

# 从变分不等式的投影收缩算法 到凸优化的分裂收缩算法<sup>1</sup>

何炳生

南京大学数学系      南京大学管理科学与工程国际研究中心

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

**摘要.** 最近几年, 我们发表了一些求解线性约束凸优化问题的分裂收缩算法. 这些算法被用来求解图像处理和机器学习等领域的优化问题, 得到了相应的认可, 论文也得到了一些正面评价和实质性的引用. 事实上, 我们在凸优化分裂收缩算法方面的成果, 与前期对变分不等式的研究有密切联系. 这篇文章, 叙述一下这些研究工作的来龙去脉. 这类研究工作所需要的数学基础, 只是微积分和线性代数的一些最基本的概念. 起关键作用的还是中学就学到的一些数学道理.

**关键词.** 变分不等式, 投影收缩算法, 凸优化的分裂收缩算法

## 1 引言

历史原因, 我18岁高中毕业, 无奈而立才上的大学. 务农的10多年, 知道了什么是优选法. 大学念的是与年龄相悖的数学, 念完就做相对容易的最优化方法. 得知优化中互补问题(变分不等式的一种特殊情形)能用来描述经济与管理方面的平衡问题, 就对变分不等式的求解产生了兴趣. 实际生活中的问题, 函数只是一种对应关系, 通常没有显式表达式, 对一个给定的点, 要获得相应的函数值又往往要通过一次实验, 代价不菲. 求解这种“黑箱函数”的问题, 需要只用函数值且少用函数值的方法. 从管理科学中来的变分不等式问题, 函数往往是黑箱的. 受华罗庚先生推广优选法的影响, 对单调变分不等式求解, 我们对只用函数值的预测-校正算法比较感兴趣. 根据算法的特点, 称这类只用函数值算法为投影收缩算法, 是用前人在 [1] 中的术语.

近年来, 压缩感知, 图像处理, 机器学习等领域 [3] 提出了大量的优化问题. 信息科学中的此类问题, 初看形式各异, 但其中不少可以归结为 (或者松弛成) 以下一些典型的线性约束凸优化问题:

- 线性约束的凸优化问题       $\min\{\theta(x) | Ax = b \text{ (or } \geq b), x \in \mathcal{X}\};$
- 两个算子的可分离凸优化       $\min\{\theta_1(x) + \theta_2(y) | Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y}\};$
- 多个算子的可分离凸优化问题       $\min\{\sum_{i=1}^m \theta_i(x_i) | \sum_{i=1}^m A_i x_i = b, x_i \in \mathcal{X}_i\}.$

凸优化问题的一阶最优性条件就是一个单调变分不等式. 用变分不等式当框架, 很容易将投影收缩算法方面的成果移植到求解上述凸优化问题, 差别只在于变分不等式求解中用投影提供的预测, 现在通过求解适当的(往往有显式解或者容易求解的)子问题实现. 利用变分不等式框架, 既能帮助我们按需设计分裂算法, 又能极大地简化收敛性证明, 包括收敛速

<sup>1</sup>本文对前言略作修改后发表在《高等学校计算数学学报》38 卷第1 期(2016) 74-96.

率,也都迎刃而解.从单调变分不等式的投影收缩算法到凸优化的分裂收缩算法,其实是同一个原理.

读懂这些算法,只要对微积分中的梯度和线性代数中的矩阵有基本了解.就个人而言,由于当农民的10多年没有新的数学往脑子里灌,使得余弦定理等一些中学知识,能够在脑子里沉淀,又在后半生的工作中用上.要说有些基本想法,还得益于一些日常生活的理念.

## 2 变分不等式

设  $\Omega \subset \mathbb{R}^n$  是一个非空闭凸集,  $F$  是  $\mathbb{R}^n \rightarrow \mathbb{R}^n$  的一个映射. 考虑单调变分不等式问题:

$$u^* \in \Omega, \quad (u - u^*)^\top F(u^*) \geq 0, \quad \forall u \in \Omega. \quad (2.1)$$

说变分不等式单调,是指其中的算子  $F$  满足  $(u - v)^\top (F(u) - F(v)) \geq 0$ .

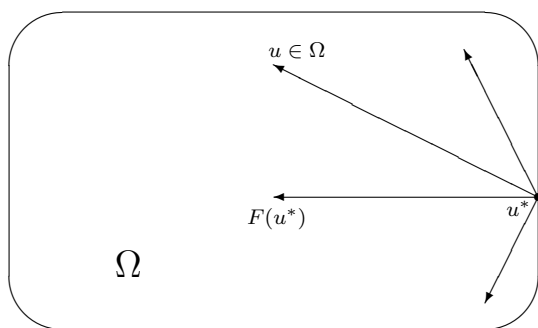


图 1. 变分不等式的几何解释

### 2.1 可微凸优化就是一个特殊的单调变分不等式

设  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  是一个可微凸函数. 考察凸优化问题

$$\min\{f(x) \mid x \in \Omega\}. \quad (2.2)$$

什么样的  $x$  才是最优点? 显然,它必须属于  $\Omega$ ,并且在这点没有下降的可行方向. (2.3)

我们用  $\nabla f(x)$  表示  $f(x)$  的梯度,并记

- $Sd(x) = \{s \in \mathbb{R}^n \mid s^\top \nabla f(x) < 0\}$ , 点  $x$  处的下降方向集;
- $Sf(x) = \{s \in \mathbb{R}^n \mid s = x' - x, x' \in \Omega\}$ , 点  $x$  处的可行方向集.

利用这些记号, (2.3) 就相当于

$$x \text{ is an optimal solution} \iff x \in \Omega \ \& \ Sf(x) \cap Sd(x) = \emptyset. \quad (2.4)$$

最优性条件 (2.4) 的等价数学形式就是

$$x \in \Omega, \quad (x' - x)^\top \nabla f(x) \geq 0, \quad \forall x' \in \Omega. \quad (2.5)$$

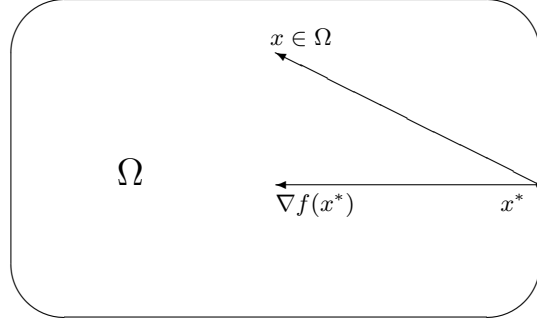


图 2. 可微凸优化与变分不等式的关系

可微凸函数最基本的性质是

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x),$$

因而也有

$$f(x) \geq f(y) + \nabla f(y)^\top (x - y).$$

两式一加就有

$$(x - y)^\top (\nabla f(x) - \nabla f(y)) \geq 0,$$

所以说凸函数的导算子是单调的. 将  $\nabla f(x)$  写成  $F(x)$ , (2.5) 就成了

$$x \in \Omega, \quad (x' - x)^\top F(x) \geq 0, \quad \forall x' \in \Omega$$

这样一个单调变分不等式. 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

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

相应与 (2.4) 的关系式就变成

$$Sf(x) \cap Sa(x) = \emptyset.$$

这表示所有可行方向都不是上升方向, 就是通常所说的 **瞎子爬山原理**.

**Lemma 2.1.** 设  $\mathcal{X} \subset \mathbb{R}^n$  是闭凸集,  $\theta(x)$  和  $f(x)$  都是凸函数, 其中  $f(x)$  可微. 记  $x^*$  是凸优化问题  $\min\{\theta(x) + f(x) \mid x \in \mathcal{X}\}$  的解. 我们有

$$x^* = \arg \min\{\theta(x) + f(x) \mid x \in \mathcal{X}\} \quad (2.6a)$$

的充分必要条件是

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^\top \nabla f(x^*) \geq 0, \quad \forall x \in \mathcal{X}. \quad (2.6b)$$

这个引理相当简单, 却是这篇文章中要用的几个基本原理之一. 我们称 (2.6b) 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

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

## 2.2 线性约束凸优化问题等价的变分不等式

设  $\theta(x)$  是  $\mathfrak{R}^n \rightarrow \mathfrak{R}$  的凸函数,  $A \in \mathfrak{R}^{m \times n}$ ,  $b \in \mathfrak{R}^m$ . 我们考虑如下的线性约束凸优化问题

$$\min\{\theta(x) \mid Ax = b \text{ (or } \geq b), x \in \mathcal{X}\}. \quad (2.7)$$

这个问题的 Lagrange 函数是定义在  $\mathcal{X} \times \Lambda$  上的

$$L(x, \lambda) = \theta(x) - \lambda^\top (Ax - b), \quad (2.8)$$

其中,  $\Lambda = \mathfrak{R}^m$  或者  $\mathfrak{R}_+^m$ , 分别对应于 (2.7) 中的线性约束为等式或者不等式. 一对  $(x^*, \lambda^*)$  如果满足

$$L_{\lambda \in \Lambda}(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L_{x \in \mathcal{X}}(x, \lambda^*)$$

就被称为 Lagrange 函数的鞍点. 换句话说, 鞍点就是同时满足

$$\begin{cases} x^* = \arg \min\{L(x, \lambda^*) \mid x \in \mathcal{X}\}, \\ \lambda^* = \arg \max\{L(x^*, \lambda) \mid \lambda \in \Lambda\}. \end{cases}$$

