




# An alternating structure-adapted Bregman proximal gradient descent algorithm for constrained nonconvex nonsmooth optimization problems and its inertial variant

Xue Gao<sup>1</sup> · Xingju Cai<sup>2</sup> · Xiangfeng Wang<sup>3</sup> · Deren Han<sup>4</sup> 

Received: 8 October 2022 / Accepted: 26 May 2023 / Published online: 24 June 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

We consider the nonconvex nonsmooth minimization problem over abstract sets, whose objective function is the sum of a proper lower semicontinuous biconvex function of the entire variables and two smooth nonconvex functions of their private variables. Fully exploiting the problem structure, we propose an alternating structure-adapted Bregman proximal (ASABP for short) gradient descent algorithm, where the geometry of the abstract set and the function is captured by employing generalized Bregman function. Under the assumption that the underlying function satisfies the Kurdyka–Łojasiewicz property, we prove that each bounded sequence generated by ASABP globally converges to a critical point. We then adopt an inertial strategy to accelerate the ASABP algorithm (IASABP), and utilize a backtracking line search scheme to find “suitable” step sizes, making the algorithm efficient and robust. The global  $O(1/K)$  sublinear convergence rate measured by Bregman distance is also established. Furthermore, to illustrate the potential of ASABP and its inertial version (IASABP), we apply them to solving the Poisson linear inverse problem, and the results are promising.

---

✉ Deren Han  
handr@buaa.edu.cn

Xue Gao  
xgao@hebut.edu.cn

Xingju Cai  
caixingju@njnu.edu.cn

Xiangfeng Wang  
xfwang@sei.ecnu.edu.cn

<sup>1</sup> Institute of Mathematics, Hebei University of Technology, Tianjin 300401, People’s Republic of China

<sup>2</sup> School of Mathematical Sciences, Jiangsu Key Lab for NSLSCS, Nanjing Normal University, Nanjing 210023, People’s Republic of China

<sup>3</sup> School of Computer Science and Technology, East China Normal University, Shanghai 200026, People’s Republic of China

<sup>4</sup> School of Mathematical Sciences, LMIB, Beihang University, Beijing 100191, People’s Republic of China

**Keywords** Proximal gradient decent · Bregman distance · Inertial · Kurdyka–Łojasiewicz property · Nonconvex nonsmooth optimization

**Mathematics Subject Classification** 90C26 · 65K05 · 47J25

### 1 Introduction

Let  $X \subset \mathbb{R}^n$  and  $Y \subset \mathbb{R}^m$  be nonempty open convex sets whose closures are denoted by  $\bar{X}, \bar{Y}$  respectively. Let  $F : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  and  $G : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  be proper lower semicontinuous nonconvex functions; and let  $H : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper lower semicontinuous biconvex function. In this paper, we consider the optimization problem

$$\min_{x \in \bar{X}, y \in \bar{Y}} \Phi(x, y) = F(x) + G(y) + H(x, y), \tag{1}$$

which captures numerous applications, such as Poisson inverse problems (see, for instance, the recent review paper [6, 8] and references therein), nonnegative matrix factorization [1, 2], parallel magnetic resonance imaging [7], multi-modal learning for image classification [29].

An intuitive method for solving problem (1) is the alternating minimization method (also known as the Gauss-Seidel method or the block coordinate descent method), which was originally proposed for the case  $X = \mathbb{R}^n, Y = \mathbb{R}^m$ ,

$$\begin{cases} x^{k+1} \in \arg \min \{ \Phi(x, y^k) : x \in \mathbb{R}^n \}, \\ y^{k+1} \in \arg \min \{ \Phi(x^{k+1}, y) : y \in \mathbb{R}^m \}. \end{cases} \tag{2}$$

If  $\Phi$  is continuously differentiable and strictly convex with respect to one argument while the other one is fixed, every limit point of the sequence  $\{(x^k, y^k)\}$  generated by scheme (2) minimizes  $\Phi$  [9, 10]. A classical technique to relax the strict convexity condition is the proximal alternating minimization method [3, 4]. However, minimizing  $\Phi(\cdot, y)$  or  $\Phi(x, \cdot)$  which is the sum of two functions per iteration is not an easy task. To alleviate the computational cost, Bolte et al. [11] considered linearizing  $H$  and proposed the proximal alternating linearized minimization (PALM) algorithm,

$$\begin{cases} x^{k+1} \in \arg \min \left\{ F(x) + \langle \nabla_x H(x^k, y^k), x - x^k \rangle + \frac{1}{2\zeta_k} \|x - x^k\|^2 : x \in \mathbb{R}^n \right\}, \\ y^{k+1} \in \arg \min \left\{ G(y) + \langle \nabla_y H(x^{k+1}, y^k), y - y^k \rangle + \frac{1}{2\varrho_k} \|y - y^k\|^2 : y \in \mathbb{R}^m \right\}. \end{cases} \tag{3}$$

The step sizes  $\zeta_k$  and  $\varrho_k$  are computed according to

$$\zeta_k \in (0, (\text{Lip}(\nabla_x H(\cdot, y^k)))^{-1}), \quad \varrho_k \in (0, (\text{Lip}(\nabla_y H(x^{k+1}, \cdot)))^{-1}),$$

where “Lip(·)” denotes the Lipschitz constant. Under the assumption that  $\Phi$  satisfies the Kurdyka–Łojasiewicz (KL) property (see Def. 2.14 for its definition) and a few other conditions, PALM globally converges to a critical point of  $\Phi$ . This scheme has great advantage when the functions  $F$  and  $G$  are simple in the sense that their proximal operators have closed-form expressions or can be evaluated via a fast scheme, while the stepsizes need to change with the iteration sequence.

When the coupling term  $H$  has a good structure, e.g., it is a quadratic function [11, 21, 27], the difficulties in solving the minimization problems of (2) and (3) may be caused by the component functions  $F$  and  $G$ . For this case, Nikolova and Tan [20] proposed the alternating

structure-adapted proximal (ASAP) gradient descent algorithm,

$$\begin{cases} x^{k+1} \in \arg \min \left\{ H(x, y^k) + \langle \nabla F(x^k), x - x^k \rangle + \frac{1}{2\bar{\zeta}} \|x - x^k\|^2 : x \in \mathbb{R}^n \right\}, \\ y^{k+1} \in \arg \min \left\{ H(x^{k+1}, y) + \langle \nabla G(y^k), y - y^k \rangle + \frac{1}{2\bar{\varrho}} \|y - y^k\|^2 : y \in \mathbb{R}^m \right\}, \end{cases} \quad (4)$$

where

$$\bar{\zeta} \in (0, (\text{Lip}(\nabla F))^{-1}), \quad \bar{\varrho} \in (0, (\text{Lip}(\nabla G))^{-1}).$$

ASAP (4) is much suitable for the case that the component function  $H$  has a good structure in the sense of being simple with respect to each variable  $x$  and  $y$  so that the proximal operators of  $H(\cdot, y)$  and  $H(x, \cdot)$  have closed-form expressions.

ASAP (4) is a good alternative to PALM (3). Nevertheless, both algorithms are restricted to the case that there is no abstract sets, while in many applications, there are abstract sets in the model (1), i.e., either  $X \neq \mathbb{R}^n$  or  $Y \neq \mathbb{R}^m$ , or both  $X \neq \mathbb{R}^n$  and  $Y \neq \mathbb{R}^m$  [6, 8]. Certainly, one can use the indicator functions  $\delta_{\bar{X}}(x)$  and  $\delta_{\bar{Y}}(y)$  to equivalently convert problem (1) to

$$\min_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} F(x) + G(y) + \bar{H}(x, y),$$

where

$$\bar{H}(x, y) = H(x, y) + \delta_{\bar{X}}(x) + \delta_{\bar{Y}}(y).$$

However, in this case, the subproblems in ASAP algorithm (4) are constrained minimization problems, which unavoidably degrades its efficiency. Another key point of the ASAP algorithm is that,  $F$  and  $G$  are assumed to be globally Lipschitz smooth on the entire spaces, which is very restrictive and excludes many applications. Furthermore, the proximal parameters  $\bar{\zeta}$  and  $\bar{\varrho}$  in (4), playing essential roles in numerical implementations, depend on the estimation of the Lipschitz constants of  $\nabla F$  and  $\nabla G$ , which is not an easy task.

There are several techniques that can be utilized to handle some of these difficulties [6, 12, 13, 15, 25]. For example, [13, 15] chose Legendre or Bregman functions to absorb the abstract sets and hence automatically kept the iterates in the abstract sets; Bauschke et al. [6] introduced a NoLips algorithm, avoiding the dependence of the Lipschitz smoothness, which is then extended to the nonconvex case [5, 12, 17, 18, 25, 26, 28]. However, the convergence of this powerful algorithm is limited to the convex case, or the nonconvex case but with the assumptions that  $X = \mathbb{R}^n, Y = \mathbb{R}^m$ . Moreover, for the nonconvex case, the Lipschitz smoothness is still required, which contradicts to the original main intention of NoLips algorithm.

In this paper, we propose an alternating structure-adapted Bregman proximal (ASABP) gradient descent algorithm,

$$\begin{cases} x^{k+1} \in \arg \min \left\{ H(x, y^k) + \langle \nabla F(x^k), x - x^k \rangle + \frac{1}{\tau_k} D_h(x, x^k) : x \in \bar{X} \right\}, \\ y^{k+1} \in \arg \min \left\{ H(x^{k+1}, y) + \langle \nabla G(y^k), y - y^k \rangle + \frac{1}{\sigma_k} D_\psi(y, y^k) : y \in \bar{Y} \right\}, \end{cases} \quad (5)$$

where we choose the generalized Bregman functions  $h$  and  $\psi$ , such that the pairs  $(F, h)$  and  $(G, \psi)$  are GL-smad on  $X$  and  $Y$  (see Definition 2.6 and Definition 2.12). By associating the generalized Bregman functions  $h$  and  $\psi$  to the objective functions  $F$  and  $G$  in a suitable way, ASABP can absorb the abstract sets and result in the subproblems unconstrained ones. Merely assuming that the underlying function satisfies the Kurdyka–Łojasiewicz (KL) property yet without the Lipschitz smoothness, we establish the global convergence to a critical point.

This result paves a new way for analyzing the NoLips algorithms in solving nonsmooth nonconvex optimization problems with abstract sets.

Numerically, the vectors  $x^k$  and  $y^k$  and the parameters  $\tau_k$  and  $\sigma_k$  influent the performance of the algorithm greatly. To improve the numerical performance, we merge an inertial strategy [22] into the proposed ASABP algorithm, leading to the inertial alternating structure-adapted Bregman proximal (IASABP) gradient descent algorithm (see Algorithm 2). The inertial stepsizes are judiciously selected by backtracking line search strategy. Under mild conditions, we can get the  $O(1/K)$  convergence rate of the sequence generated by IASABP measured by Bregman distance.

