第45卷 第4期
2026年12月

《新疆师范大学学报》(自然科学版)
Journal of Xinjiang Normal University
(Natural Sciences Edition)

Vol.45,No.4
Dec. 2026

多块可分离凸优化问题的邻近ADMM算法及收敛性

罗　秀,王湘美*,廖春芳
(贵州大学 数学与统计学院,贵州 贵阳 550025)

摘　要:交替方向乘子法(Alternating Direction Method of Multipliers,ADMM)是求解两模块可分离凸优化问题的重要方法之一,近年来取得了较大进展。文献[24]提出了一种求解多块可分离凸优化问题的twisted邻近交替方向乘子法(Twisted version of the proximal Alternating Direction Method of Multipliers,TADMM),并证明了算法的收敛性和迭代复杂度。基于文献[24]的算法,文章提出两种推广的TADMM,分别证明其收敛性和迭代复杂度。最后,将LASSO模型应用到手写数字识别问题,并运用上述三种TADMM求解。数值实验表明,三种算法均收敛,并且文章提出的第一种TADMM表现最好。

关键词:可分离凸优化;增广拉格朗日函数;邻近交替方向乘子法;收敛性;手写数字识别

中图分类号:O224　　**文献标识码**:A　　**文章编号**:1008-9659(2026)04-0085-14

文章考虑如下多块可分离的凸优化问题

$$\begin{aligned} \min_{x_i \in \chi_i} \quad & \sum_{i=1}^{m} f_i(x_i) \\ s.t. \quad & \sum_{i=1}^{m} A_i x_i = b \end{aligned} \tag{1}$$

其中,每个$\chi_i \subset R^{n_i}$是凸子集,$A_i \in R^{l \times n_i}$, $b \in R^l$, $f_i:R^{n_i} \to (-\infty,+\infty]$是下半连续真凸函数(不一定光滑)。当$m = 2$时,Glowinski等人在1975年提出交替方向乘子法[1](Alternating Direction Method of Multipliers,ADMM),算法通过引入增广拉格朗日函数,将原约束优化问题转化为无约束优化问题,并通过交替更新变量和拉格朗日乘子来逐步逼近最优解。Boyd等人[2]将ADMM应用到分布优化领域,推动了其在大规模问题中的应用[3]。Fortin[4]和Glowinski[5]证明了该算法的全局收敛性,更多研究可参考文献[6]~[12]及其参考文献。当$m \geqslant 3$时,刘浩洋[13]和Chen等人[14]给出了ADMM不收敛的反例,且文献[14]分析了其发散原因。为解决该问题,He等人[15-16]引入高斯回代纠正步,虽牺牲部分计算效率,但保证了全局收敛性。此外,为降低子问题的求解难度,一些学者提出并研究子问题中引入近端项$\frac{1}{2}(x - x^k)^T Q(x - x^k)$的近端交替方向乘子法(Proximal Alternating Direction Method of Multipliers,PADMM)。该算法最早由Eckstein[17]提出,He等人[18]进一步研究了动态惩罚系数和近端参数策略。更多关于PADMM的研究和应用详见文献[19]~[23]及其参考文献。针对多块可分离凸优化问题(1),Wang等人[24]提出一种twisted近端交替方向乘子法(Twisted version of the proximal Alternating Direction Method of Multipliers,TADMM)。该算法引入了高斯回代纠正步与近端项,同时将乘子更新步骤提前至第二步进行。理论分析表明,该算法对任意初始点均能保证收敛,且在最坏情况下具有$O(1/k)$的迭代复杂度。受以上研究启发,文章拟针对多块可分离的凸优化问题提出两种推广的TADMM,同时证明其收敛性和迭代复杂度。

[收稿日期]2025-09-15　　[修回日期]2025-09-30
[基金项目]国家自然科学基金项目(12161017;12561953)。
[作者简介]罗　秀(1998-),女,硕士研究生,主要从事优化理论方面研究,E-mail:luoxiu02@163.com.
*[通讯作者]王湘美(1972-),女,教授,主要从事优化理论方面研究,E-mail:xmwang2@gzu.edu.cn.

1 预备知识

$R^{n\times m}$是$n\times m$维欧几里德空间,记$I\in R^{n\times n}$为n阶单位矩阵。设$A\in R^{n\times n}$是实对称矩阵,分别用$\lambda_{\min}(A)$和$\lambda_{\max}(A)$表示A的最小和最大特征值,用$A>0$(或$A\geq 0$)表示A正定(或半正定)。对于任意n阶实对称矩阵$A,B\in R^{n\times n}$,$A>B$(或$A\geq B$)表示$A-B>0$(或$A-B\geq 0$). 设$A>0$,$x\in R^n$,$\|x\|_A:=\sqrt{x^TAx}$表示向量x的A-范数,$\|x\|_1$表示x的1范数。$\|\cdot\|$表示矩阵的核范数(即矩阵所有奇异值的和)。

设$F:R^n\rightrightarrows R^n$为定义在$R^n$上的集值映射,称$F$是单调的,如果对于任意的$x_1,x_2\in R^n$,有

$$\langle y_1-y_2,x_1-x_2\rangle\geqslant 0,\ \forall y_1\in F(x_1),\forall y_2\in F(x_2)$$

设函数$f:R^n\to(-\infty,+\infty]$为定义在$R^n$上凸函数,$x$为定义域$\mathrm{dom}f:=\{x\in R^n|f(x)<+\infty\}$中的一点。若向量$g\in R^n$满足

$$f(y)\geqslant f(x)+g^T(y-x),\ \forall y\in \mathrm{dom}f$$

则称g为函数f在点x处的次梯度。称集合

$$\partial f(x)=\{g|g\in R^n,f(y)\geqslant f(x)+g^T(y-x),\forall y\in \mathrm{dom}f\}$$

为f在点x处的次微分。凸函数f的次微分∂f是单调的[25]。

设$C\subset R^n$是R^n中的一个非空凸子集,其示性函数为

$$\delta(x;C)=\begin{cases}0, & x\in C\\ +\infty, & x\notin C\end{cases}$$

设$x\in C$,集合

$$N_C(x)=\{z|\langle z,y-x\rangle\leqslant 0,\forall y\in C\}$$

称为集合C在点x处的法锥(Normal Cone)。示性函数的次微分可以表示为

$$\partial\delta(x;C)=\begin{cases}N_C(x), & x\in C\\ \varnothing, & x\notin C\end{cases}$$

设矩阵$X\in R^{n\times n}$有分块

$$X=\begin{pmatrix}A & B\\ B^T & C\end{pmatrix}$$

其中,$A\in R^{k\times k}(k<n)$. 如果$|A|\neq 0$,则称矩阵$S:=C-B^TA^{-1}B$为矩阵X中A的舒尔补[26]。关于舒尔补,有以下结论。

引理1[26] 设矩阵$X\in R^{n\times n}$有分块

$$X=\begin{pmatrix}A & B\\ B^T & C\end{pmatrix}$$

其中,$A\in R^{k\times k}(k<n)$,则$X>0$当且仅当$A>0$且$S:=C-B^TA^{-1}B>0$.

2 算法及收敛性

设$m\geqslant 3$,考虑以下多模块可分离的带线性约束的凸优化问题

$$\begin{aligned}&\min_{x_i\in X_i}\sum_{i=1}^m f_i(x_i)\\ &s.t.\sum_{i=1}^m A_ix_i=b\end{aligned}\tag{2}$$

其中,每个$\chi_i\subset R^{n_i}$是凸子集;$A_i\in R^{l\times n_i}$,$b\in R^l$和$f_i:R^{n_i}\to(-\infty,+\infty]$是下半连续真凸函数。问题(2)的拉格朗日函数$L:R^{n_1}\times\cdots\times R^{n_m}\times R^l\to R$定义为

$$L(x_1,\cdots,x_m,\lambda)=\sum_{i=1}^m f_i(x_i)-\lambda^T\left(\sum_{i=1}^m A_ix_i-b\right)\forall(x_1,\cdots,x_m,\lambda)\in R^{n_1}\times\cdots\times R^{n_m}\times R^l\tag{3}$$

称点$(x_1^*,\cdots,x_m^*,\lambda^*)\in R^{n_1}\times\cdots\times R^{n_m}\times R^l$为拉格朗日函数$L$的鞍点。如果

$$L(x_1^*,\cdots,x_m^*,\lambda)\leqslant L(x_1^*,\cdots,x_m^*,\lambda^*)\leqslant L(x_1,\cdots,x_m,\lambda^*),\forall(x_1,\cdots,x_m,\lambda)\in R^{n_1}\times\cdots\times R^{n_m}\times R^l$$

