

一种部分基于 Bregman 距离的 Douglas-Rachford 分裂算法^{*}

高 思, 罗洪林

(重庆师范大学 数学科学学院, 重庆 401331)

摘要:针对非凸函数求和的最小化问题,提出了一种部分基于 Bregman 距离的 Douglas-Rachford 分裂算法。提出的算法通过引入 Bregman 距离函数,对经典 Douglas-Rachford 分裂框架中的部分子问题进行非欧几里得推广,从而扩展了该类算法在非欧几何结构下的应用。在理论分析方面,通过构造价值函数并且结合 Kurdyka-Łojasiewicz 性质,证明了算法迭代序列具有全局收敛性。在数值实验中,以非凸可行性问题为测试基准,并针对含 Bregman 距离的子问题进行了专门的数值计算设计。研究结果显示本文提出的算法在收敛速度、计算精度与数值稳定性方面均明显优于传统算法。

关键词:非凸优化;Douglas-Rachford 分裂算法;Bregman 临近映射;Kurdyka-Łojasiewicz 性质

中图分类号:O221

文献标志码:A

文章编号:1672-6693(2026)03-0030-10

Douglas-Rachford 分裂算法(后简记为 DR 算法)是一种经典的优化算法,最初由 Douglas 等人^[1]提出,用于热微分方程的数值求解。随后,Lions 等人^[2]将 DR 算法应用于 2 个闭凸函数之和的最小化问题。DR 算法能够有效处理具有竞争结构的问题,例如寻找 2 个闭凸集交集的点(可行性问题),该问题可转化为最小化 2 个集合指示函数之和的优化问题。Eckstein 等人^[3]在凸设定下进一步研究了 DR 算法,揭示了该算法与近端点算法的关系。目前,DR 算法已被广泛应用于信号处理等领域的凸优化问题研究中,详见文献[4-7]。

在非凸设定下,有关 DR 算法的加速方法已有较多研究,包括惯性 DR 算法^[8]、带松弛参数的 DR 算法^[9]、参数化 DR 算法^[10]以及惯性参数化 DR 算法^[11]。对于凸情形而言,DR 算法在过去数十年中得到了广泛应用,相关研究在收敛性证明及各种推广扩展方面取得了明显进展。关于 DR 算法的更多细节参见文献[12-13]。然而,上述研究都是在欧式情况下开展,它们在子问题计算方面显得不够灵活。有鉴于此,本文提出一种部分基于 Bregman 距离的 DR 算法(后简记为 PBDR 算法),用于求解非凸复合问题:

$$\min_{x \in \mathbf{R}^n} F(x) = f(x) + g(x), \quad (1)$$

其中: f 具有 Lipschitz 连续梯度,它的 Lipschitz 常数以 $L_{\nabla f}$ 为界,且假设 $f(x) + \frac{l}{2} \|x\|^2$ 为凸函数, $g: \mathbf{R}^n \rightarrow \overline{\mathbf{R}} = (-\infty, \infty]$ 为真下半连续函数。特别地,当 $h = \frac{1}{2} \|\cdot\|^2$ 时,PBDR 算法在忽略某些常数项后即退化为经典 DR 算法。本研究通过引入 Bregman 距离并且运用特殊设计的价值函数序列,将 DR 算法扩展至非欧几里得设定。该框架允许灵活选择 Bregman 距离中的核函数 h ,从而更好地适应目标函数的局部几何特性。通过充分利用目标函数的几何性质,PBDR 算法能够获得更优的收敛速度。

1 记号与基础知识

本文使用 $\mathbf{R}^n$ 表示 n 维欧几里得空间, $\mathbf{N}$ 表示自然数集合。 $\langle \cdot, \cdot \rangle$ 表示内积 $\langle x, y \rangle = x^T y$, $\|\cdot\|$ 表示由该内积诱导的范数 $\|x\| = \sqrt{\langle x, x \rangle}$, 其中 $x, y \in \mathbf{R}^n$, $\cdot^{(i)}$ 代表向量 $\cdot$ 的第 i 个分量。对于一个扩展实值函数 f , 它的定义域为 $\text{dom } f := \{x \in \mathbf{R}^n : f(x) < +\infty\}$ 。若 $\text{dom } f \neq \emptyset$ 且函数值不恒为 $-\infty$, 则称该函数为真函数。若函数是

^{*} 收稿日期:2025-09-24 修回日期:2026-02-23 网络出版时间:2026-04-27T14:18

资助项目:国家自然科学基金面上项目(No. 11771064);重庆市教育委员会科学技术研究计划重点项目(No. KJZD-K202500507)

第一作者简介:高思,女,研究方向为最优化理论与算法,E-mail:gaosi2025@163.com;通信作者简介:罗洪林,男,教授,博士,E-mail:luohonglin@cqnu.edu.cn

网络出版地址:https://link.cnki.net/urlid/50.1165.n.20260425.1457.002

下半连续的,则称之为闭函数。对于一个凸的、下半连续的函数 $f: \mathbf{R}^n \rightarrow \bar{\mathbf{R}} = (-\infty, \infty]$, Fenchel 共轭函数 $f^*: \mathbf{R}^n \rightarrow \bar{\mathbf{R}}$ 定义为 $f^*(\mathbf{y}) := \sup_{\mathbf{x} \in \mathbf{R}^n} \{\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x})\}$ 。对于一个真函数 $f: \mathbf{R}^n \rightarrow \bar{\mathbf{R}}$, 在 $\bar{\mathbf{x}} \in \text{dom } f$ 处的 Fréchet 次微分 (也称为正则次微分) 定义为:

$$\hat{\partial} f(\bar{\mathbf{x}}) := \left\{ \mathbf{v} \in \mathbf{R}^n : \liminf_{\mathbf{x} \rightarrow \bar{\mathbf{x}}, \mathbf{x} \neq \bar{\mathbf{x}}} \frac{f(\mathbf{x}) - f(\bar{\mathbf{x}}) - \langle \mathbf{v}, \mathbf{x} - \bar{\mathbf{x}} \rangle}{\|\mathbf{x} - \bar{\mathbf{x}}\|} \geq 0 \right\}.$$

记 $\mathbf{y} \xrightarrow{f} \mathbf{x}$ 表示 $\mathbf{y} \rightarrow \mathbf{x}$ 且 $f(\mathbf{y}) \rightarrow f(\mathbf{x})$ 。则基本次微分 (也称为极限次微分) 定义为:

$$\partial f(\bar{\mathbf{x}}) := \{ \mathbf{v} \in \mathbf{R}^n : \exists (\mathbf{x}_t, \mathbf{v}_t) \text{ s. t. } \mathbf{x}_t \xrightarrow{f} \bar{\mathbf{x}}, \mathbf{v}_t \in \hat{\partial} f(\mathbf{x}_t), \mathbf{v}_t \rightarrow \mathbf{v} \}, \quad (2)$$

其中 $t \rightarrow +\infty$ 且 $t \in \mathbf{N} \setminus \{0\}$ 。上述定义直接给出如下稳定性性质:

$$\{ \mathbf{v} \in \mathbf{R}^n : \exists \mathbf{x}_t \xrightarrow{f} \mathbf{x}, \mathbf{v}_t \rightarrow \mathbf{v}, \mathbf{v}_t \in \partial f(\mathbf{x}_t) \} \subseteq \partial f(\mathbf{x}), \quad (3)$$

同时使用记号 $\text{dom } \partial f := \{ \mathbf{x} \in \mathbf{R}^n : \partial f(\mathbf{x}) \neq \emptyset \}$ 。若 f 连续可微, 则次微分式 (2) 退化为 f 的导数 ∇f 。另一方面, 若 f 为凸函数, 则式 (2) 退化为凸分析中的经典次微分 (例如文献 [14] 中的命题 8.12), 即:

$$\partial f(\mathbf{x}) = \{ \mathbf{v} \in \mathbf{R}^n : \langle \mathbf{v}, \mathbf{y} - \mathbf{x} \rangle \leq f(\mathbf{y}) - f(\mathbf{x}), \forall \mathbf{y} \in \mathbf{R}^n \}.$$

