

求解凸优化问题的改进对称交替方向乘子法

蒋 峰, 党亚峥

(上海理工大学 管理学院 上海 20093)

摘要: 对称交替方向乘子法 (简称 S-ADMM 算法) 是求解可分离凸优化问题的一种有效方法。该算法利用目标函数的可分离性, 将原问题分解成多个极小化子问题, 然后交替求解。能否有效地求解子问题对算法的有效性有重要影响。在很多实际应用中, 不能精确地求解子问题, 或者精确求解子问题花费代价较大。为解决这一问题, 提出了一种改进的对称交替方向乘子法 (简称 MS-ADMM 算法)。与一般的 S-ADMM 算法相比, 该算法在 x 子问题中引入一个半近邻项, 近似地求解 x 子问题, 克服了之前算法的不足。在适当的假设下, 证明了其收敛性。最后, 通过数值计算说明了该算法的有效性。

关键词: 凸优化; 改进的对称交替方向乘子法; 收敛性

中图分类号: O 224 **文献标志码:** A

A modified symmetric alternating direction method of multipliers for convex optimization problems

JIANG Feng, DANG Yazheng

(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Symmetric alternating direction method of multipliers (S-ADMM) is an efficient method for convex optimization problems with separable structure. The algorithm makes use of the separability of the objective function to decompose the original problem into several minimization subproblems and to solve them alternately. Whether the subproblems can be effectively solved affects the effectiveness of the algorithm. In many practical applications, subproblems cannot be solved precisely, or the cost of solving subproblems precisely is relatively high. To solve this problem, a modified symmetric alternating direction method of multipliers (MS-ADMM) is proposed. Compared to the general symmetric ADMM, this algorithm adds a semi-proximal term to x -subproblem which is then solved approximately. This overcomes the shortcoming of the previous algorithm. The convergence of the sequence generated by the proposed algorithm is proved under some suitable assumptions. Preliminary numerical experiments illustrate the effectiveness of proposed algorithm.

Keywords: convex optimization; modified symmetric alternating direction method of multipliers;

收稿日期: 2019-05-30

基金项目: 上海市自然科学基金资助项目 (17ZR1419000); 河南省科技攻关项目 (172102310252)

第一作者: 蒋 峰 (1994-), 男, 硕士研究生. 研究方向: 机器学习. E-mail: jfjsa@163.com

通信作者: 党亚峥 (1973-), 女, 副教授. 研究方向: 算法优化、智能电网. E-mail: jgdzyz@163.com

convergence

1 问题的提出

考虑下面的凸优化问题:

$$\min \{f(x)+g(z)|Ax=z, x\in X, z\in Z\} \tag{1}$$

式中: $A\in\mathbb{R}^{m\times n}$; $X\subset\mathbb{R}^n$; $Z\subset\mathbb{R}^m$ 是给定闭凸集; m, n 为矩阵的维数; $f:\mathbb{R}^n\rightarrow\mathbb{R}\cup\{+\infty\}$ 和 $g:\mathbb{R}^m\rightarrow\mathbb{R}\cup\{+\infty\}$ 为凸函数。假设问题(1)的解集非空。

Glowinski 等^[1-2]提出了交替方向乘子法(ADMM), 并用此算法来求解问题(1), 其迭代格式如下:

$$\begin{cases} x^{k+1}=\operatorname{argmin}\left\{f(x)-\left(y^k\right)^T\left(Ax-z^k\right)+\frac{\gamma}{2}\left\|Ax-z^k\right\|^2\right\} \\ z^{k+1}=\operatorname{argmin}\left\{g(z)-\left(y^k\right)^T\left(Ax^{k+1}-z\right)+\frac{\gamma}{2}\left\|Ax^{k+1}-z\right\|^2\right\} \\ y^{k+1}=y^k-\gamma\left(Ax^{k+1}-z^{k+1}\right) \end{cases} \tag{2}$$

式中: $y\in\mathbb{R}^m$ 为拉格朗日乘子; $\gamma>0$ 是惩罚参数。与同时求解 x, z 极小化问题的拉格朗日乘子法不同, ADMM 算法先求解 x 最小化问题, 再求解 z 最小化问题, 最后再更新拉格朗日乘子 y , 这很大程度提升了求解问题(1)的效率。目前, ADMM 算法已经被广泛应用于图像处理、机器学习和金融统计等领域^[3], 在实际应用中也衍生了许多 ADMM 算法的变体。

