一种基于高等代数的线性最小二乘问题的解法

张先才 邓见光* 安妮 张足生
（东莞理工学院 网络空间安全学院，广东东莞 523808）

摘要： 自从 19 世纪初勒让德提出线性最小二乘问题以来，诸多数学家已经就它的意义、误差分析及解法进行了广泛的研究，但基于高等代数理论的解法在许多资料上只有零散的讨论，为此提出了一种基于高等代数理论的线性最小二乘问题的解法。基于严格的证明，得到了适用于数据集为列满秩和非列满秩时的解法，并进一步推算出解集中的最优最小二乘解。实验证明该算法确实可以得到最优最小二乘解，并且在数据集属性数较少的情况下优于负梯度下降法。

关键词： 线性最小二乘问题；高等代数；矩阵变换；线性方程组；最优最小二乘解
中图分类号： O241. 1 **文献标志码：** A **文章编号：** 1009－0312（2020）05－0001－07
DOI:10.16002/j.cnki.10090312.2020.05.001

对于给定的数据集，寻找一个函数（模型）使得误差平方和最小，这类问题可称为最小二乘问题[1]。最小二乘问题根据所求函数的非线性和线性特点可分为非线性最小二乘问题和线性最小二乘问题。非线性最小二乘问题比较复杂，没有解析解，通常用迭代法来计算，负梯度下降法和高斯－牛顿法可用于求解非线性最小二乘问题[2-3]。

对于线性最小二乘问题来说，有三种不同思路求解：1）以多元函数微分理论为基础，将问题转为求解代价函数的极值点，因为线性模型下的代价函数为凸函数，所以极值点为最小值，根据极值点处梯度为零的等量关系，列方程组计算得到极值点[4]；2）用迭代法按负梯度方向来逼近代价函数的极值点，即负梯度下降法；3）以高等代数理论为基础，将线性最小二乘问题看成求解不相容线性方程组的最小二乘解，进而转化为线性方程组的求解问题[5]。

对于第三种思路，目前未见从证明到设计、从实现再到评估完整的论述。而且大多资料只给出了在数据集列满秩条件下的最小二乘解，并且没有考虑最优最小二乘解。

为此本文提出了一种基于高等代数理论的线性最小二乘问题的完整解法。以高等代数为基础，推导并设计了既适用于数据集列满秩和非列满秩，又可以计算最优最小二乘解的算法。实验证明该算法确实可以得到最优最小二乘解，并且在数据集属性数较少的情况下优于负梯度下降法。

1 问题建模

本文先对线性最小二乘问题做数学描述[6]，再将线性最小二乘问题转化为求解不相容方程组的最小二乘解问题。本文用大写字母表示矩阵与集合，小写字母和希腊字母表示已知数和未知数。本文符号说明：$f(x_1, x_2, \cdots, x_n)$：实数域上 n 元函数；$\alpha_1, \alpha_2, \cdots, \alpha_n$：函数 f 的非常数项参数；β：函数 f 的常数项参数；$\varphi_1, \varphi_2, \cdots, \varphi_n$：实数域上一元函数；$D$：数据集；$(x_{i1}, x_{i2}, \cdots, x_{in}, y_i)$：数据集的第 i 条记录且全为实数，前 n 个值称为属性或输入特征，最后 1 个值称为输出特征；J：最小二乘解集；X：数据集输入特征组成的 $s \times n$ 矩阵；y：数据集输出特征组成的 s 维列向量；α：函数 f 的非常数项参数组成的 n 维列向量；$\boldsymbol{\Phi}(X)$：经 $\varphi_1, \varphi_2, \cdots, \varphi_n$ 变换后的输入特征组成的 $s \times n$ 矩阵；c：全为 1 的 n 维列向量；A：由 $\boldsymbol{\Phi}(X)$ 和 c 组成的 $s \times (n+1)$ 矩阵；γ：由 α 和 β 组成的 $(n+1)$ 维列向量；$R(\sim)$：$\sim$ 的秩；t：A 的秩；E_i：i 阶单位

收稿日期：2020－04－26
基金项目：广东省普通高校特色创新项目（2018KTSCX221）；广东省普通高校青年创新人才项目（2017KQNCX194）；东莞理工学院科技产业创新服务专项（2019ZYFWXFD02）。
作者简介：张先才（1995—），男，安徽六安人，硕士生，主要从事机器学习研究，Email：zhang. xiancai@ qq. com。
* 通讯作者：邓见光（1981—），男，河南周口人，副教授，博士，主要从事大数据、物联网研究，Email：dengjg@ dgut. edu. cn。

阵; B: $t\times(n+1-t)$ 矩阵; γ_1: t 维列向量; γ_2: $(n+1-t)$ 维列向量; z_i: 基础解系的第 i 个解的前 t 个值; k: t 维列向量; η: $(n+1-t)$ 维自由未知向量。

