第一讲 代数与矩阵基础（2）

李宇峰
liyf@nju.edu.cn
人工智能学院

代数与矩阵基础

- 向量，矩阵
- 线性空间，线性子空间
  - 零空间，像空间
- 线性相关，线性无关
  - 空间的维数与基底
- 矩阵的基本运算
  - 转置，共轭转置
  - 相乘，求逆
- 矩阵的数值特征
  - 行列式，二次型，特征值，trace，rank
- 内积与范数
  - 向量范数
  - 矩阵范数

5. 矩阵的特征值

学习特征值具有多方面的重要意义

- **矩阵相似性判定**：相似矩阵具有相同的特征值，特征值是重要的依据。相似矩阵在很多数学性质上是等价的，通过**研究特征值可以对矩阵进行分类和等价关系的判断**。

- **矩阵的秩与行列式**：矩阵的非零特征值的个数与矩阵的秩有密切关系，同时矩阵的行列式等于其所有特征值的乘积。**这为通过计算特征值来确定矩阵的秩和行列式提供了一种途径**。

- **图像识别与处理**：在图像压缩、特征提取和模式识别等方面有广泛应用。例如，在人脸识别中，通过对人脸图像矩阵进行特征值分解，可以提取出人脸的主要特征，用于识别和分类。
- **数据挖掘与机器学习**：在聚类分析、降维算法等机器学习任务中，特征值起着重要作用。如在 K-Means 聚类算法中，有时需要通过对数据的协方差矩阵的特征值分析来确定数据的分布特征，从而更好地进行聚类。

5. 矩阵的特征值

- 矩阵的特征值是n阶矩阵的最重要的数值特征之一。

而且也是一个有重要应用背景的数值特征。

- 线性变换T的输出向量与输入向量只相差一个比例因

子λ，也就是

$$Tu = \lambda u, \quad u \neq 0 \quad (5.1)$$

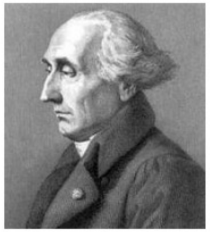
即称标量λ和向量u分别是线性变换T的特征值(eigenvalue)

和特征向量(eigenvector) 。

5. 矩阵的特征值

正式概念提出：

- **19世纪**，矩阵和向量空间概念逐渐成形。1855年，凯莱在研究线性变换下的不变量时，定义了方阵的特征方程和特征根（特征值）。1858年，他在《矩阵论的研究报告》中系统阐述了矩阵理论，也涉及特征值。1878年，Frobenius用双线性型语言建立矩阵理论，进一步发展了矩阵特征值相关内容
- **20世纪**：线性代数从计算方法转向抽象理论。希尔伯特在 1904 年使用 “eigen” 来表示特征值这一概念，使得特征值的定义更加明确和规范。数学家们对矩阵特征值的研究不断深入，将其与线性空间、线性映射等抽象概念相结合，形成了完整的理论体系。

5. 矩阵的特征值

- 当线性变换T在给定基底下的矩阵表示是一个$n \times n$的矩阵时，确定数λ和非零向量u使得

$$Au = \lambda u \qquad (5.2)$$

这个问题就是矩阵A的特征值问题，数λ就是A的一个特征值，非零向量u就是A的对应于λ的一个特征向量。

注意：特征向量不是唯一的(αu也是特征向量)。

- 由特征值的定义，A的特征值问题可以表示为，确定非零向量$u \neq 0$，使得

$$(A - \lambda I)u = 0 \quad (5.3)$$

5. 矩阵的特征值

根据线性代数知识，(5.3)式存在有非零解$u \neq 0$的充要条件是矩阵$A - \lambda I$的行列式等于零，即

$$\det(A - \lambda I) = 0 \qquad (5.4)$$

显然，只要(5.4)式中有一个特征值为零，则就有$\det(A) = 0$．这表明矩阵只要有一个特征值为零，那么这个矩阵A一定是奇异矩阵。相反，非奇异矩阵一定没有特征值为零。

5. 矩阵的特征值

矩阵A的特征值常用符号$eig(A)$来表示。下面我们给出特征值的一些有用的基本性质：

(1) 若A, B都为$n \times n$矩阵，则矩阵乘积的特征值$eig(AB) = eig(BA)$；

(2) 若A为$m \times n$矩阵，B为$n \times m$矩阵，那么$eig(AB)$和$eig(BA)$有相同的非零特征值，所不同的是零特征值的重数不一样；

(3) 逆矩阵的特征值$eig(A^{-1}) = 1/eig(A)$；

(4) 设I_n为$n \times n$单位矩阵，c为标量，则

$$eig(I_n + cA) = 1 + c \cdot eig(A)$$

$$eig(A - cI_n) = eig(A) - c;$$

5. 矩阵的特征值

(5) A的 相 异 特征值所对应的特征向量是线性无关的，就是说若向量$\lambda_i \neq \lambda_j$，向量u_i和u_j分别是对应于λ_i，λ_j的特征向量，那么u_i，u_j一定线性无关；

(6) 若A是$n \times n$实对称矩阵，则其所有特征值都是实数，其n个特征向量可以构成一个完备系，即一定存在n个互相正交的特征向量可以构成n维空间的基底。

