

三块非凸优化问题正则化交替方向法的收敛性^{*}

雷宏旺, 彭建文

(重庆师范大学 数学科学学院, 重庆 401331)

摘要:【目的】针对一类三块非凸优化问题,提出一种正则化交替方向法。【方法】为了更易求得唯一的点 $(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1})$,在原始乘子交替方向法的框架下,对 $\mathbf{x}$ 子问题和 $\mathbf{y}$ 子问题同时添加一个临近项来正则化原始子问题。【结果】在增广拉格朗日函数满足KL性质且惩罚参数充分大的条件下,由算法生成的迭代序列的任何聚点都是增广拉格朗日函数的稳定点。【结论】数值算例结果验证了此算法的有效性。

关键词:交替方向法;Kurdyka-Lojasiewicz性质;稳定点

中图分类号:O221.6

文献标志码:A

文章编号:1672-6693(2021)03-0006-10

近些年来,对于凸优化问题的ADMM方法及与之相关的纯粹理论与应用方面的研究已日趋成熟^[1-4],而当此类算法推广去解决非凸优化问题时,算法的收敛性或许无法保证。在最近的文献研究中,因为非凸优化问题的应用研究也愈来愈多,因而针对非凸优化问题所提出的ADMM方法也愈加多了起来,那么构造非凸优化问题的交替方向法及分析它的收敛性具有重要的研究价值。

Guo等人^[5]考虑如下的非凸优化问题:

$$\begin{aligned} \min & f(\mathbf{x}) + g(\mathbf{y}) + h(\mathbf{z}), \\ \text{s. t. } & \mathbf{Ax} + \mathbf{By} + \mathbf{z} = \mathbf{b}. \end{aligned} \quad (1)$$

其中: $f: \mathbf{R}^{n_1} \rightarrow \mathbf{R} \cup \{+\infty\}$ 为真下半连续函数, $g: \mathbf{R}^{n_2} \rightarrow \mathbf{R}$, $h: \mathbf{R}^s \rightarrow \mathbf{R}$ 且都为连续可微函数, ∇g 和 ∇h 是Lipschitz连续的, $\mathbf{A} \in \mathbf{R}^{s \times n_1}$, $\mathbf{B} \in \mathbf{R}^{s \times n_2}$ 为给定矩阵, $\mathbf{b} \in \mathbf{R}^s$ 为一个向量。

针对上述非凸优化问题,Guo等人^[5]提出如下的交替方向乘子法:

$$\begin{cases} \mathbf{x}^{k+1} \in \arg \min_{\mathbf{x}} \{L_{\beta}(\mathbf{x}, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k)\} \\ \mathbf{y}^{k+1} \in \arg \min_{\mathbf{y}} \{L_{\beta}(\mathbf{x}^{k+1}, \mathbf{y}, \mathbf{z}^k, \boldsymbol{\lambda}^k)\} \\ \mathbf{z}^{k+1} \in \arg \min_{\mathbf{z}} \{L_{\beta}(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}, \boldsymbol{\lambda}^k)\} \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \beta(\mathbf{Ax}^{k+1} + \mathbf{By}^{k+1} + \mathbf{z}^{k+1} - \mathbf{b}) \end{cases}. \quad (2)$$

文献[5]中在增广拉格朗日函数满足KL性质的情况下,证明了算法(2)式生成的迭代序列的任何聚点都是临界点,并给出算法的收敛率的一些充分条件。而在最近的文献[6]中,考虑了两分块的非凸优化问题:

$$\begin{aligned} \min & f(\mathbf{x}) + g(\mathbf{y}), \\ \text{s. t. } & \mathbf{Ax} - \mathbf{y} = \mathbf{0}. \end{aligned}$$

其中: $f: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{+\infty\}$ 为正常下半连续函数, $g: \mathbf{R}^{n_2} \rightarrow \mathbf{R}$ 为可微连续函数, ∇g 是Lipschitz连续且Lipschitz常数 $L > 0$, $\mathbf{A} \in \mathbf{R}^{m \times n}$ 为给定矩阵, $\mathbf{G}$ 是一个对称半正定矩阵。上述问题的ADMM迭代格式为:

$$\begin{cases} \mathbf{x}^{k+1} \in \arg \min_{\mathbf{x}} \left\{ L_{\beta}(\mathbf{x}, \mathbf{y}^k, \boldsymbol{\lambda}^k) + \frac{1}{2} \|\mathbf{x} - \mathbf{x}^k\|_{\mathbf{G}}^2 \right\} \\ \mathbf{y}^{k+1} \in \arg \min_{\mathbf{y}} \{L_{\beta}(\mathbf{x}^{k+1}, \mathbf{y}, \boldsymbol{\lambda}^k)\} \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \beta(\mathbf{Ax}^{k+1} - \mathbf{y}^{k+1}) \end{cases}.$$

* 收稿日期:2020-07-03 修回日期:2021-03-07 网络出版时间:2021-05-20 13:24

资助项目:国家自然科学基金重大项目(No. 11991024)

第一作者简介:雷宏旺,男,研究方向为最优化理论与算法,E-mail:leihongwangl@126.com;通信作者:彭建文,男,教授,博士生导师,E-mail:jwpeng2008@gmail.com

网络出版地址:https://kns.cnki.net/kcms/detail/50.1165.N.20210520.0936.004.html

针对上述算法,在增广拉格朗日函数满足 KL 性质的情况下,证明了算法生成的迭代序列的任何聚点都是临界点。因此,本文中非凸优化问题和算法可作为文献[6]的一种推广,提出一种针对三块非凸优化问题的正则化 ADMM 方法,考虑如下非凸优化问题:

$$\begin{aligned} \min & f(\mathbf{x}) + g(\mathbf{y}) + h(\mathbf{z}), \\ \text{s. t. } & \mathbf{Ax} + \mathbf{By} - \mathbf{z} = \mathbf{0}. \end{aligned} \quad (3)$$

其中: $f: \mathbf{R}^m \rightarrow \mathbf{R} \cup \{+\infty\}$ 是正常下半连续函数, $g: \mathbf{R}^n \rightarrow \mathbf{R}, h: \mathbf{R}^s \rightarrow \mathbf{R}$ 都为连续可微函数且 ∇g 和 ∇h Lipschitz 连续, Lipschitz 常数为 L_2 和 L_3 , $\mathbf{A} \in \mathbf{R}^{s \times m}$ 及 $\mathbf{B} \in \mathbf{R}^{s \times n}$ 为给定矩阵。

问题(3)的增广拉格朗日函数定义如下:

$$L_\beta(\mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\lambda}) := f(\mathbf{x}) + g(\mathbf{y}) + h(\mathbf{z}) - \langle \boldsymbol{\lambda}, \mathbf{Ax} + \mathbf{By} - \mathbf{z} \rangle + \frac{\beta}{2} \|\mathbf{Ax} + \mathbf{By} - \mathbf{z}\|^2,$$

其中: $\boldsymbol{\lambda} \in \mathbf{R}^s$ 为增广拉格朗日乘子, $\beta > 0$ 为惩罚参数。

针对问题(3),本文提出如下正则化交替方向乘法:

$$\begin{cases} \mathbf{x}^{k+1} \in \arg \min_{\mathbf{x}} \left\{ L_\beta(\mathbf{x}, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k) + \frac{1}{2} \|\mathbf{x} - \mathbf{x}^k\|_D^2 \right\} \\ \mathbf{y}^{k+1} \in \arg \min_{\mathbf{y}} \left\{ L_\beta(\mathbf{x}^{k+1}, \mathbf{y}, \mathbf{z}^k, \boldsymbol{\lambda}^k) + \frac{1}{2} \|\mathbf{y} - \mathbf{y}^k\|_G^2 \right\}, \\ \mathbf{z}^{k+1} \in \arg \min_{\mathbf{z}} \{ L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}, \boldsymbol{\lambda}^k) \} \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \beta(\mathbf{Ax}^{k+1} + \mathbf{By}^{k+1} - \mathbf{z}) \end{cases} \quad (4)$$

