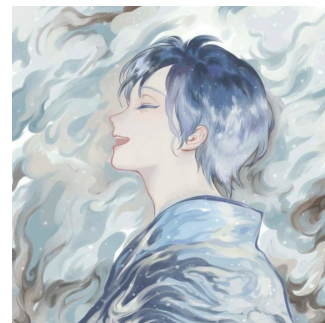




大创参考文献学习笔记

作者：Lollins

时间：November 5, 2023



改变人生的事情，你必须冒险；意义非凡的事情，大多碰巧发生；不重要的事，才有周全的计划。

前言

2023/8/14, 开始阅读老师给的三篇英文文献, 用此文补充一些看大创文献中遇到的知识点。

2023/9/19, 过去一个月了, 最近在网上看到了 `ElegantTeX` 的模板, 感觉好好看, 花了一天的时间把相关的格式和细节都改了。好看了好多, 不过字体变小了不少。

Lollins

November 5, 2023

目录

第 1 章 Matrix Theorem	1
1.1 Laplace Matrix	1
1.2 Kronecker Product	5
1.3 Frobenius Norm and Inner Product	6
1.4 Diagonally-dominant Matrix	6
1.5 Hurwitz Matrix	6
第 2 章 Convex Optimization	9
2.1 Convex Function	9
第 3 章 ODE	14
3.1 Lyapunov Stability Analysis	14
3.2 Exponentially Stable	15
3.3 Gronwall-Bellman Inequality	16
3.4 Comparison Lemma	17
3.5 Laplace Transform	17
3.5.1 Laplace Transform 概念	17
3.5.2 Laplace Transform 相关公式	17
3.5.2.1 线性叠加性	17
3.5.2.2 微分性质	18
3.5.2.3 积分性质	18
3.5.2.4 相似性质	18
3.5.2.5 平移延迟	18
3.5.2.6 位移性质	18
3.5.2.7 卷积定理	19
3.5.3 几个常用的 Laplace Transform	20
3.6 Barablat's Lemma	20
第 4 章 Calculus	21
4.1 矩阵与范数求导	21
4.1.1 梯度	21
4.1.2 矩阵微分	21
4.2 向量函数的 <i>Taylor</i> 公式	21
4.3 数量函数对向量的导数	21
第 5 章 Game Theorem	22
第 6 章 Advanced Algebra	23
6.1 Gerschgorin Circle Theorem—特征值估计	23
6.2 Jordan Form	23
6.3 矩阵特征值不等式	24
附录 A [1]	25

A.1 Terminology	25
A.2 式 (8) 中的 Hurwitz Matrix	25
A.3 式 (11) 的强凹函数不等式	25
A.4 线性化式 (19)	25
A.5 式 (20) 中 $\bar{k}B$ is Hurwitz	26
A.6 式 (20) 中的 x^* 指数稳定	26
A.7 式 (21)	26
A.8 推导出式 (28)	26
附录 B [2]	27
B.1 式 (20) 中的 r_0 取值	27

第 1 章 Matrix Theorem

1.1 Laplace Matrix

Laplace Operator

对于多元函数 $f(x_1, \dots, x_n)$ 的 Laplace Operator 为

$$\Delta f = \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2} \quad (1.1)$$

例题 1.1 三元函数 $f(x, y, z)$ 的 Laplace Operator 为

$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$$

数值微分

根据导数近似计算公式，当 Δ 近似等于 0 时，我们可以得到 f 的导数为

$$\begin{aligned} f'(x) &\approx \frac{f(x + \Delta x) - f(x)}{\Delta x} \\ f''(x) &\approx \frac{f'(x + \Delta x) - f'(x)}{\Delta x} \\ &\approx \frac{f(x + \Delta x) + f(x - \Delta x) - 2f(x)}{(\Delta x)^2} \end{aligned}$$

对二元函数 $f(x, y)$ ，我们对其使用 Laplace Operator，得到

$$\begin{aligned} \Delta f &= \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \\ &\approx \frac{f(x + \Delta x, y) + f(x - \Delta x, y) - 2f(x, y)}{(\Delta x)^2} \\ &\quad + \frac{f(x, y + \Delta y) + f(x, y - \Delta y) - 2f(x, y)}{(\Delta y)^2} \end{aligned}$$

把此二元函数离散化，为了简化，假设 x 和 y 的增量即步长为 1，即

$$\Delta x = x_{i+1} - x_i = 1, \Delta y = y_{j+1} - y_j = 1$$

点 (x_i, y_j) 处的 Laplace Operator 可以用下面的公式近似代替

$$\begin{aligned} &\frac{f(x_i + \Delta x, y_j) + f(x_i - \Delta x, y_j) - 2f(x_i, y_j)}{(\Delta x)^2} \\ &+ \frac{f(x_i, y_j + \Delta y) + f(x_i, y_j - \Delta y) - 2f(x_i, y_j)}{(\Delta y)^2} \\ &= \frac{f(x_{i+1}, y_j) + f(x_{i-1}, y_j) - 2f(x_i, y_j)}{1^2} + \frac{f(x_i, y_{j+1}) + f(x_i, y_{j-1}) - 2f(x_i, y_j)}{1^2} \\ &= f(x_{i+1}, y_j) + f(x_{i-1}, y_j) + f(x_i, y_{j+1}) + f(x_i, y_{j-1}) - 4f(x_i, y_j) \end{aligned}$$

这是非常优美的结果，它就是 (x_i, y_j) 的 4 个相邻点处的函数值之和与 4 倍的 (x_i, y_j) 点处的差值，如图 1.1 所示

示

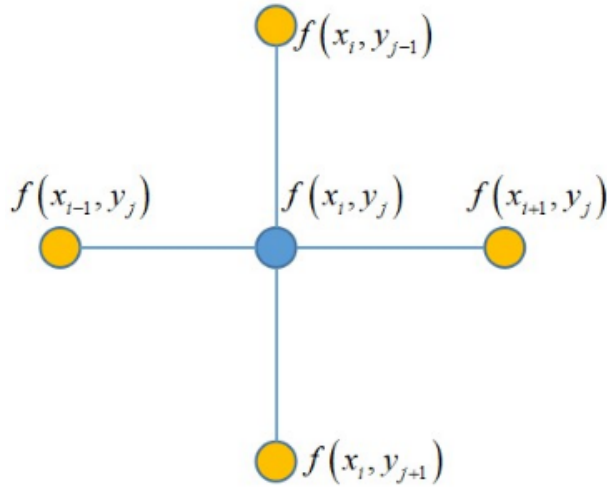


图 1.1

基于这种表示，Laplace Operator 的计算公式可以表示为

$$\begin{aligned}
 \Delta f &= f(x_{i+1}, y_j) + f(x_{i-1}, y_j) + f(x_i, y_{j+1}) + f(x_i, y_{j-1}) - 4f(x_i, y_j) \\
 &= f(x_{i+1}, y_j) - f(x_i, y_j) + f(x_{i-1}, y_j) - f(x_i, y_j) \\
 &\quad + f(x_i, y_{j+1}) - f(x_i, y_j) + f(x_i, y_{j-1}) - f(x_i, y_j) \\
 &= \sum_{(k,l) \in N(i,j)} (f(x_k, y_l) - f(x_i, y_j))
 \end{aligned} \tag{1.2}$$

其中 $N(i, j)$ 为 (x_i, y_j) 的邻居节点

图的邻接矩阵与加权度矩阵

图是一种几何结构，对它的研究起源于古老的哥尼斯堡七桥问题。一个图 $G(\text{graph})$ 由顶点和边构成，通常将顶点的集合记为 $V(\text{vertex})$ ，边的集合记为 $E(\text{edge})$ 。边由其连接的起点和终点表示。如图1.2所示，它是一个典型的图。

