最小二乘解与高斯消元法的注记

胡泓昇 [1,2]

（1. 中国科学院 数学与系统科学研究院，北京 100190；2. 北京大学 北京国际数学研究中心，北京 100871）

［**摘　要**］若在对超定线性方程组求最小二乘解的过程中使用高斯消元法（即对增广的系数矩阵作初等行变换），有时会得到正确的最小二乘解，但大多数情况下是错误的。探讨了高斯消元法和最小二乘解的一些关系。建议授课时提醒学生注意最小二乘法和求解一般的线性方程组的区别。

［**关键词**］超定线性方程组；最小二乘解；高斯消元法

［**中图分类号**］O151.26　　［**文献标识码**］C　　［**文章编号**］？？？？？？

1　引　　言

设 $\boldsymbol{A}\in M_{m\times n}(\mathbb{R})$ 为矩阵，$\boldsymbol{b}\in\mathbb{R}^m$ 为列向量，$\boldsymbol{x}=(x_1,\dots,x_n)^{\mathrm{T}}$ 为 $\mathbb{R}^n$ 中的变量，这里上标 $^{\mathrm{T}}$ 表示转置。考虑关于 $\boldsymbol{x}$ 的线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$，若 $\operatorname{rank}\boldsymbol{A}=\operatorname{rank}(\boldsymbol{A}\,\vdots\,\boldsymbol{b})$，方程组有解，可以通过对增广矩阵 $(\boldsymbol{A}\,\vdots\,\boldsymbol{b})$ 作初等行变换来简化并求解方程组，这称为高斯消元法。但如果 $\operatorname{rank}\boldsymbol{A}<\operatorname{rank}(\boldsymbol{A}\,\vdots\,\boldsymbol{b})$，方程组无解，这时希望找到 $\boldsymbol{x}\in\mathbb{R}^n$ 使得欧氏范数 $\|\boldsymbol{Ax}-\boldsymbol{b}\|$ 最小，这样的 $\boldsymbol{x}$ 称为超定线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的最小二乘解。把 $\boldsymbol{A}$ 视为线性变换，记 $\boldsymbol{d}$ 为 $\boldsymbol{b}$ 在 $\operatorname{Im}\boldsymbol{A}$ 上的正交投影。内积空间的理论告诉我们，最小二乘解全体就是 $\boldsymbol{Ax}=\boldsymbol{d}$ 的通解，也是方程组 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$（称为法方程）的通解。

既然高斯消元法可以简化线性方程组并且不改变它的解，那么对初学者来说，利用此法来简化最小二乘解的计算也是很“自然”、“合理”的。但高斯消元法会保持目标函数 $\|\boldsymbol{Ax}-\boldsymbol{b}\|$ 或最小二乘解吗？作者翻阅了许多线性代数的教材，如[1]、[2]等等，都未提及该问题。

2022 年春，作者在国科大担任线性代数课程的助教时，主讲老师布置如下的测验题：

例题 1　求下面方程组的最小二乘解：

$$\begin{pmatrix}1&-1&1\\1&-1&2\\-1&1&-1\\-1&1&-2\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}1\\-1\\0\\1\end{pmatrix}。$$

解　求得法方程 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ 为

［**收稿日期**］2022-07-20；［**修改日期**］2022-07-27

［**基金项目**］NSFC No.12171457

［**作者简介**］胡泓昇（1994-）,男，博士，现为北京大学北京国际数学研究中心博士后，从事表示论研究。Email: huhongsheng16@mails.ucas.ac.cn

$$\begin{pmatrix} 4 & -4 & 6 \\ -4 & 4 & -6 \\ 6 & -6 & 10 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}=\begin{pmatrix} -1 \\ 1 \\ -3 \end{pmatrix},$$

解得法方程的通解（即原方程组的最小二乘解）为

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}=\begin{pmatrix} 2 \\ 0 \\ -3/2 \end{pmatrix}+\lambda\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}，\ \lambda\in\mathbb{R}。$$

以上便是正确答案。但是，有一些同学的答卷是下面这样的：

观察到题目中第二个和第四个方程只相差一个正负号，因此可以去掉第四个方程（或者等价地，把增广矩阵的第二行加到第四行上），得到一个新的方程组 $\boldsymbol{Bx}=\boldsymbol{c}$ 如下，

$$\begin{pmatrix} 1 & -1 & 1 \\ 1 & -1 & 2 \\ -1 & 1 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}=\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix},$$

求得法方程 $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Bx}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{c}$ 为

$$\begin{pmatrix} 3 & -3 & 4 \\ -3 & 3 & -4 \\ 4 & -4 & 6 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix},$$

求解该法方程，即得题中所求的最小二乘解为

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}=\begin{pmatrix} 2 \\ 0 \\ -3/2 \end{pmatrix}+\lambda\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}，\ \lambda\in\mathbb{R}。$$

