Quasi-convex Feasibility Problems: Subgradient Methods and Convergence Rates

Yaohua Hu[a], Gongnong Li[a], Carisa Kwok Wai Yu[b,*], Tsz Leung Yip[c]

[a]*Shenzhen Key Laboratory of Advanced Machine Learning and Applications, College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, P. R. China*
[b]*Department of Mathematics, Statistics and Insurance, The Hang Seng University of Hong Kong, Shatin, Hong Kong*
[c]*Department of Logistics and Maritime Studies, Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*

Abstract

The feasibility problem is at the core of the modeling of many problems in various areas, and the quasi-convex function usually provides a precise representation of reality in many fields such as economics, finance and management science. In this paper, we consider the quasi-convex feasibility problem (QFP), that is to find a common point of a family of sublevel sets of quasi-convex functions, and propose a unified framework of subgradient methods for solving the QFP. This paper is contributed to establish the quantitative convergence theory, including the iteration complexity and the convergence rates, of subgradient methods with the constant/dynamic stepsize rules and several general control schemes, including the α-most violated constraints control, the s-intermittent control and the stochastic control. An interesting finding is disclosed by iteration complexity results that the stochastic control enjoys both advantages of low computational cost requirement and low iteration complexity. More importantly, we introduce a notion of Hölder-type error bound property for the QFP, and use it to establish the linear (or sublinear) convergence rates for subgradient methods to a feasible solution of the QFP. Preliminary numerical results to the multiple Cobb-Douglas productions efficiency problem indicate the powerful modeling capability of the QFP and show the high efficiency and stability of subgradient methods for solving the QFP.

Keywords: Global optimization, quasi-convex feasibility problem, subgradient method, iteration complexity, convergence rate

2010 MSC: 65K05, 90C26, 49M37

*Corresponding author

Email addresses: mayhhu@szu.edu.cn (Yaohua Hu), gnli@szu.edu.cn (Gongnong Li), carisayu@hsu.edu.hk (Carisa Kwok Wai Yu), t.l.yip@polyu.edu.hk (Tsz Leung Yip)

1. Introduction

Let $I := \{1, \ldots, m\}$ be a finite index set, and let $\{f_i : i \in I\}$ be a family of continuous functions on $\mathbb{R}^n$. The feasibility problem is to find a point $x \in \mathbb{R}^n$ such that

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

The feasibility problem is at the core of the modeling of many problems in various areas of mathematics and physical sciences, such as image recovery [11], wireless sensor networks localization [16] and gene regulatory network inference [35].

When the functions involved in (1.1) are convex, the corresponding problem is called the convex feasibility problem (CFP) that has attracted a great deal of attention in various application fields. Motivated by its extensive applications, tremendous efforts have been devoted to the development of optimization algorithms for solving the CFP; see [3, 8, 11, 37, 40] and references therein. One of the most popular approaches for solving the CFP (1.1) is the class of subgradient methods, which was originally proposed by Censor and Lent [9] with a cyclic control scheme. Many extensions of subgradient methods have been proposed by employing several control schemes, such as the parallel control, the s-intermittent control and the most violated constraint control. Various convergence properties of subgradient methods for solving the CFP have been well explored; one can refer to a review paper [3] and a recent book [40].

However, the convex function is too restrictive to many real-life problems encountered in economics, finance and management science. In contrast, the quasi-convex function usually provides a much more accurate representation of reality in economics and finance and still possesses certain desirable properties of convex function. For example, 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, 33] and references therein. In recent decades, much attention has been drawn to quasi-convex optimization; see [2, 13, 14, 18, 28, 29, 30, 33] and references therein. For the feasibility problem (1.1), Censor and Segal considered in [10] the quasi-convex feasibility problem (QFP), in which the functions involved are quasi-convex. They proposed the subgradient methods with a dynamic stepsize rule and the most violated constraint/ almost cyclic control/ parallel control schemes to solve the QFP, and established their global convergence property to a feasible solution.

In convergence theory, besides the global convergence property, the establishment of convergence rate is another important issue in guaranteeing the numerical performance of relevant algorithms. For the CFP, the linear convergence rate of subgradient methods (employing the dynamic stepsize and different controls) has been established under the Slater condition [3] or the polyhedral assumption [38]. However, to the best of our knowledge, there is limited study devoted to establishing the convergence rate of subgradient methods for solving the QFP. Moreover, the error bound property is a strictly weaker condition than the Slater condition and the polyhedral assumption, and in recent years, it plays an important role in the convergence rate analysis of various optimization algorithms; see [5, 35, 37, 39] and references therein. This motivates us to develop an error bound-based analysis for the convergence rate issue of subgradient methods for solving the QFP.

