Linear Algebra

Wenxiao Yang[*]

[*]Department of Mathematics, University of Illinois at Urbana-Champaign

2021

目录

4   Algebra Computation   **10**

1   Vector Space

1.1   Vector Space $(V, +, \times)$ (over a field $\mathbb{F}$)

A <u>vector space</u> over a field $\mathbb{F}$ is a set V w/ an operation <u>addition</u> $+ : V \times V \to V$ and an operation <u>scalar multiplication</u> $\mathbb{F} \times V \to V$

(1) Addition is associative & commutative

(2) $\exists 0 \in V$, additive identity: $0 + v = v \forall v \in V$

(3) $1v = v \forall v \in V$ (where $1 \in \mathbb{F}$ is multi. id. in $\mathbb{F}$)

(4) $\forall \alpha, \beta \in \mathbb{F}, \ v \in V, \ \alpha(\beta v) = (\alpha\beta)v$

(5) $\forall v \in V, \ (-1)v = -v$ we have $v + (-v) = 0$

(6) $\forall \alpha \in \mathbb{F}, \ v, u \in V, \ \alpha(v + u) = \alpha v + \alpha u$

(7) $\forall \alpha, \beta \in \mathbb{F}, \ v \in V, \ (\alpha + \beta)v = \alpha v + \beta v$

1.2   A field is a vector space over its subfield

Example 1. $\mathbb{K} \subset \mathbb{F}$ *is a subfield of a field* $\mathbb{F}$*. Then* $\mathbb{F}$ *is a vector space over* $\mathbb{K}$*. (Since* $\mathbb{F} \subset \mathbb{F}[x]$*, then* $\mathbb{F}[x]$ *is a vector space over* $\mathbb{F}$*.)*

1.3   Vector subspace

Suppose that V is a vector space over $\mathbb{F}$. A <u>vector subspace</u> or just <u>subspace</u> is a nonempty subset $W \subset V$ closed under addition and scalar multiplication. i.e. $v + w \in W, \ av \in W, \ \forall v, w \in W, \ a \in \mathbb{F}$.

Example 2. $\mathbb{K} \subset \mathbb{L} \subset \mathbb{F}$*, then* $\mathbb{L}$ *is a subspace of* $\mathbb{F}$ *over* $\mathbb{K}$*.*

1.4   Linear independent, Linear combination

1.5   span V, basis, dimension

A set of elements $v_1, ..., v_n \in V$ is said to **span** V if every vector $v \in V$ can be expressed as a linear combination of $v_1, ..., v_n$. If $v_1, ..., v_n$ spans and is linearly independent, then we call the set a ***basis*** for V.

Proposition 1 (Proposition 2.4.10.)**.** *Suppose* V *is a vector space over a field* $\mathbb{F}$ *having a basis* $\{v_1, ..., v_n\}$ *with* $n \geq 1$.

(i) For all $v \in V$, $v = a_1 v_1 + ... + a_n v_n$ for exactly one $(a_1, ..., a_n) \in \mathbb{F}^n$.

(ii) If $w_1, ..., w_n$ span V , then they are linearly independent.

(iii)If $w_1, ..., w_n$ are linearly independent, then they span V.

If a vector space V over $\mathbb{F}$ has a basis with n vectors, then V is said to be n-dimensional (over $\mathbb{F}$) or is said to have **dimension** n.

1.6 Standard basis vectors

$$e_1 = (1, 0, ..., 0), e_2 = (0, 1, 0, ..., 0), ..., e_n = (0, 0, ..., 0, 1) \in \mathbb{F}^n$$

are a basis for $\mathbb{F}^n$ called the **standard basis vectors**.