因为经典的 ADMM 算法难以求得子问题的精确解, 为了弥补这一缺陷, 何炳生等^[4]在 x 子问题中加入了 $\frac{1}{2}(x-x^k)^T G(x-x^k)$ 项, 得到了下面的新算法:

$$\begin{cases} x^{k+1}=\operatorname{argmin}\left\{f(x)-\left(y^k\right)^T\left(Ax-z^k\right)+\frac{\gamma}{2}\left\|Ax-z^k\right\|^2+\frac{1}{2}\left(x-x^k\right)^T G\left(x-x^k\right)\right\} \\ z^{k+1}=\operatorname{argmin}\left\{g(z)-\left(y^k\right)^T\left(Ax^{k+1}-z\right)+\frac{\gamma}{2}\left\|Ax^{k+1}-z\right\|^2\right\} \\ y^{k+1}=y^k-\gamma\left(Ax^{k+1}-z^{k+1}\right) \end{cases} \tag{3}$$

式中, G 是一个半正定矩阵。何炳生等^[4]证明了该算法的收敛速率。

从问题(1)本身来看, 变量 x 和 z 是平等的, 在设计算法时也希望能够平等对待 x 和 z 子问题。因此何炳生等^[5]提出了对称 ADMM 算法, 其迭代

格式如下:

$$\begin{cases} x^{k+1}=\operatorname{argmin}\left\{f(x)-\left(y^k\right)^T\left(Ax-z^k\right)+\frac{\gamma}{2}\left\|Ax-z^k\right\|^2\right\} \\ y^{k+\frac{1}{2}}=y^k-\gamma\left(Ax^{k+1}-z^k\right) \\ z^{k+1}=\operatorname{argmin}\left\{g(z)-\left(y^{k+\frac{1}{2}}\right)^T\left(Ax^{k+1}-z\right)+\frac{\gamma}{2}\left\|Ax^{k+1}-z\right\|^2\right\} \\ y^{k+1}=y^{k+\frac{1}{2}}-\gamma\left(Ax^{k+1}-z^{k+1}\right) \end{cases} \tag{4}$$

与原来的 ADMM 算法不同, 式(4)在每一次迭代中更新拉格朗日乘子两次。文献[4]中分析了该算法的全局收敛性, 数值计算表明该算法比原始的 ADMM 算法收敛速度更快。

然而, 在很多实际应用中, 精确地求解式(4)中的 x 子问题难以实现, 或者要付出很大代价。因此本文提出了一种改进的对称 ADMM 算法。该算法的主要创新之处是在对称 ADMM 算法的 x 极小化子问题中加入半近邻项近似求解此问题, 同时给出了收敛性证明。数值计算表明, 改进的对称 ADMM 算法的收敛速度比对称 ADMM 算法更快。

2 算 法

问题(1)等价于一个变分不等式 $VI(\Omega, F, \theta)$, 求

$$\left(x^*, z^*, y^*\right)^T=\boldsymbol{w}^* \in \Omega$$

满足

$$\theta(\boldsymbol{u})-\theta\left(\boldsymbol{u}^*\right)+\left(\boldsymbol{w}-\boldsymbol{w}^*\right)^T \boldsymbol{F}\left(\boldsymbol{w}^*\right) \geq 0, \forall \boldsymbol{w} \in \Omega \tag{5}$$

式中: $\boldsymbol{u}=\left(\begin{smallmatrix} x \\ z \end{smallmatrix}\right)$; $\boldsymbol{w}=\left(\begin{smallmatrix} x \\ z \\ y \end{smallmatrix}\right)$; $\boldsymbol{F}(\boldsymbol{w})=\left(\begin{smallmatrix} -\boldsymbol{A}^T \boldsymbol{y} \\ y \\ \boldsymbol{A} x-\boldsymbol{z} \end{smallmatrix}\right)$; $\theta(\boldsymbol{u})=f(x)+g(z)$; F 是单调的。定义 Ω^* 为 $VI(\Omega, F, \theta)$ 的解集, 在问题(1)有解的假设下, Ω^* 非空。

