

非光滑非凸 - 强拟凹鞍点问题的Bregman近端梯度算法

张 艳, 李小兵

重庆交通大学数学与统计学院, 重庆

收稿日期: 2024年12月27日; 录用日期: 2025年1月18日; 发布日期: 2025年1月29日

摘 要

针对非光滑非凸 - 强拟凹鞍点问题, 本文利用Bregman距离建立了Bregman近端梯度上升下降算法。对Bregman近端梯度上升迭代算法中, 得到内部最大化问题函数差值不等式, 从而得到近端梯度上升迭代点之间的不等式关系。对于非凸非光滑问题, 引入扰动类梯度下降序列, 得到算法的次收敛性, 当目标函数为半代数时, 得到算法的全局收敛性。

关键词

近端梯度上升下降法, Bregman距离, 非光滑非凸 - 强拟凹鞍点问题

Bregman Proximal Gradient Algorithm for Nonsmooth Nonconvex-Strongly Quasi-Concave Saddle Point Problems

Yan Zhang, Xiaobing Li

School of Mathematics and Statistics, Chongqing Jiaotong University, Chongqing

Received: Dec. 27th, 2024; accepted: Jan. 18th, 2025; published: Jan. 29th, 2025

Abstract

For the nonsmooth nonconvex-strongly quasi-concave saddle point problems, this paper establishes the Bregman proximal gradient ascent-descent algorithm by using the Bregman distance. In the Bregman proximal gradient ascent iterative algorithm, the difference inequality of the internal maximization problem function is obtained, and thus the inequality relationship between the proxi-

mal gradient ascent iterative points is derived. For nonconvex and nonsmooth problems, a perturbed gradient-like descent sequence is introduced to obtain the sub-convergence of the algorithm. When the objective function is semi-algebraic, the global convergence of the algorithm is obtained.

Keywords

Proximal Gradient Ascent-Descent Method, Bregman Distance, Nonsmooth Nonconvex-Strongly Quasi-Concave Saddle Point Problems

Copyright © 2025 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

在现代科学与工程众多领域里, 优化问题无处不在。无论是经济系统中的资源分配, 还是机器学习中的模型训练; 不管是工程控制中的参数调节, 亦或是博弈论中的策略选择, 优化理念贯穿于其中。传统的凸优化问题因其良好性质已得到广泛且深入的研究。但现实世界复杂多样, 很多实际问题无法简单归为凸优化问题。例如, 在机器学习领域, 生成对抗网络(GANs) [1]里生成器和判别器之间的博弈就呈现出非凸非光滑的极大极小问题特征; 在经济学中, 企业间竞争策略的制定也存在类似情况。并且在鲁棒优化[2]、图像处理[3]、信号处理[4]等更多领域, 非凸非光滑的极大极小问题正亟待深入探索与有效解决。近几年对于非凸非光滑的极大极小问题备受学者们的关注[5]-[7]。

本文将考虑下面的鞍点问题:

$$\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} \{F(x, y) = f(x) + c(x, y) - g(y)\} \quad (1)$$

其中 $f: \mathbb{R}^n \rightarrow (-\infty, +\infty)$ 是真下半连续函数, $c: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ 是一个连续可微函数(C^1), $g: \mathbb{R}^m \rightarrow (-\infty, +\infty)$ 是一个真凸下半连续函数, $\mathbb{R}^n, \mathbb{R}^m$ 分别表示为 n 维, m 维实向量空间。

当 $\forall x \in \mathbb{R}^n, c(x, y) - g(y)$ 关于 y 是 σ_1 -强凹函数($\sigma_1 > 0$)时, Cohen 和 Teboulle [6]提出近端梯度上升下降法(PGDA)求解问题(1), 即

$$\begin{aligned} y_{k+1} &= \text{prox}_{\beta g} (y_k + \beta \nabla_y c(x_k, y_k)) \\ x_{k+1} &= \text{prox}_{\alpha f} (x_k - \alpha \nabla_x c(x_k, w_k)) \end{aligned} \quad (2)$$

当 $w_k = y_k$ 时, 为平行近端梯度上升下降法(PPGDA); 当 $w_k = y_{k+1}$ 时, 为交替近端梯度上升下降法(APGDA)。Cohen 和 Teboulle [6]还提出了 Bregman 近端梯度下降上升法(BGDA):

$$\begin{aligned} y_{k+1} &= \arg \max_{y \in \mathbb{R}^m} \left\{ \langle \nabla_y c(x_k, y_k), y \rangle - g(y) - \frac{1}{2\beta} \|y - y_k\|^2 \right\} \\ x_{k+1} &\in \arg \min_{x \in \mathbb{R}^n} \left\{ f(x) + \langle \nabla_x c(x_k, w_k), x \rangle + \frac{1}{\alpha} D_h(x, x_k) \right\} \end{aligned} \quad (3)$$

当 $w_k = y_k$ 时, 是平行 Bregman 近端梯度下降上升法(PBGDA); 当 $w_k = y_{k+1}$ 时, 为交替近端梯度下降上升法(ABGDA)。

本文对函数 $c(x, y) - g(y)$ 的凹性条件做了改变, 将其凹性条件减弱, 变为强拟凹, 即 $\forall x \in \mathbb{R}^n, c(x, y) - g(y)$ 关于 y 是 σ_1 -强拟凹函数($\sigma_1 > 0$)。

在实际的各类现实问题情境中, 非光滑非凸-强拟凸极小极大问题的出现频率相对更高。例如, 经济学领域中的生产与成本优化问题、资源分配与福利经济学问题; 博弈论领域中的非合作博弈的均衡求解; 机器学习领域中的深度学习模型的训练优化、强化学习中的策略优化等等。

