

变分不等式框架下结构型 凸优化的分裂收缩算法

VIII. 投影收缩算法和可微凸优化问题的算法设计

中学的数理基础 必要的社会实践
普通的大学数学 一般的优化原理

何炳生 南京大学数学系

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

天元数学东北中心 2023年10月17 – 27日

1 引言

研究凸优化分裂收缩算法的统一框架, 跟我们长期研究经典单调变分不等式的投影收缩算法[1, 3, 4, 5, 6, 9, 10, 11, 14]有着密切的联系. 变分不等式是描述平衡问题的数学工具, 在管理科学和工程计算中都有广泛的应用.

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

$$\text{VI}(\Omega, F) \quad u^* \in \Omega, \quad (u - u^*)^T F(u^*) \geq 0, \quad \forall u \in \Omega \quad (1.1)$$

的求解. 变分不等式(1.1)单调, 是指其中的算子 F 满足

$$(u - v)^T (F(u) - F(v)) \geq 0, \quad \forall u, v \in \mathfrak{R}^n \text{ (或 } \Omega). \quad (1.2)$$

这里的单调算子 F 可以是非线性的, 有别于由凸优化转换而来的变分不等式中的仿射算子 F , 其中的系数矩阵是反对称矩阵, 具备性质

$$(u - v)^T (F(u) - F(v)) \geq 0, \quad \forall u, v \in \mathfrak{R}^n \text{ (或 } \Omega).$$

为了有所区别, 在这一讲的讨论中, 我们把由线性约束凸优化问题转换得来的变分不等式称为混合变分不等式, 形式为(1.1)的变分不等式称为经典变分不等式.

非线性互补问题是一类特殊的变分不等式

在第一讲中,我们就讨论过凸优化问题

$$\min\{f(x) \mid x \in \Omega\}. \quad (1.3)$$

它的最优性条件

$$x^* \in \Omega, \quad (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \in \Omega.$$

当 $\Omega = \mathfrak{R}_+^n$ 时,就是一个约束集合为非负卦限的变分不等式

$$x^* \geq 0, \quad (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \geq 0.$$

设 F 是 $\mathfrak{R}^n \rightarrow \mathfrak{R}^n$ 的一个算子. 非负卦限 \mathfrak{R}_+^n 上变分不等式的一般形式是

$$\text{VI}(\mathfrak{R}_+^n, F) \quad x \geq 0, \quad (x' - x)^T F(x) \geq 0, \quad \forall x' \geq 0. \quad (1.4)$$

非线性互补问题是最优化理论与方法中一类很重要的问题. 它的数学形式是

$$\text{(NCP)} \quad x \geq 0, \quad F(x) \geq 0, \quad x^T F(x) = 0. \quad (1.5)$$

事实上, NCP 是 $\Omega = \mathfrak{R}_+^n$ 的一类变分不等式.

定理 1 非负卦限上的变分不等式 (1.4) 和互补问题 (1.5) 是等价的.

证明 如果 x 是互补问题 (1.5) 的解, 那么有 $x \geq 0$ 和 $F(x) \geq 0$. 对于任意的 $x' \geq 0$ 有 $x'^T F(x) \geq 0$. 又因 $x^T F(x) = 0$, 所以有

$$x \geq 0, \quad (x' - x)^T F(x) = x'^T F(x) - x^T F(x) \geq 0, \quad \forall x' \geq 0.$$

所以 x 是 $\text{VI}(R_+^n, F)$ (1.4) 的一个解.

反过来, 如果 x 是 $\text{VI}(R_+^n, F)$ (1.4) 的一个解, 则 $x \geq 0$. 将 $x' = 2x$ 和 $x' = 0$ 代入

$$(x' - x)^T F(x) \geq 0,$$

得到 $\pm x^T F(x) \geq 0$. 因此 $x^T F(x) = 0$.

要证明 x 是互补问题 (1.5) 的解, 只剩下 $F(x) \geq 0$ 需要证明, 对此采用反证法. 如果 $F(x)$ 的某个分量 $F_j(x) < 0$, 我们取 x' , 使得

$$x'_i = \begin{cases} x_i, & \text{if } i \neq j \\ x_j + 1, & \text{if } i = j \end{cases}$$

这样的 $x' \geq 0$. 但 $(x' - x)^T F(x) = F_j(x) < 0$, 这与 x 是 $\text{VI}(R_+^n, F)$ (1.4) 的解矛盾.

因此, 在 $\Omega = \mathfrak{R}_+^n$ 时, 可微凸优化问题 (1.3) 与一个互补问题等价.

经典单调变分不等式的投影收缩算法也是一种预测-校正方法. 经典变分不等式框架下的投影收缩算法也为线性约束可微凸优化问题的求解提供了一些其他的途径.

这一讲的内容安排上,

- 首先介绍经典的单调变分不等式及其等价的投影方程,
- 经典变分不等式投影收缩算法和凸优化分裂收缩算法中的预测
- 经典变分不等式投影收缩算法和凸优化分裂收缩算法中的预测校正
- 投影收缩算法中的孪生方向和姊妹方法
- 投影收缩算法在求解可分离凸优化上的应用

2 投影与变分不等式的一些基本性质

投影收缩算法中的基本运算是执行向量到简单闭凸集上的投影.

2.1 投影的基本性质

用 $P_{\Omega}(\cdot)$ 表示欧氏范数下在凸集 Ω 上的投影, 也就是说

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

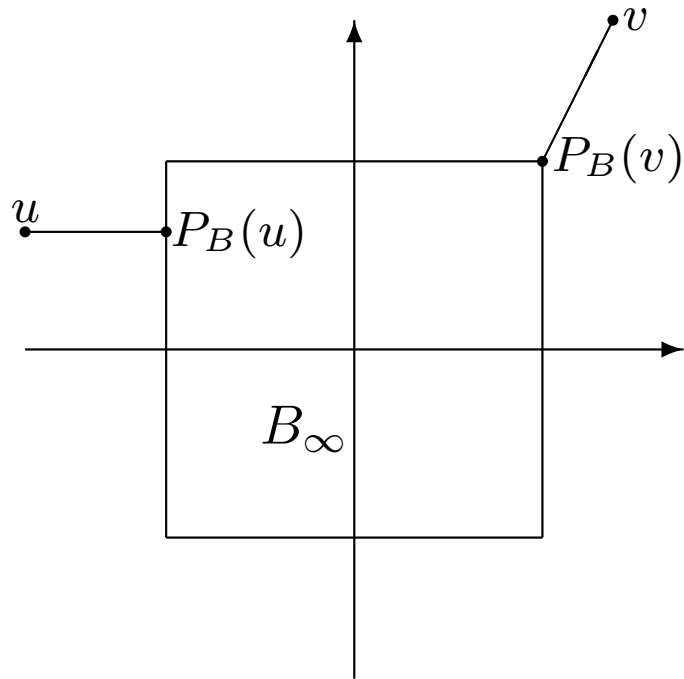
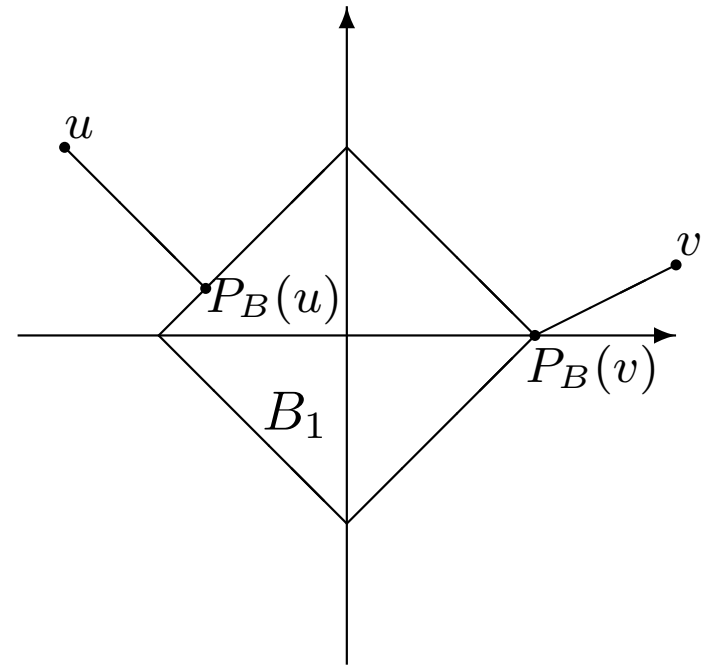
如果 $\Omega = \mathfrak{R}_+^n$ (n -维空间的非负卦限), 那么 $P_{\Omega}(v)$ 的每个分量为

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

如果 Ω 是 n -维空间中以 c 为球心半径为 r 的球, 那么

$$P_{\Omega}(v) = \begin{cases} \frac{r(v-c)}{\|v-c\|} + c, & \text{if } \|v-c\| \geq r; \\ v, & \text{otherwise.} \end{cases}$$

在 l_{∞} 和 l_1 模意义下的“单位球”上投影如下图所示:

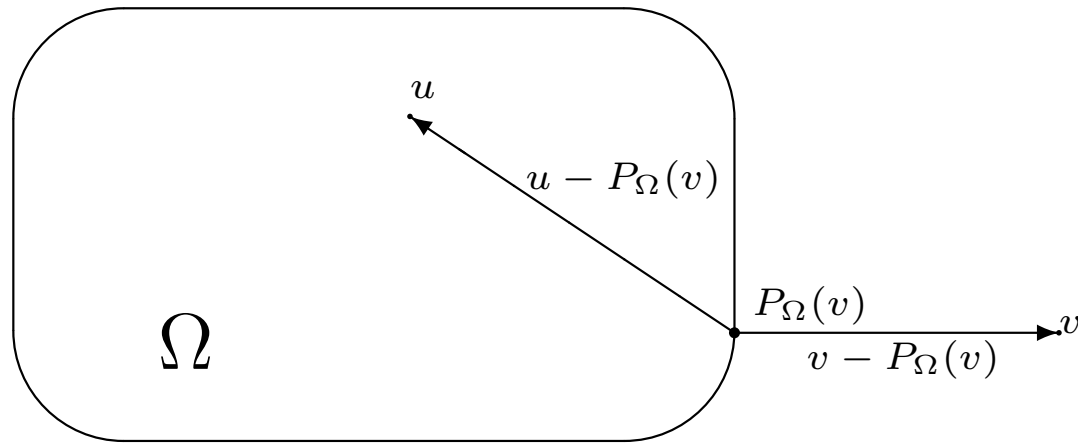
Fig. 1. Projection on B_∞ Fig. 2. Projection on B_1

通篇, 我们记

$$\tilde{u} = P_\Omega[u - \mathbb{I}F(u)].$$

引理 1 设 $\Omega \subset R^n$ 是闭凸集, 则有

$$(v - P_\Omega(v))^T (u - P_\Omega(v)) \leq 0 \quad \forall v \in R^n, \forall u \in \Omega. \quad (2.1)$$



不等式(2.1)的几何解释.

