Outline

Linear Space

A vector space (linear space) is a space consisting of a set $\boldsymbol{V}$ of vectors along with the following operations:

- Addition of two vectors: $(+)$
- Multiplication of a vector with a scalar: $(.)$

e.g., Euclidean space

Consider an n-dimensional real linear space, $\mathbb{R}^n$. A vector in $\mathbb{R}^n$ is an n-tuple such as $\boldsymbol{x} = [x_1 \quad x_2 \quad \dots \quad x_n]^T$.

A set of vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, .., \boldsymbol{x}_m\}$ is said to be linearly independent if

$$\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \dots + \alpha_m \boldsymbol{x}_m = 0. \qquad (*)$$

implies that $\alpha_i = 0$, $i = 1, 2, .., m$.

If $\boldsymbol{x}_1, \boldsymbol{x}_2, .., \boldsymbol{x}_m$ are linearly dependent, there exist some $\alpha_i \neq 0$ such that (*) holds. Suppose $\alpha_1 \neq 0$, then $x_1 = -(\alpha_2 \boldsymbol{x}_2 + \dots + \alpha_m \boldsymbol{x}_m)/\alpha_1$, i.e., $\boldsymbol{x}_1$ can be expressed as a linear combination of others.

The dimension of a linear space is defined as the maximum number of linearly independent vectors in the space.

Basis & Representation

A set of linearly independent vectors in $\mathbb{R}^n$ is a basis if any vector in $\mathbb{R}^n$ can be expressed as a unique linear combination of the set.

Let $\{\boldsymbol{q}_1, \boldsymbol{q}_2, ..., \boldsymbol{q}_n\}$ be a basis of $\mathbb{R}^n$. Then any $x \in \mathbb{R}^n$ can be expressed as

$$\boldsymbol{x} = \alpha_1 \boldsymbol{q}_1 + \alpha_2 \boldsymbol{q}_2 + ... + \alpha_n \boldsymbol{q}_n = Q\bar{\boldsymbol{x}}$$

where

$$Q = [\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad ... \quad \boldsymbol{q}_n], \quad \bar{\boldsymbol{x}} = [\alpha_1 \quad \alpha_2 \quad ... \quad \alpha_n]^T$$

$\bar{\boldsymbol{x}}$ is called the representation of $\boldsymbol{x}$ w.r.t. the basis $\{\boldsymbol{q}_1, \boldsymbol{q}_2, ..., \boldsymbol{q}_n\}$.

For every $\mathbb{R}^n$, the representation of a vector $\boldsymbol{x} = [x_1 \quad x_2 \quad ... \quad x_n]^T$ w.r.t. the orthonormal basis $\{\boldsymbol{i}_1, \boldsymbol{i}_2, ..., \boldsymbol{i}_n\}$ with

$$\boldsymbol{i}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \boldsymbol{i}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, ..., \boldsymbol{i}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

is equal to itself.

Norms of Vectors

A norm is a function $\|.\| : \mathbb{R}^n \to \mathbb{R}$ that assigns a real-valued length to each vector.

For all vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ and all scalars $\alpha \in \mathbb{R}$, a norm must satisfy

- $\|\boldsymbol{x}\| \geq 0$, and $\|\boldsymbol{x}\| = 0$ only if $\boldsymbol{x} = 0$
- $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$
- $\|\alpha \boldsymbol{x}\| = |\alpha| \|\boldsymbol{x}\|$

In general, $'p'$-norms are defined by

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

In particular, we have,

- 1-norm: $\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|$
- 2-norm: $\|\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^T \boldsymbol{x}} = \sqrt{\sum_{i=1}^{n} x_i^2}$ (Euclidean norm)
- ∞-norm: $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$

MATLAB: `norm(x,1)`, `norm(x,2)`, `norm(x,inf)`

Inner Product and Orthonormality

The inner product of vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ is defined as

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^T \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i$$

Cauchy-Schwarz inequality: $\left| \langle \boldsymbol{x}, \boldsymbol{y} \rangle \right| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|$

A vector $\boldsymbol{x}$ is said to be normalized if $\|\boldsymbol{x}\|_2 = 1$, i.e., if $\boldsymbol{x}^T \boldsymbol{x} = 1$.

