DOI: 10.15960/j.cnki.issn.1007-6093.2015.03.008

修正乘子交替方向法求解三个可分离算子的凸优化*

何炳生[1,2,†]

摘要 指出直接推广的经典乘子交替方向法对三个算子的问题不能保证收敛的原因, 并且给出将其改造成收敛算法的相应策略. 同时, 在一个统一框架下, 证明了修正的乘子交替方向法的收敛性和遍历意义下具有$O(1/t)$ 收敛速率.

关键词 凸优化，分裂收缩算法，变分不等式，统一框架，收敛速率

中图分类号 O221.2, O224

2010 数学分类号 90C25, 90C30, 90C33

Modified alternating directions method of multipliers for convex optimization with three separable functions*

HE Bingsheng[1,2,†]

Abstract In this paper, we indicate the reason of divergence, and illustrate the strategies which modify the alternating direction method of multipliers (ADMM) to a convergent one for the linearly constrained separable convex optimization with three individual functions. Finally, using a uniform framework, we give the simple proofs for the convergence and $O(1/t)$ convergence rate in the ergodic sense of the ADMM-like methods.

Keywords convex optimization, splitting contraction methods, variational inequality, uniform framework, convergence rate

Chinese Library Classification O221.4, O224

2010 Mathematics Subject Classification 90C25, 90C30, 90C33

0 引 言

经典的乘子交替方向法(Alternating Directions Method of Multipliers, ADMM) 是求解含两个可分离算子线性约束凸优化和单调变分不等式的有效工具. 人们很自然地想将这类算法推广到多个算子的问题上. 尽管直接推广的方法求解某些问题时表现不俗, 但在一般条件下要保证理论收敛不太可能. 交替方向法是求解凸优化的一类分裂收缩算法.

收稿日期：2015-05-03

*基金项目: 国家自然科学基金(No.11471156)

1. 南京大学数学系, 南京 210093；Department of Mathematics, Nanjing University, Nanjing 210093, China

2. 南京大学管理科学与工程国际研究中心, 南京 210093；International Centre of Management Science and Engineering, Nanjing University, Nanjing 210093, China

† 通信作者 E-mail: hebma@nju.edu.cn

这类方法的迭代过程中, 要求解的子问题往往是一个简单的凸优化. 它的目标函数是两个凸函数的和, 其中一个凸函数不一定光滑. 说这类子问题简单, 是指问题的解有显式表达式, 或是能求得一个符合要求的近似解. 通篇, 在分析算法收敛性时, 常常要用到下面的引理.

引理 0.1 设 $\mathcal{X} \subset \mathbb{R}^n$ 是闭凸集, $\theta(x)$ 和 $f(x)$ 都是 $\mathbb{R}^n \to \mathbb{R}$ 的凸函数, 其中 $f(x)$ 可微. 设凸优化问题 $\min\{\theta(x)+f(x)|x\in\mathcal{X}\}$ 的解集非空, 有

$$x^* = \arg\min\{\theta(x)+f(x)|x\in\mathcal{X}\} \tag{0.1a}$$

的充分必要条件是

$$x^*\in\mathcal{X}, \quad \theta(x)-\theta(x^*)+(x-x^*)^{\mathrm{T}}\nabla f(x^*)\geqslant 0, \quad \forall\, x\in\mathcal{X}. \tag{0.1b}$$

1 线性约束的凸优化

本文讨论的问题都和增广拉格朗日方法有关, 我们从线性约束的凸优化谈起. 设 $\mathcal{X} \subset \mathbb{R}^n$ 是闭凸集, $A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$. 考虑线性约束的凸优化问题

$$\min\{\theta(x) \mid Ax=b,\ x\in\mathcal{X}\}. \tag{1.1}$$

它的拉格朗日函数是定义在 $\mathcal{X}\times\mathbb{R}^m$ 上的

$$L(x,\lambda)=\theta(x)-\lambda^{\mathrm{T}}(Ax-b). \tag{1.2}$$

1.1 拉格朗日函数 (1.2) 的鞍点及与之等价的变分不等式

讨论问题(1.1)的拉格朗日函数(1.2)的鞍点. 如果点 $(x^*,\lambda^*)\in\mathcal{X}\times\mathbb{R}^m$ 满足

$$L_{\lambda\in\mathbb{R}^m}(x^*,\lambda)\leqslant L(x^*,\lambda^*)\leqslant L_{x\in\mathcal{X}}(x,\lambda^*),$$

则称为拉格朗日函数(1.2)的鞍点. 上面的不等式关系可以写成

$$\begin{cases} x^*=\arg\min\{L(x,\lambda^*)|x\in\mathcal{X}\}, \\ \lambda^*=\arg\max\{L(x^*,\lambda)|\lambda\in\mathbb{R}^m\}. \end{cases}$$

利用式(1.2)和引理 0.1, 上述优化问题的最优性条件是

$$\begin{cases} x^*\in\mathcal{X}, \quad \theta(x)-\theta(x^*)+(x-x^*)^{\mathrm{T}}(-A^{\mathrm{T}}\lambda^*)\geqslant 0, \quad \forall\, x\in\mathcal{X}, \\ \lambda^*\in\mathbb{R}^m, \quad (\lambda-\lambda^*)^{\mathrm{T}}(Ax^*-b)\geqslant 0, \quad \forall\,\lambda\in\mathbb{R}^m. \end{cases}$$

写成更紧凑的形式, 鞍点就是变分不等式

$$w^*\in\Omega, \quad \theta(x)-\theta(x^*)+(w-w^*)^{\mathrm{T}}F(w^*)\geqslant 0, \quad \forall\, w\in\Omega \tag{1.3a}$$

的解, 其中

$$w=\begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(w)=\begin{pmatrix} -A^{\mathrm{T}}\lambda \\ Ax-b \end{pmatrix}, \quad \Omega=\mathcal{X}\times\mathbb{R}^m. \tag{1.3b}$$

注意到, 这样的 F, 总满足 $(w-\bar{w})^{\mathrm{T}}(F(w)-F(\bar{w}))=0$.