对于具有多组变量的函数 f , 记 $\partial_{\mathbf{x}} f(\nabla_{\mathbf{x}} f)$ 为 f 关于变量 $\mathbf{x}$ 的次微分 (导数)。若 $\liminf_{\|\mathbf{x}\| \rightarrow \infty} f(\mathbf{x}) = \infty$, 则称函数 f 是强制的。最后, 若 $f - \frac{\sigma}{2} \|\cdot\|^2$ 是凸函数, 则称函数 f 是系数为 $\sigma > 0$ 的强凸函数。

对于闭集 $S \subseteq \mathbf{R}^n$, 它的指示函数定义为 $\delta_S(\mathbf{x}) = 0$ (若 $\mathbf{x} \in S$) 和 $\delta_S(\mathbf{x}) = \infty$ (否则)。此外, S 在 $\mathbf{x} \in S$ 处的 (极限) 法锥由 $N_S(\mathbf{x}) = \partial \delta_S(\mathbf{x})$ 给出。子集 $S \subseteq \mathbf{R}^n$ 的距离函数为:

$$\text{dist}(\cdot, S): \mathbf{R}^n \rightarrow \bar{\mathbf{R}}, \mathbf{x} \mapsto \text{dist}(\mathbf{x}, S) = \inf \{ \|\mathbf{x} - \mathbf{y}\| : \mathbf{y} \in S \},$$

若集合 S 是闭凸的, 则使用 $P_S(\mathbf{x})$ 表示 $\mathbf{x}$ 到 S 的投影。

定义 1^[15] (半代数) 1) 若存在有限多个多项式 $g_{ij}, h_{ij}: \mathbf{R}^n \rightarrow \mathbf{R}$ 使得 $S = \bigcup_{j=1}^p \bigcap_{i=1}^q \{ \mathbf{x} \in \mathbf{R}^n : g_{ij}(\mathbf{x}) = 0 \ \& \ h_{ij}(\mathbf{x}) < 0 \}$, 则称集合 $S \subseteq \mathbf{R}^n$ 为半代数集; 2) 若函数图像 $\text{gph } f = \{ (\mathbf{x}, \mathbf{y}) \in \mathbf{R}^{n+1} : f(\mathbf{x}) = \mathbf{y} \}$ 是半代数集, 则称函数 $f: \mathbf{R}^n \rightarrow \bar{\mathbf{R}}$ 为半代数函数。

定义 2^[16] (Kurdyka-Łojasiewicz (KL) 性质与 KL 函数) 若存在 $\bar{\mathbf{x}}$ 的邻域 $U, \eta \in (0, \infty]$ 和连续凹函数 $\varphi: [0, \eta) \rightarrow \mathbf{R}_+ = [0, +\infty)$ 使得:

1) $\varphi(0) = 0$ 且 φ 在 $(0, \eta)$ 上连续可微, $\varphi' > 0$;

2) 对所有满足 $f(\bar{\mathbf{x}}) < f(\mathbf{x}) < f(\bar{\mathbf{x}}) + \eta$ 的 $\mathbf{x} \in U$, 有 $\varphi'(f(\mathbf{x}) - f(\bar{\mathbf{x}})) \text{dist}(\mathbf{0}, \partial f(\mathbf{x})) \geq 1$;

则称真函数 f 在 $\bar{\mathbf{x}} \in \text{dom } \partial f$ 处具有 KL 性质。若真闭函数 f 在 $\text{dom } \partial f$ 中的所有点均满足 KL 性质, 则称该函数为 KL 函数。

注 1^[15] 设 $f: \mathbf{R}^n \rightarrow \bar{\mathbf{R}}$ 为真下半连续函数, 且 $\bar{\mathbf{x}} \in \text{dom } \partial f$ 。若 f 是半代数函数, 则它在 $\bar{\mathbf{x}}$ 处具有凹 KL 性质, 且 $\varphi(t) = \xi \cdot t^{1-\theta}$, 其中: $\xi > 0, \theta \in (0, 1)$ 。

由可微函数 $h: \mathbf{R}^n \rightarrow \mathbf{R}$ 诱导的 Bregman 距离定义为:

$$D_h: \mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R}; (\mathbf{x}, \mathbf{y}) \mapsto h(\mathbf{x}) - h(\mathbf{y}) - \langle \nabla h(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle.$$

引理 1^[17] 设 h 为 σ -强凸且具有 Lipschitz 连续梯度 (常数为 $L_{\nabla h}$)。则 $\forall \mathbf{x}, \mathbf{y} \in \mathbf{R}^n$, 满足:

$$\frac{\sigma}{2} \|\mathbf{x} - \mathbf{y}\|^2 \leq D_h(\mathbf{x}, \mathbf{y}) \leq \frac{L_{\nabla h}}{2} \|\mathbf{x} - \mathbf{y}\|^2. \quad (4)$$

2 PBDR 算法与收敛性分析

本文所有研究基于以下假设进行。

假设 1 1) $f: \mathbf{R}^n \rightarrow \bar{\mathbf{R}} = (-\infty, \infty]$ 为真函数, 且具有 Lipschitz 连续梯度, Lipschitz 常数为 $L_{\nabla f}$, $g: \mathbf{R}^n \rightarrow \bar{\mathbf{R}}$ 为真下半连续函数。2) $f(\mathbf{x}) + \frac{l}{2} \|\mathbf{x}\|^2$ 为凸函数, 且 $l \in [-L_{\nabla f}, +\infty)$ 。3) f 与 g 至少有 1 个为强制函数, 且都有下界。

本文建立的 PBDR 算法如下。

算法 1 PBDR 算法。

步骤 1, 输入初始点 $\mathbf{x}_0$ 与步长参数 λ ;

$$\text{步骤 2, } \begin{cases} \mathbf{y}_{t+1} \in \arg \min_y \left\{ f(\mathbf{y}) + \frac{1}{2\gamma} \|\mathbf{y} - \mathbf{x}_t\|^2 \right\} \\ \mathbf{z}_{t+1} \in \arg \min_z \left\{ g(\mathbf{z}) + \frac{1}{\gamma} D_h(\mathbf{z}, \mathbf{y}_{t+1}) + \frac{1}{\gamma} \langle \mathbf{x}_t - \mathbf{y}_{t+1}, \mathbf{z} \rangle \right\}; \\ \mathbf{x}_{t+1} = \mathbf{x}_t + (\mathbf{z}_{t+1} - \mathbf{y}_{t+1}) \end{cases}$$

步骤 3, 如果没有达到收敛条件, 回到步骤 1。

下面给出 1 个在证明与理解收敛性方面有很大帮助的注解。

注 2 对于算法框架有以下最优化结果:

$$\mathbf{0} = \nabla f(\mathbf{y}_{t+1}) + \frac{1}{\gamma}(\mathbf{y}_{t+1} - \mathbf{x}_t), \mathbf{0} \in \partial g(\mathbf{z}_{t+1}) + \frac{1}{\gamma}(\nabla h(\mathbf{z}_{t+1}) - \nabla h(\mathbf{y}_{t+1})) + \frac{1}{\gamma}(\mathbf{x}_t - \mathbf{y}_{t+1}). \quad (5)$$

因此 $\forall t \geq 1$, 有:

$$\mathbf{0} \in \nabla f(\mathbf{y}_t) + \partial g(\mathbf{z}_t) + \frac{1}{\gamma}(\nabla h(\mathbf{z}_t) - \nabla h(\mathbf{y}_t)). \quad (6)$$

