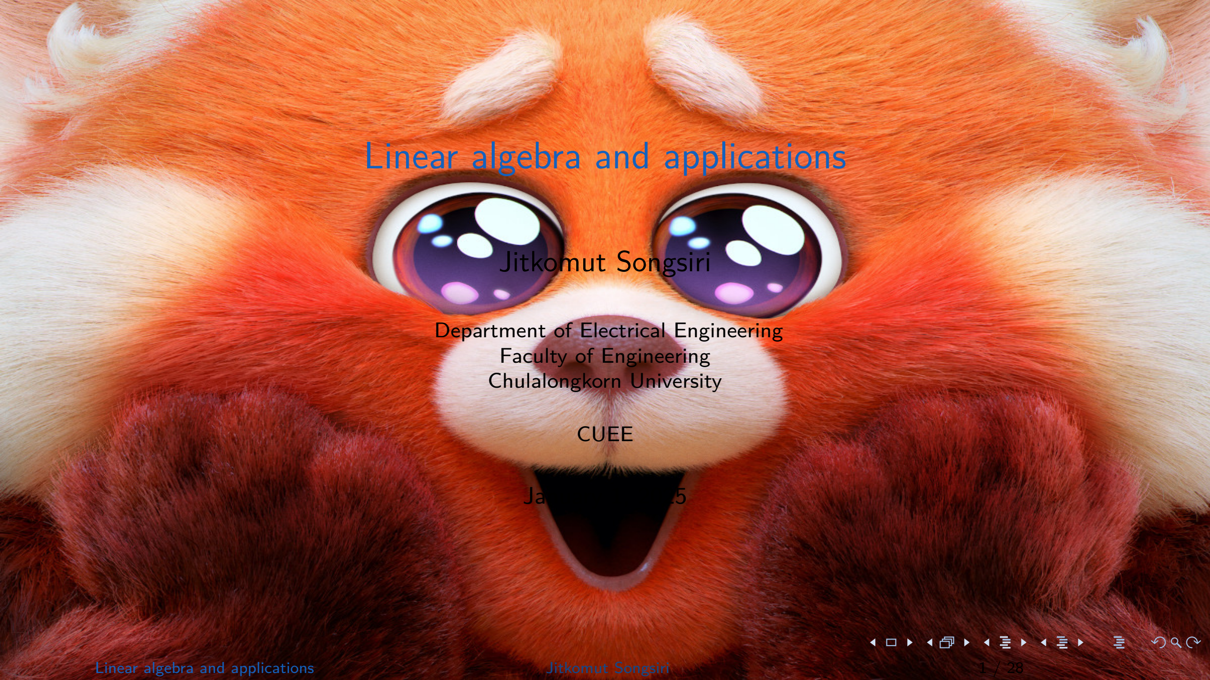


Linear algebra and applications

Jitkomut Songsiri

Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University

CUEE

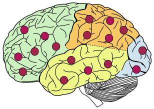
Ja[illegible]5

Outline

1. Special matrices and applications

How to read this handout

1. the note is used with lecture in EE205 (you cannot master this topic just by reading this note) – class activities include
   - graphical concepts, math derivation of details/steps in between
   - computer codes to illustrate examples
2. always read 'textbooks' after lecture
3. pay attention to the symbol ✎; you should be able to prove such ✎ result
4. each chapter has a list of references; find more formal details/proofs from in-text citations
5. almost all results in this note can be Googled; readers are encouraged to 'stimulate neurons' in your brain by proving results without seeking help from the Internet first
6. typos and mistakes can be reported to `jitkomut@gmail.com`



Special matrices and applications

Special matrices

- orthogonal matrix
- projection matrix
- permutation matrix
- symmetric matrix
- positive definite matrix

