

STOCHASTIC QUASI-SUBGRADIENT METHOD FOR STOCHASTIC QUASI-CONVEX FEASIBILITY PROBLEMS

GANG LI

Department of Mathematics, Zhejiang Sci-Tech University
Hangzhou 310018, China

MINGHUA LI

School of Mathematics and Big Data, Chongqing University of Arts and Sciences
Yongchuan, Chongqing, 402160, China

YAOHUA HU*

Shenzhen Key Laboratory of Advanced Machine Learning and Applications
College of Mathematics and Statistics, Shenzhen University
Shenzhen 518060, China

(Communicated by the associate editor name)

ABSTRACT. The feasibility problem is at the core of the modeling of many problems in various disciplines of mathematics and physical sciences, and the quasi-convex function is widely applied in many fields such as economics, finance, and management science. In this paper, we consider the stochastic quasi-convex feasibility problem (SQFP), which is to find a common point of infinitely many sublevel sets of quasi-convex functions. Inspired by the idea of a stochastic index scheme, we propose a stochastic quasi-subgradient method 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 at each iteration. Moreover, we introduce a notion of Hölder-type error bound property relative to the random control sequence for the SQFP, and use it to establish the global convergence theorem and convergence rate theory of the stochastic quasi-subgradient method. It is revealed in this paper that the stochastic quasi-subgradient method enjoys both advantages of low computational cost requirement and fast convergence feature.

2020 *Mathematics Subject Classification.* Primary: 65K05, 90C26; Secondary: 49M37.

Key words and phrases. Stochastic feasibility problem, quasi-convex programming, subgradient method, random control, convergence theory.

The first author is supported in part by the Zhejiang Provincial Natural Science Foundation of China (LY18A010030), Scientific Research Fund of Zhejiang Provincial Education Department (19060042-F) and Science Foundation of Zhejiang Sci-Tech University (19062150-Y).

The second author is supported in part by the Foundation for High-level Talents of Chongqing University of Art and Sciences (P2017SC01), Chongqing Key Laboratory of Group and Graph Theories and Applications, and Key Laboratory of Complex Data Analysis and Artificial Intelligence of Chongqing Municipal Science and Technology Commission.

The third author is supported in part by the National Natural Science Foundation of China (12071306, 11871347), Natural Science Foundation of Guangdong Province of China (2019A1515011917, 2020B1515310008, 2020A1515010372), Project of Educational Commission of Guangdong Province of China (2019KZDZX1007), Natural Science Foundation of Shenzhen (JCYJ20190808173603590) and Interdisciplinary Innovation Team of Shenzhen University.

* Corresponding author: Yaohua Hu.

1. **Introduction.** Let I be an (finite or infinite) index set, and let $\{f_i : i \in I\}$ 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

$$f_i(x) \leq 0 \quad \text{for each } i \in I. \quad (1)$$

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 [13], wireless sensor networks localization [19], radiation therapy treatment planning [11] and gene regulatory network inference [35].

When I is finite, problem (1) is the classical *deterministic feasibility problem*. In extensive practical problems, functions involved in (1) are assumed to be convex, and the corresponding problem is called the *convex feasibility problem* (CFP). Motivated by its extensive applications, tremendous efforts have been devoted to the development of optimization algorithms for solving the CFP. One of the most popular approaches is the classical *subgradient method*. Many extensions on control schemes (cyclic/parallel/most-violated control schemes) and various convergence features of subgradient methods have been devised and well explored; see [3, 11, 13, 35, 37] and references therein.

In a number of applications (e.g., the robust stabilization and control, and the integral equations system), the index set I in problem (1) is always infinite, then the corresponding problem (1) is called the *stochastic feasibility problem* [10, 15, 33]. The classical subgradient method is not implementable for the stochastic feasibility problem; particularly, the cyclic control never completes the first iteration and finding the most violated control is an intractable task because of the infinite inequalities constraints. Inspired by the idea of stochastic index scheme [8, 29, 36] in optimization, stochastic subgradient methods¹ have been proposed and investigated for solving stochastic convex feasibility problems (SCFP), in which each component function in (1) is convex; see [26, 29, 33] and references therein. In particular, the global convergence and linear convergence rate of the stochastic subgradient method to a solution of SCFP (1) (with probability 1) was established in [29] under an assumption of the error bound property; the finite-termination convergence theorem was established in [26, 33] under assumptions of the strong feasibility and the distinguishability of feasible and infeasible points.

Most literature mentioned above considered the feasibility problem (1) with convex components. However, the convex function is too restrictive in many real-life problems encountered in economics, finance and management science. Compared with the convex function, the *quasi-convex function* usually provides a much more accurate representation of realities and still possesses certain desirable properties of the convex function. For instance, the fractional function, characterized by a ratio of technical terms (e.g., efficiency), is a typical class of quasi-convex (but non-convex) functions, which has been widely applied in various areas; see [2, 34] and references therein. This leads to a significant increase of studies in quasi-convex optimization; see [2, 14, 18, 31, 34] and references therein.

