

文章编号: 0583-1431(2021)06-1005-22

文献标识码: A

非凸多分块优化部分对称正则化 交替方向乘子法

简金宝

广西民族大学数学与物理学院应用数学与人工智能研究中心
广西混杂计算与集成电路分析重点实验室 南宁 530006
E-mail: jianjb@gxu.edu.cn

刘鹏杰

广西大学数学与信息科学学院 南宁 530004
E-mail: liupengjie2019@163.com

江羨珍

广西民族大学数学与物理学院应用数学与人工智能研究中心
广西混杂计算与集成电路分析重点实验室 南宁 530006
E-mail: yl2811280@163.com

摘 要 交替方向乘子法求解两分块优化的研究已逐渐成熟和完善, 但对于非凸多分块优化的研究相对较少. 本文提出带线性约束的非凸多分块优化的部分对称正则化交替方向乘子法. 首先, 在适当的假设条件下, 包括部分对称乘子修正中参数的估值区域, 证明了算法的全局收敛性. 其次, 当增广拉格朗日函数满足 Kurdyka-Łojasiewicz (KL) 性质时, 证明了算法的强收敛性. 当 KL 性质关联函数具有特殊结构时, 保证了算法的次线性和线性收敛率. 最后, 对算法进行了初步数值试验, 结果表明算法的数值有效性.

关键词 多分块优化; 非凸优化; 交替方向乘子法; Kurdyka-Łojasiewicz 性质; 收敛率
MR(2010) 主题分类 90C26, 90C30
中图分类 O221

A Partially Symmetric Regularized Alternating Direction Method of Multipliers for Nonconvex Multi-block Optimization

Jin Bao JIAN

College of Mathematics and Physics,
Center for Applied Mathematics and Artificial Intelligence,
Guangxi Key Laboratory of Hybrid Computation and IC Design Analysis,
Guangxi University for Nationalities, Nanning 530006, P. R. China
E-mail: jianjb@gxu.edu.cn

收稿日期: 2020-06-10; 接受日期: 2020-11-10

基金项目: 国家自然科学基金 (11771383); 广西自然科学基金 (2020GXNSFDA238017, 2018GXNSFFA281007)

Peng Jie LIU

*College of Mathematics and Information Science, Guangxi University,
Nanning 530004, P. R. China
E-mail: liupengjie2019@163.com*

Xian Zhen JIANG

*College of Mathematics and Physics,
Center for Applied Mathematics and Artificial Intelligence,
Guangxi Key Laboratory of Hybrid Computation and IC Design Analysis,
Guangxi University for Nationalities, Nanning 530006, P. R. China
E-mail: yl2811280@163.com*

Abstract The researches on the alternating direction method of multiplier method (ADMM) for solving two-block optimization have been gradually mature and perfect. However, the studies on ADMM for solving nonconvex multi-block optimization are relatively few. In this paper, we first propose a partially symmetric regularized ADMM for nonconvex multi-block optimization with linear constraints. Second, under appropriate assumptions including the region of the two parameters in the updating formulas for the multiplier, the global convergence of the proposed method is proved. Third, when the augmented Lagrangian function satisfies the Kurdyka–Łojasiewicz (KL) property, the strong convergence of the method is proved. Furthermore, when the associated KL property function has a special structure, the sublinear and linear convergence rate of the method are obtained. Finally, some preliminary numerical experiments are carried out, and this shows that the proposed method is numerically effective.

Keywords multi-block optimization; nonconvex optimization; alternating direction method of multipliers; Kurdyka–Łojasiewicz property; convergence rate

MR(2010) Subject Classification 90C26, 90C30

Chinese Library Classification O221

1 引言

工程计算中, 许多优化问题都可以转化成具有线性约束的几个最小化函数和的形式, 例如压缩感知 [35, 38], 机器学习 [39], 频谱分解问题 [33], 波束形成技术 [25], 电气最优潮流 [36] 等. 由于拥有可分结构, 分裂算法在处理此类问题时具有独特优势 [4]. 线性约束两分块优化问题可描述为:

$$\begin{aligned} \min \quad & f_1(x_1) + f_2(x_2), \\ \text{s.t.} \quad & \psi(x_1, x_2) := A_1 x_1 + A_2 x_2 - b = 0. \end{aligned} \quad (1.1)$$

函数 $f_1 : \mathbb{R}^{n_1} \rightarrow \mathbb{R} \cup \{+\infty\}$ 正常下半连续, 函数 $f_2 : \mathbb{R}^{n_2} \rightarrow \mathbb{R}$ 光滑, 矩阵 $A_1 \in \mathbb{R}^{p \times n_1}$, $A_2 \in \mathbb{R}^{p \times n_2}$, 向量 $b \in \mathbb{R}^p$. 问题 (1.1) 的增广拉格朗日函数为

$$\mathcal{L}_\beta(x, \lambda) = \mathcal{L}_\beta(x_1, x_2, \lambda) = f_1(x_1) + f_2(x_2) - \langle \lambda, \psi(x) \rangle + \frac{\beta}{2} \|\psi(x)\|^2,$$

其中 $x = (x_1, x_2) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$, $\lambda \in \mathbb{R}^p$ 为线性约束的拉格朗日乘子, $\beta > 0$ 为罚参数.

交替方向乘子法 (ADMM) 是基于 Douglas–Rachford (DR) 分裂思想 [12, 27] 建立起来的求

解可分优化的一类分裂算法, 其求解问题 (1.1) 的经典迭代模式为:

$$x_1^{k+1} = \arg \min_{x_1} \{\mathcal{L}_\beta(x_1, x_2^k, \lambda^k)\}, \quad x_2^{k+1} = \arg \min_{x_2} \{\mathcal{L}_\beta(x_1^{k+1}, x_2, \lambda^k)\}, \quad \lambda^{k+1} = \lambda^k - \beta\psi(x^{k+1}).$$

在目标函数凸的情况下, 对 ADMM 的理论与算法方面的研究已较为完善 [6, 13, 14, 18, 20, 37]. Glowinski [13] 提出在拉格朗日乘子更新步中引入一松弛因子 τ , 即 $\lambda^{k+1} = \lambda^k - \tau\beta\psi(x^{k+1})$. 参数 τ 有利于算法加快收敛速度, 提高数值效果. He 等人 [20] 考虑了 ADMM 的对称版本, 并选取不同的松弛因子 r 和 s , 即

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1} \{\mathcal{L}_\beta(x_1, x_2^k, \lambda^k)\}, & \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta\psi(x_1^{k+1}, x_2^k), \\ x_2^{k+1} = \arg \min_{x_2} \{\mathcal{L}_\beta(x_1^{k+1}, x_2, \lambda^{k+\frac{1}{2}})\}, & \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta\psi(x^{k+1}). \end{cases}$$

上述算法与经典 ADMM 的最大不同是在迭代过程中加入乘子的中间更新项 $\lambda^{k+\frac{1}{2}}$, 此最早来源于 Peaceman 和 Rachford [29] 的分裂思想 (故又称 PR 分裂). 为使上述算法具有更好的理论特征, Gu 等人 [14] 在子问题中分别添加半邻近项, 并给出相应的收敛率分析, 其算法迭代格式为:

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1} \left\{ \mathcal{L}_\beta(x_1, x_2^k, \lambda^k) + \frac{1}{2} \|x_1 - x_1^k\|_P^2 \right\}, & \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta\psi(x_1^{k+1}, x_2^k), \\ x_2^{k+1} = \arg \min_{x_2} \left\{ \mathcal{L}_\beta(x_1^{k+1}, x_2, \lambda^{k+\frac{1}{2}}) + \frac{1}{2} \|x_2 - x_2^k\|_Q^2 \right\}, & \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta\psi(x^{k+1}), \end{cases}$$

其中, P 和 Q 为对称半正定矩阵, $\|x\|_P^2 = x^\top Px$.

近年来, 非凸两分块的 ADMM 及收敛性分析成果不断涌现 [2, 8, 16, 24, 26, 32, 34]. Wang 等人 [31, 32] 对 ADMM 的两个子问题都加上 Bregman 距离函数, 在 A_1 为行满秩, x_2 -子问题的 Bregman 距离函数强凸, 且 $\mathcal{L}_\beta(x_1, x_2, \lambda)$ 关于 x_1 或 x_1 -子问题的 Bregman 距离函数强凸的情况下, 分析了相应算法收敛结果. Jian 等人 [24] 针对一类特殊两分块非凸优化问题 (A_2 为单位阵, $b = 0$) 提出一种由 ADMM 衍生的正则化版本, 即对经典 ADMM 第一个子问题正则化, 但无需限制正则项强凸, 其迭代如下:

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1} \{\mathcal{L}_\beta(x_1, x_2^k, \lambda^k) + \frac{1}{2} \|x_1 - x_1^k\|_G^2\}, \\ x_2^{k+1} = \arg \min_{x_2} \{\mathcal{L}_\beta(x_1^{k+1}, x_2, \lambda^k)\}, & \lambda^{k+1} = \lambda^k - \beta(A_1 x_1^{k+1} - x_2^{k+1}). \end{cases}$$

本文考虑如下带线性约束的非凸多分块优化问题:

$$\begin{aligned} \min \quad & f(x) := \sum_{i=1}^m f_i(x_i), \\ \text{s.t.} \quad & \sum_{i=1}^m A_i x_i = b, \end{aligned} \tag{1.2}$$

其中, $x = (x_1, \dots, x_m)$, 函数 $f_i: \mathbb{R}^{n_i} \rightarrow \mathbb{R} \cup \{+\infty\}$ ($i = 1, \dots, m-1$) 均正常下半连续, $f_m: \mathbb{R}^p \rightarrow \mathbb{R}$ 光滑且其梯度 ∇f_m 为 Lipschitz 连续, 矩阵 $A_i \in \mathbb{R}^{p \times n_i}$, $i = 1, \dots, m-1$, 向量 $b \in \mathbb{R}^p$. 当 A_m 列满秩 ($A_m^\top A_m$ 可逆) 时, 模型 (1.2) 中线性约束两边同左乘 $(A_m^\top A_m)^{-1} A_m^\top$, 最后的矩阵即可转化成单位阵的情形. 在实际工程计算中也广泛存在 A_m 为单位阵这一情形, 如压缩感知 [35, 38], 稳健主成分分析 [5, 31] 等. 因此, 为便于构造算法和收敛性分析, 在模型 (1.2) 中设 A_m 为 p 阶单位阵 I_p , $A_m x_m = x_m$.

对于三分块及其以上的多分块优化问题, 相应的 ADMM 或其它分裂算法的收敛性仍有待深入研究, 甚至无法收敛 [9, 10]. Han 和 Yuan [17] 首先证明了当所有目标函数 f_i 均强凸, 且罚参数

小于某一阈值时, 由经典两分块 ADMM 直接推广的多分块 ADMM 版本具有全局收敛性. 另一方面, He, Tao 和 Yuan 通过修正并推广 ADMM, 获得多分块 ADMM. 如部分平行分裂 ALM 的预测校正法^[19], 带高斯回代 ADMM^[21, 23], 部分平行正则项的 ADMM^[22] 等.

目前, 多分块优化分裂算法研究更多聚焦于凸问题. 对于非凸多分块优化的 ADMM 或其它分裂算法的研究相对较少. Guo 等人^[15] 将非凸两分块的经典 ADMM 直接推广到多分块情形, 在 A_i 列满秩且罚参数局限于某一范围时, 分析了算法的全局收敛性. 此外, 若效益函数满足 Kurdyka-Łojasiewicz 性质, 还获得了算法的强收敛性. Wang 等人^[31] 结合 Bregman 距离, 得到了非凸多分块情形 BADMM 的收敛性.

