

DOI:10.13718/j.cnki.xsxb.2022.10.002

求解非凸优化问题的近似交替方向乘子法^①

谭秋芬, 罗洪林

重庆师范大学 数学科学学院, 重庆 401331

摘要: 考虑有界约束上具有可分结构的非凸优化问题, 提出了一种基于 ADMM 的新算法 P-ADMM (即近似 ADMM). 在基于 ADMM 的框架下, P-ADMM 在解决有界约束上的非凸子问题时, 采用梯度投影, 以此简化非凸子问题的求解, 降低运算成本, 并且通过引入一个“平滑的”(即指数加权)原始迭代序列, 在每次迭代时, 向增广拉格朗日函数中增加一个以平滑的原始迭代为中心的近似二次项, 使所得到的近似增广拉格朗日函数在每次迭代时被不精确地最小化, 在保证算法收敛性的同时也能够提升算法的收敛速度. 数值实验表明, 该算法可有效应用于求解一类非凸的船舶分布式能源管理问题.

关键词: 非凸优化; 近似 ADMM; 二次近似项; 梯度投影

中图分类号: O221.2

文献标志码: A

文章编号: 1000-5471(2022)10-0007-12

Approximate ADMM for Non-Convex Optimization Problems

TAN Qiufen, LUO Honglin

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

Abstract: In this paper, a new algorithm P-ADMM (approximate ADMM) based on ADMM is proposed for non-convex optimization problems with separable structures under bounded constraints. In the framework based on ADMM, P-ADMM adopts gradient projection to solve non-convex sub-problems with bounded constraints, so as to simplify the solving of non-convex sub-problems and reduce the cost of operation. Moreover, by introducing a “smooth” (exponential weighted) original iteration sequence, and at each iteration, by adding an approximate quadratic term centered on smooth original iteration to the augmented Lagrange function, the obtained approximate augmented Lagrange function is inaccurately minimized in each iteration, which can not only ensure the convergence of the algorithm but also improve the convergence speed of the algorithm. Numerical experiments show that this algorithm can be effectively applied to a class of non-convex ship distributed energy management problem.

Key words: non-convex optimization; approximate ADMM; quadratic approximate term; gradient projection

① 收稿日期: 2021-11-03

基金项目: 国家自然科学基金项目(11991024, 11771064); 重庆市高校创新研究群体项目(CXQT20014); 重庆市自然科学基金项目(cstc2021jcyj-msx300).

作者简介: 谭秋芬, 硕士研究生, 主要从事最优化理论与算法的研究.

通信作者: 罗洪林, 教授.

本文考虑以下线性约束优化问题

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

其中 $f(\mathbf{x})$ 是 $\mathbb{R}^{n_1} \rightarrow \mathbb{R}$ 上的非凸可微函数且

$$P = \{\mathbf{x} \in \mathbb{R}^{n_1} \mid l_i \leq x_i \leq u_i, i = 1, 2, \dots, n_1\}$$

$g(\mathbf{y})$ 是 $\mathbb{R}^{n_2} \rightarrow \mathbb{R}$ 上的凸可微函数, 矩阵 $\mathbf{A} \in \mathbb{R}^{m \times n_1}$, $\mathbf{B} \in \mathbb{R}^{m \times n_2}$. 问题(1) 有着广泛的应用^[1-4].

当问题(1) 的目标函数是凸可分时, 交替方向乘子法(ADMM) 是解决这类问题较为有效的方法^[5-8]. 若目标函数非凸, 则通过一些假设条件, ADMM 仍然可以保持较好的适用性^[9-14]. 其中, 文献[10-11] 针对非凸问题提出了 Bregman ADMM 并证明了其收敛性, 该算法在分别求解子问题时, 在原本的增广拉格朗日函数上增加了一个 Bregman 距离函数, 通过将非凸子问题进行一个凸化处理来简化原来的子问题, 进而对其进行求解. 文献[13] 提出了非凸 Proximal ADMM, 该算法在子问题中引入了近端项, 将子问题进行凸化处理, 进而使用 ADMM 进行求解, 而文献[15] 在处理非凸问题上考虑的是近似二次项, 近似二次项在保证问题凸化的同时还能使得迭代点不会偏离稳定点太多, 这为凸化处理提供了更好的方向.

当考虑问题(1) 中的有界约束时, 使用 ADMM 直接解决是比较困难的. 幸运的是, 文献[15] 在处理有界多面体上的非凸优化问题时引入了梯度投影算法, 对含有有界约束的单块优化问题进行了求解并取得了较好的结果. 受以上文献启发, 在借鉴梯度投影和交替方向乘子法(ADMM) 两种经典方法思想的基础上, 本文针对具有可分结构的非凸优化问题提出了一种新的近似交替方向乘子法(P-ADMM), 基于 ADMM 的框架在解决有界多面体上的非凸子问题时, 采用梯度投影, 以此简化非凸子问题的求解, 降低运算成本, 并且通过引入一个“平滑的”(即指数加权) 原始迭代序列, 并在每次迭代时, 向增广拉格朗日函数中增加一个以平滑的原始迭代为中心的近似二次项, 使所得到的近似增广拉格朗日函数在每次迭代时被不精确地最小化, 在保证算法的收敛性的同时也能够提升算法的收敛速度. 该算法可有效应用于求解一类非凸的船舶分布式能源管理问题.

1 预备知识

1.1 论文的符号定义

- 1) $\mathbb{R}^{n_1}$ 为 n_1 维欧氏空间, $\mathbb{R}^{n_2}$ 为 n_2 维欧氏空间.
- 2) $\mathbb{R}^{m \times n_1}$ 是 $m \times n_1$ 维实矩阵空间, $\mathbb{R}^{m \times n_2}$ 是 $m \times n_2$ 维实矩阵空间.
- 3) $[\cdot]_+$ 是约束 P 上的投影算子, 满足 $[\mathbf{x}]_+ = \operatorname{argmin}_{\bar{\mathbf{x}} \in P} \|\mathbf{x} - \bar{\mathbf{x}}\|$, 特别的, $[\mathbf{x}]_+$ 的第 i 个坐标分量满足

$$([\mathbf{x}]_+)_i = \begin{cases} l_i, & x_i < l_i \\ x_i, & x_i \in [l_i, u_i], \text{ 其中 } i \in \{1, 2, \dots, n_1\} \\ u_i, & x_i > u_i \end{cases}$$

- 4) 令 $\boldsymbol{\omega} = (\mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\lambda})$, $\boldsymbol{\omega}^t = (\mathbf{x}^t, \mathbf{y}^t, \mathbf{z}^t, \boldsymbol{\lambda}^t)$, $\tilde{\boldsymbol{\omega}} = (\mathbf{x}, \mathbf{y}, \tilde{\mathbf{y}}, \mathbf{z}, \boldsymbol{\lambda})$, $\tilde{\boldsymbol{\omega}}^t = (\mathbf{x}^t, \mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^t, \boldsymbol{\lambda}^t)$.

1.2 稳定解集合

原问题(1) 的 KKT 条件如下

$$\begin{aligned} \nabla f(\mathbf{x}^*) + \mathbf{A}^T \boldsymbol{\lambda}^* - \boldsymbol{\mu}^* + \mathbf{v}^* &= \mathbf{0} \\ \nabla g(\mathbf{y}^*) + \mathbf{B}^T \boldsymbol{\lambda}^* &= \mathbf{0} \\ \mathbf{Ax}^* + \mathbf{By}^* &= \mathbf{b} \\ \boldsymbol{\mu}^* &\geq \mathbf{0}, \mathbf{v}^* \geq \mathbf{0} \\ l_i &\leq x_i^* \leq u_i, \forall i \in \{1, 2, \dots, n_1\} \\ \mu_i^* (l_i - x_i^*) &= 0, \forall i \in \{1, 2, \dots, n_1\} \end{aligned}$$

$$\nu_i^*(x_i^* - u_i) = 0, \forall i \in \{1, 2, \dots, n_1\}$$

其中: $\lambda^*, \mu^*, \nu^*, \lambda^*$ 为乘子; μ_i^* 和 ν_i^* 表示第 i 个分量. 令 $X^* \times Y^*$ 为满足 KKT 条件的所有 (x^*, y^*) 所构成的集合, 称 $X^* \times Y^*$ 为原问题(1)的稳定解集, 并令 $W = \{(x, y, \lambda) \mid (x, y), \lambda \text{ 为一对原始对偶解}\}$.