When I is finite and functions involved in (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

¹Randomized projection methods [10, 15, 28] can be understood as a special case of stochastic subgradient methods with certain stepsize.

et al. [20, 23] proposed the quasi-subgradient methods with cyclic/parallel/most-violated/stochastic control schemes to solve the nondifferentiable QFP, and established their global convergence to a feasible solution of the QFP. However, to the best of our knowledge, there is still no paper devoted to the study of optimization algorithms for solving the *stochastic quasi-convex feasibility problem* (SQFP).

In the present paper, we consider the SQFP (1), where I is infinite and the involved functions are quasi-convex, and aim to propose a stochastic quasi-subgradient method to solve SQFP (1). In the stochastic quasi-subgradient method, the quasi-subgradients of a random (and finite) index set of component functions at the current iterate are used to construct the descent direction 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 for solving the SQFP (1). In particular, we first introduce a notion of Hölder-type error bound property relative to a random control sequence for the SQFP, which extends and loosens the Lipschitz-type error bound property introduced in [29, Assumption 2] (see Remark 3). Moreover, we use it to establish the global convergence of the stochastic quasi-subgradient method to a feasible solution of the SQFP (1) with probability 1 (see Theorem 3.10) and to quantitatively estimate the convergence rates (see Theorem 3.11). The established convergence theory extends most of existing convergence results of subgradient methods for the CFP [3, Theorems 7.18 and 7.36], the QFP [20] and the SCFP [29, Proposition 3]; see Remark 3. As far as we know, the proposed stochastic quasi-subgradient method and the established convergence theory are new in the literature of SQFP.

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 propose a stochastic quasi-subgradient method to solve SQFP (1) and investigate its global convergence theorem and convergence rate theory.

2. Notations and preliminary results. The 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 Euclidean norm $\|\cdot\|$. As usual, 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. The nonnegative orthant and the (positive) unit simplex in \mathbb{R}^n are denoted by \mathbb{R}_+^n and Δ_+^n , respectively; that is,

$$\Delta_+^n := \{\lambda \in \mathbb{R}_{++}^n : \sum_{i=1}^n \lambda_i = 1\}.$$

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

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

For $x \in \mathbb{R}^n$ and $Z \subseteq \mathbb{R}^n$, we use $\text{dist}(x, Z)$ and $\mathbb{P}_Z(x)$ to denote the Euclidean distance of x from Z and the Euclidean projection of x onto Z , respectively, that is,

$$\text{dist}(x, Z) := \inf_{z \in Z} \|x - z\| \quad \text{and} \quad \mathbb{P}_Z(x) := \arg \min_{z \in Z} \|x - z\|.$$

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

$$f(\alpha x + (1 - \alpha)y) \leq \max\{f(x), f(y)\} \quad \text{for any } x, y \in \mathbb{R}^n \text{ and } \alpha \in [0, 1].$$

The sublevel sets of f at x are denoted by

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

A convex function can be characterized by the convexity of its epigraphs, while a quasi-convex function can be characterized by the convexity of its sublevel sets. In particular, f is quasi-convex if and only if $\text{lev}_f^<(x)$ (and/or $\text{lev}_f^{\leq}(x)$) is convex for any $x \in \mathbb{R}^n$.

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 [1, 17, 22] and references therein. In particular, Kiwiel [25], Censor and Segal [12], and Hu et al. [22] utilized the following quasi-subgradient to develop and analyze quasi-subgradient methods for solving quasi-convex optimization problems.

Definition 2.1. The quasi-subdifferential of $f : \mathbb{R}^n \rightarrow \mathbb{R}$ at $x \in \mathbb{R}^n$ is defined by

$$\partial^* f(x) := \{g \in \mathbb{R}^n : \langle g, y - x \rangle \leq 0 \text{ for any } y \in \text{lev}_f^<(x)\}.$$

Any vector $g \in \partial^* f(x)$ is called a quasi-subgradient of f at x .

The nonemptiness of specific subdifferential is an essential property for a certain type of functions, e.g., the convex subdifferential for the convex functions. It was proved in [22, Lemma 2.1] that the quasi-subdifferential of a quasi-convex function is nontrivial, that is, $\partial^* f(x) \setminus \{0\} \neq \emptyset$ for each $x \in \mathbb{R}^n$. Noting by Definition 2.1 that $\partial^* f(x)$ is a normal cone to its sublevel set, it can be derived from [22, Lemma 2.1] that the quasi-subdifferential of a quasi-convex function contains at least a unit vector. This is a special property of the quasi-subdifferential beyond the convex subdifferential (for a convex function). Moreover, it was claimed in [22] that the quasi-subdifferential coincides with the convex cone hull of the convex subdifferential whenever f is convex.

