一来，N 个数据样本 $x(n), n=0,1,\cdots,N-1$ 的离散 Fourier 变换 (DFT) 就可以用向量的典范内积表示为

$$X(f)=\sum_{n=0}^{N-1} x(n)\mathrm{e}^{-\mathrm{j}(\frac{2\pi}{N})nf}=\boldsymbol{e}_{N-1}^{\mathrm{H}}\boldsymbol{x}=\langle \boldsymbol{e}_{N-1},\boldsymbol{x}\rangle$$

其中，$\boldsymbol{x}=[x(0),x(1),\cdots,x(N-1)]^{\mathrm{T}}$ 常称为数据向量。

下面是内积空间的范数具备的一般性质。

定理 1.4.1[40] 在实或复内积空间里，范数具有以下性质：

(1) $\|\mathbf{0}\|=0$，并且 $\|\boldsymbol{x}\|>0,\ \forall\ \boldsymbol{x}\neq\mathbf{0}$；

(2) $\|c\boldsymbol{x}\|=|c|\cdot\|\boldsymbol{x}\|$ 对所有向量 $\boldsymbol{x}$ 和标量 c 成立；

(3) 范数服从极化恒等式 (polarization identity)

$$\langle\boldsymbol{x},\boldsymbol{y}\rangle=\frac{1}{4}\left(\|\boldsymbol{x}+\boldsymbol{y}\|^2-\|\boldsymbol{x}-\boldsymbol{y}\|^2\right),\quad \forall\ \boldsymbol{x},\boldsymbol{y}\quad (\text{实内积空间}) \tag{1.4.4}$$

$$\langle\boldsymbol{x},\boldsymbol{y}\rangle=\frac{1}{4}\left(\|\boldsymbol{x}+\boldsymbol{y}\|^2-\|\boldsymbol{x}-\boldsymbol{y}\|^2-\mathrm{j}\,\|\boldsymbol{x}+\mathrm{j}\,\boldsymbol{y}\|^2+\mathrm{j}\|\boldsymbol{x}-\mathrm{j}\,\boldsymbol{y}\|^2\right)$$
$$\forall\ \boldsymbol{x},\boldsymbol{y}\quad (\text{复内积空间}) \tag{1.4.5}$$

(4) 范数满足平行四边形法则 (parallelogram law)

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2+\|\boldsymbol{x}-\boldsymbol{y}\|^2=2(\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2),\quad \forall\ \boldsymbol{x},\boldsymbol{y} \tag{1.4.6}$$

(5) 范数满足三角不等式 $\|\boldsymbol{x}+\boldsymbol{y}\|\leqslant\|\boldsymbol{x}\|+\|\boldsymbol{y}\|, \forall\ \boldsymbol{x},\boldsymbol{y}$；

(6) 范数服从 Cauchy-Schwartz 不等式

$$|\langle\boldsymbol{x},\boldsymbol{y}\rangle|\leqslant\|\boldsymbol{x}\|\cdot\|\boldsymbol{y}\| \tag{1.4.7}$$

等号 $|\langle\boldsymbol{x},\boldsymbol{y}\rangle|=\|\boldsymbol{x}\|\cdot\|\boldsymbol{y}\|$ 成立，当且仅当 $\boldsymbol{y}=c\boldsymbol{x}$，其中，$c$ 为某个非零常数。

下面分别介绍常数向量、函数向量和随机向量的内积与范数。

1. 常数向量的典范内积与范数

常数向量的内积通常采用典范内积，而常用的向量范数有以下几种。

(1) L_0 范数 (也称 0 范数)

$$\|\boldsymbol{x}\|_0 \stackrel{\text{def}}{=} \text{非零元素的个数} \tag{1.4.8}$$

(2) L_1 范数 (也称和范数或 1 范数)

$$\|\boldsymbol{x}\|_1 \stackrel{\text{def}}{=} \sum_{i=1}^{m}|x_i|=|x_1|+\cdots+|x_m| \tag{1.4.9}$$

(3) L_2 范数 (常称 Euclidean 范数，有时也称 Frobenius 范数)

$$\|\boldsymbol{x}\|_2=\left(|x_1|^2+\cdots+|x_m|^2\right)^{1/2} \tag{1.4.10}$$

(4) L_∞ 范数 (也称无穷范数或极大范数)

$$\|\boldsymbol{x}\|_\infty = \max\{|x_1|, \cdots, |x_m|\} \tag{1.4.11}$$

(5) L_p 范数 (也称 Hölder 范数 [294])

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^m |x_i|^p\right)^{1/p}, \qquad p \geqslant 1 \tag{1.4.12}$$

