一种惯性邻近的 Peaceman-Rachford 分裂方法

窦明圆¹, 李慧云², 刘新为^{1*}

1. 河北工业大学理学院, 天津 300401;

2. 河北工业大学控制科学与工程学院, 天津 300401

E-mail: dmypyl@163.com, lhyih@hebut.edu.cn, mathlxw@hebut.edu.cn

收稿日期: 2016-07-05; 接受日期: 2016-11-08; 网络出版日期: 2016-12-29; * 通信作者

国家自然科学基金 (批准号: 11271107, 11671116 和 91630202) 和河北省自然科学基金 (批准号: A2015202365) 资助项目

摘要 严格压缩的 Peaceman-Rachford (PR) 分裂方法是一种收敛速度快于交替方向乘子法的求解线性约束可分离凸优化问题的有效方法. 最近提出的半邻近 PR 分裂方法是严格压缩的 PR 分裂方法的一种改进方法. 基于惯性邻近交替方向乘子法的思想, 本文进一步改进了半邻近 PR 分裂方法, 提出了一种惯性邻近 PR 分裂方法. 该方法利用前两次产生的迭代点来产生新的迭代点, 可以加速半邻近 PR 分裂方法的收敛. 本文提出的方法具有一般性, 它包含严格压缩的 PR 分裂方法和半邻近 PR 分裂方法作为特殊情形. 在一定的假设下, 本文证明了该算法产生的迭代序列的渐进可行性及函数值的收敛性, 进而得到了迭代序列的全局收敛性. 最后, 本文通过数值试验说明了算法的有效性.

关键词 凸优化 半邻近分裂方法 Peaceman-Rachford 分裂方法 惯性邻近点方法

MSC (2010) 主题分类 90C25, 90C30

1 引言

我们考虑带有线性约束且目标函数具有可分离结构的凸优化问题:

$$\begin{aligned} \min \quad & \theta_1(x) + \theta_2(y) \\ \text{s.t.} \quad & Ax + By = b, \\ & x \in \mathcal{X}, \quad y \in \mathcal{Y}, \end{aligned} \quad (1.1)$$

其中 $\theta_1(x) : \mathbb{R}^{n_1} \rightarrow \mathbb{R} \cup \{+\infty\}$, $\theta_2(y) : \mathbb{R}^{n_2} \rightarrow \mathbb{R} \cup \{+\infty\}$ 是闭正常凸函数 (不一定光滑); $\mathcal{X} \subset \mathbb{R}^{n_1}$, $\mathcal{Y} \subset \mathbb{R}^{n_2}$ 是闭凸集; $A \in \mathbb{R}^{m \times n_1}$, $B \in \mathbb{R}^{m \times n_2}$, $b \in \mathbb{R}^m$ 是给定的矩阵和向量. 本文假设该凸优化问题解集非空. 可分离问题 (1.1) 在很多领域诸如图像处理和统计学习中有十分广泛的应用. 分裂方法是求解含可分离算子线性约束凸优化问题的有效工具, 交替方向乘子法 (ADMM)^[1,2] 是其中的一种方法. 最近, 为加速 ADMM 的收敛出现了一系列由交替方向乘子法衍变出的方法^[3,4].

Peaceman-Rachford 分裂方法 (简记为 PR 分裂方法) 也是一种求解问题 (1.1) 的分裂方法. 正如 Gabay 在文献 [5] 中所说, “与 ADMM 相比, PR 分裂方法具有较差的鲁棒性, 即它在更加严格的假设

英文引用格式: Dou M Y, Li H Y, Liu X W. An inertial proximal Peaceman-Rachford splitting method (in Chinese). Sci Sin Math, 2017, 47: 333–348, doi: 10.1360/N012016-00134

下才收敛. 但是, 当 PR 分裂方法收敛时, 它的收敛速度更快”. 为了在一般假设下确保 PR 分裂方法的收敛性, 2014 年, He 等^[6] 通过在 PR 分裂方法的基础上加欠定松弛因子提出了一种严格压缩的 PR 分裂方法, 并证明了全局收敛性. 此后, PR 分裂方法才得以广泛应用 (参见文献 [7, 8]). 2015 年, Gu 等^[9] 提出了半邻近的 PR 分裂方法, 在严格压缩的 PR 分裂方法的基础上, 通过加半邻近项简化了子问题的求解. 此外, 对其他改进的严格压缩的 PR 分裂方法的研究参见文献 [10–13].

最近, 一种惯性技术被应用到凸优化问题且成为了研究的热点. 这种惯性技术最早由 Alvarez^[14] 提出, 用来求解动力系统问题, 它不只是利用当前迭代信息产生新的迭代点, 而是利用当前迭代信息与上一步迭代信息产生新的迭代点, 从而更好更快地找到最优解. 随后, 这种惯性技术应用于求解极大单调算子包含问题 (参见文献 [15–17]). 2014 年, Chen 等^[18] 应用惯性技术求解混合变分不等式问题. 随着对惯性技术的不断研究, 文献中提出了很多结合惯性技术的算法 (参见文献 [19–21]). 特别地, 2015 年, Chen 等^[22] 提出了一种惯性邻近的 ADMM, 利用惯性技术加速了半邻近 ADMM 的收敛.

与 ADMM 相比, PR 分裂方法在每次迭代中更新两次 Lagrange 乘子. 基于 PR 分裂方法与 ADMM 之间的关系, 本文把惯性技术与半邻近 PR 分裂方法结合, 提出了一种惯性邻近的 PR 分裂方法. 该算法不仅包含了严格压缩的 PR 分裂方法和半邻近的 PR 分裂方法, 而且通过令邻近项矩阵等于零, 得到了惯性 PR 分裂方法. 此外, 本文在与惯性邻近 ADMM 相同的假设下, 证明了该算法的全局收敛性. 最后, 本文利用惯性邻近的 PR 分裂方法进行了数值试验. 结果表明本算法数值表现优于惯性邻近的 ADMM 和半邻近的 PR 分裂方法, 从而验证了惯性技术的加速性质.

本文中 $\|\cdot\|$ 表示 l_2 -范数. $M \succeq 0$ ($M \succ 0$) 表示 M 是对称半正定 (正定) 矩阵. 矩阵或者向量的转置用上标 “T” 表示. 对于任意的 $n \times n$ 对称半正定矩阵 M 与任意的 n 维向量 u 和 v , 定义 $\langle u, v \rangle_M = u^T M v$ 和 $\|u\|_M = \sqrt{\langle u, u \rangle_M}$. 零矩阵和零向量简单记为 0. 为简单起见, 有时把向量 $(x^T, y^T, z^T)^T$ 简记为 (x, y, z) .

2 算法

首先, 为方便起见, 令 $u = (x, y)$, $w = (x, y, \lambda)$, $v = (y, \lambda)$, $\theta(u) = \theta_1(x) + \theta_2(y)$, 及 $w^k = (x^k, y^k, \lambda^k)$, $u^k = (x^k, y^k)$, $v^k = (y^k, \lambda^k)$. 由凸优化问题的一阶最优性条件可知, 问题 (1.1) 等价于一个混合变分不等式问题: 求

$$w^* = (x^*, y^*, \lambda^*) \in \Omega := \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m$$

满足

$$\text{VI}(\Omega, F, \theta) : \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall w \in \Omega, \quad (2.1)$$

其中

$$F(w) = \begin{pmatrix} -A^T \lambda \\ -B^T \lambda \\ Ax + By - b \end{pmatrix},$$

并且 $F(w)$ 是单调的. 在这里, 定义 Ω^* 是该变分不等式问题 $\text{VI}(\Omega, F, \theta)$ 的解集, 并且在假设问题 (1.1) 的解集非空的情形下, Ω^* 非空.