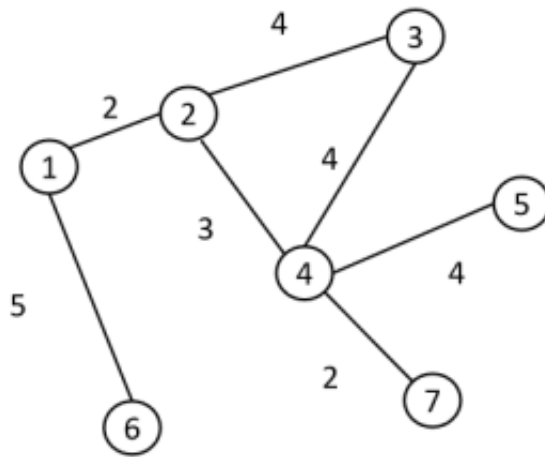


图 1.2

图的边可以有方向的，也可以是无方向的，前者被称为有向图，后者被称为无向图。邻接矩阵是用来方便存储图的结构，用线性代数方法研究图的问题。

如果一个图有 n 个顶点，其邻接矩阵 W 为 $n \times n$ 的矩阵，矩阵元素 w_{ij} 表示边 (i, j) 的权重，如果矩阵两

个顶点之间没有边连接，则记为 0。对于无向连接矩阵，满足 $w_{ij} = w_{ji}$ 。图1.2的邻接矩阵如下

$$\begin{bmatrix} 0 & 2 & 0 & 0 & 0 & 5 & 0 \\ 2 & 0 & 4 & 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 4 & 0 & 0 & 0 \\ 0 & 3 & 4 & 0 & 4 & 0 & 2 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 \end{bmatrix}$$

对于无向图，顶点的加权重是与该顶点相关的所有边的权重之和。对于无向图连接矩阵 W ，顶点 i 的加权重为 W 第 i 行元素之和

$$d_i = \sum_{j=1}^n w_{ij}$$

加权重矩阵 D 为对角矩阵，其主对角线元素为每个顶点的加权重，其他位置的元素为 0。

$$d_{ii} = d_i = \sum_{j=1}^n w_{ij} \quad (1.3)$$

对于图1.2，它的加权重矩阵如式1.4所示

$$\begin{bmatrix} 7 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 13 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix} \quad (1.4)$$

Laplace Matrix

如果将图的顶点处的值看作是函数值，则在顶点 i 处的 Laplace Operator 为

$$\Delta f_i = \sum_{j \in N_i} (f_i - f_j)$$

这里的 N_i 为顶点 i 的所有邻居顶点集合。这里我们调换了 f_i 和 f_j 的位置，和之前的 Laplace Operator 相比，相当于多了一个负号。由于图的边可以带有权重，我们可以在上面的计算公式中加上权重

$$\Delta f_i = \sum_{j \in N_i} w_{ij} (f_i - f_j) \quad (1.5)$$

这一推广如图1.3所示，图中图中红色的顶点是 i ，蓝色的顶点是它的邻居顶点，灰色的顶点是其他顶点。

如果 j 不是 i 的邻居，则 $w_{ij} = 0$ 。因此式1.5也可以写做

$$\Delta f_i = \sum_{j \in N_i} w_{ij} (f_i - f_j) = \sum_{j \in N_i} w_{ij} f_i - \sum_{j \in N_i} w_{ij} f_j = d_i f_i - \mathbf{w}_i \mathbf{f} \quad (1.6)$$

这里的 d_i 就是第 i 个节点的加权重， \mathbf{w}_i 为邻接矩阵第 i 行， \mathbf{f} 是所有顶点的值构成的列向量， $\mathbf{w}_i \mathbf{f}$ 是二者的内积。

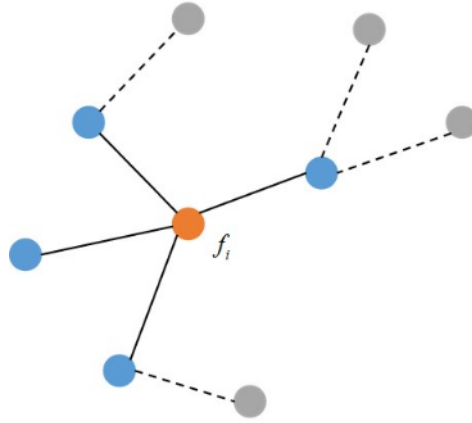


图 1.3

对图的所有顶点，我们有

$$\begin{aligned} \Delta f &= \begin{bmatrix} \Delta f_1 \\ \dots \\ \Delta f_n \end{bmatrix} = \begin{bmatrix} d_1 f_1 - \mathbf{w}_1 \mathbf{f} \\ \dots \\ d_n f_n - \mathbf{w}_n \mathbf{f} \end{bmatrix} = \begin{bmatrix} d_1 & \dots & \dots \\ \dots & \dots & \dots \\ \dots & \dots & d_n \end{bmatrix} \begin{bmatrix} f_1 \\ \dots \\ f_n \end{bmatrix} - \begin{bmatrix} \mathbf{w}_1 \\ \dots \\ \mathbf{w}_n \end{bmatrix} \begin{bmatrix} f_1 \\ \dots \\ f_n \end{bmatrix} \\ &= (\mathbf{D} - \mathbf{W}) \mathbf{f} \end{aligned} \quad (1.7)$$

我们在邻接矩阵和加权重度矩阵的基础上定义 **Laplace Matrix**。假设无向图 G 有 n 个顶点，邻接矩阵为 W ，加权重度矩阵为 D 。**Laplace Matrix** 定义为加权重度矩阵与邻接矩阵之差，即

$$L = D - W \quad (1.8)$$

则图1.2的 Laplace Matrix 为

$$\begin{bmatrix} 7 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 13 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 2 & 0 & 0 & 0 & 5 & 0 \\ 2 & 0 & 4 & 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 4 & 0 & 0 & 0 \\ 0 & 3 & 4 & 0 & 4 & 0 & 2 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 7 & -2 & 0 & 0 & 0 & -5 & 0 \\ -2 & 9 & -4 & -3 & 0 & 0 & 0 \\ 0 & -4 & 8 & -4 & 0 & 0 & 0 \\ 0 & -3 & -4 & 13 & -4 & 0 & -2 \\ 0 & 0 & 0 & -4 & 4 & 0 & 0 \\ -5 & 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & -2 & 0 & 0 & 2 \end{bmatrix}$$

显然 Laplace Matrix 的每行元素之和都为 0。

Properties

1. 对任意向量 $\mathbf{f} \in \mathbb{R}^n$ ，有

$$\mathbf{f}^T \mathbf{L} \mathbf{f} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} (f_i - f_j)^2$$

2. Laplace Matrix 是对称半正定矩阵；

3. Laplace Matrix 的最小特征值为 0，其对应的特征向量为常向量 1，即所有分量为 1，即 $L \mathbf{1}_n = \mathbf{0} \mathbf{1}_n = \mathbf{0}_n$ ；

4. Laplace Matrix 有 n 个非负实数特征值，并且满足

$$\lambda_n \geq \dots \geq \lambda_1 \geq 0$$

相关证明可以参考这篇知乎文章 [理解图的拉普拉斯矩阵](#)

1.2 Kronecker Product

定义 1.1

给定任意矩阵 $X \in \mathbb{R}^{m \times n}$ 和 $Y \in \mathbb{R}^{p \times q}$ 则矩阵 X 和 Y 的 Kronecker Product 为

$$X \otimes Y = \begin{bmatrix} x_{11}Y & x_{12}Y & \cdots & x_{1n}Y \\ x_{21}Y & x_{22}Y & \cdots & x_{2n}Y \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1}Y & x_{m2}Y & \cdots & x_{mn}Y \end{bmatrix} \in \mathbb{R}^{(mp) \times (nq)} \quad (1.9)$$

其中， \otimes 表示 Kronecker Product。



注：根据式1.9，我们可以得到

$$Y \otimes X = \begin{bmatrix} y_{11}X & y_{12}X & \cdots & y_{1q}X \\ y_{21}X & y_{22}X & \cdots & y_{2q}X \\ \vdots & \vdots & \ddots & \vdots \\ y_{p1}X & y_{p2}X & \cdots & y_{pq}X \end{bmatrix} \in \mathbb{R}^{(mp) \times (nq)}$$

这意味着 $X \otimes Y$ 和 $Y \otimes X$ 不相同，即 Kronecker Product 不存在交换律。

Properties

1. 结合律 (associativity)

$$X \otimes Y \otimes Z = (X \otimes Y) \otimes Z = X \otimes (Y \otimes Z) \quad (1.10)$$

2. 分配律 (distributivity)

$$(X + Y) \otimes Z = X \otimes Z + Y \otimes Z \quad (1.11)$$

3. 给定任意矩阵 $X \in \mathbb{R}^{m \times n}$ 和 $Y \in \mathbb{R}^{p \times q}$ ，则

$$(X \otimes Y)^T = X^T \otimes Y^T \quad (1.12)$$

4. 给定任意矩阵 $X \in \mathbb{R}^{m \times n}$ 、 $Y \in \mathbb{R}^{s \times t}$ 、 $U \in \mathbb{R}^{n \times p}$ 和 $V \in \mathbb{R}^{t \times q}$ ，则

$$(X \otimes Y)(U \otimes V) = (XU) \otimes (YV) \in \mathbb{R}^{(ms) \times (pq)} \quad (1.13)$$

