中国科学: 数学 2012 年 第 42 卷 第 5 期: 515～525
www.springerlink.com math.scichina.com

《中国科学》杂志社
SCIENCE CHINA PRESS

求解凸规划及鞍点问题定制的 PPA 算法及其收敛速率

何炳生*, 申远

南京大学数学系, 南京 210093
E-mail: hebma@nju.edu.cn, yshen7@gmail.com

收稿日期: 2011-12-29; 接受日期: 2012-04-13; * 通信作者
教育部博士点基金 (批准号: 20110091110004) 资助项目

摘要 线性约束的凸优化问题和鞍点问题的一阶最优性条件是一个单调变分不等式. 在变分不等式框架下求解这些问题, 选取适当的矩阵 G, 采用 G- 模下的 PPA 算法, 会使迭代过程中的子问题求解变得相当容易. 本文证明这类定制的 PPA 算法的误差界有 $1/k$ 的收敛速率.

关键词 凸优化 单调算子 G- 模下的 PPA 算法 收敛速率

MSC (2010) 主题分类 90C25, 90C30, 90C33

1 引言

设 $\Omega \subset \mathbb{R}^l$ 是闭凸集, 考虑如下的混合单调变分不等式: 求 u^*, 使得

$$u^* \in \Omega, \quad \theta(u) - \theta(u^*) + (u-u^*)^{\mathrm{T}} F(u^*) \geqslant 0, \quad \forall\, u \in \Omega, \tag{1.1}$$

其中 $\theta : \mathbb{R}^l \to \mathbb{R}$ 是 (不必光滑) 凸函数, $F : \mathbb{R}^l \to \mathbb{R}^l$ 是单调算子. 记问题 (1.1) 为 $\mathrm{VI}(\Omega, F, \theta)$ 并假设它的解集 (用 Ω^* 表示) 是非空的. 科学计算中一些典型问题, 可以归结为求解上述变分不等式问题.

在线性等式 (或不等式) 约束的凸优化问题

$$\min\, \{ f(x) \,|\, Ax = b \ (\text{或}\ Ax \geqslant b),\ x \in \mathcal{X} \} \tag{1.2}$$

中, 设 $f(x)$ 是凸函数, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $\mathcal{X}$ 是 $\mathbb{R}^n$ 中的闭凸集. 引入线性约束的 Lagrange 乘子 λ, 求解 (1.2) 等价于求 $u^* = (x^*, \lambda^*)$, 使得

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

其中

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

英文引用格式: He B S, Shen Y. On the convergence rate of customized proximal point algorithm for convex optimization and saddle-point problem (in Chinese). Sci Sin Math, 2012, 42(5): 515–525, doi: 10.1360/012011-1049

对不同的线性等式 (或不等式) 约束条件,

$$\Lambda=\begin{cases}\mathbb{R}^m, & \text{对等式约束 } Ax=b,\\ \mathbb{R}^m_+, & \text{对不等式约束 } Ax\geqslant b.\end{cases} \tag{1.3c}$$

由 $F(u)$ 的单调性, 变分不等式 (1.3) 是 (1.1) 的一种具体形式, 其中 $\theta(u)=f(x)$.

用全变差极小处理图像去模糊[1], 经离散化以后, 问题的数学模型是一个 min-max 问题. 设

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\{\theta_1(x)-y^{\mathrm{T}}Ax-\theta_2(y)\} \tag{1.4}$$

中 $\mathcal{X}\subset\mathbb{R}^n$, $\mathcal{Y}\subset\mathbb{R}^m$ 是闭凸集, $\theta_1:\mathbb{R}^n\to\mathbb{R}$, $\theta_2:\mathbb{R}^m\to\mathbb{R}$ 是凸函数, 矩阵 $A\in\mathbb{R}^{m\times n}$. 求解 (1.4) 等价于求 $u^*=(x^*,y^*)$ 使得

$$u^*\in\Omega,\quad (\theta_1(x)-\theta_1(x^*))+(\theta_2(y)-\theta_2(y^*))+(u-u^*)^{\mathrm{T}}F(u^*)\geqslant 0,\quad \forall\, u\in\Omega, \tag{1.5a}$$

其中

$$u=\begin{pmatrix}x\\ y\end{pmatrix},\quad F(u)=\begin{pmatrix}-A^{\mathrm{T}}y\\ Ax\end{pmatrix},\quad \Omega=\mathcal{X}\times\mathcal{Y}. \tag{1.5b}$$

这里的 $F(u)$ 依然是一个单调算子, 变分不等式 (1.5) 是 (1.1) 的一种具体形式, 其中 $\theta(u)=\theta_1(x)+\theta_2(y)$.

假设 通篇假设对给定的凸函数 $\theta_1(x)$ 和 $\theta_2(y)$ 以及任意的 $r>0$ 和 $a\in\mathbb{R}^n, b\in\mathbb{R}^m$, 子问题

$$\min\left\{\theta_1(x)+\frac{r}{2}\|x-a\|^2\,|\,x\in\mathcal{X}\right\}\quad \text{和}\quad \min\left\{\theta_2(y)+\frac{r}{2}\|y-b\|^2\,|\,y\in\mathcal{Y}\right\}$$

的求解是简单的.

