# 一些典型凸优化问题的分裂收缩算法

## 基于变分不等式和邻近点算法的统一框架

南师大数科院系列讲座　2022 年元月 4 – 9 日

何 炳 生

Homepage: maths.nju.edu.cn/~hebma

## 一. 凸优化 变分不等式 邻近点算法

凸优化问题的一阶最优性条件是一个单调变分不等式 (Variational Inequality). 变分不等式是盲人爬山判别是否到达顶点的数学表达形式. 约束优化问题, 引进乘子就有了拉格朗日函数, 线性约束凸优化问题的拉格朗日函数的鞍点等价于相应的变分不等式的解点.

## 1 盲人爬山和变分不等式

我们先比较直观的来看凸规划的最优性条件. 设 $\Omega \subset \Re^n$ 是一个凸集, $f: \Re^n \to \Re$ 是一个可微凸函数. 考察凸优化问题

$$\min\{f(x) \mid x \in \Omega\}. \tag{1.1}$$

什么样的 $x$ 才是最优点? 用最通俗的话讲,

$$\text{最优点必须属于 } \Omega\text{, 并且从这点出发的任何可行方向都不是下降方向.} \tag{1.2}$$

我们用 $\nabla f(x)$ 表示 $f(x)$ 的梯度, 对于属于 $\Omega$ 的 $x$, 记

$$Sf(x) = \{s \in \Re^n \mid s = x' - x,\ x' \in \Omega\},$$

$Sf(x)$ 就是优化问题 (1.1) 在点 $x$ 处的可行方向集；记

$$Sd(x) = \{s \in \Re^n \mid s^T \nabla f(x) < 0\},$$

那么 $Sd(x)$ 就是函数 $f(x)$ 在 $x$ 处的下降方向集. 利用这些记号, (1.2) 就相当于

$$x^* \text{ 是最优解} \quad \Longleftrightarrow \quad x^* \in \Omega \ \& \ Sf(x^*) \cap Sd(x^*) = \emptyset. \tag{1.3}$$

最优性条件 (1.3) 的等价数学形式就是

$$x^* \in \Omega, \quad (x - x^*)^T \nabla f(x^*) \ge 0, \quad \forall x \in \Omega. \tag{1.4}$$

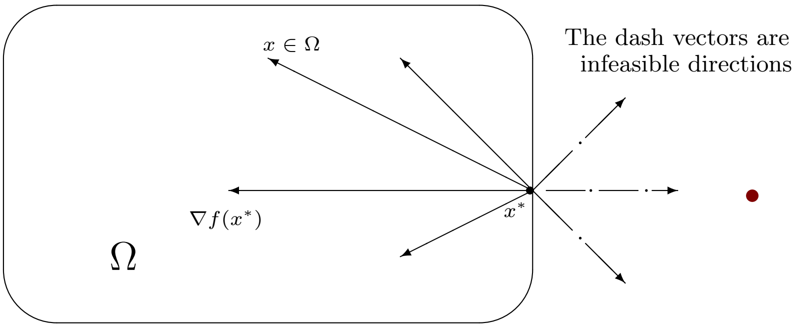



图 1. 可微凸优化与变分不等式的关系

顺便提及一下, 在求最大化的时候, 将下降方向的集合 $Sd(x)$ 改成上升方向 (ascent direction) 的集合

$$Sa(x) = \{s \in \Re^n \mid s^T \nabla f(x) > 0\}, \text{ 点 } x \text{ 处的上升方向集.}$$

相应与 (1.3) 的关系式就变成

$$x^* \in \Omega, \quad Sf(x^*) \cap Sa(x^*) = \emptyset.$$

这表示所有可行方向都不是上升方向, 就是通常所说的 **盲人爬山原理**.

将 $\nabla f(x)$ 写成 $F(x)$, (1.4) 就成了

$$x^* \in \Omega, \quad (x - x^*)^T F(x^*) \geq 0, \quad \forall x \in \Omega \tag{1.5}$$

这样一个单调变分不等式. 变分不等式 (1.5) 中的 $F(x)$, 不一定是某个函数的梯度. 换句话说, 当 $F(x)$ 可微时, $F(x)$ 的雅可比矩阵不一定对称.