证明. 首先, 根据 $P_{\Omega}(v)$ 的定义, 有

$$\|v - P_{\Omega}(v)\| \leq \|v - w\|, \quad \forall w \in \Omega. \quad (2.2)$$

注意到对任意的 $v \in \mathfrak{R}^n$, 都有 $P_{\Omega}(v) \in \Omega$, 由于 $\Omega \subset \mathfrak{R}^n$ 是闭凸集, 则对任意的 $u \in \Omega$ 和 $\theta \in (0, 1)$, 都有

$$w := \theta u + (1 - \theta)P_{\Omega}(v) = P_{\Omega}(v) + \theta(u - P_{\Omega}(v)) \in \Omega.$$

对这个 w , 利用 (2.2), 就有

$$\|v - P_{\Omega}(v)\|^2 \leq \|v - P_{\Omega}(v) - \theta(u - P_{\Omega}(v))\|^2.$$

对上式展开, 对任意的 $u \in \Omega$ 和 $\theta \in (0, 1)$, 都有

$$[v - P_{\Omega}(v)]^T [u - P_{\Omega}(v)] \leq \frac{\theta}{2} \|u - P_{\Omega}(v)\|^2.$$

令 $\theta \rightarrow 0_+$, 引理(2.1)得证. \square

在投影收缩算法的分析中, 不等式(2.1)是一个非常用的基本工具. 我们因此而称之为投影算子的工具不等式. 由(2.1), 容易证明下面的引理.

引理 2 设 $\Omega \subset \mathbb{R}^n$ 是闭凸集, 则有

$$\|P_{\Omega}(v) - P_{\Omega}(u)\| \leq \|v - u\|, \quad \forall u, v \in \mathbb{R}^n. \quad (2.3)$$

$$\|P_{\Omega}(v) - u\| \leq \|v - u\|, \quad \forall v \in \mathbb{R}^n, u \in \Omega. \quad (2.4)$$

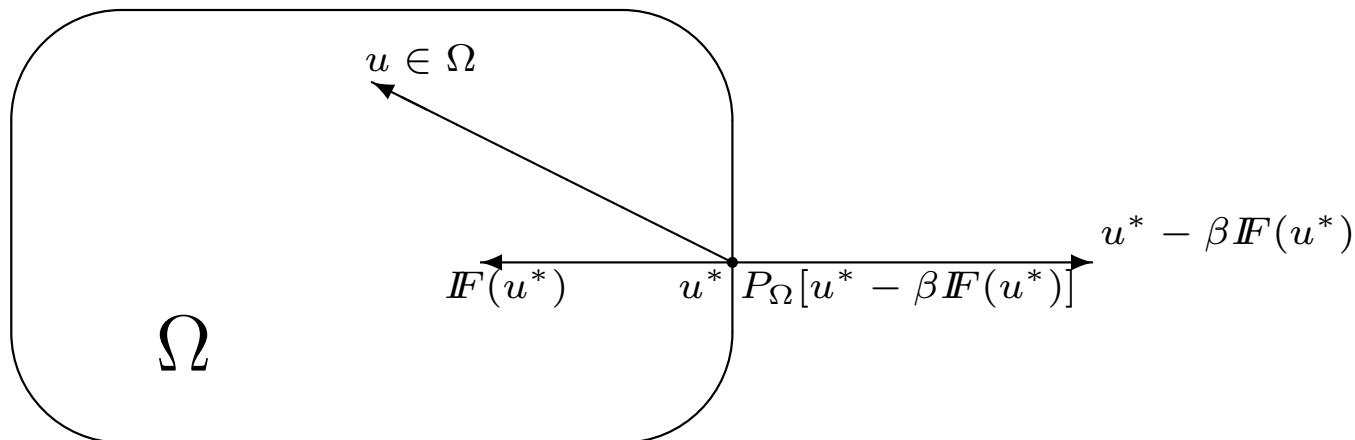
$$\|P_{\Omega}(v) - u\|^2 \leq \|v - u\|^2 - \|v - P_{\Omega}(v)\|^2, \quad \forall v \in \mathbb{R}^n, u \in \Omega. \quad (2.5)$$

我们把这些性质留给读者自己去证明.

2.2 变分不等式等价的投影方程

设变分不等式(1.1)的解集 Ω^* 非空. 我们用 u^* 表示一个确定的解点. 对任意的 $\beta > 0$, 变分不等式等价于投影方程

$$u \in \Omega^* \Leftrightarrow u = P_{\Omega}[u - \beta IF(u)].$$



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

换言之, 求解变分不等式可以归结为求

$$e(u, \beta) := u - P_{\Omega}[u - \beta IF(u)] \quad (2.6)$$

的一个零点 u^* , 后面我们会给出证明. 因此, 对确定的 $\beta > 0$, $\|e(u, \beta)\|$ 可以看作一种误差的度量函数. 为了方便, 我们往往把 $e(u, 1)$ 记成 $e(u)$.

定理 2 设 $\beta > 0$. u^* 是 $VI(\Omega, F)$ 的解当且仅当 $e(u^*, \beta) = 0$.

证明. 先证必要性. 若 u^* 是 $VI(\Omega, F)$ 的解, 则 $u^* \in \Omega$. 由于 $\Omega \subset R^n$ 是闭凸集, 利

用 (2.1) 得到

$$(v - P_{\Omega}(v))^T (u^* - P_{\Omega}(v)) \leq 0, \quad \forall v \in R^n.$$

上式中取 $v := u^* - \beta \mathbf{F}(u^*)$, 则有 $(e(u^*, \beta) - \beta \mathbf{F}(u^*))^T e(u^*, \beta) \leq 0$, 即

$$\|e(u^*, \beta)\|^2 \leq \beta e(u^*, \beta)^T \mathbf{F}(u^*). \quad (2.7)$$

另一方面, 由于 $P_{\Omega}[u^* - \beta \mathbf{F}(u^*)] \in \Omega$, 而且 u^* 是变分不等式的解, 根据 (1.1) 可以得到

$$\{P_{\Omega}[u^* - \beta \mathbf{F}(u^*)] - u^*\}^T \mathbf{F}(u^*) \geq 0,$$

即

$$e(u^*, \beta)^T \mathbf{F}(u^*) \leq 0. \quad (2.8)$$

由不等式 (2.7) 和 (2.8) 可得 $e(u^*, \beta) = 0$.

再证充分性. 取 $v = u^* - \beta \mathbf{F}(u^*)$, 利用 (2.1) 和 $e(u^*, \beta)$ 的表达式, 有

$$\{e(u^*, \beta) - \beta \mathbf{F}(u^*)\}^T \{u - P_{\Omega}[u^* - \beta \mathbf{F}(u^*)]\} \leq 0, \quad \forall u \in \Omega. \quad (2.9)$$

根据条件 $e(u^*, \beta) = 0$, 有 $u^* = P_{\Omega}(\cdot) \in \Omega$ 和 $P_{\Omega}[u^* - \beta \mathbf{F}(u^*)] = u^*$. 代入不等式 (2.9), 可以得到

$$u^* \in \Omega, \quad (u - u^*)^T \mathbf{F}(u^*) \geq 0, \quad \forall u \in \Omega,$$

即 u^* 是 $\text{VI}(\Omega, F)$ 的解. 定理得证. \square

下面的定理说明 $\|e(u, \beta)\|$ 是 β 的不减函数, 而 $\{\|e(u, \beta)\|/\beta\}$ 是 β 的不增函数. 这个简单证明只用到一元二次不等式的初等知识和工具不等式 (2.1), 它初见于 [19].

定理 3 对所有的 $u \in \mathfrak{R}^n$ 和 $\tilde{\beta} \geq \beta > 0$, 我们有

$$\|e(u, \tilde{\beta})\| \geq \|e(u, \beta)\| \quad (2.10)$$

和

$$\frac{\|e(u, \tilde{\beta})\|}{\tilde{\beta}} \leq \frac{\|e(u, \beta)\|}{\beta}. \quad (2.11)$$

证明. 设 $t = \|e(x, \tilde{\beta})\|/\|e(x, \beta)\|$, 定理的结论就相当于要证明

$$1 \leq t \leq \frac{\tilde{\beta}}{\beta}.$$

注意到它的等价表达式是 t 的一元二次不等式

$$(t - 1)(t - \frac{\tilde{\beta}}{\beta}) \leq 0 \quad (2.12)$$

的解. 首先, 由工具不等式 (2.1), 我们有

$$(v - P_{\Omega}(v))^T (P_{\Omega}(v) - w) \geq 0, \quad \forall w \in \Omega. \quad (2.13)$$

在 (2.13) 中令 $w := P_{\Omega}[u - \tilde{\beta}\mathbf{F}(u)]$ 和 $v := u - \beta\mathbf{F}(u)$, 利用 $e(u, \beta)$ 的定义和

$$P_{\Omega}[u - \beta\mathbf{F}(u)] - P_{\Omega}[u - \tilde{\beta}\mathbf{F}(u)] = e(u, \tilde{\beta}) - e(u, \beta),$$

我们得到

$$\{e(u, \beta) - \beta \mathbf{F}(u)\}^T \{e(u, \tilde{\beta}) - e(u, \beta)\} \geq 0. \quad (2.14)$$

用相应的方法(将上式中的 β 和 $\tilde{\beta}$ 互换位置), 可得

$$\{e(u, \tilde{\beta}) - \tilde{\beta} \mathbf{F}(u)\}^T \{e(u, \beta) - e(u, \tilde{\beta})\} \geq 0. \quad (2.15)$$

分别将不等式 (2.14) 和 (2.15) 乘上 $\tilde{\beta}$ 和 β , 然后再将它们相加, 我们得到

$$\{\tilde{\beta}e(u, \beta) - \beta e(u, \tilde{\beta})\}^T \{e(u, \tilde{\beta}) - e(u, \beta)\} \geq 0 \quad (2.16)$$

并有

$$\beta \|e(x, \tilde{\beta})\|^2 - (\beta + \tilde{\beta})e(x, \beta)^T e(x, \tilde{\beta}) + \tilde{\beta} \|e(x, \beta)\|^2 \leq 0.$$

对上式采用 Cauchy-Schwarz 不等式, 就有

$$\beta \|e(x, \tilde{\beta})\|^2 - (\beta + \tilde{\beta}) \|e(x, \beta)\| \cdot \|e(x, \tilde{\beta})\| + \tilde{\beta} \|e(x, \beta)\|^2 \leq 0. \quad (2.17)$$

将 (2.17) 除 $\beta \|e(x, \beta)\|^2$, 并利用 t 的定义便得

$$t^2 - \left(1 + \frac{\tilde{\beta}}{\beta}\right)t + \frac{\tilde{\beta}}{\beta} \leq 0.$$

因此不等式 (2.12) 成立, 定理得证. \square

定理 3 说明, 若以 $\|e(u, \beta)\|$ 作为停机的误差度量, 常数 $\beta > 0$ 不宜过大, 也不宜过小. 一般要结合问题的物理意义考虑.