根据 (2.6), 上述优化子问题的最优性条件是

$$\begin{cases} x^* \in \mathcal{X}, & \theta(x) - \theta(x^*) + (x - x^*)^\top (-A^\top \lambda^*) \geq 0, \quad \forall x \in \mathcal{X}, \\ \lambda^* \in \Lambda, & (\lambda - \lambda^*)^\top (Ax^* - b) \geq 0, \quad \forall \lambda \in \Lambda. \end{cases} \quad (2.9)$$

这个最优性条件就可以表示成一个单调的 (混合) 变分不等式:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^\top F(w^*) \geq 0, \quad \forall w \in \Omega, \quad (2.10a)$$

其中

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^\top \lambda \\ Ax - b \end{pmatrix} \quad \text{和} \quad \Omega = \mathcal{X} \times \Lambda. \quad (2.10b)$$

由 (2.9) 到 (2.10), 毫无疑问. 对 (2.10) 中任意的  $w \in \Omega$ , 分别取  $w = (x, \lambda^*)$  和  $w = (x^*, \lambda)$ , 就得到 (2.9). 我们称 (2.10) 为单调变分不等式, 是因为  $\theta$  凸, 并且算子  $F$  单调, 它满足

$$(w - \tilde{w})^\top (F(w) - F(\tilde{w})) \geq 0.$$

注意到, (2.10b) 中的  $F(w)$ , 恰有  $(w - \tilde{w})^\top (F(w) - F(\tilde{w})) = 0$ , 所以也是单调的.

## 2.3 可分离结构型凸优化问题等价的变分不等式

对两个算子的可分离线性结构型约束凸优化问题

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y}\}, \quad (2.11)$$

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

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^\top F(w^*) \geq 0, \quad \forall w \in \Omega, \quad (2.12a)$$

其中

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^\top \lambda \\ -B^\top \lambda \\ Ax + By - b \end{pmatrix}, \quad (2.12b)$$

和

$$\theta(u) = \theta_1(x) + \theta_2(y) \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \mathfrak{R}^m. \quad (2.12c)$$

对三个算子的可分离线性结构型约束凸优化问题

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) \mid Ax + By + Cz = b, x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\}, \quad (2.13)$$

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

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^\top F(w^*) \geq 0, \quad \forall w \in \Omega, \quad (2.14a)$$

其中

$$u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^\top \lambda \\ -B^\top \lambda \\ -C^\top \lambda \\ Ax + By + Cz - b \end{pmatrix}, \quad (2.14b)$$

和

$$\theta(u) = \theta_1(x) + \theta_2(y) + \theta_3(z) \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \mathfrak{R}^m. \quad (2.14c)$$

在变分不等式 (2.12) 和 (2.14) 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 (2.11) 和 (2.13) 相对应的单调混合变分不等式形式完全相同, 但 (2.11) 是含两个算子的可分离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) [6] 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 (2.13), 论文 [4] 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 [17, 18].

**注.** 这一节分别导出了凸优化问题 (2.7), (2.11) 和 (2.13) 相对应的单调混合变分不等式 (2.10), (2.12) 和 (2.14). 这些混合变分不等式与变分不等式 (2.1) 没有本质上的差别. 这使得我们有可能将变分不等式 (2.1) 收缩算法的成果移植来求解带线性约束的凸优化问题.

### 3 邻近点算法

邻近点算法 (Proximal Point Algorithms), 简称 PPA 算法, 是求解凸优化和单调变分不等式的一类基本算法 [23, 25]. 我们暂时不谈算法如何实现, 先说一下这个算法的基本性质.

#### 3.1 简单凸优化的邻近点算法

设  $\mathcal{X} \subset \mathfrak{R}^n$  是闭凸集.  $\theta(x) : \mathfrak{R}^n \rightarrow \mathfrak{R}$  是凸函数. 求解凸优化问题

$$\min\{\theta(x) \mid x \in \mathcal{X}\}$$

的邻近点算法：对给定的  $r > 0$  和  $x^k$ , 求得

$$\tilde{x}^k = \operatorname{Argmin}\{\theta(x) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}. \quad (3.1)$$

我们称  $\tilde{x}^k$  是  $k$ -次迭代的邻近点. 根据 (2.6), 它满足

$$\tilde{x}^k \in \mathcal{X}, \quad \theta(x) - \theta(\tilde{x}^k) \geq (x - \tilde{x}^k)^\top r(x^k - \tilde{x}^k), \quad \forall x \in \mathcal{X}.$$

$x^k$  是原问题解的充分必要条件是  $x^k = \tilde{x}^k$ . 用解点  $x^*$  替代上式中任意的  $x \in \mathcal{X}$ , 就有

$$(\tilde{x}^k - x^*)^\top r(x^k - \tilde{x}^k) \geq \theta(\tilde{x}^k) - \theta(x^*) \geq 0.$$

因此

$$(x^k - x^*)^\top (x^k - \tilde{x}^k) \geq \|x^k - \tilde{x}^k\|^2. \quad (3.2)$$

新的迭代点由

$$x^{k+1} = x^k - \gamma(x^k - \tilde{x}^k), \quad \gamma \in (0, 2) \quad (3.3)$$

给出 (通常  $\gamma \in (1, 2)$  收敛会更快些). 根据 (3.2), 由 (3.3) 生成新迭代点的序列  $\{x^k\}$  满足

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \gamma(2 - \gamma)\|x^k - \tilde{x}^k\|^2. \quad (3.4)$$

### 3.2 单调变分不等式邻近点算法的基本性质

对单调变分不等式 (2.1) 的邻近点算法, 我们在欧氏模情况下讨论.

**欧氏模下单调变分不等式的 PPA 算法.** 对给定的  $u^k$  和  $r > 0$ , 求得  $\tilde{u}^k$  使其满足

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^\top F(\tilde{u}^k) \geq (u - \tilde{u}^k)^\top r(u^k - \tilde{u}^k), \quad \forall u \in \Omega. \quad (3.5)$$

$u^k$  是问题 (2.1) 解的充分必要条件是  $u^k = \tilde{u}^k$ . 用  $u^*$  替代 (3.5) 式中任意的  $u \in \Omega$ , 就有

$$(\tilde{u}^k - u^*)^\top (u^k - \tilde{u}^k) \geq \frac{1}{r}(\tilde{u}^k - u^*)^\top F(\tilde{u}^k). \quad (3.6)$$

由  $F$  的单调性和最优点的性质得到

$$(\tilde{u}^k - u^*)^\top F(\tilde{u}^k) \geq (\tilde{u}^k - u^*)^\top F(u^*) \geq 0. \quad (3.7)$$

因此 (3.6) 的右端非负, 进而得到

$$(u^k - u^*)^\top (u^k - \tilde{u}^k) \geq \|u^k - \tilde{u}^k\|^2.$$

新的迭代点由

$$u^{k+1} = u^k - \gamma(u^k - \tilde{u}^k), \quad \gamma \in (0, 2) \quad (3.8)$$

给出. 产生的迭代点具有欧氏模下邻近点算法收敛的关键性质

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - \gamma(2 - \gamma)\|u^k - \tilde{u}^k\|^2. \quad (3.9)$$

找到一个满足 (3.5) 的  $\tilde{u}^k$  并非易事, 但 PPA 算法会给我们构建投影收缩算法提供启示.

### 3.3 单调混合变分不等式邻近点算法的基本性质

对单调（混合）变分不等式 (2.12) 的邻近点算法, 我们分别在欧氏模和  $H$ -模情况下讨论.

**欧氏模下单调(混合)变分不等式的 PPA 算法.** 对给定的  $w^k$  和  $r > 0$ , 求  $\tilde{w}^k$  使其满足

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geq (w - \tilde{w}^k)^\top r(w^k - \tilde{w}^k), \quad \forall w \in \Omega. \quad (3.10)$$

$w^k$  是问题 (2.12) 解的充分必要条件是  $w^k = \tilde{w}^k$ . 用  $w^*$  替代上式中任意的  $w \in \Omega$ , 就有

$$(\tilde{w}^k - w^*)^\top r(w^k - \tilde{w}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^\top F(\tilde{w}^k).$$

由  $F$  的单调性,

$$\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^\top F(\tilde{w}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^\top F(w^*) \geq 0. \quad (3.11)$$

上面最后一个不等式来自于最优点的性质. 进而得到

$$(w^k - w^*)^\top (w^k - \tilde{w}^k) \geq \|w^k - \tilde{w}^k\|^2.$$

由

$$w^{k+1} = w^k - \gamma(w^k - \tilde{w}^k), \quad \gamma \in (0, 2) \quad (3.12)$$

产生的迭代点. 类似与 (3.4), 迭代序列  $\{w^k\}$  有如下的关键收敛性质

$$\|w^{k+1} - w^*\|^2 \leq \|w^k - w^*\|^2 - \gamma(2 - \gamma)\|w^k - \tilde{w}^k\|^2. \quad (3.13)$$

**$H$ -模下单调(混合)变分不等式的 PPA 算法.** 对给定的  $w^k$  和正定矩阵  $H$ , 求  $\tilde{w}^k$  使得

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geq (w - \tilde{w}^k)^\top H(w^k - \tilde{w}^k), \quad \forall w \in \Omega. \quad (3.14)$$

同样,  $w^k$  是问题 (2.12) 解的充分必要条件是  $w^k = \tilde{w}^k$ . 类似与欧氏模下邻近点算法, 由 (3.12) 产生的迭代点具有相应的  $H$ -模下的收敛关键性质

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \gamma(2 - \gamma)\|w^k - \tilde{w}^k\|_H^2. \quad (3.15)$$

序列  $\{w^k\}$  说成是在  $H$ -模下 Fejér 单调的.