记L的所有鞍点集合为W^*. 由凸分析[25]可知,点$(x_1^*,\cdots,x_m^*,\lambda^*)\in W^*$当且仅当

$$\begin{cases} 0 \in \partial f_1(x_1^*) - A_1^T\lambda^* + N_{\chi_1}(x_1^*) \\ 0 \in \partial f_2(x_2^*) - A_2^T\lambda^* + N_{\chi_2}(x_2^*) \\ \vdots \\ 0 \in \partial f_m(x_m^*) - A_m^T\lambda^* + N_{\chi_m}(x_m^*) \\ 0 = A_1x_1^* + A_2x_2^* + \cdots + A_mx_m^* - b \end{cases} \tag{4}$$

设$\beta > 0$,问题(2)的增广拉格朗日函数$L_\beta: R^{n_1} \times \cdots \times R^{n_m} \times R^l \to R$定义为,对任意$(x_1,\cdots,x_m,\lambda) \in R^{n_1} \times \cdots \times R^{n_m} \times R^l$

$$L_\beta(x_1,\cdots,x_m,\lambda) = \sum_{i=1}^m f_i(x_i) - \lambda^T\left(\sum_{i=1}^m A_ix_i - b\right) + \frac{\beta}{2}\left\|\sum_{i=1}^m A_ix_i - b\right\|^2 \tag{5}$$

为了找到问题(2)的最优解(即拉格朗日函数L的鞍点),分别提出两类TADMM算法,即:TADMM1和TADMM2.

2.1　TADMM1

TADMM1的具体步骤如下面的算法1.该算法由预测步和校正步构成,是文献[24]中TADMM的推广。与TADMM相比,该算法对$\tilde{x}_1^k$的子问题增加了近端项$\frac{1}{2}(x_1 - x_1^k)^TM_1(x_1 - x_1^k)$.

算法1(TADMM1):

第0步:给定初始点$(x_1^0,x_2^0,\cdots,x_m^0,\lambda^0) \in \chi_1 \times \chi_2 \times \cdots \times \chi_m \times R^l$,参数$\beta > 0, \alpha \in (0,2)$,对称矩阵$M_i \in R^{n_i \times n_i}$,$(i = 1,2,\cdots,m)$及停机准则$\varepsilon > 0$.初始化$k:= 0$.

第1步:预测

$$\begin{cases} \tilde{x}_1^k = \arg\min\limits_{x_1\in\chi_1} L_\beta(x_1,x_2^k,\cdots,x_m^k,\lambda^k) + \frac{1}{2}(x_1 - x_1^k)^TM_1(x_1 - x_1^k) \\ \tilde{\lambda}^k = \lambda^k - \beta(A_1\tilde{x}_1^k + A_2x_2^k + \cdots + A_mx_m^k - b) \\ \tilde{x}_2^k = \arg\min\limits_{x_2\in\chi_2} L_\beta(\tilde{x}_1^k,x_2,\cdots,x_m^k,\tilde{\lambda}^k) + \frac{1}{2}(x_2 - x_2^k)^TM_2(x_2 - x_2^k) \\ \vdots \\ \tilde{x}_i^k = \arg\min\limits_{x_i\in\chi_i} L_\beta(\tilde{x}_1^k,\cdots,x_{i-1}^k,x_i,x_{i+1}^k,\cdots,\tilde{\lambda}^k) + \frac{1}{2}(x_i - x_i^k)^TM_i(x_i - x_i^k) \\ \vdots \\ \tilde{x}_m^k = \arg\min\limits_{x_m\in\chi_m} L_\beta(\tilde{x}_1^k,\cdots,x_{m-1}^k,x_m,\tilde{\lambda}^k) + \frac{1}{2}(x_m - x_m^k)^TM_m(x_m - x_m^k) \end{cases} \tag{6}$$

第2步:校正

$$\begin{cases} x_1^{k+1} = x_1^k - \alpha(x_1^k - \tilde{x}_1^k) \\ x_2^{k+1} = x_2^k - \alpha(x_2^k - \tilde{x}_2^k) \\ \vdots \\ x_m^{k+1} = x_m^k - \alpha(x_m^k - \tilde{x}_m^k) \\ \lambda^{k+1} = \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k) \end{cases} \tag{7}$$

第3步:如果$\max\left\{\frac{\|x_1^k - x_1^{k+1}\|}{1 + \|x_1^k\|},\cdots,\frac{\|x_m^k - x_m^{k+1}\|}{1 + \|x_m^k\|},\frac{\|\lambda^k - \lambda^{k+1}\|}{1 + \|\lambda^k\|}\right\} < \varepsilon$,则停止迭代;否则令$k:= k + 1$,返回第1步。

下面分析TADMM1的收敛性。为了后文分析的简便性,对任意自然数k,引入以下记号

$$u^k = \begin{pmatrix} x_1^k \\ x_2^k \\ \vdots \\ x_m^k \\ \lambda^k \end{pmatrix}, u = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ \lambda \end{pmatrix}, \tilde{u}^k = \begin{pmatrix} \tilde{x}_1^k \\ \tilde{x}_2^k \\ \vdots \\ \tilde{x}_m^k \\ \tilde{\lambda}^k \end{pmatrix} \tag{8}$$

$$Q=\begin{pmatrix} M_1 & 0 & \cdots & 0 & 0 \\ 0 & \beta A_2^TA_2+M_2 & \cdots & 0 & -A_2^T \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \beta A_m^TA_m+M_m & -A_m^T \\ 0 & -A_2 & \cdots & -A_m & \frac{1}{\beta}I \end{pmatrix} \tag{9}$$

引理2[24] 设序列$\{u^k\}_{k\geqslant1}$由TADMM1生成,则对任意正整数k均有

$$(\tilde{u}^k-u^*)^TQ(u^k-\tilde{u}^k)\geqslant0,\forall u^*\in W^* \tag{10}$$

证明 设k为正整数。由式(6)中$x_i(i=1,2,\cdots,m)$子问题的最优性条件可得

$$\begin{cases} 0\in\partial f_1(\tilde{x}_1^k)-A_1^T\lambda^k+\beta A_1^T(A_1\tilde{x}_1^k+A_2x_2^k+\cdots+A_mx_m^k-b)+M_1(\tilde{x}_1^k-x_1^k)+N_{\chi_1}(\tilde{x}_1^k) \\ \tilde{\lambda}^k=\lambda^k-\beta(A_1\tilde{x}_1^k+A_2x_2^k+\cdots+A_mx_m^k-b) \\ \vdots \\ 0\in\partial f_2(\tilde{x}_2^k)-A_2^T\tilde{\lambda}^k+\beta A_2^T(A_1\tilde{x}_1^k+A_2\tilde{x}_2^k+\cdots+A_mx_m^k-b)+M_2(\tilde{x}_2^k-x_2^k)+N_{\chi_2}(\tilde{x}_2^k) \\ 0\in\partial f_m(\tilde{x}_m^k)-A_m^T\tilde{\lambda}^k+\beta A_m^T(A_1\tilde{x}_1^k+A_2x_2^k+\cdots+A_m\tilde{x}_m^k-b)+M_m(\tilde{x}_m^k-x_m^k)+N_{\chi_m}(\tilde{x}_m^k) \end{cases} \tag{11}$$

将式(11)中的第二式代入其余子式中可得

$$\begin{cases} 0\in\partial f_1(\tilde{x}_1^k)-A_1^T\tilde{\lambda}^k+M_1(\tilde{x}_1^k-x_1^k)+N_{\chi_1}(\tilde{x}_1^k) \\ 0\in\partial f_2(\tilde{x}_2^k)-A_2^T\tilde{\lambda}^k+A_2^T(\lambda^k-\tilde{\lambda}^k)+(\beta A_2^TA_2+M_2)(\tilde{x}_2^k-x_2^k)+N_{\chi_2}(\tilde{x}_2^k) \\ \vdots \\ 0\in\partial f_m(\tilde{x}_m^k)-A_m^T\tilde{\lambda}^k+A_m^T(\lambda^k-\tilde{\lambda}^k)+(\beta A_m^TA_m+M_m)(\tilde{x}_m^k-x_m^k)+N_{\chi_m}(\tilde{x}_m^k) \end{cases} \tag{12}$$

设$u^*=(x_1^*,\cdots,x_m^*,\lambda^*)\in W^*$是拉格朗日函数的鞍点,则有

$$\begin{cases} 0\in\partial f_1(x_1^*)-A_1^T\lambda^*+N_{\chi_1}(x_1^*) \\ 0\in\partial f_2(x_2^*)-A_2^T\lambda^*+N_{\chi_2}(x_2^*) \\ \vdots \\ 0\in\partial f_m(x_m^*)-A_m^T\lambda^*+N_{\chi_m}(x_m^*) \end{cases} \tag{13}$$

