运筹学与控制论

DOI:10.11721/cqnuj20220112

求解三块可分凸优化问题的 Bregman Peaceman-Rachford 分裂法*

万家彤，彭建文

(重庆师范大学 数学科学学院，重庆 401331)

摘要:【目的】针对带有线性约束的三块可分凸优化问题，提出带有 Bregman 距离的 Peaceman-Rachford(PR)分裂法。【方法】在原始 PR 分裂法的基础上结合 Bregman 距离函数，并选择不同的松弛因子来更新拉格朗日乘子。【结果】当 Bregman 距离函数为 δ-强凸时，从变分不等式的角度建立了由算法产生的迭代序列的全局收敛性以及给出了在遍历意义下 $O(1/t)$ 的最坏收敛速率。【结论】所得结果推广了求解两块可分凸优化问题的 PR 算法，具有一定的理论意义。

关键词:凸优化;PR 分裂法;变分不等式;Bregman 距离

中图分类号:O224　　**文献标志码**:A　　**文章编号**:1672-6693(2022)01-0019-08

考虑如下两块带有线性约束的可分离凸优化问题:

$$\min\{\theta_1(\boldsymbol{x}_1)+\theta_2(\boldsymbol{x}_2)\,|\,\boldsymbol{A}_1(\boldsymbol{x}_1)+\boldsymbol{A}_2(\boldsymbol{x}_2)=\boldsymbol{b},\boldsymbol{x}_1\in\mathcal{X}_1,\boldsymbol{x}_2\in\mathcal{X}_2\},\tag{1}$$

其中:$\theta_1(\boldsymbol{x}_1):\mathbf{R}^{n_1}\to\mathbf{R}\cup(+\infty)$,$\theta_2(\boldsymbol{x}_2):\mathbf{R}^{n_2}\to\mathbf{R}\cup(+\infty)$是闭正常凸函数(不一定光滑);$\mathcal{X}_1\subset\mathbf{R}^{n_1}$,$\mathcal{X}_2\subset\mathbf{R}^{n_2}$是闭凸集;$\boldsymbol{A}_1\in\mathbf{R}^{l\times n_1}$,$\boldsymbol{A}_2\in\mathbf{R}^{l\times n_2}$,$\boldsymbol{b}\in\mathbf{R}^l$ 是给定的矩阵和向量。假设问题(1)的解集非空,则对应的增广拉格朗日函数为:

$$L_\beta(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{\lambda})=\theta_1(\boldsymbol{x}_1)+\theta_2(\boldsymbol{x}_2)-\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{A}_1\boldsymbol{x}_1+\boldsymbol{A}_2\boldsymbol{x}_2-\boldsymbol{b})+\frac{\beta}{2}\|\boldsymbol{A}_1\boldsymbol{x}_1+\boldsymbol{A}_2\boldsymbol{x}_2-\boldsymbol{b}\|^2,$$

其中:$\boldsymbol{\lambda}\in\mathbf{R}^l$ 为拉格朗日乘子,$\beta>0$ 是惩罚参数。

Gabay 和 Mercier[1]针对问题(1),最早提出交替方向乘子法(ADMM):

$$\begin{cases}\boldsymbol{x}_1^{k+1}:=\arg\min\{\theta_1(\boldsymbol{x}_1)-(\boldsymbol{\lambda}^k)^{\mathrm{T}}\boldsymbol{A}_1\boldsymbol{x}_1+\dfrac{\beta}{2}\|\boldsymbol{A}_1\boldsymbol{x}_1+\boldsymbol{A}_2\boldsymbol{x}_2^k-\boldsymbol{b}\|^2\,|\,\boldsymbol{x}_1\in\mathcal{X}_1\}\\ \boldsymbol{x}_2^{k+1}:=\arg\min\{\theta_2(\boldsymbol{x}_2)-(\boldsymbol{\lambda}^k)^{\mathrm{T}}\boldsymbol{A}_2\boldsymbol{x}_2+\dfrac{\beta}{2}\|\boldsymbol{A}_1\boldsymbol{x}_1^{k+1}+\boldsymbol{A}_2\boldsymbol{x}_2-\boldsymbol{b}\|^2\,|\,\boldsymbol{x}_2\in\mathcal{X}_2\}\text{。}\\ \boldsymbol{\lambda}^{k+1}:=\boldsymbol{\lambda}^k-\beta(\boldsymbol{A}_1\boldsymbol{x}_1^{k+1}+\boldsymbol{A}_2\boldsymbol{x}_2^{k+1}-\boldsymbol{b})\end{cases}\tag{2}$$

在 ADMM 的每次迭代中,$(\boldsymbol{x}_1,\boldsymbol{x}_2)$以 Gauss-Seidel 方式分解。因此,(2)式中的子问题可能有封闭形式的解。同时由于该算法的收敛速度较快,子问题规模较小从而计算效率较高,所以它已被广泛应用到统计学、机器学习等领域[2-5]。目前该算法已经被推广到多块区域的情形[6]。

ADMM 在目标函数 θ_1 和 θ_2 的子问题比较简单的情况下效果很好,但是如果 θ_1 和 θ_2 本身结构较为复杂,那么所产生的子问题可能并不容易求解。故对于不同的具体问题,ADMM 得到发展并产生了一些新的形式,例如线性 ADMM[7-8]、邻近 ADMM[9-10]等。Bregman 距离[11]可以看作是欧几里得距离的延伸,即:给定一个可微凸函数 ρ,对于任意的两个点 $\boldsymbol{x},\boldsymbol{y}\in\mathrm{dom}(\rho)$之间的 Bregman 距离定义为:

$$B_\rho(\boldsymbol{x},\boldsymbol{y}):=\rho(\boldsymbol{x})-\rho(\boldsymbol{y})-\langle\nabla\rho(\boldsymbol{y}),\boldsymbol{x}-\boldsymbol{y}\rangle\text{。}$$

Wang 等人[12]基于二次项的 Bregman 距离提出了 Bregman ADMM 统一框架:

* **收稿日期**:2021-07-14　**修回日期**:2022-01-10　**网络出版时间**:2022-02-09 17:36

资助项目:国家自然科学基金重大项目(No. 11991024);重庆英才-创新创业领军人才-"最优化理论与方法及其应用"创新创业示范团队项目(No. CQYC20210309536);重庆市高校创新研究群体项目(No. CXQT20014)

第一作者简介:万家彤,女,研究方向为最优化理论与算法,E-mail:1475827468@qq.com;**通信作者**:彭建文,男,教授,博士生导师,E-mail:jwpeng168@hotmail.com

网络出版地址:https://kns.cnki.net/kcms/detail/50.1165.N.20220209.1406.020.html

$$\begin{cases} \boldsymbol{x}_1^{k+1} := \arg\min\{\theta_1(\boldsymbol{x}_1) - (\boldsymbol{\lambda}^k)^{\mathrm{T}}\boldsymbol{A}_1\boldsymbol{x}_1 + \frac{\beta}{2}\|\boldsymbol{A}_1\boldsymbol{x}_1 + \boldsymbol{A}_2\boldsymbol{x}_2^k - \boldsymbol{b}\|^2 + B_\rho(\boldsymbol{x}_1, \boldsymbol{x}_1^k) \mid \boldsymbol{x}_1 \in \mathcal{X}_1\} \\ \boldsymbol{x}_2^{k+1} := \arg\min\{\theta_2(\boldsymbol{x}_2) - (\boldsymbol{\lambda}^k)^{\mathrm{T}}\boldsymbol{A}_2\boldsymbol{x}_2 + \frac{\beta}{2}\|\boldsymbol{A}_1\boldsymbol{x}_1^{k+1} + \boldsymbol{A}_2\boldsymbol{x}_2 - \boldsymbol{b}\|^2 + B_\psi(\boldsymbol{x}_2, \boldsymbol{x}_2^k) \mid \boldsymbol{x}_2 \in \mathcal{X}_2\} \\ \boldsymbol{\lambda}^{k+1} := \boldsymbol{\lambda}^k - \beta(\boldsymbol{A}_1\boldsymbol{x}_1^{k+1} + \boldsymbol{A}_2\boldsymbol{x}_2^{k+1} - \boldsymbol{b}) \end{cases}。$$