受算法(3)的启发, 对于 y -步迭代过程, 我们将 y -步中的 $\|y - y_k\|^2$ 部分同样用 Bregman 距离代替, 即

$$\begin{aligned} y_{k+1} &= \arg \max_{y \in \mathbb{R}^m} \left\{ \langle \nabla_y c(x_k, y_k), y \rangle - g(y) - \frac{1}{2\beta} D_{h_1}(y, y_k) \right\} \\ x_{k+1} &\in \arg \min_{x \in \mathbb{R}^n} \left\{ f(x) + \langle \nabla_x c(x_k, w_k), x \rangle + \frac{1}{\alpha} D_{h_2}(x, x_k) \right\} \end{aligned} \quad (4)$$

在我们的算法中, y -步迭代的正则部分($\|y - y_k\|^2$)也为 Bregman 距离, 当 Bregman 距离为欧式空间的距离时, 即为算法(3)。算法(3)是算法(4)的特殊情况, 故算法(4)更有意义。对于算法的收敛性分析本文借鉴 Cohen 和 Teboulle [6]的方法引入扰动类梯度下降序列, 得到算法的次收敛性, 对于目标函数是半代数函数时, 得到算法的全局的收敛性。

2. 将鞍点问题复合最小化问题

先给出关于问题(1)中的假设和 Moreau 近端映射的定义。

假设 1 (a) $\inf_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} F(x, y) > -\infty$ 。

(b) $c: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ 是一个连续可微函数(C^1), 并且 $c(x, y)$ 关于 y 是凹函数。另外, $\exists L_{xx}, L_{xy}, L_{yy}, L_{yx} > 0$, 使得 $\forall \bar{x}, x \in \mathbb{R}^n, \forall \bar{y}, y \in \mathbb{R}^m$, 有

$$\begin{aligned} \|\nabla_x c(x, y) - \nabla_x c(x, \bar{y})\| &\leq L_{xy} \|y - \bar{y}\| \\ \|\nabla_x c(x, y) - \nabla_x c(\bar{x}, y)\| &\leq L_{xx} \|x - \bar{x}\| \\ \|\nabla_y c(x, y) - \nabla_y c(\bar{x}, y)\| &\leq L_{yx} \|x - \bar{x}\| \\ \|\nabla_y c(x, y) - \nabla_y c(x, \bar{y})\| &\leq L_{yy} \|y - \bar{y}\| \end{aligned}$$

(c) $\forall x \in \mathbb{R}^n, c(x, y) - g(y)$ 关于 y 是 σ_1 -强拟凹连续可微函数, 且关于 y 的梯度是 L_1 -Lipschitz 连续的, 即存在一个常数 $L_1 > 0$, 使得对于 $\forall y, \bar{y} \in \mathbb{R}^m$, $\|c(x, y) - g(y) - (c(x, \bar{y}) - g(\bar{y}))\| \leq L_1 \|y - \bar{y}\|$, ($\sigma_1 \in (L_1, 2L_1]$)。

定义 1 [8] 设函数 $\varphi: \mathbb{R}^d \rightarrow (-\infty, \infty]$ 是真下半连续函数, 对 $\forall t > 0$, 则关于函数 φ 的 Moreau 近端映射为

$$\text{prox}_{t\varphi}(z) := \arg \min_{x \in \mathbb{R}^d} \left\{ \varphi(x) - \frac{1}{2t} \|x - z\|^2 \right\}.$$

由于 $\forall x \in \mathbb{R}^n, c(x, y) - g(y)$ 关于 y 是 σ_1 ($\sigma_1 > 0$)-强拟凹函数, 则 $g(y) - c(x, y)$ 关于 y 是 σ_1 -强拟凸函数。根据文献[9]中的定理 1 可以得到 $y^*(x) := \arg \max_{y \in \mathbb{R}^m} \{c(x, y) - g(y)\} = \arg \min_{y \in \mathbb{R}^m} \{g(y) - c(x, y)\}$ 是单值的,

即 $y^*(x): \mathbb{R}^n \rightarrow \mathbb{R}^m$ 是有意义的。现在定义 $\phi: \mathbb{R}^n \rightarrow \mathbb{R}$

$$\phi(x) := \max \{c(x, y) - g(y)\} = c(x, y^*(x)) - g(y^*(x)).$$

则问题(1)变为

$$\min_{x \in \mathbb{R}^n} \{\Theta(x) := f(x) + \phi(x)\}. \quad (5)$$

我们继续分析函数 y^* 与 ϕ 的性质。下面的引理是关于强拟凸函数的一个性质, 将会用来证明函数 y^* 的连续性。

引理 1 设 $\psi: \mathbb{R}^d \rightarrow \mathbb{R}$ 是 σ -强拟凸函数($\sigma > 0$), 并且是连续可微的。若 $\forall x_1, x_2 \in \mathbb{R}^d$, 且 $\psi(x_1) \leq \psi(x_2)$, 则有

$$\langle \nabla \psi(x_2), x_2 - x_1 \rangle \geq \frac{\sigma}{2} \|x_2 - x_1\|.$$

证明: 由于 ψ 是 C 上的 σ -强拟凸函数, 根据强拟凸函数的定义有:

$$\psi(\lambda x_1 + (1-\lambda)x_2) \leq \max\{\psi(x_1), \psi(x_2)\} - \lambda(1-\lambda)\frac{\sigma}{2}\|x_1 - x_2\|^2, \forall \lambda \in (0, 1).$$