The rest of this paper is organized as follows. In Sect. 2, we list some basic notations and preliminary materials. In Sect. 3, we present the new alternating structure-adapted Bregman proximal (ASABP) gradient descent algorithm and its inertial version IASABP in detail, and give the convergence analysis results. To show the ability of IASABP and ASABP in dealing with non-Lipschitz problem with abstract sets, we implement them to solve the Poisson linear inverse problem and perform numerical experiment in Sect. 4. Finally, we make some conclusions in Sect. 5.

## 2 Preliminaries

In this section, we summarize some notations and elementary facts to be used for further analysis.

The Euclidean scalar product in  $\mathbb{R}^d$  and its corresponding norm are respectively denoted by  $\langle \cdot, \cdot \rangle$  and  $\| \cdot \|$ . If  $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$  is a point-to-set mapping, its graph is defined by

$$\text{Graph } F := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : y \in F(x)\}.$$

Similarly the graph of a real-extended-valued function  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  is defined by

$$\text{Graph } f := \{(x, s) \in \mathbb{R}^d \times \mathbb{R} : s = f(x)\}.$$

Given a function  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ , its domain is

$$\text{dom } f := \{x \in \mathbb{R}^d : f(x) < +\infty\},$$

and  $f$  is proper if and only if  $\text{dom } f \neq \emptyset$ . The proximal operator associated with  $f$  is given by

$$\text{prox}_f(x) = \arg \min_{u \in \mathbb{R}^d} \left\{ f(u) + \frac{1}{2} \|u - x\|^2 \right\} \quad \text{for any } x \in \mathbb{R}^d. \tag{6}$$

The indicator function of a nonempty set  $C$  is defined by

$$\delta_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

The normal cone of a nonempty closed convex set  $C \subset \mathbb{R}^d$  at  $x \in C$  is defined by

$$N_C(x) = \{d \in \mathbb{R}^d : \langle d, y - x \rangle \leq 0, \forall y \in C\};$$

and if  $x \notin C$ , we set  $N_C(x) = \emptyset$ . For any subset  $C \subset \mathbb{R}^d$  and any  $x \in \mathbb{R}^d$ , the distance from  $x$  to  $C$ , denoted by  $\text{dist}(x, C)$ , is defined as

$$\text{dist}(x, C) := \inf_{y \in C} \|y - x\|.$$

When  $C = \emptyset$ , we set  $\text{dist}(x, C) = +\infty$  for all  $x \in \mathbb{R}^d$ .

## 2.1 Subdifferentials and partial subdifferentials

Let us first recall a few definitions and results concerning subdifferentials of a given function.

**Definition 2.1** [3] Let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper lower semicontinuous function.

- (i) For each  $x \in \text{dom } f$ , the Fréchet subdifferential of  $f$  at  $x$ , written as  $\hat{\partial} f(x)$ , is the set of vectors  $u \in \mathbb{R}^d$  which satisfy

$$\liminf_{y \neq x, y \rightarrow x} \frac{f(y) - f(x) - \langle u, y - x \rangle}{\|y - x\|} \geq 0.$$

If  $x \notin \text{dom } f$ , we set  $\hat{\partial} f(x) = \emptyset$ .

- (ii) The set

$$\partial f(x) := \left\{ u \in \mathbb{R}^d : \exists x^k \rightarrow x, f(x^k) \rightarrow f(x), u^k \in \hat{\partial} f(x^k) \rightarrow u \right\},$$

is the limiting subdifferential, or simply the subdifferential for short, of  $f$  at  $x \in \text{dom } f$ .

**Remark 2.2** [24] From Definition 2.1, we have the following conclusions.

- (i)  $\hat{\partial} f(x) \subset \partial f(x)$  for each  $x \in \mathbb{R}^d$ ; both sets are closed; and  $\hat{\partial} f(x)$  is convex while  $\partial f(x)$  is not necessarily convex.
- (ii) Let  $(x^k, u^k) \in \text{Graph } \partial f$  be a sequence that converges to  $(x, u)$ . If  $f(x^k)$  converges to  $f(x)$  as  $k \rightarrow +\infty$ , then  $(x, u) \in \text{Graph } \partial f$ .
- (iii) (Fermat’s rule) A necessary but not sufficient condition for  $x \in \mathbb{R}^d$  to be a local minimizer of  $f$  is that  $x$  is a critical point, that is,

$$0 \in \partial f(x).$$

The set of critical points of  $f$  is denoted by *crit*  $f$ .

- (iv) If  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  is proper lower semicontinuous and  $h : \mathbb{R}^d \rightarrow \mathbb{R}$  is continuously differentiable, then for any  $x \in \text{dom } f$ ,  $\partial(f + h)(x) = \partial f(x) + \nabla h(x)$ .

In first order alternating methods (e.g., scheme (5)), each iteration is equivalent to finding a minimizer of the auxiliary function, which is characterized by Fermat’s rule. Since Fermat’s rule is defined on subgradients, applying alternating schemes leads us to consider partial subgradients defined as follows.

**Definition 2.3** [20] Given  $H : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ , the subdifferential of  $H$  at  $(x^+, y^+)$  is denoted by  $\partial H(x^+, y^+)$ . In addition,  $\partial_x H(x^+, y^+)$  and  $\partial_y H(x^+, y^+)$  are the partial subdifferentials of  $H(\cdot, y^+)$  at  $x^+$  and  $H(x^+, \cdot)$  at  $y^+$  respectively.

**Remark 2.4** Note that the desired critical point is characterized by its whole subgradient, namely  $(0, 0) \in \partial H(x^*, y^*)$ . Let  $H : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  be proper function. Then

$$\partial H(x, y) \subset \partial_x H(x, y) \times \partial_y H(x, y), \quad \forall (x, y) \in \text{dom } H,$$

holds naturally [24]. If the function  $H$  also satisfies

$$\partial_x H(x, y) \times \partial_y H(x, y) \subset \partial H(x, y), \quad \forall (x, y) \in \text{dom } H, \tag{7}$$

at least for  $(x, y) = (0, 0)$ , then  $0 \in \partial_x H(x^*, y^*) \times \partial_y H(x^*, y^*)$  means that  $(x^*, y^*)$  is a critical point. (7) is automatically satisfied in the form as shown in the following example, which gives a sufficient condition.

**Example 1** Let  $H : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  be of the form

$$H(x, y) = h(x, y) + \xi(x) + \eta(y),$$

where  $h : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  is continuously differentiable;  $\xi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  and  $\eta : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  are proper lower semicontinuous functions. Then, for any  $(x, y) \in \text{dom}H = \text{dom}\xi \times \text{dom}\eta$ ,  $H$  satisfies (7).

More details and examples can refer to [20]. Besides condition (7), another property, closely linked to the closedness of the partial subdifferential, is crucial.

**Definition 2.5** [20] (Parametric closedness of the partial subdifferentials) Let  $H : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  be a function,  $\{(x^k, y^k)\}_{k \in \mathbb{N}}$  and  $(x^+, y^+) \in \text{dom}H$  such that for  $k \rightarrow +\infty$ ,

$$(x^k, y^k) \rightarrow (x^+, y^+), \quad H(x^k, y^k) \rightarrow H(x^+, y^+). \tag{8}$$

The  $x$ -partial subdifferential of  $H$  is said to be parametrically closed at  $(x^+, y^+)$  with respect to the sequence  $\{(x^k, y^k)\}_{k \in \mathbb{N}}$  if, for any sequence  $\{p_x^k\}_{k \in \mathbb{N}}$ ,

$$\partial_x H(x^k, y^k) \ni p_x^k \rightarrow p_x^+, \quad k \rightarrow +\infty,$$

then  $p_x^+ \in \partial_x H(x^+, y^+)$ . Similarly, we can define the parametric closedness of  $y$ -partial subdifferential of  $H$  at  $(x^+, y^+)$  with respect to the sequence  $\{(x^k, y^k)\}_{k \in \mathbb{N}}$ . The  $x$ -partial and  $y$ -partial subdifferentials of  $H$  are said to be parametrically closed if they are parametrically closed at any point  $(x^+, y^+) \in \text{dom}H$  with respect to any sequence satisfying (8).

**Example 2** [20] Let  $H : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  be a biconvex function continuous on its domain. Then its partial subdifferentials are parametrically closed at any point  $(x^+, y^+) \in \text{dom}H$ .

### 2.2 Generalized Bregman function and Bregman distance

To define the Bregman distance, we first give the definition of generalized Bregman function.

**Definition 2.6** (Generalized Bregman function) Let  $\Omega$  be a nonempty convex open subset of  $\mathbb{R}^d$ , and  $\overline{\Omega}$  denotes its closure. Associated with  $\Omega$ , a function  $h : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  is called a generalized Bregman function if it satisfies the following:

- (i)  $h$  is proper lower semicontinuous and convex with  $\text{dom}h \subset \overline{\Omega}$ .
- (ii)  $\text{dom}\partial h = \Omega = \text{int dom}h$ .
- (iii)  $h$  is strictly convex on  $\text{int dom}h$ .
- (iv) If  $\{x^k\} \in \Omega$  converges to  $\bar{x} \in \partial\Omega$  (the boundary of  $\Omega$ ), then  $\lim_{k \rightarrow +\infty} \langle \nabla h(x^k), u - x^k \rangle = -\infty$  for all  $u \in \Omega$ .

We denote the class of generalized Bregman functions by  $\mathcal{G}(\Omega)$ .

**Definition 2.7** (Bregman distance) Given  $h \in \mathcal{G}(\Omega)$ , the Bregman distance associated to  $h$ , denoted by  $D_h : \text{dom}h \times \text{int dom}h \rightarrow \mathbb{R}_+$ , is defined by

$$D_h(x, y) = h(x) - h(y) - \langle \nabla h(y), x - y \rangle.$$

**Remark 2.8** We give some remarks on Definition 2.6 and Definition 2.7.

- (i) The class of functions which satisfy (i) and (ii) in Definition 2.6 are the Kernel generating distance [12]. And the class of functions satisfying (i)-(iii) are Legendre functions [6].
- (ii) Definition 2.6 (ii) is equivalent to the statement that  $h$  is essentially smooth [23, 25], which means that  $\partial h(x) = \{\nabla h(x)\}$  when  $x \in \text{int dom}h$ , while  $\partial h(x) = \emptyset$  when  $x \notin \text{int dom}h$ .
- (iii) Given  $h \in \mathcal{G}(\Omega)$ . If  $h$  is continuous on  $\overline{\Omega}$ , Definition 2.6 (iv) holds if and only if  $\lim_{k \rightarrow +\infty} D_h(u, x^k) = +\infty$  for all  $u \in \Omega$  and all sequence  $\{x^k\} \in \Omega$  which converges to a point  $\bar{x} \in \partial\Omega$ .
- (iv) Note that the structural form of  $D_h$  is also useful when  $h$  is not convex, and still enjoys two simple but remarkable properties,

- The three point identity: For any  $y, z \in \text{int dom}h$  and  $x \in \text{dom}h$ , we have

$$D_h(x, z) - D_h(x, y) - D_h(y, z) = \langle \nabla h(y) - \nabla h(z), x - y \rangle.$$

- Linear additivity: For any  $\alpha, \beta \in \mathbb{R}$ , and any functions  $h_1$  and  $h_2$ , we have

$$D_{\alpha h_1 + \beta h_2}(x, y) = \alpha D_{h_1}(x, y) + \beta D_{h_2}(x, y),$$

for any  $(x, y) \in (\text{dom}h_1 \cap \text{dom}h_2)^2$  such that both  $h_1$  and  $h_2$  are differentiable at  $y$ .