针对问题(1),Peaceman 和 Rachford[13]提出了 PR 分裂法,即:

$$\begin{cases} \boldsymbol{x}_1^{k+1} := \arg\min\{\theta_1(\boldsymbol{x}_1) - (\boldsymbol{\lambda}^k)^{\mathrm{T}}\boldsymbol{A}_1\boldsymbol{x}_1 + \frac{\beta}{2}\|\boldsymbol{A}_1\boldsymbol{x}_1 + \boldsymbol{A}_2\boldsymbol{x}_2^k - \boldsymbol{b}\|^2 \mid \boldsymbol{x}_1 \in \mathcal{X}_1\} \\ \boldsymbol{\lambda}^{k+\frac{1}{2}} := \boldsymbol{\lambda}^k - \beta(\boldsymbol{A}_1\boldsymbol{x}_1^{k+1} + \boldsymbol{A}_2\boldsymbol{x}_2^k - \boldsymbol{b}) \\ \boldsymbol{x}_2^{k+1} := \arg\min\{\theta_2(\boldsymbol{x}_2) - (\boldsymbol{\lambda}^{k+\frac{1}{2}})^{\mathrm{T}}\boldsymbol{A}_2\boldsymbol{x}_2 + \frac{\beta}{2}\|\boldsymbol{A}_1\boldsymbol{x}_1^{k+1} + \boldsymbol{A}_2\boldsymbol{x}_2 - \boldsymbol{b}\|^2 \mid \boldsymbol{x}_2 \in \mathcal{X}_2\} \\ \boldsymbol{\lambda}^{k+1} := \boldsymbol{\lambda}^{k+\frac{1}{2}} - \beta(\boldsymbol{A}_1\boldsymbol{x}_1^{k+1} + \boldsymbol{A}_2\boldsymbol{x}_2^{k+1} - \boldsymbol{b}) \end{cases}。 \tag{3}$$

PR 分裂方法与经典 ADMM 的区别在于引入了乘子 $\boldsymbol{\lambda}^{k+\frac{1}{2}}$,这实际上是对拉格朗日乘子 $\boldsymbol{\lambda}$ 进行了两次迭代。当该方法收敛时,它具有比经典 ADMM 更快的收敛速度。之后,He 等人[14]又提出了一个严格收缩的 PRSM,在惩罚参数 β 前引入松弛因子 $\alpha \in (0,1)$,这样建立的 PRSM 算法在遍历和非遍历情况下均具有 $O(1/t)$ 的最坏收敛率,该算法迭代形式为:

$$\begin{cases} \boldsymbol{x}_1^{k+1} := \arg\min\{\theta_1(\boldsymbol{x}_1) - (\boldsymbol{\lambda}^k)^{\mathrm{T}}\boldsymbol{A}_1\boldsymbol{x}_1 + \frac{\beta}{2}\|\boldsymbol{A}_1\boldsymbol{x}_1 + \boldsymbol{A}_2\boldsymbol{x}_2^k - \boldsymbol{b}\|^2 \mid \boldsymbol{x}_1 \in \mathcal{X}_1\} \\ \boldsymbol{\lambda}^{k+\frac{1}{2}} := \boldsymbol{\lambda}^k - \alpha\beta(\boldsymbol{A}_1\boldsymbol{x}_1^{k+1} + \boldsymbol{A}_2\boldsymbol{x}_2^k - \boldsymbol{b}) \\ \boldsymbol{x}_2^{k+1} := \arg\min\{\theta_2(\boldsymbol{x}_2) - (\boldsymbol{\lambda}^{k+\frac{1}{2}})^{\mathrm{T}}\boldsymbol{A}_2\boldsymbol{x}_2 + \frac{\beta}{2}\|\boldsymbol{A}_1\boldsymbol{x}_1^{k+1} + \boldsymbol{A}_2\boldsymbol{x}_2 - \boldsymbol{b}\|^2 \mid \boldsymbol{x}_2 \in \mathcal{X}_2\} \\ \boldsymbol{\lambda}^{k+1} := \boldsymbol{\lambda}^{k+\frac{1}{2}} - \alpha\beta(\boldsymbol{A}_1\boldsymbol{x}_1^{k+1} + \boldsymbol{A}_2\boldsymbol{x}_2^{k+1} - \boldsymbol{b}) \end{cases}。$$

Du 等人[15]在变分不等式的框架下建立了 Symmetric Bregman ADMM 的全局收敛性,该算法的迭代形式为:

$$\begin{cases} \boldsymbol{x}_1^{k+1} := \arg\min\{\theta_1(\boldsymbol{x}_1) - (\boldsymbol{\lambda}^k)^{\mathrm{T}}\boldsymbol{A}_1\boldsymbol{x}_1 + \frac{\beta}{2}\|\boldsymbol{A}_1\boldsymbol{x}_1 + \boldsymbol{A}_2\boldsymbol{x}_2^k - \boldsymbol{b}\|^2 + B_\rho(\boldsymbol{x}_1, \boldsymbol{x}_1^k)\} \\ \boldsymbol{\lambda}^{k+\frac{1}{2}} := \boldsymbol{\lambda}^k - \alpha\beta(\boldsymbol{A}_1\boldsymbol{x}_1^{k+1} + \boldsymbol{A}_2\boldsymbol{x}_2^k - \boldsymbol{b}) \\ \boldsymbol{x}_2^{k+1} := \arg\min\{\theta_2(\boldsymbol{x}_2) - (\boldsymbol{\lambda}^{k+\frac{1}{2}})^{\mathrm{T}}\boldsymbol{A}_2\boldsymbol{x}_2 + \frac{\beta}{2}\|\boldsymbol{A}_1\boldsymbol{x}_1^{k+1} + \boldsymbol{A}_2\boldsymbol{x}_2 - \boldsymbol{b}\|^2 + B_\psi(\boldsymbol{x}_2, \boldsymbol{x}_2^k)\} \\ \boldsymbol{\lambda}^{k+1} := \boldsymbol{\lambda}^{k+\frac{1}{2}} - \alpha\beta(\boldsymbol{A}_1\boldsymbol{x}_1^{k+1} + \boldsymbol{A}_2\boldsymbol{x}_2^{k+1} - \boldsymbol{b}) \end{cases}。$$

算法(3)是针对问题(1)中目标函数为两块的情形,如果将情况推广到三块可分凸优化问题,则不能保证算法的收敛性。因此,本文考虑如下带有线性约束的三块可分凸优化问题:

$$\min\{\theta_1(\boldsymbol{x}_1) + \theta_2(\boldsymbol{x}_2) + \theta_3(\boldsymbol{x}_3) \mid \boldsymbol{A}_1\boldsymbol{x}_1 + \boldsymbol{A}_2\boldsymbol{x}_2 + \boldsymbol{A}_3\boldsymbol{x}_3 = \boldsymbol{b}, \boldsymbol{x}_i \in \mathcal{X}_i, i=1,2,3\}。 \tag{4}$$