于是, 若 $\lim_{t \rightarrow \infty} \|\mathbf{x}_{t+1} - \mathbf{x}_t\| = \lim_{t \rightarrow \infty} \|\mathbf{z}_{t+1} - \mathbf{y}_{t+1}\| = 0$, 且如果 $\{(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_t)\}$ 存在至少 1 个聚点 $(\bar{\mathbf{y}}, \bar{\mathbf{z}}, \bar{\mathbf{x}})$ 及收敛子列 $\lim_{j \rightarrow \infty} (\mathbf{y}_{t_j}, \mathbf{z}_{t_j}, \mathbf{x}_{t_j}) = (\bar{\mathbf{y}}, \bar{\mathbf{z}}, \bar{\mathbf{x}})$ 满足 $\lim_{j \rightarrow \infty} g(\mathbf{z}_{t_j}) = g(\bar{\mathbf{z}})$, 则沿该子列在式(6)中取极限并利用式(3), 不难看出 $(\bar{\mathbf{y}}, \bar{\mathbf{z}})$ 给出式(1)的一个稳定点, 即满足 $\bar{\mathbf{y}} = \bar{\mathbf{z}}$ 且 $\mathbf{0} \in \nabla f(\bar{\mathbf{z}}) + \partial g(\bar{\mathbf{z}})$ 。

2.1 PBDR 算法价值函数与参数假设

首先给出 PBDR 算法的价值函数形式, 它来源于文献[16]。与文献[16]相比, 本文中该函数的等价形式有所不同。

定义 3 设 $\gamma > 0$ 。PBDR 算法的价值函数定义为:

$$D_\gamma(\mathbf{y}, \mathbf{z}, \mathbf{x}) = f(\mathbf{y}) + g(\mathbf{z}) + \frac{1}{2\gamma}(\|\mathbf{y} - \mathbf{x}\|^2 - \|\mathbf{z} - \mathbf{x}\|^2). \quad (7)$$

下面给出价值函数的几个等价形式, 这在证明价值函数下降性中很有用处。

注 3 D_γ 可等价写为:

$$\begin{aligned} D_\gamma(\mathbf{y}, \mathbf{z}, \mathbf{x}) &= f(\mathbf{y}) + g(\mathbf{z}) + \frac{1}{2\gamma}(\|\mathbf{y} - \mathbf{z}\|^2 + 2\langle \mathbf{z} - \mathbf{x}, \mathbf{y} - \mathbf{z} \rangle) = \\ &= f(\mathbf{y}) + g(\mathbf{z}) + \frac{1}{\gamma} D_h(\mathbf{z}, \mathbf{y}) + \frac{1}{\gamma} \langle \mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{z} \rangle - \frac{1}{\gamma} \|\mathbf{z} - \mathbf{y}\|^2 - \frac{1}{\gamma} D_h(\mathbf{z}, \mathbf{y}), \end{aligned} \quad (8)$$

其中: 第 1 个等式由关系 $\|\mathbf{y} - \mathbf{x}\|^2 = \|\mathbf{y} - \mathbf{z}\|^2 + \|\mathbf{z} - \mathbf{x}\|^2 + 2\langle \mathbf{z} - \mathbf{x}, \mathbf{y} - \mathbf{z} \rangle$ 得到, 第 2 个等式则通过向式(7)中添加项 $\frac{1}{\gamma} D_h(\mathbf{z}, \mathbf{y}) + \frac{1}{\gamma} \langle \mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{z} \rangle$ 得到, 即:

$$\begin{aligned} D_\gamma(\mathbf{y}, \mathbf{z}, \mathbf{x}) &= f(\mathbf{y}) + g(\mathbf{z}) + \frac{1}{2\gamma}(\|\mathbf{y} - \mathbf{x}\|^2 - \|\mathbf{z} - \mathbf{x}\|^2) + \frac{1}{\gamma} D_h(\mathbf{z}, \mathbf{y}) + \frac{1}{\gamma} \langle \mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{z} \rangle - \\ &= \frac{1}{\gamma} D_h(\mathbf{z}, \mathbf{y}) + \frac{1}{\gamma} \langle \mathbf{y} - \mathbf{x}, \mathbf{z} - \mathbf{y} \rangle + f(\mathbf{y}) + g(\mathbf{z}) + \frac{1}{2\gamma}(\|\mathbf{y} - \mathbf{x}\|^2 - \|\mathbf{z} - \mathbf{x}\|^2) + \\ &= \frac{1}{\gamma} D_h(\mathbf{z}, \mathbf{y}) + \frac{1}{\gamma} \langle \mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{z} \rangle - \frac{1}{\gamma} D_h(\mathbf{z}, \mathbf{y}) + \frac{1}{2\gamma}(\|\mathbf{z} - \mathbf{x}\|^2 - \|\mathbf{z} - \mathbf{y}\|^2 - \|\mathbf{y} - \mathbf{x}\|^2) = \\ &= f(\mathbf{y}) + g(\mathbf{z}) + \frac{1}{\gamma} D_h(\mathbf{z}, \mathbf{y}) + \frac{1}{\gamma} \langle \mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{z} \rangle - \frac{1}{\gamma} (D_h(\mathbf{z}, \mathbf{y}) + \frac{1}{2} \|\mathbf{z} - \mathbf{y}\|^2). \end{aligned}$$

接下来给出算法收敛性分析的前提假设, 也就是收敛性需要满足的前提条件。

假设 2 1) $\sigma_h \geq 1, \gamma > 0$; 2) 令 $M_\gamma = \frac{(1+L_{\nabla h})}{2}(\gamma L_{\nabla f} + 1)^2 + \left(L_{\nabla h} + \frac{3}{2}\right)\gamma l - \frac{(1+L_{\nabla h}) + 1}{2}$, 有 $M_\gamma < 0$; 3) 算

法 1 产生的序列 $(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_t)$ 有界。

2.2 PBDR 算法的收敛性

现在便给出第1个收敛性结果,其中包括了价值函数的下降性以及子序列的收敛性。

定理1 (全局子序列收敛性)若假设1、2成立,那么 $\{D_\gamma(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_t)\}_{t \geq 1}$ 是非增的,且有

$$\lim_{t \rightarrow \infty} \|\mathbf{x}_{t+1} - \mathbf{x}_t\| = \lim_{t \rightarrow \infty} \|\mathbf{z}_{t+1} - \mathbf{y}_{t+1}\| = 0 \quad (9)$$

成立。进一步地,对任意聚点 $(\bar{\mathbf{y}}, \bar{\mathbf{z}}, \bar{\mathbf{x}})$,有 $\bar{\mathbf{z}} = \bar{\mathbf{y}}$ 以及 $\mathbf{0} \in \nabla f(\bar{\mathbf{z}}) + \partial g(\bar{\mathbf{z}})$ 成立。

证明 首先,由式(8)中的第1个关系注意到:

$$\begin{aligned} D_\gamma(\mathbf{y}_{t+1}, \mathbf{z}_{t+1}, \mathbf{x}_{t+1}) - D_\gamma(\mathbf{y}_{t+1}, \mathbf{z}_{t+1}, \mathbf{x}_t) &= \frac{1}{\gamma} (\langle \mathbf{z}_{t+1} - \mathbf{x}_{t+1}, \mathbf{y}_{t+1} - \mathbf{z}_{t+1} \rangle - \langle \mathbf{z}_{t+1} - \mathbf{x}_t, \mathbf{y}_{t+1} - \mathbf{z}_{t+1} \rangle) = \\ &= \frac{1}{\gamma} \langle \mathbf{x}_t - \mathbf{x}_{t+1}, \mathbf{y}_{t+1} - \mathbf{z}_{t+1} \rangle = \frac{1}{\gamma} \|\mathbf{x}_{t+1} - \mathbf{x}_t\|^2. \end{aligned} \quad (10)$$