1.3 基本假设

(I) $BB^T \geq \mu_0 I$, 其中 $\mu_0 > 0$, 即 B 是列满秩的.

(II) $f(x)$ 是可微的, 且满足 Lipschitz 可微条件

$$\|\nabla f(x_1) - \nabla f(x_2)\| \leq L_1 \|x_1 - x_2\|, L_1 > 0, \forall x_1, x_2 \in P$$

则可得, 存在 $\gamma < 0$, 使得

$$\langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2 \rangle \geq \gamma \|x_1 - x_2\|^2$$

(III) $g(y)$ 是强凸的, 强凸系数为 m , 即

$$g(y_1) - g(y_2) \geq \langle \nabla g(y_2), y_1 - y_2 \rangle + \frac{m}{2} \|y_1 - y_2\|^2, \forall y_1, y_2 \in \mathbb{R}^{n_2}$$

(IV) $g(y)$ 是可微的, 且满足 Lipschitz 可微条件

$$\|\nabla g(y_1) - \nabla g(y_2)\| \leq L_2 \|y_1 - y_2\|, L_2 > 0, \forall y_1, y_2 \in \mathbb{R}^{n_2}$$

(V) 存在常数 $\tilde{\sigma} > 0$, 使得

$$\inf_y \{g(y) - \tilde{\sigma} \|\nabla g(y)\|^2\} > -\infty, \forall y \in \mathbb{R}^{n_2}$$

(VI) 函数 $g(y)$ 是强制的, 即

$$\lim_{\|y\| \rightarrow +\infty} g(y) = +\infty, \forall y \in \mathbb{R}^{n_2}$$

1.4 一种近似的 ADMM

接下来介绍近似的交替方向乘子法(P-ADMM), 问题(1)的增广拉格朗日函数为

$$L(x, y, \lambda) = f(x) + g(y) + \langle \lambda, Ax + By - b \rangle + \frac{\alpha}{2} \|Ax + By - b\|^2$$

其中常数 $\alpha > 0$. 经典的 ADMM 算法处理以下两个子问题: 对于固定的 y, λ , 在有界约束 P 上极小化 $L(x, y, \lambda)$; 对于固定的 x, λ , 极小化 $L(x, y, \lambda)$. 最后, 利用原始残差 $Ax + By = b$ 更新乘子 λ . 由于有界约束的特殊性, 在求解有界非凸子问题上采用梯度投影, 且考虑到 $f(x)$ 是非凸的, 通过引入一个指数平均(或平滑)方案来生成一个额外的序列 $\{z^t\}$, 并在增广的拉格朗日函数中插入一个额外的以 z^t 为中心的二次近似项, 在保证非凸子问题凸化的同时也能够使得原始迭代点 x^t 不会偏离稳定迭代点太多.

表 1 近似交替方向乘子法

算法 P-ADMM

步骤 1. 设置参数 $\alpha > \max\left\{\frac{2L_2^2}{\mu_0(m-2\eta)}, \frac{1}{\mu_0\tilde{\sigma}}\right\} > 0, 0 < \beta \leq 1, \frac{1}{L_k} > 2c > 0, p > 0, \xi > 0$;

步骤 2. 输入初值 $x^0 \in P, y^0 \in \mathbb{R}^{n_2}, z^0 \in P, \lambda^0 \in \mathbb{R}^m$;

步骤 3. 令 $t = 0$

步骤 4. $x^{t+1} = \left[x^t - c \nabla_x \left(L(x^t, y^t, \lambda^t) + \frac{p}{2} \|x^t - z^t\|^2 \right) \right]_+$;

步骤 5. $z^{t+1} = z^t + \beta(x^{t+1} - z^t)$;

步骤 6. $y^{t+1} = \operatorname{argmin}_{y \in \mathbb{R}^{n_2}} L(x^{t+1}, y, \lambda^t)$;

步骤 7. $\lambda^{t+1} = \lambda^t + \alpha(Ax^{t+1} + By^{t+1} - b)$;

步骤 8. 若满足终止条件 $\max\{\|x^t - x^{t+1}\|, \|y^t - y^{t+1}\|, \|\lambda^t - \lambda^{t+1}\|\} \leq \xi$, 算法终止. 否则, 令 $t = t + 1$, 返回步骤 4.

注 1 该算法与传统的 ADMM 相比的优势在于, 传统的 ADMM 处理有界约束上的非凸子问题是十分困难的, 一方面是子问题的非凸性, 另一方面则是有界约束的限制, 而采用梯度投影、引入近似二次项则可以简化子问题的求解, 很好地处理了非凸性和有界约束, 使得算法更加简便, 降低了运算成本.

注 2 与文献[15] 算法相比, P-ADMM 是投影算法和 ADMM 的结合, 主要用于解决有界约束上的可分离的非凸优化问题, 而文献[15] 中的算法主要用于解决有界约束上的单块非凸优化模型, 是本文模型的一种特殊情况, 因此 P-ADMM 的应用性更加广泛.

注 3 与文献[16] 中算法 APGMM 相比, P-ADMM 在解决有界约束上的非凸子问题时不是直接采用的梯度投影, 而是在原本的增广拉格朗日函数上引入了近似二次项, 近似二次项的引入在保证非凸子问题凸化的同时也加快了算法收敛的速度(见数值实验).

2 收敛性分析

为了方便分析 P-ADMM 的收敛性, 首先给出一些关键的引理.

引理 1 若假设(I), (IV) 成立, 那么可得

$$\|\lambda^{t+1} - \lambda^t\|^2 \leq \frac{L_2^2}{\mu_0} \|y^{t+1} - y^t\|^2$$

证 首先, 通过假设(I) 可得

$$\begin{aligned} \|\mathbf{B}^T(\lambda^{t+1} - \lambda^t)\|^2 &= \langle \mathbf{B}^T(\lambda^{t+1} - \lambda^t), \mathbf{B}^T(\lambda^{t+1} - \lambda^t) \rangle = \\ &= \langle \mathbf{B}\mathbf{B}^T(\lambda^{t+1} - \lambda^t), (\lambda^{t+1} - \lambda^t) \rangle \geq \\ &= \mu_0 \|\lambda^{t+1} - \lambda^t\|^2 \end{aligned} \quad (2)$$

利用算法关于 y 在 $t+1$ 步迭代式的一阶最优性条件可得

$$\nabla g(y^{t+1}) + \langle \mathbf{B}, \lambda^t \rangle + \alpha \langle \mathbf{B}, \mathbf{A}x^{t+1} + \mathbf{B}y^{t+1} - \mathbf{b} \rangle = 0$$

再通过算法关于 λ 在 $t+1$ 步的迭代式可得 $\nabla g(y^{t+1}) = -\langle \mathbf{B}, \lambda^{t+1} \rangle$, 同理可得

$$\nabla g(y^t) = -\langle \mathbf{B}, \lambda^t \rangle$$

通过假设(IV) 可知, 函数 $g(x)$ 是梯度 Lipschitz 连续的, 则

$$\|\nabla g(y^{t+1}) - \nabla g(y^t)\| = \|\mathbf{B}^T(\lambda^{t+1} - \lambda^t)\| \leq L_2 \|y^{t+1} - y^t\| \quad (3)$$

结合(2), (3) 式可得 $\|\lambda^{t+1} - \lambda^t\|^2 \leq \frac{L_2^2}{\mu_0} \|y^{t+1} - y^t\|^2$, 故引理 1 成立.

引理 2 若假设(II) 成立, 且 $p + \gamma > 0$, 令函数

$$K(x, y, z, \lambda) = f(x) + g(y) + \langle \lambda, \mathbf{A}x + \mathbf{B}y - \mathbf{b} \rangle + \frac{\alpha}{2} \|\mathbf{A}x + \mathbf{B}y - \mathbf{b}\|^2 + \frac{p}{2} \|x - z\|^2$$

则下列结论成立:

1) 函数 $K(x, y, z, \lambda)$ 关于 x 是强凸的, 且强凸系数为 $\gamma_k = p + \gamma$, 即

$$\langle \nabla K_x(x_1, y, z, \lambda) - \nabla K_x(x_2, y, z, \lambda), x_1 - x_2 \rangle \geq (p + \gamma) \|x_1 - x_2\|^2$$

2) 函数 $K(x, y, z, \lambda)$ 关于 x 是梯度 Lipschitz 连续的, 且 Lipschitz 系数为 $L_k = p + \gamma + \alpha\sigma_1^2$, 即

$$\|\nabla K_x(x_1, y, z, \lambda) - \nabla K_x(x_2, y, z, \lambda)\| \leq (p + \gamma + \alpha\sigma_1^2) \|x_1 - x_2\|$$

其中 σ_1 为矩阵 $\mathbf{A}$ 的最大奇异值.