It is obvious that Bregman distance is, in general, not symmetric. It is thus natural to introduce a measure for the symmetry of  $D_h$ .

**Definition 2.9** [6] (Symmetry coefficient) Given  $h \in \mathcal{G}(\Omega)$ , its symmetry coefficient is defined by

$$\alpha(h) = \inf\{D_h(x, y)/D_h(y, x) | (x, y) \in \text{int dom}h \times \text{int dom}h, x \neq y\} \in [0, 1].$$

**Remark 2.10** The symmetry coefficient has the following properties:

- (i) Clearly, the closer  $\alpha(h)$  gets to 1, the more symmetric  $D_h$  is.
- (ii) For any  $x, y \in \text{int dom}h$ ,

$$\alpha(h)D_h(x, y) \leq D_h(y, x) \leq \frac{1}{\alpha(h)}D_h(x, y),$$

where we have adopted the convention that  $\frac{1}{0} = +\infty$  and  $+\infty \times r = +\infty$  for any  $r \geq 0$ .

**Lemma 2.11** Given  $h \in \mathcal{G}(\Omega)$ , for any proper lower semicontinuous convex function  $\Gamma : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  and any  $z \in \text{int dom}h$ , if

$$z_+ = \arg \min_{x \in \Omega} \{\Gamma(x) + D_h(x, z)\},$$

then for any  $x \in \text{dom}h$ ,

$$\Gamma(z_+) + D_h(z_+, z) \leq \Gamma(x) + D_h(x, z) - D_h(x, z_+).$$

### 2.3 Generalized L-smooth adaptable and extended descent lemma

We denote  $\mathcal{G}(f, h)$  the set of pair of functions  $(f, h)$  satisfies

- (i)  $h \in \mathcal{G}(\Omega)$ ,
- (ii)  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  is proper lower semicontinuous nonconvex with  $\text{dom}h \subset \text{dom}f$ , which is continuously differentiable on  $\overline{\Omega}$ .

**Definition 2.12** (Generalized L-smooth adaptable) A pair of functions  $(f, h) \in \mathcal{G}(f, h)$  is called generalized L-smooth adaptable (GL-smad) on  $\Omega$  if there exists  $L > 0$  such that  $Lh + f$  and  $Lh - f$  are convex on  $\Omega$ .

From the above definition we immediately obtain the two-sided descent lemma, which complements and extends the NoLips descent lemma derived in [6].

**Lemma 2.13** (Extended descent lemma) *The pair of functions  $(f, h) \in \mathcal{G}(f, h)$  is GL-smad on  $\Omega$  if and only if*

$$|f(x) - f(y) - \langle \nabla f(y), x - y \rangle| \leq LD_h(x, y), \quad \forall x \in \text{dom}h, \forall y \in \text{int dom}h.$$

Due to the structural form of  $D_f$ , the extended descent lemma reads equivalently as

$$|f(x) - f(z) - \langle \nabla f(y), xz \rangle + D_f(z, y)| \leq LD_h(x, y),$$

for any  $x, z \in \text{dom}h$  and any  $y \in \text{int dom}h$ .

When the function  $f$  is assumed to be convex, the convexity condition of  $Lh + f$  naturally holds. In this case, the NoLips descent lemma given in [6] is recovered. When  $h(z) = \frac{1}{2}\|z\|^2$  and consequently  $D_h(x, y) = \frac{1}{2}\|x - y\|^2$ , Lemma 2.13 would be reduced to the descent lemma [19],

$$|f(x) - f(y) - \langle \nabla f(y), x - y \rangle| \leq \frac{L}{2}\|x - y\|^2, \quad \forall x \in \text{dom}h, \forall y \in \text{int dom}h.$$

### 2.4 Kurdyka–ojasiewicz property

Let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper lower semicontinuous function. For  $-\infty < \eta_1 < \eta_2 \leq +\infty$ , define

$$[\eta_1 < f < \eta_2] := \{x \in \mathbb{R}^d : \eta_1 < f(x) < \eta_2\}.$$

**Definition 2.14** [4] (Kurdyka–Łojasiewicz property) The function  $f$  is said to have the Kurdyka–Łojasiewicz (KL) property at  $\bar{x} \in \text{dom} \partial f$  if there exist  $\eta \in (0, +\infty)$ , a neighborhood  $U$  of  $\bar{x}$  and a continuous concave function  $\varphi : [0, \eta) \rightarrow \mathbb{R}_+$  such that

- (i)  $\varphi(0) = 0$ ;
- (ii)  $\varphi$  is  $C^1$  on  $(0, \eta)$ ;
- (iii) for all  $s \in (0, \eta)$ ,  $\varphi'(s) > 0$ ;
- (iv) for all  $x$  in  $U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]$ , the Kurdyka–Łojasiewicz inequality holds, i.e.,

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

**Remark 2.15** (i) If  $f$  satisfies the KL property at each point of  $\text{dom} \partial f$ , then  $f$  is called a KL function;  
 (ii) For convenience, denote the class of functions which satisfy (i), (ii) and (iii) in Definition 2.14 as  $\Phi_\eta$ .

The following lemma gives the important uniformized KL property. More properties and examples can refer to [3, 11, 14].



**Lemma 2.16** [11] (Uniformized KL property) Let  $C$  be a compact set and let  $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper lower semicontinuous function. Assume that  $f$  is constant on  $C$  and satisfies the KL property at each point of  $C$ . Then, there exist  $\varepsilon > 0, \eta > 0$  and  $\varphi \in \Phi_\eta$  such that for any  $\bar{x} \in C$  and all  $x$  in the following intersection,

$$\{x \in \mathbb{R}^n : \text{dist}(x, C) < \varepsilon\} \cap [f(\bar{x}) < f(x) < f(\bar{x}) + \eta],$$

one has,

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

### 3 Algorithms and convergence

In this section, we first describe our alternating structure-adapted Bregman proximal (ASABP) gradient descent algorithm and its inertial version (IASABP algorithm) for solving

$$\min_{x \in \bar{X}, y \in \bar{Y}} \Phi(x, y) = F(x) + G(y) + H(x, y),$$

where  $X \subset \mathbb{R}^n$  and  $Y \subset \mathbb{R}^m$  are open convex sets with nonempty interiors, and  $\bar{X}$  and  $\bar{Y}$  are their closures, respectively. Then we present their convergence rate results and establish the global convergence result for ASABP algorithm. Throughout the rest of this section, we suppose the following assumption, which guarantees that the algorithms are well-defined.

**Assumption 3.1** (i)  $\inf\{\Phi(x, y) : (x, y) \in \bar{X} \times \bar{Y}\} > -\infty$  and  $\Phi$  is coercive.

- (ii)  $h \in \mathcal{G}(X)$  and  $\psi \in \mathcal{G}(Y)$  such that the pairs  $(F, h) \in \mathcal{G}(F, h)$  and  $(G, \psi) \in \mathcal{G}(G, \psi)$  are GL-smad on  $X$  and  $Y$  with coefficients  $L_h$  and  $L_\psi$  respectively.
- (iii) For the sequence  $\{x^k\}_{k \in \mathbb{N}} \in \text{int dom} h$  and  $x \in \text{dom} h, \|x^k - x\| \rightarrow 0 \iff D_h(x, x^k) \rightarrow 0$ . Similarly,  $\|y^k - y\| \rightarrow 0 \iff D_\psi(y, y^k) \rightarrow 0$  for any sequence  $\{y^k\}_{k \in \mathbb{N}} \in \text{int dom} \psi$  and  $y \in \text{dom} \psi$ .
- (iv)  $H : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper lower semicontinuous biconvex function.

#### 3.1 The ASABP and IASABP algorithm

We now first introduce the alternating structure-adapted Bregman proximal (ASABP) gradient descent algorithm.

---

##### Algorithm 1 ASABP

---

- 1: Start with arbitrary initial points  $x^0 \in \text{int dom} h, y^0 \in \text{int dom} \psi, 0 < \tau_k < \frac{1}{L_h}$  and  $0 < \sigma_k < \frac{1}{L_\psi}$ .
- 2: For each  $k = 0, 1, \dots, \{(x^k, y^k)\}_{k \in \mathbb{N}}$  is generated as follows:  
 Compute

$$x^{k+1} \in \arg \min \left\{ H(x, y^k) + \langle \nabla F(x^k), x - x^k \rangle + \frac{1}{\tau_k} D_h(x, x^k) : x \in \bar{X} \right\}. \quad (9)$$

Then compute

$$y^{k+1} \in \arg \min \left\{ H(x^{k+1}, y) + \langle \nabla G(y^k), y - y^k \rangle + \frac{1}{\sigma_k} D_\psi(y, y^k) : y \in \bar{Y} \right\}. \quad (10)$$


---

Under Assumption 3.1,  $x^{k+1}$  will be automatically in  $X$  and  $y^{k+1}$  will be automatically in  $Y$ . Hence, the problems (9) and (10) are essentially unconstrained optimization problems, which in many cases enable these problems to possess closed-form solutions.

Note that the efficiency of ASABP depends on the vectors  $x^k$  and  $y^k$ , and the parameters  $\tau_k$  and  $\sigma_k$ . For the first order algorithms, it was found that extrapolation can accelerate the convergence, and the adaptive line search scheme can alleviate the difficulty in choosing the parameters. Hence, we give the inertial variant of ASABP algorithm, i.e, the inertial alternating structure-adapted Bregman proximal gradient descent algorithm, which we denote as IASABP for short.