接下来,利用引理1中式(4)并结合式(8)中的第2个关系,可以得到:

$$\begin{aligned} D_\gamma(\mathbf{y}_{t+1}, \mathbf{z}_{t+1}, \mathbf{x}_t) - D_\gamma(\mathbf{y}_{t+1}, \mathbf{z}_t, \mathbf{x}_t) &= g(\mathbf{z}_{t+1}) + \frac{1}{\gamma} D_h(\mathbf{z}_{t+1}, \mathbf{y}_{t+1}) + \\ &+ \frac{1}{\gamma} \langle \mathbf{y}_{t+1} - \mathbf{x}_t, \mathbf{y}_{t+1} - \mathbf{z}_{t+1} \rangle - \frac{1}{\gamma} \left(D_h(\mathbf{z}_{t+1}, \mathbf{y}_{t+1}) + \frac{1}{2} \|\mathbf{z}_{t+1} - \mathbf{y}_{t+1}\|^2 \right) - g(\mathbf{z}_t) - \frac{1}{\gamma} D_h(\mathbf{z}_t, \mathbf{y}_{t+1}) - \\ &- \frac{1}{\gamma} \langle \mathbf{y}_{t+1} - \mathbf{x}_t, \mathbf{y}_{t+1} - \mathbf{z}_t \rangle + \frac{1}{\gamma} \left(D_h(\mathbf{z}_t, \mathbf{y}_{t+1}) + \frac{1}{2} \|\mathbf{z}_t - \mathbf{y}_{t+1}\|^2 \right) = g(\mathbf{z}_{t+1}) + \frac{1}{\gamma} D_h(\mathbf{z}_{t+1}, \mathbf{y}_{t+1}) + \\ &+ \frac{1}{\gamma} \langle \mathbf{x}_t - \mathbf{y}_{t+1}, \mathbf{z}_{t+1} \rangle - \frac{1}{\gamma} \left(D_h(\mathbf{z}_{t+1}, \mathbf{y}_{t+1}) + \frac{1}{2} \|\mathbf{z}_{t+1} - \mathbf{y}_{t+1}\|^2 \right) - g(\mathbf{z}_t) - \frac{1}{\gamma} D_h(\mathbf{z}_t, \mathbf{y}_{t+1}) - \\ &- \frac{1}{\gamma} \langle \mathbf{x}_t - \mathbf{y}_{t+1}, \mathbf{z}_t \rangle + \frac{1}{\gamma} \left(D_h(\mathbf{z}_t, \mathbf{y}_{t+1}) + \frac{1}{2} \|\mathbf{z}_t - \mathbf{y}_{t+1}\|^2 \right) \leq \\ &= \frac{1}{\gamma} \left[D_h(\mathbf{z}_t, \mathbf{y}_{t+1}) + \frac{1}{2} \|\mathbf{z}_t - \mathbf{y}_{t+1}\|^2 - \left(D_h(\mathbf{z}_{t+1}, \mathbf{y}_{t+1}) + \frac{1}{2} \|\mathbf{z}_{t+1} - \mathbf{y}_{t+1}\|^2 \right) \right] \leq \\ &= \frac{1}{\gamma} \left(\frac{1+L_{\nabla h}}{2} \|\mathbf{z}_t - \mathbf{y}_{t+1}\|^2 - \frac{1+\sigma_h}{2} \|\mathbf{x}_{t+1} - \mathbf{x}_t\|^2 \right). \end{aligned} \quad (11)$$

接下来,由式(5)中的第1个关系注意到

$$\nabla \left(f + \frac{l}{2} \|\cdot\|^2 \right) (\mathbf{y}_{t+1}) = \frac{1}{\gamma} (\mathbf{x}_t - \mathbf{y}_{t+1}) + l \mathbf{y}_{t+1},$$

由假设1的2)可知 $f + \frac{l}{2} \|\cdot\|^2$ 为凸函数,此时利用凸函数梯度的单调性,可以得到 $\forall t \geq 1$,有:

$$\begin{aligned} \left\langle \frac{1}{\gamma} (\mathbf{x}_t - \mathbf{y}_{t+1}) + l \mathbf{y}_{t+1} - \frac{1}{\gamma} (\mathbf{x}_{t-1} - \mathbf{y}_t) - l \mathbf{y}_t, \mathbf{y}_{t+1} - \mathbf{y}_t \right\rangle &\geq 0 \Rightarrow \\ \langle \mathbf{x}_t - \mathbf{x}_{t-1} - (\mathbf{y}_{t+1} - \mathbf{y}_t), \mathbf{y}_{t+1} - \mathbf{y}_t \rangle &\geq -\gamma l \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2 \Rightarrow \\ \langle \mathbf{x}_t - \mathbf{x}_{t-1}, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle &\geq \langle \mathbf{y}_{t+1} - \mathbf{y}_t, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle - \gamma l \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2 \geq (1 - \gamma l) \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2. \end{aligned}$$

因此,进一步得到:

$$\begin{aligned} \|\mathbf{y}_{t+1} - \mathbf{z}_t\|^2 &= \|\mathbf{y}_{t+1} - \mathbf{y}_t + \mathbf{y}_t - \mathbf{z}_t\|^2 = \|\mathbf{y}_{t+1} - \mathbf{y}_t - (\mathbf{x}_t - \mathbf{x}_{t-1})\|^2 = \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2 + \|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2 - \\ &- 2 \langle \mathbf{x}_t - \mathbf{x}_{t-1}, \mathbf{y}_{t+1} - \mathbf{y}_t \rangle \leq \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2 + \|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2 - 2(1 - \gamma l) \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2. \end{aligned} \quad (12)$$

将式(12)代入式(11),得到当 $t \geq 1$ 时,

$$\begin{aligned} D_\gamma(\mathbf{y}_{t+1}, \mathbf{z}_{t+1}, \mathbf{x}_t) - D_\gamma(\mathbf{y}_{t+1}, \mathbf{z}_t, \mathbf{x}_t) &\leq \\ &= \frac{1+L_{\nabla h}}{2\gamma} [(-1+2\gamma l) \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2 + \|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2] - \frac{1+\sigma_h}{2\gamma} \|\mathbf{x}_{t+1} - \mathbf{x}_t\|^2. \end{aligned} \quad (13)$$

最后,由式(7)得到:

$$\begin{aligned} D_\gamma(\mathbf{y}_{t+1}, \mathbf{z}_t, \mathbf{x}_t) - D_\gamma(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_t) &= f(\mathbf{y}_{t+1}) + \frac{1}{2\gamma} \|\mathbf{y}_{t+1} - \mathbf{x}_t\|^2 - \\ &- f(\mathbf{y}_t) - \frac{1}{2\gamma} \|\mathbf{y}_t - \mathbf{x}_t\|^2 \leq -\frac{1}{2} \left(\frac{1}{\gamma} - l \right) \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2. \end{aligned} \quad (14)$$

将式(10)、(13)和(14)相加再结合 $\sigma \geq 1$, 进一步可得 $\forall t \geq 1$, 有:

$$\begin{aligned} D_\gamma(\mathbf{y}_{t+1}, \mathbf{z}_{t+1}, \mathbf{x}_{t+1}) - D_\gamma(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_t) &\leq \frac{1}{\gamma} \|\mathbf{x}_{t+1} - \mathbf{x}_t\|^2 - \frac{1+\sigma_h}{2\gamma} \|\mathbf{x}_{t+1} - \mathbf{x}_t\|^2 + \\ &\frac{1+L_{\nabla h}}{2\gamma} [(-1+2\gamma l) \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2 + \|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2] - \frac{1}{2} \left(\frac{1}{\gamma} - l \right) \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2 \leq \\ &\frac{1}{2\gamma} [(1+L_{\nabla h})(-1+2\gamma l) - (1-\gamma l)] \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2 + \frac{(1+L_{\nabla h})}{2\gamma} \|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2, \end{aligned}$$

注意到 f 具有 Lipschitz 连续的梯度, 常数为 $L_{\nabla f}$, 且由式(5)中的第 1 个关系, 有:

$$\frac{1}{\gamma} \|\mathbf{x}_t - \mathbf{y}_{t+1} - (\mathbf{x}_{t-1} - \mathbf{y}_t)\| = \|\nabla f(\mathbf{y}_{t+1}) - \nabla f(\mathbf{y}_t)\| \leq L_{\nabla f} \|\mathbf{y}_{t+1} - \mathbf{y}_t\|,$$

因此可以得到:

$$\|\mathbf{x}_t - \mathbf{x}_{t-1}\| \leq (\gamma L_{\nabla f} + 1) \|\mathbf{y}_{t+1} - \mathbf{y}_t\|. \quad (15)$$

进一步得出 $\forall t \geq 1$, 有:

$$\begin{aligned} D_\gamma(\mathbf{y}_{t+1}, \mathbf{z}_{t+1}, \mathbf{x}_{t+1}) - D_\gamma(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_t) &\leq \\ \frac{1}{2\gamma} [(1+L_{\nabla h})(-1+2\gamma l) - 1 + \gamma l + (1+L_{\nabla h})(\gamma L_{\nabla f} + 1)^2] \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2 &= \frac{1}{\gamma} M_\gamma \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2. \end{aligned} \quad (16)$$