2.3 投影收缩算法的三个基本不等式

设 u^* 是变分不等式 $VI(\Omega, F)$ 的解. 记

$$\tilde{u} = P_{\Omega}[u - \beta F(u)] \quad (2.18)$$

根据变分不等式的定义 (1.1), 有第一个基本不等式

$$(F1) \quad (\tilde{u} - u^*)^T \beta F(u^*) \geq 0, \quad \forall u^* \in \Omega^*. \quad (2.19)$$

由于 $u^* \in \Omega$, 由 (2.18) 给出的 \tilde{u} 是 $[u - \beta F(u)]$ 在 Ω 上的投影. 在投影的基本性质不等式 (2.1) 中, 分别设 $v = u - \beta F(u)$ 和任意的属于 Ω 的 $u = u^*$, 则有

$$(F2) \quad (\tilde{u} - u^*)^T \{[u - \beta F(u)] - \tilde{u}\} \geq 0, \quad \forall u^* \in \Omega^*. \quad (2.20)$$

此外, 根据单调算子的性质, 有

$$(F3) \quad (\tilde{u} - u^*)^T \{\beta F(\tilde{u}) - \beta F(u^*)\} \geq 0, \quad \forall u^* \in \Omega^*. \quad (2.21)$$

我们把 (2.19)-(2.21) 称为投影收缩算法中的三个基本不等式 [5]. 将这里的 (2.19), (2.20) 和 (2.21) 加在一起, 就得到

$$(\tilde{u} - u^*)^T d(u, \tilde{u}) \geq 0, \quad \forall u^* \in \Omega^*. \quad (2.22)$$

其中

$$d(u, \tilde{u}) = (u - \tilde{u}) - \beta[F(u) - F(\tilde{u})]. \quad (2.23)$$

如果只将 (2.19) 和 (2.21) 加在一起, 得到的是

$$(\tilde{u} - u^*)^T \{\beta \mathbf{F}(\tilde{u})\} \geq 0, \quad \forall u^* \in \Omega^*. \quad (2.24)$$

后面我们将会定义, 这里由不同组合得到的 $d(u, \tilde{u})$ 和 $\beta \mathbf{F}(\tilde{u})$, 称为一对孪生方向.

3 投影收缩算法和分裂收缩算法中的预测

和求解混合变分不等式的分裂收缩算法一样, 求解单调经典变分不等式 (1.1) 的投影收缩算法 [5, 9, 14, 6] 是一个采用投影为预测的预测-校正方法.

3.1 求解经典变分不等式的投影收缩算法中的预测

求解经典变分不等式 (1.1) 的投影收缩算法的第 k -步迭代从给定的 u^k 开始, 通过欧氏模意义下的投影得到预测点 \tilde{u}^k , 具体公式是

$$\tilde{u}^k = P_{\Omega}[u^k - \beta_k \mathbf{F}(u^k)], \quad (3.1)$$

其中 β_k 的选择要求满足

$$\beta_k \|\mathbf{F}(u^k) - \mathbf{F}(\tilde{u}^k)\| \leq \nu \|u^k - \tilde{u}^k\|, \quad \nu \in (0, 1). \quad (3.2)$$

定义 (合格的投影预测) 求解非线性变分不等式 (1.1) 的预测-校正方法中, 对给定的常数 $\nu \in (0, 1)$, 若通过投影 (3.1) 得到的预测点 \tilde{u}^k 满足条件 (3.2), 则称其为一个合格的投影预测.

在 F 为 Lipschitz 连续的条件下, (3.2) 是能够实现的. 由于 \tilde{u}^k 可以表示成

$$\tilde{u}^k = \arg \min \left\{ \frac{1}{2} \|u - [u^k - \beta_k F(u^k)]\|^2 \mid u \in \Omega \right\},$$

根据最优性定理, 由 (3.1) 得到的 \tilde{u}^k 满足

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

进而得到

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

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

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^T \beta_k F(\tilde{u}^k) \geq (u - \tilde{u}^k)^T d(u^k, \tilde{u}^k), \quad \forall u \in \Omega, \quad (3.3)$$

其中的 $d(u^k, \tilde{u}^k)$ 是由 (2.23) 定义的.

不等式 (3.3) 对应于求解混合变分不等式的统一框架中的预测, 这里的 $d(u^k, \tilde{u}^k)$ 相当于其中的 $Q(v^k - \tilde{v}^k)$. 差别在于 $d(u^k, \tilde{u}^k)$ 并不是 $(u^k - \tilde{u}^k)$ 的线性函数.

将(3.3)中的 $u \in \Omega$ 选成任意确定的 $u^* \in \Omega^*$, 得到

$$(\tilde{u}^k - u^*)^T d(u^k, \tilde{u}^k) \geq (\tilde{u}^k - u^*)^T \beta_k \mathbb{F}(\tilde{u}^k). \quad (3.4)$$

由 \mathbb{F} 的单调性和 u^* 的最优性, 上式右端

$$(\tilde{u}^k - u^*)^T \mathbb{F}(\tilde{u}^k) \geq (\tilde{u}^k - u^*)^T \mathbb{F}(u^*) \geq 0,$$

不等式(3.4)的左端非负, 随后改写成

$$\{(u^k - u^*) - (u^k - \tilde{u}^k)\}^T d(u^k, \tilde{u}^k) \geq 0,$$

得到

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \quad (3.5)$$

根据 $d(u^k, \tilde{u}^k)$ 的表达式(2.23)和假设(3.2), 利用 Cauchy-Schwarz 不等式, 有

$$(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \geq (1 - \nu) \|u^k - \tilde{u}^k\|^2. \quad (3.6)$$

由上面两个不等式得知

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq (1 - \nu) \|u^k - \tilde{u}^k\|^2. \quad (3.7)$$

定义(上升方向) 求解变分不等式(1.1)的方法中, 假如存在一个常数 $\delta > 0$, 向量 $d(u^k, \tilde{u}^k)$ 满足关系式

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq \delta \|u^k - \tilde{u}^k\|^2, \quad \forall u^* \in \Omega^*, \quad (3.8)$$

则称其为距离函数 $\|u - u^*\|^2$ 在 u^k 处的上升方向.

因此, 由 (2.23) 给出的 $d(u^k, \tilde{u}^k)$ 是未知距离函数 $\|u - u^*\|^2$ 在 u^k 处欧氏模下的一个上升方向. 虽然我们并不知道解点在哪里, 但是沿着方向 $-d(u^k, \tilde{u}^k)$, 选取适当步长, 可以找到欧氏模下比 u^k 更靠近解集的 u^{k+1} .

3.2 求解混合变分不等式的分裂收缩算法中的预测

由线性约束的凸优化问题得到的混合变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall w \in \Omega. \quad (3.9)$$

统一框架中定义的预测为

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \quad (3.10)$$

其中矩阵 $Q^T + Q$ 是本质上正定的. 将 (3.10) 中任意的 $w \in \Omega$ 选成 $w^* \in \Omega^*$, 我们有

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

由

$$(\tilde{w}^k - w^*)^T F(\tilde{w}^k) = (\tilde{w}^k - w^*)^T F(w^*)$$

和 w^* 的最优性, (3.11) 的左端非负. 随后由它得到

$$(v^k - v^*)^T Q(v^k - \tilde{v}^k) \geq (v^k - \tilde{v}^k)^T Q(v^k - \tilde{v}^k).$$

对正定矩阵 H 和 $H^{-1}Q = M$ 上式可以表示成

$$\langle H(v^k - v^*), M(v^k - \tilde{v}^k) \rangle \geq (v^k - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \quad (3.12)$$

不等式 (3.12) 告诉我们, 在条件 $Q^T + Q$ 正定的情况下, 向量 $M(v^k - \tilde{v}^k)$ 是 H -模下未知距离函数 $\|v - v^*\|_H^2$ 在 v^k 处的一个上升方向.

这两类方法的预测中, (3.3) 和 (3.4) 分别跟 (3.10) 和 (3.11) 相对应. 对应于经典变分不等式的投影收缩算法和混合变分不等式的分裂收缩算法, 不等式 (3.5) 和 (3.12) 分别提供了相应的上升方向. 它们的右端严格大于零分别由假设 (3.2) 和 $Q^T + Q \succ 0$ 得到保证.

4 投影收缩算法和分裂收缩算法中的校正

校正是利用距离函数的下降方向(上升方向的反方向), 使得新的迭代点在某种确定的模的意义下离解集比原来的点更近一些.

4.1 两类方法中采用固定步长的校正

1. 变分不等式投影收缩算法固定步长的校正 在投影收缩算法中, 我们一般考虑欧氏模下的收缩. 用关系式 (3.5), 校正通过

$$u^{k+1} = u^k - \alpha d(u^k, \tilde{u}^k) \quad (4.1)$$

产生新的迭代点, 其中 $d(u^k, \tilde{u}^k)$ 是由 (2.23) 给出的. 下面我们讨论如何选取步长 α .