定义问题 (1.2) 的增广拉格朗日函数 $\mathcal{L}_\beta(x_1, \dots, x_m, \lambda)$:

$$\mathcal{L}_\beta(x, \lambda) = \mathcal{L}_\beta(x_1, \dots, x_m, \lambda) = \sum_{i=1}^m f_i(x_i) - \left\langle \lambda, \sum_{i=1}^m A_i x_i - b \right\rangle + \frac{\beta}{2} \left\| \sum_{i=1}^m A_i x_i - b \right\|^2. \quad (1.3)$$

基于上述分析, 本文考虑一种求解问题 (1.2) 的部分对称正则化交替方向乘子法 (PSRADMM) 迭代:

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1} \left\{ \mathcal{L}_\beta(x_1, x_2^k, \dots, x_m^k, \lambda^k) + \frac{1}{2} \|x_1 - x_1^k\|_{F_1}^2 \right\}, \end{cases} \quad (1.4a)$$

$$\begin{cases} x_2^{k+1} = \arg \min_{x_2} \left\{ \mathcal{L}_\beta(x_1^{k+1}, x_2, x_3^k, \dots, x_m^k, \lambda^k) + \frac{1}{2} \|x_2 - x_2^k\|_{F_2}^2 \right\}, \end{cases} \quad (1.4b)$$

$$\begin{cases} \vdots \\ x_{m-1}^{k+1} = \arg \min_{x_{m-1}} \left\{ \mathcal{L}_\beta(x_1^{k+1}, \dots, x_{m-2}^{k+1}, x_{m-1}, x_m^k, \lambda^k) + \frac{1}{2} \|x_{m-1} - x_{m-1}^k\|_{F_{m-1}}^2 \right\}, \end{cases} \quad (1.4c)$$

$$\begin{cases} \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta \left(\sum_{i=1}^{m-1} A_i x_i^{k+1} + x_m^k - b \right), \end{cases} \quad (1.4d)$$

$$\begin{cases} x_m^{k+1} = \arg \min_{x_m} \{ \mathcal{L}_\beta(x_1^{k+1}, \dots, x_{m-1}^{k+1}, x_m, \lambda^{k+\frac{1}{2}}) \}, \end{cases} \quad (1.4e)$$

$$\begin{cases} \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right), \end{cases} \quad (1.4f)$$

其中, $F_i \in \mathbb{R}^{n_i \times n_i}$ ($i = 1, \dots, m-1$) 均为对称正定矩阵. 对于上面 PSRADMM 迭代的初始点 $w^0 = (x_1^0, \dots, x_m^0, \lambda^0)$, 假设 $\mathcal{L}_\beta(w^0) < +\infty$.

显然, PSRADMM 迭代按

$$x_1^{k+1} \rightarrow x_2^{k+1} \rightarrow \dots \rightarrow x_{m-1}^{k+1} \rightarrow \lambda^{k+\frac{1}{2}} \rightarrow x_m^{k+1} \rightarrow \lambda^{k+1}$$

顺序更新原始变量与对偶变量, 并具有以下特征:

(i) 分别求解 x_i -子问题 ($i = 1, \dots, m$), 且求解 x_i -子问题 ($i = 1, \dots, m-1$) 时所用对偶变量信息均为 λ^k ;

(ii) 充分利用所获最新信息 $(x_1^{k+1}, x_2^{k+1}, \dots, x_{m-1}^{k+1}, \lambda^{k+\frac{1}{2}})$ 产生 x_m^{k+1} ;

(iii) 如取 $F_i = \alpha I - \beta A_i^\top A_i$, 则 (1.4) 中关于 x_i -子问题 ($i = 1, \dots, m-1$) 可简化为

$$\min_{x_i} \left\{ f_i(x_i) + \frac{\alpha}{2} \|x_i - b_i^k\|^2 \right\},$$

其中 $b_i^k \in \mathbb{R}^{n_i}$ 为某一已知量.

本文第 2 节给出一些与本文分析相关的预备知识, 第 3 节和第 4 节分别对 PSRADMM 的收敛性和收敛率进行分析论证, 第 5 节对 PSRADMM 进行初步数值试验.

2 预备知识

本文中, $\|\cdot\|$ 表示欧氏范数. 对任意 $x, y \in \mathbb{R}^n$, $\langle x, y \rangle = x^\top y$. 对于对称半正定矩阵 G , $\|x\|_G^2 = x^\top Gx$, $G \succeq (\succ) 0$ 表示 G 为对称半正定 (正定) 矩阵. $\lambda_{\min}(G)$ 与 $\lambda_{\max}(G)$ 分别表示对称矩阵 G 的最小特征值及最大特征值, 则有 $\lambda_{\min}(G)\|x\|^2 \leq \|x\|_G^2 \leq \lambda_{\max}(G)\|x\|^2$, $\forall x \in \mathbb{R}^n$.

对任意集合 $S \subseteq \mathbb{R}^n$, 任意点 $x \in \mathbb{R}^n$, 若 S 非空, 点 x 到 S 的距离记为 $d(x, S) = \inf_{y \in S} \|y - x\|$. 特别地, 若 $S = \emptyset$, 则令 $d(x, S) = +\infty$.

函数 $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ 的定义域与上图分别记作 $\text{dom}f$, $\text{epi}f$, 即

$$\text{dom}f = \{x \in \mathbb{R}^n : f(x) < +\infty\}, \quad \text{epi}f = \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq \alpha\}.$$

若 $\text{dom}f \neq \emptyset$, 则称 f 为正常函数.

定义 2.1 [30] 若函数 $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ 在 x_0 处满足 $f(x_0) \leq \liminf_{x \rightarrow x_0} f(x)$, 则称函数 f 在 x_0 处下半连续. 若 f 在每一点处均下半连续, 则称 f 为下半连续函数.

定义 2.2 [30] 设函数 $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ 正常下半连续:

(i) f 在 $x \in \text{dom}f$ 处的 Fréchet 次微分定义为

$$\hat{\partial}f(x) = \left\{ x^* \in \mathbb{R}^n : \liminf_{y \neq x, y \rightarrow x} \frac{f(y) - f(x) - \langle x^*, y - x \rangle}{\|y - x\|} \geq 0 \right\},$$

当 $x \notin \text{dom}f$ 时, 令 $\hat{\partial}f(x) = \emptyset$.

(ii) f 在 $x \in \text{dom}f$ 处的极限次微分定义为

$$\partial f(x) = \{x^* \in \mathbb{R}^n : \exists x_k \rightarrow x, f(x_k) \rightarrow f(x), \hat{x}_k \in \hat{\partial}f(x_k), \hat{x}_k \rightarrow x^*\}.$$

注 2.3 对于本文用到的几个次微分的性质 (见文 [30]), 列示如下:

(i) $\forall x \in \mathbb{R}^n$, $\hat{\partial}f(x) \subseteq \partial f(x)$, 且 $\hat{\partial}f(x)$ 为闭凸集, $\partial f(x)$ 为闭集.

(ii) 设 $x_k^* \in \partial f(x_k)$ 且 $\lim_{k \rightarrow \infty} (x_k, x_k^*) = (x, x^*)$, 则 $x^* \in \partial f(x)$, 即 $\partial f(x)$ 具有闭性.

(iii) 若 $x \in \mathbb{R}^n$ 为 f 的极小值点, 则 $0 \in \partial f(x)$. 若 $0 \in \partial f(x)$, 称 x 为函数 f 的稳定点, 函数 f 稳定点的集合记为 $\text{crit}f$.

(iv) 若 $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ 正常下半连续, $g: \mathbb{R}^n \rightarrow \mathbb{R}$ 连续可微, 则对于任意 $x \in \text{dom}f$, 有 $\partial(f+g)(x) = \partial f(x) + \nabla g(x)$.

定义 2.4 [3] (Kurdyka-Łojasiewicz 性质) 设函数 $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ 正常下半连续, $\bar{x} \in \text{dom}(\partial f) := \{x \in \mathbb{R}^n : \partial f(x) \neq \emptyset\}$. 记 $[\eta_1 < f < \eta_2] := \{x \in \mathbb{R}^n : \eta_1 < f(x) < \eta_2\}$. 若存在 $\eta \in (0, +\infty]$, \bar{x} 的某一邻域 U 以及满足如下条件的凹函数 $\varphi: [0, \eta) \rightarrow [0, +\infty)$:

(i) $\varphi(0) = 0$;

(ii) φ 在 0 处连续, 在区间 $(0, \eta)$ 上一阶连续可微;

(iii) $\varphi'(t) > 0, \forall t \in (0, \eta)$,

使得对于任意 $x \in U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]$, 如下 Kurdyka-Łojasiewicz 不等式

$$\varphi'(f(x) - f(\bar{x}))d(0, \partial f(x)) \geq 1$$

成立, 则称函数 f 在点 \bar{x} 处具有 Kurdyka-Łojasiewicz 性质 (简称 KL 性质), 同时称 φ 为 f 的 KL 性质关联函数.

记 Φ_η 是满足上述定义 2.4 中条件 (i), (ii), (iii) 的全体凹函数的集合. 若 f 在定义域 $\text{dom}f$ 的每一点处都满足 KL 性质, 则称 f 为 KL 函数.

引理 2.5 ^[3] (一致 KL 性质) 设 Ω 是一个紧集, 函数 $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ 正常下半连续. 设函数 f 在 Ω 上取固定常数, 并在 Ω 的任意点处均满足 KL 性质, 则存在 $\epsilon > 0, \eta > 0, \varphi \in \Phi_\eta$, 使得对于任意 $\bar{x} \in \Omega$ 及 $x \in \{x \in \mathbb{R}^n : d(x, \Omega) < \epsilon\} \cap [f(\bar{x}) < f(x) < f(\bar{x}) + \eta]$, 有

$$\varphi'(f(x) - f(\bar{x}))d(0, \partial f(x)) \geq 1.$$

实际应用中, 许多函数均满足 KL 性质 ^[31], 如半代数函数、强凸函数、实解析函数、次解析函数等.

引理 2.6 ^[28] 设函数 h 连续可微, 且其梯度 ∇h 为 L_h -Lipschitz 连续, 则有

$$|h(y) - h(x) - \langle \nabla h(x), y - x \rangle| \leq \frac{L_h}{2} \|y - x\|^2, \quad \forall x, y \in \mathbb{R}^n. \quad (2.1)$$

记 $w = (x, \lambda)$, 则由 $\mathcal{L}_\beta(w)$ 的定义 (1.3) 有

$$\begin{cases} \partial_{x_i} \mathcal{L}_\beta(w) = \partial f_i(x_i) - A_i^\top \lambda + \beta A_i^\top \left(\sum_{j=1}^m A_j x_j - b \right), & i = 1, \dots, m-1, \end{cases} \quad (2.2a)$$

$$\begin{cases} \partial_{x_m} \mathcal{L}_\beta(w) = \nabla f_m(x_m) - \lambda + \beta \left(\sum_{j=1}^m A_j x_j - b \right), \end{cases} \quad (2.2b)$$

$$\begin{cases} \partial_\lambda \mathcal{L}_\beta(w) = - \left(\sum_{j=1}^m A_j x_j - b \right). \end{cases} \quad (2.2c)$$

由此有下面引理.

引理 2.7 $w^* = (x^*, \lambda^*)$ 为 $\mathcal{L}_\beta(\cdot)$ 的稳定点, 即 $0 \in \partial \mathcal{L}_\beta(w^*)$, 当且仅当

$$A_i^\top \lambda^* \in \partial f_i(x_i^*) \quad (i = 1, 2, \dots, m-1), \quad \nabla f_m(x_m^*) = \lambda^*, \quad \sum_{i=1}^m A_i x_i^* - b = 0. \quad (2.3)$$

3 收敛性分析

余下的收敛性分析, 对于问题 (1.2) 及任意参数 $\theta \in \mathbb{R}$, 由 (1.3) 定义的增广拉格朗日函数 $\mathcal{L}_\beta(x, \lambda)$ 具有如下重要性质:

$$\mathcal{L}_\beta \left(x_1, \dots, x_m, \lambda - \theta \left(\sum_{i=1}^m A_i x_i - b \right) \right) = \mathcal{L}_\beta(x, \lambda) + \theta \left\| \sum_{i=1}^m A_i x_i - b \right\|^2. \quad (3.1)$$

此性质可大大简化下面的分析. 记

$$\begin{aligned} x &= (x_1, \dots, x_m), \quad x^k = (x_1^k, \dots, x_m^k), \quad x^* = (x_1^*, \dots, x_m^*); \\ w &= (x, \lambda), \quad w^k = (x^k, \lambda^k), \quad w^* = (x^*, \lambda^*). \end{aligned}$$

下面分析中, 始终假设 $\{w^k\}$ 为 PSRADMM (1.4) 所产生的迭代点列, 且有界.

假设 3.1 (i) 问题 (1.2) 中的函数 f_i ($i = 1, 2, \dots, m-1$) 均为正常下半连续, 函数 f_m 光滑且其梯度 ∇f_m 为 L_m -Lipschitz 连续, 即 $\|\nabla f_m(x) - \nabla f_m(y)\| \leq L_m \|x - y\|, \forall x, y \in \mathbb{R}^p$;

(ii) 迭代 (1.4) 中的 F_i ($i = 1, 2, \dots, m-1$) 均为对称正定矩阵.

