

求解可分离凸优化问题的惯性 近似松弛交替方向乘子法

薛中会¹, 殷倩雯², 党亚峥²

(1. 上海出版印刷高等专科学校, 上海 200093; 2. 上海理工大学 管理学院, 上海 200093)

摘要: 基于交替方向乘子法 (ADMM) 提出了一种求解可分离凸优化可行问题的惯性近似松弛交替方向乘子法 (IPR-ADMM)。新构造的算法不仅具有提高算法收敛性的优势的惯性外推项, 而且引入随机变量以随机加速新步长, 从而提高算法的灵活性。并在适当的假设下, 证明了算法的全局迭代收敛性。数值实验结果表明, 数据维数取值越大, 算法收敛越快, 越趋于稳定, 且 IPR-ADMM 算法的收敛性明显优于扩展的邻近交替方向法 (ePADM)。

关键词: 惯性近似松弛; 交替方向乘子法; 凸优化; 惯性外推; 随机加速; 全局收敛性
中图分类号: O 221.2 **文献标志码:** A

Inertia approximate relaxation alternating direction multiplier method for separable convex optimization problems

XUE Zhonghui¹, YIN Qianwen², DANG Yazheng²

(1. Shanghai Publishing and Printing College, Shanghai 200093, China; 2. Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Based on alternating direction multiplier method (ADMM), an inertial approximate relaxation alternating direction multiplier method (IPR-ADMM) was proposed to solve separable convex optimization feasible problems. The new algorithm not only integrated the advantages of inertia extrapolation term to improve the convergence of the algorithm, but also introduced random variables to accelerate the new step size randomly, so as to improve the flexibility of the algorithm. Under suitable assumptions, the global convergence of the algorithm was proved. The numerical results show that the larger the data dimension is, the faster and more stable the convergence of the algorithm is. Moreover, the convergence of IPR-ADMM algorithm is better than that of ePADM algorithm.

Keywords: inertia approximate relaxation; alternating direction multiplier method; convex optimization; inertia extrapolation; random acceleration; global convergence

1 问题的提出

一个优化问题如果满足:

$$\begin{aligned} \min & f(\mathbf{x}) \\ \text{s.t.} & g_i(\mathbf{x}) \leq 0, i = 1, 2, \dots, m \\ & h_j(\mathbf{x}) = 0, j = 1, 2, \dots, p \end{aligned}$$

收稿日期: 2021-03-30

基金项目: 国家自然科学基金资助项目 (72071130)

第一作者: 薛中会 (1974-), 男, 副教授。研究方向: 最优化理论和数学物理方程。E-mail: hnlxzh@163.com

称该问题为一个凸优化问题。其中, 目标函数 f 为凸函数, 不等式约束 g_i 也为凸函数, 而等式约束 h_j 为一个仿射函数。

一个可分离的凸优化问题为

$$\min \{\theta_1(\mathbf{x}) + \theta_2(\mathbf{y}) | \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} = \mathbf{b}, \mathbf{x} \in \mathbf{X}, \mathbf{y} \in \mathbf{Y}\} \quad (1)$$

式中: $\mathbf{X} \subset \mathbb{R}^{n_1}, \mathbf{Y} \subset \mathbb{R}^{n_2}$ 是有界闭合非空凸集; $f: \mathbf{X} \rightarrow \mathbb{R}$ 和 $g: \mathbf{Y} \rightarrow \mathbb{R}$ 是凸函数 (非必要光滑的); $\mathbf{A} \in \mathbb{R}^{m \times n_1}, \mathbf{B} \in \mathbb{R}^{m \times n_2}$ 和 $\mathbf{b} \in \mathbb{R}^m$ 。

交替方向乘子法 (ADMM) 是一种求解具有可分离的凸优化问题的重要方法。由于其处理速度快、收敛性能好, ADMM 算法在统计学习、机器学习等领域有着广泛应用, 在求解可分离凸优化问题上具有简单、灵活、实用性强的效果。其优势在于利用对偶上升算法的可分离性, 可以将大规模问题拆分成 2 个甚至多个小规模子问题, 随后交替求解分解所得的各个小规模子问题, 从而提高了求解问题的效率。ADMM 算法最早由 Glowinski 等^[1] 和 Gabay 等^[2] 提出, ADMM 算法的经典迭代步骤为

$$\begin{cases} \mathbf{x}^{k+1} = \arg \min_{\mathbf{x} \in \mathbf{X}} \left\{ \theta_1(\mathbf{x}) + \langle \lambda^k, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}^k - \mathbf{b} \rangle + \frac{\beta}{2} \|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}^k - \mathbf{b}\|^2 \right\} \\ \mathbf{y}^{k+1} = \arg \min_{\mathbf{y} \in \mathbf{Y}} \left\{ \theta_2(\mathbf{y}) + \langle \lambda^k, \mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y} - \mathbf{b} \rangle + \frac{\beta}{2} \|\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y} - \mathbf{b}\|^2 \right\} \\ \lambda^{k+1} = \lambda^k - \beta(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^{k+1} - \mathbf{b}) \end{cases} \quad (2)$$

ADMM 在每一步迭代中都能解决较简单的子问题, 并且可以分别利用 $f(\mathbf{x})$ 和 $g(\mathbf{x})$ 的结构。众所周知, 如式 (1) 具有 KKT 点, 则由式 (2) 生成的对偶序列收敛到对偶问题的最优解, 但是, 在没有附加条件的情况下, 原始迭代序列不一定收敛。为了改善原始收敛性, Eckstein^[3] 首先通过向式 (2) 的子问题添加一些二次项, 提出了邻近 ADMM 算法。经研究发现, 惯性技术也可以加入 ADMM 算法中, 在适当的假设条件下能够加速算法的收敛效果。Alvarez^[4] 最早提出惯性技术这一概念, 其基本思想是利用当前迭代和上一步迭代之间的相关联系得到下一步迭代, 这样不仅能够较快地得到所求问题的最优解, 而且在收敛性证明上也相对容易。近年来, 惯性技术被运用于邻近点算法 (PPM) 求解极大单调算子包含问题。通常情况下, 为了加速邻近点算法的收敛速度, 考虑二阶微分包含问题

$$\mathbf{w}'' + \gamma \mathbf{w}' + \nabla \psi(\mathbf{w}) = 0$$

式中, $\gamma > 0$ 是一个摩擦参数。

具体来说, Alvarez^[4] 提出了一种隐式离散化方法, 给定任意的 $\mathbf{w}^{k-1}$ 和 $\mathbf{w}^k$, 生成的下一个迭代点可以由下式得到:

$$\frac{\mathbf{w}^{k+1} - 2\mathbf{w}^k + \mathbf{w}^{k-1}}{h^2} + \gamma \frac{\mathbf{w}^{k+1} - \mathbf{w}^k}{h} + \nabla \psi(\mathbf{w}^{k+1}) = 0$$

于是可以得到算法的迭代形式为

$$\mathbf{w}^{k+1} = (\mathbf{I} + \lambda \nabla \psi)^{-1}(\mathbf{w}^k + \alpha(\mathbf{w}^k - \mathbf{w}^{k-1}))$$

其中, $\lambda = h^2/(1 + \gamma h), \alpha = 1/(1 + \gamma h)$ 。

与邻近点算法相比, 算法的迭代形式可以看作是将外推点 $\mathbf{w}^k + \alpha(\mathbf{w}^k - \mathbf{w}^{k-1})$ 应用于邻近点算法, 故称其为惯性邻近点算法。随后, 惯性技术被应用于其他算法中。2014 年, Lorenz 等^[5] 提出了惯性向前-向后分裂算法, 并在文献 [6-8] 算法的基础上结合 Nesterov 方法对算法进行改进, 即只针对 $\mathbf{y}$ -子问题使用惯性技术, 并通过数值实验证明了该算法的有效性。惯性 Douglas-Rachford 算子分裂方法^[9]、惯性三算子分离算法^[10] 和惯性 ADMM^[11] 也相继被提出。Chen 等^[12] 通过结合邻近 ADMM 算法^[13] 和惯性邻近点算法^[7-8] 提出了求解带有线性约束可分离凸优化问题的惯性邻近 ADMM 算法, 同时说明了矩阵并非一定是正定的, 也可以是半正定的, 并证明了算法的收敛性。