1.1 线性最小二乘问题

设函数 $f(x_1,x_2,\cdots,x_n)=\alpha_1\varphi_1(x_1)+\alpha_2\varphi_2(x_2)+\cdots+\alpha_n\varphi_n(x_n)+\beta$，其中 $x_1,x_2,\cdots,x_n$ 是实数，$\varphi_1,\varphi_2,\cdots,\varphi_n$ 是已知的函数（可以是非线性的），实数 $\alpha_1,\alpha_2,\cdots,\alpha_n,\beta$ 是未知的参数。

数据集 $D=\left\{(x_{11},x_{12},\cdots,x_{1n},y_1),(x_{21},x_{22},\cdots,x_{2n},y_2),\cdots,(x_{s1},x_{s2},\cdots,x_{sn},y_s)\right\}$，$D$ 中的元素称为记录，每条记录前 n 个值称为属性或输入特征，最后一个值称为输出特征，输入和输出特征都为实数。

线性最小二乘问题是要找到函数 f 合适的参数 $\alpha_1,\alpha_2,\cdots,\alpha_n,\beta$，使得函数 f 与数据集 D 的误差平方和最小。该问题描述为定义 1。

定义 1 （线性最小二乘问题）给定函数 f 和数据集 D，求 $(\alpha_1,\alpha_2,\cdots,\alpha_n,\beta)\in \boldsymbol{R}^{n+1}$，使得

$$\min_{(\alpha_1,\alpha_2,\cdots,\alpha_n,\beta)}\sum_{i=1}^{s}\left(f(x_{i1},x_{i2},\cdots,x_{in})-y_i\right)^2. \quad (1)$$

设该问题的解集为 J，解集中长度最小的解称为最优最小二乘解。求解最优最小二乘解的问题描述为定义 2。

定义 2 （求解最优最小二乘解问题）给定线性最小二乘问题的解集 J，求 $(\alpha_1,\alpha_2,\cdots,\alpha_n,\beta)\in J$，使得

$$\min_{(\alpha_1,\alpha_2,\cdots,\alpha_n,\beta)}\sqrt{{\alpha_1}^2+{\alpha_2}^2+,\cdots,+{\alpha_n}^2+\beta^2}. \quad (2)$$

1.2 转化

将数据集 D 及函数 f 的参数写成如下所示矩阵形式

$$\boldsymbol{X}=\begin{pmatrix}x_{11}&\cdots&x_{1n}\\ \vdots&\ddots&\vdots\\ x_{s1}&\cdots&x_{sn}\end{pmatrix},\ y=\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_s\end{pmatrix},\ \alpha=\begin{pmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_n\end{pmatrix},$$

其中 (X,y) 是数据集 D，它的每一行是数据集的每条记录，α 指函数 f 的非常数项的参数。另外设 $c=(1,1,\cdots,1)^T$ 且维数为 n，记 $\boldsymbol{\Phi}(X)=\begin{pmatrix}\varphi_1(x_{11})&\cdots&\varphi_n(x_{1n})\\ \vdots&\ddots&\vdots\\ \varphi_1(x_{s1})&\cdots&\varphi_n(x_{sn})\end{pmatrix}$，则:

$$\sum_{i=1}^{s}\left(f(x_{i1},x_{i2},\cdots,x_{in})-y_i\right)^2=\|\boldsymbol{\Phi}(X)\alpha+c\beta-y\|^2=\left\|(\boldsymbol{\Phi}(X),c)\begin{pmatrix}\alpha\\ \beta\end{pmatrix}-y\right\|^2,$$

则式（1）转化为式（3）

$$\min_{(\alpha\beta)^T}\left\|(\boldsymbol{\Phi}(X),c)\begin{pmatrix}\alpha\\ \beta\end{pmatrix}-y\right\|^2, \quad (3)$$

其中 $(\boldsymbol{\Phi}(X),c)$，y 是已知的，$\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$ 是未知的，式（3）表达的即是求不相容方程组 $(\boldsymbol{\Phi}(X),c)\begin{pmatrix}\alpha\\ \beta\end{pmatrix}=y$ 的最小二乘解问题。为了方便表达，本文设 $A=(\boldsymbol{\Phi}(X),c)$，$\gamma=\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$，并将 D 和矩阵 A 都称为数据集，具体所指根据上下文判断。则有

$$\boldsymbol{A}\gamma=y. \quad (4)$$

综上，线性最小二乘问题等价于求不相容方程组（4）的最小二乘解问题。

2 算法

先说明算法的整体流程，然后基于垂线最短定理，将求解不相容方程组的最小二乘解问题转化为求解线性方程组问题，再基于矩阵的初等变换，给出一个解该线性方程组的算法，在该基础上得到计算最优最小二乘解的算法。最后总结不同的数据集和线性最小二乘问题的解之间的关系，并分析算法的时间复杂度。