投影收缩算法中, 我们将条件 (3.2) 满足时, 由单位步长校正产生新迭代点

$$u^{k+1} = u^k - d(u^k, \tilde{u}^k) \quad (4.2)$$

的方法, 称为初等方法 (Primary Method). 利用 (3.5), 由简单计算可得

$$\begin{aligned} \|u^{k+1} - u^*\|^2 &= \|(u^k - u^*) - d(u^k, \tilde{u}^k)\|^2 \\ &= \|u^k - u^*\|^2 - 2(u^k - u^*)^T d(u^k, \tilde{u}^k) + \|d(u^k, \tilde{u}^k)\|^2 \\ &\leq \|u^k - u^*\|^2 - [2(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \|d(u^k, \tilde{u}^k)\|^2]. \end{aligned} \quad (4.3)$$

利用 (2.23) 和 (3.2), 可以得到

$$\begin{aligned}
& 2(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \|d(u^k, \tilde{u}^k)\|^2 \\
&= d(u^k, \tilde{u}^k)^T \{2(u^k - \tilde{u}^k) - d(u^k, \tilde{u}^k)\} \\
&= \{(u^k - \tilde{u}^k) - \beta_k[\mathbf{F}(u^k) - \mathbf{F}(\tilde{u}^k)]\}^T \{(u^k - \tilde{u}^k) + \beta_k[\mathbf{F}(u^k) - \mathbf{F}(\tilde{u}^k)]\} \\
&= \|u^k - \tilde{u}^k\|^2 - \beta_k^2 \|\mathbf{F}(u^k) - \mathbf{F}(\tilde{u}^k)\|^2 \\
&\geq (1 - \nu^2) \|u^k - \tilde{u}^k\|^2.
\end{aligned} \tag{4.4}$$

代入 (4.3), 说明由 (4.2) 产生的序列 $\{u^k\}$ 满足

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - (1 - \nu^2) \|u^k - \tilde{u}^k\|^2. \tag{4.5}$$

2. 凸优化分裂收缩算法固定步长的校正 在凸优化的分裂收缩算法中, 我们一般考虑 H -模下的收缩. 用关系式 (3.12), 校正通过

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k) \tag{4.6}$$

产生新的迭代点, 其中 $M = H^{-1}Q$. 下面我们讨论如何选取步长 α . 由简单计算可得

$$\begin{aligned}
\|v^{k+1} - v^*\|_H^2 &= \|(v^k - v^*) - \alpha M(v^k - \tilde{v}^k)\|_H^2 \\
&= \|v^k - v^*\|_H^2 - 2\alpha(v^k - v^*)^T HM(v^k - \tilde{v}^k) + \alpha^2 \|M(v^k - \tilde{v}^k)\|_H^2 \\
&\leq \|v^k - v^*\|_H^2 - \alpha((v^k - \tilde{v}^k)^T [Q^T + Q - \alpha M^T HM](v^k - \tilde{v}^k)).
\end{aligned}$$

可以得到

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \alpha \|v^k - \tilde{v}^k\|_G^2, \quad (4.7)$$

其中

$$G = Q^T + Q - \alpha M^T H M.$$

根据统一框架收敛性的要求, G 是正定矩阵.

4.2 两类方法中采计算步长的校正

1. 变分不等式投影收缩算法计算步长的校正 我们将 (4.1) 中的 u^{k+1} 记为 $u^{k+1}(\alpha)$, 表示新的迭代点依赖于步长 α . 考察与 α 相关的距离平方缩短量,

$$\vartheta_k(\alpha) = \|u^k - u^*\|^2 - \|u^{k+1}(\alpha) - u^*\|^2. \quad (4.8)$$

根据定义

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

对任意给定的确定解点 u^* , 上式表明 $\vartheta_k(\alpha)$ 是 α 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. 利用 (3.5), 对任意的 $\alpha > 0$, 有

$$\vartheta_k(\alpha) \geq q_k(\alpha), \quad (4.9)$$

其中

$$q_k(\alpha) = 2\alpha(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2. \quad (4.10)$$

既然二次函数 $q_k(\alpha)$ 是 $\vartheta_k(\alpha)$ 的一个下界函数. 使 $q_k(\alpha)$ 达到极大的 α_k^* 是

$$\alpha_k^* = \operatorname{argmax}\{q_k(\alpha)\} = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2}. \quad (4.11)$$

注意到这里的 α_k^* 是由 (3.5) 确定的, 分子是 (3.5) 的右端, 分母是 (4.1) 中 $d(u^k, \tilde{u}^k)$ 的欧氏长度的平方.

在实际计算中, 我们一般取一个松弛因子 $\gamma \in [1.2, 1.8]$, 令

$$u^{k+1} = u^k - \gamma\alpha_k^* d(u^k, \tilde{u}^k), \quad (4.12)$$

根据 (4.8) 和 (4.9), 由 (4.12) 产生的 u^{k+1} 满足

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - q_k(\gamma\alpha_k^*). \quad (4.13)$$

其中

$$\begin{aligned} q_k(\gamma\alpha_k^*) &= 2\gamma\alpha_k^*(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \gamma^2(\alpha_k^*)^2 \|d(u^k, \tilde{u}^k)\|^2 \\ &= \gamma(2 - \gamma)\alpha_k^*(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \end{aligned} \quad (4.14)$$

此外, 从 (4.4), 我们已经有 $2(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \|d(u^k, \tilde{u}^k)\|^2 > 0$, 因而根

据(4.11)得到 $\alpha_k^* > \frac{1}{2}$ (见(4.4)). 结合(3.6),

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - \frac{1}{2}\gamma(2 - \gamma)(1 - \nu)\|u^k - \tilde{u}^k\|^2. \quad (4.15)$$

不等式(4.5)和(4.15)说明, 用固定步长和计算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.

虽然投影收缩算法的这些结论都在假设(3.2)满足时才成立, 但是当 F 是 Lipschitz 连续, 采用 Armijo 法则选取适当的 β_k 实施的预测(3.1), 是能够使条件(3.2)满足的. 采用下面程序的方法我们称其为投影收缩算法-I.

求解经典单调变分不等式的投影收缩算法-I

给定 $\beta_0 = 1, \mu = 0.4, \nu = 0.9, u^0 \in \Omega$.

For $k = 0, 1, \dots$, 假如停机准则尚未满足, **do**

$$1). \tilde{u}^k = P_{\Omega}[u^k - \beta_k \mathbf{F}(u^k)],$$

$$r_k := \frac{\beta_k \|\mathbf{F}(u^k) - \mathbf{F}(\tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|},$$

$$\mathbf{while} \quad r_k > \nu, \quad \beta_k := \frac{2}{3}\beta_k * \min\{1, \frac{1}{r_k}\},$$

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

$$r_k := \frac{\beta_k \|\mathbf{F}(u^k) - \mathbf{F}(\tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|},$$

end(while)

$$2). d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k [\mathbf{F}(u^k) - \mathbf{F}(\tilde{u}^k)],$$

$$\alpha_k^* = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2},$$

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

$$3). \mathbf{if} \quad r_k \leq \mu \quad \mathbf{then} \quad \beta_k := \beta_k * 1.5, \quad \mathbf{end(if)}$$

4). 令 $\beta_{k+1} = \beta_k$ 和 $k := k + 1$, 开始新的一次迭代.

- 在做投影预测 (3.1) 的时候, 要求满足条件 (3.2). 这里取的 $\nu = 0.9$, 是经验值. 在投

影收缩算法-I的 1) 中, 当 $r_k > \nu$ 时, 用

$$\beta_k := \frac{2}{3}\beta_k * \min\{1, \frac{1}{r_k}\}$$

对 β_k 做调正. 这里的两种不同情形是: 当 $r_k \in (\nu, 1]$ 之间, 就取 $\beta_k := \frac{2}{3}\beta_k$; 当 $r_k > 1$ 时, 相当于把原来的 β_k 缩小 r_k 倍再乘上 $(2/3)$. 调正参数 β_k 以后, 重做一次投影预测, 条件 (3.2) 一般能够得到满足. 这里的 $(2/3)$ 也是经验值, 实际计算中, 也可以改成 $(3/4)$.

- 在条件 (3.2) 满足的前提下, 我们同时希望

$$\frac{\beta_k \|F(u^k) - F(\tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|}$$

不要太小. 根据我们计算的一些例子, 程序中的那句

$$\text{if } r_k \leq \mu \text{ then } \beta_k := \beta_k * 1.5, \text{ end(if)}$$

是不可缺少的. 就像信赖域方法 [16] 中的信赖域半径, 在迭代计算过程发现过小也需要增大, 其中的 1.5 也是经验值.

2. 凸优化分裂收缩算法计算步长的校正 分裂收缩算法的计算步长的校正方法的根据是 (3.12). 利用 $HM = Q$ 并考虑距离函数 $\|v - v^*\|_H^2$ 的下降, 新的迭代点 v^{k+1} 由

$$v^{k+1} = v^k - \gamma \alpha_k^* M(v^k - \tilde{v}^k), \quad (4.16a)$$

产生, 其中

$$\alpha_k^* = \frac{(v^k - \tilde{v}^k)^T Q(v^k - \tilde{v}^k)}{\|M(v^k - \tilde{v}^k)\|_H^2}. \quad (4.16b)$$

分子是 (3.12) 的右端, 分母 $\|M(v^k - \tilde{v}^k)\|_H^2$ 可以通过

$$\|M(v^k - \tilde{v}^k)\|_H^2 = (v^k - \tilde{v}^k)^T M^T H M(v^k - \tilde{v}^k) = (v^k - \tilde{v}^k)^T M^T Q(v^k - \tilde{v}^k)$$

得到. 单调收缩关系式满足

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \frac{\gamma(2-\gamma)}{2} \alpha_k^* \|v^k - \tilde{v}^k\|_{(Q^T+Q)}^2, \quad \forall v^* \in \mathcal{V}^*.$$

凸优化分裂收缩算法的统一框架源自经典单调变分不等式的投影收缩算法. 不同的是: 投影收缩算法通过投影得到预测点并提供欧氏模下的下降方向, 分裂收缩算法则是通过求解一些子问题得到 H -模下的下降方向.

5 投影收缩算法中的孪生方向和姊妹方法

我们在经典单调变分不等式求解方面发表的计算步长的投影收缩算法 [5, 9], 被工程力学界的一些学者用来解决了一些长期困扰他们的岩土工程问题 [17, 18]. 与 §4.2 中介绍的投影收缩算法-I 相对应的投影收缩算法-II, 它们分别取孪生方向之一作为寻查方向

而取相同的步长, 得到一对姊妹方法. 根据我们的数值试验 [9] 和工程界的计算实践 [17, 18], 算法-II 要比算法-I 效率高一些. 这一节, 我们专门讲一下基于同一预测的一对孪生方向和姊妹方法.

定义 (孪生方向) 求解经典变分不等式的方法中, 假如 $d(u^k, \tilde{u}^k)$ 是一个上升方向, 则称分处不等式 (3.3) 两边的

$$\beta_k \mathbb{F}(\tilde{u}^k) \quad \text{和} \quad d(u^k, \tilde{u}^k), \quad (5.1)$$

为一对孪生方向.