**引理 1.1** 设 $\mathcal{X} \subset \Re^n$ 是闭凸集, $\theta(x)$ 和 $f(x)$ 都是凸函数, 其中 $f(x)$ 在包含 $\mathcal{X}$ 的一个开集上可微. 记 $x^*$ 是凸优化问题 $\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}$ 的解. 我们有

$$x^* \in \arg\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\} \tag{1.6a}$$

的充分必要条件是

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \in \mathcal{X}. \tag{1.6b}$$

**证明:** 首先, 如果(1.6a) 正确, 那么对任意的 $x \in \mathcal{X}$, 我们有

$$\frac{\theta(x_\alpha) - \theta(x^*)}{\alpha} + \frac{f(x_\alpha) - f(x^*)}{\alpha} \geq 0, \tag{1.7}$$

其中

$$x_\alpha = (1 - \alpha)x^* + \alpha x, \quad \forall \alpha \in (0, 1].$$

因为 $\theta(x)$ 是 $x$ 的凸函数, 根据 $x_\alpha$ 的定义有

$$\theta(x_\alpha) \leq (1 - \alpha)\theta(x^*) + \alpha\theta(x),$$

并且因此

$$\theta(x)-\theta(x^*)\ge\frac{\theta(x_\alpha)-\theta(x^*)}{\alpha},\ \ \forall\alpha\in(0,1].$$

将此代入 (1.7) 的左边, 我们就有

$$\theta(x)-\theta(x^*)+\frac{f(x_\alpha)-f(x^*)}{\alpha}\ge 0,\quad \forall\alpha\in(0,1].$$

注意到上式中 $f(x_\alpha)=f(x^*+\alpha(x-x^*))$, 由于 $f(x)$ 可微, 令 $\alpha\to 0_+$, 得到

$$\theta(x)-\theta(x^*)+\nabla f(x^*)^T(x-x^*)\ge 0,\ \ \forall\, x\in\mathcal{X}.$$

这样就从 (1.6a) 得到了 (1.6b). 反过来, 因为 $f$ 是凸函数, 有

$$f(x_\alpha)\le(1-\alpha)f(x^*)+\alpha f(x),$$

这可以写成

$$f(x_\alpha)-f(x^*)\le\alpha(f(x)-f(x^*)).$$

因此, 对所有的 $\alpha\in(0,1]$, 我们有

$$f(x)-f(x^*)\ge\frac{f(x_\alpha)-f(x^*)}{\alpha}=\frac{f(x^*+\alpha(x-x^*))-f(x^*)}{\alpha},$$

取 $\alpha\to 0_+$, 有

$$f(x)-f(x^*)\ge\nabla f(x^*)^T(x-x^*).$$

将此代入 (1.6b) 的左边, 得

$$x^*\in\mathcal{X},\quad \theta(x)-\theta(x^*)+f(x)-f(x^*)\ge 0,\quad \forall\, x\in\mathcal{X},$$

因此 (1.6a) 为真. 引理得证. □

这个引理相当简单, 却是这本著作要用的几个基本原理之一, 它建立了凸优化问题和变分不等式之间的一一对应关系. 以后我们提到这个引理, 称其为 *最优性条件的基本引理*. 我们称 (1.6b) 这样的形式为单调混合变分不等式. 当然, 如果 $\theta(x)$ 也可微, 相应的条件就可以写成

$$x^*\in\mathcal{X},\quad (x-x^*)^T\big(\nabla\theta(x^*)+\nabla f(x^*)\big)\ge 0,\quad \forall\, x\in\mathcal{X}.$$

利用这个引理, 我们同样能得到凸函数一阶性质的几何解释(1.4).

# 2 线性约束凸优化问题及等价的变分不等式

设 $\mathcal{U}\subset\Re^n$ 是闭凸集, $\theta:\Re^n\to\Re$ 是(并非一定光滑的)凸函数, $\mathcal{A}\in\Re^{m\times n}$, $b\in\Re^m$. 我们考虑线性约束的凸优化问题

$$\min\{\theta(u)\mid \mathcal{A}u=b\ (\mathrm{or}\ \ge b), u\in\mathcal{U}\},\tag{2.1}$$