本文在对称 ADMM 算法的 x 子问题中加入了半近邻项, 提出了改进的对称 ADMM 算法。下面详述该算法。

步骤 1 给定半正定矩阵 G , 初始点 $\left(x^0, z^0, y^0\right) \in X \times Z \times \mathbb{R}^m, \alpha \in(0,1), \gamma>0$, 置 $k=0$ 。

步骤 2 计算

$$\begin{cases} x^{k+1} = \operatorname{argmin} \left\{ f(x) - (y^k)^T (Ax - z^k) + \frac{\gamma}{2} \|Ax - z^k\|^2 + \frac{1}{2} \|x - x^k\|_G^2 \right\} \\ y^{k+\frac{1}{2}} = y^k - \alpha\gamma (Ax^{k+1} - z^k) \\ z^{k+1} = \operatorname{argmin} \left\{ g(z) - (y^{k+\frac{1}{2}})^T (Ax^{k+1} - z) + \frac{\gamma}{2} \|Ax^{k+1} - z\|^2 \right\} \\ y^{k+1} = y^{k+\frac{1}{2}} - \alpha\gamma (Ax^{k+1} - z^{k+1}) \end{cases}$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & \sqrt{\gamma}I_m & 0 \\ 0 & 0 & \sqrt{\frac{1}{\gamma}}I_m \end{pmatrix}$$

又因为 $\alpha \in (0, 1)$, 矩阵

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & (2-\alpha) & 1 \\ 0 & 1 & \frac{1}{\alpha} \end{pmatrix}$$

是半正定矩阵, 故引理1成立。

步骤3 如果满足终止条件, 停止迭代返回运行结果; 否则, $k = k + 1$, 返回步骤2。

特别地, 当 $G = 0$ 时, 算法为对称ADMM算法。

3 收敛性分析

首先, 定义几个分析算法收敛性质需要用到的矩阵。

$$\begin{cases} H = \begin{pmatrix} 2G & 0 & 0 \\ 0 & (2-\alpha)\gamma I_m & I_m \\ 0 & I_m & \frac{1}{\alpha\gamma}I_m \end{pmatrix} \\ M = \begin{pmatrix} I_n & 0 & 0 \\ 0 & I_m & 0 \\ 0 & \alpha\gamma I_m & 2\alpha I_m \end{pmatrix} \\ Q = \begin{pmatrix} G & 0 & 0 \\ 0 & \gamma I_m & \alpha I_m \\ 0 & I_m & \frac{1}{\gamma}I_m \end{pmatrix} \end{cases} \quad (6)$$

引理1 当 $\alpha \in (0, 1)$, G 是半正定矩阵时, 式(6)定义的矩阵 H 是半正定矩阵。

证明

$$H = \begin{pmatrix} G & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & (2-\alpha)\gamma I_m & I_m \\ 0 & I_m & \frac{1}{\alpha\gamma}I_m \end{pmatrix}$$

因为 G 是半正定矩阵, 所以等号右侧的第一部分是半正定矩阵, 只需证明第二部分也是半正定即可。令

$$H_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & (2-\alpha)\gamma I_m & I_m \\ 0 & I_m & \frac{1}{\alpha\gamma}I_m \end{pmatrix}$$

因为

$$H_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sqrt{\gamma}I_m & 0 \\ 0 & 0 & \sqrt{\frac{1}{\gamma}}I_m \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & (2-\alpha)I_m & I_m \\ 0 & I_m & \frac{1}{\alpha}I_m \end{pmatrix}.$$

引理2 假设 H, M, Q 是式(6)定义的矩阵, 则有

a. $HM = Q$;

b. $Q^T + Q - M^T H M \geq \frac{1-\alpha}{2(1+\alpha)} M^T H M$ 。

证明 利用矩阵 H, Q, M 的定义, 通过简单计算可得

$$HM = \frac{1}{2} \begin{pmatrix} 2G & 0 & 0 \\ 0 & (2-\alpha)\gamma I_m & I_m \\ 0 & I_m & \frac{1}{\alpha\gamma}I_m \end{pmatrix} \begin{pmatrix} I_n & 0 & 0 \\ 0 & I_m & 0 \\ 0 & \alpha\gamma I_m & 2\alpha I_m \end{pmatrix} = \begin{pmatrix} G & 0 & 0 \\ 0 & \gamma I_m & \alpha I_m \\ 0 & I_m & \frac{1}{\gamma}I_m \end{pmatrix}$$