其中:$\theta_i: \mathbf{R}^{n_i} \to \mathbf{R} \cup \{+\infty\}(i=1,2,3)$ 是闭正常凸函数(不一定光滑);$\mathcal{X}_i \subset \mathbf{R}^{n_i}(i=1,2,3)$ 是闭凸集;$\boldsymbol{A}_i \in \mathbf{R}^{n_i}(i=1,2,3)$,$\boldsymbol{b} \in \mathbf{R}^l$ 分别为给定的列满秩矩阵和向量。假设问题(4)的解集非空,本文针对该问题提出如下带 Bregman 距离的 PR 分裂法:

$$\begin{cases} \boldsymbol{x}_1^{k+1} := \arg\min\{\theta_1(\boldsymbol{x}_1) - (\boldsymbol{\lambda}^k)^{\mathrm{T}}\boldsymbol{A}_1\boldsymbol{x}_1 + \frac{\beta}{2}\left\|\boldsymbol{A}_1\boldsymbol{x}_1 + \sum_{i=2}^{3}\boldsymbol{A}_i\boldsymbol{x}_i^k - \boldsymbol{b}\right\|^2 + B_\rho(\boldsymbol{x}_1, \boldsymbol{x}_1^k)\} \\ \boldsymbol{\lambda}^{k+\frac{1}{2}} := \boldsymbol{\lambda}^k - \alpha\beta\left(\boldsymbol{A}_1\boldsymbol{x}_1^{k+1} + \sum_{i=2}^{3}\boldsymbol{A}_i\boldsymbol{x}_i^k - \boldsymbol{b}\right) \\ \boldsymbol{x}_2^{k+1} := \arg\min\{\theta_2(\boldsymbol{x}_2) - (\boldsymbol{\lambda}^{k+\frac{1}{2}})^{\mathrm{T}}\boldsymbol{A}_2\boldsymbol{x}_2 + \frac{\beta}{2}\|\boldsymbol{A}_1\boldsymbol{x}_1^{k+1} + \boldsymbol{A}_2\boldsymbol{x}_2 + \boldsymbol{A}_3\boldsymbol{x}_3^k - \boldsymbol{b}\|^2 + B_\psi(\boldsymbol{x}_2, \boldsymbol{x}_2^k)\} \\ \boldsymbol{x}_3^{k+1} := \arg\min\{\theta_3(\boldsymbol{x}_3) - (\boldsymbol{\lambda}^{k+\frac{1}{2}})^{\mathrm{T}}\boldsymbol{A}_3\boldsymbol{x}_3 + \frac{\beta}{2}\|\boldsymbol{A}_1\boldsymbol{x}_1^{k+1} + \boldsymbol{A}_2\boldsymbol{x}_2^k + \boldsymbol{A}_3\boldsymbol{x}_3 - \boldsymbol{b}\|^2 + B_\varphi(\boldsymbol{x}_3, \boldsymbol{x}_3^k)\} \\ \boldsymbol{\lambda}^{k+1} := \boldsymbol{\lambda}^{k+\frac{1}{2}} - \gamma\beta\left(\boldsymbol{A}_1\boldsymbol{x}_1^{k+1} + \sum_{i=2}^{3}\boldsymbol{A}_i\boldsymbol{x}_i^k - \boldsymbol{b}\right) \end{cases}。 \tag{5}$$

其中：$\beta>0$，$\alpha\in(-1,1)$，$\gamma\in(-\alpha,1]$，$\alpha+\dfrac{\alpha^2}{\alpha+\gamma}<\dfrac{1}{2}$。对引入的 Bregman 距离函数做出适当假设：在拉格朗日乘子的两次更新中选择不同松弛因子，使用并行方法求解 $\boldsymbol{x}_2$-子问题、$\boldsymbol{x}_3$-子问题。

值得注意的是，当算法的参数取值发生变化时，算法(5)会退化成另外不同的算法：

1）当 $\gamma=\alpha$，$\theta_3\equiv 0$，$\boldsymbol{A}_3\equiv 0$ 时，算法(5)退化为文献[15]中的 Symmetric Bregman ADMM 算法；

2）记 $\rho(\cdot)=\dfrac{1}{2}\|\cdot\|_{G_1}^2$，$\psi(\cdot)=\dfrac{1}{2}\|\cdot\|_{G_2}^2$，$\varphi(\cdot)=\dfrac{1}{2}\|\cdot\|_{G_3}^2$，$\boldsymbol{G}_i=\overline{\boldsymbol{G}}_i-(1-\tau)\beta\boldsymbol{A}_i^{\mathrm{T}}\boldsymbol{A}_i(i=1,2,3,\overline{\boldsymbol{G}}_i\succeq 0)$，当 $\gamma\in\left(0,\dfrac{1+\sqrt{5}}{2}\right)$，$\alpha=0$，$\tau>\dfrac{8+2\max\{1-\gamma,\gamma^2-\gamma\}}{5}$时，算法(5)退化为文献[16]中的 proximal ADMM 算法；

3）当 $\alpha=0$ 且不使用并行方法求解 $\boldsymbol{x}_2$-子问题、$\boldsymbol{x}_3$-子问题时，算法(5)就退化为文献[17]中的 Bregman ADMM 算法；

4）当 $\rho\equiv 0$，$\psi(\cdot)=\dfrac{\mu\beta\|\boldsymbol{A}_1\cdot\|_2^2}{2}$，$\varphi(\cdot)=\dfrac{\mu\beta\|\boldsymbol{A}_2\cdot\|_2^2}{2}(\mu>\alpha)$，$\alpha\in(0,1)$时，算法(5)退化为文献[18]中的带有邻近正则项的 SC-PRSM 算法。

同时，本文在 Bregaman 距离函数为 δ-强凸的条件下，从变分不等式的角度去分析了由算法(5)产生的迭代序列的全局收敛性，以及该算法在遍历意义下 $O(1/t)$的最坏收敛速率。

1 预备知识

问题(4)等价于一个变分不等式问题 $VL(\Omega,F,\theta)$，也即求$(\boldsymbol{x}_1^*,\boldsymbol{x}_2^*,\boldsymbol{x}_3^*,\boldsymbol{\lambda}^*)=\boldsymbol{\omega}^*\in\Omega$，满足：

$$\theta(\boldsymbol{u})-\theta(\boldsymbol{u}^*)+(\boldsymbol{\omega}-\boldsymbol{\omega}^*)^{\mathrm{T}}F(\boldsymbol{\omega}^*)\geqslant 0,\ \forall\,\boldsymbol{\omega}\in\Omega=\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{X}_3\times\mathbf{R}^l,\tag{6}$$

其中：$\theta(\boldsymbol{u})=\theta_1(\boldsymbol{x}_1)+\theta_2(\boldsymbol{x}_2)+\theta_3(\boldsymbol{x}_3)$，$\boldsymbol{u}=\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\\ \boldsymbol{x}_3\end{pmatrix}$，$\boldsymbol{\omega}=\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\\ \boldsymbol{x}_3\\ \boldsymbol{\lambda}\end{pmatrix}$，$F(\boldsymbol{\omega})=\begin{pmatrix}-\boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{\lambda}\\ -\boldsymbol{A}_2^{\mathrm{T}}\boldsymbol{\lambda}\\ -\boldsymbol{A}_3^{\mathrm{T}}\boldsymbol{\lambda}\\ \sum_{i=1}^{3}\boldsymbol{A}_i\boldsymbol{x}_i-\boldsymbol{b}\end{pmatrix}$。

显然这里的映射 $F(\boldsymbol{\omega})$是单调的，且很容易得到：

$$(\boldsymbol{\omega}-\widetilde{\boldsymbol{\omega}})^{\mathrm{T}}(F(\boldsymbol{\omega})-F(\widetilde{\boldsymbol{\omega}}))=0,\ \forall\,\boldsymbol{\omega},\widetilde{\boldsymbol{\omega}}\in\Omega。$$

所以(6)式等价于：

$$\theta(\boldsymbol{u})-\theta(\boldsymbol{u}^*)+(\boldsymbol{\omega}-\boldsymbol{\omega}^*)^{\mathrm{T}}F(\boldsymbol{\omega})\geqslant 0,\ \forall\,\boldsymbol{\omega}\in\Omega。$$