Vectors $\boldsymbol{x}, \boldsymbol{y}$ are said to be orthogonal if $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$

A set of vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, ..\boldsymbol{x}_m\}$ is said to be orthonormal if

$$\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

Exercise 1: Prove Cauchy-Schwarz inequality. (Hint: Noting that $\langle \boldsymbol{x} + \alpha \boldsymbol{y}, \boldsymbol{x} + \alpha \boldsymbol{y} \rangle \geq 0$ for any α, choose $\alpha = -\langle \boldsymbol{y}, \boldsymbol{x} \rangle / \langle \boldsymbol{y}, \boldsymbol{y} \rangle$)

Orthonormalization

Given a set of linearly independent vectors $\{\boldsymbol{e}_1, \boldsymbol{e}_2, ..\boldsymbol{e}_m\}$, an orthonormal set $\{\boldsymbol{q}_1, \boldsymbol{q}_2, ..\boldsymbol{q}_m\}$ can be computed as

$$
\begin{aligned}
\boldsymbol{u}_1 &= \boldsymbol{e}_1 & \boldsymbol{q}_1 &= \boldsymbol{u}_1/\|\boldsymbol{u}_1\| \\
\boldsymbol{u}_2 &= \boldsymbol{e}_2 - (\boldsymbol{q}_1^T \boldsymbol{e}_2)\boldsymbol{q}_1 & \boldsymbol{q}_2 &= \boldsymbol{u}_2/\|\boldsymbol{u}_2\| \\
&\vdots \\
\boldsymbol{u}_m &= \boldsymbol{e}_m - \sum_{i=1}^{m-1} (\boldsymbol{q}_i^T \boldsymbol{e}_m)\boldsymbol{q}_i & \boldsymbol{q}_m &= \boldsymbol{u}_m/\|\boldsymbol{u}_m\|
\end{aligned}
$$

This procedure is called Schmidt orthonormalization procedure.

Let $A = [\boldsymbol{a}_1 \;\; \boldsymbol{a}_2 \;\; ... \;\; \boldsymbol{a}_m]$ where $\{\boldsymbol{a}_1, \boldsymbol{a}_2, ..\boldsymbol{a}_m\}$ is an orthonormal set. Then $A^T A = I_m$.

Column Space, Nullspace and Rank of Matrices

Column space (or range) of a matrix $A \in \mathbb{R}^{m\times n}$ is the set of all possible linear combinations of all columns of A.

Rank of A, denoted $\text{rank}(A)$, is the dimension of its column space, i.e., the number of linearly independent columns of A. It also equals the number of linearly independent rows. So $\text{rank}(A) \leq \min(m, n)$.

A vector $\boldsymbol{x}$ is called a null vector of A if $A\boldsymbol{x} = 0$. The nullspace (or kernel) of A is the set of all null vectors of A.

Nullity of A is the dimension of its nullspace, or equivalently, the maximum number of linearly independent null vectors of A.

Rank-nullity theorem: $\text{rank}(A) + \text{nullity}(A) = n$

MATLAB Commands:

- `rank(A)`: Gives the rank of matrix A
- `orth(A)`: Gives an orthogonal basis of column space of A
- `null(A)`: Gives an orthogonal basis of nullspace of A

Linear Algebraic Equations

Consider a set of linear algebraic equations:

$$Ax = y \tag{*}$$

where $A \in \mathbb{R}^{m \times n}$, $\boldsymbol{y} \in \mathbb{R}^{m \times 1}$ are given and $\boldsymbol{x} \in \mathbb{R}^n$ is the unknown vector.

Existence of a solution:

Equation (*) admits a solution $\boldsymbol{x}$ iff $\text{rank}(A) = \text{rank}\big([A\ \boldsymbol{y}]\big)$

A solution exists in (*) for every $\boldsymbol{y}$ iff $\text{rank}(A) = m$

Parameterization of solutions: Let $\boldsymbol{x}_0$ be a solution of (*).

If $\text{rank}(A) = n$, solution $\boldsymbol{x}_0$ is unique.