5. 矩阵的特征值

我们这里考虑矩阵的正定性、半正定性等都是实对称矩阵，其实这些矩阵的特性都可以用其特征值来描述：

(1) 正定矩阵 $\Leftrightarrow$ 所有特征值都大于零；

(2) 半正定矩阵 $\Leftrightarrow$ 所有特征值为非负实数；

(3) 负定矩阵 $\Leftrightarrow$ 所有特征值为负的实数；

(4) 半负定矩阵 $\Leftrightarrow$ 每个特征值为非正实数；

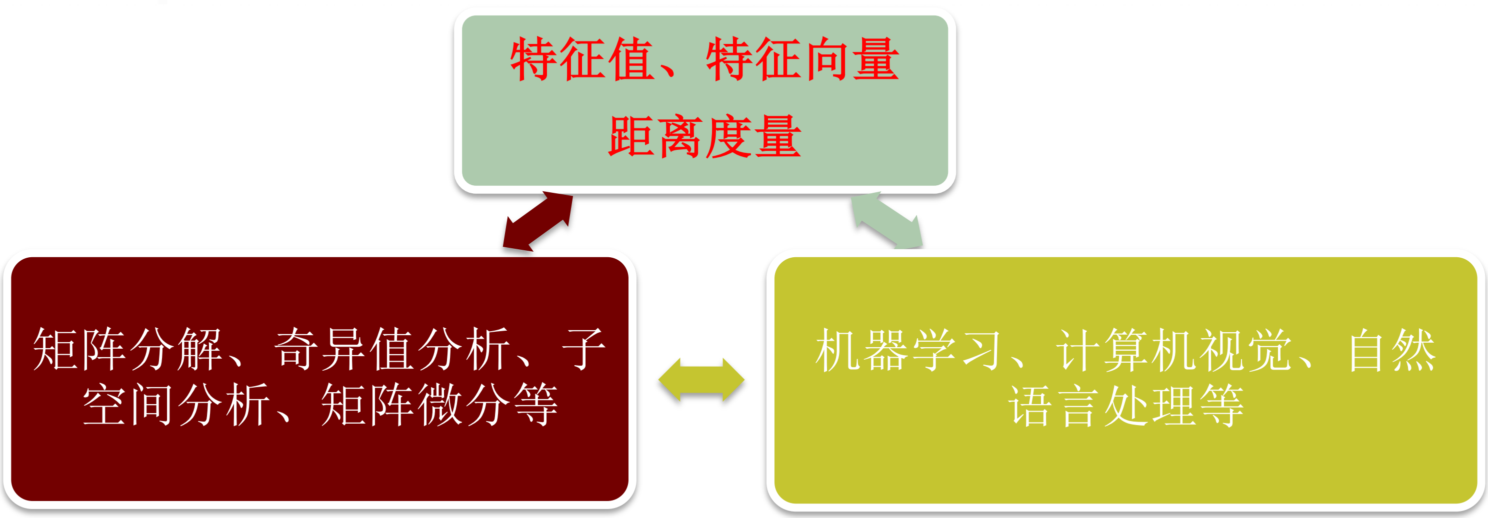
(5) 不定矩阵 $\Leftrightarrow$ 特征值有正有负的实数。

就是说，矩阵的特征值是刻画矩阵特性（奇异性、正定性等）的重要特征。

矩阵的性能指标

表 1.3.1　矩阵的性能指标

性能指标	描述的矩阵性能
二次型	矩阵的正定性与负定性
行列式	矩阵的奇异性
特征值	矩阵的奇异性、正定性和对角元素的结构
迹	矩阵对角元素之和、特征值之和
秩	行 (或列) 之间的线性无关性；线性方程组的适定性



矩阵的迹

- **特征值的统计量**：矩阵的迹等于其所有特征值之和，这一性质使得迹成为研究矩阵特征值的重要工具。通过计算矩阵的迹，可以快速得到特征值的一个重要统计量，为进一步分析特征值的分布和性质提供了便利。
  - 例如，在判断矩阵是否可对角化时，迹与特征值的关系可以作为一个重要的判断依据。若两个矩阵相似，它们具有相同的特征值，自然也具有相同的迹，所以迹可用于初步判断矩阵之间的相似关系。

- **应用例子-模型评估**：在机器学习中，一些模型评估指标与矩阵的迹有关。例如，均方误差（MSE）可以表示为预测值与真实值之间误差矩阵的迹的形式。通过最小化误差矩阵的迹，可以优化模型的参数，使得模型的预测结果更接近真实值。

矩阵的迹

矩阵的迹

定义 设A为一个$n \times n$矩阵，则称其对角线元素之和为A的迹(trace)，记作$tr(A)$. 即

$$tr(A) = a_{11} + a_{22} + \cdots + a_{nn}$$

矩阵A的迹对于A的特征值之和，即$tr(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

这说明，矩阵的迹反映矩阵所有特征值之和。

矩阵的迹

- 只有方阵才有迹，长方矩阵没有迹的概念。
- 矩阵的迹有一些基本性质：

(1) 线性性： $tr(\alpha A+\beta B)=\alpha tr(A)+\beta tr(B)$； $tr(cA)=ctr(A)$；

$$tr(A\pm B)=tr(A)\pm tr(B);$$

(2) 设$A\in R^{m\times n}$， $B\in R^{n\times m}$， 则$tr(AB)=tr(BA)$;

(3) 若A是一个$m\times n$矩阵，则$tr(A^HA)=0\ \Leftrightarrow\ A=0_{m\times n}$(零矩阵);

(4) $tr(xy^H)=y^Hx$; $x^HAx=tr(Axx^H)$;

矩阵的迹

- $\mathrm{trace}(\boldsymbol{A}+\boldsymbol{B}) =$ \trace$\boldsymbol{A} + \mathrm{trace}\ \boldsymbol{B}$

  推导

  $\mathrm{trace}(\boldsymbol{A}+\boldsymbol{B}) = \sum_{i=1}^{n}(a_{ii}+b_{ii}) = \sum_{i=1}^{n} a_{ii} + \sum_{i=1}^{n} b_{ii} = \mathrm{trace}\ \boldsymbol{A} + \mathrm{trace}\ \boldsymbol{B}$

- $\mathrm{trace}(c\boldsymbol{A}) = c(\mathrm{trace}\ \boldsymbol{A})$

  推导

  $\mathrm{trace}(c\boldsymbol{A}) = \sum_{i=1}^{n} ca_{ii} = c\sum_{i=1}^{n} a_{ii} = c(\mathrm{trace}\ \boldsymbol{A})$

- $\mathrm{trace}(\boldsymbol{AB}) = \mathrm{trace}(\boldsymbol{BA})$

  推导

  $\mathrm{trace}(\boldsymbol{AB}) = \sum_{i=1}^{n}(\boldsymbol{AB})_{ii} = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij}b_{ji}\right)$

  $\mathrm{trace}(\boldsymbol{BA}) = \sum_{j=1}^{n}(\boldsymbol{BA})_{jj} = \sum_{j=1}^{n}\left(\sum_{i=1}^{m} b_{ji}a_{ij}\right)$

矩阵的秩

- **反映矩阵本质属性：**矩阵的秩是矩阵的一个重要数字特征，它是矩阵在初等变换下的不变量，能够反映矩阵的内在结构和本质特性。**无论对矩阵进行怎样的初等行变换或列变换，其秩始终保持不变**，这为研究矩阵的各种性质和分类提供了重要依据，是矩阵理论体系中的一个关键概念。

- **刻画向量组的线性相关性：**矩阵的秩与向量组的线性相关性紧密相关。矩阵的行秩等于列秩，都等于矩阵的秩。以列向量组为例，**矩阵的秩就是其列向量组的极大线性无关组所含向量的个数**。如果矩阵的秩等于列向量的个数，那么列向量组线性无关；如果矩阵的秩小于列向量的个数，那么列向量组线性相关。因此，通过计算矩阵的秩，可以方便地判断向量组的线性相关性，进而研究向量空间的结构和性质。

