文章编号: 1000-0887(2007) 12-1483-10

© 应用数学和力学编委会, ISSN 1000-0887

求解结构型单调变分不等式的改进的邻近类分解方法*

李　敏[1], 袁晓明[2]

(1. 东南大学 经济管理学院 管理科学与工程系, 南京 210096;
2. 上海交通大学, 安泰经济与管理学院 管理科学系, 上海 200052)

(郭兴明推荐)

摘要: 邻近类分解方法首先是由Chen 和Teboulle(Math. Programming, 1994, **64**(1): 81-101) 提出用来求解凸的极小化问题· 在此基础上, 该文提出一种新方法求解具有分离结构的单调变分不等式· 其主要优点在于放松了算法中对某些参数的限制, 使得新方法更加便于计算· 在和原分解方法相同的假设下, 可以证明新方法是全局收敛的·

关　键　词: 分解; 非精确准则; 邻近; 结构型变分不等式

中图分类号: O221; O224　　文献标识码: A

引　言

变分不等式问题的基本形式为: 寻找向量 $\boldsymbol{u}^* \in \Omega$ 使得

$$(\boldsymbol{u}-\boldsymbol{u}^*)^{\mathrm{T}} F(\boldsymbol{u}^*) \geqslant 0, \qquad \forall \boldsymbol{u} \in \Omega, \tag{1}$$

其中 $F: \Omega \to R^s$ 是在非空闭凸集合 $\Omega \subseteq R^s$ 上的连续映射· 本文涉及的变分不等式问题有如下分离结构:

$$\boldsymbol{u} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}, \quad F(\boldsymbol{u}) = \begin{pmatrix} f(\boldsymbol{x}) \\ g(\boldsymbol{y}) \end{pmatrix}, \tag{2}$$

$$\Omega = \left\{ (\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x} \in X, \boldsymbol{y} \in Y, \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} = \boldsymbol{b} \right\}, \tag{3}$$

其中 X 和 Y 分别是给定的 R^n 和 R^m 上的非空闭凸子集, $\boldsymbol{A} \in R^{l\times n}$ 和 $\boldsymbol{B} \in R^{l\times m}$ 是给定的矩阵, $\boldsymbol{b} \in R^l$ 是给定的向量· 在本文中, 我们假设 $f: X \to R^n$ 和 $g: Y \to R^m$ 是连续单调算子, 问题(1) ~ (3) 的解集 Ω^* 是非空的· 在实际生活中, 有很多实例可以转化为变分不等式问题(1) ~ (3) (记为 VI(1) ~ (3)) , 例如, 在经济、交通、工程等方面的应用, 具体可见参考文献[2-6]·

VI(1) ~ (3) 的等价形式为: 寻找向量 $\boldsymbol{w}^* \in W := X \times Y \times R^l$ 使得

$$(\boldsymbol{w}-\boldsymbol{w}^*)^{\mathrm{T}} Q(\boldsymbol{w}^*) \geqslant 0, \qquad \forall \boldsymbol{w} \in W, \tag{4}$$

其中

* 收稿日期: 2006-05-22; 修订日期: 2007-09-14
基金项目: 国家自然科学基金资助项目(70671024); 国家高新技术发展(863) 计划(2006AA11Z209)
作者简介: 李敏(1980—) , 女, 江苏徐州人, 博士;
袁晓明, 博士(联系人. Tel: + 86-21-52301397; E-mail: xmyuan@ sjtu. edu. cn) .

$$w=\begin{pmatrix}x\\y\\\lambda\end{pmatrix},\quad Q(w)=\begin{pmatrix}f(x)-A^{\mathrm{T}}\lambda\\g(y)-B^{\mathrm{T}}\lambda\\Ax+By-b\end{pmatrix}. \tag{5}$$

以下, 记VI(4)～(5)为SVI(W, Q), 并将其解集记为 W^*. 因此求解VI(1)～(3)等价于寻找向量 $w^*\in W^*$.

与VI(1)～(3)密切相关的问题是凸规划:

$$\min\{f(x)+g(y):Ax-y=0\}, \tag{6}$$

其中 $f:R^n\to(-\infty,+\infty]$ 和 $g:R^m\to(-\infty,+\infty]$ 是给定的正常闭凸函数, $A\in R^{m\times n}$, 可查阅参考文献[7-8]. 由Chen和Teboulle在文献[1]中提出的邻近分解算法是求解(6)式的有效方法之一. 在文献[1]中同时提出来的原始邻近分解算法的不精确形式更为实用, 因为涉及到的子问题只要求近似求解. 为了引入非精确邻近分解方法, 对给定的函数 h 和 $\varepsilon\geqslant 0$, 我们采用如下记号:

$$\varepsilon\text{-}\arg\min h(z)=\{v:h(v)\leqslant\inf h+\varepsilon\}.$$

对任意的初始点 $(x^0,y^0,\lambda^0)\in R^n\times R^m\times R^l$, 非精确邻近分解方法按如下步骤生成迭代序列 $\{(x^k,y^k,\lambda^k)\}\subseteq R^n\times R^m\times R^l$:

算法Ⅰ(Chen 和 Teboulle)

步1　求解

$$x^{k+1}=\alpha_k\text{-}\arg\min\{f(x)-\langle\lambda^k-\beta_k(Ax^k-y^k),Ax\rangle+(1/(2\beta_k))\|x-x^k\|^2\}.$$

步2　求解

$$y^{k+1}=\rho_k\text{-}\arg\min\{g(y)+\langle\lambda^k-\beta_k(Ax^k-y^k),y\rangle+(1/(2\beta_k))\|y-y^k\|^2\}.$$