If $\text{rank}(A) < n$ and $k = n - \text{rank}(A)$, then given any real $\alpha_i, i = 1, ..., k$,

$$\boldsymbol{x} = \boldsymbol{x}_0 + \alpha_1 \boldsymbol{n}_1 + \alpha_2 \boldsymbol{n}_2 + ... + \alpha_k \boldsymbol{n}_k$$

is a solution, where $\{\boldsymbol{n}_1, \boldsymbol{n}_2, ..., \boldsymbol{n}_k\}$ is a basis of the nullspace of A.

Example: Consider

$$A = \begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 2 & 3 & 4 \\ 2 & 0 & 2 & 0 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} -4 \\ -8 \\ 0 \end{bmatrix}.$$

The 3rd column of A is the sum of the first two columns and the 4th column is 2 times the 2nd column. Hence, $\text{rank}(A) = 2$ and $\text{nullity}(A) = 2$.

Next, it can be verified that $\boldsymbol{n}_1 = [1\ 1\ -1\ 0]^T$, $\boldsymbol{n}_2 = [0\ 2\ 0\ -1]^T$ form a basis of the nullspace of A.

Further, $\boldsymbol{x}_0 = [0\ -4\ 0\ 0]^T$ is a solution. Hence the general solution can be written as

$$\boldsymbol{x} = \boldsymbol{x}_0 + \alpha_1 \boldsymbol{n}_1 + \alpha_2 \boldsymbol{n}_2$$

where α_1 and α_2 are any real numbers.

Similarity Transformation

Consider a linear equation

$$A\boldsymbol{x} = \boldsymbol{y}, \quad A \in \mathbb{R}^{n \times n} \tag{**}$$

Here, A maps $\boldsymbol{y} \in \mathbb{R}^n$ to $\boldsymbol{x} \in \mathbb{R}^n$.

Let $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{y}}$ be the representations of $\boldsymbol{x}$ and $\boldsymbol{y}$ w.r.t. another basis $\{\boldsymbol{q}_1, \boldsymbol{q}_2, ..., \boldsymbol{q}_n\}$, Then, with $Q = [\boldsymbol{q}_1 \; \boldsymbol{q}_2 \; ... \; \boldsymbol{q}_n]$, we have

$$\boldsymbol{x} = Q\bar{\boldsymbol{x}}, \quad \boldsymbol{y} = Q\bar{\boldsymbol{y}}$$

Substituting into (**), we have,

$$Q\bar{\boldsymbol{y}} = AQ\bar{\boldsymbol{x}} \;\text{ or }\; \bar{\boldsymbol{y}} = \bar{A}\bar{\boldsymbol{x}} = Q^{-1}AQ\bar{\boldsymbol{x}}$$

Here, $\bar{A} = Q^{-1}AQ$, or $A = Q\bar{A}Q^{-1}$. This is called similarity transformation, and matrices A and $\bar{A}$ are said to be similar.

Eigenvalues and Eigenvectors

Consider an $n \times n$ real matrix A.

A real or complex number λ is called an eigenvalue of A if there exists a nonzero vector $\boldsymbol{v}$ such that

$$A\boldsymbol{v} = \lambda\boldsymbol{v}.$$

Eigenvalues of A are found by computing the roots of characteristic polynomial:

$$\Delta(\lambda) = \det(\lambda I - A)$$

which is a monic polynomial (leading coefficient 1) of degree n and has n roots $\lambda_i, i = 1, .., n$, not necessarily all distinct.

Any nonzero vector $\boldsymbol{v}$ satisfying $A\boldsymbol{v} = \lambda\boldsymbol{v}$ is called an eigenvector of A associated with eigenvalue λ, and is computed by solving

$$(A - \lambda I)\boldsymbol{v} = 0.$$

Spectral radius of A is defined as

$$\rho(A) = \max\{|\lambda_1|, |\lambda_2|, ..., |\lambda_n|\}$$

and for any norm of A, $\rho(A) \leq \|A\|$.