因为 $\forall x_1, x_2 \in C$, 且 $\psi(x_1) \leq \psi(x_2)$, 则有

$$\psi(\lambda x_1 + (1-\lambda)x_2) \leq \psi(x_2) - \lambda(1-\lambda)\frac{\sigma}{2}\|x_1 - x_2\|^2,$$

$$\psi(\lambda x_1 + (1-\lambda)x_2) - \psi(x_2) \leq -\lambda(1-\lambda)\frac{\sigma}{2}\|x_1 - x_2\|^2,$$

$$\frac{\psi(\lambda x_1 + (1-\lambda)x_2) - \psi(x_2)}{\lambda} \leq -(1-\lambda)\frac{\sigma}{2}\|x_1 - x_2\|^2,$$

$$\frac{\psi(x_2 + \lambda(x_1 - x_2)) - \psi(x_2)}{\lambda} \leq -(1-\lambda)\frac{\sigma}{2}\|x_1 - x_2\|^2.$$

由于 ψ 是连续可微函数, 让上面不等式的两边的 $\lambda \rightarrow 0$, 则有

$$\langle \nabla \psi(x_2), x_1 - x_2 \rangle \leq -\frac{\sigma}{2}\|x_1 - x_2\|$$

即

$$\frac{\sigma}{2}\|x_2 - x_1\| \leq \langle \nabla \psi(x_2), x_2 - x_1 \rangle.$$

引理 2 映射 $y^*: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 是 $\frac{2L_{yx}}{\sigma_1}$ -Lipschitz 连续的, 即

$$\|y^*(x) - y^*(\bar{x})\| \leq \frac{2L_{yx}}{\sigma_1} \|x - \bar{x}\|, \forall x, \bar{x} \in \mathbb{R}^n.$$

证明: 因为 $\forall x \in \mathbb{R}^n$, $y^*(x) := \arg \max_{y \in \mathbb{R}^m} \{c(x, y) - g(y)\}$, 则

$$\forall \bar{x} \in \mathbb{R}^n, c(x, y^*(x)) - g(y^*(x)) \geq c(x, y^*(\bar{x})) - g(y^*(\bar{x}))$$

即 $\forall \bar{x} \in \mathbb{R}^n, g(y^*(x)) - c(x, y^*(x)) \leq g(y^*(\bar{x})) - c(x, y^*(\bar{x}))$ 。由于 $g(y) - c(x, y)$ 关于 y 是 σ_1 -强拟凸的, 根据引理 1 有

$$\frac{\sigma_1}{2} \|y^*(\bar{x}) - y^*(x)\|^2 \leq \langle \partial g(y^*(\bar{x})) - \nabla_y c(x, y^*(\bar{x})), y^*(\bar{x}) - y^*(x) \rangle. \quad (6)$$

由于 $y^*(x) := \arg \max_{y \in \mathbb{R}^m} \{c(x, y) - g(y)\}$, 根据一阶最优性条件有

$$0 \in \nabla_y c(x, y^*(x)) - \partial g(y^*(x)) \Rightarrow \nabla_y c(x, y^*(x)) \in \partial g(y^*(x)).$$

根据上式可以得到 $\nabla_y c(\bar{x}, y^*(\bar{x})) \in \partial g(y^*(\bar{x}))$, 再结合(6)式有

$$\begin{aligned} \frac{\sigma_1}{2} \|y^*(\bar{x}) - y^*(x)\|^2 &\leq \langle \nabla_y c(\bar{x}, y^*(\bar{x})) - \nabla_y c(x, y^*(\bar{x})), y^*(\bar{x}) - y^*(x) \rangle \\ &\leq \|\nabla_y c(\bar{x}, y^*(\bar{x})) - \nabla_y c(x, y^*(\bar{x}))\| \|y^*(\bar{x}) - y^*(x)\| \\ &\leq L_{yx} \|\bar{x} - x\| \|y^*(\bar{x}) - y^*(x)\| \\ \|y^*(\bar{x}) - y^*(x)\| &\leq \frac{2L_{yx}}{\sigma_1} \|\bar{x} - x\|, \forall \bar{x}, x \in \mathbb{R}^n. \end{aligned}$$

我们可以得到文献[6]命题 1 相同的结论: $\phi: \mathbb{R}^n \rightarrow \mathbb{R}$ 是连续函数(C^1)函数, 且梯度

$$\nabla_\phi(x) = \nabla_x c(x, y^*(x)) \text{ 是 } L_\phi = L_{xx} + \frac{2L_{xy}L_{yx}}{\sigma_1} \text{-Lipschitz 连续的, 其中 } y^*(x) = \arg \min_{y \in \mathbb{R}^m} \{g(y) - c(x, y)\}.$$

根据上面得到的结论, 复合最小化问题(5)可以用 Bregman 近端梯度法求解, 即

$$x_{k+1} = \arg \min_{x \in \mathbb{R}^n} \left\{ f(x) + \langle \nabla_x c(x_k, y^*(x_k)), x \rangle + \frac{1}{\alpha} D_h(x, x_k) \right\}.$$

但是该算法困难在于寻找 $y^*(x_k)$, 并且在迭代时, 必须要使用内循环计算, 从而导致算法变得非常复杂, 故在这种情况下将我们提出算法(4)看成近似 Bregman 近端梯度法, 不用求解

$y^*(x) = \arg \max_{y \in \mathbb{R}^m} \{c(x, y) - g(y)\}$, 直接用算法(4)中的 y -步中的 y_k 或者 y_{k+1} 代替 $y^*(x_k)$, 而 x -步就是上述

算法。其中, 算法(4)中的 y -步的步长设为 $\beta = \frac{1}{L_{yy}}$, 对于 x -步的步长 α 的选择将在下面的定理 1 中体现。

3. 收敛性分析

将鞍点问题(1)表述为复合最小化问题(5), 分析 Bregman 近端梯度上升下降法(4)的收敛性, 设 $\{x_k, y_k\}_{k \in \mathbb{N}}$ 是由算法(4)产生的序列, 本节我们得到 $\{x_k\}_{k \in \mathbb{N}}$ 收敛到函数 Θ 的临界点: 即 $\{x_k\}_{k \in \mathbb{N}} \rightarrow \bar{x} \in \text{crit } \Theta := \{x: 0 \in \partial \Theta(x)\}$ 。另外得到 $\{y_k\}_{k \in \mathbb{N}} \rightarrow y^*(\bar{x})$, 其中 $y^*(x)$ 是鞍点问题(1)的内部极大化问题的解。由于复合最小化问题(5)的目标函数是非凸非光滑的, 缺少下降性质, 借鉴文献[6]的方法引入扰动类梯度下降序列, 得到算法的次收敛性。