我们也可以从得到关系式 (2.22) 和 (2.24) 的过程中知道, 这一对孪生方向是由基本不等式的不同组合生成的.

定义 (姊妹方法) 求解经典单调变分不等式的方法中, 由同一个预测点 \tilde{u}^k 提供了一对 (5.1) 中的孪生方向 $d(u^k, \tilde{u}^k)$ 和 $\beta_k \mathbb{F}(\tilde{u}^k)$. 用相同的步长 α , 分别由

$$u^{k+1}(\alpha) = u^k - \alpha H^{-1} d(u^k, \tilde{u}^k) \quad (5.2)$$

和

$$u^{k+1}(\alpha) = \arg \min \{ \|u - [u^k - \alpha \beta_k H^{-1} \mathbb{F}(\tilde{u}^k)]\|_H^2 \mid u \in \Omega \} \quad (5.3)$$

给出新的收缩迭代点的方法称为一对 H -模下收缩的姊妹方法.

在 H 为单位阵的时候, (5.3) 的右端是欧氏模下向量 $[u^k - \alpha \beta_k \mathbb{F}(\tilde{u}^k)]$ 到 Ω 上的投影.

我们先讨论欧氏模下依赖步长 α 的姊妹方法, 分别用带下标的

$$u_I^{k+1}(\alpha) = u^k - \alpha d(u^k, \tilde{u}^k) \quad (5.4)$$

和

$$u_{II}^{k+1}(\alpha) = P_\Omega[u^k - \alpha\beta_k \mathbf{F}(\tilde{u}^k)] \quad (5.5)$$

表示不同方法依赖步长 α 的新的迭代点. 对任意给定的 $u^* \in \Omega^*$, 我们用

$$\vartheta_k(\alpha) = \|u^k - u^*\|^2 - \|u_I^{k+1}(\alpha) - u^*\|^2 \quad (5.6)$$

和

$$\zeta_k(\alpha) = \|u^k - u^*\|^2 - \|u_{II}^{k+1}(\alpha) - u^*\|^2 \quad (5.7)$$

看成是本次迭代的进步量, 它是步长 α 的函数. 根据 §4.2 的分析 (见(4.8) - (4.10)),

$$\vartheta_k(\alpha) \geq q_k(\alpha) = 2\alpha(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2. \quad (5.8)$$

下面的定理说明, 对同样的 α , $\zeta_k(\alpha)$ 的下界不小于 $\vartheta_k(\alpha)$ 的下界.

定理 4 设 $u_{II}^{k+1}(\alpha)$ 由 (5.5) 生成. 对任意的 $\alpha > 0$, 对由 (5.7) 定义的 $\zeta_k(\alpha)$ 有

$$\zeta_k(\alpha) \geq \|u_I^{k+1}(\alpha) - u_{II}^{k+1}(\alpha)\|^2 + q_k(\alpha), \quad (5.9)$$

其中 $q_k(\alpha)$ 由 (4.10) 给出.

证明. 首先, 因为 $u_{II}^{k+1}(\alpha) = P_\Omega[u^k - \alpha\beta_k \mathbf{F}(\tilde{u}^k)]$ 和 $u^* \in \Omega$, 根据投影的性

质(见(2.5)), 我们有

$$\begin{aligned} \|u_{II}^{k+1}(\alpha) - u^*\|^2 &\leq \|u^k - \alpha\beta_k \mathbf{F}(\tilde{u}^k) - u^*\|^2 \\ &\quad - \|u^k - \alpha\beta_k \mathbf{F}(\tilde{u}^k) - u_{II}^{k+1}(\alpha)\|^2, \quad \forall u^* \in \Omega^* \end{aligned} \quad (5.10)$$

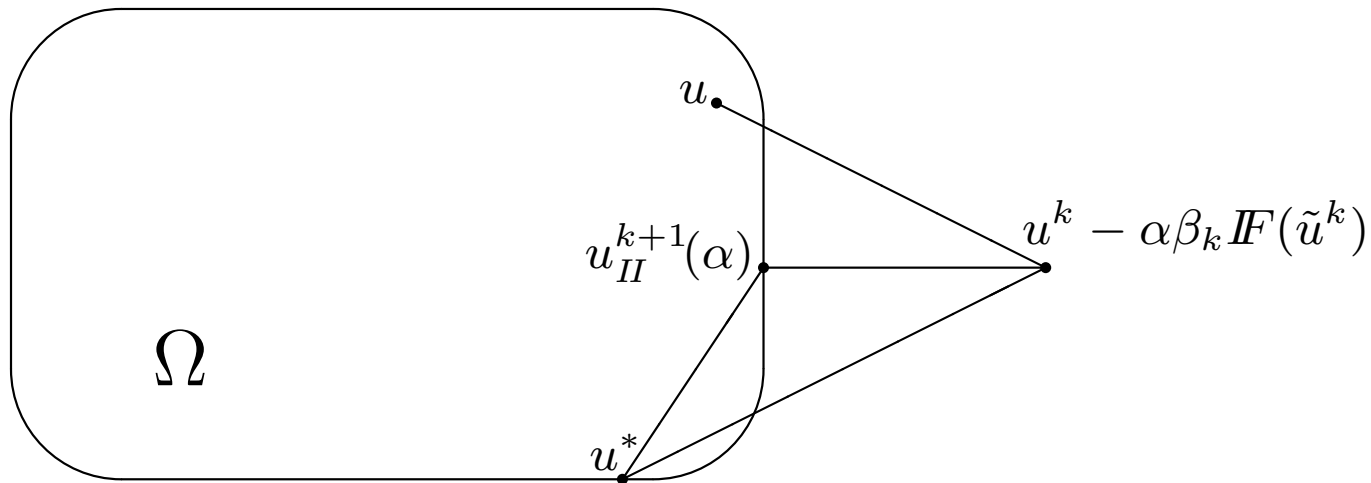


图6.1. 不等式 (5.10) 的几何解释

因此, 利用 $\zeta_k(\alpha)$ 的定义(见 (5.7)), 我们有

$$\begin{aligned}
\zeta_k(\alpha) &\geq \|u^k - u^*\|^2 - \|(u^k - u^*) - \alpha\beta_k \mathbf{F}(\tilde{u}^k)\|^2 \\
&\quad + \|(u^k - u_{II}^{k+1}(\alpha)) - \alpha\beta_k \mathbf{F}(\tilde{u}^k)\|^2 \\
&= 2\alpha(u^k - u^*)^T \beta_k \mathbf{F}(\tilde{u}^k) + 2\alpha(u_{II}^{k+1}(\alpha) - u^k)^T \beta_k \mathbf{F}(\tilde{u}^k) \\
&\quad + \|u^k - u_{II}^{k+1}(\alpha)\|^2 \\
&= \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - u^*)^T \beta_k \mathbf{F}(\tilde{u}^k). \tag{5.11}
\end{aligned}$$

将 (5.11) 中右端的最后一项 $(u_{II}^{k+1}(\alpha) - u^*)^T \beta_k \mathbf{F}(\tilde{u}^k)$ 分解成

$$(u_{II}^{k+1}(\alpha) - u^*)^T \beta_k \mathbf{F}(\tilde{u}^k) = (u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T \beta_k \mathbf{F}(\tilde{u}^k) + (\tilde{u}^k - u^*)^T \beta_k \mathbf{F}(\tilde{u}^k), \tag{5.12}$$

利用 $(\tilde{u}^k - u^*)^T \beta_k \mathbf{F}(\tilde{u}^k) \geq (\tilde{u}^k - u^*)^T \beta_k \mathbf{F}(u^*) \geq 0$, (5.12) 右端的最后一部分非负. 代入 (5.11) 的右端, 进一步得到

$$\zeta_k(\alpha) \geq \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T \beta_k \mathbf{F}(\tilde{u}^k). \tag{5.13}$$

因为 $u_{II}^{k+1}(\alpha) \in \Omega$, 用它替代 (3.3) 中的任意 $u \in \Omega$, 得到

$$(u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T \beta_k \mathbf{F}(\tilde{u}^k) \geq (u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \tag{5.14}$$

将它们代入 (5.13) 的右端, 就有

$$\zeta_k(\alpha) \geq \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \quad (5.15)$$

对上式右端, 利用 $u_I^{k+1}(\alpha)$ 和 $q_k(\alpha)$ 的表达式 (见(5.4)和(4.10)), 进一步化成

$$\begin{aligned} \zeta_k(\alpha) &\geq \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \\ &= \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - u^k)^T d(u^k, \tilde{u}^k) \\ &\quad + 2\alpha(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \\ &= \|(u^k - u_{II}^{k+1}(\alpha)) - \alpha d(u^k, \tilde{u}^k)\|^2 - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2 \\ &\quad + 2\alpha(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \\ &= \|u_I^{k+1}(\alpha) - u_{II}^{k+1}(\alpha)\|^2 + q_k(\alpha). \end{aligned}$$

这样就完成了定理结论 (5.9) 的证明. \square

定理 4 说明, $q_k(\alpha)$ 也是 $\zeta_k(\alpha)$ 的下界. 对同样的 α , $\zeta_k(\alpha)$ 优于 $\vartheta_k(\alpha)$, 除非 $u_I^{k+1}(\alpha) = u_{II}^{k+1}(\alpha)$. 在实际计算中, 这对姊妹方法分别采用校正公式

$$\text{(投影收缩算法-I)} \quad u^{k+1} = u^k - \gamma\alpha_k^* d(u^k, \tilde{u}^k) \quad (5.16)$$

和

$$\text{(投影收缩算法-II)} \quad u^{k+1} = P_\Omega[u^k - \gamma\alpha_k^* \beta_k \mathbf{F}(\tilde{u}^k)] \quad (5.17)$$

产生新的迭代点, 其中的 α_k^* 都由 (4.11) 给出. 由于 $\zeta_k(\gamma\alpha_k^*) \geq q_k(\gamma\alpha_k^*)$, 它们所产生的迭代序列 $\{u^k\}$ 都满足

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - \gamma(2 - \gamma)\alpha_k^*(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \quad (5.18)$$

事实上, 由于 $\alpha_k^* > \frac{1}{2}$, 结合关系式 (3.6), 从 (5.18) 得到

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - \frac{1}{2}\gamma(2 - \gamma)(1 - \nu)\|u^k - \tilde{u}^k\|^2. \quad (5.19)$$

这就是 §4.2 中的 (4.15). 根据上面这个不等式, 利用简单的分析术语就可以证明投影收缩算法的收敛性.

