J. Nonlinear Var. Anal. 10 (2026), No. 3, pp. 687-701
Available online at http://jnva.biemdas.com
https://doi.org/10.23952/jnva.10.2026.3.09

STOCHASTIC QUASI-SUBGRADIENT METHODS WITH BREGMAN PROJECTION FOR STOCHASTIC QUASI-CONVEX FEASIBILITY PROBLEMS

YAOHUA HU[1], JINGCHAO LI[1,*], CHEZHOU YANG[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. The feasibility problem lies at the core of modeling numerous challenges across mathematics and the physical sciences. Quasi-convex functions, widely applied in economics, finance, and management science, offer a flexible framework for such problems. This paper addresses the stochastic quasi-convex feasibility problem (SQFP), which seeks a common point within infinitely many sublevel sets of quasi-convex functions. Inspired by the idea of a stochastic index scheme and Bregman projection, we propose a stochastic quasi-subgradient method with Bregman projection to solve the SQFP, in which the quasi-subgradients of a random (and finite) index set of component quasi-convex functions at the current iterate are used to construct the descent direction and employ the Bregman projection in place of the Euclidean projection at each iteration. Furthermore, we introduce a notion of Hölder-type bounded error bound property relative to the Bregman projection and random control sequence for the SQFP. Leveraging this property, we establish a global convergence theorem and quantify convergence rates for the proposed algorithm. This paper reveals that the stochastic quasi-subgradient method with Bregman projection offers both low computational cost requirement and fast convergence feature.

Keywords. Bregman distance; Convergence rate; Quasi-convex programming; Stochastic feasibility problem; Subgradient method.

1. INTRODUCTION

Let I be an (finite or infinite) index set, and let $C \subseteq \mathbb{R}^n$ be a closed and convex set and $\{f_i : i \in I\}$ be a family of continuous real-valued functions on $\mathbb{R}^n$. The feasibility problem aims to find a point $x \in \mathbb{R}^n$ such that

$$x \in C \quad \text{and} \quad f_i(x) \leq 0 \quad \text{for each } i \in I. \tag{1.1}$$

This type of feasibility problems is at the core of the modeling of many problems in various disciplines of mathematics and physical sciences, such as image recovery [9], wireless sensor networks localization [18], radiation therapy treatment planning [11] and gene regulatory network inference [33].

*Corresponding author.
E-mail address: jingchaoli@szu.edu.cn (J. Li).
Received 2 April 2025; Accepted 8 November 2025; Published online 1 April 2026.

When I is finite, problem (1.1) is the classical *deterministic feasibility problem*. In numerous practical applications, the component functions involved in (1.1) are assumed to be convex, and the corresponding problem is called the *convex feasibility problem* (CFP). Driven by its broad applicability, significant efforts have focused on developing optimization algorithms for solving CFPs. One of the most popular approaches is the classical *projected subgradient method*. Many extensions on control schemes (cyclic/parallel/most-violated control schemes) and various convergence features of projected subgradient methods have been devised and well explored. For comprehensive discussions, we refer readers to [5, 11, 36] and references therein.

In applications such as robust stabilization, control systems, and integral equations, the index set I in problem (1.1) is typically infinite, then the corresponding problem (1.1) is called the *stochastic feasibility problem* [15, 32]. Classical projected subgradient methods are impractical here: parallel control schemes incur prohibitive computational costs, cyclic control fails to complete even a single iteration, and identifying the most violated constraint becomes intractable due to infinitely many inequalities constraints. To address these challenges, stochastic subgradient methods[1], inspired by the idea of stochastic index scheme [9, 28] in optimization, have been developed for solving stochastic convex feasibility problems (SCFP), where each component function in (1.1) is convex; see [25, 28, 32] and references therein. Notably, under an error bound assumption, (1.1) established global convergence and a linear convergence rate (with probability 1) for the stochastic subgradient method; Meanwhile, [25, 32] derived finite-termination convergence theorems under assumptions of strong feasibility and distinguishable feasible and infeasible points.

Previous studies primarily examine the feasibility problem (1.1) under the assumption of convexity. However, convex functions are overly restrictive for many real-world applications in economics, finance, and management science. In contrast, *quasi-convex functions* retain key desirable properties of convex functions while offering a more accurate representation of practical scenarios. A canonical example is the fractional function, defined as a ratio of technical metrics (e.g., efficiency ratios), which is quasi-convex and widely applied across disciplines; see [4, 34] and references therein. This versatility leads to a significant increase of studies in quasi-convex optimization; see [4, 14, 17, 34] and references therein.

When I is finite and functions involved in (1.1) are quasi-convex, the corresponding problem is called the *quasi-convex feasibility problem* (QFP), which was first introduced by Goffin et al. [16] at a differentiable case. Censor and Segal [12] and Hu et al. [19, 22] proposed the projected quasi-subgradient methods with cyclic/parallel/most-violated/stochastic control schemes to solve the non-differentiable QFP, and established their global convergence to a feasible solution of the QFP. When I is finite and functions involved in (1.1) are quasi-convex, Li et al. [27] proposed a stochastic quasi-subgradient method to solve the *stochastic quasi-convex feasibility problem* (SQFP) and established its global convergence to a feasible solution of the SQFP with probability 1.