现设 Ω^* 是该变分不等式问题 $VI(\Omega,F,\theta)$的解集，且在问题(4)解集非空的情形下 Ω^* 是非空的。

为方便后面的收敛性分析，设 $t:=\sum_{i=2}^{3}n_i$，并定义下列矩阵：

$$\widetilde{\boldsymbol{A}}:=(\boldsymbol{A}_2,\boldsymbol{A}_3),\tag{7}$$

$$\boldsymbol{M}:=\begin{pmatrix}\boldsymbol{I}_t & 0\\ -\gamma\beta\widetilde{\boldsymbol{A}} & (\alpha+\gamma)\boldsymbol{I}_l\end{pmatrix},\tag{8}$$

$$\boldsymbol{Q}:=\begin{pmatrix}\beta\mathrm{diag}(\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}}) & -\alpha\widetilde{\boldsymbol{A}}^{\mathrm{T}}\\ -\widetilde{\boldsymbol{A}} & \dfrac{1}{\beta}\boldsymbol{I}_l\end{pmatrix},\tag{9}$$

$$\mathrm{diag}(\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}}):=\begin{pmatrix}\boldsymbol{A}_2^{\mathrm{T}}\boldsymbol{A}_2 & 0\\ 0 & \boldsymbol{A}_3^{\mathrm{T}}\boldsymbol{A}_3\end{pmatrix}。\tag{10}$$

引理 1　设 $\mathrm{diag}(\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}})$是由(10)式定义的矩阵，$\widetilde{\boldsymbol{A}}$是由(7)式定义的矩阵，则有：

$$2\mathrm{diag}(\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}})\succeq\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}}。\tag{11}$$

且如果$\dfrac{2\alpha^2+\alpha\gamma}{\alpha+\gamma}<\dfrac{1}{2}$，则有：

$$\mathrm{diag}(\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}})-\left(\alpha+\frac{\alpha^2}{\alpha+\gamma}\right)\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}}\succ 0。\tag{12}$$

证明 (11)式显然是成立的。如果$\frac{2\alpha^2+\alpha\gamma}{\alpha+\gamma}\leqslant 0$,则(12)式的正定性直接由$\mathrm{diag}(\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}})$和$\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}}$的正定性可得。如果$0<\frac{2\alpha^2+\alpha\gamma}{\alpha+\gamma}<\frac{1}{2}$,结合(11)式有:

$$\mathrm{diag}(\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}})-\left(\alpha+\frac{\alpha^2}{\alpha+\gamma}\right)\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}}\succeq\left(1-2\left(\alpha+\frac{\alpha^2}{\alpha+\gamma}\right)\right)\mathrm{diag}(\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}})\succ 0。$$

所以(12)式成立。 证毕

命题 1[11] 设$\rho(\boldsymbol{x})$是可微凸函数,$B_\rho(\boldsymbol{x},\boldsymbol{y})$是 Bregman 距离。$\forall\,\boldsymbol{x},\boldsymbol{y}\in\mathrm{dom}(\rho)$,则有:

1) 等式$B_\rho(\boldsymbol{x},\boldsymbol{y})=B_\rho(\boldsymbol{y},\boldsymbol{x})$在一般条件下是不成立的;

2) $B_\rho(\boldsymbol{x},\boldsymbol{y})\geqslant 0,B_\rho(\boldsymbol{x},\boldsymbol{x})=0$;

3) 当$\boldsymbol{y}$固定时,$B_\rho(\boldsymbol{x},\boldsymbol{y})$关于$\boldsymbol{x}$是凸的;

4) 如果$\rho(\boldsymbol{x})$是δ-强凸的,则$B_\rho(\boldsymbol{x},\boldsymbol{y})\geqslant\frac{\delta}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2$。

定义 1[16] 对于给定的凸函数ρ和矩阵$\boldsymbol{A}$,若$\rho(\boldsymbol{x})-\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{2}$为凸函数,则称$\rho\succeq\boldsymbol{A}$。

命题 2[16] 如果$\rho\succeq\boldsymbol{A}$,$\forall\,\boldsymbol{x},\boldsymbol{y}\in\mathrm{dom}(\rho)$,则有$B_\rho(\boldsymbol{x},\boldsymbol{y})-\frac{(\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{y})}{2}\geqslant 0$成立。

引理 2[16] 对于可微凸函数σ,$\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d}$是给定的向量,有下式成立:

$$(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}(\nabla\sigma(\boldsymbol{c})-\nabla\sigma(\boldsymbol{d}))=B_\sigma(\boldsymbol{a},\boldsymbol{d})-B_\sigma(\boldsymbol{a},\boldsymbol{c})+B_\sigma(\boldsymbol{b},\boldsymbol{c})-B_\sigma(\boldsymbol{b},\boldsymbol{d})。$$

特别地,当$\sigma(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x}$,$\boldsymbol{H}$是某对称半正定矩阵,则:

$$(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{H}(\boldsymbol{c}-\boldsymbol{d})=\frac{1}{2}[\|\boldsymbol{a}-\boldsymbol{d}\|_{\boldsymbol{H}}^2-\|\boldsymbol{a}-\boldsymbol{c}\|_{\boldsymbol{H}}^2+\|\boldsymbol{b}-\boldsymbol{c}\|_{\boldsymbol{H}}^2-\|\boldsymbol{b}-\boldsymbol{d}\|_{\boldsymbol{H}}^2]。\tag{13}$$

命题 3[16] 对于可微凸函数ρ,ψ,φ,有下列各式成立:

$$(\boldsymbol{x}_1-\widetilde{\boldsymbol{x}}_1^k)^{\mathrm{T}}(\nabla\rho(\boldsymbol{x}_1^k)-\nabla\rho(\widetilde{\boldsymbol{x}}_1^k))=B_\rho(\boldsymbol{x}_1,\widetilde{\boldsymbol{x}}_1^k)-B_\rho(\boldsymbol{x}_1,\boldsymbol{x}_1^k)+B_\rho(\widetilde{\boldsymbol{x}}_1^k,\boldsymbol{x}_1^k),$$
$$(\boldsymbol{x}_2-\widetilde{\boldsymbol{x}}_2^k)^{\mathrm{T}}(\nabla\psi(\boldsymbol{x}_2^k)-\nabla\psi(\widetilde{\boldsymbol{x}}_2^k))=B_\psi(\boldsymbol{x}_2,\widetilde{\boldsymbol{x}}_2^k)-B_\psi(\boldsymbol{x}_2,\boldsymbol{x}_2^k)+B_\psi(\widetilde{\boldsymbol{x}}_2^k,\boldsymbol{x}_2^k),$$
$$(\boldsymbol{x}_3-\widetilde{\boldsymbol{x}}_3^k)^{\mathrm{T}}(\nabla\varphi(\boldsymbol{x}_3^k)-\nabla\varphi(\widetilde{\boldsymbol{x}}_3^k))=B_\varphi(\boldsymbol{x}_3,\widetilde{\boldsymbol{x}}_3^k)-B_\varphi(\boldsymbol{x}_3,\boldsymbol{x}_3^k)+B_\varphi(\widetilde{\boldsymbol{x}}_3^k,\boldsymbol{x}_3^k)。$$

引理 3 $\forall\,\alpha+\frac{\alpha^2}{\alpha+\gamma}<\frac{1}{2}$,其中$\alpha\in(-1,1),\gamma\in(-\alpha,1]$,设$\boldsymbol{H}:=\boldsymbol{Q}\boldsymbol{M}^{-1}$,则$\boldsymbol{H}$是正定矩阵。