采用校正公式 (5.16) 的好处是生成 u^{k+1} 不用再做投影. 实际问题中, 到 Ω 上的投影代价往往是不高的 (例如 Ω 常常是一个正卦限或者框形), 因此常采用校正公式 (5.17). 这方面的理由我们在论文 [14] 中有更详细的说明.

求解经典单调变分不等式的投影收缩算法-II

给定 $\beta_0 = 1, \nu \in (0, 1), u^0 \in \Omega$.

For $k = 0, 1, \dots$, 假如停机准则尚未满足, **do**

$$1). \tilde{u}^k = P_{\Omega}[u^k - \beta_k \mathbf{F}(u^k)],$$

$$r_k := \frac{\beta_k \|\mathbf{F}(u^k) - \mathbf{F}(\tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|},$$

$$\mathbf{while} \quad r_k > \nu, \quad \beta_k := \frac{2}{3} \beta_k * \min\{1, \frac{1}{r_k}\},$$

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

$$r_k := \frac{\beta_k \|\mathbf{F}(u^k) - \mathbf{F}(\tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|},$$

end(while)

$$2). d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k [\mathbf{F}(u^k) - \mathbf{F}(\tilde{u}^k)],$$

$$\alpha_k^* = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2},$$

$$u^{k+1} = P_{\Omega}[u^k - \gamma \alpha_k^* \beta_k \mathbf{F}(\tilde{u}^k)],$$

$$3). \mathbf{if} \quad r_k \leq \mu \quad \mathbf{then} \quad \beta_k := \beta_k * 1.5, \quad \mathbf{end(if)}$$

4). 令 $\beta_{k+1} = \beta_k$ 和 $k := k + 1$, 开始新的一次迭代.

从投影收缩算法-I到投影收缩算法-II, 只是将

$$u^{k+1} = u^k - \gamma\alpha_k^* d(u^k, \tilde{u}^k) \quad \text{改成了} \quad u^{k+1} = P_\Omega[u^k - \gamma\alpha_k^* \beta_k \mathbf{F}(\tilde{u}^k)].$$

论文 [17, 18] 中解决岩土工程问题用的就是这一节的方法. 根据我们的数值经验 [9] 和他们的计算实践, 算法-II的效率比算法-I的效率高一些.

在条件 (3.2) 满足的合格预测 (3.1) 的基础上, 校正也可以用

$$u^{k+1} = P_\Omega[u^k - \beta_k \mathbf{F}(\tilde{u}^k)]$$

实现. 这时候, 预测-校正写在一起就是

$$\begin{cases} \tilde{u}^k & = P_\Omega[u^k - \beta_k \mathbf{F}(u^k)], \\ u^{k+1} & = P_\Omega[u^k - \beta_k \mathbf{F}(\tilde{u}^k)]. \end{cases} \quad (5.20)$$

这就是人们所说的 Korpelevich 外梯度法 [15].

凸优化的分裂收缩算法的预测, 同样提供了一对孪生方向. 如何借鉴投影收缩算法中的孪生方向和姊妹方法, 我们在 [8] 中开展了一些讨论.

6 投影收缩算法在求解可分离凸优化上的应用

作为投影收缩算法在求解可分离凸优化上的应用, 我们考虑更一般两块可分离凸优化问题

$$\min \{ \theta_1(x) + \theta_2(y) \mid Ax + By = b \text{ (或 } \geq b), x \in \mathcal{X}, y \in \mathcal{Y} \}. \quad (6.1)$$

假设问题有解并且 $\theta_1(x)$ 和 $\theta_2(y)$ 分别在包含凸闭集 \mathcal{X} 和 \mathcal{Y} 的一个开集上可微. 问题 (6.1) 的拉格朗日函数是定义在 $\mathcal{X} \times \mathcal{Y} \times \Lambda$ 上的

$$L(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T (Ax + By - b),$$

其中

$$\Lambda = \begin{cases} \mathfrak{R}^m, & \text{if } Ax + By = b, \\ \mathfrak{R}_+^m, & \text{if } Ax + By \geq b. \end{cases}$$

拉格朗日函数的鞍点 $(x^*, y^*, \lambda^*) \in \mathcal{X} \times \mathcal{Y} \times \Lambda$ 满足不等式

$$L(x^*, y^*, \lambda) \leq L(x^*, y^*, \lambda^*) \leq L(x, y, \lambda^*), \quad \forall (x, y, \lambda) \in \mathcal{X} \times \mathcal{Y} \times \Lambda.$$

这意味着

$$\begin{cases} x^* \in \mathcal{X}, & L(x, y^*, \lambda^*) \geq L(x^*, y^*, \lambda^*) & \forall x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, & L(x^*, y, \lambda^*) \geq L(x^*, y^*, \lambda^*) & \forall y \in \mathcal{Y}, \\ \lambda^* \in \Lambda, & L(x^*, y^*, \lambda^*) \geq L(x^*, y^*, \lambda) & \forall \lambda \in \Lambda. \end{cases}$$

也就是说

$$\begin{cases} x^* \in \arg \min \{ \theta_1(x) + \theta_2(y^*) - (\lambda^*)^T (Ax + By^* - b) \mid x \in \mathcal{X} \}, \\ y^* \in \arg \min \{ \theta_1(x^*) + \theta_2(y) - (\lambda^*)^T (Ax^* + By - b) \mid y \in \mathcal{Y} \}, \\ \lambda^* \in \arg \max \{ \theta_1(x^*) + \theta_2(y^*) - \lambda^T (Ax^* + By^* - b) \mid \lambda \in \Lambda \}. \end{cases}$$

当 $\theta_1(x)$, $\theta_2(y)$ 可微时, 根据最优性质定理, 我们有

$$\begin{cases} x^* \in \mathcal{X}, & (x - x^*)^T (\nabla \theta_1(x^*) - A^T \lambda^*) \geq 0, & \forall x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, & (y - y^*)^T (\nabla \theta_2(y^*) - B^T \lambda^*) \geq 0, & \forall y \in \mathcal{Y}, \\ \lambda^* \in \Lambda, & (\lambda - \lambda^*)^T (Ax^* + By^* - b) \geq 0, & \forall \lambda \in \Lambda. \end{cases}$$

用记号

$$\nabla \theta_1(x) = f(x), \quad \nabla \theta_2(y) = g(y),$$

相应的紧凑形式的单调变分不等式就是

$$w^* \in \Omega, \quad (w - w^*)^T \mathbf{F}(w^*) \geq 0, \quad \forall w \in \Omega, \quad (6.2a)$$

其中

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} f(x) - A^T \lambda \\ g(y) - B^T \lambda \\ Ax + By - b \end{pmatrix}, \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Lambda. \quad (6.2b)$$

假如交替方向法中子问题求解比较困难, 而 $\theta_1(x)$ 和 $\theta_2(y)$ 的梯度随手可得并且 Lipschitz 连续, 可以考虑用投影收缩算法求解可微的线性约束凸优化问题等价的变分不等式 (6.2). 对给定的 $w^k = (x^k, y^k, \lambda^k)$, 通过

$$\begin{cases} \tilde{x}^k = P_{\mathcal{X}} \left\{ x^k - \frac{1}{r} [f(x^k) - A^T \lambda^k] \right\}, & (6.3a) \\ \tilde{y}^k = P_{\mathcal{Y}} \left\{ y^k - \frac{1}{s} [g(y^k) - B^T \lambda^k] \right\}, & (6.3b) \\ \tilde{\lambda}^k = P_{\Lambda} \left\{ \lambda^k - \beta (Ax^k + By^k - b) \right\}, & (6.3c) \end{cases}$$

得到预测点 $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$. 其中的 $r, s > 0$ 是适当选取的正常数使得

$$\|f(x^k) - f(\tilde{x}^k)\| \leq \nu r \|x^k - \tilde{x}^k\|, \quad \text{和} \quad \|g(y^k) - g(\tilde{y}^k)\| \leq \nu s \|y^k - \tilde{y}^k\|. \quad (6.4)$$

当 $f(x)$ 和 $g(y)$ Lipschitz 连续时, 这是可以办到的.

下面对预测 (6.3) 进行分析. 通过投影 (6.3a) 得到的 \tilde{x}^k 是极小化问题

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

的解, 根据最优性条件的定理, 有

$$\tilde{x}^k \in \mathcal{X} \quad (x - \tilde{x}^k)^T \left\{ \tilde{x}^k - \left[x^k - \frac{1}{r} [f(x^k) - A^T \lambda^k] \right] \right\} \geq 0, \quad \forall x \in \mathcal{X}.$$

这可以写成

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

利用 (6.2), 我们有

$$\tilde{x}^k \in \mathcal{X}, \quad (x - \tilde{x}^k)^T \left\{ \begin{array}{l} \frac{f(\tilde{x}^k) - A^T \tilde{\lambda}^k + A^T (\tilde{\lambda}^k - \lambda^k)}{+ [r(\tilde{x}^k - x^k) - (f(\tilde{x}^k) - f(x^k))]} \end{array} \right\} \geq 0, \quad \forall x \in \mathcal{X}. \quad (6.5a)$$

通过投影 (6.3b) 得到的 \tilde{y}^k , 有

$$\tilde{y}^k \in \mathcal{Y}, \quad (y - \tilde{y}^k)^T \left\{ \begin{array}{l} \frac{g(\tilde{y}^k) - B^T \tilde{\lambda}^k + B^T (\tilde{\lambda}^k - \lambda^k)}{+ [s(\tilde{y}^k - y^k) - (g(\tilde{y}^k) - g(y^k))]} \end{array} \right\} \geq 0, \quad \forall y \in \mathcal{Y}. \quad (6.5b)$$

根据 (6.3c) 得到的 $\tilde{\lambda}^k$, 我们有

$$\tilde{\lambda}^k \in \mathfrak{R}^m, \quad (\lambda - \tilde{\lambda}^k)^T \left\{ \begin{array}{l} \frac{(A\tilde{x}^k + B\tilde{y}^k - b) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)}{-A(\tilde{x}^k - x^k) - B(\tilde{y}^k - y^k)} \end{array} \right\} \geq 0, \quad \forall \lambda \in \mathfrak{R}^m. \quad (6.5c)$$

这样写,是为了让上面三个式子中下波纹线的部分合在一起就成了(6.2)中的 $\mathcal{F}(\tilde{w}^k)$.