Despite its utility, the projected subgradient method suffers from critical limitations inherent to Euclidean projections; see, e.g., [1, 2, 10]. Specifically, the Euclidean projection disrupts the algorithm's descent property, frequently inducing a zig-zagging trajectory that slows convergence. Moreover, the Euclidean projection could be computationally expensive if the feasible

[1]Randomized projection methods [15] can be understood as a special case of stochastic subgradient methods with certain stepsize.

set C is not simple. To overcome these limitations 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. This approach retains algorithmic efficiency while mitigating the drawbacks of traditional projections.

The Bregman subgradient method extends the projected subgradient method by using the Bregman projection instead of the Euclidean projection. The history of the Bregman subgradient method originates in 1983 from the mirror descent method proposed by Nemirovsky and Yudin [29], which can viewed as a Bregman subgradient method with the Kullback-Leibler divergence. Moreover, the Bregman subgradient method has 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, 10]. 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, 10] and references therein.

Inspired by the idea of stochastic index scheme [9, 28] and the Bregman projection [2, 10], this paper aims to propose a stochastic quasi-subgradient method with Bregman projection (see Algorithm 3.1) to solve the SQFP (1.1). In the proposed Algorithm 3.1, the quasi-subgradients of a random (and finite) index set of component functions at the current iterate are used to construct the descent direction and the Bregman projection is employed in place of the Euclidean projection at each iteration. Contrast to the deterministic subgradient methods, the stochastic quasi-subgradient method consumes much less computational cost at each iteration, and thus is particularly attractive in applications with numerous objectives or constraints.

The major contribution of this paper is to establish the convergence theory, including the global convergence theorem and convergence rate analysis, of the stochastic quasi-subgradient method with Bregman projection for solving the SQFP (1.1). In particular, we first introduce a notion of Hölder-type bounded error bound property relative to a random control sequence and a Bregman projection for the SQFP. Moreover, we use it to establish the global convergence of the stochastic quasi-subgradient method with Bregman projection to a feasible solution of the SQFP (1.1) with probability 1 (see Theorem 4.2) and to quantitatively estimate the convergence rates (see Theorem 4.3). The established convergence theory extends most of existing convergence results of subgradient methods for the CFP [5, Theorems 7.18 and 7.36], the QFP [19] and the SCFP [28, Proposition 3]; see Remark 4.1.

The remainder of 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. We propose a stochastic quasi-subgradient method with Bregman projection to solve SQFP (1.1) in Section 3 and investigate its global convergence theorem and convergence rate theory in Section 4.

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 studied for quasi-convex functions, defined using the "normal cone" to the level sets; see [3] and references therein. In particular, Kiwiel [24], Censor and Segal [12], and Hu et al. [19, 21] utilized a quasi-subgradient for developing 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 [21, 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, [10, 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 [19, 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., [19, 21].

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*

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

The following two lemmas are useful in the convergence analysis of subgradient methods.

Lemma 2.2 ([23, Lemma 4.1])**.** *Let $\gamma \ge 1$ and $a_i \ge 0$ for $i = 1, \cdots, n$. Then it holds that*

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

Lemma 2.3 ([31, pp. 46, Lemma 6])**.** *Let $r > 0$ and $b > 0$, and $\{u_k\} \subseteq \mathbb{R}_+$ be a sequence of nonnegative scalars such that*

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

Then it holds that

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

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, 10]. 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 non-empty 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 [13, 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

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

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: $\theta(t) := \frac{1}{2}t^2$, $\mathscr{D}_\varphi(x,y) = \frac{1}{2}\|x-y\|^2$.
(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: $\theta(t) := -\sqrt{1-t^2}$, $\mathscr{D}_\varphi(x,y) = \sum_{i=1}^n \frac{1-x_iy_i}{\sqrt{1-y_i^2}} - \sqrt{1-x_i^2}$.
(vi) ℓ_p quasi-norm: $\theta(t) := -t^p$ with $p \in (0,1)$, $\mathscr{D}_\varphi(x,y) = \sum_{i=1}^n -x_i^p + (1-p)y_i^p + px_iy_i^{p-1}$.
(vii) Fractional power: $\theta(t) := \frac{pt-t^p}{1-p}$ with $p \in (0,1)$, $\mathscr{D}_\varphi(x,y) = \sum_{i=1}^n \frac{py_i^{p-1}(x_i-y_i)-(x_i^p-y_i^p)}{1-p}$.

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, 10] for detailed examples. The following proposition recall some useful properties of the Bregman projection mapping.

Proposition 2.2 ([20, 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) $\sigma\|x - \mathscr{P}_V(g,x)\| \leq \|g\|$.

3. Stochastic Quasi-subgradient Method with Bregman Projection

This section aims to propose a stochastic quasi-subgradient method with Bregman projection to solve the stochastic quasi-convex feasibility problem.