基于惯性技术在加快收敛性方面具有很好的效果, 本文对可分离凸优化问题采用惯性技术, 同时引入随机加速的随机变量以更新步长, 提出了惯性近似松弛交替方向乘子法。在适当的假设条件下, 基于惯性邻近点法的收敛性证明了惯性邻近 ADMM 算法的收敛性。另外, 数值实验验证了新算法在实践中具有更好的数值表现。

2 预备知识

$\|\mathbf{u}\| = \sqrt{\mathbf{u}^T \mathbf{u}}$ 表示 Euclidean 范数, $\langle \cdot, \cdot \rangle$ 表示内积。令 $\Omega \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^r$ 是一个非空闭合凸子集, $P_\Omega(\mathbf{x})$ 为向量 $\mathbf{x}$ 在凸集 Ω 的投影, 即

$$P_\Omega(\mathbf{x}) = \arg \min_{\mathbf{y} \in \Omega} \|\mathbf{x} - \mathbf{y}\|$$

引理 1 令 $\zeta > 0, \zeta$ 是一个常量, 并且 $\mathbf{C}$ 是一个非空闭凸集, 那么, 当且仅当

$$\mathbf{x}^* = P_{\mathbf{C}}[\mathbf{x}^* - \zeta f(\mathbf{x}^*)]$$

$\mathbf{x}^* \in \mathbf{C}$ 是 $VI(f, \mathbf{C})$ 的最优解。

记 $e(\mathbf{x}, \zeta) = \mathbf{x} - P_{\mathbf{C}}[\mathbf{x} - \zeta f(\mathbf{x})]$ 为投影方程的残差。

$VI(f, C)$ 为等价于 $e(x, \zeta)$ 的零点。因此, 将 $\|(x, \zeta)\|$ 当作算法的终止准则。

引理 1 表明, 对于任意的 $\zeta > 0$, $\|(x, \zeta)\|$ 是一个非增函数。

引理 2^[14] 令 $\Omega \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^r$ 是一个非空闭凸集, 那么, 对任意的 $x, y \in \mathbb{R}^N$ 和 $z \in \Omega$, 有

- $\langle P_\Omega(x) - x, z - P_\Omega(x) \rangle \geq 0$;
- $\|P_\Omega(x) - P_\Omega(y)\|^2 \leq \langle P_\Omega(x) - P_\Omega(y), x - y \rangle$;
- $\|P_\Omega(x) - z\|^2 \leq \|x - z\|^2 - \|P_\Omega(x) - x\|^2$;
- $\|P_\Omega(x) - P_\Omega(y)\| \leq \|x - y\| - \|P_\Omega(x) - x + y - P_\Omega(y)\|$ 。

通常可行集 C 具有如下形式: $C = \{x | Ax + By = b, x, y \in Q\}$ 。其中, $A, B \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, Q 是 $\mathbb{R}_+^n$ 的一个简单的凸集, 那么, 基本的 $VI(f, C)$ 可以转换为结构化的变分不等式问题, 表示为 $VI(f, C)$: 找到一个 $u^* \in Q$, 使得

$$\theta(u) - \theta(u^*) + F(w^*)^T(w - w^*) \geq 0, \forall u \in Q \quad (3)$$

其中,

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

定义 1 令 f 是从 $C \subseteq \mathbb{R}^n$ 到 $\mathbb{R}^n$ 的一个映射, 若

$$(u - v)^T(f(u) - f(v)) \geq 0, \forall u, v \in C$$

则 f 在 C 上是单调的。

如果 C 是 $\mathbb{R}^n$ 上的一个紧凸集, 且 $F: C \rightarrow \mathbb{R}^n$ 是一个连续映射, 那么, 变分不等式问题(VIP)至少有 1 个解。进而可知, 若函数是单调的, 那么, 变分不等式问题的解存在且唯一。

引理 3 如果 C 是 $\mathbb{R}^n$ 的非空闭凸集, 而 $F: C \rightarrow \mathbb{R}^n$ 是一个连续映射, 那么, 对于任意给定的 $x, y \in C$, 有 $\langle x - y, F(x) - F(y) \rangle \geq 0, \forall x, y \in C$ 。

定义 2 假设序列 $\{X_n\}$ 是由 $X_1, X_2, \dots, X_n, \dots$ 组成的随机变量, 如果对于任意的 $\varepsilon > 0$, 有 $\lim_{n \rightarrow \infty} P\{X_n - X > \varepsilon\} = 0$, 那么, 序列 $\{X_n\}$ 以概率收敛到 X , 记为 $X_n \xrightarrow{P} X$ 。

引理 4 马尔可夫不等式: 对于任何 $\varepsilon > 0$ 和 $r > 0$, 有

$$P(|X| \geq \varepsilon) \leq \frac{E(|X|)^r}{\varepsilon^r}$$

引理 5 假设 $X_n \xrightarrow{P} X$, 那么, 一定有 $X_{n_k} \rightarrow X$ 。

3 惯性近似松弛 ADMM

现针对可分离凸优化问题构建惯性近似松弛

交替方向乘子法(IPR-ADMM)。

问题(1)的增广拉格朗日函数为

$$L(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b) + \frac{\beta}{2} \|Ax + By - b\|^2$$

如果 $(x^*, y^*) \in X \times Y$ 是问题的最优解, 那么, 存在一个 $\lambda^* \in \mathbb{R}^m$, 使得 (x^*, y^*, λ^*) 是一个鞍点, 满足

$$\begin{cases} L(x, y^*, \lambda^*) - L(x^*, y^*, \lambda^*) \geq 0, \forall x \in X \\ L(x^*, y, \lambda^*) - L(x^*, y^*, \lambda^*) \geq 0, \forall y \in Y \\ L(x^*, y^*, \lambda) - L(x^*, y^*, \lambda^*) \geq 0, \forall \lambda \in \mathbb{R}^m \end{cases} \quad (4)$$

因此, 由式(4)得出

$$\begin{cases} x^* = \arg \min_{x \in X} L(x, y^*, \lambda^*) \\ y^* = \arg \min_{y \in Y} L(x^*, y, \lambda^*) \\ \lambda^* = \arg \min_{\lambda \in \mathbb{R}^m} L(x^*, y^*, \lambda) \end{cases} \quad (5)$$

根据引理 1, 问题(1)的最优解可表示为

$$\begin{cases} \theta_1(x) - \theta_1(x^*) - (x - x^*)^T(A^T \lambda^*) \geq 0, \forall x \in X \\ \theta_2(y) - \theta_2(y^*) - (y - y^*)^T(B^T \lambda^*) \geq 0, \forall y \in Y \\ (\lambda - \lambda^*)^T(Ax^* + By^* - b) \geq 0, \forall \lambda \in \mathbb{R}^m \end{cases} \quad (6)$$

通过式(6)的紧凑形式, 问题(1)的解集为 W^* , 其中,

$$W := X \times Y \times \mathbb{R}^m, \theta(u) =: \theta_1(x) + \theta_2(y)$$

现介绍本文的主要算法。

算法 1 惯性近似松弛 ADMM。

a. 对于给定的一个 w^k , 令 $(x^{-1}, y^{-1}, \lambda^{-1}) = (x^0, y^0, \lambda^0)$, 有

$$\begin{aligned} (\bar{x}^k, \bar{y}^k, \bar{\lambda}^k) &= (x^k, y^k, \lambda^k) + \\ &\tau_k(x^k - x^{k-1}, y^k - y^{k-1}, \lambda^k - \lambda^{k-1}) \end{aligned} \quad (7)$$

b. 计算

$$\begin{cases} \bar{x}^k = \arg \min_{x \in X} \left\{ \theta_1(x) - \langle \bar{\lambda}^k, Ax + B\bar{y}^k - b \rangle + \frac{\beta}{2} \|Ax + B\bar{y}^k - b\|^2 + \frac{R}{2} \|x - \bar{x}^k\|^2 \right\} \\ \bar{y}^k = \arg \min_{y \in Y} \left\{ \theta_2(y) - \langle \bar{\lambda}^k, A\bar{x}^k + By - b \rangle + \frac{\beta}{2} \|A\bar{x}^k + By - b\|^2 + \frac{S}{2} \|y - \bar{y}^k\|^2 \right\} \\ \bar{\lambda}^k = \bar{\lambda}^k - \beta(A\bar{x}^k + B\bar{y}^k - b) \end{cases} \quad (8)$$

c. 计算

$$w^{k+1} = \bar{w}^k - \eta_k \alpha_k (\bar{w}^k - \bar{w}^k) \quad (9)$$

其中,

$$\begin{aligned} \xi_k &= \omega(a, b), \eta_k = \frac{1}{k} \sum_{i=1}^k \xi_k \\ \alpha_k &= \frac{\|\bar{w}^k - \bar{w}^k\|_Q^2 + (\bar{\lambda}^k - \bar{\lambda}^k)^T B(\bar{y}^k - \bar{y}^k)}{\|\bar{w}^k - \bar{w}^k\|_Q^2} \end{aligned}$$

c. 停止准则。计算

$$c = \frac{\max(|\mathbf{w}^{k+1} - \mathbf{w}^k|)}{\max(|\mathbf{w}^0 - \tilde{\mathbf{w}}^0|)} \leq 10^{-4}$$