1.2 求解问题(1.1)的增广拉格朗日乘子法

凸优化问题(1.1)的增广拉格朗日函数由它的拉格朗日函数(1.2)和一个有等式线性约束的二次函数 $\frac{\beta}{2}\|Ax-b\|^2$ 组成, 即

$$\mathcal{L}_\beta(x,\lambda)=\theta(x)-\lambda^{\mathrm{T}}(Ax-b)+\frac{\beta}{2}\|Ax-b\|^2\,(\beta>0\ \text{是等式约束的罚参数}).$$

增广拉格朗日乘子法(Augmented Lagrangian Method, ALM)的 k-次迭代从一个给定的 λ^k 开始, 通过

$$\begin{cases}x^{k+1}=\arg\min\{\mathcal{L}_\beta(x,\lambda^k)\ \big|\ x\in\mathcal{X}\},\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}-b),\end{cases}\tag{1.4}$$

完成, 为下一次迭代提供了一个新的 λ^{k+1}. 在变分不等式(1.3a)和(1.3b)中, x 是原始变量, λ 是对偶变量. 在算法(1.4)中, x^{k+1} 是根据 λ^k 计算得来的结果. 因此, 我们称 x 为算法的中间变量, λ 为核心变量.

根据优化问题的最优性条件(0.1a)和(0.1b), 算法(1.4)提供的 $w^{k+1}=(x^{k+1},\lambda^{k+1})\in\Omega$ 满足

$$\begin{cases}\theta(x)-\theta(x^{k+1})+(x-x^{k+1})^{\mathrm{T}}\{-A^{\mathrm{T}}\lambda^k+\beta A^{\mathrm{T}}(Ax^{k+1}-b)\}\geqslant 0, \quad \forall x\in\mathcal{X},\\ (\lambda-\lambda^{k+1})^{\mathrm{T}}\{(Ax^{k+1}-b)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\}\geqslant 0, \quad \forall\lambda\in\mathbb{R}^m.\end{cases}$$

进一步利用关系式 $\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}-b)$ 和式(1.3a)和(1.3b)的记号, 可以把上式写紧凑的式子:

$$w^{k+1}\in\Omega,\theta(x)-\theta(x^{k+1})+(w-w^{k+1})^{\mathrm{T}}F(w^{k+1})\geqslant(\lambda-\lambda^{k+1})^{\mathrm{T}}\frac{1}{\beta}(\lambda^k-\lambda^{k+1}),\ \ \forall w\in\Omega.\tag{1.5}$$

变分不等式(1.5)中, 如果 $\lambda^k=\lambda^{k+1}$, 那么 w^{k+1} 就是变分不等式(1.3a)的解. 将(1.5)中的任意 $w\in\Omega$ 设成一个固定的解点 w^*, 我们得到

$$(\lambda^{k+1}-\lambda^*)^{\mathrm{T}}\frac{1}{\beta}(\lambda^k-\lambda^{k+1})\geqslant\theta(x^{k+1})-\theta(x^*)+(w^{k+1}-w^*)^{\mathrm{T}}F(w^{k+1}).\tag{1.6}$$

利用 $(w^{k+1}-w^*)^{\mathrm{T}}F(w^{k+1})=(w^{k+1}-w^*)^{\mathrm{T}}F(w^*)$ 和 w^* 的最优性, 即

$$\theta(x^{k+1})-\theta(x^*)+(w^{k+1}-w^*)^{\mathrm{T}}F(w^*)\geqslant 0,$$

推得式(1.6)的右端非负, 最终得到

$$(\lambda^{k+1}-\lambda^*)^{\mathrm{T}}(\lambda^k-\lambda^{k+1})\geqslant 0.\tag{1.7}$$

在恒等式

$$\|b\|^2=\|a\|^2-\|a-b\|^2-2b^{\mathrm{T}}(a-b)$$

中, 置 $a=\lambda^k-\lambda^*$ 和 $b=\lambda^{k+1}-\lambda^*$, 并利用式(1.7), 就有

$$\|\lambda^{k+1}-\lambda^*\|^2\leqslant\|\lambda^k-\lambda^*\|^2-\|\lambda^k-\lambda^{k+1}\|^2.$$

这个关于核心变量序列 $\{\lambda^k\}$ 收缩的不等式是ALM收敛的关键保证.

2 两个可分离算子的凸优化问题

考虑两个可分离算子的线性约束凸优化问题

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\, x\in\mathcal{X}, y\in\mathcal{Y}\}. \tag{2.1}$$

它的拉格朗日函数是

$$L^2(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^{\mathrm{T}}(Ax+By-b). \tag{2.2}$$

2.1 拉格朗日函数(2.2)的鞍点及与之等价的变分不等式

与第 1.1 节中同样的分析, 函数(2.2)的鞍点$((x^*,y^*),\lambda^*)$是变分不等式

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^{\mathrm{T}}F(w^*)\geqslant 0,\quad \forall w\in\Omega \tag{2.3a}$$

的解, 其中

$$w=\begin{pmatrix}x\\y\\\lambda\end{pmatrix},\quad u=\begin{pmatrix}x\\y\end{pmatrix},\quad F(w)=\begin{pmatrix}-A^{\mathrm{T}}\lambda\\-B^{\mathrm{T}}\lambda\\Ax+By-b\end{pmatrix}, \tag{2.3b}$$
$$\theta(u)=\theta_1(x)+\theta_2(y),\quad \Omega=\mathcal{X}\times\mathcal{Y}\times\mathbb{R}^m.$$

凸优化问题(2.1)的增广拉格朗日函数由它的拉格朗日函数(2.2)和等式线性约束的二次函数 $\frac{\beta}{2}\|Ax+By-b\|^2$ 组成, 即

$$\mathcal{L}^2_\beta(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^{\mathrm{T}}(Ax+By-b)+\frac{\beta}{2}\|Ax+By-b\|^2, \tag{2.4}$$

其中 $\beta>0$ 是等式约束的罚参数. 如果我们用ALM求解问题(2.1), k-次迭代从一个给定的 λ^k 开始, 通过

$$\begin{cases}(x^{k+1},y^{k+1})=\arg\min\big\{\mathcal{L}^2_\beta(x,y,\lambda^k)\;\big|\;x\in\mathcal{X},y\in\mathcal{Y}\big\},\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b),\end{cases} \tag{2.5}$$

提供一个新的 λ^{k+1}. 此时, 迭代中的核心变量还是 λ, 还会有收缩不等式(1.2). 方法的缺点是问题(2.1)的可分离性质没有得到利用.

