关于矛盾方程组的最小二乘解

一般情况下，实数域上的线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1s}x_s = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2s}x_s = b_2 \\ \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{ns}x_s = b_n \end{cases}$$

很有可能无解，这时候这就是一个**矛盾方程组**，我们找不到一组$x_1, x_2, \cdots, x_s$使得

$$d \triangleq \sum_{i=1}^{n}(a_{i1}x_1 + \cdots + a_{is}x_s - b_i)^2 = 0$$

但是我们可以试着找一组$x_1', x_2', \cdots, x_s'$使得d尽可能小（最小），这时候$x_1', x_2', \cdots, x_s'$就使该方程组的**最小二乘解**.

令

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1s} \\ \vdots & & \\ a_{n1} & \cdots & a_{ns} \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_s \end{pmatrix}, \boldsymbol{Y} \triangleq \boldsymbol{AX} = \begin{pmatrix} \sum_{i=1}^{s} a_{1i}x_i \\ \vdots \\ \sum_{i=1}^{s} a_{ni}x_i \end{pmatrix}$$

这样$d = |Y - b|^2$就是欧式空间$V = \mathbb{R}^n$中向量Y和b的距离的平方，现在要做的就是找$X = X^0$，使得Y与b的距离最短，此时可记原方程组为$AX = b$.

令$A = (\alpha_1, \cdots, \alpha_s)$，其中$\alpha_1, \cdots, \alpha_s$是列向量，则

$$Y = AX = x_1\alpha_1 + \cdots + x_s\alpha_s \in L(\alpha_1, \cdots, \alpha_s) \triangleq W$$

设$Y^0 \in W$是b在W上的投影，则当$Y = Y^0$时Y与b的距离最短，也就是使得$Y^0 = AX^0$成立的X^0就是我们要找的方程组的最小二乘解.

我们知道Y^0是一定存在的，又由$Y^0 \in W$可以知道X^0也是一定存在的，要注意的是，虽然Y^0是唯一的，但X^0却未必是唯一的.

$$\begin{aligned} & (b - Y^0) \perp W \\ \iff & (b - Y^0) \perp \alpha_i (i = 1, \cdots, s) \\ \iff & \alpha_i^T (b - Y^0) = 0 \\ \iff & A^T (b - Y^0) = 0 \\ \iff & A^T Y^0 = A^T b \\ \iff & A^T A X^0 = A^T b \end{aligned}$$

由此，如果X^0是$AX = b$的一个最小二乘解，X^0应该是$A^TAX = A^Tb$的一个解.

称方程组

$$A^TAX = A^Tb$$

是由$AX = b$导出的**正规方程组**.

可以用线性方程组的求解理论证明正规方程组$A^TAX = A^Tb$总是有解的，即$AX = b$的最小二乘解总是存在的.

令$L(A^T)$和$L(A^TA)$分别表示A^T和A^TA的列空间，那么总有

$$L(A^TA)\subseteq L(A^T)$$

对于一个实矩阵A，总有$r(A^TA)=r(A)$，于是

$$\dim L(A^TA)=r(A^TA)=r(A)=r(A^T)=\dim L(A^T)$$

所以

$$L(A^TA)=L(A^T)$$

于是

$$A^Tb\in L(A^T)=L(A^TA)$$

也就是存在X^0使得$A^TAX^0=A^Tb$.

当A是可逆方阵得时候，线性方程组$AX=b$总有唯一的解$X=A^{-1}b$，当A为非可逆阵的时候，$AX=b$有解的充要条件为$r(A\ b)=r(A)$，这时候，是否也有某个矩阵使得解可以表示为$X=Gb$的形式呢?

定义 对于数域$\mathbb{P}$上的$m\times n$阶矩阵A，若有$n\times m$阶矩阵G，使得

$$AGA=A$$

则称G是A的一个**广义逆矩阵**，简称**广义逆**.