如果 $c < \varepsilon$, 停止; 否则, 令 $k = k + 1$ 返回 a。

为了确保收敛, 需要满足 $\xi_k = \omega(a, b)$ 是某个独立且均匀分布的扩展序列, 并且 $\xi_k = \omega(a, b)$ 的数学期望为 $E(\xi_k)$, $k = 1, 2, \dots$ 。另外, 在整个迭代过程不需要计算 $\rho(A^T A)$ 或 A^{-1} 。

4 收敛性证明

现利用变分不等式证明算法 1 的全局收敛性。根据式 (8) 的变分不等式形式生成如下形式的迭代方案:

$$\begin{cases} (\mathbf{x}' - \tilde{\mathbf{x}}^k) \{ \theta_1(\tilde{\mathbf{x}}^k) - A^T [\tilde{\lambda}^k - \beta(A\tilde{\mathbf{x}}^k + B\tilde{\mathbf{y}}^k - b)] + \\ R(\tilde{\mathbf{x}}^k - \tilde{\mathbf{x}}^k) \} \geq 0, \forall \mathbf{x}' \in X \\ (\mathbf{y}' - \tilde{\mathbf{y}}^k) \{ \theta_2(\tilde{\mathbf{y}}^k) - B^T [\tilde{\lambda}^k - \beta(A\tilde{\mathbf{x}}^k + B\tilde{\mathbf{y}}^k - b)] + \\ S(\tilde{\mathbf{y}}^k - \tilde{\mathbf{y}}^k) \} \geq 0, \forall \mathbf{y}' \in Y \\ \tilde{\lambda}^k = \tilde{\lambda}^k - \beta(A\tilde{\mathbf{x}}^k + B\tilde{\mathbf{y}}^k - b) \end{cases} \quad (10)$$

先介绍一些引理和定理用以证明算法的收敛性。

假设 1 ^[15] 假设 $\{\tau_k\}_{k=0}^\infty$ 满足:

a. 对于任意的 $k > 0$, 有 $0 \leq \tau_k \leq \tau$, 其中, $\tau \in [0, 1)$;

b. 惯性近似松弛 ADMM 产生的序列满足

$$\sum_{k=0}^{\infty} \tau_k \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2 < \infty.$$

引理 6 令 $\tilde{\mathbf{w}}^k = (\tilde{\mathbf{x}}^k, \tilde{\mathbf{y}}^k, \tilde{\lambda}^k)$, 对任意的 $\mathbf{w}^* \in W^*$, 得到 ^[16]

$$(\tilde{\mathbf{w}}^k - \mathbf{w}^*)^T Q(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k) \geq \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 + (\tilde{\lambda}^k - \tilde{\lambda}^k)^T B(\tilde{\mathbf{y}}^k - \tilde{\mathbf{y}}^k)$$

其中, Q 是一个正半定矩阵,

$$Q = \begin{pmatrix} RI_{n_1} & \mathbf{0}_{n_1 \times n_2} & \mathbf{0}_{n_1 \times m} \\ \mathbf{0}_{n_2 \times n_1} & \beta B^T B + SI_{n_2} & \mathbf{0}_{n_2 \times m} \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}$$

证明 由于 $\mathbf{w}^* \in W^*$, $\tilde{\mathbf{x}} \in X$ 和 $\tilde{\mathbf{y}} \in Y$, 可得

$$\begin{cases} (\tilde{\mathbf{x}}^k - \mathbf{x}^*)^T (\theta_1(\mathbf{x}^*) - A^T \lambda^*) \geq 0 \\ (\tilde{\mathbf{y}}^k - \mathbf{y}^*)^T (\theta_2(\mathbf{y}^*) - B^T \lambda^*) \geq 0 \end{cases} \quad (11)$$

分别将式 (10) 和式 (11) 对应相加, 由函数 θ_1 和 θ_2 的单调性, 可得

$$\begin{cases} (A\tilde{\mathbf{x}}^k - A\mathbf{x}^*)^T (\tilde{\lambda}^k - \lambda^*) + (\tilde{\mathbf{x}}^k - \tilde{\mathbf{x}}^k) R(\tilde{\mathbf{x}}^k - \mathbf{x}^*) \geq \\ \beta(A\tilde{\mathbf{x}}^k - A\mathbf{x}^*)^T (B\tilde{\mathbf{y}}^k - B\mathbf{y}^k) \\ (B\tilde{\mathbf{y}}^k - B\mathbf{y}^*)^T (\tilde{\lambda}^k - \lambda^*) + (\tilde{\mathbf{y}}^k - \tilde{\mathbf{y}}^k) S(\tilde{\mathbf{y}}^k - \mathbf{y}^*) \geq 0 \end{cases} \quad (12)$$

将式 (10) 和式 (12) 相加, 由 $A\mathbf{x}^* + B\mathbf{y}^* - b = 0$, 可得

$$\frac{1}{\beta} (\tilde{\lambda}^k - \tilde{\lambda}^k) (\tilde{\lambda}^k - \lambda^*)^T + (\tilde{\mathbf{x}}^k - \tilde{\mathbf{x}}^k) R(\tilde{\mathbf{x}}^k - \mathbf{x}^*) + (\tilde{\mathbf{y}}^k - \tilde{\mathbf{y}}^k) S(\tilde{\mathbf{y}}^k - \mathbf{y}^*) \geq \beta(A\tilde{\mathbf{x}}^k - A\mathbf{x}^*)^T (B\tilde{\mathbf{y}}^k - B\mathbf{y}^k) \quad (13)$$

在式 (13) 的两边同时加上 $\beta(B\tilde{\mathbf{y}}^k - B\mathbf{y}^*)^T (B\tilde{\mathbf{y}}^k - B\mathbf{y}^k)$, 结合矩阵 Q 的定义和 $A\mathbf{x}^* + B\mathbf{y}^* - b = 0$, 可得

$$(\tilde{\mathbf{w}}^k - \mathbf{w}^*)^T Q(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k) \geq \beta(A\tilde{\mathbf{x}}^k + B\tilde{\mathbf{y}}^k - b)^T (B\tilde{\mathbf{y}}^k - B\mathbf{y}^k) \quad (14)$$

联合 (10) 和 (14), 可得

$$(\tilde{\mathbf{w}}^k - \mathbf{w}^*)^T Q(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k) \geq (\tilde{\lambda}^k - \tilde{\lambda}^k)^T (B\tilde{\mathbf{y}}^k - B\mathbf{y}^k) \quad (15)$$

$$(\tilde{\mathbf{w}}^k - \mathbf{w}^*)^T Q(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k) \geq \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 + (\tilde{\lambda}^k - \tilde{\lambda}^k)^T (B\tilde{\mathbf{y}}^k - B\mathbf{y}^k)$$

引理 6 得证。

引理 7 由式 (8) 可知, $\tilde{\mathbf{w}}^k = (\tilde{\mathbf{x}}^k, \tilde{\mathbf{y}}^k, \tilde{\lambda}^k)$, 对任意的 $\mathbf{w}^* \in W^*$, 可得

$$\begin{aligned} \|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2 &\geq \|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2 + \\ 2(\tilde{\lambda}^k - \tilde{\lambda}^k)^T (B\tilde{\mathbf{y}}^k - B\mathbf{y}^k) &+ \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 \end{aligned} \quad (16)$$