引理 3 从给定的 $w^k = (x^k, y^k, \lambda^k)$ 出发,由(6.3)产生的预测点 $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ 满足

$$\tilde{w}^k \in \Omega, \quad (w - \tilde{w}^k)^T \mathcal{F}(\tilde{w}^k) \geq (w - \tilde{w}^k)^T d(w^k, \tilde{w}^k), \quad \forall w \in \Omega \quad (6.6a)$$

其中

$$d(w^k, \tilde{w}^k) = \begin{pmatrix} [r(x^k - \tilde{x}^k) - (f(x^k) - f(\tilde{x}^k))] + A^T(\lambda^k - \tilde{\lambda}^k) \\ [s(y^k - \tilde{y}^k) - (g(y^k) - g(\tilde{y}^k))] + B^T(\lambda^k - \tilde{\lambda}^k) \\ \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) - A(x^k - \tilde{x}^k) - B(y^k - \tilde{y}^k) \end{pmatrix}. \quad (6.6b)$$

证明 将(6.5)的三部分写在一起并利用 $\mathcal{F}(w)$ 的定义经整理便得引理之结论. \square

下面我们证明由(6.6b)给出的 $d(w^k, \tilde{w}^k)$ 是上升方向. 将(6.6)中任意的 w 设为属于 Ω^* 的 w^* ,则有

$$(\tilde{w}^k - w^*)^T d(w^k, \tilde{w}^k) \geq (\tilde{w}^k - w^*)^T \mathcal{F}(\tilde{w}^k). \quad (6.7)$$

利用 \mathcal{F} 的单调性和 w^* 的最优性,

$$(\tilde{w}^k - w^*)^T \mathcal{F}(\tilde{w}^k) \geq (\tilde{w}^k - w^*)^T \mathcal{F}(w^*) \geq 0.$$

因此

$$(w^k - w^*)^T d(w^k, \tilde{w}^k) \geq (w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k). \quad (6.8)$$

利用 $d(w^k, \tilde{w}^k)$ 的表达式 (6.6b) 和预测条件 (6.4) 得到

$$\begin{aligned}
& (w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k) \\
&= r\|x^k - \tilde{x}^k\|^2 - (x^k - \tilde{x}^k)^T (f(x^k) - f(\tilde{x}^k)) \\
&\quad + s\|y^k - \tilde{y}^k\|^2 - (y^k - \tilde{y}^k)^T (g(y^k) - g(\tilde{y}^k)) + \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 \\
&\geq (1 - \nu)(r\|x^k - \tilde{x}^k\|^2 + s\|y^k - \tilde{y}^k\|^2) + \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2. \tag{6.9}
\end{aligned}$$

由于 $\nu \in (0, 1)$, 上式右端大于 0. 这里的 (6.8) 和 (6.9) 相当于 §3 中的 (3.5) 和 (3.6). 因此, $d(w^k, \tilde{w}^k)$ 是距离函数 $\|w - w^*\|^2$ 的上升方向.

由于 $d(w^k, \tilde{w}^k)$ 是距离函数 $\|w - w^*\|^2$ 的上升方向, 根据 (6.6), 分处不等式 (6.6a) 两边的

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

为一对孪生方向. 我们可以用 H -模下的姊妹方法进行校正. 对给定的正定矩阵 H , 校正公式-I 通过

$$\text{(校正公式-I)} \quad w^{k+1} = w^k - \gamma\alpha_k^* H^{-1} d(w^k, \tilde{w}^k), \quad \gamma \in (0, 2) \tag{6.10a}$$

产生新的迭代点. 其中

$$\alpha_k^* = \frac{(w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k)}{\|H^{-1} d(w^k, \tilde{w}^k)\|_H^2}. \tag{6.10b}$$

定理 5 求解变分不等式 (6.2), 由 (6.3) 预测和 (6.10) 校正产生的序列 $\{\tilde{w}^k\}$ 和 $\{w^k\}$ 满足

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \gamma(2-\gamma)\alpha_k^* (w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k), \quad \forall w \in \Omega, \quad (6.11)$$

其中 $d(w^k, \tilde{w}^k)$ 由 (6.6b) 给出.

证明 先将 (6.10a) 中的 $\gamma\alpha_k^*$ 置为任意的 $\alpha > 0$, 并将输出记为 $w_I^{k+1}(\alpha)$, 并记

$$\vartheta_k^H(\alpha) = \|w^k - w^*\|_H^2 - \|w_I^{k+1}(\alpha) - w^*\|_H^2. \quad (6.12)$$

这样

$$\begin{aligned} \vartheta_k^H(\alpha) &\stackrel{(6.10a)}{=} \|w^k - w^*\|_H^2 - \|(w^k - w^*) - \alpha H^{-1} d(w^k, \tilde{w}^k)\|_H^2 \\ &= 2\alpha (w^k - w^*)^T d(w^k, \tilde{w}^k) - \alpha^2 \|H^{-1} d(w^k, \tilde{w}^k)\|_H^2 \\ &\stackrel{(6.8)}{\geq} 2\alpha (w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k) - \alpha^2 \|H^{-1} d(w^k, \tilde{w}^k)\|_H^2 \\ &=: q_k^H(\alpha). \end{aligned} \quad (6.13)$$

对二次函数 $q_k^H(\alpha)$ 求极值得到 (6.10b) 中的 α_k^* . 当 (6.13) 中的 $\alpha = \gamma\alpha_k^*$ 时,

$$\begin{aligned} \vartheta_k^H(\gamma\alpha_k^*) &\geq q_k^H(\gamma\alpha_k^*) \\ &= 2\gamma\alpha_k^* (w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k) - \gamma^2 (\alpha_k^*)^2 \|H^{-1} d(w^k, \tilde{w}^k)\|_H^2 \\ &= \gamma(2-\gamma)\alpha_k^* (w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k). \end{aligned}$$

证明的最后一个等式使用了 (6.10b) 中

的 $\alpha_k^* \|H^{-1}d(w^k, \tilde{w}^k)\|_H^2 = (w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k)$. 定理结论得证. \square

采用 (6.10) 校正, 理论上可以取维数相配的任何正定的矩阵. 对 (6.6) 给出的 $d(w^k, \tilde{w}^k)$, 建议取

$$H = \begin{pmatrix} rI & 0 & 0 \\ 0 & sI & 0 \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}, \quad (6.14)$$

这时, H 是正定的数量对角矩阵,

$$H^{-1}d(w^k, \tilde{w}^k) = \begin{pmatrix} (x^k - \tilde{x}^k) - \frac{1}{r} [(f(x^k) - f(\tilde{x}^k)) - A^T(\lambda^k - \tilde{\lambda}^k)] \\ (y^k - \tilde{y}^k) - \frac{1}{s} [(g(y^k) - g(\tilde{y}^k)) - B^T(\lambda^k - \tilde{\lambda}^k)] \\ (\lambda^k - \tilde{\lambda}^k) - \beta A(x^k - \tilde{x}^k) - \beta B(y^k - \tilde{y}^k) \end{pmatrix}.$$

对应于校正公式 (6.10), 我们也可以采用姊妹方法中的另一种方法校正, 由

$$\text{(校正公式-II)} \quad w^{k+1} = \arg \min \{ \|w - [w^k - \gamma \alpha_k^* H^{-1} F(\tilde{w}^k)]\|_H^2 \mid w \in \Omega \} \quad (6.15)$$

产生新的迭代点 w^{k+1} , 其中 $\gamma \in (0, 2)$, α_k^* 由 (6.10b) 提供, 可以得到跟校正方法 (6.10) 同样的收缩性质.

定理 6 求解变分不等式 (6.2), 由 (6.3) 预测和 (6.15) 校正产生的序列 $\{\tilde{w}^k\}$ 和 $\{w^k\}$ 满足

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \gamma(2-\gamma)\alpha_k^* (w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k), \quad \forall w \in \Omega, \quad (6.16)$$

其中 $d(w^k, \tilde{w}^k)$ 由 (6.6b) 给出.

证明. 先将 (6.15) 中的 $\gamma\alpha_k^*$ 置为任意的 $\alpha > 0$ 并将输出记为 $w_{II}^{k+1}(\alpha)$. 我们考察

$$\zeta_k^H(\alpha) = \|w^k - w^*\|_H^2 - \|w_{II}^{k+1}(\alpha) - w^*\|_H^2 \quad (6.17)$$

首先, 因为 $w_{II}^{k+1}(\alpha) = \arg \min \{ \|w - [w^k - \alpha H^{-1} \mathbf{F}(\tilde{w}^k)]\|_H^2 \mid w \in \Omega \}$, 根据最优性质定理, 有

$$(w - w_{II}^{k+1}(\alpha))^T H \{ w_{II}^{k+1}(\alpha) - w^k + \alpha H^{-1} \mathbf{F}(\tilde{w}^k) \} \geq 0, \quad \forall w \in \Omega.$$

将上面任意的 w 替换成 w^* , 就有不等式

$$(w_{II}^{k+1}(\alpha) - w^*)^T H \{ [w^k - \alpha H^{-1} \mathbf{F}(\tilde{w}^k)] - w_{II}^{k+1}(\alpha) \} \geq 0. \quad (6.18)$$

在恒等式

$$(a - b)^T H(c - a) = \frac{1}{2} (\|c - b\|_H^2 - \|c - a\|_H^2) - \frac{1}{2} \|a - b\|_H^2$$

中置

$$a = w_{II}^{k+1}(\alpha), \quad b = w^* \quad \text{和} \quad c = w^k - \alpha H^{-1} \mathbf{F}(\tilde{w}^k)$$

并利用 (6.18), 则有

$$\begin{aligned} & \|w^k - \alpha H^{-1} \mathbf{F}(\tilde{w}^k) - w^*\|_H^2 \\ & - \|w^k - \alpha H^{-1} \mathbf{F}(\tilde{w}^k) - w_{II}^{k+1}(\alpha)\|_H^2 - \|w_{II}^{k+1}(\alpha) - w^*\|_H^2 \geq 0. \end{aligned}$$

因此,

$$\begin{aligned} \|w_{II}^{k+1}(\alpha) - w^*\|_H^2 &\leq \|(w^k - w^*) - \alpha H^{-1} \mathbf{F}(\tilde{w}^k)\|_H^2 \\ &\quad - \|(w^k - w_{II}^{k+1}(\alpha)) - \alpha H^{-1} \mathbf{F}(\tilde{w}^k)\|_H^2. \end{aligned}$$