注释 1　L_0 范数不满足范数公理中的齐次性 $\|c\boldsymbol{x}\|_0 = |c|\,\|\boldsymbol{x}\|_0$，它只是一种虚拟的范数。然而，$L_0$ 范数在稀疏向量与稀疏表示中却起着关键的作用，详见 1.12 节。

注释 2　当 $p=2$ 时，L_p 范数与 Euclidean 范数完全等价。另外，无穷范数是 L_p 范数的极限形式，即有

$$\|\boldsymbol{x}\|_\infty = \lim_{p\to\infty}\left(\sum_{i=1}^m |x_i|^p\right)^{1/p} \tag{1.4.13}$$

范数 $\|\boldsymbol{x}\|$ 称为酉不变的，若 $\|\boldsymbol{U}\boldsymbol{x}\| = \|\boldsymbol{x}\|$ 对所有向量 $\boldsymbol{x} \in \mathbb{C}^m$ 和所有酉矩阵 $\boldsymbol{U} \in \mathbb{C}^{m\times m}$ 恒成立。

命题 1.4.1 [238]　Euclidean 范数 $\|\cdot\|_2$ 是酉不变的。

假定向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 有共同的起点 (即原点 O)，它们的端点分别为 x 和 y，则 $\|\boldsymbol{x}-\boldsymbol{y}\|_2$ 度量两个向量 $\boldsymbol{x}, \boldsymbol{y}$ 两端点 x, y 之间的标准 Euclidean 距离。特别地，非负的标量 $\langle \boldsymbol{x}, \boldsymbol{x}\rangle^{1/2}$ 称为向量 $\boldsymbol{x}$ 的 Euclidean 长度。Euclidean 长度为 1 的向量叫做归一化 (或标准化) 向量。对于任何不为零的向量 $\boldsymbol{x} \in \mathbb{C}^m$，向量 $\boldsymbol{x}/\langle \boldsymbol{x}, \boldsymbol{x}\rangle^{1/2}$ 都是归一化的，并且它与 $\boldsymbol{x}$ 同方向。

Euclidean 范数是应用最为广泛的向量范数定义。在本书后面的讨论中，如无特别声明，向量范数均指 Euclidean 范数。

利用向量的典范内积和 Euclidean 范数可以定义两个向量之间的夹角。

定义 1.4.1　两个向量之间的夹角定义为

$$\cos\theta \overset{\text{def}}{=} \frac{\langle \boldsymbol{x}, \boldsymbol{y}\rangle}{\sqrt{\langle \boldsymbol{x}, \boldsymbol{x}\rangle}\sqrt{\langle \boldsymbol{y}, \boldsymbol{y}\rangle}} = \frac{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}}{\|\boldsymbol{x}\|\cdot\|\boldsymbol{y}\|} \tag{1.4.14}$$

显然，若 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y} = 0$，则 $\theta = \pi/2$，此时，称常数向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 正交。因此，两个常数向量正交的数学定义如下。

定义 1.4.2　两个常数向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 称为正交，并记作 $\boldsymbol{x}\perp\boldsymbol{y}$，若它们的内积等于零，即 $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{x}^{\mathrm{H}}\boldsymbol{y} = 0$。

由定义知，零向量 $\mathbf{0}$ 与同一空间的任何向量都正交。

2. *函数向量的内积与范数*

若 $\boldsymbol{x}(t)$ 和 $\boldsymbol{y}(t)$ 分别是变量 t 的函数向量，则它们的内积定义为

$$\langle \boldsymbol{x}(t), \boldsymbol{y}(t)\rangle \overset{\text{def}}{=} \int_a^b \boldsymbol{x}^{\mathrm{H}}(t)\boldsymbol{y}(t)\mathrm{d}t \tag{1.4.15}$$

分别为 M 个已知模式向量 $\boldsymbol{s}_i$ 的样本均值向量和样本互协方差矩阵。于是，未知模式向量 $\boldsymbol{x}$ 到已知模式向量 $\boldsymbol{s}_i$ 之间的 Mahalanobis 距离定义为

$$D_{\mathrm{M}}(\boldsymbol{x},\boldsymbol{s}_i)=\sqrt{(\boldsymbol{x}-\boldsymbol{s}_i)^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{s}_i)} \tag{1.4.29}$$

根据近邻分类法，若

$$D_{\mathrm{M}}(\boldsymbol{x},\boldsymbol{s}_i)=\min_k D_{\mathrm{M}}(\boldsymbol{x},\boldsymbol{s}_k),\quad k=1,\cdots,M \tag{1.4.30}$$

则将未知模式向量 $\boldsymbol{x}$ 归为 $\boldsymbol{s}_i$ 所属的模式类型。

向量之间的相异度的测度不一定局限于距离函数。两个向量所夹锐角的余弦函数

$$D(\boldsymbol{x},\boldsymbol{s}_i)=\cos(\theta_i)=\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{s}_i}{\|\boldsymbol{x}\|_2\|\boldsymbol{s}_i\|_2} \tag{1.4.31}$$