首先, 根据 PSRADMM (1.4) 中各子问题的最优性必要条件, 有

$$\begin{cases} 0 \in \partial f_i(x_i^{k+1}) - A_i^\top \lambda^k + \beta A_i^\top \left(\sum_{j=1}^i A_j x_j^{k+1} + \sum_{j=i+1}^m A_j x_j^k - b \right) \\ \quad + F_i(x_i^{k+1} - x_i^k), \quad i = 1, \dots, m-1, \end{cases} \quad (3.2a)$$

$$\begin{cases} 0 = \nabla f_m(x_m^{k+1}) - \lambda^{k+\frac{1}{2}} + \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right). \end{cases} \quad (3.2b)$$

为分析 PSRADMM (1.4) 产生的 (效益函数) 序列 $\{\mathcal{L}_\beta(w^k)\}$ 的单调性, 记

$$\begin{cases} \delta_i := \delta_i(r, s, \beta) = \frac{1}{2} \lambda_{\min}(F_i) - \frac{(m+1)\beta(1-s)^2}{r+s} \lambda_{\max}(A_i^\top A_i), \quad i = 1, \dots, m-1, \end{cases} \quad (3.3a)$$

$$\begin{cases} \delta_m := \delta_m(r, s, \beta) = \frac{\beta - L_m}{2} - \frac{(m+1)[\beta^2(1-s)^2 + L_m^2]}{(r+s)\beta} - \frac{rs\beta}{r+s} \\ \quad = \frac{[r(1-2s) + s - 2(m+1)(1-s)^2]\beta^2 - (r+s)L_m\beta - 2(m+1)L_m^2}{2(r+s)\beta}, \end{cases} \quad (3.3b)$$

$$\begin{cases} \delta := \delta(r, s, \beta) = \min\{\delta_1, \dots, \delta_m\} = \min\{\delta_1(r, s, \beta), \dots, \delta_m(r, s, \beta)\}. \end{cases} \quad (3.3c)$$

引理 3.2 假设 3.1 成立, 且 (1.4d) 及 (1.4f) 中参数 r 和 s 满足 $r + s > 0$, 则

$$\mathcal{L}_\beta(w^{k+1}) \leq \mathcal{L}_\beta(w^k) - \delta \|x^{k+1} - x^k\|^2, \quad \forall k. \quad (3.4)$$

证明 首先, 由 (1.4f), (1.4d) 及 (3.1), 得

$$\mathcal{L}_\beta(w^{k+1}) = \mathcal{L}_\beta(x^{k+1}, \lambda^{k+\frac{1}{2}}) + s\beta \left\| \sum_{i=1}^m A_i x_i^{k+1} - b \right\|^2 \quad (3.5)$$

与

$$\begin{aligned} & \mathcal{L}_\beta(x_1^{k+1}, \dots, x_{m-1}^{k+1}, x_m^k, \lambda^{k+\frac{1}{2}}) \\ &= \mathcal{L}_\beta(x_1^{k+1}, \dots, x_{m-1}^{k+1}, x_m^k, \lambda^k) + r\beta \left\| \sum_{i=1}^{m-1} A_i x_i^{k+1} + x_m^k - b \right\|^2. \end{aligned} \quad (3.6)$$

另一方面, 由 (1.3) 及 (3.2b), 有

$$\begin{aligned} & \mathcal{L}_\beta(x^{k+1}, \lambda^{k+\frac{1}{2}}) - \mathcal{L}_\beta(x_1^{k+1}, \dots, x_{m-1}^{k+1}, x_m^k, \lambda^{k+\frac{1}{2}}) \\ & \stackrel{(1.3)}{=} f_m(x_m^{k+1}) - f_m(x_m^k) - \frac{\beta}{2} \|x_m^{k+1} - x_m^k\|^2 - \left\langle \lambda^{k+\frac{1}{2}} - \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right), x_m^{k+1} - x_m^k \right\rangle \\ & \stackrel{(3.2b)}{=} f_m(x_m^{k+1}) - f_m(x_m^k) - \langle \nabla f_m(x_m^{k+1}), x_m^{k+1} - x_m^k \rangle - \frac{\beta}{2} \|x_m^{k+1} - x_m^k\|^2. \end{aligned} \quad (3.7)$$

又由 ∇f_m 的 L_m -Lipschitz 连续性及 (2.1) 知

$$f_m(x_m^{k+1}) - f_m(x_m^k) - \langle \nabla f_m(x_m^{k+1}), x_m^{k+1} - x_m^k \rangle \leq \frac{L_m}{2} \|x_m^{k+1} - x_m^k\|^2.$$

将上式代入 (3.7), 得

$$\mathcal{L}_\beta(x^{k+1}, \lambda^{k+\frac{1}{2}}) - \mathcal{L}_\beta(x_1^{k+1}, \dots, x_{m-1}^{k+1}, x_m^k, \lambda^{k+\frac{1}{2}}) \leq -\frac{\beta - L_m}{2} \|x_m^{k+1} - x_m^k\|^2. \quad (3.8)$$

其次, 因为 $x_1^{k+1}, \dots, x_{m-1}^{k+1}$ 分别是 (1.4a)–(1.4c) 的最优解, 故

$$\begin{aligned} \mathcal{L}_\beta(x_1^{k+1}, x_2^k, \dots, x_m^k, \lambda^k) - \mathcal{L}_\beta(x_1^k, x_2^k, \dots, x_m^k, \lambda^k) &\leq -\frac{1}{2}\|x_1^{k+1} - x_1^k\|_{F_1}^2, \\ \mathcal{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3^k, \dots, x_m^k, \lambda^k) - \mathcal{L}_\beta(x_1^{k+1}, x_2^k, x_3^k, \dots, x_m^k, \lambda^k) &\leq -\frac{1}{2}\|x_2^{k+1} - x_2^k\|_{F_2}^2, \\ &\vdots \\ \mathcal{L}_\beta(x_1^{k+1}, \dots, x_{m-1}^{k+1}, x_m^k, \lambda^k) - \mathcal{L}_\beta(x_1^{k+1}, \dots, x_{m-2}^{k+1}, x_{m-1}^k, x_m^k, \lambda^k) &\leq -\frac{1}{2}\|x_{m-1}^{k+1} - x_{m-1}^k\|_{F_{m-1}}^2. \end{aligned}$$

将上述 $m-1$ 个不等式相加, 得

$$\begin{aligned} \mathcal{L}_\beta(x_1^{k+1}, \dots, x_{m-1}^{k+1}, x_m^k, \lambda^k) - \mathcal{L}_\beta(w^k) &\leq -\frac{1}{2} \sum_{i=1}^{m-1} \|x_i^{k+1} - x_i^k\|_{F_i}^2 \\ &\leq -\frac{1}{2} \sum_{i=1}^{m-1} \lambda_{\min}(F_i) \|x_i^{k+1} - x_i^k\|^2. \end{aligned} \quad (3.9)$$

因此, 联立关系式 (3.5), (3.6), (3.8) 及 (3.9), 得

$$\begin{aligned} \mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^k) &\stackrel{(3.5)}{=} \mathcal{L}_\beta(x^{k+1}, \lambda^{k+\frac{1}{2}}) + s\beta \left\| \sum_{i=1}^m A_i x_i^{k+1} - b \right\|^2 - \mathcal{L}_\beta(w^k) \\ &\stackrel{(3.8)}{\leq} \mathcal{L}_\beta(x_1^{k+1}, \dots, x_{m-1}^{k+1}, x_m^k, \lambda^{k+\frac{1}{2}}) \\ &\quad - \frac{\beta - L_m}{2} \|x_m^{k+1} - x_m^k\|^2 + s\beta \left\| \sum_{i=1}^m A_i x_i^{k+1} - b \right\|^2 - \mathcal{L}_\beta(w^k) \\ &\stackrel{(3.6)}{=} \mathcal{L}_\beta(x_1^{k+1}, \dots, x_{m-1}^{k+1}, x_m^k, \lambda^k) + r\beta \left\| \sum_{i=1}^{m-1} A_i x_i^{k+1} + x_m^k - b \right\|^2 \\ &\quad - \frac{\beta - L_m}{2} \|x_m^{k+1} - x_m^k\|^2 + s\beta \left\| \sum_{i=1}^m A_i x_i^{k+1} - b \right\|^2 - \mathcal{L}_\beta(w^k) \\ &\stackrel{(3.9)}{\leq} -\frac{1}{2} \sum_{i=1}^{m-1} \lambda_{\min}(F_i) \|x_i^{k+1} - x_i^k\|^2 + r\beta \left\| \sum_{i=1}^{m-1} A_i x_i^{k+1} + x_m^k - b \right\|^2 \\ &\quad - \frac{\beta - L_m}{2} \|x_m^{k+1} - x_m^k\|^2 + s\beta \left\| \sum_{i=1}^m A_i x_i^{k+1} - b \right\|^2. \end{aligned} \quad (3.10)$$

又由 (1.4d) 和 (1.4f), 易推得

$$\sum_{i=1}^{m-1} A_i x_i^{k+1} + x_m^k - b = \frac{1}{(r+s)\beta} (\lambda^k - \lambda^{k+1}) - \frac{s}{r+s} (x_m^{k+1} - x_m^k), \quad (3.11)$$

$$\sum_{i=1}^m A_i x_i^{k+1} - b = \frac{1}{(r+s)\beta} (\lambda^k - \lambda^{k+1}) + \frac{r}{r+s} (x_m^{k+1} - x_m^k). \quad (3.12)$$

另一方面, 由 (3.2b) 和 (1.4f) 得

$$\lambda^{k+1} = \nabla f_m(x_m^{k+1}) + \beta(1-s) \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right). \quad (3.13)$$

结合 ∇f_m 的 L_m -Lipschitz 连续性, 有

$$\|\lambda^{k+1} - \lambda^k\| \leq L_m \|x_m^{k+1} - x_m^k\| + \beta|1-s| \left(\sum_{i=1}^{m-1} \|A_i(x_i^{k+1} - x_i^k)\| + \|x_m^{k+1} - x_m^k\| \right). \quad (3.14)$$

进而, 由 Cauchy 不等式, 有

$$\frac{\|\lambda^{k+1} - \lambda^k\|^2}{m+1} \leq (L_m^2 + \beta^2(1-s)^2) \|x_m^{k+1} - x_m^k\|^2 + \beta^2(1-s)^2 \sum_{i=1}^{m-1} \|A_i(x_i^{k+1} - x_i^k)\|^2. \quad (3.15)$$

于是, 联立 (3.11) 与 (3.12), 并结合 (3.15) 及 $r+s > 0$, 得

$$\begin{aligned} & r\beta \left\| \sum_{i=1}^{m-1} A_i x_i^{k+1} + x_m^k - b \right\|^2 + s\beta \left\| \sum_{i=1}^m A_i x_i^{k+1} - b \right\|^2 \\ &= \frac{1}{(r+s)\beta} \|\lambda^k - \lambda^{k+1}\|^2 + \frac{rs\beta}{r+s} \|x_m^{k+1} - x_m^k\|^2 \\ &\leq \frac{(m+1)\beta(1-s)^2}{r+s} \sum_{i=1}^{m-1} \lambda_{\max}(A_i^\top A_i) \|x_i^{k+1} - x_i^k\|^2 \\ &\quad + \left(\frac{(m+1)(\beta^2(1-s)^2 + L_m^2)}{(r+s)\beta} + \frac{rs\beta}{r+s} \right) \|x_m^{k+1} - x_m^k\|^2. \end{aligned} \quad (3.16)$$

将 (3.16) 代入 (3.10), 并注意到 (3.3), 有

$$\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^k) \leq - \sum_{i=1}^m \delta_i \|x_i^{k+1} - x_i^k\|^2 \leq -\delta \sum_{i=1}^m \|x_i^{k+1} - x_i^k\|^2 = -\delta \|x^{k+1} - x^k\|^2.$$

引理获证.

由 (3.4) 知, $\delta > 0$ 是数列 $\{\mathcal{L}_\beta(w^k)\}$ 具有良好下降性的充分条件, 而这恰好可通过对参数 $(r, s, \beta, \lambda_{\min}(F_1), \dots, \lambda_{\min}(F_{m-1}))$ 合理选择和控制得以实现. 据此给出这些参数的限制条件.

假设 3.3 PSRADMM (1.4) 中的参数 (r, s, β) 以及矩阵 F_i ($i = 1, 2, \dots, m-1$) 满足:

(i) $r+s > 0$, 且 $r(1-2s) + s - 2(m+1)(1-s)^2 > 0$, 此不等式组的解集可表示为 $(r, s) \in D := D_1 \cup D_2$, 其中

$$D_1 := \left(\frac{2(m+1)(1-s)^2 - s}{1-2s}, +\infty \right) \times \left(-\infty, \frac{1}{2} \right), \quad (3.17)$$

$$D_2 := \left(-s, \frac{2(m+1)(1-s)^2 - s}{1-2s} \right) \times \left(\frac{m+1-\sqrt{m+1}}{m}, \frac{m+1+\sqrt{m+1}}{m} \right); \quad (3.18)$$

$$(ii) \quad \beta > \beta_0 := \frac{(r+s) + \sqrt{(r+s)^2 + 8(m+1)[r(1-2s) + s - 2(m+1)(1-s)^2]}}{2[r(1-2s) + s - 2(m+1)(1-s)^2]} L_m;$$

$$(iii) \quad \lambda_{\min}(F_i) > \frac{2(m+1)\beta(1-s)^2}{r+s} \lambda_{\max}(A_i^\top A_i) \quad (i = 1, 2, \dots, m-1).$$

由 (3.17), (3.18) 可见, (r, s) 有较为广泛的选取余地, 特别地, 当 $s \in (\frac{m+1}{m+2}, 1)$ 时, $(r = s, s) \in D_2$. 从而, (1.4) 中的参数 r 和 s 在此区间可取同一值.