当以上假设满足时, 对于问题 (1.1) 和 (1.4), 文献 [2] 提出了一种 G- 模下的外延 PPA 算法. 适当选取矩阵 G 时, 可使子问题的原始变量和对偶变量的求解分离实行, 从而大大简化子问题的计算. 文献 [1] 中的算法, 可以看作一种具体的 G- 模下的 PPA 算法, 该算法在图像去噪问题中取得很好的计算效果. 近年来, 优化算法的计算复杂性和收敛速率越来越为计算数学工作者所重视[3–5]. 本文研究求解 (1.1) 的 G- 模下外延 PPA 算法的计算复杂性, 证明了该算法有 $O(1/k)$ 的收敛速率.

2 G- 模下的外延 PPA 算法

邻近点算法 (proximal point algorithm, 简记为 PPA) 是优化计算的一类基本方法. 它最早由 Martinet[6] 在 1970 年提出, Rockafellar 对这类算法做了深入的研究[7,8]. 增广 Lagrange 乘子法 (augmented Lagrangian method) 是关于 Lagrange 乘子的 PPA 算法[9]. 近年来, PPA 算法在理论和算法实现方面都得到了学者们的进一步关注[10–13]. 若用经典的欧氏模下的邻近点算法求解单调变分不等式 (1.1), 对给定的 $u^k\in\mathbb{R}^l$ 和 $r>0$, PPA 算法的第 k 次迭代中的子问题是

$$\text{(PPA)}\qquad u\in\Omega,\quad \theta(u')-\theta(u)+(u'-u)^{\mathrm{T}}\{F(u)+r(u-u^k)\}\geqslant 0,\quad \forall\, u'\in\Omega. \tag{2.1}$$

设 $\tilde{u}^k$ 是 (2.1) 的解并将其作为下一个迭代点 u^{k+1}, 迭代序列 $\{u^k\}$ 满足关系式

$$\|u^{k+1}-u^*\|^2\leqslant\|u^k-u^*\|^2-\|u^k-\tilde{u}^k\|^2,\quad \forall\, u^*\in\Omega^*.$$

我们说, 对解集 Ω^*, 序列 $\{u^k\}$ 是在欧氏模下 Fejér 单调的.

因为 (2.1) 的解通常没有显式表达式, 经典的欧氏模下的邻近点算法往往不容易实现. 如果将 (2.1) 中的常数 $r>0$ 换成一个对称正定矩阵 G, 得到的就是 G- 模下的 PPA 算法. 这时, $\tilde{u}^k$ 满足

$$\text{(G-PPA)} \qquad \tilde{u}^k \in \Omega, \quad \theta(u)-\theta(\tilde{u}^k)+(u-\tilde{u}^k)^{\mathrm{T}}\{F(\tilde{u}^k)+G(\tilde{u}^k-u^k)\} \geqslant 0, \quad \forall\, u \in \Omega. \tag{2.2}$$

以 u^* 替代上式中的 u, 得到

$$(\tilde{u}^k-u^*)^{\mathrm{T}}G(u^k-\tilde{u}^k) \geqslant \theta(\tilde{u}^k)-\theta(u^*)+(\tilde{u}^k-u^*)^{\mathrm{T}}F(\tilde{u}^k). \tag{2.3}$$

由于 F 是单调算子, 因此有

$$\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{u}^k-u^*)^{\mathrm{T}}F(\tilde{u}^k) \geqslant \theta(\tilde{u}^k)-\theta(u^*)+(\tilde{u}^k-u^*)^{\mathrm{T}}F(u^*).$$

因为 $\tilde{u}^k \in \Omega$ 并且 u^* 是 $\mathrm{VI}(\Omega,F,\theta)$ 的解, 上式的右端非负, 并由此 (2.3) 的右端非负. 因此,

$$(u^k-u^*)^{\mathrm{T}}G(u^k-\tilde{u}^k) \geqslant \|u^k-\tilde{u}^k\|_G^2, \quad \forall\, u^* \in \Omega^*. \tag{2.4}$$

新的迭代点由

$$u^{k+1}=u^k-\gamma(u^k-\tilde{u}^k), \quad \gamma \in (1,2) \tag{2.5}$$

生成. 利用 (2.4) 就有

$$\begin{aligned}\|u^{k+1}-u^*\|_G^2 &= \|(u^k-u^*)-\gamma(u^k-\tilde{u}^k)\|_G^2\\ &= \|u^k-u^*\|_G^2-2\gamma(u^k-u^*)^{\mathrm{T}}G(u^k-\tilde{u}^k)+\gamma^2\|u^k-\tilde{u}^k\|_G^2\\ &\leqslant \|u^k-u^*\|_G^2-\gamma(2-\gamma)\|u^k-\tilde{u}^k\|_G^2.\end{aligned} \tag{2.6}$$

我们把对给定的 u^k, 由 (2.2) 生成的 $\tilde{u}^k$ 称为 k- 步迭代中的预测点, 再由 (2.5) 产生新的迭代点 u^{k+1} 的算法称为 G- 模下外延的 PPA 算法. 以上的分析证明了下面的定理.

