术语表

1. 矩阵统一用大写字母表示，如A, X标量用小写字母表示，如x, y;
   - 对于返回值是矩阵的函数，同理应该为$F(X)$,标量函数就记作$f(x)$
   - 另外，对于不混淆意思的$x * 1$维度矩阵，也就是**向量，允许用小写字母表示**，如x, y, x_i
2. 对于矩阵的行列数，统一用$\{m, n\}$下标表示，如$A_{\{m,n\}}$

外参数矩阵 – Camera Extrinsics: 一个$3 * 4$的矩阵T,

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} T'$$

λ **– 特征值**: 对于矩阵 A,如果满足

$$Ax = \lambda x$$

λ就是 A 的特征值，x 是特征向量

Tr **– 迹**: Tr 表示矩阵的迹，是用于求矩阵对角元素之和的算子

$$\mathrm{Tr}[X] = \sum_i X_{ii}$$

数学知识杂烩

1 矩阵论

1.1 线性空间与线性变换

1.1.1 线性空间

在矩阵论中，将在线性代数的基础上，推广向量空间R^n,一般地定义线性空间的概念。这里参考杨明教授的《矩阵论》[1] 和 Strang, Gilbert 的 《Linear Algebra and Its Applications》[2]。

定义 1.1.1.1 (线性空间 (linear space))：V是一个非空集合，F是一个数域，在其中定义两种运算, 加法($\forall \alpha, \beta \in V, \alpha + \beta \in V$)和数乘($\forall k \in F, \alpha \in V, k\alpha \in V$), 如果满足若干运算法则，就称$V$是数域$F$上的线性空间。

比如$F_{\{n,1\}} = \left\{(x_1, x_2, ..., x_n)^T\right\} \mid x_i \in F$ 是 F 上的$\{m, 1\}$维线性空间。

$F_{\{n,m\}} = \left\{A = (a_{ij})_{\{m\times n\}} | a_{ij} \in F\right\}$是 F 上的$\{m, n\}$维矩阵空间 记作$R_{\{2,2\}}$[1]

$P_n(x) = \left\{\sum_{i=0}^{n-1} a_i x_i | a_i \in R\right\}$称为多项式空间$P_{n[x]}$

定义 1.1.1.2：如果对于线性空间V 存在一组**线性无关**的向量$(\alpha_1, \alpha_2, ..., \alpha_n)$，使得空间中任意一个向量可以由它表示，$(\alpha_1, \alpha_2, ..., \alpha_n)$张成了向量空间,此时有$V$的维度

$$\dim(V) = n$$

定义 1.1.1.3 (矩阵的逆)：如果对于向量A，如果存在 B 使得$AB = I$,则称 B 是 A 的逆矩阵，记作A^{-1}，并且说 A 是可逆的，有时候也叫做非奇异矩阵[2]。并且

1. $(AB)^{-1} = B^{-1}A^{-1}$
2. A 的各列线性无关，$\dim(A) = n$

---

[1]这里与教课书上$R^{2\times 2}$的写法不同，读者注意。

[2]在数值运算中，我们很少求解矩阵的逆，计算他的计算量是行变换解方程的三倍[2]

1.1.2 内积空间

1.1.2.1 欧氏空间和酉空间

定义 1.1.2.1.1：对数域F上的线性空间$V_n(F)$,定义一个从$V_n(F)$到F的二元运算(α,β),记作$(\alpha,\beta):V_n(F)\to F$，满足以下条件

1. $(\alpha,\beta)=(\beta,\alpha)$
2. $(k\alpha,\beta)=k(\alpha,\beta)$
3. $(\alpha,\alpha)\geq 0,\quad (\alpha,\alpha)=0$仅当 $\alpha=0$

我们就称这个二元运算是线性空间的一个内积，并且定义了内积的线性空间称作内积空间$\left[V_{n(F)};(\alpha,\beta)\right]$。如果$F=R$（实数域）就称作欧氏空间；如果$F=C$（复数域）就称作酉空间。

下面举几个欧氏空间的例子

1. $[\mathbb{R}^n;(\alpha,\beta)=\alpha^T\beta]$,习惯上，我们直接将$\mathbb{R}^n$记作欧氏空间。
2. $\left[R_{\{n,n\}};(A,B)=\mathrm{Tr}[AB^T]\right]$
3. $\left[P_n[x];(f(x),g(x))=\int_0^1 f(x)\cdot g(x)\mathrm{d}x\right).]$

定义 1.1.2.1.2 ($\|\alpha\|$)：在内积空间，称

$$\|\alpha\|=\sqrt{(\alpha,\alpha)}=\alpha^T\alpha$$

为向量α的长度。在欧式空间中，也称作**欧几里得范数**,常表示为$\|\alpha\|_2$**(注意下标)**。下面给出几条它的性质

1. **(Cauchy 不等式)** $|(\alpha,\beta)|^2\leq\|\alpha\|\|\beta\|$
2. $\|\alpha+\beta\|\leq\|\alpha\|+\|\beta\|$
3. 定义α,β夹角$\theta=\arccos\dfrac{(\alpha,\beta)}{\|\alpha\|\|\beta\|}$

1.1.2.2 标准正交基

定义 1.1.2.2.1：在V_n中一组基$\{\varepsilon_1,\varepsilon_2,...,\varepsilon_n\}$满足

$$(\varepsilon_i,\varepsilon_i)=0\ \ \&\ \ \|\varepsilon_i\|=1\ \ \&\ \ \left(\varepsilon_i,\varepsilon_j\right)\neq 0\quad 0\leq i,j\leq n,i\neq j$$

就称作其为标准正交基。

从一组基$\{\alpha_1,\alpha_2,...,\alpha_n\}$转换到标准正交基$\{\beta_1,\beta_2,...,\beta_n\}$有

$$\beta_i=\alpha_i-\sum_{i=1}^{k-1}\frac{(a_k,\beta_i)}{(\beta_i.\beta_i)}\beta_i$$

证明方法参考[1] [p. 18],这里略去。用矩阵的方式表示为

$$(\alpha_1,\alpha_2,...,\alpha_n)=(\beta_1,\beta_2,...,\beta_n)\begin{bmatrix}(\beta_1,\beta_1) & (\alpha_2,\beta_1) & ... & (\alpha_n,\beta_1)\\ & (\beta_2,\beta_2) & ... & (\alpha_n,\beta_2)\\ & & ... & ...\\ & & & (\beta_n,\beta_n)\end{bmatrix} \tag{1.1}$$

1.1.2.3 线性变换

定义 1.1.2.3.1 (线性变换): 如果存在一个单射$T:V_n(F)\to V_n(F)$,满足

1. $T(\alpha+\beta)=T(\alpha)+T(\beta)$
2. $T(k\alpha)=kT(\alpha)$

就称T是一个$V_n(F)$上的线性变换。

定义 1.1.2.3.2 (线性变换的矩阵): $\forall\alpha\in V_n(F)$,取一组基$\{\alpha_1,\alpha_2,...,\alpha_n\}$

$$\alpha=x_1\alpha_1+x_2\alpha_2+...+x_n\alpha_n$$

$$[T(\alpha_1)\ \ T(\alpha_2)\ \ ...\ \ T(\alpha_n)]^T=A^T[\alpha_1\ \ \alpha_2\ \ ...\ \ \alpha_n]^T$$

称A为基$\{\alpha_1,\alpha_2,...,\alpha_n\}$下的**矩阵**。

定理 1.1.2.3.1: 如果 $\{\alpha_1,\alpha_2,...,\alpha_n\}$到$\{\beta_1,\beta_2,...,\beta_n\}$有过渡矩阵$C$

$$\text{其中}\quad [\alpha_1\ \ \alpha_2\ \ ...\ \ \alpha_n]=[\beta_1\ \ \beta_2\ \ ...\ \ \beta_n]C$$

,同时在两组基下的矩阵分为为A,B，则有

$$B=C^{-1}AC$$

定义 1.1.2.3.3 (不变子空间): 如果T是$V_n(F)$上的线性变换，W是$V_n(F)$的一个子空间，如果$T(W)\in W$，就称W是T的不变子空间。

定义 1.1.2.3.4 (正交变换): 如果T是$V_n(F)$上的线性变换，且满足$\forall\alpha,\beta\in V_n(F),(T(\alpha),T(\beta))=(\alpha,\beta)$，就称$T$是正交变换。 当空间是欧式空间就叫做正交变换，其对应矩阵我们叫做正交矩阵C,酉空间就叫做酉变换,其矩阵叫做酉矩阵U。