证明 由(8)式的定义有:$\boldsymbol{M}^{-1}=\begin{pmatrix}\boldsymbol{I}_t & 0\\ \frac{\gamma\beta}{\alpha+\gamma}\widetilde{\boldsymbol{A}} & \frac{1}{\alpha+\gamma}\boldsymbol{I}_l\end{pmatrix}$,则:

$$\boldsymbol{H}:=\boldsymbol{Q}\boldsymbol{M}^{-1}=\begin{pmatrix}\beta\mathrm{diag}(\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}}) & -\alpha\widetilde{\boldsymbol{A}}^{\mathrm{T}}\\ -\widetilde{\boldsymbol{A}} & \frac{1}{\beta}\boldsymbol{I}_l\end{pmatrix}\begin{pmatrix}\boldsymbol{I}_t & 0\\ \frac{\gamma\beta}{\alpha+\gamma}\widetilde{\boldsymbol{A}} & \frac{1}{\alpha+\gamma}\boldsymbol{I}_l\end{pmatrix}=\begin{pmatrix}\beta\mathrm{diag}(\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}})-\frac{\alpha\gamma}{\alpha+\gamma}\beta\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}} & -\frac{\alpha}{\alpha+\gamma}\widetilde{\boldsymbol{A}}^{\mathrm{T}}\\ -\frac{\alpha}{\alpha+\gamma}\widetilde{\boldsymbol{A}} & \frac{1}{(\alpha+\gamma)\beta}\boldsymbol{I}_l\end{pmatrix}。$$

所以,$\boldsymbol{H}$是对称矩阵。由(12)式和$\alpha+\gamma>0$,得:

$$\boldsymbol{H}=\begin{pmatrix}\beta\mathrm{diag}(\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}})-\left(\alpha+\frac{\alpha^2}{\alpha+\gamma}\right)\beta\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}} & 0\\ 0 & \frac{1}{2(\alpha+\gamma)\beta}\boldsymbol{I}_l\end{pmatrix}+\begin{pmatrix}\frac{2\alpha^2}{\alpha+\gamma}\beta\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}} & -\frac{\alpha}{\alpha+\gamma}\widetilde{\boldsymbol{A}}^{\mathrm{T}}\\ -\frac{\alpha}{\alpha+\gamma}\widetilde{\boldsymbol{A}} & \frac{1}{2(\alpha+\gamma)\beta}\boldsymbol{I}_l\end{pmatrix}\succeq$$
$$\begin{pmatrix}\beta\mathrm{diag}(\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}})-\left(\alpha+\frac{\alpha^2}{\alpha+\gamma}\right)\beta(\widetilde{\boldsymbol{A}}^{\mathrm{T}}\widetilde{\boldsymbol{A}}) & 0\\ 0 & \frac{1}{2(\alpha+\gamma)\beta}\boldsymbol{I}_l\end{pmatrix}\succ 0。$$

证毕

2 收敛性分析

为了更方便地分析算法(5)的收敛性,现定义如下辅助变量:

$$v^k=\begin{pmatrix}x_2^k\\x_3^k\\\lambda^k\end{pmatrix},\xi=\begin{pmatrix}\rho\\\psi\\\varphi\end{pmatrix},\tilde{\omega}^k=\begin{pmatrix}\tilde{x}_1^k\\\tilde{x}_2^k\\\tilde{x}_3^k\\\tilde{\lambda}^k\end{pmatrix}=\begin{pmatrix}x_1^{k+1}\\x_2^{k+1}\\x_3^{k+1}\\\lambda^k-\beta(A_1x_1^{k+1}+A_2x_2^k+A_3x_3^k-b)\end{pmatrix},\tilde{v}^k=\begin{pmatrix}\tilde{x}_2^k\\\tilde{x}_3^k\\\tilde{\lambda}^k\end{pmatrix}。\tag{14}$$

其中$(x_1^{k+1},x_2^{k+1},x_3^{k+1})$由算法(5)产生。根据$\tilde{\omega}^k$ 的定义，有：

$$\lambda^{k+\frac{1}{2}}=\lambda^k-\alpha(\lambda^k-\tilde{\lambda}^k)=\tilde{\lambda}^k+(\alpha-1)(\tilde{\lambda}^k-\lambda^k),\tag{15}$$

以及

$$\begin{aligned}\lambda^{k+1}&=\lambda^{k+\frac{1}{2}}-\gamma\beta(A_1\tilde{x}_1^k+A_2\tilde{x}_2^k+A_3\tilde{x}_3^k-b)=\\&\lambda^k-\alpha(\lambda^k-\tilde{\lambda}^k)-\gamma\beta(A_1\tilde{x}_1^k+A_2x_2^k+A_3x_3^k-b-A_2(x_2^k-\tilde{x}_2^k)-A_3(x_3^k-\tilde{x}_3^k))=\\&\lambda^k-\left[(\alpha+\gamma)(\lambda^k-\tilde{\lambda}^k)-\gamma\beta\sum_{i=2}^{3}A_i(x_i^k-\tilde{x}_i^k)\right],\end{aligned}\tag{16}$$

因此有 $v^{k+1}=v^k-M(v^k-\tilde{v}^k)$。

引理 4　设序列$\{\omega^k\}$由算法(5)产生，$\{\tilde{\omega}^k\}$，$\{\tilde{v}^k\}$由(14)式定义，Q 由(9)式定义，则$\{\tilde{\omega}^k\}\in\Omega$，$\forall\,\omega\in\Omega$，有下式成立：

$$\theta(u)-\theta(\tilde{u}^k)+(\omega-\tilde{\omega}^k)^{\mathrm T}F(\tilde{\omega}^k)\geqslant(v-\tilde{v}^k)^{\mathrm T}Q(v^k-\tilde{v}^k)+(u-\tilde{u}^k)^{\mathrm T}(\nabla\xi(u^k)-\nabla\xi(\tilde{u}^k)),\tag{17}$$

证明　由算法(5)中 x_1-子问题的一阶最优性条件得，$\forall\,x_1\in\mathcal{X}_1$ 存在$\tilde{x}_1^k\in\mathcal{X}_1$ 使得：

$$\theta_1(x_1)-\theta_1(\tilde{x}_1^k)+(x_1-\tilde{x}_1^k)^{\mathrm T}\{-A_1^{\mathrm T}(\lambda^k-\beta(A_1\tilde{x}_1^k+A_2x_2^k+A_3x_3^k-b))+\nabla\rho(\tilde{x}_1^k)-\nabla\rho(x_1^k)\}\geqslant0。$$

由(14)式中$\tilde{\lambda}^k=\lambda^k-\beta(A_1x_1^{k+1}+A_2x_2^k+A_3x_3^k-b)$，得：

$$\theta_1(x_1)-\theta_1(\tilde{x}_1^k)+(x_1-\tilde{x}_1^k)^{\mathrm T}[-A_1^{\mathrm T}\tilde{\lambda}^k+(\nabla\rho(\tilde{x}_1^k)-\nabla\rho(x_1^k))]\geqslant0。\tag{18}$$

对于 x_2-子问题，$\forall\,x_2\in\mathcal{X}_2$ 有$\tilde{x}_2^k\in\mathcal{X}_2$ 使得：

$$\theta_2(x_2)-\theta_2(\tilde{x}_2^k)+(x_2-\tilde{x}_2^k)^{\mathrm T}\{-A_2^{\mathrm T}\lambda^{k+\frac{1}{2}}+\beta A_2^{\mathrm T}(A_1\tilde{x}_1^k+A_2\tilde{x}_2^k+A_3x_3^k-b)+\nabla\psi(\tilde{x}_2^k)-\nabla\psi(x_2^k)\}\geqslant0。$$