也是相异度的一种有效测度。若 $\cos(\theta_i)<\cos(\theta_j),\forall j\neq i$ 成立，则认为未知模式向量 $\boldsymbol{x}$ 与样本模式向量 $\boldsymbol{s}_i$ 最相似。式 (1.4.31) 的变型

$$D(\boldsymbol{x},\boldsymbol{s}_i)=\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{s}_i}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{s}_i^{\mathrm{T}}\boldsymbol{s}_i+\boldsymbol{x}^{\mathrm{T}}\boldsymbol{s}_i} \tag{1.4.32}$$

称为 Tanimoto 测度[477]，它广泛应用于信息恢复、疾病分类、动物和植物分类等。

待分类的信号称为目标信号，分类通常是根据某种物理或几何概念进行的。令 X 为目标信号，A_i 代表第 i 类目标的分类概念。于是，采用目标–概念距离 (object-concept distance) $D(X,A_i)$ 描述与目标之间的相异度[457]，从而有类似于式 (1.4.22) 的关系

$$D(X,A_i)\leqslant D(X,A_j),\quad \forall i,j \tag{1.4.33}$$

因此，将目标信号 X 归为目标–概念距离 $D(X,A_i)$ 最小的第 i 类目标 C_i。

以上介绍了五种相异度：Euclidean 距离、Mahalanobis 距离、夹角余弦、Tanimoto 测度以及目标–概念距离。

1.4.3 矩阵的内积与范数

将向量的内积与范数加以推广，即可引出矩阵的内积与范数。

令 $m\times n$ 复矩阵 $\boldsymbol{A}=[\boldsymbol{a}_1,\cdots,\boldsymbol{a}_n]$ 和 $\boldsymbol{B}=[\boldsymbol{b}_1,\cdots,\boldsymbol{b}_n]$，将这两个矩阵分别“拉长”为 $mn\times 1$ 向量

$$\boldsymbol{a}=\mathrm{vec}(\boldsymbol{A})=\begin{bmatrix}\boldsymbol{a}_1\\ \vdots\\ \boldsymbol{a}_n\end{bmatrix},\quad \boldsymbol{b}=\mathrm{vec}(\boldsymbol{B})=\begin{bmatrix}\boldsymbol{b}_1\\ \vdots\\ \boldsymbol{b}_n\end{bmatrix}$$

$\mathrm{vec}(\boldsymbol{A})$ 称为矩阵 $\boldsymbol{A}$ 的 (列) 向量化。矩阵的向量化将在 1.11 节中作详细介绍。

矩阵的内积记作 $\langle\boldsymbol{A},\boldsymbol{B}\rangle:\mathbb{C}^{m\times n}\times\mathbb{C}^{m\times n}\to\mathbb{C}$，定义为两个“拉长向量” $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 之间的内积

$$\langle\boldsymbol{A},\boldsymbol{B}\rangle=\langle\mathrm{vec}(\boldsymbol{A}),\mathrm{vec}(\boldsymbol{B})\rangle=\sum_{i=1}^{n}\boldsymbol{a}_i^{\mathrm{H}}\boldsymbol{b}_i=\sum_{i=1}^{n}\langle\boldsymbol{a}_i,\boldsymbol{b}_i\rangle \tag{1.4.34}$$

或等价写作

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \mathrm{vec}(\boldsymbol{A})^{\mathrm{H}} \mathrm{vec}(\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}^{\mathrm{H}} \boldsymbol{B}) \tag{1.4.35}$$

式中 $\mathrm{tr}(\boldsymbol{C})$ 表示正方矩阵 $\boldsymbol{C}$ 的迹函数，定义为该矩阵对角元素之和。

令 $\mathbb{K}$ 表示一实数域或复数域，$\mathbb{K}^{m\times n}$ 表示 $m \times n$ 实数或复数矩阵的集合。

矩阵 $\boldsymbol{A} \in \mathbb{K}^{m\times n}$ 的范数记作 $\|\boldsymbol{A}\|$，它是矩阵 $\boldsymbol{A}$ 的实值函数，必须具有以下性质：

(1) 正值性：对于任何非零矩阵 $\boldsymbol{A} \neq \boldsymbol{O}$，其范数大于零，即 $\|\boldsymbol{A}\| > 0$ 若 $\boldsymbol{A} \neq \boldsymbol{O}$ (零矩阵)；并且 $\|\boldsymbol{A}\| = 0$ 当且仅当 $\boldsymbol{A} = \boldsymbol{O}$。

(2) 正比例性：对于任意 $c \in \mathbb{K}$，有 $\|c\boldsymbol{A}\| = |c| \cdot \|\boldsymbol{A}\|$。