## 4 求解变分不等式的投影收缩算法

首先介绍投影的概念. 我们用  $P_\Omega(\cdot)$  表示欧氏范数下在凸集  $\Omega$  上的投影. 也就是说, 对给定的  $v$ ,

$$P_\Omega(v) = \arg \min\{\|u - v\| \mid u \in \Omega\}.$$

特别是, 当  $\Omega = \mathfrak{R}_+^n$  ( $n$ -维空间的非负卦限), 那么  $P_\Omega(v)$  的每个分量为

$$(P_\Omega(v))_j = \begin{cases} v_j, & \text{if } v_j \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

**投影的基本性质.** 我们只列出几个这里需要的、显而易见的性质. 对这些性质证明感兴趣的读者可以参考作者主页上系列讲义 [30] 的第二讲.

**Lemma 4.1.** 设  $\Omega \subset \mathbb{R}^n$  是闭凸集, 则对任意的  $v \in \mathbb{R}^n$ , 有

$$(v - P_{\Omega}(v))^{\top}(u - P_{\Omega}(v)) \leq 0, \quad \forall u \in \Omega. \quad (4.1)$$

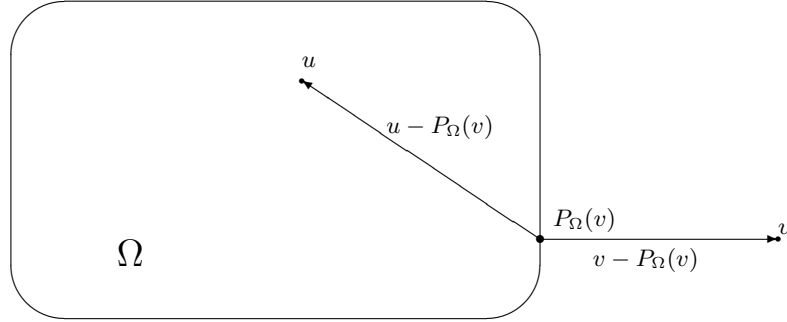


图 3. 不等式(4.1) 的几何解释

变分不等式所等价的投影方程. 利用投影的概念, 求解变分不等式 (2.1) 可以归结为求

$$e(u, \beta) := u - P_{\Omega}[u - \beta F(u)] \quad \text{的一个零点.}$$

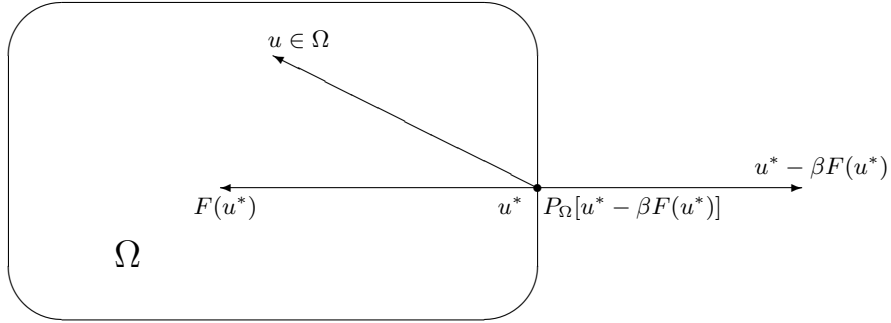


图 4.  $u^*$  是  $\text{VI}(\Omega, F)$  的解等价于  $u^* = P_{\Omega}[u^* - \beta F(u^*)]$  的几何解释

求解单调变分不等式 (2.1) 的投影收缩算法是一种预测-校正方法. 由投影提供预测, 由校正实现收缩. 下面我们分述它们的任务与效用.

#### 4.1 预测由一次投影完成

在求解单调变分不等式 (2.1) 的投影收缩算法中, 对给定的当前点  $u^k$  和  $\beta_k > 0$ , 我们利用投影

$$\tilde{u}^k = P_{\Omega}[u^k - \beta_k F(u^k)], \quad (4.2)$$

生成一个预测点  $\tilde{u}^k$ . 根据变分不等式解的性质,  $u^k = \tilde{u}^k$  是  $u^k \in \Omega^*$  的充分必要条件.

##### 4.1.1 由投影得到的预测点可以看成是一个特定变分不等式的解

在投影的基本性质 (4.1) 中, 令  $v = u^k - \beta_k F(u^k)$ , 根据 (4.2),  $\tilde{u}^k = P_{\Omega}(v)$ , 就有

$$\tilde{u}^k \in \Omega, \quad \{[u^k - \beta_k F(u^k)] - \tilde{u}^k\}^{\top}(u - \tilde{u}^k) \leq 0, \quad \forall u \in \Omega.$$



进而得到

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^\top \beta_k F(u^k) \geq (u - \tilde{u}^k)^\top (u^k - \tilde{u}^k), \quad \forall u \in \Omega.$$

两边都加上  $(u - \tilde{u}^k)^\top \{-\beta_k[F(u^k) - F(\tilde{u}^k)]\}$ , 就有

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^\top \beta_k F(\tilde{u}^k) \geq (u - \tilde{u}^k)^\top d(u^k, \tilde{u}^k), \quad \forall u \in \Omega, \quad (4.3)$$

其中

$$d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k[F(u^k) - F(\tilde{u}^k)]. \quad (4.4)$$

这种我们希望的形式. (4.3) 表明  $\tilde{u}^k$  是一个特定的变分不等式的解. 如果记  $r = 1/\beta_k$ , (4.3) 就有点像 (3.5) 的样子, 只是将  $d(u^k, \tilde{u}^k)$  去替代了 (3.5) 中的  $(u^k - \tilde{u}^k)$ .

**注.** 后面我们会看到, 对由凸优化来的变分不等式 (2.12), 通过求解一些子问题, 产生的预测点  $\tilde{w}^k$  是与 (4.3) 类似的变分不等式

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{w}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geq (w - \tilde{w}^k)^\top H(w^k - \tilde{w}^k), \quad \forall w \in \Omega,$$

或者

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{w}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geq (w - \tilde{w}^k)^\top HM(w^k - \tilde{w}^k), \quad \forall w \in \Omega,$$

的解. 这是我们能求变分不等式的方法移植到结构型凸优化的根本原因.

#### 4.1.2 预测点产生的距离函数上升方向

将 (4.3) 中任意的  $u \in \Omega$  选成  $u^*$ , 就得到

$$(\tilde{u}^k - u^*)^\top d(u^k, \tilde{u}^k) \geq \beta_k(\tilde{u}^k - u^*)^\top F(\tilde{u}^k). \quad (4.5)$$

由 (3.7), (4.5) 的右端非负, 随后得到

$$(u^k - u^*)^\top d(u^k, \tilde{u}^k) \geq (u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k). \quad (4.6)$$

**定义:** (上升方向) 对给定的对称正定矩阵  $H$ , 假如存在一个  $\delta > 0$ , 使得

$$(u^k - u^*)^\top Hd(u^k, \tilde{u}^k) \geq \delta \|u^k - \tilde{u}^k\|^2,$$

则称  $d(u^k, \tilde{u}^k)$  是未知距离函数  $\frac{1}{2}\|u - u^*\|_H^2$  在  $u^k$  处关于  $H$ -模的上升方向.

在投影 (4.2) 中, 我们假设

$$\beta_k \|F(u^k) - F(\tilde{u}^k)\| \leq \nu \|u^k - \tilde{u}^k\|, \quad \nu \in (0, 1). \quad (4.7)$$

由  $d(u^k, \tilde{u}^k)$  的表达式 (4.4) 和上述假设, 根据 Cauchy-Schwarz 不等式, 有

$$(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) \geq (1 - \nu) \|u^k - \tilde{u}^k\|^2. \quad (4.8)$$

不等式 (4.6) 和 (4.8) 告诉我们, 在条件 (4.7) 满足的情况下, 由 (4.4) 定义的  $d(u^k, \tilde{u}^k)$  是欧氏模下的一个上升方向.

## 4.2 校正产生靠近解集更近的新的迭代点

校正是利用距离函数的下降方向(上升方向的反方向), 使得新的迭代点离解集更近一些. 因此, 我们用

$$u^{k+1} = u^k - \alpha d(u^k, \tilde{u}^k) \quad (4.9)$$

产生新的迭代点, 其中  $d(u^k, \tilde{u}^k)$  是由 (4.4) 给出的. 这有点像 (3.8) 的样子, 只是将  $d(u^k, \tilde{u}^k)$  去替代了 (3.8) 中的  $(u^k - \tilde{u}^k)$ . 下面我们讨论如何选取步长  $\alpha$ .