证明 由引理 6 和式 (10), 可得

$$\begin{aligned} \|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2 &= \|\tilde{\mathbf{w}}^k - \mathbf{w}^* - (\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k)\|_Q^2 = \\ \|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2 &- 2(\tilde{\mathbf{w}}^k - \mathbf{w}^*)^T Q(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k) + \\ \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 &\leq \|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2 - 2[(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k)^T Q + \\ (\tilde{\lambda}^k - \tilde{\lambda}^k)^T (B\tilde{\mathbf{y}}^k - B\mathbf{y}^k)] &+ \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 = \\ \|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2 &- 2(\tilde{\lambda}^k - \tilde{\lambda}^k)^T (B\tilde{\mathbf{y}}^k - B\mathbf{y}^k) - \\ \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 \end{aligned}$$

式 (16) 表明, $d(\tilde{\mathbf{w}}^k, \tilde{\mathbf{w}}^k) = \tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k$ 是 $\|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2$ 的梯度下降方向。

引理 8 令 $\tilde{\mathbf{w}}^k = (\tilde{\mathbf{x}}^k, \tilde{\mathbf{y}}^k, \tilde{\lambda}^k)$, 对任意的 $\mathbf{w}^* \in W^*$, 定义如下函数:

$$\mathbf{w}^{k+1}(\alpha) = \tilde{\mathbf{w}}^k - \alpha_k(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k) \quad (17)$$

$$\theta^k(\alpha) = \|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2 - \|\mathbf{w}^{k+1}(\alpha) - \mathbf{w}^*\|_Q^2 \quad (18)$$

$$\varphi(\tilde{\mathbf{w}}^k, \tilde{\mathbf{w}}^k) = \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 + (\tilde{\lambda}^k - \tilde{\lambda}^k)^T B(\tilde{\mathbf{y}}^k - \tilde{\mathbf{y}}^k) \quad (19)$$

$$\psi^k(\alpha) = 2\alpha\varphi(\tilde{\mathbf{w}}^k, \tilde{\mathbf{w}}^k) - \alpha^2 \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 \quad (20)$$

对任意给定的 $\mathbf{w}^* = (\mathbf{x}^*, \mathbf{y}^*, \lambda^*) \in W^*$, $\alpha \geq 0$, 都存在 $\theta^k(\alpha) \geq \psi^k(\alpha)$ 。

证明 由式 (17) ~ (20) 的定义可知

$$\begin{aligned} \|\mathbf{w}^{k+1}(\alpha) - \mathbf{w}^*\|_Q^2 &= \|\tilde{\mathbf{w}}^k - \alpha_k(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k) - \mathbf{w}^*\|_Q^2 = \\ \|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2 &- 2\alpha(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k)^T Q(\tilde{\mathbf{w}}^k - \mathbf{w}^*) + \\ \alpha_k^2 \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 &\leq \|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2 - 2\alpha\varphi(\tilde{\mathbf{w}}^k, \tilde{\mathbf{w}}^k) + \\ \alpha^2 \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 &= \|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2 - \psi^k(\alpha) \end{aligned}$$

因此, 引理 8 得证, 即 $\theta^k(\alpha) \geq \psi^k(\alpha)$ 。

由引理 8 可知, $\psi^k(\alpha)$ 是 $\theta^k(\alpha)$ 的一个下界, 且 $\theta^k(\alpha) \geq \psi^k(\alpha)$ 。从 $\psi^k(\alpha)$ 的定义可知, 这是一个关于 α 的二次函数, 且它的最大值

$$\alpha_k = \frac{\varphi(\tilde{\mathbf{w}}^k, \tilde{\mathbf{w}}^k)}{\|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2} = \frac{\|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 + (\tilde{\lambda}^k - \tilde{\lambda}^k)^T \mathbf{B}(\tilde{\mathbf{y}}^k - \tilde{\mathbf{y}}^k)}{\|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2}$$

即最优步长。

引理 9 设 $\{\tau_k\}_{k=0}^\infty$ 满足假设 1, 令 $\tilde{\mathbf{w}}^k = (\tilde{\mathbf{x}}^k, \tilde{\mathbf{y}}^k, \tilde{\lambda}^k)$ (由式 (8) 得到), 对任意的 $\mathbf{w}^* \in \mathbf{W}^*$, 有 $\lim_{k \rightarrow \infty} \|\mathbf{w}^k - \mathbf{w}^*\|_Q$ 存在。

证明 由引理 3 和式 (10) 可知,

$$\begin{aligned} \langle \tilde{\mathbf{w}}^k - \mathbf{w}^*, \mathbf{Q}(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k) \rangle &\leq \theta(u^*) - \theta(\tilde{u}^k) - \\ \langle \tilde{\mathbf{w}}^k - \mathbf{w}^*, \mathbf{F}(\tilde{\mathbf{w}}^k) \rangle &\leq \theta(u^*) - \theta(\tilde{u}^k) - \\ \langle \tilde{\mathbf{w}}^k - \mathbf{w}^*, \mathbf{F}(\tilde{\mathbf{w}}^k) \rangle - \|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|^2 &\leq \\ -\|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|^2 \end{aligned} \quad (21)$$

又由式 (8) 可得

$$\begin{aligned} \langle \tilde{\mathbf{w}}^k - \mathbf{w}^*, \mathbf{Q}(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k) \rangle &= \langle \tilde{\mathbf{w}}^k - \mathbf{w}^*, \mathbf{Q}(\tilde{\mathbf{w}}^k - \mathbf{w}^k - \\ \tau_k(\mathbf{w}^k - \mathbf{w}^{k-1})) \rangle &= \langle \tilde{\mathbf{w}}^k - \mathbf{w}^*, \mathbf{Q}(\tilde{\mathbf{w}}^k - \mathbf{w}^k) \rangle - \\ \tau_k \langle \tilde{\mathbf{w}}^k - \mathbf{w}^*, \mathbf{Q}(\mathbf{w}^k - \mathbf{w}^{k-1}) \rangle \end{aligned} \quad (22)$$

定义 $h_k = \|\mathbf{w}^k - \mathbf{w}^*\|_Q^2$, $\tilde{h}_k = \|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2$, $g_k = 2\tau_k \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2$, 计算可得

$$2\langle \tilde{\mathbf{w}}^k - \mathbf{w}^*, \mathbf{Q}(\tilde{\mathbf{w}}^k - \mathbf{w}^k) \rangle = \tilde{h}_k - h_k + \|\tilde{\mathbf{w}}^k - \mathbf{w}^k\|_Q^2 \quad (23)$$

$$2\langle \tilde{\mathbf{w}}^k - \mathbf{w}^*, \mathbf{Q}(\mathbf{w}^k - \mathbf{w}^{k-1}) \rangle = h_k - h_{k-1} + \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2 + 2\langle \tilde{\mathbf{w}}^k - \mathbf{w}^k, \mathbf{Q}(\mathbf{w}^k - \mathbf{w}^{k-1}) \rangle \quad (24)$$

结合式 (21)~(24), 整理可得

$$\begin{aligned} \tilde{h}_k - h_k - \tau_k(h_k - h_{k-1}) &\leq -\|\tilde{\mathbf{w}}^k - \mathbf{w}^k\|_Q^2 + \\ 2\tau_k \langle \tilde{\mathbf{w}}^k - \mathbf{w}^k, \mathbf{Q}(\mathbf{w}^k - \mathbf{w}^{k-1}) \rangle &+ \tau_k \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2 - \\ 2\|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2 &= -\|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 + (\tau_k^2 + \tau_k) \|\mathbf{w}^k - \\ \mathbf{w}^{k-1}\|_Q^2 - 2\|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2 &\leq -\|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 + \\ 2\tau_k \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2 - 2\|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2 &\leq \\ -\|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 + 2\tau_k \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2 \end{aligned} \quad (25)$$

根据不等式 (25) 可以推测出 $\tilde{h}_k - h_k \leq \tau_k(h_k - h_{k-1}) + g_k \leq \tau_k[h_k - h_{k-1}]_+ + g_k$, 其中, 对于 $t \in \mathbb{R}$, 有 $[t]_+ := \max\{t, 0\}$ 。因此, 有

$$\begin{aligned} [\tilde{h}_k - h_k]_+ &\leq \tau_k[h_k - h_{k-1}]_+ + g_k \leq \\ \tau^{k+1}[h_0 - h_{-1}]_+ &+ \sum_{j=0}^k \tau_k^j g_{k-j} \end{aligned} \quad (26)$$