(3) 三角不等式：$\|\boldsymbol{A} + \boldsymbol{B}\| \leqslant \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$。

(4) 两个矩阵乘积的范数小于或等于两个矩阵范数的乘积，即 $\|\boldsymbol{AB}\| \leqslant \|\boldsymbol{A}\| \cdot \|\boldsymbol{B}\|$。

例 1.4.2 考查 $n \times n$ 矩阵 $\boldsymbol{A}$ 的实值函数 $f(\boldsymbol{A}) = \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|$。容易验证：

(1) $f(\boldsymbol{A}) \geqslant 0$，并且当 $\boldsymbol{A} = 0$ 即 $a_{ij} \equiv 0$ 时 $f(\boldsymbol{A}) = 0$。

(2) $f(c\boldsymbol{A}) = \sum_{i=1}^{n}\sum_{j=1}^{n}|ca_{ij}| = |c|\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}| = |c|f(\boldsymbol{A})$。

(3) $f(\boldsymbol{A}+\boldsymbol{B}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(|a_{ij} + b_{ij}|\right) \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n}\left(|a_{ij}| + |b_{ij}|\right) = f(\boldsymbol{A}) + f(\boldsymbol{B})$。

(4) 对于两个矩阵的乘积，有

$$\begin{aligned} f(\boldsymbol{AB}) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left|\sum_{k=1}^{n} a_{ik}b_{kj}\right| \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}|a_{ik}||b_{kj}| \\ &\leqslant \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{k=1}^{n}|a_{ik}|\sum_{l=1}^{n}|b_{kl}|\right) = f(\boldsymbol{A})f(\boldsymbol{B}) \end{aligned}$$

因此，实函数 $f(\boldsymbol{A}) = \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|$ 是一种矩阵范数。

矩阵的范数有三种主要类型：诱导范数、元素形式范数和 Schatten 范数。

1. 诱导范数 (induced norm)

诱导范数又称 $m \times n$ 矩阵空间上的算子范数 (operator norm)，定义为

$$\|\boldsymbol{A}\| = \max\{\|\boldsymbol{Ax}\| : \boldsymbol{x} \in \mathbb{K}^n, \|\boldsymbol{x}\| = 1\} \tag{1.4.36}$$

$$= \max\left\{\frac{\|\boldsymbol{Ax}\|}{\|\boldsymbol{x}\|} : \boldsymbol{x} \in \mathbb{K}^n, \boldsymbol{x} \neq \boldsymbol{0}\right\} \tag{1.4.37}$$

常用的诱导范数为 p-范数

$$\|\boldsymbol{A}\|_p \stackrel{\mathrm{def}}{=} \max_{\boldsymbol{x}\neq 0} \frac{\|\boldsymbol{Ax}\|_p}{\|\boldsymbol{x}\|_p} \tag{1.4.38}$$

p 范数也称 Minkowski p 范数或者 L_p 范数。特别地，$p = 1, 2, \infty$ 时，对应的诱导范数分

别为

$$\|\boldsymbol{A}\|_1 = \max_{1\leqslant j\leqslant n} \sum_{i=1}^{m} |a_{ij}| \tag{1.4.39}$$

$$\|\boldsymbol{A}\|_{\mathrm{spec}} = \|\boldsymbol{A}\|_2 \tag{1.4.40}$$

$$\|\boldsymbol{A}\|_\infty = \max_{1\leqslant i\leqslant m} \sum_{j=1}^{n} |a_{ij}| \tag{1.4.41}$$

也就是说，诱导 L_1 和 L_∞ 范数分别直接是该矩阵的各列元素绝对值之和的最大值 (最大绝对列和) 及最大绝对行和；而诱导 L_2 范数则是矩阵 $\boldsymbol{A}$ 的最大奇异值。

诱导 L_1 范数 $\|\boldsymbol{A}\|_1$ 和诱导 L_∞ 范数 $\|\boldsymbol{A}\|_\infty$ 也分别称为绝对列和范数 (column-sum norm) 及绝对行和范数 (row-sum norm)。诱导 L_2 范数习惯称为谱范数 (spectrum norm)。

例如，矩阵

$$\boldsymbol{A} = \begin{bmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \\ 7 & -8 & -9 \\ -10 & 11 & 12 \end{bmatrix}$$

的绝对列和范数与绝对行和范数分别为

$$\|\boldsymbol{A}\|_1 = \max\{22, 26, 30\} = 30, \quad \|\boldsymbol{A}\|_\infty = \max\{6, 15, 24, 33\} = 33$$

2. “元素形式”范数 (“entrywise” norm)