现在, 我们定义几个矩阵以便于下面的分析:

$$G = \begin{pmatrix} S & 0 & 0 \\ 0 & T + \frac{2-\alpha}{2}\beta B^T B & -\frac{1}{2}B^T \\ 0 & -\frac{1}{2}B & \frac{1}{2\alpha\beta}I_m \end{pmatrix}, \quad (2.2)$$

$$H = \begin{pmatrix} \frac{2-\alpha}{2}\beta B^T B & -\frac{1}{2}B^T \\ -\frac{1}{2}B & \frac{1}{2\alpha\beta}I_m \end{pmatrix}, \quad M = \begin{pmatrix} 0 & 0 \\ 0 & H \end{pmatrix}, \quad P = \begin{pmatrix} S & 0 \\ 0 & T \end{pmatrix}.$$

注意到, 当 $\alpha \in (0, 1)$ 且 S 与 T 半正定时, G 是半正定的. 此外, 对于任意 $w, w' \in \Omega$, 皆满足

$$\|w - w'\|_G^2 = \|u - u'\|_P^2 + \|v - v'\|_H^2. \quad (2.3)$$

对于问题 (1.1), He 等^[6] 通过在 PR 分裂方法的基础上加欠定松弛因子提出了严格压缩的 PR 分裂方法, 其第 k 步迭代如下:

$$\begin{aligned} x^{k+1} &= \arg \min \{L(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} &= \lambda^k - \alpha\beta(Ax^{k+1} + By^k - b), \\ y^{k+1} &= \arg \min \{L(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \alpha\beta(Ax^{k+1} + By^{k+1} - b), \end{aligned} \quad (2.4)$$

其中 $L(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b) + \frac{\beta}{2}\|Ax + By - b\|^2$ 是问题 (1.1) 的增广 Lagrange 函数, λ 为 Lagrange 乘子, $\beta > 0$ 是罚参数, $\alpha \in (0, 1)$. 随后, Gu 等^[9] 通过在严格压缩的 PR 分裂方法的基础上加半邻近项, 提出了半邻近 PR 分裂方法, 使得子问题的求解更加简便, 其第 k 步迭代如下:

$$\begin{aligned} x^{k+1} &= \arg \min \left\{ L(x, y^k, \lambda^k) + \frac{1}{2}\|x - x^k\|_S^2 \mid x \in \mathcal{X} \right\}, \\ \lambda^{k+\frac{1}{2}} &= \lambda^k - \alpha\beta(Ax^{k+1} + By^k - b), \\ y^{k+1} &= \arg \min \left\{ L(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) + \frac{1}{2}\|y - y^k\|_T^2 \mid y \in \mathcal{Y} \right\}, \\ \lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \alpha\beta(Ax^{k+1} + By^{k+1} - b), \end{aligned} \quad (2.5)$$

其中 S 和 T 是给定的半正定矩阵. 最近, Chen 等^[22] 提出了一种惯性邻近 ADMM, 利用惯性技术加速了邻近 ADMM 的收敛, 其第 k 步迭代如下:

$$\begin{aligned} (\bar{x}^k, \bar{y}^k, \bar{\lambda}^k) &= (x^k, y^k, \lambda^k) + \alpha_k(x^k - x^{k-1}, y^k - y^{k-1}, \lambda^k - \lambda^{k-1}), \\ x^{k+1} &= \arg \min \left\{ L(x, \bar{y}^k, \bar{\lambda}^k) + \frac{1}{2}\|x - \bar{x}^k\|_S^2 \mid x \in \mathcal{X} \right\}, \\ \lambda^{k+1} &= \bar{\lambda}^k - \beta(Ax^{k+1} + B\bar{y}^k - b), \\ y^{k+1} &= \arg \min \left\{ L(x^{k+1}, y, \lambda^{k+1}) + \frac{1}{2}\|y - \bar{y}^k\|_T^2 \mid y \in \mathcal{Y} \right\}. \end{aligned} \quad (2.6)$$

本文把惯性技术与半邻近 PR 分裂方法结合, 提出了惯性邻近 PR 分裂方法. 下面详细描述算法.

惯性邻近 PR 分裂方法 (IsP-PRSM)

第 0 步 给定半正定矩阵 S 和 T , $(x^0, y^0, \lambda^0) \in \Omega$, $\alpha \in (0, 1)$, $\beta > 0$, 非负参数序列 $\{\alpha_k\}_{k=0}^\infty$, 令 $(x^{-1}, y^{-1}, \lambda^{-1}) = (x^0, y^0, \lambda^0)$.

第 k 步 在终止条件满足之前, 由

$$(\bar{x}^k, \bar{y}^k, \bar{\lambda}^k) = (x^k, y^k, \lambda^k) + \alpha_k(x^k - x^{k-1}, y^k - y^{k-1}, \lambda^k - \lambda^{k-1}), \quad (2.7a)$$

$$x^{k+1} = \arg \min \left\{ L(x, \bar{y}^k, \bar{\lambda}^k) + \frac{1}{2} \|x - \bar{x}^k\|_S^2 \mid x \in \mathcal{X} \right\}, \quad (2.7b)$$

$$\lambda^{k+\frac{1}{2}} = \bar{\lambda}^k - \alpha\beta(Ax^{k+1} + B\bar{y}^k - b), \quad (2.7c)$$

$$y^{k+1} = \arg \min \left\{ L(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) + \frac{1}{2} \|y - \bar{y}^k\|_T^2 \mid y \in \mathcal{Y} \right\}, \quad (2.7d)$$

$$\lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \alpha\beta(Ax^{k+1} + By^{k+1} - b) \quad (2.7e)$$

产生 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$, 且令 $k = k + 1$.

特别地, 当 $S, T = 0$ 时, 算法为惯性 PR 分裂方法. 当 $\alpha_k = 0$ 时, 算法为半邻近 PR 分裂方法. 当 $\alpha_k = 0$ 且 $S = T = 0$ 时, 算法为严格压缩的 PR 分裂方法.

3 全局收敛性

本节证明算法 IsP-PRSM 的全局收敛性. 在证明收敛性过程中, 本文并没有用压缩收敛的统一框架证明, 而是借鉴证明惯性邻近点法的过程 (参见文献 [15, 18, 22]). 在证明过程中, 由于 PR 分裂方法并不是邻近点方法的应用, 直接用惯性邻近点方法的证明过程无法证得本文算法的全局收敛性. 此时, 本文中的引理 3.2 起到了重要作用, 它衔接了引理 3.1 与 3.3. 此外, 本文利用文献 [9] 的证明方法, 证明了引理 3.4, 从而可以证明算法产生迭代序列的渐进可行性和函数值的收敛性. 收敛性的证明基于以下假设.

假设 3.1 设 $\{\alpha_k\}_{k=0}^\infty$ 满足

- (1) 对于任意的 $k \geq 0$, 有 $0 \leq \alpha_k \leq t$, 其中 $t \in [0, 1)$;
- (2) 由迭代 (2.7) 产生的序列 $\{w^k\}_{k=0}^\infty$ 满足

$$\sum_{k=0}^{\infty} \alpha_k \|w^k - w^{k-1}\|_G^2 < \infty. \quad (3.1)$$

惯性邻近 ADMM 可以解释为混合变分不等式问题邻近点算法的应用, 而惯性邻近 PR 分裂方法只可以得到以下结果.

引理 3.1 ^[9] 给定 $w^k = (x^k, y^k, \lambda^k) \in \Omega$, 由迭代 (2.7) 产生的下一个迭代点 $w^{k+1} \in \Omega$ 满足

$$\begin{aligned} & \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T \{F(w^{k+1}) + G(w^{k+1} - \bar{w}^k)\} \\ & \geq (w^{k+1} - w)^T \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} [(1 - \alpha)\beta B(\bar{y}^k - y^{k+1}) + (1 - 2\alpha)\beta r^{k+1}], \quad \forall w \in \Omega, \end{aligned} \quad (3.2)$$

其中 $r^{k+1} = Ax^{k+1} + By^{k+1} - b$.

下面引理使得本文仅在 (3.2) 的形式下证明了引理 3.3.