由假设 2 中的 2) 可知 $M_\gamma < 0$, 且可以看到 $\{D_\gamma(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_t)\}_{t \geq 1}$ 是非增的。将式(16)从 $t=1$ 到 $t=N-1 \geq 1$ 进行求和, 可以得到:

$$D_\gamma(\mathbf{y}_N, \mathbf{z}_N, \mathbf{x}_N) - D_\gamma(\mathbf{y}_1, \mathbf{z}_1, \mathbf{x}_1) \leq \frac{1}{\gamma} M_\gamma \sum_{t=1}^{N-1} \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2. \quad (17)$$

由假设 2 中的 3) 可以得到 $\{(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_t)\}$ 存在 1 个聚点 $(\bar{\mathbf{y}}, \bar{\mathbf{z}}, \bar{\mathbf{x}})$ 且具有收敛子列 $\lim_{j \rightarrow \infty} (\mathbf{y}_{t_j}, \mathbf{z}_{t_j}, \mathbf{x}_{t_j}) = (\bar{\mathbf{y}}, \bar{\mathbf{z}}, \bar{\mathbf{x}})$, 则利用 D_γ 的下半连续性并在式(17)中取 $N=t_j$ 且令 $j \rightarrow \infty$ 取极限, 有:

$$-\infty < D_\gamma(\bar{\mathbf{y}}, \bar{\mathbf{z}}, \bar{\mathbf{x}}) - D_\gamma(\mathbf{y}_1, \mathbf{z}_1, \mathbf{x}_1) \leq \frac{1}{\gamma} M_\gamma \sum_{t=1}^{\infty} \|\mathbf{y}_{t+1} - \mathbf{y}_t\|^2.$$

由此立即得出结论: $\lim_{t \rightarrow \infty} \|\mathbf{y}_{t+1} - \mathbf{y}_t\| = 0$ 。将该结论与式(15)结合, 有 $\lim_{t \rightarrow \infty} \|\mathbf{x}_t - \mathbf{x}_{t-1}\| = 0$, 于是式(9)成立。此外, 由算法 1 步骤 2 中的第 3 个关系, 有:

$$\|\mathbf{z}_{t+1} - \mathbf{z}_t\| = \|(\mathbf{x}_{t+1} - \mathbf{x}_t + \mathbf{y}_{t+1}) - (\mathbf{x}_t - \mathbf{x}_{t-1} + \mathbf{y}_t)\| \leq 2\|\mathbf{x}_{t+1} - \mathbf{x}_t\| + \|\mathbf{y}_{t+1} - \mathbf{y}_t\|,$$

取 $t \rightarrow \infty$ 时的极限, 可以得到 $\lim_{t \rightarrow \infty} \|\mathbf{z}_{t+1} - \mathbf{z}_t\| = 0$ 。因此有:

$$\lim_{j \rightarrow \infty} (\mathbf{y}_{t_j}, \mathbf{z}_{t_j}, \mathbf{x}_{t_j}) = \lim_{j \rightarrow \infty} (\mathbf{y}_{t_j-1}, \mathbf{z}_{t_j-1}, \mathbf{x}_{t_j-1}) = (\bar{\mathbf{y}}, \bar{\mathbf{z}}, \bar{\mathbf{x}}), \quad (18)$$

由 $\mathbf{z}_t$ 作为极小化点的定义, 有:

$$g(\mathbf{z}_t) + \frac{1}{\gamma} D_h(\mathbf{z}_t, \mathbf{y}_t) + \frac{1}{\gamma} \langle \mathbf{x}_{t-1} - \mathbf{y}_t, \mathbf{z}_t \rangle \leq g(\bar{\mathbf{z}}) + \frac{1}{\gamma} D_h(\bar{\mathbf{z}}, \mathbf{y}_t) + \frac{1}{\gamma} \langle \mathbf{x}_{t-1} - \mathbf{y}_t, \bar{\mathbf{z}} \rangle,$$

沿收敛子列取极限并利用式(18), 得到 $\limsup_{j \rightarrow \infty} g(\mathbf{z}_{t_j}) \leq g(\bar{\mathbf{z}})$ 。另一方面, 根据 g 的下半连续性, 可以得到 $\liminf_{j \rightarrow \infty} g(\mathbf{z}_{t_j}) \geq g(\bar{\mathbf{z}})$ 。因此有 $\lim_{j \rightarrow \infty} g(\mathbf{z}_{t_j}) = g(\bar{\mathbf{z}})$ 。现在沿收敛子列 $\{(\mathbf{y}_{t_j}, \mathbf{z}_{t_j}, \mathbf{x}_{t_j})\}$ 在式(6)中取极限, 并利用式(3)、(6)和(9), 可以看到定理 1 的结论成立。证毕

接下来将假设函数 f, g 和 h 均为半代数函数, 从而证明如果生成的整个序列有 1 个聚点, 那么实际上整个序列是收敛的。

定理 2 (整个序列的全局收敛性) 若假设 1、2 成立, 并假设 f, g 和 h 都是半代数函数, 那么整个序列 $\{(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_t)\}$ 是收敛的。

证明 考虑 D_γ 在 $(\mathbf{y}_{t+1}, \mathbf{z}_{t+1}, \mathbf{x}_{t+1})$ 处的次微分。首先, 注意到 $\forall t \geq 0$, 利用式(8)的第 1 个关系和 $\mathbf{x}_{t+1}$ 的定义, 有:

$$\nabla_{\mathbf{x}} D_\gamma(\mathbf{y}_{t+1}, \mathbf{z}_{t+1}, \mathbf{x}_{t+1}) = \frac{1}{\gamma} (\mathbf{z}_{t+1} - \mathbf{y}_{t+1}) = \frac{1}{\gamma} (\mathbf{x}_{t+1} - \mathbf{x}_t).$$

然后, 利用式(5)中的第 1 个关系和式(7), 有:

$$\nabla_y D_\gamma(\mathbf{y}_{t+1}, \mathbf{z}_{t+1}, \mathbf{x}_{t+1}) = \nabla f(\mathbf{y}_{t+1}) + \frac{1}{\gamma}(\mathbf{y}_{t+1} - \mathbf{x}_{t+1}) = \frac{1}{\gamma}(\mathbf{x}_t - \mathbf{x}_{t+1}).$$

此外,对于关于 $\mathbf{z}$ 的次微分,由式(7)有:

$$\begin{aligned} \partial_z D_\gamma(\mathbf{y}_{t+1}, \mathbf{z}_{t+1}, \mathbf{x}_{t+1}) &= \partial g(\mathbf{z}_{t+1}) - \frac{1}{\gamma}(\mathbf{z}_{t+1} - \mathbf{x}_{t+1}) = \partial g(\mathbf{z}_{t+1}) + \frac{1}{\gamma}(\nabla h(\mathbf{z}_{t+1}) - \\ &\nabla h(\mathbf{y}_{t+1})) + \frac{1}{\gamma}(\mathbf{x}_t - \mathbf{y}_{t+1}) - \frac{1}{\gamma}(\nabla h(\mathbf{z}_{t+1}) - \nabla h(\mathbf{y}_{t+1})) - \frac{1}{\gamma}(\mathbf{x}_t - \mathbf{y}_{t+1}) - \\ &\frac{1}{\gamma}(\mathbf{z}_{t+1} - \mathbf{x}_{t+1}) \ni -\frac{1}{\gamma}(\nabla h(\mathbf{z}_{t+1}) - \nabla h(\mathbf{y}_{t+1})) - \frac{1}{\gamma}(\mathbf{x}_t - \mathbf{y}_{t+1}) - \frac{1}{\gamma}(\mathbf{z}_{t+1} - \mathbf{x}_{t+1}), \end{aligned}$$

