中国科学: 数学 2018 年 第 48 卷 第 2 期: 255～272
SCIENTIA SINICA Mathematica
综 述

凸优化和单调变分不等式收缩算法的统一框架

何炳生[1,2]

1. 南方科技大学数学系, 深圳 518055;
2. 南京大学数学系, 南京 210093
E-mail: hebma@nju.edu.cn

收稿日期: 2017-02-17; 接受日期: 2017-08-16; 网络出版日期: 2017-11-10
国家自然科学基金 (批准号: 11471156) 资助项目

摘要 线性约束的凸优化问题可以转化成一个形式更一般的单调变分不等式. 在变分不等式的框架下研究最优化问题的求解方法, 就像微积分中利用导数求函数的极值, 常常会带来很大的方便. 求解单调变分不等式的投影收缩算法有一个预测 - 校正的统一框架, 基于 "孪生方向和相同步长" 有两类花费几乎相当的算法, 计算实践证明第二类算法效率往往更高. 近年发展起来并被广泛采用的凸规划的分裂收缩算法属于一个更一般的框架, 这个框架中的预测同样提供了一对孪生方向. 迄今为止的凸规划的分裂收缩算法, 都相当于变分不等式投影收缩算法中的第一类算法. 本文指出, 利用现有的步长法则, 配上孪生方向中的另一个方向, 同样可以构造相应的第二类算法. 本文在统一框架下证明了两类算法的 $O(1/t)$ 迭代复杂性.

关键词 凸优化 单调变分不等式 投影收缩算法 分裂收缩算法 统一框架 孪生方向和相同步长

MSC (2010) 主题分类 90C25, 90C30, 90C33

1 引言

对一般形式的单调变分不等式, 本文给出一个由预测 - 校正组成的迭代算法的统一框架. 本节先介绍问题、算法和收敛性质的概貌, 然后再在以后各节详细论述与证明.

1.1 单调变分不等式

设 $\Omega \subset \mathbb{R}^l$ 是非空闭凸集, $u \in \mathbb{R}^n$, $w \in \mathbb{R}^l$ $(n \leqslant l)$, $\theta(u) : \mathbb{R}^n \to \mathbb{R}$ 是凸函数, $F(w) : \mathbb{R}^l \to \mathbb{R}^l$ 是单调算子. 本文考虑如下的变分不等式:

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

英文引用格式: He B S. A uniform framework of contraction methods for convex optimization and monotone variational inequality (in Chinese). Sci Sin Math, 2018, 48: 255–272, doi: 10.1360/N012017-00034

© 2017《中国科学》杂志社

上述问题中, u 可以是 w 本身, 也可以是 w 的部分分量. 当 $u \neq w$ 时, 通常用 u 表示向量 w 的排在前面的部分分量. 我们说 F 单调, 是指

$$(w-\bar{w})^{\mathrm{T}}(F(w)-F(\bar{w})) \geqslant 0, \quad \forall\, w, \bar{w} \in \Omega \text{ (或 } \mathbb{R}^l).$$

我们还设算子 F 是 Lipschitz 连续的, 即存在常数 $L>0$, 使得

$$\|F(w)-F(\bar{w})\| \leqslant L\|w-\bar{w}\|, \quad \forall\, w, \bar{w}.$$

不过, 我们并不要求知道 L 的大小. 此外, 通篇总设问题 (1.1) 有解, 记解集为 Ω^*. 变分不等式在管理科学[1)] 和工程问题[1–3] 上都有广泛应用, 其他方面的应用也可以参见文献 [4].

1.2 预测点和孪生方向

如果求解 (1.1) 的算法的第 k- 次迭代, 只需要有 w^k 的部分分量 v^k 就可以计算得到新的迭代点 w^{k+1}, 那么称 v 为 w 的核心变量 (essential variables), 称余下的部分向量 $w\backslash v$ 为中间变量 (intermediate variables). v 也可以是 w 本身. 当 $v \neq w$ 时, 通常用 v 表示向量 w 的排在后面的部分分量.

定义 1.1 (预测和孪生方向) 设 $\mathbb{A}$ 是求解变分不等式 (1.1) 的一个迭代方法, $\delta>0$ 和 $K>0$ 是给定的常数. 算法 $\mathbb{A}$ 的 k 次迭代从给定的 v^k 开始, 给出满足

$$\tilde{w}^k \in \Omega, \quad \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^{\mathrm{T}}F(\tilde{w}^k) \geqslant (v-\tilde{v}^k)^{\mathrm{T}}d(v^k,\tilde{v}^k), \quad \forall\, w \in \Omega \tag{1.2}$$

的向量 $\tilde{w}^k$, 同时向量 $d(v^k,\tilde{v}^k)$ 满足

$$\|d(v^k,\tilde{v}^k)\| \leqslant K\|v^k-\tilde{v}^k\| \quad \text{和} \quad (v^k-\tilde{v}^k)^{\mathrm{T}}d(v^k,\tilde{v}^k) \geqslant \delta\|v^k-\tilde{v}^k\|^2, \tag{1.3}$$

则称 $\tilde{w}^k$ 是 k 次迭代的一个合格的预测点. 这时称 (1.2) 两边的

$$d(v^k,\tilde{v}^k) \quad \text{和} \quad F(\tilde{w}^k) \tag{1.4}$$

为预测提供的欧氏模下一对孪生方向.

1.3 校正公式和孪生方法的收缩性质

设 H 是可以与向量 v 相乘的正定矩阵. 在矩阵 H 的基础上, 令

$$\mathcal{H} = \begin{pmatrix} hI & 0 \\ 0 & H \end{pmatrix}, \tag{1.5}$$

其中 $h>0$, hI 是与中间变量所对应的数量矩阵. 换句话说, 当 $w=v$ 时, $\mathcal{H}=H$. 收缩算法的校正是利用预测提供的方向, 选取适当的步长, 使得新的迭代点在 H 模 (或欧氏模) 意义下离解集更近一些.

定义 1.2 (校正和孪生方法) 迭代算法 $\mathbb{A}$ 用预测提供的孪生方向 (1.4), 通过

$$v^{k+1} = v^k - \alpha_k H^{-1} d(v^k,\tilde{v}^k), \tag{1.6a}$$

1) 何炳生. 凸优化和单调变分不等式的收缩算法. Http://maths.nju.edu.cn/~hebma 中的系列讲义.

或者

$$w^{k+1} = \arg\min\left\{\alpha_k\theta(u) + \frac{1}{2}\|w - [w^k - \alpha_k\mathcal{H}^{-1}F(\tilde{w}^k)]\|_{\mathcal{H}}^2 \;\middle|\; w \in \Omega\right\} \tag{1.6b}$$

产生新的迭代点, 其中

$$\alpha_k = \gamma\alpha_k^*, \quad \alpha_k^* = \frac{(v^k - \tilde{v}^k)^{\mathrm{T}} d(v^k, \tilde{v}^k)}{\|H^{-1}d(v^k, \tilde{v}^k)\|_H^2}, \quad \gamma \in (0, 2). \tag{1.7}$$

称 (1.6a) 和 (1.6b) 分别为第一和二类校正方法. 由同一预测 (1.2), 配上校正 (1.6a) 或者 (1.6b) 的预测 - 校正方法, 称为一对孪生方法.

性质 1.1 (主要收敛性质) 我们将会证明两种算法产生的迭代序列 $\{v^k\}$ 都满足收缩性质

$$\|v^{k+1} - v^*\|_H^2 \leqslant \|v^k - v^*\|_H^2 - \gamma(2-\gamma)\alpha_k^*\delta\|v^k - \tilde{v}^k\|^2, \quad \forall\, v^* \in \mathcal{V}^*, \tag{1.8}$$

其中 $\mathcal{V}^*$ 是 Ω^* 中对应 v 的那部分; 并且证明算法都具有 $O(1/t)$ 的迭代复杂性.

经典的单调变分不等式

$$u^* \in \Omega, \quad (u - u^*)^{\mathrm{T}} F(u^*) \geqslant 0, \quad \forall\, u \in \Omega, \tag{1.9}$$

相当于在问题 (1.1) 中取 $w = u$ 和 $\theta(u) \equiv 0$, 并在算法的预测步 (1.2) 中, 取 $v = u = w$. 在欧氏模意义下考虑收缩时, 校正 (1.6) 的两个不同公式就分别变为

$$u^{k+1} = u^k - \alpha_k d(u^k, \tilde{u}^k) \quad \text{和} \quad u^{k+1} = P_\Omega[u^k - \alpha_k F(\tilde{u}^k)].$$

文献 [5,6] 对变分不等式问题 (1.9) 用 "孪生方向 - 相同步长" 的两类方法作了比较, 理论证明和数值试验结果都表明第二类方法比第一类方法好, 只是第二类算法要多加一个投影, 本文建议在投影容易实现的情形下采用第二类方法. 利用这些方法 (参见文献 [5,7]), 中国科学院武汉岩土力学研究所的研究工作者成功解决一些岩土工程问题 (参见文献 [1,2]).