3.1. **Stochastic quasi-convex feasibility problem.** Let I be an infinite index set, and let $\{f_i : i \in I\}$ be a family of quasi-convex and continuous (possibly nondifferentiable) real-valued 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 stochastic quasi-convex feasibility problem (SQFP) that is to find a feasible point $x \in \mathbb{R}$ such that

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

Let $(I, \mathscr{F}, \Pr)$ be a complete probability space. By the formulation given in [15], an equivalent representation of the SQFP (3.1) can be written as finding a point $x \in \mathbb{R}$ such that

$$\Pr(\{\omega \in I : x \in C \cap V, f_\omega(x) \le 0\}) = 1.$$

When I is a finite index set, the SQFP (3.1) is reduced to the classical QFP [12, 16, 19].

As usual, we assume throughout the whole paper that the SQFP (3.1) is consistent, i.e., the solution set of the SQFP (3.1) is nonempty:

$$S = \{x \in C \cap V : f_i(x) \le 0, \forall i \in I\} \ne \emptyset.$$

Moreover, we always assume that each component function of the SQFP (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. Moreover, we assume

$$\beta_{\inf} := \inf_{i \in I} \beta_i > 0 \quad \text{and} \quad L_{\sup} := \sup_{i \in I} L_i < \infty.$$

3.2. **Stochastic quasi-subgradient method with Bregman projection.** One of the most popular and practical optimization algorithms for solving the feasibility problem is a class of subgradient methods; see [5, 12, 19, 36] and references therein. However, for the SQFP (3.1) where the index set I is infinite, the typical deterministic control schemes in the classical subgradient method are not implementable, e.g., the parallel/cyclic/most-violated controls consume expensive computational cost because of the infinite inequalities constraints.

To conquer the obstacle of numerous objectives or infinite constraints in application problems, the idea of the stochastic index scheme is increasingly popular and extensively used for optimization problems with a large number of component functions [6] or a large number of constraints [28]. A typical and very popular example is the stochastic gradient descent (SGD) algorithm in machine learning [9], in which only one component function is randomly selected to construct the descent direction at each iteration. It was pointed out in [19] that the stochastic control enjoys both advantages of low computational cost requirement and low (worst-case) iteration complexity.

In order to execute the idea of stochastic index scheme, the random control sequence is recalled from [25] as follows.

Definition 3.1. Let $(\Omega, \mathscr{F}, \Pr)$ be a given probability space. The sequence $\{I_k\}$ is said to be a random control sequence in I if $I_k : \Omega \to 2^I \setminus \{\emptyset\}$ are independent and identically distributed (set-valued) random variables on $(\Omega, \mathscr{F}, \Pr)$ with $M := \sup_{\omega \in \Omega, k \in \mathbb{N}} |I_k(\omega)| < \infty$.

The projected subgradient method suffers from several disadvantages arising from the Euclidean projection, e.g., the Euclidean projection destroys the nice descent property and often lead to a zig-zagging effect, resulting in slow convergence. To avoid these drawbacks of the Euclidean projection, one popular approach is to replace the Euclidean projection by a Bregman projection, which 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, 10, 20].

Integrating the idea of stochastic index scheme [9, 28] and employing the Bregman projection [10, 20] into the quasi-subgradient method [12, 19], we propose the following stochastic quasi-subgradient method with Bregman projection to solve the SQFP (3.1). In particular, the quasi-subgradients of a random index set of component functions are selected to construct the descent direction at each iteration, and the Bregman projection is employed in place of the Euclidean projection. Recall that $\{(\beta_i, L_i)\}$ are the parameters given in Assumption 3.1.

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

$$0 < \underline{v} \leq v_k \leq \overline{v} < \sigma, \tag{3.2}$$

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

$$x_{k+1} := \mathscr{P}_V \left(v_k \sum_{i \in I_k(\omega)} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i} \right)^{\frac{1}{\beta_i}} g_{k,i}, x_k \right). \tag{3.3}$$

Remark 3.1. Algorithm 3.1 provides a unified framework of stochastic subgradient methods for feasibility problems, either deterministic or stochastic, either convex or quasi-convex.

(i) When I is finite and $\{I_k\}$ is single-valued and the Bregman kernel is selected as the energy (see Example 2.1(i)), Algorithm 3.1 is reduced to the stochastic subgradient methods for solving the CFP [32] and the QFP [22].

(ii) When I is infinite and the Bregman kernel is the energy, Algorithm 3.1 is reduced to the stochastic subgradient methods for solving the SCFP [28] and SQFP [27].

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; one can refer to [1, 2, 10, 20] for the detailed examples. Moreover, it is clear by (3.3) and Proposition 2.2(i) that

$$\text{the sequence } \{x_i\}_{i>k} \text{ will terminate at } x_k \text{ whenever it enters } S. \tag{3.4}$$

4. Convergence Analysis

This section is contributed to the convergence analysis of the stochastic quasi-subgradient method with Bregman projection. To guarantee the convergence property of Algorithm 3.1, we shall assume the following condition on the weights $\{\lambda_{k,i}\}$; see [5, Remark 3.13] and [27]. A natural example is the naive weights, i.e., $\lambda_{k,i} = \frac{1}{|I_k(\omega)|}$ for each $i \in I_k(\omega)$.