并假设它有解. 问题(2.1)的拉格朗日 (Lagrange) 函数

$$L(u,\lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b) \tag{2.2}$$

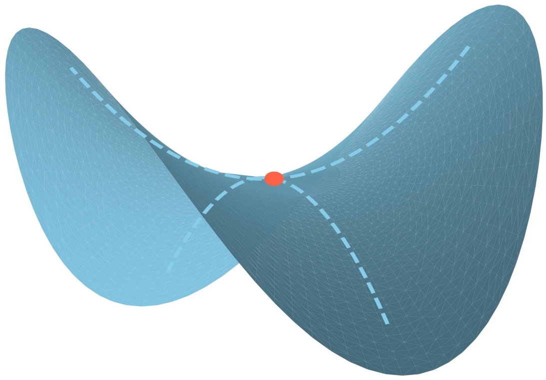
定义在 $\mathcal{U} \times \Lambda$ 上, 其中

$$\Lambda = \begin{cases} \Re^m, & \text{if } \mathcal{A}u = b, \\ \Re^m_+, & \text{if } \mathcal{A}u \ge b, \end{cases}$$

这里的 $\Re^m_+$ 表示 $\Re^m$ 中的非负卦限. 如果一对 $(u^*,\lambda^*)$ 满足

$$(u^*,\lambda^*) \in \mathcal{U} \times \Lambda, \quad L(u^*,\lambda) \le L(u^*,\lambda^*) \le L(u,\lambda^*), \quad \forall (u,\lambda) \in \mathcal{U} \times \Lambda, \tag{2.3}$$

就称它为Lagrange 函数(2.3) 的鞍点, 鞍点中的 $u^*$ 就是(2.1)的解.



定义鞍点的不等式(2.3)的两部分写开来就是

$$\begin{cases} u^* \in \mathcal{U}, & L(u,\lambda^*) - L(u^*,\lambda^*) \ge 0, & \forall\, u \in \mathcal{U}, \\ \lambda^* \in \Lambda, & L(u^*,\lambda^*) - L(u^*,\lambda) \ge 0, & \forall\, \lambda \in \Lambda. \end{cases}$$

利用拉格朗日函数(2.2)的表达式, 从上式得到下面的变分不等式组

$$\begin{cases} u^* \in \mathcal{U}, & \theta(u) - \theta(u^*) + (u-u^*)^T(-\mathcal{A}^T\lambda^*) \ge 0, & \forall\, u \in \mathcal{U}, \\ \lambda^* \in \Lambda, & (\lambda-\lambda^*)^T(\mathcal{A}u^* - b) \ge 0, & \forall\, \lambda \in \Lambda. \end{cases} \tag{2.4}$$

将上式两部分加在一起写成紧凑的形式就是变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w-w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{2.5a}$$

其中

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix}, \quad \Omega = \mathcal{U} \times \Lambda. \tag{2.5b}$$

对(2.4)中任意的 $w \in \Omega$ 分别取 $w = (u,\lambda^*)$ 和 $w = (u^*,\lambda)$, 就能反过来得到(2.3). 这样, 我们把鞍点的集合和变分不等式的解集建立了等价关系. 通篇, 我们都用 $\Omega^*$ 表示变分不等式(2.4)的解集 (因此也就是拉格朗日函数(2.2)鞍点的集合). 注意到(2.5b)中, 仿射算子

$$F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} = \begin{pmatrix} 0 & -\mathcal{A}^T \\ \mathcal{A} & 0 \end{pmatrix} \begin{pmatrix} u \\ \lambda \end{pmatrix} - \begin{pmatrix} 0 \\ b \end{pmatrix}$$

中的矩阵为反对称矩阵, 因此有$(w-\tilde{w})^T(F(w)-F(\tilde{w}))\equiv 0$. 由此而得的恒等式

$$(\tilde{w}-w^*)^T F(\tilde{w}) \equiv (\tilde{w}-w^*)^T F(w^*) \tag{2.6}$$

在算法框架的收敛性证明中常常要用到.

## 可分离结构型凸优化问题等价的变分不等式

科学计算中的许多问题可以归结为一个有两个可分离目标函数的结构型凸优化问题

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\}. \tag{2.7}$$

这相当于在 (2.1) 中, 置 $n=n_1+n_2$, $\mathcal{X}\subset\Re^{n_1}$, $\mathcal{Y}\subset\Re^{n_2}$, $\mathcal{U}=\mathcal{X}\times\mathcal{Y}$. $\theta(u)=\theta_1(x)+\theta_2(y)$, $\theta_1(x):\Re^{n_1}\to\Re$, $\theta_2(y):\Re^{n_2}\to\Re$. 矩阵 $\mathcal{A}=(A,B)$, 其中$A\in\Re^{m\times n_1}$, $B\in\Re^{m\times n_2}$. 这个问题的最优性条件同样可以表示成一个单调的（混合）变分不等式:

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^T F(w^*)\ge 0,\quad \forall w\in\Omega, \tag{2.8a}$$

其中

$$u=\begin{pmatrix} x\\ y\end{pmatrix},\quad w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad \theta(u)=\theta_1(x)+\theta_2(y), \tag{2.8b}$$

和

$$F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix},\qquad \Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m. \tag{2.8c}$$

同样, 对三个可分离目标函数线性约束的凸优化问题

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z) \mid Ax+By+Cz=b,\ x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\}, \tag{2.9}$$

通过类似的分析, 它的最优性条件同样可以表示成一个单调的（混合）变分不等式:

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^T F(w^*)\ge 0,\quad \forall w\in\Omega, \tag{2.10a}$$

其中

$$u=\begin{pmatrix} x\\ y\\ z\end{pmatrix},\quad w=\begin{pmatrix} x\\ y\\ z\\ \lambda\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ -C^T\lambda\\ Ax+By+Cz-b\end{pmatrix}, \tag{2.10b}$$

和

$$\theta(u)=\theta_1(x)+\theta_2(y)+\theta_3(z),\qquad\qquad \Omega=\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}\times\Re^m. \tag{2.10c}$$

在变分不等式 (2.8) 和 (2.10) 中, $u$ 对应的是凸优化问题中的原始变量, $w=(u,\lambda)$ 是原始变量加对偶变量. 通篇, 我们把约束集合是 $\Omega$ 的变分不等式的解集记为 $\Omega^*$.

虽然 (2.7) 和 (2.9) 相对应的单调混合变分不等式形式完全相同, 但 (2.7) 是含两个可分离目标函数的结构型约束凸优化问题, 可以用相当有效的乘子交替方向法 (Alternating directions method of multipliers) [3] 求解, 而直接将乘子交替方向法推广到求解三个算子的可分离结构型约束凸优化问题 (2.9), 论文 [2] 证明是不一定收敛的. 对多于两个算子的问题, 需要用一些基于变分不等式的预测-校正方法去处理 [6, 7, 8].

# 3 凸优化和变分不等式的邻近点算法

邻近点算法 (Proximal Point Algorithm) 简称 PPA 算法, 是求解凸优化问题

$$\min\{\theta(x)+f(x) \mid x\in\mathcal{X}\} \tag{3.1}$$

的一类基本算法[11, 12]. 为了介绍求解变分不等式(2.4)的 PPA 算法, 我们先证明一个简单的结论.

**引理 3.1** 设 $a,b\in\Re^n$, $H\in\Re^{n\times n}$ 是对称正定矩阵. 如果

$$b^T H(a-b)\ge 0,$$

则有

$$\|b\|_H^2 \le \|a\|_H^2 - \|a-b\|_H^2.$$

**证明** 对任何给定的正定矩阵$H$, 有

$$\|a\|_H^2 = \|b+(a-b)\|_H^2 = \|b\|_H^2 + 2b^T H(a-b) + \|a-b\|_H^2.$$

引理的结论可以从上面的恒等式和引理的假设条件直接得到. □

**求解凸优化问题 (3.1) 的 PPA 算法**. 第 $k$-步迭代从给定的 $x^k$ 开始, 新的迭代点

$$x^{k+1} = \operatorname{argmin}\{\theta(x)+f(x)+\frac{r}{2}\|x-x^k\|^2 \mid x\in\mathcal{X}\}, \quad (r>0). \tag{3.2}$$

PPA 算法的第 $k$-步迭代, 在优化问题原目标函数后面加上正则项 $\frac{r}{2}\|x-x^k\|^2$, 使得新的迭代点离原来的迭代点不要太远, 犹如探险过程中步步为营、稳扎稳打.