本文所设计的算法整体流程图如图 1 所示。

整体来看算法的流程分为两步：即转化和计算。在转化这步，首先将输入的数据集转化为不相容方程组的增广矩阵，再转化为线性方程组的增广矩阵；在计算这步，首先对增广矩阵做有限次行列初等变换，如果变换的结果为单位阵，则得到最小最优二乘解，如果不是，再计算最小二乘解并通过递归得到最优最小二乘解。

2.1 转化为求解线性方程组问题

垂线最短定理是将求解不相容方程组的最小二乘解转化为求解线性方程组的基础。下面给出垂直最短定理以及它的证明，该证明基于正交子空间理论[7]。

推论 1（垂线最短定理） 设 $u\in\boldsymbol{R}^n$，v_1 为向量 u 在 R_n 的子空间 V_1 上的正射影，$\|v_2\|$ 为向量 u 到 V_1 的距离，对于 V_1 中的任一向量 λ 有：$\|v_2\|=\|u-v_1\|\leqslant\|u-\lambda\|$，且等号成立当且仅当 $\lambda=v_1$。

证明 由 v_2 是 V_1 的正交补中的元素，得

$v_2 = (u - v_1) \perp V_1$。由 $v_1 \in V_1, \lambda \in V_1$，得 $v_1 - \lambda \in V_1$，所以 $(u - v_1) \perp (v_1 - \lambda)$。而 $u - \lambda = (u - v_1) + (v_1 - \lambda)$，由勾股定理得：$\|u - \lambda\|^2 = \|u - v_1\|^2 + \|v_1 - \lambda\|^2$。因为 $\|v_1 - \lambda\| \geqslant 0$，所以 $\|v_2\| = \|u - v_1\| \leqslant \|u - \lambda\|$，且等号成立当且仅当 $\lambda = v_1$。证毕。

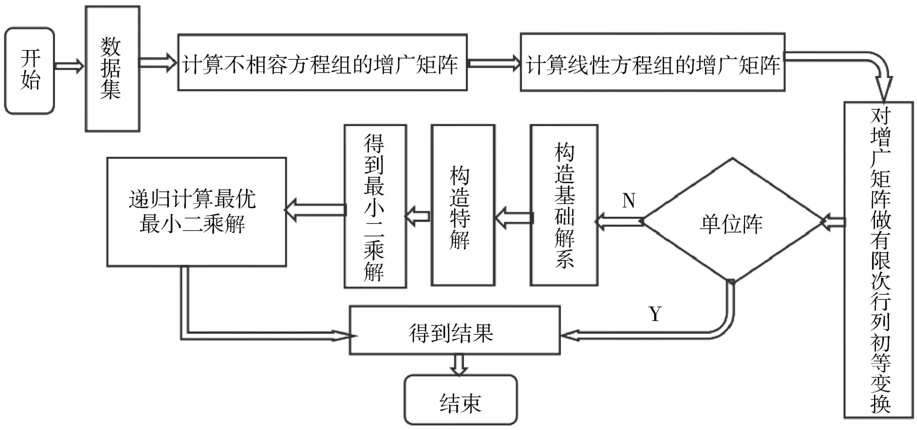

图 1　算法流程图

该推论说明了若 v_2 是向量 μ 与子空间 V_1 的最短的差向量，则 v_2 是 μ 到 V_1 的垂线。

对于不相容方程组（4），将矩阵 A 看成列向量组的形式，即 $A = (d_1, d_2, \cdots, d_{n+1})$，设 W 是该向量组生成的子空间，即 $W = \mathrm{span}\left\{d_1, d_2, \cdots, d_{n+1}\right\}$，则以下命题是等价的：

（1）j 是方程组 $A\gamma = y$ 的最小二乘解。

（2）$(y - Aj)$ 是 y 到 W 的垂线。

（3）$(y - Aj) \perp d_i (i = 1, 2, \cdots, n + 1)$。

（4）$A^T Aj = A^T y$。

证明

（1）$\Leftrightarrow$（2）：若 j 是方程组（4）的最小二乘解，则对于任意的 $w \in W$，有 $\|y - Aj\| \leqslant \|y - w\|$。即 $(y - Aj)$ 是 y 与子空间 W 最短的差向量，由垂线最短定理可知，$(y - Aj)$ 是 y 到子空间 W 的垂线。反之亦成立。

（2）$\Rightarrow$（3）：$(y - Aj)$ 为 y 到 W 的垂线，可得 $(y - Aj) \perp d_i (i = 1, 2, \cdots, n + 1)$。