所得结果和正确答案完全相同！这是巧合还是一般性的事实？如果是巧合，出现这种“巧合”的原因又是什么？

2　不可用高斯消元法求最小二乘解

对 $\boldsymbol{Ax}=\boldsymbol{b}$ 作高斯消元法，即考虑另一方程组 $\boldsymbol{PAx}=\boldsymbol{Pb}$ 的最小二乘问题，其中 $\boldsymbol{P}\in GL_m(\mathbb{R})$ 为可逆方阵，相应的法方程变为 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{PAx}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{Pb}$。以下的几个事实和例子是简单的：

(i)　若 $\boldsymbol{A}^{\mathrm{T}}$ 表示的线性变换是单射（即 $\boldsymbol{A}$ 行满秩），则对任意可逆阵 $\boldsymbol{P}$，$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ 与 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{PAx}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{Pb}$ 有相同的通解。事实上，由于 $\boldsymbol{A}$ 表示的线性变换是满射（或更一般地，由于 $\boldsymbol{b}\in\operatorname{Im}\boldsymbol{A}$），最小二乘问题已退化为普通的线性方程组问题，可以做任意的初等行变换而不改变方程组的解。

(ii) 若 $\boldsymbol{P}$ 是正交阵，则 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ 与 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{PAx}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{Pb}$ 是相同的方程组，有相同的通解。特别地，在任意的最小二乘问题中，改变方程的次序不会改变最小二乘解，因此可以作“交换两行”的初等行变换。事实上，改变方程次序不会改变目标函数 $\|\boldsymbol{Ax}-\boldsymbol{b}\|$。

(iii) “某行乘以一个非零数”会改变最小二乘解。例如

$$\begin{pmatrix}1&1\\-1&-1\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}0\\-2\end{pmatrix} \text{ 和 } \begin{pmatrix}1&1\\-2&-2\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}0\\-4\end{pmatrix}$$

的最小二乘解不同。

(iv) “把某行乘以一个数加到另一行上”会改变最小二乘解，特别地，若某些方程线性相关，不可随意去掉其中一个。例如

$$\begin{pmatrix}1&1\\1&1\\1&1\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}2\\1\\1\end{pmatrix} \text{ 和 } \begin{pmatrix}1&1\\1&1\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}2\\1\end{pmatrix}$$

的最小二乘解不同。

从上面的几个例子可以看出，除了“交换两行”的操作外，另外两类初等行变换一般会改变最小二乘解。例题 1 给出的第二种解答是错误的，得到正确答案仅是“巧合”。

3 例题 1 中出现“巧合”的原因

把例题 1 求得的最小二乘解代入原方程组中，得到 $\boldsymbol{Ax}=(1/2,-1,-1/2,1)^{\mathrm{T}}$，这个向量就是 $\boldsymbol{b}$ 在 $\operatorname{Im}\boldsymbol{A}$ 上的正交投影 $\boldsymbol{d}$，它的第二个和第四个坐标分别与 $\boldsymbol{b}=(1,-1,0,1)^{\mathrm{T}}$ 的第二、四个坐标相等。一般地，有下面的结论。

命题 1 考虑方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的最小二乘问题，记 $\boldsymbol{b}=(b_1,\ldots,b_m)^{\mathrm{T}}$ 在 $\operatorname{Im}\boldsymbol{A}$ 上的正交投影为 $\boldsymbol{d}=(d_1,\ldots,d_m)^{\mathrm{T}}$，假设 $b_{i_1}=d_{i_1}$，…，$b_{i_r}=d_{i_r}$（$1\le i_1<\cdots<i_r\le m$），则可以在 $\boldsymbol{Ax}=\boldsymbol{b}$ 的第 $i_1,\ldots,i_r$ 行内作任意的初等行变换而不改变最小二乘解。

这里“在第 $i_1,\ldots,i_r$ 行内作初等行变换”意指“第 i_k 行乘以一个非零数，$1\le k\le r$”、“把第 i_k 行乘以一个数加到第 i_j 行上，$1\le k,j\le r$”，以及相应的行对换。

证 根据第 2 节的（ii），不妨设 $i_1=1$，…，$i_r=r$。由内积空间的知识，$\boldsymbol{Ax}=\boldsymbol{b}$ 的最小二乘解为方程组 $\boldsymbol{Ax}=\boldsymbol{d}=(b_1,\ldots b_r,d_{r+1},\ldots d_m)^{\mathrm{T}}$ 的通解。