证 对函数 $K(x, y, z, \lambda)$ 关于 x 求梯度, 则

$$\nabla K_x(x, y, z, \lambda) = \nabla f(x) + \mathbf{A}^T \lambda + \alpha \mathbf{A}^T (\mathbf{A}x + \mathbf{B}y - \mathbf{b}) + p(x - z)$$

由假设(II) 知

$$\begin{aligned} \langle \nabla K_x(x_1, y, z, \lambda) - \nabla K_x(x_2, y, z, \lambda), x_1 - x_2 \rangle &= \\ \langle \nabla f(x_1) - \nabla f(x_2) + \alpha \mathbf{A}^T \mathbf{A}(x_1 - x_2) + p(x_1 - x_2), x_1 - x_2 \rangle &= \end{aligned}$$

$$\langle \nabla f(\mathbf{x}_1) - \nabla f(\mathbf{x}_2), \mathbf{x}_1 - \mathbf{x}_2 \rangle + p \|\mathbf{x}_1 - \mathbf{x}_2\|^2 + \alpha \langle \mathbf{A}^\top \mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2), \mathbf{x}_1 - \mathbf{x}_2 \rangle \geq (p + \gamma) \|\mathbf{x}_1 - \mathbf{x}_2\|^2$$

同理可得

$$\begin{aligned} & \|\nabla K_x(\mathbf{x}_1, \mathbf{y}, \mathbf{z}, \boldsymbol{\lambda}) - \nabla K_x(\mathbf{x}_2, \mathbf{y}, \mathbf{z}, \boldsymbol{\lambda})\| = \\ & \|\nabla f(\mathbf{x}_1) - \nabla f(\mathbf{x}_2) + \alpha \mathbf{A}^\top \mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2) + p(\mathbf{x}_1 - \mathbf{x}_2)\| \leq \\ & \|\nabla f(\mathbf{x}_1) - \nabla f(\mathbf{x}_2)\| + \alpha \|\mathbf{A}^\top \mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2)\| + p \|\mathbf{x}_1 - \mathbf{x}_2\| \leq \\ & (p + \gamma + \alpha \sigma_1^2) \|\mathbf{x}_1 - \mathbf{x}_2\| \end{aligned}$$

则引理 2 成立.

接下来, 构造辅助函数

$$\begin{aligned} M(\mathbf{x}, \mathbf{y}, \tilde{\mathbf{y}}, \mathbf{z}, \boldsymbol{\lambda}) = & f(\mathbf{x}) + g(\mathbf{y}) + \langle \boldsymbol{\lambda}, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} - \mathbf{b} \rangle + \frac{\alpha}{2} \|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} - \mathbf{b}\|^2 + \\ & \frac{p}{2} \|\mathbf{x} - \mathbf{z}\|^2 + \eta \|\mathbf{y} - \tilde{\mathbf{y}}\|^2 \end{aligned}$$

其中 $\eta > 0$, 证明函数列 $M(\tilde{\boldsymbol{\omega}}^t)$ 的单调递减性和有界性.

引理 3 若假设(I), (III) 和(IV) 成立, $\tilde{\boldsymbol{\omega}}^t = (\mathbf{x}^t, \mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^t, \boldsymbol{\lambda}^t)$ 为 P-ADMM 所生成的点列, 那么对任意的 t , 若 $2c < \frac{1}{L_k}$, $\alpha > \max\left\{\frac{2L_2^2}{\mu_0(m-2\eta)}, \frac{1}{\mu_0\tilde{\sigma}}\right\} > 0$, 有

$$\begin{aligned} & M(\mathbf{x}^t, \mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^t, \boldsymbol{\lambda}^t) - M(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}, \mathbf{y}^t, \mathbf{z}^{t+1}, \boldsymbol{\lambda}^{t+1}) \geq \\ & \frac{1}{2c} \|\mathbf{x}^t - \mathbf{x}^{t+1}\|^2 + \left(\frac{m}{2} - \eta - \frac{L_2^2}{\mu_0\alpha}\right) \|\mathbf{y}^t - \mathbf{y}^{t+1}\|^2 + \eta \|\mathbf{y}^t - \mathbf{y}^{t-1}\|^2 + \frac{p}{2\beta} \|\mathbf{z}^t - \mathbf{z}^{t+1}\|^2 \end{aligned}$$

注 4 当矩阵 $\mathbf{B}$ 为对称矩阵, 即 $\mathbf{B}^\top = \mathbf{B}$, 且满足 $2c < \frac{1}{L_k}$, $\alpha > \max\left\{\eta + \sqrt{\frac{2L_2^2}{\mu_0} + \eta^2}, \frac{1}{\mu_0\tilde{\sigma}}\right\} > 0$ 时,

将假设(III) 中函数 $g(\mathbf{y})$ 的强凸性弱化成凸性, 仍然能够保证辅助函数列 $M(\tilde{\boldsymbol{\omega}}^t)$ 的单调递减性.

证 由算法关于 $\mathbf{x}$ 在 $t+1$ 步的迭代式及投影定理, 可得

$$\langle \mathbf{x}^t - c \nabla_x K(\mathbf{x}^t, \mathbf{y}^t, \mathbf{z}^t, \boldsymbol{\lambda}^t) - \mathbf{x}^{t+1}, \mathbf{x}^t - \mathbf{x}^{t+1} \rangle \leq 0$$

则有

$$\langle \nabla_x K(\mathbf{x}^t, \mathbf{y}^t, \mathbf{z}^t, \boldsymbol{\lambda}^t), \mathbf{x}^t - \mathbf{x}^{t+1} \rangle \geq \frac{1}{c} \|\mathbf{x}^t - \mathbf{x}^{t+1}\|^2 \quad (4)$$

通过引理 2 的结论 1), 利用函数 $K(\mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\lambda})$ 关于 $\mathbf{x}$ 的凸性, 有

$$K(\mathbf{x}^t, \mathbf{y}^t, \mathbf{z}^t, \boldsymbol{\lambda}^t) - K(\mathbf{x}^{t+1}, \mathbf{y}^t, \mathbf{z}^t, \boldsymbol{\lambda}^t) \geq \langle \nabla_x K(\mathbf{x}^{t+1}, \mathbf{y}^t, \mathbf{z}^t, \boldsymbol{\lambda}^t), \mathbf{x}^t - \mathbf{x}^{t+1} \rangle \quad (5)$$

结合(4), (5) 式和引理 2 的结论 2) 以及 $2c < \frac{1}{L_k}$, 可得

$$\begin{aligned} & M(\mathbf{x}^t, \mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^t, \boldsymbol{\lambda}^t) - M(\mathbf{x}^{t+1}, \mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^t, \boldsymbol{\lambda}^t) = \\ & K(\mathbf{x}^t, \mathbf{y}^t, \mathbf{z}^t, \boldsymbol{\lambda}^t) - K(\mathbf{x}^{t+1}, \mathbf{y}^t, \mathbf{z}^t, \boldsymbol{\lambda}^t) \geq \langle \nabla_x K(\mathbf{x}^{t+1}, \mathbf{y}^t, \mathbf{z}^t, \boldsymbol{\lambda}^t), \mathbf{x}^t - \mathbf{x}^{t+1} \rangle = \\ & \langle \nabla_x K(\mathbf{x}^{t+1}, \mathbf{y}^t, \mathbf{z}^t, \boldsymbol{\lambda}^t) - \nabla_x K(\mathbf{x}^t, \mathbf{y}^t, \mathbf{z}^t, \boldsymbol{\lambda}^t), \mathbf{x}^t - \mathbf{x}^{t+1} \rangle + \langle \nabla_x K(\mathbf{x}^t, \mathbf{y}^t, \mathbf{z}^t, \boldsymbol{\lambda}^t), \mathbf{x}^t - \mathbf{x}^{t+1} \rangle \geq \\ & -L_k \|\mathbf{x}^t - \mathbf{x}^{t+1}\|^2 + \frac{1}{c} \|\mathbf{x}^t - \mathbf{x}^{t+1}\|^2 \geq \frac{1}{2c} \|\mathbf{x}^t - \mathbf{x}^{t+1}\|^2 \end{aligned} \quad (6)$$