（3）$\Rightarrow$（2）：因为 W 是 $\left\{d_1, d_2, \cdots, d_{n+1}\right\}$ 生成的子空间，且 $(y - Aj) \perp d_i (i = 1, 2, \cdots, n + 1)$，所以 $(y - Aj) \perp W$，即 $(y - Aj)$ 在 W 的正交补上。而 $y = (y - Aj) + Aj, Aj \in W$，由垂线的定义可得 $(y - Aj)$ 是 y 到 W 的垂线。

（3）$\Leftrightarrow$（4）：$(y - Aj) \perp d_i (i = 1, 2, \cdots, n + 1) \Leftrightarrow A^T(y - Aj) = 0 \Leftrightarrow A^T Aj = A^T y$。证毕。

由（1）和（4）可直接得到推论 2：

推论 2　j 是不相容方程组（4）的最小二乘解的充要条件是 j 适合线性方程组

$$A^T A\gamma = A^T y. \tag{5}$$

该推论将求解不相容方程组的最小二乘解问题转化为求解线性方程组的问题。设 $R(\sim)$ 表示 $\sim$ 的秩，下面证明 $R(A) = R(A^T A)$，并在此基础上说明线性方程组（5）一定有解。

推论 3　$R(A) = R(A^T A)$

证明　只要证线性方程组 $A\gamma = 0$ 和 $A^T A\gamma = 0$ 是同解方程组。

设 j 是方程组 $A\gamma = 0$ 的解，即 $Aj = 0$，则 $A^T Aj = 0$，所以 j 也是方程组 $A^T A\gamma = 0$ 的解。

设 j 是方程组 $A^T A\gamma = 0$ 的解，即 $A^T Aj = 0$，则 $j^T A^T Aj = 0$。由矩阵转置的性质可得：$(Aj)^T Aj = 0$，则 $Aj = 0$，即 j 是方程组 $A\gamma = 0$ 的解。证毕。

推论 4　线性方程组 $A^T A\gamma = A^T y$ 必有解。

证明　（该证明建立在线性方程组解的结构相关理论之上）

方程组（5）的增广矩阵为 $(A^T A, A^T y)$，由矩阵的秩的定义可知 $R(A^T A, A^T y) \geqslant R(A^T A)$。

因为两个矩阵乘积的秩比各自的秩都小，所以 $R(A^T A, A^T y) = R(A^T(A, y)) \leqslant R(A^T) = R(A)$。由推论 3 可得 $R(A^T A, A^T y) \leqslant R(A^T A)$。

由此可得 $R(A^T A, A^T y) = R(A^T A)$，即方程组（5）的增广矩阵和系数矩阵的秩相等，所以方程组（5）必有解。

2.2　求解线性方程组

本文以矩阵的初等变换为基础，给出一个求解方程组（5）的算法[8-9]。该算法具有两个优点：1）计算了 $A^T A$ 满秩和非满秩（对应于数据

集列满秩和非列满秩）情况下的解；2）虽然当 A^TA 满秩，可通过逆矩阵直接求解，即 $\gamma=(A^TA)^{-1}A^Ty$，但本文的算法在不调用求逆函数的前提下计算该情况下的解。

下面先说明数据集非列满秩时的解法，在此基础上说明列满秩情况下的解法。

2.2.1 数据集非列满秩时的解法

A^TA 是 $(n+1)\times(n+1)$ 矩阵，设 $R(A^TA)=t$，则 $t<(n+1)$。非齐次线性方程组（5）的通解等于它的一个特解加上对应齐次线性方程组 $A^TA\gamma=0$ 的通解，本文根据这个前提来说明方程组（5）的解法。

有限次行列初等变换：对 $A^TA\gamma=0$ 的系数矩阵做有限次行初等变换和交换两列的位置操作后，得到如下形式的方程组

$$\begin{pmatrix}\boldsymbol{E}_t & \boldsymbol{B}\\0 & 0\end{pmatrix}\begin{pmatrix}\gamma_1\\\gamma_2\end{pmatrix}=0\text{ ,}\qquad(6)$$

其中，$\boldsymbol{E}_t$ 为 t 阶单位阵，$\boldsymbol{B}$ 为 $t\times(n+1-t)$ 矩阵，γ_1 为 t 维未知列向量，γ_2 为 $(n+1-t)$ 维未知列向量。

具体来说，行列初等变换分为三步以得到目标形式的系数矩阵：1）对系数矩阵做有限次行初等变换，将其化为首 1 的行阶梯形；2）对行阶梯形做有限次交换两列的位置操作，使行阶梯形中首 1 的元素在其对角线上；3）对行阶梯形做有限次行初等变换，将其转化为简化阶梯形。简化阶梯形就是形如（6）的系数矩阵。

