矩阵分析及其应用

孔贺

南科大，自动化与智能制造学院

kongh@sustech.edu.cn

课程结构

- 一. 线性代数基本知识回顾（3 个学时）
- 二. 线性空间与线性映射（12 个学时）
- 三. λ-矩阵与 Jordan 标准型（8 个学时）
- 四. 内积空间与矩阵分解（12 个学时）
- 五. 范数及其他（8 个学时）

5.1 向量范数

定义：假设 $\mathbf{V}$ 是数域 $\mathbf{F}$（本章讨论实数域 $\mathbf{R}$ 或虚数域 $\mathbf{C}$）上的线性空间。对 $x \in \mathbf{V}$，我们用 $\|x\|$ 表示按照某种法则确定的与向量 x 对应的实值函数，即

$$\|\cdot\| : \mathbf{V} \to \mathbf{R}^{+}$$

我们称上述实值函数为范数，如果其满足

- 1. 非负性

$$\begin{cases} \|x\| > 0, & x \neq 0 \\ \|x\| = 0, & x = 0 \end{cases}$$

- 2. 正齐性

$$\|xk\| = \|x\| \, |k|$$

  其中，k 为数域 $\mathbf{F}$ 中的任意数

- 3. 三角不等式

$$\|x + y\| \leq \|x\| + \|y\|$$

  其中，$x, y \in \mathbf{V}$

- 向量范数的性质
  - (a)
    $$\|0\| = 0$$
  - (b)
    $$\|-x\| = \|x\|$$
  - (c)
    $$\|x - y\| \geq |\|x\| - \|y\||$$
  - (d)
    $$\|x + y\| \geq |\|x\| - \|y\||$$

常用的向量范数

- 给定 $x = [x_1, x_2, \cdots, x_n]^{\mathrm{T}} \in \mathbf{C}^n$，以及 $1 \le p < \infty$，我们称

$$\|x\|_p = \begin{cases} \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{\frac{1}{p}}, & 1 \le p < \infty \\ \max\{|x_1|, |x_2|, \cdots, |x_n|\}, & p = \infty \end{cases}$$

  为向量的 p–范数。可以证明 p–范数满足范数的定义。
- p–范数的例子
  - 1–范数（Manhattan distance）

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

  - 2–范数（亦称为欧式范数）

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

  - ∞–范数

$$\|x\|_\infty = \max\{|x_1|, |x_2|, \cdots, |x_n|\}$$

范数与长度？

回忆内积空间，向量长度的定义

3.2 向量的长度和夹角

- **定义**：假设 $\mathbf{V}$ 为酉空间，给定向量 $\alpha, \beta \in \mathbf{V}$，称 $\|\alpha\| := \sqrt{\langle \alpha, \alpha \rangle}$ 为向量的长度（模）；称 $d(\alpha, \beta) := \|\alpha - \beta\|$ 为向量 α 与 β 之间的距离。
- **定理**：假设 $\mathbf{V}$ 为酉空间，给定向量 $\alpha \in \mathbf{V}$，则其长度 $\|\alpha\|$ 具有以下性质
  - (1) 非负性：$\|\alpha\| \geq 0$，当且仅当 $\alpha = 0$ 时 $\|\alpha\| = 0$
  - (2) 齐次性：$\|\alpha k\| = |k| \, \|\alpha\|$，其中 k 为任意数
  - (3) 满足三角不等式：$\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$
  - (4) 满足 Cauchy-Schwartz 不等式：$|\langle \alpha, \beta \rangle| \leq \|\alpha\| \, \|\beta\|$
  - (5) 满足平行四边形公式：$\|\alpha + \beta\|^2 + \|\alpha - \beta\|^2 = 2(\|\alpha\|^2 + \|\beta\|^2)$

图：第三章　内积空间与矩阵分解

范数是否一定为长度，何时为长度？

- 范数是否一定为长度？
    - 例子：取
$$x=\begin{bmatrix}1\\0\end{bmatrix},\ y=\begin{bmatrix}0\\1\end{bmatrix}$$
      试验证，p–范数是否为长度？
- 需要验证什么？
- 范数何时为长度？