### 4.2.1 计算步长的收缩算法

我们将 (4.9) 中的  $u^{k+1}$  记为  $u^{k+1}(\alpha)$ , 表示新的迭代点依赖于步长  $\alpha$ . 考察与  $\alpha$  相关的距离平方缩短量,

$$\vartheta_k(\alpha) := \|u^k - u^*\|^2 - \|u^{k+1}(\alpha) - u^*\|^2. \quad (4.10)$$

根据定义

$$\begin{aligned} \vartheta_k(\alpha) &= \|u^k - u^*\|^2 - \|u^k - u^* - \alpha d(u^k, \tilde{u}^k)\|^2 \\ &= 2\alpha(u^k - u^*)^\top d(u^k, \tilde{u}^k) - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2. \end{aligned} \quad (4.11)$$

对任意给定的确定解点  $u^*$ , (4.11) 表明  $\vartheta_k(\alpha)$  是  $\alpha$  的一个二次函数. 只是  $u^*$  是未知的, 我们无法直接求  $\vartheta_k(\alpha)$  的极大.

**Theorem 4.1.** 设  $u^{k+1}(\alpha)$  由 (4.9) 生成. 对任意的  $\alpha > 0$ , 由 (4.10) 定义的  $\vartheta_k(\alpha)$  有

$$\vartheta_k(\alpha) \geq q_k(\alpha), \quad (4.12)$$

其中  $d(u^k, \tilde{u}^k)$  由 (4.4) 给出,

$$q_k(\alpha) = 2\alpha(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2. \quad (4.13)$$

**证明.** 这个结论可以从 (4.11) 利用 (4.6) 直接得到.  $\square$

定理 4.1 表明二次函数  $q_k(\alpha)$  是  $\vartheta_k(\alpha)$  的一个下界函数. 使  $q_k(\alpha)$  达到极大的  $\alpha_k^*$  是

$$\alpha_k^* = \operatorname{argmax}\{q_k(\alpha)\} = \frac{(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2}. \quad (4.14)$$

收缩算法的本意是想在每次迭代中极大化二次函数  $\vartheta_k(\alpha)$  (见 (4.11)), 由于它含有未知的  $u^*$ , 我们不得已才极大化它的下界函数  $q_k(\alpha)$ .

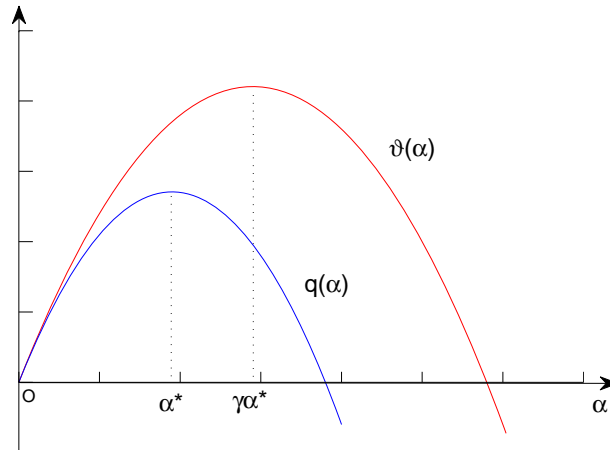


图 5. 取松弛因子  $\gamma \in [1, 2)$  的示意图

因此, 在实际计算中, 我们一般取一个松弛因子  $\gamma \in [1, 2)$ , 令

$$u^{k+1} = u^k - \gamma \alpha_k^* d(u^k, \tilde{u}^k), \quad (4.15)$$

取  $\gamma \in [1, 2)$  的理由可见相应的示意图. 根据 (4.10) 和 (4.12), 由 (4.15) 产生的  $u^{k+1}$  满足

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - q_k(\gamma \alpha_k^*).$$

由  $q_k(\alpha)$  和  $\alpha_k^*$  的定义 (分别见 (4.13) 和 (4.14)), 得到

$$\begin{aligned} q_k(\gamma \alpha_k^*) &= 2\gamma \alpha_k^* (u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) - \gamma^2 (\alpha_k^*)^2 \|d(u^k, \tilde{u}^k)\|^2 \\ &= \gamma(2 - \gamma) \alpha_k^* (u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k). \end{aligned}$$

事实上, 容易证明  $\alpha_k^* > \frac{1}{2}$  (见后面的 (4.19)). 因此, 由校正公式 (4.15) 产生的  $u^{k+1}$  满足

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - \frac{\gamma(2 - \gamma)}{2} (u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k). \quad (4.16)$$

注意到, 上式中的  $(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k)$  是 (4.6) 的右端, (4.8) 说明它大于  $(1 - \nu)\|u^k - \tilde{u}^k\|^2$ .

#### 4.2.2 固定步长的收缩算法

我们将条件 (4.7) 满足时, 用单位步长的迭代公式

$$u^{k+1} = u^k - d(u^k, \tilde{u}^k), \quad (4.17)$$

产生新迭代点的方法, 称为初等方法 (**Primary Method**). 利用 (4.6), 由简单计算可得

$$\begin{aligned} \|u^{k+1} - u^*\|^2 &= \|(u^k - u^*) - d(u^k, \tilde{u}^k)\|^2 \\ &= \|u^k - u^*\|^2 - 2(u^k - u^*)^\top d(u^k, \tilde{u}^k) + \|d(u^k, \tilde{u}^k)\|^2 \\ &\leq \|u^k - u^*\|^2 - (2(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) - \|d(u^k, \tilde{u}^k)\|^2). \end{aligned} \quad (4.18)$$

利用 (4.4) 和 (4.7), 可以得到

$$\begin{aligned} &2(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) - \|d(u^k, \tilde{u}^k)\|^2 \\ &= d(u^k, \tilde{u}^k)^\top \{2(u^k - \tilde{u}^k) - d(u^k, \tilde{u}^k)\} \\ &= \{(u^k - \tilde{u}^k) - \beta_k [F(u^k) - F(\tilde{u}^k)]\}^\top \{(u^k - \tilde{u}^k) + \beta_k [F(u^k) - F(\tilde{u}^k)]\} \\ &= \|u^k - \tilde{u}^k\|^2 - \beta_k^2 \| [F(u^k) - F(\tilde{u}^k)] \|^2 \\ &\geq (1 - \nu^2) \|u^k - \tilde{u}^k\|^2. \end{aligned} \quad (4.19)$$

代入 (4.18), 说明由 (4.17) 产生的序列  $\{u^k\}$  满足

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - (1 - \nu^2) \|u^k - \tilde{u}^k\|^2. \quad (4.20)$$

不等式 (4.16) 和 (4.20) 说明, 用计算步长和固定步长的算法产生的序列  $\{u^k\}$  都是收缩和有界的. 利用这些关键不等式, 容易证明收敛定理. 虽然这些结论都在假设 (4.7) 满足时才成立, 但这个假设可以当  $F$  是 Lipschitz 连续时在 (4.2) 中通过用 Armijo 法则选取适当的  $\beta_k$  就可以实现.

### 4.3 奇妙的孪生方向和相等步长

投影收缩算法给我们享受的美妙之处是能在校正公式中取孪生方向之一作为寻查方向而取相同的步长. 通过投影提供的预测点  $\tilde{u}^k$  满足变分不等式 (4.3).

定义: (孪生方向) 在变分不等式 (4.3) 中,

$$\tilde{u}^k \in \Omega, (u - \tilde{u}^k)^\top \beta_k F(\tilde{u}^k) \geq (u - \tilde{u}^k)^\top d(u^k, \tilde{u}^k), \quad \forall u \in \Omega,$$

其中  $d(u^k, \tilde{u}^k)$  由 (4.4) 给出. 分处不等式两端的

$$\beta_k F(\tilde{u}^k) \quad \text{和} \quad d(u^k, \tilde{u}^k),$$

我们称之为为一对孪生方向.

§4.2 的校正公式 (4.9) 中采用  $d(u^k, \tilde{u}^k)$  为方向, 这里我们用  $\beta_k F(\tilde{u}^k)$  替代它, 并附加一次投影. 同样用

$$u_{II}^{k+1}(\alpha) = P_\Omega[u^k - \alpha \beta_k F(\tilde{u}^k)], \quad (4.21)$$

为依赖于步长  $\alpha$  的新的迭代点. 为了区别与由 (4.9) 生成的  $u^{k+1}$ , 我们在 (4.21) 中用带下标的  $u_{II}^{k+1}(\alpha)$  表示. 对任意给定的  $u^* \in \Omega^*$ , 我们将

$$\zeta_k(\alpha) = \|u^k - u^*\|^2 - \|u_{II}^{k+1}(\alpha) - u^*\|^2 \quad (4.22)$$

看成是本次迭代的进步量, 它是步长  $\alpha$  的函数. 我们不能直接极大化  $\zeta_k(\alpha)$ , 因为它含有我们要求的  $u^*$ . 下面的定理说明, 对同样的  $\alpha$ ,  $\zeta_k(\alpha)$  ‘优于’ (4.12) 中的  $\vartheta_k(\alpha)$ .

**Theorem 4.2.** 设  $u^{k+1}(\alpha)$  由 (4.21) 生成. 对任意的  $\alpha > 0$ , 由 (4.22) 定义的  $\zeta_k(\alpha)$  有

$$\zeta_k(\alpha) \geq q_k(\alpha) + \|(u^k - u_{II}^{k+1}(\alpha)) - \alpha d(u^k, \tilde{u}^k)\|^2, \quad (4.23)$$

其中  $d(u^k, \tilde{u}^k)$  和  $q_k(\alpha)$  分别由 (4.4) 和 (4.13) 给出.

