

# 求解一类复合优化的惯性近似Bregman 近端梯度算法

房宛玉

青岛大学数学与统计学院, 山东 青岛

收稿日期: 2026年4月19日; 录用日期: 2026年5月13日; 发布日期: 2026年5月22日

## 摘要

本文提出了一种惯性近似Bregman近端梯度算法针对目标函数梯度不满足全局Lipschitz连续的复合优化问题。通过利用相对光滑来代替全局Lipschitz条件, 引入惯性外推机制以提升收敛速度, 并结合一种回溯线搜索, 确保非凸环境下的数值稳定性。此外, 我们采用了一种近似Bregman距离的新度量, 使得子问题的求解比现有Bregman类算法更为简单。在收敛性证明中, 在目标函数满足Kurdyka-Łojasiewicz (KL)性质的假设下, 证明了算法生成的序列具有有限长度, 并全局收敛至广义平稳点。

## 关键词

复合优化, 惯性外推, Bregman距离, Kurdyka-Łojasiewicz性质

# Inertial Approximate Bregman Proximal Gradient Algorithms for a Class of Composite Optimization Problems

Wanyu Fang

School of Mathematics and Statistics, Qingdao University, Qingdao Shandong

Received: April 19, 2026; accepted: May 13, 2026; published: May 22, 2026

## Abstract

This paper proposes an inertial approximate Bregman proximal gradient algorithm for composite optimization problems where the gradient of the objective function does not satisfy global Lipschitz continuity. By utilizing relative smoothness to replace the global Lipschitz condition, an inertial extrapolation mechanism is introduced to improve the convergence rate, and combined with a backline

search, numerical stability in non-convex settings is ensured. Furthermore, we adopt a new metric for the approximate Bregman distance, which simplifies the solution of the subproblems compared to existing Bregman-type algorithms. In the convergence proof, under the assumption that the objective function satisfies the Kurdyka-Łojasiewicz (KL) property, we prove that the sequence generated by the algorithm is of finite length and converges globally to a generalized steady point.

## Keywords

Composite Optimization, Inertial Extrapolation, Bregman Distance, Kurdyka-Łojasiewicz Property

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

本文主要研究以下形式的复合优化问题

$$\min_{x \in \text{cl}C} \varphi(x) := f(x) + g(x), \quad (1.1)$$

其中  $f(x): \mathbb{R}^n \rightarrow (-\infty, +\infty]$  是一个连续可微函数(可能非凸),  $g(x): \mathbb{R}^n \rightarrow (-\infty, +\infty]$  是一个凸函数, 其中  $\text{cl}C$  表示  $C \subset \mathbb{R}^n$  的闭包, 即一个非空的开凸集。这类问题具有广泛的应用背景, 例如, 图像与信号处理、矩阵分解、稀疏信号重建、图像去噪, 机器学习等[1]-[11]。

求解此类复合优化问题的最经典算法之一是近端梯度法(Proximal Gradient, PG)及其变体[12]-[16]。其理论基础早在 1970 年代由 Rockafellar 等确立[17], 为了加速 PG 的收敛, Polyak (1964)引入重球方法(Heavy-ball method)作为惯性技术的早期形式[18], Nesterov 进一步提出突破性惯性加速技术[19], 并由 Beck 和 Teboulle 扩展为 FISTA 算法[20], 在凸情况下实现  $O(1/k^2)$  的收敛率。然而, 无论是标准 PG 算法还是 FISTA 算法, 其核心收敛性保证都严重依赖于一个基本假设, 光滑项  $\nabla f$  必须是全局 Lipschitz 连续的。在许多实际的现代应用中, 例如相位恢复、非负矩阵分解以及  $l_p$  范数正则化问题, 损失函数的梯度通常会随着变量的增大或靠近边界而趋于无穷, 使得这一全局 Lipschitz 条件并不成立。

为克服全局 Lipschitz 的理论瓶颈, Bregman 提出[21] Bregman 距离, 是 2 范数平方距离的推广, 在使用时具有更大的灵活性。Nemirovski 和 Yudin 在镜像下降算法中推广 Bregman 距离[22], Auslender 和 Teboulle 进一步将其应用于近端方法。近年来, Lu 等[23]和 Bolte 等[24]引入相对光滑概念。基于此的 Bregman 近端梯度法(BPG)摒弃全局 Lipschitz 假设, 用一般性的核函数  $\phi$  所诱导的精确 Bregman 距离替代传统的欧氏距离。随后, Zhou 等人在没有假设全局 Lipschitz 的条件下, 将 BPG 的收敛性分析进行了简化[25]。在此基础上, 基于 Bregman 距离的各类优化算法不断涌现, 并被广泛应用于多种复杂实际场景, 例如随机次梯度训练[26]、非负矩阵分解[27]、块坐标平滑优化[28]、DC 规划[29]以及去中心化优化[30]等。同时, 为了进一步提升收敛速度, 带有外推和惯性机制的 Bregman 算法也被提出[30]-[34]。

然而, 尽管上述基于 Bregman 距离的算法在理论和应用方面取得了显著成就, 但在实际计算过程中, 单步迭代需求解包含精确 Bregman 距离的子问题, 对于非二次的复杂核函数, 该子问题呈现非线性特征, 并且缺乏闭式解, 导致内层计算求解复杂。另一方面, 带有外推机制与非单调线搜索的加速技术在常规非凸优化领域备受关注。但是, 将惯性技术引入非凸优化时会破坏目标函数的单调下降性质, 导致理论

分析极其困难。综上所述,如何在缺乏全局 Lipschitz 条件下,将惯性外推机制与 Bregman 距离结合,并保证算法在非凸环境下的全局收敛性仍是一个难题。

因此,针对上述理论与计算的双重挑战,本文提出了一种惯性近似 Bregman 近端梯度算法。首先,我们在子问题中引入了近似 Bregman 距离,将复杂的非线性子问题转化为计算高效的标准二次规划(QP)或邻近算子求解。其次,为提高算法的计算效率,我们引入一种惯性机制和回溯线搜索策略,确保算法在非凸优化问题中的应用,并在满足 Kurdyka-Łojasiewicz 条件下给出算法的全局收敛性分析。

本文的组织结构如下,第 2 章给出相关预备知识;第 3 章给出相对光滑下的算法框架和参数选择;第 4 章给出算法的全局收敛性分析;第 5 章给出相关的数值实验;最后在第 6 章总结全文。

## 2 预备知识

首先,我们给出一些将在论文中使用的符号和定义。

本文在  $n$  维实欧氏空间  $\mathbb{R}^n$  中进行讨论。 $\mathbb{R}_+$  和  $\mathbb{R}_{++}$  分别表示非负实数集和正实数集。我们使用  $\langle \cdot, \cdot \rangle$  表示  $\mathbb{R}^n$  中的标准内积,  $\|\cdot\|$  表示诱导的  $l_2$  范数。对于任意点  $x \in \mathbb{R}^n$  和半径  $\delta > 0$ , 以  $x$  为中心的开球记为  $B(x, \delta)$ 。集合  $S \subset \mathbb{R}^n$  的内部和闭包分别记为  $\text{int}(S)$  和  $\text{cl}(S)$ 。点  $x$  到集合  $S$  的距离定义为  $\text{dist}(x, S) := \inf_{y \in S} \|x - y\|$ 。对于扩展实值函数  $f(x): \mathbb{R}^n \rightarrow (-\infty, +\infty]$ , 其有效域定义为  $\text{dom } f := \{x \in \mathbb{R}^n \mid f(x) < +\infty\}$ 。若  $\text{dom } f \neq \emptyset$  且对所有  $x \in \mathbb{R}^n$  均有  $f(x) > -\infty$ , 则称  $f$  为真函数。如果对于任意收敛到  $x$  的序列  $\{x^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n$  (即  $x^k \rightarrow x$ ), 均满足  $f(x) \leq \liminf_{k \rightarrow \infty} f(x^k)$ , 则称  $f$  在  $x$  处是下半连续的(Lower semicontinuous, l.s.c.)。对于给定的实数  $\alpha \in \mathbb{R}$ , 函数  $f$  的  $\alpha$ -下水平集记为  $\text{lev}_\alpha f := \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$ 。如果对于任意  $\alpha \in \mathbb{R}$ , 集合  $\text{lev}_\alpha f$  都是有界的, 则称函数  $f$  是水平有界的。

**定义 2.1** [35] 设  $f$  为一个适当且下半连续的函数。

(i)  $f$  在  $x \in \text{dom } f$  处的 Fréchet 次微分记作  $\hat{\partial}f(x)$

$$\hat{\partial}f(x) = \left\{ u \in \mathbb{R}^n \mid \liminf_{y \rightarrow x, y \neq x} \frac{f(y) - f(x) - \langle u, y - x \rangle}{\|y - x\|} \geq 0 \right\}.$$

当  $x \notin \text{dom } f$  时,  $\hat{\partial}f(x) = \emptyset$ 。

(ii)  $f$  在  $x \in \mathbb{R}^n$  处的极限次微分定义为

$$\partial f(x) := \left\{ u \in \mathbb{R}^n : \exists x^k \rightarrow x, f(x^k) \rightarrow f(x), \widehat{u^k} \partial \in f(x^k) \rightarrow u, k \rightarrow \infty \right\}.$$

其中符号  $x^k \xrightarrow{f} x$  表示  $x^k \rightarrow x$  且  $f(x^k) \rightarrow f(x)$ , 定义极限次微分的有效域为  $\text{dom } \partial f := \{x \in \mathbb{R}^n \mid \partial f(x) \neq \emptyset\}$ 。

(iii) 若  $0 \in \partial f(x)$ , 则称  $x$  是  $f$  的一个稳定点,  $f$  所有稳定点的集合记作  $\text{crit}(f)$ 。

一般情况下,  $\hat{\partial}f(x) \subset \partial f(x)$  恒成立。当  $f(x)$  为凸函数时, 极限次微分  $\partial f(x)$  退化为凸分析中的经典次微分。对于本文考虑的复合优化问题(1.1), 根据次微分积分法则, 目标函数  $\varphi(x)$  的次微分严格满足可加性, 即  $\partial \varphi(x) \equiv \nabla f(x) + \partial g(x)$ 。