2.2 求解问题(2.1)的乘子交替方向法

用ADMM[1,2]求解问题(2.1), k-次迭代从给定的(y^k,λ^k)开始, 通过

$$\begin{cases}x^{k+1}=\arg\min\{\mathcal{L}^2_\beta(x,y^k,\lambda^k)\,|\,x\in\mathcal{X}\},\\ y^{k+1}=\arg\min\{\mathcal{L}^2_\beta(x^{k+1},y,\lambda^k)\,|\,y\in\mathcal{Y}\},\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b)\end{cases} \tag{2.6}$$

完成, 为下一次迭代提供了一个新的(y^{k+1},λ^{k+1}).

- 在式(2.6)中, x-子问题和 y-子问题是利用了问题(2.1)的可分离性质分别求解的.
- 从式(2.5)到式(2.6), ADMM 可以看作是松弛了的ALM.

ADMM 方法(2.6)求解问题(2.1), x^{k+1} 是由给定的(y^k,λ^k)通过计算得到的, 因此是中间变量. 我们把变量 w 去掉中间变量剩余的部分称为核心变量.

在这个 ADMM 方法中, 核心变量则是 $v=(y,\lambda)$. y 和 λ 分别是核心变量中的原始部分和对偶部分. 由方法(2.6)生成的序列 $\{v^k=(y^k,\lambda^k)\}$ 满足

$$\|v^{k+1}-v^*\|_H^2 \leqslant \|v^k-v^*\|_H^2 - \|v^k-v^{k+1}\|_H^2, \tag{2.7}$$

其中

$$v=\begin{pmatrix} y\\ \lambda\end{pmatrix}, \qquad H=\begin{pmatrix}\beta B^{\mathrm{T}}B & 0\\ 0 & \frac{1}{\beta}I\end{pmatrix}.$$

不等式(2.7)表明核心变量序列 $\{v^k=(y^k,\lambda^k)\}$ 收缩, 它是保证ADMM收敛的关键不等式. 一个比较简单的证明可以见文献 [3] 的第 11 讲.

ADMM 被认为是求解含两个可分离算子的凸优化问题非常有效的方法[4], 近年来在图像处理、稀疏优化中发现大量应用. 交替方向法(2.6)可以解释为核心变量的原始部分 y 跟核心变量的对偶部分 λ 对弈. ADMM 收敛的原因是核心变量 $v=(y,\lambda)$中的原始部分 y 和对偶部分 λ 是对等的. 我们证明了 ADMM 同时具有遍历意义和非遍历意义下$O(1/t)$的收敛性质[5,6].

3 三个可分离算子的凸优化问题

我们继续考虑三个可分离算子的线性约束凸优化问题

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z) \mid Ax+By+Cz=b,\, x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\}. \tag{3.1}$$

它的拉格朗日函数是

$$L^3(x,y,z,\lambda)=\theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^{\mathrm{T}}(Ax+By+Cz-b). \tag{3.2}$$

3.1 拉格朗日函数(3.2)的鞍点及与之等价的变分不等式

同样的分析告知我们, 拉格朗日函数(3.2)的鞍点 $((x^*,y^*,z^*),\lambda^*)$是变分不等式

$$w^*\in\Omega, \quad \theta(u)-\theta(u^*)+(w-w^*)^{\mathrm{T}}F(w^*)\geqslant 0, \quad \forall w\in\Omega \tag{3.3a}$$

的解, 其中

$$w=\begin{pmatrix} x\\ y\\ z\\ \lambda\end{pmatrix}, \quad u=\begin{pmatrix} x\\ y\\ z\end{pmatrix}, \quad F(w)=\begin{pmatrix} -A^{\mathrm{T}}\lambda\\ -B^{\mathrm{T}}\lambda\\ -C^{\mathrm{T}}\lambda\\ Ax+By+Cz-b\end{pmatrix}, \tag{3.3b}$$

$$\theta(u)=\theta_1(x)+\theta_2(y)+\theta_3(z), \quad \Omega=\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}\times\mathbb{R}^m.$$

凸优化问题(3.1)的增广拉格朗日函数由它的拉格朗日函数(3.2)和等式线性约束的二次函数 $\frac{\beta}{2}\|Ax+By+Cz-b\|^2$ 组成, 即

$$\mathcal{L}_\beta^3(x,y,z,\lambda)=\theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^{\mathrm{T}}(Ax+By+Cz-b)+\frac{\beta}{2}\|Ax+By+Cz-b\|^2. \quad (3.4)$$

如果我们不将 y 和 z 分离, 而是把(y,z)捆在一起, 采用 ADMM求解, 迭代公式就是

$$\begin{cases} x^{k+1}=\arg\min\big\{\mathcal{L}_\beta^3(x,y^k,z^k,\lambda^k)\bigm| x\in\mathcal{X}\big\},\\ (y^{k+1},z^{k+1})=\arg\min\big\{\mathcal{L}_\beta^3(x^{k+1},y,z,\lambda^k)\bigm| y\in\mathcal{Y},z\in\mathcal{Z}\big\},\\ \lambda^{k+1}=\lambda^k-\beta\big(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b\big). \end{cases} \quad (3.5)$$

这样方法是收敛的. 缺点是原问题本来具备的 y 和 z 的可分离性质没有得到充分利用.

3.2 直接推广的 ADMM 对三个可分离算子的问题不能保证收敛

人们自然想到直接推广 ADMM 来求解三个可分离算子的问题. 如果我们用直接推广的ADMM求解问题(3.1), k-次迭代从一个给定的(y^k,z^k,λ^k)开始, 通过

$$\begin{cases} x^{k+1}=\arg\min\big\{\mathcal{L}_\beta^3(x,y^k,z^k,\lambda^k)\bigm| x\in\mathcal{X}\big\},\\ y^{k+1}=\arg\min\big\{\mathcal{L}_\beta^3(x^{k+1},y,z^k,\lambda^k)\bigm| y\in\mathcal{Y}\big\},\\ z^{k+1}=\arg\min\big\{\mathcal{L}_\beta^3(x^{k+1},y^{k+1},z,\lambda^k)\bigm| z\in\mathcal{Z}\big\},\\ \lambda^{k+1}=\lambda^k-\beta\big(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b\big), \end{cases} \quad (3.6)$$

给出新的 $(y^{k+1},z^{k+1},\lambda^{k+1})$. 为有讨论的共同语言, 我们建议只把不加任何附加条件的算法(3.6)称做 ADMM 的直接推广. 在任何一点与算法(3.6)不同的, 都是某种修正的格式.