The Hölder condition of order β was used in [27] to provide a fundamental property of the quasi-subgradient, and plays an important role in the establishment of a basic inequality in convergence analysis of subgradient-type algorithms for quasi-convex optimization problems; see, e.g., [21, 22, 27]. The Hölder condition with order 1 is reduced to the Lipschitz condition, and this property holds for very broad classes of functions in economics and management science with various values of $\beta \leq 1$.

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

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

The following lemma extends a fundamental property of quasi-convex optimization in [27, Proposition 2.1] to the quasi-convex feasibility problem under the Hölder condition.

Lemma 2.3 ([20, Lemma 2.1]). *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a quasi-convex and continuous function, X be a closed and convex set, and let $S := \{x \in X : f(x) \leq 0\}$. Let*

$0 < \beta \leq 1$ 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 $y \in X \setminus S$, it holds that

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

We end this section by recalling the following two lemmas, which will be useful in convergence analysis of the stochastic quasi-subgradient method.

Lemma 2.4 ([24, Lemma 4.1]). *Let $\gamma \geq 1$ and $a_i \geq 0$ for $i = 1, \dots, n$. Then it holds that*

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

Lemma 2.5 ([32, pp. 46, Lemma 6]). *Let $r > 0$ and $b > 0$, and $\{u_k\} \subseteq \mathbb{R}_+$ be a sequence of nonnegative scalars such that*

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

Then it holds that

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

3. 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 and $C \subseteq \mathbb{R}^n$ 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

$$x \in C \quad \text{and} \quad f_i(x) \leq 0 \quad \text{for each } i \in I. \quad (2)$$

Let (I, \mathcal{F}, \Pr) be a complete probability space. By the formulation given in [10, 15], an equivalent representation of the SQFP can be written as finding a feasible point $x \in \mathbb{R}$ such that

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

When I is a finite index set, problem (2) is reduced to the classical QFP [12, 16, 20].

As usual, we assume throughout the whole paper that the SQFP is consistent, i.e., the solution set of the SQFP (2) is nonempty:

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

Moreover, we always assume that each component function of the SQFP (2) 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

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

3.1. Stochastic quasi-subgradient method. One of the most popular and practical optimization algorithms for solving the feasibility problem is a class of subgradient methods; see [3, 12, 20, 37] and references therein. However, for the SQFP (2) where the index set I is infinite, the typical deterministic control schemes in the classical subgradient method are not implementable. Particularly, the parallel control consumes expensive computational cost, the cyclic control never completes the first iteration, and finding the most violated control is an intractable task 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 [4] or a large number of constraints [29]. A typical and very popular example is the stochastic gradient descent (SGD) algorithm in machine learning [8], in which only one component function is randomly selected to construct the descent direction at each iteration. It was pointed out in [20] that the stochastic control enjoys both advantages of low computational cost requirement and low (worst-case) iteration complexity.

Inspired by the idea of stochastic index scheme, we will propose a stochastic quasi-subgradient method for solving the SQFP (2). In the stochastic quasi-subgradient method, the random control sequence is recalled from [26] as follows.

Definition 3.2. Let $(\Omega, \mathcal{F}, \Pr)$ be a given probability space. The sequence $\{I_k\}$ is said to be a random control sequence in I if $I_k : \Omega \rightarrow 2^I \setminus \{\emptyset\}$ are independent and identically distributed (set-valued) random variables on $(\Omega, \mathcal{F}, \Pr)$ with $M := \sup_{\omega \in \Omega, k \in \mathbb{N}} |I_k(\omega)| < \infty$.

Remark 1. As described in Definition 3.2, $\{I_k(\omega)\}$ is a nonadaptive control sequence in I for each $\omega \in \Omega$. The terminology ‘‘identically distributed’’ means that

$$\Pr(\{\omega \in \Omega : I_k(\omega) = J\}) = \Pr(\{\omega \in \Omega : I_n(\omega) = J\})$$

for any $k, n \in \mathbb{N}$ and any nonempty $J \subseteq I$ with $|J| \leq M$; and the terminology ‘‘independent’’ means that

$$\Pr\left(\bigcap_{k \in K} \{\omega \in \Omega : I_k(\omega) = J_k\}\right) = \prod_{k \in K} \Pr(\{\omega \in \Omega : I_k(\omega) = J_k\})$$

for any finite index set K and any nonempty $J_k \subseteq I$ with $|J_k| \leq M$.

Integrating the idea of stochastic index scheme [26, 29] into the quasi-subgradient method [12, 20], we propose the following stochastic quasi-subgradient method to solve the SQFP (2). In particular, the quasi-subgradients of a random index set of component functions are selected to construct the descent direction at each iteration. Recall that $\{(\beta_i, L_i)\}$ are the parameters given in Assumption 3.1.