**Algorithm 2** IASABP

- 1: Start with arbitrary initial points  $x^{-1} = x^0 \in \text{int dom}h$ ,  $y^{-1} = y^0 \in \text{int dom}\psi$ ,  $0 < \tau_k < \frac{1}{L_h}$  and  $0 < \sigma_k < \frac{1}{L_\psi}$ . Fix  $0 < \eta < 1$  and  $0 < \theta < 1$ .
- 2: For each  $k = 0, 1, \dots$ ,  $\{(x^k, y^k)\}_{k \in \mathbb{N}}$  is generated as follows:

- (i) Take  $\alpha_k^0 \in [0, 1)$  and  $C_k^1 = \frac{\frac{1}{\tau_k}}{\frac{1}{\tau_k} + L_h}$ . Find the smallest nonnegative integer  $j$  satisfying

$$D_h(x^k, x^k + \eta^j \alpha_k^0 (x^k - x^{k-1})) \leq \theta C_k^1 D_h(x^k, x^{k-1}). \tag{11}$$

Set  $\alpha_k = \eta^j \alpha_k^0$ . Compute

$$\bar{x}^k = x^k + \alpha_k (x^k - x^{k-1}),$$

then

$$x^{k+1} = \arg \min \left\{ H(x, y^k) + \langle \nabla F(\bar{x}^k), x - \bar{x}^k \rangle + \frac{1}{\tau_k} D_h(x, \bar{x}^k) : x \in \bar{X} \right\}. \tag{12}$$

- (ii) Take  $\beta_k^0 \in [0, 1)$  and  $C_k^2 = \frac{\frac{1}{\sigma_k}}{\frac{1}{\sigma_k} + L_\psi}$ . Find the smallest nonnegative integer  $i$  satisfying

$$D_\psi(y^k, y^k + \eta^i \beta_k^0 (y^k - y^{k-1})) \leq \theta C_k^2 D_\psi(y^k, y^{k-1}). \tag{13}$$

Set  $\beta_k = \eta^i \beta_k^0$ . Compute

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

then

$$y^{k+1} = \arg \min \left\{ H(x^{k+1}, y) + \langle \nabla G(\bar{y}^k), y - \bar{y}^k \rangle + \frac{1}{\sigma_k} D_\psi(y, \bar{y}^k) : y \in \bar{Y} \right\}. \tag{14}$$

**Remark 3.1** (i) Notice that (11) and (13) hold naturally when taking  $\alpha_k^0 = 0$  and  $\beta_k^0 = 0$  for all  $k \geq 0$ . In this case, the inertial strategy makes no difference, i.e.,  $\bar{x}^k = x^k$ ,  $\bar{y}^k = y^k$ . IASABP algorithm then reduces to the ASABP algorithm.

- (ii) (Finite termination of line search) There exists  $J_k \in \mathbb{N}$  in step (11) such that for any  $j \geq J_k$ ,  $\alpha_k = \eta^j \alpha_k^0$  satisfies

$$D_h(x^k, x^k + \alpha_k (x^k - x^{k-1})) \leq \theta C_k^1 D_h(x^k, x^{k-1}).$$

Similarly, we can get  $\beta_k$  via finite number of step (13).

We then verify that the ASABP and IASABP algorithms are well defined. To this end, choose arbitrarily  $u \in \text{int dom}h$  and stepsize  $0 < \tau < \frac{1}{L_h}$ . Fixed  $y \in \text{int dom}\psi$ , we define the Bregman proximal mapping,

$$\begin{aligned} \mathcal{B}_\tau(u; y) &= \arg \min \left\{ H(x, y) + \langle \nabla F(u), x - u \rangle + \frac{1}{\tau} D_h(x, u) : x \in \bar{X} \right\} \\ &= \arg \min \left\{ H(x, y) + \langle \nabla F(u), x - u \rangle + \frac{1}{\tau} D_h(x, u) : x \in \mathbb{R}^n \right\}, \end{aligned}$$

where the second equality follows from the fact that  $\text{dom}h \subset \bar{X}$ . Observe that  $x^{k+1} \in \mathcal{B}_{\tau_k}(u^k; y^k)$ , where  $u^k = x^k$  in ASABP algorithm and  $u^k = \bar{x}^k$  for IASABP algorithm. Hence from Proposition 3.2, (9) and (12) are well defined, and so are the (10) and (14).

**Proposition 3.2** *Suppose that Assumption 3.1 holds. Let  $y \in \text{int dom}\psi$ ,  $u \in \text{int dom}h$  and  $0 < \tau < \frac{1}{L_h}$ . Then the set  $\mathcal{B}_\tau(u; y)$  is a nonempty compact single valued mapping from  $\text{int dom}h$  to  $\text{int dom}h$ .*

**Proof** Fix any  $y \in \text{int dom}\psi$ ,  $u \in \text{int dom}h$  and  $0 < \tau < \frac{1}{L_h}$ . For any  $x \in \mathbb{R}^n$ , we define

$$\Phi_h(x, u; y) = H(x, y) + F(u) + G(y) + \langle \nabla F(u), x - u \rangle + \frac{1}{\tau} D_h(x, u),$$

so that  $\mathcal{B}_\tau(u; y) = \arg \min_{x \in \mathbb{R}^n} \Phi_h(x, u; y)$ . Since  $h$  is strictly convex and  $H$  is biconvex as assumed in Assumption 3.1, the objective  $\Phi_h(x, u; y)$  have at most one minimizer. Taking into account that  $(F, h)$  is GL-smad with  $L_h$ ,

$$\begin{aligned} \Phi_h(x, u; y) &= \Phi(x, y) - \left( F(x) - F(u) - \langle \nabla F(u), x - u \rangle - \frac{1}{\tau} D_h(x, u) \right) \\ &\geq \Phi(x, y). \end{aligned}$$

According to Assumption 3.1 (i), it holds that

$$\lim_{\|x\| \rightarrow +\infty} \Phi_h(x, u; y) \geq \lim_{\|x\| \rightarrow +\infty} \Phi(x, y) = +\infty.$$

Since  $\Phi_h$  is also proper and lower semicontinuous, invoking Weierstrass’s theorem, it follows that the set  $\mathcal{B}_\tau(u; y)$  is nonempty and compact. Furthermore, the optimality condition for  $\mathcal{B}_\tau(u; y)$  implies that  $\partial h(\mathcal{B}_\tau(u; y))$  must be nonempty, and Definition 2.6 indicates that  $\mathcal{B}_\tau(u; y)$  belongs to  $\text{int dom}h$ . □

### 3.2 Rate of convergence

In this subsection, we establish the global sublinear rate of convergence for ASABP and IASABP, measured by Bregman distance. Using the extended descent lemma, we obtain the following key estimation for the objective function  $\Phi$ , which will play an important role in deriving our main convergence results.

**Lemma 3.3** *Suppose that Assumption 3.1 holds. Let  $\{(x^k, y^k)\}_{k \in \mathbb{N}}$  be the sequence generated by IASABP, then we have*

$$\begin{aligned} &\Phi(x^{k+1}, y^{k+1}) + \frac{\alpha(h)}{\tau_k} D_h(x^{k+1}, x^k) + \frac{\alpha(\psi)}{\sigma_k} D_\psi(y^{k+1}, y^k) \\ &\leq \Phi(x^k, y^k) + \frac{\theta}{\tau_k} D_h(x^k, x^{k-1}) + \frac{\theta}{\sigma_k} D_\psi(y^k, y^{k-1}), \end{aligned}$$

where  $\alpha(h)$  and  $\alpha(\psi)$  are the symmetric coefficients of  $h, \psi$  respectively.

**Proof** Applying Lemma 2.11 with  $\Gamma(x) = \tau_k (H(x, y^k) + \langle \nabla F(\bar{x}^k), x - \bar{x}^k \rangle)$ , it yields that for any  $x \in \text{dom}h$ ,

$$\begin{aligned} & H(x^{k+1}, y^k) - H(x, y^k) \\ & \leq \langle \nabla F(\bar{x}^k), x - x^{k+1} \rangle + \frac{1}{\tau_k} D_h(x, \bar{x}^k) - \frac{1}{\tau_k} D_h(x, x^{k+1}) - \frac{1}{\tau_k} D_h(x^{k+1}, \bar{x}^k) \\ & \leq F(x) - F(x^{k+1}) + L_h D_h(x, \bar{x}^k) + D_F(x^{k+1}, \bar{x}^k) \\ & \quad + \frac{1}{\tau_k} D_h(x, \bar{x}^k) - \frac{1}{\tau_k} D_h(x, x^{k+1}) - \frac{1}{\tau_k} D_h(x^{k+1}, \bar{x}^k) \\ & \leq F(x) - F(x^{k+1}) + \left( \frac{1}{\tau_k} + L_h \right) D_h(x, \bar{x}^k) \\ & \quad - \frac{1}{\tau_k} D_h(x, x^{k+1}) - \left( \frac{1}{\tau_k} - L_h \right) D_h(x^{k+1}, \bar{x}^k), \end{aligned}$$

where the second inequality follows from Lemma 2.13 and the last inequality follows from the GL-smad property of  $(F, h)$ . Particularly, taking  $x = x^k$ , together with  $0 < \tau_k < \frac{1}{L_h}$  and the nonnegativity of Bregman distance  $D_h$ , we obtain

$$\begin{aligned} & H(x^{k+1}, y^k) + F(x^{k+1}) - \left( H(x^k, y^k) + F(x^k) \right) \\ & \leq \left( \frac{1}{\tau_k} + L_h \right) D_h(x^k, \bar{x}^k) - \frac{1}{\tau_k} D_h(x^k, x^{k+1}). \end{aligned}$$

If  $\bar{x}^k = x^k$  we get  $D_h(x^k, \bar{x}^k) = 0$ . Hence,

$$H(x^{k+1}, y^k) + F(x^{k+1}) + \frac{1}{\tau_k} D_h(x^k, x^{k+1}) \leq H(x^k, y^k) + F(x^k).$$

If  $\bar{x}^k \neq x^k$ , according to (11), we have  $D_h(x^k, \bar{x}^k) \leq \theta C_k^1 D_h(x^k, x^{k-1})$ . Recalling the definition of  $C_k^1$ , we have

$$\begin{aligned} & H(x^{k+1}, y^k) + F(x^{k+1}) + \frac{1}{\tau_k} D_h(x^k, x^{k+1}) \\ & \leq H(x^k, y^k) + F(x^k) + \left( \frac{1}{\tau_k} + L_h \right) \theta C_k^1 D_h(x^k, x^{k-1}) \\ & \leq H(x^k, y^k) + F(x^k) + \frac{\theta}{\tau_k} D_h(x^k, x^{k-1}). \end{aligned}$$

Consequently, for any case we conclude that for any  $k \in \mathbb{N}$ ,

$$\begin{aligned} & H(x^{k+1}, y^k) + F(x^{k+1}) + \frac{1}{\tau_k} D_h(x^k, x^{k+1}) \\ & \leq H(x^k, y^k) + F(x^k) + \frac{\theta}{\tau_k} D_h(x^k, x^{k-1}). \end{aligned} \tag{15}$$

We can obtain the following inequality in a similar way,

$$\begin{aligned} & H(x^{k+1}, y^{k+1}) + G(y^{k+1}) + \frac{1}{\sigma_k} D_\psi(y^k, y^{k+1}) \\ & \leq H(x^{k+1}, y^k) + G(y^k) + \frac{\theta}{\sigma_k} D_\psi(y^k, y^{k-1}). \end{aligned} \tag{16}$$

Adding (15) and (16) and we can obtain that for any  $k \in \mathbb{N}$ ,

$$\begin{aligned} &\Phi(x^{k+1}, y^{k+1}) + \frac{1}{\tau_k} D_h(x^k, x^{k+1}) + \frac{1}{\sigma_k} D_\psi(y^k, y^{k+1}) \\ &\leq \Phi(x^k, y^k) + \frac{\theta}{\tau_k} D_h(x^k, x^{k-1}) + \frac{\theta}{\sigma_k} D_\psi(y^k, y^{k-1}). \end{aligned}$$