Except the deterministic control schemes mentioned above, the idea of the stochastic index scheme is increasingly popular in optimization algorithms and applications; e.g., first-order algorithms with random projection in large-scale network optimization problems [36], incremental subgradient methods with random component selection in distributed optimization problems [24] and stochastic gradient descent algorithms in machine learning [6]. The stochastic control scheme was also applied in subgradient methods for solving the CFP in [31]. However, to the best of our knowledge, the stochastic control scheme has not been employed in subgradient methods for solving the QFP yet.

In this paper, we consider the QFP (1.1), where the involved functions are quasi-convex and continuous, and study the subgradient methods (see Algorithms 3.1 and 3.2) for solving the QFP in a unified framework, which covers most types of control schemes discussed in the literature. The main contribution of the present paper is to establish the quantitative convergence theory, including the iteration complexity and the convergence rate, of subgradient methods with two typical stepsize rules and several general control schemes for solving the QFP. In particular, the constant stepsize and the dynamic stepsize are considered (the constant stepsize is simple and practical in implementation and avoid the difficulty in estimating the Lipschitz modulus as in the dynamic stepsize), and the α-most violated constraints control, the s-intermittent control and the stochastic control are discussed in this paper.

In convergence analysis, we first establish as a by-product the global convergence theorem and derive the (worst-case) iteration complexity of subgradient methods; concretely, subgradient methods with a constant stepsize converge to an approximate feasible solution of the QFP within a tolerance expressed by the stepsize, while subgradient methods with a dynamic stepsize converge to an exact feasible solution; see Theorems 3.1, 3.3, 3.7 and 3.9. More importantly, we introduce a notion of the Hölder-type error bound property for the QFP and use it to explore the linear (or sublinear) convergence rates of subgradient methods to a feasible solution of the QFP; see Theorems 3.2, 3.4, 3.8 and 3.10. The established theorems not only extend subgradient methods in [10] to the constant stepsize rule and the more general control schemes, but also improve the global convergence results to the quantitative complexity and convergence rate. As far as we know, the establishment of convergence rates of subgradient methods are new in the literature of QFP.

Moreover, the iteration complexity and convergence rates of the subgradient method with a stochastic control are presented in terms of the expectation of violation and the expectation of distance from the feasible solution set in Theorems 3.5, 3.6, 3.11 and 3.12, respectively. This paper seems to be the first attempt to investigate the subgradient method with the stochastic control for solving the QFP, and interestingly, provides a theoretical evidence for the benefit of the stochastic control that it enjoys both advantages of low computational cost requirement and low (worst-case) iteration complexity; see Remark 3.10 for explanation.

Finally, we formulate the multiple Cobb-Douglas production efficiency problem (M-CDPE) [7] as an application of the QFP, also reformulate it as a CFP. Preliminary numerical results indicate the advantage on the modeling capability of the QFP over the CFP and show the high efficiency and stability of subgradient methods with both the constant and dynamic stepsize rules for solving the QFP; especially the stochastic control for large-scale problem-s. This study may deliver a new approach for finding a feasible (optimal) solution of the large-scale MCDPE.

The present paper is organized as follows. In Section 2, we present the notations and preliminary lemmas which will be used in this paper. In Section 3, we provide a unified framework of subgradient methods with two stepsize rules and several control schemes to solve the QFP and establish the quantitative iteration complexity and convergence rates. The application to the MCDPE and the numerical results are presented in Section 4. A conclusion is given in Section 5. The technical proofs for the global convergence and iteration complexity are deferred to the supplementary material.

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

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

Moreover, we use the notation that $a^+ := \max\{a, 0\}$ for any $a \in \mathbb{R}$, define the positive part 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 $\frac{0}{0} = 0$, $\sum_{i\in\emptyset} a_i = 0$ and $\cup_{i\in\emptyset} I_i = \emptyset$ for any sequence of scalars $\{a_i\}$ and any family of index sets $\{I_i\}$. For $x \in \mathbb{R}^n$ and $Z \subseteq \mathbb{R}^n$, the Euclidean distance of x from Z and the Euclidean projection of x onto Z are respectively defined by

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

The normal cone of Z at x is defined by

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

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

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

The sublevel sets of f at x are denoted by

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

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. The following equivalent characterization of a quasi-convex function is well-known.