Eigenvalues and Eigenvectors

Let $\lambda_{i_1}, \lambda_{i_2}, ..., \lambda_{i_\eta}$ be the distinct eigenvalues of A. Then, the CP can be written as

$$\Delta(\lambda) = \det(\lambda I - A) = (\lambda - \lambda_{i_1})^{n_1}.(\lambda - \lambda_{i_2})^{n_2}...(\lambda - \lambda_{i_\eta})^{n_\eta}$$

where $\sum_{k=1}^{\eta} n_k = n$.

Algebraic multiplicity of a distinct eigenvalue λ_{i_k} is the number of times it appears as a root of CP.

Geometric multiplicity of a distinct eigenvalue λ_{i_k} is the number of linearly independent eigenvectors associated with it. It is equal to the nullity of $\lambda_{i_k} I - A$.

Exercise 2: Find the eigenvalues, eigenvectors and multiplicities of the following:

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

Eigenvalues and Eigenvectors

Suppose A has n linearly independent eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, .., \boldsymbol{v}_n$ associated with eigenvalues $\lambda_1, \lambda_2, .., \lambda_n$ (not necessarily distinct). Then, with $T = [\boldsymbol{v}_1 \; \boldsymbol{v}_2 \; ... \; \boldsymbol{v}_n]$,

$$T^{-1} A\, T = \mathrm{diag}\big(\lambda_1, \lambda_2, .., \lambda_n\big) = D \quad \text{or} \quad A = T\, D\, T^{-1} = T\, D\, T^{-1}$$

D is called the diagonal form of A and the above transformation is called diagonalization.

Matrix A is diagonalizable if the algebraic and geometric multiplicities of each of its distinct eigenvalue are equal.

In general, there exists a nonsingular transformation T such that

$$T^{-1} A\, T = J = \mathrm{diag}\big(J_1, J_2, .., J_\eta\big)$$

where

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{bmatrix}, \quad i = 1, .., \eta$$

are called Jordan blocks. J is called the Jordan Canonical form of A.

Quadratic Forms and Matrix Definiteness

For a vector $x \in \mathbb{R}^n$, the square of the Euclidean norm is

$$\|x\|^2 = x^T x.$$

If S is any nonsingular transformation, the vector Sx has a norm squared $(Sx)^T Sx = x^T S^T Sx$. Letting $P = S^T S$, we write

$$\|x\|_P^2 = x^T P x$$

as the norm squared of Sx. $\|x\|_P^2$ is also referred to as the norm of x with respect to P.

An expression of the form

$$x^T Q x$$

where x is a vector is called a quadratic form.

A quadratic form is a generalization of the scalar square in higher dimensions.

Quadratic Forms and Matrix Definiteness

A quadratic form $x^T Qx$, with a real matrix Q is equal to $x^T Q_s x$ where Q_s is a real symmetric matrix.

To see that this is true, note that any real square matrix Q, we can write

$$Q = Q_s + Q_a$$

where

$Q_s = \frac{Q+Q^T}{2}$ [symmetric part, $Q_s^T = Q_s$]
$Q_a = \frac{Q-Q^T}{2}$ [antisymmetric part, $Q_a^T = -Q_a$]

If a quadratic form $x^T Ax$ has A antisymmetric, then, it must be equal to zero since $x^T Ax$ is a scalar and

$$x^T Ax = (x^T Ax)^T = x^T A^T x = -x^T Ax$$

Hence, for a general square Q,

$$x^T Qx = x^T (Q_s + Q_a)x = x^T Qx.$$

Therefore, WLOG, we consider a real symmetric matric q in a quadratic form.

Quadratic Forms and Matrix Definiteness

A real matrix Q is symmetric if $Q = Q^T$.

All eigenvalues of a real symmetric matrix Q are real.