Assumption 4.1. There exists $\mu > 0$ such that $\min_{i \in I_k(\omega)} \lambda_{k,i} \geq \mu$ for any $\omega \in \Omega$ and $k \in \mathbb{N}$.

4.1. **Basic inequality.** The basic inequality shows a significant property and plays a key tool in convergence analysis of subgradient methods. The basic inequality of Algorithm 3.1 is presented in the following lemma, which is able to derive some basic properties of the stochastic quasi-subgradient method with Bregman projection.

Lemma 4.1. *Let $x \in S$, $\omega \in \Omega$, and let $\{x_k\}$ be a sequence generated by Algorithm 3.1. Suppose that Assumptions 3.1 and 4.1 hold. Then the following assertions are true.*

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

$$\mathscr{D}_\varphi(x, x_{k+1}) \le \mathscr{D}_\varphi(x, x_k) - \underline{v}\left(1 - \frac{\overline{v}}{\sigma}\right)\mu L_{\sup}^{-\frac{2}{\beta_{\inf}}} \sum_{i \in I_k(\omega)} \left(f_i^+(x_k)\right)^{\frac{2}{\beta_i}}. \tag{4.1}$$

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

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

Proof. (i) Fix $k \in \mathbb{N}$. We will claim that

$$\mathscr{D}_\varphi(x, x_{k+1}) \le \mathscr{D}_\varphi(x, x_k) - v_k\left(1 - \frac{v_k}{\sigma}\right) \sum_{i \in I_k(\omega)} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{2}{\beta_i}}. \tag{4.2}$$

Granting this, (4.1) directly follows by (3.2) and Assumptions 3.1 and 4.1, and thus assertion (i) is proved.

To show (4.2), without loss of generality, we assume that $x_k \notin S$; otherwise, $f_i^+(x_k) = 0$ for each $i \in I$, and thus, (4.2) is satisfied automatically by (3.4). By the three point identity (2.3) and due to the fact that $\mathscr{D}_\varphi \ge 0$, we have

$$\begin{aligned} \mathscr{D}_\varphi(x, x_{k+1}) - \mathscr{D}_\varphi(x, x_k) &= -\mathscr{D}_\varphi(x_{k+1}, x_k) + \langle x - x_{k+1}, \nabla\varphi(x_k) - \nabla\varphi(x_{k+1})\rangle \\ &\le \langle x - x_{k+1}, \nabla\varphi(x_k) - \nabla\varphi(x_{k+1})\rangle \\ &\le v_k \sum_{i \in I_k(\omega)} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{1}{\beta_i}} \langle g_{k,i}, x - x_{k+1}\rangle, \end{aligned} \tag{4.3}$$

where the last inequality follows from (3.3) and Proposition 2.2(ii). Due to Assumption 3.1 and the fact that $g_{k,i} \in \mathbb{S}$, we obtain by Lemma 2.1 that

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

and then

$$\sum_{i \in I_k(\omega)} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{1}{\beta_i}} \langle g_{k,i}, x_k - x\rangle \ge \sum_{i \in I_k(\omega)} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i}\right)^{\frac{2}{\beta_i}}. \tag{4.4}$$

Due to the fact that $g_{k,i} \in \mathbb{S}$ again, we obtain by Proposition 2.2(iv) that

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

and hence

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

(thanks to $\{\lambda_{k,i}\}_{i\in I_k(\omega)} \in \Delta_+^{|I_k(\omega)|}$ and the convexity of $\|\cdot\|^2$). Combining this with (4.4), (4.3) is reduced to (4.2), as desired.

(ii) (4.2), together with (3.2), 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 (4.1) and (3.2) that

$$\sum_{k=1}^{\infty}\sum_{i\in I_k(\omega)} \left(f_i^+(x_k)\right)^{\frac{2}{\beta_i}} \le \frac{1}{\underline{v}\left(1-\frac{\overline{v}}{\sigma}\right)\mu} L_{\sup}^{\frac{2}{\beta_{\inf}}} \mathscr{D}_\varphi(x,x_1) < \infty.$$

This shows that $\lim_{k\to\infty}\sum_{i\in I_k(\omega)}\left(f_i^+(x_k)\right)^{\frac{2}{\beta_i}} = 0$. The proof is complete. □

The error bound property [30] plays an important role and has been extensively used in convergence analysis of various numerical algorithms; see [8, 28, 19, 35] and references therein. Below, we introduce a notion of the Hölder-type bounded error bound property relative to a Bregman distance and a random control sequence for the SQFP (3.1). In particular, associated with the random control sequence $\{I_k\}$, the sigma-field $\{\mathscr{F}_k\}$ records the history of the method, that is,

$$\mathscr{F}_k := \{x_1, I_1(\omega), \dots, I_{k-1}(\omega)\} \quad \text{for each } k\in\mathbb{N}. \tag{4.5}$$