$HM = Q$ 得证。再次使用 H, Q, M 的定义, 有

$$M^T H M = M^T Q = \begin{pmatrix} G & 0 & 0 \\ 0 & \gamma(1+\alpha)I_m & 2\alpha I_m \\ 0 & 2\alpha I_m & \frac{2\alpha}{\gamma}I_m \end{pmatrix} \quad (7)$$

由式(6)和式(7)可得

$$Q^T + Q - M^T H M = (1-\alpha) \begin{pmatrix} \frac{1}{1-\alpha}G & 0 & 0 \\ 0 & \gamma I_m & I_m \\ 0 & I_m & \frac{2}{\gamma}I_m \end{pmatrix} \quad (8)$$

此外

$$D = 2(1+\alpha) \begin{pmatrix} \frac{1}{1-\alpha}G & 0 & 0 \\ 0 & \gamma I_m & I_m \\ 0 & I_m & \frac{2}{\gamma}I_m \end{pmatrix} - M^T H M = \begin{pmatrix} \frac{1+3\alpha}{1-\alpha}G & 0 & 0 \\ 0 & \gamma(1+\alpha)I_m & 2I_m \\ 0 & 2I_m & \frac{4+2\alpha}{\gamma}I_m \end{pmatrix} \quad (9)$$

因为

$$\mathbf{D} = \begin{pmatrix} \frac{1+3\alpha}{1-\alpha}\mathbf{G} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \gamma(1+\alpha)\mathbf{I}_m & 2\mathbf{I}_m \\ 0 & 2\mathbf{I}_m & \frac{4+2\alpha}{\gamma}\mathbf{I}_m \end{pmatrix} \quad (10)$$

因为 $\alpha \in (0, 1)$, $\mathbf{G}$ 是半正定矩阵, 所以式(10)中等号右侧第一部分是正定矩阵。又因为矩阵

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & \gamma(1+\alpha) & 2 \\ 0 & 2 & \frac{4+2\alpha}{\gamma} \end{pmatrix} \geq 0$$

因此 $\mathbf{D}$ 是半正定矩阵。所以

$$\begin{pmatrix} \frac{1}{1-\alpha}\mathbf{G} & 0 & 0 \\ 0 & \gamma\mathbf{I}_m & \mathbf{I}_m \\ 0 & \mathbf{I}_m & \frac{2}{\gamma}\mathbf{I}_m \end{pmatrix} \geq \frac{1}{2(1+\alpha)}\mathbf{M}^T\mathbf{H}\mathbf{M} \quad (11)$$

根据式(8)和式(11)有

$$\mathbf{Q}^T + \mathbf{Q} - \mathbf{M}^T\mathbf{H}\mathbf{M} \geq \frac{1-\alpha}{2(1+\alpha)}\mathbf{M}^T\mathbf{H}\mathbf{M}$$

设 $\{\mathbf{w}^k\}$ 是由改进的对称 ADMM 算法产生的序列, 定义一个新的序列

$$\tilde{\mathbf{w}}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{z}^k \\ \tilde{y}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ z^{k+1} \\ y^k - \gamma(\mathbf{A}x^{k+1} - z^k) \end{pmatrix} \quad (12)$$

由式(12)和 $\mathbf{M}$ 的定义可得

$$\mathbf{w}^{k+1} = \mathbf{w}^k - \mathbf{M}(\mathbf{w}^k - \tilde{\mathbf{w}}^k) \quad (13)$$

引理 3 设 $\{\mathbf{w}^k\}$ 是由改进的对称 ADMM 产生的序列, $\mathbf{Q}$ 为式(6)定义的矩阵。那么对于 $\forall \mathbf{w} \in \Omega$, 有

$$\theta(\mathbf{u}) - \theta(\tilde{\mathbf{u}}^k) + (\mathbf{w} - \tilde{\mathbf{w}}^k)^T \{ \mathbf{F}(\tilde{\mathbf{w}}^k) - \mathbf{Q}(\mathbf{w}^k - \tilde{\mathbf{w}}^k) \} \geq 0, \quad \tilde{\mathbf{w}}^k \in \Omega \quad (14)$$

