QUASI-SUBGRADIENT METHODS WITH BREGMAN DISTANCE FOR QUASI-CONVEX FEASIBILITY PROBLEMS

YAOHUA HU[1], JINGCHAO LI[1], YANYAN LIU[1], CARISA KWOK WAI YU[2,*]

[1] *School of Mathematical Sciences, Shenzhen University, Shenzhen 518060, China.*
[2] *Department of Mathematics, Statistics and Insurance, The Hang Seng University of Hong Kong, Shatin, Hong Kong.*

Abstract. In this paper, we consider the quasi-convex feasibility problem (QFP), which is to find a common point of a family of sublevel sets of quasi-convex functions. By employing the Bregman projection mapping, we propose a unified framework of Bregman quasi-subgradient methods for solving the QFP. This paper is contributed to establish the convergence theory, including the global convergence, iteration complexity and convergence rates, of the Bregman quasi-subgradient methods with several general control schemes, including the α-most violated constraints control and the s-intermittent control. Moreover, we introduce a notion of the Hölder-type bounded error bound property relative to the Bregman distance for the QFP, and use it to establish the linear (or sublinear) convergence rates for Bregman quasi-subgradient methods to a feasible solution of the QFP.

Keywords. Quasi-convex feasibility problem; Subgradient method; Bregman distance; Convergence analysis.

1. Introduction

The feasibility problem is to find a point $x \in \mathbb{R}^n$ such that

$$x \in C \quad \text{and} \quad f_i(x) \leq 0 \quad \text{for all } i = 1, \ldots, m, \tag{1.1}$$

where $\{f_i : i = 1, \ldots, m\}$ is a family of continuous functions on $\mathbb{R}^n$ and $C \subseteq \mathbb{R}^n$ is a closed and convex set. The feasibility problem is at the core of the modeling of many problems in various areas of mathematics and physical sciences, such as image recovery [15], radiation therapy treatment planning [11], wireless sensor networks localization [17, 18], and gene regulatory network inference [30, 31].

Problem (1.1) is called the convex feasibility problem (CFP) when the involved functions are convex, which has attracted a great deal of attention in various application fields. However, the convex function is too restrictive and not accurate enough to characterize many real-life problems encountered in economics, finance and management science. In contrast, the quasi-convex function usually provides a much more accurate representation of reality in economics and finance and still possesses certain desirable properties of the convex function. In recent decades, much attention has been drawn to quasi-convex optimization [4, 16, 32] and quasi-convex feasibility problem (QFP) [13, 20, 27], in which the functions involved in (1.1) are quasi-convex.

*Corresponding author.

E-mail addresses: mayhhu@szu.edu.cn (Y. Hu), jingchaoli@szu.edu.cn (J. Li), 2100201002@email.szu.edu.cn (Y. Liu), carisayu@hsu.edu.hk (C. K. W. Yu).

Received April 12, 2023; Accepted × ×, 202×.

Motivated by its extensive applications, tremendous efforts have been devoted to the development of optimization algorithms for solving the feasibility problem (1.1); see [5, 10, 15, 34] and references therein. One of the most popular approaches is the class of projected subgradient methods, which was originally proposed by Censor and Lent [12] with a cyclic control scheme. Many extensions of projected subgradient methods have been proposed by employing several control schemes, and various convergence properties of subgradient methods for solving the CFP [5, 34] and the QFP [13, 19, 20] have been well explored, including the global convergence, iteration complexity and linear convergence rate to a feasible solution.

Unfortunately, the projected subgradient method suffers from several disadvantages arising from the Euclidean projection; see, e.g., [1, 2, 9]. In particular, the Euclidean projection destroys the nice descent property and might often lead to a zig-zagging effect, resulting in the slow convergence. Moreover, the Euclidean projection could be computationally expensive if the feasible set C is not simple. To avoid these disadvantages arising from the Euclidean projection, one popular approach is to replace the Euclidean projection by the Bregman projection, which is a proximal mapping on the subgradient-linearized function at current iterate with the Bregman distance in place of the Euclidean distance.

The Bregman subgradient method is an extension of projected subgradient method, in which the Bregman projection is adopted in place of the Euclidean projection in the projected subgradient method. The history of the Bregman subgradient method originates in 1983 from the mirror descent method proposed by Nemirovsky and Yudin [28], which can viewed as a Bregman subgradient method with the Kullback-Leibler divergence [6]. Moreover, the Bregman subgradient method enjoys several advantages: (i) it requires only first-order information, (ii) for certain types of constraints and suitable Bregman distance, it generates simple iterative schemes, and (iii) it exhibits a nearly dimension independent computational complexity in terms of the problem's dimension; see, e.g., [1, 2, 6, 7]. Motivated by these advantages of Bregman subgradient methods, their convergence theory and iteration complexity have been well studied for constrained convex and quasi-convex optimization problems; see [1, 2, 9] and references therein.

In this paper, inspired by the idea of Bregman projection, we propose the Bregman quasi-subgradient methods for solving the QFP (1.1) in a unified framework (see Algorithm 3.1), which covers most types of control schemes discussed in the literature. The main contribution of the present paper is to establish the convergence theory, including the global convergence, iteration complexity and convergence rates, of Bregman quasi-subgradient methods with several general control schemes for solving the QFP. In particular, the α-most violated constraints control and the s-intermittent control are considered in this paper. In convergence analysis, we first establish the global convergence of Bregman quasi-subgradient methods to a feasible solution of the QFP; see Theorems 3.1 and 3.4. Furthermore, we derive their (worst-case) iteration complexity to obtain an approximate feasible solution; see Theorems 3.2 and 3.5. More importantly, we introduce a notion of the Hölder-type bounded error bound property relative to the Bregman distance for the QFP and use it to explore the linear (or sublinear) convergence rates of Bregman quasi-subgradient methods to a feasible solution of the QFP; see Theorems 3.3 and 3.6.

The present paper is organized as follows. In Section 2, we present the notations and some preliminary lemmas which will be used in this paper. In Section 3, we provide a unified framework of Bregman quasi-subgradient methods with general control schemes to solve the QFP and establish the convergence theory.

2. Notations and preliminary results

Notations used in the present paper are standard in the n-dimensional Euclidean space $\mathbb{R}^n$ with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$. For $x\in\mathbb{R}^n$ and $r>0$, we use $\mathbb{B}(x,r)$ to denote the closed ball centered at x with radius r, and use $\mathbb{S}$ to denote the unit sphere centered at the origin. For a convex set $Z\subseteq\mathbb{R}^n$ and $x\in Z$, the normal cone of Z at x is defined by

$$\mathrm{N}_Z(x):=\{y\in\mathbb{R}^n:\langle y,z-x\rangle\le 0 \text{ for any } z\in Z\}.$$

As usual, we use $\mathbb{R}^m_+$ and $\mathbb{R}^m_{++}$ to denote the nonnegative orthant and the positive orthant of $\mathbb{R}^m$, respectively. The positive simplex in $\mathbb{R}^m$ is denoted by Δ^m_+, that is,

$$\Delta^m_+:=\{\lambda\in\mathbb{R}^m_{++}:\sum_{i=1}^m\lambda_i=1\}.$$

Moreover, we use the notation that $a^+:=\max\{a,0\}$ for any $a\in\mathbb{R}$, define the positive part function of $f:\mathbb{R}^n\to\mathbb{R}$ by

$$f^+(x):=\max\{f(x),0\}\quad\text{for any } x\in\mathbb{R}^n,$$

and adopt the convention that $\frac{0}{0}=0$ and $\cup_{i\in\emptyset}I_i=\emptyset$ for any family of index sets $\{I_i\}$.