通过算法关于 $\mathbf{z}$ 在 $t+1$ 步的迭代式得到

$$\begin{aligned} & M(\mathbf{x}^{t+1}, \mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^t, \boldsymbol{\lambda}^t) - M(\mathbf{x}^{t+1}, \mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^{t+1}, \boldsymbol{\lambda}^t) = \\ & \frac{p}{2} (\|\mathbf{x}^{t+1} - \mathbf{z}^t\|^2 - \|\mathbf{x}^{t+1} - \mathbf{z}^{t+1}\|^2) = \frac{p}{2} (\mathbf{z}^{t+1} - \mathbf{z}^t)^\top ((\mathbf{x}^{t+1} - \mathbf{z}^t) + (\mathbf{x}^{t+1} - \mathbf{z}^{t+1})) = \end{aligned}$$

$$\frac{p}{2} \left(\frac{2}{\beta} - 1 \right) \| \mathbf{z}^t - \mathbf{z}^{t+1} \|^2 \geq \frac{p}{2\beta} \| \mathbf{z}^t - \mathbf{z}^{t+1} \|^2 \quad (7)$$

由假设(III) 可知

$$g(\mathbf{y}^t) - g(\mathbf{y}^{t+1}) \geq \langle \nabla g(\mathbf{y}^{t+1}), \mathbf{y}^t - \mathbf{y}^{t+1} \rangle + \frac{m}{2} \| \mathbf{y}^t - \mathbf{y}^{t+1} \|^2 \quad (8)$$

利用算法关于 $\mathbf{y}$ 在 $t+1$ 步迭代式的一阶最优性条件得

$$\nabla g(\mathbf{y}^{t+1}) + \langle \mathbf{B}, \boldsymbol{\lambda}^t \rangle + \alpha \langle \mathbf{B}, \mathbf{A}\mathbf{x}^{t+1} + \mathbf{B}\mathbf{y}^{t+1} - \mathbf{b} \rangle = 0$$

再利用算法关于 $\boldsymbol{\lambda}$ 在 $t+2$ 步的迭代式可得

$$\nabla g(\mathbf{y}^{t+1}) = -\langle \mathbf{B}, \boldsymbol{\lambda}^{t+1} \rangle \quad (9)$$

结合(8),(9) 式可得

$$\begin{aligned} & M(\mathbf{x}^{t+1}, \mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^{t+1}, \boldsymbol{\lambda}^t) - M(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}, \mathbf{y}^t, \mathbf{z}^{t+1}, \boldsymbol{\lambda}^t) = \\ & g(\mathbf{y}^t) - g(\mathbf{y}^{t+1}) + \langle \boldsymbol{\lambda}^t, \mathbf{B}(\mathbf{y}^t - \mathbf{y}^{t+1}) \rangle + \eta \| \mathbf{y}^t - \mathbf{y}^{t-1} \|^2 - \eta \| \mathbf{y}^{t+1} - \mathbf{y}^t \|^2 + \\ & \frac{\alpha}{2} \| \mathbf{A}\mathbf{x}^{t+1} + \mathbf{B}\mathbf{y}^t - \mathbf{b} \|^2 - \frac{\alpha}{2} \| \mathbf{A}\mathbf{x}^{t+1} + \mathbf{B}\mathbf{y}^{t+1} - \mathbf{b} \|^2 \geq \\ & \langle -\mathbf{B}^T \boldsymbol{\lambda}^{t+1}, \mathbf{y}^t - \mathbf{y}^{t+1} \rangle + \frac{m}{2} \| \mathbf{y}^t - \mathbf{y}^{t+1} \|^2 + \langle \boldsymbol{\lambda}^t, \mathbf{B}(\mathbf{y}^t - \mathbf{y}^{t+1}) \rangle + \\ & \frac{\alpha}{2} \| \mathbf{A}\mathbf{x}^{t+1} + \mathbf{B}\mathbf{y}^t - \mathbf{b} \|^2 - \frac{\alpha}{2} \| \mathbf{A}\mathbf{x}^{t+1} + \mathbf{B}\mathbf{y}^{t+1} - \mathbf{b} \|^2 + \eta \| \mathbf{y}^t - \mathbf{y}^{t-1} \|^2 - \eta \| \mathbf{y}^{t+1} - \mathbf{y}^t \|^2 = \\ & \langle \boldsymbol{\lambda}^t - \boldsymbol{\lambda}^{t+1}, \mathbf{B}(\mathbf{y}^t - \mathbf{y}^{t+1}) \rangle + \frac{m}{2} \| \mathbf{y}^t - \mathbf{y}^{t+1} \|^2 + \frac{\alpha}{2} \| \mathbf{A}\mathbf{x}^{t+1} + \mathbf{B}\mathbf{y}^{t+1} - \mathbf{b} + \mathbf{B}(\mathbf{y}^t - \mathbf{y}^{t+1}) \|^2 - \\ & \frac{\alpha}{2} \| \mathbf{A}\mathbf{x}^{t+1} + \mathbf{B}\mathbf{y}^{t+1} - \mathbf{b} \|^2 + \eta \| \mathbf{y}^t - \mathbf{y}^{t-1} \|^2 - \eta \| \mathbf{y}^{t+1} - \mathbf{y}^t \|^2 = \\ & \langle \boldsymbol{\lambda}^t - \boldsymbol{\lambda}^{t+1}, \mathbf{B}(\mathbf{y}^t - \mathbf{y}^{t+1}) \rangle + \frac{\alpha}{2} \| \frac{1}{\alpha} (\boldsymbol{\lambda}^{t+1} - \boldsymbol{\lambda}^t) + \mathbf{B}(\mathbf{y}^t - \mathbf{y}^{t+1}) \|^2 - \frac{1}{2\alpha} \| \boldsymbol{\lambda}^{t+1} - \boldsymbol{\lambda}^t \|^2 + \\ & \left(\frac{m}{2} - \eta \right) \| \mathbf{y}^t - \mathbf{y}^{t+1} \|^2 + \eta \| \mathbf{y}^t - \mathbf{y}^{t-1} \|^2 = \\ & \frac{\alpha}{2} \| \mathbf{B}(\mathbf{y}^t - \mathbf{y}^{t+1}) \|^2 + \left(\frac{m}{2} - \eta \right) \| \mathbf{y}^t - \mathbf{y}^{t+1} \|^2 + \eta \| \mathbf{y}^t - \mathbf{y}^{t-1} \|^2 \geq \\ & \left(\frac{m}{2} - \eta \right) \| \mathbf{y}^t - \mathbf{y}^{t+1} \|^2 + \eta \| \mathbf{y}^t - \mathbf{y}^{t-1} \|^2 \end{aligned} \quad (10)$$

最后, 通过算法关于 $\boldsymbol{\lambda}$ 在 $t+1$ 步的迭代式可得

$$\begin{aligned} & M(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}, \mathbf{y}^t, \mathbf{z}^{t+1}, \boldsymbol{\lambda}^t) - M(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}, \mathbf{y}^t, \mathbf{z}^{t+1}, \boldsymbol{\lambda}^{t+1}) = \\ & -\alpha \| \mathbf{A}\mathbf{x}^{t+1} + \mathbf{B}\mathbf{y}^{t+1} - \mathbf{b} \|^2 = -\frac{1}{\alpha} \| \boldsymbol{\lambda}^{t+1} - \boldsymbol{\lambda}^t \|^2 \end{aligned}$$

再由引理 1 可得

$$M(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}, \mathbf{y}^t, \mathbf{z}^{t+1}, \boldsymbol{\lambda}^t) - M(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}, \mathbf{y}^t, \mathbf{z}^{t+1}, \boldsymbol{\lambda}^{t+1}) \geq -\frac{L_2^2}{\mu_0 \alpha} \| \mathbf{y}^{t+1} - \mathbf{y}^t \|^2 \quad (11)$$

将(6),(7),(10),(11) 式相加, 则有

$$\begin{aligned} & M(\mathbf{x}^t, \mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^t, \boldsymbol{\lambda}^t) - M(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}, \mathbf{y}^t, \mathbf{z}^{t+1}, \boldsymbol{\lambda}^{t+1}) \geq \\ & \frac{1}{2c} \| \mathbf{x}^t - \mathbf{x}^{t+1} \|^2 + \left(\frac{m}{2} - \eta - \frac{L_2^2}{\mu_0 \alpha} \right) \| \mathbf{y}^t - \mathbf{y}^{t+1} \|^2 + \eta \| \mathbf{y}^t - \mathbf{y}^{t-1} \|^2 + \frac{p}{2\beta} \| \mathbf{z}^t - \mathbf{z}^{t+1} \|^2 \end{aligned}$$

故引理 3 成立.