其中: $\mathbf{D} \in \mathbf{R}^{m \times m}, \mathbf{G} \in \mathbf{R}^{n \times n}$ 为对称矩阵。

1 预备知识

任给集合 $S \subseteq \mathbf{R}^n$ 以及任给的点 $\mathbf{x} \in \mathbf{R}^n$, 记 $d(\mathbf{x}, S)$ 为 $\mathbf{x}$ 到 S 的距离, 定义为: $d(\mathbf{x}, S) := \inf_{\mathbf{y} \in S} \|\mathbf{y} - \mathbf{x}\|$, 若 $S = \emptyset$, 对任意的 $\mathbf{x}$, 令 $d(\mathbf{x}, S) := +\infty$ 。设函数 $f: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{+\infty\}$, 有效域和上图分别定义为: $\text{dom } f := \{\mathbf{x} \in \mathbf{R}^n : f(\mathbf{x}) < +\infty\}$, $\text{epi } f := \{(\mathbf{x}, \alpha) \in \mathbf{R}^n \times \mathbf{R} : f(\mathbf{x}) \leq \alpha\}$ 。若 $\text{dom } f \neq \emptyset$, 则函数 f 为正常函数; 若 $\text{epi } f$ 为闭集, 那么函数 f 为下半连续函数。

定义 1^[7] 若函数 $f: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{+\infty\}$ 为一个正常下半连续函数, 则有:

1) 函数 f 在 $\mathbf{x} \in \text{dom } f$ 处的 Frechet 次微分, 定义为所有满足下面条件的 $\mathbf{x}^*$ 的集合, 记作 $\hat{\partial} f(\mathbf{x})$,

$$\liminf_{\mathbf{y} \neq \mathbf{x}, \mathbf{y} \rightarrow \mathbf{x}} \frac{f(\mathbf{y}) - f(\mathbf{x}) - \langle \mathbf{x}^*, \mathbf{y} - \mathbf{x} \rangle}{\|\mathbf{y} - \mathbf{x}\|} \geq 0.$$

若 $\mathbf{x} \notin \text{dom } f$, 令 $\hat{\partial} f(\mathbf{x}) := \emptyset$;

2) 函数 f 在 $\mathbf{x} \in \text{dom } f$ 处的极限次微分, 记作 $\partial f(\mathbf{x})$, 定义为:

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

注 1 从上述定义, 可以得到:

1) $\hat{\partial} f(\mathbf{x}) \subseteq \partial f(\mathbf{x}), \forall \mathbf{x} \in \mathbf{R}^n$, 其中 $\hat{\partial} f(\mathbf{x})$ 为闭凸集, 而 $\partial f(\mathbf{x})$ 仅为闭集;

2) 设 $(\mathbf{x}_k, \mathbf{x}_k^*) \in \text{Graph } \partial f$ 且收敛到 $(\mathbf{x}, \mathbf{x}^*)$ 。由 $\partial f(\mathbf{x})$ 的定义, 当 $k \rightarrow +\infty$ 时, 若 $f(\mathbf{x}_k)$ 收敛到 $f(\mathbf{x})$, 则有: $(\mathbf{x}, \mathbf{x}^*) \in \text{Graph } \partial f$;

3) $\mathbf{x} \in \mathbf{R}^n$ 为函数 f 稳定点的必要条件是: $\mathbf{0} \in \partial f(\mathbf{x})$;

4) 如果 $f: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{+\infty\}$ 是正常下半连续函数, $g: \mathbf{R}^n \rightarrow \mathbf{R}$ 是连续可微函数, 那么对于 $\forall \mathbf{x} \in \text{dom } f$, 都有 $\partial(f+g)(\mathbf{x}) = \partial f(\mathbf{x}) + \nabla g(\mathbf{x})$ 。

若点 $\mathbf{x}$ 满足注 1 中的结论(3), 则称点 $\mathbf{x}$ 为临界点或稳定点, 函数 f 的所有临界点或稳定点的集合记作 $\text{crit } f$ 。

定义 2^[8] (Kurdyka-Lojasiewicz 性质) 设函数 $f: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{+\infty\}$ 是正常下半连续函数, 令 $-\infty < \eta_1 < \eta_2 \leq +\infty$, 记 $[\eta_1 < f < \eta_2] := \{\mathbf{x} \in \mathbf{R}^n : \eta_1 < f(\mathbf{x}) < \eta_2\}$, 称函数 f 在 $\mathbf{x}^* \in \text{dom } \partial f$ 具有 Kurdyka-Lojasiewicz 性质, 若存

在 $\eta \in (0, +\infty]$, $\mathbf{x}^*$ 的某邻域 U 及连续的凹函数 $\varphi: [0, \eta] \rightarrow \mathbf{R}_+$, 使得:

- 1) $\varphi(0) = 0$;
- 2) φ 在 $(0, \eta)$ 上连续可微并且 φ 在 0 处是连续的;
- 3) $\varphi'(s) > 0, \forall s \in (0, \eta)$;
- 4) 对 $\forall \mathbf{x} \in U \cap [f(\mathbf{x}^*) < f < f(\mathbf{x}^*) + \eta]$, KL 不等式都成立: $\varphi'(f(\mathbf{x}) - f(\mathbf{x}^*))d(\mathbf{0}, \partial f(\mathbf{x})) \geq 1$ 。

定义 3 $\mathbf{w}^* = (\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*, \boldsymbol{\lambda}^*)$ 为增广拉格朗日函数的稳定点, 即 $\mathbf{0} \in \partial L_\beta(\mathbf{w}^*)$, 当且仅当:

$$\begin{cases} \mathbf{A}^T \boldsymbol{\lambda}^* \in \partial f(\mathbf{x}^*) \\ \nabla g(\mathbf{y}^*) = \mathbf{B}^T \boldsymbol{\lambda}^* \\ \nabla h(\mathbf{z}^*) = -\boldsymbol{\lambda}^* \\ \mathbf{A}\mathbf{x}^* + \mathbf{B}\mathbf{y}^* - \mathbf{z}^* = \mathbf{0} \end{cases}.$$

引理 1^[9] (一致 KL 性质) 设 Ω 为紧集, 函数 $f: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{+\infty\}$ 是正常下半连续函数. 假设函数 f 在 Ω 上为常值并且在 Ω 的每一点都满足 KL 性质, 则存在 $\varepsilon > 0, \eta > 0, \varphi \in \varphi_\eta$, 使得对 $\forall \bar{\mathbf{x}} \in \Omega$ 及 $\forall \mathbf{x}$ 属于下述交集: $\{\mathbf{x} \in \mathbf{R}^n : d(\mathbf{x}, \Omega) < \varepsilon\} \cap [f(\bar{\mathbf{x}}) < f < f(\bar{\mathbf{x}}) + \eta]$, 有: $\varphi'(f(\mathbf{x}) - f(\bar{\mathbf{x}}))d(\mathbf{0}, \partial f(\mathbf{x})) \geq 1$ 。

引理 2^[10] 设函数 $\psi: \mathbf{R}^n \rightarrow \mathbf{R}$ 是连续可微函数且 $\nabla \psi$ 是 Lipschitz 连续的, 即存在 Lipschitz 常数 $L_\psi > 0$, 如果对任意 $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ 有: $|\psi(\mathbf{y}) - \psi(\mathbf{x}) - \langle \nabla \psi(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle| \leq \frac{L_\psi}{2} \|\mathbf{y} - \mathbf{x}\|^2$ 。