Algorithm 3.3. Select an initial point $x_1 \in \mathbb{R}^n$ and a sequence of stepsizes $\{v_k\} \subseteq (0, +\infty)$ satisfying

$$0 < \underline{v} \leq v_k \leq \bar{v} < 2, \quad (3)$$

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 $g_{k,i} \in \partial^* f_i(x_k) \cap \mathbb{S}$ for each $i \in I_k(\omega)$, and update x_{k+1} by

$$x_{k+1} := \mathbb{P}_C \left(x_k - v_k \sum_{i \in I_k(\omega)} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i} \right)^{\frac{1}{\beta_i}} g_{k,i} \right). \quad (4)$$

Remark 2. Algorithm 3.3 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, Algorithm 3.3 is reduced to the stochastic subgradient methods for solving the CFP [33] and the QFP [23].

(ii) When I is infinite, each f_i is convex and $\{I_k\}$ is single-valued, Algorithm 3.3 is reduced to the stochastic subgradient methods for solving the SCFP [29].

The remainder of this paper is contributed to the convergence analysis of the stochastic quasi-subgradient method (i.e., Algorithm 3.3). To guarantee the convergence property of Algorithm 3.3, we shall assume the following condition on the weights $\{\lambda_{k,i}\}$; see [3, Remark 3.13]. A natural example is the naive weights, i.e., $\lambda_{k,i} = \frac{1}{|I_k(\omega)|}$ for each $i \in I_k(\omega)$, which satisfies Assumption 3.4 automatically.

Assumption 3.4. There exists $\mu > 0$ such that $\min_{i \in I_k(\omega)} \lambda_{k,i} \geq \mu$ for any $\omega \in \Omega$ and $k \in \mathbb{N}$.

3.2. Basic inequality. The basic inequality shows a significant property and plays a key tool in convergence analysis of subgradient methods. We start from the basic inequality of Algorithm 3.3, which is able to derive some basic properties of the stochastic quasi-subgradient method.

Lemma 3.5. *Let $x \in S$, $\omega \in \Omega$, and let $\{x_k\}$ be a sequence generated by Algorithm 3.3. Suppose that Assumptions 3.1 and 3.4 hold. Then the following assertions are true.*

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

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 - \underline{v}(2 - \bar{v})\mu L_{\sup}^{-\frac{2}{\beta_{\inf}}} \sum_{i \in I_k(\omega)} (f_i^+(x_k))^{\frac{2}{\beta_i}}. \quad (5)$$

(ii) *$\{\|x_k - x\|\}$ is monotonically decreasing, and hence $\{x_k\}$ is bounded.*

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

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

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 - v_k(2 - v_k) \sum_{i \in I_k(\omega)} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i} \right)^{\frac{2}{\beta_i}}. \quad (6)$$

Granting this, (5) directly follows by (3) and Assumptions 3.1 and 3.4, and thus assertion (i) is proved.

To show (6), without loss of generality, we assume that $x_k \notin S$; otherwise, $f_i^+(x_k) = 0$ for each $i \in I$ and then (4) generates $x_{k+1} = x_k$, hence (6) is satisfied automatically. By (4) and the nonexpansive property of the projection operator, we obtain that

$$\begin{aligned} \|x_{k+1} - x\|^2 &\leq \|x_k - v_k \sum_{i \in I_k(\omega)} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i} \right)^{\frac{1}{\beta_i}} g_{k,i} - x\|^2 \\ &= \|x_k - x\|^2 - 2v_k \sum_{i \in I_k(\omega)} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i} \right)^{\frac{1}{\beta_i}} \langle x_k - x, g_{k,i} \rangle \\ &\quad + v_k^2 \left\| \sum_{i \in I_k(\omega)} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i} \right)^{\frac{1}{\beta_i}} g_{k,i} \right\|^2. \end{aligned} \quad (7)$$

Note by $x \in S$ that $f_i(x) \leq 0$ for each $i \in I$. Then, for each $i \in I_k(\omega)$ such that $f_i(x_k) > 0$, it follows from Lemma 2.3 that $\langle x_k - x, g_{k,i} \rangle \geq \left(\frac{f_i^+(x_k)}{L_i} \right)^{\frac{1}{\beta_i}}$; otherwise,

$f_i^+(x_k) = 0$. Hence we conclude that

$$\sum_{i \in I_k(\omega)} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i} \right)^{\frac{1}{\beta_i}} \langle x_k - x, g_{k,i} \rangle \geq \sum_{i \in I_k(\omega)} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i} \right)^{\frac{2}{\beta_i}}. \quad (8)$$

On the other side, we obtain by the convexity of $\|\cdot\|^2$ that

$$\left\| \sum_{i \in I_k(\omega)} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i} \right)^{\frac{1}{\beta_i}} g_{k,i} \right\|^2 \leq \sum_{i \in I_k(\omega)} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i} \right)^{\frac{2}{\beta_i}},$$

(thanks to $\{\lambda_{k,i}\}_{i \in I_k(\omega)} \subseteq \Delta_+^{|I_k(\omega)|}$ and $g_{k,i} \in \mathbb{S}$). This, together with (7) and (8), deduces (6), as desired.