如果A的秩为r，令

$$A=P\begin{pmatrix}E_r & O\\ O & O\end{pmatrix}Q$$

其中P,Q分别为$m\times m,n\times n$阶的可逆阵，则A的所有广义逆为

$$G=Q^{-1}\begin{pmatrix}E_r & C\\ D & F\end{pmatrix}P^{-1}$$

其中C,D,F分别为任意的$r\times(n-r),(m-r)\times r,(m-r)\times(n-r)$阶矩阵.

$$\begin{aligned}AGA&=P\begin{pmatrix}E_r & O\\ O & O\end{pmatrix}QQ^{-1}\begin{pmatrix}E_r & C\\ D & F\end{pmatrix}P^{-1}P\begin{pmatrix}E_r & O\\ O & O\end{pmatrix}Q\\&=P\begin{pmatrix}E_r & O\\ O & O\end{pmatrix}\begin{pmatrix}E_r & C\\ D & F\end{pmatrix}\begin{pmatrix}E_r & O\\ O & O\end{pmatrix}Q\\&=P\begin{pmatrix}E_r & O\\ O & O\end{pmatrix}Q\\&=A\end{aligned}$$

所以所有这样形式的G都是A的广义逆.

设G是A的一个广义逆，而且

$$QGP=\begin{pmatrix}B & C\\ D & F\end{pmatrix}$$

其中B是$r\times r$阶方阵，F是$(m-r)\times(n-r)$阶矩阵，于是

$$
\begin{aligned}
AGA &= P\begin{pmatrix} E_r & O \\ O & O \end{pmatrix} QQ^{-1}\begin{pmatrix} B & C \\ D & F \end{pmatrix} P^{-1}P\begin{pmatrix} E_r & O \\ O & O \end{pmatrix} Q \\
&= P\begin{pmatrix} B & O \\ O & O \end{pmatrix} Q
\end{aligned}
$$

由

$$
A = P\begin{pmatrix} E_r & O \\ O & O \end{pmatrix} Q,\ AGA = A
$$

可以知道$B = E_r$，进而有

$$
G = Q^{-1}\begin{pmatrix} E_r & C \\ D & F \end{pmatrix} P^{-1}
$$

也就是，A的所有广义逆都有这样的形式.

现在用广义逆给出方程组有解的充要条件和通解公式.

引理 设A为$m \times n$阶矩阵，G为$n \times m$阶矩阵，则$AGA = A$当且仅当对于任一列向量X_0以及$b = AX_0$，有$AGb = b$.

证 如果有$AGA = A$，很显然有$AGAX_0 = AX_0$，也就是$AGb = b$.

反过来，如果有$AGb = b$，就有$(AGA - A)X_0 = O$，特别地，令

$$
X_0 = \alpha_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} (i = 1, \cdots, n)
$$

那么就有

$$
AGA - A = (AGA - A)E = \sum_{i=1}^{n}(AGA - A)\alpha_i = O
$$

（这里其实并不是求和，只是将$\{\alpha_i\}$按顺序并置构成了单位矩阵E，同时n个$O_{n\times 1}$也并置成了$O_{n\times n}$。）

引理 如果G是A的广义逆，那么线性方程组有解当且仅当$AGb = b$.

证 当$AGb = b$时，$X = Gb$就是$AX = b$的解.

反之，如果$X = X_0$是$AX = b$的解，因为G是A的广义逆，有$AGb = b$.

定理 设G是$m \times n$阶矩阵A的一个广义逆，则当$AX = b$有解时，通解可以表示为

$$
X = Gb + (E_n - GA)Y
$$

其中Y是任意一个n维列向量.

证 既然$AX = b$有解，就有$AGb = b$，于是

$$
A(Gb + (E_n - GA)Y) = AGb + (A - AGA)Y = b
$$

也就是$X = Gb + (E_n - GA)Y$是$AX = b$的解.