引理 4 若假设(I) 和(IV) 成立, $\tilde{\boldsymbol{\omega}}^t = (\mathbf{x}^t, \mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^t, \boldsymbol{\lambda}^t)$, 序列 $\{\tilde{\boldsymbol{\omega}}^t\}$ 由 P-ADMM 产生, 若 $\alpha >$

$\max\left\{\frac{2L_2^2}{\mu_0(m-2\eta)}, \frac{1}{\mu_0\tilde{\sigma}}\right\} > 0$, 那么 $\{\tilde{\omega}^t\}$ 是有界的, 且函数列 $\{M(\tilde{\omega}^t)\}$ 有下界.

证 首先, 通过完全平方可得

$$\begin{aligned} M(\mathbf{x}^t, \mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^t, \boldsymbol{\lambda}^t) &= f(\mathbf{x}^t) + g(\mathbf{y}^t) + \langle \boldsymbol{\lambda}^t, \mathbf{A}\mathbf{x}^t + \mathbf{B}\mathbf{y}^t - \mathbf{b} \rangle + \frac{\alpha}{2} \|\mathbf{A}\mathbf{x}^t + \mathbf{B}\mathbf{y}^t - \mathbf{b}\|^2 + \\ &\quad \frac{p}{2} \|\mathbf{x}^t - \mathbf{z}^t\|^2 + \eta \|\mathbf{y}^t - \mathbf{y}^{t-1}\|^2 = \\ &\quad f(\mathbf{x}^t) + g(\mathbf{y}^t) + \frac{p}{2} \|\mathbf{x}^t - \mathbf{z}^t\|^2 + \frac{\alpha}{2} \|\mathbf{A}\mathbf{x}^t + \mathbf{B}\mathbf{y}^t - \mathbf{b} + \frac{1}{\alpha}\boldsymbol{\lambda}^t\|^2 - \\ &\quad \frac{1}{2\alpha} \|\boldsymbol{\lambda}^t\|^2 + \eta \|\mathbf{y}^t - \mathbf{y}^{t-1}\|^2 \end{aligned} \quad (12)$$

再利用假设(I) 和(IV), 可得

$$\|\boldsymbol{\lambda}^t\|^2 \leq \frac{1}{\mu_0} \|\mathbf{B}^\top \boldsymbol{\lambda}^t\|^2 = \frac{1}{\mu_0} \|\nabla g(\mathbf{y}^t)\|^2$$

因此

$$\frac{1}{2\alpha} \|\boldsymbol{\lambda}^t\|^2 \leq \frac{1}{\mu_0\alpha} \|\nabla g(\mathbf{y}^t)\|^2 \quad (13)$$

将(13) 式代入(12) 式可得

$$\begin{aligned} M(\mathbf{x}^t, \mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^t, \boldsymbol{\lambda}^t) &\geq f(\mathbf{x}^t) + (g(\mathbf{y}^t) - \tilde{\sigma} \|\nabla g(\mathbf{y}^t)\|^2) + \left(\tilde{\sigma} - \frac{1}{\mu_0\alpha}\right) \|\nabla g(\mathbf{y}^t)\|^2 + \\ &\quad \frac{p}{2} \|\mathbf{x}^t - \mathbf{z}^t\|^2 + \frac{\alpha}{2} \|\mathbf{A}\mathbf{x}^t + \mathbf{B}\mathbf{y}^t - \mathbf{b} + \frac{1}{\alpha}\boldsymbol{\lambda}^t\|^2 + \eta \|\mathbf{y}^t - \mathbf{y}^{t-1}\|^2 \geq \\ &\quad f(\mathbf{x}^t) + (g(\mathbf{y}^t) - \tilde{\sigma} \|\nabla g(\mathbf{y}^t)\|^2) \end{aligned} \quad (14)$$

由引理 3 可知, 函数列 $\{M(\tilde{\omega}^t)\}$ 是单调递减的, 则有

$$M(\tilde{\omega}^1) \geq M(\tilde{\omega}^t) \quad (15)$$

由于函数 $f(\mathbf{x})$ 是有界闭集 P 上的可微函数, 则函数 $f(\mathbf{x})$ 是有界的, 由于 $\alpha > \max\left\{\frac{2L_2^2}{\mu_0(m-2\eta)}, \frac{1}{\mu_0\tilde{\sigma}}\right\} > 0$, 结合(14), (15) 式可知, 序列 $\{\mathbf{x}^t\}$, $\{\|\nabla g(\mathbf{y}^t)\|\}$ 和 $\{\|\mathbf{y}^t - \mathbf{y}^{t-1}\|\}$ 都是有界的, 通过(13) 式进一步可知函数列 $\{\boldsymbol{\lambda}^t\}$ 也是有界的. 由假设(VI) 可知 $\inf_{\mathbf{y}} g(\mathbf{y}) > -\infty$, 则序列 $\{\mathbf{y}^t\}$ 是有界的. 因此, 序列 $\tilde{\omega}^t$ 是有界的. 由假设(VI) 和(14) 式有 $M(\tilde{\omega}^t) > -\infty$. 综上所述, 序列 $\{M(\tilde{\omega}^t)\}$ 是有下界的.

将 P-ADMM 的迭代点 $\mathbf{x}^t, \mathbf{y}^t, \mathbf{z}^t, \boldsymbol{\lambda}^t$ 记为 $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}, \hat{\boldsymbol{\lambda}}$, 并将算法中 $\mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\lambda}$ 的更新变量定义为 $\hat{\mathbf{x}}^+, \hat{\mathbf{y}}^+, \hat{\mathbf{z}}^+, \hat{\boldsymbol{\lambda}}^+$, 则

$$\begin{aligned} \hat{\mathbf{x}}^+ &= \left[\hat{\mathbf{x}} - c \nabla_{\mathbf{x}} \left(L(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\boldsymbol{\lambda}}) + \frac{p}{2} \|\hat{\mathbf{x}} - \hat{\mathbf{z}}\|^2 \right) \right]_+ \\ \hat{\mathbf{z}}^+ &= \hat{\mathbf{z}} + \beta(\hat{\mathbf{x}}^+ - \hat{\mathbf{z}}) \\ \hat{\mathbf{y}}^+ &= \argmin_{\mathbf{y} \in \mathbb{R}^{n_2}} L(\hat{\mathbf{x}}^+, \mathbf{y}, \hat{\boldsymbol{\lambda}}) \\ \hat{\boldsymbol{\lambda}}^+ &= \hat{\boldsymbol{\lambda}} + \alpha(\mathbf{A}\hat{\mathbf{x}}^+ + \mathbf{B}\hat{\mathbf{y}}^+ - \mathbf{b}) \end{aligned} \quad (16)$$

下面的引理说明, 若算法停止, 则可以找到一对原始对偶解.

引理 5 若

$$(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}, \hat{\boldsymbol{\lambda}}) = (\hat{\mathbf{x}}^+, \hat{\mathbf{y}}^+, \hat{\mathbf{z}}^+, \hat{\boldsymbol{\lambda}}^+)$$

那么 $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\boldsymbol{\lambda}}) \in W$ 是原问题的一对解.