定理 2.1　G- 模下外延的 PPA 算法产生的序列 $\{u^k=(x^k,\lambda^k)\}$ 满足

$$\|u^{k+1}-u^*\|_G^2 \leqslant \|u^k-u^*\|_G^2-\gamma(2-\gamma)\|u^k-\tilde{u}^k\|_G^2, \quad \forall\, u^* \in \Omega^*. \tag{2.7}$$

定理 2.1 说明, 对解集 Ω^*, 序列 $\{u^k\}$ 是在 G- 模下 Fejér 单调的. G- 模下外延的 PPA 算法是迭代点到解集越来越近的收缩算法. 在 (2.2) 中, 如果 $\tilde{u}^k=u^k$, 那么 u^k 就是原问题的解. 因此,

$$\text{我们把 } \|u^k-\tilde{u}^k\|_G^2 \text{ 看作在当前点 } u^k \text{ 处的一种误差度量.} \tag{2.8}$$

在迭代式 (2.5) 中, γ 是松弛因子. 为了说明取 $\gamma \in (1,2)$ 的作用, 我们考察与 γ 相关的 G- 模下距离平方的缩短量

$$\vartheta(\gamma)=\|u^k-u^*\|_G^2-\|u^{k+1}-u^*\|_G^2. \tag{2.9}$$

利用 (2.6) 有

$$\vartheta(\gamma)=\|u^k-u^*\|_G^2-\|u^{k+1}-u^*\|_G^2 \geqslant (2\gamma-\gamma^2)\|u^k-\tilde{u}^k\|_G^2.$$

我们得到 $\vartheta(\gamma)$ 的一个下界

$$q(\gamma)=(2\gamma-\gamma^2)\|u^k-\tilde{u}^k\|_G^2.$$

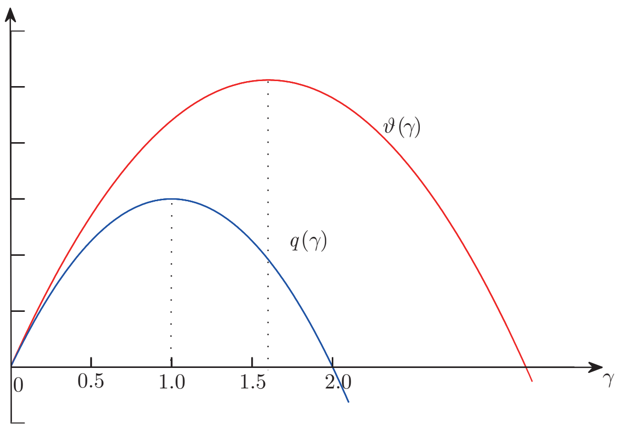

图 1　取 $\gamma \in [1,2)$ 的示意图

$q(\gamma)$ 是 γ 的二次函数, 它在 $\gamma^*=1$ 取得极大值. 注意到我们的本意是要在 G- 模下极大化 $\vartheta(\gamma)$ (见 (2.9)), 由于它含有未知的 u^*, 不得已才保守地去极大化它的下界函数 $q(\gamma)$. 取 $\gamma \in (1,2)$ 的理由可见图 1 中的示意.

3　定制的 PPA 方法

采用定制的 PPA 算法 (customized PPA), 可以将复杂的子问题分解化简求解. 所谓定制的 PPA 算法, 就是根据问题的需要, 选择适当的正定矩阵 G, 使产生预测点的过程 (2.2) 可以分解实行, 并且子问题都是在本文的假设条件下容易求解的问题. 显然, 通过 (2.2) 求预测点是 G- 模下外延的 PPA 算法的关键步骤. 文献 [2] 中的例子说明这是可以办到的. 我们只需说明预测点 $\tilde{u}^k$ 如何生成.

3.1　约束凸优化问题

对约束凸优化问题 (1.2), 若采用欧氏模下的 PPA 方法, 对给定的 (x^k,λ^k), 变分不等式 (1.3) 子问题 (2.1) 的解 $(\tilde{x}^k,\tilde{\lambda}^k)\in\Omega$ 要求满足

$$f(x)-f(\tilde{x}^k)+\begin{pmatrix} x-\tilde{x}^k \\ \lambda-\tilde{\lambda}^k \end{pmatrix}^{\mathrm{T}}\left\{\begin{pmatrix} -A^{\mathrm{T}}\tilde{\lambda}^k \\ A\tilde{x}^k-b \end{pmatrix}+r\begin{pmatrix} \tilde{x}^k-x^k \\ \tilde{\lambda}^k-\lambda^k \end{pmatrix}\right\}\geqslant 0, \quad \forall\,(x,\lambda)\in\Omega.$$

直接求得满足上式的 $(\tilde{x}^k,\tilde{\lambda}^k)\in\Omega$, 一般是办不到的. 假如将上式中的 r 换成矩阵

$$G=\begin{pmatrix} rI_n & -A^{\mathrm{T}} \\ -A & sI_m \end{pmatrix}, \tag{3.1}$$

就相当于要求 $(\tilde{x}^k,\tilde{\lambda}^k)\in\Omega$ 使得对所有的 $(x,\lambda)\in\Omega$ 都有

$$f(x)-f(\tilde{x}^k)+\begin{pmatrix} x-\tilde{x}^k \\ \lambda-\tilde{\lambda}^k \end{pmatrix}^{\mathrm{T}}\left\{\begin{pmatrix} -A^{\mathrm{T}}\tilde{\lambda}^k \\ A\tilde{x}^k-b \end{pmatrix}+\begin{pmatrix} r(\tilde{x}^k-x^k)-A^{\mathrm{T}}(\tilde{\lambda}^k-\lambda^k) \\ -A(\tilde{x}^k-x^k)+s(\tilde{\lambda}^k-\lambda^k) \end{pmatrix}\right\}\geqslant 0. \tag{3.2}$$