1.2 Jordan 标准形

1.2.1 Jordan 标准形及其求解

形如$J(\lambda)=\begin{bmatrix}\lambda & 1 & & \\ & \lambda & 1 & \\ & & \ddots & 1\\ & & & \lambda\end{bmatrix}$的矩阵叫做$r$阶 Jordan 块，若干个 Jordan 块组成的矩阵叫做 Jordan 矩阵，如下

$$J = \begin{bmatrix} J_1(\lambda_1) & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_m(\lambda_n) \end{bmatrix}$$

在复数域上，每个方阵矩阵都可以相似于 Jordan 矩阵，并且唯一（若不计较 Jordan 的排列次序），即$A = S^{-1}JS$，其中S是可逆矩阵，J是 Jordan 矩阵。

定义 1.2.1.1 (特征值和特征向量)：对于定义在E 上的变换 T, 如果 $\exists\lambda$

$$T(\mu) = \lambda\mu$$

这里的λ就是T的特征值，μ是特征向量。

定义 1.2.1.2 (直和 $\oplus$):

$$A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} = \operatorname{diag}(A, B)$$

定义 1.2.1.3 (矩阵的核 $\operatorname{Null}(A)$):

$$\operatorname{Null}(A) = \{v \in V : Av = 0\}.$$

定理 1.2.1.1 (秩-零化度定理)：$\operatorname{rank} A + \operatorname{nullity}\ A$(A 的特征值空间维度$) = n$

$$[A|E]^T = \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

对其做行初等变换，将 A 的下几行消元成 0，对应的右边行即为一个解。可以理解为其核空间每一个维度可以消解 A 空间一个冗余维度。

首先我们可以求解出其所有特征值。

代数重数gm

$$|\lambda I - A| = \sum_{i=0}^{k} (\lambda - \lambda_i)^{(a_i)}$$

然后对于每一个特征值,$\dim(\operatorname{Null}(\lambda_i I - A))$是多少就有多少 jordan 块，同时我们也要求解出其核空间的向量,记作x,然后使用 jordan 链进行递归求解直到无解,求解每一个 jordan 块的大小。

$$A\underbrace{\begin{bmatrix}p_1 & p_2 & p_3 & p_4\end{bmatrix}}_{P} = \begin{bmatrix}p_1 & p_2 & p_3 & p_4\end{bmatrix}\begin{bmatrix}\lambda & 1 & 0 & 0\\ 0 & \lambda & 1 & 0\\ 0 & 0 & \lambda & 1\\ 0 & 0 & 0 & \lambda\end{bmatrix} = \begin{bmatrix}\lambda p_1 & \lambda p_2 + p_1 & \lambda p_3 + p_2 & \lambda p_4 + p_3\end{bmatrix}$$

$$(A-\lambda E)p_1 = 0 \quad (1)$$
$$(A-\lambda E)p_2 = P_1 \quad (2)$$
$$\cdots$$
$$(A-\lambda E)p_n = P_{n-1} \quad (n)$$

上式叫做jordan链，直到$(k+1)$无法解出。则这个特征向量对应的jordan块的大小确定。

例 1.2.1.1: 求可逆矩阵 P 和 jordan 矩阵 J 使得$A = P^{-1}JP$

$$A = \begin{bmatrix}-3 & 3 & -2\\ -7 & 6 & -3\\ 1 & -1 & 2\end{bmatrix}$$

$|\lambda I - A| = 0 \Rightarrow (\lambda-1)(\lambda-2)^2$,确定其有两个 jordan 矩阵$J_1, J_2$组成，对于$J_1$，向量$\alpha_1 = \begin{bmatrix}1 & 2 & 1\end{bmatrix}^T$。

求解$(2I-A)x = 0$,只有一个特征向量$\alpha_2 = \begin{bmatrix}-1 & -1 & 1\end{bmatrix}^T$，所以$J_2$只由一个 jordan 块构成。根据 jordan 链求出$\beta_1 = (-1,-2,0)^T$。所以得

$$P = \begin{bmatrix}\alpha_1 & \alpha_2 & \beta_1\end{bmatrix}$$

1.2.2 最小多项式

定义 1.2.2.1 (矩阵多项式): 首先定义$g(\lambda) = \sum_{i=0}^{m} a_i\lambda^i$,那么就可以定义矩阵 $g(A) = \sum_{i=0}^{m} a_i A^i$ 为 A 的矩阵多项式。给出如下性质

1. $Ax = \lambda x \Rightarrow g(A)x = g(\lambda)x$
2. $P^{-1}AP = B \Rightarrow P^{-1}g(A)P = g(B)$
3. $A = \mathrm{diag}(A_1, A_2, ..., A_n) \Rightarrow g(A) = \mathrm{diag}(g(A_1), g(A_2), ..., g(A_n))$
4. A的特征多项式$(f(\lambda) = |\lambda I - A| = \sum_{i=0}^{m} a_i\lambda^i)$就是$A$的化零多项式$(g(A) = 0)$。

证 *of 4*: 左式λ换A。 ■

例 1.2.2.1 (求 A^m): 假设$g(A) = A^m \Rightarrow g(\lambda) = \lambda^m = \mathrm{Hl}(\lambda) + \sum b_i\lambda_i \Rightarrow g(A) = \sum b_i A_i$

定义 1.2.2.2 (最小多项式)：A的所有化零多项式中次数最低,$a_0=0$（首项系数为零）的最小多项式，记作$m_r(\lambda)$。和特征多项式有相同的根，重数可以不同。

> **定理 1.2.2.1**：一个矩阵可以对角化的充分必要条件是$m_r(\lambda)$的所有重数为 1.
>
> $$m_{r(\lambda)}=\Pi(\lambda-a_i)\quad a_i\text{互不相等}$$

1.3 矩阵分解

1.3.1 常见

1.3.1.1 矩阵的三角分解

设$A\in F_{\{n,n\}}$

1. $L,U\in F_{\{n,n\}}$分别为下三角和上三角矩阵，$A=LU$称作LU分解
2. $L,V\in F_{\{n,n\}}$分别为对角元素为 1 的下三角和上三角矩阵，D是对角矩阵，称作LDV分解

例 1.3.1.1.1 (求解$P=\begin{bmatrix}2&2&3\\4&7&7\\-2&4&5\end{bmatrix}$的 LU 和 PDV 分解)：这里简单给出一个定理，

$$\underset{\uparrow\ \text{行初等变换}}{H}[P|E]=[HP|P]\qquad\begin{bmatrix}P\\--\\E\end{bmatrix}L=\begin{bmatrix}PL\\--\\\underset{\uparrow\ \text{列初等变换}}{L}\end{bmatrix}$$

将其写成增广矩阵$[P\mid E]\Rightarrow[U\mid L']$,做行初等变换将 P 消成上三角矩阵。$L'A=U\Rightarrow A=(L')^{-1}U=LU$

对U建立$\begin{bmatrix}P\\--\\E\end{bmatrix}$，做列初等变换，$\begin{bmatrix}P\\--\\E\end{bmatrix}\Rightarrow\begin{bmatrix}D\\--\\V'\end{bmatrix}$,有$UV'=D\Rightarrow U=D(V')^{-}1=DV$

接下来我们将分析一个矩阵有 LU,LDV 分解的条件，并且注意，矩阵的LDV分解并不唯一。

对于A,设 k 阶顺序余子式$\Delta_i=\left|A_{[0:i;0:i]}\right|$，如果$\mathrm{Rank}(A)=k$,其$1-k$阶顺序余子式不为 0，则可 LU 分解，1到$(n-1)$阶顺序余子式不为 0，则可 LDV 分解。

1.3.1.2 满秩分解

定理 1.3.1.2.1 ($R_{行}(A) = R_{列}(A) = R(A)$): 行的秩表示为行向量张成向量空间的维度，列的秩表示为行向量张成向量空间的维度。

证: 根据高斯变换，得到行的秩等于主元个数等于列空间维度。 ■

定理 1.3.1.2.2: 任意不为零的矩阵$A_{\{m,n\}}$都有满秩分解

$$A_{\{m,n\}} = B_{\{m,r\}}C_{\{r,n\}}$$

求解方法①$B'[A \mid E] \Rightarrow [\underbrace{C'}_{\text{行阶梯型}} \mid B'] \Rightarrow A = [B_{\{m,r\}} \quad 0_{\{m,n-r\}}]\begin{bmatrix} C_{\{r,n\}} \\ 0_{\{n-r,n\}} \end{bmatrix} = B_{\{m,r\}}C_{\{r,n\}}$

② 先化为 Hermite 标准形H(带主元行阶梯型),找到 A 中主元列取出构成$B = (\alpha_1, \alpha_2, ..., \alpha_n)$,$H$中非零行构成$C$