根据假设 $\mathbf{w}^0 = \mathbf{w}^{-1}$, 且令 $p_k = h_k - h_{k-1}$ 。可得 $p_0 = [p_0]_+ = 0$ 和 $g_0 = 0$ 。根据式 (26), 有

$$\sum_{k=0}^{\infty} [p_0]_+ \leq \frac{1}{1-\tau} \sum_{k=0}^{\infty} g_k = \frac{1}{1-\tau} \sum_{k=1}^{\infty} g_k < \infty \quad (27)$$

其中, 第 2 个不等式由假设 1 的 b 得到。

令 $q_k := h_k - \sum_{j=1}^k [g_j]_+$, 由式 (27) 和 $g_k \geq 0$, 可得 q_k 是有下界的, 即

$$\begin{aligned} q_{k+1} &= h_{k+1} - [g_{k+1}]_+ - \sum_{j=1}^k [g_j]_+ \leq \\ h_{k+1} - g_{k+1} - \sum_{j=1}^k [g_j]_+ &= h_k - \sum_{j=1}^k [g_j]_+ = q_k \end{aligned}$$

当 $k \rightarrow \infty$, 序列 $\{q_k\}$ 收敛, 有极限

$$\lim_{k \rightarrow \infty} h_k = \sum_{j=1}^{\infty} [g_j]_+ + \lim_{k \rightarrow \infty} q_k$$

因此, $\{h_k\}$ 收敛, 即 $\lim_{k \rightarrow \infty} \|\mathbf{w}^k - \mathbf{w}^*\|_Q$ 存在。证毕。

引理 10 假设 $\{\tau_k\}_{k=0}^\infty$ 是满足假设 1 的非负参数序列, 且序列 $\{\mathbf{w}^k\}$ 是由算法 1 生成。如果 $0 < a < E(\xi_k) = \rho < b$, 则 $\sum_{k=0}^{\infty} \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 \leq \infty$, 有 $\lim_{k \rightarrow \infty} \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q = 0$ 。

证明 由引理 6、假设 1 和式 (9),

$$\begin{aligned} \|\mathbf{w}^{k+1} - \mathbf{w}^*\|_Q^2 &= \|\tilde{\mathbf{w}}^k - \eta_k \alpha_k (\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k) - \mathbf{w}^*\|_Q^2 = \\ \|\tilde{\mathbf{w}}^k - \mathbf{w}^*\|_Q^2 - 2\eta_k \alpha_k (\tilde{\mathbf{w}}^k - \mathbf{w}^*) \mathbf{Q}(\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k) &+ \\ \eta_k^2 \alpha_k^2 \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 &\leq \|\mathbf{w}^k - \mathbf{w}^*\|_Q^2 + \\ 2(\mathbf{w}^k - \mathbf{w}^*) (\tau_k(\mathbf{w}^k - \mathbf{w}^{k-1})) + \tau_k^2 \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2 &+ \\ (\tilde{\lambda}^k - \tilde{\lambda}^k)^T \mathbf{B}(\tilde{\mathbf{y}}^k - \tilde{\mathbf{y}}^k) + \eta_k^2 \alpha_k^2 \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 &\leq \\ \|\mathbf{w}^k - \mathbf{w}^*\|_Q^2 + 2(\mathbf{w}^k - \mathbf{w}^*) (\tau_k(\mathbf{w}^k - \mathbf{w}^{k-1})) &+ \\ \tau_k^2 \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2 - \frac{1}{2} \eta_k (2 - \eta_k) \alpha_k (\|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 &+ \\ (\tilde{\lambda}^k - \tilde{\lambda}^k)^T \mathbf{B}(\tilde{\mathbf{y}}^k - \tilde{\mathbf{y}}^k)) \end{aligned} \quad (28)$$

可得

$$\begin{aligned} (\mathbf{w}^k - \mathbf{w}^*) (\tau_k(\mathbf{w}^k - \mathbf{w}^{k-1})) &= \frac{1}{2} \|\mathbf{w}^k - \mathbf{w}^*\|_Q^2 - \\ \frac{1}{2} \|\mathbf{w}^{k-1} - \mathbf{w}^*\|_Q^2 + \frac{1}{2} \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2 \end{aligned} \quad (29)$$

又因为当 $\tau_k < 1$ 时, $\tau_k^2 < \tau_k$, 将式 (29) 代入式 (28), 可得

$$\begin{aligned} \|\mathbf{w}^{k+1} - \mathbf{w}^*\|_Q^2 &\leq \|\mathbf{w}^k - \mathbf{w}^*\|_Q^2 + \|\mathbf{w}^k - \mathbf{w}^*\|_Q^2 - \|\mathbf{w}^{k-1} - \\ \mathbf{w}^*\|_Q^2 + \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2 + \tau_k^2 \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2 &- \\ \frac{1}{2} \eta_k (2 - \eta_k) \alpha_k \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 &\leq \|\mathbf{w}^k - \mathbf{w}^*\|_Q^2 + \\ \tau_k p_k + (\tau_k + \tau_k^2) \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2 - \\ \frac{1}{2} \eta_k (2 - \eta_k) \alpha_k \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 &\leq \|\mathbf{w}^k - \mathbf{w}^*\|_Q^2 + \\ [p_k]_+ + 2\tau_k \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2 - \\ \frac{1}{2} \eta_k (2 - \eta_k) \alpha_k \|\tilde{\mathbf{w}}^k - \tilde{\mathbf{w}}^k\|_Q^2 \end{aligned} \quad (30)$$

记 $c = \frac{1}{2}\eta_k(2-\eta_k)\alpha_k$, 式(30)可转变为

$$\|\tilde{\mathbf{w}}^k - \bar{\mathbf{w}}^k\|_Q^2 \leq \frac{1}{c} (\|\mathbf{w}^k - \mathbf{w}^*\|_Q^2 - \|\mathbf{w}^{k+1} - \mathbf{w}^*\|_Q^2 + [p_k]_+ + 2\tau_k \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2) \quad (31)$$

由于 $\xi_1, \xi_2, \dots, \xi_k, \dots$ 是独立的且均匀分布, 可得 $\eta_k, \mathbf{w}^k, \tilde{\mathbf{w}}^k, \tau_k \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2$ 也是独立的, 且由于

$$E(\eta_k) = E\left(\frac{1}{k} \sum_{i=1}^k \xi_i\right) = \frac{1}{k} \sum_{i=1}^k E(\xi_i) = \rho, E(\alpha) > \frac{1}{2}$$

有

$$E(\|\tilde{\mathbf{w}}^k - \bar{\mathbf{w}}^k\|_Q^2) \leq E(\|\mathbf{w}^k - \mathbf{w}^*\|_Q^2 - \|\mathbf{w}^{k+1} - \mathbf{w}^*\|_Q^2 + [p_k]_+ + 2\tau_k \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2) \leq mE(\|\mathbf{w}^k - \mathbf{w}^*\|_Q^2 - \|\mathbf{w}^{k+1} - \mathbf{w}^*\|_Q^2 + [p_k]_+ + 2\tau_k \|\mathbf{w}^k - \mathbf{w}^{k-1}\|_Q^2) \quad (32)$$

其中, $m = E\left(\frac{1}{c}\right) = \frac{4}{\rho(2-\rho)}$.

k 取 $0 \sim \infty$, 对式(32)的两边求和, 结合式(27)、假设1和引理9, 引理10得证。

定理1 假设 $\{\tau_k\}_{k=0}^\infty$ 是满足假设1的非负参数序列, 且序列 $\{\mathbf{w}^k\}$ 是由算法1生成的序列, 则有

a. $\sum_{k=1}^\infty \|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{y}}^k - b\|^2 < \infty$, 因此, 有 $\lim_{k \rightarrow \infty} E(\|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{y}}^k - b\|) = 0$;

b. $\theta_1(\tilde{\mathbf{x}}^k) + \theta_2(\tilde{\mathbf{y}}^k)$ ($k \rightarrow \infty$ 时) 收敛到问题(1)的最优值。

证明 a. 由引理10可得, $\sum_{k=0}^\infty \|\tilde{\mathbf{w}}^k - \bar{\mathbf{w}}^k\|_g^2 \leq \infty$,

因此, $\sum_{k=0}^\infty \|\tilde{\lambda}^k - \bar{\lambda}^k\|^2 < \infty$.