步3　更新

$$\lambda^{k+1}=\lambda^k-\beta_k(Ax^{k+1}-y^{k+1}).$$

这里参数序列 $\{\alpha_k\}$ 和 $\{\rho_k\}$ 满足:

$$\forall k,\ \alpha_k,\rho_k\geqslant 0,\qquad \sum_{k=0}^{\infty}\sqrt{\alpha_k}<\infty,\ \sum_{k=0}^{\infty}\sqrt{\rho_k}<\infty, \tag{7}$$

对任给的 $0<\varepsilon\leqslant\min(1/3,1/(2\|A\|+1))$, $\{\beta_k\}$ 满足:

$$\varepsilon\leqslant\beta_k\leqslant\min\left(\frac{1-\varepsilon}{2},\frac{1-\varepsilon}{2\|A\|}\right),\qquad\forall k\geqslant 0. \tag{8}$$

算法Ⅰ不仅结合了交替方向法(见参考文献[4,5,9-13])和邻近点算法(见参考文献[8,14])的优点, 而且更加适合并行计算. 另外, 考虑到精确求解子问题往往代价较高, 而算法Ⅰ允许近似求解子问题, 这就启发我们可以通过放松不精确准则来设计更加有效的算法. 对算法Ⅰ的改进可以参见文献[15-16]: 在文献[15]中, 引入Bregman函数, Euclidean范数被更一般的范数所代替, 进而提出广义邻近分解算法; 在文献[16]中, 通过引入对数二次邻近项来标准化子问题, 提出求解结构型凸规划和单调变分不等式的内点算法.

显然, 算法Ⅰ的有效性在很大程度上依赖于参数 α_k、ρ_k 和 β_k 的选取. 因此从计算的角度来看, 有必要深入研究参数的选取. 在算法Ⅰ中对参数 α_k 和 ρ_k 的约束表明 $\alpha_k\to 0$, $\rho_k\to 0$, 这意味着在迭代过程中对子问题的求解精度要求越来越高. 而对参数 β_k 的选取需要预先知道 $\|A\|$ 的范围, 这种估计往往需要花费很大的代价. 基于以上讨论, 在这篇文章中, 我们提出一种新算法, 它在一般的假设条件下有效地放松了算法Ⅰ中对参数的限制. 需要指出的是, 为了保证算法的收敛性, 我们要增加一小部分计算量. 从实用的角度看, 相对于算法Ⅰ, 新的

算法有较大的改进•

本文的结构如下: 第 1 节详细介绍了新算法; 第 2 节主要证明了由新方法生成的迭代点的收缩性, 这些性质对收敛性分析很重要; 第 3 节讨论了收敛性; 最后在第 4 节给出一些结论•

1 新方法的基本框架

对给定的迭代点 $(\boldsymbol{x}^k, \boldsymbol{y}^k, \lambda^k) \in R^n \times R^m \times R^l$, 新方法的 $(k+1)$ 次迭代包含两步• 第一步用改进的邻近分解算法生成预测点 $(\tilde{\boldsymbol{x}}^k, \tilde{\boldsymbol{y}}^k, \tilde{\lambda}^k) \in X \times Y \times R^l$; 第二步用极小的代价校正预测点, 生成新的迭代点 $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}, \lambda^{k+1}) \in R^n \times R^m \times R^l$•

对两个对称正定矩阵 $\boldsymbol{M}$ 和 $\boldsymbol{N}$, 记号 $\boldsymbol{M} \succ \boldsymbol{N}$ ($\boldsymbol{M} \succeq \boldsymbol{N}$) 表示 $\boldsymbol{M} - \boldsymbol{N}$ 是正定的(半正定的)• $\|\boldsymbol{v}\|_{\boldsymbol{M}} = \sqrt{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{v}}$ 表示向量 $\boldsymbol{v}$ 的 $\boldsymbol{M}$ 范数• 我们假设 $\boldsymbol{H} \in R^{l\times l}$, $\boldsymbol{R} \in R^{n\times n}$ 和 $\boldsymbol{S} \in R^{m\times m}$ 是满足条件 $\boldsymbol{R} \succ 2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{A}$ 和 $\boldsymbol{S} \succ 2\boldsymbol{B}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{B}$ 的对称正定矩阵• 因此存在正常数 τ, 使得

$$\boldsymbol{R} \succeq (2+\tau)\boldsymbol{A}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{A} + \frac{\tau}{2}\boldsymbol{R}, \quad \boldsymbol{S} \succeq (2+\tau)\boldsymbol{B}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{B} + \frac{\tau}{2}\boldsymbol{S}• \tag{9}$$

为简单起见, 理论上不妨假设 $\boldsymbol{H}$、$\boldsymbol{R}$ 和 $\boldsymbol{S}$ 是常数矩阵• 已有的自适应调整参数矩阵的技术(参见文献[10-11, 17])很容易被推广到新方法上去•

1.1 预测步

令正常数 τ 满足(9) 式, 非负常数序列 $\{\nu_k\}$ 是有界的, 即:

$$\sup_k \nu_k = \nu \in (0, 1)• \tag{10}$$

预测步:

步 0 给定 $\varepsilon > 0$, $\boldsymbol{w}^0 = (\boldsymbol{x}^0, \boldsymbol{y}^0, \lambda^0) \in X \times Y \times R^l$• 令 $k = 0$•