1.7 Linear transformation

Given two vector spaces V and W over $\mathbb{F}$ a **linear transformation** is a function $T : V \to W$ such that for all $a \in \mathbb{F}$ and $v, w \in V$, we have

$$T(av) = aT(v) \ and \ T(v + w) = T(v) + T(w)$$

Proposition 2 (Proposition 2.4.15.)**.** *If V and W are vector spaces and $v_1, ..., v_n$ is a basis for V then any function from $\{v_1, ..., v_n\} \to W$ extends uniquely to a linear transformation $V \to W$.*

Any $v \in V$, $\exists (a_1, ..., a_n)$ s.t. $v = a_1 v_1 + ... + a_n v_n$. Then $T(v) = T(a_1 v_1 + ... + a_n v_n) = a_1 T(v_1) + ... + a_n T(v_n)$

1.8 一个线性变换对应一个矩阵, 线性变换矩阵相乘仍为线性变换矩阵

Corollary 1 (Corollary 2.4.16.)**.** *If $v_1, ..., v_n$ is a basis for a vector space V and $w_1, ..., w_n$ is a basis for a vector space W (both over $\mathbb{F}$), then any linear transformation $T : V \to W$ determines (and is determined by) the $m \times n$ matrix:*

$$A = A(T) = \begin{bmatrix} A_{11} & A_{12} & ... & A_{1n} \\ A_{21} & A_{22} & ... & A_{2n} \\ \vdots & \vdots & ... & \vdots \\ A_{m1} & A_{m2} & ... & A_{mn} \end{bmatrix}$$

$$\begin{bmatrix} w_1 & \cdots & w_m \end{bmatrix}^T = A \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}^T$$

$\mathcal{L}(V, M)$ denotes the set of all linear transformations from V to W; $M_{m \times n}(\mathbb{F})$ the set of $m \times n$ matrix with entries in $\mathbb{F}$. $T \to A(T)$ defines a *bijection* $\mathcal{L}(V, M) \to M_{m \times n}(\mathbb{F})$. $A(T)$ **represents the linear transformation** T.

Proposition 3 (Proposition 2.4.19)**.** *Suppose that V , W, and U are vector spaces over $\mathbb{F}$, with fixed chosen bases. If $T : V \to W$ and $S : W \to U$ are linear transformations represented by matrices $A = A(T)$ and $B = B(S)$, then $ST = S \circ T : V \to U$ is a linear transformation represented by the matrix $BA = B(S)A(T)$.*

1.9 GL(V): invertible linear transformations $V \to V$

Given a vector space V over F, we let $GL(V) \subset \mathcal{L}(V, V)$ denote the subset of **invertible linear transformations**.

$$GL(V) = \{T \in \mathcal{L}(V, V) | T \text{ is a bijection}\} = \mathcal{L}(V, V) \cap Sym(V)$$

2 Eigenvalues Related

2.1 Eigenvalues, Eigenvectors

A vector x is an **eigenvector** of a matrix A if Ax is parallel to x, that is if $Ax = \lambda x$ for some number $\lambda \in \mathbb{R}$. The number λ is called an **eigenvalue** of A.

i.e. the root of $(A - \lambda I_n)x = 0 \Leftrightarrow det(A - \lambda I_n) = 0$

Theorem 1. *If an $n \times n$ matrix A has n linearly independent eigenvectors $x_1, ..., x_n$ corresponding to eigenvalues $\lambda_1, ..., \lambda_n$, then $A = PDP^{-1}$ where D is diagonal with entries $\lambda_1, ..., \lambda_n$, and P has columns $x_1, ..., x_n$.*

2.2 Jacobian matrix

Suppose $\mathbf{f} : \mathbf{R}^n \to \mathbf{R}^m$ is a function such that each of its first-order partial derivatives exist on $\mathbf{R}^n$. This function takes a point $\mathbf{x} \in \mathbf{R}^n$ as input and produces the vector $\mathbf{f}(\mathbf{x}) \in \mathbf{R}^m$ as output. Then the Jacobian matrix of $\mathbf{f}$ is defined to be an $m \times n$ matrix, denoted by $\mathbf{J}$, whose (i, j) th entry is $\mathbf{J}_{ij} = \frac{\partial f_i}{\partial x_j}$, or explicitly

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla^{\mathrm{T}} f_1 \\ \vdots \\ \nabla^{\mathrm{T}} f_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

where $\nabla^{\mathrm{T}} f_i$ is the transpose (row vector) of the gradient of the i component.