Combining with Remark 2.10 (ii) completes the proof. □

Equipped with Lemma 3.3, we introduce the following useful notations,

$$u_\gamma^k := (x^k, y^k, x^{k-1}, y^{k-1}), \quad \underline{\tau} \leq \inf\{\tau_k\} \leq \sup\{\tau_k\} \leq \bar{\tau}, \quad \underline{\sigma} \leq \inf\{\sigma_k\} \leq \sup\{\sigma_k\} \leq \bar{\sigma}, \quad \forall k \in \mathbb{N}.$$

And we define an auxiliary sequence by

$$J(u_\gamma^k) = \Phi(x^k, y^k) + M_1 D_h(x^k, x^{k-1}) + M_2 D_\psi(y^k, y^{k-1}),$$

where  $M_1 = \frac{1}{2} \left( \frac{\theta}{\underline{\tau}} + \frac{\alpha(h)}{\bar{\tau}} \right)$  and  $M_2 = \frac{1}{2} \left( \frac{\theta}{\underline{\sigma}} + \frac{\alpha(\psi)}{\bar{\sigma}} \right)$ . Choosing  $\theta < \min \left\{ \frac{\tau \alpha(h)}{\bar{\tau}}, \frac{\sigma \alpha(\psi)}{\bar{\sigma}} \right\}$ , we have

$$\delta_1 = \min \left\{ M_1 - \frac{\theta}{\underline{\tau}}, \frac{\alpha(h)}{\bar{\tau}} - M_1 \right\} > 0 \quad \text{and} \quad \delta_2 = \min \left\{ M_2 - \frac{\theta}{\underline{\sigma}}, \frac{\alpha(\psi)}{\bar{\sigma}} - M_2 \right\} > 0. \tag{17}$$

**Lemma 3.4** *Suppose that Assumption 3.1 holds. Let  $\{z_\gamma^k = (x^k, y^k)\}_{k \in \mathbb{N}}$ ,  $\{\bar{z}_\gamma^k = (\bar{x}^k, \bar{y}^k)\}_{k \in \mathbb{N}}$  be the sequences generated by IASABP.*

(i) *The sequence  $\{J(u_\gamma^k)\}_{k \in \mathbb{N}}$  is monotonically nonincreasing and in particular for all  $k \geq 0$ ,*

$$\begin{aligned} &\delta_\gamma \left( D_h(x^k, x^{k-1}) + D_\psi(y^k, y^{k-1}) + D_h(x^{k+1}, x^k) + D_\psi(y^{k+1}, y^k) \right) \\ &\leq J(u_\gamma^k) - J(u_\gamma^{k+1}), \end{aligned} \tag{18}$$

where  $\delta_\gamma = \min\{\delta_1, \delta_2\}$  and  $\delta_1, \delta_2$  are defined as (17). Moreover,  $\{J(u_\gamma^k)\}_{k \in \mathbb{N}}$  converges to some finite value, denoted by  $J^*$ .

(ii) *We have*

$$\sum_{k=0}^{+\infty} \left( D_h(x^{k+1}, x^k) + D_\psi(y^{k+1}, y^k) \right) < +\infty,$$

and hence  $\lim_{k \rightarrow +\infty} D_h(x^{k+1}, x^k) = 0$ ,  $\lim_{k \rightarrow +\infty} D_\psi(y^{k+1}, y^k) = 0$ . Moreover,  $\lim_{k \rightarrow +\infty} \|z_\gamma^{k+1} - z_\gamma^k\| = 0$  and  $\lim_{k \rightarrow +\infty} \|z_\gamma^k - \bar{z}_\gamma^k\| = 0$ .

(iii) (Convergence rate) *For any  $K \geq 0$ , it holds that*

$$\min_{0 \leq k \leq K} \left\{ D_h(x^{k+1}, x^k) + D_\psi(y^{k+1}, y^k) \right\} \leq \frac{1}{\delta_\gamma(K+1)} (J(u_\gamma^0) - J^*).$$

**Proof** (i) Choose  $\theta < \min \left\{ \frac{\tau \alpha(h)}{\bar{\tau}}, \frac{\sigma \alpha(\psi)}{\bar{\sigma}} \right\}$  and  $\delta_1, \delta_2$  as in (17), then (18) follows immediately from Lemma 3.3. Meanwhile,  $\Phi$  is assumed to be bounded below and  $J$  is also bounded below. Hence  $\{J(u_\gamma^k)\}_{k \in \mathbb{N}}$  converges to some real number  $J^*$ .

(ii) Using the inequality (18), we have for all  $k \geq 0$ ,

$$D_h(x^{k+1}, x^k) + D_\psi(y^{k+1}, y^k) \leq \frac{1}{\delta_\gamma} \left( J(u_\gamma^k) - J(u_\gamma^{k+1}) \right).$$

For  $K \geq 0$ , summing from  $k = 0$  to  $K$  and using the statement (i) yields

$$\begin{aligned} \sum_{k=0}^K \left( D_h(x^{k+1}, x^k) + D_\psi(y^{k+1}, y^k) \right) &\leq \frac{1}{\delta_{\mathcal{J}}} \left( J(u_{\mathcal{J}}^0) - J(u_{\mathcal{J}}^{K+1}) \right) \\ &\leq \frac{1}{\delta_{\mathcal{J}}} \left( J(u_{\mathcal{J}}^0) - J^* \right) < +\infty. \end{aligned} \tag{19}$$

Taking the limit as  $K \rightarrow +\infty$  leads to

$$\sum_{k=0}^{+\infty} \left( D_h(x^{k+1}, x^k) + D_\psi(y^{k+1}, y^k) \right) < +\infty,$$

which deduces that  $\lim_{k \rightarrow +\infty} \|z_{\mathcal{J}}^k - z_{\mathcal{J}}^{k+1}\| = 0$ . And from IASABP algorithm, we have

$$\lim_{k \rightarrow +\infty} \|z_{\mathcal{J}}^k - \bar{z}_{\mathcal{J}}^k\| = 0.$$

(iii) Using (19) yields

$$\sum_{k=0}^K \left( D_h(x^{k+1}, x^k) + D_\psi(y^{k+1}, y^k) \right) \leq \frac{1}{\delta_{\mathcal{J}}} \left( J(u_{\mathcal{J}}^0) - J^* \right).$$

Hence, we obtain

$$\min_{0 \leq k \leq K} \left\{ D_h(x^{k+1}, x^k) + D_\psi(y^{k+1}, y^k) \right\} \leq \frac{1}{\delta_{\mathcal{J}}(K+1)} \left( J(u_{\mathcal{J}}^0) - J^* \right).$$

□

Similar to Lemma 3.3 and Lemma 3.4, we have the following sublinear rate of convergence for ASABP.

**Lemma 3.5** *Suppose that Assumption 3.1 holds. Let  $\{(x^k, y^k)\}_{k \in \mathbb{N}}$  be the sequence generated by ASABP. Then*

(i) *The sequence  $\{\Phi(x^k, y^k)\}_{k \in \mathbb{N}}$  is monotonically nonincreasing and in particular for all  $k \geq 0$ ,*

$$\begin{aligned} \Phi(x^{k+1}, y^{k+1}) &\leq \Phi(x^k, y^k) - \frac{1}{\tau_k} D_h(x^k, x^{k+1}) - \frac{1}{\sigma_k} D_\psi(y^k, y^{k+1}) \\ &\leq \Phi(x^k, y^k) - \frac{1}{\bar{\tau}} D_h(x^k, x^{k+1}) - \frac{1}{\bar{\sigma}} D_\psi(y^k, y^{k+1}) \\ &\leq \Phi(x^k, y^k) - \delta \left( D_h(x^k, x^{k+1}) + D_\psi(y^k, y^{k+1}) \right), \end{aligned} \tag{20}$$

where  $\delta = \min\{\frac{1}{\bar{\tau}}, \frac{1}{\bar{\sigma}}\}$ . And  $\{\Phi(x^k, y^k)\}_{k \in \mathbb{N}}$  converges to some finite value, denoted by  $\Phi^*$ .

(ii) (Convergence rate) *For any  $K \geq 0$ , it holds that*

$$\min_{0 \leq k \leq K} \left\{ D_h(x^k, x^{k+1}) + D_\psi(y^k, y^{k+1}) \right\} \leq \frac{1}{\delta(K+1)} \left( \Phi(x^0, y^0) - \Phi^* \right).$$

### 3.3 Global convergence of ASABP

The analysis in Sect. 3.2 can only give the monotonically nonincreasing property of the potential function of ASABP and IASABP algorithms. In this subsection, without the Lipschitz continuity requirement on the gradients of  $F$  and  $G$ , we aim to prove that the sequence generated by ASABP algorithm is a Cauchy sequence, which hence converges to a critical point of model (1). It is worth noticing that all the NoLips algorithms for the nonconvex case [5, 12, 18, 25], the global convergence results were limited to the case that  $X = \mathbb{R}^n$ ,  $Y = \mathbb{R}^m$  and the Lipschitz smoothness condition holds.

We start our analysis by specializing the assumptions on the coupling term  $H$ .

**Assumption 3.2** (i) The subdifferential of  $H$  obeys

$$\partial_x H(x, y) \times \partial_y H(x, y) \subset \partial H(x, y), \quad \forall (x, y) \in \text{dom}H.$$

(ii) The domain of  $H$  is closed.

Moreover,  $H$  satisfies one of the following conditions.

(iii)  $H$  is continuous on its domain.

(iv)  $H : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  has the form

$$H(x, y) = \phi(x, y) + w(x),$$

where  $w : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is continuously differentiable on its domain;  $\phi : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  is continuous on  $\text{dom}H$ , such that  $\phi(\cdot, y)$  is differentiable on  $\mathbb{R}^n$  of gradient  $\nabla_x \phi$  for any  $y \in \mathcal{D}_y = \{y \in \mathbb{R}^m | (x, y) \in \text{dom}H\} \neq \emptyset$ . For each bounded subset  $\mathcal{B}_x \times \mathcal{B}_y \subset \text{dom}H$ , there exists  $\xi > 0$  such that, for any  $\bar{x} \in \mathcal{B}_x$  and any  $(y, \bar{y}) \in \mathcal{B}_y^2$ ,

$$\|\nabla_x \phi(\bar{x}, y) - \nabla_x \phi(\bar{x}, \bar{y})\| \leq \xi \|y - \bar{y}\|.$$

**Remark 3.6** We give some remarks on Assumption 3.2.

(i) Since  $F$  and  $G$  are continuous differentiable on  $\bar{X}$  and  $\bar{Y}$  respectively, Assumption 3.2 (i) is equivalent to having  $\partial_x \Phi(x, y) \times \partial_y \Phi(x, y) \subset \partial \Phi(x, y)$  for any  $(x, y) \in \bar{X} \times \bar{Y} \cap \text{dom}H$ , which is nonempty and closed.

(ii) Assumption 3.2 (iii) and (iv) aim at guaranteeing that the cluster points of the sequence generated by the algorithm are critical points. However, they are sufficient but not necessary, and the results presented in this paper can be established for a broader class of functions  $H$ .

(iii) Note that for Assumption 3.2 (iv), the regularity assumptions on  $H$  are not symmetric in  $x$  and  $y$ . In particular,  $\phi(x, \cdot)$  does not need to be differentiable.