为了处理目标函数中非全局 Lipschitz 连续的梯度, 我们引入基于核函数的非欧氏度量框架。设  $C$  是  $\mathbb{R}^n$  中的非空开凸子集, 引入核生成距离和 Bregman 距离。

**定义 2.2** [21] 设  $C$  是非空且凸的开集, 函数  $\phi: \mathbb{R}^n \rightarrow (-\infty, +\infty]$  被称为关于  $C$  的核生成距离, 记作  $\phi \in \mathcal{G}(C)$ , 若其满足以下条件

(i)  $\phi$  是适当的、下半连续且凸的函数, 满足  $\text{dom } \phi \subset \text{cl } C$  且  $\text{dom } \partial \phi = C$ 。

(ii)  $\phi$  在  $\text{int } \text{dom } \phi \equiv C$  上是  $C^1$  的。

**定义 2.3** [21] 给定  $\phi \in \mathcal{G}(C)$ , Bregman 距离  $D_\phi: \text{dom}\phi \times C \rightarrow \mathbb{R}_+$  定义为

$$D_\phi(x, y) = \phi(x) - \phi(y) - \nabla\phi(y), x - y.$$

若  $\phi(x) = \frac{1}{2}\|x\|^2$ , 则  $D_\phi(x, y) = \frac{1}{2}\|x - y\|^2$ 。此外, 对于  $d$  阶正定矩阵  $M$ , 令  $\|x\|_M := \sqrt{\langle x, Mx \rangle}$ , 并取  $\bar{\phi}(x) = \|x\|_M^2$ , 则  $D_{\bar{\phi}}(x, y) = \|x - y\|_M^2$ 。特别地, 如果  $M$  是单位矩阵, 那么  $D_\phi(x, y)$  为欧几里得平方距离。另外, Bregman 距离  $D_\phi$  并非严格意义上的距离, 因为它不满足对称性和三角不等式。由于  $\phi$  的凸性, 对于任意  $x \in \text{dom}\phi$  和  $y \in \text{int dom}\phi$ , 有  $D_\phi(x, y) \geq 0$ 。此外, 当  $\phi$  严格凸时,  $D_\phi(x, y) = 0$  当且仅当  $x = y$ 。

现在我们来定义光滑适应性, 也称为相对光滑性[21]。

**定义 2.4** [24] 考虑函数对  $(f, \phi)$ , 满足  $\phi \in \mathcal{G}(\mathbb{R}^n)$ , 且  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  是适当的下半连续函数, 并在  $C = \text{int dom}\phi$  上连续可微, 若存在常数  $L > 0$  使得函数  $L_\phi - f$  和  $L_\phi + f$  在  $\mathbb{R}^n$  上均是凸函数, 则称函数对  $(f, \phi)$  在  $\mathbb{R}^n$  上满足  $L$ -smad。

**引理 2.1** 函数对  $(f, \phi)$  是  $L$ -smad 的, 当且仅当对于任意  $x, y \in C$ , 成立如下不等式

$$f(x) \leq f(y) + \nabla f(y), x - y + LD_\phi(x, y).$$

利用二阶泰勒展开, 若  $\phi$  为  $C^2$  且强凸, 对于任意  $y \in C$ , 使得  $y + d \in C$  的方向  $d$ , 有

$$f(y + d) \leq f(y) + \nabla f(y), d + L\tilde{D}_\phi(y + d, y).$$

**定义 2.5** (去奇异函数) 满足如下条件的函数  $\psi: [0, \eta] \rightarrow \mathbb{R}^+$  称为去奇异函数

1.  $\psi$  在  $(0, \eta)$  上是凹函数且连续可微; 2.  $\psi$  在 0 点连续且  $\psi(0) = 0$ ; 3.  $\Psi'(x) > 0, \forall x \in (0, \eta)$ , 记  $\Xi_\eta$  为定义在  $[0, \eta]$  上的所有去奇异函数所构成的集合。

**定义 2.6** [36] 设  $\Psi: \mathbb{R}^n \rightarrow (-\infty, +\infty)$  是一个适当的下半连续函数。若对于点  $\hat{x} \in \text{dom}\partial\Psi$ , 存在常数  $\eta \in (0, +\infty)$ 、 $\hat{x}$  的一个邻域  $U$ , 以及去奇异函数  $\psi \in \Xi_\eta$ , 使得对于所有满足  $\Psi(\hat{x}) < \Psi(x) < \Psi(\hat{x}) + \eta$  的  $x \in U$ , 均严格满足如下 KL 不等式, 则称  $\Psi$  在  $\hat{x}$  处满足 KL 性质。

$$\psi'(\Psi(x) - \Psi(\hat{x})) \cdot \text{dist}(0, \partial\Psi(x)) \geq 1,$$

如果  $\Psi$  在其极限次微分的有效域  $\text{dom}\partial\Psi$  内的每一点处都满足 KL 性质, 则称  $\Psi$  为一个 KL 函数。

特别地, 当去奇异函数取形式  $\psi(s) = cs^{1-\theta}$  (其中  $c > 0$  且  $\theta \in [0, 1]$ ) 时, 参数  $\theta$  被称为 KL 指数, 它决定了算法在稳态点附近的渐近收敛速率。

**引理 2.2** [3] 假设  $f: \mathbb{R}^n \rightarrow (-\infty, +\infty)$  是一个真且下半连续的函数, 设  $\Omega \subset \mathbb{R}^n$  为一个紧集。如果  $f$  在  $\Omega$  上取值恒定, 且  $f$  在  $\Omega$  中的每一点都满足 KL 性质, 则存在常数  $\varepsilon > 0$ 、 $\eta > 0$  以及函数  $\psi \in \Xi_\eta$ , 使得对于任意  $\hat{x} \in \Omega$  以及任意满足  $\text{dist}(x, \Omega) < \varepsilon$  且  $f(\hat{x}) < f(x) < f(\hat{x}) + \eta$  的  $x \in \mathbb{R}^n$ , 如下一致化不等式均成立

$$\psi'(f(x) - f(\hat{x})) \text{dist}(0, \partial f(x)) \geq 1.$$

### 3. 惯性近似 Bregman 近端梯度算法

为求解问题(1.1)我们根据文献[37]提出一种惯性近似 Bregman 近端梯度算法, 在求解复合优化问题时, 经典的 Bregman 邻近梯度映射在点  $x \in C$  处的定义为

$$T_\lambda(x) := \arg \min_{u \in C} \left\{ \nabla f(x), u - x + g(u) + \frac{1}{\lambda} D_\phi(u, x) \right\}$$

其中  $\lambda > 0$  称为步长。虽然该映射通过相对平滑框架克服了全局 Lipschitz 梯度的限制, 但对于一般的核函数  $\phi$  和正则项  $g$ , 精确的 Bregman 距离通常会会导致一个非线性的复杂优化模型, 使得  $T_\lambda(x)$  极难获得闭

式解。为了解决这一问题，我们利用核函数的光滑性，在当前点  $x$  附近对  $\phi(u)$  进行二阶泰勒展开，得到精确距离的二次逼近

$$\begin{aligned} D_\phi(u, x) &= \phi(u) - \phi(x) - \langle \nabla \phi(x), u - x \rangle \\ &= \phi(x) + \langle \nabla \phi(x), u - x \rangle + \frac{1}{2} \langle \nabla^2 \phi(x)(u - x), u - x \rangle - \phi(x) - \langle \nabla \phi(x), u - x \rangle \\ &= \frac{1}{2} \langle \nabla^2 \phi(x)(u - x), u - x \rangle =: \tilde{D}_\phi(u, x), \end{aligned}$$

我们称这个非负的二次型项  $\tilde{D}_\phi(u, x)$  为(二阶)近似 Bregman 距离。基于上述近似，我们不再求解原先复杂的非线性映射，而是考虑在  $x \in C$  处计算近似 Bregman 邻近梯度映射，其定义转化为

$$\tilde{T}_\lambda(x) := \arg \min_{u \in C} \left\{ \nabla f(x), u - x + g(u) + \frac{1}{\lambda} \tilde{D}_\phi(u, x) \right\}.$$

借助函数  $g$  的凸性以及海森矩阵  $\nabla^2 \phi(x)$  的半正定性，该映射成功将子问题转化为一个标准的凸二次规划问题。进一步地，当核函数  $\phi$  是强凸函数且不存在正则项(即  $g \equiv 0$ )时，目标函数严格凸，可以通过令其对  $u$  的导数为零，直接获得极为简洁的闭式解

$$\tilde{T}_\lambda(x) = x - \lambda \left[ \nabla^2 \phi(x) \right]^{-1} \nabla f(x).$$

倘若选取的  $\phi$  不具备强凸性，我们只需引入一个微小的二次扰动，取  $\phi_0 := \phi + \frac{\sigma}{2} \|\cdot\|^2$  (其中  $\sigma > 0$ ) 作为新的距离生成核即可。特别地，当选取  $\phi = \frac{1}{2} \|\cdot\|^2$  (即  $\nabla^2 \phi(x) = I$ ) 时，近似映射  $\tilde{T}_\lambda(x)$  便等价于经典的欧几里德标准邻近映射，这种二阶近似策略将计算复杂度大幅降低。然而，仅有高效的子问题求解机制并不足以应对病态的非凸优化问题。在高维数据中目标函数的特殊结构易使一阶无记忆下降算法出现震荡或停滞，因此我们在上述近似框架中引入了 Nesterov 的惯性外推机制。计算一个外推点  $y^k$ ,

$$y^k = x^k + \beta_k (x^k - x^{k-1}),$$

参数  $\beta_k = \frac{k-1}{k+2}$ 。通过将当前迭代点与前一次迭代点进行仿射组合，惯性项能够有效累积历史梯度的动量信息，抑制震荡并帮助算法快速冲出平坦区域。