2.3 Hessian matrix

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a function taking as input a vector $\mathbf{x} \in \mathbb{R}^n$ and outputting a scalar $f(\mathbf{x}) \in \mathbb{R}$. If all second partial derivatives of f exist and are continuous over the domain of the function, then the Hessian matrix $\mathbf{H}$ of f is a square $n \times n$ matrix, usually defined and arranged as follows:

$$\mathbf{H}_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix},$$

or, by stating an equation for the coefficients using indices i and j,

$$(\mathbf{H}_f)_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

The Hessian matrix is a symmetric matrix, since the hypothesis of continuity of the second derivatives implies that the order of differentiation does not matter (Schwarz's theorem).

The determinant of the Hessian matrix is called the Hessian determinant.

2.4 Positive Definite Matrices

2.4.1 Definition

We say that a symmetric $n \times n$ matrix A is:

(1). **positive semidefinite** (written $A \succeq 0$) if $x^T A x \geq 0$ for all x.

(2). **positive definite** (written $A \succ 0$) if $x^T A x > 0$ for all $x \neq 0$.

(3). **negative semidefinite** (written $A \preceq 0$) if $x^T A x \leq 0$ for all x.

(4). **negative definite** (written $A \prec 0$) if $x^T A x < 0$ for all $x \neq 0$.

(5). **indefinite** (not written in any particular way) if none of the above apply.

$x^T A x$ is a function of x called the quadratic form associated to A.

Note: $A^T A$ is **positive semidefinite**, since $x^T A^T A x = \|Ax\|^2 \geq 0$.

2.4.2 Diagonal matrix situation

$$D = \begin{bmatrix} d_1 & 0 & ... & 0 \\ 0 & d_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & d_n \end{bmatrix}$$

Lemma 1. *If $d_1, ... d_n$ are all nonnegative, then $D \succeq 0$;*

If $d_1, ... d_n$ are all positive, then $D \succ 0$;

If $d_1, ... d_n$ are all nonpositive, then $D \preceq 0$;

If $d_1, ... d_n$ are all negative, then $D \prec 0$;

2.4.3 Using eigenvalues

If A is an $n \times n$ symmetric matrix, then it can be factored as

$$A = Q^T \Lambda Q = Q^T \begin{bmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \lambda_n \end{bmatrix} Q$$

where $\lambda_1, ..., \lambda_n$ are the eigenvalues of A and the columns of Q are the corresponding eigenvectors.

We can get $x^T A x = x^T Q^T \Lambda Q x = (Qx)^T \Lambda (Qx)$

If we substitute $y = Qx$:

$x^T A x = y^T \Lambda y = \lambda_1 y_1^2 + \lambda_2 y_2^2 + ... + \lambda_n y_n^2$

Theorem 2.

If $\lambda_1, ...\lambda_n$ are all nonnegative, then symmetric matrix $A \succeq 0$;

If $\lambda_1, ...\lambda_n$ are all positive, then $A \succ 0$;

If $\lambda_1, ...\lambda_n$ are all nonpositive, then $A \preceq 0$;

If $\lambda_1, ...d\lambda_n$ are all negative, then $A \prec 0$;

if it has both positive and negative eigenvalues, then A is indefinite

2.4.4 Sylvester's Criterion

Consider a $n \times n$ matrix A:

$$A = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nn} \end{bmatrix}$$

Denote its $k \times k$ submatrix $A^{(k)}$:

$$A^{(k)} = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1k} \\ a_{21} & a_{22} & ... & a_{2k} \\ ... & ... & ... & ... \\ a_{k1} & a_{k2} & ... & a_{kk} \end{bmatrix}$$

Let $\Delta_k = det(A^{(k)})$

$$det(A - xI) = (\lambda_1 - x)(\lambda_2 - x)...(\lambda_n - x)$$

by setting $x = 0$ we get $det(A) = \lambda_1\lambda_2...\lambda_n$. When $A \succ 0$, all the eigenvalues are positive, so $det(A) > 0$ as well.