证明 通过推导 x 极小化问题的最优性条件可得

$$\begin{aligned} f(x) - f(\tilde{x}^k) + (x - \tilde{x}^k)^T \cdot \\ \{ \mathbf{A}^T [\gamma(\mathbf{A}\tilde{x}^k - z^k) - y^k] + \mathbf{G}(\tilde{x}^k - x^k) \} \geq 0, \\ \forall x \in X \end{aligned} \quad (15)$$

因为 $\tilde{y}^k = y^k - \gamma(\mathbf{A}\tilde{x}^k - z^k)$, 上式可被写为

$$\begin{aligned} f(x) - f(\tilde{x}^k) + (x - \tilde{x}^k)^T \cdot \\ \{ -\mathbf{A}^T \tilde{y}^k + \mathbf{G}(\tilde{x}^k - x^k) \} \geq 0, \quad \forall x \in X \end{aligned} \quad (16)$$

同理, 根据 z 极小化问题的最优性条件, 可得

$$\begin{aligned} g(z) - g(\tilde{z}^k) + (z - \tilde{z}^k)^T \cdot \\ \{ -\gamma(\mathbf{A}x^{k+1} - \tilde{z}^k) + y^{k+\frac{1}{2}} \} \geq 0, \quad \forall z \in Z \end{aligned} \quad (17)$$

由式(12)的定义可知

$$\begin{aligned} -\gamma(\mathbf{A}x^{k+1} - \tilde{z}^k) + y^{k+\frac{1}{2}} = \\ \tilde{y}^k + \alpha(\tilde{y}^k - y^k) + \gamma(\tilde{z}^k - z^k) \end{aligned} \quad (18)$$

所以, 式(17)变为

$$\begin{aligned} g(z) - g(\tilde{z}^k) + (z - \tilde{z}^k)^T \cdot \\ \{ \tilde{y}^k + \gamma(\tilde{z}^k - z^k) + \alpha(\tilde{y}^k - y^k) \} \geq 0, \quad \forall z \in Z \end{aligned} \quad (19)$$

由式(12)可得

$$(\mathbf{A}\tilde{x}^k - \tilde{z}^k) + (\tilde{z}^k - z^k) + \frac{1}{\gamma}(\tilde{y}^k - y^k) = 0 \quad (20)$$

结合式(16)、式(19)和式(20), 得

$$\begin{aligned} \theta(\mathbf{u}) - \theta(\tilde{\mathbf{u}}^k) + (\mathbf{w} - \tilde{\mathbf{w}}^k)^T \cdot \\ \left\{ \begin{pmatrix} -\mathbf{A}^T \tilde{y}^k \\ \tilde{y}^k \\ \mathbf{A}\tilde{x}^k - \tilde{z}^k \end{pmatrix} + \begin{pmatrix} \mathbf{G}(\tilde{x}^k - x^k) \\ \gamma(\tilde{z}^k - z^k) + \alpha(\tilde{y}^k - y^k) \\ (\tilde{z}^k - z^k) + \frac{1}{\gamma}(\tilde{y}^k - y^k) \end{pmatrix} \right\} \geq 0 \end{aligned} \quad (21)$$

因为

$$\begin{pmatrix} \mathbf{G}(\tilde{x}^k - x^k) \\ \gamma(\tilde{z}^k - z^k) + (\tilde{y}^k - y^k) \\ (\tilde{z}^k - z^k) + \frac{1}{\gamma}(\tilde{y}^k - y^k) \end{pmatrix} = -\mathbf{Q}(\mathbf{w}^k - \tilde{\mathbf{w}}^k)$$

所以式(21)可被写成

$$\theta(\mathbf{u}) - \theta(\tilde{\mathbf{u}}^k) + (\mathbf{w} - \tilde{\mathbf{w}}^k)^T \{ \mathbf{F}(\tilde{\mathbf{w}}^k) - \mathbf{Q}(\mathbf{w}^k - \tilde{\mathbf{w}}^k) \} \geq 0$$