5. 给定矩阵 $X \in \mathbb{R}^{m \times n}$ 和 $Y \in \mathbb{R}^{p \times q}$ 都是非奇异的，则

$$(X \otimes Y)^{-1} = X^{-1} \otimes Y^{-1} \quad (1.14)$$

还有一些特殊性质（懒打公式了，先从网上找了图片截屏放进去），如图1.4所示

【性质6】迹： $\text{tr}(X \otimes Y) = \text{tr}(X) \cdot \text{tr}(Y)$, $X \in \mathbb{R}^{m \times m}$, $Y \in \mathbb{R}^{n \times n}$

【性质7】F-范数： $\|X \otimes Y\|_F = \|X\|_F \cdot \|Y\|_F$

【性质8】 ℓ_2 范数： $\|x \otimes y\|_2 = \|x\|_2 \cdot \|y\|_2$

【性质9】行列式： $\det(X \otimes Y) = \det(X)^n \cdot \det(Y)^m$, $X \in \mathbb{R}^{m \times m}$, $Y \in \mathbb{R}^{n \times n}$

【性质10】秩： $\text{rank}(X \otimes Y) = \text{rank}(X) \cdot \text{rank}(Y)$

图 1.4

本节主要参考了知乎的 [代数基础 | Kronecker 积](#)，文章中给出了详细的证明。

1.3 Frobenius Norm and Inner Product

定义 1.2

对于矩阵 $A \in C^{m \times n}$, 其 Frobenius Norm 为

$$\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{\frac{1}{2}} \quad (1.15)$$

其 Frobenius Inner Product 为

$$\langle A, B \rangle_F = \sum_{j=1}^m \sum_{i=1}^n A_{ij}^* B_{ij} \quad (1.16)$$



又由于

$$\begin{aligned} \text{tr}(A^H B) &= \sum_{i=1}^n (A^H B)_{ii} \\ &= \sum_{i=1}^n \sum_{k=1}^m A_{ik}^H B_{ki} \\ &= \sum_{i=1}^n \sum_{k=1}^m A_{ki}^* B_{ki} \end{aligned}$$

故我们可以得到

$$\langle A, B \rangle_F = \text{tr}(A^H B)$$

本节主要参考了 [Frobenius 范数和内积的关系](#)

1.4 Diagonally-dominant Matrix

定义 1.3 (Strictly Diagonally-dominant Matrix)

对于一个矩阵 $A_{n \times n}$, 满足 $\forall i = 1, 2, \dots, n$, 有

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}| \quad (1.17)$$

那么我们称 $A_{n \times n}$ 为 Strictly Diagonally-dominant Matrix。



定理 1.1

如果矩阵 A 严格对角占优, 则 A 非奇异。



具体的证明可以参考这篇文章[严格对角占优矩阵非奇异](#)。

1.5 Hurwitz Matrix

定义 1.4 (Hurwitz Matrix)

给定一个多项式 (多项式的所有根都有负实部)

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \dots + a_1 z + a_0 \quad (1.18)$$

和 n 阶矩阵

$$H = \begin{pmatrix} a_{n-1} & a_{n-3} & a_{n-5} & \cdots & \cdots & \cdots & 0 & 0 & 0 \\ a_n & a_{n-2} & a_{n-4} & & & & \vdots & \vdots & \vdots \\ 0 & a_{n-1} & a_{n-3} & & & & \vdots & \vdots & \vdots \\ \vdots & a_n & a_{n-2} & \ddots & & & 0 & \vdots & \vdots \\ \vdots & 0 & a_{n-1} & & \ddots & & a_0 & \vdots & \vdots \\ \vdots & \vdots & a_n & & & \ddots & a_1 & 0 & \vdots \\ \vdots & \vdots & 0 & & & & a_2 & a_0 & \vdots \\ \vdots & \vdots & \vdots & & & & a_3 & a_1 & 0 \\ 0 & 0 & 0 & \cdots & \cdots & \cdots & a_4 & a_2 & a_0 \end{pmatrix} \quad (1.19)$$

那么称矩阵 H 为多项式 p 的 Hurwitz Matrix。

p 多项式的所有根都有负实部 $\iff H$ 的所有顺序主子式大于 0。



例题 1.2 在 Mathematica 中调用函数指令，可以得到 $p(x) = 5 + 4x + 3x^2$ 和 $p(x) = 5 + 4x + 3x^2 + 2x^3 + x^4$ 的 2 阶与 4 阶的 Hurwitz Matrix，分别如图 1.5 和图 1.6 所示（原本想 \LaTeX 嵌入 mma 代码，奈何网上资料太少加上自己能力有限，嵌入失败，就截图放里面了）

```
ResourceFunction["HurwitzMatrix"][5 + 4 x + 3 x^2, x] // MatrixForm
```

[41] ✓ 0.1s Wolfram Language

$$\begin{pmatrix} 4 & 0 \\ 3 & 5 \end{pmatrix}$$

图 1.5

```
ResourceFunction["HurwitzMatrix"][5 + 4 x + 3 x^2 + 2 x^3 + x^4, x] // MatrixForm
```

[43] ✓ 0.1s Wolfram Language

$$\begin{pmatrix} 2 & 4 & 0 & 0 \\ 1 & 3 & 5 & 0 \\ 0 & 2 & 4 & 0 \\ 0 & 1 & 3 & 5 \end{pmatrix}$$

图 1.6

定理 1.2

求解 n 阶方阵的全部顺序主子式 D_i ， D_i 表示矩阵的第 i 阶主子式。

1. 如果所有主子式 D_i 均为正，则系统稳定；
2. 如果存在某个主子式 D_i 为零，但其后续主子式均为正，则系统稳定；
3. 如果任意一个主子式 D_i 为负，则系统不稳定。



定理 1.3

A 为 Hurwitz Matrix 的充分必要条件为，对于 \forall 正定矩阵 Q ，存在满足方程

$$PA + A^T P = -Q$$

的正定矩阵 P ，且 P 是方程的唯一解。



第 2 章 Convex Optimization

2.1 Convex Function

基本的定义与性质

定义 2.1 (Convex Function)

对于函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 若满足 $\text{dom}(f) \subset \mathbb{R}^n$ 是凸集, $\forall t \in [0, 1], \forall \mathbf{x}, \mathbf{y} \in \text{dom}(f)$, 有

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y}) \quad (2.1)$$

则这个函数是一个凸函数。如果等号处处不成立, 则称它是一个严格凸函数 (Strictly)。

定义 2.2 (Concave Functions)

若 $-f$ 是一个凸函数, 那么 f 所定义的函数就是凹函数。如果等号处处不成立, 则称它是一个严格凹函数。

注 1: 凸函数与凹函数不是二元对立的关系, 比如 $y = ax + bx \in \mathbb{R}$, 它是凸函数也是凹函数

注 2: 无论是凸函数, 还是凹函数, 一定要求定义域为凸集。而且没有凹集的说法

定义 2.3 (First-order Characterization of Convex Functions)

如果 f 一阶可导, $\text{dom}(f)$ 是凸的, 并且 $\forall \mathbf{x}, \mathbf{y} \in \text{dom}(f)$, 有

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) \quad (2.2)$$

则称它是一个凸函数。

其实定义2.3可以看成 *Taylor* 公式的一阶展开。

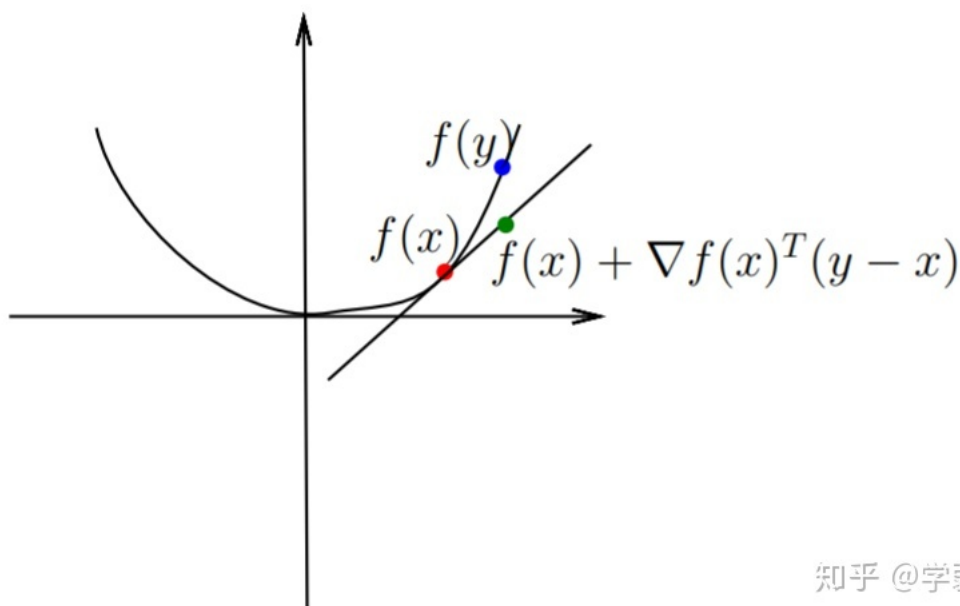


图 2.1

对于这个一阶条件定义, 可以参照图2.1来理解。

它给了我们一个重要的启示: 如果我们假设 $\nabla f(\mathbf{x}) = 0$, 那么无论 \mathbf{y} 是什么, 都有 $f(\mathbf{y}) \geq f(\mathbf{x})$ 。潜在的意