因为对系数矩阵做行初等变换不改变方程组的解，交换两列的位置只改变未知量的位置，所以方程组（6）与原齐次线性方程组 $A^TA\gamma=0$ 为同解方程组，只不过解的位置会根据所做的交换两列位置的操作发生变化。为了方便讨论，这里假设没有做列操作。

构造方程组（6）的一个基础解系：设 $\boldsymbol{E}_{(n+1-t)}=(e_1,e_2,\cdots,e_{(n+1-t)})$ 为 $(n+1-t)$ 阶单位阵（注意与前文的 $\boldsymbol{E}_t$ 的阶数不同），将 $e_1,e_2,\cdots,e_{(n+1-t)}$ 作为基础解系中每个解的后 $(n+1-t)$ 个值，这就满足了线性无关的条件。因为

$$\begin{pmatrix}\boldsymbol{E}_t & \boldsymbol{B}\\0 & 0\end{pmatrix}\begin{pmatrix}\gamma_1\\\gamma_2\end{pmatrix}=0\Leftrightarrow\boldsymbol{E}_t\gamma_1+\boldsymbol{B}\gamma_2=0\Leftrightarrow\gamma_1=-\boldsymbol{B}\gamma_2\text{ ,}$$

所以，该基础解系中每个解的前 t 个值为：

$$(z_1,z_2,\cdots,z_{(n+1-t)})=-\boldsymbol{B}(e_1,e_2,\cdots,e_{(n+1-t)})=-\boldsymbol{B}\text{ ,}$$

所以构造的基础解系为：

$$\begin{pmatrix}z_1,z_2,\cdots,z_{(n+1-t)}\\e_1,e_2,\cdots,e_{(n+1-t)}\end{pmatrix}=\begin{pmatrix}-\boldsymbol{B}\\\boldsymbol{E}_{(n+1-t)}\end{pmatrix}.$$

构造方程组（5）的一个特解：对方程组（5）的增广矩阵做同样的变换，这里同样指的是对系数矩阵 (A^TA,A^Ty) 做什么变换，就对增广矩阵做什么变换。可得如下形式方程组

$$\begin{pmatrix}\boldsymbol{E}_t & \boldsymbol{B}\\0 & 0\end{pmatrix}\begin{pmatrix}\gamma_1\\\gamma_2\end{pmatrix}=\begin{pmatrix}k\\0\end{pmatrix}\Leftrightarrow\boldsymbol{E}_t\gamma_1+\boldsymbol{B}\gamma_2=k\Leftrightarrow$$

$$\gamma_1=-\boldsymbol{B}\gamma_2+k\text{ ,}$$

其中 k 为 t 维列向量，令 $\gamma_2=0$，则 $\gamma_1=k$，所以方程组（5）的一个特解为 $\begin{pmatrix}k\\0\end{pmatrix}$。

在实际的算法中，只对增广矩阵 (A^TA,A^Ty) 而不是系数矩阵 A^TA 做了有限次初等行列变换。因为其已经包含了系数矩阵的结果。

结果　基于以上三步的讨论，当非满秩时，线性方程组（5）的通解为：

$$\gamma=\begin{pmatrix}-\boldsymbol{B}\\\boldsymbol{E}_{(n+1-t)}\end{pmatrix}\eta+\begin{pmatrix}k\\0\end{pmatrix}\text{ ,}\qquad(7)$$

其中 η 为 $(n+1-t)$ 维自由未知向量。

2.2.2 数据集列满秩时的解法

满秩与非满秩的区别在于：1）对系数矩阵做有限次行列初等变换后，满秩情况下的系数矩阵会变成单位阵；2）满秩情况下齐次方程的通解为 0 向量。所以只要对增广矩阵做有限次行列初等变换，得到如下形式的方程组：

$$\boldsymbol{E}_{(n+1)}\gamma=k\text{ ,}$$

其中 $\boldsymbol{E}_{(n+1)}$ 为 $(n+1)$ 阶单位阵，k 为 $(n+1)$ 维列向量（与前文的 k 不矛盾，因为 $R(A^TA)=t=n+1$）。然后直接可得该情况下的解 $\gamma=k$。

2.3 计算最优最小二乘解

本文基于解集的结构来计算最优最小二乘解。当 A^TA 满秩时最小二乘解唯一，所以该解也是最优最小二乘解。当 A^TA 非满秩时，基于（7）式，可得解集 J：

$$J=\left\{\begin{pmatrix}-\boldsymbol{B}\\\boldsymbol{E}_{(n+1-t)}\end{pmatrix}\eta+\begin{pmatrix}k\\0\end{pmatrix}\,\middle|\,\eta\in\boldsymbol{R}^{(n+1-t)}\right\}.$$