相当长的一段时间内, 人们不知道算法(3.6)是否收敛. 直觉让我们认为不能保证收敛. 原因是, 用算法(3.6)处理三个算子的问题, 核心变量 $v=(y,z,\lambda)$中(y,z)是核心变量的原始部分, λ 是核心变量的对偶部分. 对核心变量的原始部分中的 y 和 z 两部分, 算法(3.6)是不公平的. 在求解 y-子问题时, 只能用(x^{k+1},z^k,λ^k)的信息, 因为 z^k 尚未更新; 而在求解 z-子问题时, 因为新的 y^{k+1} 已经有了, 就用$(x^{k+1},y^{k+1},\lambda^k)$的信息. 用直接推广的 ADMM 来求解三个可分离算子的问题(3.1), 很多实际计算还是成功的. 但对三个或者三个以上算子的问题有不收敛的例子[7].

4 改造 ADMM 成求解三个算子问题的收敛方法

既然算法(3.6)对一般的三个算子的问题(3.1)不能保证收敛的原因是它在求解子问题时对原始核心变量中的 y 和 z 两部分不公平, 因此, 我们考虑改造 ADMM 成求解三个算子问题的收敛方法的路, 就自然成了以下两条:

(i) 将算法(3.6)提供的 y^{k+1} 和 z^{k+1} 改造, 对算法(3.6)造成的不公平进行适当找补;

(ii) 对算法(3.6)本身进行改造, 使得它处理原始核心变量中 y 和 z 两部分时公平.

4.1 找补校正的方法

第一条路是对算法(3.6)生成的 y^{k+1} 和 z^{k+1} 进行找补(校正). 注意到, 略去子问题中的常数项, 直接推广的交替方向法(3.6)可以写成它的等价形式

$$\begin{cases} x^{k+1} = \arg\min\{\theta_1(x) - (\lambda^k)^{\mathrm{T}}(Ax) + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\}, \\ y^{k+1} = \arg\min\{\theta_2(y) - (\lambda^k)^{\mathrm{T}}(By) + \frac{\beta}{2}\|Ax^{k+1} + By + Cz^k - b\|^2 \mid y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\theta_3(z) - (\lambda^k)^{\mathrm{T}}(Cz) + \frac{\beta}{2}\|Ax^{k+1} + By^{k+1} + Cz - b\|^2 \mid z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{cases} \tag{4.1}$$

从式(4.1)可以看出, 开始执行 k-次迭代需要的是 (By^k, Cz^k, λ^k). 因此, 为下一次迭代我们也只要提供 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$. 这组 (By^{k+1}, Cz^{k+1})可以由

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \end{pmatrix} := \begin{pmatrix} By^k \\ Cz^k \end{pmatrix} - \mu \begin{pmatrix} I & -I \\ 0 & I \end{pmatrix} \begin{pmatrix} B(y^k - y^{k+1}) \\ C(z^k - z^{k+1}) \end{pmatrix}, \quad \mu \in (0,1] \tag{4.2}$$

提供. 这是变动最小、最简单的一种校正. 在实际计算中[8], 我们取 $\mu = [0.9, 0.95]$, 称这个方法为带高斯回代的交替方向法. 要在式(4.2)的右端取个小于 1 的常数 μ, 是因为从式(3.5)到式(3.6), 松弛了的是核心变量中原始部分(y 和 z)的子问题求解. 预测时给了过多的自由, 在总体找补的时候, 加个适当的缩小量 $\mu \in [0.9, 0.95]$是合理的.

4.2 强制平等的方法

按照第二条路, 对算法(3.6)进行改造, 使得它公平处理原始核心变量中 y 和 z 两部分. 一个简单的想法就是在求解 z-子问题时, 即使 y^{k+1} 已经有了, 我们还是只用 y^k 的信息. 这样就导致了下面的算法:

$$\begin{cases} x^{k+1} = \arg\min\big\{\mathcal{L}_\beta^3(x, y^k, z^k, \lambda^k) \bigm| x \in \mathcal{X}\big\}, \\ y^{k+1} = \arg\min\big\{\mathcal{L}_\beta^3(x^{k+1}, y, z^k, \lambda^k) \bigm| y \in \mathcal{Y}\big\}, \\ z^{k+1} = \arg\min\big\{\mathcal{L}_\beta^3(x^{k+1}, y^k, z, \lambda^k) \bigm| z \in \mathcal{Z}\big\}, \\ \lambda^{k+1} = \lambda^k - \beta\big(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b\big). \end{cases}$$

与收敛的方法(3.5)比较, 相当于将式(3.5)中的

$$(y^{k+1}, z^{k+1}) = \arg\min\big\{\mathcal{L}_\beta^3(x^{k+1}, y, z, \lambda^k) \bigm| y \in \mathcal{Y}, z \in \mathcal{Z}\big\},$$

改成了

$$\begin{cases} y^{k+1} = \arg\min\big\{\mathcal{L}_\beta^3(x^{k+1}, y, z^k, \lambda^k) \bigm| y \in \mathcal{Y}\big\}, \\ z^{k+1} = \arg\min\big\{\mathcal{L}_\beta^3(x^{k+1}, y^k, z, \lambda^k) \bigm| z \in \mathcal{Z}\big\}. \end{cases}$$

这样做, 原问题的可分离的性质利用了, 对 y 和 z 也公平了, 但也有点过于自由. 我们在平等平行处理 y 和 z 子问题的时候, 目标函数后面都额外再加一个正则项, 迭代法变成

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta^3(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta^3(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau\beta}{2}\|B(y - y^k)\|^2 \mid y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\mathcal{L}_\beta^3(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau\beta}{2}\|C(z - z^k)\|^2 \mid z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{cases} \tag{4.3}$$

其中 $\tau > 1$. 加正则项的意义就是在过度自由的情况下再加上个自我约束机制. 略去算法(4.3)中 y 和 z 子问题的常数项, 方法也可以改写成如下等价形式[9]:

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta^3(x, y^k, z^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b), \\ y^{k+1} = \arg\min\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^{\mathrm{T}} By + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\theta_3(z) - (\lambda^{k+\frac{1}{2}})^{\mathrm{T}} Cz + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \,|\, z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{cases} \tag{4.4}$$

其中 $\mu = \tau + 1 > 2$. 对这两个算法等价性有兴趣的读者, 只要验证它们 y-子问题和 z-子问题的最优性条件相同即可. 我们在 2010 年就提出了算法(4.4), 随即被 Stanley Osher[10]在降维问题上采用. 这一节提到的对三个算子问题改造过的 ADMM 类算法, 是文献 [8,9] 中处理多个算子方法的特例. 详细的收敛性证明见文献 [11].