这是可以分解实现的. 对 (3.2) 中的 (x,λ), x 取固定的 $\tilde{x}^k$, λ 取任意的 $\lambda\in\Lambda$, 则 (3.2) 式的对偶部分就成了

$$\tilde{\lambda}^k\in\Lambda, \quad (\lambda-\tilde{\lambda}^k)^{\mathrm{T}}((Ax^k-b)+s(\tilde{\lambda}^k-\lambda^k))\geqslant 0, \quad \forall\,\lambda\in\Lambda.$$

注意到, 这样的 $\tilde{\lambda}^k$ 可以通过显式表达式

$$\tilde{\lambda}^k = P_\Lambda\left[\lambda^k - \frac{1}{s}(Ax^k - b)\right]$$

直接给出. 再对 (3.2) 中的 (x,λ), λ 取固定的 $\tilde{\lambda}^k$, x 取任意的 $x \in \mathcal{X}$, 则 (3.2) 式的原始部分就成了

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

由于 $\tilde{\lambda}^k$ 已知, $\tilde{x}^k$ 可以通过求解极小化问题

$$\min\left\{f(x) + \frac{r}{2}\left\|x - \left[x^k + \frac{1}{r}\,A^{\mathrm{T}}(2\tilde{\lambda}^k - \lambda^k)\right]\right\|^2 \,\middle|\, x \in \mathcal{X}\right\} \tag{3.3}$$

得到. 根据假设条件, 这是一个容易求解的子问题. 换言之, 对给定的 $u^k = (x^k, \lambda^k)$, 求得了满足 (3.2) 的 $(\tilde{x}^k, \tilde{\lambda}^k)$, 它是 G- 模下 PPA 子问题

$$\tilde{u}^k \in \Omega, \quad f(x) - f(\tilde{x}^k) + (u - \tilde{u}^k)^{\mathrm{T}}\{F(\tilde{u}^k) + G(\tilde{u}^k - u^k)\} \geqslant 0, \quad \forall\, u \in \Omega \tag{3.4}$$

的解, 其中对称矩阵 G 当

$$rs > \|A^{\mathrm{T}}A\|$$

时是正定的. 在诸如矩阵完整化等问题中[14], $\|A^{\mathrm{T}}A\|$ 的估算是容易的. 要实现 (3.2), 具体做法是:

对给定的 $u^k = (x^k, \lambda^k)$, 按等式约束问题和不等式约束问题分别用

$$\tilde{\lambda}^k = \lambda^k - \frac{1}{s}(Ax^k - b), \tag{3.5a}$$

和

$$\tilde{\lambda}^k = \left[\lambda^k - \frac{1}{s}(Ax^k - b)\right]_+ \tag{3.5b}$$

给出预测点的对偶部分 $\tilde{\lambda}^k$. 然后, 通过求解

$$\min\left\{f(x) + \frac{r}{2}\left\|x - \left[x^k + \frac{1}{r}\,A^{\mathrm{T}}(2\tilde{\lambda}^k - \lambda^k)\right]\right\|^2 \,\middle|\, x \in \mathcal{X}\right\}, \tag{3.6}$$

得到预测点的原始部分 $\tilde{x}^k$.

我们也可以用另外的顺序产生预测点 $\tilde{u}^k = (\tilde{x}^k, \tilde{\lambda}^k)$. 对给定的 (x^k, λ^k), 求 $(\tilde{x}^k, \tilde{\lambda}^k) \in \Omega$, 使得对所有的 $(x,\lambda) \in \Omega$ 都有

$$f(x) - f(\tilde{x}^k) + \begin{pmatrix} x - \tilde{x}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^{\mathrm{T}} \left\{\begin{pmatrix} -A^{\mathrm{T}}\tilde{\lambda}^k \\ A\tilde{x}^k - b \end{pmatrix} + \begin{pmatrix} r(\tilde{x}^k - x^k) + A^{\mathrm{T}}(\tilde{\lambda}^k - \lambda^k) \\ A(\tilde{x}^k - x^k) + s(\tilde{\lambda}^k - \lambda^k) \end{pmatrix}\right\} \geqslant 0. \tag{3.7}$$

它同样有 (3.4) 的形式, 只是其中对称矩阵 G 变成

$$G = \begin{pmatrix} rI_n & A^{\mathrm{T}} \\ A & sI_m \end{pmatrix}.$$

同样, 当 $rs > \|A^{\mathrm{T}}A\|$ 时, G 是正定的. 要实现 (3.7), 具体做法是:

对给定的 $u^k = (x^k, \lambda^k)$, 首先通过求解

$$\min\left\{f(x) + \frac{r}{2}\left\|x - \left[x^k + \frac{1}{r}A^{\mathrm{T}}\lambda^k\right]\right\|^2 \,\middle|\, x \in \mathcal{X}\right\},$$

得到预测点的原始部分 $\tilde{x}^k$. 然后, 根据问题是等式约束或不等式约束分别用

$$\tilde{\lambda}^k = \lambda^k - \frac{1}{s}(A(2\tilde{x}^k - x^k) - b),$$

和

$$\tilde{\lambda}^k = \left[\lambda^k - \frac{1}{s}(A(2\tilde{x}^k - x^k) - b)\right]_+$$

给出预测点的对偶部分 $\tilde{\lambda}^k$.