则基于解集 J 求解最优最小二乘解的问题可描述为：

$$\min_{\eta}\left\|\begin{pmatrix}-\boldsymbol{B}\\\boldsymbol{E}_{(n+1-t)}\end{pmatrix}\eta+\begin{pmatrix}k\\0\end{pmatrix}\right\|^2\text{ ,}$$

其中 $\begin{pmatrix}-\boldsymbol{B}\\\boldsymbol{E}_{(n+1-t)}\end{pmatrix}$ 和 $\begin{pmatrix}k\\0\end{pmatrix}$ 已知，这就相当于求形如方程组 $\begin{pmatrix}-\boldsymbol{B}\\\boldsymbol{E}_{(n+1-t)}\end{pmatrix}\eta=-\begin{pmatrix}k\\0\end{pmatrix}$ 的最小二乘解，所以通过递归调用就可得到最优最小二乘解（注意应从

计算线性方程组的增广矩阵这一步开始）。另外 $\begin{pmatrix}-B\\E_{(n+1-t)}\end{pmatrix}$ 是列满秩的，则方程组 $\begin{pmatrix}-B\\E_{(n+1-t)}\end{pmatrix}\eta=-\begin{pmatrix}k\\0\end{pmatrix}$ 的最小二乘解唯一，所以此时最优最小二乘解是唯一的。

2.4 数据集与最小二乘解的关系

基于上述的讨论及推论3，对于线性最小二乘问题，不同的数据集与解的关系总结如表1所示。

表1 数据集 A 与最小二乘解的关系

	$A^TA\gamma=A^Ty$	最小二乘解	最优最小二乘解
A 列满秩	A^TA 满秩	唯一	唯一
A 非列满秩	A^TA 非满秩	无穷个	唯一

2.5 时间复杂度

算法输入的规模是 $s\times(n+1)$，即 s 条记录，每条记录有 n 个输入特征和1个输出特征。算法时间开销主要在以下三步：1）计算 $A^TA\gamma=A^Ty$ 的增广矩阵；2）对该增广矩阵做有限次行列初等变换；3）递归计算最优最小二乘解。下面分析这三步的时间复杂度。

对于第一步，A 是 $s\times(n+1)$ 矩阵，计算 A^TA 及 A^Ty 的复杂度为 $O(s\times n^2+s\times n)=O(s\times n^2)$；对于第二步，增广矩阵是 $n\times(n+1)$ 矩阵，将其化成首1的行阶梯形的复杂度如式（8）所示，交换两列及化成简化阶梯形的复杂度如式（9）所示，所以第二步复杂度为 $O(n^3)$。对于第三步，输入矩阵的规模为 $(n+1-t)\times n$，且只递归一次，所以复杂度为 $O(n^3)$。

$$O\left(\sum_{i=0}^{n-1}(n-i)^2\right)=O\left(\frac{n(n+1)(2n+1)}{6}\right)=O(n^3)\ ,\tag{8}$$

$$O\left(n^2+\sum_{i=0}^{n-1}(n-i)^2\right)=O(n^3)\ .\tag{9}$$

综上所述，算法的时间复杂度 $T(s\times n)=O(s\times n^2)+O(n^3)$。可见算法的时间开销随数据记录数成线性增长，随数据属性数成立方增长。所以该算法适用于数据属性数较少的情况。

3 实验

将负梯度下降法作为对照，来评估算法在误差、时间开销及解的长度方面的性能。用于实验的笔记本电脑（RedmiBook14）配备了型号为2.10GHz Intel Core i3－8145U CPU的处理器，8 GB的内存及Win10操作系统。我们在Matlab上实现了基于高等代数的算法以及基于负梯度下降的算法，分别称为高代法（Advanced Algebraic Method，AAM）和负梯度下降法（Negative Gradient Descent Method，NGDM），其中负梯度下降法学习率设为－0.000 01，循环次数设为2 000，初始解服从［0，2］上的均匀分布。

实验的数据集按照式（10）生成，其中，δ 是服从［0，1］上均匀分布的误差项。在 s 条记录及 n 个属性的生成数据集中，参数（$\alpha_1,\alpha_2,\cdots,\alpha_n,\beta$）及（$x_1,x_2,\cdots,x_n$）由服从［0，1］及［0，10］上均匀分布的总体采样一次及 s 次得到，y 由式（10）计算得到：

$$y=\alpha_1x_1+\alpha_2x_2+\cdots+\alpha_nx_n+\beta+\delta\ .\tag{10}$$

本文将数据集特征数（输入特征数与输出特征数之和）占特征数与记录数之和的比例称为特征占比（即为 $\frac{n+1}{n+1+s}$）。图2展示了数据规模一定时（$s\times(n+1)=90\ 000$），高代法在不同特征占比下的时间开销。曲线的增长趋势说明了特征数比记录数更加影响时间开销。