5 统一框架下的算法描述和收敛条件

我们将要求解的问题都看成形如式(3.3)的变分不等式, 也就是

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^{\mathrm{T}} F(w^*) \geqslant 0, \quad \forall w \in \Omega.$$

核心变量是 v, 是在 w 中去掉中间变量 x 以后的部分分量. 把算法都归结为如下的预测-校正方法的统一框架:

[预测] 对给定的 v^k, 求得一个 $\tilde{w}^k$, 使得

$$\tilde{w}^k \in \Omega, \ \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^{\mathrm{T}} F(\tilde{w}^k) \geqslant (v - \tilde{v}^k)^{\mathrm{T}} Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{5.1a}$$

其中 Q 不一定对称, 但是要求 $Q^{\mathrm{T}} + Q$ 正定.

[校正] 给出新迭代点 v^{k+1} 的校正公式为

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{5.1b}$$

将收敛性条件也归纳成下面两条:

对算法(5.1a)和(5.1b)中的矩阵 Q 和 M, 有正定矩阵 H, 使得

$$HM = Q. \tag{5.2a}$$

并且

$$G = Q^{\mathrm{T}} + Q - M^{\mathrm{T}} HM \succ 0\,(\text{至少半正定}). \tag{5.2b}$$

在这个统一框架和相应的收敛性条件下, 能够帮助我们构造新的收敛算法[12], 证明收敛性就变得特别容易, 收敛速率的分析也迎刃而解. 这一节只验证第 4 节中修正的方法, 可以纳入算法框架(5.1a)和(5.1b), 并且满足收敛性条件(5.2a)和(5.2b). 证明见第 6 节.

5.1 找补校正的方法

我们用式(4.1)的输出 $x^{k+1}, y^{k+1}, z^{k+1}$ 定义辅助的预测变量 $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k)$, 其中

$$(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k) = (x^{k+1}, y^{k+1}, z^{k+1}), \tag{5.3a}$$

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b). \tag{5.3b}$$

根据优化问题的最优性条件(0.1a)和(0.1b), 用 $\tilde{x}^k = x^{k+1}$ 和式(5.3b), 可以将式(4.1)中 x-子问题的最优性条件写成

$$\tilde{x}^k \in \mathcal{X},\ \ \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^{\mathrm{T}}(-A^{\mathrm{T}}\tilde{\lambda}^k) \geqslant 0, \quad \forall x \in \mathcal{X}. \tag{5.4a}$$

根据同样的道理, 对 y-子问题和 z-子问题的最优性条件, 分别有

$$\tilde{y}^k \in \mathcal{Y},\ \ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^{\mathrm{T}}(-B^{\mathrm{T}}\tilde{\lambda}^k) \geqslant (y - \tilde{y}^k)^{\mathrm{T}}\beta B^{\mathrm{T}}B(y^k - \tilde{y}^k), \quad \forall y \in \mathcal{Y}, \tag{5.4b}$$

和

$$\begin{aligned} \tilde{z}^k \in \mathcal{Z}, \quad & \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^{\mathrm{T}}\big(-C^{\mathrm{T}}\tilde{\lambda}^k\big) \\ & \geqslant (z - \tilde{z}^k)^{\mathrm{T}}\big(\beta C^{\mathrm{T}}B(\tilde{y}^k - y^k) + \beta C^{\mathrm{T}}C(\tilde{z}^k - z^k)\big), \quad \forall z \in \mathcal{Z}. \end{aligned} \tag{5.4c}$$

根据关系式(5.3b), 有

$$\begin{aligned} \tilde{\lambda}^k \in \mathbb{R}^m, \quad & (\lambda - \tilde{\lambda}^k)^{\mathrm{T}}(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b) \\ & \geqslant (\lambda - \tilde{\lambda}^k)^{\mathrm{T}}\big\{-B(y^k - \tilde{y}^k) - C(z^k - \tilde{z}^k) + \tfrac{1}{\beta}(\lambda^k - \tilde{\lambda}^k)\big\}, \quad \forall \lambda \in \mathbb{R}^m. \end{aligned} \tag{5.4d}$$

利用变分不等式(3.3a)和(3.3b), 我们可以将式(5.4a)~(5.4d)中的结果表述成下面的引理.

引理 5.1 设 $(x^{k+1}, y^{k+1}, z^{k+1})$ 根据给定的 v^k 由式(4.1)生成. 那么, 由式(5.3a)和式(5.3b)定义的预测点 $\tilde{w}^k$ 满足

$$\tilde{w}^k \in \Omega,\ \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^{\mathrm{T}}F(\tilde{w}^k) \geqslant (v - \tilde{v}^k)^{\mathrm{T}}Q(v^k - \tilde{v}^k),\ \forall\, w \in \Omega, \tag{5.5}$$

其中

$$Q = \begin{pmatrix} \beta B^{\mathrm{T}}B & 0 & 0 \\ \beta C^{\mathrm{T}}B & \beta C^{\mathrm{T}}C & 0 \\ -B & -C & \frac{1}{\beta}I_m \end{pmatrix}. \tag{5.6}$$

我们接着来给出校正矩阵. 首先, 利用式(5.3a)和式(5.3b), 由(4.1)生成无需校正的 λ^{k+1} 可以写成

$$\begin{aligned} \lambda^{k+1} &= \lambda^k - \beta\big(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b\big) \\ &= \lambda^k - \big[-\beta B(y^k - \tilde{y}^k) - \beta C(z^k - \tilde{z}^k) + (\lambda^k - \tilde{\lambda}^k)\big]. \end{aligned} \tag{5.7}$$