证 由(16) 式以及 $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}, \hat{\boldsymbol{\lambda}}) = (\hat{\mathbf{x}}^+, \hat{\mathbf{y}}^+, \hat{\mathbf{z}}^+, \hat{\boldsymbol{\lambda}}^+)$, 可得到以下等式

$$\begin{aligned}\nabla f(\hat{\mathbf{x}}) + \mathbf{A}^T \hat{\boldsymbol{\lambda}} + \alpha \mathbf{A}^T (\mathbf{A} \hat{\mathbf{x}} + \mathbf{B} \hat{\mathbf{y}} - \mathbf{b}) + p(\hat{\mathbf{x}} - \hat{\mathbf{z}}) - \boldsymbol{\mu} + \mathbf{v} &= 0 \\ \nabla g(\hat{\mathbf{y}}^+) + \mathbf{B}^T \hat{\boldsymbol{\lambda}} + \alpha \mathbf{B}^T (\mathbf{A} \hat{\mathbf{x}}^+ + \mathbf{B} \hat{\mathbf{y}}^+ - \mathbf{b}) &= 0 \\ \hat{\mathbf{x}}^+ &= \hat{\mathbf{z}} \\ \mathbf{A} \hat{\mathbf{x}}^+ + \mathbf{B} \hat{\mathbf{y}}^+ - \mathbf{b} &= 0 \\ \boldsymbol{\mu} &\geq 0, \mathbf{v} \geq 0 \\ \hat{\mathbf{x}} &\in P\end{aligned}$$

进一步简化得到

$$\begin{aligned}\nabla f(\hat{\mathbf{x}}) + \mathbf{A}^T \hat{\boldsymbol{\lambda}} - \boldsymbol{\mu} + \mathbf{v} &= 0 \\ \nabla g(\hat{\mathbf{y}}) + \mathbf{B}^T \hat{\boldsymbol{\lambda}} &= 0 \\ \mathbf{A} \hat{\mathbf{x}} + \mathbf{B} \hat{\mathbf{y}} - \mathbf{b} &= 0 \\ \boldsymbol{\mu} &\geq 0, \mathbf{v} \geq 0 \\ \hat{\mathbf{x}} &\in P\end{aligned}$$

由于 $\hat{\mathbf{x}} \in P$, 则有如下等式成立

$$\begin{aligned}\mu_i(l_i - \hat{x}_i) &= 0, \forall i \\ \nu_i(\hat{x}_i - u_i) &= 0, \forall i\end{aligned}$$

则 $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\boldsymbol{\lambda}})$ 满足原问题(1) 的 KKT 条件, 即有 $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\boldsymbol{\lambda}}) \in W$ 为原问题的一对解.

推论 1 若假设(I), (III) 和(IV) 成立, 对于任意 $\epsilon > 0$, 存在 $\delta(\epsilon) > 0$, 且 $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}, \hat{\boldsymbol{\lambda}}), (\hat{\mathbf{x}}^+, \hat{\mathbf{y}}^+, \hat{\mathbf{z}}^+, \hat{\boldsymbol{\lambda}}^+)$ 为 P-ADMM 产生的相邻两个迭代点, 满足

$$\max\{\|\hat{\mathbf{x}} - \hat{\mathbf{x}}^+\|, \|\hat{\mathbf{y}} - \hat{\mathbf{y}}^+\|, \|\hat{\mathbf{z}} - \hat{\mathbf{x}}^+\|, \|\hat{\boldsymbol{\lambda}} - \hat{\boldsymbol{\lambda}}^+\|\} < \delta(\epsilon)$$

我们可得

$$\max\{\text{dist}(\hat{\mathbf{z}}, X^*), \text{dist}((\hat{\mathbf{x}}, \hat{\mathbf{y}}), X^* \times Y^*)\} < \epsilon$$

证 反证, 假设存在序列 $\{\hat{\mathbf{x}}^k\}, \{\hat{\mathbf{y}}^k\}, \{\hat{\mathbf{z}}^k\}, \{\hat{\boldsymbol{\lambda}}^k\}$ 满足对于某个 $\epsilon_0 > 0$ 和任意的 $\delta > 0$, 当

$$\max\{\|\hat{\mathbf{x}}^k - (\hat{\mathbf{x}}^k)^+\|, \|\hat{\mathbf{y}}^k - (\hat{\mathbf{y}}^k)^+\|, \|\hat{\mathbf{z}}^k - (\hat{\mathbf{x}}^k)^+\|, \|\hat{\boldsymbol{\lambda}}^k - (\hat{\boldsymbol{\lambda}}^k)^+\|\} < \delta$$

成立时, 即

$$\lim_{k \rightarrow \infty} \|\hat{\mathbf{x}}^k - (\hat{\mathbf{x}}^k)^+\| = \lim_{k \rightarrow \infty} \|\hat{\mathbf{y}}^k - (\hat{\mathbf{y}}^k)^+\| = \lim_{k \rightarrow \infty} \|\hat{\mathbf{z}}^k - (\hat{\mathbf{x}}^k)^+\| = \lim_{k \rightarrow \infty} \|\hat{\boldsymbol{\lambda}}^k - (\hat{\boldsymbol{\lambda}}^k)^+\| = 0$$

则有

$$\text{dist}(\hat{\mathbf{x}}^k, X^*) > 0, \text{dist}((\hat{\mathbf{z}}^k, \hat{\mathbf{y}}^k), X^* \times Y^*) > 0, \forall k = 1, 2, 3 \dots \quad (17)$$

通过引理 4 可知, 由算法 P-ADMM 产生的迭代序列是有界的, 则存在子列 $\{\hat{\mathbf{x}}^{k_i}\}, \{\hat{\mathbf{y}}^{k_i}\}, \{\hat{\mathbf{z}}^{k_i}\}, \{\hat{\boldsymbol{\lambda}}^{k_i}\}$ 满足

$$\hat{\mathbf{x}}^{k_i} \rightarrow \bar{\mathbf{x}}, \hat{\mathbf{y}}^{k_i} \rightarrow \bar{\mathbf{y}}, \hat{\mathbf{z}}^{k_i} \rightarrow \bar{\mathbf{z}}, \hat{\boldsymbol{\lambda}}^{k_i} \rightarrow \bar{\boldsymbol{\lambda}}$$

进而可以得到

$$\bar{\mathbf{x}} = \bar{\mathbf{x}}^+, \bar{\mathbf{y}} = \bar{\mathbf{y}}^+, \bar{\mathbf{z}} = \bar{\mathbf{z}}^+, \bar{\boldsymbol{\lambda}} = \bar{\boldsymbol{\lambda}}^+$$

则有

$$(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}, \bar{\boldsymbol{\lambda}}) = (\bar{\mathbf{x}}^+, \bar{\mathbf{y}}^+, \bar{\mathbf{z}}^+, \bar{\boldsymbol{\lambda}}^+)$$

由引理 5 可得 $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\boldsymbol{\lambda}}) \in W$, 即

$$(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in X^* \times Y^*, \bar{\mathbf{x}} = \bar{\mathbf{z}} \in X^*$$

则有

$$\text{dist}((\bar{\mathbf{x}}, \bar{\mathbf{y}}), X^* \times Y^*) = 0, \text{dist}(\hat{\mathbf{z}}^k, X^*) = 0 \quad (18)$$

通过(17) 式可得

$$\text{dist}((\hat{\mathbf{x}}^{k_i}, \hat{\mathbf{y}}^{k_i}), X^* \times Y^*) > 0, \text{dist}(\hat{\mathbf{z}}^{k_i}, X^*) > 0$$

则有

$$\lim_{i \rightarrow \infty} \text{dist}((\hat{\mathbf{x}}^{k_i}, \hat{\mathbf{y}}^{k_i}), X^* \times Y^*) = \text{dist}((\bar{\mathbf{x}}, \bar{\mathbf{y}}), X^* \times Y^*) > 0$$

$$\lim_{i \rightarrow \infty} \text{dist}(\hat{\mathbf{z}}^{k_i}, X^*) = \text{dist}(\bar{\mathbf{z}}, X^*) > 0$$

这与(18) 式矛盾, 则推论 1 成立.