1.3.1.3 可对角化矩阵的谱分解

$$A = P^{-1}\operatorname{diag}(\underbrace{\lambda_1\ ,..., \lambda_1}_{r_1个}, \lambda_2, ..., \underbrace{\lambda_n}_{r_n个})P = \underbrace{P^{-1}\sum\left(\lambda_i I_{r_i}\right)P}_{谱分解}$$

其中λ_i表述矩阵的特征值，r_i表示特征值的重数。并且我们有

$$\sum I_{r_i} = E \ \ \&\ \ I_{r_i}I_{r_j} = 0(i \neq j) \ \ \&\ \ I_{r_i}^2 = I_{r_i}^2 \ \ \&\ \ I_{r_i}^T = I_{r_i}$$

1.3.2 Schur 分解与正规矩阵

定理 1.3.2.1 (UR 分解): 对于一个可逆矩阵A,存在

$$A = U(\text{ 可逆矩阵})\ R\ (\text{上三角矩阵})$$

证: 根据 式(1.1),可知 ■

定理 1.3.2.2: 如果A 列满秩，则

$$A_{\{m,k\}} = Q_{\{m,k\}}R_{\{k,k\}} \quad \text{Q 的列向量是 A 列空间的标准正交基}$$

定义 1.3.2.1 (正规矩阵): 对于矩阵$A_{\{m,n\}}$,如果$A^HA = AA^H$ 就称A是正规矩阵。

常见的对角矩阵，对称反对称($A^T = \pm A$),Hermite 矩阵与反 Hermite 矩阵($A^H = \pm A$),正交矩阵与酉矩阵($AA^T = A^TA = I, AA^H = A^HA = I$)都是正规矩阵。

定理 1.3.2.3：对于矩阵$A_{\{n,n\}}$,其是正规矩阵的充要条件是A相似于对角矩阵，即

$$\exists U_{\{n,n\}} \quad U^H A U = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$$

推论充分必要条件

①A 的特征向量是空间的标准正交基。

②有**谱分解**$A = \sum_{i=1}^{s} \lambda_i I_i$,

1.3.3 矩阵的奇异值分解

矩阵的奇异值分解是在线性动态系统的辨识,最佳逼近问题,实验数据处理,数字 图像存储中应用广泛的一种分解。

定理 1.3.3.1：对于$AA^H \in C_{\{m,n\}}$ $A^H A \in C_{\{m,n\}}$, 有如下性质

1. $\mathrm{Rank}(AA^H) = \mathrm{Rank}(A^H A) = \mathrm{Rank}(A)$ （Use 定理 1.3.1.2.1）
2. $\lambda_i \neq 0$ ，$AA^H x = \lambda_i x \Rightarrow A^H A x = \lambda_i x$
3. 均为半正定矩阵，行满秩$(AA^H)_{\{m,m\}}$ 正定，列满秩$(A^H A)_{\{n,n\}}$ 正定。

证 *of 2*：$|AA^H - \lambda_i E| = 0 = \left|(AA^H - \lambda_i E)^T\right| = |A^H A - \lambda_i E|$ ■

定义 1.3.3.1 (奇异值)：对于矩阵$A_{\{m,n\}}$,其奇异值是AA^H的特征值的平方根

$$\sigma_i = \sqrt{\lambda_i} \ \& \ (AA^H)x = \lambda_i x$$

其奇异值个数等于矩阵的秩。

定理 1.3.3.2：For $A_{\{m,n\}}$,其奇异值可分解为

$$A = U_{\{m,m\}} \Sigma_{\{m,n\}} V_{\{n,n\}}^H$$

其中U, V是酉矩阵,

$$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, ..., \sigma_w) \oplus 0_{\{m-w,n-w\}} \ \& \ w = \mathrm{rank}(A^H A) \ \& \ \sigma_i \neq 0$$

证: 对于$A^H A$ 因为其是半正定矩阵，所以有

$$V^H A^H A V = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_k) \oplus 0$$

其中$V = [v_1 \ v_2 \ ... \ v_n]$ 是$A^H A$标准正交的特征向量

$$
\begin{aligned}
(Av_i, Av_j) = (Av_j)^H Av_i = v_j^H (A^H Av_i) = v_j^H (\lambda_i v_i) = \lambda_i \left(\overset{\substack{0 \ \& \ i \neq j \\ 1 \ \& \ i = j}}{v_j^H v_i} \right) \\
= 0 \ \& \ (i \neq j) \\
= \lambda_i = \sigma_i^2 \ \& \ (i = j, \lambda_i \neq 0)
\end{aligned}
$$

设$u_i = \frac{1}{\sigma_i} Av_i \Rightarrow Av_i = \sigma_i u \ \& \ i = 1, 2, ..., \mathrm{rank}(A)$. 所以

$$
\begin{aligned}
AV = A[v_1 \ \ v_2 \ \ ... \ \ v_n] = [\sigma_1 u_1 \ \ \sigma_2 u_2 \ \ ... \ \ \sigma_r u_r \ \ 0 \ \ ... \ \ 0] \\
= [u_1 \ \ u_2 \ \ ... \ \ u_r \ \ 0_{\{1,n-r\}}] \,\mathrm{diag}(\sigma_1, \sigma_2, ..., \sigma_r, 0, ..., 0)
\end{aligned}
$$

$$
A = U \begin{bmatrix} \Delta & \\ & 0 \end{bmatrix} V^{-1} = U \Sigma V^H = \sum_{i=1}^{r} (u_i \sigma_i v_i^H)
$$

后面这一项可以取$k < r$,做近似压缩。 ∎

定理 1.3.3.3 (极分解): 对于$A_{\{n,n\}}$,有$A = P_{\{n,n\}}(\text{半正定})\ Q_{\{n,n\}}(\text{酉矩阵})$

证: $A = U\Sigma V^H = U\Sigma(E)V^H = U\Sigma(U^H U)V^H = (U\Sigma U^H)(UV^H) = PQ$ ∎

P 是缩放, Q 是旋转。所以得名极分解。

定义 1.3.3.2 (对称、实对称、正定、半正定、海森矩阵):

名称	**定义**	注释
实对称矩阵	$A \in \mathbb{R}^{n,n} \ \ A = A^T$	反对称$A = -A^T$
Hermite(厄米特)矩阵	$A \in \mathbb{C}^{n,n} \ \ A = A^H$	反 Hermite$A = -A^H$
正交矩阵	$T \in \mathbb{R} \ \ (\alpha, \beta) = (T(\alpha), T(\beta))$	$AA^T = E, A^{-1} = A^T$
酉矩阵	$T \in \mathbb{C} \ \ (\alpha, \beta) = (T(\alpha), T(\beta))$	$A^H A = AA^H = E$
正规矩阵	$AA^H = A^H A$	与正交矩阵相似的矩阵
正定矩阵	$\forall x \in \mathbb{R}^n, x^T Ax > 0$	正定矩阵的特征值都是正的
半正定矩阵	$\forall x \in \mathbb{R}^n, x^T Ax \geq 0$	特征值都是非负的
海森矩阵	$H_{\{i,j\}} = \frac{\partial^2 f}{\partial x_i \partial x_j}$	二阶偏导数矩阵

1.3.3.1 最小二乘应用

参考[3],我们简述其在最小二乘问题上的应用。考虑

$$
\min \|Ac - y\|_2 \ \& \ (A \text{ 为数据矩阵 } c \text{ 为参数矩阵})
$$

比如对于 $f(x) = c_1x + c_2x^2 + c_3x^3$

$$A = \begin{bmatrix} x_1 & x_1^2 & ... & x_1^3 \\ x_2 & x_2^2 & ... & x_2^3 \\ & ... & ... & \\ x_n & x_n^2 & ... & x_n^3 \end{bmatrix} \ \& \ c = \begin{bmatrix} c_1 \\ c_2 \\ ... \\ c_n \end{bmatrix} y = \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix}$$

首先我们知道$A = U\sum V^H \ \& \ U$是酉矩阵, $\|U\| = 1$,所以

$$\|y - Ac\|_2^2 = \left\|U^H(y - Ac)\right\|_2^2 = \left\|U^H\left(y - U\sum V^Hc\right)\right\|_2^2$$