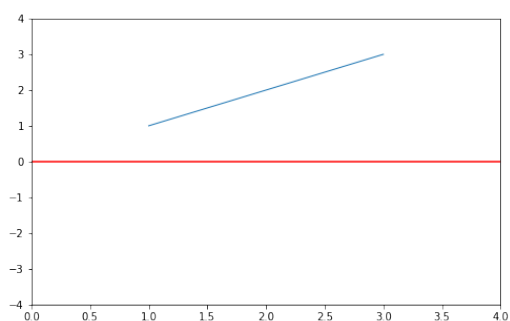
思是，对于驻点为 0 的 x 点，我们取到了最小值。因此正如机器学习界经常说的一句话：当一个问题被证明是一个凸优化问题，那么基本上就可以算解决了。

定义 2.4 (Second-order Characterization of Convex Functions)

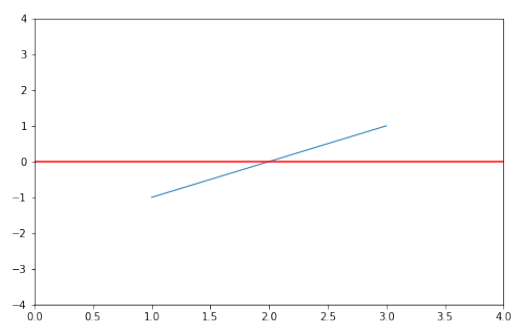
如果 f 是二阶可微的， $\text{dom}(f)$ 是凸的，并且 $\forall x \in \text{dom}(f), \nabla^2 f(x) \succeq 0$ 。那么 f 就是一个凸函数。



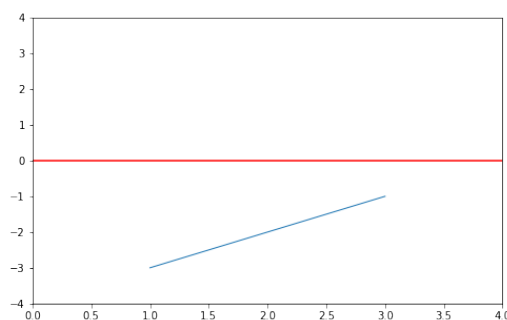
二阶导数（即 $\nabla^2 f$ ）是用来衡量一阶导数的变化率。考虑 $\text{dom}(f) \subset \mathbb{R}$ 的一维情况，对应 $\nabla^2 f \geq 0$ 画出 ∇f 图像的三种情况，如图 2.2 所示，对应的 f 图像如 2.3 所示，我们可以发现图 2.3 都为凸函数。



(a) ∇f 在 x 轴上方



(b) ∇f 穿过 x 轴



(c) ∇f 在 x 轴下方

图 2.2: ∇f 图像的三种情况

推论 2.1

如果 f 是凸函数，那么 $\forall x, y$ ，我们有

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq 0 \quad (2.3)$$



证明 1、从一维的角度理解， $\nabla^2 f \geq 0 \Rightarrow \nabla f$ 为增函数，即证！

2、根据公式 2.2，我们有

$$\begin{cases} f(y) \geq f(x) + \nabla f(x)^T (y - x) \\ f(x) \geq f(y) + \nabla f(y)^T (x - y) \end{cases}$$

将两式相加，即得

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq 0$$

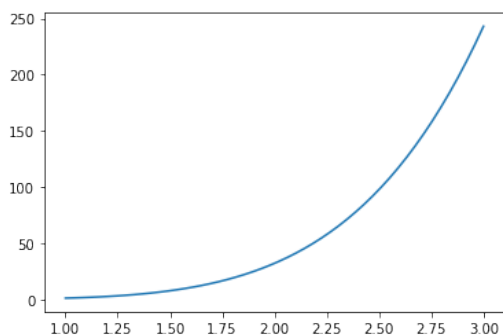
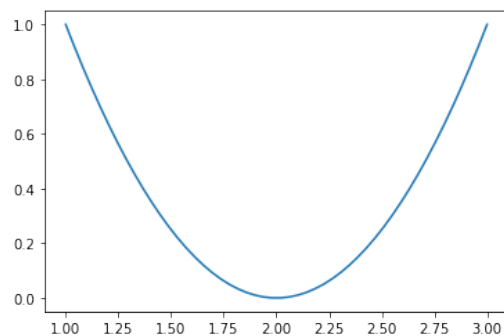
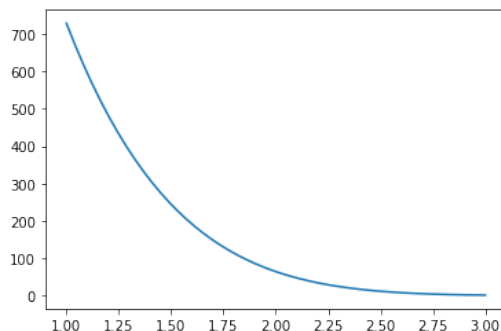
证毕！

注：如果一个函数是严格凸的，并不能推出 $\nabla^2 f > 0$ ，一个反例就是 $f(x) = x^4$

命题 2.1

给定函数 $f(x)$ ，如果 $g(t) = f(x + tv)$ 是一个关于 t 的凸函数，则 $f(x)$ 是凸函数。反之亦然。



(a) 对应 ∇f 在 x 轴上方(b) 对应 ∇f 穿过 x 轴(c) 对应 ∇f 在 x 轴下方图 2.3: f 图像的三种情况

证明 这里我们仅必要性，充分性同理可以推出。根据凸函数的性质，我们可以得到

$$g(\theta t_1 + (1 - \theta)t_2) \leq \theta g(t_1) + (1 - \theta)g(t_2) \quad (2.4)$$

再由 $g(t) = f(\mathbf{x} + t\mathbf{v})$ ，带入式2.4，得

$$\begin{aligned} f(\mathbf{x} + (\theta t_1 + (1 - \theta)t_2)\mathbf{v}) &= f(\theta(\mathbf{x} + t_1\mathbf{v}) + (1 - \theta)(\mathbf{x} + t_2\mathbf{v})) \\ &\leq \theta f(\mathbf{x} + t_1\mathbf{v}) + (1 - \theta)f(\mathbf{x} + t_2\mathbf{v}) \end{aligned}$$

证毕！

引理 2.1

对于函数 $g_1(\mathbf{x}), \dots, g_2(\mathbf{x})$ 都是凸的，那么它们的非负凸组合得到的函数也是凸的。对于凹函数也有类似的结论。



引理 2.2

设 $f(\mathbf{x}, \mathbf{y})$ 是一个凸函数，那么函数 $g(\mathbf{x}) = \sup_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})$ 也是一个凸函数。



复合函数的凸性与应用

所谓的复合函数，就是形如 $f(x) = h(g(x))$ ，我们假设它的性质足够好，具有二阶可导性。那么我们就可以得到

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x) \quad (2.5)$$

定理 2.1 (Rules for Composite Convex Functions)

设 f, g, h 二阶可导, 且 $f(x) = h(g(x))$, 那么

1. 如果 h 为凸函数, 并且不降, g 为凸函数, 那么 f 为凸函数.
2. 如果 h 为凸函数, 不增, g 为凹函数, 那么 f 为凸函数.
3. 如果 h 为凹函数, 不降, g 为凹函数, 那么 f 为凹函数.
4. 如果 h 为凹函数, 不增, g 为凸函数, 那么 f 为凹函数.