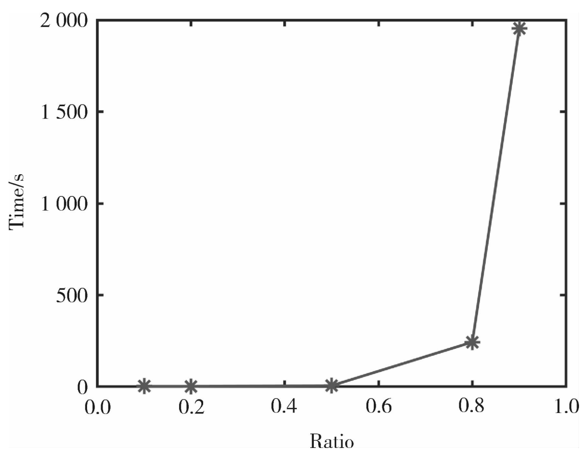

图2 不同特征占比的时间开销

本文用平均绝对百分比误差（MAPE）来作为误差的指标。图3展示了两种算法在数据集记录数为1 000的前提下，不同属性数的MAPE。由图3可知不同属性数下高代法的误差总是小于负梯度下降法，说明在误差方面高代法优于负梯度下降法。另外，因为数据集的误差项服从0到1的均匀分布，属性越多其相对误差越小，所以图3曲线呈下降趋势。

图4展示了两种算法在数据集记录数为1 000的前提下，不同数据集属性数的时间开销结果。

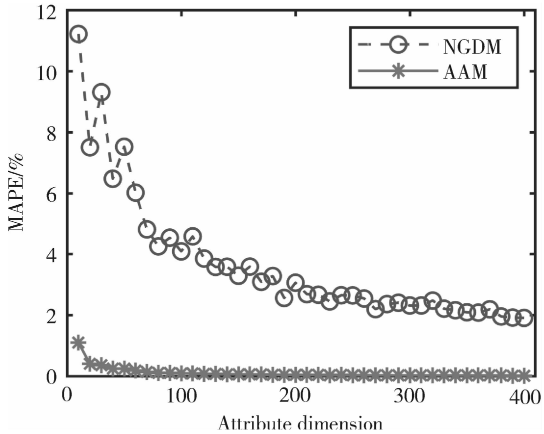

图3　误差对比

由图4 可知属性数小于355 时高代法更快，大于355 时负梯度下降法更快，说明高代法适用于数据集属性数较少的情况，这验证了本文的复杂度分析。应该注意的是两种算法的分界点不是固定的，与数据集、负梯度下降法的循环次数以及学习率有关。图5 展示了两种算法在数据集属性数为350 的前提下，不同数据集记录数的时间开销结果，由曲线的增长趋势可知，两种算法的记录数与时间开销呈线性增长关系。

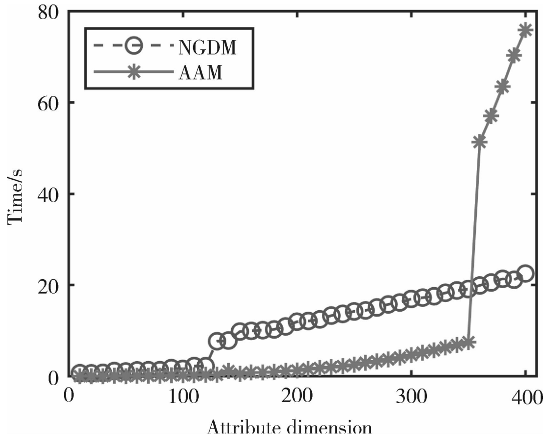

图4　不同属性数的时间开销对比

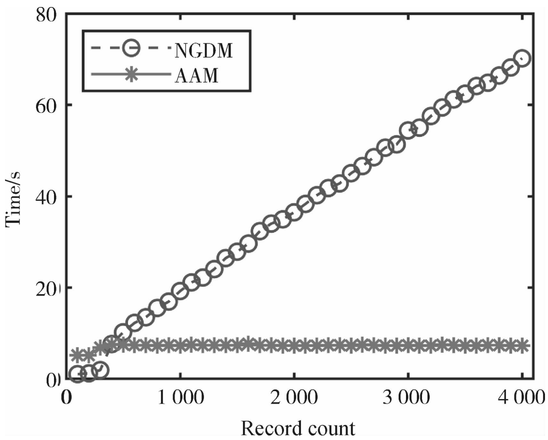

图5　不同记录数的时间开销对比

图6 和图7 展示了数据集非列满秩时（记录数为100，属性数为110 到400），两种方法所得解的长度及 MAPE。由图6 可知在数据集非列满秩时，高代法所得的解总是更短，由图7 可知高代法的误差依然小于负梯度下降法。这说明高代法在数据集非列满秩下确实可以得到更短的最小二乘解。