结合式(12)和式(13),并由法锥的单调性,则存在$y_i^k\in\partial f_i(\tilde{x}_i^k)(i=2,3,\cdots,m)$和$y_i^*\in\partial f_i(x_i^*)(i=1,2,\cdots,m)$,使得

$$\begin{cases} \left\langle -y_1^k+y_1^*+A_1^T(\tilde{\lambda}^k-\lambda^*)+M_1(x_1^k-\tilde{x}_1^k),\tilde{x}_1^k-x_1^*\right\rangle\geqslant0 \\ \left\langle -y_2^k+y_2^*+A_2^T(\tilde{\lambda}^k-\lambda^*)+z_2^k,\tilde{x}_2^k-x_2^*\right\rangle\geqslant0 \\ \vdots \\ \left\langle -y_m^k+y_m^*+A_m^T(\tilde{\lambda}^k-\lambda^*)+z_m^k,\tilde{x}_m^k-x_m^*\right\rangle\geqslant0 \end{cases} \tag{14}$$

其中,$z_i^k=-A_i^T(\lambda^k-\tilde{\lambda}^k)+(\beta A_i^TA_i+M_i)(x_i^k-\tilde{x}_i^k)(i=2,3,\cdots,m)$. 又由次梯度$\partial f_i(i=1,2,\cdots,m)$的单调性,可得

$$\begin{cases} \left\langle A_1^T(\tilde{\lambda}^k-\lambda^*)+M_1(x_1^k-\tilde{x}_1^k),\tilde{x}_1^k-x_1^*\right\rangle\geqslant\langle y_1^k-y_1^*,\tilde{x}_1^k-x_1^*\rangle\geqslant0 \\ \left\langle A_2^T(\tilde{\lambda}^k-\lambda^*)+z_2^k,\tilde{x}_2^k-x_2^*\right\rangle\geqslant\langle y_2^k-y_2^*,\tilde{x}_2^k-x_2^*\rangle\geqslant0 \\ \vdots \\ \left\langle A_m^T(\tilde{\lambda}^k-\lambda^*)+z_m^k,\tilde{x}_m^k-x_m^*\right\rangle\geqslant\langle y_m^k-y_m^*,\tilde{x}_m^k-x_m^*\rangle\geqslant0 \end{cases} \tag{15}$$

将以上不等式累加可得

$$\sum_{i=1}^m\langle\tilde{\lambda}^k-\lambda^*,A_i\tilde{x}_i^k-A_ix_i^*\rangle+\langle M_1(x_1^k-\tilde{x}_1^k),\tilde{x}_1^k-x_1^*\rangle+\sum_{i=2}^m\left\langle -A_i^T(\lambda^k-\tilde{\lambda}^k),\tilde{x}_i^k-x_i^*\right\rangle$$
$$+\sum_{i=2}^m\langle(\beta A_i^TA_i+M_i)(x_i^k-\tilde{x}_i^k),\tilde{x}_i^k-x_i^*\rangle\geqslant0 \tag{16}$$

再由式(4)中的最后一个等式,有

$$\sum_{i=1}^{m}(A_i\tilde{x}_i^k - A_ix_i^*) = \sum_{i=1}^{m}A_i\tilde{x}_i^k - b = (A_1\tilde{x}_1^k + \sum_{i=2}^{m}A_ix_i^k - b) + \sum_{i=2}^{m}A_i(\tilde{x}_i^k - x_i^k) = \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) + \sum_{i=2}^{m}A_i(\tilde{x}_i^k - x_i^k)$$

然后结合式(16)可得

$$\left\langle \tilde{\lambda}^k - \lambda^*, \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k)\right\rangle + \sum_{i=2}^{m}\langle \tilde{\lambda}^k - \lambda^*, -A_i(x_i^k - \tilde{x}_i^k)\rangle + \sum_{i=2}^{m}\left\langle -A_i^T(\lambda^k - \tilde{\lambda}^k), \tilde{x}_i^k - x_i^*\right\rangle$$

$$+\langle M_1(x_1^k - \tilde{x}_1^k), \tilde{x}_1^k - x_1^*\rangle + \sum_{i=2}^{m}\langle(\beta A_i^TA_i + M_i)(x_i^k - \tilde{x}_i^k), \tilde{x}_i^k - x_i^*\rangle \geqslant 0$$

因此，根据矩阵Q的定义，式(10)成立，证明完毕。

下面，给出一些条件以确保矩阵Q是正定的。为此，设

$$\delta_{\min} = \min_{2\leqslant i\leqslant m}\{\lambda_{\min}(A_i^TA_i)\}, \delta_{\max} = \max_{2\leqslant i\leqslant m}\{\lambda_{\max}(A_i^TA_i)\}\text{和}\bar{\delta}_{\min} = \min_{2\leqslant i\leqslant m}\{\lambda_{\min}(M_i)\}$$

假设

$$\bar{\delta}_{\min} > \beta[(m-1)\delta_{\max} - \delta_{\min}] \tag{17}$$

则矩阵Q是正定的。事实上，令$C = \frac{1}{\beta}I$

$$A = \begin{pmatrix} \beta A_2^TA_2 + M_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta A_m^TA_m + M_m \end{pmatrix}\text{和}B = \begin{pmatrix} -A_2 \\ \vdots \\ -A_m \end{pmatrix}$$

则矩阵Q可以写为

$$Q = \begin{pmatrix} M_1 & 0 \\ 0 & Q_1 \end{pmatrix}, \text{其中}Q_1 = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$$

注意，Q_1中A的舒尔补为

$$\begin{aligned} S = C - B^TA^{-1}B &= \frac{1}{\beta}I - \sum_{i=2}^{m}A_i(\beta A_i^TA_i + M_i)A_i^T \\ &= \sum_{i=2}^{m}\left[\frac{1}{(m-1)\beta}I - A_i(\beta A_i^TA_i + M_i)A_i^T\right] \\ &\geq \sum_{i=2}^{m}\left[\frac{1}{(m-1)\beta}I - \frac{1}{\beta\delta_{\min} + \bar{\delta}_{\min}}A_iA_i^T\right] \\ &\geq \sum_{i=2}^{m}\left[\frac{1}{(m-1)\beta} - \frac{\delta_{\max}}{\beta\delta_{\min} + \bar{\delta}_{\min}}\right]I \end{aligned}$$

根据式(17)，$S > 0$. 因此，由引理1可知Q_1是正定的，同理上文所述的矩阵Q也是正定的。

引理3[24]　设序列$\{u^k\}_{k\geqslant 1}$由TADMM1生成，则有

$$\|u^{k+1} - u^*\|_Q^2 \leqslant \|u^k - u^*\|_Q^2 - (2-\alpha)\alpha\|u^k - \tilde{u}^k\|_Q^2, \forall u^* \in W^* \tag{18}$$

证明　设$u^* \in W$，k为正整数。由引理2可知$(\tilde{u}^k - u^*)^TQ(u^k - \tilde{u}^k) \geqslant 0$，又

$$(\tilde{u}^k - u^*)^TQ(u^k - \tilde{u}^k) = (u^k - u^*)^TQ(u^k - \tilde{u}^k) - \|u^k - \tilde{u}^k\|_Q^2$$

所以

$$(u^k - u^*)^TQ(u^k - \tilde{u}^k) \geqslant \|u^k - \tilde{u}^k\|_Q^2$$

由TADMM1中的修正步，得$u^{k+1} = u^k - \alpha(u^k - \tilde{u}^k)$，于是

$$\begin{aligned} \|u^{k+1} - u^*\|_Q^2 &= \|(u^k - u^*) + (u^{k+1} - u^k)\|_Q^2 \\ &= \|(u^k - u^*) - \alpha(u^k - \tilde{u}^k)\|_Q^2 \\ &= \|u^k - u^*\|_Q^2 - 2\alpha(u^k - u^*)^TQ(u^k - \tilde{u}^k) + \alpha^2\|(u^k - \tilde{u}^k)\|_Q^2 \\ &\leqslant \|u^k - u^*\|_Q^2 - (2-\alpha)\alpha\|(u^k - \tilde{u}^k)\|_Q^2 \end{aligned}$$

证毕。

定理1 设序列$\{u^k\}_{k\geqslant 1}$由TADMM1生成,假设矩阵Q是正定,则$\{u^k\}_{k\geqslant 1}$收敛到$u^\infty\in W^*$.