将 $m\times n$ 矩阵先按照列堆栈的形式，排列成一个 $mn\times 1$ 向量，然后采用向量的范数定义，即得到矩阵的范数。由于这类范数是使用矩阵的元素表示的，故称为元素形式范数。元素形式范数是下面的 p 矩阵范数

$$\|\boldsymbol{A}\|_p \stackrel{\text{def}}{=} \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^p\right)^{1/p} \tag{1.4.42}$$

以下是三种典型的元素形式 p 范数：

(1) L_1 范数 (和范数) $(p=1)$

$$\|\boldsymbol{A}\|_1 \stackrel{\text{def}}{=} \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}| \tag{1.4.43}$$

(2) Frobenius 范数 $(p=2)$

$$\|\boldsymbol{A}\|_{\mathrm{F}} \stackrel{\text{def}}{=} \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2\right)^{1/2} \tag{1.4.44}$$

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

$$\|\boldsymbol{A}\|_\infty = \max_{i=1,\cdots,m;\, j=1,\cdots,n} \{|a_{ij}|\} \tag{1.4.45}$$

Frobenius 范数可以视为向量的 Euclidean 范数对按照矩阵各列依次排列的“拉长向量” $\boldsymbol{x}=[a_{11},\cdots,a_{m1},a_{12},\cdots,a_{m2},\cdots,a_{1n},\cdots,a_{mn}]^{\mathrm{T}}$ 的推广。矩阵的 Frobenius 范数有时也称 Euclidean 范数、Schur 范数、Hilbert-Schmidt 范数或者 L_2 范数。

Frobenius 范数又可写作迹函数的形式

$$\|\boldsymbol{A}\|_{\mathrm{F}} \stackrel{\text{def}}{=} \langle \boldsymbol{A},\boldsymbol{A}\rangle^{1/2}=\sqrt{\operatorname{tr}\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right)} \tag{1.4.46}$$

由正定的矩阵 $\boldsymbol{\Omega}$ 进行加权的 Frobenius 范数

$$\|\boldsymbol{A}\|_{\Omega}=\sqrt{\operatorname{tr}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Omega}\boldsymbol{A})} \tag{1.4.47}$$

称为 Mahalanobis 范数。

Schatten 范数就是用矩阵的奇异值定义的范数，将在第 5 章 (奇异值分析) 中介绍。

注意，向量 $\boldsymbol{x}$ 的 L_p 范数 $\|\boldsymbol{x}\|_p$ 相当于该向量的长度。当矩阵 $\boldsymbol{A}$ 作用于长度为 $\|\boldsymbol{x}\|_p$ 的向量 $\boldsymbol{x}$ 时，得到线性变换结果为向量 $\boldsymbol{Ax}$，其长度为 $\|\boldsymbol{Ax}\|_p$。线性变换矩阵 $\boldsymbol{A}$ 可视为一线性放大器算子。因此，比率 $\|\boldsymbol{Ax}\|_p/\|\boldsymbol{x}\|_p$ 提供了线性变换 $\boldsymbol{Ax}$ 相对于 $\boldsymbol{x}$ 的放大倍数，而矩阵 $\boldsymbol{A}$ 的 p 范数 $\|\boldsymbol{A}\|_p$ 是由 $\boldsymbol{A}$ 产生的最大放大倍数。类似地，放大器算子 $\boldsymbol{A}$ 的最小放大倍数由

$$\min|\boldsymbol{A}|_p \stackrel{\text{def}}{=} \min_{\boldsymbol{x}\neq 0}\frac{\|\boldsymbol{Ax}\|_p}{\|\boldsymbol{x}\|_p} \tag{1.4.48}$$

给出。比率 $\|\boldsymbol{A}\|_p/\min|\boldsymbol{A}|_p$ 描述放大器算子 $\boldsymbol{A}$ 的“动态范围”。

若 $\boldsymbol{A},\boldsymbol{B}$ 是 $m\times n$ 矩阵，则矩阵的范数具有以下性质

$$\|\boldsymbol{A}+\boldsymbol{B}\|+\|\boldsymbol{A}-\boldsymbol{B}\|=2(\|\boldsymbol{A}\|^2+\|\boldsymbol{B}\|^2)\quad(\text{平行四边形法则}) \tag{1.4.49}$$

$$\|\boldsymbol{A}+\boldsymbol{B}\|\cdot\|\boldsymbol{A}-\boldsymbol{B}\|\leqslant\|\boldsymbol{A}\|^2+\|\boldsymbol{B}\|^2 \tag{1.4.50}$$

以下是矩阵的内积与范数之间的关系 [238]。

(1) Cauchy-Schwartz 不等式

$$|\langle\boldsymbol{A},\boldsymbol{B}\rangle|^2\leqslant\|\boldsymbol{A}\|^2\|\boldsymbol{B}\|^2 \tag{1.4.51}$$