对凸优化问题等价的变分不等式 (1.1), 已有的收缩算法 (如文献 [8–11]) 都是第一类算法 (1.6a) 的特例, 这类算法收敛性证明相对容易, 但不一定是在花费相当的情形下最好的方法. 第二类算法 (1.6b) 往往会取得更好的收敛效果. 本文给出求解 (1.1) 的统一框架, 它包含了文献 [5,6,12,13] 对变分不等式 (1.9) 的求解方法, 又有新的拓展, 对岩土工程中的第二类变分不等式[3] 也提供了求解方法.

2 预备知识

本文需要的预备知识是常识性的, 这里列出两个引理.

引理 2.1 设 $\Omega \subset \mathbb{R}^n$ 是闭凸集, $\theta(u)$ 和 $f(u)$ 都是 $\mathbb{R}^n \to \mathbb{R}$ 的凸函数. 如果 $f(u)$ 可微并且 $\min\{\theta(u) + f(u) \mid u \in \Omega\}$ 有解, 那么,

$$\tilde{u} \in \arg\min\{\theta(u) + f(u) \mid u \in \Omega\} \tag{2.1a}$$

的充分必要条件是

$$\tilde{u} \in \Omega, \quad \theta(u) - \theta(\tilde{u}) + (u - \tilde{u})^{\mathrm{T}}\nabla f(\tilde{u}) \geqslant 0, \quad \forall\, u \in \Omega. \tag{2.1b}$$

这里略去引理 2.1 的简单证明, 有兴趣的读者可以在本页脚注2) 的第 1.3 小节中找到这个证明. 利用引理 2.1, 我们可以得到下面的结论, 它在本文第 6 节的收敛性证明中要用到.

2) 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. Http://maths.nju.edu.cn/~hebma 中的 "My Talk".

引理 2.2 设 $\Omega \subset \mathbb{R}^n$ 是闭凸集, $\theta(u)$ 是 $\mathbb{R}^n \to \mathbb{R}$ 的凸函数, $H \in \mathbb{R}^{n\times n}$ 是正定矩阵. 对任意给定的 $q \in \mathbb{R}^n$ 和 $s > 0$, 如果

$$\tilde{u} = \arg\min\left\{s\theta(u) + \frac{1}{2}\|u-q\|_H^2 \mid u \in \Omega\right\}, \tag{2.2}$$

那么有

$$\|\tilde{u}-u\|_H^2 \leqslant \{2s\theta(u) + \|u-q\|_H^2\} - \{2s\theta(\tilde{u}) + \|\tilde{u}-q\|_H^2\}, \quad \forall\, u \in \Omega. \tag{2.3}$$

证明 由于 $\tilde{u}$ 是极小化问题 $\min\{s\theta(u) + \frac{1}{2}\|u-q\|_H^2 \mid u \in \Omega\}$ 的解, 根据引理 2.1, 有 $\tilde{u} \in \Omega$,

$$s\theta(u) - s\theta(\tilde{u}) + (u-\tilde{u})^{\mathrm{T}} H(\tilde{u}-q) \geqslant 0, \quad \forall\, u \in \Omega. \tag{2.4}$$

将恒等式

$$(u-\tilde{u})^{\mathrm{T}} H(\tilde{u}-q) = \frac{1}{2}(\|u-q\|_H^2 - \|\tilde{u}-q\|_H^2) - \frac{1}{2}\|u-\tilde{u}\|_H^2$$

用到 (2.4) 中, 就会得到

$$s\theta(u) - s\theta(\tilde{u}) + \frac{1}{2}(\|u-q\|_H^2 - \|\tilde{u}-q\|_H^2) - \frac{1}{2}\|u-\tilde{u}\|_H^2 \geqslant 0, \quad \forall\, u \in \Omega.$$

引理的结论 (2.3) 可以从上面的不等式直接得到. □

3 与变分不等式 (1.1) 等价的凸优化问题

变分不等式问题 (1.1) 包含的问题类型更广泛. 除了经典的变分不等式 (1.9) 是 (1.1) 的特例, 工程力学中有些问题 (参见文献 [3]) 可以表述为

$$u^* \in \Omega, \quad \theta(u) - \theta(u^*) + (u-u^*)^{\mathrm{T}} F(u^*) \geqslant 0, \quad \forall\, u \in \Omega, \tag{3.1}$$

其中凸函数 $\theta(u) \neq 0$ 非光滑. 此外, 线性约束的凸优化问题都可以归结为形式为 (1.1) 的变分不等式. 我们以三个可分离算子的线性约束凸优化问题

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) \mid Ax + By + Cz = b,\, x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\} \tag{3.2}$$

为例, 其中 $\theta_1(x) : \mathbb{R}^{n_1} \to \mathbb{R}$, $\theta_2(y) : \mathbb{R}^{n_2} \to \mathbb{R}$, $\theta_3(z) : \mathbb{R}^{n_3} \to \mathbb{R}$ 是 (不一定光滑的) 凸函数; $A \in \mathbb{R}^{m\times n_1}$, $B \in \mathbb{R}^{m\times n_2}$, $C \in \mathbb{R}^{m\times n_3}$ 和 $b \in \mathbb{R}^m$ 分别是给定的矩阵和向量; $\mathcal{X} \subset \mathbb{R}^{n_1}$, $\mathcal{Y} \subset \mathbb{R}^{n_2}$ 和 $\mathcal{Z} \subset \mathbb{R}^{n_3}$ 是给定的凸集. 记 $n = n_1 + n_2 + n_3$. 问题 (3.2) 的 Lagrange 函数是

$$L(x,y,z,\lambda) = \theta_1(x) + \theta_2(y) + \theta_3(z) - \lambda^{\mathrm{T}}(Ax + By + Cz - b), \tag{3.3}$$

其中 (x,y,z) 是原始变量, λ 是对偶变量. 如果一对 $((x^*,y^*,z^*),\lambda^*) \in (\mathcal{X}\times\mathcal{Y}\times\mathcal{Z})\times\mathbb{R}^m$ 满足

$$L_{\lambda\in\mathbb{R}^m}(x^*,y^*,z^*,\lambda) \leqslant L(x^*,y^*,z^*,\lambda^*) \leqslant L_{x\in\mathcal{X},y\in\mathcal{Y},z\in\mathcal{Z}}(x,y,z,\lambda^*),$$

则称其为 Lagrange 函数的鞍点. 上面的不等式关系可以写成

$$\begin{cases} x^* \in \mathcal{X}, & L(x,y^*,z^*,\lambda^*) - L(x^*,y^*,z^*,\lambda^*) \geqslant 0, \quad \forall\, x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, & L(x^*,y,z^*,\lambda^*) - L(x^*,y^*,z^*,\lambda^*) \geqslant 0, \quad \forall\, y \in \mathcal{Y}, \\ z^* \in \mathcal{Z}, & L(x^*,y^*,z,\lambda^*) - L(x^*,y^*,z^*,\lambda^*) \geqslant 0, \quad \forall\, z \in \mathcal{Z}, \\ \lambda^* \in \mathbb{R}^m, & L(x^*,y^*,z^*,\lambda^*) - L(x^*,y^*,z^*,\lambda) \geqslant 0, \quad \forall\, \lambda \in \mathbb{R}^m, \end{cases}$$

也就是

$$\begin{cases} x^* = \arg\min\{L(x, y^*, z^*, \lambda^*) \mid x \in \mathcal{X}\}, \\ y^* = \arg\min\{L(x^*, y, z^*, \lambda^*) \mid y \in \mathcal{Y}\}, \\ z^* = \arg\min\{L(x^*, y^*, z, \lambda^*) \mid z \in \mathcal{Z}\}, \\ \lambda^* = \arg\max\{L(x^*, y^*, z^*, \lambda) \mid \lambda \in \mathbb{R}^m\}. \end{cases}$$

利用 Lagrange 函数的具体形式 (3.3) 和引理 2.1, 上述优化问题的最优性条件是

$$\begin{cases} x^* \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^*) + (x - x^*)^{\mathrm{T}}(-A^{\mathrm{T}}\lambda^*) \geqslant 0, \quad \forall\, x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^*) + (y - y^*)^{\mathrm{T}}(-B^{\mathrm{T}}\lambda^*) \geqslant 0, \quad \forall\, y \in \mathcal{Y}, \\ z^* \in \mathcal{Z}, \quad \theta_3(z) - \theta_3(z^*) + (z - z^*)^{\mathrm{T}}(-C^{\mathrm{T}}\lambda^*) \geqslant 0, \quad \forall\, z \in \mathcal{Z}, \\ \lambda^* \in \mathbb{R}^m, \quad (\lambda - \lambda^*)^{\mathrm{T}}(Ax^* + By^* + Cz^* - b) \geqslant 0, \quad \forall\, \lambda \in \mathbb{R}^m. \end{cases} \tag{3.4}$$

显然, 上式的紧凑形式就是形如 (1.1) 的变分不等式

$$u^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^{\mathrm{T}}F(w^*) \geqslant 0, \quad \forall\, w \in \Omega, \tag{3.5a}$$

其中

$$w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^{\mathrm{T}}\lambda \\ -B^{\mathrm{T}}\lambda \\ -C^{\mathrm{T}}\lambda \\ Ax + By + Cz - b \end{pmatrix}, \tag{3.5b}$$

$$\theta(u) = \theta_1(x) + \theta_2(y) + \theta_3(z), \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \mathbb{R}^m. \tag{3.5c}$$

从 (3.5) 回到 (3.4), 只要分别令 (3.5) 中那个任意的 $w \in \Omega$ 为 $w = (x, y^*, z^*, \lambda^*), \ldots, (x^*, y^*, z^*, \lambda)$ 就能得到. 注意到 (3.5b) 中的 F 总满足 $(w - \bar{w})^{\mathrm{T}}(F(w) - F(\bar{w})) = 0$, 因此 F 也是单调的. 在变分不等式框架中构造算法[14] 极大地简化了收敛性分析. 得到越来越多的学者认可 (参见文献 [15]). 顺便提及一下, 用直接推广的交替方向法求解含三个算子的问题 (3.2) 是不能保证收敛的 (参见文献 [16]).