证明 由引理3可知序列$\{u^k\}_{k\geqslant 1}$有界,且

$$\lim_{k\to+\infty}\|u^k-\tilde{u}^k\|_Q=0 \tag{19}$$

因此,序列$\{\tilde{u}^k\}$也有界。令u^∞是$\{\tilde{u}^k\}$的一个聚点,$\{\tilde{u}^{k_j}\}$是$\{\tilde{u}^k\}$收敛到u^∞的一个子列,即$\lim_{j\to+\infty}u^{k_j}=u^\infty$. 由最优性条件

$$\begin{cases}0\in\partial f_1(\tilde{x}_1^{k_j})-A_1^T\tilde{\lambda}^{k_j}+A_1^T(\lambda^{k_j}-\tilde{\lambda}^{k_j})+(\beta A_1^TA_1+M_1)(\tilde{x}_1^{k_j}-x_1^{k_j})+N_{\chi_1}(\tilde{x}_1^{k_j})\\0\in\partial f_2(\tilde{x}_2^{k_j})-A_2^T\tilde{\lambda}^{k_j}+A_2^T(\lambda^{k_j}-\tilde{\lambda}^{k_j})+(\beta A_2^TA_2+M_2)(\tilde{x}_2^{k_j}-x_2^{k_j})+N_{\chi_2}(\tilde{x}_2^{k_j})\\\vdots\\0\in\partial f_m(\tilde{x}_m^{k_j})-A_m^T\tilde{\lambda}^{k_j}+A_m^T(\lambda^{k_j}-\tilde{\lambda}^{k_j})+(\beta A_m^TA_m+M_m)(\tilde{x}_m^{k_j}-x_m^{k_j})+N_{\chi_m}(\tilde{x}_m^{k_j})\\0=A_1\tilde{x}_1^{k_j}+A_2\tilde{x}_2^{k_j}+\cdots+A_m\tilde{x}_m^{k_j}-b\end{cases} \tag{20}$$

令$k\to\infty$可得

$$\begin{cases}0\in\partial f_1(x_1^\infty)-A_1^T\lambda^\infty+N_{\chi_1}(x_1^\infty)\\0\in\partial f_2(x_2^\infty)-A_2^T\lambda^\infty+N_{\chi_2}(x_2^\infty)\\\vdots\\0\in\partial f_m(x_m^\infty)-A_m^T\lambda^\infty+N_{\chi_m}(x_m^\infty)\\0=A_1x_1^\infty+A_2x_2^\infty+\cdots+A_mx_m^\infty-b\end{cases} \tag{21}$$

这表明,u^∞是$L(x_1,x_2,\cdots,x_m,\lambda)$的一个鞍点,所以由引理2可知$\|u^{k+1}-u^\infty\|_Q\leqslant\|u^k-u^\infty\|_Q$,又因为$\lim_{j\to+\infty}u^{k_j}=u^\infty$,所以$\lim_{k\to+\infty}u^k=u^\infty$. 证毕。

2.2 TADMM2算法

TADMM2的具体步骤如以下算法2. 该算法同样由预测步和校正步构成,也是文献[24]中TADMM的推广。和TADMM1有类似的更新策略,不同之处在于将拉格朗日乘子的更新放在了第一步,并且在更新某一个变量时,其他变量采用原来的值。这样在第1步中的变量$x_1,x_2,\cdots,x_m$可以同时更新。

算法2(TADMM2):

第0步:给定初始点$(x_1^0,x_2^0,\cdots,x_m^0,\lambda^0)\in\chi_1\times\chi_2\times\cdots\times\chi_m\times R^l$, 参数$\beta>0,\alpha\in(0,2)$, 对称矩阵$M_i\in R^{n_i\times n_i}(i=1,2,\cdots,m)$及停机准则$\varepsilon>0$. 初始化$k:=0$.

第1步:预测

$$\begin{cases}\tilde{\lambda}^k=\lambda^k-\beta(A_1x_1^k+A_2x_2^k+\cdots+A_mx_m^k-b)\\\tilde{x}_1^k=\arg\min\limits_{x_1\in\chi_1}L_\beta(x_1,x_2^k,\cdots,x_m^k,\tilde{\lambda}^k)+\dfrac{1}{2}(x_1-x_1^k)^TM_1(x_1-x_1^k)\\\tilde{x}_2^k=\arg\min\limits_{x_2\in\chi_2}L_\beta(x_1^k,x_2,\cdots,x_m^k,\tilde{\lambda}^k)+\dfrac{1}{2}(x_2-x_2^k)^TM_2(x_2-x_2^k)\\\vdots\\\tilde{x}_i^k=\arg\min\limits_{x_i\in\chi_i}L_\beta(x_1^k,\cdots,x_{i-1}^k,x_i,x_{i+1}^k,\cdots,\tilde{\lambda}^k)+\dfrac{1}{2}(x_i-x_i^k)^TM_i(x_i-x_i^k)\\\vdots\\\tilde{x}_m^k=\arg\min\limits_{x_m\in\chi_m}L_\beta(x_1^k,\cdots,x_{m-1}^k,x_m,\tilde{\lambda}^k)+\dfrac{1}{2}(x_m-x_m^k)^TM_m(x_m-x_m^k)\end{cases} \tag{22}$$

第2步:校正

$$\begin{cases}x_1^{k+1}=x_1^k-\alpha(x_1^k-\tilde{x}_1^k)\\x_2^{k+1}=x_2^k-\alpha(x_2^k-\tilde{x}_2^k)\\\vdots\\x_m^{k+1}=x_m^k-\alpha(x_m^k-\tilde{x}_m^k)\\\lambda^{k+1}=\lambda^k-\alpha(\lambda^k-\tilde{\lambda}^k)\end{cases} \tag{23}$$

第3步：如果 $\max\left\{\frac{\|x_1^k - x_1^{k+1}\|}{1+\|x_1^k\|},\cdots,\frac{\|x_m^k - x_m^{k+1}\|}{1+\|x_m^k\|},\frac{\|\lambda^k - \lambda^{k+1}\|}{1+\|\lambda^k\|}\right\} < \varepsilon$，则停止迭代；否则令 $k:=k+1$，返回第1步。

TADMM2的收敛性的证明与TADMM1类似，省略下述结论的详细证明。由于其迭代格式不同，在引理4中会用到以下矩阵 P

$$P=\begin{pmatrix} \beta A_1^T A_1 + M_1 & 0 & \cdots & 0 & -A_1^T \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \beta A_m^T A_m + M_m & -A_m^T \\ -A_1 & -A_2 & \cdots & -A_m & \frac{1}{\beta} I \end{pmatrix} \tag{24}$$

矩阵 P 是对称的，和讨论矩阵 Q 正定条件类似，通过恰当地选择矩阵 $\{M_i\}(i=1,2,\cdots,m)$，可以使矩阵 P 正定。

引理4　设序列 $\{u^k\}_{k\geqslant 1}$ 由TADMM2生成，则有

$$(\tilde{u}^k - u^*)^T P(u^k - \tilde{u}^k) \geqslant 0, \forall u^* \in W^* \tag{25}$$

引理5　设序列 $\{u^k\}_{k\geqslant 1}$ 由TADMM2生成，则有

$$\|u^{k+1} - u^*\|_P^2 \leqslant \|u^k - u^*\|_P^2 - (2-\alpha)\alpha\|u^k - \tilde{u}^k\|_P^2, \forall u^* \in W^* \tag{26}$$

定理2　设序列 $\{u^k\}_{k\geqslant 1}$ 由TADMM2生成，且矩阵 P 是正定的，则 $\{u^k\}_{k\geqslant 1}$ 收敛到 $u^\infty \in W^*$.

3　迭代复杂度

下面讨论TADMM1和TADMM2的迭代复杂度。

定理3　设序列 $\{u^k\}_{k\geqslant 1}$ 由TADMM1生成，且矩阵 Q 是正定的，则对于任意的整数 $k>0$，有

$$\|u^k - \tilde{u}^k\|_Q^2 \leqslant \frac{1}{(k+1)(2-\alpha)\alpha}\|u^0 - u^*\|_Q^2, \forall u^* \in W \tag{27}$$

证明　设 k 为正整数。设 $z_i^k(i=2,3,\cdots,m)$ 和 $y_i^k \in \partial f_i(\tilde{x}_i^k)\,(i=1,2,\cdots,m)$ 如引理2的证明中所定义。根据引理2证明中的式(12)，有

$$\begin{cases} \langle -y_1^k + A_1^T\tilde{\lambda}^k, x_1 - \tilde{x}_1^k\rangle \leqslant 0, \forall x_1 \in \chi_1 \\ \langle -y_2^k + A_2^T\tilde{\lambda}^k + z_2^k, x_2 - \tilde{x}_2^k\rangle \leqslant 0, \forall x_2 \in \chi_2 \\ \vdots \\ \langle -y_m^k + A_m^T\tilde{\lambda}^k + z_m^k, x_m - \tilde{x}_m^k\rangle \leqslant 0, \forall x_m \in \chi_m \\ \left\langle -\sum_{i=1}^m A_i\tilde{x}_i^k + b + \sum_{i=2}^m [-A_i(x_i^k - \tilde{x}_i^k)] + \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k), \lambda - \tilde{\lambda}^k \right\rangle \leqslant 0 \end{cases} \tag{28}$$