**引理 3.2** *PPA* 方法(3.2)产生的序列 $\{x^k\}$ 满足

$$\|x^{k+1}-x^*\|^2 \le \|x^k-x^*\|^2 - \|x^k-x^{k+1}\|^2, \quad \forall x^*\in\mathcal{X}^*. \tag{3.3}$$

**证明**. 因为 $x^{k+1}$ 是(3.2)的最优解, 根据引理 1.1 有

$$x^{k+1}\in\mathcal{X},\ \ \theta(x)-\theta(x^{k+1})+(x-x^{k+1})^T\{\nabla f(x^{k+1})+r(x^{k+1}-x^k)\}\ge 0,\ \ \forall x\in\mathcal{X}.$$

令上式中任意的 $x=x^*$, 得到

$$(x^{k+1}-x^*)^T r(x^k-x^{k+1}) \ge \theta(x^{k+1})-\theta(x^*)+(x^{k+1}-x^*)^T\nabla f(x^{k+1}). \tag{3.4}$$

由于 $f$ 是凸的, $(x^{k+1}-x^*)^T\nabla f(x^{k+1}) \ge (x^{k+1}-x^*)^T\nabla f(x^*)$. 因此有

$$\begin{aligned}&\theta(x^{k+1})-\theta(x^*)+(x^{k+1}-x^*)^T\nabla f(x^{k+1})\\ \ge\;& \theta(x^{k+1})-\theta(x^*)+(x^{k+1}-x^*)^T\nabla f(x^*)\ge 0.\end{aligned}$$

上式中最后一个不等式的根据是(1.6b). 将上面的结果代入 (3.4) 的右端, 得到

$$(x^{k+1}-x^*)^T(x^k-x^{k+1})\ge 0. \tag{3.5}$$

令 $a=x^k-x^*$, $b=x^{k+1}-x^*$ 并利用引理 3.1, 便有

$$\|x^{k+1}-x^*\|^2\le \|x^k-x^*\|^2-\|x^k-x^{k+1}\|^2.$$

其中 $x^*$ 表示问题 (3.1)的任何一个解点. 引理得证. □

上式是证明凸优化 PPA 算法收敛的关键不等式.

**求解变分不等式 (2.5) 的 $H$-模下的 PPA 算法.**
设 $H\succ 0$ 为给定的正定矩阵, 求解变分不等式(2.5)的 $H$-模下的PPA 算法的第 $k$-步从已知的 $w^k$ 出发, 求得的新迭代点 $w^{k+1}$ 使得

$$\begin{aligned} w^{k+1}\in\Omega,\quad \theta(u)-\theta(u^{k+1})&+(w-w^{k+1})^TF(w^{k+1})\\ &\ge (w-w^{k+1})^TH(w^k-w^{k+1}),\ \ \forall\, w\in\Omega.\end{aligned}\tag{3.6}$$

显然, 如果在(3.6)中有 $w^k=w^{k+1}$, 那么 $w^{k+1}$ 就是变分不等式问题(2.4)的解. 有时, 我们也把 (3.6) 写成等价的形式

$$w^{k+1}\in\Omega,\ \theta(u)-\theta(u^{k+1})+(w-w^{k+1})^T\{F(w^{k+1})+H(w^{k+1}-w^k)\}\ge 0,\ \forall\, w\in\Omega.$$

其中的 $H(w^{k+1}-w^k)$ 是二次函数 $\frac{1}{2}\|w-w^k\|_H^2$ 在 $w^{k+1}$ 处的梯度, 这也是把(3.6)称为求解变分不等式(2.4)的$H$-模下邻近点算法的原因.

**引理 3.3** 对给定的正定矩阵 $H\succ 0$ 和向量 $w^k$, 如果 $w^{k+1}$ 是由(3.6)提供的, 那么有

$$\|w^{k+1}-w^*\|_H^2\le \|w^k-w^*\|_H^2-\|w^k-w^{k+1}\|_H^2,\quad \forall w^*\in\Omega^*. \tag{3.7}$$

**证明** 将(3.6)中任意确定的 $w\in\Omega$ 设为 $w^*$, 就有

$$(w^{k+1}-w^*)^TH(w^k-w^{k+1})\ge \theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^{k+1}). \tag{3.8}$$

利用 $(w^{k+1}-w^*)^TF(w^{k+1})=(w^{k+1}-w^*)^TF(w^*)$ (见(2.6)), 可以把(3.8)的右端改写为与它等同的

$$\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^*).$$

由 $w^{k+1}\in\Omega$ 和 $w^*$ 是(2.4)的解, 上式非负. 因此从(3.8)得到

$$(w^{k+1}-w^*)^TH(w^k-w^{k+1})\ge 0.$$

在引理 3.1 中令 $a = w^k - w^*$ 和 $b = w^{k+1} - w^*$, 就从上式得到引理的结论. $\Box$

引理 3.3 说明迭代点离解集里的每一点都越来越近. 具有性质(3.7)的序列 $\{w^k\}$ 被称为在 $H$-模下 Fejér 单调的, 它是证明迭代序列 $\{w^k\}$ 收敛的关键不等式. 对变分不等式(2.5), 我们在 [10] 中首次明确提出了 $H$-模下的 PPA 算法. 用 PPA 求解凸优化问题, 方法简单, 收敛性证明也特别容易被理解 [1].