2 收敛性分析

始终假设由算法产生的迭代点 $\{\mathbf{w}^k := (\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k)\}$ 有界, 由算法的最优性条件有:

$$\begin{cases} \mathbf{A}^T \boldsymbol{\lambda}^{k+1} + \beta \mathbf{A}^T (\mathbf{B}\mathbf{y}^{k+1} - \mathbf{B}\mathbf{y}^k) + \beta \mathbf{A}^T (\mathbf{z}^k - \mathbf{z}^{k+1}) - \mathbf{D}(\mathbf{x}^{k+1} - \mathbf{x}^k) \in \partial f(\mathbf{x}^{k+1}) \\ \mathbf{B}^T \boldsymbol{\lambda}^{k+1} + \beta \mathbf{B}^T (\mathbf{z}^k - \mathbf{z}^{k+1}) - \mathbf{G}(\mathbf{y}^{k+1} - \mathbf{y}^k) = \nabla g(\mathbf{y}^{k+1}) \\ -\boldsymbol{\lambda}^{k+1} = \nabla h(\mathbf{z}^{k+1}), \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \beta(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^{k+1} - \mathbf{z}^{k+1}) \end{cases}.$$
 (5)

下面给出本文收敛性分析需要的假设条件:

假设 设函数 $f: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{+\infty\}$ 是正常下半连续函数, $g: \mathbf{R}^m \rightarrow \mathbf{R}$ 以及 $h: \mathbf{R}^p \rightarrow \mathbf{R}$ 为连续可微函数且 $\nabla g, \nabla h$ 是 Lipschitz 连续的, Lipschitz 常数为 $L_1 > 0$ 和 $L_2 > 0$, 令 $L := \max\{L_1, L_2\}$ 。假设以下条件都是满足的: 1) $\beta > \max\{2L, \frac{L}{\mu}\}$; 2) $\exists \mu > 0$, 使得 $\mathbf{A}^T \mathbf{A} \geq \mu \mathbf{I}, \mathbf{B}^T \mathbf{B} \geq \mu \mathbf{I}$, 令:

$$\delta := \min\left\{\frac{\beta-L}{2} - \frac{L^2}{\beta}, \frac{\beta\mu-L}{2}\right\} > 0;$$
 (6)

3) $\mathbf{D} \geq 0, \mathbf{G} \geq 0, \mathbf{D} + \mathbf{A}^T \mathbf{A} > 0$ 。为了方便, 令 $\{\mathbf{v}^k := (\mathbf{y}^k, \mathbf{z}^k)\}$, 下面的引理表明序列 $\{L_\beta(\mathbf{w}^k)\}$ 单调下降。

引理 3 设 $\{\mathbf{w}^k := (\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k)\}$ 为算法 (3) 产生的迭代序列, 则有:

$$L_\beta(\mathbf{w}^{k+1}) \leq L_\beta(\mathbf{w}^k) - \delta \|\mathbf{v}^{k+1} - \mathbf{v}^k\|^2 - \|\mathbf{y}^{k+1} - \mathbf{y}^k\|_G^2 - \frac{1}{2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_D^2。$$

证明 由增广拉格朗日函数 $L_\beta(\cdot)$ 的定义可知:

$$\begin{aligned} L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \boldsymbol{\lambda}^{k+1}) &= \\ L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \boldsymbol{\lambda}^k) + \langle \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k+1}, \mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^{k+1} - \mathbf{z}^{k+1} \rangle &= \\ L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \boldsymbol{\lambda}^k) + \frac{1}{\beta} \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k+1}\|^2, \end{aligned}$$
 (7)

并且:

$$\begin{aligned} L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^k, \boldsymbol{\lambda}^k) - L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \boldsymbol{\lambda}^k) &= \\ h(\mathbf{z}^k) - h(\mathbf{z}^{k+1}) + \langle \boldsymbol{\lambda}^k, \mathbf{z}^k - \mathbf{z}^{k+1} \rangle + \frac{\beta}{2} \|\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^{k+1} - \mathbf{z}^k\|^2 - \frac{\beta}{2} \|\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^{k+1} - \mathbf{z}^{k+1}\|^2. \end{aligned}$$
 (8)

因为 ∇h 是 Lipschitz 连续的, 并且 Lipschitz 常数 $L_2 \leq L$, 因此, 由引理 2 以及由最优性条件的第 3 个式子可得:

$$h(\mathbf{z}^k) - h(\mathbf{z}^{k+1}) \geq \langle -\boldsymbol{\lambda}^{k+1}, \mathbf{z}^k - \mathbf{z}^{k+1} \rangle - \frac{L}{2} \|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2. \quad (9)$$

将(8)式代入(7)式得:

$$\begin{aligned} & L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^k, \boldsymbol{\lambda}^k) - L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \boldsymbol{\lambda}^k) \geq \\ & \langle \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k+1}, \mathbf{z}^k - \mathbf{z}^{k+1} \rangle + \frac{\beta}{2} \|\mathbf{Ax}^{k+1} + \mathbf{By}^{k+1} - \mathbf{z}^k\|^2 - \frac{\beta}{2} \|\mathbf{Ax}^{k+1} + \mathbf{By}^{k+1} - \mathbf{z}^{k+1}\|^2 - \frac{L}{2} \|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2. \end{aligned} \quad (10)$$

由最优性条件的第4个式子知, $\mathbf{Ax}^{k+1} + \mathbf{By}^{k+1} - \mathbf{z}^k = \frac{1}{\beta}(\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k+1}) + (\mathbf{z}^{k+1} - \mathbf{z}^k)$, 因此:

$$\langle \boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k+1}, \mathbf{z}^k - \mathbf{z}^{k+1} \rangle + \frac{\beta}{2} \|\mathbf{Ax}^{k+1} + \mathbf{By}^{k+1} - \mathbf{z}^k\|^2 = \frac{\beta}{2} \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 + \frac{1}{2\beta} \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k+1}\|^2. \quad (11)$$

将(11)式代入(10)式可得:

$$L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^k, \boldsymbol{\lambda}^k) - L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \boldsymbol{\lambda}^k) \geq \frac{\beta-L}{2} \|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2. \quad (12)$$

由 $-\boldsymbol{\lambda}^{k+1} = \nabla h(\mathbf{z}^{k+1})$ 以及假设条件可知:

$$\frac{1}{\beta} \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 \leq \frac{L^2}{\beta} \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2. \quad (13)$$

因此, 由(7), (12)和(13)式可得:

$$L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \boldsymbol{\lambda}^{k+1}) \leq L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^k, \boldsymbol{\lambda}^k) - \left(\frac{\beta-L}{2} - \frac{L^2}{\beta} \right) \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2.$$

类似地:

$$\begin{aligned} & L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k) - L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^k, \boldsymbol{\lambda}^k) = \\ & g(\mathbf{y}^k) - g(\mathbf{y}^{k+1}) + \langle \boldsymbol{\lambda}^k, \mathbf{By}^{k+1} - \mathbf{By}^k \rangle + \frac{\beta}{2} \|\mathbf{Ax}^{k+1} + \mathbf{By}^k - \mathbf{z}^k\|^2 - \frac{\beta}{2} \|\mathbf{Ax}^{k+1} + \mathbf{By}^{k+1} - \mathbf{z}^k\|^2. \end{aligned} \quad (14)$$

因为 ∇g 是 Lipschitz 连续且 Lipschitz 常数 $L_1 \leq L$, 故由引理2和最优性条件的第2个式子可知:

$$g(\mathbf{y}^k) - g(\mathbf{y}^{k+1}) \geq \langle \mathbf{B}^T \boldsymbol{\lambda}^k - \beta \mathbf{B}^T (\mathbf{Ax}^{k+1} + \mathbf{By}^{k+1} - \mathbf{z}^k) - \mathbf{G}(\mathbf{y}^{k+1} - \mathbf{y}^k), \mathbf{y}^k - \mathbf{y}^{k+1} \rangle - \frac{L}{2} \|\mathbf{y}^k - \mathbf{y}^{k+1}\|^2. \quad (15)$$