将(15)式中 $\lambda^{k+\frac{1}{2}}=\tilde{\lambda}^k+(\alpha-1)(\tilde{\lambda}^k-\lambda^k)$和$\tilde{\lambda}^k=\lambda^k-\beta(A_1x_1^{k+1}+A_2x_2^k+A_3x_3^k-b)$代入上式可得：

$$\theta_2(x_2)-\theta_2(\tilde{x}_2^k)+(x_2-\tilde{x}_2^k)^{\mathrm T}\{-A_2^{\mathrm T}\tilde{\lambda}^k-\alpha A_2^{\mathrm T}(\tilde{\lambda}^k-\lambda^k)+\beta A_2^{\mathrm T}A_2(\tilde{x}_2^k-x_2^k)+\nabla\psi(\tilde{x}_2^k)-\nabla\psi(x_2^k)\}\geqslant0。\tag{19}$$

同理，$\forall\,x_3\in X_3$ 可得：

$$\theta_3(x_3)-\theta_3(\tilde{x}_3^k)+(x_3-\tilde{x}_3^k)^{\mathrm T}\{-A_3^{\mathrm T}\tilde{\lambda}^k-\alpha A_3^{\mathrm T}(\tilde{\lambda}^k-\lambda^k)+\beta A_3^{\mathrm T}A_3(\tilde{x}_3^k-x_3^k)+\nabla\varphi(\tilde{x}_3^k)-\nabla\varphi(x_3^k)\}\geqslant0。\tag{20}$$

最后，将$\tilde{\lambda}^k=\lambda^k-\beta(A_1x_1^{k+1}+A_2x_2^k+A_3x_3^k-b)$改写成下式：

$$(\lambda-\tilde{\lambda}^k)^{\mathrm T}\left[(A_1\tilde{x}_1^k+A_2\tilde{x}_2^k+A_3\tilde{x}_3^k-b)-\sum_{i=2}^{3}A_i(\tilde{x}_i^k-x_i^k)+\frac{1}{\beta}(\tilde{\lambda}^k-\lambda^k)\right]\geqslant0,\ \forall\,\lambda\in\mathbf{R}^l。\tag{21}$$

于是结合(18)～(21)式可得(17)式成立。　　证毕

引理 5　设序列$\{\omega^k\}$由算法(5)产生，$G=Q^{\mathrm T}+Q-M^{\mathrm T}HM$，$\{\tilde{\omega}^k\}$，$\{\tilde{v}^k\}$由(14)式定义，则 $\forall\,\omega\in\Omega$，有下式成立：

$$\theta(u)-\theta(\tilde{u}^k)+(\omega-\tilde{\omega}^k)^{\mathrm T}F(\omega)\geqslant T(\omega,\omega^{k+1})-T(\omega,\omega^k)+T^{\delta}(\tilde{\omega}^k,\omega^k)。\tag{22}$$

其中：$T(\omega_1,\omega_2)=B_\xi(u_1,u_2)+\dfrac{1}{2}\|v_1-v_2\|_H^2$，$T^{\delta}(\omega_1,\omega_2)=B_\xi(u_1,u_2)+\dfrac{1}{2}\|v_1-v_2\|_G^2$。

证明　将 $Q=HM$，$M(v^k-\tilde{v}^k)=v^k-v^{k+1}$代入(17)式，再根据 $F(\omega)$的单调性得：

$$\begin{aligned}&\theta(u)-\theta(\tilde{u}^k)+(\omega-\tilde{\omega}^k)^{\mathrm T}F(\omega)\geqslant\theta(u)-\theta(\tilde{u}^k)+(\omega-\tilde{\omega}^k)^{\mathrm T}F(\tilde{\omega}^k)\geqslant\\&(v-\tilde{v}^k)^{\mathrm T}H(v^k-v^{k+1})+B_\xi(u,\tilde{u}^k)-B_\xi(u,u^k)+B_\xi(\tilde{u}^k,u^k),\end{aligned}\tag{23}$$

由(13)式得：

$$(v-\tilde{v}^k)^{\mathrm T}H(v^k-v^{k+1})=\frac{1}{2}(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2)+\frac{1}{2}(\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2),\tag{24}$$

因为 $G=Q^{\mathrm T}+Q-M^{\mathrm T}HM=H-(M^{\mathrm T}-I)H(M-I)$，则：

$$\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2=\|v^k-\tilde{v}^k\|_H^2-\|v^k-\tilde{v}^k-(v^k-v^{k+1})\|_H^2=$$

$$\|\boldsymbol{v}^k-\tilde{\boldsymbol{v}}^k\|_{\boldsymbol{H}}^2-\|\boldsymbol{v}^k-\tilde{\boldsymbol{v}}^k-\boldsymbol{M}(\boldsymbol{v}^k-\tilde{\boldsymbol{v}}^k)\|_{\boldsymbol{H}}^2=(\boldsymbol{v}^k-\tilde{\boldsymbol{v}}^k)^{\mathrm{T}}(\boldsymbol{H}-(\boldsymbol{M}^{\mathrm{T}}-\boldsymbol{I})\boldsymbol{H}(\boldsymbol{M}-\boldsymbol{I}))(\boldsymbol{v}^k-\tilde{\boldsymbol{v}}^k)=\|\tilde{\boldsymbol{v}}^k-\boldsymbol{v}^k\|_{\boldsymbol{G}}^2。\quad(25)$$

结合(23)~(25)式可得(22)式成立。 证毕

因为 $\boldsymbol{Q}^{\mathrm{T}}+\boldsymbol{Q}=\begin{pmatrix}2\beta\mathrm{diag}(\tilde{\boldsymbol{A}}^{\mathrm{T}}\tilde{\boldsymbol{A}}) & -(\alpha+1)\tilde{\boldsymbol{A}}^{\mathrm{T}}\\ -(\alpha+1)\tilde{\boldsymbol{A}} & \frac{2}{\beta}\boldsymbol{I}_l\end{pmatrix}$,以及:

$$\boldsymbol{M}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{M}=\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{M}=\begin{pmatrix}\beta\mathrm{diag}(\tilde{\boldsymbol{A}}^{\mathrm{T}}\tilde{\boldsymbol{A}}) & -\tilde{\boldsymbol{A}}^{\mathrm{T}}\\ -\alpha\tilde{\boldsymbol{A}} & \frac{1}{\beta}\boldsymbol{I}_l\end{pmatrix}\begin{pmatrix}\boldsymbol{I}_t & 0\\ -\gamma\beta\tilde{\boldsymbol{A}} & (\alpha+\gamma)\boldsymbol{I}_l\end{pmatrix}=\begin{pmatrix}\beta\mathrm{diag}(\tilde{\boldsymbol{A}}^{\mathrm{T}}\tilde{\boldsymbol{A}})+\gamma\beta\tilde{\boldsymbol{A}}^{\mathrm{T}}\tilde{\boldsymbol{A}} & -(\alpha+\gamma)\tilde{\boldsymbol{A}}^{\mathrm{T}}\\ -(\alpha+\gamma)\tilde{\boldsymbol{A}} & \frac{\alpha+\gamma}{\beta}\boldsymbol{I}_l\end{pmatrix}。$$

则由矩阵 $\boldsymbol{G}$ 的定义,得:

$$\boldsymbol{G}=\boldsymbol{Q}^{\mathrm{T}}+\boldsymbol{Q}-\boldsymbol{M}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{M}=\boldsymbol{M}^{\mathrm{T}}\boldsymbol{H}+\boldsymbol{H}\boldsymbol{M}-\boldsymbol{M}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{M}=\boldsymbol{H}-(\boldsymbol{M}^{\mathrm{T}}-\boldsymbol{I})\boldsymbol{H}(\boldsymbol{M}-\boldsymbol{I})=$$

$$\begin{pmatrix}\beta\mathrm{diag}(\tilde{\boldsymbol{A}}^{\mathrm{T}}\tilde{\boldsymbol{A}})-\gamma\beta\tilde{\boldsymbol{A}}^{\mathrm{T}}\tilde{\boldsymbol{A}} & -(1-\gamma)\tilde{\boldsymbol{A}}^{\mathrm{T}}\\ -(1-\gamma)\tilde{\boldsymbol{A}} & \frac{2-\alpha-\gamma}{\beta}\boldsymbol{I}_l\end{pmatrix}。$$