故引理 3 得证。

引理 4 设 $\{\mathbf{w}^k\}$ 是由改进的对称 ADMM 算法产生的序列, $\{\tilde{\mathbf{w}}^k\}$ 是式(12)定义的序列, $\mathbf{H}$, $\mathbf{M}$, $\mathbf{Q}$ 是式(6)定义的矩阵。那么对于 $\forall \mathbf{w} \in \Omega$, 有

$$\begin{aligned} (\mathbf{w} - \tilde{\mathbf{w}}^k)^T \mathbf{Q}(\mathbf{w}^k - \tilde{\mathbf{w}}^k) \geq \\ \frac{1}{2} (\|\mathbf{w} - \mathbf{w}^k\|_{\mathbf{H}}^2 - \|\mathbf{w} - \mathbf{w}^{k+1}\|_{\mathbf{H}}^2) + \\ \frac{1-\alpha}{2(1+\alpha)} \|\mathbf{w}^k - \mathbf{w}^{k+1}\|_{\mathbf{H}}^2 \end{aligned} \quad (22)$$

证明 对于同一空间中的向量 $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ 和具有适当维度的矩阵 $\mathbf{H}$, 满足

$$\begin{aligned} (\mathbf{a} - \mathbf{b})^T \mathbf{H}(\mathbf{c} - \mathbf{d}) = \frac{1}{2} (\|\mathbf{a} - \mathbf{d}\|_{\mathbf{H}}^2 - \|\mathbf{a} - \mathbf{c}\|_{\mathbf{H}}^2) + \\ \frac{1}{2} (\|\mathbf{c} - \mathbf{b}\|_{\mathbf{H}}^2 - \|\mathbf{d} - \mathbf{b}\|_{\mathbf{H}}^2) \end{aligned}$$

令

$$\mathbf{a} = \mathbf{w}, \quad \mathbf{b} = \tilde{\mathbf{w}}^k, \quad \mathbf{c} = \mathbf{w}^k, \quad \mathbf{d} = \mathbf{w}^{k+1}$$

上式可写为

$$\begin{aligned} & (w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k) = \\ & \frac{1}{2} \left(\|w - w^{k+1}\|_H^2 - \|w - w^k\|_H^2 \right) + \\ & \frac{1}{2} \left(\|w^k - \tilde{w}^k\|_H^2 - \|w^{k+1} - \tilde{w}^k\|_H^2 \right) \end{aligned} \quad (23)$$

因为 $w^{k+1} = w^k - M(w^k - \tilde{w}^k)$, 所以有

$$\begin{aligned} & \|w^k - \tilde{w}^k\|_H^2 - \|w^{k+1} - \tilde{w}^k\|_H^2 = \\ & \|w^k - \tilde{w}^k\|_H^2 - \left\| (w^k - \tilde{w}^k) - (w^k - w^{k+1}) \right\|_H^2 = \\ & \|w^k - \tilde{w}^k\|_H^2 - \left\| (w^k - \tilde{w}^k) - M(w^k - \tilde{w}^k) \right\|_H^2 = \\ & (w^k - \tilde{w}^k)^T (Q^T + Q - M^T H M) (w^k - \tilde{w}^k) \geq \\ & \frac{1-\alpha}{2(1+\alpha)} (w^k - \tilde{w}^k)^T (Q^T + Q - M^T H M) (w^k - \tilde{w}^k) = \\ & \frac{1-\alpha}{2(1+\alpha)} \|w^k - w^{k+1}\|_H^2 \end{aligned} \quad (24)$$

将式(24)代入式(23), 式(22)成立。故引理4得证。

定理1 假设 $\{w^k\}$ 和 $\{\tilde{w}^k\}$ 是由改进的对称 ADMM 算法产生的序列, 那么对于 $\forall w \in \Omega$, 有

$$\begin{aligned} & \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \geq \\ & \frac{1}{2} \left(\|w - w^k\|_H^2 - \|w - w^{k+1}\|_H^2 \right) + \\ & \frac{1-\alpha}{2(1+\alpha)} \|w^k - w^{k+1}\|_H^2 \end{aligned} \quad (25)$$

证明 根据式(14)和式(22)得

$$\begin{aligned} & \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq \\ & \frac{1}{2} \left(\|w - w^k\|_H^2 - \|w - w^{k+1}\|_H^2 \right) + \\ & \frac{1-\alpha}{2(1+\alpha)} \|w^k - w^{k+1}\|_H^2 \end{aligned} \quad (26)$$