Definition 4.1. The SQFP (3.1) is said to satisfy the Hölder-type bounded error bound property of order $p>0$ relative to the Bregman distance $\mathscr{D}_\varphi$ and the random control sequence $\{I_k\}$ if, for any $r>0$ such that $S\cap\mathbb{B}(0,r)\neq\emptyset$, there exists $\eta := \eta(r) > 0$ such that

$$\mathscr{D}_\varphi^p(S,x) \le \eta\,\mathbb{E}\left\{\sum_{i\in I_k(\omega)} f_i^+(x)|\mathscr{F}_k\right\} \quad \text{for any } x\in C\cap V\cap\mathbb{B}(0,r) \text{ and } k\in\mathbb{N}. \tag{4.6}$$

Remark 4.1. In the case when $I_k(\omega) = \{\omega_k\}$ is single-valued and the Bregman kernel is selected as the energy and $p=\frac{1}{2}$, the error bound property (4.6) is reduced to [28, Assumption 2]:

$$\operatorname{dist}(x,S) \le \eta\,\mathbb{E}\left\{f_{\omega_k}^+(x)|\mathscr{F}_k\right\} \quad \text{for any } x\in C \text{ and } k\in\mathbb{N}, \tag{4.7}$$

which was used in [28] to explore the stochastic subgradient method for the SCFP. Clearly, the Hölder-type error bound property of order $p\ge\frac{1}{2}$ extends and loosens the condition (4.7), because the larger the order p, the less restrictive the condition.

To establish the convergence theory of the stochastic quasi-subgradient method with Bregman projection, we shall assume the Hölder-type bounded error bound property on the SQFP (3.1).

Assumption 4.2. The SQFP (3.1) satisfies the Hölder-type bounded error bound property of order $p\ge\frac{1}{2}$ relative to the with Bregman projection $\mathscr{D}_\varphi$ and the random control sequence $\{I_k\}$.

By virtue of the basic inequality (4.1) and the Hölder-type bounded error bound property, we provide the basic inequality in terms of conditional expectation for Algorithm 3.1.

Lemma 4.2. *Let $x \in S$, and let $\{x_k\}$ be generated by Algorithm 3.1 and $\{\mathscr{F}_k\}$ be defined by (4.5). Suppose that Assumptions 3.1, 4.1 and 4.2 hold, and let $\rho := \underline{v}\left(1-\frac{\overline{v}}{\sigma}\right)\mu M\left(\eta M L_{\sup}\right)^{-\frac{2}{\beta_{\inf}}}$. Then there exists $N \in \mathbb{N}$ such that the following basic inequality holds*

$$\mathbb{E}\{\mathscr{D}_\varphi(x,x_{k+1})|\mathscr{F}_k\} \le \mathscr{D}_\varphi(x,x_k) - \rho\, \mathscr{D}_\varphi^{\frac{2p}{\beta_{\inf}}}(S,x_k) \quad \text{for each } k \ge N. \tag{4.8}$$

Proof. By assumptions made in this lemma, Lemma 4.1 is applicable, and hence (4.1) holds for each $k \in \mathbb{N}$. Taking the conditional expectation of (4.1) with respect to $\mathscr{F}_k$, it follows that

$$\mathbb{E}\{\mathscr{D}_\varphi(x,x_{k+1})|\mathscr{F}_k\} \le \mathscr{D}_\varphi(x,x_k) - \underline{v}\left(1-\frac{\overline{v}}{\sigma}\right)\mu L_{\sup}^{-\frac{2}{\beta_{\inf}}}\mathbb{E}\left\{\sum_{i\in I_k(\omega)}\left(f_i^+(x_k)\right)^{\frac{2}{\beta_i}}|\mathscr{F}_k\right\}. \tag{4.9}$$

Below, we estimate the conditional expectation term at the right hand side of (4.9). Indeed, by Lemma 4.1(iii), there exists $N \in \mathbb{N}$ such that $\sum_{i\in I_k(\omega)}\left(f_i^+(x_k)\right)^{\frac{2}{\beta_i}} < 1$ for each $k \ge N$; consequently, $f_i^+(x_k) < 1$ for each $k \ge N$ and $i \in I_k(\omega)$. Fix $k \ge N$, and note by Assumption 3.1 that $\beta_i \ge \beta_{\inf}$ for each $i \in I$. Then we obtain that

$$\sum_{i\in I_k(\omega)}\left(f_i^+(x_k)\right)^{\frac{2}{\beta_i}} \ge \sum_{i\in I_k(\omega)}\left(f_i^+(x_k)\right)^{\frac{2}{\beta_{\inf}}}. \tag{4.10}$$

Noting by Assumption 3.1 that $\frac{2}{\beta_{\inf}} > 1$, we can apply Lemma 2.2 (with $f_i^+(x_k)$, $\frac{2}{\beta_{\inf}}$ in place of a_i, γ) to achieve that

$$\sum_{i\in I_k(\omega)}\left(f_i^+(x_k)\right)^{\frac{2}{\beta_{\inf}}} \ge |I_k(\omega)|^{1-\frac{2}{\beta_{\inf}}}\left(\sum_{i\in I_k(\omega)} f_i^+(x_k)\right)^{\frac{2}{\beta_{\inf}}} \ge M^{1-\frac{2}{\beta_{\inf}}}\left(\sum_{i\in I_k(\omega)} f_i^+(x_k)\right)^{\frac{2}{\beta_{\inf}}}$$