$A \succ 0 \Rightarrow \mathbf{u}^T A\mathbf{u} > 0$ for all $\mathbf{u}$:

Set $\mathbf{u} = [0, 0..., 0, u_{k+1}, u_{k+2}, ..., u_n]$

Then we can simplify the quadratic form for A to the quadratic form for $A^{(k)}$. Therefore, we expect $A^{(k)} \succ 0 \Rightarrow \Delta_k > 0$ for all k.

Theorem 3.

$A \succ 0$ iff $\Delta_i > 0$ $\forall i = 1, ..., n$

$A \prec 0$ iff $(-1)^i\Delta_i > 0$ $\forall i = 1, ..., n$

A is indefinite if the first Δ_k(noozero) that breaks each pattern respectively is the wrong sign.

3 Euclidean geometry basics

3.1 Norm

3.1.1 Vector's Norm

Vector $x \in \mathbb{R}^n$-n-dim Euclidean space

$$x = (x_1, \ldots, x_n) \equiv \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}^\top = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Norm of x, $\|x\|$ satisfies properties:

(a) $\|x\| \geqslant 0$

(b) $\|x\| = 0 \Leftrightarrow x = 0$

(c) $\|cx\| = |c|\|x\|$, for $c \in \mathbb{R}$

(d) $\|x + y\| \leqslant \|x\| + \|y\| \longleftarrow$ Triangle Ineq.

Enclidean Norm (default $\rho = 2$): $\|x\| = \sqrt{x^\top x} = \sqrt{\sum_{i=1}^n x_i^2}$

Other norms:

1. l_1-norm : $\|x\|_1 = \sum_{i=1}^n |x_i|$

2. l_ρ-norm : $\|x\|_\rho = \sqrt[\rho]{\sum_{i=1}^n |x_i|^\rho}$

3. Supremum norm or l_∞-norm : $\|x\|_\infty = \max_i |x_i|$

3.1.2 Matrix's Norm

$A \in \mathbb{R}^{n \times m}$ is a matrix

$\|Ax\| \leqslant \|A\|\|x\|, \|AB\| \leqslant \|A\|\|B\|$

Default is $\rho = 1$: $\|A\| = \max_{\|x\|=1} \|Ax\|$. 即找到最大的绝对值和的"列"。

$\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$ (Frobenius norm)

$\|A\|_1 = \max_j \sum_{i=1}^n |A_{ij}|$ [Π]

$\|A\|_\infty = \max_2 \sum_{j=1}^n |A_{ij}|$ [□]

$\|A\|_2 = \max_k \sigma_k, \sigma_k$ is the singurai veruee of A

$\|A\| = \max\left(\frac{\|A \times \|}{\|\times\|}\right) \Rightarrow \|A\| \geqslant \frac{\|A \times \|}{\|\times\|}$

$\|A \times \| \leqslant \|A\|\|x\|$

3.2 Euclidean distance, inner product

Euclidean distance on $\mathbb{R}^n$:

$$|x - y| = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2}$$

Euclidean inner product:

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n = x^T y$$

Two important results for Euchidean norm:

1) Pythagorean Theorem: If $x^\top y = 0$,

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

2) Cauchy - Schwarz Inequality:

$$\left| x^\top y \right| \leqslant \|x\|\|y\|$$

$" = "$ iff $x = \alpha y$ for some $\alpha \in \mathbb{R}$

3.3 Isometry

An **isometry** of $\mathbb{R}^n$ is a bijection $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ that preserves distance, which means,

$$|\Phi(x) - \Phi(y)| = |x - y|, \ \forall x, y \in \mathbb{R}^n$$

We use $Isom(\mathbb{R}^n)$ denotes the set of all isometries of $\mathbb{R}^n$,

$$Isom(\mathbb{R}^n) = \{\Phi : \mathbb{R}^n \to \mathbb{R}^n | |\Phi(x) - \Phi(y)| = |x - y|, \ \forall x, y \in \mathbb{R}^n\}$$

Proposition 4. $\Phi, \Psi \in Isom(\mathbb{R}^n)$, *then* $\Phi \circ \Psi, \Phi^{-1} \in Isom(\mathbb{R}^n)$

证明.