此外, 矩阵 Q 可以写成分块的形式

$$Q = \begin{pmatrix} \beta Q_0 & 0 \\ -\mathcal{A} & \frac{1}{\beta}I_m \end{pmatrix}, \quad 其中 \quad Q_0 = \begin{pmatrix} B^{\mathrm{T}}B & 0 \\ C^{\mathrm{T}}B & C^{\mathrm{T}}C \end{pmatrix}, \quad \mathcal{A} = (B, C).$$

我们记

$$D_0 = \begin{pmatrix} B^{\mathrm{T}}B & 0 \\ 0 & C^{\mathrm{T}}C \end{pmatrix}, \quad \text{则有} \quad Q_0^{\mathrm{T}} + Q_0 = D_0 + \mathcal{A}^{\mathrm{T}}\mathcal{A}.$$

利用这些符号和式(5.7), 校正公式就是

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k),$$

其中

$$M = \begin{pmatrix} \mu Q_0^{-T}D_0 & 0 \\ -\beta\mathcal{A} & I \end{pmatrix}. \tag{5.8}$$

这样, 我们对第 4.1 节中的找补校正方法, 纳入了预测-校正方法框架(5.1a)和(5.1b).

最后我们来验证收敛性条件. 对式(5.5)中的矩阵 M, 它的逆矩阵是

$$M^{-1} = \begin{pmatrix} \frac{1}{\mu}D_0^{-1}Q_0^{\mathrm{T}} & 0 \\ \frac{1}{\mu}\beta\mathcal{A}D_0^{-1}Q_0^{\mathrm{T}} & I \end{pmatrix}. \tag{5.9}$$

是否有正定矩阵 H 使得 $HM = Q$, 只要验证矩阵

$$H = QM^{-1} = \begin{pmatrix} \beta Q_0 & 0 \\ -\mathcal{A} & \frac{1}{\beta}I_m \end{pmatrix}\begin{pmatrix} \frac{1}{\mu}D_0^{-1}Q_0^{\mathrm{T}} & 0 \\ \frac{1}{\mu}\beta\mathcal{A}D_0^{-1}Q_0^{\mathrm{T}} & I \end{pmatrix} = \begin{pmatrix} \frac{1}{\mu}\beta Q_0 D_0^{-1}Q_0^{\mathrm{T}} & 0 \\ 0 & \frac{1}{\beta}I \end{pmatrix}$$

是否正定. 事实上, 矩阵 H 是正定的. 进而我们验证矩阵 G 的正定性.

根据式(5.2b), 利用 $HM = Q$ 和 $Q_0^{\mathrm{T}} + Q_0 = D_0 + \mathcal{A}^{\mathrm{T}}\mathcal{A}$, 得到

$$G = \begin{pmatrix} (1-\mu)\beta(D_0 + \mathcal{A}^{\mathrm{T}}\mathcal{A}) & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix} \succ 0 \quad (\text{因为 } \mu \in (0,1)).$$

至此, 证明了找补校正的方法纳入框架(5.1a)和(5.1b), 相应的收敛性条件(5.2a)和(5.2b)也得到满足.

最后解释找补校正方法叫做带高斯回代的交替方向法的原因. 利用式(5.9)中 M^{-1} 的表达式, 原始变量的校正可以写成

$$Q_0^T\begin{pmatrix} y^{k+1} - y^k \\ z^{k+1} - z^k \end{pmatrix} = \mu D_0 \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{z}^k - z^k \end{pmatrix}. \tag{5.10}$$

注意到式(5.10)中的矩阵 Q_0^{T} 是一个上三角矩阵, 这样先后求得 z^{k+1}, y^{k+1} 的过程我们把它称为高斯回代. 由于式(5.10)的具体形式是

$$\begin{pmatrix} B^{\mathrm{T}} & 0 \\ 0 & C^{\mathrm{T}} \end{pmatrix}\begin{pmatrix} I & I \\ 0 & I \end{pmatrix}\begin{pmatrix} B(y^{k+1} - y^k) \\ C(z^{k+1} - z^k) \end{pmatrix} = \mu \begin{pmatrix} B^{\mathrm{T}} & 0 \\ 0 & C^{\mathrm{T}} \end{pmatrix}\begin{pmatrix} B(\tilde{y}^k - y^k) \\ C(\tilde{z}^k - z^k) \end{pmatrix},$$

并且开始下一次迭代只要预先有$(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$, 这组$(By^{k+1}, Cz^{k+1})$可以通过

$$\begin{pmatrix} I & I \\ 0 & I \end{pmatrix}\begin{pmatrix} B(y^{k+1} - y^k) \\ C(z^{k+1} - z^k) \end{pmatrix} = \mu \begin{pmatrix} B(\tilde{y}^k - y^k) \\ C(\tilde{z}^k - z^k) \end{pmatrix}$$

求得. 它的等价形式是

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \end{pmatrix} = \begin{pmatrix} By^{k} \\ Cz^{k} \end{pmatrix} - \mu \begin{pmatrix} I & -I \\ 0 & I \end{pmatrix} \begin{pmatrix} B(y^k - \tilde{y}^k) \\ C(z^k - \tilde{z}^k) \end{pmatrix}.$$

换句话说, 校正过程中并不需要求逆矩阵.