有了长度，为什么还要引进范数

- 引入范数，使得我们可以对两个向量之间的距离做定义
  - 假设 $\|\cdot\|: \mathbf{V} \to \mathbf{R}^+$ 是向量空间上的范数，那么对任意的 $x, y \in \mathbf{V}$，记

    $$d(x, y) = \|x - y\|$$

    为 x, y 之间的距离，也即

    $$d(\cdot, \cdot) = \mathbf{V} \times \mathbf{V} \to \mathbf{R}^+$$

    决定了 $\mathbf{V}$ 上的二元函数。
- 定义距离的好处
  - 定性
  - 定量

2D 中 p–范数的模样

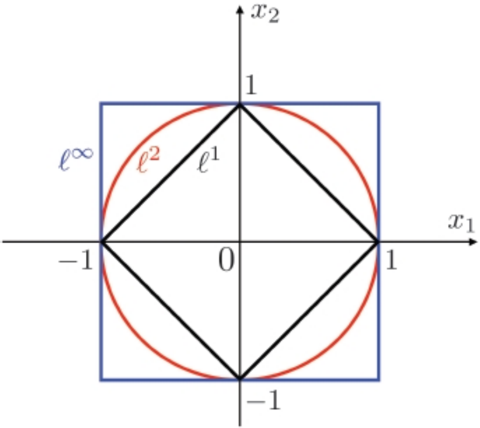

Figure 2.2. Contour curves ($\|x\|_p = 1$) of ℓ^1, ℓ^2, ℓ^∞ norms.

图: Masaaki Nagahara, Sparsity Methods for Systems and Control, 2020

向量范数的等价性

虽然一个向量的不同范数有不同的值，但是它们之间有重要的关系。例如在考虑向量序列收敛性的时候，它们表现出一致性，即范数的等价性。

- 定理：假设 $\mathbf{V}$ 是数域 $\mathbf{F}$ 上的 n 维线性空间。对任意的 $x \in \mathbf{V}$，$\|x\|_\alpha$ 和 $\|x\|_\beta$ 为任意两种向量范数（不限于 p–范数），则总存在两个正实数 c_1, c_2 使得

$$c_1 \|x\|_\beta \le \|x\|_\alpha \le c_2 \|x\|_\beta$$

5.2 矩阵范数

对于任意一个维数为 $m \times n$ 的矩阵 A，用 $\|A\| \to \mathbf{R}^+$ 来表示按照某种法则确定与其对应的实值函数。我们称上述实值函数为 A 的矩阵范数，如果其满足

- 1. 非负性

$$\begin{cases} \|A\| > 0, & A \neq 0 \\ \|A\| = 0, & A = 0 \end{cases}$$

- 2. 正齐性

$$\|Ak\| = \|A\| \, |k|$$

其中，k 为数域 $\mathbf{F}$ 中的任意数

- 3. 三角不等式，即对任意同维数的矩阵 A, B，则有

$$\|A + B\| \leq \|A\| + \|B\|$$

- 4. 矩阵乘法的相容性，即若矩阵 A, B 可乘，则有

$$\|AB\| \leq \|A\| \, \|B\|$$

矩阵范数举例

- 1. 向量 1–范数向矩阵的推广

$$\|A\| = \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|$$

- 2. 向量 2–范数向矩阵的推广（又称 Frobenius 范数）

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}$$

- 3. Frobenius 范数的性质
  - (a) 假设 $A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$，则有 $\|A\|_F^2 = \sum_{i=1}^{n}\|a_i\|_2^2$
  - (b) $\|A\|_F^2 = tr(A^{\mathrm{H}}A) = \sum_{i=1}^{n}\lambda_i(A^{\mathrm{H}}A)$
  - (c) 对任何 m 阶酉矩阵 U 和 n 阶酉矩阵 V，都有

$$\|A\|_F = \|UA\|_F = \left\|A^{\mathrm{H}}\right\|_F = \|AV\|_F = \|UAV\|_F$$

范数的相容性与诱导范数

给定 $x \in \mathbf{C}^n$，$A \in \mathbf{C}^{m\times n}$ 及其范数 $\|A\|$。若将 Ax 与 x 作为矩阵，则有 $\|Ax\| \leq \|A\| \, \|x\|$。更一般性的，若取向量范数 $\|x\|_\alpha$ 与 $\|Ax\|_\alpha$，则不等式