定义 2 [6] 设 $\Theta: \mathbb{R}^n \rightarrow (-\infty, \infty]$ 是真下半连续函数, 若序列 $\{(x_k, v_k)\}_{k \geq 0} \subseteq \text{dom } \Theta \times \mathbb{R}_+$ 满足下面的三个条件:

(a) 扰动充分下降性质: 存在一个常数 $c_1 > 0$ 使得对于任意的 $k \in \mathbb{N}$ 有

$$c_1 \left(\|x_{k+1} - x_k\|^2 + v_k^2 \right) \leq \left(\Theta(x_k) + \frac{1}{2} v_k^2 \right) - \left(\Theta(x_{k+1}) + \frac{1}{2} v_{k+1}^2 \right).$$

(b) 迭代间隙上的扰动次梯度下界: 存在一个常数 $c_2 > 0$ 使得对于任意的 $k \in \mathbb{N}$, 这里存在

$\xi_{k+1} \in \partial \Theta(x_{k+1})$, 且 ξ_{k+1} 满足 $\|\xi_{k+1}\| \leq c_2 (\|x_{k+1} - x_k\| - v_k)$ 。

(c) 设 $\{x_k\}_{k \in \mathbb{N}}$ 是一个子序列且收敛到 $\bar{x}$, 则 $\limsup_{k \in \mathbb{N}} \Theta(x_k) \leq \Theta(\bar{x})$ 。

则 $\{(x_k, v_k)\}_{k \geq 0}$ 是关于 Θ 的扰动类梯度下降序列。

下面给出在分析算法收敛性时的一些假设。

假设 2 (a) Bregman 函数 $h_1: \mathbb{R}^m \rightarrow \mathbb{R}$ 是 σ_{h_1} -强凸函数并且是连续可微函数, 即 h_1 是 L_{h_1} -光滑函数 ($\sigma_{h_1} = L_{h_1}$):

(b) Bregman 函数 $h_2: \mathbb{R}^n \rightarrow \mathbb{R}$ 是 1-强凸函数, 并且是连续可微函数, 即 h_2 是 L_{h_2} -光滑函数;

(c) $\forall x \in \mathbb{R}^n, (-c(x, y), h_1(y))$ 关于 y 是 $\frac{L_{yy}}{2}$ -光滑自适应的, 即 $\forall y, \bar{y} \in \mathbb{R}^m$, 存在 $\frac{L_{yy}}{2} > 0$, 有

$$-c(x, y) \leq -c(x, \bar{y}) + \langle -\nabla_y c(x, \bar{y}), y - \bar{y} \rangle + \frac{L_{yy}}{2} D_{h_1}(y, \bar{y});$$

(d) 算法(4)的 x -步迭代过程是有意义的, 即

$$x_{k+1} \in \arg \min_{x \in \mathbb{R}^n} \left\{ f(x) + \langle \nabla_x c(x_k, w_k), x \rangle + \frac{1}{\alpha} D_{h_2}(x, x_k) \right\} \neq \emptyset.$$

在分析收敛性之前我们先给出一个与文献[10]中的引理 2.6 类似的结论, 即内部最大化问题函数差值不等式。

引理 3 设 $y \in \text{dom}(g)$, $(-c(x_k, y), h_1(y))$ 关于 y 是 $\frac{L_{yy}}{2}$ -光滑自适应的, 则

$$\Gamma_k(y_{k+1}) - \Gamma_k(y) \leq \frac{L_{yy}}{2} (D_{h_1}(y, y_k) - D_{h_1}(y, y_{k+1})). \quad (7)$$

证明: 近端梯度上升下降算法中的 y -步迭代过程

$$y_{k+1} = \text{prox}_{\frac{1}{L_{yy}}g} \left(y_k + \frac{1}{L_{yy}} \nabla_y c(x_k, y_k) \right). \quad (8)$$

是解决问题(1)的内部最大问题的步骤, 该最大化问题可以表述为极小化问题

$$\min_{y \in \mathbb{R}^m} \{ \Gamma_k(y) := g(y) - c(x_k, y) \}.$$

根据 Moreau 近端映射的定义, (8)式变为

$$\begin{aligned} y_{k+1} &= \arg \min_{y \in \mathbb{R}^m} \left\{ g(y) + \frac{L_{yy}}{2} \left\| y - \left(y_k + \frac{1}{L_{yy}} \nabla_y c(x_k, y_k) \right) \right\|^2 \right\} \\ &= \arg \min_{y \in \mathbb{R}^m} \left\{ g(y) - \langle \nabla_y c(x_k, y_k), y - y_k \rangle + \frac{L_{yy}}{2} \|y - y_k\|^2 + \frac{1}{2L_{yy}} (\nabla_y c(x_k, y_k))^2 \right\} \end{aligned}$$

由于 $\frac{1}{2L_{yy}} (\nabla_y c(x_k, y_k))^2$ 关于 y 是常数, 则有

$$y_{k+1} = \arg \min_{y \in \mathbb{R}^m} \left\{ g(y) - \langle \nabla_y c(x_k, y_k), y - y_k \rangle + \frac{L_{yy}}{2} \|y - y_k\|^2 \right\}.$$