A real symmetric matrix Q is:

- Positive definite ($Q \succ 0$) if $x^T Qx > 0$ for all nonzero x.
- Positive semi-definite ($Q \succeq 0$) if $x^T Qx \geq 0$ for all nonzero x.
- Negative semi-definite ($Q \preceq 0$) if $x^T Qx \leq 0$ for all nonzero x.
- Negative definite ($Q \prec 0$) if $x^T Qx < 0$ for all nonzero x.
- Indefinite if $x^T Qx > 0$ for some x and $x^T Qx < 0$ for some other x

We can test for definiteness independently of the vectors x. If $\lambda_i, i = 1, 2, ..$ are the eigenvalues of Q, then

- $Q \succ 0$ if all $\lambda_i > 0$
- $Q \succeq 0$ if all $\lambda_i \geq 0$
- $Q \preceq 0$ if all $\lambda_i \leq 0$
- $Q \prec 0$ if all $\lambda_i < 0$

Quadratic Forms and Matrix Definiteness

A practical test for definiteness: Let $Q = [q_{ij}] \in \mathbb{R}^{n \times n}$.

A $k \times k$ matrix formed by deleting $n - k$ rows and same $n - k$ columns of Q is called a principal submatrix (of order k) of Q.

The determinant of a principal submatrix is called a principal minor.

The leading/principal minors of Q are the principal minors:

$$m_1 = q_{11}$$

$$m_2 = \begin{vmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{vmatrix}$$

$$m_3 = \begin{vmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{32} \\ q_{31} & q_{32} & q_{33} \end{vmatrix}, \ldots.$$

In terms of leading minors, we have

- $Q \succ 0$ if $m_i > 0$ for all i
- $Q \prec 0$ if $m_i < 0$ for all odd i and if $m_i > 0$ for all even i
- $Q \succeq 0$ if all principal minors are non-negative.
- $Q \preceq 0$ if $-Q \succeq 0$.

Quadratic Forms and Matrix Definiteness

Rayleigh-Ritz Inequality For a real symmetric matrix A, its quadratic form satisfies

$$\lambda_{min}(A)\, x^T x \leq x^T A x \leq \lambda_{max}(A)\, x^T x$$

where $\lambda_{min}(.)$ and $\lambda_{max}(.)$ denote the smallest and largest eigenvalues of (.).

With symmetric matrices A and B, inequality $A \succeq B$ (or $B \succeq A$) means $A - B \succeq 0$ (or $B - A \succeq 0$).

Matrix Norms

Consider an $m \times n$ real matrix represented as

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \ddots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

The induced norm of A is defined as

$$\|A\| = \sup_{\boldsymbol{x} \neq 0} \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \sup_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|$$

This is called induced norm as it is defined thorough the norm of $\boldsymbol{x}$.

- $\|A\|_1 = \max_j \sum_{i=1}^{m} |a_{ij}|$ [Maximum 'column sum']
- $\|A\|_2 = \bar{\sigma}(A) = \sqrt{\lambda_{max}(A^* A)}$ [Maximum singular value]
- $\|A\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$ [Maximum 'row sum']

Matrix Norms

The Frobenius norm (not an induced norm) of A is defined as

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2} = \sqrt{\mathrm{tr}(A^* A)}$$

Matrix norm properties:

- $\|A\| \geq 0$, and $\|A\| = 0$ only if $A = 0$.
- $\|A + B\| \leq \|A\| + \|B\|$
- $\|\alpha A\| = |\alpha| \|A\|$, for any scalar α

Singular Value Decomposition

For any real or complex $m \times n$ matrix A, there exist an $m \times m$ matrix U, an $n \times n$ matrix V and an $m \times n$ matrix Σ such that

$$A = U\Sigma V^*, \quad \Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}$$

where $U^*U = I$, $V^*V = I$ (unitary or orthogonal matrices) and

$$\Sigma_1 = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_p \end{bmatrix}_{p \times p},$$

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p$$

$\sigma_{max}(AB) \leq \sigma_{max}(A).\sigma_{max}(B)$

MATLAB Command: `[U S V] = svd(X)`

Some Matrix Relationships

Some identities

$$
\begin{aligned}
AB &\neq BA \quad \text{(not commutative)} \\
(AB)^T &= B^T A^T \\
(AB)^* &= B^* A^* \quad \textbf{*}\text{: complex conjugate} \\
(AB)^{-1} &= B^{-1} A^{-1} \quad (A, B \text{ invertible})
\end{aligned}
$$