A function $f:\mathbb{R}^n\to\mathbb{R}$ is said to be convex, σ-strongly convex and quasi-convex if

$$f(\alpha x+(1-\alpha)y)\le\alpha f(x)+(1-\alpha)f(y),$$
$$f(\alpha x+(1-\alpha)y)\le\alpha f(x)+(1-\alpha)f(y)-\frac{\sigma}{2}\alpha(1-\alpha)\|x-y\|^2,\ \text{and}$$
$$f(\alpha x+(1-\alpha)y)\le\max\{f(x),f(y)\},$$

respectively, for any $x,y\in\mathbb{R}^n$ and $\alpha\in[0,1]$. The sublevel sets of f at x are denoted by

$$\mathrm{lev}^{<}_f(x):=\{y\in\mathbb{R}^n:f(y)<f(x)\}\quad\text{and}\quad\mathrm{lev}^{\le}_f(x):=\{y\in\mathbb{R}^n:f(y)\le f(x)\}.$$

A convex function can be characterized by the convexity of its epigraph, while the geometrical interpretation for a quasi-convex function is characterized by the convexity of its sublevel sets.

Proposition 2.1. *$f:\mathbb{R}^n\to\mathbb{R}$ is quasi-convex if and only if $\mathrm{lev}^{<}_f(x)$ (and/or $\mathrm{lev}^{\le}_f(x)$) is convex for each $x\in\mathbb{R}^n$.*

2.1. Quasi-subdifferential.

The subdifferential of a quasi-convex function plays an important role in quasi-convex optimization. Several specific types of subdifferentials have been introduced and explored for quasi-convex functions that are defined via the "normal cone" to the level sets; see [3, 20] and references therein. In particular, Kiwiel [25], Censor and Segal [13], and Hu et al. [21, 22] utilized a quasi-subgradient for developing and analyzing quasi-subgradient methods.

Definition 2.1. Let $h:\mathbb{R}^n\to\mathbb{R}$ be a quasi-convex function, and let $x\in\mathbb{R}^n$. The quasi-subdifferential of h at x is defined by

$$\partial^{\mathrm{Q}}h(x):=\mathrm{N}_{\mathrm{lev}^{<}_h(x)}(x)=\{g:\langle g,y-x\rangle\le 0\text{ for any } y\in\mathrm{lev}^{<}_h(x)\}.$$

It was shown in [22, Lemma 2.1] that each quasi-convex and upper semicontinuous function has a nontrivial quasi-convex subdifferential; particularly, $\partial^Q f(x)$ contains at least a unit vector since it is a normal cone to its sublevel set. From Definition 2.1, the quasi-subgradient is not easy to calculate via estimating a normal vector to the level set. Alternatively, [9, Proposition 3] provides a practical approach for calculating a quasi-subgradient by computing the gradient at a differentiable point, or the limit of gradients close to a nondifferentiable point.

The notion of Hölder condition has been widely used in economics and management science. In particular, the Hölder condition of order 1 is reduced to the Lipschitz condition.

Definition 2.2. Let $0<\beta\le 1$ and $L>0$. The function $f:\mathbb{R}^n\to\mathbb{R}$ is said to satisfy the Hölder condition of order β with modulus L on $\mathbb{R}^n$ if

$$|f(x)-f(y)|\le L\|x-y\|^{\beta}\quad \text{for any } x,y\in\mathbb{R}^n.$$

The Hölder condition was used to provide the following fundamental property of the quasi-subgradient in [26, Proposition 2.1] and [20, Lemma 2.1], which plays an important role in the establishment of a basic inequality in convergence analysis of subgradient-type algorithms for quasi-convex optimization; see, e.g., [21, 22, 23].

Lemma 2.1. *Let $f:\mathbb{R}^n\to\mathbb{R}$ be a quasi-convex and continuous function, and C be a closed and convex set, and let $S:=\{x\in C: f(x)\le 0\}$. Let $0<\beta\le 1$ and $L>0$, and suppose that f satisfies the Hölder condition of order β with modulus L on $\mathbb{R}^n$. Then for any $x\in S$ and $y\in C\setminus S$, it holds that*

$$f(y)\le L\langle g,y-x\rangle^{\beta}\quad \textit{for each } g\in\partial^Q f(y)\cap\mathbb{S}.$$

The following two lemmas are useful in the convergence analysis of subgradient methods, which are taken from [24, Lemma 4.1] and [29, pp. 46, Lemma 6], respectively.

Lemma 2.2. *Let $a_i\ge 0$ for $i=1,2,\ldots,n$. Then the following assertions are true.*

(i) *If $\gamma\in(0,1]$, then*

$$\left(\sum_{i=1}^n a_i\right)^{\gamma}\le\sum_{i=1}^n a_i^{\gamma}\le n\left(\sum_{i=1}^n a_i\right)^{\gamma}.$$

(ii) *If $\gamma\in[1,\infty)$, then*

$$\frac{1}{n^{\gamma-1}}\left(\sum_{i=1}^n a_i\right)^{\gamma}\le\sum_{i=1}^n a_i^{\gamma}\le\left(\sum_{i=1}^n a_i\right)^{\gamma}.$$

Lemma 2.3. *Let $r>0$ and $b>0$, and $\{u_k\}\subseteq\mathbb{R}_+$ be a sequence of nonnegative scalars such that*

$$u_{k+1}\le u_k-bu_k^{1+r}\quad \textit{for each } k\in\mathbb{N}.$$

Then it holds that

$$u_k\le u_0\left(1+ru_0^r bk\right)^{-\frac{1}{r}}\quad \textit{for each } k\in\mathbb{N}.$$

2.2. **Bregman distance.** Bregman distance is a type of non-Euclidean distance-like functions, which has been widely used in the type of Bregman subgradient methods [1, 2, 9]. Let $\varphi:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ be a Legendre function and $C\subseteq\mathbb{R}^n$ be a closed and convex set with nonempty interior, satisfying the following conditions.

(a) φ is proper, lower semicontinuous and convex with $\operatorname{dom}\varphi\subseteq C$ and $\operatorname{dom}\nabla\varphi=\operatorname{int}C$.

(b) φ is σ-strongly convex and continuous on $\operatorname{dom}\varphi$, and continuously differentiable on $\operatorname{int}C$.