**Theorem 3.7** Suppose that Assumptions 3.1 and 3.2 hold. If the sequence  $\{z^k = (x^k, y^k)\}_{k \in \mathbb{N}}$  generated by ASABP converges to  $\hat{z} = (\hat{x}, \hat{y})$ , then it is a critical point of the problem (1) with the abstract sets  $\bar{X}, \bar{Y}$ , i.e.,

$$\exists \hat{H} \in \partial H(\hat{z}), \quad 0 \in \hat{H} + \nabla F(\hat{x}) \times \nabla G(\hat{y}) + N_{\bar{X} \times \bar{Y}}(\hat{z}),$$

where  $N_{\bar{X} \times \bar{Y}}(\hat{z})$  denotes the normal cone of  $\bar{X} \times \bar{Y}$  at  $\hat{z}$ .

**Proof** The first order optimality condition of the subproblem (9) shows that there exists  $\hat{H}_x^{k+1} \in \partial_x H(x^{k+1}, y^k)$ ,

$$\left\langle x - x^{k+1}, \hat{H}_x^{k+1} + \nabla F(x^k) + \frac{1}{\tau_k} \left( \nabla h(x^{k+1}) - \nabla h(x^k) \right) \right\rangle \geq 0, \quad \forall x \in \bar{X}, \quad (21)$$

or equivalently,

$$0 \in \partial_x H(x^{k+1}, y^k) + \nabla F(x^k) + \frac{1}{\tau_k} \left( \nabla h(x^{k+1}) - \nabla h(x^k) \right) + N_{\bar{X}}(x^{k+1}).$$

We first investigate the behavior of the following term,

$$\langle x - x^{k+1}, \nabla h(x^{k+1}) - \nabla h(x^k) \rangle.$$

Fix some arbitrary  $x \in \bar{X}$ . We consider the following two possible cases.

The first case is that the sequence  $\{D_h(x, x^k)\}$  converges. For this case, the three point identity (see Remark 2.8) yields

$$\begin{aligned} & \lim_{k \rightarrow +\infty} \langle \nabla h(x^{k+1}) - \nabla h(x^k), x - x^{k+1} \rangle \\ &= \lim_{k \rightarrow +\infty} \left( D_h(x, x^k) - D_h(x, x^{k+1}) \right) - \lim_{k \rightarrow +\infty} D_h(x^{k+1}, x^k) \\ &= 0. \end{aligned}$$

Combining with the facts that  $H$  is biconvex and  $F$  is continuously differentiable on  $\bar{X}$ , passing to the limit in the optimality condition (21), we obtain

$$\langle x - \hat{x}, \partial_x H(\hat{z}) + \nabla F(\hat{x}) \rangle \geq 0, \quad \forall x \in \bar{X},$$

i.e.,

$$0 \in \partial_x H(\hat{z}) + \nabla F(\hat{x}) + N_{\bar{X}}(\hat{x}). \tag{22}$$

In particular, if  $x^k \rightarrow \hat{x} \in X$ , one has that  $\lim_{k \rightarrow +\infty} \|\nabla h(x^{k+1}) - \nabla h(x^k)\| = 0$  because of the continuity of  $\nabla h$  on  $X$ . The problem (1) in this case is equivalent to an unconstrained problem and (22) holds naturally. Certainly, the sequence  $\{D_h(x, x^k)\}$  must be convergent for this situation.

The second case is that the sequence  $\{D_h(x, x^k)\}$  fails to converge. For this case, since  $\{D_h(x, x^k)\}$  is nonnegative, it cannot be monotonically decreasing. Thus, there are infinitely many  $k \in \mathbb{N}$  such that  $D_h(x, x^{k+1}) \geq D_h(x, x^k)$ . Let  $\{k_l\} \subset \mathbb{N}$  be a subsequence with  $D_h(x, x^{k_l+1}) \geq D_h(x, x^{k_l})$  for arbitrary  $l$ . Then for all  $x \in \bar{X}$ ,

$$\begin{aligned} 0 &\geq \limsup_{l \rightarrow +\infty} \left( D_h(x, x^{k_l}) - D_h(x, x^{k_l+1}) \right) \\ &= \limsup_{l \rightarrow +\infty} \left\langle \nabla h(x^{k_l+1}) - \nabla h(x^{k_l}), x - x^{k_l+1} \right\rangle + \lim_{l \rightarrow +\infty} D_h(x^{k_l+1}, x^{k_l}) \\ &= \limsup_{l \rightarrow +\infty} \left\langle \nabla h(x^{k_l+1}) - \nabla h(x^{k_l}), x - x^{k_l+1} \right\rangle. \end{aligned}$$

This means that

$$\nabla h(x^{k_l+1}) - \nabla h(x^{k_l}) \in N_{\bar{X}}(x^{k_l+1}).$$

Hence,

$$\left( \nabla h(x^{k_l+1}) - \nabla h(x^{k_l}) \right) + N_{\bar{X}}(x^{k_l+1}) \subset N_{\bar{X}}(x^{k_l+1}).$$

Since the corresponding subsequence also converges to  $\hat{x}$ , passing to the limit  $l \rightarrow +\infty$  in the optimality condition,  $0 \in \partial_x H(\hat{z}) + \nabla F(\hat{x}) + N_{\bar{X}}(\hat{x})$  holds. Especially, if  $x^k \rightarrow \hat{x} \in \partial X$  (the boundary of  $X$ ), for any  $x \in X$ , the sequence  $\{D_h(x, x^k)\}$  fails to converge due to Remark 2.8 (iii). It provides a sufficient condition for this case.



Similarly, we have

$$0 \in \partial_y H(\hat{z}) + \nabla G(\hat{y}) + N_{\bar{Y}}(\hat{y}).$$

Combining with Assumption 3.2, we complete the proof. □

Denote the set of cluster points of the sequence  $\{z^k = (x^k, y^k)\}_{k \in \mathbb{N}}$  generated by ASABP by  $\Omega^*$ , i.e.,

$$\Omega^* = \{z^* = (x^*, y^*) : \exists \text{ an increasing sequence of integer } \{k_l\}_{l \in \mathbb{N}}, \text{ such that } z^{k_l} \rightarrow z^* \text{ as } l \rightarrow +\infty\}.$$

**Lemma 3.8** *Suppose that Assumptions 3.1 and 3.2 hold. Let  $\{z^k = (x^k, y^k)\}_{k \in \mathbb{N}}$  be the sequence generated by ASABP which is assumed to be bounded. Then the following assertions hold.*

(i)  $\Omega^*$  is a nonempty compact set, and

$$\lim_{k \rightarrow +\infty} \text{dist}(z^k, \Omega^*) = 0. \tag{23}$$

(ii)  $\lim_{k \rightarrow +\infty} \Phi(z^k) = \Phi(z^*)$  and  $\Phi$  is constant on  $\Omega^*$ .

**Proof** (i)  $\Omega^*$  is nonempty because the sequence  $\{z^k\}_{k \in \mathbb{N}}$  generated by ASABP is bounded. By definition, it is trivial that the set  $\Omega^*$  is compact. We now prove (23) by contradiction. Suppose that  $\{\text{dist}(z^k, \Omega^*)\}_{k \in \mathbb{N}}$  does not converge to zero. Then there exist  $C > 0$  and an increasing sequence  $\{k_j\}_{j \in \mathbb{N}}$  such that, for any  $j \in \mathbb{N}$ ,  $\text{dist}(z^{k_j}, \Omega^*) > C$ . However, since  $\{z^{k_j}\}_{j \in \mathbb{N}}$  is a subsequence of the bounded sequence  $\{z^k\}_{k \in \mathbb{N}}$ , it has a convergent subsequence  $\{z^{k_{j_n}}\}_{n \in \mathbb{N}}$  with limit  $z^* \in \Omega^*$ . Then,

$$C < \text{dist}(z^{k_{j_n}}, \Omega^*) \leq \|z^{k_{j_n}} - z^*\| \rightarrow 0,$$

which leads to a contradiction.

(ii) Taking an arbitrary point  $z^* \in \Omega^*$ , then there exists a subsequence  $\{z^{k_j}\}_{j \in \mathbb{N}}$  of  $\{z^k\}_{k \in \mathbb{N}}$  converging to  $z^*$ . Notice that  $(x^{k_j}, y^{k_j}) \in \bar{X} \times \bar{Y} \cap \text{dom}H$ , which is assumed to be closed. Hence,  $(x^*, y^*) \in \bar{X} \times \bar{Y} \cap \text{dom}H$ , and by continuity one has

$$\lim_{k \rightarrow +\infty} \Phi(z^{k_j}) = \Phi(z^*).$$

Since  $\{\Phi(z^k)\}_{k \in \mathbb{N}}$  is a convergent sequence by Lemma 3.5, this proves assertion (ii). □

To establish the convergence of the sequence generated by ASABP, we need some extra assumptions on the Bregman functions  $h$  and  $\psi$ , which are also required in [5, 12, 18, 25].

**Assumption 3.3** (i)  $h, \psi$  are strongly convex functions with coefficients  $\mu_h, \mu_\psi$  respectively.

(ii)  $\nabla h, \nabla \psi$  are Lipschitz continuous with coefficients  $L_1, L_2$  respectively on any bounded subsets.

**Remark 3.9** (i) It is worth noticing that Assumption 3.3 (i) is not actually stringent. For example, Burg’s entropy  $h(x) = -\log x$  with  $\text{dom}h = (0, +\infty)$  is a generalized Bregman function which is not strongly convex. However, regularized Burg’s entropy  $\hat{h}(x) = -\mu \log x + \frac{\sigma}{2}x^2$  with  $\sigma, \mu > 0$ , is strongly convex with coefficient  $\sigma$  on  $\text{dom}h = (0, +\infty)$ .

- (ii) Combining linear additivity of Bregman distance (see Remark 2.8), Assumption 3.1 (ii) and Assumption 3.3 (ii),  $\nabla F$  is Lipschitz continuous with coefficient  $L_h L_1$  on any bounded subset in  $\bar{X}$ . Similarly,  $\nabla G$  is Lipschitz continuous with coefficient  $L_\psi L_2$  on any bounded subset in  $\bar{Y}$ .

**Proposition 3.10** *Suppose that Assumptions 3.1, 3.2 with (iv) and Assumption 3.3 hold. Let  $\{z^k = (x^k, y^k)\}_{k \in \mathbb{N}}$  be the sequence generated by ASABP which is assumed to be bounded. For all  $k \geq 0$ , define*

$$A^{k+1} = (A_x^{k+1}, A_y^{k+1}),$$

where

$$\begin{cases} A_x^{k+1} = \nabla F(x^{k+1}) - \nabla F(x^k) - \frac{1}{\tau_k} (\nabla h(x^{k+1}) - \nabla h(x^k)) + \nabla_x \phi(x^{k+1}, y^{k+1}) - \nabla_x \phi(x^{k+1}, y^k) \\ \quad \in \nabla F(x^{k+1}) - \nabla F(x^k) - \frac{1}{\tau_k} (\nabla h(x^{k+1}) - \nabla h(x^k)) + \partial_x H(x^{k+1}, y^{k+1}) - \partial_x H(x^{k+1}, y^k), \\ A_y^{k+1} = \nabla G(y^{k+1}) - \nabla G(y^k) - \frac{1}{\sigma_k} (\nabla \psi(y^{k+1}) - \nabla \psi(y^k)). \end{cases}$$

Then  $A^{k+1} \in \partial \Phi(z^{k+1})$  and there exists  $\hat{\delta} > 0$  such that

$$\|A^{k+1}\| \leq \hat{\delta} \|z^k - z^{k+1}\|,$$

where  $\hat{\delta} = \sqrt{2} \max \left\{ L_h L_1 + \frac{L_1}{\tau}, L_\psi L_2 + \frac{L_2}{\sigma} + \xi \right\}$ .