把(15)式代入(14)式, 再由假设条件知:

$$\begin{aligned} & L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k) - L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^k, \boldsymbol{\lambda}^k) \geq \\ & -\beta \langle \mathbf{Ax}^{k+1} + \mathbf{By}^{k+1} - \mathbf{z}^k, \mathbf{By}^k - \mathbf{By}^{k+1} \rangle + \|\mathbf{y}^k - \mathbf{y}^{k+1}\|_G^2 + \frac{\beta}{2} \|\mathbf{Ax}^{k+1} + \mathbf{By}^k - \mathbf{z}^k\|^2 - \frac{\beta}{2} \|\mathbf{Ax}^{k+1} + \mathbf{By}^{k+1} - \mathbf{z}^k\|^2 - \frac{L}{2} \|\mathbf{y}^k - \mathbf{y}^{k+1}\|^2 \geq \\ & \frac{\beta\mu-L}{2} \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 + \|\mathbf{y}^{k+1} - \mathbf{y}^k\|_G^2. \end{aligned} \quad (16)$$

又因为 $\mathbf{x}^{k+1}$ 是第1个子问题的最优解, 所以有:

$$L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k) \leq L_\beta(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k) - \frac{1}{2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_D^2. \quad (17)$$

故由(12), (16)和(17)式可得:

$$\begin{aligned} & L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \boldsymbol{\lambda}^{k+1}) \leq \\ & L_\beta(\mathbf{x}^{k+1}, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k) - \frac{\beta\mu-L}{2} \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 - \left(\frac{\beta-L}{2} - \frac{L^2}{\beta} \right) \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 - \|\mathbf{y}^{k+1} - \mathbf{y}^k\|_G^2 \leq \\ & L_\beta(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k) - \delta \|\mathbf{v}^{k+1} - \mathbf{v}^k\|^2 - \|\mathbf{y}^{k+1} - \mathbf{y}^k\|_G^2 - \frac{1}{2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_D^2. \end{aligned} \quad \text{证毕}$$

引理4 设 $\{\mathbf{w}^k\}$ 为算法(3)产生的迭代序列, 假定序列有界, 则有: $\sum_{k=0}^{+\infty} \|\mathbf{w}^{k+1} - \mathbf{w}^k\|^2 < +\infty$.

证明 因为 $\{\mathbf{w}^k\}$ 是有界的, 则存在一个子序列 $\{\mathbf{w}^{k_j}\}$ 使得 $\mathbf{w}^{k_j} \rightarrow \mathbf{w}^*$, 又由于 g 与 h 的连续性以及 f 的下半连续性, 则 $L_\beta(\cdot)$ 是下半连续的, 因此有: $L_\beta(\mathbf{w}^*) \leq \liminf_{k_j \rightarrow +\infty} L_\beta(\mathbf{w}^{k_j})$, 即 $L_\beta(\mathbf{w}^{k_j})$ 有下界. 由引理3知 $L_\beta(\cdot)$ 是单调递减的, 因此 $L_\beta(\mathbf{w}^{k_j})$ 收敛且 $L_\beta(\mathbf{w}^k) \geq L_\beta(\mathbf{w}^*)$, 由引理3知:

$$L_\beta(\mathbf{w}^{k+1}) \leq L_\beta(\mathbf{w}^k) - \delta \|\mathbf{v}^{k+1} - \mathbf{v}^k\|^2 - \|\mathbf{y}^{k+1} - \mathbf{y}^k\|_G^2 - \frac{1}{2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_D^2, \quad (18)$$

则对(18)式关于 $k=0, 1, \dots, +\infty$ 求和有:

$$\sum_{k=0}^{+\infty} \delta \|v^{k+1} - v^k\|^2 - \sum_{k=0}^{+\infty} \|y^{k+1} - y^k\|_G^2 - \frac{1}{2} \sum_{k=0}^{+\infty} \|x^{k+1} - x^k\|_D^2 \leq \sum_{k=0}^{+\infty} (L_\beta(w^k) - L_\beta(w^{k+1})) < +\infty.$$

由于 $D \geq 0, G \geq 0, \delta > 0$, 则:

$$\begin{cases} \sum_{k=0}^{+\infty} \|v^{k+1} - v^k\|^2 < +\infty, \\ \sum_{k=0}^{+\infty} \|y^{k+1} - y^k\|_G^2 < +\infty, \\ \sum_{k=0}^{+\infty} \|x^{k+1} - x^k\|_D^2 < +\infty. \end{cases} \quad (19)$$

因此, $\sum_{k=0}^{+\infty} \|y^{k+1} - y^k\|^2 < +\infty, \sum_{k=0}^{+\infty} \|z^{k+1} - z^k\|^2 < +\infty$. 又由(12)式知, $\sum_{k=0}^{+\infty} \|\lambda^{k+1} - \lambda^k\|^2 < +\infty$.

又因 $\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - z^{k+1}), \lambda^k = \lambda^{k-1} - \beta(Ax^k + By^k - z^k)$. 故:

$$\lambda^{k+1} - \lambda^k = (\lambda^k - \lambda^{k-1}) + \beta(Ax^k - Ax^{k+1}) + \beta(By^k - By^{k+1}) + \beta(z^{k+1} - z^k),$$

则有:

$$\|\beta(Ax^k - Ax^{k+1})\|^2 \leq 4(\|\lambda^{k+1} - \lambda^k\|^2 + \|\lambda^k - \lambda^{k-1}\|^2 + \beta^2 \|B\|^2 \|y^{k+1} - y^k\|^2 + \beta^2 \|z^{k+1} - z^k\|^2) < +\infty. \quad (20)$$

因此 $\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\|_{A^T A}^2 < +\infty$, 考虑到(19)式, 则有 $\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\|_{A^T A + D}^2 < +\infty$. 因 $A^T A + D > 0$, 可得

$$\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\|^2 < +\infty, \text{ 故 } \sum_{k=0}^{+\infty} \|w^{k+1} - w^k\|^2 < +\infty. \quad \text{证毕}$$

引理 5 设 $\{w^k := (x^k, y^k, z^k, \lambda^k)\}$ 为算法(3)产生的迭代序列, 则存在 $M > 0$, 使得:

$$d(0, \partial L_\beta(w^{k+1})) \leq M(\|y^{k+1} - y^k\| + \|z^{k+1} - z^k\| + \|z^k - z^{k-1}\|).$$

证明 由增广拉格朗日函数 $L_\beta(\cdot)$ 的定义可得:

$$\begin{cases} \partial_x L_\beta(w^{k+1}) = \partial f(x^{k+1}) - A^T \lambda^{k+1} + \beta A^T (Ax^{k+1} + By^{k+1} - z^{k+1}) + D(x^{k+1} - x^k) \\ \partial_y L_\beta(w^{k+1}) = \nabla g(y^{k+1}) - B^T \lambda^{k+1} + \beta B^T (Ax^{k+1} + By^{k+1} - z^{k+1}) + G(y^{k+1} - y^k) \\ \partial_z L_\beta(w^{k+1}) = \nabla f(x^{k+1}) + \lambda^{k+1} - \beta(Ax^{k+1} + By^{k+1} - z^{k+1}) \\ \partial_\lambda L_\beta(w^{k+1}) = -(Ax^{k+1} + By^{k+1} - z^{k+1}) \end{cases}.$$

联系(4)式可得:

$$\begin{cases} A^T (\lambda^k - \lambda^{k+1}) + \beta A^T (By^{k+1} - By^k) + \beta A^T (z^k - z^{k+1}) - D(x^{k+1} - x^k) \in \partial_x L_\beta(w^{k+1}) \\ B^T (\lambda^k - \lambda^{k+1}) + \beta B^T (z^k - z^{k+1}) - G(y^{k+1} - y^k) \in \partial_y L_\beta(w^{k+1}) \\ \lambda^{k+1} - \lambda^k \in \partial_z L_\beta(w^{k+1}) \\ \frac{1}{\beta} (\lambda^{k+1} - \lambda^k) \in \partial_\lambda L_\beta(w^{k+1}) \end{cases}.$$