$$\|Ax\|_\alpha \leq \|A\| \, \|x\|_\alpha$$

是否依然成立？

- 定义：设 $\|x\|_\alpha$ 、$\|A\|_\beta$ 分别是向量、矩阵范数。若对于任何的矩阵 A 与向量 x 都有

  $$\|Ax\|_\alpha \leq \|A\|_\beta \, \|x\|_\alpha$$

  则称 $\|A\|_\beta$ 为与向量范数 $\|x\|_\alpha$ 相容的矩阵范数。
- 定理：设 $\|x\|_\alpha$ 为向量范数，定义

  $$\|A\|_\gamma = \max_{x\in\mathbf{C}^n,\ x\neq 0} \frac{\|Ax\|_\alpha}{\|x\|_\alpha} \tag{1}$$

  则 $\|A\|_\gamma$ 满足矩阵范数定义，且其与向量范数 $\|x\|_\alpha$ 相容。
- 定义：由式（1）所定义的矩阵范数称为由向量范数 $\|x\|_\alpha$ 所诱导的范数，简称诱导范数或者算子范数。

矩阵的 p–范数

- **定义**：由向量 p–范数所诱导的矩阵范数称为矩阵的 p–范数，即

$$\|A\|_p = \max_{x \in \mathbf{C}^n,\ x \neq 0} \frac{\|Ax\|_p}{\|x\|_p} \tag{2}$$

- **定理**：给定 $A = (a_{ij})_{m\times n} \in \mathbf{C}^{m\times n}$，则
  - (a) $\|A\|_1 = \max_j \left(\sum_{i=1}^m |a_{ij}|\right),\ j = 1, \cdots, n$ （A 的列和范数）
  - (b) $\|A\|_2 = \max_j \left(\lambda_j(A^{\mathrm{H}}A)\right)^{\frac{1}{2}},\ j = 1, \cdots, n$，其中，$\lambda_j(A^{\mathrm{H}}A)$ 表示矩阵 $A^{\mathrm{H}}A$ 的第 j 个特征值。$\|A\|_2$ 经常被称为谱范数，其是矩阵 A 的最大正奇异值
  - (c) $\|A\|_\infty = \max_i \left(\sum_{j=1}^n |a_{ij}|\right),\ i = 1, \cdots, m$ （A 的行和范数）

矩阵的谱半径

- 定义：给定 $A \in \mathbf{C}^{n\times n}$，且 A 的 n 个特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$，则我们称

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

为 A 的谱半径。

- 定理：给定 $A \in \mathbf{C}^{n\times n}$

$$\rho(A) \leq \|A\|$$

其中，$\|A\|$ 是 A 的任何一种范数。

5.3 范数应用举例

矩阵序列与极限

- 定义：给定矩阵序列 $\{A_k\}$，其中 $A_k = (a_{ij}^k)_{m\times n} \in \mathbf{C}^{m\times n}$。若 $m \times n$ 个数列 $\left\{a_{ij}^k\right\}$，其中 $i = 1, \cdots, m; j = 1, \cdots, m$，都收敛，则称矩阵序列 $\{A_k\}$ 收敛。若
$$\lim_{k\to\infty} a_{ij}^k = a_{ij}$$
则有
$$\lim_{k\to\infty} A_k = A = (a_{ij})_{m\times n}$$
进而称 A 是矩阵序列 $\{A_k\}$ 的极限。
- 定理：矩阵序列 $\{A_k\}$ 收敛于 A 的充要条件是
$$\lim_{k\to\infty} \|A_k - A\| = 0$$
其中 $\|A_k - A\|$ 为任何一种矩阵范数。

说明：一个收敛的矩阵序列的极限唯一。

n 阶方阵的幂组成的矩阵序列的收敛性

给定 $A \in \mathbf{C}^{n\times n}$，研究 $A, A^2, \cdots, A^k, \cdots$ 的收敛性

- 引理：给定 $A \in \mathbf{C}^{n\times n}$，若对于某一种矩阵范数满足

$$\|A\| < 1$$

则有

$$\lim_{k\to\infty} A^k = 0$$

- 定理：给定矩阵序列 $A, A^2, \cdots, A^k, \cdots$，则

$$\lim_{k\to\infty} A^k = 0$$

的充要条件是

$$\rho(A) < 1$$

矩阵级数

- **定义**：给定 $A_k = (a_{ij}^k)_{m\times n} \in \mathbf{C}^{m\times n}$，若 $m \times n$ 个常数项级数

$$\sum_{k=0}^{\infty} a_{ij}^k = a_{ij}^0 + a_{ij}^1 + \cdots + a_{ij}^k + \cdots$$