Associated to the kernel φ, the Bregman distance $\mathscr{D}_\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ is defined by

$$\mathscr{D}_\varphi(x,y) := \begin{cases} \varphi(x) - \varphi(y) - \langle \nabla\varphi(y), x-y \rangle, & \forall x \in C,\, y \in \operatorname{int}C, \\ +\infty, & \text{otherwise}. \end{cases} \tag{2.1}$$

By the strong convexity of kernel φ, one has that

$$\mathscr{D}_\varphi(x,y) \ge \frac{\sigma}{2}\|x-y\|^2 \quad \text{for any } x \in C,\, y \in \operatorname{int}C. \tag{2.2}$$

Moreover, Bregman distance $\mathscr{D}_\varphi$ enjoys a remarkable three point identity [14, Lemma 3.1] that

$$\mathscr{D}_\varphi(z,y) + \mathscr{D}_\varphi(y,x) - \mathscr{D}_\varphi(z,x) = \langle \nabla\varphi(x) - \nabla\varphi(y), z-y \rangle. \tag{2.3}$$

The Bregman distance from a point $x \in \mathbb{R}^n$ to a set $Z \subseteq \mathbb{R}^n$ is defined by

$$\mathscr{D}_\varphi(Z,x) := \inf_{z \in Z} \mathscr{D}_\varphi(z,x).$$

Example 2.1. Separable Bregman distances are the most commonly used in the literature. In detail, when $C := \prod_{i=1}^n C_i$ is of separable structure, the Legendre function is written as the summation of one-dimensional functions

$$\varphi(x) := \sum_{i=1}^n \theta(x_i),$$

where $\theta : \mathbb{R} \to \mathbb{R}_+ \cup \{+\infty\}$ satisfies conditions (a) and (b) on C_i. By the separable structure and (2.1), one has $\mathscr{D}_\varphi(x,y) = \sum_{i=1}^n \mathscr{D}_\theta(x_i,y_i)$. Several popular Bregman kernels are described as follows, as well as the generated Bregman distances. In particular, the Euclidean distance, the Kullback-Leibler divergence and the Itakura-Saito divergence are Bregman distances generated by the energy, the Boltzmann-Shannon entropy and the Burg entropy, respectively.

(i) Energy: $\theta(t) := \frac{1}{2}t^2$, $\operatorname{dom}\theta = \mathbb{R}$ and $\mathscr{D}_\varphi(x,y) = \frac{1}{2}\|x-y\|^2$.

(ii) Boltzmann-Shannon entropy: $\theta(t) := t\log t$, $\mathscr{D}_\varphi(x,y) = \sum_{i=1}^n x_i \log\frac{x_i}{y_i} - x_i + y_i$.

(iii) Burg entropy: $\theta(t) := -\log t$, $\mathscr{D}_\varphi(x,y) = \sum_{i=1}^n \log\frac{y_i}{x_i} + \frac{x_i}{y_i} - 1$.

(iv) Fermi-Dirac entropy: $\theta(t) := t\log t + (1-t)\log(1-t)$, $\mathscr{D}_\varphi(x,y) = \sum_{i=1}^n x_i \log\frac{x_i}{y_i} + (1-x_i)\log\frac{1-x_i}{1-y_i}$.

(v) Hellinger distance: $\theta(t) := -\sqrt{1-t^2}$, $\mathscr{D}_\varphi(x,y) = \sum_{i=1}^n \frac{1-x_i y_i}{\sqrt{1-y_i^2}} - \sqrt{1-x_i^2}$.

(vi) ℓ_p quasi-norm: $\theta(t) := -t^p$ with $p \in (0,1)$, $\mathscr{D}_\varphi(x,y) = \sum_{i=1}^n -x_i^p + (1-p)y_i^p + p x_i y_i^{p-1}$.

(vii) Fractional power: $\theta(t) := \frac{pt-t^p}{1-p}$ with $p \in (0,1)$, $\mathscr{D}_\varphi(x,y) = \sum_{i=1}^n \frac{p y_i^{p-1}(x_i-y_i) - (x_i^p - y_i^p)}{1-p}$.

Given Bregman distance $\mathscr{D}_\varphi$ and a closed and convex set V, the Bregman projection mapping $\mathscr{P}_V : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is defined by

$$\mathscr{P}_V(g,x) := \arg\min_{z \in V} \langle g, z \rangle + \mathscr{D}_\varphi(z,x) \quad \text{for each } g \in \mathbb{R}^n, x \in \operatorname{int}C. \tag{2.4}$$

Thanks to the property of Bregman kernel φ, $\mathscr{P}_V(\cdot,\cdot)$ is well-defined and is a single-valued mapping with images in $\operatorname{int}C$ (see also [2]). For some choices of φ and V (such as the polyhedron and the nonnegative orthant), the Bregman projection mapping (2.4) can be computed via a closed formula; one can refer to [1, 2, 8, 9] for detailed examples.

Proposition 2.2. *Let $x \in \operatorname{int}C \cap V$ and $g \in \mathbb{R}^n$. The following assertions are true.*

(i) $\mathscr{P}_V(0,x) = x$.
(ii) $\langle g + \nabla\varphi(\mathscr{P}_V(g,x)) - \nabla\varphi(x), z - \mathscr{P}_V(g,x)\rangle \geq 0$ *for each* $z \in V$.
(iii) $\sigma\|x - \mathscr{P}_V(g,x)\|^2 \leq \langle g, x - \mathscr{P}_V(g,x)\rangle$.
(iv) $\sigma\|x - \mathscr{P}_V(g,x)\| \leq \|g\|$.

Proof. Assertions (i) and (ii) directly follow from the definition and optimality condition of (2.4) and (2.1), respectively; assertion (iv) immediately follows from assertion (iii) and the Cauchy-Schwarz inequality. Hence it only remains to show assertion (iii). To this end, it follows from the σ-strongly convex of Bregman kernel φ that

$$\langle \nabla\varphi(x) - \nabla\varphi(y), x - y\rangle \geq \sigma\|x - y\|^2 \quad \text{for any } x, y \in \operatorname{int} C.$$

With $\mathscr{P}_V(g,x)$ in place of y and by assertion (ii) (with x in place of z), the above inequality is reduced to assertion (iii). The proof is complete. $\square$

3. Bregman quasi-subgradient method

Let $I := \{1,2,\ldots,m\}$, and let $\{f_i : i \in I\}$ be a family of quasi-convex and continuous functions defined on $\mathbb{R}^n$, $C \subseteq \mathbb{R}^n$ be a closed and convex set with non-empty interior and $V \subseteq \mathbb{R}^n$ be a closed and convex set. In the present paper, we consider the quasi-convex feasibility problem (QFP) that is to find a feasible point $x \in \mathbb{R}^n$ such that

$$x \in C \cap V \quad \text{and} \quad f_i(x) \leq 0 \quad \text{for each } i \in I. \tag{3.1}$$

As usual, we assume that the QFP is consistent, i.e., the solution set of the QFP is nonempty:

$$S := \{x \in C \cap V : f_i(x) \leq 0, \forall i \in I\} \neq \varnothing.$$

Furthermore, we always assume that each component function of the QFP (3.1) satisfies a Hölder condition as in the following assumption.

Assumption 3.1. For each $i \in I$, f_i satisfies the Hölder condition of order $\beta_i \in (0,1]$ with modulus $L_i \in (0,\infty)$ on $C \cap V$. Moreover, we write

$$\beta_{\min} := \min_{i\in I} \beta_i, \quad \beta_{\max} := \max_{i\in I} \beta_i, \quad L_{\max} := \max_{i\in I} L_i. \tag{3.2}$$

In particular, the QFP (3.1) is said to be homogeneous if $\beta_{\min} = \beta_{\max}$.

One of the most popular algorithms for solving the feasibility problem (3.1) is a class of subgradient methods; see [5, 20] and references therein. By employing the Bregman projection mapping and inspired by the idea of subgradient methods, we propose the Bregman quasi-subgradient methods for solving the QFP (3.1) in a general framework, stated as follows.