类似地，存在 $y_i^{k+1} \in \partial f_i(\tilde{x}_i^{k+1})\,(i=1,2,\cdots,m)$ 使得

$$\begin{cases} \langle -y_1^{k+1} + A_1^T\tilde{\lambda}^{k+1}, x_1 - \tilde{x}_1^{k+1}\rangle \leqslant 0, \forall x_1 \in \chi_1 \\ \langle -y_2^{k+1} + A_2^T\tilde{\lambda}^{k+1} + z_2^{k+1}, x_2 - \tilde{x}_2^{k+1}\rangle \leqslant 0, \forall x_2 \in \chi_2 \\ \vdots \\ \langle -y_m^{k+1} + A_m^T\tilde{\lambda}^{k+1} + z_m^{k+1}, x_m - \tilde{x}_m^{k+1}\rangle \leqslant 0, \forall x_m \in \chi_m \\ \left\langle -\sum_{i=1}^m A_i\tilde{x}_i^{k+1} + b + \sum_{i=2}^m [-A_i(x_i^{k+1} - \tilde{x}_i^{k+1})] + \frac{1}{\beta}(\lambda^{k+1} - \tilde{\lambda}^{k+1}), \lambda - \tilde{\lambda}^{k+1} \right\rangle \leqslant 0 \end{cases} \tag{29}$$

在式(28)和式(29)中分别令 $(x_1,x_2,\cdots,x_m)^T = (\tilde{x}_1^{k+1},\tilde{x}_2^{k+1},\cdots,\tilde{x}_m^{k+1})^T$ 和 $(x_1,x_2,\cdots,x_m)^T = (\tilde{x}_1^k,\tilde{x}_2^k,\cdots,\tilde{x}_m^k)^T$，则有

$$\begin{cases}\langle -y_1^k + A_1^T\tilde{\lambda}^k, \tilde{x}_1^{k+1} - \tilde{x}_1^k\rangle \leqslant 0, \forall x_1 \in \chi_1 \\ \langle -y_2^k + A_2^T\tilde{\lambda}^k + z_2^k, \tilde{x}_2 - \tilde{x}_2^k\rangle \leqslant 0, \forall x_2 \in \chi_2 \\ \vdots \\ \langle -y_m^k + A_m^T\tilde{\lambda}^k + z_m^k, \tilde{x}_2 - \tilde{x}_m^k\rangle \leqslant 0, \forall x_m \in \chi_m \\ \left\langle -\sum_{i=1}^m A_i\tilde{x}_i^k + b + \sum_{i=2}^m [-A_i(x_i^k - \tilde{x}_i^k)] + \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k), \tilde{\lambda}^{k+1} - \tilde{\lambda}^k\right\rangle \leqslant 0\end{cases} \tag{30}$$

和

$$\begin{cases}\langle -y_1^{k+1} + A_1^T\tilde{\lambda}^{k+1}, \tilde{x}_1^k - \tilde{x}_1^{k+1}\rangle \leqslant 0, \forall x_1 \in \chi_1 \\ \langle -y_2^{k+1} + A_2^T\tilde{\lambda}^{k+1} + z_2^{k+1}, \tilde{x}_2^k - \tilde{x}_2^{k+1}\rangle \leqslant 0, \forall x_2 \in \chi_2 \\ \vdots \\ \langle -y_m^{k+1} + A_m^T\tilde{\lambda}^{k+1} + z_m^{k+1}, \tilde{x}_m^k - \tilde{x}_m^{k+1}\rangle \leqslant 0, \forall x_m \in \chi_m \\ \left\langle -\sum_{i=1}^m A_i\tilde{x}_i^{k+1} + b + \sum_{i=2}^m [-A_i(x_i^{k+1} - \tilde{x}_i^{k+1})] + \frac{1}{\beta}(\lambda^{k+1} - \tilde{\lambda}^{k+1}), \tilde{\lambda}^k - \tilde{\lambda}^{k+1}\right\rangle \leqslant 0\end{cases} \tag{31}$$

由式(30)和式(31)可得

$$\begin{aligned}&\sum_{i=1}^m \langle y_i^k - y_1^{k+1}, \tilde{x}_i^k - \tilde{x}_i^{k+1}\rangle + \sum_{i=1}^m \left\langle -A_i^T(\tilde{\lambda}^k - \tilde{\lambda}^{k+1}), \tilde{x}_i^k - \tilde{x}_i^{k+1}\right\rangle + \left\langle \sum_{i=1}^m (A_i\tilde{x}_i^k - A_i\tilde{x}_i^{k+1}), \tilde{\lambda}^k - \tilde{\lambda}^{k+1}\right\rangle \\ &\leqslant \sum_{i=2}^m \langle z_i^k - z_1^{k+1}, \tilde{x}_i^k - \tilde{x}_i^{k+1}\rangle + \sum_{i=2}^m \left\langle \left\{-A_i[(x_i^k - \tilde{x}_i^k) - (x_i^{k+1} - \tilde{x}_i^{k+1})]\right\}, \tilde{\lambda}^k - \tilde{\lambda}^{k+1}\right\rangle \\ &+\left\langle \frac{1}{\beta}\Big[(\lambda^k - \tilde{\lambda}^k) - (\lambda^{k+1} - \tilde{\lambda}^{k+1})\Big], \tilde{\lambda}^k - \tilde{\lambda}^{k+1}\right\rangle \\ &+\sum_{i=1}^m \left\langle (\beta A_i^T A_i + M_i)[(x_i^k - \tilde{x}_i^k) - (x_i^{k+1} - \tilde{x}_i^{k+1})], \tilde{x}_i^k - \tilde{x}_i^{k+1}\right\rangle \\ &+\sum_{i+2}^m \left\langle \left\{-A_i[(x_i^k - \tilde{x}_i^k) - (x_i^{k+1} - \tilde{x}_i^{k+1})]\right\}, \tilde{\lambda}^k - \tilde{\lambda}^{k+1}\right\rangle \\ &+\left\langle \frac{1}{\beta}\Big[(\lambda^k - \tilde{\lambda}^k) - (\lambda^{k+1} - \tilde{\lambda}^{k+1})\Big], \tilde{\lambda}^k - \tilde{\lambda}^{k+1}\right\rangle \\ &= (\tilde{u}^k - \tilde{u}^{k+1})^T Q[(u^k - \tilde{u}^k)(u^{k+1}\tilde{u}^{k+1})]\end{aligned} \tag{32}$$

由次微分 $\partial f_i(i = 1,2,\cdots,m)$ 的单调性,有

$$\sum_{i=1}^m \langle y_i^k - y_i^{k+1}, \tilde{x}_i^k - \tilde{x}_i^{k+1}\rangle \geqslant 0$$

此外,又由

$$\sum_{i=1}^m \left\langle -A_i^T(\tilde{\lambda}^k - \tilde{\lambda}^{k+1}), \tilde{x}_i^k - \tilde{x}_i^{k+1}\right\rangle + \left\langle \sum_{i=1}^m (A_i\tilde{x}_i^k - A_i\tilde{x}_i^{k+1}), \tilde{\lambda}^k - \tilde{\lambda}^{k+1}\right\rangle = 0$$

式(32)表明

$$(\tilde{u}^k - \tilde{u}^{k+1})^T Q[(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})] \geqslant 0 \tag{33}$$

则

$$(\tilde{u}^k - \tilde{u}^{k+1})^T Q[(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})] \geqslant \|(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\|_Q^2 \tag{34}$$

因为

$$\|a\|_Q^2 - \|b\|_Q^2 = 2a^TQ(a - b) - \|a - b\|_Q^2 \tag{35}$$

在式(35)中,令 $a = u^k - \tilde{u}^k$ 和 $a = u^{k+1} - \tilde{u}^{k+1}$,则可得

$$\|u^k - \tilde{u}^k\|_Q^2 - \|u^{k+1} - \tilde{u}^{k+1}\|_Q^2 = \frac{2}{\alpha}(u^k - u^{k+1})^T Q[(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})] - \|(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\|_Q^2 \tag{36}$$

由式(33)和式(36)可得

$$\|u^k - \tilde{u}^k\|_Q^2 - \|u^{k+1} - \tilde{u}^{k+1}\|_Q^2 \geqslant \frac{2-\alpha}{\alpha}\|(u^k - \tilde{u}^k) - (u^{k+1} - \tilde{u}^{k+1})\|_Q^2 \geqslant 0$$