在融合了二阶近似与惯性加速之后，算法的单调性与稳定性面临着新的挑战。由于  $\tilde{D}_\phi(u, x)$  仅仅是精确 Bregman 距离的局部二阶逼近，对于任意给定的  $u, x \in \mathbb{R}^n$ ，全局不等式  $D_\phi(u, x) \leq \tilde{D}_\phi(u, x)$  或  $D_\phi(u, x) \geq \tilde{D}_\phi(u, x)$  并不总是成立。这意味着如果我们盲目地接受近似子问题给出的全步长方向  $d^k$ ，目标函数值可能会发生不可控的上升。正是基于这一深刻的理论动机，我们在算法中嵌入了回溯线搜索过程。通过寻找一个合适的步长  $t_k \in (0, 1)$ ，使得如下充分下降条件得以满足

$$\varphi(y^k + t_k d^k) \leq \varphi(y^k) + \alpha t_k \left( \langle \nabla f(y^k), d^k \rangle + g(y^k + d^k) - g(y^k) \right). \quad (3.1)$$

结合子问题的最优性条件，该线搜索本质上保证了目标函数沿着方向  $d^k$  能够获得与近似 Bregman 距离成正比的严格下降量

$$\varphi(x^{k+1}) \leq \varphi(y^k) - \frac{\alpha t_k}{2\lambda} \langle \nabla^2 \phi(y^k) d^k, d^k \rangle.$$

这不仅自适应地弥补了近似度量带来的截断误差，更重要的是，依据凸集的线段拓扑性质，外推点  $y^k$  与边界候选点  $y^k + d^k$  之间的严格凸组合地将下一步迭代点  $x^{k+1}$  “拉回”到开集  $C$  内部。基于上述分析，

我们提出惯性近似 Bregman 近端梯度算法, 详见算法 1。

---

算法 1: 惯性近似 Bregman 近端梯度算法

---

1. 输入: 存在  $C^2$  函数  $\phi \in G(C)$ , 使得对  $(f, \phi)$  在  $C$  上  $L$ -smad 的。

选取初始点  $x^0 \in \text{int dom } \phi$ ,  $0 < \lambda < \frac{2\sigma}{L_f}$ ,  $\alpha \in (0, 1)$ ,  $\eta \in (0, 1)$ ,  $\beta_k \in [0, \bar{\beta}]$ ,  $t_k = 1$ ,

$$\bar{\beta} < \frac{-\frac{M_\phi}{\lambda} + \sqrt{\left(\frac{M_\phi}{\lambda}\right)^2 + 2L_f\left(\frac{\sigma}{\lambda} - \frac{L_f}{2}\right)}}{L_f}.$$

2. 对于  $k = 0, 1, 2, \dots$ , 计算

$$y^k = x^k + \beta_k(x^k - x^{k-1})$$

在  $y^k$  处, 利用二阶近似 Bregman 距离构造局部二次规划子问题, 求解最优更新方向  $d^k$

$$d^k := \arg \min_{d \in \mathbb{R}^n, y^k + d \in \text{cl } C} \left\{ \langle \nabla f(y^k), d \rangle + g(y^k + d) + \frac{1}{2\lambda} \langle \nabla^2 \phi(y^k) d, d \rangle \right\}$$

3. 当  $\Phi(y^k + t_k d^k) > \Phi(y^k) + \alpha t_k (\langle \nabla f(y^k), d^k \rangle + g(y^k + d^k) - g(y^k))$  时,

执行  $t_k \leftarrow \eta t_k$

4. 更新  $x^{k+1} = y^k + t_k d^k$ 。

---

#### 4. 算法的收敛性分析

这一章节我们将给出算法的收敛性证明, 首先给出相关的假设条件。

**假设 4.1** 设  $C \subseteq \mathbb{R}^n$  为非空开凸集, 函数  $f: \mathbb{R}^n \rightarrow (-\infty, +\infty)$  在  $C$  上连续可微, 即  $f \in C^1(C)$ ; 函数  $g: \mathbb{R}^n \rightarrow (-\infty, +\infty)$  为适当的、下半连续且凸函数, 并满足  $\text{dom } g \cap C \neq \emptyset$ 。

**假设 4.2** 核函数  $\phi: \mathbb{R}^n \rightarrow (-\infty, +\infty)$  为适当的下半连续函数, 并满足  $\phi \in C^2(C)$ , 其有效定义域满足

$$\text{dom } \phi \subseteq \text{cl } C, \text{ dom } \partial \phi = C.$$

此外, 核函数  $\phi$  在  $C$  上是  $\sigma$ -强凸的, 其中  $\sigma > 0$ , 即对任意  $x \in C$  有  $\nabla^2 \phi(x) \succcurlyeq \sigma I$ 。

**假设 4.3** 目标函数  $\varphi$  在  $\text{cl } C$  上有下界, 即  $\varphi^* := \inf_{x \in \text{cl } C} \varphi(x) > -\infty$ , 给定初始点  $x^0 \in C$ , 其初始下水平集

$\Omega_0 := \{x \in \text{cl } C \mid \varphi(x) \leq \varphi(x^0)\}$  为有界紧集。

**假设 4.4** 存在一个紧致凸集  $\Omega \subset C$  满足  $\Omega_0 \subset \text{int } \Omega$ , 使得对所有  $k \geq 1$ , 均有外推点  $y^k \in \Omega$ ;

由于  $\phi \in C^2(C)$ , 故函数  $\nabla^2 \phi$  在紧致集  $\Omega$  上连续, 存在常数  $M_\phi > 0$  使得  $\|\nabla^2 \phi(x)\| \leq M_\phi, \forall x \in \Omega$ ;

在紧致集  $\Omega$  上  $\nabla f$  局部 Lipschitz 连续, 存在常数  $L_f > 0$ , 使得  $\|\nabla f(x) - \nabla f(y)\| \leq L_f \|x - y\|, \forall x, y \in \Omega$ 。

**假设 4.5** 存在常数  $L > 0$ , 使得  $f$  相对于  $\phi$  在  $C$  上是  $L$ -相对光滑的。

**命题 4.1** 在假设 4.2 与假设 4.4 成立的条件下, 对于任意  $x, y \in \Omega$ , 精确 Bregman 距离  $D_\phi(x, y)$  与其二阶近似 Bregman 距离  $\tilde{D}_\phi(x, y)$  在集合  $\Omega$  存在  $c_1 = \sigma/M_\phi, c_2 = M_\phi/\sigma$  满足

$$c_1 \tilde{D}_\phi(x, y) \leq D_\phi(x, y) \leq c_2 \tilde{D}_\phi(x, y).$$

**证** 由假设 4.2, 函数  $\phi$  在集合  $\Omega$  上是  $\sigma$ -强凸的, 以及假设 4.4,  $\phi$  在  $\Omega$  上有一致上界  $M_\phi$ 。因此对于任意  $z \in \Omega$ , 其 Hessian 矩阵满足

$$\sigma I \preceq \nabla^2 \phi(z) \preceq M_\phi I. \tag{4.2}$$

根据二阶近似 Bregman 距离定义, 由于  $y \in \Omega$ , 对式(4.2)左右同时作用于向量  $x - y$ , 可得

$$\frac{\sigma}{2} \|x - y\|^2 \leq \tilde{D}_\phi(x, y) \leq \frac{M_\phi}{2} \|x - y\|^2. \tag{4.3}$$

根据带有积分余项形式的泰勒展开公式,  $D_\phi(x, y) = \int_0^1 (1-s) \langle \nabla^2 \phi(y + s(x-y))(x-y), x-y \rangle ds$ .

由于集合  $\Omega$  为凸集, 对于任意  $s \in [0, 1]$ , 有  $y + s(x-y) \in \Omega$ . 因此在积分区间内 Hessian 同样满足矩阵界(4.2), 于是

$$\int_0^1 (1-s) \sigma \|x - y\|^2 ds \leq D_\phi(x, y) \leq \int_0^1 (1-s) M_\phi \|x - y\|^2 ds, \\ \frac{\sigma}{2} \|x - y\|^2 \leq D_\phi(x, y) \leq \frac{M_\phi}{2} \|x - y\|^2. \tag{4.4}$$

由公式(4.3)与公式(4.4)可得

$$D_\phi(x, y) \leq \frac{M_\phi}{2} \|x - y\|^2 = \frac{M_\phi}{\sigma} \left( \frac{\sigma}{2} \|x - y\|^2 \right) \leq \frac{M_\phi}{\sigma} \tilde{D}_\phi(x, y).$$

同理,

$$D_\phi(x, y) \geq \frac{\sigma}{2} \|x - y\|^2 = \frac{\sigma}{M_\phi} \left( \frac{M_\phi}{2} \|x - y\|^2 \right) \geq \frac{\sigma}{M_\phi} \tilde{D}_\phi(x, y),$$

$$\frac{\sigma}{M_\phi} \tilde{D}_\phi(x, y) \leq D_\phi(x, y) \leq \frac{M_\phi}{\sigma} \tilde{D}_\phi(x, y).$$

**推论 4.1** 在假设 4.5 成立的条件下, 对任意  $x, y \in \Omega$  成立

$$f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + \frac{LM_\phi}{2} \|x - y\|^2. \tag{4.5}$$

**证** 由公式(4.4)可知  $D_\phi(x, y) \leq \frac{M_\phi}{2} \|x - y\|^2$ , 带入上述不等式即可。

**引理 4.1** 假设 4.1~4.5 成立, 设参数  $\alpha \in (0, 1)$ ,  $\lambda > 0$ , 且惯性参数  $\beta_k \geq 0$ , 若步长  $t$  满足

$$t \leq \min \left\{ 1, \frac{\sigma(1-\alpha)}{\lambda LM_\phi} \right\},$$

则 iABPG 中的回溯线搜索条件(3.1)必在有限步内终止且存在统一下界  $t_{\min} \in (0, 1)$  使得对所有  $k$ , 满足

$$t_k \geq t_{\min} := \eta \min \left\{ 1, \frac{\sigma(1-\alpha)}{\lambda LM_\phi} \right\} > 0. \tag{4.6}$$

**证** 在第  $k$  步, 近似 Bregman 距离为  $\tilde{D}_\phi(x, y^k) := \frac{1}{2} \langle \nabla^2 \phi(y^k)(x - y^k), x - y^k \rangle$ , 由假设 4.2, 结合下半连续凸函数  $g(x)$ , 目标函数子问题是强凸的, 所以最优解  $x^{k+1}$  存在且唯一, 记  $d^k := x^{k+1} - y^k$ . 方向导数  $\Delta_k := \langle \nabla f(y^k), d^k \rangle + g(y^k + d^k) - g(y^k)$ .