引理 3.2 设 $\{w^k\}_{k=0}^\infty$ 由迭代 (2.7) 产生, 则有

$$-(1 - \alpha)\beta(r^{k+1})^T(B(\bar{y}^k - y^{k+1})) - (1 - 2\alpha)\beta\|r^{k+1}\|^2 \leq \frac{1}{2}\|w^{k+1} - \bar{w}^k\|_M^2. \quad (3.3)$$

证明 由 $2a^T b \leq \|a\|^2 + \|b\|^2$ 有

$$-(1-\alpha)\beta(r^{k+1})^T(B(\bar{y}^k - y^{k+1})) - (1-2\alpha)\beta\|r^{k+1}\|^2 \leq \frac{3\alpha-1}{2}\beta\|r^{k+1}\|^2 + \frac{1-\alpha}{2}\beta\|B(y^{k+1} - \bar{y}^k)\|^2.$$

定义

$$D = \begin{pmatrix} I_{n_2} & 0 \\ \alpha\beta B & 2\alpha\beta I_m \end{pmatrix}, \quad (3.4)$$

则

$$D^T H D = \begin{pmatrix} (1-\alpha)\beta B^T B & 0 \\ 0 & 2\alpha\beta I_m \end{pmatrix}. \quad (3.5)$$

由 (2.7c) 和 (2.7e) 可知, $\bar{\lambda}^k = \lambda^{k+1} + 2\alpha\beta r^{k+1} + \alpha\beta B(\bar{y}^k - y^{k+1})$, 则

$$\bar{v}^k - v^{k+1} = \begin{pmatrix} \bar{y}^k - y^{k+1} \\ \bar{\lambda}^k - \lambda^{k+1} \end{pmatrix} = D \begin{pmatrix} \bar{y}^k - y^{k+1} \\ r^{k+1} \end{pmatrix}. \quad (3.6)$$

结合 (3.5), 经过简单计算可得

$$\begin{aligned} \frac{1}{2}\|v^{k+1} - \bar{v}^k\|_H^2 &= \frac{1}{2}((y^{k+1} - \bar{y}^k)^T, (r^{k+1})^T) D^T H D \begin{pmatrix} y^{k+1} - \bar{y}^k \\ r^{k+1} \end{pmatrix} \\ &= \frac{1-\alpha}{2}\beta\|B(y^{k+1} - \bar{y}^k)\|^2 + \alpha\beta\|r^{k+1}\|^2. \end{aligned}$$

因为当 $\alpha \in (0, 1)$ 时有 $\frac{3\alpha-1}{2} < \alpha$, 所以,

$$\frac{1}{2}\|v^{k+1} - \bar{v}^k\|_H^2 > \frac{3\alpha-1}{2}\beta\|r^{k+1}\|^2 + \frac{1-\alpha}{2}\beta\|B(y^{k+1} - \bar{y}^k)\|^2.$$

又由 H 和 M 的定义可知,

$$\frac{1}{2}\|v^{k+1} - \bar{v}^k\|_H^2 = \frac{1}{2}\|w^{k+1} - \bar{w}^k\|_M^2.$$

因此, 结合上面的不等式即可得证. $\square$

下面两个引理对于收敛性的证明起到了至关重要的作用.

引理 3.3 设 $\{\alpha_k\}_{k=0}^\infty$ 满足假设 3.1, 令 $\{w^k\}_{k=0}^\infty$ 是由迭代 (2.7) 产生的序列, 则对任意的 $w^* \in \Omega^*$, $\lim_{k \rightarrow \infty} \|w^k - w^*\|_G$ 存在.

证明 令 (2.1) 中的 $w = w^{k+1}$, (3.2) 中的 $w = w^*$, 将两不等式相加, 且由 F 的单调性, 有

$$(w^{k+1} - w^*)^T G(w^{k+1} - \bar{w}^k) + (w^{k+1} - w^*)^T \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} [(1-\alpha)\beta B(\bar{y}^k - y^{k+1}) + (1-2\alpha)\beta r^{k+1}] \leq 0,$$

整理得

$$(w^{k+1} - w^*)^T G(w^{k+1} - \bar{w}^k) + (1-\alpha)\beta(r^{k+1})^T B(\bar{y}^k - y^{k+1}) + (1-2\alpha)\beta\|r^{k+1}\|^2 \leq 0. \quad (3.7)$$

由 (2.7) 知 $\bar{w}^k = w^k + \alpha_k(w^k - w^{k-1})$, 则

$$\begin{aligned} (w^{k+1} - w^*)^T G(w^{k+1} - \bar{w}^k) &= (w^{k+1} - w^*)^T G(w^{k+1} - w^k - \alpha_k(w^k - w^{k-1})) \\ &= (w^{k+1} - w^*)^T G(w^{k+1} - w^k) - \alpha_k(w^{k+1} - w^*)^T G(w^k - w^{k-1}). \end{aligned} \quad (3.8)$$

定义 $\varphi_k = \frac{1}{2}\|w^k - w^*\|_G^2$, 经过简单计算可得

$$\begin{aligned} (w^{k+1} - w^*)^T G(w^{k+1} - w^k) &= \varphi_{k+1} - \varphi_k + \frac{1}{2}\|w^{k+1} - w^k\|_G^2, \\ (w^{k+1} - w^*)^T G(w^k - w^{k-1}) &= \varphi_k - \varphi_{k-1} + \frac{1}{2}\|w^k - w^{k-1}\|_G^2 + (w^{k+1} - w^k)^T G(w^k - w^{k-1}). \end{aligned}$$

结合 (3.3)、(3.7) 和 (3.8), 经过整理得

$$\begin{aligned} \varphi_{k+1} - \varphi_k - \alpha_k(\varphi_k - \varphi_{k-1}) &\leq -\frac{1}{2}\|w^{k+1} - w^k - \alpha_k(w^k - w^{k-1})\|_G^2 + \frac{1}{2}(\alpha_k + \alpha_k^2)\|w^k - w^{k-1}\|_G^2 \\ &\quad + \frac{1}{2}\|w^{k+1} - \bar{w}^k\|_M^2. \end{aligned} \quad (3.9)$$

因为 $\bar{w}^k = w^k + \alpha_k(w^k - w^{k-1})$, 且当 $\alpha_k \in [0, 1)$ 时 $\frac{1}{2}(\alpha_k + \alpha_k^2) \leq \alpha_k$, 所以, (3.9) 变为

$$\varphi_{k+1} - \varphi_k - \alpha_k(\varphi_k - \varphi_{k-1}) \leq -\frac{1}{2}\|w^{k+1} - \bar{w}^k\|_G^2 + \alpha_k\|w^k - w^{k-1}\|_G^2 + \frac{1}{2}\|w^{k+1} - \bar{w}^k\|_M^2. \quad (3.10)$$

又由矩阵 M 与 G 的关系知, $\|w^{k+1} - \bar{w}^k\|_M^2 \leq \|w^{k+1} - \bar{w}^k\|_G^2$, 代入 (3.10) 可得

$$\varphi_{k+1} - \varphi_k - \alpha_k(\varphi_k - \varphi_{k-1}) \leq \alpha_k\|w^k - w^{k-1}\|_G^2. \quad (3.11)$$

令 $\theta_k = \varphi_k - \varphi_{k-1}$, $\delta_k = \alpha_k\|w^k - w^{k-1}\|_G^2$, 则由 (3.11) 可得 $\theta_{k+1} \leq \alpha_k\theta_k + \delta_k \leq t[\theta_k]_+ + \delta_k$, 其中 $[t]_+ = \max\{t, 0\}$. 因此有

$$[\theta_{k+1}]_+ \leq t[\theta_k]_+ + \delta_k \leq t^{k+1}[\theta_0]_+ + \sum_{j=0}^k t^j \delta_{k-j}.$$

注意到 $w^0 = w^{-1}$, 则 $\theta_0 = [\theta_0]_+ = 0$, $\delta_0 = 0$. 因此, 由上式和假设 3.1 可知,

$$\sum_{k=0}^{\infty} [\theta_k]_+ \leq \frac{1}{1-\alpha} \sum_{k=0}^{\infty} \delta_k = \frac{1}{1-\alpha} \sum_{k=1}^{\infty} \delta_k < \infty. \quad (3.12)$$