L2 范数[3]

$$\begin{aligned}
&= \left\|U^Hy - \sum V^Hc\right\|_2^2 = \left\|D - \sum Z\right\|_2^2 \\
&= \left\|\left(d_1 \ d_2 \ ... \ d_n\right)^T - \left(\sigma_1z_1 \ ... \ \sigma z_r \ 0_{\{1,n-r\}}\right)^T\right\|_2^2 \\
&= \left\|\left(d_1 - \sigma_1z_1 \ d_2 - \sigma_2z_2 \ ... \ d_r - \sigma_rz_r \ d_{r+1} \ ... \ d_n\right)^T\right\|_2^2 \\
&= \sum_{i=1}^{r}(d_i - \sigma_iz_i)^2 + \sum_{i=r+1}^{n} d_i^2
\end{aligned}$$

因为$A = U\sum V^H$ A为定值, $U, \sum, V_H$ 也为定值，所以d也为定值

$$\min \|y - Ac\|_2^2 = \min \sum_{i=1}^{r}(d_i - \sigma_iz_i)^2 \Rightarrow z_i = \frac{d_i}{\sigma_i} \ \& \ (i = 1, 2, ..., r)$$

$$V^Hc_i = z_i \rightarrow c_i = Vz_i \ \& \ (i = 1, 2, ..., r)$$

并且$[r_1, n]$ 的c_i 不影响，可任取。

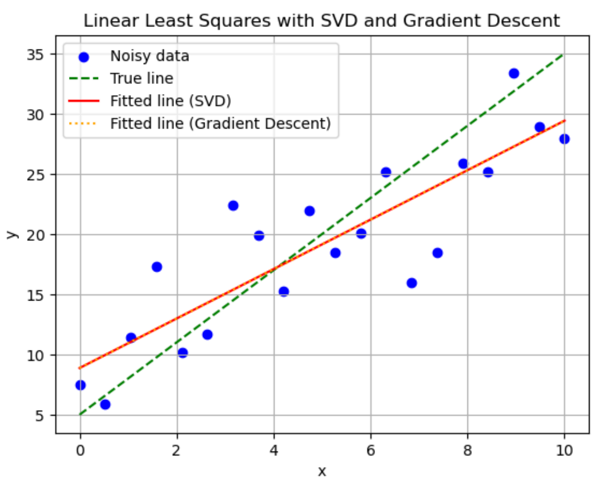

图 1.1: 通过 SVD 求解线性最小二乘问题的例子并和梯度下降法比较

[3]L2 范数表示欧几里得空间上符合直觉的向量长度$\|x\|_2 = \sqrt{x^Tx} = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$

1.4 矩阵的广义逆

1.4.1 减号广义逆

定义 1.4.1.1 (逆的推广): for $A_{\{m,n\}}$

$$A^{-1}_{\{m,m\}}A_{\{m,n\}} \quad (\text{左逆}) \qquad A_{\{m,n\}}A^{-1}_{\{n,n\}} \quad (\text{右逆})$$

左逆列满秩($m \geq n$)，右逆行满秩($m \leq n$)。

$$(\text{左逆求解})\, P_{\{n,m\}}\begin{bmatrix}A_{\{m,n\}} & E_{\{m,m\}}\end{bmatrix} = \begin{bmatrix}E_{\{m,m\}} & \cdots & P_{\{m,m\}}\end{bmatrix} = \begin{bmatrix}E_{\{m,m\}} & \cdots & P_{\{n,m\}} \\ & & \cdots\end{bmatrix}$$

$$(\text{右逆求解})\begin{bmatrix}A_{\{m,n\}} \\ E_{\{n,n\}}\end{bmatrix}P_{\{n,m\}} = \begin{bmatrix}E_{\{m,m\}} \\ \cdots \\ P_{\{n,n\}}\end{bmatrix} = \begin{bmatrix}E_{\{m,m\}} & \\ \cdots & \\ P_{\{n,m\}} & \cdots\end{bmatrix}$$

定义 1.4.1.2 (减号广义逆): 对于$A_{\{m,n\}} \in \mathbb{C}^{n,n}$ 存在$G_{\{n,m\}} \in \mathbb{C}^{n,n}$ 使得 $AGA = A$,就称G是A的减号广义逆。 $G \in A\{1\}$ & $A\{1\} = \{A_1^-, A_2^-, ..., A_n^-\}$

$G \in A\{1\}$**的充分必要条件**

$$\text{PAQ} = \text{diag}(I_r, 0) \;\&\; G = Q\begin{bmatrix}I_r & U \\ V & W\end{bmatrix}P \;\&\; UVW\text{大小合适，任意矩阵}$$

证: ①必要性, $P(AGA)Q = \text{diag}(I_r, 0)\begin{bmatrix}I_r & U \\ V & W\end{bmatrix}\text{diag}(I_r, 0) = \text{diag}(I_r, 0) = PAQ$

②充分性 $PAQ = P(AGA)Q = \text{diag}(I_r, 0)Q^{-1}GP^{-1}\,\text{diag}(I_r, 0) = \text{diag}(I_r, 0)$

$$\begin{bmatrix}I_r & \\ & 0\end{bmatrix}Q^{-1}GP^{-1}\begin{bmatrix}I_r & \\ & 0\end{bmatrix} = \begin{bmatrix}I_r & \\ & 0\end{bmatrix} \Rightarrow Q^{-1}GP^{-1} = \begin{bmatrix}E_{\{r,r\}} & U \\ V & W\end{bmatrix}$$

所以$G = Q\begin{bmatrix}E_{\{r,r\}} & U \\ V & W\end{bmatrix}P$ ■

至于求解过程就是

$$\begin{bmatrix}A_{\{m,n\}} & E_{\{m,m\}} \\ E_{\{n,n\}} & 0\end{bmatrix} \Rightarrow (\text{做行列初等变换}) \Rightarrow \begin{bmatrix}\text{diag}(E_r, 0) & P_{\{m,m\}} \\ Q_{\{n,n\}} & 0\end{bmatrix}$$

1.4.2 M-P 逆

定义 1.4.2.1 (M-P 逆): 对于$A_{\{m,n\}}$,其 M-P 逆$A^+ = G$, 有(任意矩阵都存在，并且唯一)

$$AGA = A \;\&\; GAG = G \;\&\; (AG)^H = AG \;\&\; (GA)^H = GA$$

不做具体赘述了，记住在 python 中可以使用 `np.linalg.pinv` 求解就足够了。

> **定理 1.4.2.1 (最佳的最小二乘解)**:
> $$\min \|Ax-b\|_2^2 = (x = A^+ b)$$

1.5 矩阵分析

1.5.1 矩阵函数及其应用

1.6 计算机视觉中的应用

我们会用到的向量$\in \mathbb{R}^2\mathbb{R}^3$ [4]。

1.6.1 $\mathbb{R}^3$中的变换（相机相关）

定义 1.6.1.1 (齐次坐标)：对于n维度向量$x = (x_1, x_2, ..., x_n)$，其在$n+1$维向量空间中的齐次向量（坐标）为$(x_1, x_2, ..., x_n, 1)^T$,称为$n$维齐次坐标，记作$\tilde{x}$。他对于$\mathbb{R}^2\mathbb{R}^3$中的变换非常有用, 它统一了旋转、平移和缩放。(但是注意齐次坐标并不能相加或者相乘, 只能乘以$T_{\{n+1,n+1\}}$来做变换)。

比如对于一个向量p,一个混合变换$p' = Ap + t$,使用分块矩阵的知识，也可以看作是

$$p' = \begin{bmatrix} A_{\{n,n\}} & t_{\{1,1\}} \\ B_{\{1,n\}} & s_{\{1,1\}} \end{bmatrix} \begin{bmatrix} p \\ 1 \end{bmatrix} = T\ \mathrm{Hm}(p) \quad \&\Big(t=0, B=0_{\{1,n\}}\Big)$$

这里的T,我们称作透视矩阵(Transform Matrix)。T的第一行代表的是以$(x,y,z)^T$为基的坐标系下的变化，T_{21}一般直接以0,T_{22}也可以表示缩放性质。

变换	条件	自由度
欧几里得变换	$A=R\ \&\ s=1\ \&\ B=0_{\{1,3\}}$	6
缩放变换	$A=cR\ \&\ s=1\ \&\ B=0_{\{1,3\}}$	7
纯缩放变换	$A=cE\ \&\ s=1\ \&\ B=0_{\{1,3\}}, t=0_{\{3,1\}}$	1
仿射变换	$s=1\ \&\ B=0_{\{1,3\}}$	15
射影变换	None	12