矩阵的秩

- **在计算机图形学中的应用：**在三维图形的变换和渲染中，常常需要对矩阵进行各种操作来实现图形的平移、旋转、缩放等变换。矩阵的秩可以用来判断变换矩阵是否是满秩的，即是否可逆。如果变换矩阵是满秩的，那么对应的变换是可逆的，这意味着图形可以通过逆变换恢复到原来的状态，保证了图形变换的准确性和可恢复性。

- **在信号处理中的应用：**在信号处理中，经常需要对信号进行采样和处理，将信号表示为矩阵的形式。矩阵的秩可以用来分析信号的特征和冗余度。例如，在压缩感知中，通过分析信号矩阵的秩，可以找到信号的稀疏表示，从而实现对信号的高效压缩和恢复。

- **在数据分析与机器学习中的应用：**在数据分析中，数据常常以矩阵的形式存储和处理。矩阵的秩可以用来判断数据的相关性和冗余性。如果数据矩阵的秩较低，说明数据中存在较多的线性相关性，存在冗余信息，可以通过降维等方法对数据进行处理，去除冗余，提高数据处理的效率和准确性。在机器学习的一些算法中，如主成分分析（PCA），也是基于对数据矩阵的秩的分析，来找到数据的主要特征和变化方向，实现数据的降维和特征提取。

矩阵的秩

- 前面我们讲到，向量系的线性相关和线性无关这些概念是非常重要的，它反映了向量系的一个本质特性。向量系线性相关意味着这个向量系中至少有一个向量可以由其他向量线性表示出来，一般地说这向量系中的数据有冗余。

- 一组m维n个向量组成的向量系$u_1, u_2, \ldots, u_n$线性无关，是指只有当n个系数$\alpha_1, \alpha_2, \ldots, \alpha_n$全为零时，其线性组合

$$\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n = 0$$

才成立。若向量系是线性无关的，意味着这个向量系中的任何一个向量不能由其他向量线性表示出来。

矩阵的秩

- 若由n个向量组成的向量系中，至少能确定由$r(r \le n)$个向量组成的向量系是线性无关的，而任何$r+1$个向量都是线性相关的，那就称这r个向量组成的向量系是极大线性无关组。对于给定的向量系，极大线性无关组并不唯一，但其所含向量的个数r必定是唯一的。

定义：设A是一个$m \times n$矩阵，如果该矩阵的列向量系的极大线性无关组所含向量的个数是$r\ (\le \min(m, n))$，则称矩阵A的秩($rank$)是r，记作$rank(A)=r$.

矩阵的秩

$$A = \begin{bmatrix} 2 & 4 & 1 & 3 \\ -1 & -2 & 1 & 0 \\ 0 & 0 & 2 & 2 \\ 3 & 6 & 2 & 5 \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} -2 & -4 & 4 \\ 2 & -8 & 0 \\ 8 & 4 & -12 \end{bmatrix}$$

上述两个矩阵的秩分别为多少？

矩阵的秩

$$A = \begin{bmatrix} 2 & 4 & 1 & 3 \\ -1 & -2 & 1 & 0 \\ 0 & 0 & 2 & 2 \\ 3 & 6 & 2 & 5 \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} -2 & -4 & 4 \\ 2 & -8 & 0 \\ 8 & 4 & -12 \end{bmatrix}$$

上述两个矩阵的秩分别为多少？

2，2

矩阵的秩

- 可以证明A的行向量系的极大线性无关组所含向量的个数也是r。所以对于给定的矩阵A它的秩是唯一的。

- 矩阵的秩是矩阵数值特征中的重要特征。设$A \in C^{m\times n}$，则A的秩$rank(A)$有一些基本性质：

(1) $rank(A)$是一个正整数；

(2) $rank(A) \le \ min(m,n)$，就是说$rank(A)$不大于A的行数与列数；

(3) $rank(A^H) = rank(A^T) = rank(A)$；

矩阵的秩

(4) 若$c \neq 0$, 则$rank(cA) = rank(A)$;

(5) $rank(AB) \leq \min\{rank(A), rank(B)\}$ （思考题）

(6) 对A左乘或者右乘一个非奇异矩阵后，其秩保存不变。

即设$B \in C^{m\times m}$, $C \in C^{n\times n}$都为非奇异矩阵，则

$$rank(BA) = rank(AC) = rank(BAC) = rank(A);$$

(7) $rank(A^T A) = rank(AA^T) = rank(A)$;

$$rank(A^H A) = rank(AA^H) = rank(A).$$

矩阵的秩

(6) 对A左乘或者右乘一个非奇异矩阵后，其秩保存不变。

即设$B \in C^{m\times m}, C \in C^{n\times n}$都为非奇异矩阵，则

$$rank(BA) = rank(AC) = rank(BAC) = rank(A);$$

证明：根据性质5，$rank(AB) \le \min\{rank(A), rank(B)\}$

可知 $rank(BA) \le \min\{rank(A), rank(B)\} \le rank(A)$

此外 $rank(A) = rank(B$^-1$BA) \le rank(BA)$

从而 $rank(A) = rank(BA)$

矩阵的秩

- 根据矩阵秩的大小又可以将矩阵分成以下4类：

(1) 满秩(full rank)矩阵，秩为n的$n \times n$方阵，满秩矩阵就是非奇异矩阵；

(2) 秩亏缺(rank deficient)矩阵，$rank(A_{m\times n}) < min(m,n)$；

(3) 行满秩(full row rank)矩阵，$rank(A_{m\times n}) = m$, 行满秩矩阵一定是方阵或“宽”矩阵($m \le n$)；

(4) 列满秩(full column rank)矩阵，$rank(A_{m\times n}) = n$, 列满秩矩阵一定是方阵或“高”矩阵($m \ge n$).

矩阵的秩

- 根据矩阵$A \in R^{m \times n}$的秩的大小，线性方程组可以分为下面

三种类型：

(1) 适定线性方程组 若m=n且$rank(A) = n$, 即A为非奇异矩阵，则称$Ax = b$为适定方程组(well-determined systems)，适定方程组存在唯一解；

(2) 亚定线性方程组 若$m < n$，则称$Ax = b$为亚定线性方程组(under-determined systems), 这时如果$rank(A) = m$, 则方程组$Ax = b$一定有解，且有无穷多组解。

矩阵的秩

(3) 超定线性方程组 若$m>n$, 则称$Ax = b$为超定方程组(over-determined systems)。

超定方程组有可能有解、有可能无解。

- 如果线性方程组有解，则称方程组是相容线性方程组(consistent systems)，相容性条件是$rank(A) = rank(A{:}\,b)$, 也就是增广矩阵$(A{:}\,b)$的秩等于矩阵A的秩。