Algorithm 3.1. *Select an initial point* $x_1 \in C \cap V$ *and a sequence of stepsizes* $\{v_k\} \subseteq (0,+\infty)$ *satisfying*

$$0 < \underline{v} \leq v_k \leq \overline{v} < \sigma. \tag{3.3}$$

For each $k \in \mathbb{N}$, *having* $x_k \in \mathbb{R}^n$, *we select a nonempty index set* $I_k \subseteq I$ *and weights* $\{\lambda_{k,i}\}_{i\in I_k} \in \Delta_+^{|I_k|}$, *calculate* $g_{k,i} \in \partial^Q f_i(x_k) \cap \mathbb{S}$ *for each* $i \in I_k$, *and update* x_{k+1} *by*

$$x_{k+1} := \mathscr{P}_V\left(v_k \sum_{i\in I_k} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{1}{\beta_i}} g_{k,i}, x_k\right). \tag{3.4}$$

Algorithm 3.1 provides a unified framework of Bregman subgradient methods for solving the QFP. The main computational task in Algorithm 3.1 is the Bregman projection mapping (2.4). As mentioned above, for some choices of φ and V, the Bregman projection mapping (2.4) can be computed via a closed formula, and thus the resulting Algorithm 3.1 is particularly attractive. In the special case when the Bregman kernel is chosen as the energy (see Example 2.1(i)), the Bregman distance/projection is reduced to the Euclidean distance/projection, and then the Bregman quasi-subgradient method is reduced to the projected quasi-subgradient method [20] for solving the QFP. Moreover, it is clear by (3.4) and Proposition 2.2(i) that

$$\text{the sequence } \{x_i\}_{i>k} \text{ will terminate at } x_k \text{ whenever it enters } S. \tag{3.5}$$

The following lemma provides a basic inequality of Algorithm 3.1, which shows a significant property and plays a key tool in convergence analysis of Bregman quasi-subgradient methods.

Lemma 3.1. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1. Then the following basic inequality holds for each $x \in S$ and $k \in \mathbb{N}$ that*

$$\mathscr{D}_\varphi(x, x_{k+1}) \le \mathscr{D}_\varphi(x, x_k) - v_k\left(1 - \frac{v_k}{\sigma}\right)\sum_{i\in I_k}\lambda_{k,i}\left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{2}{\beta_i}}. \tag{3.6}$$

Proof. Fix $x \in S$ and $k \in \mathbb{N}$. We assume, without loss of generality, that $x_k \notin S$; otherwise, $f_i^+(x_k) = 0$ for each $i \in I$, and thus, (3.6) is satisfied automatically by (3.5). By the three point identity (2.3) and since $\mathscr{D}_\varphi \ge 0$, we obtain

$$\begin{aligned}
\mathscr{D}_\varphi(x, x_{k+1}) - \mathscr{D}_\varphi(x, x_k) &= -\mathscr{D}_\varphi(x_{k+1}, x_k) + \langle x - x_{k+1}, \nabla\varphi(x_k) - \nabla\varphi(x_{k+1})\rangle \\
&\le \langle x - x_{k+1}, \nabla\varphi(x_k) - \nabla\varphi(x_{k+1})\rangle \\
&\le v_k \sum_{i\in I_k}\lambda_{k,i}\left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{1}{\beta_i}}\langle g_{k,i}, x - x_{k+1}\rangle,
\end{aligned} \tag{3.7}$$

where the last inequality follows from (3.4) and Proposition 2.2(ii). Due to Assumption 3.1 and $g_{k,i} \in \mathbb{S}$, we obtain by Lemma 2.1 and Proposition 2.2(iv) the following inequalities

$$\langle g_{k,i}, x_k - x\rangle \ge \left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{1}{\beta_i}},$$

and

$$\langle g_{k,i}, x_k - x_{k+1}\rangle \le \|x_k - x_{k+1}\| \le \frac{v_k}{\sigma}\sum_{i\in I_k}\lambda_{k,i}\left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{1}{\beta_i}}, \tag{3.8}$$

respectively. Consequently, one has

$$\sum_{i\in I_k}\lambda_{k,i}\left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{1}{\beta_i}}\langle g_{k,i}, x_k - x\rangle \ge \sum_{i\in I_k}\lambda_{k,i}\left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{2}{\beta_i}},$$

and

$$\sum_{i\in I_k}\lambda_{k,i}\left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{1}{\beta_i}}\langle g_{k,i}, x_k - x_{k+1}\rangle \le \frac{v_k}{\sigma}\left(\sum_{i\in I_k}\lambda_{k,i}\left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{1}{\beta_i}}\right)^2 \le \frac{v_k}{\sigma}\sum_{i\in I_k}\lambda_{k,i}\left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{2}{\beta_i}},$$

(thanks to $\{\lambda_{k,i}\}_{i\in I_k} \in \Delta_+^{|I_k|}$ and the convexity of $\|\cdot\|^2$). Combining these two inequalities with (3.7) is reduced to (3.6). The proof is complete. $\square$

To guarantee the convergence property of the Bregman quasi-subgradient method, we shall assume the following condition on parameters; see [5, Remark 3.13].

Assumption 3.2. There exist $\mu > 0$ such that $\min_{i\in I_k} \lambda_{k,i} \geq \mu$ for each $k \in \mathbb{N}$.

By virtue of the basic inequality, we show some basic properties of Algorithm 3.1 under Assumption 3.2.

Lemma 3.2. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1 with $\{\lambda_{k,i}\}$ satisfying Assumption 3.2, and $x \in S$. Then the following assertions are true.*

(i) *It holds for each $k \in \mathbb{N}$ that*

$$\mathscr{D}_\varphi(x, x_{k+1}) \leq \mathscr{D}_\varphi(x, x_k) - \mu \underline{\nu}\left(1 - \frac{\bar{\nu}}{\sigma}\right) L_{\max}^{-\frac{2}{\beta_{\min}}} \sum_{i\in I_k} \left(f_i^+(x_k)\right)^{\frac{2}{\beta_i}}. \tag{3.9}$$

(ii) *$\{\mathscr{D}_\varphi(x, x_k)\}$ is monotonically decreasing, and hence $\{x_k\}$ is bounded.*

(iii) $\lim_{k\to\infty} \sum_{i\in I_k} \left(f_i^+(x_k)\right)^{\frac{2}{\beta_i}} = 0$.

Proof. (i) By (3.2), (3.3) and Assumption 3.2, (3.9) directly follows from (3.6).

(ii) (3.9) shows that the sequence $\{\mathscr{D}_\varphi(x, x_k)\}$ is monotonically decreasing, and hence is bounded. Consequently by (2.2), $\{\|x_k - x\|\}$ is bounded, and so as is $\{x_k\}$.

(iii) It follows from (3.9) that

$$\sum_{k=1}^{\infty} \sum_{i\in I_k} \left(f_i^+(x_k)\right)^{\frac{2}{\beta_i}} \leq \frac{1}{\underline{\nu}\left(1 - \frac{\bar{\nu}}{\sigma}\right)\mu} L_{\max}^{\frac{2}{\beta_{\min}}} \mathscr{D}_\varphi(x, x_1) < \infty.$$

This implies that $\lim_{k\to\infty} \sum_{i\in I_k} \left(f_i^+(x_k)\right)^{\frac{2}{\beta_i}} = 0$.

$\square$

The control sequence of index sets $\{I_k\}$ plays a central role in guaranteeing convergence property and numerical performance of subgradient methods for solving the feasibility problems. In this paper, we will investigate the convergence theory for Algorithm 3.1 with two general but popular control schemes; see, e.g., [5, 20].