Orthogonal matrix

a *real* matrix $U \in \mathbf{R}^{n \times n}$ is called **orthogonal** if

$$UU^T = U^T U = I$$

properties: ✎

- an orthogonal matrix is special case of unitary for real matrices
- an orthogonal matrix is always invertible and $U^{-1} = U^T$
- columns vectors of U are mutually orthogonal
- norm is preserved under an orthogonal transformation: $\|Ux\|_2^2 = \|x\|_2^2$

example:

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Applications

1. rotation: in $\mathbf{R}^3$, rotate a vector x by the angle θ around the z-axis

$$w = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \triangleq U \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

   where U is orthogonal
2. eigenvectors of symmetric matrices are orthogonal (more detail later)
3. Q in QR decomposition is orthogonal
4. orthogonal matrices are used to whiten the data (transform correlated random vector to uncorrelated random vector)
5. discrete Fourier transform (DFT): $y = Wx$ where W is unitary (equivalence of orthogonal matrix in complex)

Unitary matrix

a *complex* matrix $U \in \mathbf{C}^{n\times n}$ is called **unitary** if

$$U^*U = UU^* = I, \qquad (U^* \triangleq \bar{U}^T)$$

example: let $z = e^{-i2\pi/3}$

$$U = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1 & 1 \\ 1 & z & z^2 \\ 1 & z^2 & z^4 \end{bmatrix} = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{-i2\pi/3} & e^{-i4\pi/3} \\ 1 & e^{-i4\pi/3} & e^{-i8\pi/3} \end{bmatrix}$$

facts: ✎

- a unitary matrix is always invertible and $U^{-1} = U^*$
- columns vectors of U are mutually orthogonal
- 2-norm is preserved under a unitary transformation: $\|Ux\|_2^2 = (Ux)^*(Ux) = \|x\|_2^2$

Example: Discrete Fourier transform (DFT)

DFT of the length-N time-domain sequence $x[n]$ is defined by

$$X[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] e^{-i2\pi kn/N}, \quad 0 \leq k \leq N-1$$

define $z = e^{-i2\pi/N}$, we can write the DFT in a matrix form as

$$\begin{bmatrix} X[0] \\ X[1] \\ X[2] \\ \vdots \\ X[N-1] \end{bmatrix} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & z^1 & z^2 & \cdots & z^{N-1} \\ 1 & z^2 & z^4 & \cdots & z^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z^{N-1} & z^{2(N-1)} & \cdots & z^{(N-1)(N-1)} \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ x[2] \\ \vdots \\ x[N-1] \end{bmatrix}$$

or $\mathbf{X} = \mathbf{D}\mathbf{x}$ where $\mathbf{D}$ is called the **DFT matrix** and is unitary ($\therefore \mathbf{x} = \mathbf{D}^*\mathbf{X}$)

Unitary property of DFT

the columns of DFT matrix are of the form:

$$\phi_k = (1/\sqrt{N}) \begin{bmatrix} 1 & e^{-i2\pi k/N} & e^{-i2\pi k\cdot 2/N} & \cdots & e^{-i2\pi k(N-1)/N} \end{bmatrix}^T$$

use $\langle \phi_l, \phi_k \rangle = \phi_k^* \phi_l$ and apply the sum of geometric series:

$$\langle \phi_l, \phi_k \rangle = \frac{1}{N} \sum_{n=0}^{N-1} e^{i2\pi(k-l)n/N} = \frac{1}{N} \cdot \frac{1 - e^{i2\pi(k-l)}}{1 - e^{i2\pi(k-l)/N}}$$

the columns of DFT matrix are therefore *orthogonal*

$$\langle \phi_l, \phi_k \rangle = \begin{cases} 1, & \text{for } k = l + rN, \quad r = 0, 1, 2, \ldots \\ 0, & \text{for } k \neq l \end{cases}$$

Projection matrix

$P \in \mathbf{R}^{n \times n}$ is said to be a **projection** matrix if $P^2 = P$ (aka **idempotent**)