当  $d^k = 0$  时, 条件(3.1)显然成立。

当  $d^k \neq 0$  时, 由子问题的一阶最优性必要条件可知, 存在次梯度  $\xi^{k+1} \in \partial g(x^{k+1})$  使得

$$\nabla f(y^k) + \xi^{k+1} + \frac{1}{\lambda} \nabla^2 \phi(y^k) d^k = 0, \tag{4.7}$$

利用凸函数  $g$  的次梯度不等式  $g(y^k + d^k) - g(y^k) \leq \langle \xi^{k+1}, d^k \rangle$ , 将其代入  $\Delta_k$  的定义中, 并结合假设 4.2 中的  $\nabla^2 \phi(y^k) \geq \sigma I$  以及等式(4.7)可得  $\langle \nabla f(y^k), d^k \rangle + g(y^k + d^k) - g(y^k) \leq -\frac{1}{\lambda} \langle \nabla^2 \phi(y^k) d^k, d^k \rangle \leq -\frac{\sigma}{\lambda} \|d^k\|^2$ 。

由于当前  $d^k \neq 0$ , 该方向导数  $\Delta_k$  严格为负。由命题 4.1,

$$f(y^k + td^k) \leq f(y^k) + t \langle \nabla f(y^k), d^k \rangle + \frac{LM_\phi}{2} t^2 \|d^k\|^2.$$

其次, 由  $g$  的凸性,  $g(y^k + td^k) = g((1-t)y^k + t(y^k + d^k)) \leq (1-t)g(y^k) + tg(y^k + d^k)$ 。

将两式相加, 得到

$$\varphi(y^k + td^k) \leq \varphi(y^k) + t \left( \langle \nabla f(y^k), d^k \rangle + g(y^k + d^k) - g(y^k) \right) + \frac{LM_\phi}{2} t^2 \|d^k\|^2, \quad (4.8)$$

若步长  $t$  满足  $t \leq \frac{2\sigma(1-\alpha)}{\lambda LM_\phi}$ , 则  $\frac{LM_\phi}{2} t^2 \|d^k\|^2 \leq (1-\alpha)t \left( -\frac{\sigma}{\lambda} \|d^k\|^2 \right)$ , 带入公式(4.8)可得

$$\varphi(y^k + td^k) \leq \varphi(y^k) + \alpha t \left( \langle \nabla f(y^k), d^k \rangle + g(y^k + d^k) - g(y^k) \right).$$

因此当  $t \leq \min \left\{ 1, \frac{2\sigma(1-\alpha)}{\lambda LM_\phi} \right\}$  时, 回溯线搜索条件(3.1)必在有限步内终止。由于回溯线搜索通过

$t \leftarrow \eta t, \eta \in (0, 1)$  逐步缩小步长, 因此在有限次迭代后落入该安全区间内, 可得

$$t_k \geq \min \left\{ 1, \eta \frac{2\sigma(1-\alpha)}{\lambda LM_\phi} \right\}.$$

**引理 4.2** 假设 4.1~4.5 成立, 则存在常数  $C_1 > C_2 \geq 0$ , 使得对于任意  $k \geq 1$ , 如下不等式成立

$$\varphi(x^{k+1}) - \varphi(x^k) \leq -C_1 \|x^{k+1} - x^k\|^2 + C_2 \|x^k - x^{k-1}\|^2,$$

其中  $C_1 := \frac{\sigma}{\lambda} - \frac{L_f}{2} - \frac{M_\phi \bar{\beta}}{2\lambda}$ ,  $C_2 := \frac{L_f \bar{\beta}^2}{2} + \frac{M_\phi \bar{\beta}}{2\lambda}$ 。

**证** 根据假设 4.4

$$f(x^{k+1}) \leq f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L_f}{2} \|x^{k+1} - x^k\|^2. \quad (4.9)$$

由子问题的一阶最优性条件可得,  $\xi^{k+1} = -\nabla f(y^k) - \frac{1}{\lambda} \nabla^2 \phi(y^k) d^k$ 。

由于  $g(x)$  是凸函数,  $g(x^{k+1}) - g(x^k) \leq \langle \xi^{k+1}, x^{k+1} - x^k \rangle$ 。将上述  $\xi^{k+1}$  的表达式代入, 得到

$$g(x^{k+1}) - g(x^k) \leq \left\langle -\nabla f(y^k) - \frac{1}{\lambda} \nabla^2 \phi(y^k) d^k, x^{k+1} - x^k \right\rangle. \quad (4.10)$$

将等式(4.9)与(4.10)相加可得

$$\begin{aligned} \varphi(x^{k+1}) - \varphi(x^k) &\leq \langle \nabla f(x^k) - \nabla f(y^k), x^{k+1} - x^k \rangle \\ &\quad - \left[ \frac{1}{\lambda} \langle \nabla^2 \phi(y^k) d^k, x^{k+1} - x^k \rangle \right] + \frac{L_f}{2} \|x^{k+1} - x^k\|^2. \end{aligned}$$

根据 Cauchy-Schwarz 不等式、 $L_f$ -Lipschitz 连续性以及惯性定义,

$$\begin{aligned} \langle \nabla f(x^k) - \nabla f(y^k), x^{k+1} - x^k \rangle &\leq \|\nabla f(x^k) - \nabla f(y^k)\| \cdot \|x^{k+1} - x^k\| \\ &\leq L_f \|x^k - y^k\| \cdot \|x^{k+1} - x^k\| \\ &= L_f \beta_k \|x^k - x^{k-1}\| \cdot \|x^{k+1} - x^k\|, \end{aligned}$$

利用 Young 不等式和  $x^{k+1} - y^k = (x^{k+1} - x^k) - \beta_k(x^k - x^{k-1})$ , 可知

$$\langle \nabla f(x^k) - \nabla f(y^k), x^{k+1} - x^k \rangle \leq \frac{L_f}{2} \|x^{k+1} - x^k\|^2 + \frac{L_f \beta_k^2}{2} \|x^k - x^{k-1}\|^2, \tag{4.11}$$

$$\begin{aligned} &\frac{1}{\lambda} \langle \nabla^2 \phi(y^k)(x^{k+1} - y^k), x^{k+1} - x^k \rangle \\ &= \frac{1}{\lambda} \langle \nabla^2 \phi(y^k)(x^{k+1} - x^k), x^{k+1} - x^k \rangle - \frac{\beta_k}{\lambda} \langle \nabla^2 \phi(y^k)(x^k - x^{k-1}), x^{k+1} - x^k \rangle. \end{aligned}$$

由假设 4.2 可知核函数  $\Phi$  是  $\sigma$ -强凸的, 即  $\nabla^2 \Phi(y^k) \succeq \sigma I$ , 因此

$$-\frac{1}{\lambda} \langle \nabla^2 \phi(y^k)(x^{k+1} - x^k), x^{k+1} - x^k \rangle \leq -\frac{\sigma}{\lambda} \|x^{k+1} - x^k\|^2.$$

由假设 4.4  $\nabla^2 \phi(y^k) \preceq M_\phi I$ , 通过 Cauchy-Schwarz、Young 不等式可得

$$\begin{aligned} \frac{\beta_k}{\lambda} \langle \nabla^2 \phi(y^k)(x^k - x^{k-1}), x^{k+1} - x^k \rangle &\leq \frac{\beta_k}{\lambda} \|\nabla^2 \phi(y^k)\| \cdot \|x^k - x^{k-1}\| \cdot \|x^{k+1} - x^k\| \\ &\leq \frac{M_\phi \beta_k}{\lambda} \|x^k - x^{k-1}\| \cdot \|x^{k+1} - x^k\| \\ &\leq \frac{M_\phi \beta_k}{2\lambda} \|x^{k+1} - x^k\|^2 + \frac{M_\phi \beta_k}{2\lambda} \|x^k - x^{k-1}\|^2. \end{aligned}$$

$$-\frac{1}{\lambda} \langle \nabla^2 \phi(y^k)(x^{k+1} - y^k), x^{k+1} - x^k \rangle \leq -\frac{\sigma}{\lambda} \|x^{k+1} - x^k\|^2 + \frac{M_\phi \beta_k}{2\lambda} \|x^{k+1} - x^k\|^2 + \frac{M_\phi \beta_k}{2\lambda} \|x^k - x^{k-1}\|^2. \tag{4.12}$$

将等式(4.11)与(4.12)相加可得

$$\varphi(x^{k+1}) - \varphi(x^k) \leq -\left(\frac{\sigma}{\lambda} - L_f - \frac{M_\phi \beta_k}{2\lambda}\right) \|x^{k+1} - x^k\|^2 + \left(\frac{L_f \beta_k^2}{2} + \frac{M_\phi \beta_k}{2\lambda}\right) \|x^k - x^{k-1}\|^2,$$

记括号内常数分别为  $C_1$  和  $C_2$ , 由  $\beta_k \leq \bar{\beta}$ ,  $C_1 > C_2 \geq 0$ , 引理 4.2 得证。

**引理 4.3** 定义序列的增广状态变量  $z^k := (x^k, x^{k-1})$ ,  $\mathcal{E}(z^k) := \varphi(x^k) + \delta \|x^k - x^{k-1}\|^2$ 。

(i) 存在常数  $\delta > 0$  与  $\gamma > 0$ , 使得序列  $\{\mathcal{E}(z^k)\}$  严格单调递减, 即

$$\mathcal{E}(z^{k+1}) - \mathcal{E}(z^k) \leq -\gamma \|x^{k+1} - x^k\|^2 \leq 0$$

(ii) 算法生成的序列  $\{x^k\}$  与  $\{y^k\}$  均有界, 且始终包含于紧集  $\Omega$  中;