若对 $\boldsymbol{Ax}=\boldsymbol{b}$ 的前 r 行作初等行变换，即考虑方程组 $\boldsymbol{PAx}=\boldsymbol{Pb}$ 的最小二乘问题，其中

$$\boldsymbol{P}=\begin{pmatrix}\boldsymbol{P}_1&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{I}_{m-r}\end{pmatrix}，\boldsymbol{P}_1\in GL_r(\mathbb{R})，\boldsymbol{I}_{m-r} \text{ 为 } m-r \text{ 阶单位阵。}$$

任取 $\boldsymbol{A}$ 的列向量 $\boldsymbol{\alpha}$，则 $\boldsymbol{\alpha} \perp \left(\boldsymbol{b}-\boldsymbol{d}\right)$。而 $\boldsymbol{b}-\boldsymbol{d}$ 的前 r 个坐标都是零，所以 $\boldsymbol{P\alpha} \perp \boldsymbol{P}\left(\boldsymbol{b}-\boldsymbol{d}\right)$，因此 $\boldsymbol{Pb}$ 在 $\operatorname{Im}\boldsymbol{PA}$ 上的正交投影为 $\boldsymbol{Pd}$。从而 $\boldsymbol{PAx}=\boldsymbol{Pb}$ 的最小二乘解全体为方程组 $\boldsymbol{PAx}=\boldsymbol{Pd}$ 的通解，亦为 $\boldsymbol{Ax}=\boldsymbol{d}$ 的通解。命题得证。

从上面的证明中，或者从几何直观上，不难得到如下更一般的叙述。

命题 2 $\boldsymbol{A}$，$\boldsymbol{b}$，$\boldsymbol{d}$ 如上，设 $\boldsymbol{P} \in GL_m\left(\mathbb{R}\right)$，则 $\boldsymbol{Ax}=\boldsymbol{b}$ 和 $\boldsymbol{PAx}=\boldsymbol{Pb}$ 有相同的最小二乘解当且仅当对 $\boldsymbol{A}$ 的任意列向量 $\boldsymbol{\alpha}$ 都成立 $\boldsymbol{P\alpha} \perp \boldsymbol{P}\left(\boldsymbol{b}-\boldsymbol{d}\right)$。

注 在例题 1 中，$\boldsymbol{b}$ 和其投影 $\boldsymbol{d}$ 的第二个和第四个坐标分别相等。根据命题 1，可以通过“把第二行加到第四行上”以舍去第四个方程。但不可把第二个和第四个方程都舍去，否则得到的新的方程组将有不同的最小二乘解。

4 结 论

一般情况下，对超定线性方程组使用高斯消元法会改变最小二乘解，甚至方程组中有两个相同的方程时都不能随便去掉其中一个，因此不可用此法简化最小二乘解的计算。鉴于众多教材都忽略了这个问题，且初学者或因熟于求解普通的线性方程组而不易意识到这样的错误，建议授课老师在讲课过程中提醒学生注意。

当然，关于如何简化最小二乘法的计算前人已有诸多研究，如[3]、[4]、[5]、[6]以及它们所引的文献等等。由于作者对应用和计算领域一无所知，本文也就只能到此为止。

致谢 感谢方明老师在中国科学院大学 2021-2022 学年春季学期线性代数 II-B05 课程上的授课，也感谢班上的全体同学，他们在这一学期给身为助教的作者带来了许多欢乐和悲伤。作者也十分感谢相关文献对本文的启发以及审稿专家提出的宝贵意见。

[参 考 文 献]

[1] 席南华. 基础代数（第二卷）[M]. 北京: 科学出版社, 2018.
[2] 姚慕生, 吴泉水, 谢启鸿. 高等代数学[M]. 3 版. 上海: 复旦大学出版社, 2014.
[3] LAWSON C L, HANSON R J. Solving least squares problems[M]. Philadelphia: Society for Industrial and Applied Mathematics (SIAM), 1995.
[4] CLINE A K. An elimination method for the solution of linear least squares problems[J]. SIAM Journal on Numerical Analysis, 1973, 10: 283-289.
[5] 桂冰. 用分块的高斯-约当方法求解最小二乘问题[J]. 工科数学，1994，10(2): 124-125.
[6] 张忠桢，胡荣强. 直交投影与最小二乘问题[J]. 工科数学，1996，12(2): 111-113.

A Note on Least Squares Solutions and Gaussian Eliminations

HU Hong-sheng [1,2]
(1. Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China;
2. Beijing International Center of Mathematics Research, Peking University, Beijing, 100871, China)

Abstract: If we use Gaussian eliminations (i.e. do elementary row operations on the augmented matrix) to find least squares solutions for overdetermined linear equations, sometimes we will obtain the correct solutions. However, in general it yields wrong answers. We discuss some relations between Gaussian eliminations and least squares solutions, intending to remind beginners to pay attention to their difference.

Key words: overdetermined linear equations; least squares solution; Gaussian elimination