4 预测点和孪生方向

本节讨论合格预测点的性质并以典型问题为例说明合格预测点如何产生.

4.1 合格预测点的性质

合格的预测点由第 1.2 小节中的定义 1.1 给出. 将 (1.2) 中任意的 $w \in \Omega$ 选成 w^*, 就得到

$$(\tilde{v}^k - v^*)^{\mathrm{T}} d(v^k, \tilde{v}^k) \geqslant \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^{\mathrm{T}} F(\tilde{w}^k). \tag{4.1}$$

由 F 的单调性, $(\tilde{w}^k - w^*)^{\mathrm{T}} F(\tilde{w}^k) \geqslant (\tilde{w}^k - w^*)^{\mathrm{T}} F(w^*)$, 可得

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

由于 $w^* \in \Omega^*$, 根据 (1.1), 上式右端非负, 因此有

$$\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^{\mathrm{T}} F(\tilde{w}^k) \geqslant 0. \tag{4.2}$$

这说明 (4.1) 的右端非负. 随后就有 $(\tilde{v}^k - v^*)^{\mathrm{T}} d(v^k, \tilde{v}^k) \geqslant 0$. 据此得到

$$(v^k - v^*)^{\mathrm{T}} d(v^k, \tilde{v}^k) \geqslant (v^k - \tilde{v}^k)^{\mathrm{T}} d(v^k, \tilde{v}^k). \tag{4.3}$$

对任意给定的 v^*, $H(v^k - v^*)$ 可以看作 $\frac{1}{2}\|v - v^*\|_H^2$ 在 v^k 处的梯度. 因此, 上式可以看作

$$\nabla\left(\frac{1}{2}\|v - v^*\|_H^2\right)\bigg|_{v=v^k} H^{-1} d(v^k, \tilde{v}^k) \geqslant (v^k - \tilde{v}^k)^{\mathrm{T}} d(v^k, \tilde{v}^k).$$

条件 (1.3) 说明, 当 $v^k \neq \tilde{v}^k$ 时上式右端大于 0. 因此, (4.3) 表示 $H^{-1} d(v^k, \tilde{v}^k)$ 是距离函数 $\frac{1}{2}\|v - v^*\|_H^2$ 在 v^k 处的一个上升方向.

此外, 如果 $w^k \in \Omega$, 那么将它代入 (1.2) 就有

$$\theta(u^k) - \theta(\tilde{u}^k) + (w^k - \tilde{w}^k)^{\mathrm{T}} F(\tilde{w}^k) \geqslant (v^k - \tilde{v}^k)^{\mathrm{T}} d(v^k, \tilde{v}^k).$$

将上式和不等式 (4.2) 相加, 就有

$$\theta(u^k) - \theta(u^*) + (w^k - w^*)^{\mathrm{T}} F(\tilde{w}^k) \geqslant (v^k - \tilde{v}^k)^{\mathrm{T}} d(v^k, \tilde{v}^k). \tag{4.4}$$

注意到不等式 (4.3) 和 (4.4) 有相同的右端. 不同之处是, 推导 (4.4) 时需要有 $w^k \in \Omega$.

4.2 预测点的产生

针对不同类型的问题, 产生合格预测点的方法很多. 我们以单调变分不等式 (1.9) 和由凸优化问题 (3.2) 导致的变分不等式 (3.5) 为例说明预测点的生成方式.

4.2.1 通过投影提供单调变分不等式 (1.9) 的预测点

在求解单调变分不等式 (1.9) 的投影收缩算法中 (参见文献 [7,12]), 对给定的当前点 u^k 和 $\beta_k > 0$, 我们利用投影

$$\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)], \tag{4.5}$$

生成一个预测点 $\tilde{u}^k$. 换句话说,

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

根据引理 2.1, 有

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

即

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^{\mathrm{T}} F(u^k) \geqslant (u - \tilde{u}^k)^{\mathrm{T}} \frac{1}{\beta_k}(u^k - \tilde{u}^k), \quad \forall\, u \in \Omega.$$

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

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^{\mathrm{T}} F(\tilde{u}^k) \geqslant (u - \tilde{u}^k)^{\mathrm{T}} d(u^k, \tilde{u}^k), \quad \forall\, u \in \Omega, \tag{4.6}$$

其中

$$d(u^k,\tilde{u}^k)=\frac{1}{\beta_k}(u^k-\tilde{u}^k)-[F(u^k)-F(\tilde{u}^k)]. \tag{4.7}$$

根据对 F 的 Lipschitz 连续假设, $d(u^k,\tilde{u}^k)$ 满足 (1.3) 的前半部分要求. 在投影 (4.5) 中, 我们可以假设 $\beta_k>0$, 使得

$$\beta_k\|F(u^k)-F(\tilde{u}^k)\|\leqslant\nu\|u^k-\tilde{u}^k\|,\quad \nu\in(0,1). \tag{4.8}$$

利用 Armijo 类法则, 可以使得

$$\beta_{\max}\geqslant\sup\{\beta_k\}\geqslant\inf\{\beta_k\}\geqslant\beta_{\min}>0.$$

由 $d(u^k,\tilde{u}^k)$ 的表达式 (4.7) 和条件 (4.8), 根据 Cauchy-Schwarz 不等式, 有

$$(u^k-\tilde{u}^k)^{\mathrm{T}}d(u^k,\tilde{u}^k)\geqslant\frac{1}{\beta_k}(1-\nu)\|u^k-\tilde{u}^k\|^2\geqslant\frac{1}{\beta_{\max}}(1-\nu)\|u^k-\tilde{u}^k\|^2. \tag{4.9}$$

这样的 $d(u^k,\tilde{u}^k)$ 满足 (1.3) 的后半部分要求. 因此, 由 (4.5) 给出的 $\tilde{u}^k$ 是一个合格的预测点. 除此以外, 也可以像文献 [13] 中通过近似求解一个辅助子问题得到合格的预测点. 求解经典的单调变分不等式 (1.9) 的投影收缩算法通常由这种方法得到预测点. 由投影得到预测点的收缩方法, 称之为投影收缩算法[17]. 然而, 人们往往忽视了采用同样步长的第二类校正方法[18,19].

4.2.2 通过求解优化子问题得到 (3.5) 的预测点

变分不等式 (3.5) 是问题 (1.1) 的一个具体例子, 它是由凸优化问题 (3.2) 导出的. 注意到凸优化问题 (3.2) 的 Lagrange 函数是 (3.3), 相应的增广 Lagrange 函数 (augmented Lagrange function) 是

$$\mathcal{L}_\beta(x,y,z,\lambda)=L(x,y,z,\lambda)+\frac{\beta}{2}\|Ax+By+Cz-b\|^2,$$

其中 $\beta>0$ 是等式约束 $Ax+By+Cz=b$ 的罚参数. 换句话说,

$$\mathcal{L}_\beta(x,y,z,\lambda)=\theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^{\mathrm{T}}(Ax+By+Cz-b)+\frac{\beta}{2}\|Ax+By+Cz-b\|^2. \tag{4.10}$$

对给定的 $v^k=(y^k,z^k,\lambda^k)$, 先按 x-y-z 的顺序求得 $\tilde{u}^k=(\tilde{x}^k,\tilde{y}^k,\tilde{z}^k)$:

$$\begin{cases}\tilde{x}^k=\arg\min\left\{\mathcal{L}_\beta(x,y^k,z^k,\lambda^k)\mid x\in\mathcal{X}\right\},\\ \tilde{y}^k=\arg\min\left\{\mathcal{L}_\beta(\tilde{x}^k,y,z^k,\lambda^k)\mid y\in\mathcal{Y}\right\},\\ \tilde{z}^k=\arg\min\left\{\mathcal{L}_\beta(\tilde{x}^k,\tilde{y}^k,z,\lambda^k)\mid z\in\mathcal{Z}\right\},\end{cases} \tag{4.11a}$$