(iii)  $\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\| = 0$ 。

**证** (i) 由引理 4.2 可得

$$\begin{aligned} \mathcal{E}(z^{k+1}) - \mathcal{E}(z^k) &= \left(\varphi(x^{k+1}) + \delta \|x^{k+1} - x^k\|^2\right) - \left(\varphi(x^k) + \delta \|x^k - x^{k-1}\|^2\right) \\ &= \left[\varphi(x^{k+1}) - \varphi(x^k)\right] + \delta \|x^{k+1} - x^k\|^2 - \delta \|x^k - x^{k-1}\|^2 \\ &\leq (-C_1 + \delta) \|x^{k+1} - x^k\|^2 + (C_2 - \delta) \|x^k - x^{k-1}\|^2. \end{aligned}$$

选取  $\delta \in (C_2, C_1)$  并令  $\gamma = C_1 - \delta > 0$ , 则  $\mathcal{E}(z^{k+1}) - \mathcal{E}(z^k) \leq -\gamma \|x^{k+1} - x^k\|^2 \leq 0$ 。

(ii) 由于序列  $\{\mathcal{E}(z^k)\}$  单调递减, 故  $\mathcal{E}(z^{k+1}) \leq \mathcal{E}(z^k) \leq \mathcal{E}(z^0)$ 。对任意  $k \geq 1$  有

$$\varphi(x^k) \leq \varphi(x^k) + \delta \|x^k - x^{k-1}\|^2 = \mathcal{E}(z^k) \leq \mathcal{E}(z^0) = \varphi(x^0).$$

我们可得  $\varphi(x^k) \leq \varphi(x^0), \forall k \geq 1$ 。根据假设 4.3 的定义, 因此  $\{x^k\} \subset \Omega_0$ 。再由假设 4.4,  $\Omega_0 \subset \text{int } \Omega$ , 且外推点满足  $y^k \in \Omega$ , 故

$$\{x^k\} \subset \Omega_0 \subset \text{int } \Omega, \{y^k\} \subset \Omega.$$

(iii) 对  $\mathcal{E}(z^{k+1}) - \mathcal{E}(z^k) \leq -\gamma \|x^{k+1} - x^k\|^2$  从  $k=0$  累加到  $N$

$$\sum_{k=0}^N \gamma \|x^{k+1} - x^k\|^2 \leq \sum_{k=0}^N (\mathcal{E}(z^k) - \mathcal{E}(z^{k+1})) = \mathcal{E}(z^0) - \mathcal{E}(z^{N+1}),$$

由前面可知序列有界, 且目标函数  $\varphi$  有下界  $\varphi^*$ 。同时  $\delta \|x^{N+1} - x^N\|^2 \geq 0$ , 因此  $\mathcal{E}(z^{N+1}) \geq \varphi(x^{N+1}) \geq \varphi^*$ 。带入上式可得  $\sum_{k=0}^N \gamma \|x^{k+1} - x^k\|^2 \leq \mathcal{E}(z^0) - \varphi^* < \infty$ , 故

$$\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\| = 0.$$

**引理 4.4** 存在常数  $C_3 > 0$  使得

$$\text{dist}(0, \partial \mathcal{E}(z^{k+1})) \leq C_3 (\|x^{k+1} - x^k\| + \|x^k - x^{k-1}\|).$$

**证**  $\mathcal{E}(z^k) = \varphi(x^k) + \delta \|x^k - x^{k-1}\|^2$  的子梯度

$$\partial \mathcal{E}(z^{k+1}) = \begin{pmatrix} \nabla f(x^{k+1}) + \partial g(x^{k+1}) + 2\delta(x^{k+1} - x^k) \\ -2\delta(x^{k+1} - x^k) \end{pmatrix},$$

从引理 4.1 中最优性条件  $\xi^{k+1} \in \partial g(x^{k+1})$

$$\xi^{k+1} = -\nabla f(y^k) - \frac{1}{\lambda} \nabla^2 \phi(y^k)(x^{k+1} - y^k),$$

取  $w^{k+1} \in \partial \mathcal{E}(z^{k+1})$ , 利用子问题最优性条件产生的次梯度  $\xi^{k+1} \in \partial g(x^{k+1})$

$$\xi^{k+1} = -\nabla f(y^k) - \frac{1}{\lambda} \nabla^2 \Phi(y^k)(x^{k+1} - y^k)$$

代入上式并结合  $L_f$  和  $M_\phi$  的界限, 得

$$\|\nabla f(x^{k+1}) + \xi^{k+1}\| \leq L_f \|x^{k+1} - y^k\| + \frac{M_\phi}{\lambda} \|x^{k+1} - y^k\| = \left(L_f + \frac{M_\phi}{\lambda}\right) \|x^{k+1} - y^k\|.$$

根据  $\|x^{k+1} - y^k\| \leq \|x^{k+1} - x^k\| + \bar{\beta} \|x^k - x^{k-1}\|$ , 可得

$$\begin{aligned} \text{dist}(0, \partial \mathcal{E}(z^{k+1})) &\leq \left(L_f + \frac{M_\phi}{\lambda}\right) (\|x^{k+1} - x^k\| + \bar{\beta} \|x^k - x^{k-1}\|) + 4\delta \|x^{k+1} - x^k\| \\ &= \left(L_f + \frac{M_\phi}{\lambda} + 4\delta\right) \|x^{k+1} - x^k\| + \bar{\beta} \left(L_f + \frac{M_\phi}{\lambda}\right) \|x^k - x^{k-1}\| \end{aligned}$$

令  $C_3 = \max \left\{ L_f + \frac{M_\phi}{\lambda} + 4\delta, \bar{\beta} \left( L_f + \frac{M_\phi}{\lambda} \right) \right\}$ , 引理得证。

**假设 4.6** 目标函数  $\varphi(x) = f(x) + g(x)$  在其定义域上满足 Kurdyka-Łojasiewicz (KL) 性质。

**定理 4.1** 在假设 4.1~4.6 成立下, 序列具有有限长度, 即  $\sum_{k=1}^{\infty} \|x^{k+1} - x^k\| < \infty$ , 且强收敛至广义平稳点  $x^*$ 。

证 由引理 4.3 可知, 序列  $\{\mathcal{E}(z^k)\}$  单调递减且有下界, 设序列  $\{\mathcal{E}(z^k)\}$  递减收敛于极限  $\mathcal{E}^*$ 。若某步  $\mathcal{E}(z^k) = \mathcal{E}^*$ , 算法已在平稳点终止。

假设对所有  $k$  均有  $\mathcal{E}(z^k) > \mathcal{E}^*$ , 由目标函数满足 KL 性质, 存在足够大的  $k_0 \in \mathbb{N}$ , 使得对所有  $k \geq k_0$ , 存在连续凹函数  $\psi$  满足

$$\psi'(\mathcal{E}(z^{k+1}) - \mathcal{E}^*) \cdot \text{dist}(0, \partial \mathcal{E}(z^{k+1})) \geq 1.$$

记  $E_{k+1} := \psi(\mathcal{E}(z^{k+1}) - \mathcal{E}^*) - \psi(\mathcal{E}(z^{k+2}) - \mathcal{E}^*)$ , 凹函数  $\psi$  的性质,  $E_{k+1} \geq \psi'(\mathcal{E}(z^{k+1}) - \mathcal{E}^*) \cdot (\mathcal{E}(z^{k+1}) - \mathcal{E}(z^{k+2}))$ 。

根据引理 4.3 中  $\mathcal{E}(z^{k+1}) - \mathcal{E}(z^{k+2}) \geq \gamma \|x^{k+2} - x^{k+1}\|^2$  和引理 4.4 可得  $E_{k+1} \geq \frac{\gamma \|x^{k+2} - x^{k+1}\|^2}{C_3 (\|x^{k+1} - x^k\| + \|x^k - x^{k-1}\|)}$ 。

$$\|x^{k+2} - x^{k+1}\| \leq \sqrt{\frac{C_3}{\gamma} E_{k+1} \cdot (\|x^{k+1} - x^k\| + \|x^k - x^{k-1}\|)}$$

根据 Young 不等式,  $\|x^{k+2} - x^{k+1}\| \leq \frac{1}{4} \|x^{k+1} - x^k\| + \frac{1}{4} \|x^k - x^{k-1}\| + \frac{C_3}{\gamma} E_{k+1}$

对上式从  $k=1$  到  $N$  累加, 记部分和  $S_N = \sum_{k=k_0}^N \|x^{k+2} - x^{k+1}\|$ , 则

$$S_N \leq \frac{1}{4} \sum_{k=k_0}^N \|x^{k+1} - x^k\| + \frac{1}{4} \sum_{k=k_0}^N \|x^k - x^{k-1}\| + \frac{C_3}{\gamma} \sum_{k=k_0}^N E_{k+1},$$

根据  $\sum_{k=k_0}^N \|x^{k+1} - x^k\| \leq S_N + \|x^{k_0+1} - x^{k_0}\|$ ,  $\sum_{k=k_0}^N \|x^k - x^{k-1}\| \leq S_N + \|x^{k_0+1} - x^{k_0}\| + \|x^{k_0} - x^{k_0-1}\|$ , 代入后合并  $S_N$  项

$$\left(1 - \frac{1}{4} - \frac{1}{4}\right) S_N = \frac{1}{2} S_N \leq \frac{1}{2} \|x^{k_0+1} - x^{k_0}\| + \frac{1}{4} \|x^{k_0} - x^{k_0-1}\| + \frac{C_3}{\gamma} \sum_{k=k_0}^N E_{k+1}$$

由于  $\sum_{k=k_0}^N E_{k+1} = \psi(\mathcal{E}(z^{k_0+1}) - \mathcal{E}^*) - \psi(\mathcal{E}(z^{N+2}) - \mathcal{E}^*) \leq \psi(\mathcal{E}(z^{k_0+1}) - \mathcal{E}^*)$  为有限常数, 令  $N \rightarrow \infty$ , 可知  $\sum_{k=1}^{\infty} \|x^{k+1} - x^k\| < \infty$  即序列  $\{x^k\}$  是一个 Cauchy 序列。因此, 它全局收敛到唯一的极限点  $x^*$ 。由于序列收敛, 差分趋于零, 结合惯性公式亦有外推点  $y^k \rightarrow x^*$ 。由引理 4.1 中子问题的最优性条件

$$-\frac{1}{\lambda t_k} \nabla^2 \phi(y^k)(x^{k+1} - y^k) - \nabla f(y^k) \in \partial g(y^k + d^k),$$