3.2 Min-max 问题

与 min-max 问题 (1.4) 等价的变分不等式 (1.5), 是 (1.1) 的一种特殊情况, 因而可以用 G- 模下的 PPA 方法来求解. 事实上, 文献 [1,15] 提到的方法都可以看作用 G- 模下的 PPA 方法求解 (1.4). 对给定的 $u^k = (x^k, y^k)$, 求 $\tilde{u}^k = (\tilde{x}^k, \tilde{y}^k) \in \Omega$, 使得

$$\theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \end{pmatrix}^{\mathrm{T}} \left\{ \begin{pmatrix} -A^{\mathrm{T}}\tilde{y}^k \\ A\tilde{x}^k \end{pmatrix} + \begin{pmatrix} r(\tilde{x}^k - x^k) - A^{\mathrm{T}}(\tilde{y}^k - y^k) \\ -A(\tilde{x}^k - x^k) + s(\tilde{y}^k - y^k) \end{pmatrix} \right\} \geqslant 0, \quad \forall\, u \in \Omega. \tag{3.8}$$

由于 $\theta(u) = \theta_1(x) + \theta_2(y)$, 实现 (3.8) 的具体做法是:

对给定的 (x^k, y^k), 通过求解 y 的子问题

$$\tilde{y}^k = \mathrm{Argmin}\left\{\theta_2(y) + y^{\mathrm{T}}Ax^k + \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\right\} \tag{3.9a}$$

得到 $\tilde{y}^k$, 然后通过求解一个关于 x 的子问题

$$\tilde{x}^k = \mathrm{Argmin}\left\{\theta_1(x) + \frac{r}{2}\left\|x - \left[x^k + \frac{1}{r}A^{\mathrm{T}}(2\tilde{y}^k - y^k)\right]\right\|^2 \,\middle|\, x \in \mathcal{X}\right\} \tag{3.9b}$$

得到 $\tilde{x}^k$.

当然, G- 模下 PPA 预测点 $(\tilde{x}^k, \tilde{y}^k) \in \mathcal{X} \times \mathcal{Y}$ 也可以通过先 x, 后 y 的顺序完成:

对给定的 (x^k, y^k), 通过求解 x 的子问题

$$\tilde{x}^k = \mathrm{Argmin}\left\{\theta_1(x) - x^{\mathrm{T}}A^{\mathrm{T}}y^k + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\right\} \tag{3.10a}$$

得到 $\tilde{x}^k$, 然后通过求解一个关于 y 的子问题

$$\tilde{y}^k = \mathrm{Argmin}\left\{\theta_2(y) + \frac{s}{2}\left\|y - \left[y^k - \frac{1}{s}A(2\tilde{x}^k - x^k)\right]\right\|^2 \,\middle|\, y \in \mathcal{Y}\right\} \tag{3.10b}$$

得到 $\tilde{y}^k$.

由 (3.10) 生成的 $(\tilde{x}^k, \tilde{y}^k) \in \mathcal{X} \times \mathcal{Y}$, 对一切 求 $(x, y) \in \mathcal{X} \times \mathcal{Y}$, 都有

$$\theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \end{pmatrix}^{\mathrm{T}} \left\{ \begin{pmatrix} -A^{\mathrm{T}}\tilde{y}^k \\ A\tilde{x}^k \end{pmatrix} + \begin{pmatrix} r(\tilde{x}^k - x^k) + A^{\mathrm{T}}(\tilde{y}^k - y^k) \\ A(\tilde{x}^k - x^k) + s(\tilde{y}^k - y^k) \end{pmatrix} \right\} \geqslant 0, \quad \forall\, u \in \Omega. \tag{3.11}$$

无论是 (3.8) 还是 (3.11), 都可以写成

$$\tilde{u}^k \in \Omega, \quad (u-\tilde{u}^k)^{\mathrm{T}}\{F(\tilde{u}^k)+G(\tilde{u}^k-u^k)\} \geqslant 0, \quad \forall\, u \in \Omega \tag{3.12}$$

的形式, 所不同的只是它们的矩阵 G 分别是

$$G=\begin{pmatrix} rI_n & -A^{\mathrm{T}} \\ -A & sI_m \end{pmatrix} \quad \text{和} \quad G=\begin{pmatrix} rI_n & A^{\mathrm{T}} \\ A & sI_m \end{pmatrix}.$$

它们都在当 $rs>\|A^{\mathrm{T}}A\|$ 时正定.

4 收敛速率

定制的 G- 模下的 PPA 算法通过选择特定的正定矩阵 G, 大大简化了子问题的计算. 该算法在图像处理中[1,15] 取得令人满意的结果, 最近也引起一批研究图像的学者的重视与推广[16,17]. 一个好的方法, 需要有好的理论结果. 我们利用 (2.8) 中定义的误差度量, 研究该算法的复杂性, 指出其具有 $O(1/k)$ 的收敛速率.