只需要将每个条件带入到 $f''(x)$ 中, 观察它的正负即可验证。下面给出条件 1 的证明

$$h'' \geq 0, h' \geq 0, g'' \geq 0 \Rightarrow f'' \geq 0 \Rightarrow f \text{ 为凸函数}$$

幸运的是, 虽然定理 2.1 只是一维的情况, 但是其推广到多维。

强凸与强光滑性

定义 2.5 (Strong Convexity)

若函数 $f(x) - \frac{m}{2}\|x\|^2$ 是一个凸函数, 那么 $f(x)$ 就是一个凸性量为 m 的强凸函数。



定义 2.5 可以等价于公式 2.6, 下面给出证明

$$\begin{aligned} f(\theta x_1 + (1-\theta)x_2) &\leq \theta f(x_1) + (1-\theta)f(x_2) \\ &\quad - \frac{m}{2}\theta(1-\theta)\|x_1 - x_2\|^2 \end{aligned} \quad (2.6)$$

证明 由于 $f(x) - \frac{m}{2}\|x\|^2$ 是一个凸函数, 根据凸函数的性质, 我们有

$$\begin{aligned} f(\theta x_1 + (1-\theta)x_2) - \frac{m}{2}\|\theta x_1 + (1-\theta)x_2\|^2 &\leq \theta(f(x_1) - \frac{m}{2}\|x_1\|^2) \\ &\quad + (1-\theta)(f(x_2) - \frac{m}{2}\|x_2\|^2) \end{aligned} \quad (2.7)$$

其实我们只需要证明??的等式形式就行了, 即证

$$\begin{aligned} &\frac{m}{2}\|\theta x_1 + (1-\theta)x_2\|^2 - \frac{m}{2}[\theta\|x_1\|^2 + (1-\theta)\|x_2\|^2] \\ &= -\frac{m}{2}\theta(1-\theta)\|x_1 - x_2\|^2 \end{aligned} \quad (2.8)$$

根据 2-范数的定义展开, 得

$$\begin{aligned} \|\theta x_1 + (1-\theta)x_2\|^2 &= \sum_{i=1}^n (\theta x_{1i} + (1-\theta)x_{2i})^2 \\ \|x_1\|^2 &= \sum_{i=1}^n x_{1i}^2 \\ \|x_2\|^2 &= \sum_{i=1}^n x_{2i}^2 \\ \|x_1 - x_2\|^2 &= \sum_{i=1}^n (x_{1i} - x_{2i})^2 \end{aligned}$$

将上面得式子带入公式 2.8 中, 证毕!

注: 没有额外加标识的范数都是 2-范数。

定义 2.6 (Smoothness)

如果 $\nabla f(x)$ 满足 $\forall x, y, \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, L > 0$, 那么称它具有 L 的光滑性度量。



下面介绍一些相关的等价定理。

定理 2.2

以下几个性质是等价的：

- 1、 f 是强凸的，且凸性度量为 m ；
- 2、 $(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^T(\mathbf{x} - \mathbf{y}) \geq m\|\mathbf{x} - \mathbf{y}\|^2, \forall \mathbf{x}, \mathbf{y}$ ；
- 3、 $\nabla^2 f(\mathbf{x}) \succeq mI, \forall \mathbf{x}$
- 4、 $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{m}{2}\|\mathbf{y} - \mathbf{x}\|^2$



注：如果 f 是强凹函数，那么 $-f$ 为强凸函数。

定理之间的互推略过，可以参考本文的最后。

定理 2.3

以下几个性质是等价的：

- 1、 $\nabla f(\mathbf{x})$ 是 Lipschitz 连续的，且光滑性度量为 L ；
- 2、 $(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^T(\mathbf{x} - \mathbf{y}) \leq L\|\mathbf{x} - \mathbf{y}\|^2, \forall \mathbf{x}, \mathbf{y}$ ；
- 3、 $\nabla^2 f(\mathbf{x}) \preceq LI, \forall \mathbf{x}$
- 4、 $f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|^2$



强凸和强光滑性是非常好的函数性质，而且就定理2.2和2.3就可以看出，它们利用了二阶信息，对函数分别提供了一个上界和一个下界。

本节主要参考了[凸优化 | 笔记整理（2）——凸函数，强凸函数及相关拓展](#)

第3章 ODE

3.1 Lyapunov Stability Analysis

What is Lyapunov Stability?

早在 1892 年，俄国有一个叫李雅普诺夫的学者发表了一篇著名的文章《运动稳定性一般》问题，建立了关于运动稳定的一般理论，光看这个文章的名字就不一般，也确实，在尔后百余年，这个理论在数学、力学和控制理论中全面开花，已经成为稳定性研究方向的基础性理论，俄罗斯人对于数学上和工程上的直觉确实令人赞叹。



图 3.1

李雅普诺夫稳定性理论研究的是在扰动下平衡点的稳定性问题，可以将问题分为以下四种情况（如图3.2所示）：

1. 如果平衡状态 x_e 受到扰动后，仍然停留在 x_e 附近，我们就称 x_e 在李雅普诺夫意义下是稳定的（Lyapunov stable）；
2. 如果平衡状态 x_e 受到扰动后，最终都会收敛到 x_e ，我们就称 x_e 在李雅普诺夫意义下是渐进稳定的（Asymptotically stable）；
3. 如果平衡状态 x_e 受到任何扰动后，最终都会收敛到 x_e ，我们就称 x_e 在李雅普诺夫意义下是大范围内渐进稳定的（Asymptotically stable in large）；
4. 如果平衡状态 x_e 受到某种扰动后，状态开始偏离 x_e ，我们就称 x_e 在李雅普诺夫意义下是不稳定的（Unstable）。

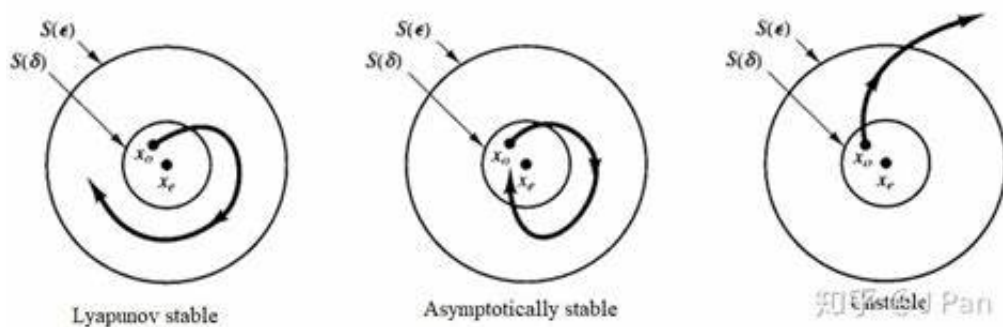


图 3.2: Lyapunov Stability

李雅普诺夫第二法

第二法比较天才，来源于一个朴素的想法：稳定的系统能量总是不断被耗散的，李雅普诺夫通过定义一个标量函数 $V(\mathbf{x})$ （通常能代表广义能量）来分析稳定性。这种方法的避免了直接求解方程，也没有进行近似线性化，所以也一般称之为直接法。

定理 3.1

如果标量函数 $V(\mathbf{x})$ 满足：

- $V(\mathbf{x}) = 0$ if and only if $\mathbf{x} = 0$
- $V(\mathbf{x}) > 0$ if and only if $\mathbf{x} \neq 0$
- $\dot{V}(\mathbf{x}) = \frac{d}{dt}V(\mathbf{x}) = \sum_{i=1}^n \frac{\partial V}{\partial x_i} f_i(\mathbf{x}) \leq 0$ when $\mathbf{x} \neq 0$

则称系统在李雅普诺夫意义下是稳定的，特别的，若 $\mathbf{x} \neq 0$ 时，有 $\dot{V}(\mathbf{x}) < 0$ ，则系统是渐进稳定的。

本节主要参考了[如何理解李雅普诺夫稳定性分析](#)

3.2 Exponentially Stable

定义 3.1 (Exponentially Stable)

假设稳定点 $\mathbf{x}^* = 0$ 为 Exponentially Stable 平衡点，那么满足存在两个正数 k 和 λ ， $D = \{\mathbf{x} \in R^n \mid \|\mathbf{x}\| < r\}$ ， $\forall t \geq t_0, \forall \mathbf{x}(t_0) \in D$ ，使得

$$\|\mathbf{x}(t)\| \leq k\|\mathbf{x}(t_0)\|e^{-\lambda(t-t_0)} \quad (3.1)$$

其中 λ 为指数收敛率， k 为超调量。

定理 3.2

假设 $\mathbf{x} = 0$ 为非线性系统 $\dot{\mathbf{x}} = f(\mathbf{x})$ 的平衡点，且 $f(\mathbf{x})$ 在 $\mathbf{x} = 0$ 的邻域是连续可微的。令

$$A = \left. \frac{\partial f}{\partial \mathbf{x}} \right|_{\mathbf{x}=0} \quad (3.2)$$