根据 $\alpha\in(-1,1)$,$\gamma\in(-\alpha,1]$和(11)式,得:

$$\boldsymbol{G}=\begin{pmatrix}\beta\mathrm{diag}(\tilde{\boldsymbol{A}}^{\mathrm{T}}\tilde{\boldsymbol{A}})-\beta\tilde{\boldsymbol{A}}^{\mathrm{T}}\tilde{\boldsymbol{A}} & 0\\ 0 & \frac{1-\alpha}{\beta}\boldsymbol{I}_l\end{pmatrix}+\begin{pmatrix}(1-\gamma)\beta\tilde{\boldsymbol{A}}^{\mathrm{T}}\tilde{\boldsymbol{A}} & -(1-\gamma)\tilde{\boldsymbol{A}}^{\mathrm{T}}\\ -(1-\gamma)\tilde{\boldsymbol{A}} & \frac{1-\gamma}{\beta}\boldsymbol{I}_l\end{pmatrix}\succeq$$

$$\begin{pmatrix}\beta\mathrm{diag}(\tilde{\boldsymbol{A}}^{\mathrm{T}}\tilde{\boldsymbol{A}})-\beta\tilde{\boldsymbol{A}}^{\mathrm{T}}\tilde{\boldsymbol{A}} & 0\\ 0 & \frac{1-\alpha}{\beta}\boldsymbol{I}_l\end{pmatrix}\succeq\begin{pmatrix}-\frac{1}{2}\beta\tilde{\boldsymbol{A}}^{\mathrm{T}}\tilde{\boldsymbol{A}} & 0\\ 0 & \frac{1-\alpha}{\beta}\boldsymbol{I}_l\end{pmatrix}。$$

由上面分析可以看出 $\boldsymbol{G}$ 是不定的。假定 ψ,φ 关于常数 σ 是强凸的,并且 σ 满足 $\sigma\tilde{\boldsymbol{I}}^{\mathrm{T}}\tilde{\boldsymbol{I}}-\frac{1}{2}\beta\tilde{\boldsymbol{A}}^{\mathrm{T}}\tilde{\boldsymbol{A}}\succ 0$,则有:$\boldsymbol{\eta}+\begin{pmatrix}0 & 0\\ 0 & \boldsymbol{H}\end{pmatrix}\succ 0$,$\boldsymbol{\eta}+\begin{pmatrix}0 & 0\\ 0 & \boldsymbol{G}\end{pmatrix}\succ 0$。

注 这里 $\boldsymbol{\eta}=(\rho,\psi,\varphi,0)^{\mathrm{T}}$ 可以看作是 $\boldsymbol{\omega}$ 的函数。

定理1 设序列$\{\boldsymbol{\omega}^k\}$由算法(5)产生,$\{\tilde{\boldsymbol{\omega}}^k\}$,$\{\tilde{\boldsymbol{v}}^k\}$由(14)式定义,则:$T(\boldsymbol{\omega}^*,\boldsymbol{\omega}^{k+1})\leqslant T(\boldsymbol{\omega}^*,\boldsymbol{\omega}^k)-T^{\delta}(\tilde{\boldsymbol{\omega}}^k,\boldsymbol{\omega}^k)$,$\forall\,\boldsymbol{\omega}^*\in\Omega^*$。

证明 在(22)式中设 $\boldsymbol{\omega}=\boldsymbol{\omega}^*\in\Omega^*$,并使用(7)式得:

$$\theta(\tilde{\boldsymbol{u}}^k)-\theta(\boldsymbol{u}^*)+(\tilde{\boldsymbol{\omega}}^k-\boldsymbol{\omega}^*)^{\mathrm{T}}F(\boldsymbol{\omega}^*)+T(\boldsymbol{\omega}^*,\boldsymbol{\omega}^{k+1})\leqslant T(\boldsymbol{\omega}^*,\boldsymbol{\omega}^k)-T^{\delta}(\tilde{\boldsymbol{\omega}}^k,\boldsymbol{\omega}^k)。$$

因为 $\theta(\tilde{\boldsymbol{u}}^k)-\theta(\boldsymbol{u}^*)+(\tilde{\boldsymbol{\omega}}^k-\boldsymbol{\omega}^*)^{\mathrm{T}}F(\boldsymbol{\omega}^*)\geqslant 0$,所以 $T(\boldsymbol{\omega}^*,\boldsymbol{\omega}^{k+1})\leqslant T(\boldsymbol{\omega}^*,\boldsymbol{\omega}^k)-T^{\delta}(\tilde{\boldsymbol{\omega}}^k,\boldsymbol{\omega}^k)$。 证毕

定理2 $\theta(\tilde{\boldsymbol{u}}_n)-\theta(\boldsymbol{u})+(\tilde{\boldsymbol{\omega}}_n-\boldsymbol{\omega})^{\mathrm{T}}F(\boldsymbol{\omega})\leqslant\frac{1}{n+1}T(\boldsymbol{\omega},\boldsymbol{\omega}^0)$,$\forall\,\boldsymbol{\omega}\in\Omega$,其中 $\tilde{\boldsymbol{\omega}}_n:=\frac{1}{n+1}\sum_{k=0}^{n}\tilde{\boldsymbol{\omega}}^k\in\Omega$。如果 $\boldsymbol{\eta}+\begin{pmatrix}0 & 0\\ 0 & \boldsymbol{H}\end{pmatrix}\succ 0$,$\boldsymbol{\eta}+\begin{pmatrix}0 & 0\\ 0 & \boldsymbol{G}\end{pmatrix}\succ 0$,则有$\{\boldsymbol{\omega}^k\}$收敛到 $\boldsymbol{\omega}^{\infty}\in\Omega^*$。

证明 对(22)式求和可得:

$$(n+1)\theta(\boldsymbol{u})-\sum_{k=0}^{n}\theta(\tilde{\boldsymbol{u}}^k)+\Big((n+1)\boldsymbol{\omega}-\sum_{k=0}^{n}\tilde{\boldsymbol{\omega}}^k\Big)^{\mathrm{T}}F(\boldsymbol{\omega})+T(\boldsymbol{\omega},\boldsymbol{\omega}^0)\geqslant 0,$$

由 $\theta(\cdot)$的凸性可得:

$$\theta(\tilde{\boldsymbol{u}}_n)=\theta\Big(\frac{1}{n+1}\sum_{k=0}^{n}\tilde{\boldsymbol{u}}^k\Big)\leqslant\frac{1}{n+1}\sum_{k=0}^{n}\theta(\tilde{\boldsymbol{u}}^k),$$

所以

$$\theta(\boldsymbol{u})-\theta(\tilde{\boldsymbol{u}}_n)+(\boldsymbol{\omega}-\tilde{\boldsymbol{\omega}}_n)^{\mathrm{T}}F(\boldsymbol{\omega})+\frac{1}{n+1}T(\boldsymbol{\omega},\boldsymbol{\omega}^0)\geqslant 0,$$

即 $\theta(\tilde{\boldsymbol{u}}_n)-\theta(\boldsymbol{u})+(\tilde{\boldsymbol{\omega}}_n-\boldsymbol{\omega})^{\mathrm{T}}F(\boldsymbol{\omega})\leqslant\frac{1}{n+1}T(\boldsymbol{\omega},\boldsymbol{\omega}^0)$。