1.6.1.1 欧几里得变换、缩放变换

欧几里得变换包括绕原点旋转，平移，轴对称，中心对称（正交变换），在欧几里得变换中的T_A是正交矩阵，只有三个自由度（之后三维空间刚体变换会讲到)。

[4] $\mathbb{R}^n$表示 n 维向量空间

定义 (正交矩阵): [4] 正交矩阵(Orthogonal Matrix)是指矩阵的转置等于逆矩阵的矩阵，也就是说$A^{-1} = A^T$，并且可以推证$|A| = 1$。 其中，$\det(A) = 1$的时候叫做**旋转矩阵**，$\det(A) = -1$的时候叫做**瑕旋转矩阵**（瑕旋转是旋转加上镜射。镜射也是一种瑕旋转）。所有$A_{\{n,n\}}$构成一个**正交群**。

至于缩放变换呢,略去介绍。

1.6.1.2 小孔成像模型

我们直接俄导一下基于小孔成像的相机模型，也就是三维物体在二维计算机屏幕上的表示。

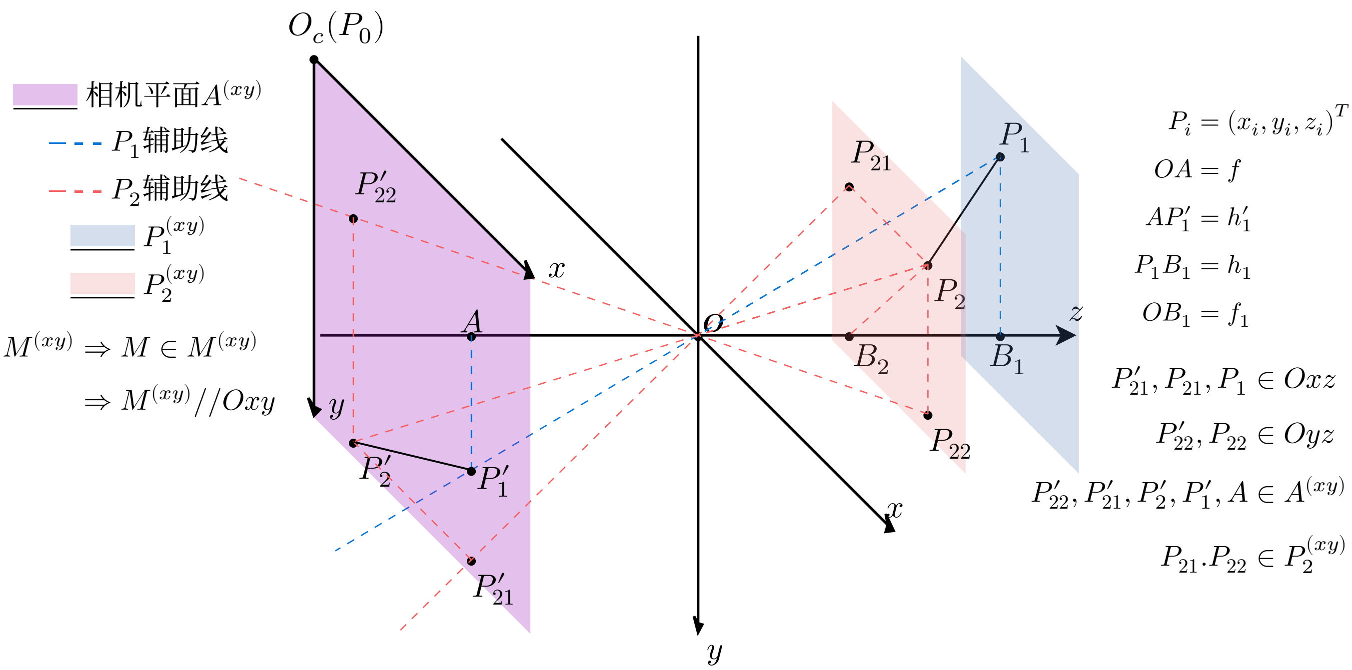

图 1.2: 相机模型示意图

如图 2 中所示， 对于任意一点，相当于关于原点的一个对称。对于大部分教程，都教大家用相似三角形求解,你可以在上图中寻找对应的三角形。诚然，这就是一个**纯缩放变换**，只需要求出缩放系数c可。

$$c = -\frac{f}{z_i}$$

在现代相机中，我们对像做了预处理，使得原本成倒像的小孔成像可以正确成像，成像平面于$A'^{(xy)}$平面（**实际像平面**），所以实际上$c_a = -c = \dfrac{f}{z_i}$。

同时我们定义像素坐标系如图中所示，以$O_c = (x_0, y_0, z_0 = f)^T$为原点，$x, y$轴如图所示的坐标系（实际上它应该定义在实际像平面上）。假设$\alpha = c_a k_x, \beta = c_a k_y, \gamma = c_a k_z$, ($k_x, k_y$为两个尺度因子，表示从米变换为像素)从$P_i$到$P'_i$有变换

$$\underbrace{\widetilde{P'_i}_a}_{\text{实际像平面}} = \begin{bmatrix} \alpha & 0 & 0 & x_0 \\ 0 & \beta & 0 & y_0 \\ 0 & 0 & \gamma & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tilde{P}_i$$

定义$u = \widetilde{P'_i}_a(0), v = \widetilde{P'_i}_a(1)$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\alpha}{z_i} & 0 & x_0 \\ 0 & \frac{\beta}{z_i} & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\alpha}{z_i} & 0 & \frac{x_0}{z_i} \\ 0 & \frac{\beta}{z_i} & \frac{y_0}{z_i} \\ 0 & 0 & \frac{1}{z_i} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \frac{1}{z_i} \underbrace{\begin{bmatrix} \alpha & \overset{c\ \text{缩放因子}^5}{\boxed{0}} & x_0 \\ 0 & \beta & y_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Camera Intrinsics (内参数矩阵)}} \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \frac{1}{z_i} K p_i$$

同时我们的相机平面并不一定这么理想，它可能是各种姿态，我们需要先将初始的p_i^s 转换到可以求解

$$p_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tilde{p_i} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} T' \tilde{p_i^s} = T_{\{3,4\}} \tilde{p_i^s}$$

我们将这里的$T_{\{3,4\}}$称作相机的 Camera Extrinsics (外参数矩阵)。综上所述有

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z_i} K T \tilde{p_i^s} = K T \underbrace{\begin{bmatrix} \frac{x_i}{z_i} \\ \frac{y_i}{z_i} \\ 1 \end{bmatrix}}_{\text{位于归一化平面}}$$

1.6.1.3 畸变模型

我们先来简单介绍一下几何光学中的成像模型以及像差理论。[5]

实际的光学系统总是与理想的光学系统存在很大差异（如不同的孔径和视场），一个物点发出的光线经过实际光线汇聚后，其实不再汇聚于一点，而是一个**弥散斑**。对于单色光而言，从几何光学的角度分析，常见的像差有球差，惠差，场曲，畸变，像散。

其中，只有畸变影响几何形状，其他像差影响成像的清晰度（这个我们无能为力，选择好的镜头，合适的焦距，孔径，使用距离）。

畸变分为径向畸变（由透镜形状引起的畸变）和切向畸变（透镜和 CCD 不共面导致的畸变）

[5]**【缩放因子c】**代表图像轴u, v的的非垂直性,表示u, v的夹角偏离 90° 的情况。理想情况下，图像平面上的 x 轴和 y 轴是垂直的$c = 0$，但在某些特殊情况下，可能由于镜头设计或制造偏差，它们并非严格垂直。一般来说它等于 0，Opencv 中也是将其设置为 0 进行的标定。

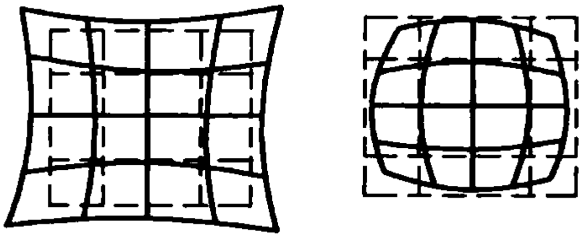

图 1.3: 径向畸变的几种类型(b)枕形畸变 (c)桶形畸变

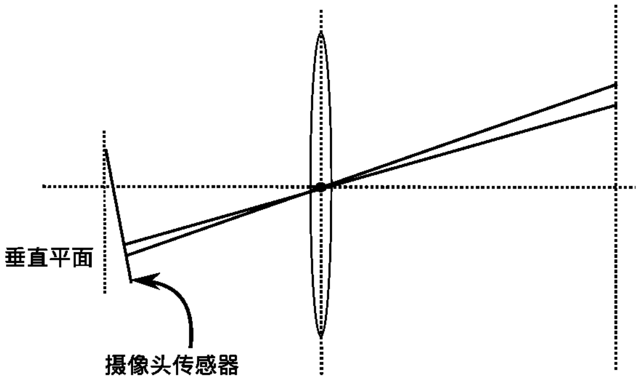

图 1.4: 切向畸变产生原因

我们将$(u,v)^T$表示成极坐标形式，即$(r,\theta)^T$。

对于径向畸变而言, 主要是随着r的增大, 畸变程度增大, 其矫正公式如下(使用k_1就足够)

$$x_{corrected} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$

$$y_{corrected} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$

对于切向畸变,引入p_1, p_2

$$x_{corrected} = x + 2p_1 xy + p_2(r^2 + 2x^2)$$

$$y_{corrected} = y + p_1(r^2 + 2y^2) + 2p_2 xy$$

然后通过相机标定标定相机内参和畸变系数即可。

1.6.1.4 双目相机

单个相机是无法确定z的, 通常我们采用双目相机或者 RGBD 相机来解决这个问题。双目相机的原理是显然的，对于单个相机，我们能确定图像上一点在相机坐标系下一条直线上，只要两个直线有交点，物点可知。