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：$\langle\boldsymbol{A},\boldsymbol{B}\rangle=0\ \Rightarrow\ \|\boldsymbol{A}+\boldsymbol{B}\|^2=\|\boldsymbol{A}\|^2+\|\boldsymbol{B}\|^2$

(3) 极化恒等式

$$\operatorname{Re}(\langle\boldsymbol{A},\boldsymbol{B}\rangle)=\frac{1}{4}\left(\|\boldsymbol{A}+\boldsymbol{B}\|^2-\|\boldsymbol{A}-\boldsymbol{B}\|^2\right) \tag{1.4.52}$$

$$\operatorname{Re}(\langle\boldsymbol{A},\boldsymbol{B}\rangle)=\frac{1}{2}\left(\|\boldsymbol{A}+\boldsymbol{B}\|^2-\|\boldsymbol{A}\|^2-\|\boldsymbol{B}\|^2\right) \tag{1.4.53}$$

式中 $\operatorname{Re}(\langle\boldsymbol{A},\boldsymbol{B}\rangle)$ 表示 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B}$ 的实部。

$\min\{m,n\}$。若 $p\times q$ 矩阵 $\boldsymbol{B}$ 是 $\boldsymbol{A}$ 的子矩阵，其奇异值 $\gamma_1\geqslant\cdots\geqslant\gamma_{\min\{p,q\}}$，则

$$\sigma_i\geqslant\gamma_i,\quad i=1,\cdots,\min\{p,q\} \tag{5.2.23}$$

并且

$$\gamma_i\geqslant\sigma_{i+(m-p)+(n-q)},\qquad i\leqslant\min\{p+q-m,p+q-n\} \tag{5.2.24}$$

矩阵的奇异值与矩阵的范数、行列式、条件数、特征值等有着密切的关系。

1\. 奇异值与范数的关系

矩阵 $\boldsymbol{A}$ 的谱范数等于 $\boldsymbol{A}$ 的最大奇异值，即

$$\|\boldsymbol{A}\|_{\rm spec}=\sigma_1 \tag{5.2.25}$$

注意到矩阵 $\boldsymbol{A}$ 的 Frobenius 范数 $\|\boldsymbol{A}\|_{\rm F}$ 是酉不变的，即 $\|\boldsymbol{U}^{\rm H}\boldsymbol{A}\boldsymbol{V}\|_{\rm F}=\|\boldsymbol{A}\|_{\rm F}$，故有

$$\|\boldsymbol{A}\|_{\rm F}=\left[\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right]^{1/2}=\|\boldsymbol{U}^{\rm H}\boldsymbol{A}\boldsymbol{V}\|_{\rm F}=\|\boldsymbol{\Sigma}\|_{\rm F}=\sqrt{\sigma_1^2+\sigma_2^2+\cdots+\sigma_r^2} \tag{5.2.26}$$

即是说，任何一个矩阵的 Frobenius 范数等于该矩阵所有非零奇异值平方和的正平方根。

2\. 奇异值与行列式的关系

设 $\boldsymbol{A}$ 是 $n\times n$ 正方矩阵。由于酉矩阵的行列式之绝对值等于 1，所以由定理 5.2.1 有

$$|\det(\boldsymbol{A})|=|\det\boldsymbol{\Sigma}|=\sigma_1\sigma_2\cdots\sigma_n \tag{5.2.27}$$

若所有 σ_i 都不等于零，则 $|\det(\boldsymbol{A})|\neq 0$，这表明 $\boldsymbol{A}$ 是非奇异的。若至少有一个 $\sigma_i(i>r)$ 等于零，则 $\det(\boldsymbol{A})=0$，即 $\boldsymbol{A}$ 奇异。这就是之所以把全部 σ_i 值统称为奇异值的原因。

3\. 奇异值与条件数的关系

对于一个 $m\times n$ 矩阵 $\boldsymbol{A}$，其条件数也可以利用奇异值定义为

$$\text{cond}(\boldsymbol{A})=\sigma_1/\sigma_p,\qquad p=\min\{m,n\} \tag{5.2.28}$$

由定义式 (5.2.28) 可以看出，条件数是一个大于或等于 1 的正数，因为 $\sigma_1\geqslant\sigma_p$。显然，由于至少有一个奇异值 $\sigma_p=0$，故奇异矩阵的条件数为无穷大，而条件数虽然不是无穷大，但很大时，就称 $\boldsymbol{A}$ 是接近奇异的。这意味着，当条件数很大时，$\boldsymbol{A}$ 的行向量或列向量的线性相关性很强。另由定义式 (5.1.8) 易知，正交或酉矩阵 $\boldsymbol{V}$ 的条件数等于 1。从这个意义上讲，正交或酉矩阵是“理想条件”的。式 (5.2.28) 也可用作条件数 $\text{cond}(\boldsymbol{A})$ 的评价。