6. 内积与范数

矩阵内积

- **度量向量空间中向量的关系**
  - **角度与相似性**：在向量空间中，矩阵内积可用于计算向量之间的夹角。通过内积公式，能得到两向量夹角的余弦值，进而衡量向量间的相似程度。在文本分类等自然语言处理任务中，可将文本表示为向量，通过计算向量内积来判断文本的相似性，若内积越大，夹角越小，文本内容就越相似。
  - **正交性判断**：当两个向量的内积为 0 时，它们相互正交。在信号处理中，例如将复杂信号分解为相互正交的基信号之和，利用向量的正交性，能使信号分解和处理更高效，像傅里叶变换就是基于三角函数系的正交性。
- **定义空间中的几何结构**
  - **诱导范数**：矩阵内积可以诱导出范数，为向量空间赋予了度量结构，用于衡量向量的长度或大小。
  - **构建希尔伯特空间**：内积空间是构建希尔伯特空间的基础，希尔伯特空间在泛函分析、量子力学等领域有广泛应用，为研究函数空间、算子理论等提供了有力工具。
- **优化与最小二乘法**
  - **在曲线拟合中**：对于给定的一组数据点，要找到一条最佳拟合曲线，可转化为求解一个线性方程组的最小二乘问题。通过定义误差向量的内积，构建目标函数，利用内积的性质来求解使误差最小的参数，从而得到最优拟合曲线。

6. 内积与范数

矩阵范数

- **衡量矩阵的大小和“能量”**
  - **数值大小度量**：矩阵范数提供了一种衡量矩阵“大小”的方式，能帮助判断矩阵在数值上的规模。在数值计算中，当评估矩阵运算结果的大小时，范数可直观反映矩阵元素的综合规模，如在迭代算法中，通过监测矩阵范数的变化来判断算法是否收敛。
  - **能量刻画**：在某些物理和工程问题中，矩阵范数可用于刻画系统的“能量”。如在控制系统中，系统的能量可与相关矩阵的范数建立联系，通过分析矩阵范数来评估系统的能量状态和稳定性。
- **分析矩阵的性质和行为**
  - **可逆性判断**：对于方阵A，若$||A|| < 1$，则 I-A 可逆，其中 I 为单位矩阵。利用矩阵范数的这一性质，可在一定程度上判断矩阵的可逆性，为求解线性方程组等问题提供理论支持。
  - **特征值估计**：矩阵范数与矩阵的特征值存在一定关系，如谱范数等于矩阵的最大奇异值，而奇异值与特征值紧密相关。通过矩阵范数可以对特征值的范围进行估计，为研究矩阵的特征值分布提供了一种途径。
- **评估数值算法的稳定性和误差**
  - **算法稳定性分析**：在数值算法中，如矩阵的求逆、矩阵分解等算法，矩阵范数用于分析算法的稳定性。若算法在计算过程中，矩阵范数的增长是有界的，则算法相对稳定；反之，可能导致计算结果误差过大甚至算法失效。
  - **误差估计**：在数值计算中，通过矩阵范数可以对计算结果的误差进行估计。如在求解线性方程组时，若存在误差，可利用矩阵范数来估计解的误差范围，为精度控制提供依据。

6. 内积与范数

1. 向量的内积与范数

- 两个实n维向量$x=(x_1,x_2,\ldots,x_n)^T, y=(y_1,y_2,\ldots,y_n)^T$之间的内积(inner product)记为$(x,y)$。定义为

$$(x,y)=x_1y_1+x_2y_2+\cdots+x_ny_n=\sum_{i=1}^{n}x_iy_i=x^Ty$$

- 两个n维复向量$x=(x_1,x_2,\ldots,x_n)^T, y=(y_1,y_2,\ldots,y_n)^T$之间的内积也记为$(x,y)$, 定义为

$$(x,y)=x_1^*y_1+x_2^*y_2+\cdots+x_n^*y_n=\sum_{i=1}^{n}x_i^*y_i=x^Hy$$

6. 内积与范数

- 一个关于向量的函数 $\rho(x)$，如其为x的度量，则必须满足以下三个条件:
  - $\rho(x) \geq 0$ 且 $\rho(x) = 0$ 当且仅当 $x = 0$
  - $\rho(ax) = |a|\rho(x), a$ 为任意常数
  - 对于任意 $\boldsymbol{x}$ 和 $\boldsymbol{y}$, 以下不等式成立

$$\rho(x + y) \leq \rho(x) + \rho(y)$$

6. 内积与范数

- 若两个实向量之间定义了内积，我们就可以利用这个内积，来定义两个实向量夹角的余弦：

$$cos\theta = \frac{\langle x, y\rangle}{\sqrt{\langle x, x\rangle}\sqrt{\langle y, y\rangle}}$$

- 定义了内积的线性空间就是内积空间，实的内积空间称为欧几里德空间(Euclidean)。有了内积的定义以后，我们还要用到向量的长度(length)、距离(distance)和邻域(neighborhood)等测度。

6. 内积与范数

- 向量$x=(x_1,x_2,\dots,x_n)^T\in C^n$的最常用的范数是*Euclidean*范数，或者称$L2$-范数(norm), 定义为

$$\|x\|_2=\sqrt{|x_1|^2+|x_2|^2+\cdots+|x_n|^2}=\sqrt{\langle x,x\rangle}$$

这就是常用的向量x长度的度量，用内积来表示长度，$|x_i|$表示复数x_i的模，若x_i是实数，则$|x_i|$表示实数x_i的绝对值。有这个长度的定义，前面定义实向量x,y的夹角就可以表示成

$$cos\theta=\frac{\langle x,y\rangle}{\|x\|_2\|y\|_2}$$

6. 内积与范数

- $L2$-范数还可以度量两个向量之间的距离

$$d(x, y) = \|x - y\|_2 = \sqrt{|x_1 - y_1|^2 + |x_2 - y_2|^2 + \cdots + |x_n - y_n|^2}$$

以及一个向量的ε-邻域($\varepsilon > 0$)

$$N_\varepsilon(x) = \{y \mid \|y - x\|_2 \le \varepsilon\}$$

邻域的概念是最优化理论和方法中的一个重要概念，因为当我们讨论一种优化算法的性能时，局部最优（在某个邻域最优）比全局最优（在整个定义域最优）更方便进行比较和确定。

6. 内积与范数

- 对于向量$x=(x_1,x_2,\dots,x_n)^T$而言，综合起来有下列常用的范数：

(1) L_0范数(也称0- “范数”)

$$\|x\|_0 \triangleq \text{非零元素的个数}$$

(2) L_1范数(也称和范数或1-范数)

$$\|x\|_1 \triangleq \sum_{i=1}^{n}|x_i| = |x_1|+|x_2|+\cdots+|x_n|$$

(3) L_2范数(常称*Euclidean*范数或2-范数)

$$\|x\|_2 \triangleq \sqrt{\sum_{i=1}^{n}|x_i|^2} = (|x_1|^2+|x_2|^2+\cdots+|x_n|^2)^{\frac{1}{2}}$$

