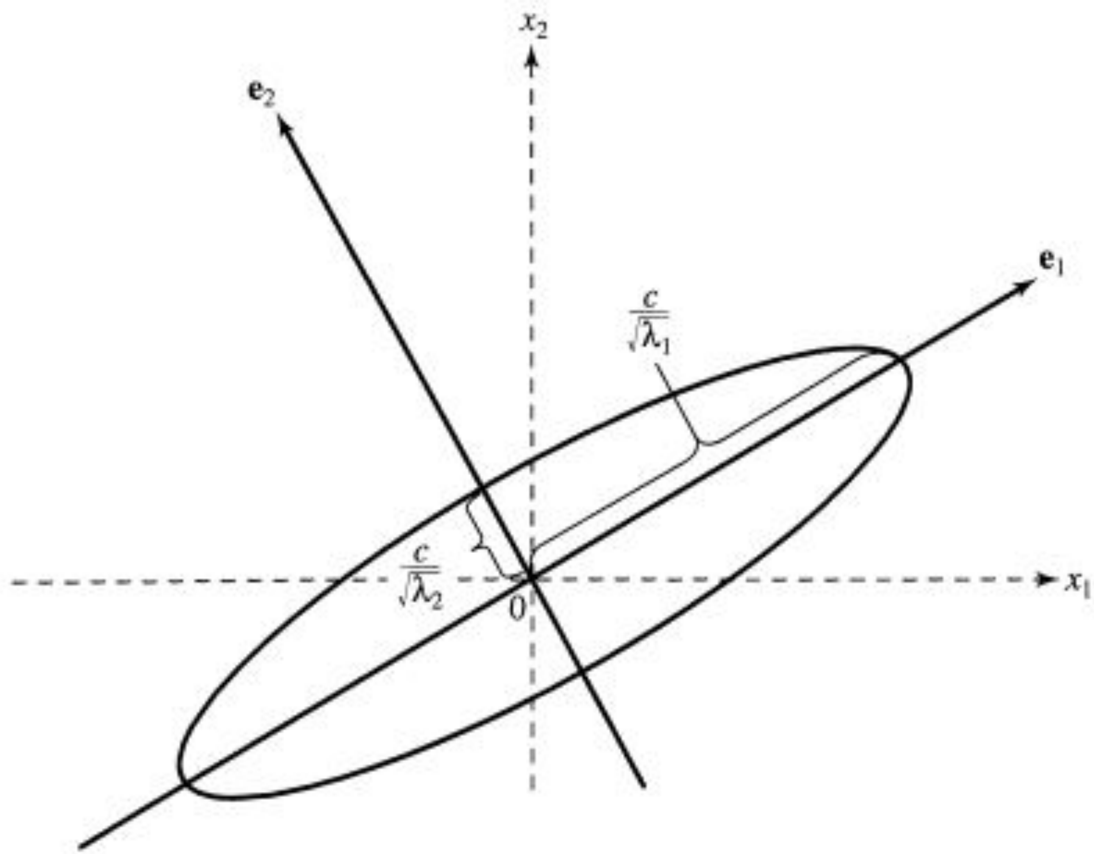
By the spectral decomposition,

$$\mathbf{A} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1^\top + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^\top \text{ so } \mathbf{x}^\top \mathbf{A} \mathbf{x} = \lambda_1 \left(\mathbf{x}^\top \mathbf{e}_1 \right)^2 + \lambda_2 \left(\mathbf{x}^\top \mathbf{e}_2 \right)^2$$

Now, $c^2 = \lambda_1 y_1^2 + \lambda_2 y_2^2$ is an ellipse in $y_1 = \mathbf{x}^\top \mathbf{e}_1$ and $y_2 = \mathbf{x}^\top \mathbf{e}_2$ because $\lambda_1, \lambda_2 > 0$ when $\mathbf{A}$ is positive definite.

Geometrical understanding

- We easily verify that $\mathbf{x} = c\lambda_1^{-1/2}\mathbf{e}_1$ satisfies $\mathbf{x}^\top \mathbf{A}\mathbf{x} = \lambda_1 \left(c\lambda_1^{-1/2}\mathbf{e}_1^\top \mathbf{e}_1\right)^2 = c^2$. Similarly, $\mathbf{x} = c\lambda_2^{-1/2}\mathbf{e}_2$ gives the appropriate distance in the $\mathbf{e}_2$ direction.