所以$\{\|u^k - \tilde{u}^k\|_Q^2\}$单调递减。又由引理3中的式(18),有

$$\sum_{k=0}^{+\infty}(2-\alpha)\alpha\|u^k - \tilde{u}^k\|_Q^2 \leqslant \|u^0 - u^*\|_Q^2, \forall u^* \in W^*$$

所以

$$(k+1)(2-\alpha)\alpha\|u^k - \tilde{u}^k\|_Q^2 \leqslant \|u^0 - u^*\|_Q^2, \forall u^* \in W^*$$

从而式(24)成立,证毕。

类似地,关于算法TADMM2,有以下迭代复杂度估计。

定理4　设序列$\{u^k\}_{k\geqslant 1}$由TADMM2生成,且矩阵P是正定的。则对于任意的整数$k>0$,有

$$\|u^k - \tilde{u}^k\|_P^2 \leqslant \frac{1}{(k+1)(2-\alpha)\alpha}\|u^0 - u^*\|_P^2, \forall u^* \in W \tag{37}$$

定理3和定理4表明,算法TADMM1和TADMM2的迭代复杂度估计与文献[24]中TADMM类似。

4　数值实验

4.1　Fused-LASSO模型

文献[27]中的Fused-LASSO模型是用来求解稀疏数据拟合的经典模型之一。本节实验将该模型运用到手写数字1和7的识别问题,分别应用TADMM1、TADMM2及文献[24]中的TADMM来求解该问题对应的Fused-LASSO模型,并比较其数值结果。所有数值实验均在Python 3.12.0环境中实现,运行平台为配备8GB内存、1.8 GHz主频的英特尔酷睿主机。

下面先介绍Fused-LASSO模型[27]及其子问题的求解。Fused-LASSO模型为

$$\min_{x\in R^m} \frac{1}{2}\|Ax-b\|_2^2 + \mu_1\|x\|_1 + \mu_2\sum_{i=2}^{m}|x_i - x_{i-1}| \tag{38}$$

其中,$A\in R^{n\times m}$是给定的矩阵,$\mu_1,\mu_2>0$为正则参数。将模型(38)转化为以下三模块可分离凸优化问题

$$\begin{aligned}&\min_{x,y\in R^m,z\in R^{m-1}} f_1(x)+f_2(y)+f_3(z)\\ &s.t. A_1x + A_2y + A_3z = 0\end{aligned} \tag{39}$$

其中,$f_1:=\frac{1}{2}\|A(\cdot)-b\|_2^2, f_2=\mu_1\|\cdot\|_1, f_3=\mu_2\|\cdot\|_1, A_1=\begin{pmatrix}I_{m\times m}\\ L_{(m-1)\times m}\end{pmatrix}, A_2=\begin{pmatrix}-I_{m\times m}\\ 0_{(m-1)\times m}\end{pmatrix}, A_3=\begin{pmatrix}0_{m\times(m-1)}\\ -I_{(m-1)\times(m-1)}\end{pmatrix}$,$I$为单位阵,

$$L=\begin{pmatrix}-1 & 1 & 0 & \cdots & 0 & 0\\ 0 & -1 & 1 & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & -1 & 1 & 0\\ 0 & 0 & \cdots & 0 & -1 & 1\end{pmatrix}$$

对应的增广拉格朗日函数$L_\beta:R^m\times R^m\times R^{m-1}\times R^{2m-1}\to R$定义为,对任意$(x,y,z,\lambda)\in R^m\times R^m\times R^{m-1}\times R^{2m-1}$,

$$L_\beta(x,y,z,\lambda) = \min_{x,y,z}\frac{1}{2}\|Ax-b\|_2^2 + \mu_1\|y\|_1 + \mu_2\|z\|_1 + \lambda^T(A_1x+A_2y+A_3z) + \frac{\beta}{2}\|A_1x+A_2y+A_3z\|^2$$

其中,$\lambda\in R^{2m-1}$为拉格朗日乘子,$\beta>0$为二次罚项的系数。在运用TADMM1、TADMM2及TADMM求解问题(39)时,令$M_1=M_2=M_3=vI$,其中,$v>0$是一个常数。下面是TADMM1子问题解的显式表达式

(1)x-子问题

$$\begin{aligned}\tilde{x}^k &= \arg\min_x L_\beta(x,y^k,z^k,\lambda^k) + \frac{1}{2}\|x-x^k\|_{M_1}^2\\ &\Leftrightarrow \nabla x\left(L(x,y^k,z^k,\lambda^k) + \frac{1}{2}\|x-x^k\|_{M_1}^2\right)\Big|_{\tilde{x}^k} = 0\end{aligned}$$

$$\Leftrightarrow A^T(Ax-b)-A_1^T\lambda^k+\beta A_1^T(A_1x+A_2y^k+A_3z^k)+v(x-x^k)=0$$
$$\Leftrightarrow \tilde{x}^k=(A^TA+\beta A_1^TA_1+vI)^{-1}(A^Tb+A_1^T\lambda^k+vx^k-\beta(A_1^TA_2y^k+A_1^TA_3z^k))$$

(2)y–子问题

$$\tilde{y}^k=\arg\min_y L_\beta(\tilde{x}^k,y,z^k,\tilde{\lambda}^k)+\frac{1}{2}\|y-y^k\|_{M_2}^2$$
$$\Leftrightarrow \tilde{y}^k=\arg\min_y\left\{\mu_1\|y\|_1-\langle\tilde{\lambda}^k,A_1\tilde{x}^k+A_2y+A_3z^k\rangle+\frac{\beta}{2}\|A_1\tilde{x}^k+A_2y+A_3z^k\|_2^2+\frac{1}{2}\|y-y^k\|_{M_2}^2\right\}$$
$$\Leftrightarrow \tilde{y}^k=\arg\min_y\left\{\mu_1\|y\|_1-\langle\tilde{\lambda}^k,A_2^T\tilde{\lambda}^k+\beta\tilde{x}^k+vy^k\rangle+\frac{\beta+v}{2}\|y\|^2\right\}$$
$$\Leftrightarrow \tilde{y}^k=\arg\min_y\left\{\mu_1\|y\|_1+\frac{\beta+1}{2}\|y-\frac{1}{\beta+v}(A_2^T\tilde{\lambda}^k+\beta\tilde{x}^k+vy^k)\|^2\right\}$$
$$\Leftrightarrow \tilde{y}^k=\frac{1}{\beta+v}\mathrm{Sgn}(A_2^T\tilde{\lambda}^k+\beta\tilde{x}^k+vy^k)\mathrm{Max}\left\{|A_2^T\tilde{\lambda}^k+\beta\tilde{x}^k+vy^k|-\mu_1 1_m,\mathbf{0}\right\}$$

其中
$$\mathrm{Sgn}(t):=\begin{cases}1,t>0\\0,t=0,\ \forall t\in R\\-1,t<0\end{cases}$$

(3)z–子问题

$$\tilde{z}^k=\arg\min_z L_\beta(\tilde{x}^k,y^k,z,\tilde{\lambda}^k)+\frac{1}{2}\|z-z^k\|_{M_3}^2$$
$$\Leftrightarrow \tilde{z}^k=\arg\min_z\left\{\mu_2\|z\|_1-\langle\tilde{\lambda}^k,A_1\tilde{x}^k+A_2y^k+A_3z\rangle+\frac{\beta}{2}\|A_1\tilde{x}^k+A_2y^k+A_3z\|_2^2+\frac{1}{2}\|z-z^k\|_{M_3}^2\right\}$$
$$\Leftrightarrow \tilde{z}^k=\arg\min_z\left\{\mu_2\|z\|_1-\langle\tilde{\lambda}^k,A_3^T\tilde{\lambda}^k-\beta A_3^TA_1\tilde{x}^k+vz^k\rangle+\frac{\beta+v}{2}\|z\|^2\right\}$$
$$\Leftrightarrow \tilde{z}^k=\arg\min_z\left\{\mu_2\|z\|_1+\frac{\beta+v}{2}\|z-\frac{1}{\beta+1}(A_3^T\tilde{\lambda}^k-\beta A_3^TA_1\tilde{x}^k+vz^k)\|^2\right\}$$
$$\Leftrightarrow \tilde{z}^k=\frac{1}{\beta+v}\mathrm{Sgn}(A_3^T\tilde{\lambda}^k-\beta A_3^TA_1\tilde{x}^k+vz^k)\mathrm{Max}\left\{|A_3^T\tilde{\lambda}^k-\beta A_3^TA_1\tilde{x}^k+vz^k|-\mu_2 1_{m-1},\mathbf{0}\right\}$$