将上面等式中的 $\|y - y_k\|^2$ 变为 Bregman 距离就是本文提出的算法的 y -步迭代过程, 即

$$y_{k+1} = \arg \min_{y \in \mathbb{R}^m} \left\{ g(y) - \langle \nabla_y c(x_k, y_k), y - y_k \rangle + \frac{L_{yy}}{2} D_{h_1}(y, y_k) \right\}.$$

由于 $c(x_k, y_k)$ 关于 y 是常数, 则有

$$y_{k+1} = \arg \min_{y \in \mathbb{R}^m} \left\{ g(y) - c(x_k, y_k) - \langle \nabla_y c(x_k, y_k), y - y_k \rangle + \frac{L_{yy}}{2} D_{h_1}(y, y_k) \right\}. \quad (9)$$

设

$$Q(y, y_k) = g(y) - c(x_k, y_k) - \langle \nabla_y c(x_k, y_k), y - y_k \rangle + \frac{L_{yy}}{2} D_{h_1}(y, y_k),$$

$$l_c(y, y_k) = -c(x_k, y) + c(x_k, y_k) + \langle \nabla_y c(x_k, y_k), y - y_k \rangle.$$

对于(9)式, 根据一阶最优性条件有

$$0 \in \partial g(y_{k+1}) - \nabla_y c(x_k, y_k) + \frac{L_{yy}}{2} (\nabla h_1(y_{k+1}) - \nabla h_1(y_k)) + N_{\text{dom}(g)}(y_{k+1}).$$

设 $\gamma \in \partial g(y_{k+1})$, 则有

$$-\left(\gamma - \nabla_y c(x_k, y_k) + \frac{L_{yy}}{2} (\nabla h_1(y_{k+1}) - \nabla h_1(y_k)) \right) \in N_{\text{dom}(g)}(y_{k+1}).$$

$\forall y \in \text{dom}(g)$, 根据法锥的定义有

$$\begin{aligned} \left\langle y_{k+1} - y, -\left(\gamma - \nabla_y c(x_k, y_k) + \frac{L_{yy}}{2} (\nabla h_1(y_{k+1}) - \nabla h_1(y_k)) \right) \right\rangle &\geq 0, \\ \left\langle \gamma - \nabla_y c(x_k, y_k) + \frac{L_{yy}}{2} (\nabla h_1(y_{k+1}) - \nabla h_1(y_k)), y - y_{k+1} \right\rangle &\geq 0, \\ \langle \gamma, y - y_{k+1} \rangle &\geq \left\langle \nabla_y c(x_k, y_k) - \frac{L_{yy}}{2} (\nabla h_1(y_{k+1}) - \nabla h_1(y_k)), y - y_{k+1} \right\rangle. \end{aligned} \quad (10)$$

因为 g 是凸函数, 且 $\gamma \in \partial g(y_{k+1})$, 则有

$$\langle g(y) - g(y_{k+1}) \rangle \geq \langle \gamma, y - y_{k+1} \rangle, \quad (11)$$

结合(10)和(11)两个不等式则有

$$\langle g(y) - g(y_{k+1}) \rangle \geq \left\langle \nabla_y c(x_k, y_k) - \frac{L_{yy}}{2} (\nabla h_1(y_{k+1}) - \nabla h_1(y_k)), y - y_{k+1} \right\rangle.$$

$$\begin{aligned} \Gamma_k(y_{k+1}) - Q(y_{k+1}, y_k) &= g(y_{k+1}) - c(x_k, y_{k+1}) - g(y_{k+1}) + c(x_k, y_k) + \langle \nabla_y c(x_k, y_k), y_{k+1} - y_k \rangle - \frac{L_{yy}}{2} D_{h_1}(y_{k+1}, y_k) \\ &= -c(x_k, y_{k+1}) + c(x_k, y_k) + \langle \nabla_y c(x_k, y_k), y_{k+1} - y_k \rangle - \frac{L_{yy}}{2} D_{h_1}(y_{k+1}, y_k) \end{aligned} \quad (12)$$

由于 $(-c(x_k, y), h_1(y))$ 关于 y 是 $\frac{L_{yy}}{2}$ -光滑自适应的, 则有

$$\begin{aligned} -c(x_k, y_{k+1}) - (-c(x_k, y_k)) - \langle -\nabla_y c(x_k, y_k), y_{k+1} - y_k \rangle &\leq \frac{L_{yy}}{2} D_{h_1}(y_{k+1}, y_k), \\ -c(x_k, y_{k+1}) + c(x_k, y_k) + \langle \nabla_y c(x_k, y_k), y_{k+1} - y_k \rangle &\leq \frac{L_{yy}}{2} D_{h_1}(y_{k+1}, y_k), \end{aligned} \quad (13)$$

结合(12)和(13)不等式就有 $\Gamma_k(y_{k+1}) - Q(y_{k+1}, y_k) \leq 0$, 那么

$$\begin{aligned}
\Gamma_k(y) - \Gamma_k(y_{k+1}) &\geq \Gamma_k(y) - Q(y_{k+1}, y_k) \\
&= g(y) - c(x_k, y) - g(y_{k+1}) + c(x_k, y_k) + \langle \nabla_y c(x_k, y_k), y_{k+1} - y_k \rangle - \frac{L_{yy}}{2} D_{h_1}(y_{k+1}, y_k) \\
&= -c(x_k, y) + c(x_k, y_k) + \langle \nabla_y c(x_k, y_k), y - y_k \rangle - \langle \nabla_y c(x_k, y_k), y - y_k \rangle \\
&\quad + \langle \nabla_y c(x_k, y_k), y_{k+1} - y_k \rangle - \frac{L_{yy}}{2} D_{h_1}(y_{k+1}, y_k) + g(y) - g(y_{k+1}) \\
&= l_c(y, y_k) - \langle \nabla_y c(x_k, y_k), y - y_{k+1} \rangle - \frac{L_{yy}}{2} D_{h_1}(y_{k+1}, y_k) + g(y) - g(y_{k+1}) \\
\Gamma_k(y) - \Gamma_k(y_{k+1}) &\geq l_c(y, y_k) - \langle \nabla_y c(x_k, y_k), y - y_{k+1} \rangle - \frac{L_{yy}}{2} D_{h_1}(y_{k+1}, y_k) \\
&\quad + \left\langle \nabla_y c(x_k, y_k) - \frac{L_{yy}}{2} (\nabla h_1(y_{k+1}) - \nabla h_1(y_k)), y - y_{k+1} \right\rangle \\
&= l_c(y, y_k) - \frac{L_{yy}}{2} D_{h_1}(y_{k+1}, y_k) + \frac{L_{yy}}{2} \langle \nabla h_1(y_k) - \nabla h_1(y_{k+1}), y - y_{k+1} \rangle
\end{aligned}$$