定义:

$$\begin{aligned} (x_1^{k+1}, y_1^{k+1}, z_1^{k+1}, \lambda_1^{k+1}) := & \\ & \left(A^T (\lambda^k - \lambda^{k+1}) + \beta A^T (By^{k+1} - By^k) + \beta A^T (z^k - z^{k+1}) - D(x^{k+1} - x^k), \right. \\ & \left. B^T (\lambda^k - \lambda^{k+1}) + \beta B^T (z^k - z^{k+1}) - G(y^{k+1} - y^k), \lambda^{k+1} - \lambda^k, \frac{1}{\beta} (\lambda^{k+1} - \lambda^k) \right), \end{aligned}$$

所以有 $(x_1^{k+1}, y_1^{k+1}, z_1^{k+1}, \lambda_1^{k+1}) \in \partial L_\beta(w^{k+1})$, 则存在 $m_1 > 0$ 使得:

$$\|x_1^{k+1}, y_1^{k+1}, z_1^{k+1}, \lambda_1^{k+1}\| \leq m_1 (\|x^{k+1} - x^k\| + \|y^{k+1} - y^k\| + \|z^{k+1} - z^k\| + \|\lambda^{k+1} - \lambda^k\|). \quad (21)$$

由于 $A^T A > 0$, 则对于(20)式, 存在 $m_2 > 0$ 使得:

$$\|x^{k+1} - x^k\| \leq m_2 (\|y^{k+1} - y^k\| + \|z^{k+1} - z^k\| + \|\lambda^{k+1} - \lambda^k\| + \|\lambda^k - \lambda^{k-1}\|). \quad (22)$$

因为 $(x_1^{k+1}, y_1^{k+1}, z_1^{k+1}, \lambda_1^{k+1}) \in \partial L_\beta(w^{k+1})$, 再由(12), (21)和(22)式, 对 $k \geq 1$, 则存在 $M > 0$, 使得:

$$d(0, \partial L_\beta(w^{k+1})) \leq \|x_1^{k+1}, y_1^{k+1}, z_1^{k+1}, \lambda_1^{k+1}\| \leq M(\|y^{k+1} - y^k\| + \|z^{k+1} - z^k\| + \|z^k - z^{k-1}\|). \quad \text{证毕}$$

定理 1 设 $\{w^k := (x^k, y^k, z^k, \lambda^k)\}$ 为算法(3)产生的迭代序列并假定有界,令 Ω 为所有极限点的集合,则有:

- 1) Ω 为非空紧集,当 $k \rightarrow +\infty$ 时, $d(w^k, \Omega) \rightarrow 0$;
- 2) $\Omega \subseteq \text{crit } L_\beta$;
- 3) $L_\beta(\cdot)$ 在 Ω 上为常数,则 $\lim_{k \rightarrow +\infty} L_\beta(w^k) = L_\beta(w^*)$ 。

证明 1) 由 Ω 和 $d(w^k, \Omega)$ 定义可推得结论成立。

2) 令 $w^* := (x^*, y^*, z^*, \lambda^*) \in \Omega$, 则存在 $\{w^k\}$ 的子序列 $\{w^{k_j}\}$ 收敛至 w^* , 由引理 4 可知当 $k \rightarrow +\infty$ 时有:

$$\|w^{k+1} - w^k\| \rightarrow 0, \quad (23)$$

因而有 $(x^{k_{j+1}}, y^{k_{j+1}}, z^{k_{j+1}}, \lambda^{k_{j+1}}) \rightarrow (x^*, y^*, z^*, \lambda^*)$ 。

因为 x^{k+1} 是 $\min L_\beta(x, y^k, z^k, \lambda^k) + \frac{1}{2} \|x - x^k\|_D^2$ 的最优解, 则有:

$$L_\beta(x^{k+1}, y^k, z^k, \lambda^k) + \frac{1}{2} \|x^{k+1} - x^k\|_D^2 \leq L_\beta(x^*, y^k, z^k, \lambda^k) + \frac{1}{2} \|x^* - x^k\|_D^2,$$

即为:

$$L_\beta(x^{k+1}, y^k, z^k, \lambda^k) \leq L_\beta(x^*, y^k, z^k, \lambda^k) + \frac{1}{2} \|x^* - x^k\|_D^2 - \frac{1}{2} \|x^{k+1} - x^k\|_D^2. \quad (24)$$

由(23), (24)式以及 $L_\beta(\cdot)$ 对于 y, z 和 λ 都是连续的, 则有:

$$\limsup_{j \rightarrow +\infty} L_\beta(x^{k_{j+1}}, y^{k_{j+1}}, z^{k_{j+1}}, \lambda^{k_{j+1}}) = \limsup_{j \rightarrow +\infty} L_\beta(x^{k_{j+1}}, y^{k_j}, z^{k_j}, \lambda^{k_j}) \leq L_\beta(x^*, y^*, z^*, \lambda^*). \quad (25)$$

另一方面, 由增广拉格朗日函数的下半连续性, 有:

$$\liminf_{j \rightarrow +\infty} L_\beta(x^{k_{j+1}}, y^{k_{j+1}}, z^{k_{j+1}}, \lambda^{k_{j+1}}) \geq L_\beta(x^*, y^*, z^*, \lambda^*). \quad (26)$$

由(25), (26)式可得: $\lim_{k_j \rightarrow +\infty} L_\beta(w^{k_{j+1}}) = L_\beta(w^*)$, 因而有 $\lim_{k_j \rightarrow +\infty} f(x^{k_{j+1}}) = f(x^*)$ 。然后再通过(4)式中 $k_j \rightarrow +\infty$, 则有:

$$\begin{cases} A^T \lambda^* \in \partial f(x^*) \\ \nabla g(y^*) = B^T \lambda^* \\ \nabla h(z^*) = -\lambda^* \\ Ax^* + By^* - z^* = 0 \end{cases}.$$

因此, 由定义 3 可知, $w^* \in \text{crit } L_\beta(\cdot)$, 即 $\Omega \subseteq \text{crit } L_\beta$ 。

3) 对任意点 $(x^*, y^*, z^*, \lambda^*) \in \Omega$, 则存在子序列 $\{w^{k_j}\}$ 收敛至 w^* , 又因为 $\lim_{j \rightarrow +\infty} L_\beta(w^{k_{j+1}}) = L_\beta(w^*)$, 且 $\{L_\beta(w^k)\}_{k \in \mathbb{N}}$ 是单调递减的, 则 $\lim_{k \rightarrow +\infty} L_\beta(w^k) = L_\beta(w^*)$, 所以 $L_\beta(\cdot)$ 在 Ω 上为常数。证毕

定理 2 设 $\{w^k := (x^k, y^k, z^k, \lambda^k)\}$ 为算法(3)产生的迭代序列并假定有界, $A^T A \geq \mu I$, $L_\beta(\cdot)$ 在 Ω 上满足 KL 性质, 则有:

$$1) \sum_{k=0}^{+\infty} \|w^{k+1} - w^k\| < +\infty;$$

2) $\{w^k\}$ 收敛至 $L_\beta(\cdot)$ 的一个稳定点。

证明 1) 由定理 1 可知, $\lim_{k \rightarrow +\infty} L_\beta(w^k) = L_\beta(w^*)$, 因而考虑两种情况。

第一种: 存在整数 k_0 , 使得 $L_\beta(w^{k_0}) = L_\beta(w^*)$, 由引理 3 可知, 对 $\forall k > k_0$ 有:

$$\delta \|v^{k+1} - v^k\|^2 + \|y^{k+1} - y^k\|_G^2 + \frac{1}{2} \|x^{k+1} - x^k\|_D^2 \leq L_\beta(w^k) - L_\beta(w^{k+1}) \leq L_\beta(w^{k_0}) - L_\beta(w^*) = 0,$$

所以对 $\forall k > 0, v^{k+1} = v^k, \|x^{k+1} - x^k\|_D^2 = 0, \|y^{k+1} - y^k\|_G^2 = 0$ 。