(thanks to $|I_k(\omega)| \le M$ as in Definition 3.1). This, together with (4.10), implies that

$$\sum_{i\in I_k(\omega)}\left(f_i^+(x_k)\right)^{\frac{2}{\beta_i}} \ge M^{1-\frac{2}{\beta_{\inf}}}\left(\sum_{i\in I_k(\omega)} f_i^+(x_k)\right)^{\frac{2}{\beta_{\inf}}}.$$

Then by the elementary of probability theory and the convexity of $t^{\frac{2}{\beta_{\inf}}}$ on $\mathbb{R}_+$ (as $\frac{2}{\beta_{\inf}} > 1$), we obtain that

$$\begin{aligned}\mathbb{E}\left\{\sum_{i\in I_k(\omega)}\left(f_i^+(x_k)\right)^{\frac{2}{\beta_i}}|\mathscr{F}_k\right\} &\ge M^{1-\frac{2}{\beta_{\inf}}}\mathbb{E}\left\{\left(\sum_{i\in I_k(\omega)} f_i^+(x_k)\right)^{\frac{2}{\beta_{\inf}}}|\mathscr{F}_k\right\}\\ &\ge M^{1-\frac{2}{\beta_{\inf}}}\left(\mathbb{E}\left\{\sum_{i\in I_k(\omega)} f_i^+(x_k)|\mathscr{F}_k\right\}\right)^{\frac{2}{\beta_{\inf}}}.\end{aligned}$$

Note by Assumption 4.2 that (4.6) is satisfied; this, together with Lemma 4.1(ii) and (4.9), yields that

$$\mathbb{E}\{\mathscr{D}_\varphi(x,x_{k+1})|\mathscr{F}_k\} \le \mathscr{D}_\varphi(x,x_k) - \underline{v}\left(1-\frac{\overline{v}}{\sigma}\right)\mu M\left(\eta M L_{\sup}\right)^{-\frac{2}{\beta_{\inf}}}\mathscr{D}_\varphi^{\frac{2p}{\beta_{\inf}}}(S,x_k).$$

That is, (4.8) is shown to hold by the definition of ρ, and the proof is complete. □

4.2. **Global convergence theorem.** This subsection aims to establish the global convergence theorem of the stochastic quasi-subgradient method with Bregman projection for the SQFP (3.1). To this end, we recall the following supermartingale convergence theorem, which is useful in the establishment of global convergence theorem.

Theorem 4.1 ([7, pp. 148])**.** *Let $\{Y_k\}$, $\{Z_k\}$ and $\{W_k\}$ be three sequences of random variables, and let $\{\mathscr{F}_k\}$ be a sequence of sets of random variables such that $\mathscr{F}_k \subseteq \mathscr{F}_{k+1}$ for any $k \in \mathbb{N}$. Suppose for any $k \in \mathbb{N}$ that*

(a) *Y_k, Z_k and W_k are functions of nonnegative random variables in $\mathscr{F}_k$;*
(b) *$\mathbb{E}\{Y_{k+1} \mid \mathscr{F}_k\} \leq Y_k - Z_k + W_k$;*
(c) *$\sum_{k=1}^{\infty} W_k < \infty$.*

Then $\sum_{k=1}^{\infty} Z_k < \infty$ and $\{Y_k\}$ converges to a nonnegative random variable with probability 1.

By virtue of the basic inequality in Lemma 4.2 and the supermartingale convergence theorem, we establish the global convergence theorem of the stochastic quasi-subgradient method with Bregman projection as follows.

Theorem 4.2. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1. Suppose that Assumptions 3.1, 4.1 and 4.2 hold. Then $\{x_k\}$ converges to a feasible solution of SQFP (3.1) with probability 1.*

Proof. By assumptions made in this theorem, Lemma 4.2 is applicable to ensuring (4.8). Then, by applying Theorem 4.1 (to $\|x_k - x\|$, $\rho\, \mathscr{D}_{\varphi}^{\frac{2p}{\beta_{\inf}}}(S, x_k)$, 0 in place of Y_k, Z_k, W_k), we obtain that

$$\{\|x_k - x\|\} \text{ is convergent and } \sum_{k=0}^{\infty} \mathscr{D}_{\varphi}^{\frac{2p}{\beta_{\inf}}}(S, x_k) < \infty \text{ with probability 1.}$$

Consequently, one has $\lim_{k\to\infty} \mathscr{D}_{\varphi}(S, x_k) = 0$, and then by (2.2) that $\lim_{k\to\infty} \operatorname{dist}(x_k, S) = 0$, and hence $\{x_k\}$ has a cluster point $x^* \in S$, with probability 1. This, together with Lemma 4.1(ii) and (2.2), shows that $\{x_k\}$ converges to $x^* \in S$ with probability 1. The proof is complete. □

Remark 4.2. Theorem 4.2 shows the global convergence (with probability 1) of the stochastic quasi-subgradient method with Bregman projection to a feasible solution of SQFP (1.1) under the assumptions of Hölder continuity and Hölder-type error bound property. It extends [22, Theorem 4.2] and [20, Theorems 3.1 and 3.4] to the infinite inequalities constraints situation and the general stochastic control. However, a slight defect in Theorem 4.2 is that an addition assumption of Hölder-type error bound property is required to guarantee the global convergence property. This is because the condition expectation on the second term of the right-hand side of (4.1) can be estimated in finite inequalities constraints situation, but cannot in the case of infinite constraints, without Assumption 4.2.