步 1 寻找向量 $\tilde{\boldsymbol{x}}^k \in X$ 和 $\xi_x^k \in R^n$, 使得

$$(\boldsymbol{x} - \tilde{\boldsymbol{x}}^k)^{\mathrm{T}}\left\{f(\tilde{\boldsymbol{x}}^k) - \boldsymbol{A}^{\mathrm{T}}[\lambda^k - \boldsymbol{H}(\boldsymbol{A}\tilde{\boldsymbol{x}}^k + \boldsymbol{B}\boldsymbol{y}^k - \boldsymbol{b})] + \boldsymbol{R}(\tilde{\boldsymbol{x}}^k - \boldsymbol{x}^k + \xi_x^k)\right\} \geqslant 0, \quad \forall \boldsymbol{x} \in X, \tag{11}$$

其中

$$\|\xi_x^k\|_{\boldsymbol{R}}^2 \leqslant \nu_k \|\boldsymbol{x}^k - \tilde{\boldsymbol{x}}^k\|_{\boldsymbol{R}_a}^2, \quad \boldsymbol{R}_a = \boldsymbol{R} - (2+\tau)\boldsymbol{A}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{A}• \tag{12}$$

步 2 寻找向量 $\tilde{\boldsymbol{y}}^k \in Y$ 和 $\xi_y^k \in R^m$, 使得

$$(\boldsymbol{y} - \tilde{\boldsymbol{y}}^k)^{\mathrm{T}}\left\{g(\tilde{\boldsymbol{y}}^k) - \boldsymbol{B}^{\mathrm{T}}[\lambda^k - \boldsymbol{H}(\boldsymbol{A}\tilde{\boldsymbol{x}}^k + \boldsymbol{B}\tilde{\boldsymbol{y}}^k - \boldsymbol{b})] + \boldsymbol{S}(\tilde{\boldsymbol{y}}^k - \boldsymbol{y}^k + \xi_y^k)\right\} \geqslant 0, \quad \forall \boldsymbol{y} \in Y, \tag{13}$$

其中

$$\|\xi_y^k\|_{\boldsymbol{S}}^2 \leqslant \nu_k \|\boldsymbol{y}^k - \tilde{\boldsymbol{y}}^k\|_{\boldsymbol{S}_b}^2, \quad \boldsymbol{S}_b = \boldsymbol{S} - (2+\tau)\boldsymbol{B}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{B}• \tag{14}$$

步 3 更新 $\tilde{\lambda}^k$

$$\tilde{\lambda}^k = \lambda^k - \boldsymbol{H}(\boldsymbol{A}\tilde{\boldsymbol{x}}^k + \boldsymbol{B}\tilde{\boldsymbol{y}}^k - \boldsymbol{b})• \tag{15}$$

步 4 验证收敛性:

如果 $\|\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k\|_\infty < \varepsilon$, 则停止迭代, $\boldsymbol{w}^k$ 为近似解点; 否则转步 5• □

注 1.1 从(9)、(12) 和(14) 式知, $\boldsymbol{R}_a \succeq (\tau/2)\boldsymbol{R}$ 和 $\boldsymbol{S}_b \succeq (\tau/2)\boldsymbol{S}$• 因此非精确准则(12) 式和(14) 式中的矩阵范数都是有定义的•

注 1.2 非精确准则(12) 和(14) 式允许求解子问题(11) 和(13) 的相对误差被常数 $\nu \in (0, 1)$ 限制, 与(7) 式相比降低了子问题的难度• 这种非精确准则可以有效地减少工作量, 因此更为实用• 对于这种非精确准则的延伸, 可参见文献[14, 18]•

注 1.3 (8)式中对 β_k 进行了严格的限制，这使得在很多情况下，由于 $\|A\|$ 的近似界无法估计，导致算法Ⅰ不能实现• 新方法中，由于邻近参数矩阵 R 和 S 以及乘子矩阵 H 都是对称正定的常数矩阵，没有其他的限制条件，因此，很好地解决了这个问题•

注 1.4 算法Ⅰ有很多优点，包括便于并行计算，继承了交替方向法和邻近点算法的优点等• 而这些优点，在新算法的预测步中都得以保留•

1.2 校正步

我们采用如下记号:

$$G=\begin{pmatrix} R & 0 & 0\\ 0 & S & 0\\ 0 & 0 & H^{-1}\end{pmatrix}_{(n+m+l)\times(n+m+l)},\quad \xi=\begin{pmatrix}\xi_x\\ \xi_y\\ 0\end{pmatrix}• \tag{16}$$

根据之前的假设，可以推出 G 是对称正定的•

校正步:

步 5 通过校正 w^k 得到新的迭代点 w^{k+1}:

$$w^{k+1}:=w^k-\alpha_k d(w^k,\tilde{w}^k,\xi^k), \tag{17}$$

其中

$$\alpha_k=\gamma_k\alpha_k^*,\quad \gamma_k\in[\gamma_L,\gamma_U]\subseteq(0,2),\quad \alpha_k^*:=\frac{\varphi(w^k,\tilde{w}^k,\xi^k)}{\|d(w^k,\tilde{w}^k,\xi^k)\|_G^2}, \tag{18}$$

$$\begin{aligned}\varphi(w^k,\tilde{w}^k,\xi^k):=&(Ax^k+By^k-b)^{\mathrm T}H[A(x^k-\tilde{y}^k)+B(y^k-\tilde{y}^k)]+\\&(w^k-\tilde{w}^k)^{\mathrm T}Gd(w^k,\tilde{w}^k,\xi^k)\end{aligned} \tag{19}$$

和

$$d(w^k,\tilde{w}^k,\xi^k):=(w^k-\tilde{w}^k)-\xi^k• \tag{20}$$

步 6 令 $k:=k+1$，转步 1• □

注 1.5 在第 2 节中我们将证明，若 w^* 为任一解点，$-d(w^k,\tilde{w}^k,\xi^k)$ 则是 $\|w-w^*\|_G^2$ 在点 w^k 处的下降方向；而 α_k 是沿着这个方向的且带有松弛因子 γ_k 的最优步长• 校正步中的主要计算是矩阵和向量的乘积运算• 因此而增加的工作量是很小的•