然后通过

$$\tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k+Cz^k-b) \tag{4.11b}$$

定义 $\tilde{\lambda}^k$, 这样生成了 $\tilde{w}^k=(\tilde{x}^k,\tilde{y}^k,\tilde{z}^k,\tilde{\lambda}^k)$. (4.11a) 中的子问题分别是只含变量 x、y 和 z 的凸优化问题, 本文不讨论如何具体求解这些子问题. 利用增广 Lagrange 函数的表达式 (见 (4.10)) 和引理 2.1, 可知 (4.11a) 的 x- 子问题的最优性条件可以写成

$$\tilde{x}^k\in\mathcal{X},\quad \theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^{\mathrm{T}}\{-A^{\mathrm{T}}\lambda^k+\beta A^{\mathrm{T}}(A\tilde{x}^k+By^k+Cz^k-b)\}\geqslant 0.$$

再根据 (4.11b) 中 $\tilde{\lambda}^k$ 的定义, 上面的不等式可以写成

$$\tilde{x}^k \in \mathcal{X}, \quad \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^{\mathrm{T}}\{-A^{\mathrm{T}}\tilde{\lambda}^k\} \geqslant 0, \quad \forall\, x \in \mathcal{X}. \tag{4.12a}$$

类似地, (4.11a) 中的 y- 和 z- 子问题的最优性条件可以分别写成

$$\tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^{\mathrm{T}}\{-B^{\mathrm{T}}\tilde{\lambda}^k + \beta B^{\mathrm{T}}B(\tilde{y}^k - y^k)\} \geqslant 0, \quad \forall\, y \in \mathcal{Y} \tag{4.12b}$$

和

$$\tilde{z}^k \in \mathcal{Z}, \quad \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^{\mathrm{T}}\{-C^{\mathrm{T}}\tilde{\lambda}^k + \beta C^{\mathrm{T}}B(\tilde{y}^k - y^k) + \beta C^{\mathrm{T}}C(\tilde{z}^k - z^k)\} \geqslant 0, \quad \forall\, z \in \mathcal{Z}. \tag{4.12c}$$

此外, 由 $\tilde{\lambda}^k$ 的表达式 (4.11b) 得到的等式

$$(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b) - B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0$$

也可以写成变分不等式的形式

$$\begin{aligned}\tilde{\lambda}^k \in \mathbb{R}^m, \quad & (\lambda - \tilde{\lambda}^k)^{\mathrm{T}}(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b) + (\lambda - \tilde{\lambda}^k)^{\mathrm{T}}\Big(- B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\Big) \\ & \geqslant 0, \quad \forall\, \lambda \in \mathbb{R}^m.\end{aligned} \tag{4.12d}$$

利用 (3.5) 中的表达式, 将 (4.12a)、(4.12b)、(4.12c) 和 (4.12d) 写在一起就是

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^{\mathrm{T}}F(\tilde{w}^k) \geqslant (v - \tilde{v}^k)^{\mathrm{T}}Q(v^k - \tilde{v}^k), \quad \forall\, w \in \Omega, \tag{4.13}$$

其中

$$v = \begin{pmatrix} y \\ z \\ \lambda \end{pmatrix} \quad \text{和} \quad Q = \begin{pmatrix} \beta B^{\mathrm{T}}B & 0 & 0 \\ \beta C^{\mathrm{T}}B & \beta C^{\mathrm{T}}C & 0 \\ -B & -C & \frac{1}{\beta}I_m \end{pmatrix}. \tag{4.14}$$

对应于定义 1.1 中的 $d(v^k, \tilde{v}^k)$, 这里有

$$d(v^k, \tilde{v}^k) = Q(v^k - \tilde{v}^k). \tag{4.15}$$

Q 是一个矩阵, 条件 (1.3) 的第一部分显然满足. 此外,

$$\begin{aligned}(v^k - \tilde{v}^k)^{\mathrm{T}}d(v^k, \tilde{v}^k) &= \frac{1}{2}\{(v^k - \tilde{v}^k)^{\mathrm{T}}(Q^{\mathrm{T}} + Q)(v^k - \tilde{v}^k)\} \\ &= \frac{1}{2}(v^k - \tilde{v}^k)^{\mathrm{T}}\left(\begin{pmatrix} \beta B^{\mathrm{T}}B & 0 & 0 \\ 0 & \beta C^{\mathrm{T}}C & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} + \begin{pmatrix} \beta B^{\mathrm{T}}B & \beta B^{\mathrm{T}}C & -B^{\mathrm{T}} \\ \beta C^{\mathrm{T}}B & \beta C^{\mathrm{T}}C & -C^{\mathrm{T}} \\ -B & -C & \frac{1}{\beta}I_m \end{pmatrix}\right)(v^k - \tilde{v}^k) \\ &= \frac{1}{2}\Big(\beta\|B(y^k - \tilde{y}^k)\|^2 + \beta\|C(z^k - \tilde{z}^k)\|^2 + \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2\Big) \\ &\quad + \frac{1}{2}\Big(\beta\|B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k) - \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k)\|^2\Big) \\ &\geqslant \frac{1}{2}\Big(\beta\|B(y^k - \tilde{y}^k)\|^2 + \beta\|C(z^k - \tilde{z}^k)\|^2 + \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2\Big).\end{aligned}$$

因此, 在 B 和 C 列满秩的条件下, 条件 (1.3) 的第二部分满足. 因此, 由 (4.11) 给出的 $\tilde{w}^k$ 是一个合格的预测点. 求解凸优化的分裂收缩算法通常是用类似的方法生成预测点.

5 以 $H^{-1}d(v^k,\tilde{v}^k)$ 为搜索方向的第一类方法

产生新迭代点的校正公式由 (1.6) 给出, 我们先讨论第一类方法. 为讨论如何选取步长 α, 我们把第一类方法产生新迭代点的校正公式 (1.6a) 写成依赖 α 的

$$v_I^{k+1}(\alpha)=v^k-\alpha H^{-1}d(v^k,\tilde{v}^k), \tag{5.1}$$

希望每次迭代得到尽可能大的 "进步量". 通过定义

$$\vartheta_k(\alpha)=\|v^k-v^*\|_H^2-\|v_I^{k+1}(\alpha)-v^*\|_H^2, \tag{5.2}$$

研究如何选取适当的 $\alpha=\alpha_k$, 使得 $\vartheta_k(\alpha_k)$ 尽可能大. 为了下面的讨论, 定义

$$q_k(\alpha)=2\alpha(v^k-\tilde{v}^k)^{\mathrm{T}}d(v^k,\tilde{v}^k)-\alpha^2\|H^{-1}d(v^k,\tilde{v}^k)\|_H^2. \tag{5.3}$$

5.1 计算步长的第一类方法

根据 (5.2) 和 (5.1), 得到

$$\begin{aligned}\vartheta_k(\alpha)&=\|v^k-v^*\|_H^2-\|(v^k-v^*)-\alpha H^{-1}d(v^k,\tilde{v}^k)\|_H^2\\&=2\alpha(v^k-v^*)^{\mathrm{T}}d(v^k,\tilde{v}^k)-\alpha^2\|H^{-1}d(v^k,\tilde{v}^k)\|_H^2.\end{aligned} \tag{5.4}$$

对任意给定的 v^*, (5.4) 表明 $\vartheta_k(\alpha)$ 是 α 的一个二次函数. 只是 v^* 是 w^* 的部分向量, 是未知的, 我们无法直接求 $\vartheta_k(\alpha)$ 的极大.

定理 5.1 设 $v_I^{k+1}(\alpha)$ 由 (1.6a) 生成. 对任意的 $\alpha>0$, 由 (5.2) 定义的 $\vartheta_k(\alpha)$, 有

$$\vartheta_k(\alpha)\geqslant q_k(\alpha), \tag{5.5}$$

其中 $q_k(\alpha)$ 由 (5.3) 给出.

证明 这个结论可以从 (5.4) 和 (4.3) 直接得到. □

定理 5.1 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 α_k^* 是

$$\alpha_k^*=\mathrm{argmax}\{q_k(\alpha)\}=\frac{(v^k-\tilde{v}^k)^{\mathrm{T}}d(v^k,\tilde{v}^k)}{\|H^{-1}d(v^k,\tilde{v}^k)\|_H^2}. \tag{5.6}$$

这个最优步长可以根据寻查方向 $d(v^k,\tilde{v}^k)$ 相应的 (4.3) 的左右两端的信息确定. 根据条件 (1.3), 存在

$$\alpha_{\min}=\lambda_{\min}(H)\left(\frac{\delta}{K^2}\right)>0, \tag{5.7}$$

其中 $\lambda_{\min}(H)$ 是正定矩阵 H 的最小特征值, 使得 $\alpha_k^*\geqslant\alpha_{\min}$, $\forall\,k\geqslant 0$. 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 (5.4)), 由于它含有未知的 v^*, 我们不得已才极大化它的下界二次函数 $q_k(\alpha)$. 对任意的 $\gamma\in(0,2)$, 用

$$v_I^{k+1}=v^k-\gamma\alpha_k^*H^{-1}d(v^k,\tilde{v}^k), \tag{5.8}$$

产生新的迭代点. 根据 (5.2) 和 (5.5), 由 (5.8) 产生的 v_I^{k+1} 满足

$$\|v_I^{k+1}-v^*\|_H^2\leqslant\|v^k-v^*\|_H^2-q_k(\gamma\alpha_k^*). \tag{5.9}$$