令 $m_k = \varphi_k - \sum_{j=1}^k [\theta_j]_+$, 由 (3.12) 和 $\varphi_k \geq 0$ 可得 m_k 有下界. 另一方面,

$$m_{k+1} = \varphi_{k+1} - [\theta_{k+1}]_+ - \sum_{j=1}^k [\theta_j]_+ \leq \varphi_{k+1} - \theta_{k+1} - \sum_{j=1}^k [\theta_j]_+ = \varphi_k - \sum_{j=1}^k [\theta_j]_+ = m_k,$$

即 m_k 是非增的, 故 $\{m_k\}$ 收敛. 由于

$$\lim_{k \rightarrow \infty} \varphi_k = \sum_{j=1}^{\infty} [\theta_j]_+ + \lim_{k \rightarrow \infty} m_k,$$

因此 $\{\varphi_k\}$ 收敛, 即 $\lim_{k \rightarrow \infty} \|w^k - w^*\|_G$ 存在. □

引理 3.4 设 $\{\alpha_k\}_{k=0}^\infty$ 满足假设 3.1, 令 $\{w^k\}_{k=0}^\infty$ 是由迭代 (2.7) 产生的序列, 则

$$\sum_{k=0}^{\infty} \|w^{k+1} - \bar{w}^k\|_G^2 < \infty.$$

因此, $\lim_{k \rightarrow \infty} \|w^{k+1} - \bar{w}^k\|_G = 0$.

证明 由 (3.7) 有

$$(w^* - w^{k+1})^T G(w^{k+1} - \bar{w}^k) \geq (1 - 2\alpha)\beta \|r^{k+1}\|^2 + (1 - \alpha)\beta (r^{k+1})^T B(\bar{y}^k - y^{k+1}).$$

由于 $\|a\|_G^2 - \|b\|_G^2 = \|a - b\|_G^2 + 2b^T G(a - b)$, 令 $a = w^* - \bar{w}^k$, $b = w^* - w^{k+1}$, 则

$$\|\bar{w}^k - w^*\|_G^2 - \|w^{k+1} - w^*\|_G^2 = \|\bar{w}^k - w^{k+1}\|_G^2 + 2(w^* - w^{k+1})^T G(w^{k+1} - \bar{w}^k).$$

结合上面两个式子可知,

$$\|\bar{w}^k - w^*\|_G^2 - \|w^{k+1} - w^*\|_G^2 \geq \|\bar{w}^k - w^{k+1}\|_G^2 + 2(1 - 2\alpha)\beta \|r^{k+1}\|^2 + 2(1 - \alpha)\beta (r^{k+1})^T B(\bar{y}^k - y^{k+1}).$$

由 (2.3) 和 $\|v^{k+1} - \bar{v}^k\|_H^2 = (1 - \alpha)\beta \|B(\bar{y}^k - y^{k+1})\|^2 + 2\alpha\beta \|r^{k+1}\|^2$, 上式可整理为

$$\begin{aligned} \|\bar{w}^k - w^*\|_G^2 - \|w^{k+1} - w^*\|_G^2 &\geq \|\bar{u}^k - u^{k+1}\|_P^2 + (1 - \alpha)\beta \|B(\bar{y}^k - y^{k+1})\|^2 + 2(1 - \alpha)\beta \|r^{k+1}\|^2 \\ &\quad + 2(1 - \alpha)\beta (r^{k+1})^T B(\bar{y}^k - y^{k+1}). \end{aligned}$$

定义

$$K = \begin{pmatrix} (1 - \alpha)\beta B^T B & (1 - \alpha)\beta B^T \\ (1 - \alpha)\beta B & 2(1 - \alpha)\beta I_m \end{pmatrix}.$$

结合 (3.4) 和 (3.6), 上式可化简为

$$\|\bar{w}^k - w^*\|_G^2 - \|w^{k+1} - w^*\|_G^2 \geq \|\bar{u}^k - u^{k+1}\|_P^2 + (\bar{v}^k - v^{k+1})^T D^{-T} K D^{-1} (\bar{v}^k - v^{k+1}).$$

经过计算可得

$$\begin{aligned} K - c \cdot D^T H D &= \begin{pmatrix} (1 - c)(1 - \alpha)\beta B^T B & (1 - \alpha)\beta B^T \\ (1 - \alpha)\beta B & 2(1 - \alpha - c\alpha)\beta I_m \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\beta} B^T & 0 \\ 0 & \sqrt{\beta} I_m \end{pmatrix} \begin{pmatrix} (1 - c)(1 - \alpha) I_m & (1 - \alpha) I_m \\ (1 - \alpha) I_m & 2(1 - \alpha - c\alpha) I_m \end{pmatrix} \begin{pmatrix} \sqrt{\beta} B & 0 \\ 0 & \sqrt{\beta} I_m \end{pmatrix}. \end{aligned}$$

注意到, 取 $c = \frac{1 - \sqrt{1 - 2\alpha(1 - \alpha)}}{2\alpha} \in (0, 1)$ 时, 矩阵

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

是半正定的, 此时 $K \geq c \cdot D^T H D$. 因此,

$$\|\bar{w}^k - w^*\|_G^2 - \|w^{k+1} - w^*\|_G^2 \geq c \|\bar{w}^k - w^{k+1}\|_G^2. \quad (3.13)$$

经计算可得

$$(w^k - w^*)^T G(w^k - w^{k-1}) = \frac{1}{2} \|w^k - w^*\|_G^2 - \frac{1}{2} \|w^{k-1} - w^*\|_G^2 + \frac{1}{2} \|w^k - w^{k-1}\|_G^2. \quad (3.14)$$

由于当 $\alpha_k < 1$ 时 $\alpha_k^2 < \alpha_k$, 再由 $\bar{w}^k$ 的定义及 (3.13) 和 (3.14), 有

$$\begin{aligned} \|w^{k+1} - \bar{w}^k\|_G^2 &\leq \frac{1}{c} [\|w^k - w^*\|_G^2 - \|w^{k+1} - w^*\|_G^2 + \alpha_k \theta_k + 2\alpha_k \|w^k - w^{k-1}\|_G^2] \\ &\leq \frac{1}{c} [(\|w^k - w^*\|_G^2 - \|w^{k+1} - w^*\|_G^2) + [\theta_k]_+ + 2\alpha_k \|w^k - w^{k-1}\|_G^2]. \end{aligned}$$

两边从 $k = 0$ 到 $k = \infty$ 求和, 结合 (3.12)、引理 3.3 和假设 3.1, 即可得证. $\square$

下面的定理证明了算法产生序列的渐进可行性和目标函数值的收敛性.

定理 3.1 设 $\{\alpha_k\}_{k=0}^\infty$ 满足假设 3.1, $\alpha \in (0, 1)$, $\beta > 0$, 令 $\{w^k\}_{k=0}^\infty$ 是由迭代 (2.7) 产生的序列, 则有

- (1) $\sum_{k=1}^\infty \|Ax^k + By^k - b\|^2 < \infty$, 因此, $\lim_{k \rightarrow \infty} \|Ax^k + By^k - b\| = 0$;
- (2) 当 $k \rightarrow \infty$ 时, 目标函数值 $\theta_1(x^k) + \theta_2(y^k)$ 收敛到问题 (1.1) 的最优值.