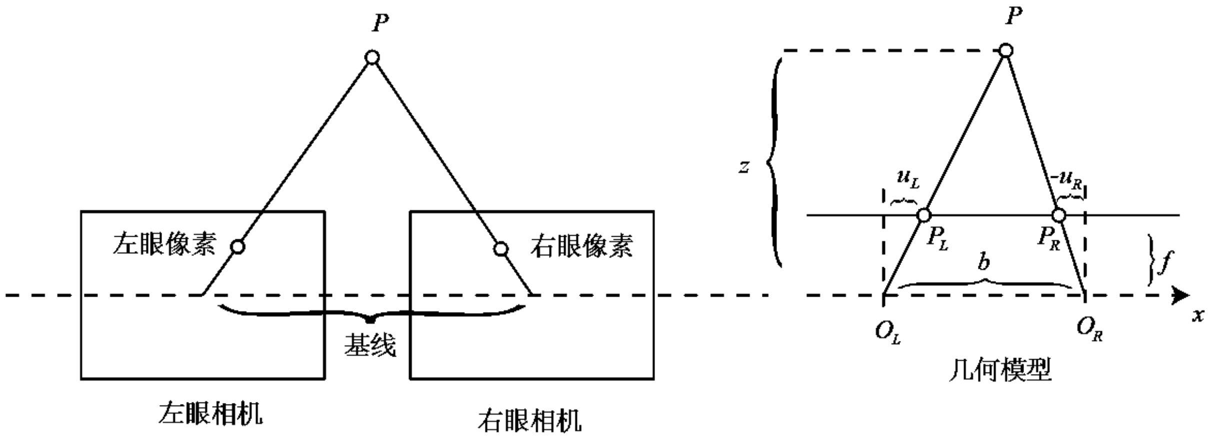

图 1.5: 双目相机原理

双目相机一般由左眼和右眼两个水平放置的相机组成，如上图所示。根据相似三角形原理，

$$\frac{z-f}{z} = \frac{b-U_L+U_R}{b} \Rightarrow z = \frac{fb}{d}, \quad d = u_L - u_R.$$

这里 d 为左右图的横坐标之差,称为视差(Disparity)，求解它需要知道左边相机像素在右边相机上对应哪一个位置。

1.6.2 李群和李代数

定义 1.6.2.1 (群)：一个集合G和一个二元运算$*$，如果满足以下条件，就称G是一个群。

1. 封闭性：对于任意$a, b \in G$，有$a * b \in G$
2. 结合律：对于任意$a, b, c \in G$，有$(a * b) * c = a * (b * c)$
3. 单位元：存在一个元素$e \in G$，使得对于任意$a \in G$，有$e * a = a * e = a$
4. 逆元：对于任意$a \in G$，存在一个元素$b \in G$，使得$a * b = b * a = e$

1.6.3 三维空间刚体变换

1.6.3.1 旋转矩阵

上一章节有详细的讲解，掠过。

1.6.3.2 旋转向量

旋转矩阵有九个量，但实际上作为正交矩阵，它只有三个自由度，并且还被约束为正交矩阵，有时会使求解变得困难。我们可以用旋转向量来表示旋转矩阵。

1.6.3.3 欧拉角

1.6.3.4 四元数

2 矩阵导数

2.1 基本介绍

从一个最基本的函数$f(x) = ax$ 开始，显然

$$\frac{\partial}{\partial x} f(x) = a$$

拓展 $f(x) = \sum_i a_i x_i = a^T x \quad \& \quad A = [a_1, a_2, ..., a_n]^T \quad X = [x_1, x_2, ..., x_n]$

$$\frac{\partial}{\partial \underset{n\text{维向量}}{x}} \overset{\text{标量}\sum_i^n a_i x_i}{f(x)} = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \cdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{bmatrix} = a \tag{2.2}$$

注意到这里$\frac{\partial}{\partial x} f(x)$的维度等于$x$的维度。那如果我们继续拓展呢? x是任意维度矩阵的时候，结果又会怎么样呢?

> **定理 2.1.1**：$\frac{\partial}{\partial X} f(x)$应和$X$的行列数相同（$f(x)$是标量的时候），并且
>
> $$\left[\frac{\partial}{\partial X} f(x)\right]_{ij} = \frac{\partial}{\partial X_{ij}} f(x)$$

$$\frac{\partial f}{\partial X^T} = \frac{\partial f}{\partial X}^T \tag{2.3}$$

并且我们继续分析式(2.2) 它表明，至少对于我们给出的例子 有$\frac{\partial a^T x}{\partial x} = a$，那对于所有呢？显然它也成立

> **定理 2.1.2**：如果存在一个多元标量函数$f(x) = \text{迹}(\mathrm{Tr})[A_T X]$，那么
>
> $$\left[\frac{\partial f}{\partial X}\right] = A$$

定理 2.1.3 (Tr 性质)：我们给出一些关于迹的性质，其中矩阵 A 和 B 的大小适当，且 c 是一个标量值。

1. $\mathrm{Tr}[A+B]=\mathrm{Tr}[A]+\mathrm{Tr}[B]$
2. $\mathrm{Tr}[cA]=c\,\mathrm{Tr}[A]$
3. $\mathrm{Tr}[AB]=\mathrm{Tr}[BA]=\sum_{i=1}^{n}\sum_{k=1}^{n}A_{ik}B_{ki}$
4. $\mathrm{Tr}[A^T]=\mathrm{Tr}[A]$
5. $\mathrm{Tr}[A_1A_2...A_n]=\mathrm{Tr}[A_nA_{n-1}...A_1]$
6. $\mathrm{Tr}[A^TB]=\sum A_{ij}B_{ij}$
7. $\mathrm{Tr}[A]=\sum$ 特征值 (λ)
8. $f(x)=\mathrm{Tr}[f(x)]=\mathrm{Tr}[f(x)^T]\quad\&\text{ When }\ f(x)_{\{1,1\}}$

证 *of* 定理 *2.1.3(3)* $\mathrm{Tr}[AB]=\mathrm{Tr}[BA]$:

$$AB_{ij}=\sum_{k=1}^{n}A_{ik}B_{kj}\Rightarrow AB_{ii}=\sum_{k=1}^{n}A_{ik}B_{ki}$$

$$\sum_{i=1}^{n}\sum_{k=1}^{n}A_{ik}B_{ki}=\sum_{i=1}^{n}\sum_{k=1}^{n}B_{ik}A_{ki}\tag{2.4}$$

$ik\ ki$的情况都会出现，所以显然式(2.4) 成立，证毕。 ■

2.2 标量矩阵导数$f(x)$

2.2.1 矩阵微分

定义 2.2.1.1:

$$dX=\begin{bmatrix}dX_{11} & dX_{12} & ... & dX_{1n}\\ dX_{21} & dX_{22} & ... & dX_{2n}\\ ... & ... & ... & ...\\ dX_{11} & dX_{12} & ... & dX_{1n}\end{bmatrix}\tag{2.5}$$

矩阵里面的每一个如dX_{11},其实

定理 2.2.1.1:

$$d\,\mathrm{Tr}[A]=\mathrm{Tr}[dA]$$

证:

$$\mathrm{Tr}[dA] = \sum_i dA_{ii} = \sum_i dA_{ii} = d\sum_i A_{ii} = d\,\mathrm{Tr}[A]$$