由 $q_k(\alpha)$ 和 α_k^* 的定义 (分别见 (5.3) 和 (5.6)), 以及 $\alpha_k = \gamma\alpha_k^*$ 得到

$$q_k(\alpha_k) = 2\alpha_k(v^k - \tilde{v}^k)^{\mathrm{T}} d(v^k, \tilde{v}^k) - \alpha_k^2 \|H^{-1} d(v^k, \tilde{v}^k)\|_H^2 = (2-\gamma)\alpha_k (v^k - \tilde{v}^k)^{\mathrm{T}} d(v^k, \tilde{v}^k). \tag{5.10}$$

所以, 根据 (1.3), 由校正公式 (5.8) 产生的 v_I^{k+1} 满足

$$\|v_I^{k+1} - v^*\|_H^2 \leqslant \|v^k - v^*\|_H^2 - \gamma(2-\gamma)\alpha_k^* \delta \|v^k - \tilde{v}^k\|^2.$$

由 $\alpha_k^* \geqslant \alpha_{\min}$ 和 (5.7), 便有

$$\|v_I^{k+1} - v^*\|_H^2 \leqslant \|v^k - v^*\|_H^2 - \gamma(2-\gamma)\lambda_{\min}(H)\left(\frac{\delta^2}{K^2}\right)\|v^k - \tilde{v}^k\|^2. \tag{5.11}$$

不等式 (5.11) 说明, 序列 $\{v^k\}$ 都是收缩和有界的. 利用这个不等式, 容易证明方法的收敛性. 在实际计算中, 我们一般取一个松弛因子 $\gamma \in [1, 2)$, 其理由可见图 1.

5.2 采用固定步长的第一类方法

如果问题 (1.1) 是由结构型凸优化问题导致的, 它生成预测点的一般形式是 (4.13), 寻查方向是 $d(v^k, \tilde{v}^k) = Q(v^k - \tilde{v}^k)$ (见 (4.15)). 设

$$M = H^{-1}Q,$$

在实际计算中, 人们往往取一个固定的 α, 迭代采取

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k), \tag{5.12}$$

其中 α 是常数, 满足 $0 < \alpha < \alpha_{\max}$, 这里

$$\alpha_{\max} = \arg\max\{\alpha \mid Q^{\mathrm{T}} + Q - \alpha M^{\mathrm{T}} H M \succeq 0\}.$$

由于 (见 (5.3))

$$\begin{aligned} q_k(\alpha) &= 2\alpha(v^k - \tilde{v}^k)^{\mathrm{T}} Q(v^k - \tilde{v}^k) - \alpha^2 \|M(v^k - \tilde{v}^k)\|_H^2 \\ &= \alpha\{(v^k - \tilde{v}^k)^{\mathrm{T}}[Q^{\mathrm{T}} + Q - \alpha M^{\mathrm{T}} H M](v^k - \tilde{v}^k)\} \\ &\geqslant \alpha(\alpha_{\max} - \alpha)(v^k - \tilde{v}^k)^{\mathrm{T}} M^{\mathrm{T}} H M (v^k - \tilde{v}^k), \end{aligned} \tag{5.13}$$

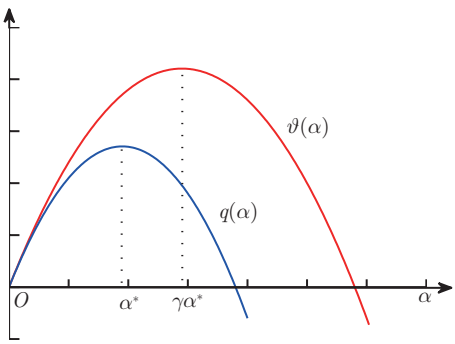

图 1 取松弛因子 $\gamma \in [1, 2)$ 的示意图

根据 (5.9) 和 (5.10), 有

$$\|v^{k+1}-v^*\|_H^2 \leqslant \|v^k-v^*\|_H^2-\alpha(\alpha_{\max}-\alpha)\|v^k-\tilde{v}^k\|_{M^{\mathrm{T}}HM}^2.$$

文献 [6,8–11,14,20,21] 中给出的求解凸优化的分裂收缩方法, 都采用固定步长的方式, 因为简单, 更容易被他人接受和采用. 例如, 文献 [22] 就用文献 [10] 中的方法成功求解数据科学中的问题. 如果 Q 本身就是一个对称正定矩阵, 那么就可以令 $H=Q$, M 为单位矩阵, 方法就变成邻近点算法 (PPA) [23,24] 或者定制的邻近点算法 (customized PPA) [8,25].

6 以 $\mathcal{H}^{-1}F(\tilde{w}^k)$ 为搜索方向的第二类方法

为讨论第二类方法中如何选取步长 α, 我们把第二类方法产生新迭代点的校正公式 (1.6b) 写成依赖 α 的

$$w_{II}^{k+1}(\alpha)=\arg\min\left\{\alpha\theta(u)+\frac{1}{2}\|w-[w^k-\alpha\mathcal{H}^{-1}F(\tilde{w}^k)]\|_{\mathcal{H}}^2 \;\middle|\; w\in\Omega\right\}. \tag{6.1}$$

注意到 v 为 w 的核心变量部分, 同样定义

$$\zeta_k(\alpha)=\|v^k-v^*\|_H^2-\|v_{II}^{k+1}(\alpha)-v^*\|_H^2. \tag{6.2}$$

特别需要指出的是, 第一类方法的最优步长是根据 (4.3) 的左右两端的信息确定的. 而对于第二类算法, 最优步长不是类似地根据 (4.4) 的左右两端的信息确定的, 而是直接采用相应的第一类算法的步长. 下面的定理说明, 在 (1.6b) 中取与 (1.6a) 中同样的 α_k, (6.2) 中的 $\zeta_k(\alpha)$ '不差于' (5.2) 中的 $\vartheta_k(\alpha)$.

定理 6.1 设 $w_{II}^{k+1}(\alpha)$ 由 (6.1) 生成. 对任意的 $\alpha>0$, 由 (6.2) 定义的 $\zeta_k(\alpha)$, 有

$$\zeta_k(\alpha)\geqslant q_k(\alpha)+\|v_I^{k+1}(\alpha)-v_{II}^{k+1}(\alpha)\|_H^2, \tag{6.3}$$

其中 $v_I^{k+1}(\alpha)$ 和 $q_k(\alpha)$ 分别由 (5.1) 和 (5.3) 给出.

证明 我们先考虑 $\|w^k-w^*\|_{\mathcal{H}}^2-\|w_{II}^{k+1}(\alpha)-w^*\|_{\mathcal{H}}^2$. 由于 $w_{II}^{k+1}(\alpha)$ 是 (6.1) 的解, 在引理 2.2 的 (2.2) 中令 $s=\alpha$, $q=w^k-\alpha\mathcal{H}^{-1}F(\tilde{w}^k)$, 有

$$\begin{aligned}\|w_{II}^{k+1}(\alpha)-w\|_{\mathcal{H}}^2 \leqslant{}& (2\alpha\theta(u)+\|w-[w^k-\alpha\mathcal{H}^{-1}F(\tilde{w}^k)]\|_{\mathcal{H}}^2)\\ &-(2\alpha\theta(u_{II}^{k+1})+\|w_{II}^{k+1}(\alpha)-[w^k-\alpha\mathcal{H}^{-1}F(\tilde{w}^k)]\|_{\mathcal{H}}^2),\quad \forall\, w\in\Omega.\end{aligned}$$

对上式中任意的 $w\in\Omega$, 取 $w=w^*$, 就得到

$$\begin{aligned}\|w_{II}^{k+1}(\alpha)-w^*\|_{\mathcal{H}}^2 \leqslant{}& (2\alpha\theta(u^*)+\|w^k-\alpha\mathcal{H}^{-1}F(\tilde{w}^k)-w^*\|_{\mathcal{H}}^2)\\ &-(2\alpha\theta(u_{II}^{k+1})+\|w^k-\alpha\mathcal{H}^{-1}F(\tilde{w}^k)-w_{II}^{k+1}(\alpha)\|_{\mathcal{H}}^2).\end{aligned} \tag{6.4}$$

因此,

$$\begin{aligned}\|w^k-w^*\|_{\mathcal{H}}^2-\|w_{II}^{k+1}(\alpha)-w^*\|_{\mathcal{H}}^2 \geqslant{}& \|w^k-w^*\|_H^2-\|(w^k-w^*)-\alpha\mathcal{H}^{-1}F(\tilde{w}^k)\|_{\mathcal{H}}^2\\ &+\|(w^k-w_{II}^{k+1}(\alpha))-\alpha\mathcal{H}^{-1}F(\tilde{w}^k)\|_{\mathcal{H}}^2+2\alpha(\theta(u_{II}^{k+1})-\theta(u^*)).\end{aligned}$$