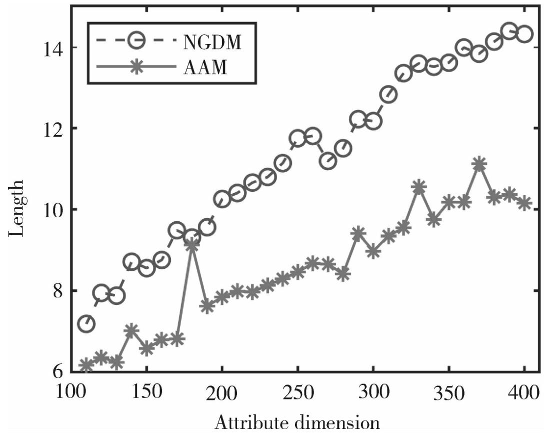

图6　非列满秩下解的长度

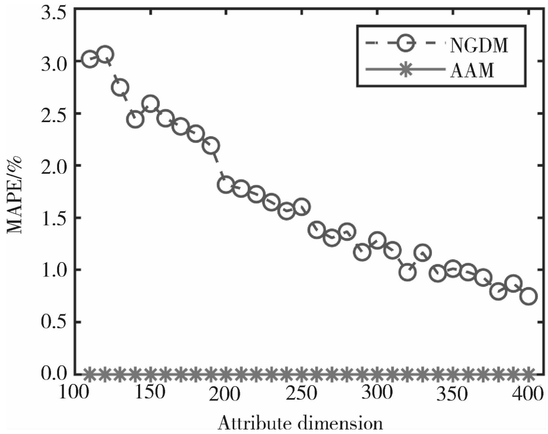

图7　非列满秩下的误差

4　结语

线性最小二乘问题应用广泛，但是基于高等代数的经典解法缺乏较为系统的论述。本文提出了一种基于高等代数理论的线性最小二乘问题的完整解法。以高等代数为基础，推导并设计了既适用于数据集列满秩和非列满秩，又可以计算最优最小二乘解的算法。通过实验说明了该算法误差总是小于负梯度下降法；该算法的时间开销对数据集属性数更加敏感，在属性数较少的情况下优于负梯度下降法；在数据集非列满秩下，该算

法相比负梯度下降法可以得到更短的最小二乘解。

参考文献

[1] 陈希孺. 最小二乘法的历史回顾与研究现状[J]. 中国科学院研究生院学报,1998,15(1):5-12.
[2] 江世宏. 计算方法[M]. 北京: 科学出版社,2014.
[3] 郑洲顺,普乐. 非线性最小二乘问题的一种迭代解法[J]. 数学理论与应用,2002,22(1):43-45.
[4] 茆诗松,程依明. 概率论与数理统计[M]. 北京: 高等教育出版社,2011:44.
[5] 同济大学应用数学系. 高等代数与解析几何[M]. 北京: 高等教育出版社,2005:286-290.
[6] 王学锋,王江涛. 线性流形上埃尔米特自反矩阵反问题的最小二乘解及其最佳逼近[J]. 东莞理工学院学报,2015,22(5):8-14.
[7] 陈之辉,于荣格. 欧氏空间的子空间的正交系与正交补[J]. 沧州师范学院学报,2019,35(1):1-4.
[8] 石擎天,黄坤阳. 线性方程组求解及应用[J]. 教育教学论坛,2020,12(12):325-327.
[9] 张新发. 初等变换的关系及可逆矩阵的分解[J]. 大学数学,2003,19(2):82-85.

A Solution to the Linear Least Squares Problem Based on Advanced Algebra

ZHANG Xiancai　DENG Jianguang*　AN Ni　ZHANG Zusheng
(School of Cyberspace Science, Dongguan University of Technology, Dongguan 523808, China)

Abstract Since Legendre proposed the linear least squares problem in the early 19th century, many mathematicians have conducted extensive research on its significance, error analysis, and solutions, but solutions based on higher algebra theory have only been scattered discussion on many materials. To this end, a solution to the linear least squares problem based on higher algebra theory is proposed. Based on strict proofs, a solution suitable for the data set with column full rank and non-column full rank is obtained, and the optimal least squares solution in the solution set is further deduced. Experiments show that the algorithm can indeed obtain the optimal least squares solution, and it is better than the negative gradient descent method when the number of attributes in the data set is small.

Key words linear least squares problem; advanced algebra; matrix transformation; linear equations; optimal least squares solution