定理 1 若假设(I), (III) 和(IV) 成立, 序列 $\{\boldsymbol{\omega}^t\}$ 为 P-ADMM 所产生, 则下列结论成立

- 1) $\lim_{t \rightarrow \infty} (\|\mathbf{x}^t - \mathbf{x}^{t+1}\| + \|\mathbf{y}^t - \mathbf{y}^{t+1}\| + \|\mathbf{z}^t - \mathbf{z}^{t+1}\| + \|\boldsymbol{\lambda}^t - \boldsymbol{\lambda}^{t+1}\|) = 0$
- 2) $\lim_{t \rightarrow \infty} \|\mathbf{x}^{t+1} - \mathbf{z}^t\| = 0$
- 3) $\lim_{t \rightarrow \infty} \text{dist}(\mathbf{z}^t, X^*) = 0, \lim_{t \rightarrow \infty} \text{dist}((\mathbf{x}^t, \mathbf{y}^t), X^* \times Y^*) = 0, \lim_{t \rightarrow \infty} \text{dist}((\mathbf{x}^t, \mathbf{y}^t, \boldsymbol{\lambda}^t), W) = 0$

证 1) 由引理 3 可知

$$M(\mathbf{x}^t, \mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^t, \boldsymbol{\lambda}^t) - M(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}, \mathbf{y}^t, \mathbf{z}^{t+1}, \boldsymbol{\lambda}^{t+1}) \geq \frac{1}{2c} \|\mathbf{x}^t - \mathbf{x}^{t+1}\|^2 + \left(\frac{m}{2} - \eta - \frac{L_2^2}{\mu_0 \alpha}\right) \|\mathbf{y}^t - \mathbf{y}^{t+1}\|^2 + \eta \|\mathbf{y}^t - \mathbf{y}^{t-1}\|^2 + \frac{p}{2\beta} \|\mathbf{z}^t - \mathbf{z}^{t+1}\|^2 \geq 0$$

由引理 4 知函数列 $\{M(\boldsymbol{\omega}^t)\}$ 有下界, 则极数 $\sum_{t=1}^{\infty} M(\mathbf{x}^t, \mathbf{y}^t, \mathbf{y}^{t-1}, \mathbf{z}^t, \boldsymbol{\lambda}^t) - M(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}, \mathbf{y}^t, \mathbf{z}^{t+1}, \boldsymbol{\lambda}^{t+1})$ 是收敛的, 因此

$$\sum_{t=1}^{\infty} \|\mathbf{x}^t - \mathbf{x}^{t+1}\|^2 < \infty, \sum_{t=1}^{\infty} \|\mathbf{y}^t - \mathbf{y}^{t+1}\|^2 < \infty, \sum_{t=1}^{\infty} \|\mathbf{z}^t - \mathbf{z}^{t+1}\|^2 < \infty$$

进而可得到

$$\lim_{t \rightarrow \infty} \|\mathbf{x}^t - \mathbf{x}^{t+1}\| = 0, \lim_{t \rightarrow \infty} \|\mathbf{y}^t - \mathbf{y}^{t+1}\| = 0, \lim_{t \rightarrow \infty} \|\mathbf{z}^t - \mathbf{z}^{t+1}\| = 0$$

通过定理 1 可知

$$\|\boldsymbol{\lambda}^{t+1} - \boldsymbol{\lambda}^t\|^2 \leq \frac{L_2^2}{\mu_0} \|\mathbf{y}^{t+1} - \mathbf{y}^t\|^2$$

则可得

$$\lim_{t \rightarrow \infty} \|\boldsymbol{\lambda}^t - \boldsymbol{\lambda}^{t+1}\| = 0$$

结论 1) 成立.

2) 由于

$$\lim_{t \rightarrow \infty} \|\mathbf{z}^t - \mathbf{z}^{t+1}\| = 0$$

结合 $\mathbf{z}$ 的第 $t+1$ 次迭代式 $\mathbf{z}^{t+1} = \mathbf{z}^t + \beta(\mathbf{x}^{t+1} - \mathbf{z}^t)$ 可得

$$\lim_{t \rightarrow \infty} \|\mathbf{x}^{t+1} - \mathbf{z}^t\| = 0$$

则结论 2) 成立.

3) 由 1), 2) 两个结论可得

$$\lim_{t \rightarrow \infty} \|\mathbf{x}^t - \mathbf{x}^{t+1}\| = \lim_{t \rightarrow \infty} \|\mathbf{y}^t - \mathbf{y}^{t+1}\| = \lim_{t \rightarrow \infty} \|\mathbf{z}^t - \mathbf{x}^{t+1}\| = \lim_{t \rightarrow \infty} \|\boldsymbol{\lambda}^t - \boldsymbol{\lambda}^{t+1}\| = 0$$

结合推论 1 有

$$\begin{aligned}\lim_{t \rightarrow \infty} \text{dist}(\mathbf{z}^t, X^*) &= 0 \\ \lim_{t \rightarrow \infty} \text{dist}((\mathbf{x}^t, \mathbf{y}^t), X^* \times Y^*) &= 0 \\ \lim_{t \rightarrow \infty} \text{dist}((\mathbf{x}^t, \mathbf{y}^t, \boldsymbol{\lambda}^t), W) &= 0\end{aligned}$$

则结论 3) 成立.

3 数值实验

本文考虑用 P-ADMM 解决一类非凸的船舶分布式能源管理问题, 在文献[17] 中, 考虑具有多个代理的船舶动力系统, 且各代理根据各自的需求、供给相互协调. 首先将船舶能量管理问题以多智能体系统的形式重新制定, 然后利用交替方向乘子法(ADMM) 寻找模型中满足共识要求的最经济的工作点, 本文考虑如下可分非凸光滑的能量成本函数

$$\begin{aligned}\min_{\mathbf{x}} F(\mathbf{x}) &= \sum_{i=1}^D [x_i^2 - 10\cos(2\pi x_i) + 10] \\ \text{s. t. } \mathbf{x} &\in M\end{aligned}\tag{19}$$

其中

$$M:=\{\mathbf{x} \in \mathbb{R}^D \mid -5.12 \leq x_i \leq 5.12, i=1,2,\cdots,D\}$$

$F(\mathbf{x})$ 为能量成本函数, x_i 为共识变量, D 为该问题的维度, 问题(19) 可等价转化成如下问题

$$\begin{aligned}\min_{\mathbf{x}, \mathbf{y}} G(\mathbf{x}) + H(\mathbf{y}) &= \sum_{i=1}^D [10 - 10\cos(2\pi x_i)] + \sum_{i=1}^D y_i^2 \\ \text{s. t. } \mathbf{x} &= \mathbf{y}, \mathbf{x} \in M\end{aligned}\tag{20}$$

问题(20) 所对应的增广拉格朗日函数为

$$L(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}) = G(\mathbf{x}) + H(\mathbf{y}) + \boldsymbol{\lambda}^\top (\mathbf{x} - \mathbf{y}) + \frac{\alpha}{2} \|\mathbf{x} - \mathbf{y}\|^2$$

利用 P-ADMM 求解问题(20).

本文实验在配置英特尔 Core i5-6200U 2.30 GHz 处理器、4.00 GB 内存和 Windows 10(64 位) 操作系统的计算机上进行, 其中 t 为迭代步数. 算法的参数设置如下: $p=8\ 000$, $c=\frac{1}{5\ 000}$, $D=50$, $\alpha=100$, $\mathbf{y}^0$, $\mathbf{z}^0, \boldsymbol{\lambda}^0$ 都是维数 50 的零向量, $\mathbf{x}^0 \in \mathbb{R}^{50}$ 为均匀分布在 $[-5.12, 5.12]$ 上的向量, 终止条件

$$r_{\text{Error}} = \max\{\|\mathbf{x}^t - \mathbf{x}^{t+1}\|, \|\mathbf{y}^t - \mathbf{y}^{t+1}\|, \|\boldsymbol{\lambda}^t - \boldsymbol{\lambda}^{t+1}\|\} \leq 10^{-8}$$

图 1 展示了在 P-ADMM 算法下船舶电力系统能量成本的走势情况, 其中 $G(\mathbf{x}^t) + H(\mathbf{y}^t)$ 为成本函数, t 为迭代步. 通过图 1 可以看出, 能量成本值从开始的下降逐渐趋于稳定, 最后达到收敛. 同时, 观察发现选择较大的 β , 会使收敛速度加快, 即算法可以在更小的迭代步数下完成收敛, 减少了算法迭代的次数.

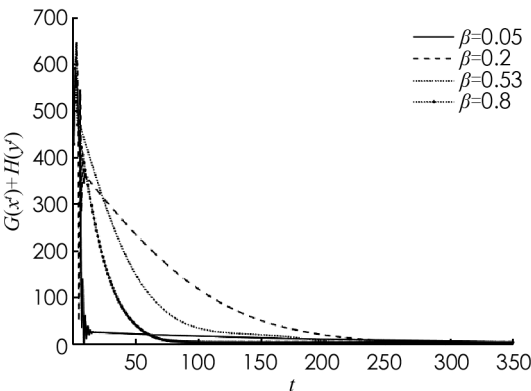


图 1 成本函数

图 2 和图 3 分别展示了在参数 $\beta=0.2$ 的条件下 $\|\mathbf{x}^t - \mathbf{y}^t\|$ 和 $\|\mathbf{x}^{t+1} - \mathbf{z}^t\|$ 的下降情况, 进一步验证了算法 P-ADMM 解决问题(20)的可行性. 从图 2,3 中可知, $\|\mathbf{x}^t - \mathbf{y}^t\|$ 和 $\|\mathbf{x}^{t+1} - \mathbf{z}^t\|$ 一开始下降得非常快, 再逐渐趋于平稳并达到收敛.