证明 (1) 由 (2.7c) 和 (2.7e) 可知, $\lambda^{k+1} = \bar{\lambda}^k - \alpha\beta(Ax^{k+1} + B\bar{y}^k - b) - \alpha\beta r^{k+1}$. 因此,

$$\begin{aligned} Ax^{k+1} + By^{k+1} - b &= Ax^{k+1} + B\bar{y}^k - b + B(y^{k+1} - \bar{y}^k) \\ &= \frac{1}{\alpha\beta}(\bar{\lambda}^k - \lambda^{k+1}) - r^{k+1} + B(y^{k+1} - \bar{y}^k), \end{aligned}$$

故 $Ax^{k+1} + By^{k+1} - b = \frac{1}{2\alpha\beta}(\bar{\lambda}^k - \lambda^{k+1}) + \frac{B}{2}(y^{k+1} - \bar{y}^k)$. 所以,

$$\|Ax^{k+1} + By^{k+1} - b\|^2 = \left\| \frac{B}{2}(y^{k+1} - \bar{y}^k) - \frac{1}{2\alpha\beta}(\lambda^{k+1} - \bar{\lambda}^k) \right\|^2. \quad (3.15)$$

由 G 的定义可知,

$$\begin{aligned} \|w^{k+1} - \bar{w}^k\|_G^2 &= \|x^{k+1} - \bar{x}^k\|_S^2 + \|y^{k+1} - \bar{y}^k\|_T^2 + \frac{1}{2\alpha\beta} \|\lambda^{k+1} - \bar{\lambda}^k\|^2 \\ &\quad + 2(2 - \alpha)\beta \left\| \frac{B}{2}(y^{k+1} - \bar{y}^k) \right\|^2 - 2 \left[\frac{B}{2}(y^{k+1} - \bar{y}^k) \right]^T (\lambda^{k+1} - \bar{\lambda}^k). \end{aligned} \quad (3.16)$$

由于 $0 < \alpha < 1$, $\beta > 0$, 故 $2(2 - \alpha) > 2\alpha$. 所以,

$$\begin{aligned} &\frac{1}{2\alpha\beta} \|\lambda^{k+1} - \bar{\lambda}^k\|^2 + 2(2 - \alpha)\beta \left\| \frac{B}{2}(y^{k+1} - \bar{y}^k) \right\|^2 - 2 \left[\frac{B}{2}(y^{k+1} - \bar{y}^k) \right]^T (\lambda^{k+1} - \bar{\lambda}^k) \\ &> \frac{1}{2\alpha\beta} \|\lambda^{k+1} - \bar{\lambda}^k\|^2 + 2\alpha\beta \left\| \frac{B}{2}(y^{k+1} - \bar{y}^k) \right\|^2 - 2 \left[\frac{B}{2}(y^{k+1} - \bar{y}^k) \right]^T (\lambda^{k+1} - \bar{\lambda}^k) \\ &= 2\alpha\beta \left\| \frac{B}{2}(y^{k+1} - \bar{y}^k) - \frac{1}{2\alpha\beta}(\lambda^{k+1} - \bar{\lambda}^k) \right\|^2. \end{aligned} \quad (3.17)$$

由 (3.15)–(3.17)、 S 和 T 的半正定性可知,

$$\|Ax^{k+1} + By^{k+1} - b\|^2 < \frac{1}{2\alpha\beta} \|w^{k+1} - \bar{w}^k\|_G^2.$$

由引理 3.4 知,

$$\sum_{k=1}^\infty \|Ax^{k+1} + By^{k+1} - b\|^2 < \sum_{k=1}^\infty \frac{1}{2\alpha\beta} \|w^{k+1} - \bar{w}^k\|_G^2 < \infty,$$

故

$$\lim_{k \rightarrow \infty} \|Ax^k + By^k - b\| = 0.$$

(2) 令 $w^* = (x^*, y^*, \lambda^*) \in \Omega^*$, (2.1) 中取 $w = (x^k, y^k, \lambda^*)$, 则有

$$\theta_1(x^k) + \theta_2(y^k) - \theta_1(x^*) - \theta_2(y^*) \geq (x^k - x^*)^T A^T \lambda^* + (y^k - y^*)^T B^T \lambda^* = (\lambda^*)^T (Ax^k + By^k - b). \quad (3.18)$$

由于 $\lim_{k \rightarrow \infty} (Ax^k + By^k - b) = 0$, 因此, 由 (3.18) 可知,

$$\liminf_{k \rightarrow \infty} (\theta_1(x^k) + \theta_2(y^k)) \geq \theta_1(x^*) + \theta_2(y^*). \quad (3.19)$$

另一方面, 令 (3.2) 中 $w = w^*$, 经过简单计算整理得

$$\begin{aligned} \theta_1(x^*) + \theta_2(y^*) - \theta_1(x^{k+1}) - \theta_2(y^{k+1}) &\geq -(\lambda^*)^T (Ax^{k+1} + By^{k+1} - b) + (w^{k+1} - w^*)^T G(w^{k+1} - \bar{w}^k) \\ &\quad + (1 - \alpha)\beta(r^{k+1})^T (B(\bar{y}^k - y^{k+1})) + (1 - 2\alpha)\beta\|r^{k+1}\|^2. \end{aligned} \quad (3.20)$$

由引理 3.3 可知, $\{\|w^k - w^*\|_G\}_{k=0}^\infty$ 有界. 由引理 3.4 和定理 3.1(1) 可知,

$$\lim_{k \rightarrow \infty} -(\lambda^*)^T (Ax^{k+1} + By^{k+1} - b) + (w^{k+1} - w^*)^T G(w^{k+1} - \bar{w}^k) = 0.$$

由引理 3.4 及 G 与 M 的关系, 有 $\lim_{k \rightarrow \infty} \|w^{k+1} - \bar{w}^k\|_M^2 = 0$, 再结合 (3.3) 和 (3.20) 可得

$$\limsup_{k \rightarrow \infty} (\theta_1(x^k) + \theta_2(y^k)) \leq \theta_1(x^*) + \theta_2(y^*). \quad (3.21)$$

由 (3.19) 和 (3.21), 结论得证. $\square$

为得到迭代序列 $\{w^k\}_{k=0}^\infty$ 的全局收敛性, 首先给出以下引理.

引理 3.5 ^[22] 令 $\{w^k\}_{k=0}^\infty$ 是由迭代 (2.7) 产生的序列, 则对任意 $w^* \in \Omega^*$, 有

$$(A(x^{k+1} - x^*))^T (\lambda^{k+1} - \lambda^* + (2\alpha - 1)\beta r^{k+1} + (1 - \alpha)\beta B(y^{k+1} - \bar{y}^k)) \geq (x^{k+1} - x^*)^T S(x^{k+1} - \bar{x}^k). \quad (3.22)$$

以下定理在与惯性 ADMM 算法相同的假设下, 给出了迭代序列 $\{w^k\}_{k=0}^\infty$ 的收敛性.

定理 3.2 设 $\{\alpha_k\}_{k=0}^\infty$ 满足假设 3.1, 令 $\{w^k\}_{k=0}^\infty$ 是由 (2.7) 产生的序列, 则 $\{(S + \beta A^T A)x^k\}_{k=1}^\infty$, $\{(T + \beta B^T B)y^k\}_{k=1}^\infty$ 和 $\{\lambda^k\}_{k=1}^\infty$ 均有界. 此外, 若 $S + \beta A^T A \succ 0$, $T + \beta B^T B \succ 0$, 则存在唯一点 $w^* \in \Omega^*$, 使得 $\lim_{k \rightarrow \infty} w^k = w^*$.