(ii) It is a direct consequence of (5) showed in assertion (i).

(iii) It follows from (5) that

$$\begin{aligned} \sum_{k=1}^{\infty} \sum_{i \in I_k(\omega)} (f_i^+(x_k))^{\frac{2}{\beta_i}} &\leq \frac{1}{\underline{\nu}(2-\bar{\nu})\mu} L_{\text{sup}}^{\frac{2}{\beta_{\text{inf}}}} \left(\|x_1 - x\|^2 - \lim_{k \rightarrow \infty} \|x_k - x\|^2 \right) \\ &\leq \frac{1}{\underline{\nu}(2-\bar{\nu})\mu} L_{\text{sup}}^{\frac{2}{\beta_{\text{inf}}}} \|x_1 - x\|^2 < \infty. \end{aligned}$$

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

The error bound property [30] plays an important role and has been extensively used in convergence analysis of various numerical algorithms; see [6, 7, 22, 29, 35] and references therein. Below, we introduce a notion of the Hölder-type error bound property relative to a random control sequence for the SQFP (2). In particular, to the random control sequence $\{I_k\}$, the sigma-field $\{\mathcal{F}_k\}$ records the history of the method, that is,

$$\mathcal{F}_k := \{x_1, I_1(\omega), \dots, I_{k-1}(\omega)\} \quad \text{for each } k \in \mathbb{N}. \quad (9)$$

Definition 3.6. The SQFP (2) is said to satisfy the Hölder-type error bound property of order $p \geq 1$ relative to the random control sequence $\{I_k\}$ if there exists $\eta > 0$ such that

$$\text{dist}^p(x, S) \leq \eta \mathbb{E} \left\{ \sum_{i \in I_k(\omega)} f_i^+(x) | \mathcal{F}_k \right\} \quad \text{for any } x \in C \text{ and } k \in \mathbb{N}. \quad (10)$$

In particular, the SQFP (2) is said to satisfy the (Lipschitz-type) error bound property relative to the random control sequence $\{I_k\}$ if (10) holds with $p = 1$.

Remark 3. (i) In the case when $I_k(\omega) = \{\omega_k\}$ is single-valued and $p = 1$, the error bound property (10) is reduced to [29, Assumption 2]:

$$\text{dist}(x, S) \leq \eta \mathbb{E} \{ f_{\omega_k}^+(x) | \mathcal{F}_k \} \quad \text{for any } x \in C \text{ and } k \in \mathbb{N}, \quad (11)$$

which was used in [29] to explore the stochastic subgradient method for the SCFP. Clearly, the Hölder-type error bound property of order $p \geq 1$ extends and loosens the condition (11) (when x is close to S so that $\text{dist}(x, S) < 1$), because the larger the order p , the less restrictive the condition.

(ii) It was shown in [29] that (11) is a quite general assumption in the scenario of CFP. For example, when I is finite, (11) can be ensured by the linear regularity of constraint sets [3], the weak sharp minima property [9], and the global error bound of inequality system [30]; when I is arbitrary (finite or infinite), (11) is satisfied if

provided the Slater condition of the solution set; see [29, pp. 231]. Recalling from assertion (i), the Hölder-type error bound property (10) is weaker than (11), and hence can be guaranteed by regular conditions mentioned above.

To establish the convergence theory of the stochastic quasi-subgradient method, we shall assume the Hölder-type error bound property on the SQFP (2).

Assumption 3.7. The SQFP (2) satisfies the Hölder-type error bound property of order $p \geq 1$ relative to the random control sequence $\{I_k\}$.

By virtue of the basic inequality (5) and the Hölder-type error bound property, we provide the basic inequality in terms of conditional expectation for Algorithm 3.3.

Lemma 3.8. *Let $x \in S$, and let $\{x_k\}$ be a sequence generated by Algorithm 3.3 and $\{\mathcal{F}_k\}$ be defined by (9). Suppose that Assumptions 3.1, 3.4 and 3.7 hold, and let $\rho := \underline{\nu}(2 - \bar{\nu})\mu M (\eta M L_{\text{sup}})^{-\frac{2}{\beta_{\text{inf}}}}$. Then there exists $N \in \mathbb{N}$ such that the following basic inequality holds*

$$\mathbb{E}\{\|x_{k+1} - x\|^2 | \mathcal{F}_k\} \leq \|x_k - x\|^2 - \rho \text{dist}^{\frac{2p}{\beta_{\text{inf}}}}(x_k, S) \quad \text{for each } k \geq N. \quad (12)$$