引理 4.1 设 $\tilde{u}$ 和 $\tilde{v}$ 分别是子问题 (2.2) 对给定的 u 和 v 的解. 我们有

$$(u-v)^{\mathrm{T}}G(\tilde{u}-\tilde{v}) \geqslant \|\tilde{u}-\tilde{v}\|_G^2. \tag{4.1}$$

证明 对给定的 u, $\tilde{u}$ 是子问题 (2.2) 的解, 所以有

$$\tilde{u}\in\Omega, \quad \theta(u')-\theta(\tilde{u})+(u'-\tilde{u})^{\mathrm{T}}\{F(\tilde{u})+G(\tilde{u}-u)\} \geqslant 0, \quad \forall\, u'\in\Omega.$$

由于 $\tilde{v}\in\Omega$, 用 $\tilde{v}$ 代上式的 u', 就有

$$\theta(\tilde{v})-\theta(\tilde{u})+(\tilde{v}-\tilde{u})^{\mathrm{T}}\{F(\tilde{u})+G(\tilde{u}-u)\} \geqslant 0.$$

对上式交换 u 与 v 的位置,

$$\theta(\tilde{u})-\theta(\tilde{v})+(\tilde{u}-\tilde{v})^{\mathrm{T}}\{F(\tilde{v})+G(\tilde{v}-v)\} \geqslant 0.$$

将上面两个不等式相加, 就有

$$(\tilde{u}-\tilde{v})^{\mathrm{T}}\{(F(\tilde{v})-F(\tilde{u}))+G(\tilde{v}-v)+G(u-\tilde{u})\} \geqslant 0.$$

经整理, 得

$$(u-v)^{\mathrm{T}}G(\tilde{u}-\tilde{v}) \geqslant \|\tilde{u}-\tilde{v}\|_G^2+(\tilde{u}-\tilde{v})^{\mathrm{T}}(F(\tilde{u})-F(\tilde{v})).$$

由于 F 是单调的, 从上式就得到引理的结论. □

引理 4.1 的结论说明, G- 模下的 PPA 算法是非扩张的. 根据该结论, 可以证明以下引理.

引理 4.2 设 $\tilde{u}$ 和 $\tilde{v}$ 分别是子问题 (2.2) 对给定的 u 和 v 的解. 我们有

$$(u-v)^{\mathrm{T}}G\{(u-\tilde{u})-(v-\tilde{v})\} \geqslant \|(u-\tilde{u})-(v-\tilde{v})\|_G^2. \tag{4.2}$$

证明 首先, (4.1) 可以改写成

$$(\tilde{u}-\tilde{v})^{\mathrm{T}}G\{(u-\tilde{u})-(v-\tilde{v})\}\geqslant 0.$$

上式两边都加上

$$\{(u-\tilde{u})-(v-\tilde{v})\}^{\mathrm{T}}G\{(u-\tilde{u})-(v-\tilde{v})\},$$

即得到引理的结论. □

基于引理 4.2 的结论, 可以得到以下定理.

定理 4.1 设 $\{u^k=(x^k,\lambda^k)\}$ 是 G- 模下外延的 PPA 算法产生的序列. G- 模下外延的 PPA 算法产生的误差界函数 $\{\|u^k-\tilde{u}^k\|_G\}$ 是单调下降的. 对所有的 $k>0$, 有

$$\|u^{k+1}-\tilde{u}^{k+1}\|_G^2\leqslant\|u^k-\tilde{u}^k\|_G^2-\frac{2-\gamma}{\gamma}\|(u^k-\tilde{u}^k)-(u^{k+1}-\tilde{u}^{k+1})\|_G^2. \tag{4.3}$$

证明 在恒等式 $\|a\|_G^2-\|b\|_G^2=2a^{\mathrm{T}}G(a-b)-\|a-b\|_G^2$ 中分别令 $a=(u^k-\tilde{u}^k)$ 和 $b=(u^{k+1}-\tilde{u}^{k+1})$, 我们有

$$\begin{aligned}&\|u^k-\tilde{u}^k\|_G^2-\|u^{k+1}-\tilde{u}^{k+1}\|_G^2\\&\quad=2(u^k-\tilde{u}^k)^{\mathrm{T}}G\{(u^k-\tilde{u}^k)-(u^{k+1}-\tilde{u}^{k+1})\}-\|(u^k-\tilde{u}^k)-(u^{k+1}-\tilde{u}^{k+1})\|_G^2.\end{aligned} \tag{4.4}$$

在 (4.2) 中令 $u=u^k$ 和 $v=u^{k+1}$, 得到

$$(u^k-u^{k+1})^{\mathrm{T}}G\{(u^k-\tilde{u}^k)-(u^{k+1}-\tilde{u}^{k+1})\}\geqslant\|(u^k-\tilde{u}^k)-(u^{k+1}-\tilde{u}^{k+1})\|_G^2. \tag{4.5}$$

将 (见 (2.5))

$$(u^k-u^{k+1})=\gamma(u^k-\tilde{u}^k)$$

代入 (4.5), 得到

$$(u^k-\tilde{u}^k)^{\mathrm{T}}G\{(u^k-\tilde{u}^k)-(u^{k+1}-\tilde{u}^{k+1})\}\geqslant\frac{1}{\gamma}\|(u^k-\tilde{u}^k)-(u^{k+1}-\tilde{u}^{k+1})\|_G^2. \tag{4.6}$$

将 (4.6) 代入 (4.4) 就直接得到

$$\|u^k-\tilde{u}^k\|_G^2-\|u^{k+1}-\tilde{u}^{k+1}\|_G^2\geqslant\left(\frac{2}{\gamma}-1\right)\|(u^k-\tilde{u}^k)-(u^{k+1}-\tilde{u}^{k+1})\|^2.$$

定理的结论可以从上式直接得到. □

定理 4.1 的结论说明 (2.8) 中定义的误差度量是单调下降的. 结合定理 2.1 中的结论, 我们有:

定理 4.2 设 $\{u^k=(x^k,\lambda^k)\}$ 是 G- 模下外延的 PPA 算法产生的序列. 对任意的 $k>0$, 经过 k 次迭代以后, 误差函数 $\|u^k-\tilde{u}^k\|_G^2$ 满足

$$\|u^k-\tilde{u}^k\|_G^2\leqslant\frac{1}{\gamma(2-\gamma)(k+1)}\|u^0-u^*\|_G^2,\quad\forall\ u^*\in\Omega^*. \tag{4.7}$$

证明 首先, 由 (2.7) 式得

$$\gamma(2-\gamma)\sum_{i=0}^{\infty}\|u^i-\tilde{u}^i\|_G^2\leqslant\|u^0-u^*\|_G^2,\quad\forall\ u^*\in\Omega^*. \tag{4.8}$$

再由定理 4.1, 序列 $\{\|u^i-\tilde{u}^i\|_G^2\}$ 是单调非递增的. 所以, 我们有

$$(k+1)\|u^k-\tilde{u}^k\|_G^2 \leqslant \sum_{i=0}^{k}\|u^i-\tilde{u}^i\|_G^2. \tag{4.9}$$

定理的结论 (4.7) 可以从 (4.8) 和 (4.9) 直接得到. □

该定理说明了 (2.8) 中定义的误差度量有 $O(1/k)$ 的收敛速率. 我们接着证明算法在遍历意义下的收敛速率. 首先, 直接利用 [18] 中的定理 2.1 刻画 (1.1) 的解集.

定理 4.3 混合变分不等式 (1.1) 的解集是凸的并可以表示成

$$\Omega^* = \bigcap_{u\in\Omega}\{\tilde{u}\in\Omega : (\theta(u)-\theta(\tilde{u})) + (u-\tilde{u})^{\mathrm{T}}F(u)\geqslant 0\}. \tag{4.10}$$

定理 4.3 说明, 对给定的 $\epsilon>0$, 如果 $\tilde{u}\in\Omega$ 满足

$$\theta(u)-\theta(\tilde{u})+(u-\tilde{u})^{\mathrm{T}}F(u)\geqslant -\epsilon, \quad \forall\, u\in\Omega,$$

它就是混合变分不等式 (1.1) 的 ϵ- 近似解. 换言之, 我们可以致力于寻找 $\tilde{u}\in\Omega$, 使得

$$\theta(\tilde{u})-\theta(u)+(\tilde{u}-u)^{\mathrm{T}}F(u)\leqslant \epsilon, \quad \forall\, u\in\Omega. \tag{4.11}$$

证明遍历意义下的收敛速率, 主要利用 (2.2) 和 (2.5). 由 (2.2) 和 F 的单调性, 我们有 $\tilde{u}^k\in\Omega$ 和

$$\theta(u)-\theta(\tilde{u}^k)+(u-\tilde{u}^k)^{\mathrm{T}}F(u)\geqslant (u-\tilde{u}^k)^{\mathrm{T}}G(u^k-\tilde{u}^k), \quad \forall\ u\in\Omega.$$

对上式右端利用 (2.5) 就有

$$\theta(u)-\theta(\tilde{u}^k)+(u-\tilde{u}^k)^{\mathrm{T}}F(u)\geqslant \frac{1}{\gamma}(u-\tilde{u}^k)^{\mathrm{T}}G(u^k-u^{k+1}), \quad \forall\ u\in\Omega. \tag{4.12}$$

再对 $(u-\tilde{u}^k)^{\mathrm{T}}G(u^k-u^{k+1})$ 利用恒等式

$$(a-b)^{\mathrm{T}}G(c-d)=\frac{1}{2}(\|a-d\|_G^2-\|a-c\|_G^2)+\frac{1}{2}(\|c-b\|_G^2-\|d-b\|_G^2),$$

并在其中令 $a=u$, $b=\tilde{u}^k$, $c=u^k$ 和 $d=u^{k+1}$, 就从 (4.12) 式得到

$$\theta(u)-\theta(\tilde{u}^k)+(u-\tilde{u}^k)^{\mathrm{T}}F(u)+\frac{1}{2\gamma}(\|u-u^k\|_G^2-\|u-u^{k+1}\|_G^2)\geqslant\frac{1}{2\gamma}(\|u^k-\tilde{u}^k\|_G^2-\|u^{k+1}-\tilde{u}^k\|_G^2).$$

再次利用 (2.5), 得到上式右端非负. 因此有下面的关键不等式:

$$\theta(u)-\theta(\tilde{u}^k)+(u-\tilde{u}^k)^{\mathrm{T}}F(u)+\frac{1}{2\gamma}(\|u-u^k\|_G^2-\|u-u^{k+1}\|_G^2)\geqslant 0, \quad \forall\ u\in\Omega. \tag{4.13}$$

据此, 我们证明算法在遍历意义下的收敛速率.