6. 内积与范数

(4) L_∞范数(也称无穷范数)

$$\|x\|_\infty \triangleq \max\{|x_1|, |x_2|, \ldots, |x_n|\}$$

(5) L_p范数(也称*Holder*范数)

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

6. 内积与范数

关于 p- 范数的一个经典结论是 **Hölder** 不等式:

$$|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}| \leqslant \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q, \quad \frac{1}{p}+\frac{1}{q}=1. \tag{2.2.2}$$

一个非常重要的特殊情形是 **Cauchy-Schwarz** 不等式:

$$|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}| \leqslant \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2. \tag{2.2.3}$$

$\mathbb{R}^n$ 上的所有范数都是等价的，也就是说，若 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是 $\mathbb{R}^n$ 上的两个范数，则存在正常数 c_1 和 c_2 使得

$$c_1\|\boldsymbol{x}\|_\alpha \leqslant \|\boldsymbol{x}\|_\beta \leqslant c_2\|\boldsymbol{x}\|_\alpha \tag{2.2.4}$$

对于所有的 $\boldsymbol{x}\in\mathbb{R}^n$ 都成立. 例如，如果 $\boldsymbol{x}\in\mathbb{R}^n$，那么我们有

$$\|\boldsymbol{x}\|_2 \leqslant \|\boldsymbol{x}\|_1 \leqslant \sqrt{n}\|\boldsymbol{x}\|_2, \tag{2.2.5}$$

$$\|\boldsymbol{x}\|_\infty \leqslant \|\boldsymbol{x}\|_2 \leqslant \sqrt{n}\|\boldsymbol{x}\|_\infty, \tag{2.2.6}$$

$$\|\boldsymbol{x}\|_\infty \leqslant \|\boldsymbol{x}\|_1 \leqslant n\|\boldsymbol{x}\|_\infty. \tag{2.2.7}$$

6. 内积与范数

例6.2 设$x = (1, 2, -3)^T$, 试求$\|x\|_0, \|x\|_1, \|x\|_2, \|x\|_\infty, \|x\|_p$.

6. 内积与范数

例6.2 设$x = (1, 2, -3)^T$, 试求$\|x\|_0, \|x\|_1, \|x\|_2, \|x\|_\infty, \|x\|_p$.

解： $\|x\|_0 = 3$;

$$\|x\|_1 = |1| + |2| + |-3| = 6;$$

$$\|x\|_2 = (|1|^2 + |2|^2 + |-3|^2)^{\frac{1}{2}} = \sqrt{14};$$

$$\|x\|_\infty = \max\{|1|, |2|, |-3|\} = 3;$$

$$\|x\|_p = (|1|^p, |2|^p, |-3|^p)^{\frac{1}{p}}, \quad p > 1$$

6. 内积与范数

2.函数空间的内积与范数

- 若$x(t)$和$y(t)$都是闭区间$[a, b]$上的连续函数，则它们的内积：

$$\langle x(t), y(t)\rangle \triangleq \int_a^b x(t)y(t)dt$$

变量t可以是时间变量、频率变量或者空间变量。

函数$x(t)$的范数$\|x(t)\|$定义为：$\|x(t)\| \triangleq \sqrt{\int_a^b x(t)^2 dt}$

函数$x(t)$和$y(t)$的夹角的余弦是：

$$cos\theta \triangleq \frac{\langle x, y\rangle}{\sqrt{\langle x, x\rangle}\sqrt{\langle y, y\rangle}} = \frac{\int_a^b x(t)y(t)dt}{\|x(t)\| \cdot \|y(t)\|}$$

6. 内积与范数

有了内积的定义，我们就可以引入矩阵论中最重要的一个概念，就是正交性的概念。

- 如果两个非零向量x, y的内积等于零，也就是$\langle x, y\rangle = 0$，我们就称向量x与y是正交的。
- 正交性的概念可以在不同的层面上来解释：