- P is a linear transformation from $\mathbf{R}^n$ to a subspace of $\mathbf{R}^n$, denoted as S
- columns of P are the projections of standard basis vectors and S is the range of P
- if P is applied twice on a vector in S, it gives the same vector

examples: identity and

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, \quad \begin{bmatrix} 3 & -6 \\ 1 & -2 \end{bmatrix}, \quad I - X(X^TX)^{-1}X^T \text{ (in regression)}$$

properties: ✎

- eigenvalues of P are all equal to 0 or 1
- $I - P$ is also idempotent
- if $P \neq I$, then P is singular

Orthogonal projection matrix

a matrix $P \in \mathbf{R}^{n \times n}$ is called an **orthogonal projection** matrix if

$$P^2 = P = P^T$$

properties:

- P is bounded, *i.e.*, $\|Px\| \leq \|x\|$

$$\|Px\|_2^2 = x^T P^T P x = x^T P^2 x = x^T P x \leq \|Px\| \|x\|$$

- if P is an orthogonal projection onto a line spanned by a unit vector u,

$$P = uu^T$$

  (we see that $\mathbf{rank}(P) = 1$ as the dimension of a line is 1)
- another example: $P = X(X^T X)^{-1} X^T$ for any matrix X – (in regression)

Permutation

a **permutation** matrix P is a square matrix that has exactly one entry of 1 in each row and each column and has zero elsewhere

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

facts: ✎

- P is obtained by interchanging any two rows (or columns) of an identity matrix
- PA results in permuting rows in A, and AP gives permuting columns in A
- $P^T P = I$, so $P^{-1} = P^T$ (simple)
- the modulus of all eigenvalues of P is one, *i.e.*, $|\lambda_i(P)| = 1$
- a multiplication of P with vectors or matrix has no flop count (just swap rows/columns)

Linear function

given $w \in \mathbf{R}^n$ and let $x \in \mathbf{R}^n$ be a vector variable

a **linear function** $f : \mathbf{R}^n \to \mathbf{R}$ is given by

$$f(x) = w^T x = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$$

(✎ review its linear properties, *i.e.*, superposition)

an **affine function** is a linear function plus a constant: $f(x) = w^T x + b$

- $\frac{\partial f}{\partial x_i} = w_i$ gives the rate of change of f in x_i direction
- the set $\{x \mid w^T x + b = \text{ constant }\}$ is a hyperplane in $\mathbf{R}^n$ with the normal vector w
- linear functions are used in linear regression model and linear classifier

Energy form

given a (real) square matrix A, an energy form is a quadratic function of vector x:

$$f : \mathbf{R}^n \to \mathbf{R}, \quad f(x) = x^T A x = \sum_i \sum_j a_{ij} x_i x_j$$

- $x^T A x$ is the same as the energy form using $(A + A^T)/2$ as the coefficient because

$$x^T A x = (x^T A x)^T = \frac{x^T (A + A^T) x}{2}$$

- using $A = \frac{A+A^T}{2} + \frac{A-A^T}{2}$, we can later on assume that an energy form requires only the symmetric part of A
- reverse question: given an energy form, can you determine what A is ?

$$x_1^2 + 2x_2^2 + 3x_3^2 - x_1 x_2 + 2x_2 x_3 \quad \triangleq \quad x^T A x$$

Energy form and completing the square

recall how to complete the square:

$$x_1^2 + 3x_2^2 + 14x_1x_2 = (x_1 + 7x_2)^2 - 46x_2^2$$

given these matrices, expand the energy form and complete the square

$$A = \begin{bmatrix} 4 & 6 \\ 6 & 13 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix}, \quad C = \begin{bmatrix} 4 & 6 \\ 6 & -4 \end{bmatrix}$$