将此代入 (6.17), 我们有

$$\begin{aligned} \zeta_k^H(\alpha) &\geq \|w^k - w^*\|_H^2 - \|(w^k - w^*) - \alpha H^{-1} \mathbf{F}(\tilde{w}^k)\|_H^2 \\ &\quad + \|(w^k - w_{II}^{k+1}(\alpha)) - \alpha H^{-1} \mathbf{F}(\tilde{w}^k)\|_H^2 \\ &= 2\alpha(w^k - w^*)^T \mathbf{F}(\tilde{w}^k) + 2\alpha(w_{II}^{k+1}(\alpha) - w^k)^T \mathbf{F}(\tilde{w}^k) \\ &\quad + \|w^k - w_{II}^{k+1}(\alpha)\|_H^2 \\ &= \|w^k - w_{II}^{k+1}(\alpha)\|_H^2 + 2\alpha(w_{II}^{k+1}(\alpha) - w^*)^T \mathbf{F}(\tilde{w}^k). \end{aligned} \quad (6.19)$$

将 (6.19) 中右端的最后一项 $(w_{II}^{k+1}(\alpha) - w^*)^T \mathbf{F}(\tilde{w}^k)$ 分解成

$$(w_{II}^{k+1}(\alpha) - w^*)^T \mathbf{F}(\tilde{w}^k) = (w_{II}^{k+1}(\alpha) - \tilde{w}^k)^T \mathbf{F}(\tilde{w}^k) + (\tilde{w}^k - w^*)^T \mathbf{F}(\tilde{w}^k),$$

利用

$$(\tilde{w}^k - w^*)^T \mathbf{F}(\tilde{w}^k) \geq (\tilde{w}^k - w^*)^T \mathbf{F}(w^*) \geq 0,$$

上式右端的最后一部分非负. 代入 (6.19) 的右端, 进一步得到

$$\zeta_k^H(\alpha) \geq \|w^k - w_{II}^{k+1}(\alpha)\|_H^2 + 2\alpha(w_{II}^{k+1}(\alpha) - \tilde{w}^k)^T \mathbf{F}(\tilde{w}^k). \quad (6.20)$$

因为 $w_{II}^{k+1}(\alpha) \in \Omega$, 用它替代 (6.6a) 中的任意 $w \in \Omega$, 得到

$$(w_{II}^{k+1}(\alpha) - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w_{II}^{k+1}(\alpha) - \tilde{w}^k)^T d(w^k, \tilde{w}^k). \quad (6.21)$$

将它们代入 (6.20) 的右端, 就有

$$\zeta_k^H(\alpha) \geq \|w^k - w_{II}^{k+1}(\alpha)\|_H^2 + 2\alpha(w_{II}^{k+1}(\alpha) - \tilde{w}^k)^T d(w^k, \tilde{w}^k).$$

对上式右端, 进一步化成

$$\begin{aligned} \zeta_k^H(\alpha) &\geq \|w^k - w_{II}^{k+1}(\alpha)\|_H^2 + 2\alpha(w_{II}^{k+1}(\alpha) - \tilde{w}^k)^T d(w^k, \tilde{w}^k) \\ &= \|w^k - w_{II}^{k+1}(\alpha)\|_H^2 + 2\alpha(w_{II}^{k+1}(\alpha) - w^k)^T d(w^k, \tilde{w}^k) \\ &\quad + 2\alpha(w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k) \\ &= \|(w^k - w_{II}^{k+1}(\alpha)) - \alpha H^{-1} d(w^k, \tilde{w}^k)\|_H^2 - \alpha^2 \|H^{-1} d(w^k, \tilde{w}^k)\|_H^2 \\ &\quad + 2\alpha(w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k) \\ &= \|w_I^{k+1}(\alpha) - w_{II}^{k+1}(\alpha)\|_H^2 - \alpha^2 \|H^{-1} d(w^k, \tilde{w}^k)\|_H^2 \\ &\quad + 2\alpha(w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k) \\ &\geq 2\alpha(w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k) - \alpha^2 \|H^{-1} d(w^k, \tilde{w}^k)\|_H^2. \end{aligned} \quad (6.22)$$

在上式中取 $\alpha = \gamma\alpha_k^*$ 并利用 (6.10b) 中的

$$\alpha_k^* \|H^{-1} d(w^k, \tilde{w}^k)\|_H^2 = (w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k),$$

从(6.22)得到

$$\zeta_k^H(\gamma\alpha_k^*) \geq \gamma(2-\gamma)\alpha_k^*(w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k).$$

这样就完成了定理结论(6.16)的证明. \square

我们关心校正公式-II(6.15)如何实现. 由于(6.14)中的 H 是与 $\Omega = \mathcal{X} \times \mathcal{Y} \times \Lambda$ 相对应的分块数量矩阵. 利用可分离结构, 我们有

$$\begin{aligned} & \min\{\|w - [w^k - \alpha_k H^{-1} F(\tilde{w}^k)]\|_H^2 \mid w \in \Omega\} \\ &= \min \left\{ \left\| \begin{array}{l} x - [x^k - \alpha_k \frac{1}{r}(f(\tilde{x}^k) - A^T \tilde{\lambda}^k)] \\ y - [y^k - \alpha_k \frac{1}{s}(g(\tilde{y}^k) - B^T \tilde{\lambda}^k)] \\ \lambda - [\lambda^k - \alpha_k \beta(A\tilde{x}^k + B\tilde{y}^k - b)] \end{array} \right\|_H^2 \mid \begin{array}{l} x \in \mathcal{X} \\ y \in \mathcal{Y} \\ \lambda \in \Lambda \end{array} \right\} \\ &= \min \left\{ \begin{array}{l} r\|x - [x^k - \alpha_k \frac{1}{r}(f(\tilde{x}^k) - A^T \tilde{\lambda}^k)]\|^2 \mid x \in \mathcal{X} \\ +s\|y - [y^k - \alpha_k \frac{1}{s}(g(\tilde{y}^k) - B^T \tilde{\lambda}^k)]\|^2 \mid y \in \mathcal{Y} \\ +\frac{1}{\beta}\|\lambda - [\lambda^k - \alpha_k \beta(A\tilde{x}^k + B\tilde{y}^k - b)]\|^2 \mid \lambda \in \Lambda \end{array} \right\}. \end{aligned}$$

因此,

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \operatorname{argmin}\{\|x - [x^k - \alpha_k \frac{1}{r}(f(\tilde{x}^k) - A^T \tilde{\lambda}^k)]\|^2 \mid x \in \mathcal{X}\} \\ \operatorname{argmin}\{\|y - [y^k - \alpha_k \frac{1}{s}(g(\tilde{y}^k) - B^T \tilde{\lambda}^k)]\|^2 \mid y \in \mathcal{Y}\} \\ \operatorname{argmin}\{\|\lambda - [\lambda^k - \alpha_k \beta(A\tilde{x}^k + B\tilde{y}^k - b)]\|^2 \mid \lambda \in \Lambda\} \end{pmatrix}.$$

换句话说,

$$\begin{aligned} w^{k+1} &= \operatorname{arg min}\{\|w - [w^k - \alpha_k H^{-1} \mathbf{F}(\tilde{w}^k)]\|_H^2 \mid w \in \Omega\} \\ &= \operatorname{arg min}\{\|w - [w^k - \alpha_k H^{-1} \mathbf{F}(\tilde{w}^k)]\|^2 \mid w \in \Omega\} \\ &= P_\Omega[w^k - \alpha_k H^{-1} \mathbf{F}(\tilde{w}^k)]. \end{aligned} \quad (6.23)$$

这里, 我们需要再一次强调, 上式只有当 H 是形如 (6.14) 的正定分块数量矩阵时才成立. 因此, 求解变分不等式 (6.2), 由 (6.3) 预测, 通过 (6.15) 校正是由

$$\begin{cases} x^{k+1} = P_{\mathcal{X}}\left\{x^k - \frac{\alpha_k}{r}[f(\tilde{x}^k) - A^T \tilde{\lambda}^k]\right\}, & (6.24a) \\ y^{k+1} = P_{\mathcal{Y}}\left\{y^k - \frac{\alpha_k}{s}[g(\tilde{y}^k) - B^T \tilde{\lambda}^k]\right\}, & (6.24b) \\ \lambda^{k+1} = P_{\Lambda}\left\{\lambda^k - \alpha_k \beta(A\tilde{x}^k + B\tilde{y}^k - b)\right\}, & (6.24c) \end{cases}$$

实现的. 运算执行的主要是分别在 \mathcal{X} , \mathcal{Y} 和 Λ 上的欧氏模下的投影.

References

- [1] B.S. He, A projection and contraction method for a class of linear complementarity problems and its application in convex quadratic programming, *Applied Mathematics and Optimization*, **25**, 247–262, 1992.
- [2] 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.
- [3] B.S. He, A new method for a class of linear variational inequalities, *Math. Progr.*, **66**, 137-144, 1994.
- [4] B.S. He, Solving a class of linear projection equations, *Numerische Mathematik*, **68**, 71-80, 1994.
- [5] B.S. He, A class of projection and contraction methods for monotone variational inequalities, *Applied Mathematics and Optimization*, **35**, 69-76, 1997.
- [6] 何炳生. 从变分不等式的投影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.
- [7] 何炳生. 我和乘子交替方向法 20 年. *运筹学学报*, 2018, 22: 1–31.
- [8] 何炳生. 凸优化和单调变分不等式收缩算法的统一框架. *中国科学：数学*, 2018, 48: 255–272.
- [9] He B S and Liao L Z. Improvements of some projection methods for monotone nonlinear variational inequalities, *Journal of Optimization Theory and Applications*, 2002, 112: 111–128.
- [10] 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.*, 2012, 51: 649-679.

- [11] 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.*, 2012, 51: 681-708.
- [12] B. S. He and X. M. Yuan, On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers, *Numerische Mathematik*, 130 (2015), 567-577.
- [13] B. S. He and X. M. Yuan, On construction of splitting contraction algorithms in a prediction-correction framework for separable convex optimization, [arXiv 2204.11522 \[math.OC\]](https://arxiv.org/abs/2204.11522)
- [14] 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.*, 2004, 27: 247-267.
- [15] G. M. Korpelevich. The extragradient method for finding saddle points and other problems, *Ekonomika i Matematicheskie Metody*, 1976, 12: 747-756.
- [16] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Verlag, New York, 1999.
- [17] H. Zheng, F. Liu and X.L. Du, Complementarity problem arising from static growth of multiple cracks and MLS-based numerical manifold method, *Computer Methods in Applied Mechanics and Engineering*, 2015, 295: 150–171.
- [18] H. Zheng, P. Zhang and X.L. Du, Dual form of discontinuous deformation analysis, *Computer Methods in Applied Mechanics and Engineering*, 2016, 305: 196–216.
- [19] T. Zhu and Z. G. Yu, A simple proof for some important properties of the projection mapping. *Math. Inequal. Appl.* **7**, 453–456, 2004.