**证明.** 首先, 因为  $u^{k+1}(\alpha) = P_\Omega[u - \alpha \beta_k F(\tilde{u}^k)]$  和  $u^* \in \Omega$ , 根据投影的性质和余弦定理, 有

$$\|u^{k+1}(\alpha) - u^*\|^2 \leq \|u^k - \alpha \beta_k F(\tilde{u}^k) - u^*\|^2 - \|u^k - \alpha \beta_k F(\tilde{u}^k) - u_{II}^{k+1}(\alpha)\|^2. \quad (4.24)$$

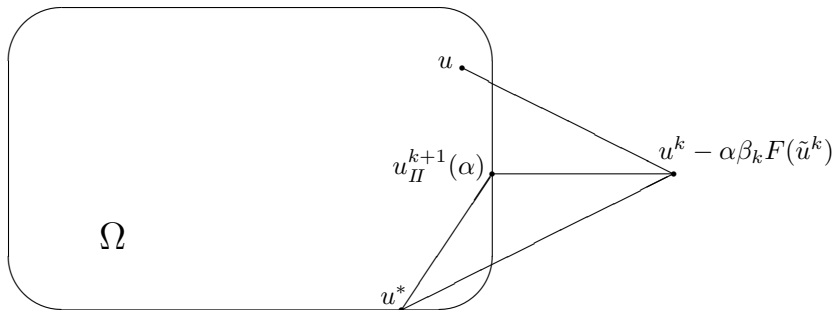


图 6. 不等式 (4.24) 的几何解释

因此, 利用  $\zeta_k(\alpha)$  的定义(见 (4.22)), 我们有

$$\begin{aligned}\zeta_k(\alpha) &\geq \|u^k - u^*\|^2 - \|(u^k - u^*) - \alpha\beta_k F(\tilde{u}^k)\|^2 + \|(u^k - u_H^{k+1}(\alpha)) - \alpha\beta_k F(\tilde{u}^k)\|^2 \\ &= 2\alpha(u^k - u^*)^\top \beta_k F(\tilde{u}^k) + 2\alpha(u_H^{k+1}(\alpha) - u^k)^\top \beta_k F(\tilde{u}^k) + \|u^k - u_H^{k+1}(\alpha)\|^2 \\ &= \|u^k - u_H^{k+1}(\alpha)\|^2 + 2\alpha(u_H^{k+1}(\alpha) - u^*)^\top \beta_k F(\tilde{u}^k).\end{aligned}\quad (4.25)$$

将 (4.25) 中右端的最后一项  $(u_H^{k+1}(\alpha) - u^*)^\top \beta_k F(\tilde{u}^k)$  分解成

$$(u_H^{k+1}(\alpha) - u^*)^\top \beta_k F(\tilde{u}^k) = (u_H^{k+1}(\alpha) - \tilde{u}^k)^\top \beta_k F(\tilde{u}^k) + (\tilde{u}^k - u^*)^\top \beta_k F(\tilde{u}^k),$$

利用 (3.7), 上式右端的最后一部分非负. 代入 (4.25) 的右端, 进一步得到

$$\zeta_k(\alpha) = \|u^k - u_H^{k+1}(\alpha)\|^2 + 2\alpha(u_H^{k+1}(\alpha) - \tilde{u}^k)^\top \beta_k F(\tilde{u}^k).\quad (4.26)$$

因为  $u_H^{k+1}(\alpha) \in \Omega$ , 用它替代 (4.3) 中的任意  $u \in \Omega$ , 得到

$$(u_H^{k+1}(\alpha) - \tilde{u}^k)^\top \beta_k F(\tilde{u}^k) \geq (u_H^{k+1}(\alpha) - \tilde{u}^k)^\top d(u^k, \tilde{u}^k).\quad (4.27)$$

将它们代入 (4.26) 的右端, 就有

$$\zeta_k(\alpha) \geq \|u^k - u_H^{k+1}(\alpha)\|^2 + 2\alpha(u_H^{k+1}(\alpha) - \tilde{u}^k)^\top d(u^k, \tilde{u}^k).\quad (4.28)$$

对上式右端, 利用  $q_k(\alpha)$  的形式 (见(4.13)), 就化成

$$\begin{aligned}&\|u^k - u_H^{k+1}(\alpha)\|^2 + 2\alpha(u_H^{k+1}(\alpha) - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) \\ &= \|u^k - u_H^{k+1}(\alpha)\|^2 + 2\alpha(u_H^{k+1}(\alpha) - u^k)^\top d(u^k, \tilde{u}^k) + 2\alpha(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) \\ &= \|(u^k - u_H^{k+1}(\alpha)) - \alpha d(u^k, \tilde{u}^k)\|^2 - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2 + 2\alpha(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) \\ &= \|(u^k - u_H^{k+1}(\alpha)) - \alpha d(u^k, \tilde{u}^k)\|^2 + q_k(\alpha).\end{aligned}$$

这样就完成了定理结论 (4.23) 的证明.  $\square$

定理 4.2 说明,  $q_k(\alpha)$  也是  $\zeta_k(\alpha)$  的下界. 对同样的  $\alpha$ ,  $\zeta_k(\alpha)$  优于  $\vartheta_k(\alpha)$ . 在实际计算中, 采用校正公式

$$(\text{收缩算法-1}) \quad u_I^{k+1} = u^k - \gamma\alpha_k^* d(u^k, \tilde{u}^k)\quad (4.29)$$

或者

$$(\text{收缩算法-2}) \quad u_H^{k+1} = P_\Omega[u^k - \gamma\alpha_k^* \beta_k F(\tilde{u}^k)]\quad (4.30)$$

产生新的迭代点  $u^{k+1}$ , 其中的  $\alpha_k^*$  都由 (4.14) 给出. 采用校正公式 (4.29), 它的好处是生成  $u^{k+1}$  不用再做投影. 实际问题中, 到  $\Omega$  上的投影代价往往不高 (例如  $\Omega$  常常是一个正卦限或者框形), 因此常采用校正公式 (4.30). 这方面的理由我们在论文 [22] 中有更详细的说明.

对投影收缩算法, (4.3) 式表明  $\tilde{u}^k$  是一个特定的变分不等式的解, 这是工作基础. 定理 4.1 的证明只是用到 (4.6), 而它是由投影的性质 (4.1) 得到的. 综观定理 4.2 的证明, (4.27) 是用  $u_H^{k+1}(\alpha)$  替代 (4.3) 中的任意  $u \in \Omega$  得来的. 此外, 我们需要的仅仅是

- 由余弦定理得到的性质 (4.24)—钝角三角形长边的平方大于两条短边的平方和.

这些我们在中学就已经学过, 大学只是告诉我们如何表述更合乎常规要求. 根据 (4.3) 提供的孪生方向, 在采用相同步长的两个方法 (4.29) 和 (4.30) 中, 如果用记号  $u_I^{k+1}(\alpha) = u^k - \alpha d(u^k, \tilde{u}^k)$ , (4.12) 和 (4.23) 就可以分别改写成

$$\vartheta_k(\alpha) = \|u^k - u^*\|^2 - \|u_I^{k+1}(\alpha) - u^*\|^2 \geq q_k(\alpha),$$

和

$$\zeta_k(\alpha) = \|u^k - u^*\|^2 - \|u_{II}^{k+1}(\alpha) - u^*\|^2 \geq q_k(\alpha) + \|u_I^{k+1}(\alpha) - u_{II}^{k+1}(\alpha)\|^2.$$

其中  $q_k(\alpha)$  由 (4.13) 给出. 二十多年前的这个发现, 是至今我们还自以为有点漂亮的结果.

**Theorem 4.3.** 设  $VI(\Omega, F)$  的解集  $\Omega^*$  非空, 则由本节提到的投影收缩算法产生的序列  $\{u^k\}$  收敛到  $VI(\Omega, F)$  的某个解点  $u^* \in \Omega^*$ .

关于变分不等式的投影收缩算法的主要论文, 线性变分不等式方面的是 [8, 9], 非线性变分不等式方面的是 [10, 11, 14, 15, 21, 24]. 处理不同类型问题的投影收缩算法之间的内在联系可参阅 [27]. 对这一节的内容想做更进一步的了解, 参考作者主页上系列讲义《凸优化和单调变分不等式的收缩算法》[30] 的第二与第三讲就更加省时. 此外, 10 年前在结构型变分不等式的交替方向法 [12, 13] 方面的工作, 对近年用交替方向法 (Alternating Direction Method of Multipliers) 求解分离结构的凸优化问题有实质性的帮助 [16, 17, 18].

## 5 线性约束的凸优化问题 — 定制的邻近点算法

在 §4 中讨论的求解变分不等式 (2.1) 的投影收缩算法中, 预测点  $\tilde{u}^k$  是通过投影 (4.2) 得到的.  $\tilde{u}^k$  是一个特定的变分不等式 (4.3) 的解.

对线性约束凸优化问题, 我们得到了与其等价的混合单调变分不等式 (2.10) 和 (2.12). 用变分不等式作为框架, 通过求解一些 (刻意构造的) 简单子问题生成预测点  $\tilde{w}^k$ , 使其是变分不等式

$$\tilde{w}^k \in \Omega, \quad \theta(w) - \theta(\tilde{w}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geq (v - \tilde{v}^k)^\top H(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \quad (5.1)$$

的解, 其中右端的矩阵  $H$  是因需而凑成的对称正定矩阵. 前面已经介绍过,  $u$  对应的是凸优化问题中的原始变量,  $w$  是原始变量加对偶变量. 向量  $v$ , 根据算法要求, 可以是向量  $w$  本身, 也可以是  $w$  的一个子向量. 有的求解可分离变量凸优化的方法中, 迭代只要有了  $w$  的部分分量就可以开始, 一次迭代也以得到这些新的部分变量而完成. 我们把这些变量记成  $v$ , 称为迭代方法中的核心变量. 预测点中  $\tilde{w}^k \setminus \tilde{v}^k$  那部分分量是根据  $v^k$  计算得到的, 因此, 把  $w \setminus v$  的那部分分量称为中间变量. 将  $\Omega^*$  中对应于核心变量  $v$  的部分记为  $\mathcal{V}^*$ .