证明 由引理 3.3 可知, 对任意 $w^* \in \Omega^*$, $\lim_{k \rightarrow \infty} \|w^k - w^*\|_G$ 存在, 因此, $\{Gw^k\}_{k=0}^\infty$ 一定有界. 又从 G 的定义可知, $\{Sx^k\}_{k=0}^\infty$ 和 $\{By^k - \frac{1}{\alpha\beta}\lambda^k\}_{k=0}^\infty$ 都有界, 从而, $\{\alpha\beta By^k - \lambda^k\}_{k=0}^\infty$ 有界. 因为

$$\lim_{k \rightarrow \infty} Ay^k + \frac{1}{\alpha\beta}\lambda^k + By^k - \frac{1}{\alpha\beta}\lambda^k = \lim_{k \rightarrow \infty} Ax^k + By^k = -b,$$

所以 $\{Ax^k + \frac{1}{\alpha\beta}\lambda^k\}_{k=0}^\infty$ 有界. 同样地, $\{\alpha\beta Ay^k + \lambda^k\}_{k=0}^\infty$ 有界. 此外, 由 (3.22) 得

$$\begin{aligned} &(A(x^k - x^*))^T (\lambda^k - \lambda^* + (2\alpha - 1)\beta r^k + (1 - \alpha)\beta B(y^k - \bar{y}^{k-1})) \\ &\geq (x^k - x^*)^T S(x^k - \bar{x}^{k-1}) \geq -\frac{1}{2}(\|x^k - x^*\|_S^2 + \|x^k - \bar{x}^{k-1}\|_S^2). \end{aligned} \quad (3.23)$$

由于 $\{Sx^k\}_{k=0}^\infty$ 有界, 所以,

$$(A(x^k - x^*))^T (\lambda^k - \lambda^* + (2\alpha - 1)\beta r^k + (1 - \alpha)\beta B(y^k - \bar{y}^{k-1}))$$

有下界. 此外,

$$\begin{aligned} & \|\alpha\beta A(x^k - x^*)\|^2 + \|(\lambda^k - \lambda^*) + (2\alpha - 1)\beta r^k + (1 - \alpha)\beta B(y^k - \bar{y}^{k-1})\|^2 \\ & \leq 2\|\alpha\beta Ax^k + \lambda^k - (\alpha\beta Ax^* + \lambda^*)\|^2 + 2\|(2\alpha - 1)\beta r^k + (1 - \alpha)\beta B(y^k - \bar{y}^{k-1})\|^2 \\ & \quad - 2\alpha\beta(A(x^k - x^*))^T(\lambda^k - \lambda^* + (2\alpha - 1)\beta r^k + (1 - \alpha)\beta B(y^k - \bar{y}^{k-1})). \end{aligned} \quad (3.24)$$

由引理 3.4 知,

$$\lim_{k \rightarrow \infty} \|(w^k - \bar{w}^{k-1})\|_G^2 = 0.$$

结合 (2.3), 有 $\lim_{k \rightarrow \infty} \|(v^k - \bar{v}^{k-1})\|_H^2 = 0$. 又由定理 3.1(1) 知,

$$\lim_{k \rightarrow \infty} r^k = 0,$$

所以, $\lim_{k \rightarrow \infty} B(y^k - \bar{y}^{k-1}) = 0$, 且

$$\lim_{k \rightarrow \infty} 2\|(2\alpha - 1)\beta r^k + (1 - \alpha)\beta B(y^k - \bar{y}^{k-1})\|^2 = 0. \quad (3.25)$$

由 (3.23)–(3.25) 可知, $\{\alpha\beta Ax^k\}_{k=0}^\infty$ 有界. 由 $\{\alpha\beta Ax^k + \lambda^k\}_{k=0}^\infty$ 的有界性可知, $\{\lambda^k\}_{k=0}^\infty$ 和 $\{Ax^k\}_{k=0}^\infty$ 有界. 由 $\lim_{k \rightarrow \infty} (Ax^k + By^k - b) = 0$ 知 $\{By^k\}_{k=0}^\infty$ 有界. 再由 $\{Gw^k\}_{k=0}^\infty$ 有界可知 $\{Ty^k\}_{k=0}^\infty$ 有界. 因此, $\{(S + \beta A^T A)x^k\}_{k=0}^\infty$, $\{\lambda^k\}_{k=0}^\infty$ 和 $\{(T + \beta B^T B)y^k\}_{k=0}^\infty$ 都有界.

当 $S + \beta A^T A \succ 0$, $T + \beta B^T B \succ 0$ 时, $\{w^k\}_{k=0}^\infty$ 有界, 故存在收敛子列 $\{w^{k_j}\}$. 设 w^* 是它的任意极限点, 则 $w^{k_j} \rightarrow w^*$ ($k \rightarrow \infty$). 由于 Ω 是闭集, 故 $w^* \in \Omega$. 此外, 注意到 $G(w^{k_j+1} - \bar{w}^{k_j}) \rightarrow 0$, $r^{k_j+1} \rightarrow 0$, $B(\bar{y}^{k_j} - y^{k_j+1}) \rightarrow 0$, 则令 (3.2) 中 $k = k_j \rightarrow \infty$, 可得

$$\theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0.$$

由 w 的任意性可知 $w^* \in \Omega^*$, 即 $\{w^k\}_{k=0}^\infty$ 的任意极限点都在 Ω^* 中. 最后经过简单证明可证 $\{w^k\}_{k=0}^\infty$ 极限的唯一性^[21]. $\square$

注意到定理 3.2 中的 $S + \beta A^T A \succ 0$ 和 $T + \beta B^T B \succ 0$ 是很容易实现的. 因为当 $S = T = 0$ 时, 为保证 (2.7b) 和 (2.7d) 解的唯一性, 必须要求 A 和 B 是列满秩的, 此时 $S + \beta A^T A \succ 0$ 和 $T + \beta B^T B \succ 0$ 显然成立. 此外, 定理 3.1 和 3.2 是在 α_k 满足假设 3.1 的条件下给出的. 事实上, 基于先前迭代信息来动态地选取 α_k 满足假设 3.1 不难实现. 例如, 选取 $\alpha_k \in [0, \bar{\alpha}_k]$, 其中 $\bar{\alpha}_k = \min\{t, 1/(k|x^k - x^{k-1}|)^2\}$ ^[14].

4 数值试验

本节所有的数值测试都是在 MATLAB-R2010a 中进行的, 运行环境为联想电脑 (Intel(R) Core (TM) i32130 CPU 3.39 GHZ, 3.33 GB).

4.1 最小二乘半定规划问题

本小节把本文算法应用到最小二乘半定规划问题^[23]

$$\min \left\{ \frac{1}{2} \|X - C\|_F^2 \mid X \in S_+^n \cap S_B \right\},$$

其中 C 是给定的 $n \times n$ 对称矩阵, $S_B = \{H \in \mathbb{R}^{n \times n} \mid H_L \leq H \leq H_U\}$, 其中 H_L 和 H_U 是 $n \times n$ 对称矩阵, 且 S_+^n 是对称半正定矩阵锥, $\|\cdot\|_F$ 是标准的 F -范数. 特别地, 它可以写成 (1.1) 的形式:

$$\min \frac{1}{2}\|X - C\|_F^2 + \frac{1}{2}\|Y - C\|_F^2, \quad \text{s.t.} \quad X - Y = 0, \quad X \in S_+^n, \quad Y \in S_B.$$

用本文提出的算法求解该问题时, 每次迭代过程中的子问题都有显式解, 因此在数值测试过程中取 $S = T = 0$, 即利用惯性 PR 分裂方法 (IPRSM) 解决该问题. 相应地, 也分别应用了严格压缩 PR 分裂方法 (SC-PRSM) 和惯性 ADMM (IADMM) [22] 求解该问题. 用 IPRSM 求解该问题时产生的 x, y -子问题求解如下:

$$\begin{aligned} X^{k+1} &= \arg \min \left\{ \frac{1}{2}\|X - C\|_F^2 - \text{Tr}(\bar{Z}^k X) + \frac{\beta}{2}\|X - \bar{Y}^k\|_F^2 \mid X \in S_+^n \right\}, \\ Y^{k+1} &= \arg \min \left\{ \frac{1}{2}\|Y - C\|_F^2 + \text{Tr}(Z^{k+\frac{1}{2}} X) + \frac{\beta}{2}\|X^{k+1} - Y\|_F^2 \mid Y \in S_B \right\}. \end{aligned}$$

x -子问题等价于

$$X^{k+1} = P_{S_+^n} \left\{ \frac{1}{1+\beta}(\beta \bar{Y}^k + \bar{Z}^k + C) \right\},$$

y -子问题等价于

$$Y^{k+1} = P_{S_B} \left\{ \frac{1}{1+\beta}(\beta X^{k+1} - Z^{k+\frac{1}{2}} + C) \right\},$$

其中 $P_{S_+^n}$ 和 P_{S_B} 分别是 S_+^n 和 S_B 上的投影.