Since Φ, Ψ are bijections, so is $\Phi \circ \Psi$. Moreover,

$$|\Phi \circ \Psi(x) - \Phi \circ \Psi(y)| = |\Phi(\Psi(x)) - \Phi(\Psi(y))| = |\Psi(x) - \Psi(y)| = |x - y|$$

Since $id \in Isom(\mathbb{R}^n)$,

$$|x - y| = |id(x) - id(y)| = |\Phi \circ \Phi^{-1}(x) - \Phi \circ \Phi^{-1}(y)| = |\Phi^{-1}(x) - \Phi^{-1}(y)|$$

$\square$

3.4 Linear isometries i.e. orthogonal group

There is a matrix $A \in GL(n, \mathbb{R})$ i.e. a *invertible linear transofrmations $T_A : \mathbb{R}^n \to \mathbb{R}^n$* is given by $T_A(v) = Av$.

$$T_A(v) \cdot T_A(w) = (Av) \cdot (Aw) = (Av)^t(Aw) = v^t A^t A w$$

$$A^t A = I \Leftrightarrow T_A(v) \cdot T_A(w) = v \cdot \Leftrightarrow T_A \in Isom(\mathbb{R}^n)$$

We define the all isometries in *invertible linear transofrmations $\mathbb{R}^n \to \mathbb{R}^n$* as **orthogonal group**

$$O(n) = \{A \in GL(n, \mathbb{R}) | A^t A = I\} \subset GL(n, \mathbb{R})$$

3.5 Special orthogonal group

$O(n)$ are the matrices representing linear isometries of $\mathbb{R}^n$. $1 = det(I) = det(A^t A) = det(A^t)det(A) = det(A)^2 \Rightarrow det(A) = 1$ or $det(A) = -1$. We use **special orthogonal group** represents A with $det(A) = 1$,

$$SO(n) = \{A \in O(n) | det(A) = 1\}$$

3.6 translation

Define a *translation* by $v \in \mathbb{R}^n$,

$$\tau_v : \mathbb{R}^n \to \mathbb{R}^n, \ \tau_v(x) = x + v$$

Note 1 (Exercise 2.5.3). $\forall v \in \mathbb{R}^n, \tau_v$ *is an isometry.*

证明. $|\tau_v(x) - \tau_v(y)| = |(x + v) - (y + v)| = |x - y|$ □

3.7 All isometries can be represented by a composition of *a translation* and *an orthogonal transformation*

Since *the composition of isometries is an isometry,* $\forall A \in O(n)$ and $v \in \mathbb{R}^n$, the composition

$$\Phi_{A,v}(x) = \tau_v(T_A(x)) = Ax + v$$

is an isometry. **which could account for all isometries**.

Theorem 4. $Isom(\mathbb{R}^n) = \{\Phi_{A,v} | A \in O(n), v \in \mathbb{R}^n\}$

4 Algebra Computation

4.1 Random Vectors

Mean:

$$\mu = \mathbb{E}(\mathbf{Z}) = \begin{pmatrix} \mathbb{E}(Z_1) \\ \mathbb{E}(Z_2) \\ \cdots \\ \mathbb{E}(Z_m) \end{pmatrix}$$

Variance-Covariance matrix Σ:

$$\Sigma_{m \times m} = Cov(\mathbf{Z}) = \mathbb{E}((\mathbf{Z} - \mu)(\mathbf{Z} - \mu)^T) = \begin{bmatrix} Var(Z_1) & \cdots & Cov(Z_1, Z_m) \\ \cdots & \cdots & \cdots \\ Cov(Z_m, Z_1) & \cdots & Var(Z_m) \end{bmatrix}$$