注 1.6 注意到即使新的迭代点 w^{k+1} 不在集合 W 内，新算法仍然是可行的• 令 $P_{W,G}(\cdot)$ 表示 G-模下在集合 W 上的投影，即:

$$P_{W,G}[v]=\arg\min\left\{\|v-u\|_G,\ u\in W\right\},$$

若取

$$w^{k+1}=P_{W,G}[w^k-\alpha_k d(w^k,\tilde{w}^k,\xi^k)] \tag{21}$$

代替(17)式作为新的迭代点，理论上更合理• 但是可能会因为投影而增加工作量• 在投影可以简单计算的情况下(例如：X 和 Y 是非负卦限或者是框约束)，我们通常用(21)式来生成新的迭代点 w^{k+1}•

2 新方法的收缩性

为了证明收敛性，首先我们要研究由新算法生成的迭代序列的收缩性• 我们要着重从理论上说明选取 $-d(w^k,\tilde{w}^k,\xi^k)$ 作为下降方向和 α_k 作为最优步长的理由•

命题 2.1 令 $w^*=(x^*,y^*,\lambda^*)\in W^*$ 是 SVI(W,Q) 的任一个解点• 对给定的 w^k，令 $\tilde{w}^k$ 由(11)～(15)式生成，$\varphi(w^k,\tilde{w}^k,\xi^k)$ 和 $d(w^k,\tilde{w}^k,\xi^k)$ 分别由(19)式和(20)式定义• 我们有

$$(w^k-w^*)^{\mathrm T}Gd(w^k,\tilde{w}^k,\xi^k)\geqslant\varphi(w^k,\tilde{w}^k,\xi^k), \tag{22}$$

其中 G 由(16)式定义.

证明 令 $w^k=(x^k,y^k,\lambda^k)\in W$, $w^*=(x^*,y^*,\lambda^*)\in W^*$. 在(4)式中, 令 $w=(x^k,y^*,\lambda^*)$, 有

$$(x^k-x^*)^{\mathrm{T}}[f(x^*)-A^{\mathrm{T}}\lambda^*]\geqslant 0. \tag{23}$$

在(11)式中令 $x=x^*$, λ^k 用 $\lambda^k+H(Ax^k+By^k-b)$ (参考(15)式)代替, 可以得到

$$(x^k-x^*)^{\mathrm{T}}\{-f(x^k)+A^{\mathrm{T}}\lambda^k-A^{\mathrm{T}}H[A(x^k-x^k)+B(y^k-y^k)]+ R(x^k-x^k-\xi_x^k)\}\geqslant 0, \tag{24}$$

(23)式+(24)式, 并且利用 f 的单调性可得

$$(x^k-x^*)^{\mathrm{T}}\{A^{\mathrm{T}}(\lambda^k-\lambda^*)-A^{\mathrm{T}}H[A(x^k-x^k)+B(y^k-y^k)]+ R(x^k-x^k-\xi_x^k)\}\geqslant 0.$$

类似地, 我们有

$$(y^k-y^*)^{\mathrm{T}}\{B^{\mathrm{T}}(\lambda^k-\lambda^*)-B^{\mathrm{T}}H[A(x^k-x^k)+B(y^k-y^k)]+ S(y^k-y^k-\xi_y^k)\}\geqslant 0.$$

因为 $Ax^*+By^*=b$, $Ax^k+By^k-b=H^{-1}(\lambda^k-\lambda^k)$, (可参考(15)式), 把以上2个不等式加起来可以得到

$$(w^k-w^*)^{\mathrm{T}}G(w^k-w^k-\xi^k)\geqslant (Ax^k+By^k-b)^{\mathrm{T}}H[A(x^k-x^k)+B(y^k-y^k)].$$

注意到

$$w^k-w^*=(w^k-w^*)-(w^k-w^k).$$

利用 $d(w^k,w^k,\xi^k)$ 和 $\varphi(w^k,w^k,\xi^k)$ 的表达式, 命题得证. □

命题 2.2 假设 $\varphi(w^k,w^k,\xi^k)$ 和 $d(w^k,w^k,\xi^k)$ 分别由(19)式和(20)式定义, 可以证明存在常数 $c>0$, 使得

$$\varphi(w^k,w^k,\xi^k)\geqslant\frac{1}{2}\|d(w^k,w^k,\xi^k)\|_G^2+c\|w^k-w^k\|_G^2,\qquad \forall k\geqslant 0. \tag{25}$$

证明 首先注意到以下这个恒等式

$$2(w^k-w^k)^{\mathrm{T}}Gd(w^k,w^k,\xi^k)\overset{(20)}{=} \|d(w^k,w^k,\xi^k)\|_G^2+\|w^k-w^k\|_G^2-\|\xi^k\|_G^2, \tag{26}$$

利用 Cauchy-Schwarz 不等式和 $Ax^k+By^k-b=H^{-1}(\lambda^k-\lambda^k)$ (参见(15)式), 可以得到

$$2(Ax^k+By^k-b)^{\mathrm{T}}HA(x^k-x^k)\geqslant -\frac{1}{2+\tau}\|\lambda^k-\lambda^k\|_{H^{-1}}^2-(2+\tau)\|x^k-x^k\|_{A^{\mathrm{T}}HA}^2 \tag{27}$$