注意到 (5.1) 的右端是一个类似与 (4.3) 的式子. 有了它, 我们就可以像 §4 中那样构造收缩算法. 本文只以引言中的第一类问题为例, 更多的材料可参考作者主页上的报告 [29].

### 5.1 变分不等式的解作为预测点

对 §2.2 中提到的凸优化问题 (2.7), 最优解  $x^*$  和相应的最优 Lagrange 乘子  $\lambda^*$  一起,  $(x^*, \lambda^*)$  是 Lagrange 函数

$$L(x, \lambda) = \theta(x) - \lambda^\top (Ax - b),$$

在  $\Omega = \mathcal{X} \times \Lambda$  上的鞍点. 对给定的  $(x^k, \lambda^k)$ , 通过

$$\tilde{x}^k = \arg \min \{ L(x, \lambda^k) + \frac{r}{2} \|x - x^k\|^2 \mid x \in \mathcal{X} \} \quad (5.2a)$$

和

$$\tilde{\lambda}^k = \arg \max \{ L([2\tilde{x}^k - x^k], \lambda) - \frac{s}{2} \|\lambda - \lambda^k\|^2 \mid \lambda \in \Lambda \}, \quad (5.2b)$$

得到预测点  $(\tilde{x}^k, \tilde{\lambda}^k)$ . 根据 (2.6), 优化问题 (5.2) 的最优性条件是

$$\begin{cases} \tilde{x}^k \in \mathcal{X}, & \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^\top (-A^\top \lambda^k + r(\tilde{x}^k - x^k)) \geq 0, \quad \forall x \in \mathcal{X}, \\ \tilde{\lambda}^k \in \Lambda, & (\lambda - \tilde{\lambda}^k)^\top (A(2\tilde{x}^k - x^k) - b + s(\tilde{\lambda}^k - \lambda^k)) \geq 0, \quad \forall \lambda \in \Lambda. \end{cases}$$

上式可以写成我们需要的形式

$$\theta(x) - \theta(\tilde{x}^k) + \begin{pmatrix} x - \tilde{x}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^\top \left\{ \begin{pmatrix} -A^\top \tilde{\lambda}^k \\ A\tilde{x}^k - b \end{pmatrix} + \begin{pmatrix} r(\tilde{x}^k - x^k) + A^\top(\tilde{\lambda}^k - \lambda^k) \\ A(\tilde{x}^k - x^k) + s(\tilde{\lambda}^k - \lambda^k) \end{pmatrix} \right\} \geq 0, \quad \forall w \in \Omega.$$

利用 (2.10) 的形式, 上式说明由 (5.2) 产生的  $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k)$  是变分不等式

$$\tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geq (w - \tilde{w}^k)^\top H(w^k - \tilde{w}^k), \quad \forall w \in \Omega. \quad (5.3)$$

的解, 这相当于 (5.1) 中取  $u = x, v = w$  的一种具体形式. 其中对称矩阵

$$H = \begin{pmatrix} rI_n & A^\top \\ A & sI_m \end{pmatrix},$$

当  $rs > \|A^\top A\|$  时,  $H$  是正定的. 将 (5.1) 中任意的  $w \in \Omega$  选成  $w^*$ , 就得到

$$(\tilde{v}^k - v^*)^\top H(v^k - \tilde{v}^k) \geq \theta(\tilde{w}^k) - \theta(w^*) + (\tilde{w}^k - w^*)^\top F(\tilde{w}^k), \quad (5.4)$$

利用 (3.11), (5.4) 的右端非负. 由 (5.4) 进而得到

$$(v^k - v^*)^\top H(v^k - \tilde{v}^k) \geq (v^k - \tilde{v}^k)^\top H(v^k - \tilde{v}^k) = \|v^k - \tilde{v}^k\|_H^2. \quad (5.5)$$

因此  $(v^k - \tilde{v}^k)$  是  $H$ -模下的上升方向. 在定制的 PPA 算法中, 预测点满足变分不等式 (5.1). 我们特别要注意到 (5.1) 中的矩阵  $H$  是对称正定的. 为了达到这个目的, 我们往往用构造性的方法去设计迭代过程中的子问题. 在上面的例子中, 我们刻意构造了子问题 (5.2b), 使得 (5.3) 中的矩阵对称, 进而通过适当选取  $r$  和  $s$  使之正定.

## 5.2 校正产生 $H$ -模下靠近解集更近的新迭代点

根据 §4.1.2 中关于上升方向的定义, (5.5) 表示由 (5.1) 产生的预测点  $\tilde{v}^k$  提供了  $H$ -模下的一个上升方向  $d(v^k, \tilde{v}^k) = v^k - \tilde{v}^k$ . 利用这个方向, 新的迭代点由

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

生成, 其中步长  $\alpha$  待定. 利用 (5.5), 类似 §4.2.1 中的分析得到

$$\begin{aligned}
& \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\
&= \|v^k - v^*\|_H^2 - \|v^k - v^* - \alpha(v^k - \tilde{v}^k)\|_H^2 \\
&= 2\alpha(v^k - v^*)^\top H(v^k - \tilde{v}^k) - \alpha^2\|v^k - \tilde{v}^k\|_H^2 \\
&\geq (2\alpha - \alpha^2)\|v^k - \tilde{v}^k\|_H^2.
\end{aligned}$$

上式右端的关于  $\alpha$  的二次函数在  $\alpha_k^* = 1$  达到最大. 由于  $\alpha_k^* \equiv 1$ , 新的迭代点可以通过

$$v^{k+1} = v^k - \gamma(v^k - \tilde{v}^k), \quad \gamma \in (0, 2) \quad (5.6)$$

给出. 通常, 在  $[1, 2)$  取松弛因子  $\gamma$ , 例如  $\gamma = 1.5$ . 我们把这类方法称为定制的邻近点算法 (Customized PPA). 如果取  $\gamma \equiv 1$ , 所述方法就是 Chambolle 和 Pock 的 C-P 方法 [3]. 定制的 PPA 提供了一个宽大的框架, 更多的关于定制的 PPA 算法的例子可参阅论文 [7, 22]. 定制的邻近点算法和 C-P 方法计算效果比较也可以在作者系列讲义 [30] 的第四讲中可以找到.

## 6 线性约束的凸优化问题 — 基于变分不等式的预测-校正方法

在 §5 中, 对源自凸优化的单调混合变分不等式 (2.10) 和 (2.12), (5.1) 中的预测点  $\tilde{w}^k$  是按需 (为了凑成一个对称正定的矩阵  $H$ ) 通过刻意构造特殊的子问题生成的. 有些已有的 (不一定收敛的) 方法, 把它们通过求解一些简单的子问题生成的点看做预测点  $\tilde{w}^k$ , 恰是变分不等式

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geq (v - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \quad (6.1)$$

的解, 其中  $Q$  并不对称, 但是  $Q^\top + Q$  是正定矩阵. 上面 (6.1) 中的向量  $v$ , 根据算法要求, 可以是  $w$  的一个子向量, 也可以是  $w$  本身. 这里给出的例子中,  $v = w$ . 在 [29] 中给出的许多例子,  $v$  是  $w$  的真子向量.

### 6.1 变分不等式的解作为预测点

在基于变分不等式的预测-校正方法中, 预测点满足变分不等式 (6.1). 注意到 (6.1) 中的矩阵  $Q$  是非对称的. 这种预测点一般是通过已知的方法定义的. 同样以求解 §2.2 中提到的凸优化问题 (2.7) 为例. 对给定的  $(x^k, \lambda^k)$ , 我们将通过

$$\tilde{x}^k = \arg \min \{L(x, \lambda^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\} \quad (6.2a)$$

和

$$\tilde{\lambda}^k = \arg \max \{L(\tilde{x}^k, \lambda) - \frac{s}{2}\|\lambda - \lambda^k\|^2 \mid \lambda \in \Lambda\}, \quad (6.2b)$$

得到的  $(\tilde{x}^k, \tilde{\lambda}^k)$  称为预测点. 直接将  $(\tilde{x}^k, \tilde{\lambda}^k)$  作为新的迭代点, 令  $(x^{k+1}, \lambda^{k+1}) = (\tilde{x}^k, \tilde{\lambda}^k)$ , 就得到 [26] 中的“原始-对偶混合梯度法” (Primal-Dual Hybrid Gradient Algorithm, 简



称 PDHG), 这个方法在没有额外条件时并不一定收敛. 根据 (2.6), 优化问题 (6.2) 的最优性条件是

$$\begin{cases} \tilde{x}^k \in \mathcal{X}, & \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^\top (-A^\top \lambda^k + r(\tilde{x}^k - x^k)) \geq 0, \quad \forall x \in \mathcal{X}, \\ \tilde{\lambda}^k \in \Lambda, & (\lambda - \tilde{\lambda}^k)^\top ((A\tilde{x}^k - b) + s(\tilde{\lambda}^k - \lambda^k)) \geq 0, \quad \forall \lambda \in \Lambda. \end{cases}$$