再利用 Bregman 距离三点恒等式则有

$$\begin{aligned}
\Gamma_k(y) - \Gamma_k(y_{k+1}) &\geq l_c(y, y_k) - \frac{L_{yy}}{2} D_{h_1}(y_{k+1}, y_k) + \frac{L_{yy}}{2} (D_{h_1}(y, y_{k+1}) + D_{h_1}(y_{k+1}, y_k) - D_{h_1}(y, y_k)) \\
&= l_c(y, y_k) + \frac{L_{yy}}{2} (D_{h_1}(y, y_{k+1}) - D_{h_1}(y, y_k)) \\
\Gamma_k(y_{k+1}) - \Gamma_k(y) &\leq -l_c(y, y_k) + \frac{L_{yy}}{2} (D_{h_1}(y, y_k) - D_{h_1}(y, y_{k+1})),
\end{aligned}$$

因为 $c(x, y)$ 关于 y 是凹函数, 则

$$\begin{aligned}
-l_c(y, y_k) &= c(x_k, y) - c(x_k, y_k) - \langle \nabla_y c(x_k, y_k), y - y_k \rangle \leq 0, \\
\Gamma_k(y_{k+1}) - \Gamma_k(y) &\leq \frac{L_{yy}}{2} (D_{h_1}(y, y_k) - D_{h_1}(y, y_{k+1})).
\end{aligned}$$

通过上面引理 3 的结论, 得到近端梯度上升迭代点之间的不等式关系, 即下面的引理 4。

引理 4 (y_k 与 y_{k+1} 之间的关系) 假设 Bregman 函数 $h_1: \mathbb{R}^m \rightarrow \mathbb{R}$ 满足假设 2(a), 令 $\kappa = \frac{L_{yy} L_{h_1}}{2(\sigma_1 - L_1)}$, 则对任意的 $k \in \mathbb{N}$ 有

$$\|y_{k+1} - y^*(x_{k+1})\| \leq \sqrt{\frac{\kappa}{\kappa+1}} \|y^*(x_k) - y_k\| + \frac{2L_{yx}}{\sigma_1} \|x_k - x_{k+1}\| \quad (14)$$

$$\|y_{k+1} - y^*(x_k)\| \leq \sqrt{\frac{\kappa}{\kappa+1}} \left(\frac{2L_{yx}}{\sigma_1} \|x_k - x_{k-1}\| + \|y_k - y^*(x_{k-1})\| \right) \quad (15)$$

$$\|y_{k+1} - y^*(x_{k+1})\|^2 \leq \frac{\kappa(1+1/2\kappa)}{\kappa+1} \|y^*(x_k) - y_k\|^2 + \frac{(1+2\kappa)4L_{yx}^2}{\sigma_1^2} \|x_k - x_{k+1}\|^2 \quad (16)$$

$$\|y_{k+1} - y^*(x_k)\|^2 \leq \frac{\kappa(1+1/2\kappa)}{\kappa+1} \left(\|y^*(x_{k-1}) - y_k\|^2 + \frac{4\kappa L_{yx}^2}{\sigma_1^2} \|x_{k-1} - x_k\|^2 \right) \quad (17)$$

证明: $y_{k+1} = \text{prox}_{\frac{1}{L_{yy}}g} \left(y_k + \frac{1}{L_{yy}} \nabla_y c(x_k, y_k) \right)$ 是解决问题(1)内部最大化的近端梯度上升步骤, 将其表述为下列最小化问题:

$\min_{y \in \mathbb{R}^m} \{ \Gamma_k(y) := g(y) - c(x_k, y) \}$, 则根据假设 1(d)可知 $\Gamma_k(y)$ 是 σ_1 -强拟凸函数, 其唯一极小元为 $y^*(x_k) := \arg \min \Gamma_k(y)$, 根据引理 1 有:

$$\frac{\sigma_1}{2} \|y_{k+1} - y^*(x_k)\|^2 \leq \langle \nabla \Gamma_k(y_{k+1}), y_{k+1} - y^*(x_k) \rangle. \quad (18)$$

另外, 根据假设 1(d)和下降引理[11]有

$$\Gamma_k(y^*(x_k)) - \Gamma_k(y_{k+1}) \leq \langle \nabla \Gamma_k(y_{k+1}), y^*(x_k) - y_{k+1} \rangle + \frac{L_1}{2} \|y_{k+1} - y^*(x_k)\|^2. \quad (19)$$

结合(18)和(19)两个不等式, 则有

$$\begin{aligned} \frac{\sigma_1}{2} \|y_{k+1} - y^*(x_k)\|^2 &\leq \Gamma_k(y_{k+1}) - \Gamma_k(y^*(x_k)) + \frac{L_1}{2} \|y_{k+1} - y^*(x_k)\|^2, \\ \left(\frac{\sigma_1}{2} - \frac{L_1}{2} \right) \|y_{k+1} - y^*(x_k)\|^2 &\leq \Gamma_k(y_{k+1}) - \Gamma_k(y^*(x_k)). \end{aligned}$$