5.2 强制平等的方法

为了利用统一框架(5.1a)和(5.1b)及其相应的收敛性条件(5.2a)和(5.2b), 我们以式(4.4)的输出 x^{k+1}, y^{k+1}, z^{k+1} 定义辅助的预测变量 $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k)$, 其中

$$(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k) = (x^{k+1}, y^{k+1}, z^{k+1}), \tag{5.11a}$$

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b). \tag{5.11b}$$

根据优化问题的最优性条件(0.1a)和(0.1b), 用 $\tilde{x}^k = x^{k+1}$和式(5.11b), 可以将式(4.4)中 x-子问题的最优性条件写成

$$\tilde{x}^k \in \mathcal{X}, \;\; \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^{\mathrm{T}}(-A^{\mathrm{T}}\tilde{\lambda}^k) \geqslant 0, \quad \forall x \in \mathcal{X}. \tag{5.12a}$$

由于 $\tilde{\lambda}^k = \lambda^{k+\frac{1}{2}}$, 式(4.4)中 y-子问题和 z-子问题是平行求解的, 它们的最优性条件分别有

$$\tilde{y}^k \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^{\mathrm{T}}(-B^{\mathrm{T}}\tilde{\lambda}^k) \geqslant (y - \tilde{y}^k)^{\mathrm{T}}\mu\beta B^{\mathrm{T}}B(y^k - \tilde{y}^k), \; \forall y \in \mathcal{Y} \tag{5.12b}$$

和

$$\tilde{z}^k \in \mathcal{Z}, \;\; \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^{\mathrm{T}}(-C^{\mathrm{T}}\tilde{\lambda}^k) \geqslant (z - \tilde{z}^k)^{\mathrm{T}}\mu\beta C^{\mathrm{T}}C(\tilde{z}^k - z^k), \; \forall z \in \mathcal{Z}. \tag{5.12c}$$

同样, 由式(5.11b)得到

$$\begin{aligned} \tilde{\lambda}^k \in \mathbb{R}^m, \qquad & (\lambda^k - \tilde{\lambda}^k)^{\mathrm{T}}(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b) \\ & \geqslant (\lambda^k - \tilde{\lambda}^k)^{\mathrm{T}}\big\{-B(y^k - \tilde{y}^k) - C(z^k - \tilde{z}^k) + \tfrac{1}{\beta}(\lambda^k - \tilde{\lambda}^k)\big\}, \; \forall \lambda \in \mathbb{R}^m. \end{aligned} \tag{5.12d}$$

利用变分不等式(3.3a)和(3.3b), 我们可以将式(5.12a)$\sim$(5.12d)中的结果表述成下面的引理.

引理 5.2 设$(x^{k+1}, y^{k+1}, z^{k+1})$根据给定的 v^k 由式(4.4)生成. 那么, 由式(5.11a)和式(5.11b)定义的预测点 $\tilde{w}^k$ 满足

$$\tilde{w}^k \in \Omega, \;\; \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^{\mathrm{T}}F(\tilde{w}^k) \geqslant (v - \tilde{v}^k)^{\mathrm{T}}Q(v^k - \tilde{v}^k), \quad \forall\, w \in \Omega, \tag{5.13}$$

其中

$$Q = \begin{pmatrix} \mu\beta B^{\mathrm{T}}B & 0 & 0 \\ 0 & \mu\beta C^{\mathrm{T}}C & 0 \\ -B & -C & \frac{1}{\beta}I_m \end{pmatrix}. \tag{5.14}$$

我们接着来找出校正矩阵. 在迭代法(4.4)中不含校正, 为了用统一框架证明算法的收敛性, 我们根据式(5.11a)和式(5.11b)定义了 $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k)$. 将这个 $\tilde{w}^k$ 看做预测点,

再找出用 $v^k=(y^k,z^k,\lambda^k)$和 $\tilde v^k=(\tilde y^k,\tilde z^k,\tilde\lambda^k)$ 表示 $v^{k+1}=(y^{k+1},z^{k+1},\lambda^{k+1})$的关系式. 由于

$$y^{k+1}=\tilde y^k,\qquad z^{k+1}=\tilde z^k, \tag{5.15a}$$

$$\begin{aligned}\lambda^{k+1}&=\lambda^k-\beta\big(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b\big)\\&=\lambda^k-\big[-\beta B(y^k-\tilde y^k)-\beta C(z^k-\tilde z^k)+(\lambda^k-\tilde\lambda^k)\big],\end{aligned} \tag{5.15b}$$

这个校正公式可以写成

$$\begin{pmatrix}y^{k+1}\\z^{k+1}\\\lambda^{k+1}\end{pmatrix}=\begin{pmatrix}y^k\\z^k\\\lambda^k\end{pmatrix}-\begin{pmatrix}I&0&0\\0&I&0\\-\beta B&-\beta C&I_m\end{pmatrix}\begin{pmatrix}y^k-\tilde y^k\\z^k-\tilde z^k\\\lambda^k-\tilde\lambda^k\end{pmatrix}.$$

换句话说, 我们有

$$v^{k+1}=v^k-M(v^k-\tilde v^k),\quad \text{其中}\quad M=\begin{pmatrix}I&0&0\\0&I&0\\-\beta B&-\beta C&I_m\end{pmatrix}. \tag{5.16}$$

这样, 我们对第 4.2 节中强制平等的方法, 纳入了预测-校正方法框架(5.1a)和(5.1b).

最后我们来验证收敛性条件. 要问是否有正定矩阵 H 使得 $HM=Q$, 可以验证

$$H=QM^{-1}=\begin{pmatrix}\mu\beta B^{\mathrm T}B&0&0\\0&\mu\beta C^{\mathrm T}C&0\\0&0&\frac1\beta I\end{pmatrix}$$

是正定的. 进而我们验证矩阵 G 的正定性. 根据定义和 $Q=HM$, 进行一些基本的矩阵运算就能得到

$$G=\begin{pmatrix}(\mu-2)\beta B^{\mathrm T}B&0&0\\0&(\mu-2)\beta C^{\mathrm T}C&0\\0&0&\frac1\beta I\end{pmatrix}+\beta\begin{pmatrix}B^{\mathrm T}B&-B^{\mathrm T}C&0\\-C^{\mathrm T}B&C^{\mathrm T}C&0\\0&0&0\end{pmatrix}. \tag{5.17}$$

上式右端的最后一个矩阵半正定. 所以, 由于 $\mu>2$, 矩阵 G 是正定的. 我们把第 4.2 节中的强制平等的方法, 纳入了预测-校正方法框架(5.1a)和(5.1b)后, 证明了它满足收敛性条件(5.2a)和(5.2b).

6 收缩性质和收敛速率的证明

对形如式(3.3a)和(3.3b)的变分不等式, 这一节利用统一框架(5.1a)和(5.1b)和条件(5.2a) 和(5.2b)证明收敛性.