当且仅当 A 是 Hurwitz Matrix(见定义1.4) 时， $\mathbf{x} = 0$ 为非线性系统的 Exponentially Stable 平衡点。

定理 3.3 (Exponential stability theorem)

假设 $\mathbf{x} = 0$ 为非线性系统 $\dot{\mathbf{x}} = f(\mathbf{x}, t)$ 的 Exponentially Stable 平衡点， $\forall t \geq t_0, \forall \mathbf{x}(t_0) \in D$ ，存在函数 V 满足以下不等式：

$$\begin{aligned} c_1\|\mathbf{x}\|^2 &\leq V(t, \mathbf{x}) \leq c_2\|\mathbf{x}\|^2 \\ \frac{\partial V}{\partial t} + \frac{\partial V}{\partial \mathbf{x}} f(t, \mathbf{x}) &\leq -c_3\|\mathbf{x}\|^2 \\ \left\| \frac{\partial V}{\partial \mathbf{x}} \right\| &\leq c_4\|\mathbf{x}\| \end{aligned} \quad (3.3)$$

其中 c_1, c_2, c_3, c_4 ，为正常数。

3.3 Gronwall-Bellman Inequality

定理 3.4 (Gronwall-Bellman Inequality)

设 $I = [a, b]$, α, β, u 为 I 上的实值函数, α, u 是 I 上的连续函数, β 是非负的。如果 u 满足

$$u(t) \leq \alpha(t) + \int_a^t \beta(s)u(s)ds, \quad \forall t \in I \quad (3.4)$$

那么

$$u(t) \leq \alpha(t) + \int_a^t \alpha(s)\beta(s) \exp\left(\int_s^t \beta(r)dr\right) ds, \quad t \in I \quad (3.5)$$

如果 α 是非递减函数, 我们可以得到

$$u(t) \leq \alpha(t) \exp\left(\int_a^t \beta(s)ds\right), \quad t \in I \quad (3.6)$$

证明 我们定义一个函数

$$v(s) = \exp\left(-\int_a^s \beta(r)dr\right) \int_a^s \beta(r)u(r)dr, \quad s \in I \quad (3.7)$$

求导, 得

$$v'(s) = \left(\underbrace{u(s) - \int_a^s \beta(r)u(r)}_{\leq \alpha(s)} \right) \beta(s) \exp\left(-\int_a^s \beta(r)dr\right), \quad s \in I \quad (3.8)$$

积分, 得

$$v(t) \leq \int_a^t \alpha(s)\beta(s) \exp\left(-\int_a^s \beta(r)dr\right) ds \quad (3.9)$$

再根据式3.7, 得

$$\begin{aligned} \int_a^t \beta(s)u(s)ds &= \exp\left(\int_a^t \beta(r)dr\right) v(t) \\ &\leq \int_a^t \alpha(s)\beta(s) \exp\left(\underbrace{\int_a^t \beta(r)dr - \int_a^s \beta(r)dr}_{=\int_s^t \beta(r)dr}\right) ds \end{aligned} \quad (3.10)$$

把式3.10代入式3.4, 即得式3.5。

再根据 α 为非递减函数, 我们可以得到 $\alpha(s) \leq \alpha(t)$, 带入式3.5, 即得

$$\begin{aligned} u(t) &\leq \alpha(t) + \left(-\alpha(t) \exp\left(\int_s^t \beta(r)dr\right) \right) \Big|_{s=a}^{s=t} \\ &= \alpha(t) \exp\left(\int_a^t \beta(r)dr\right), \quad t \in I. \end{aligned}$$

即证式3.6。

证毕!

3.4 Comparison Lemma

引理 3.1 (Comparison Lemma)

考虑一个标量微分方程

$$\dot{u} = f(t, u), \quad u(t_0) = u_0 \quad (3.11)$$

对于所有 $t \geq 0$ 和所有 $u \in J \subset \mathbb{R}$, $f(t, u)$ 对于 t 连续可微, 且对于 u 是局部 Lipschitz 的。设 $[t_0, T)$ (T 可以是无限的) 是解 $u(t)$ 存在的最大区间, 并且假设对于所有 $t \in [t_0, T)$, 有 $u(t) \in J$ 。设 $v(t)$ 是连续函数, 其上右导数 $D^+v(t)$ 对于所有 $t \in [t_0, T)$, $v(t) \in J$ 满足微分不等式

$$D^+v(t) \leq f(t, v(t)), \quad v(t_0) \leq u_0 \quad (3.12)$$

那么, 对于所有 $t \in [t_0, T)$, 有 $v(t) \leq u(t)$ 。



3.5 Laplace Transform

3.5.1 Laplace Transform 概念

定义 3.2 (Laplace Transform)

给定函数 $f(t)$ 在 $t \geq 0$ 上, f 的 Laplace Transform 为 F , F 写作

$$F(s) = \mathcal{L}[f(t)] = \int_0^\infty e^{-st} f(t) dt \quad (3.13)$$

其中 $s = \sigma + \omega i$, σ 和 ω 为实数, i 为虚数单位。



定义 3.3 (Inverse Transforms)

如果 $F(s) = \mathcal{L}[f(t)]$, 我们称 f 为 F 的 Inverse Transforms, 写作

$$f(t) = \mathcal{L}^{-1}[F(s)] = \frac{1}{2\pi i} \lim_{T \rightarrow \infty} \int_{\gamma-iT}^{\gamma+iT} e^{st} F(s) ds \quad (3.14)$$



3.5.2 Laplace Transform 相关公式

定理 3.5 (Laplace Transform 存在定理)

对于函数 $f(t)$ 满足下面两个条件:

- 在 $t \geq 0$ 的任何有限区间上分段连续;
- 在 $t \rightarrow +\infty$ 时, 存在常数 $k > 0$ 和 c 使得 $|f(t)| \leq ke^{ct}$ 。

则函数 $f(t)$ 的 Laplace Transform 存在。



3.5.2.1 线性叠加性

$$\mathcal{L}[af(t) + bg(t)] = a\mathcal{L}[f(t)] + b\mathcal{L}[g(t)] \quad (3.15)$$

易证, 跳过证明。

3.5.2.2 微分性质

$$\mathcal{L}[f^{(n)}(t)] = s^n \mathcal{L}[f(t)] - s^{n-1}f(0) - s^{n-2}f'(0) - \dots - f^{(n-1)}(0) \quad (3.16)$$

证明

$$\mathcal{L}[f^{(n)}(t)] = \int_0^\infty f^{(n)}(t) e^{-st} dt$$

分部积分, 得

$$\begin{aligned} \int_0^\infty f^{(n)}(t) e^{-st} dt &= \int_0^\infty e^{-st} df^{(n-1)}(t) \\ &= -f^{(n-1)}(0) + s \int_0^\infty f^{(n-1)}(t) e^{-st} dt \\ &= -f^{(n-1)}(0) + s \mathcal{L}[f^{(n-1)}(t)] \end{aligned}$$

递推, 即得式3.16。

3.5.2.3 积分性质

$$\mathcal{L}\left[\underbrace{\int_0^t dt \int_0^t dt \dots \int_0^t f(t) dt}_{n\text{重}}\right] = \frac{1}{s^n} \mathcal{L}[f(t)] \quad (3.17)$$

证明 令 $F(t) = \underbrace{\int_0^t dt \int_0^t dt \dots \int_0^t f(t) dt}_{n\text{重}}$, 则式3.17可以写成 $\mathcal{L}[F(t)] = \frac{1}{s^n} \mathcal{L}[F^{(n)}(t)]$, 其中 $F(0) = F^{(1)}(0) = \dots = F^{(n-1)}(0) = 0$ 。

带入式3.16, 即证。

3.5.2.4 相似性质

$$\mathcal{L}[f(at)] = \frac{1}{a} F\left(\frac{s}{a}\right) \quad (3.18)$$

相似性质又称为尺度变换, 使用 Laplace Transform 展开, 易证。

3.5.2.5 平移延迟

$$\mathcal{L}[f(t - \tau)] = e^{-\tau s} F(s) \quad (3.19)$$

使用 Laplace Transform 展开, 再使用换元法, 易证。

3.5.2.6 位移性质

$$\mathcal{L}[e^{at} f(t)] = F(s - a) \quad (3.20)$$

使用 Laplace Transform 展开, 易证。

3.5.2.7 卷积定理

定义 3.4 (卷积)

$$f(t) * g(t) = \int_0^t f(\tau)g(t-\tau)d\tau \quad (3.21)$$

根据上面卷积的定义，我们可以得到

$$\mathcal{L}[f(t) * g(t)] = \mathcal{L}[f(t)] \cdot \mathcal{L}[g(t)] \quad (3.22)$$