$A \succeq 0$: given $a \geq 0$, $b^2 = a$, in matrix case : $B^2 = A$, such B is the square root of $A \succeq 0$. Problem: B? Solution: $A = U\Lambda U^T$. Let $B = U\Lambda^{\frac{1}{2}}U^T$, thus $B^2 = U\Lambda^{\frac{1}{2}}U^T U\Lambda^{\frac{1}{2}}U^T = U\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}U^T = U\Lambda U^T$ since $U^T U = I_n$. Problem is this B unique? if $B_2 = -B$, then $B_2^2 = (-B)^2 = A$. Or usually we use the positive semi-definite square root.

Square-root matrix

$$B_3 = (U_1, \cdots, U_n)\begin{pmatrix} \sqrt{\lambda_1} & & & & \\ & -\sqrt{\lambda_2} & & & \\ & & \sqrt{\lambda_3} & & \\ & & & \ddots & \\ & & & & \sqrt{\lambda_n} \end{pmatrix}\begin{pmatrix} U_1^T \\ U_2^T \\ \vdots \\ U_n^T \end{pmatrix}$$

$B_3 = A$ as well so $\begin{cases} B = U\Lambda^{\frac{1}{2}}U^T \\ B \succeq 0 \end{cases} \Rightarrow B$ is unique.

The *square-root matrix*, of a positive semi-definite matrix $\mathbf{A}$,

$$\mathbf{A}^{1/2} = \sum_{i=1}^{k} \sqrt{\lambda_i}\mathbf{u}_i\mathbf{u}_i^\top = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top$$

has the following properties:

- $\left(\mathbf{A}^{1/2}\right)^\top = \mathbf{A}^{1/2}$ (that is, $\mathbf{A}^{1/2}$ is symmetric).
- $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$.
- $\left(\mathbf{A}^{1/2}\right)^{-1} = \sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}}\mathbf{u}_i\mathbf{u}_i^\top = \mathbf{U}\mathbf{\Lambda}^{-1/2}\mathbf{U}^\top$, where $\mathbf{\Lambda}^{-1/2}$ is a diagonal matrix with $1/\sqrt{\lambda_i}$ as the i th diagonal element.
- $\mathbf{A}^{1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1/2}\mathbf{A}^{1/2} = \mathbf{I}$, and $\mathbf{A}^{-1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1}$, where $\mathbf{A}^{-1/2} = \left(\mathbf{A}^{1/2}\right)^{-1}$.

Trace

Let $\mathbf{A} = \{a_{ij}\}$ be a $k \times k$ square matrix. The *trace* of the matrix $\mathbf{A}$, written $\mathrm{tr}(\mathbf{A})$, is the sum of the diagonal elements; that is, $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{k} a_{ii}$. Let $\mathbf{A}$ and $\mathbf{B}$ be $k \times k$ matrices and c be a scalar.

have A's generally no.

- $\mathrm{tr}(c\mathbf{A}) = c\,\mathrm{tr}(\mathbf{A})$
- $\mathrm{tr}(\mathbf{A} \pm \mathbf{B}) = \mathrm{tr}(\mathbf{A}) \pm \mathrm{tr}(\mathbf{B})$ $\quad tr(ABC) \neq tr(CBA)$ because $tr(ABC) = tr(BCA)$
- $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ $\Rightarrow tr(A_1A_2A_3\cdots A_{k-1}A_k) = tr(A_kA_1A_2\cdots A_{k-1}) = tr(A_{k-1}A_k\cdots A_{k-2})$
- $\mathrm{tr}\left(\mathbf{B}^{-1}\mathbf{AB}\right) = \mathrm{tr}(\mathbf{A})$ $\quad tr(AA^T) = tr(A^TA) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2,\ A \in \mathbb{R}^{m\times n}$
- $\mathrm{tr}\left(\mathbf{AA}^\top\right) = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}^2$
- If $\mathbf{A}$ is symmetric, $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{k} \lambda_i$

Frobenius norm: $\|A\|_F = \sqrt{tr(AA^T)} = \sqrt{\sum\sum a_{ij}^2}$, F-norm.

Spectral norm: $\|A\|_2 = \sup_{\|x\|=1} \|Ax\|$ for $A \in \mathbb{R}^{m\times n}$ = largest singular value of A.

$\|A\|_F = \sqrt{\sum_{i=1}^{m,n|m,} s_i^2}$ ↗ singular values.

$\sum_{i=1}^{r} r_i u_i u_i^T$

$\|A - B\|_F$

$rank(B) = r.$