另一方面，如果$X = X'$是$AX = b$的解，有

$$
X' = Gb + X' - Gb = Gb + X' - GAX' = Gb + (E_n - GA)X'
$$

特别地，当$b=O$的时候，有

推论 如果G是A的广义逆，那么$AX=O$的通解可以表示为

$$X=(E_n-GA)Y$$

其中Y是任意n维列向量.

我们已经知道，当$AX=b$为矛盾方程组的时候，$AX=b$的最小二乘解为$A^TAX=A^Tb$的解，这时候，利用广义逆，最小二乘解可以表示为

$$X=GA^Tb+(E_n-GA^TA)Y$$

其中G是A^TA的广义逆，Y是任意向量.

一种特殊的广义逆

设A是复数域$\mathbb{C}$上的$m\times n$阶矩阵，如果有$\mathbb{C}$上的$n\times m$阶矩阵满足

(1) $AGA=A$

(2) $GAG=G$

(3) $(AG)^H=AG$

(4) $(GA)^H=GA$

则称G为A的Moore-Penrose逆，简称M-P逆.

很明显G是A的广义逆，也就是说，M-P逆是比广义逆强的概念，而且M-P还具有对称性，如果G是A的广义逆，那么A也是G的广义逆.

设$A\in\mathbb{C}_{m\times n}$的秩为$r$，取$\beta_1,\beta_2,\cdots,\beta_r$为$A$的列极大线性无关组，则

$$A=(\alpha_1,\alpha_2,\cdots,\alpha_n)$$

的任一列可以表示为$\beta_1,\beta_2,\cdots,\beta_r$的线性组合，即

$$\alpha_i=\sum_{j=1}^{r}c_{ji}\beta_j\quad(c_{ij}\in\mathbb{C},\mathrm{i}=1,\cdots,\mathrm{n})$$

于是有

$$A=(\alpha_1,\alpha_2,\cdots,\alpha_n)=(\beta_1,\beta_2,\cdots,\beta_r)\begin{pmatrix}c_{11}&\cdots&c_{1n}\\ \vdots\\ c_{r1}&\cdots&c_{rn}\end{pmatrix}=BC$$

其中

$$B=(\beta_1,\beta_2,\cdots,\beta_r),\ C=\begin{pmatrix}c_{11}&\cdots&c_{1n}\\ \vdots\\ c_{r1}&\cdots&c_{rn}\end{pmatrix}$$

由定义$r(B)=r$，又因为

$$r = r(A) = r(BC) \leq r(C) \leq \min\{r, n\} \leq r$$

有$r(C) = r$，于是，分解式

$$A = BC$$

中，B和C分别是列满秩矩阵，和行满秩矩阵，这样的分解被称为**满秩分解**.

比较容易地，可以得到

$$r(B^H B) = r(C^H C) = r$$

因此$B^H B, C^H C$都是r阶可逆方阵，令

$$G = C^H (CC^H)^{-1} (B^H B)^{-1} B^H$$

可以验证

$$\begin{aligned}
AGA &= BCC^H (CC^H)^{-1} (B^H B)^{-1} B^H BC = BC = A \\
GAG &= C^H (CC^H)^{-1} (B^H B)^{-1} B^H BCC^H (CC^H)^{-1} (B^H B)^{-1} B^H \\
&= C^H (CC^H)^{-1} (B^H B)^{-1} B^H = G \\
GA &= C^H (CC^H)^{-1} (B^H B)^{-1} B^H BC = C^H (CC^H)^{-1} C = (GA)^H \\
(AG)^H &= G^H A^H = B (B^H B)^{-1} (CC^H)^{-1} CC^H B^H \\
&= B^H (BB^H)^{-1} B = AG \\
AG &= BCC^H (CC^H)^{-1} (B^H B)^{-1} B^H = B^H (BB^H)^{-1} B = (AG)^H
\end{aligned}$$

因此，G是A的M-P逆.