**Proof** The first order optimality conditions of (9) and (10) yield

$$0 \in \partial_x H(x^{k+1}, y^k) + \nabla F(x^k) + \frac{1}{\tau_k} (\nabla h(x^{k+1}) - \nabla h(x^k)),$$

and

$$0 \in \partial_y H(x^{k+1}, y^{k+1}) + \nabla G(y^k) + \frac{1}{\sigma_k} (\nabla \psi(y^{k+1}) - \nabla \psi(y^k)),$$

respectively. Following from these relations and Assumption 3.2, it is clear that  $A^{k+1} \in \partial \Phi(z^{k+1})$ . Combining with Assumption 3.3, we get that

$$\begin{aligned} \|A_x^{k+1}\| &\leq \left\| \nabla F(x^{k+1}) - \nabla F(x^k) - \frac{1}{\tau_k} (\nabla h(x^{k+1}) - \nabla h(x^k)) \right\| + \xi \|y^{k+1} - y^k\| \\ &\leq \left( L_h L_1 + \frac{L_1}{\tau} \right) \|x^{k+1} - x^k\| + \xi \|y^{k+1} - y^k\|. \end{aligned}$$

Similarly, we can get

$$\begin{aligned} \|A_y^{k+1}\| &= \left\| \nabla G(y^{k+1}) - \nabla G(y^k) - \frac{1}{\sigma_k} (\nabla \psi(y^{k+1}) - \nabla \psi(y^k)) \right\| \\ &\leq \left( L_\psi L_2 + \frac{L_2}{\sigma} \right) \|y^{k+1} - y^k\|. \end{aligned}$$

Rearranging the terms and completing the proof. □

We are now ready for proving the main result of this section.

**Theorem 3.11** *Let the objective function  $\Phi$  be a KL function. Suppose that Assumptions 3.1, 3.2 with (iv) and Assumption 3.3 hold. Let  $\{z^k = (x^k, y^k)\}_{k \in \mathbb{N}}$  be the sequence generated by ASABP which is assumed to be bounded. Then*

$$\sum_{k=0}^{+\infty} \|z^{k+1} - z^k\| < +\infty,$$

and as a sequence, we have  $\{z^k\}_{k \in \mathbb{N}}$  converges to a critical point of model (1) with the abstract sets  $\bar{X}, \bar{Y}$ .

**Proof** From Lemma 3.8, we get that

$$\lim_{k \rightarrow +\infty} \Phi(z^k) = \Phi(z^*).$$

Then we consider the following two cases.

First, if there exists an integer  $k'$  for which  $\Phi(z^{k'}) = \Phi(z^*)$ , according to the strongly convexity of  $h$  and  $\psi$  and (20), we have that for any  $k > k'$ ,

$$\begin{aligned} \bar{\delta} \|z^k - z^{k+1}\|^2 &\leq \delta \left( D_h(x^{k+1}, x^k) + D_\psi(y^{k+1}, y^k) \right) \\ &\leq \Phi(z^k) - \Phi(z^{k+1}), \end{aligned} \tag{24}$$

where  $\bar{\delta} = \delta \min \left\{ \frac{\mu_h}{2}, \frac{\mu_\psi}{2} \right\}$ . Combining with the monotonically nonincreasing of  $\Phi(z^k)$ , one can obtain that  $z^{k+1} = z^k$  for any  $k \geq k'$ , and the assertion holds.

Then we consider the case that  $\Phi(z^k) > \Phi(z^*)$  for all  $k \geq 0$ . Since  $\lim_{k \rightarrow +\infty} \Phi(z^k) = \Phi(z^*)$ , it follows that for any  $\eta > 0$ , there exists a nonnegative integer  $k_0$  such that  $\Phi(z^k) < \Phi(z^*) + \eta$  for all  $k \geq k_0$ . From (23), we know that  $\lim_{k \rightarrow +\infty} \text{dist}(z^k, \Omega^*) = 0$ . This means that for any  $\varepsilon > 0$ , there exists a positive integer  $k_1$  such that  $\text{dist}(z^k, \Omega^*) < \varepsilon$  for all  $k \geq k_1$ . Consequently, for all  $k > l := \max\{k_0, k_1\}$ ,

$$\text{dist}(z^k, \Omega^*) < \varepsilon, \quad \Phi(z^*) < \Phi(z^k) < \Phi(z^*) + \eta.$$

Applying Lemma 2.16 and Lemma 3.8, we deduce that for any  $k > l$ ,

$$\varphi' \left( \Phi(z^k) - \Phi(z^*) \right) \text{dist}(0, \partial \Phi(z^k)) \geq 1.$$

From Proposition 3.10,

$$\varphi' \left( \Phi(z^k) - \Phi(z^*) \right) \geq \frac{1}{\hat{\delta} \|z^k - z^{k-1}\|}. \tag{25}$$

On the other hand, from the concavity of  $\varphi$ , we get that

$$\begin{aligned} &\varphi \left( \Phi(z^k) - \Phi(z^*) \right) - \varphi \left( \Phi(z^{k+1}) - \Phi(z^*) \right) \\ &\geq \varphi' \left( \Phi(z^k) - \Phi(z^*) \right) \left( \Phi(z^k) - \Phi(z^{k+1}) \right). \end{aligned} \tag{26}$$

For convenience, we define

$$\Delta_k = \varphi \left( \Phi(z^k) - \Phi(z^*) \right) - \varphi \left( \Phi(z^{k+1}) - \Phi(z^*) \right).$$

Combining (24) with (25) and (26) yields that, for any  $k > l$ ,

$$\Delta_k \geq \frac{\bar{\delta} \|z^k - z^{k+1}\|^2}{\hat{\delta} \|z^k - z^{k-1}\|},$$

or equivalently,

$$\|z^k - z^{k+1}\| \leq \sqrt{\frac{\hat{\delta}}{\delta} \Delta_k} \|z^k - z^{k-1}\|^{\frac{1}{2}}.$$

Using the fact that  $2\sqrt{\alpha\beta} \leq \alpha + \beta$  for all  $\alpha, \beta > 0$ , we have

$$2\|z^k - z^{k+1}\| \leq \|z^k - z^{k-1}\| + \frac{\hat{\delta}}{\delta} \Delta_k. \tag{27}$$

Summing up (27) for  $k = l + 1, \dots, m$  yields

$$\begin{aligned} 2 \sum_{k=l+1}^m \|z^k - z^{k+1}\| &\leq \sum_{k=l+1}^m \|z^{k-1} - z^k\| + \sum_{k=l+1}^m \frac{\hat{\delta}}{\delta} \Delta_k \\ &= \sum_{k=l+1}^m \|z^{k-1} - z^k\| + \frac{\hat{\delta}}{\delta} \left( \varphi \left( \Phi(z^{l+1}) - \Phi(z^*) \right) - \varphi \left( \Phi(z^{m+1}) - \Phi(z^*) \right) \right) \\ &\leq \sum_{k=l+1}^m \|z^{k-1} - z^k\| + \frac{\hat{\delta}}{\delta} \varphi \left( \Phi(z^{l+1}) - \Phi(z^*) \right), \end{aligned}$$

where the last inequality due to  $\varphi(\Phi(z^{m+1}) - \Phi(z^*)) \geq 0$ . Letting  $m \rightarrow +\infty$  yield

$$\sum_{k=l+1}^{+\infty} \|z^{k+1} - z^k\| \leq \|z^{l+1} - z^l\| + \frac{\hat{\delta}}{\delta} \varphi \left( \Phi(z^{l+1}) - \Phi(z^*) \right).$$

Equipped with the fact that  $\bar{\delta}$  is bounded below, we have

$$\sum_{k=0}^{+\infty} \|z^{k+1} - z^k\| < +\infty,$$

and the sequence  $\{z^k\}_{k \in \mathbb{N}}$  is a Cauchy sequence. Combining with Theorem 3.7,  $\{z^k\}_{k \in \mathbb{N}}$  converges to a critical point of model (1) with the abstract sets  $\bar{X}, \bar{Y}$ . □

### 4 Numerical experiment

In this section, we numerically verify the ability of ASABP and IASABP analyzed in previous sections. All experiments are performed in MATLAB R2014a on a 64-bit PC with an Intel Core i7-8550U CPU (1.80GHz) and 16GB of RAM.

We consider the Poisson linear inverse problems [6], which can be conveniently described as follows. Given a matrix  $A \in \mathbb{R}_+^{m \times n}$  modeling the experimental protocol, and  $b \in \mathbb{R}_+^m$  the vector of measurement, the goal is to reconstruct the signal or image  $x \in \mathbb{R}_+^n$  from the noisy measurement  $b$  such that  $Ax \simeq b$ . Often these problems can be represented as a minimize problem like

$$\min\{d(b, Ax) + \lambda g(x) : x \in \mathbb{R}_+^n\}, \tag{28}$$

where  $\lambda > 0$  is used to control the trade-off between matching the data fidelity criteria and the weight given to its regularizer, and  $d(\cdot, \cdot)$  denotes a convex proximity measure between two vectors. Poisson linear inverse problems emerged in many fields, like astronomy, nuclear medicine, electronic microscopy, statistical and image science [6, 8, 16]. Therefore, the design of methods for such problems has been studied intensively.

By introducing an auxiliary variable  $y \in \mathbb{R}^n$ , we can solve (28) approximately according to the following optimization problem

$$\min_{x \in \mathbb{R}^n_+, y \in \mathbb{R}^n} F(x) + G(y) + H(x, y), \tag{29}$$

by defining

$$F(x) = d(b, Ax), \quad G(y) = \lambda g(y), \quad H(x, y) = \frac{\mu}{2} \|x - y\|^2,$$

where  $\mu$  is a positive penalization parameter.