- $x^T A x =$
- $x^T B x =$
- $x^T C x =$

Quadratic function

given $P \in \mathbf{R}^{n\times n}, q \in \mathbf{R}^n, r \in \mathbf{R}$, a **quadratic** function $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is of the form

$$f(x) = (1/2)x^T Px + q^T x + r$$

- $x^T Px$ is aka an energy form (due to the quadratic form that appears in the energy/power of some physical variables)

$$\text{electrical power} = i^2 R, \quad \text{kinetic energy} = \frac{1}{2}mv^2, \quad \text{energy stored in spring} = \frac{1}{2}kx^2$$

- the contour shape of f depends on the property of P (positive definite, indefinite, magnitude of eigenvalues, direction of eigenvectors) – as we will learn shortly

Symmetric matrix

definition: a (real) square matrix A is said to be **symmetric** if $A = A^T$

notation: $A \in \mathbf{S}^n$

examples:

$$\begin{bmatrix} X & Y \\ Y^T & Z \end{bmatrix} \text{ with symmetric } X, Z, \quad A = \mathbf{E}[XX^T] \text{ (correlation matrix)}$$

✎ **basic facts:**

- for any (rectangular) matrix A, AA^T and A^TA are always symmetric
- if A is symmetric and invertible, then A^{-1} is symmetric
- if A is invertible, then AA^T and A^TA are also invertible

Properties of symmetric matrix

spectral theorem: if A is a real symmetric matrix then the following statements hold

1. all eigenvalues of A are real
2. all eigenvectors of A are orthogonal
3. A admits a decomposition

$$A = UDU^T$$

   where $U^TU = UU^T = I$ (U is unitary) and a diagonal D contains $\lambda(A)$
4. for any x, we have

$$\lambda_{\min}(A)\|x\|_2^2 \;\leq\; x^TAx \;\leq\; \lambda_{\max}(A)\|x\|_2^2$$

   the first (and second) inequalities are tight when x is the eigenvector corresponding to $\lambda_{\min}$ (and $\lambda_{\max}$ respectively)

Proofs

1. assume $Ax = \lambda x$ and λ, x could be complex, denote $x^* = \bar{x}^T$

$$\begin{aligned} (x^*Ax)^* &= x^*A^*x = x^*Ax = x^*\lambda x = \lambda x^*x \\ &= (x^*\lambda x)^* = \bar{\lambda}x^*x \end{aligned}$$

since $x^*x \neq 0$, we must have $\lambda = \bar{\lambda}$

2. assume $Ax_1 = \lambda_1 x_1$ and $Ax_2 = \lambda_2 x_2$ (now all (λ_i, x_i) are real)

$$\begin{aligned} x_2^T Ax_1 &= x_2^T \lambda_1 x_1 = \lambda_1 x_2^T x_1 \\ &= x_1^T Ax_2 = x_1^T \lambda_2 x_2 = \lambda_2 x_1^T x_2 \end{aligned}$$

equating two terms give $(\lambda_1 - \lambda_2)x_2^T x_1 = 0$

for simple case, we can assume that λ_i's are distinct, so $x_2^T x_1 = 0$ $(x_2 \perp x_1)$

Exercises

1. for $x, y \in \mathbf{R}^n$, are xy^T, xx^T, yx^T symmetric?
2. for a diagonal matrix D, is $D + xx^T$ symmetric?
3. if A, B are symmetric, so is $A + B$?
4. how many distinct entries in a symmetric matrix of size n?
5. if A is symmetric and B is rectangular, is BAB^T symmetric?
6. if A is symmetric and invertible, is A^{-1} symmetric?
7. find conditions on A, B, C, D so that the block matrix: $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is symmetric

Positive definite matrix

definition: a symmetric matrix A is **positive semidefinite**, written as $A \succeq 0$ if

$$x^T A x \geq 0, \quad \forall x \in \mathbf{R}^n$$

and is said to be **positive definite**, written as $A \succ 0$ if

$$x^T A x > 0, \quad \text{for all } \textit{nonzero } x \in \mathbf{R}^n$$

❋ the curly $\succeq$ symbol is used with matrices (to differentiate it from $\geq$ for scalars)

example: $A_1 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \succeq 0$ and $A_2 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \succ 0$ because

$$x^T A_1 x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + x_2^2 - 2x_1x_2 = (x_1 - x_2)^2 \geq 0$$

$$x^T A_2 x = (x_1 - x_2)^2 + x_2^2 > 0, \quad \forall x \neq 0$$

exercise: ✎ check positive semidefiniteness of matrices on page 16

How to test if $A \succeq 0$?