又由式(8)可知, $\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{y}}^k - b = \frac{1}{\beta}(\tilde{\lambda}^k - \bar{\lambda}^k)$

因此, $\sum_{k=1}^\infty \|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{y}}^k - b\|^2 = \sum_{k=1}^\infty \left\| \frac{1}{\beta}(\tilde{\lambda}^k - \bar{\lambda}^k) \right\|^2 < \infty$, 即 $\lim_{k \rightarrow \infty} E(\|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{y}}^k - b\|) = 0$.

b. 令 $\mathbf{w}^* = (\mathbf{x}^*, \mathbf{y}^*, \lambda^*) \in \mathbf{W}^*$, 在式(3)中, 令 $\tilde{\mathbf{w}}^k = (\tilde{\mathbf{x}}^k, \tilde{\mathbf{y}}^k, \lambda^*)$, 由 $\mathbf{F}$ 的定义可得

$$\theta_1(\tilde{\mathbf{x}}^k) + \theta_2(\tilde{\mathbf{y}}^k) - \theta_1(\mathbf{x}^*) - \theta_2(\mathbf{y}^*) \geq \langle \lambda^*, \mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{y}}^k - b \rangle$$

因此, 有

$$\liminf_{k \rightarrow \infty} (\theta_1(\tilde{\mathbf{x}}^k) + \theta_2(\tilde{\mathbf{y}}^k)) \geq \theta_1(\mathbf{x}^*) + \theta_2(\mathbf{y}^*) \quad (33)$$

另一方面, 令 $\mathbf{w} = \mathbf{w}^*$, 式(6)等价于式(34)。

$$\theta(u^*) - \theta(\tilde{u}^k) + \langle \mathbf{w}^* - \tilde{\mathbf{w}}^k, \mathbf{F}(\tilde{\mathbf{w}}^k) + \mathbf{Q}(\tilde{\mathbf{w}}^k - \bar{\mathbf{w}}^k) \rangle \geq 0 \quad (34)$$

结合式(34)

$$\lim_{k \rightarrow \infty} E(\|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{y}}^k - b\|) = 0 \text{ 和 } \sum_{k=0}^\infty \|\tilde{\mathbf{w}}^k - \bar{\mathbf{w}}^k\|_g^2 \leq \infty$$

有

$$\limsup_{k \rightarrow \infty} \theta_1(\tilde{\mathbf{x}}^k) + \theta_2(\tilde{\mathbf{y}}^k) \leq \theta_1(\mathbf{x}^*) + \theta_2(\mathbf{y}^*) \quad (35)$$

结合式(33)和式(35), 定理1的b得证。

定理2 如果 $0 < a < E(\xi_k) = \rho < b$, 令 $\tilde{\mathbf{w}}^k = (\tilde{\mathbf{x}}^k, \tilde{\mathbf{y}}^k, \tilde{\lambda}^k)$ 是由式(8)生成, 那么, 对任意的 $\mathbf{w}^*$, $\mathbf{w}^* \in \mathbf{W}^*$ 都是问题(1)的最优解。如果 $R > 0, S + \beta \mathbf{B}^T \mathbf{B} > 0$, 序列 $\{\mathbf{w}^k\}$ 依概率收敛到 $\mathbf{w}^\infty \in \mathbf{W}^*$ 。

证明 对任意的 $\mathbf{w}^* \in \mathbf{W}^*$, 由引理9可知, $\lim_{k \rightarrow \infty} \|\mathbf{w}^k - \mathbf{w}^*\|_Q$ 存在。因此, 可得序列 $\{\mathbf{Q}\mathbf{w}^k\}$ 是有界的。同时由矩阵 $\mathbf{Q}$ 的定义, 有 $\{\mathbf{R}\mathbf{x}^k\} \{(S + \beta \mathbf{B}^T \mathbf{B})\mathbf{y}^k\} \left\{ \frac{1}{\beta} \lambda^k \right\}$ 是有界的。当 $R > 0, S + \beta \mathbf{B}^T \mathbf{B} > 0$ 时, 序列 $\{E(\mathbf{w}^k)\}$ 是有界的。由式(32)可知, $\{E(\tilde{\mathbf{w}}^k)\}$ 也是有界的, 且 $\lim_{k \rightarrow \infty} E(\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\|_Q^2) = 0$ 。

由定理1和序列 $\{E(\mathbf{w}^k)\} \{E(\tilde{\mathbf{w}}^k)\}$ 是有界的, 可知 $\lim_{k \rightarrow \infty} E(\|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{y}}^k - b\|_Q^2 + \|\mathbf{B}(\mathbf{y}^k - \tilde{\mathbf{y}}^k)\|_Q^2) = 0 \quad (36)$

由 $\{E(\tilde{\mathbf{w}}^k)\}$ 是有界的, 可知它至少有1个聚点。令 $E(\mathbf{w}^\infty)$ 是 $\{E(\mathbf{w}^k)\}$ 其中的1个聚点, 那么, 有子序列 $\{E(\tilde{\mathbf{w}}^{k_j})\}$ 收敛到 $E(\mathbf{w}^\infty)$, 且子序列 $\{\tilde{\mathbf{w}}^{k_j}\}$ 依概率收敛到 $\mathbf{w}^\infty \in \mathbf{W}^*$, 记作 $\tilde{\mathbf{w}}^{k_j} \xrightarrow{P} \mathbf{w}^\infty$ 。

由引理5可知, 子序列 $\{\tilde{\mathbf{w}}^{k_j}\}$ 几乎必然收敛到 $\mathbf{w}^\infty$ 。因此, 它的子序列可标记为 $\{\tilde{\mathbf{w}}^{k_j}\}$, 记作

$$\lim_{j \rightarrow \infty} \tilde{\mathbf{w}}^{k_j} = \mathbf{w}^\infty \text{ a.s.} \quad (37)$$

由式(7), (36)和(37)可知,

$$\begin{cases} \lim_{j \rightarrow \infty} (\mathbf{x} - \tilde{\mathbf{x}}^{k_j})^T \{\theta_1(\tilde{\mathbf{x}}^{k_j}) - \mathbf{A}^T \tilde{\lambda}^{k_j}\} \geq 0, \forall \mathbf{x} \in \mathbf{X} \\ \lim_{j \rightarrow \infty} (\mathbf{y} - \tilde{\mathbf{y}}^{k_j})^T \{\theta_2(\tilde{\mathbf{y}}^{k_j}) - \mathbf{B}^T \tilde{\lambda}^{k_j}\} \geq 0, \forall \mathbf{y} \in \mathbf{Y} \\ \lim_{j \rightarrow \infty} (\mathbf{A}\tilde{\mathbf{x}}^{k_j} + \mathbf{B}\tilde{\mathbf{y}}^{k_j} - b) = 0 \end{cases}$$

$$\text{那么, 有 } \begin{cases} (\mathbf{x} - \mathbf{x}^\infty)^T \{\theta_1(\mathbf{x}^\infty) - \mathbf{A}^T \lambda^\infty\} \geq 0, \forall \mathbf{x} \in \mathbf{X} \\ (\mathbf{y} - \mathbf{y}^\infty)^T \{\theta_2(\mathbf{y}^\infty) - \mathbf{B}^T \lambda^\infty\} \geq 0, \forall \mathbf{y} \in \mathbf{Y} \\ \mathbf{A}\mathbf{x}^\infty + \mathbf{B}\mathbf{y}^\infty - b = 0 \end{cases}$$

因此, $\mathbf{w}^\infty \in \mathbf{W}^*$ 。

又由 $\lim_{j \rightarrow \infty} E(\tilde{\mathbf{w}}^{k_j}) = E(\mathbf{w}^\infty)$, 可知 $\lim_{l \rightarrow \infty} E(\|\tilde{\mathbf{w}}^{k_l} - \mathbf{w}^\infty\|_Q^2) = 0$ 。

根据式(31)可知, 对任意的 $\varepsilon > 0$, 存在 $l > 0$, 可得

$$E(\|\mathbf{w}^{k_l} - \tilde{\mathbf{w}}^{k_l}\|_Q^2) < \frac{\varepsilon}{2} \quad (38)$$

$$E(\|\tilde{\mathbf{w}}^{k_l} - \mathbf{w}^\infty\|_Q^2) < \frac{\varepsilon}{2} \quad (39)$$