Adopting the above model (29), we take the first component in the objective function based on the Boltzmann-Shannon entropy,

$$d(b, Ax) = \sum_{i=1}^m \{(Ax)_i - b_i \log(Ax)_i\}.$$

It is easy to find that  $F(x) = d(b, Ax)$  has no globally Lipschitz continuous gradient [6], but satisfies GL-samd condition with a generalized Bregman function called Burg’s entropy, denoted as

$$h(x) = - \sum_{j=1}^n \log x_j, \quad \text{dom}h = \mathbb{R}^n_{++},$$

and so now the Bregman distance is given by

$$D_h(x, z) = \sum_{j=1}^n \left\{ \frac{x_j}{z_j} - \log \left( \frac{x_j}{z_j} \right) - 1 \right\},$$

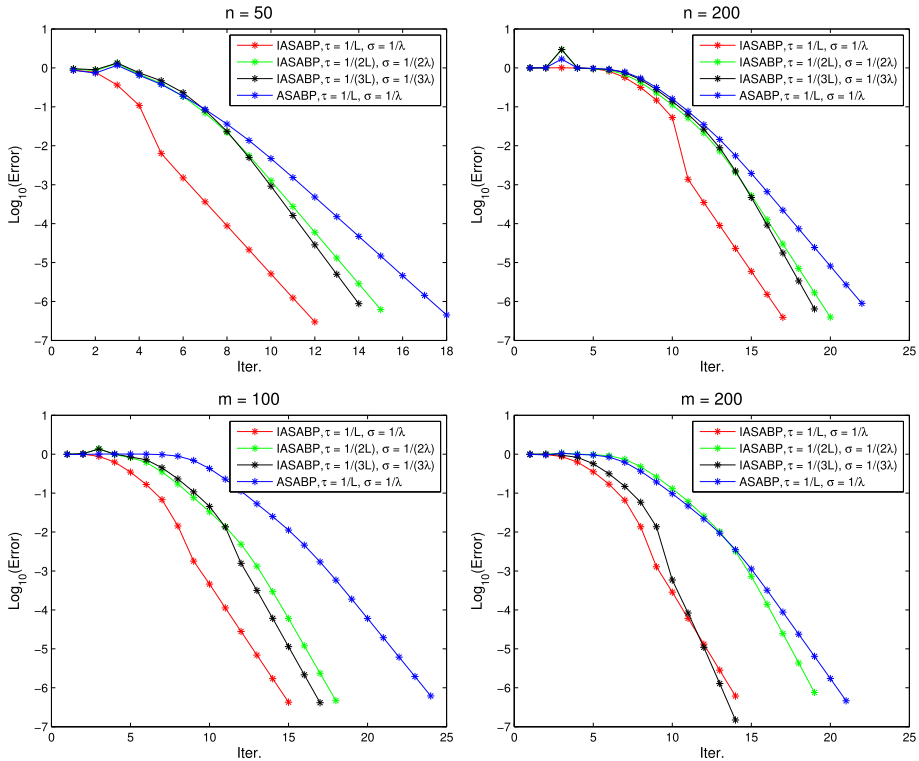
and for any  $L_h$  satisfying  $L_h \geq \|b\|_1 = \sum_{i=1}^m b_i$ , the function  $L_h h - F$  is convex on  $\mathbb{R}^n_+$ . In this case, the iterates in subproblem (9) and (12) reduce to solve  $n$  one-dimensional convex problems. Note that Assumption 3.3 does not hold for this case.

We consider Tikhonov regularization in model (29), i.e., the regularizer is  $G(y) = \lambda g(y) = \frac{\lambda}{2} \|y\|^2$ . Take Energy  $\psi = \frac{1}{2} \|y\|^2$ , which corresponding Bregman distance is defined as  $D_\psi(y, z) = \frac{1}{2} \|y - z\|^2$ , and for any  $L_\psi$  satisfying  $L_\psi \geq \lambda$ , the function  $L_\psi \psi - G$  is convex on  $\mathbb{R}^n$ . It is worth noticing that the objective function in model (29) in this case is KL function according to [3, 11].

Now we give the setting for the model to be tested. For the model (29), the entries of  $A \in \mathbb{R}^{m \times n}_+$  and  $b \in \mathbb{R}^m_+$  are generated following independent uniform distributions over the interval  $[0, 1]$ ; we set  $\lambda = 1$  and the penalization parameter  $\mu$  as large as the conditions permit. We initialize all the algorithms at the vector generated by uniform distribution over the interval  $(0, 1)$ , and terminate it when  $Error := \frac{\|x^k - x^{k-1}\|}{\max\{1, \|x^k\|\}} \leq 10^{-6}$ . We also terminate the algorithms when the number of iteration hits 100.

As these methods can be applied to overdetermined ( $m > n$ ) and underdetermined ( $m < n$ ) problems, we have preformed numerical tests on both cases. As commented before, the main parameters in IASABP are setting as follows:  $\eta$  is a random number between 0.90 and 0.95 and  $\theta = 0.99$ ;  $\alpha_k^0 = \beta_k^0 = 1$  for all  $k$ ;  $L_h = \|b\|_1$  and  $L_\psi = \lambda$ .

For this Poisson linear inverse problem, Fig. 1 records the trend of the residual functions of ASABP and IASABP with respect to the overdetermined and underdetermined situations. We can see that IASABP converges faster than ASABP. It shows that the IASABP algorithm has an advantage over ASABP, giving an interesting option for real problems.



**Fig. 1** Poisson linear inverse problem. First row:  $m = 1000$ ; Second row:  $n = 1000$

### 5 Conclusions

We have proposed an alternating structure-adapted Bregman proximal (ASABP) gradient descent algorithm for solving the nonconvex nonsmooth minimization problems with abstract sets. An improved version with inertial accelerated scheme and line search strategy (IASABP) was also presented. Under mild conditions, we proved that both ASABP and IASABP possess an  $O(1/K)$  sublinear rate of convergence, measured by Bregman distance. Furthermore, if the objective function has Kurdyka–Łojasiewicz property, we proved that the sequence generated by ASABP converges to the critical point of the constrained problem. Theoretically, our results are extension of the recent NoLips algorithms, which are either limited to the convex case, or the nonconvex case but there is no abstract sets. Numerically, our results avoid the estimation of the Lipschitz constants, an annoying task in implementation. We applied ASABP and IASABP to Poisson linear inverse problem, and their performance is very promising.

**Acknowledgements** The authors are grateful to the editor and the anonymous reviewer for their valuable comments and suggestions toward the improvement of this paper. This work was supported by National Natural Science Foundation of China (12201173, 12131004, 12126603).

**Data availability** All data generated or analysed during this study are included in the paper.

## References

1. Ahoosh, M., Hien, L., Gillis, N., Patrinos, P.: A block inertial Bregman proximal algorithm for nonsmooth nonconvex problems with application to symmetric nonnegative matrix tri-factorization. *J. Optim. Theory Appl.* **190**, 234–258 (2021)
2. Ahoosh, M., Hien, L., Gillis, N., Patrinos, P.: Multi-block Bregman proximal alternating linearized minimization and its application to orthogonal nonnegative matrix factorization. *Comput. Optim. Appl.* **79**, 681–775 (2021)
3. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality. *Math. Oper. Res.* **35**, 438–457 (2010)
4. Attouch, H., Bolte, J., Svaiter, B.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized gauss-seidel methods. *Math. Program.* **137**, 91–129 (2013)
5. Bauschke, H., Bolte, J., Chen, J., Teboulle, M., Wang, X.: On linear convergence of non-Euclidean gradient methods without strong convexity and Lipschitz gradient continuity. *J. Optim. Theory Appl.* **182**, 1068–1087 (2019)
6. Bauschke, H., Bolte, J., Teboulle, M.: A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications. *Math. Oper. Res.* **42**, 330–348 (2016)
7. Benning, M., Betcke, M., Ehrhardt, M., Schonlieb, C.: Choose your path wisely: gradient descent in a Bregman distance framework. *SIAM J. Imaging Sci.* **14**, 814–843 (2021)
8. Bertero, M., Boccacci, P., Desiderà, G., Vicidomini, G.: Image deblurring with Poisson data: from cells to galaxies. *Inverse Probl.* **25**, 123006 (2009)
9. Bertsekas, D.: Nonlinear programming. *J. Oper. Res. Soc.* **48**, 334–334 (1997)
10. Bertsekas, D., Tsitsiklis, J.: Parallel and Distributed Computation: Numerical Methods. Athena Scientific (2015)
11. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.* **146**, 459–494 (2014)
12. Bolte, J., Sabach, S., Teboulle, M., Vaisbourd, Y.: First order methods beyond convexity and Lipschitz gradient continuity with applications to quadratic inverse problems. *SIAM J. Optim.* **28**, 2131–2151 (2018)
13. Censor, Y., Iusem, A., Zenios, S.: An interior point method with Bregman functions for the variational inequality problem with paramonotone operators. *Math. Program.* **81**, 373–400 (1998)
14. Gao, X., Cai, X., Han, D.: A Gauss-Seidel type inertial proximal alternating linearized minimization for a class of nonconvex optimization problems. *J. Global Optim.* **76**, 863–887 (2020)
15. Han, D.: A new hybrid generalized proximal point algorithm for variational inequality problems. *J. Global Optim.* **26**, 125–140 (2003)
16. Hohage, T., Werner, F.: Inverse problems with Poisson data: statistical regularization theory, applications and algorithms. *Inverse Probl.* **32**, 093001 (2016)
17. Lu, H., Freund, R., Nesterov, Y.: Relatively smooth convex optimization by first-order methods, and applications. *SIAM J. Optim.* **28**, 333–354 (2018)
18. Mukkamala, M., Ochs, P., Pock, T., Sabach, S.: Convex-concave backtracking for inertial Bregman proximal gradient algorithms in nonconvex optimization. *SIAM J. Math. Data Sci.* **2**, 658–682 (2020)
19. Nesterov, Y.: Introductory Lectures on Convex Optimization: A Basic Course. Springer (2003)
20. Nikolova, M., Tan, P.: Alternating structure-adapted proximal gradient descent for nonconvex nonsmooth block-regularized problems. *SIAM J. Optim.* **29**, 2053–2078 (2019)
21. Pock, T., Sabach, S.: Inertial proximal alternating linearized minimization (iPALM) for nonconvex and nonsmooth problems. *SIAM J. Imaging Sci.* **9**, 1756–1787 (2016)
22. Polyak, B.: Some methods of speeding up the convergence of iteration methods. *USSR Comput. Maths. Math. Phys.* **4**, 1–17 (1964)
23. Rockafellar, R.: Convex Analysis. Princeton University Press, Princeton (1970)
24. Rockafellar, R., Wets, R.: Variational Analysis. Springer (2009)
25. Teboulle, M.: A simplified view of first order methods for optimization. *Math. Program.* **170**, 67–96 (2018)
26. Wu, Z., Li, C., Li, M., Lim, A.: Inertial proximal gradient methods with bregman regularization for a class of nonconvex optimization problems. *J. Global Optim.* **79**, 617–644 (2021)
27. Xu, Y., Yin, W.: A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM J. Imaging Sci.* **6**, 1758–1789 (2013)

28. Zhang, H., Dai, Y., Guo, L., Peng, W.: Proximal-like incremental aggregated gradient method with linear convergence under Bregman distance growth conditions. *Math. Oper. Res.* **46**, 61–81 (2021)
29. Zhang, X., Zhang, X., Li, X., Li, Z., Wang, S.: Classify social image by integrating multi-modal content. *Multimed. Tools. Appl.* **77**, 7469–7485 (2018)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.