Definition 3.1. Let $\alpha \in (0,1]$ and $s \in \mathbb{N}$, and let $\{x_k\}$ be a sequence generated by Algorithm 3.1. We say that $\{I_k\}$ is

(a) the α-most violated constraints control if, for each $k \in \mathbb{N}$, there exists $i_k \in I_k$ such that

$$f_{i_k}^+(x_k) \geq \alpha \max_{i\in I} f_i^+(x_k).$$

(b) the s-intermittent control if

$$I = I_k \cup I_{k+1} \cup \cdots \cup I_{k+s-1} \quad \text{for each } k \in \mathbb{N}.$$

For the remainder of this paper, we assume that Assumptions 3.1 and 3.2 are always satisfied, and establish the convergence theory, including the global convergence, iteration complexity and convergence rates, of Algorithm 3.1 with the α-most violated constraints control and the s-intermittent control, respectively.

3.1. The α-most violated constraints control.

3.1.1. *Global convergence.* We first establish the global convergence for the Bregman quasi-subgradient method with the α-most violated constraints control.

Theorem 3.1. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ being the α-most violated constraints control. Then $\{x_k\}$ converges to a feasible solution of the QFP (3.1).*

Proof. Note by Lemma 3.2(ii) that $\{x_k\}$ is bounded, and thus, must have a cluster point, denoted by $\bar{x}$. By definition of the α-most violated constraints control (cf. Definition 3.1(a)), for each $k \in \mathbb{N}$,

$$\text{there exists } i_k \in I_k \text{ such that } f_{i_k}^+(x_k) \geq \alpha \max\nolimits_{i\in I} f_i^+(x_k). \tag{3.10}$$

Then one has by (3.2) and Lemma 3.2(iii) that

$$\lim_{k\to\infty}\left(\alpha \max_{i\in I} f_i^+(x_k)\right)^{\frac{2}{\beta_{\min}}} \leq \lim_{k\to\infty}\left(f_{i_k}^+(x_k)\right)^{\frac{2}{\beta_{i_k}}} \leq \lim_{k\to\infty}\sum_{i\in I_k}\left(f_i^+(x_k)\right)^{\frac{2}{\beta_i}} = 0;$$

consequently, $\lim_{k\to\infty}\max_{i\in I} f_i^+(x_k) = 0$, and thus the cluster point $\bar{x} \in S$ by the continuity of each f_i. Recall from Lemma 3.2(ii) that $\{\mathscr{D}_\varphi(\bar{x}, x_k)\}$ is monotonically decreasing and converging to 0. Hence we conclude that $\{x_k\}$ converges to $\bar{x} \in S$ due to (2.2), and the proof is complete. □

3.1.2. *Iteration complexity.* Given $\delta > 0$, the iteration complexity of a particular algorithm is to estimate the number of iterations K required by the algorithm to obtain an approximate δ-feasible solution, that is,

$$\min_{1\leq k\leq K}\max_{i\in I} f_i^+(x_k) \leq \delta.$$

Theorem 3.2. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ being the α-most violated constraints control. Let $\delta > 0$ and $K_{\mathrm{m}} := \frac{\mathscr{D}_\varphi(S,x_1)}{\mu\underline{v}\left(1-\frac{\bar{v}}{\sigma}\right)}\left(\frac{L_{\max}}{\alpha\delta}\right)^{\frac{2}{\beta_{\min}}}$. Then*

$$\min_{1\leq k\leq K_{\mathrm{m}}}\max_{i\in I} f_i^+(x_k) \leq \delta.$$

Proof. Proving by contradiction, we assume that $\max_{i\in I} f_i^+(x_k) > \delta$ for each $1 \leq k \leq K_{\mathrm{m}}$. Then it follows from (3.9) (taking $x := \arg\min_{z\in S}\mathscr{D}_\varphi(z,x_k)$) and (3.10) that, for each $1\leq k\leq K_{\mathrm{m}}$,

$$\begin{aligned}\mathscr{D}_\varphi(S,x_{k+1}) &\leq \mathscr{D}_\varphi(S,x_k) - \mu\underline{v}\left(1-\frac{\bar{v}}{\sigma}\right)L_{\max}^{-\frac{2}{\beta_{\min}}}\left(f_{i_k}^+(x_k)\right)^{\frac{2}{\beta_{i_k}}}\\ &\leq \mathscr{D}_\varphi(S,x_k) - \mu\underline{v}\left(1-\frac{\bar{v}}{\sigma}\right)\left(\frac{\alpha}{L_{\max}}\max_{i\in I} f_i^+(x_k)\right)^{\frac{2}{\beta_{\min}}}.\end{aligned} \tag{3.11}$$

This, together with the assumption that $\max_{i\in I} f_i^+(x_k) > \delta$, implies that

$$\mathscr{D}_\varphi(S,x_{k+1}) < \mathscr{D}_\varphi(S,x_k) - \mu\underline{v}\left(1-\frac{\bar{v}}{\sigma}\right)\left(\frac{\alpha\delta}{L_{\max}}\right)^{\frac{2}{\beta_{\min}}}.$$

Summing the above inequality over $k = 1,\ldots,K_{\mathrm{m}}$, we derive that

$$0 \leq \mathscr{D}_\varphi(S,x_{K_{\mathrm{m}}+1}) < \mathscr{D}_\varphi(S,x_1) - K_{\mathrm{m}}\mu\underline{v}\left(1-\frac{\bar{v}}{\sigma}\right)\left(\frac{\alpha\delta}{L_{\max}}\right)^{\frac{2}{\beta_{\min}}},$$

which yields a contradiction with the definition of K_{m}. The proof is complete. □

3.1.3. *Convergence rate analysis.* The establishment of convergence rate is significant in guaranteeing the numerical performance of relevant algorithms. The error bound property plays an important role in convergence rate analysis of numerical algorithms; see, e.g., [20, 31, 33]. Below we introduce the Hölder-type error bound relative to a Bregman distance for the QFP (3.1), and use it to establish convergence rates of Bregman quasi-subgradient methods.

Definition 3.2. The inequality system (3.1) is said to satisfy the Hölder-type bounded error bound property of order $q>0$ relative to the Bregman distance $\mathscr{D}_\varphi$ if, for any $r>0$ such that $S\cap\mathbb{B}(0,r)\neq\emptyset$, there exists $\kappa(r)>0$ such that

$$\mathscr{D}_\varphi^q(S,x)\leq\kappa(r)\max_{i\in I}f_i^+(x)\quad\text{for each } x\in C\cap V\cap\mathbb{B}(0,r). \tag{3.12}$$

In particular, the inequality system (3.1) is said to satisfy the (Lipschitz-type) bounded error bound property relative to the Bregman distance $\mathscr{D}_\varphi$ if (3.12) holds with $q=1$.