引理 3.4 若假设 3.1 和假设 3.3 成立, 则由 (3.3) 定义的 $\delta_i = \delta_i(r, s, \beta) > 0$ ($i = 1, 2, \dots, m$), $\delta = \delta(r, s, \beta) > 0$. 进而, 数列 $\{\mathcal{L}_\beta(w^k)\}$ 单调递减.

证明 首先, 由 (3.3a) 和假设 3.3 (iii) 知 $\delta_i > 0$ ($i = 1, 2, \dots, m-1$). 其次, 由 (3.3b) 和假设 3.3 (i) (ii) 知 $\delta_m > 0$ 成立. 最后, 由 (3.3c) 知 $\delta > 0$ 成立. 进而, 由 (3.4) 知数列 $\{\mathcal{L}_\beta(w^k)\}$ 单调递减. 证毕.

引理 3.5 若假设 3.1 和假设 3.3 成立, 且 $\{w^k\}$ 有界, 则 $\sum_{k=0}^{+\infty} \|w^{k+1} - w^k\|^2 < +\infty$.

证明 根据对 $\{w^k\}$ 的有界假设知, $\{w^k\}$ 至少存在一个聚点 w^* . 设子列 $w^{k_j} \rightarrow w^*$. 由于 f_i ($i = 1, 2, \dots, m-1$) 均下半连续且 f_m 连续, 故函数 $\mathcal{L}_\beta(\cdot)$ 下半连续, 因此 $\mathcal{L}_\beta(w^*) \leq \liminf_{k_j \rightarrow +\infty} \mathcal{L}_\beta(w^{k_j})$. 这意味着 $\{\mathcal{L}_\beta(w^{k_j})\}$ 有下界. 又由引理 3.4 知 $\{\mathcal{L}_\beta(w^{k_j})\}$ 单调递减, 进而 $\{\mathcal{L}_\beta(w^{k_j})\}$ 收敛. 因此 $\{\mathcal{L}_\beta(w^k)\}$ 单调有收敛子列, 进而其整列收敛, 且 $\mathcal{L}_\beta(w^k) \geq \mathcal{L}_\beta(w^*)$. 另外, 由 (3.4) 有

$$\delta \|x^{k+1} - x^k\|^2 \leq \mathcal{L}_\beta(w^k) - \mathcal{L}_\beta(w^{k+1}), \quad \forall k = 0, 1, 2, \dots$$

上式对 $k = 0, 1, \dots, n$ 求和, 并注意到 $\mathcal{L}_\beta(w^0) < +\infty$, 有

$$\delta \sum_{k=0}^n \|x^{k+1} - x^k\|^2 \leq \mathcal{L}_\beta(w^0) - \mathcal{L}_\beta(w^{n+1}) \leq \mathcal{L}_\beta(w^0) - \mathcal{L}_\beta(w^*) < +\infty.$$

由于 $\delta > 0$, 于是 $\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\|^2 < +\infty$. 从而 $\sum_{k=0}^{+\infty} \|x_i^{k+1} - x_i^k\|^2 < +\infty$ ($i = 1, \dots, m$). 结合 (3.15) 立知 $\sum_{k=0}^{+\infty} \|\lambda^{k+1} - \lambda^k\|^2 < +\infty$. 因此 $\sum_{k=0}^{+\infty} \|w^{k+1} - w^k\|^2 < +\infty$ 获证.

定理 3.6 (全局收敛性) 若假设 3.1 和假设 3.3 成立, 且 $\{w^k\}$ 有界, 则

(i) 序列 $\{w^k\}$ 全体聚点集 $S(w^0)$ 为非空紧集, 且 $d(w^k, S(w^0)) \rightarrow 0$;

(ii) $S(w^0) \subseteq \text{crit } \mathcal{L}_\beta$;

(iii) $\{\mathcal{L}_\beta(w^k)\}$ 整列收敛, 且 $\lim_{k \rightarrow +\infty} \mathcal{L}_\beta(w^k) = \inf_k \mathcal{L}_\beta(w^k) = \mathcal{L}_\beta(w^*)$, $\forall w^* \in S(w^0)$. 从而 $\mathcal{L}_\beta(\cdot)$ 在 $S(w^0)$ 上取有限值且为常数.

证明 (i) 利用 $\{w^k\}$ 有界性及 $S(w^0)$ 的定义直接可证.

(ii) 设 $w^* \in S(w^0)$, 则存在 $\{w^k\}$ 的子列 $\{w^{k_j}\}$ 收敛到 $w^* := (x^*, \lambda^*)$. 由引理 3.5 得 $\lim_{k \rightarrow +\infty} \|w^{k+1} - w^k\| = 0$, 进而 $\lim_{k_j \rightarrow +\infty} w^{k_j+1} = w^*$. 于是, 根据 (1.4d) 知 $\{\lambda^{k_j+\frac{1}{2}}\}$ 收敛, 记其极限为 λ^{**} . 于是, 在 (1.4d) 及 (1.4f) 中令 $k = k_j \rightarrow \infty$ 取极限, 得

$$\lambda^{**} = \lambda^* - r\beta \left(\sum_{i=1}^{m-1} A_i x_i^* + x_m^* - b \right), \quad \lambda^* = \lambda^{**} - s\beta \left(\sum_{i=1}^{m-1} A_i x_i^* + x_m^* - b \right).$$

结合 $(r+s)\beta > 0$ 可知

$$\sum_{i=1}^{m-1} A_i x_i^* + x_m^* - b = 0, \quad \lambda^{**} = \lambda^*. \quad (3.19)$$

因此, x^* 为问题 (1.2) 的可行点. 由 PSRADMM (1.4) 中关于 x_i -子问题 ($i = 1, \dots, m-1$), 有

$$\begin{aligned} & f_i(x_i^{k_j+1}) - \langle \lambda^{k_j}, A_i x_i^{k_j+1} \rangle + \frac{\beta}{2} \left\| \sum_{q=1}^i A_q x_q^{k_j+1} + \sum_{q=i+1}^m A_q x_q^{k_j} - b \right\|^2 + \frac{1}{2} \|x_i^{k_j+1} - x_i^{k_j}\|_{F_i}^2 \\ & \leq f_i(x_i^*) - \langle \lambda^{k_j}, A_i x_i^* \rangle + \frac{\beta}{2} \left\| \sum_{q=1}^{i-1} A_q x_q^{k_j+1} + A_i x_i^* + \sum_{q=i+1}^m A_q x_q^{k_j} - b \right\|^2 + \frac{1}{2} \|x_i^* - x_i^{k_j}\|_{F_i}^2. \end{aligned}$$

结合 $\lim_{k_j \rightarrow +\infty} w^{k_j} = \lim_{k_j \rightarrow +\infty} w^{k_j+1} = w^*$, 可得 $\limsup_{k_j \rightarrow +\infty} f_i(x_i^{k_j+1}) \leq f_i(x_i^*)$. 进而, 结合 $f_i(x_i)$ 的下半连续性有

$$\lim_{k_j \rightarrow +\infty} f_i(x_i^{k_j+1}) = f_i(x_i^*), \quad i = 1, 2, \dots, m-1. \quad (3.20)$$

进一步, 结合 ∂f_i ($i = 1, 2, \dots, m-1$) 的闭性及 ∇f_m 的连续性, 在 (3.2) 中令 $k = k_j \rightarrow \infty$ 取极限, 并结合 (3.19) 可得

$$A_i^\top \lambda^* \in \partial f_i(x_i^*) \quad (i = 1, 2, \dots, m-1), \quad \nabla f_m(x_m^*) = \lambda^*, \quad \sum_{i=1}^m A_i x_i^* - b = 0.$$

因此, 由引理 2.7 知 $w^* \in \text{crit } \mathcal{L}_\beta$, 从而 $S(w^0) \subseteq \text{crit } \mathcal{L}_\beta$ 获证.

(iii) 设 $w^* \in S(w^0)$, 则存在 $\{w^k\}$ 的子列 $\{w^{k_j}\}$ 收敛到 $w^* := (x^*, \lambda^*)$. 结合 (1.3), (3.20) 及 f_m 的连续性, 有

$$\lim_{k_j \rightarrow +\infty} \mathcal{L}_\beta(w^{k_j+1}) = \mathcal{L}_\beta(w^*). \quad (3.21)$$

结合 $\{\mathcal{L}_\beta(w^k)\}$ 的单调下降性可知, $\{\mathcal{L}_\beta(w^k)\}$ 整列收敛. 记 $a = \lim_{k \rightarrow +\infty} \mathcal{L}_\beta(w^k) < +\infty$, 于是由 (3.21) 有

$$\mathcal{L}_\beta(w^*) \equiv a, \quad \forall w^* \in S(w^0).$$

证毕.

引理 3.7 若假设 3.1 成立, $r + s > 0$. 令

$$\begin{cases} \varepsilon_i^{k+1} = A_i^\top (\lambda^k - \lambda^{k+1}) + \beta A_i^\top \left(\sum_{j=i+1}^m A_j (x_j^{k+1} - x_j^k) \right) \\ \quad - F_i(x_i^{k+1} - x_i^k), \quad i = 1, \dots, m-1, \end{cases} \quad (3.22a)$$

$$\varepsilon_m^{k+1} = \frac{s}{r+s} (\lambda^k - \lambda^{k+1}) + \frac{rs\beta}{r+s} (x_m^{k+1} - x_m^k), \quad (3.22b)$$

$$\varepsilon_{m+1}^{k+1} = \frac{1}{(r+s)\beta} (\lambda^{k+1} - \lambda^k) - \frac{r}{r+s} (x_m^{k+1} - x_m^k), \quad (3.22c)$$

则 $\varepsilon^{k+1} := (\varepsilon_1^{k+1}, \dots, \varepsilon_m^{k+1}, \varepsilon_{m+1}^{k+1}) \in \partial \mathcal{L}_\beta(w^{k+1})$, 且存在 $\zeta > 0$, 使得

$$d(0, \partial \mathcal{L}_\beta(w^{k+1})) \leq \zeta \|x^{k+1} - x^k\|. \quad (3.23)$$

证明 首先, 由 (3.2a) 知, 存在 $\xi_i^{k+1} \in \partial f_i(x_i^{k+1})$, 使得

$$\xi_i^{k+1} - A_i^\top \lambda^k + \beta A_i^\top \left(\sum_{j=1}^i A_j x_j^{k+1} + \sum_{j=i+1}^m A_j x_j^k - b \right) + F_i(x_i^{k+1} - x_i^k) = 0, \quad i = 1, \dots, m-1.$$

结合 (3.22a) 和 (2.2a), 有 ($i = 1, \dots, m-1$),

$$\begin{aligned} \varepsilon_i^{k+1} &= A_i^\top (\lambda^k - \lambda^{k+1}) + \beta A_i^\top \left(\sum_{j=i+1}^m A_j (x_j^{k+1} - x_j^k) \right) - F_i(x_i^{k+1} - x_i^k) \\ &\quad + \left\{ \xi_i^{k+1} - A_i^\top \lambda^k + \beta A_i^\top \left(\sum_{j=1}^i A_j x_j^{k+1} + \sum_{j=i+1}^m A_j x_j^k - b \right) + F_i(x_i^{k+1} - x_i^k) \right\} \\ &= \xi_i^{k+1} - A_i^\top \lambda^{k+1} + \beta A_i^\top \left(\sum_{j=1}^m A_j x_j^{k+1} - b \right) \in \partial_{x_i} \mathcal{L}_\beta(w^{k+1}). \end{aligned}$$

另外, 将 (1.4f) 中的 $\lambda^{k+\frac{1}{2}}$ 代入 (3.2b), 得

$$s\beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right) = \nabla f_m(x_m^{k+1}) - \lambda^{k+1} + \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right).$$

结合 (3.22b), (3.12) 及 (2.2b), 有

$$\begin{aligned}\varepsilon_m^{k+1} &= s\beta \left\{ \frac{1}{(r+s)\beta} (\lambda^k - \lambda^{k+1}) + \frac{r}{r+s} (x_m^{k+1} - x_m^k) \right\} = s\beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right) \\ &= \nabla f_m(x_m^{k+1}) - \lambda^{k+1} + \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right) = \partial_{x_m} \mathcal{L}_\beta(w^{k+1}).\end{aligned}$$

由 (3.22c), (3.12) 及 (2.2c), 有

$$\varepsilon_{m+1}^{k+1} = \frac{1}{(r+s)\beta} (\lambda^{k+1} - \lambda^k) - \frac{r}{r+s} (x_m^{k+1} - x_m^k) = - \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right) = \partial_\lambda \mathcal{L}_\beta(w^{k+1}).$$

至此, $\varepsilon^{k+1} \in \partial \mathcal{L}_\beta(w^{k+1})$ 获证.