又因为 $\|\lambda^{k+1} - \lambda^k\| \leq L \|z^{k+1} - z^k\|$, 则 $\lambda^{k+1} = \lambda^k$ 。由 $\|\beta(Ax^k - Ax^{k+1})\|^2 \geq \beta^2 \mu \|x^{k+1} - x^k\|^2$ 和(20)式可知, 对任意 $k > k_0 + 1$, 有 $x^{k+1} = x^k$, 因此 $w^{k+1} = w^k$ 。

第二种: 假设 $L_\beta(w^k) > L_\beta(w^*)$, 因为 $d(w^k, \Omega) \rightarrow 0$, 则有对 $\forall \varepsilon > 0, \exists k_1 > 0$, s. t. $\forall k > k_1$, 有 $d(w^k, \Omega) < \varepsilon$;

由于 $L_\beta(w^k) \rightarrow L_\beta(w^*)$, 则 $\forall \eta > 0, \exists k_2 > 0$, s. t. $\forall k > k_2$, 有 $L_\beta(w^*) < L_\beta(w^k) < L_\beta(w^*) + \eta$, 因此对 $\forall \varepsilon, \eta > 0$, 当 $k > k' := \max\{k_1, k_2\}$ 时, 同时有:

$$d(\mathbf{w}^k, \Omega) < \epsilon, L_\beta(\mathbf{w}^*) < L_\beta(\mathbf{w}^k) < L_\beta(\mathbf{w}^*) + \eta.$$

因为 $L_\beta(\cdot)$ 在 Ω 上满足 KL 性质, 由引理 1 可知, $\varphi'(L_\beta(\mathbf{w}^k) - L_\beta(\mathbf{w}^*))d(\mathbf{0}, \partial L_\beta(\mathbf{w}^k)) \geq 1$ 。又因为:

$$L_\beta(\mathbf{w}^k) - L_\beta(\mathbf{w}^{k+1}) = (L_\beta(\mathbf{w}^k) - L_\beta(\mathbf{w}^*)) - (L_\beta(\mathbf{w}^{k+1}) - L_\beta(\mathbf{w}^*)),$$

再由 φ 的凹性, 有:

$$\varphi(L_\beta(\mathbf{w}^k) - L_\beta(\mathbf{w}^*)) - \varphi(L_\beta(\mathbf{w}^{k+1}) - L_\beta(\mathbf{w}^*)) \geq \varphi'(L_\beta(\mathbf{w}^k) - L_\beta(\mathbf{w}^*))(L_\beta(\mathbf{w}^k) - L_\beta(\mathbf{w}^{k+1})).$$

令 $\Delta_{p,q} = \varphi(L_\beta(\mathbf{w}^p) - L_\beta(\mathbf{w}^*)) - \varphi(L_\beta(\mathbf{w}^q) - L_\beta(\mathbf{w}^*))$, 且 $\varphi'(L_\beta(\mathbf{w}^k) - L_\beta(\mathbf{w}^*)) > 0$, 所以有:

$$L_\beta(\mathbf{w}^k) - L_\beta(\mathbf{w}^{k+1}) \leq \frac{\Delta_{k,k+1}}{\varphi'(L_\beta(\mathbf{w}^k) - L_\beta(\mathbf{w}^*))} \leq d(\mathbf{0}, \partial L_\beta(\mathbf{w}^k)) \Delta_{k,k+1}.$$

由引理 5 可知, $L_\beta(\mathbf{w}^k) - L_\beta(\mathbf{w}^{k+1}) \leq M(\|\mathbf{y}^k - \mathbf{y}^{k-1}\| + \|\mathbf{z}^k - \mathbf{z}^{k-1}\| + \|\mathbf{z}^{k-1} - \mathbf{z}^{k-2}\|) \Delta_{k,k+1}$, 再由引理 3 和上式, 对 $\forall k > k'$ 可得:

$$\begin{aligned} & \delta \|\mathbf{v}^{k+1} - \mathbf{v}^k\|^2 + \|\mathbf{y}^{k+1} - \mathbf{y}^k\|_G^2 + \frac{1}{2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_D^2 \leq \\ & M(\|\mathbf{y}^k - \mathbf{y}^{k-1}\| + \|\mathbf{z}^k - \mathbf{z}^{k-1}\| + \|\mathbf{z}^{k-1} - \mathbf{z}^{k-2}\|) \Delta_{k,k+1}. \end{aligned}$$

所以有:

$$\|\mathbf{v}^{k+1} - \mathbf{v}^k\| \leq \sqrt{\frac{M}{\delta}} (\|\mathbf{y}^k - \mathbf{y}^{k-1}\| + \|\mathbf{z}^k - \mathbf{z}^{k-1}\| + \|\mathbf{z}^{k-1} - \mathbf{z}^{k-2}\|)^{\frac{1}{2}} \Delta_{k,k+1}^{\frac{1}{2}}.$$

上式也等价于 $3\|\mathbf{v}^{k+1} - \mathbf{v}^k\| \leq 2(\|\mathbf{y}^k - \mathbf{y}^{k-1}\| + \|\mathbf{z}^k - \mathbf{z}^{k-1}\| + \|\mathbf{z}^{k-1} - \mathbf{z}^{k-2}\|)^{\frac{1}{2}} \left(\frac{3}{2} \sqrt{\frac{M}{\delta}} \Delta_{k,k+1}^{\frac{1}{2}} \right)$, 再由基本不等式, $\forall k > k'$ 有:

$$3\|\mathbf{v}^{k+1} - \mathbf{v}^k\| \leq (\|\mathbf{y}^k - \mathbf{y}^{k-1}\| + \|\mathbf{z}^k - \mathbf{z}^{k-1}\| + \|\mathbf{z}^{k-1} - \mathbf{z}^{k-2}\|) + \frac{9M}{4\delta} \Delta_{k,k+1}, \quad (27)$$

对(27)式取 $k = k' + 1, k' + 2, \dots, s$ 求和, 有:

$$3 \sum_{k=k'+1}^s \|\mathbf{v}^{k+1} - \mathbf{v}^k\| \leq \sum_{k=k'+1}^s (\|\mathbf{y}^k - \mathbf{y}^{k-1}\| + \|\mathbf{z}^k - \mathbf{z}^{k-1}\| + \|\mathbf{z}^{k-1} - \mathbf{z}^{k-2}\|) + \frac{9M}{4\delta} \Delta_{k'+1,s+1}, \quad (28)$$

对(28)式, 再由柯西不等式可得:

$$\begin{aligned} & \frac{3\sqrt{2}}{2} \sum_{k=k'+1}^s \|\mathbf{y}^{k+1} - \mathbf{y}^k\| + \frac{3\sqrt{2}}{2} \sum_{k=k'+1}^s \|\mathbf{z}^{k+1} - \mathbf{z}^k\| \leq \\ & \sum_{k=k'+1}^s (\|\mathbf{y}^k - \mathbf{y}^{k-1}\| + \|\mathbf{z}^k - \mathbf{z}^{k-1}\| + \|\mathbf{z}^{k-1} - \mathbf{z}^{k-2}\|) + \frac{9M}{4\delta} \Delta_{k'+1,s+1}, \end{aligned}$$

结合 $\varphi(L_\beta(\mathbf{w}^{s+1}) - L_\beta(\mathbf{w}^*)) > 1$, 对上式取极限, 令 $s \rightarrow +\infty$, 有:

$$\begin{aligned} & \left(\frac{3\sqrt{2}}{2} - 1 \right) \sum_{k=k'+1}^{+\infty} \|\mathbf{y}^{k+1} - \mathbf{y}^k\| + \left(\frac{3\sqrt{2}}{2} - 2 \right) \sum_{k=k'+1}^s \|\mathbf{z}^{k+1} - \mathbf{z}^k\| \leq \\ & \|\mathbf{y}^{k'+1} - \mathbf{y}^{k'}\| + 2\|\mathbf{z}^{k'+1} - \mathbf{z}^{k'}\| + \|\mathbf{z}^{k'} - \mathbf{z}^{k'-1}\| + \frac{9M}{4\delta} \varphi(L_\beta(\mathbf{w}^{k'+1}) - L_\beta(\mathbf{w}^*)). \end{aligned}$$