Theorem 3.3. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ being the α-most violated constraints control. Suppose that (3.1) satisfies the Hölder-type bounded error bound property of order $q>0$ relative to the Bregman distance $\mathscr{D}_\varphi$. Then the following assertions are true.*

(i) *If $2q=\beta_{\min}$, then $\{x_k\}$ converges to a feasible solution $\bar{x}\in S$ at a linear rate; particularly, there exist $c\geq 0$ and $\tau\in(0,1)$ such that*

$$\|x_k-\bar{x}\|\leq c\tau^k\quad\textit{for each } k\in\mathbb{N}. \tag{3.13}$$

(ii) *If $2q>\beta_{\min}$, then $\{x_k\}$ converges to a feasible solution $\bar{x}\in S$ at a sublinear rate; particularly, there exists $c\geq 0$ such that*

$$\|x_k-\bar{x}\|\leq ck^{-\frac{\beta_{\min}}{2(2q-\beta_{\min})}}\quad\textit{for each } k\in\mathbb{N}. \tag{3.14}$$

Proof. By Lemma 3.2(ii) that $\{x_k\}$ is bounded, there exists $r>0$ such that $\{x_k\}\subseteq\mathbb{B}(0,r)$. Then by the assumption of the Hölder-type bounded error bound property of order $q>0$ relative to the Bregman distance $\mathscr{D}_\varphi$, there exists $\kappa>0$ such that

$$\mathscr{D}_\varphi^q(S,x_k)\leq\kappa\max_{i\in I}f_i^+(x_k)\quad\text{for each } k\in\mathbb{N}. \tag{3.15}$$

This, together with (3.11), yields that

$$\mathscr{D}_\varphi(S,x_{k+1})\leq\mathscr{D}_\varphi(S,x_k)-\rho\,\mathscr{D}_\varphi^{\frac{2q}{\beta_{\min}}}(S,x_k)\quad\text{for each } k\geq N,$$

where $\rho:=\mu\underline{v}\left(1-\frac{\bar{v}}{\sigma}\right)\left(\frac{\alpha}{\kappa L_{\max}}\right)^{\frac{2}{\beta_{\min}}}$. Hence we derive that there exists $c\geq 0$ such that

$$\mathscr{D}_\varphi(S,x_k)\leq c(1-\rho)^k \text{ if } 2q=\beta_{\min};\ \ \mathscr{D}_\varphi(S,x_k)\leq ck^{-\frac{\beta_{\min}}{2q-\beta_{\min}}} \text{ if } 2q>\beta_{\min}, \tag{3.16}$$

by applying Lemma 2.3 (with $\mathscr{D}_\varphi(S,x_k)$, ρ, $\frac{2q}{\beta_{\min}}-1$ in place of u_k, b, r), for each $k\in\mathbb{N}$.

On the other side, fix $l>k$ and $x:=\arg\min_{z\in S}\mathscr{D}_\varphi(z,x_k)$. We have by (2.2) that

$$\|x_l-x_k\|^2\leq 2(\|x_l-x\|^2+\|x_k-x\|^2)\leq\frac{4}{\sigma}(\mathscr{D}_\varphi(x,x_l)+\mathscr{D}_\varphi(x,x_k))\leq\frac{8}{\sigma}\mathscr{D}_\varphi(S,x_k)$$

(thanks to Lemma 3.2(ii)). Hence by the convergence of $\{x_l\}$ to $\bar{x} \in S$ as shown in Theorem 3.1, we obtain

$$\|x_k - \bar{x}\|^2 = \lim_{l\to\infty} \|x_l - x_k\|^2 \leq \frac{8}{\sigma} \mathscr{D}_\varphi(S, x_k). \tag{3.17}$$

This, together with (3.16), implies (3.13) and (3.14), respectively. The proof is complete. □

3.2. **The s-intermittent control.**

3.2.1. *Global convergence.*

Theorem 3.4. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ being the s-intermittent control. Then $\{x_k\}$ converges to a feasible solution of the QFP* (3.1).

Proof. Fix $x \in S$ and $k \in \mathbb{N}$. We assume, without loss of generality, that $x_{sk} \notin S$; otherwise, the conclusion holds automatically by (3.5). By Lemma 3.2(i), we obtain inductively that

$$\mathscr{D}_\varphi(x, x_{s(k+1)}) \leq \mathscr{D}_\varphi(x, x_{sk}) - \mu \underline{v} \left(1 - \frac{\overline{v}}{\sigma}\right) L_{\max}^{-\frac{2}{\beta_{\min}}} \sum_{j=0}^{s-1} \sum_{i \in I_{sk+j}} \left(f_i^+(x_{sk+j})\right)^{\frac{2}{\beta_i}}. \tag{3.18}$$

Below we estimate the second term on the right hand side of (3.18) in terms of $\max_{i\in I} f_i^+(x_{sk})$. Let $i_k \in I$ be the most violated index of inequality system (3.1) at x_{sk}, that is,

$$f_{i_k}^+(x_{sk}) = \max_{i\in I} f_i^+(x_{sk}) > 0. \tag{3.19}$$

By the definition of the s-intermittent control (cf. Definition 3.1(b)), there exists $j_k \in [0, s-1]$ such that $i_k \in I_{sk+j_k}$. By the Hölder condition as in Assumption 3.1, one has

$$f_{i_k}(x_{sk}) \leq f_{i_k}(x_{sk+j_k}) + L_{\max} \|x_{sk+j_k} - x_{sk}\|^{\beta_{i_k}}. \tag{3.20}$$

Since $i_k \in I_{sk+j_k}$, we get

$$f_{i_k}(x_{sk+j_k}) \leq \sum_{i\in I_{sk+j_k}} f_i^+(x_{sk+j_k}) \leq \sum_{j=0}^{s-1} \sum_{i\in I_{sk+j}} f_i^+(x_{sk+j}). \tag{3.21}$$

On the other side, in view of Algorithm 3.1, we obtain by (3.8) that

$$\|x_{k+1} - x_k\| \leq \frac{v_k}{\sigma} \sum_{i\in I_k} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{1}{\beta_i}} \leq \sum_{i\in I_k} \left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{1}{\beta_i}}$$

(thanks to (3.3) and $\lambda_{k,i} \leq 1$) for each $k \in \mathbb{N}$. Then we derive inductively that

$$\|x_{sk+j_k} - x_{sk}\| \leq \sum_{j=0}^{s-1} \sum_{i\in I_{sk+j}} \left(\frac{f_i^+(x_{sk+j})}{L_i}\right)^{\frac{1}{\beta_i}};$$

consequently, one has by Lemma 2.2(i) (as $\beta_{i_k} \leq 1$) that

$$\|x_{sk+j_k} - x_{sk}\|^{\beta_{i_k}} \leq \sum_{j=0}^{s-1} \sum_{i\in I_{sk+j}} \left(\frac{f_i^+(x_{sk+j})}{L_i}\right)^{\frac{\beta_{i_k}}{\beta_i}} \leq L_{\max}^{-\frac{\beta_{\min}}{\beta_{\max}}} \sum_{j=0}^{s-1} \sum_{i\in I_{sk+j}} \left(f_i^+(x_{sk+j})\right)^{\frac{\beta_{i_k}}{\beta_i}}$$

(by (3.2)). This, together with (3.19)-(3.21), deduces that

$$\max_{i\in I} f_i^+(x_{sk}) \leq \sum_{j=0}^{s-1} \sum_{i\in I_{sk+j}} f_i^+(x_{sk+j}) + L_{\max}^{1-\frac{\beta_{\min}}{\beta_{\max}}} \sum_{j=0}^{s-1} \sum_{i\in I_{sk+j}} \left(f_i^+(x_{sk+j})\right)^{\frac{\beta_{i_k}}{\beta_i}}. \tag{3.22}$$