Affine Transformation
(1)

$$\mathbf{W} = \mathbf{a}_{n \times 1} + \mathbf{B}_{n \times m} \mathbf{Z}_{m \times 1}$$

$$\mathbb{E}(\mathbf{W}) = \mathbf{a} + \mathbf{B}\mu, \ Cov(\mathbf{W}) = \mathbf{B}\Sigma\mathbf{B}^T$$

(2)

$$\mathbf{W} = \mathbf{v}^T\mathbf{Z} = v_1 Z_1 + ... + v_m Z_m$$

$$\mathbb{E}(\mathbf{W}) = \mathbf{v}^T\mu = \sum_{i=1}^{m} v_i \mu_i$$

$$Var(\mathbf{W}) = \mathbf{v}^T\Sigma\mathbf{v} = \sum_{i=1}^{m} v_i^2 Var(Z_i) + 2\sum_{i<j} v_i v_j Cov(Z_i, Z_j)$$

i.e. $\mathbb{E}(\mathbf{AZ}) = \mathbf{A}\mathbb{E}(Z); \ Var(\mathbf{AZ}) = \mathbf{A}Var(\mathbf{Z})\mathbf{A}^T$

(3)

$$Cov(\mathbf{AX}, \mathbf{BY}) = \mathbb{E}[(\mathbf{AX}-\mathbf{A}\mathbb{E}(X))(\mathbf{BY}-\mathbf{B}\mathbb{E}(Y))^T] = \mathbf{A}\mathbb{E}[(\mathbf{X}-\mathbb{E}(X))(\mathbf{Y}-\mathbb{E}(Y))^T]\mathbf{B}^T = \mathbf{A}Cov(\mathbf{X}, \mathbf{Y})\mathbf{B}^T$$

4.2 矩阵求导

https://zhuanlan.zhihu.com/p/24709748

https://blog.csdn.net/daaikuaichuan/article/details/80620518

Vector by vector:

图 1: Denominator layout means $x \in \mathbb{R}^{n \times 1}$

$$\frac{\partial u}{\partial x^T} = \left(\frac{\partial u^T}{\partial x}\right)^T$$

$$\frac{\partial u^T v}{\partial x} = \frac{\partial u^T}{\partial x}v + \frac{\partial v^T}{\partial x}u^T$$

$$\frac{\partial u v^T}{\partial x} = \frac{\partial u}{\partial x}v^T + u\frac{\partial v^T}{\partial x}$$

$$\frac{\partial x^T x}{\partial x} = 2x$$

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

11

where $x, u, v \in \mathbb{R}^{n \times 1}$

Note:

$$\frac{d\|Aw - b\|^2}{dw} = \frac{d(Aw - b)^T(Aw - b)}{dw} = \frac{d(Aw - b)^T}{dw}(Aw - b) + \frac{d(Aw - b)^T}{dw}(Aw - b) = 2A(Aw - b)$$

Matrix by vector:

$$\frac{\partial AB}{\partial x} = \frac{\partial A}{\partial x}B + A\frac{\partial B}{\partial x}$$

Matrix by matrix:

$$\frac{\partial u^T X v}{\partial X} = uv^T$$

$$\frac{\partial u^T X^T X u}{\partial X} = 2Xuu^T$$

$$\frac{\partial[(Xu - v)^T(Xu - v)]}{\partial X} = 2(Xu - v)u^T$$

Trace (迹):

$$tr(a) = a$$

$$tr(AB) = tr(BA)$$

$$tr(ABC) = tr(CAB) = tr(BCA)$$

$$\frac{\partial tr(AB)}{\partial A} = B^T$$

$$tr(A) = tr(A^T)$$

$$\frac{\partial tr(ABA^T C)}{\partial A} = CAB + C^T AB^T$$

参考文献

[1] MATH 417: Christopher J Leininger Introduction to Abstract Algebra (Draft) 2017.

[2] MATH 484

[3] ECE 490

[4] STAT 425

[5] CS/MATH 357