上式可以写成我们需要的形式

$$\theta(x) - \theta(\tilde{x}^k) + \begin{pmatrix} x - \tilde{x}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^\top \left\{ \begin{pmatrix} -A^\top \tilde{\lambda}^k \\ A\tilde{x}^k - b \end{pmatrix} + \begin{pmatrix} r(\tilde{x}^k - x^k) + A^\top(\tilde{\lambda}^k - \lambda^k) \\ s(\tilde{\lambda}^k - \lambda^k) \end{pmatrix} \right\} \geq 0, \quad \forall w \in \Omega.$$

利用 (2.10) 的形式, 上式说明由 (6.2) 产生的  $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k)$  是变分不等式

$$\tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geq (w - \tilde{w}^k)^\top Q(w^k - \tilde{w}^k), \quad \forall w \in \Omega. \quad (6.3)$$

的解, 这是 (6.1) 中取  $u = x$ ,  $v = w$  的一种具体形式, 其中矩阵

$$Q = \begin{pmatrix} rI_n & A^\top \\ 0 & sI_m \end{pmatrix} \quad (6.4)$$

是非对称的. 当  $rs > \frac{1}{4}\|A^\top A\|$  时,  $Q^\top + Q$  就是正定的, 因为

$$\begin{aligned} w^\top Qw &= r\|x\|^2 + s\|\lambda\|^2 + \lambda^\top Ax \geq r\|x\|^2 + s\|\lambda\|^2 - \frac{1}{2}(2s\|\lambda\|^2 + \frac{1}{2s}\|Ax\|^2) \\ &\geq \left(r - \frac{\|A^\top A\|}{4s}\right)\|x\|^2. \end{aligned}$$

## 6.2 校正产生 $H$ -模下靠近解集更近的新迭代点

将 (6.1) 中任意的  $w \in \Omega$  选成  $w^*$ , 就分别得到

$$(\tilde{v}^k - v^*)^\top Q(v^k - \tilde{v}^k) \geq \theta(\tilde{w}^k) - \theta(w^*) + (\tilde{w}^k - w^*)^\top F(\tilde{w}^k). \quad (6.5)$$

利用 (3.11), (6.5) 的右端非负. 因此

$$(v^k - v^*)^\top Q(v^k - \tilde{v}^k) \geq (v^k - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k). \quad (6.6)$$

由于  $2v^\top Qv = v^\top (Q^\top + Q)v$ , 根据 §4.1.2 中关于上升方向的定义, 当矩阵  $Q^\top + Q \succ 0$  时, 上式表示由 (6.1) 产生的预测点  $\tilde{v}^k$  提供了欧氏模下的一个上升方向  $d(v^k, \tilde{v}^k) = Q(v^k - \tilde{v}^k)$ .

假设矩阵  $Q$  有分解式

$$Q = HM, \quad \text{使得} \quad H \succ 0, \quad (6.7)$$

那么 (6.6) 可以写成

$$(v^k - v^*)^\top HM(v^k - \tilde{v}^k) \geq (v^k - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k). \quad (6.8)$$

当  $Q^\top + Q \succ 0$  时,  $M(v^k - \tilde{v}^k)$  就是距离函数在  $H$ -模下的一个上升方向.

我们按 (6.7) 的要求, 做  $Q = HM$  的分解. 例如, 对 (6.4) 中的  $Q$ ,

$$H = \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix}, \quad M = \begin{pmatrix} I_n & \frac{1}{r}A^\top \\ 0 & I_m \end{pmatrix}$$

就是一种满足要求的分解. 然后也可以像投影收缩算法中那样, 通过两种不同的校正方式给出新的迭代点.

**每次迭代都用计算步长的收缩算法.** 由 (6.1) 产生的预测点  $\tilde{v}^k$ , 我们得到了  $H$ -模下的上升方向  $M(v^k - \tilde{v}^k)$  (见(6.8)). 利用这个方向, 新的迭代点由

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

生成, 其中步长  $\alpha$  待定. 利用 (6.8), 类似 §4.2.1 中的分析得到

$$\begin{aligned} & \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ &= \|v^k - v^*\|_H^2 - \|(v^k - v^*) - \alpha M(v^k - \tilde{v}^k)\|_H^2 \\ &= 2\alpha(v^k - v^*)^\top HM(v^k - \tilde{v}^k) - \alpha^2 \|M(v^k - \tilde{v}^k)\|_H^2 \\ &\geq 2\alpha(v^k - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k) - \alpha^2 \|M(v^k - \tilde{v}^k)\|_H^2. \end{aligned} \quad (6.9)$$

上式右端的关于  $\alpha$  的二次函数在

$$\alpha_k^* = \frac{(v^k - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k)}{\|M(v^k - \tilde{v}^k)\|_H^2} \quad (6.10)$$

达到最大. 新的迭代点可以通过

$$v^{k+1} = v^k - \gamma \alpha_k^* M(v^k - \tilde{v}^k), \quad \gamma \in (0, 2) \quad (6.11)$$

给出. 通常取  $\gamma \in [1, 2)$ . 将 (6.10) 和 (6.11) 代入 (6.9) 的右端, 得到序列  $\{v^k\}$  满足

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \gamma(2 - \gamma)\alpha_k^*(v^k - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k). \quad (6.12)$$

**每次迭代都用固定步长的收缩算法.** 我们要求  $rs > \|A^\top A\|$ , 用校正公式

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \quad (6.13)$$

生成新的迭代点. 由于  $Q = HM$ , 这时 (6.1) 就可以改写成

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geq (v - \tilde{v}^k)^\top H(v^k - v^{k+1}), \quad \forall w \in \Omega, \quad (6.14)$$

对 (6.14) 式右端的  $(v - \tilde{v}^k)^\top H(v^k - v^{k+1})$ , 用恒等式

$$(a - b)^\top H(c - d) = \frac{1}{2}(\|a - d\|_H^2 - \|a - c\|_H^2) + \frac{1}{2}(\|c - b\|_H^2 - \|d - b\|_H^2), \quad (6.15)$$

进行处理, 就有

$$\begin{aligned} & \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \\ & \geq \frac{1}{2}(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) + \frac{1}{2}(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2), \quad \forall w \in \Omega. \end{aligned} \quad (6.16)$$

对 (6.16) 右端的最后一部分, 利用 (6.13), 我们有

$$\begin{aligned} & \frac{1}{2}(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2) = \frac{1}{2}(\|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2) \\ & = \frac{1}{2}((v^k - \tilde{v}^k)^\top (2HM - M^\top HM)(v^k - \tilde{v}^k)) = \frac{1}{2}\|v^k - \tilde{v}^k\|_G^2, \end{aligned}$$

其中

$$G = Q^\top + Q - M^\top HM.$$

将上述结果代入 (6.16) 中, 就有

$$\begin{aligned} & \tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \\ & \geq \frac{1}{2}(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) + \frac{1}{2}\|v^k - \tilde{v}^k\|_G^2, \quad \forall w \in \Omega. \end{aligned} \quad (6.17)$$

再用  $w^*$  替代 (6.17) 式中任意的  $w \in \Omega$  并利用单调性, 就会得到

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \quad (6.18)$$

由于  $rs > \|A^\top A\|$ , 因此

$$\begin{aligned} G & = Q^\top + Q - M^\top HM \\ & = \begin{pmatrix} 2rI_n & A^\top \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} I_n & 0 \\ \frac{1}{r}A & I_m \end{pmatrix} \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix} \begin{pmatrix} I_n & \frac{1}{r}A^\top \\ 0 & I_m \end{pmatrix} \\ & = \begin{pmatrix} rI_n & 0 \\ 0 & sI_m - \frac{1}{r}AA^\top \end{pmatrix} \succ 0. \end{aligned}$$

因为  $G$  正定, 序列  $\{w^k\}$  是在  $H$ -模下收缩的. (6.18) 是方法收敛的关键不等式.

**注.** 这一节关于固定步长方法的收敛性, 主要根据是 (6.14), 它是由生成预测点的变分不等式 (6.1) 和校正公式 (6.13) 得来的. 证明的工具是 (6.15), 是一个对中学生来说也毫无困难的恒等式. 利用 (6.17), 我们可以进一步得到

$$\tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geq \frac{1}{2}(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2), \quad \forall w \in \Omega.$$

有了上式, 通过对  $k = 0, 1, \dots$  累加, 算法在遍历意义下的收敛速率 [19] 就能马上得到了.