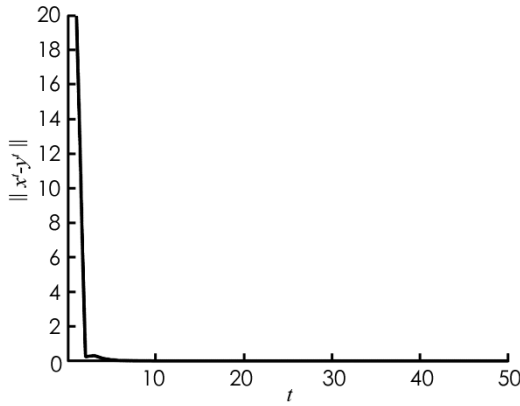
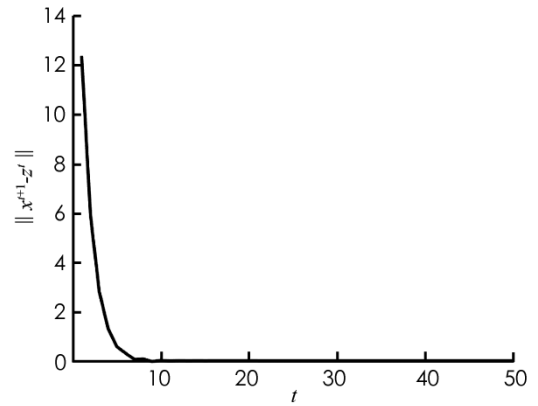
图 2 $\|\mathbf{x}^t - \mathbf{y}^t\|$ 图 3 $\|\mathbf{x}^{t+1} - \mathbf{z}^t\|$

图 4 是基于问题(20)将 P-ADMM 和 APGMM^[17] 在参数 $\beta=0.8$ 的条件下进行了比较. 从图 4 中可以看到算法 P-ADMM 相对于算法 APGMM 在下降趋势和收敛步数上都具有较大的优势. APGMM 算法开始迭代是波动的、不稳定的. 考虑到目标函数的非凸性, 我们在运行算法时, 选取 $\alpha=100$, $\alpha=1\,000$ 两种情况进行比较. 由图 4 可知算法 P-ADMM 的下降速度快, 能够在更小的迭代步达到收敛.

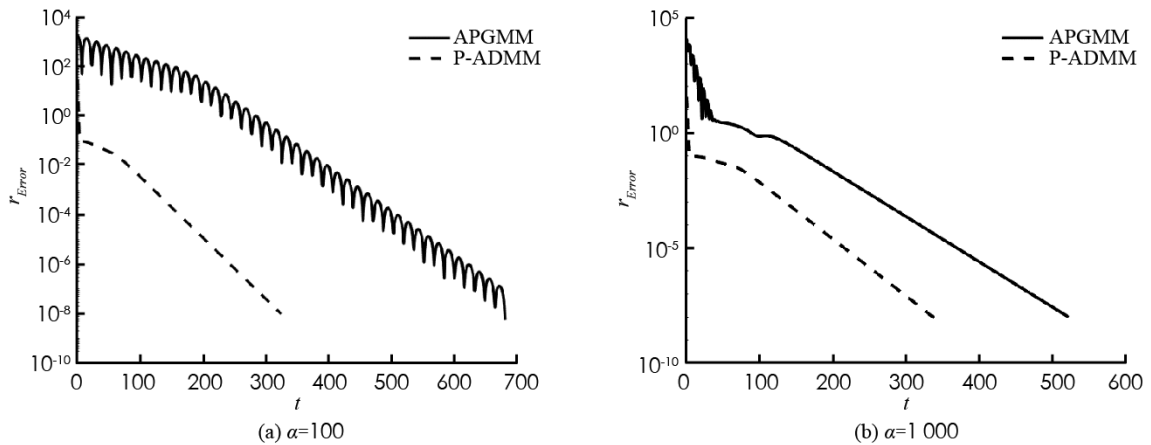


图 4 基于问题(20)的算法对比实验

4 结论

本文在梯度投影算法和 ADMM 基础上, 针对可分结构的非凸优化问题, 提出了一种新的 P-ADMM, 该算法具有良好的收敛性, 简化了子问题的求解, 降低了运算成本, 且近似二次项的引入在保证算法的收敛性的同时也能够提升算法的收敛速度. 该算法在一类非凸的船舶能源管理问题中也得到了有效应用. P-ADMM 在梯度投影时, 选取的步长是满足条件的定值, 但如何优化梯度投影的步长以提高算法速率需要进一步的研究.

参考文献:

- [1] SCHUBIGER M, BANJAC G, LYGEROS J. GPU Acceleration of ADMM for Large-Scale Quadratic Programming [J]. Journal of Parallel and Distributed Computing, 2020, 144: 55-67.
- [2] CHEN G, YANG Q. An ADMM-Based Distributed Algorithm for Economic Dispatch in Islanded Microgrids [J]. IEEE

- Transactions on Industrial Informatics, 2018, 14(9): 3892-3903.
- [3] LAI K X, ILLINDALA M S. A Distributed Energy Management Strategy for Resilient Shipboard Power System [J]. Applied Energy, 2018, 228: 821-832.
- [4] XU Y Y, YIN W T, WEN Z W, et al. An Alternating Direction Algorithm for Matrix Completion with Nonnegative Factors [J]. Frontiers of Mathematics in China, 2012, 7(2): 365-384.
- [5] GABAY D, MERCIER B. A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximation [J]. Computers & Mathematics With Applications, 1976, 2(1): 17-40.
- [6] CAI X J, HAN D R, YUAN X M. On the Convergence of the Direct Extension of ADMM for Three-Block Separable Convex Minimization Models with one Strongly Convex Function [J]. Computational Optimization and Applications, 2017, 66(1): 39-73.
- [7] NISHIHARA R, LESSARD L, RECHT B, et al. A General Analysis of the Convergence of ADMM [EB/OL]. (2015-2-6) [2021-8-12]. <https://doi.org/10.48550/arXiv.1502.02009>.
- [8] DENG W, YIN W T. On the Global and Linear Convergence of the Generalized Alternating Direction Method of Multipliers [J]. Journal of Scientific Computing, 2016, 66(3): 889-916.
- [9] WANG H H, BANERJEE A. Online Alternating Direction Method [EB/OL]. (2013-7-10) [2021-8-12]. <https://arxiv.org/abs/1206.6448>.
- [10] WANG F H, XU Z B, XU H K. Convergence of Bregman Alternating Direction Method with Multipliers for Nonconvex Composite Problems [EB/OL]. (2014-12-5) [2021-8-12]. <https://arxiv.org/abs/1410.8625>.
- [11] CHEN C H, HE B S, YE Y Y, et al. The Direct Extension of ADMM for Multi-Block Convex Minimization Problems is not Necessarily Convergent [J]. Mathematical Programming, 2016, 155(1/2): 57-79.
- [12] WANG F H, CAO W F, XU Z B. Convergence of Multi-Block Bregman ADMM for Nonconvex Composite Problems [J]. Science China Information Sciences, 2018, 61(12): 1-12.
- [13] BOT R I, NGUYEN D K. The Proximal Alternating Direction Method of Multipliers in the Nonconvex Setting: Convergence Analysis and Rates [J]. Mathematics of Operations Research, 2020, 45(2): 682-712.
- [14] PENG Z, CHEN J L, ZHU W X. A Proximal Alternating Direction Method of Multipliers for a Minimization Problem with Nonconvex Constraints [J]. Journal of Global Optimization, 2015, 62(4): 711-728.
- [15] ZHANG J W, LUO Z Q. A Proximal Alternating Direction Method of Multiplier for Linearly Constrained Nonconvex Minimization [J]. SIAM Journal on Optimization, 2020, 30(3): 2272-2302.
- [16] GAO X, ZHANG S Z. First-Order Algorithms for Convex Optimization with Nonseparable Objective and Coupled Constraints [J]. Journal of the Operations Research Society of China, 2017, 5(2): 131-159.
- [17] MOHAMED M A, CHABOK H, AWWAD E M, et al. Stochastic and Distributed Scheduling of Shipboard Power Systems Using MFOA-ADMM [J]. Energy, 2020, 206: 118041.