By Lemma 3.2(iii), there exists $N \in \mathbb{N}$ such that $\max_{i\in I_k} f_i^+(x_k) < 1$ for each $k \geq N$. Fix $k \geq \frac{N}{s}$. Hence we derive by (3.22) that

$$\max_{i\in I} f_i^+(x_{sk}) \leq (1+L_{\max}) \sum_{j=0}^{s-1} \sum_{i\in I_{sk+j}} \left(f_i^+(x_{sk+j})\right)^{\frac{\beta_{\min}}{\beta_i}};$$

and then by Lemma 2.2(ii) (as $\frac{2}{\beta_{\min}} > 1$) that

$$\left(\max_{i\in I} f_i^+(x_{sk})\right)^{\frac{2}{\beta_{\min}}} \leq (1+L_{\max})^{\frac{2}{\beta_{\min}}} (ms)^{\frac{2}{\beta_{\min}}-1} \sum_{j=0}^{s-1} \sum_{i\in I_{sk+j}} \left(f_i^+(x_{sk+j})\right)^{\frac{2}{\beta_i}}.$$

This, together with (3.18), implies that

$$\mathscr{D}_\varphi(x, x_{s(k+1)}) \leq \mathscr{D}_\varphi(x, x_{sk}) - \gamma \left(\max_{i\in I} f_i^+(x_{sk})\right)^{\frac{2}{\beta_{\min}}}, \tag{3.23}$$

where we write $\gamma := \mu\underline{v}\left(1-\frac{\bar{v}}{\sigma}\right)(L_{\max}+L_{\max}^2)^{-\frac{2}{\beta_{\min}}}(ms)^{1-\frac{2}{\beta_{\min}}}$, for each $k \geq \frac{N}{s}$. Clearly, this indicates that $\lim_{k\to\infty} \max_{i\in I} f_i^+(x_{sk}) = 0$, and thus the cluster point of $\{x_{sk}\}$ falls in S by the continuity of each f_i. This, together with Lemma 3.2(ii), shows that $\{x_k\}$ converges to this cluster point in S. The proof is complete. □

3.2.2. *Iteration complexity.*

Theorem 3.5. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ being the s-intermittent control. Suppose that (3.1) is homogeneous, i.e., $\beta_{\min} = \beta_{\max} := \beta$. Let $\delta > 0$ and $K_c := \frac{\mathscr{D}_\varphi(S,x_1)}{m\mu\underline{v}\left(1-\frac{\bar{v}}{\sigma}\right)}\left(\frac{2msL_{\max}}{\delta}\right)^{\frac{2}{\beta}}$. Then*

$$\min_{1\leq k\leq K_c} \max_{i\in I} f_i^+(x_k) \leq \delta.$$

Proof. By the homogeneous assumption that $\beta_{\min} = \beta_{\max} = \beta$, (3.22) is reduced to

$$\max_{i\in I} f_i^+(x_{sk}) \leq 2 \sum_{j=0}^{s-1} \sum_{i\in I_{sk+j}} f_i^+(x_{sk+j}).$$

Consequently, we have by Lemma 2.2(ii) (as $\frac{2}{\beta} > 1$) that

$$\left(\max_{i\in I} f_i^+(x_{sk})\right)^{\frac{2}{\beta}} \leq 2^{\frac{2}{\beta}} (ms)^{\frac{2}{\beta}-1} \sum_{j=0}^{s-1} \sum_{i\in I_{sk+j}} \left(f_i^+(x_{sk+j})\right)^{\frac{2}{\beta}},$$

and then, (3.18) (taking $x := \arg\min_{z\in S} \mathscr{D}_\varphi(z, x_{sk})$) is reduced to

$$\mathscr{D}_\varphi(S, x_{s(k+1)}) \leq \mathscr{D}_\varphi(S, x_{sk}) - \mu\underline{v}\left(1-\frac{\bar{v}}{\sigma}\right)(2L_{\max})^{-\frac{2}{\beta}}(ms)^{1-\frac{2}{\beta}}\left(\max_{i\in I} f_i^+(x_{sk})\right)^{\frac{2}{\beta}}. \tag{3.24}$$

Proving by contradiction, we assume that $\max_{i\in I} f_i^+(x_k) > \delta$ for each $1 \leq k \leq K_c$. Then it follows from (3.24) that, for each $1 \leq k \leq \frac{K_c}{s}$,

$$\mathscr{D}_\varphi(S, x_{s(k+1)}) < \mathscr{D}_\varphi(S, x_{sk}) - \mu\underline{v}\left(1-\frac{\bar{v}}{\sigma}\right)(2L_{\max})^{-\frac{2}{\beta}}(ms)^{1-\frac{2}{\beta}}\delta^{\frac{2}{\beta}}.$$

Summing the above inequality over $k = 1, \ldots, \frac{K_c}{s}$ and by the monotonically decreasing property of $\{\mathscr{D}_\varphi(S, x_k)\}$ (cf. Lemma 3.2(ii)), we derive that

$$0 \le \mathscr{D}_\varphi(S, x_{K_c+s}) < \mathscr{D}_\varphi(S, x_1) - \frac{K_c}{s}\mu\underline{v}\left(1 - \frac{\overline{v}}{\sigma}\right)(2L_{\max})^{-\frac{2}{\beta}}(ms)^{1-\frac{2}{\beta}}\delta^{\frac{2}{\beta}},$$

which yields a contradiction with the definition of K_c. The proof is complete. □

3.2.3. *Convergence rate analysis.*

Theorem 3.6. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ being the s-intermittent control. Suppose that (3.1) satisfies the Hölder-type bounded error bound property of order q relative to the Bregman distance $\mathscr{D}_\varphi$. Then the following assertions are true.*

(i) *If $2q = \beta_{\min}$, then $\{x_k\}$ converges to a feasible solution $\bar{x} \in S$ at a linear rate; particularly, there exist $c \ge 0$ and $\tau \in (0,1)$ such that*

$$\|x_k - \bar{x}\| \le c\tau^k \quad \text{for each } k \in \mathbb{N}. \tag{3.25}$$

(ii) *If $2q > \beta_{\min}$, then $\{x_k\}$ converges to a feasible solution $\bar{x} \in S$ at a sublinear rate; particularly, there exists $c \ge 0$ such that*

$$\|x_k - \bar{x}\| \le ck^{-\frac{\beta_{\min}}{2(2q-\beta_{\min})}} \quad \text{for each } k \in \mathbb{N}. \tag{3.26}$$

Proof. We obtain by (3.23) (taking $x := \arg\min_{z\in S} \mathscr{D}_\varphi(z, x_{sk})$) that there exist $\gamma > 0$ and $N \in \mathbb{N}$ such that

$$\mathscr{D}_\varphi(S, x_{s(k+1)}) \le \mathscr{D}_\varphi(S, x_{sk}) - \gamma\left(\max_{i\in I} f_i^+(x_{sk})\right)^{\frac{2}{\beta_{\min}}}, \tag{3.27}$$

holds for each $k \ge \frac{N}{s}$. By Lemma 3.2(ii) and the assumption of the Hölder-type bounded error bound property of order q relative to the Bregman distance $\mathscr{D}_\varphi$, there exists $\tau > 0$ such that (3.15) is satisfied. This, together with (3.27), yields

$$\mathscr{D}_\varphi(S, x_{s(k+1)}) \le \mathscr{D}_\varphi(S, x_{sk}) - \rho\, \mathscr{D}_\varphi^{\frac{2q}{\beta_{\min}}}(S, x_k) \quad \text{for each } k \ge \frac{N}{s},$$

where we write $\rho := \gamma\kappa^{-\frac{2}{\beta_{\min}}}$. Hence we derive that there exists $c \ge 0$ such that

$$\mathscr{D}_\varphi(S, x_k) \le c(1-\rho)^k \text{ if } 2q = \beta_{\min}; \quad \mathscr{D}_\varphi(S, x_k) \le ck^{-\frac{\beta_{\min}}{2q-\beta_{\min}}} \text{ if } 2q > \beta_{\min},$$

by Lemma 2.3 (with $\mathscr{D}_\varphi(S, x_{sk})$, ρ, $\frac{2q}{\beta_{\min}} - 1$ in place of u_k, b, r), for each $k \in \mathbb{N}$. These, together with (3.17) and the decreasing property of $\{\|x_k - \bar{x}\|\}$ as in Lemma 3.2(ii), imply (3.25) and (3.26), respectively. The proof is complete. □