■

标量形式的微分和导数可以用下式表示

$$df = \frac{\partial f}{\partial x}dx$$

对于矩阵，有如下定理

定理 2.2.1.2：

$$df(X) = \mathrm{Tr}\left[\frac{\partial f}{\partial X}^T dX\right]^6 \tag{2.6}$$

$$f(x) = \mathrm{Tr}[f(x)] = \mathrm{Tr}[f(x)^T] \tag{2.7}$$

证：For 式(2.6)

$$df(X) = \sum_{ij}\frac{\partial f}{\partial X_{ij}}dX_{ij}$$

使用式(2.4)

$$\begin{aligned}\mathrm{Tr}\left[\frac{\partial f}{\partial X}^T dX\right] &= \sum_i\sum_k\left[\left(\frac{\partial f}{\partial X}^T\right)_{ik}(dX)_{ki}\right] \\ &= \sum_i\sum_k\left[\left(\frac{\partial f}{\partial X}\right)_{ki}(dX)_{ki}\right]\end{aligned}$$

For 式(2.7),只需要意识到$f(x)_{\{1,1\}}$即可。证毕。 ■

定理 2.2.1.3：对于矩阵微分，有如下几个性质及延伸性质

1. $d(A+B) = d(A) + d(B)$
2. $d(cA) = cd(A)$
3. $d(AB) = d(A)B + Ad(B)$

[6]这里的$f(X)$, f小写，应该为标量函数

4. $d(X^T) = d(X)^T$
5. $\frac{\partial x^T Ax}{\partial x} = Ax + A^T x$
6. $\mathrm{d}(X^{-1}) = -X^{-1}\mathrm{d}XX^{-1}$ (if X^{-1}exists)

证 *of 3* $\boldsymbol{d(AB) = d(A)B + Ad(B)}$:

$$\begin{aligned} \text{LEFT} = df(AB) &= d\sum_k A_{ik}B_{ki} \\ &= \sum_k d(A_{ik})B_{ki} + A_{ik}d(B_{ki}) \end{aligned}$$

$$\text{RIGHT}=[d(A)B + Ad(B)]_{ij} = \sum_k d(A)_{ik}B_{ki} + \sum_k d(B)_{ik}A_{ki} = \text{LEFT}$$

■

证 *of 4:* $\boldsymbol{d(X^T) = d(X)^T}$: 使用式(2.5) ■

证 *of 5:* $\boldsymbol{\frac{\partial x^T Ax}{\partial x} = Ax + A^T x}$:

$$\begin{aligned} d(x^T Ax) &= d\,\mathrm{Tr}[x^T Ax] \\ &= \mathrm{Tr}[d(x^T)Ax + x^T d(Ax)] \\ &= \mathrm{Tr}\Big[d(x)^T Ax + x^T \underbrace{d(A)}_{0}\, x + x^T Ad(x)\Big] \\ (\mathrm{Tr}[A+B] = \mathrm{Tr}[A] + \mathrm{Tr}[B]) &= \mathrm{Tr}[d(x)^T Ax] + \mathrm{Tr}[x^T Ad(x)] \\ (\mathrm{Tr}(A^T) = \mathrm{Tr}[A]) &= \mathrm{Tr}[x^T A^T d(x)] + \mathrm{Tr}[x^T Ad(x)] \\ \text{Then} \Rightarrow \frac{\partial x^T Ax}{\partial x} &= (x^T A^T + x^T A)^T \\ &= Ax + Ax^T \end{aligned}$$

■

证 *of 6* $dX^{-1} = -X^{-1}dXX^{-1}$:

$$\begin{aligned} d(I) = 0 = d(XX^{-1}) &= d(X)X^{-1} + Xd(X^{-1}) \\ d(X^{-1}) &= -X^{-1}d(X)X^{-1} \end{aligned}$$

■

2.2.2 标量对矩阵求导的求解过程

从上面的例子中，我们可以得到对于一个$f(x)$求解$\frac{\partial f(x)}{\partial x}$的固定步骤

$df(x) = d\,\mathrm{Tr}[f(x)] = \mathrm{Tr}[df(x)]$ or $df(x) = d\,\mathrm{Tr}[f(x)^T] = \mathrm{Tr}[df(x)^T]$
见式(2.7)
使用定理 2.1.3 中关于 trace 的各种性质，化简成$\sum_i \mathrm{Tr}[U_i d(x)]$的形式
如式(2.6)，所以$\dfrac{\partial f(x)}{\partial x} = \sum_i (U_i)^T$

3 最优化理论

所谓最优化问题，就是求解 $\min f(x)\ \&\ x \in K$

3.1 最小二乘的几种解法

$$\min F(x) = \|f(x)\|_2^2 = \|b - A\hat{x}\|$$

3.1.1 牛顿法

$$\begin{aligned} f(x) &= f(x_0)(x-x_0)^0 + f'(x_0)(x-x_0) + \left(\frac{f''(x_0)}{2!}\right)(x-x_0)^2 + \dots \\ &\approx f(x_0)(x-x_0)^0 + f'(x_0)(x-x_0) + \frac{f''(x_0)(x-x_0)^2}{2} \end{aligned} \tag{3.8}$$

$$g(x) = \frac{\mathrm{d}}{\mathrm{d}x} f(x) = f'(x_0) + f''(x_0)(x-x_0) = g(x)$$

$g(x) = 0, x = x_1$的时候是极值,

$$x_1 - x_0 = \frac{f'(x)}{f''(x)} \Rightarrow x_1 = x_0 - \frac{f'(x)}{f''(x)}$$

3.1.2 梯度下降法

找到可微函数的局部最小值

给定**初始值** $x_0, k = 0$
while
  计算出$\frac{\partial f}{\partial x}$, cost
  $x = x - \alpha\left(\frac{\partial f}{\partial x}\right)$
  if cost $< \varepsilon$
    break
end

$$
\begin{aligned}
&\theta(j)' = \theta(j) - \alpha\frac{\partial}{\partial x_j}f(x) \ \& \ \text{每一个参数都更新完再下一步} \\
&\theta(j) = \theta(j)'
\end{aligned}
\tag{3.9}
$$

容易走出锯齿路线，从而增加迭代次数。

3.1.3 高斯牛顿法

$$
f(x_k + \Delta x) \approx f(x_k) + J(x_k)^T \Delta x
$$

我们的目标函数可以近似为:

$$
min\frac{1}{2} \parallel f(x_k) + J(x_k)^T \Delta x \parallel_2^2
$$

现在，我们将近似的目标函数展开:

$$
\begin{aligned}
M(\Delta X) = \frac{1}{2}f(x+\Delta X)^2 &= \frac{1}{2}\left(f(x) + J^T\Delta X\right)^T\left(f(x) + J^T\Delta X\right) \\
&= \frac{1}{2}\left[\underbrace{f(x)^T + \Delta X^T J}_{\text{E}}\right]\left[\underbrace{f(x) + J^T\Delta X}_{\text{F}}\right] \\
&= \frac{1}{2}\left[\underbrace{\|f(x)\|^2}_{\text{A}} + \underbrace{f(x)^T J^T \Delta X}_{\text{B}} + \underbrace{\Delta X^T J f(x)}_{\text{C}} + \underbrace{\Delta X^T J J^T \Delta X}_{\text{D}}\right]
\end{aligned}
\tag{3.10}
$$

$$
\begin{aligned}
dC &= d\,\mathrm{Tr}[f(x)^T J^T \Delta X] = \mathrm{Tr}[d(f(x)^T J^T \Delta X)] \\
&= \mathrm{Tr}[f(x)^T J^T d\Delta X] \Rightarrow \frac{\partial C}{\partial \Delta X} = J f(x)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial M(\Delta X)}{\partial \Delta X} &= \frac{1}{2}\left(0 + Jf(X) + Jf(x) + \overbrace{2(JJ^T\Delta X)}^{\text{定理 2.2.1.3(5)}}\right) \\
&= Jf(x) + JJ^T\Delta X = 0
\end{aligned}
$$

$$
H\Delta X = g \ \& \ H = JJ^T \ \& \ g = -Jf(x)
$$

所以高斯牛顿法的步骤是

给定**初始值** $x_0, k = 0$
while
  计算出$J(x_k), f(x_k) \ \& \ H = JJ^T \ \& \ g = -Jf(x)$
  使用$H\Delta X = g$求出Δx
  if $\Delta x < \varepsilon$
    break
  else
    $x_{k++} = x_k + \Delta x$

end

它的缺点是要求H矩阵可逆，而且计算量大，有时候可能无解。并且Δx过大会导致其局部近似不精确，严重的时候，可能无法保证迭代收敛。同时也会锯齿状增大迭代次数（和梯度下降一样）。

3.1.3.1 列文伯格-马夸特法(LM)