4.3. **Convergence rate analysis.** The establishment of convergence rate is significant in guaranteeing the numerical performance of relevant algorithms. This part is devoted to the convergence rate analysis for the stochastic quasi-subgradient method with Bregman projection.

Theorem 4.3. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1. Suppose that Assumptions 3.1, 4.1 and 4.2 hold. Then the following assertions are true.*

(i) *If* $2p = \beta_{\inf}$, *then there exist* $c \geq 0$ *and* $\tau \in (0,1)$ *such that*

$$\mathbb{E}\{\mathrm{dist}(x_k, S)\} \leq c\tau^k \quad \textit{for each } k \in \mathbb{N}. \tag{4.11}$$

(ii) *If* $2p > \beta_{\inf}$, *then there exists* $c \geq 0$ *such that*

$$\mathbb{E}\{\mathrm{dist}(x_k, S)\} \leq ck^{-\frac{\beta_{\inf}}{2(2p-\beta_{\inf})}} \quad \textit{for each } k \in \mathbb{N}. \tag{4.12}$$

Proof. By the made assumptions, Lemma 4.2 is applicable. Let $N \in \mathbb{N}$ and $\rho > 0$ be given in Lemma 4.2, and fix $k \geq N$ and $x := \arg\min_{z\in S} \mathscr{D}_\varphi(z, x_k)$. Then (4.8) is reduced to

$$\mathbb{E}\{\mathscr{D}_\varphi(S, x_{k+1})|\mathscr{F}_k\} \leq \mathscr{D}_\varphi(S, x_k) - \rho\, \mathscr{D}_\varphi^{\frac{2p}{\beta_{\inf}}}(S, x_k). \tag{4.13}$$

Taking the expectation on (4.13), we derive by the elementary of probability theory and the convexity of $t^{\frac{2p}{\beta_{\inf}}}$ on $\mathbb{R}_+$ (as $2p \geq \beta_{\inf}$) that

$$\mathbb{E}\left\{\mathscr{D}_\varphi(S, x_{k+1})\right\} \leq \mathbb{E}\left\{\mathscr{D}_\varphi(S, x_k)\right\} - \rho\, \left(\mathbb{E}\left\{\mathscr{D}_\varphi(S, x_k)\right\}\right)^{\frac{2p}{\beta_{\inf}}} \quad \text{for each } k \geq N. \tag{4.14}$$

(i) Suppose that $2p = \beta_{\inf}$. Then one has by (4.14) that

$$\mathbb{E}\left\{\mathscr{D}_\varphi(S, x_{k+1})\right\} \leq (1-\rho)\mathbb{E}\left\{\mathscr{D}_\varphi(S, x_k)\right\} \quad \text{for each } k \geq N.$$

Let $\tau := \sqrt{1-\rho}$ and $c := \sqrt{\frac{2}{\sigma} \max_{k=1\ldots,N} \left\{(1-\rho)^{-k}\mathbb{E}\left\{\mathscr{D}_\varphi(S, x_k)\right\}\right\}}$. Then it follows from (2.2) that

$$\mathbb{E}\left\{\mathrm{dist}^2(x_k, S)\right\} \leq \frac{2}{\sigma}\mathbb{E}\left\{\mathscr{D}_\varphi(S, x_k)\right\} \leq c^2\tau^{2k} \quad \text{for each } k \in \mathbb{N}.$$

This, together with the fact that

$$(\mathbb{E}\{\mathrm{dist}(x_k, S)\})^2 \leq \mathbb{E}\left\{\mathrm{dist}^2(x_k, S)\right\}, \tag{4.15}$$

deduces (4.11), and thus assertion (i) is proved.