(1) 数学定义，表明这两个向量的内积等于零；

(2) 几何解释，也称为是垂直的，表明两个向量的夹角是90°，并且一个向量到另一个向量的投影为零；

(3) 物理意义，两个向量正交意味着一个向量不含另一个向量的任何成分，不存在相互作用和干扰。

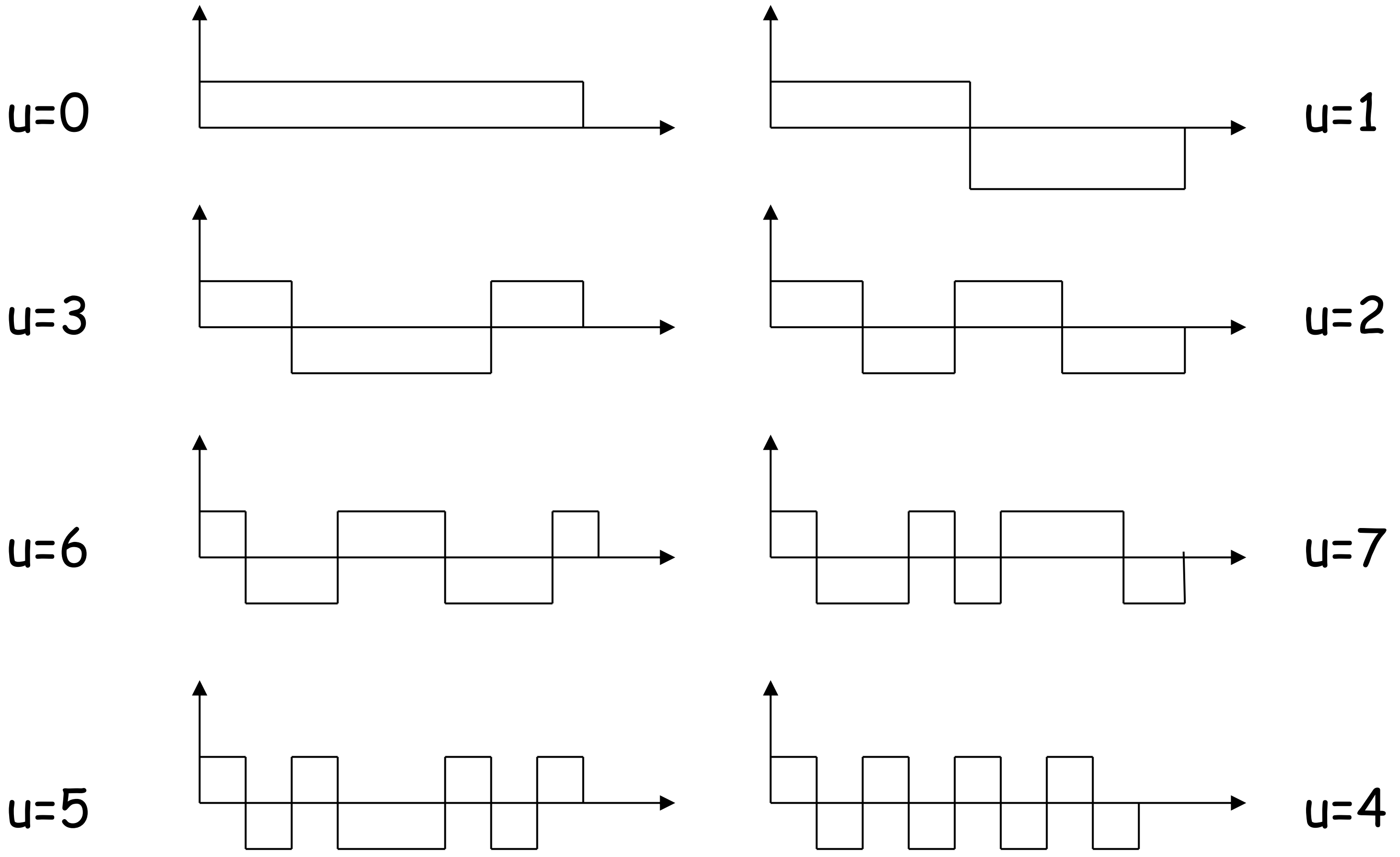
6. 内积与范数

给定一组n维向量系$x_1, x_2, \dots, x_m (m \le n)$, 如果两两正交满足：

$$\langle x_i, x_j \rangle = \begin{cases} \alpha_i > 0 & i = j \\ 0 & i \ne j \end{cases} \quad i, j = 1,2, \dots, m$$

则称向量$x_1, x_2, \dots, x_m (m \le n)$是一组正交向量系，若还有$\|x_i\| = 1,\ i = 1,2, \dots, m$, 则称该向量系是一组标准正交向量系。

正交基



矩阵范数

- A ***matrix norm*** is a function $\|*\|$ from the set of complex matrices into R that satisfies the following properties:

1. $\|A\| \geq 0$ and $\|A\| = 0 \iff A = 0.$
2. $\|aA\| = |a| \cdot \|A\|$ for all scalars a.
3. $\|A + B\| \leq \|A\| + \|B\|$ for matrices of the same size.
4. $\|A \cdot B\| \leq \|A\| \cdot \|B\|.$

矩阵范数

矩阵范数

将向量范数的概念推广到矩阵，作为矩阵大小的度量。

定义6.2 设$\|\cdot\|$是一种向量范数

$$\|\boldsymbol{A}\|_p = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^p\right)^{1/p}$$

$$\| \boldsymbol{A} \|_p = \sup_{\boldsymbol{x}\neq 0} \left\| \boldsymbol{A}\left(\frac{\boldsymbol{x}}{\| \boldsymbol{x} \|_p}\right)\right\|_p = \max_{\|\boldsymbol{x}\|_p=1} \| \boldsymbol{A}\boldsymbol{x} \|_p .$$

称之为由向量范数派生的矩阵算子范数 (induced norm)

矩阵范数

以下是三种常用的 p 范数:

(1) L_1 范数 (和范数) ($p=1$)

$$\|\boldsymbol{A}\|_1 \stackrel{\text{def}}{=} \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|$$

(2) Frobenius 范数 ($p=2$)

$$\|\boldsymbol{A}\|_{\mathrm{F}} \stackrel{\text{def}}{=} \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2}$$

(3) 最大范数 (max norm) 即 $p=\infty$ 的 p 范数，定义为

$$\|\boldsymbol{A}\|_\infty = \max_{i=1,2,\cdots,m;\, j=1,2,\cdots,n}\{|a_{ij}|\}$$

矩阵范数

2.3.3 矩阵 2-范数

矩阵 1-范数和矩阵 ∞-范数的良好特性是它们容易计算，计算量是 $O(n^2)$ 的，见 (2.3.9) 和 (2.3.10). 2-范数的计算要复杂得多.

定理 2.3.1 *假设 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，则存在一个单位 2-范数 n 维向量 $\boldsymbol{z}$ 使得 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{z} = \mu^2\boldsymbol{z}$，其中 $\mu = \| \boldsymbol{A} \|_2$.*

证明 设 $\boldsymbol{z} \in \mathbb{R}^n$ 是满足 $\| \boldsymbol{A}\boldsymbol{z} \|_2 = \| \boldsymbol{A} \|_2$ 的单位向量. 由于在 $\boldsymbol{z}$ 处使函数

$$g(\boldsymbol{x}) = \frac{1}{2}\frac{\| \boldsymbol{A}\boldsymbol{x} \|_2^2}{\| \boldsymbol{x} \|_2^2} = \frac{1}{2}\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}$$

取到最大值，所以它满足 $\nabla g(\boldsymbol{z}) = 0$，其中 ∇g 是 g 的梯度. 冗长的微分计算表明，对于 $i = 1:n$ 有

$$\frac{\partial g(\boldsymbol{z})}{\partial z_i} = \left[(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}) \sum_{j=1}^{n} [\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}]_{ij} z_j - (\boldsymbol{z}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{z}) z_i \right] \Big/ (\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z})^2 .$$

用向量符号可表示为 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{z} = (\boldsymbol{z}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{z})\boldsymbol{z}$. 令 $\mu = \| \boldsymbol{A}\boldsymbol{z} \|_2$ 可知定理成立. $\square$

矩阵范数

学习矩阵范数界定关系具有多方面的重要意义，主要体现在以下几个领域：

理论分析

- **建立统一框架**：矩阵范数有多种定义方式，如算子范数、元素范数等，不同范数适用于不同的场景和问题。通过研究它们之间的界定关系，能够在各种范数之间建立起联系，形成一个更统一、完整的理论框架，有助于更全面地理解矩阵的性质和相关数学结构。
- **深化对矩阵性质理解**：不同的矩阵范数从不同角度刻画了矩阵的“大小”或“能量”等特性。例如，ℓ_1范数突出了矩阵元素绝对值之和的特性，ℓ_2范数与矩阵的能量和欧几里得结构相关，而ℓ_∞范数关注的是矩阵元素绝对值的最大值。了解它们之间的界定关系，可以更深入地理解矩阵在不同度量下的性质以及这些性质之间的内在联系，为进一步研究矩阵理论提供基础。
- **推导矩阵相关定理**：在矩阵理论的发展中，许多重要的定理和结论都依赖于对矩阵范数的分析。矩阵范数之间的界定关系常常作为关键工具，用于推导和证明其他重要的矩阵性质和定理，如矩阵的可逆性条件、特征值的分布范围等。