Proof. By assumptions made in this lemma, Lemma 3.5 is applicable, and hence (5) holds for each $k \in \mathbb{N}$. Taking the conditional expectation of (5) with respect to \mathcal{F}_k , it follows that

$$\mathbb{E}\{\|x_{k+1} - x\|^2 | \mathcal{F}_k\} \leq \|x_k - x\|^2 - \underline{\nu}(2 - \bar{\nu})\mu L_{\text{sup}}^{-\frac{2}{\beta_{\text{inf}}}} \mathbb{E} \left\{ \sum_{i \in I_k(\omega)} (f_i^+(x_k))^{\frac{2}{\beta_i}} | \mathcal{F}_k \right\}. \quad (13)$$

Below, we estimate the conditional expectation term at the right hand side of (13). Indeed, by Lemma 3.5(iii), there exists $N \in \mathbb{N}$ such that $\sum_{i \in I_k(\omega)} (f_i^+(x_k))^{\frac{2}{\beta_i}} < 1$ for each $k \geq N$; consequently, $f_i^+(x_k) < 1$ for each $k \geq N$ and $i \in I_k(\omega)$. Fix $k \geq N$, and note by Assumption 3.1 that $\beta_i \geq \beta_{\text{inf}}$ for each $i \in I$. Then we obtain that

$$\sum_{i \in I_k(\omega)} (f_i^+(x_k))^{\frac{2}{\beta_i}} \geq \sum_{i \in I_k(\omega)} (f_i^+(x_k))^{\frac{2}{\beta_{\text{inf}}}}. \quad (14)$$

Noting by Assumption 3.1 that $\frac{2}{\beta_{\text{inf}}} > 1$, we can apply Lemma 2.4 (with $f_i^+(x_k)$, $\frac{2}{\beta_{\text{inf}}}$ in place of a_i , γ) to achieve that

$$\begin{aligned} \sum_{i \in I_k(\omega)} (f_i^+(x_k))^{\frac{2}{\beta_{\text{inf}}}} &\geq |I_k(\omega)|^{1 - \frac{2}{\beta_{\text{inf}}}} \left(\sum_{i \in I_k(\omega)} f_i^+(x_k) \right)^{\frac{2}{\beta_{\text{inf}}}} \\ &\geq M^{1 - \frac{2}{\beta_{\text{inf}}}} \left(\sum_{i \in I_k(\omega)} f_i^+(x_k) \right)^{\frac{2}{\beta_{\text{inf}}}} \end{aligned}$$

(thanks to $|I_k(\omega)| \leq M$ as in Definition 3.2). This, together with (14), implies that

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

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

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

Note by Assumption 3.7 that (10) is satisfied; this, together with (13), yields that

$$\mathbb{E} \{ \|x_{k+1} - x\|^2 | \mathcal{F}_k \} \leq \|x_k - x\|^2 - \underline{\nu}(2 - \bar{\nu})\mu M (\eta M L_{\text{sup}})^{-\frac{2}{\beta_{\text{inf}}}} \text{dist}^{\frac{2p}{\beta_{\text{inf}}}}(x_k, S).$$

That is, (12) is shown to hold by the definition of ρ , and the proof is complete. \square

3.3. Global convergence theorem. This subsection aims to establish the global convergence theorem of the stochastic quasi-subgradient method for the SQFP (2). To this end, we recall the following supermartingale convergence theorem, which is useful in the establishment of global convergence theorem.

Theorem 3.9 ([5, pp. 148]). *Let $\{Y_k\}$, $\{Z_k\}$ and $\{W_k\}$ be three sequences of random variables, and let $\{\mathcal{F}_k\}$ be a sequence of sets of random variables such that $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$ 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 \mathcal{F}_k ;
- (a) $\mathbb{E} \{ Y_{k+1} | \mathcal{F}_k \} \leq Y_k - Z_k + W_k$;
- (a) $\sum_{k=1}^{\infty} W_k < \infty$.

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 3.8 and the supermartingale convergence theorem, we establish the global convergence theorem of the stochastic quasi-subgradient method as follows.

Theorem 3.10. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.3. Suppose that Assumptions 3.1, 3.4 and 3.7 hold. Then $\{x_k\}$ converges to a feasible solution of SQFP (2) with probability 1.*

Proof. By assumptions made in this theorem, Lemma 3.8 is applicable to ensuring (12). Then, by applying Theorem 3.9 (to $\|x_k - x\|$, $\rho \text{dist}^{\frac{2p}{\beta_{\text{inf}}}}(x_k, S)$, 0 in place of Y_k, Z_k, W_k as $k \geq N$), we obtain that

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

Consequently, $\lim_{k \rightarrow \infty} \text{dist}(x_k, S) = 0$, and hence $\{x_k\}$ has a cluster point falling in S , with probability 1. This, together with Lemma 3.5(ii), shows that $\{x_k\}$ converges to this cluster point in S with probability 1. The proof is complete. \square

Remark 4. Theorem 3.10 shows the global convergence (with probability 1) of the stochastic quasi-subgradient method to a feasible solution of SQFP (1) under the assumptions of Hölder continuity and Hölder-type error bound property. It extends [23, Theorem 4.2] to the infinite inequalities constraints situation and the general

stochastic control. However, a slight defect in Theorem 3.10 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 (5) can be estimated in finite inequalities constraints situation, but cannot in the case of infinite constraints, without Assumption 3.7.