因此, $\sum_{k=0}^{+\infty} \|\mathbf{y}^{k+1} - \mathbf{y}^k\| < +\infty$, $\sum_{k=0}^{+\infty} \|\mathbf{z}^{k+1} - \mathbf{z}^k\| < +\infty$ 。

因为 $\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\| \leq L\|\mathbf{z}^{k+1} - \mathbf{z}^k\|$, 则有: $\sum_{k=0}^{+\infty} \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\| < +\infty$ 。又因为:

$$\begin{aligned} \|\mathbf{x}^{k+1} - \mathbf{x}^k\| & \leq \frac{2}{\beta} \sqrt{\frac{1}{\mu}} (\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 + \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k-1}\|^2 + \beta^2 \|\mathbf{B}\|^2 \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 + \beta^2 \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2)^{\frac{1}{2}} \leq \\ & \frac{2}{\beta} \sqrt{\frac{1}{\mu}} (\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\| + \|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k-1}\| + \beta \|\mathbf{B}\| \|\mathbf{y}^{k+1} - \mathbf{y}^k\| + \beta \|\mathbf{z}^{k+1} - \mathbf{z}^k\|), \end{aligned}$$

因此 $\sum_{k=1}^{+\infty} \|\mathbf{x}^{k+1} - \mathbf{x}^k\| < +\infty$ 。由于

$$\|\mathbf{w}^{k+1} - \mathbf{w}^k\| \leq \|\mathbf{x}^{k+1} - \mathbf{x}^k\| + \|\mathbf{y}^{k+1} - \mathbf{y}^k\| + \|\mathbf{z}^{k+1} - \mathbf{z}^k\| + \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|,$$

所以有: $\sum_{k=0}^{+\infty} \|\mathbf{w}^{k+1} - \mathbf{w}^k\| < +\infty$ 。

2) 由结论 1) 可知 $\{w^k\}$ 是一个柯西列, 因此它是收敛的, 由定理 1 的结论 2) 中可知, $\Omega \subseteq \text{crit } L_\beta$, 因而 $\{w^k\}$ 收敛至 $L_\beta(\cdot)$ 的一个稳定点。证毕

引理 6 设 $\{w^k := (x^k, y^k, z^k, \lambda^k)\}$ 为算法 (3) 产生的迭代序列, 假设: $\inf_z \left\{ h(z) - \frac{1}{4L} \|\nabla h(z)\|^2 \right\} = \bar{h} > -\infty$, $\liminf_{\|x\| \rightarrow +\infty} f(x) = +\infty$, $\liminf_{\|y\| \rightarrow +\infty} g(y) = +\infty$, 则 $\{w^k\}$ 有界。

证明 由引理 3 可知, 对 $\forall k \geq 1$, 有 $L_\beta(x^k, y^k, z^k, \lambda^k) \leq L_\beta(x^1, y^1, z^1, \lambda^1) < +\infty$, 结合 $\nabla h(z^k) = -\lambda^k$, 有:

$$\begin{aligned} L_\beta(x^1, y^1, z^1, \lambda^1) &\geq f(x^k) + g(y^k) + h(z^k) - \langle \lambda^k, Ax^k + By^k - z^k \rangle + \frac{\beta}{2} \|Ax^k + By^k - z^k\|^2 = \\ &f(x^k) + g(y^k) + h(z^k) - \frac{1}{2\beta} \|\lambda^k\|^2 + \frac{\beta}{2} \left\| Ax^k + By^k - z^k - \frac{1}{\beta} \lambda^k \right\|^2 \geq \\ &f(x^k) + g(y^k) + \bar{h} + \left(\frac{1}{4L} - \frac{1}{2\beta} \right) \|\lambda^k\|^2 + \frac{\beta}{2} \left\| Ax^k + By^k - z^k - \frac{1}{\beta} \lambda^k \right\|^2. \end{aligned}$$

由假设条件可知, $\inf_x f(x) > -\infty$, $\inf_y g(y) > -\infty$ 。由于 $\beta > 2L$, 则对 $\forall k \geq 1$, 有 $\{x^k\}$, $\{y^k\}$, $\{\lambda^k\}$ 和 $\left\{ \frac{\beta}{2} \left\| Ax^k + By^k - z^k - \frac{1}{\beta} \lambda^k \right\|^2 \right\}$ 均有界, 因此 $\{z^k\}$ 有界, 那么 $\{w^k\}$ 有界。证毕

3 数值实验

为了检验此算法的有效性, 考虑压缩感知中的优化问题:

$$\begin{aligned} \min & \|x\|_0, \\ \text{s. t. } & Px = b. \end{aligned}$$

其中: $P \in \mathbf{R}^{n \times m}$ 是测量矩阵, $b \in \mathbf{R}^n$ 为观测数据, $\|x\|_0$ 表示 x 中非零元素的个数。

针对上述问题, 可将问题中的 l_0 范数松弛到 l_1 范数和 $l_{1/2}$ 范数, 则可分别导出如下凸优化问题和非凸优化问题:

$$\begin{aligned} \min & \gamma \|x\|_1 + \|y\|^2, \\ \text{s. t. } & Px - y = b. \end{aligned} \quad (29)$$

以及

$$\begin{aligned} \min & \gamma \|x\|_{\frac{1}{2}} + \|y\|^2, \\ \text{s. t. } & Px - y = b. \end{aligned} \quad (30)$$

(29), (30) 式这两个优化问题可推广为如下三块优化问题:

$$\begin{aligned} \min & \gamma \|x\|_1 + \|y\|^2 + \|z\|^2, \\ \text{s. t. } & Px + By - z = b. \end{aligned} \quad (31)$$

以及

$$\begin{aligned} \min & \gamma \|x\|_{\frac{1}{2}} + \|y\|^2 + \|z\|^2, \\ \text{s. t. } & Px + By - z = b. \end{aligned} \quad (32)$$

其中: $B \in \mathbf{R}^{n \times n}$, 且负单位矩阵, $b = 0$ 。将算法 (3) 应用到问题 (27) 式, 取 $D = \alpha I - \beta P^T P$, $G = \alpha I - \beta B^T B$, 有:

$$\begin{cases} x^{k+1} \in S\left(x^k - \frac{\beta}{\alpha} P^T P x^k + \frac{1}{\alpha} P^T (\beta y^k + \beta z^k + \lambda^k); \frac{\gamma}{\alpha}\right) \\ y^{k+1} = ((2 + \alpha)I)^{-1} ((\lambda^k B^T - \beta B^T (P x^{k+1} - z^k)) + (\alpha I - \beta B^T B) y^k) \\ z^{k+1} = \frac{\beta (P x^{k+1} + B y^{k+1}) - \lambda^k}{2 + \beta} \\ \lambda^{k+1} = \lambda^k - \beta (P x^{k+1} - y^{k+1} - z^{k+1}) \end{cases}, \quad (33)$$

其中: $S(\cdot, u)$ 为软阈值算子^[11], 定义为:

$$S(x, u) = \begin{cases} x + u, & x \leq -u \\ 0, & |x| < u \\ x - u, & x \geq u \end{cases}.$$