最后, 由定义 (3.22) 知, 存在 $\zeta_1 > 0$, 使得

$$\|\varepsilon^{k+1}\| \leq \zeta_1 \sum_{i=1}^m \|x_i^{k+1} - x_i^k\| + \zeta_1 \|\lambda^{k+1} - \lambda^k\|.$$

另外, 由 (3.14) 知, 存在 $\zeta_2 > 0$, 使得

$$\|\lambda^{k+1} - \lambda^k\| \leq \zeta_2 \sum_{i=1}^m \|x_i^{k+1} - x_i^k\|.$$

上述两式结合 $\varepsilon^{k+1} \in \partial \mathcal{L}_\beta(w^{k+1})$, 有

$$d(0, \partial \mathcal{L}_\beta(w^{k+1})) \leq \|\varepsilon^{k+1}\| \leq \zeta_1(1 + \zeta_2) \sum_{i=1}^m \|x_i^{k+1} - x_i^k\| \leq \sqrt{m} \zeta_1(1 + \zeta_2) \|x^{k+1} - x^k\|.$$

引理获证.

定理 3.8 (强收敛性) 若假设 3.1 及假设 3.3 成立, 序列 $\{w^k\}$ 有界, 且 $\mathcal{L}_\beta(\cdot)$ 是 KL 函数, 则 $\sum_{k=0}^{+\infty} \|w^{k+1} - w^k\| < +\infty$, 且 $\{w^k\}$ 整列收敛到 $\mathcal{L}_\beta(\cdot)$ 的稳定点.

证明 根据定理 3.6, 对于任意 $w^* \in S(w^0)$, 有 $\mathcal{L}_\beta(w^k) \rightarrow \mathcal{L}_\beta(w^*) = \inf \mathcal{L}_\beta(w^k)$. 下面分两种情况完成证明.

(i) 设存在正整数 k_0 , 使得 $\mathcal{L}_\beta(w^{k_0}) = \mathcal{L}_\beta(w^*)$. 此时由 (3.4) 及 $\{\mathcal{L}_\beta(w^k)\}$ 的单调下降性, 可得

$$\delta \|x^{k+1} - x^k\|^2 \leq \mathcal{L}_\beta(w^k) - \mathcal{L}_\beta(w^{k+1}) \leq \mathcal{L}_\beta(w^{k_0}) - \mathcal{L}_\beta(w^*) = 0, \quad \forall k \geq k_0.$$

于是 $x^{k+1} = x^k$, $\forall k \geq k_0$. 结合 (3.14) 知 $\lambda^{k+1} = \lambda^k$, $\forall k \geq k_0$. 于是有 $w^{k+1} = w^{k_0}$, $\forall k \geq k_0$, $w^* = w^{k_0}$. 结论成立.

(ii) 假设 $\mathcal{L}_\beta(w^k) > \mathcal{L}_\beta(w^*)$, $\forall k$. 首先, 由定理 3.6(iii) 知, 正常下半连续函数 $\mathcal{L}_\beta(\cdot)$ 在非空紧集 $S(w^0)$ 上为常数. 进而由引理 2.5 知, $\mathcal{L}_\beta(\cdot)$ 满足一致 KL 性质. 对于引理 2.5 一致 KL 性质的参数 ϵ 和 η , 由于 $d(w^k, S(w^0)) \rightarrow 0$ 及 $\mathcal{L}_\beta(w^*) = \inf \mathcal{L}_\beta(w^k)$, 存在正整数 \tilde{k} , 使得

$$d(w^k, S(w^0)) < \epsilon, \quad \mathcal{L}_\beta(w^*) < \mathcal{L}_\beta(w^k) < \mathcal{L}_\beta(w^*) + \eta, \quad \forall k > \tilde{k}.$$

因此, 由引理 2.5 得

$$\varphi'(\mathcal{L}_\beta(w^k) - \mathcal{L}_\beta(w^*)) d(0, \partial \mathcal{L}_\beta(w^k)) \geq 1, \quad k > \tilde{k}.$$

其次, 由 KL 性质关联函数 φ 的凹性及 $\{\mathcal{L}_\beta(w^k)\}$ 的单调下降性, 有

$$\begin{aligned} & \varphi(\mathcal{L}_\beta(w^k) - \mathcal{L}_\beta(w^*)) - \varphi(\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^*)) \\ & \geq \varphi'(\mathcal{L}_\beta(w^k) - \mathcal{L}_\beta(w^*))(\mathcal{L}_\beta(w^k) - \mathcal{L}_\beta(w^{k+1})) \geq 0. \end{aligned}$$

结合 $\varphi'(\mathcal{L}_\beta(w^k) - \mathcal{L}_\beta(w^*)) > 0$ 及 $d(0, \partial \mathcal{L}_\beta(w^k)) \leq \zeta \|x^k - x^{k-1}\|$, 得

$$\begin{aligned} \mathcal{L}_\beta(w^k) - \mathcal{L}_\beta(w^{k+1}) & \leq \frac{\varphi(\mathcal{L}_\beta(w^k) - \mathcal{L}_\beta(w^*)) - \varphi(\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^*))}{\varphi'(\mathcal{L}_\beta(w^k) - \mathcal{L}_\beta(w^*))} \\ & \leq \zeta \|x^k - x^{k-1}\| [\varphi(\mathcal{L}_\beta(w^k) - \mathcal{L}_\beta(w^*)) - \varphi(\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^*))]. \end{aligned}$$

为简便, 记 $\Delta_{p,q} = \varphi(\mathcal{L}_\beta(w^p) - \mathcal{L}_\beta(w^*)) - \varphi(\mathcal{L}_\beta(w^q) - \mathcal{L}_\beta(w^*))$. 结合引理 3.2 和上述不等式, 有

$$\delta \|x^{k+1} - x^k\|^2 \leq \zeta \|x^k - x^{k-1}\| \Delta_{k,k+1}, \quad \forall k > \tilde{k}.$$

从而

$$\|x^{k+1} - x^k\| \leq \sqrt{\|x^k - x^{k-1}\| \left(\frac{\zeta \Delta_{k,k+1}}{\delta} \right)}, \quad \forall k > \tilde{k}.$$

由上式及不等式 $2\sqrt{ab} \leq a + b$ ($\forall a, b > 0$), 得

$$2\|x^{i+1} - x^i\| \leq \|x^i - x^{i-1}\| + \frac{\zeta}{\delta} \Delta_{i,i+1}, \quad \forall i > \tilde{k}. \quad (3.24)$$

将不等式 (3.24) 从 $i = k + 1 (\geq \tilde{k} + 1)$ 到 $i = q$ 进行求和, 得

$$2 \sum_{i=k+1}^q \|x^{i+1} - x^i\| \leq \sum_{i=k+1}^q \|x^i - x^{i-1}\| + \frac{\zeta}{\delta} \Delta_{k+1,q+1}.$$

由此及 $\varphi(\mathcal{L}_\beta(w^{q+1}) - \mathcal{L}_\beta(w^*)) \geq 0$, 进一步有

$$\sum_{i=k+1}^q \|x^{i+1} - x^i\| + \|x^{q+1} - x^q\| \leq \|x^{k+1} - x^k\| + \frac{\zeta}{\delta} \varphi(\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^*)).$$

在上式令 $q \rightarrow +\infty$ 取极限, 并注意到 $\|x^{q+1} - x^q\| \rightarrow 0$ (引理 3.5), 有

$$\sum_{i=k+1}^{+\infty} \|x^{i+1} - x^i\| \leq \|x^{k+1} - x^k\| + \frac{\zeta}{\delta} \varphi(\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^*)), \quad \forall k \geq \tilde{k}. \quad (3.25)$$

特别地, 有

$$\sum_{i=\tilde{k}+1}^{+\infty} \|x^{i+1} - x^i\| \leq \|x^{\tilde{k}+1} - x^{\tilde{k}}\| + \frac{\zeta}{\delta} \varphi(\mathcal{L}_\beta(w^{\tilde{k}+1}) - \mathcal{L}_\beta(w^*)) < +\infty.$$

故 $\sum_{i=0}^{+\infty} \|x^{i+1} - x^i\| < +\infty$. 结合 (3.14) 知 $\sum_{i=0}^{+\infty} \|\lambda^{i+1} - \lambda^i\| < +\infty$. 最后, 由 $w^i = (x^i, \lambda^i)$ 知

$$\sum_{i=0}^{+\infty} \|w^{i+1} - w^i\| \leq \sum_{i=0}^{+\infty} \|x^{i+1} - x^i\| + \sum_{i=0}^{+\infty} \|\lambda^{i+1} - \lambda^i\| < +\infty.$$

从而 $\{w^k\}$ 为 Cauchy 列, 进而收敛. 再由定理 3.6 (ii) 知 $\{w^k\}$ 收敛到 $\mathcal{L}_\beta(\cdot)$ 的稳定点. 证毕.

以上的全局收敛性和强收敛性等结论均需 $\{w^k\}$ 有界, 这一假设在非凸分裂算法收敛性分析中普遍被采用, 如文 [7, 8, 16, 24, 26, 31, 32, 34]. 下面给出一个充分条件保证 $\{w^k\}$ 有界.

引理 3.9 若假设 3.1, 假设 3.3 及下述条件 (i)–(iii) 成立, 则由 PSRADMM 产生的点列 $\{w^k\}$ 有界.

- (i) f_i ($i = 1, 2, \dots, m-1$) 均为强制函数, 即 $\liminf_{\|x_i\| \rightarrow +\infty} f_i(x_i) = +\infty$;
- (ii) 松弛因子 $s \in ((m+1 - \sqrt{m+1})/m, (m+1 + \sqrt{m+1})/m)$;
- (iii) 函数 $\bar{f}_m(x_m) := f_m(x_m) - \frac{\|\nabla f_m(x_m)\|^2}{2(2s-1)\beta}$ 有下界且是强制的, 即

$$\inf \bar{f}_m(x_m) > -\infty, \quad \liminf_{\|x_m\| \rightarrow +\infty} \bar{f}_m(x_m) = +\infty.$$

证明 由 $\{\mathcal{L}_\beta(w^k)\}$ 单调递减知 $\mathcal{L}_\beta(w^k) \leq \mathcal{L}_\beta(w^0) < +\infty$. 结合 (3.13), 有

$$\begin{aligned} \mathcal{L}_\beta(w^0) &\geq \mathcal{L}_\beta(w^k) \\ &= \sum_{i=1}^m f_i(x_i^k) - \left\langle \lambda^k, \sum_{i=1}^m A_i x_i^k - b \right\rangle + \frac{\beta}{2} \left\| \sum_{i=1}^m A_i x_i^k - b \right\|^2 \\ &\stackrel{(3.13)}{=} \sum_{i=1}^m f_i(x_i^k) - \left\langle \nabla f_m(x_m^k), \sum_{i=1}^m A_i x_i^k - b \right\rangle + \frac{(2s-1)\beta}{2} \left\| \sum_{i=1}^m A_i x_i^k - b \right\|^2 \\ &= \sum_{i=1}^m f_i(x_i^k) - \frac{\|\nabla f_m(x_m^k)\|^2}{2(2s-1)\beta} + \frac{(2s-1)\beta}{2} \left\| \sum_{i=1}^m A_i x_i^k - b - \frac{\nabla f_m(x_m^k)}{(2s-1)\beta} \right\|^2 \\ &= \sum_{i=1}^{m-1} f_i(x_i^k) + \bar{f}_m(x_m^k) + \frac{(2s-1)\beta}{2} \left\| \sum_{i=1}^m A_i x_i^k - b - \frac{\nabla f_m(x_m^k)}{(2s-1)\beta} \right\|^2. \end{aligned}$$

另外, 由 f_i 为正常下半连续的强制函数, 易知 $\inf_{x_i} f_i(x_i) > -\infty$ ($i = 1, \dots, m-1$). 因此有

$$\begin{cases} \sum_{i=1}^{m-1} f_i(x_i^k) + \bar{f}_m(x_m^k) + \frac{(2s-1)\beta}{2} \left\| \sum_{i=1}^m A_i x_i^k - b - \frac{\nabla f_m(x_m^k)}{(2s-1)\beta} \right\|^2 \leq \mathcal{L}_\beta(w^0) < +\infty, \\ \inf_{x_i} f_i(x_i) > -\infty \quad (i = 1, \dots, m-1), \quad \inf_{x_m} \bar{f}_m(x_m) > -\infty, \\ s \in ((m+1 - \sqrt{m+1})/m, (m+1 + \sqrt{m+1})/m) \quad \left(\Rightarrow s > \frac{1}{2} \right). \end{cases}$$

由此不难依次推知, $\{x_i^k\}$ ($i = 1, \dots, m$) 及 $\{\sum_{i=1}^m A_i x_i^k - b - \frac{\nabla f_m(x_m^k)}{(2s-1)\beta}\}$ 均有界, 从而 $\{\nabla f_m(x_m^k)\}$ 有界, 进一步结合 (3.13) 知 $\{\lambda^k\}$ 也有界. 因此, $\{w^k\}$ 的有界性获证. 证毕.