进一步对上式右端化简, 得到

$$\|w^k-w^*\|_{\mathcal{H}}^2-\|w_{II}^{k+1}(\alpha)-w^*\|_{\mathcal{H}}^2$$

$$
\begin{aligned}
&\geqslant 2\alpha(w^k-w^*)^{\mathrm{T}}F(\tilde{w}^k)+2\alpha(w_{II}^{k+1}(\alpha)-w^k)^{\mathrm{T}}F(\tilde{w}^k)+\|w^k-w_{II}^{k+1}(\alpha)\|_{\mathcal{H}}^2+2\alpha(\theta(u_{II}^{k+1})-\theta(u^*))\\
&=2\alpha(w_{II}^{k+1}(\alpha)-w^*)^{\mathrm{T}}F(\tilde{w}^k)+\|w^k-w_{II}^{k+1}(\alpha)\|_{\mathcal{H}}^2+2\alpha(\theta(u_{II}^{k+1})-\theta(u^*)).
\end{aligned}\tag{6.5}
$$

将 (6.5) 中右端的第一项 $(w_{II}^{k+1}(\alpha)-u^*)^{\mathrm{T}}F(\tilde{w}^k)$ 按需要分拆:

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

因为 $w_{II}^{k+1}(\alpha)\in\Omega$, 在 (1.2) 中令 $w=w_{II}^{k+1}(\alpha)$ 就得到

$$
(w_{II}^{k+1}(\alpha)-\tilde{w}^k)^{\mathrm{T}}F(\tilde{w}^k)\geqslant(v_{II}^{k+1}(\alpha)-\tilde{v}^k)^{\mathrm{T}}d(v^k,\tilde{v}^k)+(\theta(\tilde{u}^k)-\theta(u_{II}^{k+1}(\alpha))).\tag{6.6}
$$

由于

$$
(\tilde{w}^k-w^*)^{\mathrm{T}}F(\tilde{w}^k)\geqslant(\tilde{w}^k-w^*)^{\mathrm{T}}F(w^*)\quad\text{和}\quad\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^{\mathrm{T}}F(w^*)\geqslant 0,
$$

我们有

$$
(\tilde{w}^k-w^*)^{\mathrm{T}}F(\tilde{w}^k)\geqslant\theta(u^*)-\theta(\tilde{u}^k).\tag{6.7}
$$

将 (6.6) 和 (6.7) 相加, 得到

$$
(w_{II}^{k+1}(\alpha)-w^*)^{\mathrm{T}}F(\tilde{w}^k)\geqslant(v_{II}^{k+1}(\alpha)-\tilde{v}^k)^{\mathrm{T}}d(v^k,\tilde{v}^k)+(\theta(u^*)-\theta(u_{II}^{k+1})).\tag{6.8}
$$

将它代入 (6.5) 的右端, 就有

$$
\|w^k-w^*\|_{\mathcal{H}}^2-\|w_{II}^{k+1}(\alpha)-w^*\|_{\mathcal{H}}^2\geqslant\|w^k-w_{II}^{k+1}(\alpha)\|_{\mathcal{H}}^2+2\alpha(v_{II}^{k+1}(\alpha)-\tilde{v}^k)^{\mathrm{T}}d(v^k,\tilde{v}^k).
$$

在 v 是 w 的真子向量的情形下, 在上式中令 $h\to 0_+$ (见 (1.5) 中的 $\mathcal{H}$), 得到

$$
\|v^k-v^*\|_H^2-\|v_{II}^{k+1}(\alpha)-v^*\|_H^2\geqslant\|v^k-v_{II}^{k+1}(\alpha)\|_H^2+2\alpha(v_{II}^{k+1}(\alpha)-\tilde{v}^k)^{\mathrm{T}}d(v^k,\tilde{v}^k).\tag{6.9}
$$

对上式右端, 利用 $q_k(\alpha)$ 的形式 (见 (5.3)), 就化成

$$
\begin{aligned}
\zeta_k(\alpha)&=\|v^k-v^*\|_H^2-\|v_{II}^{k+1}(\alpha)-v^*\|_H^2\\
&\geqslant\|v^k-v_{II}^{k+1}(\alpha)\|_H^2+2\alpha(v_{II}^{k+1}(\alpha)-\tilde{v}^k)^{\mathrm{T}}d(v^k,\tilde{v}^k)\\
&=\|v^k-v_{II}^{k+1}(\alpha)\|_H^2+2\alpha(v_{II}^{k+1}(\alpha)-v^k)^{\mathrm{T}}d(v^k,\tilde{v}^k)+2\alpha(v^k-\tilde{v}^k)^{\mathrm{T}}d(v^k,\tilde{v}^k)\\
&=\|(v^k-v_{II}^{k+1}(\alpha))-\alpha H^{-1}d(v^k,\tilde{v}^k)\|_H^2-\alpha^2\|H^{-1}d(v^k,\tilde{v}^k)\|_H^2+2\alpha(v^k-\tilde{v}^k)^{\mathrm{T}}d(v^k,\tilde{v}^k)\\
&=\|(v^k-v_{II}^{k+1}(\alpha))-\alpha H^{-1}d(v^k,\tilde{v}^k)\|_H^2+q_k(\alpha).
\end{aligned}
$$

由于 $v^k-\alpha H^{-1}d(v^k,\tilde{v}^k)$ 就是 $v_I^{k+1}(\alpha)$ (见 (5.1)), 这样就完成了定理结论 (6.3) 的证明. □

定理 6.1 说明, $q_k(\alpha)$ 也是 $\zeta_k(\alpha)$ 的下界. 比较 (6.3) 和 (5.5), 可以发现, 对同样的 α, $\zeta_k(\alpha)$ 优于 $\vartheta_k(\alpha)$. 在实际计算中, 由于 v 才是核心变量, 我们采用校正公式

$$
\text{(收缩算法 -1)}\quad v_I^{k+1}=v^k-\gamma\alpha_k^*H^{-1}d(v^k,\tilde{v}^k),\tag{6.10}
$$

或者

$$
\text{(收缩算法 -2)}\quad v_{II}^{k+1}=\arg\min\left\{\gamma\alpha_k^*\theta_v(u)+\frac{1}{2}\|v-[v^k-\gamma\alpha_k^*H^{-1}F_v(\tilde{w}^k)]\|_H^2\;\middle|\;v\in\mathcal{V}\right\}\tag{6.11}
$$

给出新的迭代点 v^{k+1}, 其中 α_k^* 都由 (5.6) 给出, $\theta_v(u)$ 和 $F_v(\tilde{w}^k)$ 分别是可分离函数 $\theta(u)$ 和向量 $F(\tilde{w}^k)$ 中对应核心变量 v 的那一部分. 例如, 对于第 4.2.2 小节中三个可分离算子的优化问题, 因为 $w=(x,y,z,\lambda)$, $v=(y,z,\lambda)$, 所以相应地有

$$\theta_v(u)=\theta_2(y)+\theta_3(z) \quad 和 \quad F_v(w)=\begin{pmatrix} -B^{\mathrm{T}}\lambda \\ -C^{\mathrm{T}}\lambda \\ Ax+By+Cz-b \end{pmatrix}.$$

采用收缩算法 -2, 校正得到 v^{k+1} 需要求解的问题 (6.11) 的形式是

$$\min\left\{\theta_v(u)+\frac{r}{2}\|v-d\|_H^2 \mid v\in\mathcal{V}\right\}, \tag{6.12}$$

也就是形如 (2.2) 的问题. 对可分离结构的变分不等式问题, 求解 (6.12) 往往是容易操作的.

7 遍历意义下的迭代复杂性

为了证明算法的迭代复杂性, 我们需要下面的定理对 (1.1) 的解集做新的刻画.

定理 7.1 变分不等式 (1.1) 的解集是凸的并且可以表示成

$$\Omega^*=\bigcap_{w\in\Omega}\{\tilde{w}\in\Omega:(\theta(u)-\theta(\tilde{u}))+(w-\tilde{w})^{\mathrm{T}}F(w)\geqslant 0\}. \tag{7.1}$$

证明 证明方法与文献 [4, 定理 2.3.5] 类似. 有兴趣的读者可以参见文献 [21, 定理 2.1] 中的详细证明. □

定理 7.1 说明

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

与

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

是等价的. 我们用后者定义变分不等式 (1.1) 的近似解. 对给定的 $\epsilon>0$, 如果 $\tilde{w}\in\Omega$ 满足

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

其中

$$\mathcal{D}_{(\tilde{w})}=\{w\in\Omega \mid \|w-\tilde{w}\|\leqslant 1\} \tag{7.2}$$

就称作变分不等式 (1.1) 的 ϵ 近似解. 换句话说, 对给定的 $\epsilon>0$, 问经过多少次迭代, 能够得到一个 $\tilde{w}\in\Omega$, 使得

$$\tilde{w}\in\Omega,\quad \sup_{w\in\mathcal{D}_{(\tilde{w})}}\{\theta(\tilde{u})-\theta(u)+(\tilde{w}-w)^{\mathrm{T}}F(w)\}\leqslant\epsilon. \tag{7.3}$$

下面的两个引理, 对采用同一预测不同校正的两类方法, 我们证明了它们的同一性质. 由 F 的单调性, 我们有 $(w-\tilde{w}^k)^{\mathrm{T}}F(w)\geqslant(w-\tilde{w}^k)^{\mathrm{T}}F(\tilde{w}^k)$. 将其用到 (1.2) 的左端, 得到

$$\tilde{w}^k\in\Omega,\quad \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^{\mathrm{T}}F(w)\geqslant(v-\tilde{v}^k)^{\mathrm{T}}d(v^k,\tilde{v}^k),\quad \forall\, w\in\Omega. \tag{7.4}$$

这个不等式在下面的两个引理证明中都要用到.