引理 3.2 和引理 3.3 分别提供了求解凸优化问题 (3.1)和变分不等式 (2.5)的 PPA算法的关键不等式. 下面我们证明变分不等式 PPA算法的收敛性.

**定理 3.1** 设 $\{w^k\}$ 是求解变分不等式 (2.5) 的 $H$-模下的 *PPA* 算法 (3.6) 产生的迭代序列. 那么 $\{w^k\}$ 收敛于变分不等式的某个解点.

**证明** 由引理 3.3, 我们有

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2, \quad \forall w^* \in \Omega^* \tag{3.9}$$

和

$$\sum_{k=0}^{\infty} \|w^k - w^{k+1}\|_H^2 \le \|w^0 - w^*\|_H, \quad \forall w^* \in \Omega^*.$$

因此迭代序列 $\{w^k\}$ 包含在一个有界闭集里并有

$$\lim_{k\to\infty} \|w^k - w^{k+1}\|_H^2 = 0. \tag{3.10}$$

设 $\{w^{k_j}\}$ 是 $\{w^k\}$ 的收敛于 $w^\infty$ 的子列, 由 (3.6), 有

$$\begin{aligned} w^{k_j} \in \Omega, \quad \theta(u) - \theta(u^{k_j}) &+ (w - w^{k_j})^T F(w^{k_j}) \\ &\ge (w - w^{k_j})^T H(w^{k_j-1} - w^{k_j}), \ \ \forall\, w \in \Omega. \end{aligned}$$

对上式求极限并利用 (3.10), 得到

$$w^\infty \in \Omega, \quad \theta(u) - \theta(u^\infty) + (w - w^\infty)^T F(w^\infty \ge 0, \quad \forall w \in \Omega.$$

因此, $w^\infty$ 是变分不等式 (2.5) 的解. 由于 $w^\infty \in \Omega^*$, 根据 (3.9), 有

$$\|w^{k+1} - w^\infty\|_H^2 \le \|w^k - w^\infty\|_H^2.$$

因此迭代序列只能收敛于这个 $w^\infty$. $\Box$

定理 3.1 证明中用到的基本程式, 在以后的章节中还会经常用到.

# 参考文献

[1] X.J. Cai, G.Y. Gu, B.S. He and X.M. Yuan, A proximal point algorithms revisit on the alternating direction method of multipliers, Science China Mathematics, 56 (2013), 2179-2186.

[2] C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, *The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent*, to appear in *Mathematical Programming, Series A*.

[3] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[4] Gu G Y, He B S and Yuan X M. Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach. Comput. Optim. Appl., 2014, 59: 135-161.

[5] B. S. He, PPA-like contraction methods for convex optimization: a framework using variational inequality approach. J. Oper. Res. Soc. China **3** (2015) 391 – 420.

[6] B. S. He, Parallel splitting augmented Lagrangian methods for monotone structured variational inequalities, *Computational Optimization and Applications* **42**(2009), 195–212.

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

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

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

[10] B.S. He and X. M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, SIAM Journal on Imaging Science, **5**, 119-149, 2012.

[11] B. Martinet, Regularisation, d'inéquations variationelles par approximations succesives, *Rev. Francaise d'Inform. Recherche Oper.*, **4**, 154-159, 1970.

[12] R.T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Cont. Optim.*, **14**, 877-898, 1976.

[13] 何炳生. 论求解变分不等式的一些投影收缩算法, 计算数学, 1996, 18: 54–60.

[14] 何炳生. 从变分不等式的统一收缩算法到凸优化的分裂收缩算法, 高等学校计算数学学报, 2016, 38: 74–96.

[15] 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, 见 http://maths.nju.edu. cn/~hebma 中的《My Talk》.

[16] 何炳生. 凸优化和单调变分不等式的收缩算法, 见 http://maths.nju.edu.cn/~hebma 中的系列讲义.