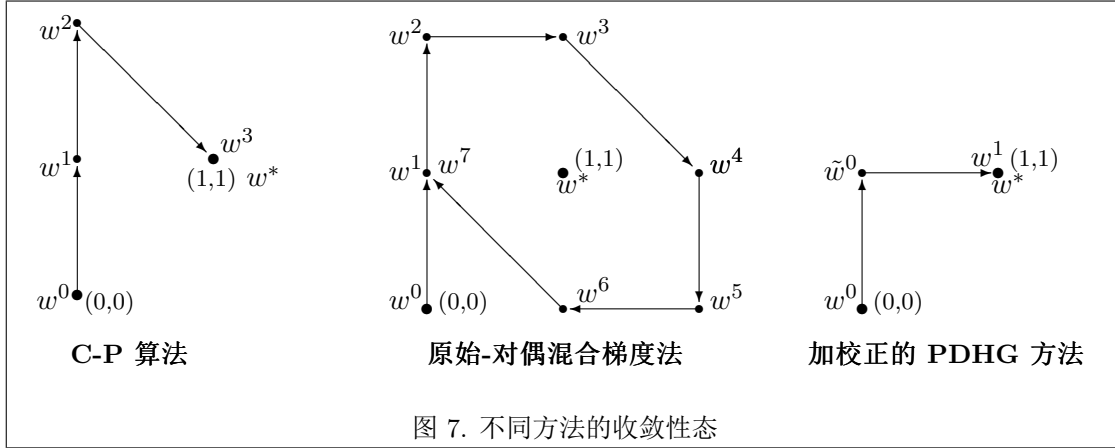
## 7 结论与展望

这篇文章介绍了凸优化的分裂收缩算法与变分不等式投影收缩算法的关系. 对投影收缩算法的理解深刻一些, 处理可分离结构凸优化的手段就多一些. 基于变分不等式的框架, 除了可以用 §5 中定制的邻近点算法, 也可以用 §6 中的预测-校正方法. 对于 §6 中的预测-校正方法, 从理论上讲, 校正不是可有可无的. 如果略去了校正, 就变成了 [26] 中的 PDHG 方法. 用 PDHG 算法有一些成功的算例, 但理论上并不保证收敛. 下面的例子可以说明.

**例** 我们以线性规划  $\min\{x \mid x = 1, x \geq 0\}$  作为约束凸优化问题 (2.7) 的例子.

- 相应的变分不等式 (2.10) 中, 变量  $w = (x, \lambda)$ , 鞍点是  $w^* = (x^*, \lambda^*) = (1, 1)$ .
- 在子问题中要用的参数, 我们取  $r = s = 1$ . 初始点都取  $w^0 = (x^0, \lambda^0) = (0, 0)$ .
- C-P (Chambolle and Pock) 算法 [3] 是在定制邻近点算法的校正(5.6) 中取  $\gamma = 1$ , 等于将(5.2)的输出直接作为新的迭代点;
- 原始-对偶混合梯度法(Zhu and Chan [26] 的PDHG), 总是直接将 (6.2) 的输出  $\tilde{w}^k$  作为新的迭代点  $w^{k+1}$ , 方法并不保证收敛;
- 加校正的原始-对偶混合梯度法就是 §6 中的预测-校正方法, 它是将 (6.2) 的输出  $\tilde{w}^k$  作为预测点, 再用 (6.13) 校正生成新的迭代点  $w^{k+1}$ , 这个例子表明这样做比 C-P 算法收敛更快一些.

我们将不同方法的收敛性态在图 7 中表明.



我们把要讨论的优化问题都归结为 (2.14) 这样的变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^\top F(w^*) \geq 0, \quad \forall w \in \Omega.$$

求解这样的问题, 我们采用如下的预测-校正方法框架:

[预测]. 对给定的  $v^k$ , 求得预测点  $\tilde{w}^k \in \Omega$ , 使其满足

$$\theta(u) - \theta(\tilde{w}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geq (v - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \quad (7.1a)$$

其中矩阵  $Q$  不一定对称, 但是  $Q^\top + Q$  正定.

[校正]. 新的核心变量  $v^{k+1}$  由

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k), \quad (7.1b)$$

给出, 其中  $\alpha > 0$  为常数.

收敛性条件为

对给定的矩阵  $Q$  和  $M$ , 有一个正定矩阵  $H$ , 使得

$$HM = Q. \quad (7.2a)$$

并且, 矩阵

$$G = Q^T + Q - \alpha M^T H M \succeq 0. \quad (7.2b)$$

在这个框架和条件下, 构造算法和证明收敛性都相当容易[16]. 对三个算子以上的可分离变量问题改造的 ADMM 方法 [17, 18], 收敛性证明用这个框架 [28] 就更简洁容易. 掌握了求解变分不等式的投影收缩算法中的优秀成果, 对求解信息科学中稀疏问题会有很大帮助. 譬如说, §4.3 中, 孪生方向和相同步长, 这类结果如何移植到凸优化的分裂算法中. 对不同结构的凸优化问题, 如何合理构造产生预测点的方法, 使得预测点是 (5.1) 和 (6.1) 这样的子变分不等式的解, 并且使因此而构造的算法效率更高.

需要说明的只是, 这里介绍的都是一阶算法. 定制的邻近点算法是一个容易实现的邻近点算法, 但它还是属于邻近点算法, 通常邻近点算法在处理问题时的弱点这里有时也会显现. 此外, 参数如何选择, 有些还与所处理的问题量纲有关, 需要对具体问题具体讨论. 优化界的共识是, 没有一个方法是对所有的问题都是最好的. 这里讨论的一阶分裂算法, 也被他人在求解一些实际问题中采用, 例如 [2, 3, 5].

了解我们的工作, 这是最初级的读本. 中级读本 是作者的报告 [29], 再详细一点的是作者的讲义 [30]. 我参与的最新工作, 一般都可以在 Optimization Online 以及香港浸会大学数学系袁晓明博士 (Xiaoming Yuan) 的主页上找到.

## References

- [1] E. Blum and W. Oettli, *Mathematische Optimierung, Econometrics and Operations Research XX*, Springer Verlag, 1975.
- [2] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foun. Trends Mach. Learn.*, **3**, 1-122, 2010.
- [3] A. Chambolle, T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, *J. Math. Imaging Vison*, **40**, 120-145, 2011.
- [4] 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, *Mathematical Programming, Series A* **155**. 57-79, 2015.
- [5] E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, *IEEE Trans. Imag. Process.*, **21**(7), 3239-3252, 2012.
- [6] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.
- [7] G.Y. Gu, B.S. He and X.M. Yuan, Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach, *Comput. Optim. Appl.*, **59**, 135-161, 2014.
- [8] B.S. He, A new method for a class of linear variational inequalities, *Math. Progr.*, **66**, 137-144, 1994.

- [9] B.S. He, Solving a class of linear projection equations, *Numerische Mathematik*, **68**, 71–80, 1994.
- [10] B.S. He, A class of projection and contraction methods for monotone variational inequalities, *Appl. Math. & Optim.*, **35**, 69–76, 1997.
- [11] B.S. He, and L.Z. Liao, Improvements of some projection methods for monotone nonlinear variational inequalities, *JOTA* **112**, 111–128, 2002.
- [12] B.S. He, L.Z. Liao, D.R. Han, and H. Yang, A new inexact alternating directions method for monotone variational inequalities, *Math. Progr.*, **92**, 103–118, 2002.
- [13] B.S. He, L.Z. Liao, and M.J. Qian, Alternating projection based prediction-correction methods for structured variational inequalities, *J. Comput. Math.*, **24**, 693–710, 2006.
- [14] B.S. He, L.Z. Liao, and X. Wang, Proximal-like contraction methods for monotone variational inequalities in a unified framework I: Effective quadruplet and primary methods, *Comput. Optim. Appl.*, **51**, 649-679, 2012
- [15] B.S. He, L.Z. Liao, and X. Wang, Proximal-like contraction methods for monotone variational inequalities in a unified framework II: General methods and numerical experiments, *Comput. Optim. Appl.*, **51**, 681-708, 2012
- [16] B. S. He, H. Liu, Z.R. Wang and X.M. Yuan, A strictly contractive Peaceman-Rachford splitting method for convex programming, *SIAM Journal on Optimization* **24**, 1011-1040, 2014.
- [17] 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**, 313-340, 2012.
- [18] 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.
- [19] B. S. He and X. M. Yuan, On the  $O(1/n)$  convergence rate of the alternating direction method, *SIAM J. Numerical Analysis* **50**, 700-709, 2012.
- [20] B.S. He and X.M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective. *SIAM J. Imaging Science*, **5**, 119-149, 2012.
- [21] B.S. He, X.M. Yuan and J.J.Z. Zhang, Comparison of two kinds of prediction-correction methods for monotone variational inequalities, *Comput. Optim. Appl.*, **27**, 247-267, 2004
- [22] B.S. He, X.M. Yuan and W.X. Zhang, A customized proximal point algorithm for convex minimization with linear constraints, *Comput. Optim. Appl.*, **56**, 559-572, 2013.
- [23] B. Martinet, Regularisation, d'inéquations variationnelles par approximations succesives, *Rev. Francaise d'Inform. Recherche Oper.*, **4**, 154-159, 1970.
- [24] D.F. Sun, A class of iterative methods for solving nonlinear projection equations, *JOTA*, **91**, 123-140, 1996.
- [25] R.T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Cont. Optim.*, **14**, 877-898, 1976.
- [26] M. Zhu and T. F. Chan, An efficient primal-dual hybrid gradient algorithm for total variation image restoration, *CAM Report 08-34*, UCLA, USA, 2008.
- [27] 何炳生, 论求解单调变分不等式的一些投影收缩算法, *计算数学*, **18**, 54-60, 1996.
- [28] 何炳生, 生活理念对设计优化算法的帮助—以改造ADMM求解三个可分离算子问题为例, 见 <http://math.nju.edu.cn/~hebma> 中的《My Talk》
- [29] 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://math.nju.edu.cn/~hebma> 中的《My Talk》
- [30] 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.