引理 7.1 设 $\{\tilde{w}^k\}$ 和 $\{v^k\}$ 是算法中分别由预测 (1.2) 和第一类校正方法 (1.6a) 产生的序列, 则对所有的 k 都有

$$\alpha_k(\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^{\mathrm{T}}F(w)) \geqslant \frac{1}{2}(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2)+\frac{1}{2}q_k(\alpha_k), \quad \forall\, w\in\Omega, \tag{7.5}$$

其中 $q_k(\cdot)$ 由 (5.3) 给出.

证明 由 (1.6a), 有 $d(v^k,\tilde{v}^k)=\frac{1}{\alpha_k}H(v^k-v^{k+1})$. 代入 (7.4), 得到

$$\alpha_k(\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^{\mathrm{T}}F(w)) \geqslant (v-\tilde{v}^k)^{\mathrm{T}}H(v^k-v^{k+1}), \quad \forall\, w\in\Omega.$$

因此, 要证明引理的结论, 我们只需证明

$$(v-\tilde{v}^k)^{\mathrm{T}}H(v^k-v^{k+1}) \geqslant \frac{1}{2}(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2)+\frac{1}{2}q_k(\alpha_k). \tag{7.6}$$

对 (7.6) 左端的 $(v-\tilde{v}^k)^{\mathrm{T}}H(v^k-v^{k+1})$ 使用恒等式

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

我们得到

$$(v-\tilde{v}^k)^{\mathrm{T}}H(v^k-v^{k+1})=\frac{1}{2}(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2)+\frac{1}{2}(\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2). \tag{7.7}$$

再次利用 $v^{k+1}=v^k-\alpha_kH^{-1}d(v^k,\tilde{v}^k)$ 和 $q_k(\cdot)$ 的定义, 马上得到

$$\begin{aligned}\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2 &= \|v^k-\tilde{v}^k\|_H^2-\|(v^k-\tilde{v}^k)-\alpha_kH^{-1}d(v^k,\tilde{v}^k)\|_H^2\\ &= 2\alpha_k(v^k-\tilde{v}^k)^{\mathrm{T}}d(v^k,\tilde{v}^k)-\alpha_k^2\|H^{-1}d(v^k,\tilde{v}^k)\|_H^2\\ &= q_k(\alpha_k).\end{aligned}$$

代入 (7.7) 的右端就得到 (7.6), 引理得证. □

引理 7.1 是证明第一类算法收敛速率的基础. 对第二类方法, 我们也有相同的结论.

引理 7.2 设 $\{\tilde{w}^k\}$ 和 $\{v^k\}$ 是算法中分别由预测 (1.2) 和第二类校正方法 (1.6b) 产生的序列, 则对所有的 k 都有

$$\alpha_k(\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^{\mathrm{T}}F(w)) \geqslant \frac{1}{2}(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2)+\frac{1}{2}q_k(\alpha_k), \quad \forall\, w\in\Omega, \tag{7.8}$$

其中 $q_k(\cdot)$ 由 (5.3) 给出.

证明 把 (7.8) 的左端 $\alpha_k(\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^{\mathrm{T}}F(w))$ 分拆成

$$\begin{aligned}\alpha_k(\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^{\mathrm{T}}F(w)) &= \alpha_k(\theta(u^{k+1})-\theta(\tilde{u}^k)+(w^{k+1}-\tilde{w}^k)^{\mathrm{T}}F(w))\\ &\quad+\alpha_k(\theta(u)-\theta(u^{k+1})+(w-w^{k+1})^{\mathrm{T}}F(w))\end{aligned} \tag{7.9}$$

分别处理. 先用预测的性质处理 (7.9) 右端的第一部分 $\alpha_k(\theta(u^{k+1})-\theta(\tilde{u}^k)+(w^{k+1}-\tilde{w}^k)^{\mathrm{T}}F(w))$. 因为 $w^{k+1}\in\Omega$, 将 (7.4) 中的 w 设成 w^{k+1}, 就得

$$\begin{aligned}&\alpha_k(\theta(u^{k+1})-\alpha\theta(\tilde{u}^k)+(w^{k+1}-\tilde{w}^k)^{\mathrm{T}}F(w))\\ &\quad\geqslant \alpha_k(v^{k+1}-\tilde{v}^k)^{\mathrm{T}}d(v^k,\tilde{v}^k)\end{aligned}$$

$$= \alpha_k(v^k - \tilde{v}^k)^{\mathrm{T}} d(v^k, \tilde{v}^k) - \alpha_k(v^k - v^{k+1})^{\mathrm{T}} d(v^k, \tilde{v}^k). \tag{7.10}$$

对 (7.10) 右端的第二个交叉项用 Cauchy-Schwarz 不等式得到

$$-\alpha_k(v^k - v^{k+1})^{\mathrm{T}} d(v^k, \tilde{v}^k) \geqslant -\frac{1}{2}\|v^k - v^{k+1}\|_H^2 - \frac{1}{2}\alpha_k^2 \|H^{-1} d(v^k, \tilde{v}^k)\|_H^2.$$

将上式代入 (7.10) 的右端, 并利用 $q_k(\alpha_k)$ 的定义, 就有

$$\alpha_k(\theta(u^{k+1}) - \alpha\theta(\tilde{u}^k) + (w^{k+1} - \tilde{w}^k)^{\mathrm{T}} F(w)) \geqslant \frac{1}{2} q_k(\alpha_k) - \frac{1}{2}\|v^k - v^{k+1}\|_H^2. \tag{7.11}$$

再用校正的性质处理 (7.9) 右端的第二部分 $\alpha_k(\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^{\mathrm{T}} F(w))$. 因为 w^{k+1} 是问题 (1.6b) 的最优解, 根据引理 2.1, 有

$$\alpha_k(\theta(u) - \theta(u^{k+1})) + (w - w^{k+1})^{\mathrm{T}} \mathcal{H}\{(w^{k+1} - w^k) + \alpha_k \mathcal{H}^{-1} F(w^{k+1})\} \geqslant 0.$$

利用单调性, $(w - w^{k+1})^{\mathrm{T}} F(w) \geqslant (w - w^{k+1})^{\mathrm{T}} F(w^{k+1})$, 从上式可以得到

$$\alpha_k(\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^{\mathrm{T}} F(w)) \geqslant (w - w^{k+1})^{\mathrm{T}} \mathcal{H}(w^k - w^{k+1}).$$

在 v 是 w 的真子向量的情形下, 令 $h \to 0_+$ (见 (1.5) 中的 $\mathcal{H}$), 得到

$$\alpha_k(\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^{\mathrm{T}} F(w)) \geqslant (v - v^{k+1})^{\mathrm{T}} H(v^k - v^{k+1}).$$

对上式右端利用恒等式 $a^{\mathrm{T}} H b = \frac{1}{2}\{\|a\|_H^2 - \|a - b\|_H^2 + \|b\|_H^2\}$, 我们得到

$$\alpha_k(\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^{\mathrm{T}} F(w)) \geqslant \frac{1}{2}(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) + \frac{1}{2}\|v^k - v^{k+1}\|_H^2. \tag{7.12}$$

将 (7.11) 和 (7.12) 相加, 得到 (7.8) 即完成了证明. □

对两类不同算法分别有引理 7.1 和 7.2, 它们的结论完全相同. 由于 $q_k(\alpha_k) > 0$, 我们从 (7.5) (和与其相同的 (7.8)) 得到证明迭代复杂度 $O(1/t)$ 的关键不等式:

$$\alpha_k(\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^{\mathrm{T}} F(w)) + \frac{1}{2}\|v - v^k\|_H^2 \geqslant \frac{1}{2}\|v - v^{k+1}\|_H^2, \quad \forall\, w \in \Omega. \tag{7.13}$$

定理 7.2 设 $\{\tilde{w}^k\}$ 和 $\{v^k\}$ 为算法中分别由预测 (1.2) 和校正 (1.6a) (或 (1.6b)) 产生的序列, 则对所有的 $t > 0$ 都有 $\tilde{w}_t \in \Omega$ 和

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^{\mathrm{T}} F(w) \leqslant \frac{1}{2\Upsilon_t}\|v - v^0\|_H^2, \quad \forall\, w \in \Omega, \tag{7.14}$$

其中

$$\tilde{w}_t = \frac{1}{\Upsilon_t}\sum_{k=0}^{t} \alpha_k \tilde{w}^k, \quad \Upsilon_t = \sum_{k=0}^{t} \alpha_k. \tag{7.15}$$