其中的包含关系可以由式(5)的第2个关系得到。设 $\boldsymbol{\zeta}_0 = (\nabla h(\mathbf{z}_{t+1}) - \nabla h(\mathbf{y}_{t+1})) + (\mathbf{x}_t - \mathbf{y}_{t+1}) + (\mathbf{z}_{t+1} - \mathbf{x}_{t+1})$, 那么有:

$$\begin{aligned} \|\boldsymbol{\zeta}_0\| &\leq \|\nabla h(\mathbf{z}_t) - \nabla h(\mathbf{y}_t)\| + \|\mathbf{z}_t - \mathbf{y}_t\| + \|\mathbf{x}_t - \mathbf{x}_{t-1}\| \leq \\ (L_{\nabla h} \|\mathbf{z}_t - \mathbf{y}_t\| + \|\mathbf{z}_t - \mathbf{y}_t\| + \|\mathbf{x}_t - \mathbf{x}_{t-1}\|) &= (L_{\nabla h} + 2) \|\mathbf{x}_t - \mathbf{x}_{t-1}\| \leq (L_{\nabla h} + 2)(\gamma L_{\nabla f} + 1) \|\mathbf{y}_{t+1} - \mathbf{y}_t\|, \end{aligned}$$

其中最后1个不等式由式(15)得到。上述关系意味着

$$\begin{aligned} \text{dist}(\mathbf{0}, \partial D_\gamma(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_t)) &= \\ \inf\{\|\boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\zeta}\| : (\boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\zeta}) \in (\nabla_x D_\gamma(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_t), \nabla_y D_\gamma(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_t), \partial_z D_\gamma(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_t))\} &\leq \\ \left\| \frac{1}{\gamma}(\mathbf{x}_t - \mathbf{x}_{t-1}, \mathbf{x}_t - \mathbf{x}_{t-1}, -\boldsymbol{\zeta}_0) \right\| &= \frac{1}{\gamma} \sqrt{\|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2 + \|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2 + \|\boldsymbol{\zeta}_0\|^2} \leq \\ \frac{(\gamma L_{\nabla f} + 1)}{\gamma} \sqrt{2 + (L_{\nabla h} + 2)^2} \|\mathbf{y}_{t+1} - \mathbf{y}_t\|, \end{aligned}$$

即存在 $\tau > 0$ 使得当 $t \geq 1$ 时,有:

$$\text{dist}(\mathbf{0}, \partial D_\gamma(\mathbf{y}_t, \mathbf{z}_t, \mathbf{x}_t)) \leq \tau \|\mathbf{y}_{t+1} - \mathbf{y}_t\|.$$

后续可采用文献[16]中定理2的证明方法,即可得出相应结果。

证毕

注4 本文的定理2与文献[16]中定理2的主要区别在于对价值函数 D_γ 次微分的处理,其他方面得出的结论一致。因此,仅需验证是否存在参数 τ (即文献[16]定理2中的式(28)是否成立),后续证明过程即可沿用。

2.3 PBDR 算法的步长规则

下面给出步长规则,以此来满足前面提到的假设2要求。在此之前为简化书写,令:

$$a = \frac{(1 + L_{\nabla h})L_{\nabla f}^2}{2}, b = (1 + L_{\nabla h})L_{\nabla f} + \left(L_{\nabla h} + \frac{3}{2}\right)L, c = -\frac{1}{2}.$$

命题1 (步长规则) 若假设1成立,取 $\sigma_h \geq 1$,令 $\gamma^* = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$; 则 $\gamma^* > 0$, 且 $\forall 0 < \gamma <$

$\min\left\{\frac{1}{L_{\nabla f}}, \gamma^*\right\}$, 都满足假设2。

证明 首先可以得到 $a > 0$, 因此 $-4ac > 0$, 也就是说 $\sqrt{b^2 - 4ac} > |b|$ 。因此可得 $\gamma^* > 0$, 此时有:

1) 由 σ 与 γ 选取方式,可以得到对假设2中1)的满足;

2) M_γ 可以重写为 $a\gamma^2 + b\gamma + c$, 容易得到这是一个关于 γ 开口向上的二次函数,且可以得到判别式 $\Delta = b^2 - 4ac > 0$, 以及当 $\gamma = 0$ 有 $M_\gamma = c < 0$ 。因此对于 $0 < \gamma < \gamma^*$, 有 $M_\gamma < 0$, 即满足假设2中的2);

3) 运用 $0 < \gamma < \frac{1}{L_{\nabla f}}$ 以及假设1中3),再结合文献[16]中定理4的证明方式,即可满足假设2中的3)。

综上,假设2满足。

证毕

注5 命题1中提到的参数设置是常见的,只需要选择具有梯度 Lipschitz 的强凸 h , 满足 $\sigma \geq 1$ 即可。比如在文献[15]中提到的令 $h(\mathbf{x}) = \alpha \sqrt{1 + \|\mathbf{x}\|^2} + \frac{\beta}{2} \|\mathbf{x}\|^2$, 此时只需要 $\beta \geq 1$ 即可。

3 子问题研究与数值实验

本文子问题设计与数值实验都是针对以下非凸可行性问题:

$$\min_{x \in D} \frac{1}{2} \text{dist}^2(x, C),$$

对应于式(1)就是令 $f = \frac{1}{2} \text{dist}^2(\cdot, C)$, $g = \delta_D$, 其中: $C = \{x \in \mathbf{R}^n : Ax = d\}$, $D = \{x \in \mathbf{R}^n : \|x\|_0 \leq r, \|x\| \leq R\}$,

$A \in \mathbf{R}^{m \times n}$, $d \in \mathbf{R}^m$, $r \in \mathbf{N} \setminus \{0\}$, $R > 0$, 对 $(x \in \mathbf{R}^n) \|x\|_0 = \sum_{i=1}^n |\text{sgn}(x^{(i)})|$, 其中 $|\text{sgn}(0)| = 0$, 否则 $|\text{sgn}(\cdot)| = 1$.

很显然可以得到 C 是半代数集合。集合 D 可以表示为:

$$D = \{x \in \mathbf{R}^n : \|x\|_0 \leq r\} \cap \{x \in \mathbf{R}^n : \|x\|^2 \leq R^2\},$$

这是半代数集合的交集;由文献[18]的公式[27(d)]可得 D 也是一个半代数集合。因此 f 与 g 都是半代数函数,也就是满足 KL 性质。因此在 PBDR 算法框架下收敛性可以得到保证。

实验涉及的代码在 MATLAB R2022a 中实现,数值实验运行环境为 64 位 Windows 系统,计算机硬件配置为 Intel(R) Core(TM) i5-8300H 2.30 GHz 处理器及 16 GB 内存。

3.1 子问题设计

对于 f 部分,由文献[16]可得 $y_{t+1} = \frac{1}{\gamma}(x_t + \gamma P_C(x_t))$, 其中 P_C 为对于集合 C 的投影算子。对于 g 部分,

取 $h(x) = \alpha \sqrt{1 + \|x\|^2} + \frac{\beta}{2} \|x\|^2$ 。并且对 $u, v \in \mathbf{R}^n$ 定义 $p(u, v) = v - u - \nabla h(u)$, 以及定义解算子为:

$$\begin{aligned} T_\gamma(u, v) &= \arg \min_{z \in \mathbf{R}^n} \left\{ g(z) + \frac{1}{\gamma} D_h(z, u) + \frac{1}{\gamma} \langle v - u, z \rangle \right\} = \arg \min_{z \in \mathbf{R}^n} \{ \gamma g(z) + \langle v - u, z \rangle + D_h(z, u) \} = \\ &= \arg \min_{z \in \mathbf{R}^n} \{ \gamma g(z) + \langle v - u, z \rangle + h(z) - \langle \nabla h(u), z \rangle \} = \arg \min_{z \in \mathbf{R}^n} \{ \gamma g(z) + \langle p(u, v), z \rangle + h(z) \}. \end{aligned} \quad (19)$$

显然, g 的 Bregman 近端子问题对应于式(19), 其中 $u = y_{t+1}$, $v = x_t$ 。根据文献[15]中的命题 6.2, g 的子问题解可直接由以下引理得到。