注 3.10 现在我们分析一些具体问题以验证引理 3.9 中的条件. 例如, 我们考虑 $l_{\frac{1}{2}}$ 正则化问题^[35] $\min\{c\|x_1\|_{\frac{1}{2}}^{\frac{1}{2}} + \frac{1}{2}\|x_2\|^2 \mid Dx_1 - x_2 - b = 0\}$, 其中 c 是正则化参数, $\|x_1\|_{\frac{1}{2}} = (\sum_{j=1}^n |x_1^j|)^{\frac{1}{2}}$. 令 $f_1(x_1) = c\|x_1\|_{\frac{1}{2}}^{\frac{1}{2}}$, $f_2(x_2) = \frac{1}{2}\|x_2\|^2$. 当 $s \in ((3 - \sqrt{3})/2, (3 + \sqrt{3})/2)$ 时, 存在充分大的 $\beta > \min\{\beta_0, 1/(2s-1)\}$, 有

$$\bar{f}_2(x_2) := f_2(x_2) - \frac{\|\nabla f_2(x_2)\|^2}{2(2s-1)\beta} = \frac{1}{2}\|x_2\|^2 - \frac{1}{2(2s-1)\beta}\|x_2\|^2 = \frac{(2s-1)\beta-1}{2(2s-1)\beta}\|x_2\|^2 > -\infty.$$

结合 $\liminf_{\|x_1\| \rightarrow +\infty} f_1(x_1) = +\infty$, $\liminf_{\|x_2\| \rightarrow +\infty} \bar{f}_2(x_2) = +\infty$, 从而引理 3.9 中的条件成立.

4 收敛率分析

本节将在强收敛前提下, 进一步讨论 PSRADMM 的收敛率. 其主要结果如下.

定理 4.1 假设 $\{w^k\}$ 是由 PSRADMM 迭代产生的序列, 且其收敛到 $\{w^*\}$. 设 $\mathcal{L}_\beta(\cdot)$ 在 w^* 处满足 KL 性质, 且关联函数 $\varphi(t) = ct^{1-\theta}$, $\theta \in [0, 1)$, $c > 0$, 则下述结论成立:

- (i) 若 $\theta = 0$, 则序列 $\{w^k\}$ 在有限次迭代后收敛, 即存在 k , 使得 $w^k = w^*$;
- (ii) 若 $\theta \in (0, \frac{1}{2}]$, 则存在 $\tau \in [0, 1)$, 使得 $\|w^k - w^*\| = O(\tau^k)$, 即 PSRADMM 是线性收敛的;
- (iii) 若 $\theta \in (\frac{1}{2}, 1)$, 则 $\|w^k - w^*\| = O(k^{\frac{1-\theta}{1-2\theta}})$, 即 PSRADMM 是次线性收敛的.

证明 对于 $\theta = 0$ 的情形, 有 $\varphi(t) = ct$, $\varphi'(t) = c$. 若 $w^k \neq w^*$, $\forall k$, 则对于充分大的 k , 由 KL 性质知, 并且 $cd(0, \partial\mathcal{L}_\beta(w^k)) \geq 1$, 这与 (3.23) 矛盾.

下面假设 $\theta > 0$, 并记 $\Delta_k := \sum_{i=k}^{+\infty} \|x^{i+1} - x^i\|$, 则由 (3.25) 有

$$\Delta_{k+1} \leq (\Delta_k - \Delta_{k+1}) + \frac{\zeta}{\delta} \varphi(\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^*)), \quad \forall k \geq \tilde{k}. \quad (4.1)$$

注意到 $\mathcal{L}_\beta(\cdot)$ 在 w^* 处满足 KL 性质, 有

$$\varphi'(\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^*))d(0, \partial\mathcal{L}_\beta(w^{k+1})) \geq 1, \quad \forall k \geq \tilde{k}.$$

又因 $\varphi(t) = ct^{1-\theta}$, 上述不等式等价于

$$(\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^*))^\theta \leq c(1-\theta)d(0, \partial\mathcal{L}_\beta(w^{k+1})), \quad \forall k \geq \tilde{k}. \quad (4.2)$$

又由 (3.23) 有

$$d(0, \partial\mathcal{L}_\beta(w^{k+1})) \leq \zeta\|x^{k+1} - x^k\| = \zeta(\Delta_k - \Delta_{k+1}). \quad (4.3)$$

结合 (4.2) 和 (4.3), 存在 $\gamma > 0$, 使得

$$\varphi(\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^*)) = c(\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^*))^{1-\theta} \leq \gamma(\Delta_k - \Delta_{k+1})^{\frac{1-\theta}{\theta}}, \quad \forall k \geq \tilde{k}.$$

结合 (4.1) 得

$$\Delta_{k+1} \leq (\Delta_k - \Delta_{k+1}) + \frac{\zeta}{\delta} \gamma (\Delta_k - \Delta_{k+1})^{\frac{1-\theta}{\theta}}, \quad \forall k \geq \tilde{k}. \quad (4.4)$$

下面基于不等式 (4.4), 借助 Attouch 和 Bolte^[1] 的有关结论, 进一步完成证明.

- 若 $\theta \in (0, \frac{1}{2}]$, 则由文 [1] 知, 存在 $c_1 > 0$ 及 $\tau \in [0, 1)$, 使得

$$\|x^k - x^*\| \leq c_1 \tau^k. \quad \text{从而 } \|x_i^k - x_i^*\| = O(\tau^k), \quad i = 1, \dots, m. \quad (4.5)$$

- 若 $\theta \in (\frac{1}{2}, 1)$, 则由文 [1] 知, 存在 $c_2 > 0$, 使得

$$\|x^k - x^*\| \leq c_2 k^{\frac{1-\theta}{1-2\theta}}. \quad \text{从而 } \|x_i^k - x_i^*\| = O(k^{\frac{1-\theta}{1-2\theta}}), \quad i = 1, \dots, m. \quad (4.6)$$

另外, 由 (3.13) 和 (2.3) 及 ∇f_m 的 L_m -Lipschitz 连续性, 可知

$$\begin{aligned} \|\lambda^k - \lambda^*\| &= \left\| (\nabla f_m(x_m^k) - \nabla f_m(x_m^*)) + \beta(1-s) \left(\sum_{i=1}^m A_i(x_i^k - x_i^*) \right) \right\| \\ &= O\left(\sum_{i=1}^m \|x_i^k - x_i^*\| \right). \end{aligned} \quad (4.7)$$

最后, 结合 (4.5)–(4.7), 立知结论 (ii) 和 (iii) 成立. 证毕.

5 数值试验

为验证本文所提算法 PSRADMM 的有效性, 我们考虑将其应用于求解压缩感知和稳健主成分分析的非凸优化问题, 并分别与 RADMM^[24] 和 BADMM^[31] 进行数值比较. 测试程序均在

MATLAB R2014a 上运行, 采用 Windows 10 操作系统, DELL 笔记本电脑 (Inter(R) Core(TM) i5-5200U CPU 2.20 GHz), 4 GB 内存.

5.1 压缩感知

压缩感知中的数学模型为:

$$\min \|x_1\|_0, \quad \text{s.t. } Dx_1 = b, \quad (5.1)$$

其中 $D \in \mathbb{R}^{m \times n}$ 为测量矩阵, $b \in \mathbb{R}^m$ 为观测数据, $\|x_1\|_0$ 表示 x_1 中非零元素的个数, 也称 l_0 范数. 众所周知, 问题 (5.1) 是 NP 难问题. 实际应用中, 常将问题 (5.1) 中 l_0 范数松弛到 $l_{\frac{1}{2}}$ 范数, 从而导出如下非凸优化问题 [24, 35]:

$$\min c\|x_1\|_{\frac{1}{2}} + \frac{1}{2}\|x_2\|^2, \quad \text{s.t. } Dx_1 - x_2 = b, \quad (5.2)$$

其中 c 是正则化参数. 问题 (5.2) 的增广拉格朗日函数为

$$\mathcal{L}_\beta(x_1, x_2, \lambda) = c\|x_1\|_{\frac{1}{2}} + \frac{1}{2}\|x_2\|^2 - \langle \lambda, Dx_1 - x_2 - b \rangle + \frac{\beta}{2}\|Dx_1 - x_2 - b\|^2.$$

将 PSRADMM (1.2) 的两分块情形应用到求解问题 (5.2) 中, 有

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1} \left\{ \mathcal{L}_\beta(x_1, x_2^k, \lambda^k) + \frac{1}{2}\|x_1 - x_1^k\|_{F_1}^2 \right\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Dx_1^{k+1} - x_2^k - b), \\ x_2^{k+1} = \arg \min_{x_2} \{ \mathcal{L}_\beta(x_1^{k+1}, x_2, \lambda^{k+\frac{1}{2}}) \}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(Dx_1^{k+1} - x_2^{k+1} - b). \end{cases} \quad (5.3)$$

简化上述过程 (5.3), 取 $F_1 = \mu_1 I - \beta D^\top D$, 可得到如下形式的迭代步骤:

$$\begin{cases} x_1^{k+1} = \mathcal{H} \left(x_1^k - \frac{\beta}{\mu_1} D^\top Dx_1^k + \frac{1}{\mu_1} D^\top (\beta x_2^k + \lambda^k + \beta b); \frac{2c}{\mu_1} \right), \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Dx_1^{k+1} - x_2^k - b), \\ x_2^{k+1} = \frac{1}{1+\beta} (\beta Dx_1^{k+1} - \lambda^{k+\frac{1}{2}} - \beta b), \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(Dx_1^{k+1} - x_2^{k+1} - b), \end{cases} \quad (5.4)$$

其中 $\mathcal{H}(\cdot; \mu)$ 为半收缩算子 [35].

此外, 将 RADMM [24] 应用于求解问题 (5.2), 令 $G = \mu_2 I - \beta D^\top D$, 类似于 (5.3) 的处理, 其迭代形式如下:

$$\begin{cases} x_1^{k+1} = \mathcal{H} \left(x_1^k - \frac{\beta}{\mu_2} D^\top Dx_1^k + \frac{1}{\mu_2} D^\top (\beta x_2^k + \lambda^k + \beta b); \frac{2c}{\mu_2} \right), \\ x_2^{k+1} = \frac{1}{1+\beta} (\beta Dx_1^{k+1} - \lambda^k - \beta b), \\ \lambda^{k+1} = \lambda^k - \beta(Dx_1^{k+1} - x_2^{k+1} - b). \end{cases} \quad (5.5)$$

在数值试验中, 我们选择 $D_{ij} \sim \mathcal{N}(0, 1)$, 标准化列使其具有单位 $l_{\frac{1}{2}}$ 范数. 观测数据 $b = Dx_1^* + v$, 其中 x_1^* 中包含 100 个正态分布的非零分量, $v \sim \mathcal{N}(0, 10^{-3}I)$. 正则化参数为 $c = 0.1\|D^\top b\|_\infty$. 初始值 x_1^0, x_2^0, λ^0 均取零向量. 此外, PSRADMM 与 RADMM 中相同的参数选取一致, 即 $\mu_1 = \mu_2 = 20, \beta = 12$. PSRADMM 中松弛因子 $r = s = 0.85$. 当前后两迭代点的变化满足

$$e^k := \frac{\max\{\|x^k - x^{k-1}\|, \|y^k - y^{k-1}\|, \|\lambda^k - \lambda^{k-1}\|\}}{\max\{\|x^{k-1}\|, \|y^{k-1}\|, \|\lambda^{k-1}\|, 1\}} \leq 10^{-4}$$

时, 终止程序.

我们分别用 PSRADMM (5.4) 与 RADMM (5.5) 求解问题 (5.2), 并选取了 5 个不同维数的 D 进行测试. 表 1 中详细报告了 PSRADMM 与 RADMM 的主要数值测试结果, 其中 iter, tcpu(s), obj-fval 分别表示迭代次数, 以秒为单位的 CPU 计算时间和目标函数值.

D 的维数		PSRADMM			RADMM		
m	n	iter	tcpu(s)	obj-fval	iter	tcpu(s)	obj-fval
200	280	399	1.51	15.46	588	2.22	15.41
300	400	203	1.72	16.11	439	3.54	15.80
400	480	197	3.07	15.45	249	3.75	15.36
500	600	186	5.51	14.37	299	8.28	14.29
600	680	204	7.19	17.50	287	8.81	17.47
总计		1189	19.00	—	1862	26.60	—

表 1 PSRADMM 与 RADMM 的数值试验结果

此外, 为了更直观地展示测试比较结果, 我们绘制了 iter 及 tcpu 关于 D 维数的变化趋势. 从表 1 及图 1 可见, PSRADMM 在解决问题 (5.2) 时明显优于 RADMM.