收敛，其中 $i = 1, \cdots, m; j = 1, \cdots, m$，那么称矩阵级数

$$\sum_{k=0}^{\infty} A_k = A_1 + A_1 + \cdots + A_k + \cdots$$

收敛。若上面的 $m \times n$ 个常数项级数都绝对收敛，则称矩阵级数绝对收敛。

- **定理**：给定 $A_k = (a_{ij}^k)_{m\times n} \in \mathbf{C}^{m\times n}$，则矩阵级数 $\sum_{k=0}^{\infty} A_k$ 绝对收敛的充要条件是正项数项级数 $\sum_{k=0}^{\infty} \|A_k\|$ 收敛，其中 $\|A_k\|$ 为任何一种矩阵范数。

矩阵的幂级数

- **定义**：给定 $A=(a_{ij})_{m\times n}\in\mathbf{C}^{n\times n}$，称形如

$$\sum_{k=0}^{\infty} c_k A^k = c_0 I_n + c_1 A + c_2 A^2 + \cdots + c_k A^k + \cdots$$

的矩阵级数为矩阵幂级数。

- **定理**：若矩阵 A 的某一种范数在幂级数

$$c_0 + c_1 x + c_2 x^2 + \cdots + c_k x^k + \cdots$$

的收敛域内，则矩阵幂级数

$$c_0 I_n + c_1 A + c_2 A^2 + \cdots + c_k A^k + \cdots$$

绝对收敛。

5.4 矩阵函数与微积分

矩阵指数函数

- **定理**：给定 $A \in \mathbf{C}^{n\times n}$ 矩阵级数

$$\sum_{k=0}^{\infty} \frac{A^k}{k!}$$

收敛，也即矩阵序列

$$S_n = \sum_{k=0}^{n} \frac{A^k}{k!}$$

是收敛的，其中 $n = 1, \cdots, \infty$。

- **定义**：

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}, \ e^{At} = \sum_{k=0}^{\infty} \frac{(At)^k}{k!}$$

- **定理**：若 $AB = BA$，则 $e^{A+B} = e^A e^B$

向量与矩阵值函数

给定 $A(t):[a,b]\to \mathbf{C}^{m\times n}$，也记为 $A(t)=[a_{ij}(t)]_{m\times n}$。

- 定义
  - 极限：$\lim\limits_{t\to t_0} A(t)=A$；$\lim\limits_{t\to t_0}\|A(t)-A\|=0$
  - 导数：$A'(t_0)=\frac{dA(t)}{dt}\big|_{t=t_0}$；$\lim\limits_{\Delta t\to 0}\frac{A(t_0+\Delta t)-A(t_0)}{\Delta t}=A'(t_0)$
  - 积分：$\int_a^b A(t)dt=W$
- 定理：
  - 极限

$$\lim_{t\to t_0} A(t)=\left[\lim_{t\to t_0} a_{ij}(t)\right]_{m\times n}$$

  - 微分

$$A'(t)=\left[a'_{ij}(t)\right]_{m\times n}$$

  - 积分

$$\int_a^b A(t)dt=\left[\int_a^b a_{ij}(t)dt\right]_{m\times n}$$

几点性质

- (1)
$$(A(t)B(t))' = A'(t)B(t) + A(t)B'(t)$$
- (2)
$$(A^{-1}(t))' = -A^{-1}(t)(A(t))'A^{-1}(t)$$
- (3)
$$(e^{At})' = Ae^{At} = e^{At}A$$

矩阵微分方程

- 给定 $A \in \mathbf{R}^{n\times n}$，以及

$$\begin{cases} \dot{x}(t) = Ax(t) \\ x(0) = x_0, \ t \geq 0 \end{cases}$$

  求其解？
- 定义：矩阵 $A \in \mathbf{R}^{n\times n}$ 被称为 Hurwitz 渐进稳定，如果 $Re\lambda(A) < 0$
- 问题：给定 $A \in \mathbf{R}^{n\times n}$，如何判断其稳定性？