其中,1_m,1_{m-1}分别表示分量均为1的m和$m-1$维的向量,$\mathbf{0}$表示零向量。类似地,也可以写出TADMM2和TADMM子问题的显式解。

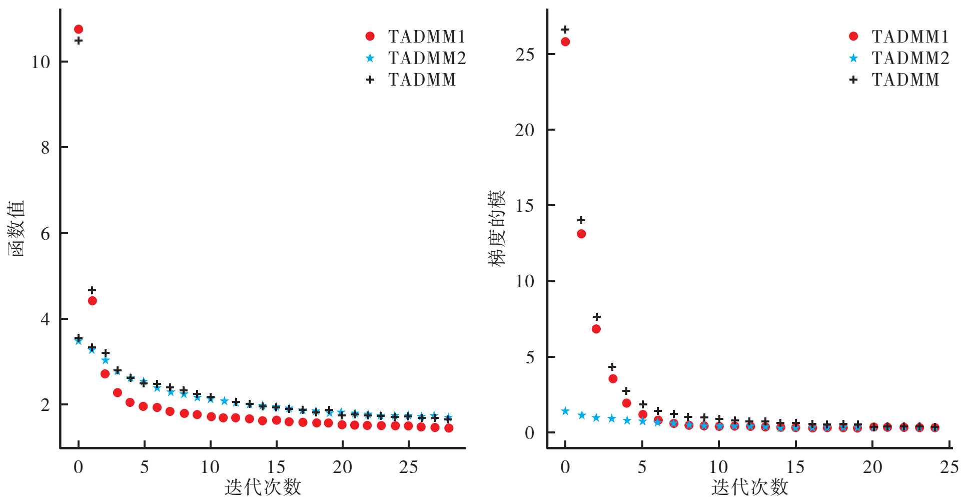
4.2 实验结果

实验中使用的数据集为MNIST数据集,由7万张图片组成,每张图片由28 × 28维的0~9的手写数字图片构成,每张图片均为黑底白字的形式,黑底用0表示,白字用0~1之间的浮点数表示,越接近1,颜色越白。将28 × 28维的图片矩阵拉直,得到1 × 784维的向量。实验中,随机选择500张数字为1和7的图片进行实验,即问题(43)中n=500, m=784.

分别运用TADMM1和TADMM2及文献[24]的算法TADMM求解问题(43). 参数设置如下:TADMM1:$\alpha=1.5,\beta=1,\mu_1=\mu_2=0.01,v=1.5$;TADMM2:$\alpha=0.9,\beta=1,\mu_1=\mu_2=0.01,v=1.5$;TADMM:$\alpha=0.5,\beta=0.5,\mu_1=\mu_2=0.01, v=1.5$. 迭代初始值为:$x_0$, y_0, z_0, λ_0均取为零向量,停机准则为$\max\left\{\frac{\|x^k-x^{k+1}\|}{1+\|x^k\|},\frac{\|y^k-y^{k+1}\|}{1+\|y^k\|},\frac{\|z^k-z^{k+1}\|}{1+\|z^k\|}\right\}<\varepsilon$. 表1是三种算法在不同停机准则下的数值比较。图1、图2是三种算法分别迭代30次和迭代50次的函数值和梯度的模的比较。图3分别是三种算法识别数字1和7的热图(其中,a、e是测试数字,b、f对应算法TADMM1的热图,c、g对应算法TADMM2的热图,d、h对应算法TADMM的热图)。实验结果表明,三种算法均具有收敛性,能有效识别出数字1和数字7. 表1和图1、图2表明,在手写数字识别问题中,与其余两种算法比较,通常算法TADMM1效率最高。

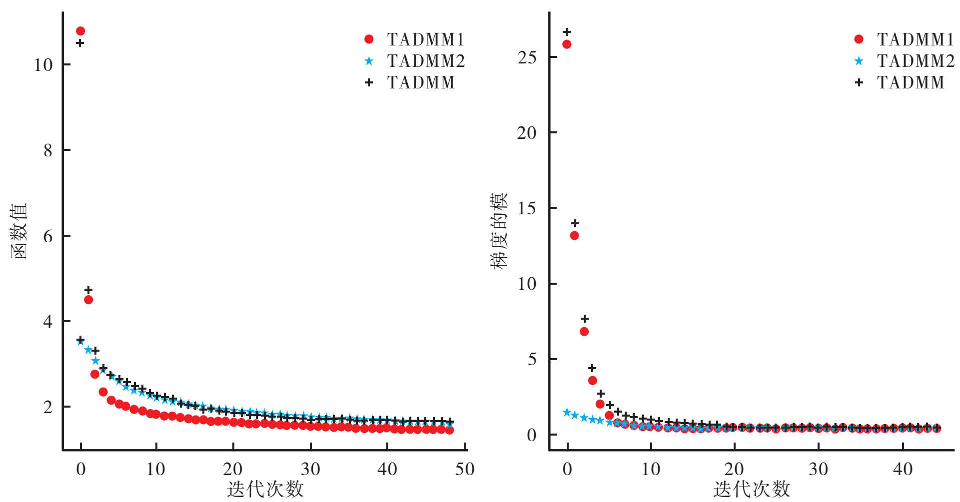
表 1　三种算法在不同停机准则下的数值比较

ε	算法	梯度的模	函数值	迭代次数	耗时(秒)
	TADMM1	0.38946	1.47632	29	4.364
0.010	TADMM2	0.36463	1.66849	33	6.020
	TADMM	0.44343	1.66539	31	5.586
	TADMM1	0.37182	1.45423	32	4.905
0.009	TADMM2	0.35858	1.64764	35	6.270
	TADMM	0.41050	1.63221	34	6.056
	TADMM1	0.34744	1.36183	53	8.216
0.005	TADMM2	0.28582	1.49887	58	9.969
	TADMM	0.30812	1.48637	56	9.787
	TADMM1	0.32169	1.26019	117	18.139
0.002	TADMM2	0.24586	1.34630	130	21.370
	TADMM	0.23593	1.33483	128	21.038