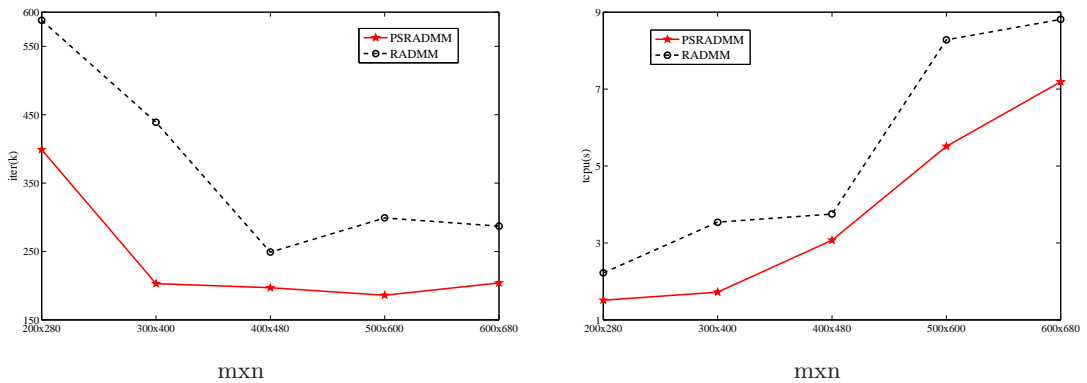


图 1 iter (左) 及 tcpu (右) 关于 D 维数的变化趋势

5.2 稳健主成分分析

稳健主成分分析是经典的数据分析与数据降维技术之一, 在工程技术领域有较为广泛的应用. 本文从矩阵分解应用考虑如下三分块稳健主成分分析非凸优化模型 [5, 31]:

$$\min_{L, S, T} \|L\|_* + \gamma \|S\|_1 + \frac{\omega}{2} \|T - M\|_F^2 \quad \text{s.t. } L + S = T. \quad (5.6)$$

在模型 (5.6) 中, $M \in \mathbb{R}^{p \times n}$ 是给定的观测矩阵, $L, S, T \in \mathbb{R}^{p \times n}$ 为决策矩阵 (变量), 低秩项 $\|L\|_* := \sum_{i=1}^{\min(p, n)} |\sigma_i(L)|^{1/2}$ ($\sigma_i(L)$ 为低秩矩阵 L 的奇异值), 稀疏项 $\|S\|_1 := \sum_{i=1}^p \sum_{j=1}^n |S_{ij}|$, γ 为 $\|L\|_*$ 与 $\|S\|_1$ 之间的协调参数, ω 是关于噪声水平的罚参数, $\|\cdot\|_F$ 表示矩阵的 Frobenius 范数, 即

$$\|A = (a_{ij})\|_F = \sqrt{\sum_{i=1}^p \sum_{j=1}^n |a_{ij}|^2}.$$

问题 (5.6) 显然是本文讨论模型 (1.2) 的三分块情形. 将向量空间 \mathbb{R}^n 中的 l_2 范数和内积运算分别拓广为矩阵空间 $\mathbb{R}^{p \times n}$ 中的 Frobenius 范数和

$$\langle A, B \rangle = \sum_{i=1}^p \sum_{j=1}^n a_{ij} b_{ij}.$$

问题 (5.6) 的增广拉格朗日函数可表述为

$$\mathcal{L}_\beta(L, S, T, \lambda) = \|L\|_* + \gamma \|S\|_1 + \frac{\omega}{2} \|T - M\|_F^2 - \langle \lambda, L + S - T \rangle + \frac{\beta}{2} \|L + S - T\|_F^2. \quad (5.7)$$

取 $F_1 = F_2 = \nu I$ (单位阵), 则 PSRADMM (1.4) 求解 (5.6) 的迭代格式为:

$$\begin{cases} L^{k+1} = \arg \min_L \left\{ \mathcal{L}_\beta(L, S^k, T^k, \lambda^k) + \frac{\nu}{2} \|L - L^k\|_F^2 \right\}, \\ S^{k+1} = \arg \min_S \left\{ \mathcal{L}_\beta(L^{k+1}, S, T^k, \lambda^k) + \frac{\nu}{2} \|S - S^k\|_F^2 \right\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(L^{k+1} + S^{k+1} - T^k), \\ T^{k+1} = \arg \min_T \{ \mathcal{L}_\beta(L^{k+1}, S^{k+1}, T, \lambda^{k+\frac{1}{2}}) \}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(L^{k+1} + S^{k+1} - T^{k+1}). \end{cases} \quad (5.8)$$

参照文献 [11, 35], 并注意 (5.8) 中 T -子问题的光滑二次特征, 迭代 (5.8) 可进一步简化为:

$$\begin{cases} L^{k+1} = \arg \min_L \left\{ \|L\|_* - \langle \lambda^k, L \rangle + \frac{\beta}{2} \|L + S^k - T^k\|_F^2 + \frac{\nu}{2} \|L - L^k\|_F^2 \right\} \\ \quad = \mathcal{H} \left(\frac{\beta(T^k - S^k) + \lambda^k + \nu L^k}{\beta + \nu}, \frac{1}{\beta + \nu} \right), \\ S^{k+1} = \arg \min_S \left\{ \gamma \|S\|_1 - \langle \lambda^k, S \rangle + \frac{\beta}{2} \|L^{k+1} + S - T^k\|_F^2 + \frac{\nu}{2} \|S - S^k\|_F^2 \right\} \\ \quad = \mathcal{S} \left(\frac{\beta(T^k - L^{k+1}) + \lambda^k + \nu S^k}{\beta + \nu}, \frac{\gamma}{\beta + \nu} \right), \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(L^{k+1} + S^{k+1} - T^k), \\ T^{k+1} = \arg \min_T \left\{ \frac{\omega}{2} \|T - M\|_F^2 + \langle \lambda^{k+\frac{1}{2}}, T \rangle + \frac{\beta}{2} \|L^{k+1} + S^{k+1} - T\|_F^2 \right\} \\ \quad = \frac{\beta(L^{k+1} + S^{k+1}) + \omega M - \lambda^{k+\frac{1}{2}}}{\omega + \beta}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(L^{k+1} + S^{k+1} - T^{k+1}), \end{cases} \quad (5.9)$$

其中, $\mathcal{H}(\cdot, \frac{1}{\beta+\nu})$ 为半收缩算子 [35], $\mathcal{S}(\cdot, \frac{\gamma}{\beta+\nu})$ 为软收缩算子 [11].

另外, 将 BADMM [31] 应用于求解问题 (5.6), 按文献 [31], 取如下迭代格式:

$$\begin{cases} L^{k+1} = \arg \min_L \left\{ \mathcal{L}_\beta(L, S^k, T^k, \lambda^k) + \frac{\rho}{2} \|L - L^k\|_F^2 \right\}, \\ S^{k+1} = \arg \min_S \left\{ \mathcal{L}_\beta(L^{k+1}, S, T^k, \lambda^k) + \frac{\rho}{2} \|S - S^k\|_F^2 \right\}, \\ T^{k+1} = \arg \min_T \left\{ \mathcal{L}_\beta(L^{k+1}, S^{k+1}, T, \lambda^k) + \frac{\rho}{2} \|T - T^k\|_F^2 \right\}, \\ \lambda^{k+1} = \lambda^k - \beta(L^{k+1} + S^{k+1} - T^{k+1}). \end{cases}$$

类似于迭代 (5.8) 的处理, 上面 BADMM 迭代可进一步简化为:

$$\begin{cases} L^{k+1} = \mathcal{H}\left(\frac{\beta(T^k - S^k) + \lambda^k + \rho L^k}{\beta + \rho}, \frac{1}{\beta + \rho}\right), \\ S^{k+1} = \mathcal{S}\left(\frac{\beta(T^k - L^{k+1}) + \lambda^k + \rho S^k}{\beta + \rho}, \frac{\gamma}{\beta + \rho}\right), \\ T^{k+1} = \frac{\beta(L^{k+1} + S^{k+1}) + \omega M - \lambda^k + \rho T^k}{\omega + \beta + \rho}, \\ \lambda^{k+1} = \lambda^k - \beta(L^{k+1} + S^{k+1} - T^{k+1}). \end{cases} \quad (5.10)$$

在数值试验中, 取观测矩阵

$$M = L^* + S^* + N,$$

其中 L^* 与 S^* 分别是通过问题 (5.6) 想要恢复的原始低秩矩阵和稀疏矩阵, N 是高斯噪声矩阵. 下述中, rank 表示低秩矩阵 L 的秩, spr 表示稀疏矩阵 S 的稀疏率. Matlab 程序中产生矩阵 M 的脚本文件如下:

- $L = \text{randn}(p, \text{rank}) * \text{randn}(\text{rank}, n);$
- $S = \text{zeros}(p, n); q = \text{randperm}(p * n); K = \text{round}(\text{spr} * p * n); S(q(1 : K)) = \text{randn}(K, 1);$
- $\varrho = 0;$ % 无高斯噪声情况; $\varrho = 0.01;$ % 带高斯噪声情况; $N = \text{randn}(p, n) * \varrho;$
- $T = L + S; M = T + N.$

特别地, 令 $p = n = 100$, 其分解问题由易到难分别考虑:

$$(\text{rank}, \text{spr}) = (1, 0.05), (5, 0.05), (10, 0.05), (20, 0.05), (1, 0.1), (5, 0.1), (10, 0.1), (20, 0.1).$$

在测试问题 (5.6) 中, 选取参数 $\gamma = \frac{0.1}{\sqrt{p}}$, $\omega = 10^3$. 根据经验和测试观测, 在 (5.8) 和 (5.9) 中, 设置参数 $(r, s) = (-0.1, 1.05) \in D_2$ (见 (3.18) 式), $\beta = 3.11$, $\nu = 0.07$. 而在 BADMM 迭代 (5.10) 中, 参数选取与原文献 [31] 一致, 即 $\beta = \rho = 0.3$. 测试中, 矩阵 L, S, T 的初始值均取为零矩阵. 如前后两迭代点之间的相对变化率

$$\text{RelChg} := \frac{\|(L^{k+1}, S^{k+1}, T^{k+1}) - (L^k, S^k, T^k)\|_F}{\|(L^k, S^k, T^k)\|_F + 1} \leq 10^{-7},$$

终止程序. 记 $\hat{L}, \hat{S}, \hat{T}$ 为相应算法求解测试问题 (5.6) 获得的数值解, 采用相对误差 RelErr 衡量算法的恢复效果, 其定义为

$$\text{RelErr} := \frac{\|(\hat{L}, \hat{S}, \hat{T}) - (L^*, S^*, T^*)\|_F}{\|(L^*, S^*, T^*)\|_F + 1}.$$

分别用 PSRADMM (5.9) 和 BADMM (5.10) 对模型 (5.6) 无高斯噪声和带高斯噪声情形进行数值测试, 主要结果报告于表 2, 其中 Iter, Time 分别表示迭代次数和 CPU 计算时间 (秒). 由表 2 可见, 当 S 的稀疏率 spr 增大时, 大部分对相误差 RelErr 略有增加, 但增幅不大; 但当引入高斯噪声 $\varrho = 0.01$ 时, 相对误差 RelErr 呈三个数量级 (10^3) 迅速增加. 这表明当分解问题由易到难时, 算法恢复性能下降. 从总迭代次数和总时间这两个重要计算指标看, 本文 PSRADMM 远优于 BADMM, 节省近 50%. 从相对误差看, 其亦不逊色于 BADMM.