3.4. Convergence rate analysis. The establishment of convergence rate is significant in guaranteeing the numerical performance of relevant algorithms. This part is devoted to the establishment of convergence rates for the stochastic quasi-subgradient method.

Theorem 3.11. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.3. Suppose that Assumptions 3.1, 3.4 and 3.7 hold. Then the following assertions are true.*

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

$$\mathbb{E} \{ \text{dist}(x_k, S) \} \leq c\tau^k \quad \text{for each } k \in \mathbb{N}. \quad (15)$$

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

$$\mathbb{E} \{ \text{dist}(x_k, S) \} \leq ck^{-\frac{\beta_{\text{inf}}}{2(p-\beta_{\text{inf}})}} \quad \text{for each } k \in \mathbb{N}. \quad (16)$$

Proof. By the made assumptions, Lemma 3.8 is applicable. Let $N \in \mathbb{N}$ and $\rho > 0$ be given in Lemma 3.8, and fix $k \geq N$ and $x := \mathbb{P}_S(x_k)$. Then (12) is reduced to

$$\mathbb{E} \{ \text{dist}^2(x_{k+1}, S) \mid \mathcal{F}_k \} \leq \text{dist}^2(x_k, S) - \rho \text{dist}^{\frac{2p}{\beta_{\text{inf}}}}(x_k, S). \quad (17)$$

Taking the expectation on (17), we derive by the elementary of probability theory and the convexity of $t^{\frac{p}{\beta_{\text{inf}}}}$ on \mathbb{R}_+ (as $p \geq 1 \geq \beta_{\text{inf}}$) that

$$\mathbb{E} \{ \text{dist}^2(x_{k+1}, S) \} \leq \mathbb{E} \{ \text{dist}^2(x_k, S) \} - \rho \left(\mathbb{E} \{ \text{dist}^2(x_k, S) \} \right)^{\frac{p}{\beta_{\text{inf}}}} \quad \text{for each } k \geq N. \quad (18)$$

(i) Suppose that $p = \beta_{\text{inf}}$. Then one has by (18) that

$$\mathbb{E} \{ \text{dist}^2(x_{k+1}, S) \} \leq (1 - \rho) \mathbb{E} \{ \text{dist}^2(x_k, S) \} \quad \text{for each } k \geq N.$$

Let $\tau := \sqrt{1 - \rho}$ and $c := \sqrt{\max_{k=1, \dots, N} \{ (1 - \rho)^{-k} \mathbb{E} \{ \text{dist}^2(x_k, S) \} \}}$. Then it follows that

$$\mathbb{E} \{ \text{dist}^2(x_k, S) \} \leq c^2 \tau^{2k} \quad \text{for each } k \in \mathbb{N}.$$

This, together with the fact that

$$\left(\mathbb{E} \{ \text{dist}(x_k, S) \} \right)^2 \leq \mathbb{E} \{ \text{dist}^2(x_k, S) \}, \quad (19)$$

deduces (15), and thus assertion (i) is proved.

(ii) Suppose that $p > \beta_{\text{inf}}$. Then, by applying Lemma 2.5 (with $\mathbb{E} \{ \text{dist}^2(x_k, S) \}$, ρ , $\frac{p}{\beta_{\text{inf}}} - 1$ in place of u_k , b , r) to (18), we obtain that there exists $c \geq 0$ such that

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

This, together with (19), implies (16). The proof is complete. \square

Remark 5. (i) Theorem 3.11 quantitatively estimates the convergence rates of the stochastic quasi-subgradient method under some mild conditions. Particularly, Theorem 3.11(i) shows a linear convergence rate for the stochastic quasi-subgradient method if each quasi-convex function in (2) is Lipschitz continuous and the SQF-P satisfies the Lipschitz-type error bound property, that is Assumptions 3.1 and 3.7 with $p = \beta_i \equiv 1$. Theorem 3.11(ii) exhibits a sublinear convergence rate

$\mathcal{O}(k^{-\frac{\beta_{\text{inf}}}{2(p-\beta_{\text{inf}})}})$ for the SQFP satisfying the general Hölder continuity and Hölder-type error bound property.

(ii) By Remarks 2 and 3, Theorem 3.11 extends most of existing convergence results of subgradient methods for feasibility problems. In particular, when I is finite, Theorem 3.11 is applicable to establish the linear/sublinear convergence rates of subgradient methods for the CFP [3, Theorems 7.18 and 7.36] and the QFP [20]. When I is infinite, Theorem 3.11 is able to show the linear convergence rate of the stochastic subgradient method for the SCFP [29, Proposition 3].