Determinants

$$
\begin{aligned}
\det(AB) &= \det(BA) = \det(A)\det(B) \\
\det(A^T) &= \det(A) \\
\det(\alpha A) &= \alpha^n \det(A) \\
\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \det(A)\det(B - CA^{-1}D) \quad \text{(if } A \text{ is nonsingular)} \\
\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \det(B)\det(A - DB^{-1}C) \quad \text{(if } B \text{ is nonsingular)}
\end{aligned}
$$

Kronecker Product

The Kronecker product of two matrices $A = [a_{ij}] \in \mathbb{R}^{m \times n}$ and $B = [b_{ij}] \in \mathbb{R}^{p \times q}$ is

$$A \otimes B = [a_{ij} B] = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{bmatrix} \in \mathbb{R}^{mp \times nq}$$

Some properties:

- $(A + B) \otimes C = A \otimes B + B \otimes C$
- $(A \otimes B) \otimes C = A \otimes (B \otimes C)$
- $(A \otimes B)^T = A^T \otimes B^T$
- $(A \otimes C)(B \otimes D) = AB \otimes CD$
- $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

Kronecker Product

The Kronecker product is useful in vector-related computations.

Let $A = [a_1 \ a_2 \ .... \ a_n]$ be a matrix with columns $a_1, a_2, ..., a_n$. We define

$$vec(A) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots a_n. \end{bmatrix}$$

The $vec(.)$ operator converts a matrix to a vector.

An identity that is often useful is

$$vec(ABD) = (D^T \otimes A)vec(B).$$

For example, consider a matrix recursion

$$P_{k+1} = AP_kA^T + A_1P_kA_1^T$$

Using the above identity, we have

$$P_k = H^k vec(P_0)$$

where $H = A \otimes A + A_1 \otimes A_1$.

Matrix Inversion

Given a $m \times n$ nonsingular A and vectors u and v satisfying $v^* A^{-1} u \neq -1$,

$$(A + uv^*)^{-1} = A^{-1} \frac{A^{-1} uv^* A^{-1}}{1 + v^* A^{-1} u}$$

Sherman-Morrison-Woodbury formula (Matrix inversion lemma)

$$(A + BCD)^{-1} = A^{-1} - A^{-1} B (DA^{-1} B + C^{-1})^{-1} DA^{-1}$$

In particular, $(I + AB)^{-1} A = A(I + BA)^{-1}$

Block inversion

$$\begin{bmatrix} A & 0 \\ C & B \end{bmatrix} = \begin{bmatrix} A^{-1} & 0 \\ -B^{-1} CA^{-1} & B^{-1} \end{bmatrix}$$

$$\begin{bmatrix} A & D \\ 0 & B \end{bmatrix} = \begin{bmatrix} A^{-1} & -A^{-1} DB^{-1} \\ 0 & B^{-1} \end{bmatrix}$$

$$\begin{bmatrix} A & D \\ C & B \end{bmatrix} = \begin{bmatrix} A^{-1}(I + D\Delta^{-1} CA^{-1}) & A^{-1} D\Delta^{-1} \\ -\Delta^{-1} CA^{-1} & \Delta^{-1} \end{bmatrix}$$

if A is invertible, where $\Delta = B - CA^{-1} D$.

Algebraic Matrix Equation

Linear matrix equations are extensions of linear vector equations.

An algebraic matrix equation of the form

$$XD + EX + F = 0$$

has a solution iff $\lambda_i(D) + \lambda_j(E) \neq 0, i = 1, 2, .., n; j = 1, 2, ..., m$

Algebraic matrix equations such as Lyapunov equation, Sylvester equation, Riccati equation are frequently used in control and estimation.

Vector/Matrix Calculus

Let $x = [x_1 \ x_2 \ .. \ x_n] \in \mathbb{R}^n$ be a vector, $s \in \mathbb{R}$ be a scalar and $f(s)$ be a m-valued function of x.