(ii) Suppose that $2p > \beta_{\inf}$. Then, by applying Lemma 2.3 (with $\mathbb{E}\left\{\mathscr{D}_\varphi(S, x_k)\right\}$, ρ, $\frac{2p}{\beta_{\inf}} - 1$ in place of u_k, b, r) to (4.14), we obtain by (2.2) that there exists $c \geq 0$ such that

$$\mathbb{E}\left\{\mathrm{dist}^2(x_k, S)\right\} \leq \frac{2}{\sigma}\mathbb{E}\left\{\mathscr{D}_\varphi(S, x_k)\right\} \leq c^2k^{-\frac{\beta_{\inf}}{(2p-\beta_{\inf})}} \quad \text{for each } k \in \mathbb{N}.$$

This, together with (4.15), implies (4.12). The proof is complete. □

Remark 4.3. (i) Theorem 4.3 quantitatively estimates the convergence rates of the stochastic quasi-subgradient method with Bregman projection under some mild conditions. Particularly, Theorem 4.3(i) shows a linear convergence rate for the stochastic quasi-subgradient method with Bregman projection if each quasi-convex function in (3.1) is Lipschitz continuous and the SQFP satisfies the error bound property with $p = \frac{1}{2}$, that is Assumptions 3.1 and 4.2 with $2p = \beta_i \equiv 1$. Theorem 4.3(ii) exhibits a sublinear convergence rate $\mathscr{O}(k^{-\frac{\beta_{\inf}}{2(2p-\beta_{\inf})}})$ for the SQFP satisfying the general Hölder continuity and Hölder-type error bound property.

(ii) By Remark 3.1, Theorem 4.3 extends most of existing convergence results of subgradient methods for feasibility problems. In particular, when I is finite, Theorem 4.3 is applicable to establish the linear/sublinear convergence rates of subgradient methods for the CFP [5, Theorems 7.18 and 7.36] and the QFP [19]. When I is infinite, Theorem 4.3 is able to show the linear convergence rate of the stochastic subgradient method for the SCFP [28, Proposition 3].

Acknowledgments

The authors are grateful to the editor and the anonymous reviewers for their valuable comments toward the improvement of this paper. The first author was supported by the National Natural Science Foundation of China (12571327, 12222112, 12426311, and 32170655), Natural Science Foundation of Guangdong Provincial (2023A1515012395), Project of Educational Commission of Guangdong Province (2023ZDZX1017), Shenzhen Science and Technology Program (RCJC20221008092753082, RCYX20231211090222026, and JCYJ20241202124209011), Shenzhen Medical Research Fund (B2502001), and Research Team Cultivation Program of Shenzhen University (2023QNT011). 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/P02/21, UGC/FDS14/P07/22, and UGC/FDS14/P05/23).

REFERENCES

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

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

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

[4] M. Avriel, W. 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 (1996), 367–426.

[6] D. P. Bertsekas, Convex Optimization ang Algorithms, Athena Scientific, Massachusetts, 2015.

[7] D. P. Bertsekas and J. N. Tsitsiklis, Neuro-Dynamic Programming, Athena Scientific, Belmont, 1996.

[8] J. Bolte, T. P. Nguyen, J. Peypouquet, and B. W. Suter, From error bounds to the complexity of first-order descent methods for convex functions, Mathematical Programming, 165 (2017), 471–507.

[9] L. Bottou, F. E. Curtis, and J. Nocedal, Optimization methods for large-scale machine learning, SIAM Review, 60 (2018), 223–311.

[10] 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 (2022), 249–271.

[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 (2005), 2071–2084.

[12] Y. Censor and A. Segal, Algorithms for the quasiconvex feasibility problem, Journal of Computational and Applied Mathematics, 185 (2005), 34–50.

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

[14] J. Crouzeix, J.E.Martinez-Legaz, and M.Volle, Generalized Convexity, Generalized Monotonicity, Springer, Dordrecht, 1998.

[15] S. D. Flam, Successive averages of firmly nonexpansive mappings, Mathematics of Operations Research, 20 (1995), 497–512.

[16] J.-L. Goffin, Z. Q. Luo, and Y. Ye, On the complexity of a column generation algorithm for convex or quasiconvex feasibility problems, In: Large Scale Optimization: State of the Art, pp. 182–191. Kluwer Academic Publishers, 1994.

[17] N. Hadjisavvas, S. Komlósi, and S. Schaible, Handbook of Generalized Convexity and Generalized Monotonicity, Springer, New York, 2005.

[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 (2016), 1207–1235.

[19] 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 (2022), 45–58.

[20] Y. Hu, J. Li, Y. Liu, and C. K. W. Yu, Quasi-subgradient methods with Bregman distance for quasi-convex feasibility problems, Journal of Nonlinear and Variational Analysis, 8 (2024), 381–395.

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

[22] Y. Hu, C. K. W. Yu, and X. Yang, Incremental quasi-subgradient methods for minimizing the sum of quasi-convex functions, Journal of Global Optimization, 75 (2019), 1003–1028.

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

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

[25] V. I. Kolobov, S. Reich, and R. Zalas, Finitely convergent deterministic and stochastic methods for solving convex feasibility problems, Mathematical Programming, 194 (2022), 1163–1183.

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

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

[28] A. Nedić, Random algorithms for convex minimization problems, Mathematical Programming, 129 (2021), 225–253.

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

[30] J.-S. Pang, Error bounds in mathematical programming, Mathematical Programming, 79 (1987), 299–332.

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

[32] B. T. Polyak, Random algorithms for solving convex inequalities, Studies in Computational Mathematics, 8 (2001), 409–422.

[33] 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 (2021), bbab311.

[34] I. Stancu-Minasian, Fractional Programming, Kluwer Academic Publisher, Dordrecht, 1997.

[35] 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 (2019), 327–358.

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