证明 令 $u = t - \tau$,

$$\begin{aligned} \mathcal{L}[f(t) * g(t)] &= \int_0^\infty dt \int_0^t e^{-st} f(\tau)g(t-\tau)d\tau \\ &= \int_{-\tau}^\infty du \int_0^t e^{-s(u+\tau)} f(\tau)g(u)d\tau \\ &= \mathcal{L}[f(t)] \cdot \mathcal{L}[g(t)] \end{aligned}$$

上面的证明有点问题，哪天有时间我再改改，证明参考图3.3。

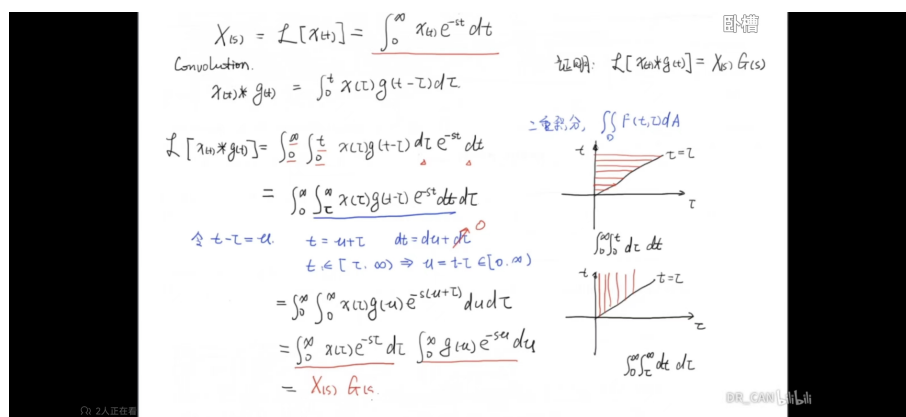


图 3.3

3.5.3 几个常用的 Laplace Transform

$$\mathcal{L}[1] = \frac{1}{s}$$

$$\mathcal{L}[e^{at}] = \frac{1}{s-a}$$

$$\mathcal{L}[t^n e^{at}] = \frac{\Gamma(n+1)}{(s-a)^{n+1}} \{n > -1\}$$

$$\mathcal{L}[t^n] = \frac{\Gamma(n+1)}{s^{n+1}} \{n > -1\}$$

$$\mathcal{L}[\sin at] = \frac{a}{s^2 + a^2}$$

$$\mathcal{L}[\cos at] = \frac{s}{s^2 + a^2}$$

$$\mathcal{L}[\sin hat] = \frac{a}{s^2 - a^2}$$

$$\mathcal{L}[\cos hat] = \frac{s}{s^2 - a^2}$$

$$\mathcal{L}[t \sin at] = \frac{2as}{(s^2 + a^2)^2}$$

$$\mathcal{L}[t \cos at] = \frac{s^2 - a^2}{(s^2 + a^2)^2}$$

3.6 Barbrat's Lemma

定理 3.6 (Theorem 1)

- $t \geq 0$ 时 f 一致连续;
- $\lim_{t \rightarrow +\infty} \int_0^t f(\tau) d\tau$ 存在且有限;

那么我们可以得到 $\lim_{t \rightarrow +\infty} f(t) = 0$ 。



定理 3.7 (Theorem 2)

$\lim_{t \rightarrow +\infty} V$ 收敛, 且 \dot{V} 一致连续, 那么 $\lim_{t \rightarrow +\infty} \dot{V} = 0$ 。



定理 3.8 (Theorem 3)

$\lim_{t \rightarrow +\infty} V$ 收敛, 且 \ddot{V} 有界, 那么 $\lim_{t \rightarrow +\infty} \dot{V} = 0$



定理 3.9 (Theorem 4 (Lyapunov-Like Lemma))

- $V(t, x)$ 有下界;
- $\dot{V}(t, x)$ 半负定;
- $\dot{V}(t, x)$ 一致连续;

那么 $\lim_{t \rightarrow +\infty} \dot{V} = 0$ 。



第 4 章 Calculus

4.1 矩阵与范数求导

4.1.1 梯度

定理 4.1

假设 \mathbf{x} 为 n 维向量，在微分多元函数时经常使用以下规则：

- 1、对于所有 $\mathbf{A} \in \mathbb{R}^{m \times n}$ ，都有 $\nabla_{\mathbf{x}} \mathbf{A} \mathbf{x} = \mathbf{A}^T$
- 2、对于所有 $\mathbf{A} \in \mathbb{R}^{n \times m}$ ，都有 $\nabla_{\mathbf{x}} \mathbf{x}^T \mathbf{A} = \mathbf{A}$
- 3、对于所有 $\mathbf{A} \in \mathbb{R}^{n \times n}$ ，都有 $\nabla_{\mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{x} = (\mathbf{A} + \mathbf{A}^T) \mathbf{x}$
- 4、 $\nabla_{\mathbf{x}} \|\mathbf{x}\|^2 = \nabla_{\mathbf{x}} \mathbf{x}^T \mathbf{x} = 2\mathbf{x}$



同样，对于任何矩阵 \mathbf{X} ，都有 $\nabla_{\mathbf{x}} \|\mathbf{X}\|_F^2 = 2\mathbf{X}$ 。正如我们之后将看到的，梯度对于设计深度学习中的优化算法有很大用处。

4.1.2 矩阵微分

推论 4.1

假设 \mathbf{x} 是 t 的函数，则

$$\frac{d(\mathbf{x}^T \mathbf{A} \mathbf{x})}{dt} = (d\mathbf{x})^T \mathbf{A} \mathbf{x} + \mathbf{x}^T \mathbf{A} (d\mathbf{x}) \quad (4.1)$$



4.2 向量函数的 Taylor 公式

对于向量 \mathbf{x}, \mathbf{x}_0 和标量函数 $f(\mathbf{x})$ ，我们有

$$f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^T \nabla f(\mathbf{x}_0) + \dots \quad (4.2)$$

4.3 数量函数对向量的导数

定义 4.1

设 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ ， $f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$ 是以向量 \mathbf{x} 为自变量的数量函数，即为 n 元函数，则规定数量函数 $f(\mathbf{x})$ 对于向量 \mathbf{x} 的导数为

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right)^T \quad (4.3)$$



第 5 章 Game Theorem

第 6 章 Advanced Algebra

6.1 Gerschgorin Circle Theorem—特征值估计

定理 6.1 (圆盘第一定理)

设 A 是 n 阶复矩阵, $A = (a_{ij})_{n \times n}$, 则 A 的特征值 z 在复平面上的下列圆盘 (又称戈氏圆盘) 中:

$$|z - a_{ii}| \leq R_i, i = 1, 2, \dots, n$$

其中 R_i 为 A 的第 i 行元素去掉 a_{ii} 后的绝对值之和, 即

$$R_i = \sum_{j \neq i}^n |a_{ij}| = |a_{i1}| + \dots + |a_{i,i-1}| + |a_{i,i+1}| + \dots + |a_{in}|$$



引理 6.1

设

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

是 n 次复系数多项式, 则 $f(x)$ 的 n 个根 $\lambda_1, \lambda_2, \dots, \lambda_n$ 都是 $a_n, a_{n-1}, \dots, a_1, a_0$ 的连续函数。



定理 6.2 (圆盘第二定理)

设矩阵 $A = (a_{ij})_{n \times n}$ 的 n 戈氏圆盘分成若干个连通区域, 若其中一个连通区域含有 k 个戈氏圆盘, 则有且只有 k 个特征值落在这个连通区域内 (若两个戈氏圆盘重合, 需计重数; 又若特征值为重根, 也计重数)。



本节主要参考了特征值的估计——圆盘定理 (Gerschgorin (戈氏) 圆盘第一定理) 和特征值的估计——圆盘定理 (Gerschgorin (戈氏) 圆盘第二定理), 在文章中会有具体的证明。

6.2 Jordan Form

定义 6.1 (Jordan Form)