引理 6.1 求解变分不等式(3.3a)和(3.3b), 如果条件(5.2a)和(5.2b)满足, 由算法(5.1a)和(5.1b)生成的点满足

$$\begin{aligned}&\tilde w^k\in\Omega,\quad \theta(u)-\theta(\tilde u^k)+(w-\tilde w^k)^{\mathrm T}F(\tilde w^k)\\&\qquad\geqslant\frac12\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\frac12\|v^k-\tilde v^k\|_G^2,\quad\forall w\in\Omega.\end{aligned} \tag{6.1}$$

证明 由式(5.1b)生成新的迭代点. 由于 $Q=HM$, 这时式(5.1a)就可以改写成

$$\tilde{w}^k\in\Omega,\ \ \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^{\mathrm{T}}F(\tilde{w}^k)\geqslant(v-\tilde{v}^k)^{\mathrm{T}}H(v^k-v^{k+1}),\quad \forall w\in\Omega,\tag{6.2}$$

对式(6.2)右端的 $(v-\tilde{v}^k)^{\mathrm{T}}H(v^k-v^{k+1})$, 用恒等式

$$(a-b)^{\mathrm{T}}H(c-d)=\frac{1}{2}\big(\|a-d\|_H^2-\|a-c\|_H^2\big)+\frac{1}{2}\big(\|c-b\|_H^2-\|d-b\|_H^2\big)\tag{6.3}$$

进行处理, 就有

$$\begin{aligned}&\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^{\mathrm{T}}F(\tilde{w}^k)\\ \geqslant&\frac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\frac{1}{2}\big(\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2\big),\quad\forall w\in\Omega.\end{aligned}\tag{6.4}$$

对式(6.4)右端的最后一部分, 利用式(5.1b), $HM=Q$ 和式(5.2b)中对矩阵 G 的定义, 我们有

$$\begin{aligned}\frac{1}{2}\big(\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2\big)&=\frac{1}{2}\big(\|v^k-\tilde{v}^k\|_H^2-\|(v^k-\tilde{v}^k)-M(v^k-\tilde{v}^k)\|_H^2\big)\\&=\frac{1}{2}\big((v^k-\tilde{v}^k)^{\mathrm{T}}(HM+M^{\mathrm{T}}H-M^{\mathrm{T}}HM)(v^k-\tilde{v}^k)\big)\\&=\frac{1}{2}\|v^k-\tilde{v}^k\|_G^2.\end{aligned}$$

将上述结果代入式(6.4)右端, 就得到式(6.1), 引理 6.1 得证.

定理 6.1 对变分不等式问题(3.3a)和(3.3b), 由算法框架(5.1a)和(5.1b)生成的核心变量序列 $\{v^k\}$, 在条件(5.2a)和(5.2b)满足的时候, 有收缩性质

$$\|v^{k+1}-v^*\|_H^2\leqslant\|v^k-v^*\|_H^2-\|v^k-\tilde{v}^k\|_G^2,\quad\forall\, v^*\in\mathcal{V}^*.\tag{6.5}$$

证明 用 w^* 替代式(6.1)中任意的 $w\in\Omega$ 并利用单调性, 就得到定理的结论.

因为 G 正定, 序列 $\{v^k\}$ 是在 H-模下收缩的, 式(6.5)是方法收敛的关键不等式. 此外, 利用式(6.1), 马上得到 $O(1/t)$ 收敛速率的结论.

定理 6.2 对变分不等式问题(3.3a)和(3.3b), 算法(5.1a)和(5.1b)在条件(5.2a)和(5.2b)满足的时候有遍历意义下的 $O(1/t)$ 收敛速率, 即对任意正整数 t, 都有

$$\theta(\tilde{u}_t)-\theta(u)+(\tilde{w}_t-w)^{\mathrm{T}}F(w)\leqslant\frac{1}{2(t+1)}\|v-v^0\|_H^2,\quad\forall w\in\Omega,\tag{6.6}$$

其中 $\tilde{w}_t=\dfrac{1}{(t+1)}\Big(\sum\limits_{k=0}^{t}\tilde{w}^k\Big)$是预测序列的算术平均.

证明 首先, 利用 $(w-\tilde{w}^k)^{\mathrm{T}}F(\tilde{w}^k)=(w-\tilde{w}^k)^{\mathrm{T}}F(w)$, 由式(6.1)可以得到

$$\theta(\tilde{u}^k)-\theta(u)+(\tilde{w}^k-w)^{\mathrm{T}}F(w)+\frac{1}{2}\|v-v^{k+1}\|_H^2\leqslant\frac{1}{2}\|v-v^k\|_H^2,\ \ \forall w\in\Omega.\tag{6.7}$$

将上式从 $k=0,1,\cdots,t$ 累加, 再利用凸函数的性质

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

就得到式(6.6). 定理得证.

参 考 文 献

[1] Gabay D, Applications of the method of multipliers to variational inequalities, *Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems*, edited by M. Fortin and R. Glowinski, North Holland, Amsterdam, The Netherlands, 1983: 299-331.

[2] Glowinski R, *Numerical Methods for Nonlinear Variational Problems* [M]. New York: Springer-Verlag, 1984.

[3] 何炳生. 凸优化和单调变分不等式的收缩算法 [EB/OL]. [2015-04-10]. http://math.nju.edu.cn/~hebma.

[4] Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers [J]. *Foun Trends Mach Learn*, 2010, **3**: 1-122.

[5] He B S, Yuan X M. On the $O(1/n)$ convergence rate of the alternating direction method [J]. *SIAM J. Numerical Analysis*, 2012, **50**: 700-709.

[6] He B S, Yuan X M. On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers [J]. *Numerische Mathematik*.

[7] 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]. to appear in *Mathematical Programming, Series A*.

[8] He B S, Tao M, Yuan X M. Alternating direction method with Gaussian back substitution for separable convex programming [J]. *SIAM Journal on Optimization*, 2012, **22**: 313-340.

[9] He B S, Tao M, Yuan X M. A splitting method for separable convex programming [J]. *IMA Journal of Numerical Analysis*, 2015, **31**: 394-426.

[10] Esser E, Möller M, Osher S, et al. A convex model for non-negative matrix factorization and dimensionality reduction on physical space [J]. *IEEE Trans. Imag. Process.*, 2012, **21**(7): 3239-3252.

[11] 何炳生. 凸优化的一阶分裂算法—变分不等式为工具的统一框架 [EB/OL]. [2015-04-10]. http://math.nju.edu.cn/~hebma.

[12] 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**: 1011-1040.