因此, 对任意的 $k > k_l$, 由式(38)和式(39)可得

$$\begin{aligned} E(\|\mathbf{w}^k - \mathbf{w}^\infty\|_Q^2) &\leq E(\|\mathbf{w}^{k_l} - \mathbf{w}^\infty\|_Q^2) \leq \\ &E(\|\mathbf{w}^{k_l} - \tilde{\mathbf{w}}^{k_l}\|_Q^2) + E(\|\tilde{\mathbf{w}}^{k_l} - \mathbf{w}^\infty\|_Q^2) < \varepsilon \end{aligned} \quad (40)$$

记作 $\lim_{k \rightarrow \infty} \|\mathbf{w}^k - \mathbf{w}^\infty\|_Q^2 = 0$ 。

由引理 4 和式 (40)，对任意的 $\varepsilon > 0$ ，存在 $l > 0$ ，可知

$$P(\|\mathbf{w}^k - \mathbf{w}^\infty\|_Q^2 \geq \varepsilon) \leq \frac{E(\|\mathbf{w}^k - \mathbf{w}^\infty\|_Q^2)}{\varepsilon} < \varepsilon$$

因此， $\lim_{k \rightarrow \infty} P(\|\mathbf{w}^k - \mathbf{w}^\infty\|_Q^2 \geq \varepsilon) = 0$ ，并且序列 $\{\mathbf{w}^k\}$ 依概率收敛到 $\mathbf{w}^\infty$ ，记作 $\mathbf{w}^k \xrightarrow{P} \mathbf{w}^\infty$ 。

5 数值实验

数值实验所用软件为 Matlab 2017b，电脑配置为 Intel 四核 i7 2.4GHz CPU，并在 Vista 操作系统上运行 8GB RAM。

例 1 首先考虑财务和统计问题^[15]，

$$\min \left\{ \frac{1}{2} \|\mathbf{X} - \mathbf{C}\|_F^2 \mid \mathbf{X} \in S_+^n \cap S_B \right\} \quad (41)$$

其中， $S_B = \{\mathbf{H} \in \mathbb{R}^{n \times n} \mid \mathbf{H}^L \geq \mathbf{H} \geq \mathbf{H}^U\}$ 。

通过引入一个辅助变量 $\mathbf{Y}$ ，使得 $\mathbf{X} - \mathbf{Y} = 0$ ，问题 (41) 可以被重构为一个可分离的凸优化问题，

$$\min \left\{ \frac{1}{2} \|\mathbf{X} - \mathbf{C}\|_F^2 + \frac{1}{2} \|\mathbf{Y} - \mathbf{C}\|_F^2 \mid \mathbf{X} = \mathbf{Y}, \right. \\ \left. \mathbf{X} \in S_+^n, \mathbf{Y} \in S_B \right\} \quad (42)$$

显然，问题 (42) 是问题 (1) 的一个特殊形式，其中， $\rho(\mathbf{A}^T \mathbf{A}) = \rho(\mathbf{B}^T \mathbf{B}) = 1$ 。

令 $\boldsymbol{\lambda} \in \mathbb{R}^{n \times n}$ 是线性约束 $\mathbf{X} - \mathbf{Y} = 0$ 的拉格朗日乘子。对于一个给定的 $(\mathbf{X}^k, \mathbf{Y}^k, \boldsymbol{\lambda}^k)$ ，算法 1 产生的新的迭代 $(\mathbf{X}^{k+1}, \mathbf{Y}^{k+1}, \boldsymbol{\lambda}^{k+1})$ 由如下公式生成：

$$\left\{ \begin{aligned} \tilde{\mathbf{X}}^k &= \arg \min_{\mathbf{X} \in S_+^n} \left\{ \frac{1}{2} \|\mathbf{X} - \mathbf{C}\|^2 + \langle \tilde{\boldsymbol{\lambda}}^k, \mathbf{X} - \tilde{\mathbf{Y}}^k \rangle + \right. \\ &\quad \left. \frac{\beta}{2} \|\mathbf{X} - \tilde{\mathbf{Y}}^k\|^2 + \frac{R}{2} \|\mathbf{X} - \tilde{\mathbf{X}}^k\|^2 \right\} \\ \tilde{\mathbf{Y}}^k &= \arg \min_{\mathbf{Y} \in S_B} \left\{ \frac{1}{2} \|\mathbf{Y} - \mathbf{C}\|^2 + \langle \tilde{\boldsymbol{\lambda}}^k, \tilde{\mathbf{X}}^k - \mathbf{Y} \rangle + \right. \\ &\quad \left. \frac{\beta}{2} \|\tilde{\mathbf{X}}^k - \mathbf{Y}\|^2 + \frac{S}{2} \|\mathbf{Y} - \tilde{\mathbf{Y}}^k\|^2 \right\} \\ \tilde{\boldsymbol{\lambda}}^k &= \tilde{\boldsymbol{\lambda}}^k - \beta(\tilde{\mathbf{X}}^k - \tilde{\mathbf{Y}}^k) \end{aligned} \right. \quad (43)$$

迭代式 (43) 的 $\mathbf{X}$ -子问题通过奇异值分解 (SVD) 进行求解，它承担每次迭代过程中的主要计算负荷。迭代式 (43) 的 $\mathbf{Y}$ -子问题是一个投影，有如下形式：

$$\tilde{\mathbf{Y}}^k = P_{S_B} \left\{ \frac{1}{S+2} (\tilde{\mathbf{X}}^k - \tilde{\boldsymbol{\lambda}}^k + \mathbf{C} + S \tilde{\mathbf{Y}}^k) \right\}$$

其中， P_{S_B} 表示由欧几里得范数到集合 S_B 的投影。

据文献 [16]，在数值实验中，对于任意的 i, j ，设 $\mathbf{C} = (c_{ij})_{n \times n}$ 是一个满足 $c_{ij} \in (-1, 1)$ 的随机矩阵。对于每个给定 $\mathbf{C} = (c_{ij})_{n \times n}$ ，将测试 20 个随机实例。将 $\mathbf{X}^0$ 和 $\mathbf{Y}^0$ 设置为 $n \times n$ 独立性矩阵， $\boldsymbol{\lambda}^0$ 是一个 n 维零向量，并且将 $\varepsilon = \frac{\max(|\mathbf{w}^{k+1} - \mathbf{w}^k|)}{\max(|\mathbf{w}^0 - \tilde{\mathbf{w}}^0|)} \leq 10^{-4}$ 记为算法的终止准则。数值实验结果如表 1~4 所示。

表 1 参数 R, S 不同取值的数值结果比较

Tab.1 Comparison of numerical results on different values of parameters R and S

算法	$R=0.01, S=0.01$		$R=0.1, S=0.1$		$R=1, S=1$	
阶数 n	迭代次数	CPU时间/s	迭代次数	CPU时间/s	迭代次数	CPU时间/s
50	138	1.290	52	0.340	52	0.276
100	211	2.785	81	1.193	48	0.703
200	278	14.618	83	4.390	22	1.477

表 2 参数 τ_k 不同取值的数值结果比较

Tab.2 Comparison of numerical results with different values of parameters τ_k

算法	$\tau_k=0.3$		$\tau_k=0.5$		$\tau_k=0.8$	
阶数 n	迭代次数	CPU时间/s	迭代次数	CPU时间/s	迭代次数	CPU时间/s
50	31	0.512	52	0.340	47	0.410
100	47	0.758	81	1.193	117	1.831
200	83	4.224	83	4.390	151	7.959

表 3 期望值 ρ 不同取值的数值结果比较

Tab.3 Comparison of numerical results with different values of expected value ρ

算法	$\rho=1.7$		$\rho=1.8$		$\rho=1.9$	
阶数 n	迭代次数	CPU时间/s	迭代次数	CPU时间/s	迭代次数	CPU时间/s
50	51	0.299	53	0.540	52	0.340
100	82	1.112	86	1.056	81	1.193
200	108	5.910	85	4.337	83	4.390

表 4 例 1 的数值结果 ($\rho=1.9$)

Tab.4 Numerical results of example 1 ($\rho=1.9$)

算法阶数 n	IPR-ADMM		ePADM	
	k	s	k	s
50	52	0.340	97	0.391
100	81	1.193	134	1.252
200	83	4.390	189	8.084