算法设计与优化

- **算法收敛性分析**：在设计和分析迭代算法（如求解线性方程组的迭代法、矩阵分解算法等）时，需要判断算法是否收敛。矩阵范数之间的界定关系可以帮助建立算法的收敛条件，通过在不同范数下分析迭代序列的性质，利用范数的界定关系来推导算法在更一般情况下的收敛性，从而为算法的有效性提供理论保证。
- **优化算法性能**：在优化问题中，目标函数和约束条件可能涉及矩阵范数。了解矩阵范数之间的界定关系，可以在不同范数之间进行灵活转换，选择更便于计算和分析的范数形式，从而优化算法的计算复杂度和性能。例如，在一些基于范数的正则化方法中，通过合理利用范数界定关系，可以选择合适的范数来平衡模型的拟合能力和复杂度，提高模型的泛化性能。

矩阵范数

学习矩阵范数界定关系具有多方面的重要意义，主要体现在以下几个领域：

误差估计

- **多范数误差评估**：在数值计算中，由于各种原因（如数据测量误差、计算截断误差等），会产生误差。不同的问题可能在不同的范数下有更直观或更易于计算的误差估计。学习矩阵范数之间的界定关系，可以在已知一种范数下的误差估计时，推导出其他范数下的误差范围，从而更全面、准确地评估计算结果的误差。
- **控制误差传播**：在复杂的数值计算过程中，误差可能会在不同的计算步骤和矩阵操作中传播。通过矩阵范数之间的界定关系，可以分析误差在不同范数下的传播规律，进而采取相应的措施来控制误差的增长，提高计算结果的可靠性。

数值稳定性分析

- **评估算法稳定性**：数值稳定性是衡量算法在数值计算中抵抗误差干扰能力的重要指标。矩阵范数之间的界定关系可以用于分析算法在不同范数下对误差的敏感性，从而更全面地评估算法的数值稳定性。如果一种算法在某种范数下对误差的放大作用较小，但在另一种范数下可能较大，通过范数界定关系可以揭示这种潜在的不稳定性，为改进算法或选择更合适的算法提供依据。
- **选择合适范数**：在实际计算中，根据具体问题的特点和对数值稳定性的要求，选择合适的矩阵范数至关重要。矩阵范数之间的界定关系有助于在不同范数之间进行权衡和选择，使得在保证计算精度的前提下，尽可能提高算法的数值稳定性和计算效率。

矩阵范数

该定理表明 $\| \boldsymbol{A} \|_2^2$ 是多项式 $p(\lambda) = \det(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} - \lambda \boldsymbol{I})$ 的一个零点. 具体来说,

$$\| \boldsymbol{A} \|_2 = \sqrt{\lambda_{\max}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})}.$$

我们将在第 7 章和第 8 章进一步讨论特征值. 现在我们只须注意计算矩阵 2-范数是需要迭代的, 而且比计算 1-范数和 ∞-范数更复杂. 幸运的是, 如果只是要得到 $\| \boldsymbol{A} \|_2$ 的数量级估计, 那么可以利用 (2.3.7), (2.3.8), (2.3.11), (2.3.12).

作为范数分析的另一个例子, 下面是估计 2-范数的一个方便结论.

推论 2.3.2 假设 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, 则 $\| \boldsymbol{A} \|_2 \leqslant \sqrt{\| \boldsymbol{A} \|_1 \| \boldsymbol{A} \|_\infty}$.

证明 若 $\boldsymbol{z} \neq \boldsymbol{0}$ 使得 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{z} = \mu^2 \boldsymbol{z}$, 其中 $\mu = \| \boldsymbol{A} \|_2$, 则 $\mu^2 \| \boldsymbol{z} \|_1 = \| \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{z} \|_1 \leqslant \| \boldsymbol{A}^{\mathrm{T}} \|_1 \| \boldsymbol{A} \|_1 \| \boldsymbol{z} \|_1 = \| \boldsymbol{A} \|_\infty \| \boldsymbol{A} \|_1 \| \boldsymbol{z} \|_1$. □

矩阵范数

F-范数和 p-范数（特别是 $p=1,2,\infty$ 的情形）满足一些不等式，在矩阵计算的分析中经常使用这些不等式. 假设 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$，我们有

$$\|\boldsymbol{A}\|_2 \leqslant \|\boldsymbol{A}\|_F \leqslant \sqrt{\min\{m,n\}}\,\|\boldsymbol{A}\|_2, \tag{2.3.7}$$

$$\max|a_{ij}| \leqslant \|\boldsymbol{A}\|_2 \leqslant \sqrt{mn}\max|a_{ij}|, \tag{2.3.8}$$

$$\|\boldsymbol{A}\|_1 = \max_{1\leqslant j\leqslant n}\sum_{i=1}^{m}|a_{ij}|, \tag{2.3.9}$$

$$\|\boldsymbol{A}\|_\infty = \max_{1\leqslant i\leqslant m}\sum_{j=1}^{n}|a_{ij}|, \tag{2.3.10}$$

$$\frac{1}{\sqrt{n}}\|\boldsymbol{A}\|_\infty \leqslant \|\boldsymbol{A}\|_2 \leqslant \sqrt{m}\,\|\boldsymbol{A}\|_\infty, \tag{2.3.11}$$

$$\frac{1}{\sqrt{m}}\|\boldsymbol{A}\|_1 \leqslant \|\boldsymbol{A}\|_2 \leqslant \sqrt{n}\,\|\boldsymbol{A}\|_1. \tag{2.3.12}$$

假设 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $1\leqslant i_1\leqslant i_2\leqslant m$, $1\leqslant j_1\leqslant j_2\leqslant n$，则

$$\|\boldsymbol{A}(i_1:i_2, j_1:j_2)\|_p \leqslant \|\boldsymbol{A}\|_p. \tag{2.3.13}$$

以上关系式的证明均留作练习.