定理 4.4 对任意的整数 $t>0$, 定制的 PPA 会提供一个 $\tilde{u}_t\in\Omega$ 使得

$$(\theta(\tilde{u}_t)-\theta(u))+(\tilde{u}_t-u)^{\mathrm{T}}F(u)\leqslant\frac{1}{2\gamma(t+1)}\|u-u^0\|_G^2, \quad \forall\ u\in\Omega, \tag{4.14}$$

其中

$$\tilde{u}_t=\frac{1}{t+1}\sum_{k=0}^{t}\tilde{u}^k.$$

证明 对不等式 (4.13) 从 $k=0$ 加到 $k=t$, 有

$$\sum_{k=0}^{t}(\theta(u)-\theta(\tilde{u}^k))+\left((t+1)u-\sum_{k=0}^{t}\tilde{u}^k\right)^{\mathrm{T}}F(u)+\frac{1}{2\gamma}(\|u-u^0\|_G^2-\|u^{t+1}-u^0\|_G^2)\geqslant 0,\quad \forall\, u\in\Omega.$$

整理得

$$\left(\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k)-\theta(u)\right)+\left(\frac{1}{t+1}\sum_{k=0}^{t}\tilde{u}^k-u\right)^{\mathrm{T}}F(u)\leqslant\frac{\|u-u^0\|_G^2}{2\gamma(t+1)},\quad \forall\, u\in\Omega. \tag{4.15}$$

根据 $\tilde{u}_t$ 的表达式和 $\theta(u)$ 的凸性, 有

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

因此从 (4.15) 得到

$$(\theta(\tilde{u}_t)-\theta(u))+(\tilde{u}_t-u)^{\mathrm{T}}F(u)\leqslant\frac{\|u-u^0\|_G^2}{2\gamma(t+1)},\quad \forall\, u\in\Omega.$$

定理得证. □

对任何给定的有界闭集 $\mathcal{D}\subset\Omega$, 我们定义

$$d=\sup\{\|u-u^0\|_G^2\mid u\in\mathcal{D}\},$$

定制的 PPA 算法经过 t 次迭代, 就能得到一个 $\tilde{u}\in\Omega$, 使得

$$\sup_{u\in\mathcal{D}}\{\theta(\tilde{u})-\theta(u))+(\tilde{u}-u)^{\mathrm{T}}F(u)\}\leqslant\frac{d}{2\gamma(t+1)}.$$

对照 (4.11) 式, 我们得到定制的 PPA 算法遍历意义下的 $\mathcal{O}(1/t)$ 收敛速率.

致谢 作者感谢匿名审稿人提出的宝贵意见与建议, 这些对提高文章质量起到了很大帮助.

参考文献

1 Chambolle A, Pock T. A first-order primal-dual algorithm for convex problems with applications to imaging. J Math Imaging Vis, 2011, 40: 120–145

2 He B S, Yuan X M. A contraction method with implementable proximal regularization for linearly constrained convex programming. Available at: http://www.optimization-online.org/DB_HTML/2010/11/2817.html

3 Nemirovski A. Prox-method with rate of convergence $O(1/t)$ for variational inequality with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. SIAM J Optim, 2005, 15: 229–251

4 Nesterov Y. Smooth minimization of non-smooth functions. Math Program Ser A, 2005, 103: 127–152

5 Nesterov Y. Gradient methods for minimzing composite objective function. CORE discussion paper, 2007/76. Université Catholique de Louvain, Belgium

6 Martinet B. Regularisation, d'inéquations variationelles par approximations succesives. Rev Francaise d'Inform Recherche Oper, 1970, 4: 154–159

7 Rockafellar R T. Monotone operators and the proximal point algorithm. SIAM J Control Optim, 1976, 14: 877–898

8 Rockafellar R T. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math Oper Res, 1976, 1: 97–116

9 Eckstein J, Bertsekas D P. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math Program, 1992, 55: 293–318

10 Ha C. A generalization of the proximal point algorithm. SIAM J Control Optim, 1990, 28: 503–512

11 Zhao X Y, Sun D F, Toh K C. A Newton-CG augmented Lagrangian method for semidefinite programming. SIAM J Optim, 2010, 20: 1737–1765

12 Gu G Y, He B S, Yuan X M. Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a uniform approach. Available at: http://www.optimization-online.org/DB_FILE/2011/10/3220.pdf
13 Cai X J, Gu G Y, He B S, Yuan X M. A relaxed customized proximal point algorithm for separable convex programming. Available at: http://www.optimization-online.org/DB_FILE/2011/08/3141.pdf
14 Chen C H, He B S, Yuan X M. Matrix completion via alternating direction methods. IMA J Numer Anal, 2012, 32: 227–245
15 He B S, Yuan X M. Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. SIAM J Imaging Sci, 2012, 5: 119–149
16 Condat L. A generic first-order primal-dual method for convex optimization involving Lipschitzian, proximable and linear composite terms. Image Team of the GREYC laboratory, A joint CNRS-UCBN-ENSICAEN research unit in Caen, France, 2011
17 Pock T, Chambolle A. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. Institute for Computer Graphics and Vision. Graz University of Technology, Austria, 2011
18 He B S, Yuan X M. On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. SIAM J Numer Anal, 2012, 50: 700-709

On the convergence rate of customized proximal point algorithm for convex optimization and saddle-point problem

HE BingSheng & SHEN Yuan

Abstract The first-order optimality condition of linearly constrained convex optimization and saddle-point problems is a monotone variational inequality. Under the framework of variational inequality, by properly selecting a matrix G, the subproblems of proximal point algorithm under G-norm can be solved easily during the iteration progress. We derive the $O(1/k)$ convergence rate of such customized proximal point algorithm.

Keywords **convex optimization, monotone operator, proximal point algorithm under G-norm, convergence rate**

MSC(2010) **90C25, 90C30, 90C33**

doi: 10.1360/012011-1049