表 1 为参数 R, S 取不同值时的数值实验结果; 表 2 为参数 τ_k 取不同值时的数值实验结果; 表 3 为期望值 ρ 取不同值时的数值实验结果。其中, n 表示不同的维数, 取 50, 100, 200。显然, 从表 1~3 可以看出, R, S 的取值越大, 迭代次数越少, 算法收敛所消耗的时间也越少; 而参数 τ_k 的值越小, 算法表现出的数值性能越好。此外, 随机变量的不同期望值 ρ 产生的迭代次数相近, 在 $\rho = 1.9$ 时表现相对较好。综上, 惯性技术和随机变量更新步长都有利于加速算法的收敛。

表 4 为不同维数下分别应用 IPR-ADMM 和 ePADM^[49] 解决该问题所用的时间和迭代次数。其中, $n=50, 100, 200$ 。 s 为迭代所用的时间, k 为迭代次数。惯性近似松弛 ADMM 算法中参数 $\tau_k = 0.5$, 随机变量 η_k 的期望值 $\rho = 1.9$ 。

图 1 为 $n=50, 100, 200$ 的条件下, IPR-ADMM 和 ePADM 算法的对比结果。其中, 横轴表示迭代次数, 纵轴表示停止准则, 即收敛停止时间。显

然, 由表 4 和图 1 可以看出, 算法 1 的性能明显比 ePADM 算法好, 因为它的迭代次数和计算时间要少得多; 并且从图 1 看出, n 的取值越大, 算法收敛越快, 越趋于稳定。

结果表明, 算法 1 对于解决问题 (1) 是有效的, 而且算法 1 的性能更良好, 实验结果展现了加速策略的有效性。

6 结 论

通过应用 PPM 算法求解 ADMM 分解的子问题, 并使用惯性外推项, 构建了一种用于求解线性约束可分离凸问题的乘数的惯性近似交替方向方法, 而且使用随机变量来加快收敛速度。在适当的假设下, 证明了该方法的全局收敛性。数值结果表明, 该算法是有效的, 收敛效果优于现有算法。

参考文献:

[1] GLOWINSKI R, MARROCCO A. Analyse numerique du champ magnetique d'un alternateur par elements finis et Sur-relaxation ponctuelle non lineaire[J]. *Computer methods in applied mechanics and engineering*, 1974, 3(1): 55–85.

[2] 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.

[3] ECKSTEIN J. Some saddle-function splitting methods for convex programming[J]. *Optimization Methods and Software*, 1994, 4(1): 75–83.

[4] ALVAREZ F. On the minimizing property of a second order dissipative system in Hilbert spaces[J]. *SIAM Journal on Control and Optimization*, 2000, 38(4): 1102–1119.

[5] LORENZ D A, POCK T. An inertial forward-backward algorithm for monotone inclusions[J]. *Journal of Mathematical Imaging and Vision*, 2015, 51(2): 311–325.

[6] MOUDAFI A, OLINY M. Convergence of a splitting inertial proximal method for monotone operators[J]. *Journal of Computational and Applied Mathematics*, 2003, 155(2): 447–454.

[7] ALVAREZ F, ATTOUCH H. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping[J]. *Set-Valued Analysis*, 2001, 9(1/2): 3–11.

[8] ATTOUCH H, CABOT A. Convergence of a relaxed

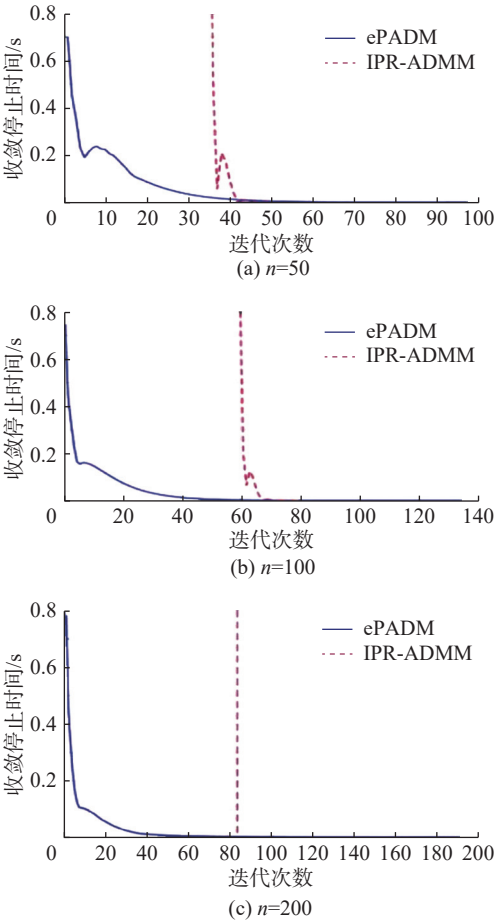


图 1 ePADM 和 IPR-ADMM 算法对比 ($\rho=1.9$)

Fig. 1 Comparison of ePADM and IPR-ADMM algorithms ($\rho=1.9$)

- inertial proximal algorithm for maximally monotone operators[J]. *Mathematical Programming*, 2020, 184(1): 243–287.
- [9] BOT R I, CSETNEK E R, HENDRICH C. Inertial Douglas-rachford splitting for monotone inclusion problems[J]. *Applied Mathematics and Computation*, 2015, 256: 472–487.
- [10] JIA H, LIU S, DANG Y. An inertial accelerated algorithm for solving split feasibility problem with multiple output sets[J]. *Journal of Mathematics*, 2021, 2021: 1–12.
- [11] CUI F Y, TANG Y C, YANG Y. An inertial three-operator splitting algorithm with applications to image inpainting[J]. *Application of Set-Valued Analysis and Optimization*, 2019, 1(2): 113–134.
- [12] CHEN C H, CHAN R H, MA S Q, et al. Inertial proximal ADMM for linearly constrained separable convex optimization[J]. *SIAM Journal on Imaging Sciences*, 2015, 8(4): 2239–2267.
- [13] ATTOUCH H, SOUEYCATT M. Augmented lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. Applications to games, PDE's and control[J]. *Pacific Journal of Optimization*, 2008, 5(1): 17–37.
- [14] WANG Y L, YANG J F, YIN W T, et al. A new alternating minimization algorithm for total variation image reconstruction[J]. *SIAM Journal on Imaging Sciences*, 2008, 1(3): 248–272.
- [15] TAO M, YUAN X M, HE B S. Solving a class of matrix minimization problems by linear variational inequality approaches[J]. *Linear Algebra and its Applications*, 2011, 434(11): 2343–2352.
- [16] HE B S, LIAO L Z, WANG X. Proximal-like contraction methods for monotone variational inequalities in a unified framework I: Effective quadruplet and primary methods[J]. *Computational Optimization and Applications*, 2012, 5(1): 649–679.

(编辑: 石 瑛)

[上接第 176 页]

- [11] HILIBRAND A S, ROBBINS M. Adjacent segment degeneration and adjacent segment disease: the consequences of spinal fusion?[J]. *The Spine Journal*, 2004, 4(S6): S190–S194.
- [12] MARTIN B I, MIRZA S K, COMSTOCK B A, et al. Reoperation rates following lumbar spine surgery and the influence of spinal fusion procedures[J]. *Spine*, 2007, 32(3): 382–387.
- [13] 曾鹏, 王正勇, 滕奇志. 一种基于区域生长算法的脊椎椎体提取方法 [J]. *科学技术与工程*, 2014, 14(6): 222–225.
- [14] 盛孙仁, 张美超, 朱青安. 中上胸椎皮质骨通道螺钉固定的生物力学有限元仿真研究 [J]. *医用生物力学*, 2019, 34(3): 251–255.
- [15] 李孝林. 基于 CT 图像应用 Mimics 软件快速构建人体胸腰段骨骼有限元模型 [J]. *中国组织工程研究与临床康复*, 2009, 13(39): 7619–7622.
- [16] 邹任玲, 李雨辰, 胡秀枋, 等. 用于脊柱后路的可旋动椎弓根螺钉及固定系统 [P]. 中国, CN112263315A. 2021-01-26.
- [17] MAGERL F P. Stabilization of the lower thoracic and lumbar spine with external skeletal fixation[J]. *Clinical Orthopaedics and Related Research*, 1984(189): 125–141.
- [18] KRAG M H, VAN HAL M E, BEYNNON B D. Placement of transpedicular vertebral screws close to anterior vertebral cortex[J]. *Spine*, 1989, 14(8): 879–883.
- [19] 程哲, 何由, 王卫国, 等. 胸腰段爆裂性骨折内固定治疗的生物力学特点 [J]. *医用生物力学*, 2019, 34(5): 486–492.

(编辑: 石 瑛)