矩阵范数

例6.4 求矩阵A的范数$\|A\|_1$，　$\|A\|_\infty$，　$\|A\|_F$

$$A=\begin{array}{cc}\begin{pmatrix}1&2&0\\-1&2&-1\\0&1&1\end{pmatrix}&\begin{matrix}3\\4\\2\end{matrix}\\\begin{matrix}2&5&2\end{matrix}&\end{array}$$

矩阵范数

例6.4 求矩阵A的范数$\|A\|_1$， $\|A\|_\infty$， $\|A\|_F$

$$A=\begin{array}{cc}\begin{pmatrix}1&2&0\\-1&2&-1\\0&1&1\end{pmatrix} & \begin{array}{c}3\\4\\2\end{array}\\ \begin{array}{ccc}2&5&2\end{array} & \end{array}$$

解：

$$\|A\|_1=\sum_{j=1}^{n}\sum_{i=1}^{m}\left|a_{ij}\right| \ =9;$$

$$\|A\|_\infty= \ \max_{1\le i\le m}\max_{1\le j\le n}\left|a_{ij}\right|=2;$$

$$\|A\|_F=\sqrt{4+4+1+1+1+1+1}=\sqrt{13}$$

矩阵范数

以下是矩阵的内积与范数之间的关系[50]：

(1) Cauchy-Schwartz 不等式

$$|\langle \boldsymbol{A}, \boldsymbol{B} \rangle|^2 \leqslant \|\boldsymbol{A}\|^2 \|\boldsymbol{B}\|^2$$

等号成立，当且仅当 $\boldsymbol{A} = c\boldsymbol{B}$，其中，$c$ 是某个复常数。

(2) Pathagoras 定理：$\langle \boldsymbol{A}, \boldsymbol{B} \rangle = 0 \;\Rightarrow\; \|\boldsymbol{A} + \boldsymbol{B}\|^2 = \|\boldsymbol{A}\|^2 + \|\boldsymbol{B}\|^2$。

对于 $\boldsymbol{A} \in \mathbb{C}^{m\times n}, \boldsymbol{x} \in \mathbb{C}^n$，有

$$\|\boldsymbol{A}\boldsymbol{x}\|_{\mathrm{F}} \leqslant \|\boldsymbol{A}\|_{\mathrm{F}} \|\boldsymbol{x}\|_{\mathrm{F}}, \text{ 即 } \|\boldsymbol{A}\boldsymbol{x}\|_2 \leqslant \|\boldsymbol{A}\|_2 \|\boldsymbol{x}\|_2.$$

矩阵范数

学习矩阵扰动具有多方面的重要意义，主要体现在以下几个领域：

数值分析

- **评估算法稳定性**：在数值算法中，如矩阵求逆、求解线性方程组等，输入矩阵的微小扰动可能会导致计算结果产生较大变化。通过研究矩阵扰动，可以分析算法对于扰动的敏感性，评估算法的稳定性。例如，对于一个给定的线性方程组$Ax = b$，若系数矩阵A存在扰动ΔA，解的变化情况能帮助判断求解算法在面对实际问题中数据不确定性时的可靠性。
- **误差分析**：在实际计算中，由于计算机的有限精度等原因，矩阵元素不可避免地会存在一定误差，这些误差可视为矩阵的扰动。研究矩阵扰动有助于量化这些误差对最终计算结果的影响，从而为数值计算提供误差估计，确定计算结果的精度范围。

线性系统理论

- **系统稳定性分析**：在控制理论中，线性时不变系统的稳定性常通过状态矩阵的特征值来判断。矩阵的微小扰动可能会使特征值发生变化，进而影响系统的稳定性。通过研究矩阵扰动对特征值的影响，可以分析系统在参数摄动等情况下的稳定性鲁棒性，即系统抵抗参数变化而保持稳定的能力。
- **系统性能评估**：对于线性系统，矩阵扰动可以用来研究系统参数变化对系统性能指标的影响，如系统的响应速度、稳态误差等。这有助于在系统设计阶段，考虑参数不确定性对系统性能的潜在影响，从而设计出更可靠、性能更优的控制系统。

矩阵范数

2.3.4 扰动与逆矩阵

我们经常使用范数来量化扰动的影响，或者证明矩阵序列收敛到某个特定极限. 作为这些范数应用的例证，我们把 $\boldsymbol{A}^{-1}$ 的变化表示为 $\boldsymbol{A}$ 的变化的函数.

引理 2.3.3 假设 $\boldsymbol{F} \in \mathbb{R}^{n\times n}$ 且 $\|\boldsymbol{F}\|_p < 1$，则 $\boldsymbol{I}-\boldsymbol{F}$ 非奇异，

$$(\boldsymbol{I}-\boldsymbol{F})^{-1} = \sum_{k=0}^{\infty} \boldsymbol{F}^k$$

且

$$\|(\boldsymbol{I}-\boldsymbol{F})^{-1}\|_p \leqslant \frac{1}{1-\|\boldsymbol{F}\|_p}.$$

证明 假设 $\boldsymbol{I}-\boldsymbol{F}$ 是奇异的，则存在某个非零向量 $\boldsymbol{x}$ 使得 $(\boldsymbol{I}-\boldsymbol{F})\boldsymbol{x}=\boldsymbol{0}$. 这样，由 $\|\boldsymbol{x}\|_p = \|\boldsymbol{F}\boldsymbol{x}\|_p$ 推出 $\|\boldsymbol{F}\|_p \geqslant 1$，产生了矛盾. 所以 $\boldsymbol{I}-\boldsymbol{F}$ 非奇异. 为了获得其逆矩阵的表达式，考虑恒等式

$$\left(\sum_{k=0}^{N} \boldsymbol{F}^k\right)(\boldsymbol{I}-\boldsymbol{F}) = \boldsymbol{I}-\boldsymbol{F}^{N+1}.$$

矩阵范数

由于 $\| F \|_p < 1$，所以由 $\| F^k \|_p \leqslant \| F \|_p^k$ 可得 $\lim_{k\to\infty} F^k = 0$. 于是

$$\left(\lim_{N\to\infty}\sum_{k=0}^{N} F^k\right)(I - F) = I.$$

从而我们有

$$(I - F)^{-1} = \lim_{N\to\infty}\sum_{k=0}^{N} F^k = \sum_{k=0}^{\infty} F^k.$$

由此不难证明

$$\| (I - F)^{-1} \|_p \leqslant \sum_{k=0}^{\infty} \| F \|_p^k = \frac{1}{1 - \| F \|_p}.$$

这就完成了引理的证明. □

定理 2.3.4 假设 A 非奇异，$r \equiv \| A^{-1}E \|_p < 1$，则 $A + E$ 非奇异，

$$\| (A + E)^{-1} - A^{-1} \|_p \leqslant \frac{\| E \|_p \| A^{-1} \|_p^2}{1 - r}.$$

代数与矩阵基础

- 向量，矩阵
- 线性空间，线性子空间
  - 零空间，像空间
- 线性相关，线性无关
  - 空间的维数与基底
- 矩阵的基本运算
  - 转置，共轭转置
  - 相乘，求逆
- 矩阵的数值特征
  - 行列式，二次型，特征值，trace，rank
- 内积与范数
  - 向量范数
  - 矩阵范数