为了避免其迭代次数过长的缺点，在高斯牛顿的基础上进行优化，提出一个信赖区域。

$$\rho = \frac{f(x+\Delta x) - f(x)}{J^T \Delta x}$$

如果它接近一就不需要更改，如果过大就需要缩小步长，如果过小就需要增大步长，这样的话，就可以动态调整步长了。最优化问题变为

$$\min \frac{1}{2} \left\| f(x) + J^T \Delta x \right\|_2^2 \;\; s.t \;\; \|D\Delta x\|_2^2 < \mu$$

构建拉格朗日函数[7]

$$L(\Delta x, \lambda) = \frac{1}{2} \left\| f(x) + J^T \Delta x \right\| + \lambda(\|D\Delta x\|_2^2 - \mu)$$

$$\frac{\partial L(\Delta x, \lambda)}{\partial \Delta x} = Jf(x) + JJ^T \Delta X + \lambda \frac{\partial(\Delta x^T D^T D \Delta x)}{\partial \Delta x} = Jf(x) + JJ^T \Delta x + \lambda D^T D \Delta x$$

之前的H变为$JJ^T + \lambda D^T D$,求解步骤变为

给定**初始值** $x_0, k = 0$
while
  计算出$J(x_k), f(x_k)$ & $H = JJ^T + \lambda D^T D$ & $g = -Jf(x)$
  计算出Δx
  计算$\rho_k = \frac{f(x+\Delta x) - f(x)}{J^T \Delta x}$
  if $\rho_k < \frac{1}{4}$
    $\Delta x_k = \frac{1}{4} \Delta x_k$
  else
    if $\rho_k > \frac{3}{4}$
      $\Delta x_k = \min(2\Delta x_k, \mu)$
    else
      $\Delta x_k = \Delta x_k$

[7] 目标函数为 f 在约束条件 g 下的极值与其拉格朗日函数的极值相同。λ 被称为拉格朗日算子

if $\rho_k > \xi$

  $x_{k+1} = x_k + \Delta x$

else

  $x_{k+1} = x_k$

k=k+1

if $\Delta x_k < \varepsilon$

  break

end

3.1.3.2 随机梯度下降法

3.1.4 Example

4 概率论+随机过程

4.1 随机变量

定义 4.1.1: 定义从 n 中取出 r 个元素的排列数为$C_n^r = \dfrac{n!}{(n-r)!}r!$

定义 4.1.2 (样本空间、随机事件、概率频率，常见公理):

名称	**定义**	注释
样本空间Ω	全部事件的空间	**所有可能的结果**
事件A	样本空间的**子集**	**可能发生的事件**
概率P	事件发生的可能性	**事件发生的可能性**
事件独立	$P(\Pi A_i) = \Pi P(A_i)$	
条件概率	$P(A\|B) = P(AB)/P(B)$	

一些基本概率公理

$$\text{(全概率公式)}\ P(A) = \sum_{i=1}^{n} P(AB_i)\ \ \&\ \ B_i\ \text{是样本空间的分割}$$

$$\text{(乘法公式)}\ P(AB) = P(A|B)P(B)$$

$$\text{(贝叶斯公式)}\ P(B_i|A) = \frac{P(AB_i)}{P(A)} = P(B_i|A) = \frac{P(AB_i)}{\sum_{i=1}^{n} P(AB_i)} = \frac{P(AB_i)}{P(A|B_i)P(B_i)}$$

常见的分布如下

5 最优化

5.1 拉格朗日乘数法

一个函数的极值可以通过驻点来求解，驻点不一定是极值点，有可能是鞍点。

5.2 PCA

先推导二维的 PCA

计算一维数据的方差

$$s^2(X)=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

对于二维数据有协方差

$$\operatorname{cov}(X,Y)=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

协方差矩阵 C 其实是$(X_1=X-\bar{X})$

$$C=\frac{1}{n-1}X_1^TX_1$$

定义一个投影方向$v=(x_0,y_0)$ st $x_0^2+y_0^2=1$

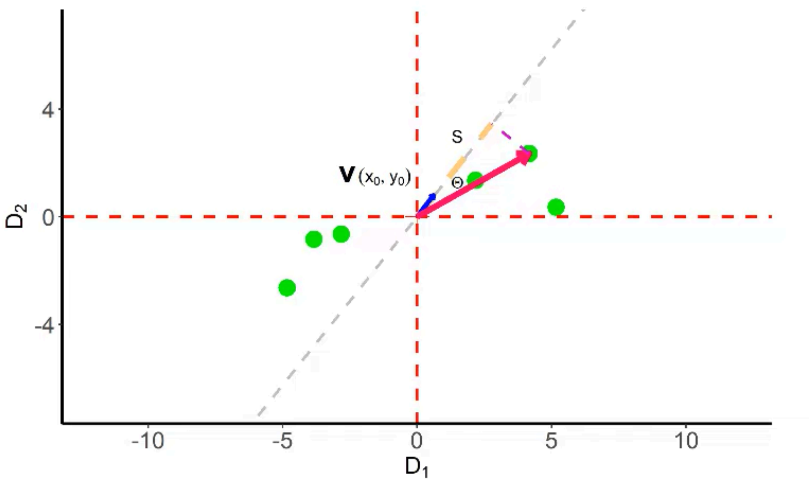

图 5.6: 示意图

$$S=\vec{v}\vec{x}=|\vec{x}|\cos\theta$$

在这个方向投影后的长度的方差其实可以表示为

$$\frac{1}{n-1}\sum S^2=\frac{1}{n-1}(\vec{v}X_1^T)(\vec{v}X_1^T)^T=\vec{v}\frac{X_1^TX_1}{n-1}\vec{v}^T=\vec{v}C\vec{v}^T$$

最优化方差最小值

$$J=\vec{v}C\vec{v}^T\quad\text{st}\quad\vec{v}\vec{v}^T=1$$

$$F(\vec{v})=\vec{v}C\vec{v}^T-\lambda(1-\vec{v}\vec{v}^T)$$

$$\frac{\partial}{\partial v}F(\vec{v}) = 0 \Rightarrow 2Cv^T - 2\lambda v^T = 0$$

$$\left(\frac{\partial}{\partial v}f(x)v = f(x)^T, \frac{\partial}{\partial v}A^TvA = Av + A^Tv\right), C = C^T$$

C 是实对称矩阵并且使用矩阵求导定理 2.2.1.3

$$Cv^T = \lambda v$$

对 C 做特征值分解，得到两个特征向量 PC1,PC2

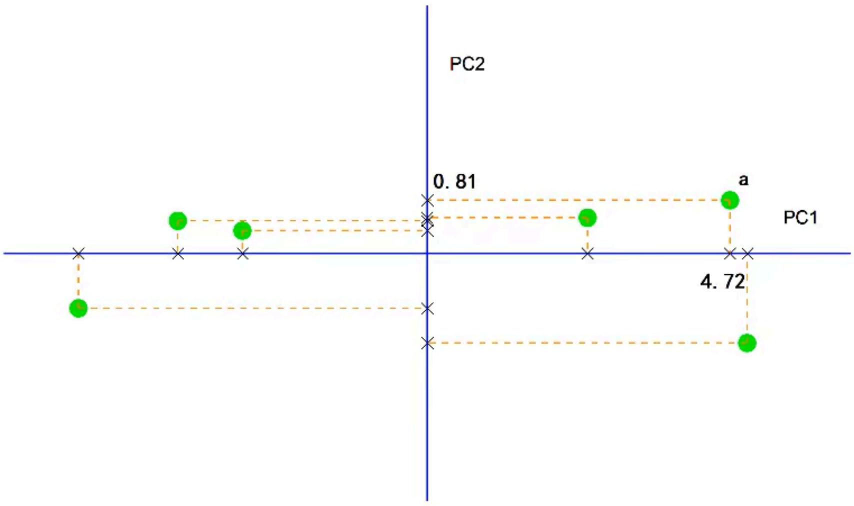

图 5.7: PC1 PC2 示意图

TODOs

6 一些其他概率论知识

参考文献

1. 杨明 & 刘先忠. 矩阵论. (华中科技大学出版社, 2003).
2. Strang, G. *Linear Algebra and Its Applications 4th ed.* (2012).
3. Akritas, A. G. & Malaschonok, G. I. Applications of Singular-Value Decomposition (SVD). *Mathematics and Computers in Simulation* **67,** 15–31 (2004)
4. 维基百科. 正交矩阵. https://zh.wikipedia.org/w/index.php?title=%E6%AD%A3%E4%BA%A4%E7%9F%A9%E9%98%B5&oldid=84256160 (2024)
5. 李湘宁, 贾宏志, 张荣福 & others. 工程光学. 108–110 (科学出版社, 2022).