和

$$2(Ax^k+By^k-b)^{\mathrm{T}}HB(y^k-y^k)\geqslant -\frac{1}{2+\tau}\|\lambda^k-\lambda^k\|_{H^{-1}}^2-(2+\tau)\|y^k-y^k\|_{B^{\mathrm{T}}HB}^2. \tag{28}$$

(27)式+(28)式, 可以得到

$$2(Ax^k+By^k-b)^{\mathrm{T}}H[A(x^k-x^k)+B(y^k-y^k)]\geqslant -\frac{2}{2+\tau}\|\lambda^k-\lambda^k\|_{H^{-1}}^2-(2+\tau)\|x^k-x^k\|_{A^{\mathrm{T}}HA}^2-(2+\tau)(y^k-y^k)_{B^{\mathrm{T}}HB}^2.$$

把上式和(26)式代入(19)式得

$$2\varphi(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)\overset{(16)}{\geqslant}\|d(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)\|_{\boldsymbol{G}}^2+\|\boldsymbol{x}^k-\boldsymbol{x}^k\|_{\boldsymbol{R}_a}^2+\|\boldsymbol{y}^k-\boldsymbol{y}^k\|_{\boldsymbol{S}_b}^2+\frac{\tau}{2+\tau}\|\lambda^k-\lambda^k\|_{\boldsymbol{H}^{-1}}^2-\|\xi^k\|_{\boldsymbol{G}}^2,\tag{29}$$

其中 $\boldsymbol{R}_a$ 和 $\boldsymbol{S}_b$ 分别由(12)式和(14)式定义. 另外,

$$\|\xi^k\|_{\boldsymbol{G}}^2\overset{(16)}{=}\|\xi_x^k\|_{\boldsymbol{R}}^2+\|\xi_y^k\|_{\boldsymbol{S}}^2\leqslant\nu(\|\boldsymbol{x}^k-\boldsymbol{x}^k\|_{\boldsymbol{R}_a}^2+\|\boldsymbol{y}^k-\boldsymbol{y}^k\|_{\boldsymbol{S}_b}^2),$$

其中的不等式可由(12)式和(14)式得出.

因此, 由 $\boldsymbol{R}_a\succeq(\tau/2)\boldsymbol{R}$, $\boldsymbol{S}_b\succeq(\tau/2)\boldsymbol{S}$ 和(29)式可以得到

$$\varphi(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)\geqslant\frac{1}{2}\|d(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)\|_{\boldsymbol{G}}^2+\frac{\tau(1-\nu)}{4}(\|\boldsymbol{x}^k-\boldsymbol{x}^k\|_{\boldsymbol{R}}^2+\|\boldsymbol{y}^k-\boldsymbol{y}^k\|_{\boldsymbol{S}}^2)+\frac{\tau}{2(2+\tau)}\|\lambda^k-\lambda^k\|_{\boldsymbol{H}^{-1}}^2.\tag{30}$$

令 $c=\min\left\{\tau(1-\nu)/4,\ \tau/(2(2+\tau))\right\}$, 命题由(30)式得出. □

命题 2.1 和 2.2 表明 $-d(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)$ 是 $\|\boldsymbol{w}-\boldsymbol{w}^*\|_{\boldsymbol{G}}^2$ 在点 $\boldsymbol{w}^k$ 的下降方向. 因此, 我们用(17)式进行校正. 从计算的观点看, 研究如何沿着下降方向选取最优步长是有意义的. 若记带有步长 α 的新的迭代点为

$$\boldsymbol{w}^{k+1}(\alpha)=\boldsymbol{w}^k-\alpha d(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k),$$

对任给的解点 $\boldsymbol{w}^*\in W^*$ 和 $\boldsymbol{w}^k$, 用记号

$$\Theta_k(\alpha):=\|\boldsymbol{w}^k-\boldsymbol{w}^*\|_{\boldsymbol{G}}^2-\|\boldsymbol{w}^{k+1}(\alpha)-\boldsymbol{w}^*\|_{\boldsymbol{G}}^2\tag{31}$$

来描述每次迭代获得的进步. 以下这个定理可以用来解释为什么最优步长 α_k 由(18)式来定义.

定理 2.1 令 $\boldsymbol{w}^*$ 是解集 W^* 内的任意一点. 对给定的点 $\boldsymbol{w}^k$, 令 $\boldsymbol{w}^k=(\boldsymbol{x}^k,\boldsymbol{y}^k,\lambda^k)$ 由(11)式~(15)式生成, $\varphi(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)$ 和 $\Theta_k(\alpha)$ 分别由(19)式和(31)式定义. 我们有

$$\Theta_k(\alpha)\geqslant\psi_k(\alpha):=2\alpha\varphi(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)-\alpha^2\|d(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)\|_{\boldsymbol{G}}^2.\tag{32}$$

证明 既然 $\boldsymbol{w}^{k+1}(\alpha)=\boldsymbol{w}^k-\alpha d(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)$, 有

$$\|\boldsymbol{w}^{k+1}(\alpha)-\boldsymbol{w}^*\|_{\boldsymbol{G}}^2=\|\boldsymbol{w}^k-\boldsymbol{w}^*\|_{\boldsymbol{G}}^2-2\alpha(\boldsymbol{w}^k-\boldsymbol{w}^*)^{\mathrm{T}}\boldsymbol{G}d(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)+\alpha^2\|d(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)\|_{\boldsymbol{G}}^2.$$

因而

$$\Theta_k(\alpha)=\|\boldsymbol{w}^k-\boldsymbol{w}^*\|_{\boldsymbol{G}}^2-\|\boldsymbol{w}^{k+1}(\alpha)-\boldsymbol{w}^*\|_{\boldsymbol{G}}^2=2\alpha(\boldsymbol{w}^k-\boldsymbol{w}^*)^{\mathrm{T}}\boldsymbol{G}d(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)-\alpha^2\|d(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)\|_{\boldsymbol{G}}^2.\tag{33}$$