由于  $\|x^{k+1} - y^k\| \rightarrow 0$  且  $\nabla^2 \phi$  在紧集  $\Omega$  上有界, 且  $t_k \geq t_{\min}^* > 0$ , 故左侧首项趋于 0。又因  $f \in C^1$  及  $g$  的次微分算子具有闭图像属性, 令  $k \rightarrow \infty$ , 可得  $0 - \nabla f(x^*) \in \partial g(x^*)$ , 即极限点  $x^*$  是目标函数  $\phi$  的广义平稳点。

## 5. 数值实验

例 5.1 首先我们考虑非负线性系统[38]并进行最小化

$$\min_{x \in \mathbb{R}_+^n} D_{KL}(Ax, b) + \theta_1 \|x\|,$$

其中  $D_{KL}(x, y) = \sum_{i=1}^m \left( x_i \log \frac{x_i}{y_i} + y_i - x_i \right)$  是 Kullback-Leibler 散度,  $A \in \mathbb{R}_+^{m \times n}$ ,  $b \in \mathbb{R}_+^m$ ,  $\theta_1 > 0$ 。令

$f(x) = D_{KL}(Ax, b)$ ,  $g(x) = \theta_1 \|x\|$ 。我们选取文献[20]的 PGL 算法、[38]中的 BPG 和[37]中的 ABPG 算法作为对比。所有代码均在 Intel(R) Core(TM) i7-10510U CPU @ 1.80 GHz (2.30 GHz), Matlab (R2020b)上运行。为保证对比的公平性, 算法一、ABPG、BPG 及 PGL 在内的所有对比算法均采用原生 MATLAB 代

码实现，不依赖任何商业求解器。其中矩阵  $A$  为列随机归一化矩阵， $b = Ax^*$ ，我们从 i.i.d.  $\mathcal{N}(0,1)$  分布中生成了真实值  $x^*$ ，稀疏度 10%，问题规模  $n = 400$ ， $m = 1000$ ， $\theta = 0.01$ ， $L_f = 1$ ， $\lambda = 0.9$ ， $\alpha = 10^{-4}$ ， $\eta = 0.5$ ， $\beta = 0.8$ ， $\sigma = 1$ ，算法停止条件是达到最大迭代次数 2000 或  $\|x^k - x^{k-1}\| \leq 10^{-6}$ ，选取核函数  $\phi_{\text{BPG}}(x) = \sum_{i=1}^n x_i \log x_i$ ， $\phi_1 = \phi_{\text{ABPG}}(x) = \sum_{i=1}^n x_i \log x_i + \frac{1}{2} \|x\|_2^2$ 。

图 1 展示了四种算法在目标函数值随迭代次数的变化曲线，可以看出，本文提出的算法 1 的收敛速度明显优于 ABPG、BPG 和 PGL，在约 200 次迭代内目标函数值已下降至  $10^{-6}$ ，而其他三种算法在相同迭代次数下仍保持较高残差。图 2 进一步给出了相对误差随 CPU 时间的变化。算法 1 在约 0.5 秒内即可将相对误差降至  $10^{-2}$  以下，而 ABPG、BPG 和 PGL 在相同 CPU 时间内误差仍高于算法 1 约一个数量级。PGL 虽最终能达到与算法 1 相近的精度，但耗时远高于算法 1。此外本实验通过对高维  $n = 2000$   $m = 5000$

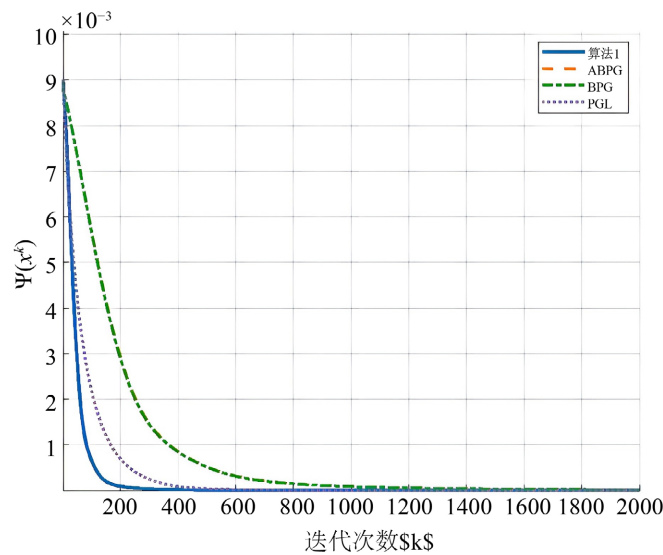


Figure 1. Variation of the objective function  
图 1. 目标函数变化

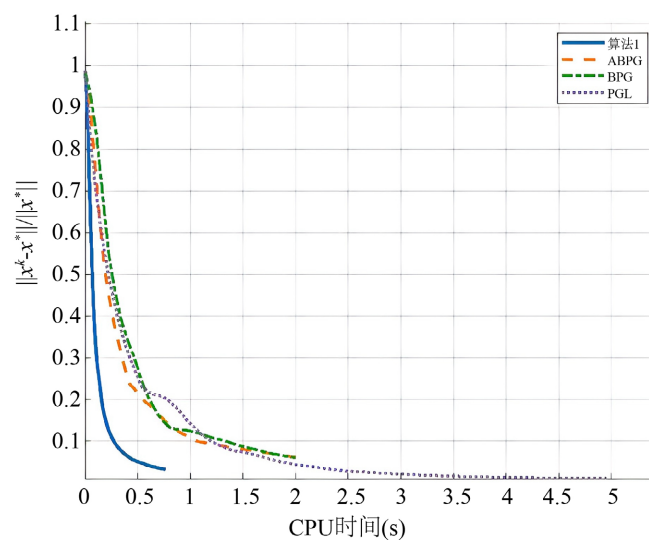


Figure 2. Relative error  
图 2. 相对误差

维度的非负线性系统进行数值模拟，对比了四种不同性质的优化算法。实验结果如图 3。在图 3 可以清晰观察到，引入惯性项的算法 1 曲线下降最为陡峭，在 2000 次迭代内达到了最低的相对误差。相比之下，未加入加速项的 ABPG 虽然也利用了 Bregman 几何结构，但其收敛效率明显滞后；PGL 算法以及 BPG 算法表现最弱。另一方面为考察惯性参数  $\beta$  对算法 1 性能的影响，我们选取不同的  $\beta$  参数进行计算速率分析，结果如图 4 当  $\beta = 0.8$  和  $\beta = 0.9$  时，算法在 CPU 时间约 1.5 秒内即可将相对误差降至  $10^{-2}$  以下，且曲线平稳下降； $\beta = 1.1$  虽初期下降较快，但后期出现轻微振荡； $\beta = 0.1$  和  $\beta = 0.4$  收敛明显较慢；而  $\beta = 1.5$  则出现强烈的数值振荡，收敛性变差。实验表明， $\beta \in [0.8, 0.9]$  是本问题的最优区间，既能有效加速又能保持数值稳定性，与理论上  $\beta \in [0, \bar{\beta}]$  的分析一致。

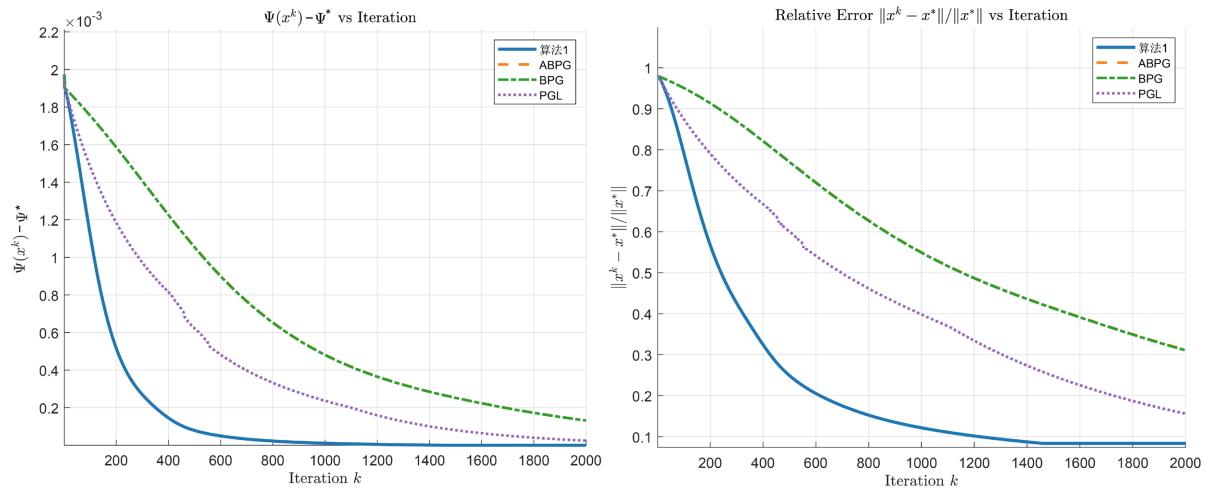


Figure 3.  $n = 2000$   $m = 5000$

图 3.  $n = 2000$   $m = 5000$

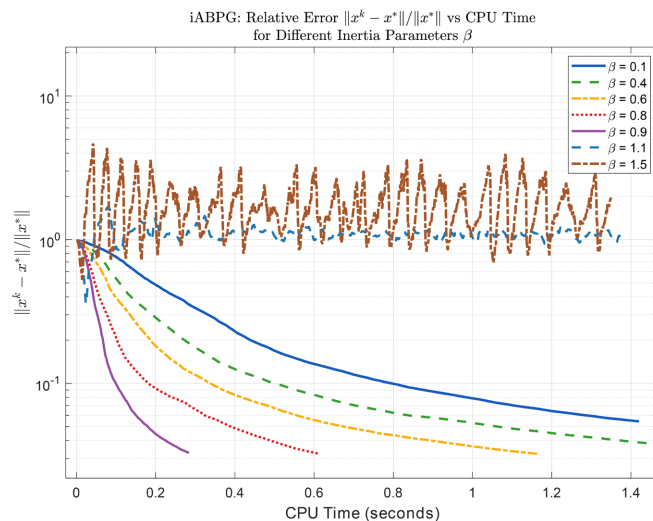


Figure 4. Computational efficiency for different  $\beta$

图 4. 不同  $\beta$  的计算效率

例 5.2 我们考虑稀疏  $l_p$  正则化最小二乘问题[39] [40]