在进行数值试验时, 对于该问题中数据的选取参考了文献 [21]. C 是随机产生的对角元属于 $(0, 2)$ 而非对角元属于 $(-1, 1)$ 的对称矩阵, H_U 是对角元为 1 非对角元属于 $(-1, 1)$ 的对称矩阵. 初始值皆设为零矩阵. 对于该问题, 作了两组数值试验. 第一组比较 IADMM 与 IPRSM (试验时记为 IPRSM1). 取 α_k 为常数 0.28 [22], 取 IPRSM1 中 $\alpha = 0.5$. 第二组比较 SC-PRSM 与 IPRSM (试验时记为 IPRSM2). 取 $\alpha = 0.8$ (文献 [6] 说明了 SC-PRSM 中 $\alpha \in [0.8, 0.9]$ 时数值表现最好), 取 $\alpha_k = 0.1$. 两组试验皆取 $\beta = 10$. 终止条件为

$$\max(\max(|X_{i,j}^{k+1} - X_{i,j}^k|, |Y_{i,j}^{k+1} - Y_{i,j}^k|), |Z_{i,j}^{k+1} - Z_{i,j}^k|) \leq 1e - 4.$$

表 1 为不同维数下分别应用 SC-PRSM、IPRSM1 和 IADMM、IPRSM2 解决该问题所用的时间和迭代次数, 其中 Iter. 表示迭代次数, CPU (s) 表示 CPU 时间 (以秒为单位). 从表 1 中可以看出, 本文提出的算法数值表现要好于 SC-PRSM 和 IADMM, 并且随着维数的增大, 本文算法的优势越明显.

表 1 最小二乘半定规划问题的数值结果

n	IADMM		IPRSM1		SC-PRSM		IPRSM2	
	Iter.	CPU (s)	Iter.	CPU (s)	Iter.	CPU (s)	Iter.	CPU (s)
100	32	0.270	28	0.243	27	0.263	27	0.206
200	31	1.363	26	1.141	26	1.090	24	0.981
300	31	3.229	27	2.665	26	2.744	24	1.585
500	32	11.563	27	9.574	27	9.140	25	7.188
800	33	47.594	29	41.055	28	38.251	26	33.677
1,000	35	96.988	30	80.637	30	77.261	27	69.347

4.2 计算机断层成像问题

计算机断层成像问题^[24, 25] 来源于医学图像处理问题. 计算机断层成像技术是使用 X 射线从许多不同的方向穿过身体, 通过从不同方向得到的 X 射线图像重建人体横截面图像的过程, 其中透射信息可以由 Radon 变换表示. 若令 $R \in \mathbb{R}^{m \times n}$ 为 Radon 变换, $x \in \mathbb{R}^n$ 为被测量的人体横截面图像, ϵ 为噪音, 则 X 射线图像为 $x^0 = Rx + \epsilon$. 此外, 断层成像模型等价于以下模型:

$$\min \mu \|\nabla x\|_1 + \frac{1}{2} \|Rx - x^0\|^2, \quad (4.1)$$

其中 $\mu > 0$ 是正则化参数, $\nabla^T = [\nabla_x^T, \nabla_y^T]^T$ 是离散梯度算子, $\nabla_x : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $\nabla_y : \mathbb{R}^n \rightarrow \mathbb{R}^n$ 分别是水平和垂直方向上的有限差分算子, 且 $\|\nabla x\|_1$ 为 x 的全变差半范数 (参见文献 [8]). 此外, 假定 ∇^T 满足周期性边界条件, 因此可以通过离散 Fourier 变换对角化.

通过引入辅助变量 $y \in \mathbb{R}^{2n}$, 则 (4.1) 等价于

$$\begin{aligned} \min \quad & \mu \|y\|_1 + \frac{1}{2} \|Rx - x^0\|^2 \\ \text{s.t.} \quad & \nabla x = y. \end{aligned} \quad (4.2)$$

令 $\theta_1(x) = \frac{1}{2} \|Rx - x^0\|^2$, $\theta_2(y) = \mu \|y\|_1$, $b = 0$, $A = \nabla$, $B = -I$, 则该问题是问题 (1.1) 的特殊情形. 在利用 SC-PRSM 求解 (4.2) 时, 得到的 x -子问题可以简化为

$$(\beta \nabla^T \nabla + R^T R)x^{k+1} = \beta \nabla^T \left(y^k + \frac{1}{\beta} \bar{\lambda}^k \right) + R^T x^0. \quad (4.3)$$

理论上, 当 (4.3) 中 x^{k+1} 的系数矩阵是对称正定阵时, x^{k+1} 可以直接求解得到. 但是对于计算机断层成像问题来讲, 我们无法测量到 x^0 完整数据, 那么矩阵 R 不是方阵, 故无法通过 Fourier 变换对角化. 此时, 直接求解 x -子问题是比较困难的. 因此, 为方便求解, 可以通过加邻近项的方法近似求解子问题. 利用 (2.7) 求解问题 (4.2), 产生的 x, y -子问题如下:

$$\begin{aligned} x^{k+1} &= \arg \min \left\{ \frac{1}{2} \|Rx - x^0\|^2 + \frac{\beta}{2} \left\| \nabla x - \bar{y}^k - \frac{1}{\beta} \bar{\lambda}^k \right\|^2 + \frac{1}{2} \|x - \bar{x}^k\|_S^2 \right\}, \\ y^{k+1} &= \arg \min \left\{ \mu \|y\|_1 + \frac{\beta}{2} \left\| \nabla x^{k+1} - y - \frac{1}{\beta} \bar{\lambda}^{k+\frac{1}{2}} \right\|^2 \right\}. \end{aligned}$$

取 $S = \tau I - R^T R - \beta \nabla^T \nabla$, $\tau > \|R^T R + \beta \nabla^T \nabla\|$, 则 x -子问题可化简为

$$x^{k+1} = \bar{x}^k - \frac{1}{\tau} R^T (R \bar{x}^k - x^0) - \frac{1}{\tau} \nabla^T (\beta \nabla \bar{x}^k - \beta \bar{y}^k - \bar{\lambda}^k).$$

注意到 y -子问题的闭形式解为 $y^{k+1} = \text{shrink}_{\frac{\mu}{\beta}}(\nabla x^{k+1} - \frac{\lambda^{k+\frac{1}{2}}}{\beta})$, 其中 $\text{shrink}_{\frac{\mu}{\beta}}(t) = t - \min\{\frac{\mu}{\beta}, \|t\|\} \frac{t}{\|t\|}$ ^[26].

我们测试了 Shepp-Logan 模型图像 (128×128), 通过标准方差为 1 的白 Gauss 噪音对图像进行模糊化, 图 1 分别是原图像和滤波反投影 (FBP) 图像^[24].

为恢复 FBP 图像, 我们分别应用半邻近 ADMM (sP-ADMM)、惯性半邻近 ADMM (IsP-ADMM)、半邻近 PRSM (sPSC-PRSM) 和 IsP-PRSM 进行图像处理, 且初始图像取 FBP 图. 对于该问题, 同样作了两组试验. 一组比较 sP-ADMM、IsP-ADMM 和 IsP-PRSM, 参数序列 α_k 取为常数 0.28. 对于 IsP-PRSM, 取步长 $\alpha = 0.7$; 另外一组比较 sPSC-PRSM 和 IsP-PRSM, 取 $\alpha = 0.85$, 对于 IsP-PRSM,

参数序列 α_k 取为常数 0.15. 两组试验皆取 $\beta = 0.1$, $\mu = 0.06$. 对于参数 τ , 采用以下更新准则:

$$\tau^{k+1} = \begin{cases} \tau^k, & \text{mod}(k, 20) \neq 0, \\ \max\{\tau^k \times 0.6, 0.8\}, & \text{mod}(k, 20) = 0, \quad k \leq 100, \\ \min\{\tau^k \times 1.01, 5\}, & \text{mod}(k, 20) = 0, \quad k > 100, \end{cases}$$

其中, 为了使 S 半正定, 须有 $\tau \geq \|R^T R + \beta \nabla^T \nabla\|$, 因此取 $\tau^0 = 1.7$. 本文利用改进的信噪比 (ISNR) 来衡量被修复图像的质量, 定义 ISNR 为

$$\text{ISNR} = 10 \log_{10} \frac{\|x^0 - x\|_2^2}{\|x - \hat{x}\|_2^2},$$

其中 x 是原始图像, x^0 是失真图像, $\hat{x}$ 是被修复图像. 信噪比越高则说明图像恢复得越好. 图 2 和 3 是用不同方法恢复出来的 Shepp-Logan 模型图像. 图 4 是信噪比与时间的关系图. 图 5 和 6 是目标函数值与时间的关系图. 从图中可以看出, 本文算法恢复出来的图像质量好于其他方法, 因此该算法与其他方法相比是有一定竞争力的.