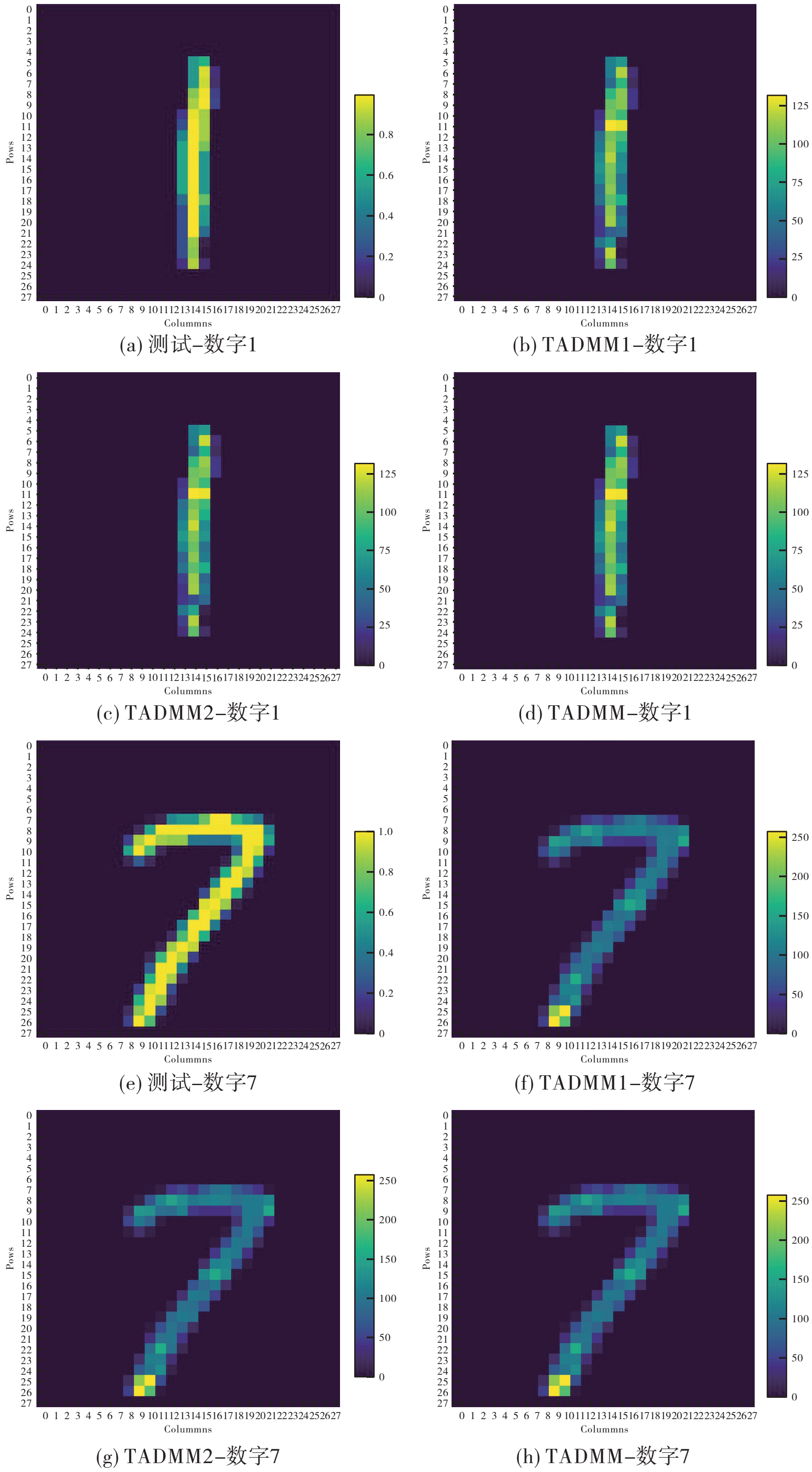
(a) 函数值–迭代次数　　(b) 梯度的模–迭代次数

图 1　迭代次数为 30 时,三种算法的函数值和梯度的模比较



(a) 函数值–迭代次数　　(b) 梯度的模–迭代次数

图 2　迭代次数为 50 时,三种算法的函数值和梯度的模比较



(a) 测试-数字1 (b) TADMM1-数字1

(c) TADMM2-数字1 (d) TADMM-数字1

(e) 测试-数字7 (f) TADMM1-数字7

(g) TADMM2-数字7 (h) TADMM-数字7

图3 三种算法的手写数字识别热图

参考文献:

[1] GLOWINSKI R, MARROCCO A. Sur L'approximation, Par Éléments Finis D'ordren, et La Résolution, Par Pénalisation-Dualité

D'une Classe de Problèmes de Dirichlet Nonlinéares[J]. Revue Française d'Automatique, Informatique, Recherche Opérationnelle. Analyse Numérique, 1975, 9(R2): 41-76.

[2] BOYD S, PARIKH N, CHU E, et al. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers[J]. Foundations and Trends® in Machine Learning, 2011, 3(01): 1-122.

[3] RYU E K, YIN W. Large-scale Convex Optimization: Algorithms and Analyses via Monotone Operators [M]. Cambridge: Cambridge University Press, 2023.

[4] FORTIN M, GLOWINSKI R. Augmented Lagrangian Methods[M]. Amsterdam: North-Holland, 1983.

[5] GLOWINSKI R. Lectures on Numerical Methods for Nonlinear Variational Problems[M]. Bombay: Tata Institute of Fundamental Research, 1980.

[6] CHAN T F, GLOWINSKI R. Finite Element Approximation and Iterative Solution of a Class of Mildly Nonlinear Elliptic Equations[R]. Stanford: Stanford University, 1978.

[7] ECKSTEIN J. Augmented Lagrangian and Alternating Direction Methods for Convex Optimization: A Tutorial and Some Illustrative Computational Results[R]. New Jersey: Rutgers University, 2012.

[8] GABAY D, MERCIER B. A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite-element Approximations[J]. Computers & Mathematics with Applications, 1976, 2: 17-40.

[9] HE B S, LIU H, WANG Z R, et al. A Strictly Peaceman-Rachford Splitting Method for Convex Programming[J]. SIAM Journal on Optimization, 2014, 24: 1011-1040.

[10] LIN Z, CHEN M, WU L, et al. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-rank Matrices[J]. Mathematical Programming, 2010.

[11] SUN H, TAI X C, YUAN J. Efficient and Convergent Preconditioned ADMM for the Potts Models[J]. SIAM Journal on Scientific Computing, 2021: B455-B478. doi: 10.1137/20M1343956.

[12] XIONG Z, KANG J, NIYATO D, et al. Cloud/edge Computing Service Management in Blockchain Networks: Multi-leader Multi-follower Game-based ADMM for Pricing[J]. IEEE Transactions on Services Computing, 2019: 1.

[13] 刘浩洋,卢将,李勇锋,等. 最优化:建模,算法与理论[M]. 北京:高等教育出版社,2020.

[14] CHEN C H, HE B S, YE Y Y, et al. The Direct Extension of ADMM for Multi-block Convex Minimization Problems is not Necessarily Convergent[J]. Mathematical Programming, 2016, 155(1-2): 57-79.

[15] HE B S, TAO M, YUAN X M. Alternating Direction Method with Gaussian Back Substitution for Separable Convex Programming[J]. SIAM Journal on Optimization, 2012, 22: 313-340.

[16] HE B S, TAO M, YUAN X M. Convergence Rate and Iteration Complexity on the Alternation Direction Method of Multipliers with a Substitution Procedure for Separable Convex Programming[R]. Manuscript.

[17] ECKSTEIN J. Some Saddle-function Splitting Methods for Convex Programming[J]. Optimization Methods and Software, 1994, 4: 75-83.

[18] HE B S, LIAO L Z, HAN D, et al. A New Inexact Alternating Direction Method for Monotone Variational Inequalities [J]. Mathematical Programming, 2002, 92: 103-118.

[19] XU M H, WU T. A Class of Linearized Proximal Alternating Direction Methods [J]. Journal of Optimization Theory and Applications, 2011, 155: 321-337.

[20] FAZEL M, PONG T K, SUN D F, et al. Hankel Matrix Rank Minimization with Application to System Identification and Realization[J]. SIAM Journal on Matrix Analysis and Applications, 2013, 34: 946-977.

[21] CHEN G, TEBOULLE M. A Proximal-based Decomposition Method for Convex Minimization Problems [J]. Mathematical Programming, 1994, 64: 81-101.

[22] HONG M Y, LUO Z Q. On the Linear Convergence of the Alternating Direction Method of Multipliers[R]. arXiv: 1208.3922, 2012.

[23] YANG J, ZHANG Y. Alternating Direction Algorithms for Problems in Compressive Sensing [J]. SIAM Journal on Scientific Computing, 2011, 33: 250-278.

[24] WANG J J, SONG W. An Algorithm Twisted from Generalized ADMM for Multi-block Separable Convex Minimization Models[J]. Journal of Computational and Applied Mathematics, 2017, 309: 342-358.

[25] ROCKAFELLAR R T. Convex Analysis[M]. Princeton: Princeton University Press, 1970.

[26] BOYD S, VANDENBERGHE L. Convex Optimization[M]. Cambridge: Cambridge University Press, 2004.

[27] TIBSHIRANI R, SAUNDERS M, ROSSET S, et al. Sparsity and Smoothness via the Fused Lasso [J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2005, 67(01): 91-108.

Proximal ADMMs for Multi-block Separable Convex Optimizations and Convergences

LUO Xiu, WANG Xiang-mei*, LIAO Chun-fang

(*College of Mathematics and Statistics, Guizhou University, Guiyang, Guizhou, 550025, China*)

Abstract: The Alternating Direction Method of Multipliers (ADMM) is one of the important methods for solving two-block separable convex optimization problems and has made significant progress in recent years. Recently, Wang et al. proposed a Twisted version of the proximal Alternating Direction Method of Multipliers (TADMM) to solve multi-block separable problems (involving three or more blocks). Building on this work, the present study introduces two new TADMM variants, and systematically analyzes their convergence behaviors and iteration complexities. Finally, these algorithms are implemented to solve LASSO models in the handwritten digit recognition problems. Numerical experiments confirm the convergence of all three methods (the original TADMM and the two proposed variants) and further demonstrate the first variant proposed herein outperforms the other two by a significant margin.

Keywords: Separable convex optimization; Augmented Lagrangian function; Proximal ADMM; Convergence; Handwritten digit recognition problem

---

(上接第84页)

Large Deviation Principle of Hilfer Fractional Stochastic Delay Differential Equation

CHEN Juan, GU Hai-bo*, YAN Yu-tao

(*College of Mathematical Sciences, Xinjiang Normal University, Urumqi, Xinjiang, 830017, China*)

Abstract: This paper investigates the Large Deviation Principle (LDP) for a class of Hilfer fractional stochastic delay differential equations with multiplicative noise. Based on the equivalence between the LDP and the Laplace principle, the weak convergence method is employed to establish the LDP for such equations by verifying the compactness of the skeleton equation and the convergence of the perturbed system. A financial market model is presented as an example, and numerical simulations are conducted to illustrate how the LDP can be used to analyze the asymptotic behavior of the system under small noise perturbations, thereby validating the theoretical results.

Keywords: Fractional calculus; Stochastic differential equation; Large Deviation Principle; Weak convergence