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|^2 + \frac{\theta_p}{p} \|x\|_p^p,$$

其中  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}$ ,  $\theta_p > 0$ 。所有代码均在 Intel(R) Core(TM) i7-10510U CPU @ 1.80 GHz (2.30 GHz), Matlab (R2020b)上运行。当  $p \in (1, 2)$  时, 目标函数的梯度  $\nabla f$  不满足全局 Lipschitz 连续。其中  $A \in \mathbb{R}^{m \times n}$ ,  $b = Ax^*$ , 真实值  $x^*$  服从高斯随机分布, 稀疏度 5%, 问题规模  $n = 1000$ ,  $m = 2000$ 。具体参数如下  $p = 1.1, 3.0$ ,  $\theta_p = 0.05$ , 统一设定初始步长  $L_f = \|A\|_2^2$ ,  $\lambda = \frac{1}{L_f}$ ,  $\lambda_{\text{BPG}} = \frac{\mu_p}{L_f}$ ,  $\beta = 0.3$ ,  $\eta = 0.9$ ,  $\sigma = 1$ , 线搜索参数  $\alpha = 0.99$ 。此外, 涉及的近端映射子问题均由 Newton-Raphson 迭代法求解, 收敛精度设为  $10^{-12}$ 。算法停止条件是达到最大迭代次数 1000 或  $\|x^k - x^{k-1}\| \leq 10^{-6}$ , 核函数  $\phi(x) = \frac{1}{2} \|x\|_2^2 + \frac{1}{p} \|x\|_p^p$ 。

图 5 展示了四种算法的目标函数值  $\phi(x^k)$  和相对误差  $\|x^k - x^*\| / \|x^*\|$  随迭代次数的变化曲线。当  $p = 1.1$  时, 算法 1 在约 300 次迭代内已将目标函数值降至  $10^0$  量级, 相对误差降至  $10^{-2}$  以下, 而 ABPG、BPG 和 PGL 下降明显缓慢, BPG 甚至在 1000 次迭代后仍保持较高残差, PGL 则陷入平台期, 而算法 1 通过

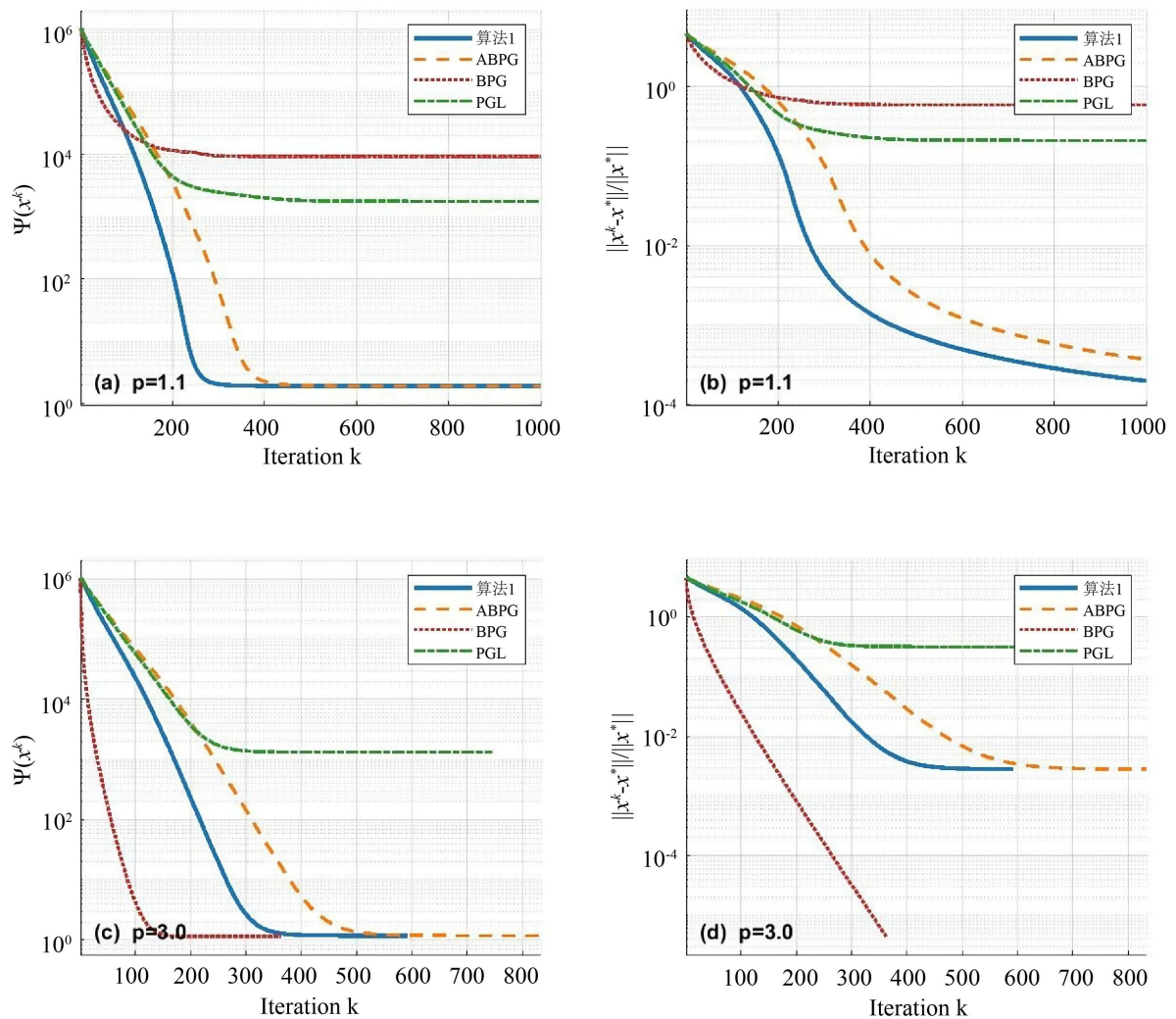
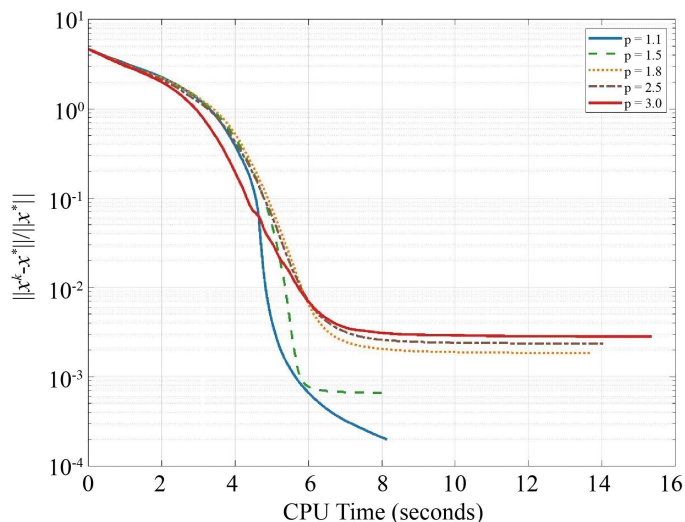


Figure 5. Algorithm comparison for different  $p$  values  
图 5. 不同  $p$  值的算法对比



**Figure 6.** Relative error reduction time under different  $p$  values  
**图 6.** 不同  $p$  值下的相对误差缩减时间

近似 Bregman 距离与惯性外推实现了加速。图 6 进一步给出了算法 1 在不同  $p$  值 1.1、1.5、1.8、2.5、3.0 下的相对误差随 CPU 时间的变化曲线。当  $p$  接近 1 时，算法 1 在约 8 秒内即可将相对误差降至  $10^{-4}$  量级，收敛速度远快于较大  $p$  值的情形；随着  $p$  增大，曲线趋于平缓但仍保持单调下降。这表明算法 1 对  $p$  的变化具有较好的鲁棒性，尤其在最具有挑战性的非 Lipschitz 问题展现出明显优势。

## 6. 结论与展望

本文提出了一种近似 Bregman 近端梯度算法及其带惯性变，用于求解复合非凸非光滑优化问题。该算法通过二阶近似 Bregman 距离将子问题转化为闭式可解形式，仅需相对光滑性而无需全局 Lipschitz 连续性即可保证全局收敛，并在 KL 性质下证明了算法的全局收敛性。数值实验在两类典型问题上进行，非负线性系统(KL 正则化)和稀疏  $l_p$  正则化最小二乘问题。结果表明，算法 1 在收敛速度、计算效率和数值稳定性上均显著优于 ABPG、BPG 和 PGL，尤其在梯度非 Lipschitz 的非凸情形中优势最为突出。

尽管本文提出的算法在解决缺乏全局 Lipschitz 连续性的复合优化问题上表现很好，但仍存在一些局限性。例如近似 Bregman 子问题中当问题维度  $n$  极大时，求解二次规划子问题复杂度可达  $O(n^3)$ ，在深度学习、三维医学图像重建等超大规模场景下难以承受，为此可以考虑引入 L-BFGS 或拟牛顿近似技术来替代海森矩阵的计算，以提升大规模问题的扩展性。其次，全局收敛理论高度依赖静态参数  $\lambda$  与  $\beta$ ，而这些参数需预先精确获知局部 Lipschitz 常数、相对光滑常数及核函数曲率界，在黑盒问题与深度神经网络等场景中难以精确评估，后续可引入 Barzilai-Borwein 步长等自适应策略与动态惯性机制实现参数自动调整以增强鲁棒性。最后，本算法仍依赖于全梯度的计算与精确的线搜索机制，在大规模任务中全梯度计算代价高昂，且引入随机梯度后因噪声方差会使现有线搜索失效，基于单步能量下降的 Lyapunov 理论体系需重构，如何将本文的惯性截断与近似 Bregman 距离融入随机方差缩减框架中，如 SVRG, SAGA 并在期望意义下的收敛性理论，将是我们下一步重点研究的方向。