由定理 1 知$\{\boldsymbol{\omega}^k\}$是有界的且$\lim\limits_{k\to\infty}T^{\delta}(\tilde{\boldsymbol{\omega}}^k,\boldsymbol{\omega}^k)=0$，因为$\boldsymbol{\eta}+\begin{pmatrix}0 & 0\\0 & \boldsymbol{H}\end{pmatrix}\succ 0$，$\boldsymbol{\eta}+\begin{pmatrix}0 & 0\\0 & \boldsymbol{G}\end{pmatrix}\succ 0$，则$\exists\,\pi,\pi_{\delta}>0$，使得：

$$T(\boldsymbol{\omega}^1,\boldsymbol{\omega}^2)\geqslant\pi\ \|\boldsymbol{\omega}^1-\boldsymbol{\omega}^2\|_2^2,\ T^{\delta}(\boldsymbol{\omega}^1,\boldsymbol{\omega}^2)\geqslant\pi_{\delta}\ \|\boldsymbol{\omega}^1-\boldsymbol{\omega}^2\|_2^2,$$

所以$\lim\limits_{k\to\infty}\|\tilde{\boldsymbol{\omega}}^k-\boldsymbol{\omega}^k\|_2=0$。设$\boldsymbol{\omega}^{\infty}$是$\{\boldsymbol{\omega}^k\}$的一个聚点，$\{\boldsymbol{\omega}^{k_j}\}$是$\{\boldsymbol{\omega}^k\}$的子序列且收敛到$\boldsymbol{\omega}^{\infty}$。

因为$\lim\limits_{k\to\infty}\|\tilde{\boldsymbol{\omega}}^k-\boldsymbol{\omega}^k\|_2=0$，所以$\tilde{\boldsymbol{\omega}}^{k_j}\to\boldsymbol{\omega}^{\infty}$，设$k=k_j$代入(17)式且令$j\to\infty$，得：

$$\theta(\boldsymbol{u})-\theta(\boldsymbol{u}^{\infty})+(\boldsymbol{\omega}-\boldsymbol{\omega}^{\infty})^{\mathrm{T}}F(\boldsymbol{\omega}^{\infty})\geqslant 0,\ \forall\,\boldsymbol{\omega}\in\Omega,$$

因此$\boldsymbol{\omega}^{\infty}\in\Omega^*$，$T(\boldsymbol{\omega}^{k+1},\boldsymbol{\omega}^{\infty})\leqslant T(\boldsymbol{\omega}^k,\boldsymbol{\omega}^{\infty})-T^{\delta}(\tilde{\boldsymbol{\omega}}^k,\boldsymbol{\omega}^k)$。设$\boldsymbol{\omega}^{\rho}$是$\{\boldsymbol{\omega}^k\}$的其余任意聚点，则

$$T(\boldsymbol{\omega}^{\rho},\boldsymbol{\omega}^{\infty})\leqslant\lim_{k\to\infty}\inf\ T(\boldsymbol{\omega}^k,\boldsymbol{\omega}^{\infty})\leqslant T(\boldsymbol{\omega}^{k_j},\boldsymbol{\omega}^{\infty})=T(\boldsymbol{\omega}^{\infty},\boldsymbol{\omega}^{\infty})=0,$$

所以$\boldsymbol{\omega}^{\rho}=\boldsymbol{\omega}^{\infty}$，即$\boldsymbol{\omega}^{\infty}$是唯一的聚点。　　证毕

3 小结

本文针对带线性约束的三块可分凸优化问题，提出一种带 Bregman 距离的 PR 分裂法，该算法推广了文献[16]中的 Symmetric Bregman ADMM 算法，同时该算法可视作多种 PR 分裂算法的统一。当 Bregman 距离函数为 δ-强凸时，本文在变分不等式的框架下建立了由此算法产生的迭代序列的全局收敛性以及该算法在遍历意义下 $O(1/t)$的最坏收敛速率。

参考文献：

[1] GABAY D, MERCIER B. A dual algorithm for the solution of nonlinear variational problems via finite element approximations[J]. Computers and Mathematics with Applications, 1976, 2(1): 17-40.

[2] DAI Y H, HAN D R, YUAN X M, et al. A sequential updating scheme of the lagrange multiplier for separable convex programming[J]. Mathematics of Computation, 2017, 86(303): 315-343.

[3] HAN D R, KONG W W, ZHANG W X. A partial splitting augmented lagrangian method for low patch-rank image decomposition[J]. Journal of Mathematical Imaging and Vision, 2015, 51(1): 145-160.

[4] SCHAEFFER H, OSHER S. A low patch-rank interpretation of texture[J]. SIAM Journal on Imaging Sciences, 2013, 6(1): 226-262.

[5] BECK A, TEBOULLE M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. SIAM Journal Imaging Sciences, 2009, 2(1): 183-202.

[6] CHEN C H, LI M, LIU X, et al. On the convergence of multi-block alternating direction method of multipliers and block coordinate descent method[EB/OL]. (2015-09-09)[2021-07-14]. http://www.optimization-online.org/DB_HTML/2015/08/5046.html.

[7] GAO B, MA F. Symmetric alternating direction method with indefinite proximal regularization for linearly constrained convex optimization[J]. Journal of Optimization Theory and Applications, 2018, 176(1): 178-204.

[8] WANG X F, YUAN X M. The linearized alternating direction method of multipliers for dantzig selector[J]. SIAM Journal on Scientific Computing, 2012, 34(5), 2792-2811.

[9] CHEN G, TEBOULE M. A proximal-based decomposition method for convex minimization problems[J]. Mathematical Programming, 1994, 64(1): 81-101.

[10] WU Z M, LIU F X, LI M. A proximal Peaceman-Rachford splitting method for solving the multi-block separable convex minimization problems[J]. International Journal of Computer Mathematics, 2019, 96(4): 708-728.

[11] BREGMAN L M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming[J]. USSR Computational Mathematics & Mathematical Physics, 1967, 7(3): 200-217.

[12] WANG H H, BANERJEE A. Bregman alternating direction method of multipliers[C]//MICHAEL I J, YANN L C, SARA A S. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2014.

[13] PEACEMAN D W, RACHFORD H H. The numerical solution of parabolic and elliptic differential equations[J]. Journal of the Society for Industrial & Applied Mathematics, 1955, 3(1): 28-41.

[14] HE B S, LIU H, WANG Z R, et al. A strictly contractive Peaceman-Rachford splitting method for convex programming[J]. SIAM Journal on Optimization, 2014, 24(3): 1011-1040.

[15] DU P B, JIANG H. On the convergence of Bregman ADMM with variational inequality[J]. IEEE Acess, 2020, 8(2020): 29608-

29615.

[16] SUN M, SUN H C. Improved proximal ADMM with partially parallel splitting for multi-block separable convex programming [J]. Journal of Applied Mathematics and Computing, 2018, 58(1/2): 151-181.

[17] 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.

[18] ROLAND G, STANLEY J O, WOTAO Y. Splitting methods in communication, imaging, science, and engineering [M]. Switzerland: Springer International Publishing, 2016.

Operations Research and Cybernetics

Bregman Peaceman-Rachford Splitting Method for Three-Block Separable Convex Optimization Problems

WAN Jiatong, PENG Jianwen

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

Abstract: [**Purposes**]For the three-block separable convex optimization problem with linear constraints, a Peaceman-Rachford (PR) splitting method with Bregman distance is proposed. [**Methods**]This method combines the Bregman distance function on the basis of the original PR splitting method, and selects different relaxation factors to update the Lagrangian multiplier. [**Findings**]When the Bregman distance function is δ-strongly convex, the global convergence of the iterative sequence generated by the algorithm is established from the perspective of variational inequality and the worst-case $O(1/t)$ convergence rate in the ergodic sense is given. [**Conclusions**]These results generalize the PR algorithm for solving the two-piece separable convex optimization problem, which has certain theoretical significance.

Keywords: convex optimization; Peaceman-Rachford splitting method; variational inequality; Bregman distance

(责任编辑　许　甲)