因为 Bregman 函数 h_1 是 σ_{h_1} -强凸函数并且是连续可微函数 L_{h_1} -光滑的, 则会有以下结论:

$$D_{h_1}(y^*(x_k), y_k) \leq \frac{L_{h_1}}{2} \|y^*(x_k) - y_k\|^2 \text{ 和 } D_{h_1}(y^*(x_k), y_{k+1}) \geq \frac{\sigma_{h_1}}{2} \|y^*(x_k) - y_{k+1}\|^2$$

则引理 3 中的不等式(7)放缩为

$$\begin{aligned} \Gamma_k(y_{k+1}) - \Gamma_k(y^*(x_k)) &\leq \frac{L_{yy}}{2} \left(\frac{L_{h_1}}{2} \|y^*(x_k) - y_k\|^2 - \frac{\sigma_{h_1}}{2} \|y^*(x_k) - y_{k+1}\|^2 \right) \\ \left(\frac{\sigma_1}{2} - \frac{L_1}{2} \right) \|y_{k+1} - y^*(x_k)\|^2 &\leq \Gamma_k(y_{k+1}) - \Gamma_k(y^*(x_k)) \leq \frac{L_{yy}}{2} \left(\frac{L_{h_1}}{2} \|y^*(x_k) - y_k\|^2 - \frac{\sigma_{h_1}}{2} \|y^*(x_k) - y_{k+1}\|^2 \right) \\ (2(\sigma_1 - L_1) + \sigma_{h_1} L_{yy}) \|y_{k+1} - y^*(x_k)\|^2 &\leq L_{yy} L_{h_1} \|y^*(x_k) - y_k\|^2 \|y_{k+1} - y^*(x_k)\| \leq \sqrt{\frac{L_{yy} L_{h_1}}{2(\sigma_1 - L_1) + \sigma_{h_1} L_{yy}}} \|y^*(x_k) - y_k\| \end{aligned}$$

由于 $\sigma_{h_1} = L_{h_1}$, 令 $\kappa = \frac{L_{yy} L_{h_1}}{2(\sigma_1 - L_1)}$, 则有

$$\begin{aligned} \|y_{k+1} - y^*(x_k)\| &\leq \sqrt{\frac{\kappa}{\kappa+1}} \|y^*(x_k) - y_k\| \\ \|y_{k+1} - y^*(x_{k+1})\| &= \|y_{k+1} - y^*(x_k) + y^*(x_k) - y^*(x_{k+1})\| \\ &\leq \|y_{k+1} - y^*(x_k)\| + \|y^*(x_k) - y^*(x_{k+1})\| \\ &\leq \sqrt{\frac{\kappa}{\kappa+1}} \|y^*(x_k) - y_k\| + \frac{2L_{yx}}{\sigma_1} \|x_k - x_{k+1}\| \end{aligned}$$

$$\begin{aligned}
\|y_{k+1} - y^*(x_k)\| &\leq \sqrt{\frac{\kappa}{\kappa+1}} \|y^*(x_k) - y_k\| = \sqrt{\frac{\kappa}{\kappa+1}} \|y^*(x_k) - y^*(x_{k-1}) + y^*(x_{k-1}) - y_k\| \\
&\leq \sqrt{\frac{\kappa}{\kappa+1}} (\|y^*(x_k) - y^*(x_{k-1})\| + \|y^*(x_{k-1}) - y_k\|) \\
&\leq \sqrt{\frac{\kappa}{\kappa+1}} \left(\frac{2L_{yx}}{\sigma_1} \|x_k - x_{k-1}\| + \|y^*(x_{k-1}) - y_k\| \right)
\end{aligned}$$

故(14)和(15)得证。在不等式 $(a+b)^2 \leq (1+\mu)a^2 + (1+\mu^{-1})b^2, (\mu > 0)$ 中, 令 $\mu = \frac{1}{2\kappa}$ 将其应用到(14)和(15)两个不等式, 就会得到(16)和(17)两个不等式。

这是对于算法的 y -步的分析, 我们选取满足假设 2(a)的 Bregman 函数 h_1 得到文献[6]中引理 6 同样的结论。关于 x 步的次梯度有界性和利用 x 步得到函数值差距的结论也会得到与文献[6]类似的结论。

引理 5 [6] (给出函数 ϕ 的光滑自适应性质) 设 $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$, 且

$\forall x \in \mathbb{R}^n, \varphi(x) := M_{xx} h_2(x) + (2L_{xy}L_{yx}/\sigma_1) \|x\|^2$, 则 (ϕ, φ) 是 1-光滑自适应的, 即

$$\phi(\bar{x}) - \phi(x) - \langle \nabla \phi(x), \bar{x} - x \rangle \leq D_\varphi(\bar{x}, x).$$

引理 6 [6] (x 步函数值差值) 设 $(x_k, w_k)_{k \in \mathbb{N}}$ 是由近端梯度算法产生的序列, 则 $\forall k \geq 0$, 有

$$\Theta(x_{k+1}) - \Theta(x_k) \leq \frac{1}{2} \left(L_{xy}^2 - M_\phi - \frac{1}{\alpha} \right) \|x_{k+1} - x_k\|^2 + \frac{1}{2} \|w_k - y^*(x_k)\|^2,$$

其中 $M_\phi = M_{xx} + \frac{4L_{xy}L_{yx}}{\sigma_1}$ 。

引理 7 [6] (x 步的次梯度有界性) 设 $(x_k, w_k)_{k \in \mathbb{N}}$ 是由近端梯度算法产生的序列, 则存在 M 大于 0 使得 $\forall k \geq 0, \exists \xi_{k+1} \in \partial \Theta(x_{k+1})$, 且满足不等式 $\|\xi_{k+1}\| \leq M (\|x_{k+1} - x_k\| + \|w_k - y^*(x_k)\|)$, 其中

$$M = \max \left\{ \frac{L_{h_1}}{\alpha} + L_{xx} + \frac{2L_{xy}L_{yx}}{\sigma_1}, L_{xy} \right\}.$$