矩阵 J 除了主对角线和主对角线上方元素之外, 其余都是 0, 且主对角线上方的对角线的系数若不为 0 只能为 1。且这 1 的左方和下方的系数 (都在主对角线上) 有相同的值。

$$J = \text{Diag}(J_1, J_2, \dots, J_r) = \begin{pmatrix} \boxed{J_1} & 0 & \dots & 0 \\ 0 & \boxed{J_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \boxed{J_r} \end{pmatrix} \quad (6.1)$$

其中 J_1, J_2, \dots, J_r 为 Jordan Block。易知对角矩阵是一种特殊的 Jordan 标准型矩阵。



定义 6.2 (Jordan Block)

形如矩阵 $J(\lambda, t)$ 的形式被称为 Jordan Block。

$$J(\lambda, t) = \begin{pmatrix} \lambda & 0 & \cdots & 0 & 0 & 0 \\ 1 & \lambda & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \lambda & 0 \\ 0 & 0 & \cdots & 0 & 1 & \lambda \end{pmatrix}_{(t \times t)} \quad (6.2)$$



例题 6.1 对于矩阵 J ，它的 Jordan Block 为 $\begin{bmatrix} 1 \end{bmatrix}$ ， $\begin{bmatrix} 2 \end{bmatrix}$ ， $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ 。

$$J = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

6.3 矩阵特征值不等式

定理 6.3

对于任意向量 x 和实对称矩阵 A ，我们有

$$\lambda_{\min}(A) \leq \frac{x^T A x}{x^T x} \leq \lambda_{\max}(A) \quad (6.3)$$



证明 存在正交矩阵 T ，满足

$$T^T A T = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

令 $x = T y$ ，我们可以得到

$$\frac{x^T A x}{x^T x} = \frac{y^T \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} y}{y^T y} = \frac{\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2}{y_1^2 + y_2^2 + \cdots + y_n^2}$$

放缩，即得

$$\lambda_{\min}(A) \frac{y_1^2 + y_2^2 + \cdots + y_n^2}{y_1^2 + y_2^2 + \cdots + y_n^2} \leq \frac{x^T A x}{x^T x} \leq \lambda_{\max}(A) \frac{y_1^2 + y_2^2 + \cdots + y_n^2}{y_1^2 + y_2^2 + \cdots + y_n^2}$$

证毕！

附录 A [1]

A.1 Terminology

1. penetration:n. 渗透
2. exploit:vt. 利用
3. utilize:vt. 利用
4. asynchronous:adj. 异步的
5. payoff:n. 收益
6. saddle point:n. 鞍点
7. permutation:n. 排列
8. consensus:n. 一致
9. protocol:n. 协议
10. collaboratively:adv. 合作的
11. convergence:n. 收敛
12. adoption:n. 采用
13. successively:n. 相继的
14. auxiliary:adj. 辅助的

A.2 式 (8) 中的 Hurwitz Matrix

试证明 $-(L \otimes I_{N \times N} + B_0)$ 是 Hurwitz Matrix

证明 对角线上的元素为 $-a_i - \sum_{j=1}^N a_{ij}$, $R_i = \sum_{j=1}^N |a_{ij}|$, 根据定理6.1, 我们可以得到

$$\lambda_i \leq -a_i - \sum_{j=1}^N a_{ij} + R_i = -a_i \leq 0$$

特征值都小于等于 0, 故其为 Hurwitz Matrix。

A.3 式 (11) 的强凹函数不等式

如果 f 是强凹函数, 那么 $-f$ 就是一个强凸函数, 再根据定理2.2 中的第 2 个性质, 即毕!

A.4 线性化式 (19)

我们取 $\mathbf{x} - \mathbf{x}_0 \rightarrow 0$, 根据公式4.2, 我们有

$$\begin{aligned} \frac{\partial f_i(\mathbf{x})}{\partial x_i} &= \frac{\partial f_i(\mathbf{x}^*)}{\partial x_i} + (\mathbf{x} - \mathbf{x}^*)^T \left(\frac{\partial^2 f_i(\mathbf{x}^*)}{\partial x_i \partial \mathbf{x}} \right) \\ &= (\mathbf{x} - \mathbf{x}^*)^T \left(\frac{\partial^2 f_i(\mathbf{x}^*)}{\partial x_i \partial \mathbf{x}} \right) \end{aligned}$$

证毕!

A.5 式 (20) 中 \bar{k} Bis Hurwitz

根据定理6.1, 我们知道

$$|\lambda_i - \frac{\partial^2 f_i}{\partial x_i^2}| \leq R_i$$

即

$$\frac{\partial^2 f_i}{\partial x_i^2} - R_i \leq \lambda_i \leq \frac{\partial^2 f_i}{\partial x_i^2} + R_i$$

而

$$|\frac{\partial^2 f_i}{\partial x_i^2}| > |R_i|, \frac{\partial^2 f_i}{\partial x_i^2} < 0$$

所以 $\lambda_i < 0, i = 1, 2, \dots, N$, 证毕!

A.6 式 (20) 中的 x^* 指数稳定

根据定理3.2即得。

A.7 式 (21)

根据定理3.3可得, 其中, 因为论文是时不变系统, 所以第二个公式没有对 t 求导。

A.8 推导出式 (28)

已知 $V = cW_1(\mathbf{x}) + (1-c)\bar{\mathbf{y}}^T P_1 \bar{\mathbf{y}}$, 试证明 $\bar{c}_1 \|\mathbf{z}\|^2 \leq V \leq \bar{c}_2 \|\mathbf{z}\|^2$, 再证明 $\|\mathbf{z}(t)\| \leq \sqrt{\frac{\bar{c}_2}{\bar{c}_1}} e^{-\frac{\delta \lambda_{\min}(B_1)}{2\bar{c}_2} t} \|\mathbf{z}(0)\|$ 。

证明 先证明第一个式子不等式左边, 结合式 (21) 的第一个式子

$$\begin{aligned} \|V\| &\geq c c_1 \|\mathbf{x} - \mathbf{x}^*\|^2 + (1-c) \lambda_{\min}(P) \|\bar{\mathbf{y}}\|^2 \\ &\geq \bar{c}_1 \|\mathbf{z}\|^2 \text{ (注 } \|\mathbf{z}\|^2 = \|\mathbf{x} - \mathbf{x}^*\|^2 + \|\bar{\mathbf{y}}\|^2 \text{)} \end{aligned}$$

同理, 不等式右边也成立。

再证第二个不等式,

$$\begin{aligned} \dot{V} &\leq \bar{c}_2 \|\dot{\mathbf{z}}\|^2 \\ &\leq -\delta \lambda_{\min}(B_1) \|\mathbf{z}\|^2 \end{aligned}$$

积分, 即得 $\|\mathbf{z}(t)\| \leq \sqrt{\frac{\bar{c}_2}{\bar{c}_1}} e^{-\frac{\delta \lambda_{\min}(B_1)}{2\bar{c}_2} t} \|\mathbf{z}(0)\|$

证毕!

附录 B [2]

B.1 式 (20) 中的 r_0 取值

如图B.1所示，考虑一个不等式 $A \leq f(x) \leq B$ ，其中 $A(2, 4), B(8, 9)$ ，那么对于 $x_1, x_2 \in [2, 8]$ ，我们都有 $\frac{f(x_1)}{f(x_2)} \leq \frac{B}{A}$ 。所以，我们有 $\frac{V(t_1)}{V(t_2)} \leq r_0$ 。

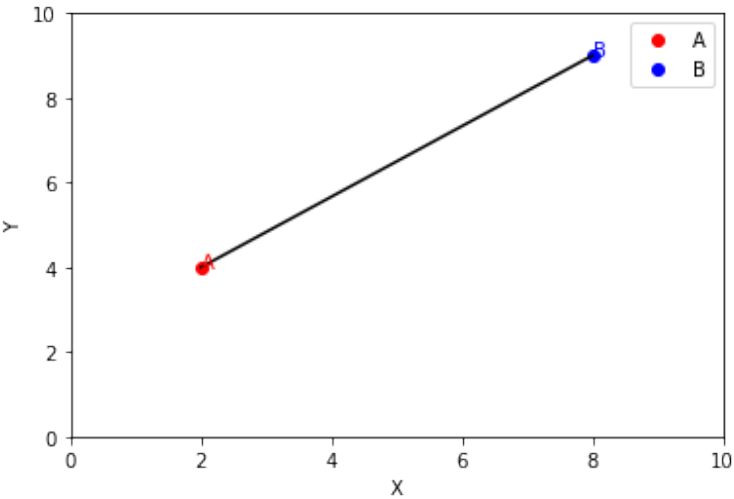


图 B.1