把命题 2.1 的结论(参见(22)式)应用到(33)式最右端的第 1 项, 我们有

$$\Theta_k(\alpha)\geqslant2\alpha\varphi(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)-\alpha^2\|d(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)\|_{\boldsymbol{G}}^2\overset{(32)}{=}\psi_k(\alpha),$$

定理得证. □

既然 $\psi_k(\alpha)$ 是由新的迭代点所取得进步的下界估计, 我们把它称作利益函数. 定理 2.1 启发我们极大化 $\psi_k(\alpha)$ 来加速收敛性. 注意到 $\psi_k(\alpha)$ 是关于 α 的二次函数, 因此可以取

$$\alpha_k^*=\frac{\varphi(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)}{\|d(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)\|_{\boldsymbol{G}}^2},\quad\text{并且有 }\psi_k(\alpha_k^*)=\alpha_k^*\varphi(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k).\tag{34}$$

由计算经验知, 为加速收敛, 有必要对理论上获得的最优步长 α_k^* 添加松弛因子. 接下来的这个定理告诉我们如何来选取松弛因子.

定理 2.2 令 $\boldsymbol{w}^*$ 是解集 W^* 内的任意一点, γ 是正常数, α_k^* 由(34)式定义. 对给定的点 $\boldsymbol{w}^k$, 令 $\boldsymbol{w}^k$ 由(11)式~(15)式生成, $\boldsymbol{w}^{k+1}(\lambda\alpha_k^*)=\boldsymbol{w}^k-\gamma\alpha_k^*d(\boldsymbol{w}^k,\boldsymbol{w}^k,\xi^k)$, 即沿着下降方向

$-d(w^k, w^k, \xi^k)$ 的步长为 $\gamma\alpha_k^*$，我们有

$$\|w^{k+1}(\gamma\alpha_k^*)-w^*\|_G^2 \leqslant \|w^k-w^*\|_G^2-\frac{\gamma(2-\gamma)}{2}\left(\frac{1}{2}\|d(w^k,w^k,\xi^k)\|_G^2+c\|w^k-w^k\|_G^2\right),\qquad \forall k\geqslant 0. \tag{35}$$

证明　首先注意到

$$\phi_k(\gamma\alpha_k^*) \overset{(32)}{=} 2\gamma\alpha_k^*\varphi(w^k,w^k,\xi^k)-(\gamma^2\alpha_k^*)\alpha_k^*(\|d(w^k,w^k,\xi^k)\|_G^2) \overset{(34)}{=} \gamma(2-\gamma)\alpha_k^*\varphi(w^k,w^k,\xi^k).$$

因此，由(32)式可以得到

$$\|w^{k+1}(\gamma\alpha_k^*)-w^*\|_G^2 \leqslant \|w^k-w^*\|_G^2-\gamma(2-\gamma)\alpha_k^*\varphi(w^k,w^k,\xi^k).$$

由命题2.2和 α_k^* 的定义(参见(34)式)，可知

$$\alpha_k^* = \frac{\varphi(w^k,w^k,\xi^k)}{\|d(w^k,w^k,\xi^k)\|_G^2} > \frac{1}{2}. \tag{36}$$

因此由(25)式可知(35)式成立. □

定理2.2表明理论上任一个 $\gamma\in(0,2)$ 可以保证新的迭代点更靠近解点，因此验证了(18)式中选择 $\gamma_k\in[\gamma_L,\gamma_U]\subseteq(0,2)$ 的理由.

3　新方法的收敛性

定理2.2表明由新算法生成的序列 $\{w^k\}$ 对解集 W^* 是 F jer 单调的. 本节主要讨论新算法的收敛性.

由定理2.2，我们可以得到

$$\|w^{k+1}-w^*\|_G \leqslant \|w^k-w^*\|_G.$$

上式说明序列 $\{w^k\}$ 是有界的以及

$$\lim_{k\to\infty}\|w^k-w^*\|_G=\mu<\infty. \tag{37}$$

因此，若在(35)式的两端都取极限，可以得到

$$\|d(w^k,w^k,\xi^k)\|_G\to 0,\quad \|w^k-w^k\|_G\to 0,$$

即

$$\|x^k-x^k-\xi_x^k\|_R\to 0,\quad \|y^k-y^k-\xi_y^k\|_S\to 0 \tag{38}$$

和

$$\|x^k-x^k\|_R\to 0,\quad \|y^k-y^k\|_S\to 0,\quad \|\lambda^k-\lambda^k\|_{H^{-1}}\to 0. \tag{39}$$

既然序列 $\{w^k\}$ 是有界的，至少存在一个极限点 w^∞，即存在一个子序列 $\{w^{k_j}\}$ 收敛到 w^∞. 实际上 w^∞ 是SVI(W,Q) 的一个解点. 为了证明这个结论，首先我们给出一个简单的引理.

引理3.1　对给定的点 w^k，令 w^k 由(11)式~(15)式生成，对任给的 $w\in W$，我们有

$$(x-x^k)^{\mathrm T}[f(x^k)-A^{\mathrm T}\lambda^k] \geqslant (x^k-x)^{\mathrm T}\left\{A^{\mathrm T}H[A(x^k-x^k)+B(y^k-y^k)]-R(x^k-x^k-\xi_x^k)\right\}$$

和

$$(y-y^k)^{\mathrm T}[g(y^k)-B^{\mathrm T}\lambda^k] \geqslant (y^k-y)^{\mathrm T}\left\{B^{\mathrm T}H[A(x^k-x^k)+B(y^k-y^k)]-S(y^k-y^k-\xi_y^k)\right\}.$$

证明　由(11)，(13)和(15)式得出的结论是成立的. □

下面我们证明 w^∞ 是SVI(W,Q) 的解点.