REFERENCES

- [1] D. Aussel and M. Pištěk, Limiting normal operator in quasiconvex analysis, *Set-Valued and Variational Analysis*, **23** (2015), 669–685.
- [2] M. Avriel, W. Diewert, S. Schaible and I. Zang, *Generalized Concavity*, Plenum Press, New York, 1988.
- [3] H. H. Bauschke and J. M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM Review*, **38** (1996), 367–426.
- [4] D. P. Bertsekas, *Convex Optimization and Algorithms*, Athena Scientific, Massachusetts, 2015.
- [5] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, Belmont, 1996.
- [6] 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.
- [7] J. M. Borwein, G. Li and L. Yao, Analysis of the convergence rate for the cyclic projection algorithm applied to basic semialgebraic convex sets, *SIAM Journal on Optimization*, **24** (2014), 498–527.
- [8] L. Bottou, F. E. Curtis and J. Nocedal, Optimization methods for large-scale machine learning, *SIAM Review*, **60** (2018), 223–311.
- [9] J. V. Burke and M. C. Ferris, Weak sharp minima in mathematical programming, *SIAM Journal on Control and Optimization*, **31** (1993), 1340–1359.
- [10] D. Butnariu and S. D. Flăm, Strong convergence of expected-projection methods in Hilbert spaces, *Numerical Functional Analysis and Optimization*, **16** (1995), 601–636.
- [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** (2006), 34–50.
- [13] P. L. Combettes, The convex feasibility problem in image recovery, *Advances in Imaging and Electron Physics*, **95** (1996), 155–270.
- [14] J. Crouzeix, J. E. Martinez-Legaz and M. Volle, *Generalized Convexity, Generalized Monotonicity*, Springer, Dordrecht, 1998.
- [15] S. D. Flăm, 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* (eds. W. W. Hager, D. W. Hearn and P. M. Pardalos), Kluwer Academic Publishers, Dordrecht, (1994), 182–191.
- [17] H. J. Greenberg and W. P. Pierskalla, Quasiconjugate functions and surrogate duality, *Cahiers Centre Études Recherche Opérationnelle*, **15** (1973), 437–448.
- [18] N. Hadjisavvas, S. Komlósi, and S. Schaible, *Handbook of Generalized Convexity and Generalized Monotonicity*, Springer, New York, 2005.
- [19] 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.
- [20] Y. Hu, G. Li, C. K. W. Yu and T. L. Yip, Quasi-convex feasibility problems: Subgradient methods and convergence rates, Preprint, 2020.
- [21] Y. Hu, J. Li and C. K. W. Yu, Convergence rates of subgradient methods for quasi-convex optimization problems, *Computational Optimization and Applications*, **77** (2020), 183–212.

- [22] 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.
- [23] 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.
- [24] X. Huang and X. Yang, A unified augmented lagrangian approach to duality and exact penalization, *Mathematics of Operations Research*, **28** (2003), 533–552.
- [25] K. C. Kiwiel, Convergence and efficiency of subgradient methods for quasiconvex minimization, *Mathematical Programming*, **90** (2001), 1–25.
- [26] V. I. Kolobov, S. Reich and R. Zalas, Finitely convergent deterministic and stochastic methods for solving convex feasibility problems, preprint, [arXiv:math/1905.05660](https://arxiv.org/abs/math/1905.05660).
- [27] I. V. Konnov, On convergence properties of a subgradient method, *Optimization Methods and Software*, **18** (2003), 53–62.
- [28] I. Necoara, P. Richtárik and A. Patrascu, Randomized projection methods for convex feasibility: Conditioning and convergence rates, *SIAM Journal on Optimization*, **29** (2019), 2814–2852.
- [29] A. Nedić, Random algorithms for convex minimization problems, *Mathematical Programming*, **129** (2011), 225–253.
- [30] J.-S. Pang, Error bounds in mathematical programming, *Mathematical Programming*, **79** (1997), 299–332.
- [31] E. A. Papa Quiroz, L. M. Ramirez and P. R. Oliveira, An inexact proximal method for quasiconvex minimization, *European Journal of Operational Research*, **246** (2015), 721–729.
- [32] B. T. Polyak, *Introduction to Optimization*, Optimization Software, New York, 1987.
- [33] B. T. Polyak, Random algorithms for solving convex inequalities, *Studies in Computational Mathematics*, **8** (2001), 409–422.
- [34] I. Stancu-Minasian, *Fractional Programming*, Kluwer Academic Publisher, Dordrecht, 1997.
- [35] 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** (2017), 055017.
- [36] M. Wang and D. P. Bertsekas, Stochastic first-order methods with random constraint projection, *SIAM Journal on Optimization*, **26** (2016), 681–717.
- [37] A. J. Zaslavski, *Approximate Solutions of Common Fixed-Point Problems*, Springer, Switzerland, 2016.

Received February 2021; revised June 2021.

E-mail address: ligang@zstu.edu.cn

E-mail address: minghuali20021848@163.com

E-mail address: mayhhu@szu.edu.cn