## 参考文献

- [1] Nikolova, M., Ng, M.K., Zhang, S. and Ching, W. (2008) Efficient Reconstruction of Piecewise Constant Images Using Nonsmooth Nonconvex Minimization. *SIAM Journal on Imaging Sciences*, **1**, 2-25. <https://doi.org/10.1137/070692285>

- [2] Bian, W. and Chen, X. (2015) Linearly Constrained Non-Lipschitz Optimization for Image Restoration. *SIAM Journal on Imaging Sciences*, **8**, 2294-2322. <https://doi.org/10.1137/140985639>
- [3] Bolte, J., Sabach, S. and Teboulle, M. (2014) Proximal Alternating Linearized Minimization for Nonconvex and Nonsmooth Problems. *Mathematical Programming*, **146**, 459-494. <https://doi.org/10.1007/s10107-013-0701-9>
- [4] Donoho, D.L. (2006) Compressed Sensing. *IEEE Transactions on Information Theory*, **52**, 1289-1306. <https://doi.org/10.1109/tit.2006.871582>
- [5] Wang, Y., Yang, J., Yin, W. and Zhang, Y. (2008) A New Alternating Minimization Algorithm for Total Variation Image Reconstruction. *SIAM Journal on Imaging Sciences*, **1**, 248-272. <https://doi.org/10.1137/080724265>
- [6] Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. (2011) Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends® in Machine Learning*, **3**, 1-122. <https://doi.org/10.1561/22000000016>
- [7] 文再文, 印卧涛, 刘歆, 等. 压缩感知和稀疏优化简介[J]. 运筹学学报, 2012, 16(3): 49-64.
- [8] Osher, S., Burger, M., Goldfarb, D., Xu, J. and Yin, W. (2005) An Iterative Regularization Method for Total Variation-Based Image Restoration. *Multiscale Modeling & Simulation*, **4**, 460-489. <https://doi.org/10.1137/040605412>
- [9] Yin, W., Osher, S., Goldfarb, D. and Darbon, J. (2008) Bregman Iterative Algorithms for  $\ell_1$ -Minimization with Applications to Compressed Sensing. *SIAM Journal on Imaging Sciences*, **1**, 143-168. <https://doi.org/10.1137/070703983>
- [10] Shalev-Shwartz, S. and Zhang, T. (2013) Accelerated Proximal Stochastic Dual Coordinate Ascent for Regularized Loss Minimization. arXiv:1309.2375
- [11] Xiao, L. and Zhang, T. (2014) A Proximal Stochastic Gradient Method with Progressive Variance Reduction. *SIAM Journal on Optimization*, **24**, 2057-2075. <https://doi.org/10.1137/140961791>
- [12] Pappas, P. (2017) A Multilevel Proximal Gradient Algorithm for a Class of Composite Optimization Problems. *SIAM Journal on Scientific Computing*, **39**, S681-S701. <https://doi.org/10.1137/16m1082299>
- [13] Zhang, N., Liu, X. and Li, Q. (2026) Parameterized Proximal-Gradient Algorithms for L1/L2 Sparse Signal Recovery. *Applied and Computational Harmonic Analysis*, **82**, Article 101835. <https://doi.org/10.1016/j.acha.2025.101835>
- [14] Malitsky, Y. and Mishchenko, K. (2024) Adaptive Proximal Gradient Method for Convex Optimization. *Advances in Neural Information Processing Systems*, **37**, 100670-100697. <https://doi.org/10.52202/079017-3193>
- [15] 郇旭东. 复合凸优化的快速邻近点算法[J]. 计算数学, 2020, 42(4): 385-404.
- [16] Zhang, H. and Xu, Z. (2024) An Alternating Proximal Gradient Algorithm for Nonsmooth Nonconvex-Linear Minimax Problems with Coupled Linear Constraints. *Journal of the Operations Research Society of China*, 1-19. <https://doi.org/10.1007/s40305-024-00550-3>
- [17] Rockafellar, R.T. (1976) Monotone Operators and the Proximal Point Algorithm. *SIAM Journal on Control and Optimization*, **14**, 877-898. <https://doi.org/10.1137/0314056>
- [18] Polyak, B.T. (1963) Gradient Methods for Minimizing Functionals. *Zhurnal vychislitel'noi matematiki i matematicheskoi fiziki*, **3**, 643-653.
- [19] Nesterov, Y. (1983) A Method for Solving the Convex Programming Problem with Convergence Rate  $O(1/k^2)$ . *Doklady Akademii Nauk SSSR*, **269**, 543-547.
- [20] Beck, A. and Teboulle, M. (2009) A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, **2**, 183-202. <https://doi.org/10.1137/080716542>
- [21] Bregman, L.M. (1967) The Relaxation Method of Finding the Common Point of Convex Sets and Its Application to the Solution of Problems in Convex Programming. *USSR Computational Mathematics and Mathematical Physics*, **7**, 200-217. [https://doi.org/10.1016/0041-5553\(67\)90040-7](https://doi.org/10.1016/0041-5553(67)90040-7)
- [22] Nemirovskij, A.S. and Yudin, D.B. (1983) Problem Complexity and Method Efficiency in Optimization.
- [23] Lu, H., Freund, R.M. and Nesterov, Y. (2018) Relatively Smooth Convex Optimization by First-Order Methods, and Applications. *SIAM Journal on Optimization*, **28**, 333-354. <https://doi.org/10.1137/16m1099546>
- [24] Bolte, J., Sabach, S., Teboulle, M. and Vaisbourd, Y. (2018) First Order Methods Beyond Convexity and Lipschitz Gradient Continuity with Applications to Quadratic Inverse Problems. *SIAM Journal on Optimization*, **28**, 2131-2151. <https://doi.org/10.1137/17m1138558>
- [25] Zhou, Y., Liang, Y. and Shen, L. (2019) A Simple Convergence Analysis of Bregman Proximal Gradient Algorithm. *Computational Optimization and Applications*, **73**, 903-912. <https://doi.org/10.1007/s10589-019-00092-y>
- [26] Ding, K. and Toh, K. (2025) Stochastic Bregman Subgradient Methods for Nonsmooth Nonconvex Optimization Problems. *Journal of Optimization Theory and Applications*, **206**, Article No. 67. <https://doi.org/10.1007/s10957-025-02749-7>

- [27] Takahashi, S., Tanaka, M. and Ikeda, S. (2025) Majorization-Minimization Bregman Proximal Gradient Algorithms for NMF with the Kullback-Leibler Divergence. *Journal of Optimization Theory and Applications*, **208**, Article No. 14. <https://doi.org/10.1007/s10957-025-02833-y>
- [28] Li, X. and Bian, W. (2024) Smoothing Randomized Block-Coordinate Proximal Gradient Algorithms for Nonsmooth Nonconvex Composite Optimization. *Numerical Algorithms*, **100**, 395-424. <https://doi.org/10.1007/s11075-024-01965-y>
- [29] Takahashi, S., Fukuda, M. and Tanaka, M. (2022) New Bregman Proximal Type Algorithms for Solving DC Optimization Problems. *Computational Optimization and Applications*, **83**, 893-931. <https://doi.org/10.1007/s10589-022-00411-w>
- [30] Pu, W., Zhang, J., Zhou, R., Fu, X. and Hong, M. (2024) A Smoothed Bregman Proximal Gradient Algorithm for Decentralized Nonconvex Optimization. *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seoul, 14-19 April 2024, 8911-8915. <https://doi.org/10.1109/icassp48485.2024.10448285>
- [31] Zhang, X., Barrio, R., Martinez, M.A., Jiang, H. and Cheng, L. (2019) Bregman Proximal Gradient Algorithm with Extrapolation for a Class of Nonconvex Nonsmooth Minimization Problems. *IEEE Access*, **7**, 126515-126529. <https://doi.org/10.1109/access.2019.2937005>
- [32] Wu, Z., Li, C., Li, M. and Lim, A. (2021) Inertial Proximal Gradient Methods with Bregman Regularization for a Class of Nonconvex Optimization Problems. *Journal of Global Optimization*, **79**, 617-644. <https://doi.org/10.1007/s10898-020-00943-7>
- [33] 赵静, 郭晨正. 非凸非光滑优化问题的两步惯性 Bregman 邻近交替线性极小化算法[J]. 数学物理学报, 2024, 44(6): 1630-1651.
- [34] 马玉敏, 蔡邢菊, 张海萍, 等. 求解一类非凸非光滑两块优化问题的带不同外推参数的邻近交替线性极小化算法[J]. 计算数学, 2026, 48(1): 102-122.
- [35] Rockafellar, R.T. and Wets, R.J.B. (1998) *Variational Analysis*. Springer.
- [36] Attouch, H., Bolte, J., Redont, P. and Soubeyran, A. (2010) Proximal Alternating Minimization and Projection Methods for Nonconvex Problems: An Approach Based on the Kurdyka-Łojasiewicz Inequality. *Mathematics of Operations Research*, **35**, 438-457. <https://doi.org/10.1287/moor.1100.0449>
- [37] Takahashi, S. and Takeda, A. (2025) Approximate Bregman Proximal Gradient Algorithm for Relatively Smooth Nonconvex Optimization. *Computational Optimization and Applications*, **90**, 227-256. <https://doi.org/10.1007/s10589-024-00618-z>
- [38] Bauschke, H.H., Bolte, J. and Teboulle, M. (2017) A Descent Lemma Beyond Lipschitz Gradient Continuity: First-Order Methods Revisited and Applications. *Mathematics of Operations Research*, **42**, 330-348. <https://doi.org/10.1287/moor.2016.0817>
- [39] Chung, J. and Gazzola, S. (2019) Flexible Krylov Methods for  $\ell_p$  Regularization. *SIAM Journal on Scientific Computing*, **41**, S149-S171. <https://doi.org/10.1137/18m1194456>
- [40] Wen, F., Liu, P., Liu, Y., Qiu, R.C. and Yu, W. (2016) Robust Sparse Recovery for Compressive Sensing in Impulsive Noise Using  $\ell_p$ -Norm Model Fitting. 2016 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, 20-25 March 2016, 4643-4647. <https://doi.org/10.1109/icassp.2016.7472557>