Theorem: $A \succeq 0$ if and only if all eigenvalues of A are non-negative

($A \succ 0$ if and only if $\lambda(A) > 0$)

Sylvester's criterion: if every principal minor of A (including $\det A$) is non-negative then $A \succeq 0$

proof in Horn Theorem 7.2.5

example 1: $A = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \succ 0$ because

- eigenvalues of A are 0.38 and 2.61 (real and positive)
- the principle minors are 1 and $\begin{vmatrix} 1 & -1 \\ -1 & 2 \end{vmatrix} = 1$ (all positive)

example 2: $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \succeq 0$ because eigenvalues of A are 0 and 3

Properties of positive definite matrix

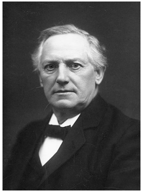
1. if $A \succeq 0$ then all the diagonal terms of A are nonnegative
2. if $A \succeq 0$ then all the leading blocks of A are positive semidefinite
3. if $A \succeq 0$ then $BAB^T \succeq 0$ for any B ✍ (exercise)
4. if $A \succeq 0$ and $B \succeq 0$, then so is $A + B$

Gram matrix

for an $m \times n$ matrix A with columns $a_1, \ldots, a_n$, the product $G = A^T A$ is called the **Gram matrix**

Gram matrix is positive semidefinite

Jørgen Pedersen Gram



$$G = A^T A = \begin{bmatrix} a_1^T a_1 & a_1^T a_2 & \cdots & a_1^T a_n \\ a_2^T a_1 & a_2^T a_2 & \cdots & a_2^T a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^T a_1 & a_n^T a_2 & \cdots & a_n^T a_n \end{bmatrix}$$

$$x^T G x = x^T A^T A x = \|Ax\|^2 \geq 0, \quad \forall x$$

- if A has zero nullspace then $Ax = 0 \leftrightarrow x = 0$; this implies that $A^T A \succ 0$
- let X be a data matrix, partitioned in N rows as x_k^T's; we typically encounter $G = \frac{X^T X}{N} = \frac{1}{N} \sum_{k=1}^{N} x_k x_k^T$ as the **sample covariance matrix**

Exercises

1. check if each of the following is positive definite

$$A_1 = \begin{bmatrix} 2 & 4 \\ 4 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -2 & 1 \\ 1 & 3 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 4 & -1 & 0 \\ -1 & 2 & 2 \\ 0 & 2 & 3 \end{bmatrix}$$

2. is a diagonal matrix always positive semidefinite?
3. for $x \in \mathbf{R}^n$ and I is the identify
   1. is $I + xx^T$ positive semidefinite?
   2. is $I - xx^T$ positive semidefinite?
   3. is xx^T positive semidefinite?
4. find conditions on a, b, c so that

$$\begin{bmatrix} 2 & a & b \\ a & 1 & -1 \\ b & -1 & c \end{bmatrix}$$

is positive definite

Numerical exercises

generate each of these matrices *randomly* and check its properties

1. orthogonal: check determinant and eigenvalues
2. orthogonal projection: check eigenvalues
3. permutation: check the eigenvalues, its inverse and transpose
4. symmetric: check eigenvalues and eigenvectors
5. positive definite: check eigenvalues, eigenvalues of leading diagonal blocks,

relate what you numerically found to the properties of these matrices

References

1. W.K. Nicholson, *Linear Algebra with Applications*, McGraw-Hill, 2006
2. G. Strang, *Linear Algebra and Learning from Data*, Wellesley-Cambridge Press, 2019