证明 首先, 因为对每个 $k > 0$, 都有 $\tilde{w}^k \in \Omega$, (7.15) 表明 $\tilde{w}_t$ 是它们的凸组合, 所以 $\tilde{w}_t \in \Omega$. 将对应于 $k = 0, 1, \ldots, t$ 的不等式 (7.13) 相加, 得到

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

利用记号 $\tilde{w}_t$, 上式可以写成

$$\frac{1}{\Upsilon_t}\sum_{k=0}^{t}\alpha_k\theta(\tilde{u}^k)-\theta(u)+(\tilde{w}_t-w)^{\mathrm{T}}F(w)\leqslant\frac{1}{2\Upsilon_t}\|v-v^0\|_H^2,\quad\forall\,w\in\Omega. \tag{7.16}$$

由于 $\theta(u)$ 是凸函数, 并由 (7.15) 可知,

$$\tilde{u}_t=\frac{1}{\Upsilon_t}\sum_{k=0}^{t}\alpha_k\tilde{u}^k$$

是 $\tilde{u}^0,\tilde{u}^1,\ldots,\tilde{u}^t$ 的凸组合, 因此有

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

将其代入 (7.16), 就得到了定理的结论. □

根据 (5.7), 对所有的 $k>0$, 有 $\alpha_k^*\geqslant\gamma\alpha_{\min}$. 由 $\alpha_k=\gamma\alpha_k^*$, 有

$$\Upsilon_t=\sum_{k=0}^{t}\alpha_k\geqslant(t+1)\gamma\alpha_{\min}.$$

代入 (7.14) 并利用 $\mathcal{D}_{(\tilde{w})}\subset\Omega$ (见 (7.2)) 就得到

$$\theta(\tilde{u}_t)-\theta(u)+(\tilde{w}_t-w)^{\mathrm{T}}F(w)\leqslant\frac{1}{t+1}\cdot\frac{\|v-v^0\|_H^2}{2\gamma\alpha_{\min}},\quad\forall\,w\in\mathcal{D}_{(\tilde{w})}.$$

结合 (7.3), 这就证明了方法的 $O(1/t)$ 迭代复杂度.

8 结论与建议

本文对凸优化和单调变分不等式的收缩算法给出统一框架, 证明了其中的孪生方法具有相同的 $O(1/t)$ 迭代复杂性. 这些方法包括了经典单调变分不等式的投影收缩算法、变分不等式框架下定制的 PPA 算法[11,26]、乘子交替方向法[27,28] 和处理多个可分离算子的分裂收缩算法[9,10].

经典的单调变分不等式 (1.9) 方面, 已有的计算步长的算法属于第一类算法[7,18,19]. 采用相同步长的孪生方法[5,11–13], 以及研究孪生算法迭代复杂性的文献 [29] 的结果, 都包含在这个框架. 对由凸优化问题转换来的变分不等式, 按需定制的邻近点算法 (customized PPA) [11,14,25], 改进的乘子交替方向法[20,30], 以及处理多个可分离算子的分裂收缩算法[9,10,14], 属于第一类算法. 这些方法能被工程界和相关学科领域重视和采用 (参见文献 [15,22]), 是机理简单的原因.

已有的计算实践证明, 对经典的单调变分不等式 (1.9), 处理同一个问题, 第二类算法更快. 基于这个理由, 对由凸优化问题转换来的变分不等式, 在子问题 (6.12) 容易求解的情形下, 我们建议采用孪生的第二类算法, 除了提高计算效率, 同时还使迭代序列 $\{w^k\}$ 一直都在 Ω 内.

致谢 感谢匿名审稿人提出的宝贵意见与建议.

参考文献

1 Zheng H, Liu F, Du X. Complementarity problem arising from static growth of multiple cracks and MLS-based numerical manifold method. Comput Methods Appl Mech Engrg, 2015, 295: 150–171

2 Zheng H, Zhang P, Du X. Dual form of discontinuous deformation analysis. Comput Methods Appl Mech Engrg, 2016, 305: 196–216
3 郑宏, 葛修润. 节理岩体的变分不等式模型. 中国科学 B 辑, 1995, 25: 986–993
4 Facchinei F, Pang J S. Finite-Dimensional Variational Inequalities and Complementarity Problems, vol. I. New York: Springer-Verlag, 2003
5 He B S, Liao L Z. Improvements of some projection methods for monotone nonlinear variational inequalities. J Optim Theory Appl, 2002, 112: 111–128
6 He B, Yuan X, Zhang J J Z. Comparison of two kinds of Prediction-Correction methods for monotone variational inequalities. Comput Optim Appl, 2004, 27: 247–267
7 He B S. A class of projection and contraction methods for monotone variational inequalities. Appl Math Optim, 1997, 35: 69–76
8 Gu G, He B, Yuan X. Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: A unified approach. Comput Optim Appl, 2014, 59: 135–161
9 He B, Tao M, Yuan X. Alternating direction method with Gaussian back substitution for separable convex programming. SIAM J Optim, 2012, 22: 313–340
10 He B S, Tao M, Yuan X M. A splitting method for separable convex programming. IMA J Numer Anal, 2015, 31: 394–426
11 He B, Yuan X, Zhang W. A customized proximal point algorithm for convex minimization with linear constraints. Comput Optim Appl, 2013, 56: 559–572
12 He B, Liao L Z, Wang X. 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
13 He B, Liao L Z, Wang X. 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
14 He B, Yuan X. Convergence analysis of Primal-Dual algorithms for a Saddle-Point problem: From contraction perspective. SIAM J Imaging Sci, 2012, 5: 119–149
15 Pock T, Chambolle A. Diagonal preconditioning for first-order primal-dual algorithms in convex optimization. In: Proceedings of the 2011 IEEE International Conference on the Computer Vision (ICCV). Barcelona: IEEE, 2011, 1762–1769
16 Chen C H, He B S, Ye Y, et al. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. Math Program, 2015, 155: 57–79
17 Blum E, Oettli W. Mathematische Optimierung, Econometrics and Operations Research. New York: Springer, 1975
18 Solodov M V, Tseng P. Modified projection-type methods for monotone variational inequalities. SIAM J Control Optim, 1996, 34: 1814–1830
19 Sun D. A class of iterative methods for solving nonlinear projection equations. J Optim Theory Appl, 1996, 91: 123–140
20 He B S, Liu H, Wang Z R, et al. A strictly contractive Peaceman-Rachford splitting method for convex programming. SIAM J Optim, 2014, 24: 1011–1040
21 He B, Yuan X. On the $O(1/n)$ convergence rate of the douglas-rachford alternating direction method. SIAM J Numer Anal, 2012, 50: 700–709
22 Esser E, Möller M, Osher S, et al. A convex model for non-negative matrix factorization and dimensionality reduction on physical space. IEEE Trans Image Process, 2012, 21: 3239–3252
23 Martinet B. Régularisation, d'inéquations variationelles par approximations succesives. Rev Franc Inform Rech Oper, 1970, 4: 154–159
24 Rockafellar R T. Monotone operators and the proximal point algorithm. SIAM J Control Optim, 1976, 14: 877–898
25 何炳生, 申远. 求解凸规划及鞍点问题定制的 PPA 算法及其收敛速率. 中国科学: 数学, 2012, 42: 515–525
26 He B S. PPA-like contraction methods for convex optimization: A framework using variational inequality approach. J Oper Res Soc China, 2015, 3: 391–420
27 Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Mach Learn, 2010, 3: 1–122
28 Glowinski R. Numerical Methods for Nonlinear Variational Problems. New York-Berlin-Heidelberg-Tokyo: Springer-Verlag, 1984
29 Cai X, Gu G, He B. On the $O(1/t)$ convergence rate of the projection and contraction methods for variational inequalities with Lipschitz continuous monotone operators. Comput Optim Appl, 2014, 57: 339–363
30 Cai X J, Gu G Y, He B S, et al. A proximal point algorithms revisit on the alternating direction method of multipliers. Sci China Math, 2013, 56: 2179–2186

A uniform framework of contraction methods for convex optimization and monotone variational inequality

Bingsheng He

Abstract Linearly constrained convex optimization problems can be reformulated as monotone variational inequalities, which include linearly constrained convex optimization problems as special cases. Usually, deriving and studying optimization methods in the framework of variational inequalities are often more convenient. Based on the twin directions and identical step size, there exists a unified predictor-corrector framework for projection-contraction methods solving monotone variational inequalities. According to the twin directions, the unified framework falls into two classes with similar computational cost in each iteration, while totally the second class outperforms the first as confirmed in numerical experiments. Most recently developed splitting contraction methods belong to a more general framework, in which similar twin directions can be generated. However, until now, all splitting contraction methods can be seen as generalizations of the methods in the first class of the unified predictor-corrector framework. In this paper, the generalization of methods in the second class is derived using existing step size and the other generalized direction. The $O(1/t)$ iteration complexity is proved in the generalized unified framework.

Keywords **convex optimization, monotone variational inequality, projection-contraction method, splitting contraction method, unified framework, twin directions and same step size**

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

doi: 10.1360/N012017-00034