Acknowledgments

The authors are grateful to the editor and the anonymous reviewers for their valuable comments and suggestions toward the improvement of this paper. The first author was supported by the National Natural Science Foundation of China (12222112, 12071306, 32170655), Project of Educational Commission of Guangdong Province (2021KTSCX103), Shenzhen Science and Technology Program (RCJC20221008092753082). The second author was supported by the Science and Technology Planning Project of Shenzhen Municipality (20220810155530001).

The forth author was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/FDS14/P03/20, UGC/FDS14/P02/21).

REFERENCES

[1] A. Auslender and M. Teboulle. Interior gradient and proximal methods for convex and conic optimization. *SIAM Journal on Optimization*, 16(3):697–725, 2006.

[2] A. Auslender and M. Teboulle. Projected subgradient methods with non-Euclidean distances for non-differentiable convex minimization and variational inequalities. *Mathematical Programming*, 120:27–48, 2009.

[3] D. Aussel and N. Hadjisavvas. Adjusted sublevel sets, normal operator, and quasi-convex programming. *SIAM Journal on Optimization*, 16(2):358–367, 2005.

[4] M. Avriel, W. E. Diewert, S. Schaible, and I. Zang. *Generalized Concavity*. Plenum Press, New York, 1988.

[5] H. H. Bauschke and J. M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Review*, 38(3):367–426, 1996.

[6] A. Beck and M. Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.

[7] A. Ben-tal, T. Margalit, and A. Nemirovski. The ordered subsets mirror descent optimization method with applications to tomography. *SIAM Journal on Optimization*, 12(1):79–108, 2001.

[8] R. Burachik and J. Dutta. Inexact proximal point methods for variational inequality problems. *SIAM Journal on Optimization*, 20(5):2653–2678, 2010.

[9] R. S. Burachik, Y. Hu, and X. Yang. Interior quasi-subgradient method with non-Euclidean distances for constrained quasi-convex optimization problems in hilbert spaces. *Journal of Global Optimization*, 83(2):249–271, 2022.

[10] C. Byrne. *Iterative Optimization in Inverse Problems*. CRC Press, New York, 2014.

[11] Y. Censor, T. Elfving, N. Kopf, and T. Bortfeld. The multiple-sets split feasibility problem and its applications for inverse problems. *Inverse Problems*, 21(6):2071–2084, 2005.

[12] Y. Censor and A. Lent. Cyclic subgradient projections. *Mathematical Programming*, 24(1):233–235, 1982.

[13] Y. Censor and A. Segal. Algorithms for the quasiconvex feasibility problem. *Journal of Computational and Applied Mathematics*, 185(1):34–50, 2006.

[14] G. Chen and M. Teboulle. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM Journal on Optimization*, 3(3):538–543, 1993.

[15] P. L. Combettes. The convex feasibility problem in image recovery. in *Advances in Imaging and Electron Physics*, volume 95, pages 155 – 270, 1996.

[16] N. Hadjisavvas, S. Komlósi, and S. Schaible. *Handbook of Generalized Convexity and Generalized Monotonicity*. Springer-Verlag, New York, 2005.

[17] Y. Hu, C. Li, J. Wang, X. Yang, and L. Zhu. Linearized proximal algorithms with adaptive stepsizes for convex composite optimization with applications. *Applied Mathematics and Optimization*, 87:52, 2023.

[18] Y. Hu, C. Li, and X. Yang. On convergence rates of linearized proximal algorithms for convex composite optimization with applications. *SIAM Journal on Optimization*, 26(2):1207–1235, 2016.

[19] Y. Hu, G. Li, M. Li, and C. K. W. Yu. Multiple-sets split quasi-convex feasibility problems: Adaptive subgradient methods with convergence guarantee. *Journal of Nonlinear and Variational Analysis*, 6(2):15–33, 2022.

[20] Y. Hu, G. Li, C. K. W. Yu, and T. L. Yip. Quasi-convex feasibility problems: Subgradient methods and convergence rates. *European Journal of Operational Research*, 298(1):45–58, 2022.

[21] Y. Hu, J. Li, and C. K. W. Yu. Convergenece rates of subgradient methods for quasi-convex optimization problems. *Computational Optimization and Applications*, 77(1):183–212, 2020.

[22] Y. Hu, X. Yang, and C.-K. Sim. Inexact subgradient methods for quasi-convex optimization problems. *European Journal of Operational Research*, 240(2):315–327, 2015.

[23] Y. Hu, C. K. W. Yu, and C. Li. Stochastic subgradient method for quasi-convex optimization problems. *Journal of Nonlinear and Convex Analysis*, 17(4):711–724, 2016.

[24] X. X. Huang and X. Q. Yang. A unified augmented Lagrangian approach to duality and exact penalization. *Mathematics of Operations Research*, 28(3):533–552, 2003.

[25] K. C. Kiwiel. Convergence and efficiency of subgradient methods for quasiconvex minimization. *Mathematical Programming*, 90:1–25, 2001.

[26] I. V. Konnov. On convergence properties of a subgradient method. *Optimization Methods and Software*, 18(1):53–62, 2003.

[27] G. Li, M. Li, and Y. Hu. Stochastic quasi-subgradient method for stochastic quasi-convex feasibility problems. *Discrete & Continuous Dynamical Systems - S*, 15(4):713–725, 2022.

[28] A. S. Nemirovsky and D. B. Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley, New York, 1983.

[29] B. T. Polyak. *Introduction to Optimization*. Optimization Software, New York, 1987.

[30] J. Qin, Y. Hu, J.-C. Yao, R. W. T. Leung, Y. Zhou, Y. Qin, and J. Wang. Cell fate conversion prediction by group sparse optimization method utilizing single-cell and bulk OMICs data. *Briefings in Bioinformatics*, 22(6):bbab311, 2021.

[31] J. Wang, Y. Hu, C. Li, and J.-C. Yao. Linear convergence of CQ algorithms and applications in gene regulatory network inference. *Inverse Problems*, 33(5):055017, 2017.

[32] C. K. W. Yu, Y. Hu, X. Yang, and S. K. Choy. Abstract convergence theorem for quasi-convex optimization problems with applications. *Optimization*, 68(7):1289–1304, 2019.

[33] M.-C. Yue, Z. Zhou, and A. M.-C. So. A family of inexact SQA methods for non-smooth convex minimization with provable convergence guarantees based on the Luo–Tseng error bound property. *Mathematical Programming*, 174(1):327–358, 2019.

[34] A. J. Zaslavski. *Approximate Solutions of Common Fixed-Point Problems*. Springer, Switzerland, 2016.