将算法(3)应用到问题(32)式,取定同样的矩阵 $\mathbf{D}, \mathbf{G}$, 有:

$$\begin{cases} \mathbf{x}^{k+1} \in H\left(\mathbf{x}^k - \frac{\beta}{\alpha} \mathbf{P}^T \mathbf{P} \mathbf{x}^k + \frac{1}{\alpha} \mathbf{P}^T (\beta \mathbf{y}^k + \beta \mathbf{z}^k + \boldsymbol{\lambda}^k); \frac{2\gamma}{\alpha}\right) \\ \mathbf{y}^{k+1} = ((2+\alpha)\mathbf{I})^{-1} ((\boldsymbol{\lambda}^k \mathbf{B}^T - \beta \mathbf{B}^T (\mathbf{P} \mathbf{x}^{k+1} - \mathbf{z}^k) + (\alpha \mathbf{I} - \beta \mathbf{B}^T \mathbf{B}) \mathbf{y}^k) \\ \mathbf{z}^{k+1} = \frac{\beta (\mathbf{P} \mathbf{x}^{k+1} + \mathbf{B} \mathbf{y}^{k+1}) - \boldsymbol{\lambda}^k}{2+\beta} \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \beta (\mathbf{P} \mathbf{x}^{k+1} - \mathbf{y}^{k+1} - \mathbf{z}^{k+1}) \end{cases}, \quad (34)$$

其中: $H(\cdot, u) = \{h_u(x_1), h_u(x_2), \dots, h_u(x_n)\}$ 为半阈值算子^[12], 定义如下:

$$h_u(x_i) = \begin{cases} \frac{2x_i}{3} \left(1 + \cos \frac{2}{3}(\pi - \varphi(|x_i|))\right), & |x_i| > \frac{\sqrt[3]{54}}{4} u^{\frac{2}{3}}, \\ 0, & \text{其他} \end{cases}$$

其中: $\varphi(x_i) = \arccos\left(\frac{u}{8} \left(\frac{x_i}{3}\right)^{-\frac{3}{2}}\right)$ 。

为了简便, 分别记 RSADMM 和 RHADMM 简记算法为(33), (34)式。下面给出数值实验中相关参数的选取, $n=500, m=600$, 矩阵 $\mathbf{P} \in \mathbf{R}^{500 \times 600}$ 随机生成服从 $N(0, 1)$ 正态分布按列单位化所得, 正则化参数 $\gamma=0.0015$, $\beta=0.6, \alpha=2.5$ 。如图 1~3 所示, 由均方来度量最优解与迭代点的关系, 即 $\text{MSE}(\|\mathbf{x}^* - \mathbf{x}^k\|) = \frac{1}{m} \|\mathbf{x}^* - \mathbf{x}^k\|$, $\text{MSE}(\|\mathbf{y}^* - \mathbf{y}^k\|) = \frac{1}{n} \|\mathbf{y}^* - \mathbf{y}^k\|$, $\text{MSE}(\|\mathbf{z}^* - \mathbf{z}^k\|) = \frac{1}{n} \|\mathbf{z}^* - \mathbf{z}^k\|$, 其中 $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) = (\mathbf{0}, \mathbf{0}, \mathbf{0})$ 为(31), (32)式的最优解。

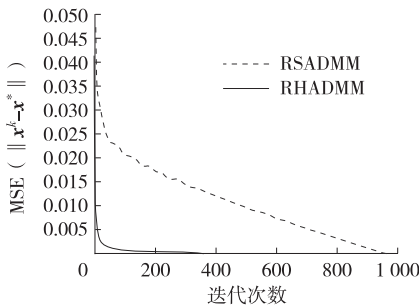


图 1 x 子问题的执行情况

Fig. 1 Implementation of the x sub-problem

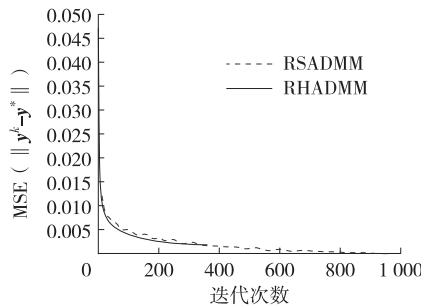


图 2 y 子问题的执行情况

Fig. 2 Implementation of the y sub-problem

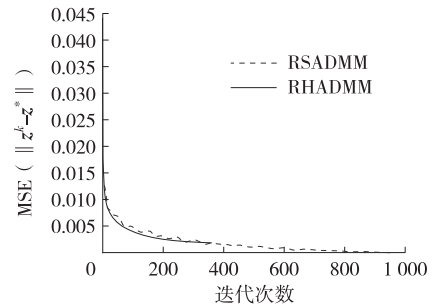


图 3 z 子问题的执行情况

Fig. 3 Implementation of the z sub-problem

两个算法的程序在 Matlab R2018a 上运行, 运行的 PC 配置为 Windows 7 系统, Inter(R) Core(TM) i5-4210U CPU 2.40 GHz, 4.00 GB 内存。数值算例的结果说明, 算法(3)是稳定有效的, 3 个序列都较好的接近到真正的解, 并且可以得到 RHADMM 的收敛速度明显比 RSADMM 更快。

4 小结

本文针对一类三块非凸优化问题, 提出了一种正则化交替方向法, 在增广拉格朗日函数满足 KL 性质且惩罚参数充分大的条件下, 由该算法生成的迭代序列的任何聚点都是增广拉格朗日函数的稳定点, 最后, 由数值算例, 验证了算法的有效性。

参考文献:

- [1] MONTEIRO, RENATO D C, SVAITER B F. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers[J]. SIAM Journal on Optimization, 2013, 23(1): 475-507.
- [2] HE B S, YUAN X M. On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers[J]. Numerische Mathematik, 2015, 130(3): 567-577.
- [3] HAN D R, SUN D, ZHANG L. Linear rate convergence of the alternating direction method of multipliers for convex composite

- quadratic and semi-definite programming[J]. Automatic Control IEEE Transactions on, 2015, 60(3): 644-658.
- [4] SHEFI R, TEBoulLE M. Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization[J]. SIAM Journal on Optimization, 24(24): 269-297, 2014.
- [5] GUO K, HAN D R, WANG Z W, WU T T. Convergence of ADMM for multi-block nonconvex separable optimization models [J]. Frontiers of Mathematics in China, 2017, 12(5): 1139-1162.
- [6] JIAN J B, ZHANG Y, CHAO M T. A regularized alternating direction method of multipliers for a class of nonconvex problems [J]. Journal of Inequalities and Applications, 2019, 2019(1): 1-16.
- [7] ROCKAFELLAR R T, WETS R J B. Variational analysis[M]. Berlin: Springer, 1998.
- [8] ATTOUCH H, BOLTE J, REDONT P, SOUBEYRAN A. Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality[J]. Mathematics of Operations Research, 2010, 35(2): 438-457.
- [9] ATTOUCH H, BOLTE J, SVAITER B F. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods[J]. Mathematical Programming, 2013, 137(1/2): 91-129.
- [10] NESTEROV Y. Introductory lectures on convex optimization: a basic course[M]. New York: Springer, 2004.
- [11] DAUBECHIES I, DEFRISE M, DE M C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint[J]. Communications on Pure and Applied Mathematics, 2004, 57(11): 1413-1457.
- [12] XU Z B, CHANG X, XU F, ZHANG H. L-1/2 regularization: a thresholding representation theory and a fast solver[J]. IEEE Transactions on Neural Networks and Learning Systems, 2012, 23(7): 1013-1027.

Operations Research and Cybernetics

Convergence of Regularized Alternating Direction Method of Multipliers for Three-Block Nonconvex Optimization Problems

LEI Hongwang, PENG Jianwen

(School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China)

Abstract: [Purposes] For a class of three-block nonconvex optimization problems, a regularized alternating direction method of multipliers is proposed. [Methods] In order to find the unique point $(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1})$ more easily, in the framework of the original multiplier alternating direction method, a square neighbor is added to the $\mathbf{x}$ -subproblem and the $\mathbf{y}$ -subproblem to regularize the original subproblem. [Findings] Under the condition that the augmented Lagrangian function satisfies the K-L property, any convergence point of the iterative sequence generated by the algorithm is a critical point. [Conclusions] Numerical experimental result shows that this method is efficient.

Keywords: Alternating direction method of multipliers; K-L property; critical point

(责任编辑 陈 乔)