令扰动序列 v_k : $v_k = \sqrt{s} \|y_k - y^*(x_k)\|$ ($w_k = y_k$ 时), $v_k = \sqrt{t \|x_k - x_{k-1}\|^2 + s \|y_k - y^*(x_{k-1})\|^2}$ ($w_k = y_{k+1}$ 时), 其中 s, t 为大于 0 的实数。引入的扰动序列 v_k 满足扰动类梯度下降序列的 3 个条件, 其证明方法与文献[6]相同。故可以得到算法的次收敛性。若问题(1)的目标函数是半代数函数, 则得到算法的全局收敛性。下面先给出定理 1, 定理 1 给出算法(4) x -步的步长 α 的选择。并且在恰当的步长选择下, 得到算法的收敛性, 即定理 2。

定理 1 设由算法(4)产生的序列 $\{x_k, y_k\}_{k \geq 0}$ 有界, 并且设 $\alpha < \frac{1}{L}$, 其中 α 为算法(4)中 x -步的算法步长。

$L := L_{xy}^2 + M_\phi + \frac{2(2\kappa^2 + 3\kappa + 1)L_{yx}^2}{\sigma_1^2}$ ($w_k = y_{k+1}$ 时), $L := L_{xy}^2 + M_\phi + \frac{2(2\kappa^2 + \kappa)L_{yx}^2}{\sigma_1^2}$ ($w_k = y_k$ 时), 则 $\forall \varepsilon > 0$, $\exists k = O(\varepsilon^{-2})$ 使得

$$\|x_{k+1} - x_k\| \leq \varepsilon, \text{ dist}(0, \partial \Theta(x_k)) \leq \varepsilon, \|y_k - y^*(x_k)\| \leq \varepsilon.$$

定理 2 设由算法(4)产生的序列 $\{x_k, y_k\}_{k \geq 0}$ 有界, 假设 $\alpha < \frac{1}{L}$, 则下面两个结论成立:

(1) 设 Ω 是序列 $\{x_k\}_{k \geq 0}$ 的聚点集合, 则 Ω 是一个非空紧集; 且 $\Omega \subseteq \text{crit } \Theta$; $\lim_{x \rightarrow \infty} \text{dist}(x_k, \Omega) = 0$; Θ 在 Ω 上是有限的常数; 设 $\{x_k\}_{k \in \mathbb{N}}$ 是 $\{x_k\}_{k \geq 0}$ 的子序列, 并且 $\{x_k\}_{k \in \mathbb{N}}$ 收敛到 $\bar{x} \in \Omega$, 则 $\{y_k\}_{k \geq 0}$ 的子序列

$\{y_k\}_{k \in \mathbb{K} \subset \mathbb{N}}$ 收敛到 $y^*(\bar{x})$ 。

(2) 另外再假设函数 f, g, c 是半代数函数, 则有 $\sum_{k=1}^{\infty} \|x_{k+1} - x_k\| < \infty$, 并且 $\{x_k\}_{k \geq 0}$ 收敛到点 $\bar{x} \in \text{crit} \Theta$; $\{y_k\}_{k \geq 0}$ 收敛到 $y^*(\bar{x})$ 。

定理 1, 定理 2 的证明过程参考文献[6]的引理 4 和定理 1。

基金项目

这项研究部分由重庆市研究生联合培养基地建设项目(JDLHYJD2021016)、重庆市高校科技创新团队建设项目(CXQT21021)资助。

参考文献

- [1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., *et al.* (2014) Generative Adversarial Nets. *Communications of the ACM*, **63**, 139-144.
- [2] Nemirovskiĭ, A.S., El Ghaoui, L. and Ben-Tal, A. (2009) Robust Optimization (Princeton Series in Applied Mathematics). Princeton University Press.
- [3] Phillips, D. (1994) Image Processing in C: Analyzing and Enhancing Digital Images. R & D Publications.
- [4] Orfanidis, S.J. (1995) Introduction to Signal Processing. Prentice-Hall, Inc.
- [5] Jiang, J. and Chen, X. (2023) Optimality Conditions for Nonsmooth Nonconvex-Nonconcave Min-Max Problems and Generative Adversarial Networks. *SIAM Journal on Mathematics of Data Science*, **5**, 693-722. <https://doi.org/10.1137/22m1482238>
- [6] Cohen, E. and Teboulle, M. (2024) Alternating and Parallel Proximal Gradient Methods for Nonsmooth, Nonconvex Minimax: A Unified Convergence Analysis. *Mathematics of Operations Research*. <https://doi.org/10.1287/moor.2022.0294>
- [7] Lin, T., Jin, C. and Jordan, M. (2020) On Gradient Descent Ascent for Nonconvex-Concave Minimax Problems. *International Conference on Machine Learning*, 13-18 July 2020, 6083-6093.
- [8] Rockafellar, R. and Wets, J. (2004) Variational Analysis. Springer.
- [9] Lara, F. (2022) On Strongly Quasiconvex Functions: Existence Results and Proximal Point Algorithms. *Journal of Optimization Theory and Applications*, **192**, 891-911. <https://doi.org/10.1007/s10957-021-01996-8>
- [10] Beck, A. and Teboulle, M. (2009) Gradient-Based Algorithms with Applications to Signal-Recovery Problems. In: Palomar, D.P. and Eldar, Y.C., Eds., *Convex Optimization in Signal Processing and Communications*, Cambridge University Press, 42-88. <https://doi.org/10.1017/cbo9780511804458.003>
- [11] Beck, A. (2017) First-Order Methods in Optimization. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611974997>