考虑超定方程 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$。此时，由于 $\boldsymbol{A}^{\rm H}\boldsymbol{A}$ 的奇异值分解为

$$\boldsymbol{A}^{\rm H}\boldsymbol{A}=\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^{\rm H} \tag{5.2.29}$$

式中，$\mathcal{P}_\Omega : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ 是到指标集 Ω 的投影

$$[\mathcal{P}_\Omega(\boldsymbol{D})]_{ij} = \begin{cases} D_{ij}, & (i,j)\in\Omega \\ 0, & \text{否则} \end{cases} \tag{5.6.4}$$

若令 $\boldsymbol{D} = \boldsymbol{A} + \boldsymbol{E}$ 中的“噪声矩阵” $\boldsymbol{E} \in \mathbb{R}^{m\times n}$ 的平均平方元素值 $\sigma^2 = \frac{1}{mn}\|\boldsymbol{E}\|_2^2$，并且被恢复的矩阵为 $\boldsymbol{X}$，则矩阵恢复分为以下四种类型[174]：

(1) 低秩矩阵 $\boldsymbol{A}$ 的精确恢复 (exact recovery) $\hat{\boldsymbol{X}} = \boldsymbol{A}^*$。

(2) 低秩矩阵 $\boldsymbol{A}$ 的近精确恢复 (near-exact recovery) $\frac{1}{mn}\|\hat{\boldsymbol{X}} - \boldsymbol{A}^*\|_2^2 \leqslant \epsilon\cdot\sigma^2$。

(3) 低秩矩阵 $\boldsymbol{A}$ 的逼近恢复 (approximate recovery) $\frac{1}{mn}\|\hat{\boldsymbol{X}} - \boldsymbol{A}^*\|_2^2 \leqslant \epsilon\cdot \mathrm{scale}(\boldsymbol{A})$，其中 $\mathrm{scale}(\boldsymbol{A}) = \frac{1}{mn}\|\boldsymbol{A}\|_{\mathrm{F}}^2$ (平均平方元素幅值) 或者 $\mathrm{scale}(\boldsymbol{A}) = \|\boldsymbol{A}\|_\infty = \max\{A_{ij}\}$ (最大元素幅值)。

(4) 样本矩阵 $\boldsymbol{D}$ 的逼近恢复 $\frac{1}{mn}\|\hat{\boldsymbol{X}} - \boldsymbol{D}\|_2^2 \leqslant \sigma^2 + \epsilon\cdot\mathrm{scale}(\boldsymbol{A})$。

注意，精确和近精确恢复要求低秩矩阵 $\boldsymbol{A}$ 满足严格的非相干性假设，而一般的低秩矩阵很难满足这一条件。因此，对于不满足严格的非相干条件的低秩矩阵，只能实现逼近恢复。

式 (5.6.3) 所示矩阵完备问题是一个 NP 难题 (NP-hard problem)。为了使矩阵完备问题可解，必须将矩阵的秩最小化予以松弛。这一松弛与矩阵的 Schatten 范数密切相关。

若 $\boldsymbol{U} \in \mathbb{C}^{m\times m}$ 和 $\boldsymbol{V} \in \mathbb{C}^{n\times n}$ 是两个酉矩阵，满足 $\|\boldsymbol{A}\| = \|\boldsymbol{UAV}\|$ 的范数称为酉不变范数 (unitarily invariant norms)[505, 314]。

令矩阵 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ 有奇异值分解 $\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{H}}$。显然，$\|\boldsymbol{A}\| = \|\boldsymbol{U}^{\mathrm{H}}\boldsymbol{AV}\| = \|\boldsymbol{\Sigma}\|$ 是一酉不变范数。

令 $\boldsymbol{\sigma} = [\sigma_1,\cdots,\sigma_k]^{\mathrm{T}}, k = \min\{m,n\}$ 表示全部奇异值组成的向量，则酉不变范数 $\|\boldsymbol{A}\| = \|\boldsymbol{\Sigma}\|$ 可以用奇异值向量的范数形式定义：$\|\boldsymbol{A}\| = \|\boldsymbol{\sigma}\|$。特别地，称

$$\|\boldsymbol{A}\|_p = \|\boldsymbol{\sigma}\|_p = \left(\sum_{i=1}^{\min\{m,n\}} \sigma_i^p\right)^{1/p} \tag{5.6.5}$$

是矩阵 $\boldsymbol{A}$ 的 Schatten p 范数。

最常用的 Schatten 范数是 $p = 1, 2, \infty$ 三种情况：

(1) $p = 1$ 时的 Schatten 范数称为核范数 (nuclear norm)，定义为一矩阵的所有奇异值之和

$$\|\boldsymbol{A}\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i = \mathrm{tr}\left(\sqrt{\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}}\right) \tag{5.6.6}$$