引理 2^[15] 设 $g = \delta_D$, 其中 $D = \{x \in \mathbf{R}^n : \|x\|_0 \leq r, \|x\| \leq R\}$, $r \leq n$ 为整数, $R > 0$ 为实数。定义

$$\bar{x} = \begin{cases} \mathbf{0}, & \text{若 } p(u, v) = \mathbf{0}; \\ -\bar{s} \|H_r(p(u, v))\|^{-1} H_r(p(u, v)), & \text{若 } p(u, v) \neq \mathbf{0}. \end{cases}$$

其中: $H_r(x)$ 为 $x \in \mathbf{R}^n$ 的硬阈(hard-threshold)算子(参数为 r), 定义如下:

$$(\forall i \in \{1, \dots, n\}) H_r^{(i)}(x) = \begin{cases} x^{(i)}, & \text{若 } |x^{(i)}| \text{ 在 } |x^{(1)}|, \dots, |x^{(n)}| \text{ 中排前 } r \text{ 位}; \\ 0, & \text{否则}. \end{cases}$$

$\bar{s}$ 为 $\min_{0 \leq s \leq R} \left\{ \varphi(s) = \alpha \sqrt{1 + s^2} + \frac{\beta}{2} s^2 - \|H_r(p(u, v))\| s \right\}$ 的唯一解, 且若 $\|H_r(p(u, v))\| \geq \alpha(1 + R^2)^{-\frac{1}{2}} R + \beta R$, 则

$\bar{s} = R$; 否则 $\bar{s} \in (0, R)$ 为 $\alpha(1 + s^2)^{-\frac{1}{2}} s + \beta s - \|H_r(p(u, v))\| = 0$ 的唯一解。那么可以得到 $\bar{x} \in T_\gamma(u, v)$ 。

3.2 数值实验

本文将对 PBDR 算法、BiFRB(Bregman inertial forward-reflected-backward)算法^[15]和 DR 算法^[16]进行基准测试, 其中 R 取值为 1 000 和 10^6 , 且 $r = \left\lceil \frac{m}{5} \right\rceil$ 。PBDR 算法和 DR 算法分别采用命题 1 和文献[16]第 5 节中

给出的步长规则实现, 并在满足 $\frac{\max\{\|x_t - x_{t-1}\|, \|y_t - y_{t-1}\|, \|z_t - z_{t-1}\|\}}{\max\{1, \|x_{t-1}\|, \|y_{t-1}\|, \|z_{t-1}\|\}} < 10^{-10}$ 这一条件时终止。

此外, PBDR 算法与 BiFRB 算法都使用核函数 $h(x) = 0.1 \cdot \sqrt{1 + \|x\|^2} + 2.51 \cdot \frac{\|x\|^2}{2}$ 。BiFRB 算法的步长和终止条件采用文献[15]中第 7 节命题 4.2 中的设置。关于惯性参数, BiFRB 算法使用 $\eta = 0.9$ 。最后, 每种算法均从原点初始化, 并运用文献[16]中注 4 描述的步长启发式方法。

图 1 描述的是函数值随迭代次数的平均变化情况, 表 1 表示不同参数下算法运行的仿真结果。实验的问题数据通过随机生成的矩阵 $A \in \mathbf{R}^{m \times n}$ 生成, 其中的元素服从标准 Gauss 分布。对于每个规模为 (m, n) 的问题, 均随机生成 50 个实例, 最后再报告平均迭代次数的最小上限和终止时的最小目标值。

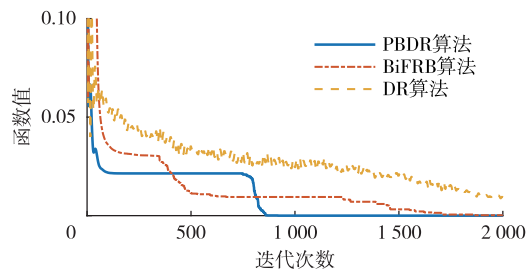
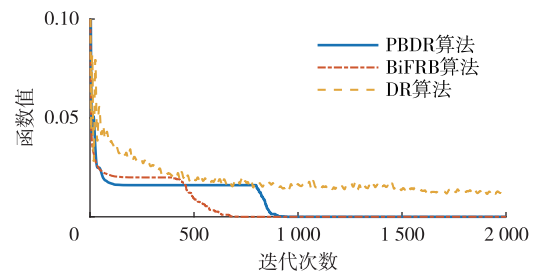
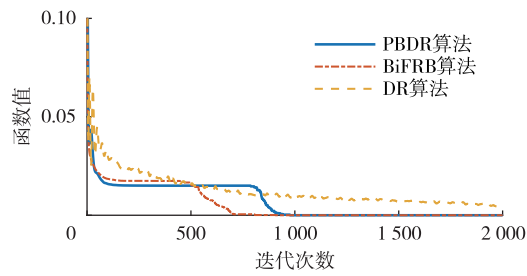
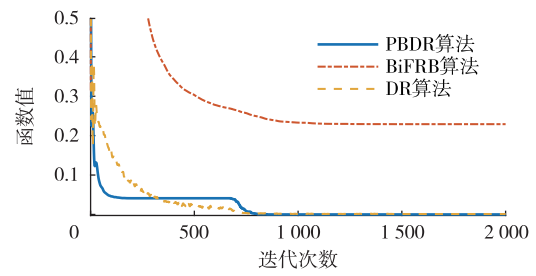
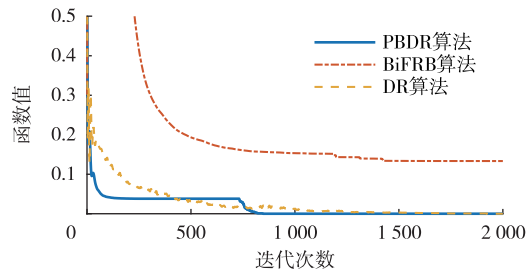
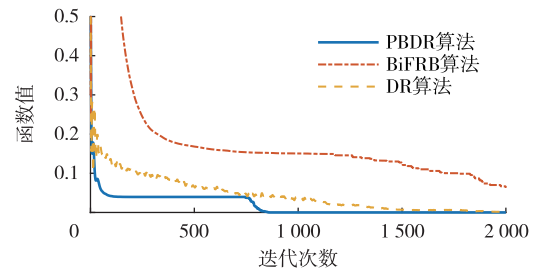
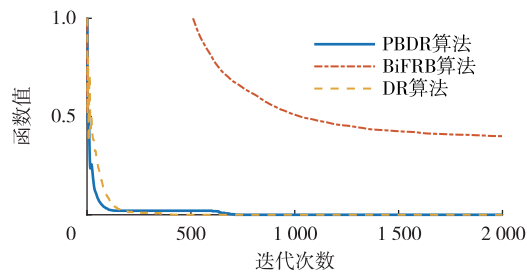
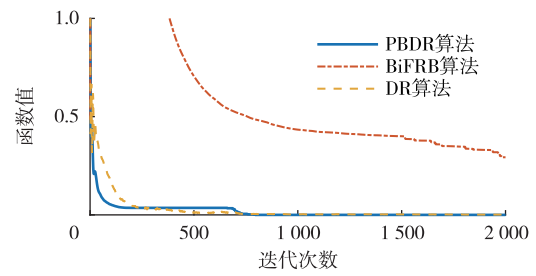
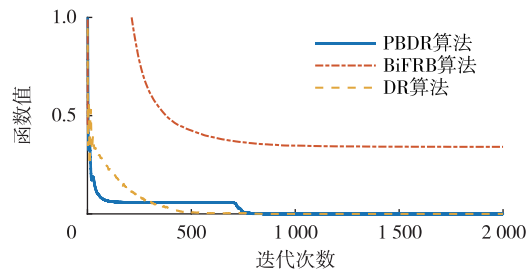
a 算法比较情况 ($m=100, n=4\ 000$)b 算法比较情况 ($m=100, n=5\ 000$)c 算法比较情况 ($m=100, n=6\ 000$)d 算法比较情况 ($m=200, n=4\ 000$)e 算法比较情况 ($m=200, n=5\ 000$)f 算法比较情况 ($m=200, n=6\ 000$)g 算法比较情况 ($m=300, n=4\ 000$)h 算法比较情况 ($m=300, n=5\ 000$)i 算法比较情况 ($m=300, n=6\ 000$)图1 函数值在经过50次重复后于迭代过程($R=10^6$)中的平均变化Fig. 1 Average change of function values over iteration ($R=10^6$) after 50 repetition