因为 $F(\cdot)$ 是单调的, 则

$$(w - \tilde{w}^k)^T \{F(w) - F(\tilde{w}^k)\} \geq 0 \quad (27)$$

将式(26)和式(27)相加, 可得式(25)。

定理2 假设 $\{w^k\}$ 和 $\{\tilde{w}^k\}$ 是由改进的对称 ADMM 算法产生的序列, $\alpha \in (0, 1)$, $\forall w^* \in \Omega^*$, 则有

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \frac{1-\alpha}{2(1+\alpha)} \|w^k - w^{k+1}\|_H^2$$

证明 令 $w = w^*$, 式(25)变为

$$\begin{aligned} & \|w^k - w^*\|_H^2 - \frac{1-\alpha}{2(1+\alpha)} \|w^k - w^{k+1}\|_H^2 \geq \\ & 2 \left\{ \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \right\} + \\ & \|w^{k+1} - w^*\|_H^2 \end{aligned} \quad (28)$$

因为 $w^* \in \Omega^*$, 所以

$$\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \geq 0$$

因此

$$\begin{aligned} & \|w^{k+1} - w^*\|_H^2 \leq \\ & \|w^k - w^*\|_H^2 - \frac{1-\alpha}{2(1+\alpha)} \|w^k - w^{k+1}\|_H^2 \end{aligned} \quad (29)$$

4 数值计算

将改进的对称 ADMM 算法应用于 LASSO 问题^[6-9]

$$\min \left\{ \frac{1}{2} \|Px - b\|^2 + F\|x\|_1 \right\}$$

式中, $P \in \mathbb{R}^{m \times n}$ 。特别地, 它可以写成式(1)的形式^[10-12]

$$\min \left\{ \frac{1}{2} \|Px - b\|^2 + u\|z\|_1 |x - z| = 0 \right\}$$

在进行数值计算时, 选取不同维数的矩阵 P , 并且规范化 P 中所有列。取 $a \in \mathbb{R}^n$ 为随机变量, ε 是噪音, 那么 $b = Pa + \varepsilon$, 正则化参数 $u = 0.01 \|P^T b\|$ 。取 $n = 4\,000$, $\alpha = 0.9$, $\gamma = 1$, $G = \frac{3}{2} \gamma I_n - \gamma A^T A$ 。此外, x, y, z 的初始值皆为零。

算法的收敛准则参照文献[4]。定义初始残差 $p^k = \gamma \|z^k - z^{k+1}\|$ 和对偶残差 $d^k = \frac{1}{\gamma} \|u^k - u^{k+1}\|$ 。对于改进的对称 ADMM 算法, 其收敛准则为

$$\max \{d^k, p^k\} \leq \sqrt{n} \delta$$

式中, $\delta = 10^{-4}$ 。

表1为应用对称 ADMM 算法和改进的对称 ADMM 算法解决该问题的结果, 其中 m 表示矩阵 P 的维数, k 表示迭代次数。从表1可以看出, 当矩阵 P 的维数较低时, 改进的对称 ADMM 算法明显快于对称 ADMM 算法; 而当 P 的维数较高时, 对比 CPU 时间发现, 改进的对称 ADMM 算法的数值表现优势更大。

为了进一步观察两种算法的收敛性, 比较了初始残差和对偶残差随迭代次数的变化情况。从图1和图2可以直观地发现, 尽管在算法迭代的某些阶段, 对称 ADMM 算法的初始残差、对偶残差减小得更快, 但是本文提出的算法先于对称 ADMM 算法达到收敛条件, 因此改进的对称 ADMM 算法更高效。

表 1 LASSO 问题数值结果
Tab.1 Numerical results for LASSO

算法	m	$k/\text{次}$	CPU 时间/s
对称 ADMM	5 000	28	29.67
	6 000	32	34.13
	8 000	53	59.93
改进的对称 ADMM	5 000	11	11.70
	6 000	14	13.61
	8 000	25	27.91