式中，$\boldsymbol{B} = \sqrt{\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}}$ 满足 $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B} = \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$。

(2) $p = 2$ 时的 Schatten 范数与 Frobenius 范数等价

$$\|\boldsymbol{A}\|_2 = \|\boldsymbol{A}\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2} = \sqrt{\mathrm{tr}\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right)} = \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2 \tag{5.6.7}$$

(3) $p=\infty$ 时的 Schatten 范数与诱导 l_2 范数 (谱范数) 相同，即 $\|\boldsymbol{A}\|_\infty=\sigma_{\max}(\boldsymbol{A})$。

因此，核范数、Frobenius 范数和谱范数都是酉不变范数。

下面是精确地恢复低秩矩阵 $\boldsymbol{A}$ 和稀疏矩阵 $\boldsymbol{E}$ 的两个要点：

(1) 大多数的低秩矩阵都可以由样本元素的集合精确恢复，这些集合甚至可以只有少得惊人的元素数目。

(2) 秩 r 的矩阵 $\boldsymbol{M}\in\mathbb{R}^{n\times n}$ 可以通过求解下列优化问题完好地恢复

$$\min\|\boldsymbol{X}\|_*\quad \text{subject to } X_{ij}=M_{ij},\quad (i,j)\in\Omega \tag{5.6.8}$$

只要样本元素的个数

$$m\geqslant Cn^{6/5}r\log n \tag{5.6.9}$$

对某个正的常数 C 成立。式 (5.6.8) 中，$\|\boldsymbol{X}\|_*=\sum_i\sigma_i(\boldsymbol{X})$ 表示矩阵的核范数即矩阵所有奇异值之和。

令 $\boldsymbol{P}_\Omega$ 是到矩阵 $\boldsymbol{X}$ 的列空间的正交投影矩阵

$$\boldsymbol{P}_\Omega=\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^\dagger\boldsymbol{X}^{\mathrm{T}},\quad \boldsymbol{X}\in\Omega$$

则有 $\boldsymbol{P}_\Omega\boldsymbol{X}=\boldsymbol{X},\boldsymbol{X}\in\Omega$，相应的元素形式为

$$[\boldsymbol{P}_\Omega\boldsymbol{X}]_{ij}=\begin{cases}X_{ij}, & X_{ij}\in\Omega\\ 0, & X_{ij}\notin\Omega\end{cases} \tag{5.6.10}$$

于是，核范数最小化问题式 (5.6.8) 可以等价写作

$$\min\|\boldsymbol{X}\|_*\quad \text{subject to } \boldsymbol{P}_\Omega\boldsymbol{X}=\boldsymbol{P}_\Omega\boldsymbol{M} \tag{5.6.11}$$

为了求解矩阵完备问题，Candes 等人[88]提出了稳健主分量分析 (robust principal component analysis) 法：将 NP 难题的秩最小化松弛为核范数的最小化，利用主分量追踪 (principal component pursuit)，通过求解约束最小化问题

$$\min_{\boldsymbol{A},\boldsymbol{E}} f(\boldsymbol{A},\boldsymbol{E})=\|\boldsymbol{A}\|_*+\lambda\|\boldsymbol{E}\|_1\quad \text{subject to } \boldsymbol{D}=\boldsymbol{A}+\boldsymbol{E} \tag{5.6.12}$$

从数据矩阵 $\boldsymbol{D}\in\mathbb{R}^{m\times n}$ 恢复一个未知的低秩矩阵 $\boldsymbol{A}$ 与一个未知的稀疏矩阵 $\boldsymbol{E}$。式中，$\|\boldsymbol{A}\|_*=\sum\limits_i^{\min\{m,n\}}\sigma_i(\boldsymbol{A})$ 表示矩阵的核范数即所有奇异值之和，反映低秩矩阵的代价或者成本；$\|\boldsymbol{E}\|_1=\sum\limits_{i=1}^m\sum\limits_{j=1}^n|E_{ij}|$ 为矩阵 $\boldsymbol{E}$ 的所有元素的绝对值之和，描述稀疏矩阵的代价；常数 $\lambda>0$ 的作用是平衡低秩要求与稀疏要求之间的矛盾。

在没有关于稀疏模式与/或矩阵秩的附加信息的情况下，矩阵分解 $\boldsymbol{D}=\boldsymbol{A}+\boldsymbol{E}$ 无疑是一个病态问题，存在着低秩与稀疏之间的不确定性，从而引发下面两个可辨识性问题：

(1) 低秩矩阵本身有可能是非常稀疏的。

(2) 稀疏矩阵的非零元素有可能只集中在矩阵的某个列，该列元素就有可能否定低秩矩阵的对应列的元素，从而改变低秩矩阵的秩。