引理 3.2　假设 $\boldsymbol{w}^{\infty}$ 是由新方法生成的序列 $\{\boldsymbol{w}^k\}$ 的极限点, 那么 $\boldsymbol{w}^{\infty}$ 是 SVI(W, Q) 的解点•

证明　既然 $\boldsymbol{w}^{\infty}$ 是 $\{\boldsymbol{w}^k\}$ 的极限点, 那么存在子序列 $\{\boldsymbol{w}^{k_j}\}$ 收敛到 $\boldsymbol{w}^{\infty}$• 注意到

$$\lim_{k\to\infty} \| \boldsymbol{w}^k - \boldsymbol{w}^k \|_{\boldsymbol{G}} = 0,$$

那么有

$$\lim_{j\to\infty} \| \boldsymbol{w}^{k_j} - \boldsymbol{w}^{\infty} \|_{\boldsymbol{G}} = 0•$$

因为 $\{\boldsymbol{w}^{k_j}\} \subseteq W$ 且 W 是闭集, 我们有 $\boldsymbol{w}^{\infty} \in W$•

对子序列 $\{\boldsymbol{w}^{k_j}\}$ 应用引理 3.1 并结合(15) 式, 对任给的 $\boldsymbol{x} \in X$ 和 $\boldsymbol{y} \in Y$ 有

$$\begin{cases} (\boldsymbol{x} - \boldsymbol{x}^{k_j})^{\mathrm{T}}[f(\boldsymbol{x}^{k_j}) - \boldsymbol{A}^{\mathrm{T}}\lambda^{k_j}] \geqslant \\ \qquad (\boldsymbol{x}^{k_j} - \boldsymbol{x})^{\mathrm{T}}\left\{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{H}[\boldsymbol{A}(\boldsymbol{x}^{k_j} - \boldsymbol{x}^{k_j}) + \boldsymbol{B}(\boldsymbol{y}^{k_j} - \boldsymbol{y}^{k_j})] - \boldsymbol{R}(\boldsymbol{x}^{k_j} - \boldsymbol{x}^{k_j} - \xi_x^{k_j})\right\}, \\ (\boldsymbol{y} - \boldsymbol{y}^{k_j})^{\mathrm{T}}[g(\boldsymbol{y}^{k_j}) - \boldsymbol{B}^{\mathrm{T}}\lambda^{k_j}] \geqslant \\ \qquad (\boldsymbol{y}^{k_j} - \boldsymbol{y})^{\mathrm{T}}\left\{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{H}[\boldsymbol{A}(\boldsymbol{x}^{k_j} - \boldsymbol{x}^{k_j}) + \boldsymbol{B}(\boldsymbol{y}^{k_j} - \boldsymbol{y}^{k_j})] - \boldsymbol{S}(\boldsymbol{y}^{k_j} - \boldsymbol{y}^{k_j} - \xi_y^{k_j})\right\}, \\ \boldsymbol{A}\boldsymbol{x}^{k_j} + \boldsymbol{B}\boldsymbol{y}^{k_j} - \boldsymbol{b} = \boldsymbol{H}^{-1}(\lambda^{k_j} - \lambda^{k_j})• \end{cases}$$

固定 $\boldsymbol{x} \in X$ 和 $\boldsymbol{y} \in Y$, 对上式两端取极限并应用(38)、(39) 式, 有

$$\begin{cases} (\boldsymbol{x} - \boldsymbol{x}^{\infty})^{\mathrm{T}}[f(\boldsymbol{x}^{\infty}) - \boldsymbol{A}^{\mathrm{T}}\lambda^{\infty}] \geqslant \boldsymbol{0}, & \forall \boldsymbol{x} \in X, \\ (\boldsymbol{y} - \boldsymbol{y}^{\infty})^{\mathrm{T}}[f(\boldsymbol{y}^{\infty}) - \boldsymbol{B}^{\mathrm{T}}\lambda^{\infty}] \geqslant \boldsymbol{0}, & \forall \boldsymbol{y} \in Y, \\ \boldsymbol{A}\boldsymbol{x}^{\infty} + \boldsymbol{B}\boldsymbol{y}^{\infty} - \boldsymbol{b} = \boldsymbol{0}, \end{cases}$$

上式表明 $\boldsymbol{w}^{\infty}$ 是 SVI(W, Q) 的解点•　　□

定理 3.1　由新算法生成的迭代序列 $\{\boldsymbol{w}^k\}$ 收敛到 SVI(W, Q) 的解点•

证明　序列 $\{\boldsymbol{w}^k\}$ 至少存在一个极限点 $\boldsymbol{w}^{\infty}$, 即存在一个子序列 $\{\boldsymbol{w}^{k_j}\}$ 收敛到 $\boldsymbol{w}^{\infty}$• 由引理 3.2 知 $\boldsymbol{w}^{\infty}$ 是 SVI(W, Q) 的解点• 在(37) 式中用 $\boldsymbol{w}^{\infty}$ 代替 $\boldsymbol{w}^*$, 可得

$$\lim_{k\to\infty} \| \boldsymbol{w}^k - \boldsymbol{w}^{\infty} \|_{\boldsymbol{G}} = \lim_{j\to\infty} \| \boldsymbol{w}^{k_j} - \boldsymbol{w}^{\infty} \|_{\boldsymbol{G}} = 0•$$