设X是A的另一个M-P逆，则有

$$\begin{aligned}
X &= XAX = X(AX)^H = XX^H A^H = XX^H (AGA)^H \\
&= XX^H A^H (AG)^H = X(AX)^H (AG)^H = XAXAG \\
&= XAG = X(AGA)G = XAGAG = (XA)^H (GA)^H G \\
&= (GAXA)^H G = (GA)^H G = GAG \\
&= G
\end{aligned}$$

所以，A的M-P逆是唯一的,

定理 设$A \in \mathbb{C}_{m \times n}$，$r(A) = r$，$A = BC$是$A$的满秩分解，则$A$有唯一的M-P逆

$$A^+ = C^H (CC^H)^{-1} (B^H B)^{-1} B^H$$

注意，当$A \in \mathbb{R}_{m \times n}$时，$A^+$存在且存在于$\mathbb{R}_{n \times m}$中，而且，与可逆阵运算不同的是，一般$(AB)^+ \neq B^+ A^+$.

现在来讨论如何用M-P逆给出矛盾方程组的最小二乘解的公式刻画.

现在只在实数域上讨论矛盾方程组的最小二乘解问题，所以有$A^H = A^T$.

定理 设$A \in \mathbb{R}_{m \times n}$，则有

（i）$AX = b$的最小二乘解的通式可以表示为

$$X = A^+ b + (E_n - A^+ A) Y$$

其中Y是任一n维实向量.

（ii）A^+b是唯一的极小最小二乘解，即作为向量长度在最小二乘解中是最小的.

证 （i）$X=X_0$是$AX=b$的最小二乘解当且仅当X_0是正规方程组$A^TAX=A^b$的解，于是有

$$X_0=(A^TA)^+A^Tb+(E_n-(A^TA)^+(A^TA))Y=A^+b+(E_n-A^+A)Y$$

其中利用了$A^+=(A^TA)^+A^T$，Y是任一n维实向量.

（ii）当$Y=O$时，由（i）即得$X_0=A^+b$是$AX=b$的一个最小二乘解，与任一最小二乘解比较

$$X-X_0=(E_n-A^+A)Y$$

于是有

$$\begin{aligned}(X-X_0,X_0)&=X_0^T(X-X_0)=(A^+b)^T(E_n-A^+A)Y\\&=b^T(A^+)^TY-b^T(A^+)^TA^+AY\\&=b^T(A^+)^TY-b^T(A^+AA^+)^TY\\&=b^T(A^+)^TY-b^T(A^+)^TY=0\end{aligned}$$

即$X_0\perp(X-X_0)$，由勾股定理，有

$$|X|^2=|X_0+(X-X_0)|^2=|X_0|^2+|X-X_0|^2\geq|X_0|^2$$

从而X_0是最小二乘解中长度最小的.

假设X_0'是另一个极小最小二乘解，由极小性的定义，$|X_0|=|X_0'|$，又由（i），存在向量Y_0使得

$$X_0'=X_0+(E_n-A^+A)Y_0$$

并且$X_0\perp(E_n-A^+A)Y_0$，由此得$X_0=X_0'$，也就是说，极小的最小二乘解是唯一的.

需要注意的是，广义逆是对任一数域讨论的，而M-P逆是针对复数域的，最小二乘解理论则只针对实数域建立.

事实上，最小二乘解理论可以在复数域上对矛盾方程组进行类似的研究，并得到类似的结论.

由最小二乘解的定义，对$AX=b$的任一最小二乘解X_1，AX_1是不变的，而且

$$|AX_1-b|=\min\{|AX-b|:X\in\mathbb{R}_n\}$$

这个值越小，说明X_1越接近$AX=b$的一般解，即体现了最小二乘解的“误差”，称AX_1-b为方程组$AX=b$的**残差向量**，$|AX_1-b|$为残差.

当X_1是$AX=b$的解时，残差向量和残差退化为零向量和代数零.