图 1 Shepp-Logan 模型图像与它的 FBP 图像

图 2 sP-ADMM、IsP-ADMM 和 IsP-PRSM 恢复出的 Shepp-Logan 模型图像, 此时 $\alpha_k = 0.28$, $\alpha = 0.7$

图 3 sPSC-PRSM 和 IsP-PRSM 恢复出的 Shepp-Logan 模型图像, 此时 $\alpha_k = 0.15$, $\alpha = 0.85$

图 4 信噪比与时间的关系图

图 5 (a) sP-ADMM、IsP-ADMM 和 IsP-PRSM 的目标函数值与时间的关系图; (b) (a) 的放大图

图 6 (a) sPSC-PRSM 和 IsP-PRSM 的目标函数值与时间的关系图; (b) (a) 的放大图

5 总结

本文提出了一种求解带有线性约束且目标函数具有可分离结构的凸优化问题的惯性邻近 PR 分裂算法, 该方法是在半邻近严格压缩的 PR 分裂方法的基础上加入惯性技术得到的方法, 证明了算法的全局收敛性. 数值试验表明, 本文算法改进了严格压缩的 PR 分裂方法和惯性邻近 ADMM, 但算法收敛速度的证明及如何选取 α_k 使算法更快有待进一步研究.

参考文献

- 1 Gabay D, Mercier B. A dual algorithm for the solution of nonlinear variational problems via finite-element approximations. *Comput Math Appl*, 1976, 2: 17–40
- 2 Glowinski R. On alternating direction methods of multipliers: A historical perspective. In: *Modeling, Simulation and Optimization for Science and Technology*. Dordrecht: Springer, 2014, 59–82
- 3 Fang E X, He B S, Liu H, et al. Generalized alternating direction method of multipliers: New theoretical insights and applications. *Math Program Comput*, 2015, 7: 149–187
- 4 Xu M H, Wu T. A class of linearized proximal alternating direction methods. *J Optim Theory Appl*, 2011, 151: 321–337
- 5 Gabay D. Applications of the method of multipliers to variational inequalities. *Stud Math Appl*, 1983, 15: 299–331
- 6 He B S, Liu H, Wang Z R, et al. A strictly contractive Peaceman-Rachford splitting method for convex programming. *SIAM J Optim*, 2014, 24: 1011–1040
- 7 Sun M, Liu J. A proximal Peaceman-Rachford splitting method for compressive sensing. *Appl Math Comput*, 2016, 50: 349–363
- 8 Li X X, Yuan X M. A proximal strictly contractive Peaceman-Rachford splitting method for convex programming with application to imaging. *SIAM J Imaging Sci*, 2015, 8: 1332–1365
- 9 Gu Y, Jiang B, Han D R. A semi-proximal-based strictly contractive Peaceman-Rachford splitting method. *ArXiv: 1506.02221*, 2015
- 10 He B S, Ma F, Yuan X M. On the step size of symmetric alternating direction method of multipliers. [Http://www.optimization-online.org](http://www.optimization-online.org), 2015
- 11 Sun M, Liu J. Generalized Peaceman-Rachford splitting method for separable convex programming with applications to image processing. *J Appl Math Comput*, 2016, 51: 605–622
- 12 Li M, Yuan X M. A strictly contractive Peaceman-Rachford splitting method with logarithmic-quadratic proximal regularization for convex programming. *Math Oper Res*, 2015, 40: 842–858

- 13 Chen Y N, Dai Y H, Han D R. Fiber orientation distribution estimation using a Peaceman-Rachford splitting method. *SIAM J Imaging Sci*, 2016, 9: 573–604
- 14 Alvarez F. On the minimizing property of a second order dissipative system in Hilbert space. *SIAM J Control Optim*, 2000, 38: 1102–1119
- 15 Alvarez F, Attouch H. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Anal*, 2001, 9: 3–11
- 16 Moudafi A, Elisabeth E. Approximate inertial proximal methods using the enlargement of maximal monotone operators. *Int J Pure Appl Math*, 2003, 5: 283–299
- 17 Lorenz D A, Pock T. An inertial forward-backward algorithm for monotone inclusions. *J Math Imaging Vision*, 2015, 51: 311–325
- 18 Chen C H, Ma S Q, Yang J F. A general inertial proximal point algorithm for mixed variational inequality problem. *SIAM J Optim*, 2014, 25: 2120–2142
- 19 Attouch H, Peyrouquet J, Redont P. A dynamical approach to an inertial forward-backward algorithm for convex minimization. *SIAM J Optim*, 2014, 24: 232–256
- 20 Ochs P, Brox T, Pock T. iPiasco: Inertial proximal algorithm for strongly convex optimization. *J Math Imaging Vision*, 2015, 53: 1–11
- 21 Chambolle A, Pock T. On the ergodic convergence rates of a first-order primal-dual algorithm. *Math Program*, 2016, 159: 253–287
- 22 Chen C H, Chan R H, Ma S Q, et al. Inertial proximal ADMM for linearly constrained separable convex optimization. *SIAM J Imaging Sci*, 2015, 8: 2239–2267
- 23 Cai X J, Gu G Y, He B S, et al. A relaxed customized proximal point algorithm for separable convex programming. [Http://www.math.hkbu.edu.hk/~xmyuan/Paper/PPA-ADM-CGHY.pdf](http://www.math.hkbu.edu.hk/~xmyuan/Paper/PPA-ADM-CGHY.pdf), 2011
- 24 Kak A C, Slaney M. *Principles of Computerized Tomographic Imaging*. New York: IEEE, 1988
- 25 阮秋琦, 阮宇智. *数字图像处理*. 第 3 版. 北京: 电子工业出版社, 2011
- 26 Ewout V, Friedlander P. Probing the pareto frontier for basis pursuit solutions. *SIAM J Sci Comput*, 2008, 31: 890–912

An inertial proximal Peaceman-Rachford splitting method

DOU MingYuan, LI HuiYun & LIU XinWei

Abstract The strictly contractive Peaceman-Rachford splitting method (SC-PRSM) is a very efficient first-order approach for linearly constrained separable convex optimization problems, and its convergence rate is faster than that of ADMM. Recently, a semi-proximal Peaceman-Rachford splitting method was proposed to modify SC-PRSM by introducing semi-proximal terms to the subproblems. In this paper, motivated by the idea of the inertial proximal ADMM, we improve the semi-proximal Peaceman-Rachford splitting method by employing an inertial technique to accelerate its convergence, i.e., at each iteration the semi-proximal Peaceman-Rachford splitting method is applied to a point extrapolated at the current iterate in the direction of last movement. The proposed algorithmic framework is very general so that SC-PRSM and semi-proximal Peaceman-Rachford splitting methods are covered as special cases. Based on the asymptotic feasibility of the iterative sequence and the convergence of the function values, we establish the convergence of the whole sequence generated by the proposed algorithm under very mild assumptions. We demonstrate the efficiency of the inertial extrapolation step via experimental results on least squares semi-definite programming problems and total variation based medical image reconstruction problems.

Keywords convex programming, semi-proximal splitting method, Peaceman-Rachford splitting method, inertial proximal point algorithm

MSC(2010) 90C25, 90C30

doi: 10.1360/N012016-00134