定理得证•　　□

注 3.1　由(36)式知 $\alpha_k^* > 1/2$• 取 $\gamma_k := (1/\alpha_k^*) \in (0, 2)$, 即 $\alpha_k = 1$, 这对应新算法的特殊情形:

$$\boldsymbol{w}^{k+1} = \boldsymbol{w}^k + \xi^k• \tag{40}$$

由定理 2.1 知

$$\begin{aligned} \Theta_k(1) := \ & \| \boldsymbol{w}^k - \boldsymbol{w}^* \|_{\boldsymbol{G}}^2 - \| \boldsymbol{w}^k + \xi^k - \boldsymbol{w}^* \|_{\boldsymbol{G}}^2 \geqslant \\ & 2\varphi(\boldsymbol{w}^k, \boldsymbol{w}^k, \xi^k) - \| d(\boldsymbol{w}^k, \boldsymbol{w}^k, \xi^k) \|_{\boldsymbol{G}}^2• \end{aligned} \tag{41}$$

根据命题 2.2, 由(41) 式可知

$$\| \boldsymbol{w}^k - \boldsymbol{w}^* \|_{\boldsymbol{G}}^2 - \| \boldsymbol{w}^k + \xi^k - \boldsymbol{w}^* \|_{\boldsymbol{G}}^2 \geqslant 2c \| \boldsymbol{w}^k - \boldsymbol{w}^k \|_{\boldsymbol{G}}^2,$$

这表明方法(40) 是收敛的•

4　结　　论

本文提出一种求解结构型单调变分不等式的改进的邻近类分解方法• 鉴于对算法中的参数约束已经大大放松, 因此从计算的角度看, 子问题更容易求解• 我们以后的研究目标包括如何设计带有实用非精确准则的有效数值算法以及如何充分利用已有的邻近分解算法的优点来设计混合算法• 例如本文中的校正步有可能和其他算法(如: 外梯度算法) 结合来得到进一步的改进•

[参 考 文 献]

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

[2] Bertsekas D P, Gafni E M. Projection method for variational inequalities with applications to the traffic assignment problem[J]. Mathematical Programming Study, 1982, **17**: 139-159.

[3] Dafermos S. Traffic equilibrium and variational inequalities[J]. Transportation Science, 1980, **14**(1): 42-54.

[4] Eckstein J, Fukushima M. Some reformulation and applications of the alternating directions method of multipliers[A]. In: Hager W W, Hearn D W, Pardalos P M, Eds. Large Scale Optimization: State of the Art[C]. Amsterdam, Holland: Kluwer Academic Publishers, 1994, 115-134.

[5] Fukushima M. Application of the alternating directions method of multipliers to separable convex programming problems[J]. Computational Optimization and Applications, 1992, **1**(1): 93-111.

[6] Nagurney A, Ramanujam P. Transportation network policy modeling with goal targets and generalized penalty functions[J]. Transportation Science, 1996, **30**(1): 3-13.

[7] Glowinski R, Tallec P L. Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics[M]. SIAM Studies in Applied Mathematics. Philadelphia, PA: SIAM, 1989.

[8] Rockafellar R T. Convex Analysis[M]. Princeton, NJ: Princeton University Press, 1970.

[9] Gabay D, Mercier B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation[J]. Computers and Mathematics With Applications, 1976, **2**(1): 17-40.

[10] He B S, Yang H, Wang S L. Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities[J]. Journal of Optimization Theory and Applications, 2000, **106**(2): 337-356.

[11] He B S, Liao L-Z, Han D R, et al. A new inexact alternating directions method for monotone variational inequalities[J]. Mathematical Programming, 2002, **92**(1): 103-118.

[12] Tseng P. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities[J]. SIAM Journal on Control and Optimization, 1991, **29**(1): 119-138.

[13] Tseng P. Alternating projection-proximal methods for convex programming and variational inequalities[J]. SIAM Journal on Optimization, 1997, **7**(4): 951-965.

[14] He B S, Liao L-Z, Yang Z H. A new approximate proximal point algorithm for maximal monotone operator[J]. Science in China, Ser A, 2003, **46**(2): 200-206.

[15] Kyono M, Fukushima M. Nonlinear proximal decomposition method for convex programming[J]. Journal of Optimization Theory and Applications, 2000, **106**(2): 357-372.

[16] Auslender A, Teboulle M. Entropic proximal decomposition methods for convex programs and variational inequalities[J]. Mathematical Programming, 2001, **91**(1): 33-47.

[17] Kontogiorgis S, Meyer R R. A variable-penalty alternating directions method for convex optimization [J]. Mathematical Programming, 1998, **83**(1): 29-53.

[18] Solodov M V, Svaiter B F. A unified framework for some inexact proximal point algorithms[J]. Numerical Functional Analysis and Optimization, 2001, **22**(7/8): 1013-1035.

Improved Proximal-Based Decomposition Method for Structured Monotone Variational Inequalities

LI Min[1], YUAN Xiao-ming[2]

(1. Department of Management Science and Engineering, School of Economics and Management, Southeast University, Nanjing 210096, P.R. China;

2. Department of Management Science, Antai College of Economics and Management, Shanghai Jiaotong University, Shanghai 200052, P.R. China)

Abstract: The proximal-based decomposition method was originally proposed by Chen and Teboulle (Math. Programming, 1994, **64**(1): 81-101) for solving convex minimization problems. This paper extended to solve monotone variational inequalities associated with separable structures with the improvements that the restrictive assumptions on the involved parameters are much relaxed, and thus makes it practical to solve the involved subproblems easily. Without additional assumptions, global convergence of the new method is proved under the same mild assumptions on the problem's data as the original method.

Key words: decomposition; inexact criterion; proximal; structured variational inequality