Data		PSRADMM				BADMM			
(rank, spr)		Iter	Time	RelChg	RelErr	Iter	Time	RelChg	RelErr
无高斯 噪声情况 $\varrho = 0$	(1, 0.05)	78	0.55	8.998915e-08	4.591931e-06	160	0.96	5.655816e-08	4.624883e-06
	(1, 0.1)	105	0.65	9.628669e-08	5.071259e-06	181	0.98	7.720516e-08	4.895803e-06
	(5, 0.05)	127	0.72	9.904375e-08	1.919833e-06	282	1.53	5.494119e-08	1.945444e-06
	(5, 0.1)	190	1.08	8.071810e-08	2.197917e-06	302	1.63	8.935297e-08	2.191404e-06
	(10, 0.05)	207	1.18	7.216220e-08	1.465440e-06	377	1.97	6.958812e-08	1.483898e-06
	(10, 0.1)	237	1.35	8.623880e-08	1.734421e-06	411	2.16	9.603930e-08	1.741659e-06
	(20, 0.05)	314	1.74	7.177763e-08	1.639117e-06	549	2.89	9.949435e-08	2.142006e-06
	(20, 0.1)	473	2.60	9.468629e-08	1.578699e-06	680	3.55	9.736016e-08	2.544968e-06
总计、 误差均值		-	1731	9.87	-	2542	15.67	-	2.696258e-06
带高斯 噪声情况 $\varrho = 0.01$	(1, 0.05)	1523	8.04	9.888607e-08	9.837687e-03	3085	15.98	9.987357e-08	9.837669e-03
	(1, 0.1)	1557	8.20	9.991197e-08	9.589790e-03	3192	16.13	9.965625e-08	9.589768e-03
	(5, 0.05)	1436	7.40	9.985193e-08	4.496118e-03	2753	13.89	9.997921e-08	4.496206e-03
	(5, 0.1)	1390	7.24	9.996991e-08	4.519106e-03	2938	14.74	9.992376e-08	4.519091e-03
	(10, 0.05)	1333	6.82	9.990830e-08	3.367146e-03	2374	11.95	9.999978e-08	3.367330e-03
	(10, 0.1)	1451	7.47	9.996544e-08	3.452542e-03	2536	12.69	9.991182e-08	3.452290e-03
	(20, 0.05)	1448	7.45	9.994107e-08	2.774133e-03	2460	12.32	9.995598e-08	2.774523e-03
	(20, 0.1)	1755	9.09	9.995555e-08	3.118335e-03	2696	13.67	9.976440e-08	3.118750e-03
总计、 误差均值		-	11893	61.71	-	51443	111.37	-	5.144453e-03

表 2 PSRADMM 与 BADMM 的数值试验结果

另外, 为更直观反映在 PSRADMM 和 BADMM 迭代中, 相对变化率 RelChg 和相对误差 RelErr 的变化趋势和稳定性, 在图 2 和图 3 中, 我们给出了模型的 rank = 10, spr = 0.1 情形, 对应于无高斯噪声 ($\varrho = 0$) 和带高斯噪声 ($\varrho = 0.01$) 时, RelChg 和 RelErr 的变化曲线.

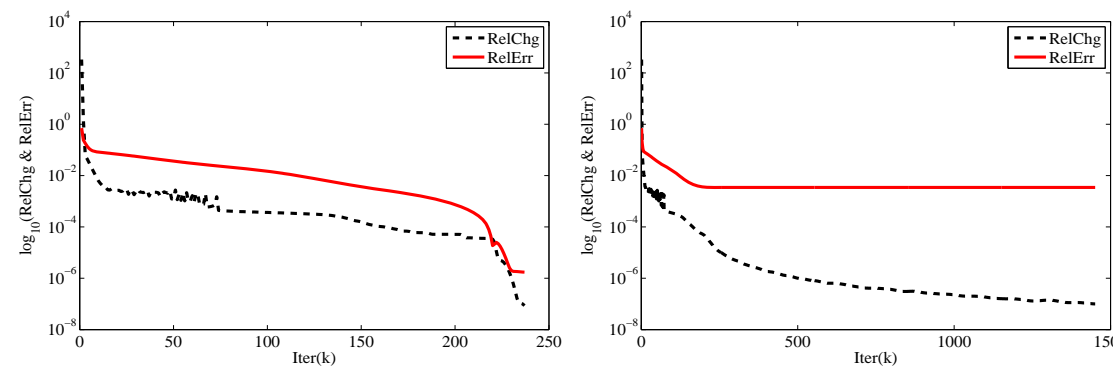


图 2 PSRADMM 无高斯噪声情形 (左) 与带高斯噪声情形 (右) RelChg 和 RelErr 的比较

从图 2 可见, 总体而言, 无论是无高斯噪声还是带高斯噪声, 随着 PSRADMM 迭代的进行, 衡量算法恢复效果的相对误差 RelErr 均能较好地逼近到 0. 具体而言, 对于无高斯噪声情况, PSRADMM 的 RelChg 和 RelErr 均能快速收敛到 0, 不发生“锯齿”现象. 而对于带高斯噪声情形, 在约 220 次迭代后, RelErr 一直维持在 $3.45\text{e-}03$ 至 $3.46\text{e-}03$ 之间, 出现严重的“锯齿”现象. 只是相对变化率 RelChg 在缓慢改进. 从图 3 可见, 无论求解无高斯噪声还是带高斯噪声的问题, PSRADMM 的相对误差趋向 0 的速度均优于 BADMM.

从上面的数值测试结果与分析可见, 对于所进行的稳健主成分分析系列测试问题, 本文建立

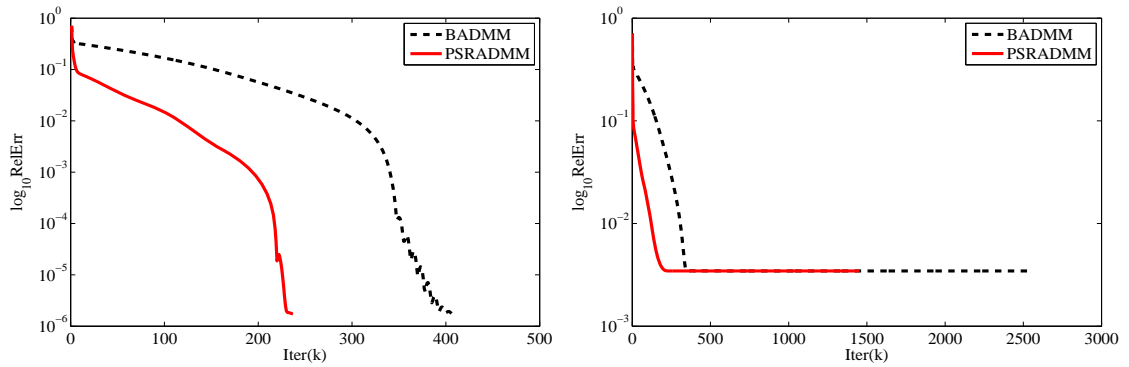


图3 PSRADMM 与 BADMM 无高斯噪声情形 (左) 与带高斯噪声情形 (右) 的 RelErr 比较

的 PSRADMM 的数值效果要明显优于 BADMM, 这很大程度上得益于部分对称乘子修改技术 (1.4d) 的引入.

参 考 文 献

- [1] Attouch H., Bolte J., On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, *Math. Program.*, 2009, **116**(1–2): 5–16.
- [2] Bai J., Liang J., Guo K., et al., Accelerated symmetric ADMM and its applications in signal processing, arXiv:1906.12015, 2019.
- [3] Bolte J., Daniilidis A., Lewis A., The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems, *SIAM J. Optim.*, 2007, **17**(4): 1205–1223.
- [4] Boyd S., Parikh N., Chu E., et al., Distributed optimization and statistical learning via the alternating direction method of multipliers, *Found. Trends Mach. Learn.*, 2011, **31**: 1–122.
- [5] Candès E., Li X., Ma Y., et al., Robust principal component analysis? *J. ACM*, 2011, **58**(3): 1–37.
- [6] Chao M., Cheng C., Zhang H., A linearized alternating direction method of multipliers with substitution procedure, *Asia Pac. J. Oper. Res.*, 2015, **32**(3): 1550011.
- [7] Chao M., Deng Z., Jian J., Convergence of linear Bregman ADMM for nonconvex and nonsmooth problems with nonseparable structure, *Complexity*, 2020, 6237942.
- [8] Chao M., Zhang Y., Jian J., An inertial proximal alternating direction method of multipliers for nonconvex optimization, *Int. J. Comput. Math.*, <https://doi.org/10.1080/00207160.2020.1812585>, 2020.
- [9] Chen C., He B., Ye Y., et al., The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent, *Math. Program.*, 2016, **155**(1–2): 57–79.
- [10] Chen C., Some notes on the divergence example for multi-block alternating direction method of multipliers (in Chinese), *Oper. Res. Trans.*, 2019, **23**(3): 135–140.
- [11] Daubechies I., Defrise M., De Mol C., An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Commun. Pure Appl. Math.*, 2003, **57**(11): 1413–1457.
- [12] Douglas J., Rachford Jr.H.H., On the numerical solution of the heat conduction problem in two and three space variables, *Trans. Amer. Math. Soc.*, 1956, **82**(2): 421–439.
- [13] Glowinski R., Numerical Methods for Nonlinear Variational Problems, Springer-Verlag, New York, 1984.
- [14] Gu Y., Jiang B., Han D., A semi-proximal-based strictly contractive Peaceman–Rachford splitting method, arXiv: 1506.02221, 2015.
- [15] Guo K., Han D., Wang D., et al., Convergence of ADMM for multi-block nonconvex separable optimization models, *Front. Math. China*, 2017, **12**(5): 1139–1162.
- [16] Guo K., Han D., Wu T., Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints, *Int. J. Comput. Math.*, 2017, **94**(8): 1653–1669.
- [17] Han D., Yuan X., A note on the alternating direction method of multipliers, *J. Optim. Theory Appl.*, 2012, **155**(1): 227–238.
- [18] Han D., Yuan X., Local linear convergence of the alternating direction method of multipliers for quadratic

- programs, *SIAM J. Numer. Anal.*, 2013, **51**(6): 3446–3457.
- [19] He B., Parallel splitting augmented Lagrangian methods for monotone structured variational inequalities, *Comput. Optim. Appl.*, 2009, **42**(2): 195–212.
- [20] He B., Ma F., Yuan X., Convergence study on the symmetric version of ADMM with larger step sizes, *SIAM J. Imaging Sci.*, 2016, **9**(3): 1467–1501.
- [21] He B., Tao M., Yuan X., Alternating direction method with Gaussian back substitution for separable convex programming, *SIAM J. Optim.*, 2012, **22**(2): 313–340.
- [22] He B., Tao M., Yuan X., A splitting method for separable convex programming, *IMA J. Numer. Anal.*, 2015, **35**(1): 394–426.
- [23] He B., Yuan X., Linearized alternating direction method with Gaussian back substitution for separable convex programming, *Numer. Algebr. Control Optim.*, 2013, **3**(2): 247–260.
- [24] Jian J., Zhang Y., Chao M., A regularized alternating direction method of multipliers for a class of nonconvex problems, *J. Inequal. Appl.*, 2019, **2019**(1): 193.
- [25] Li Z., Dai Y., Optimal beamformer design in the reverberant environment (in Chinese), *Sci. Sin. Math.*, 2016, **46**(06): 877–892.
- [26] Liu M., Jian J., An ADMM-based SQP method for separably smooth nonconvex optimization, *J. Inequal. Appl.*, 2020, **2020**: 81.
- [27] Lions P., Mercier B., Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.*, 1979, **16**(6): 964–979.
- [28] Nesterov Y., Introductory Lectures on Convex Optimization: A Basic Course, Springer Science & Business Media, Boston, 2013.
- [29] Peaceman D., Rachford H., The numerical solution of parabolic and elliptic differential equations, *J. Soci. Ind. Appl. Math.*, 1955, **3**(1): 28–41.
- [30] Rockafellar R., Wets R., Variational Analysis, Springer Science & Business Media, Berlin, 2009.
- [31] Wang F., Cao W., Xu Z., Convergence of multi-block Bregman ADMM for nonconvex composite problems, *Sci. China Inform. Sci.*, 2018, **61**(12): 122101.
- [32] Wang F., Xu Z., Xu H., Convergence of Bregman alternating direction method with multipliers for nonconvex composite problems, arXiv:1410.8625, 2014.
- [33] Wang Y., Fan S., Weighted l_1 -norm constrained sparse regularization and alternating directions method of multipliers for seismic time-frequency analysis (in Chinese), *Sci. Sin. Math.*, 2018, **48**(03): 457–470.
- [34] Wu Z., Li M., Wang D., Han D., A symmetric alternating direction method of multipliers for separable nonconvex minimization problems, *Asia Pac. J. Oper. Res.*, 2017, **34**(06): 1750030.
- [35] Xu Z., Chang X., Xu F., et al., $l_{\frac{1}{2}}$ regularization: a thresholding representation theory and a fast solver, *IEEE T. Neur. Net. Lear.*, 2012, **23**(7): 1013–1027.
- [36] Yang L., Luo J., Xu Y., et al., A distributed dual consensus ADMM based on partition for DC-DOPF with carbon emission trading, *IEEE T. Ind. Inform.*, 2020, **16**(3): 1858–1872.
- [37] Yang W., Han D., Linear convergence of the alternating direction method of multipliers for a class of convex optimization problems, *SIAM J. Numer. Anal.*, 2016, **54**(2): 625–640.
- [38] Zeng J., Fang J., Xu Z., Sparse SAR imaging based on $l_{\frac{1}{2}}$ regularization, *Sci. China Inform. Sci.*, 2012, **55**: 1755–1775.
- [39] Zhu M., Mihic K., Ye Y., On a randomized multi-block ADMM for solving selected machine learning problems, arXiv:1907.01995, 2019.