表 1 本文中 3 种算法的对比
Tab. 1 Comparison of 3 algorithms in the article

规模			PBDR 算法		BiFRB 算法		DR 算法	
R	m	n	iter	fval _{min}	iter	fval _{min}	iter	fval _{min}
1 000	100	4 000	824	6.46E-24	757	2.44E-23	2 415	3.21E-25
	100	5 000	855	1.11E-23	529	2.99E-04	2 802	2.56E-24
	100	6 000	872	1.68E-03	637	2.33E-03	2 436	2.19E-24
	200	4 000	750	3.15E-24	2 761	2.13E-03	1 211	4.47E-26
	200	5 000	773	3.02E-24	2 741	8.73E-03	1 470	2.11E-24
	200	6 000	810	3.52E-24	2 557	9.04E-06	1 760	7.16E-25
	300	4 000	679	1.13E-23	3 415	4.00E-21	927	5.75E-25
	300	5 000	727	9.12E-24	4 063	3.01E-03	1 052	4.56E-24
	300	6 000	760	5.58E-24	2 884	5.63E-02	1 200	9.16E-25
10 ⁶	100	4 000	830	8.13E-25	854	1.02E-23	2 522	2.57E-26
	100	5 000	847	2.72E-03	550	2.21E-23	2 986	7.86E-25
	100	6 000	857	5.19E-03	619	1.37E-05	2 318	1.65E-24
	200	4 000	744	1.23E-23	2 699	2.66E-02	1 237	2.18E-25
	200	5 000	786	4.94E-24	2 663	1.70E-07	1 514	1.11E-24
	200	6 000	810	8.62E-24	2 482	1.85E-02	1 610	1.09E-25
	300	4 000	685	9.52E-24	3 325	6.09E-02	930	4.68E-25
	300	5 000	731	1.21E-23	3 897	4.22E-05	1 056	1.67E-24
	300	6 000	771	1.25E-23	2 944	5.06E-02	1 203	1.10E-24

注:iter 和 fval_{min} 分别表示平均迭代次数的最小上限和迭代终止时的最小目标值。

通过图 1 和表 1 可以看出:

- 1) 与 BiFRB 算法相比,PBDR 算法在 $m \geq 200$ 时更具有优势,它的迭代次数仅有 BiFRB 算法迭代次数的 30%左右;
- 2) DR 算法在终止时的函数值较 PBDR 算法更优,但在收敛速度上的表现不如 PBDR 算法,而且 PBDR 算法的迭代次数仅有 DR 算法迭代次数的 50%左右;
- 3) PBDR 算法的收敛迭代次数不会因 m 、 n 和 R 的变动而出现很大波动,相对于另外 2 个算法而言更加稳定;
- 4) PBDR 算法与 DR 算法皆可有效收敛。

4 结束语

本文基于非凸可行性问题,运用 Bregman 距离对 DR 算法的其中一个子问题进行了非欧几里得拓展,提出了 PBDR 算法。在理论方面,本文建立了 PBDR 值函数,证明了价值函数的下降性,最终以此建立了 PBDR 算法的收敛性,并且最后给出了收敛性假设需要满足的步长准则。针对数值运用方面,本文运用引理 2 给出子问题求解方式,并且与现有算法进行了横向比较。最终的数据显示,在参数设置较大(如 $m \geq 200$)情况下,PBDR 算法相对于现有算法有着明显优势,迭代速度提升了 40%左右。

参考文献:

[1] Douglas J, Rachford H H. On the numerical solution of heat conduction problems in two and three space variables[J]. Transactions of the American Mathematical Society, 1956, 82 (2): 421-439.

[2] Lions P L, Mercier B. Splitting algorithms for the sum of two nonlinear operators[J]. SIAM Journal on Numerical Analysis, 1979, 16 (6): 964-979.

[3] Eckstein J, Bertsekas D P. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal

- monotone operators[J]. *Mathematical Programming*, 1992, 55 (1/2/3): 293-318.
- [4] He B S, Yuan X M. On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method[J]. *SIAM Journal on Numerical Analysis*, 2012, 50 (2): 700-709.
- [5] Combettes P L, Pesquet J C. A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2007, 1 (4): 564-574.
- [6] Patrinos P, Stella L, Bemporad A. Douglas-Rachford splitting: complexity estimates and accelerated variants[C/OL]// 53rd IEEE Conference on Decision and Control. Los Angeles: IEEE, 2014: 4234-4239. <https://ieeexplore.ieee.org/document/7040049>.
- [7] Gandy S, Recht B, Yamada I. Tensor completion and low- n -rank tensor recovery via convex optimization[J]. *Inverse Problems*, 2011, 27 (2): 025010.
- [8] Feng J K, Zhang H B, Zhang K L, et al. An inertial Douglas-Rachford splitting algorithm for nonconvex and nonsmooth problems[J]. *Concurrency and Computation: Practice and Experience*, 2023, 35 (17): e6343.
- [9] Themelis A, Patrinos P. Douglas-Rachford splitting and ADMM for nonconvex optimization: tight convergence results[J]. *SIAM Journal on Optimization*, 2020, 30 (1): 149-181.
- [10] Bian F M, Zhang X Q. A parameterized Douglas-Rachford splitting algorithm for nonconvex optimization[J]. *Applied Mathematics and Computation*, 2021, 410: 126425.
- [11] Lu T L, Zhang X. An inertial parametric Douglas-Rachford splitting method for nonconvex problems[J]. *Mathematics*, 2024, 12 (5): 675.
- [12] Lindstrom S B, Sims B. Survey: sixty years of Douglas-Rachford[J]. *Journal of the Australian Mathematical Society*, 2021, 110 (3): 333-370.
- [13] Han D R. A survey on some recent developments of alternating direction method of multipliers[J]. *Journal of the Operations Research Society of China*, 2022, 10 (1): 368.
- [14] Rockafellar R T, Wets R J B. *Variational analysis*[M]. Heidelberg: Springer, 1998.
- [15] Wang X F, Wang Z Y. A Bregman inertial forward-reflected-backward method for nonconvex minimization[J]. *Journal of Global Optimization*, 2024, 89 (2): 327-354.
- [16] Li G Y, Pong T K. Douglas-Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems[J]. *Mathematical Programming*, 2016, 159 (1/2): 371-401.
- [17] Bauschke H H, Combettes P L. *Convex analysis and monotone operator theory in Hilbert spaces*[M]. Cham: Springer International Publishing, 2017.
- [18] Bauschke H H, Luke D R, Phan H M, et al. Restricted normal cones and sparsity optimization with affine constraints[J]. *Foundations of Computational Mathematics*, 2014, 14 (1): 63-83.

Operations Research and Cybernetics

A Douglas-Rachford Splitting Algorithm Partially Based on Bregman Distance

GAO Si, LUO Honglin

(School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China)

Abstract: This paper proposes a Douglas-Rachford splitting algorithm partially based on Bregman distance (referred to as PBDR algorithm hereinafter) for the minimization problem of the sum of non-convex functions. By introducing the Bregman distance function, this algorithm non-Euclidean generalizes some subproblems in the classical Douglas-Rachford splitting framework, thereby extending the application of this type of algorithm in non-Euclidean geometric structures. In terms of theoretical analysis, this paper constructs a value function and combines the Kurdyka-Łojasiewicz property to prove the global convergence of the algorithm's iterative sequence. In numerical experiments, this paper takes non-convex feasibility problems as the test benchmark and designs special numerical calculations for subproblems involving Bregman distance. The research results show that the PBDR algorithm is significantly superior to traditional algorithms in terms of convergence speed, computational accuracy, and numerical stability.

Keywords: non-convex optimization; Douglas-Rachford splitting algorithm; Bregman proximity mapping; Kurdyka-Łojasiewicz property

(责任编辑 方 兴)