The derivative of x w.r.t. s is given by

$$\frac{dx}{ds} = \begin{bmatrix} dx_1/ds \\ dx_2/ds \\ \vdots \\ dx_n/ds \end{bmatrix}$$

If s is a function of x, then the gradient of s w.r.t. x is defined as

$$s_x \doteq \frac{\partial s}{\partial x} = \begin{bmatrix} \partial s/\partial x_1 \\ \partial s/\partial x_2 \\ \vdots \\ \partial s/\partial x_n \end{bmatrix}$$

The total differential of s is

$$ds = \frac{\partial s}{\partial x}^T dx = \sum_{i=1}^{n} \frac{\partial s}{\partial x_i} dx_i$$

Vector/Matrix Calculus

The Hessian of s w.r.t. x is the second derivative

$$s_{xx} \doteq \left[\frac{\partial^2 s}{\partial x_i \partial x_j}\right]$$

which is a symmetric $n \times n$ matrix.
The Jacobian of the vector-valued function $f(x)$ is defined as

$$\begin{aligned} f_x \doteq \frac{\partial f}{\partial x} &= \begin{bmatrix} \partial f/\partial x_1 & \partial f/\partial x_2 & \dots & \partial f/\partial x_n \end{bmatrix} \\ &= \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \dots & \partial f_1/\partial x_n \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \dots & \partial f_2/\partial x_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial f_m/\partial x_1 & \partial f_m/\partial x_2 & \dots & \partial f_m/\partial x_n \end{bmatrix} \end{aligned}$$

The total differential of f can be written as

$$df = \frac{\partial f}{\partial x} dx = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i$$

Vector/Matrix Calculus

Chain rule

Let f and y be vector functions of x. Then

$$\frac{\partial}{\partial x}(f^T y) = f_x^T y + y_x^T f$$

Let $A(t)$ and $B(t)$ be matrices with entries differentiable by t. Then,

$$\frac{d}{dt}[A(t)B(t)] = \left[\frac{d}{dt}A(t)\right]B(t) + A(t)\left[\frac{d}{dt}B(t)\right]$$

An example of the use: If $A(t)$ is nonsingular,

$$\frac{d}{dt}[A^{-1}(t)] = -A^{-1}(t)\left[\frac{d}{dt}A(t)\right]A^{-1}(t)$$

(Proof:) Note that $A(t)A^{-1}(t) = I$. The result follows by differentiating both sides of this equation.

Vector/Matrix Calculus

Some useful gradients:

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

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

If Q is symmetric,

$$\frac{\partial}{\partial x}(x^T Q x) = 2Qx$$

Some useful Hessians and Jacobians

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

If Q is symmetric,

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

And,

$$\frac{\partial}{\partial x}(Ax) = A$$

Vector/Matrix Calculus

Matrix differential equation:

Consider a differential equation

$$\dot{X} = XD(t) + E(t)X + F(t), \quad X(t_0) = X_0, \quad X \in \mathbb{R}^{m \times n}$$

It has a unique solution is given by

$$X(t) = \Phi_1(t, t_0) X_0 \Phi_2(t, t_0) + \int_{t_0}^{t} \Phi_1(t, \tau) F(\tau) d\tau$$

where

$$\begin{aligned}
\frac{\partial \Phi_1}{\partial t} &= E(t)\Phi_1, \quad \Phi_1(\tau, \tau) = I_m, \\
\frac{\partial \Phi_2}{\partial t} &= \Phi_2 D(t), \quad \Phi_2(\tau, \tau) = I_n,
\end{aligned}$$

Quadratic Forms and Matrix Definiteness

Exercise 3: For each of the following, state whether it is a valid quadratic form and, if yes, write it in the form $x^T Qx$.

$$
\begin{aligned}
f_1(x_1, x_2) &= 10x^2 + 9x_2^2 + x_1x_2 \\
f_2(x_1, x_2) &= x_1^2 + x_2^2 + 2\sqrt{x_1, x_2} \\
f_3(x_1, x_2) &= -100x_1x_2 \\
f_4(x_1, x_2) &= (x_1 - x_2)^2 \\
f_5(x_1, x_2, x_3) &= x_1^2 + 2x_2^2 - 7x_3^2 - 4x_1x_2 + 8x_1x_3
\end{aligned}
$$

Exercise 4: For each quadratic form in Exercise 3, state the definiteness (positive definite, negative definite etc.) of the matrix Q that defines the form.