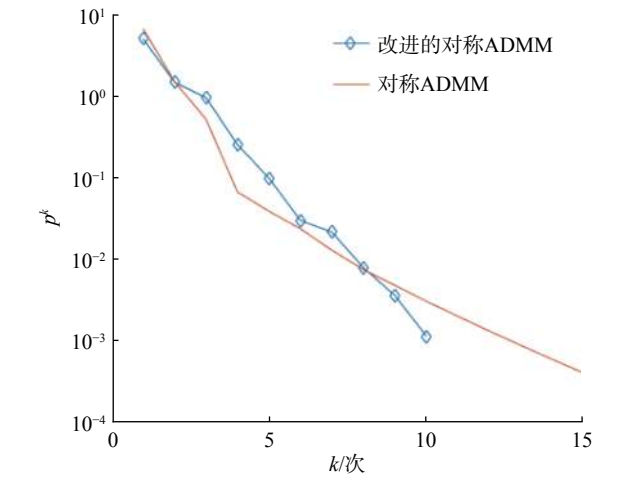


图 1 初始残差的变化情况
Fig. 1 Evolution of primal residual

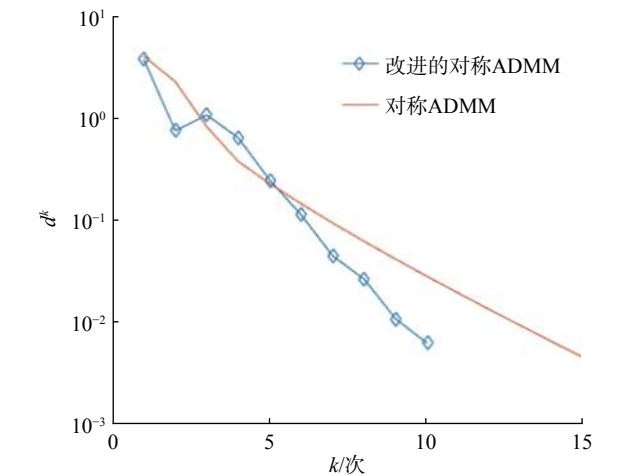


图 2 对偶残差的变化情况
Fig. 2 Evolution of dual residual

5 结 论

提出了一种求解目标函数具有可分离结构的凸优化问题的改进的对称 ADMM 算法，并证明了其收敛性。该方法的基本思想是在 x 子问题中引入一个半近邻项，从而达到加快其收敛速度的目

的。数值实验表明，该算法在求解高维的 LASSO 问题时相对于对称 ADMM 算法具有明显的优势，但是如何选择最优的 α 还需要进一步研究。

参考文献:

[1] GABAY D, MERCIER B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation[J]. Computers & Mathematics with Applications, 1976, 2(1): 17–40.

[2] CHAN T F, GLOWINSKI R. Finite element approximation and iterative solution of a class of mildly non-linear elliptic equations[R]. Stanford, CA: Stanford University, 1978.

[3] 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, 2010, 3(1): 1–122.

[4] HE B S, LIU H, WANG Z R, et al. A strictly contractive Peaceman-Rachford splitting method for convex programming[J]. SIAM Journal on Optimization, 2014, 24(3): 1011–1040.

[5] HE B S, YUAN X M. On the $O(1/n)$ convergence rate of the Douglas –Rachford alternating direction method[J]. SIAM Journal on Numerical Analysis, 2012, 50(2): 700–709.

[6] DENG W, YIN W. On the global and linear convergence of the generalized alternating direction method of multipliers[J]. Journal of Scientific Computing, 2016, 66(3): 889–916.

[7] 刘柳, 陶大程. Lasso 问题的最新算法研究[J]. 数据采集与处理, 2015, 30(1): 35–46.

[8] 何炳生. 我和乘子交替方向法 20 年[J]. 运筹学学报, 2018, 22(1): 1–31.

[9] 李静, 全晓云, 王金甲. 基于交替方向乘子法的广义交互 LASSO 模型用于肝脏疾病分类[J]. 生物医学工程学杂志, 2017, 34(3): 350–356.

[10] SUN M, LIU J. Generalized Peaceman-Rachford splitting method for separable convex programming with applications to image processing[J]. Journal of Applied Mathematics and Computing, 2016, 51(1/2): 605–622.

[11] XIE J, LIAO A, Yang X. An inexact alternating direction method of multipliers with relative error criteria[J]. Optimization Letters, 2017, 11(3):583-596.

[12] WEN Z, GOLDFARB D, YIN W, et al. Alternating direction augmented Lagrangian methods for semidefinite programming[J]. Mathematical Programming Computation, 2010, 2(3/4):203-230.