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 any $x \in \mathbb{R}^n$.*

The convex subdifferential $\partial f(x) := \{g \in \mathbb{R}^n : f(y) \ge f(x) + \langle g, y - x\rangle, \forall y \in \mathbb{R}^n\}$ might be empty for the quasi-convex function (e.g., $f(x) = x^3$ at the origin). Hence, the introduction of (nonempty) subdifferential of quasi-convex functions is an important issue in quasi-convex optimization. Several specific types of quasi-subdifferentials have been introduced and explored for quasi-convex functions; see [1, 12, 18] and references therein. In particular, Kiwiel [21], Censor and Segal [10], and Hu et al. [18] employed the following quasi-subgradient, defined as a normal vector to its strict sublevel set, in their concerned subgradient methods.

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

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

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. The following proposition is taken from [18, Lemma 2.1], saying that the quasi-subdifferential of f is nontrivial whenever f is quasi-convex.

Proposition 2.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be quasi-convex. Then $\partial^* f(\cdot) \setminus \{0\} \neq \emptyset$.*

Since $\partial^* f(x)$ is a normal cone to the sublevel set of f at x, it follows from Proposition 2.2 that the quasi-subdifferential of a quasi-convex function contains at least a unit vector. This is a specific property of the quasi-subdifferential beyond the convex subdifferential. Moreover, it was claimed in [18] that

$$\partial^* f(\cdot) = \operatorname{cone}(\partial f(\cdot)) \quad \text{whenever } f \text{ is convex,} \tag{2.1}$$

where $\operatorname{cone}(X)$ denotes the convex cone hull of X.

The notion of Hölder continuity has been widely studied in harmonic analysis and fractional analysis and extensively applied in economics and management science. In particular, the Hölder condition of order 1 is reduced to the Lipschitz condition, which is equivalent to the bounded subgradient assumption commonly used in convergence study of subgradient methods for convex optimization problems; see, e.g., [4, 24, 32].

Definition 2.2. *Let $0 < \beta \leq 1$ 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 X if*

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

The Hölder condition was used to provide a fundamental property of the quasi-subgradient in [22, Proposition 2.1], which plays an important role in the establishment of a basic inequality in convergence analysis of subgradient-type methods for quasi-convex optimization problems; see, e.g., [14, 17, 18, 19]. Below we extend [22, Proposition 2.1] to the quasi-convex inequality system.

Lemma 2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ 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 X. Then, for any $x \in S$ and $y \in X \setminus S$, it holds that*

$$f(y) \leq L\,\langle g, y - x\rangle^\beta \quad \textit{for any } g \in \partial^* f(y) \cap \mathbb{S}. \tag{2.2}$$

Proof. By assumptions of this lemma, one can check that f^+ is quasi-convex and continuous, satisfies the Hölder condition of order β with modulus L on X, $\partial^* f^+(x) = \partial^* f(x)$ for any

$x \in X \setminus S$ and $S = \arg\min_{x\in X} f^+(x)$ (one can also refer to [19, Lemma 2.2 and Section 4.1]). Then, for any $x \in S$ and $y \in X \setminus S$ (i.e., $f^+(x) = 0$ and $f^+(y) = f(y)$), by applying [22, Proposition 2.1] (to f^+), we obtain

$$f^+(y) - f^+(x) \le L \langle g, y - x\rangle^\beta \quad \text{for any } g \in \partial^* f^+(y) \cap \mathbb{S};$$

consequently, (2.2) is satisfied. The proof is complete. □

We end this section by recalling the following lemmas, which will be useful in the convergence rate analysis of subgradient methods. In particular, they are taken from [20, Lemma 4.1] and [17, Lemma 2.2], respectively.

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

(i) *If $\gamma \in [1,\infty)$, then $\frac{1}{n^{\gamma-1}} \left(\sum_{i=1}^n a_i\right)^\gamma \le \sum_{i=1}^n a_i^\gamma \le \left(\sum_{i=1}^n a_i\right)^\gamma$.*

(ii) *If $\gamma \in (0,1]$, then $\left(\sum_{i=1}^n a_i\right)^\gamma \le \sum_{i=1}^n a_i^\gamma \le n\left(\sum_{i=1}^n a_i\right)^\gamma$.*

Lemma 2.3. *Let $r > 0$, $a > 0$, $b \ge 0$, and let $\{u_k\}$ be a sequence of nonnegative scalars such that*

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

(i) *If $b = 0$, then $u_{k+1} \le u_1 \left(1 + rau_1^r k\right)^{-\frac{1}{r}}$ for each $k \in \mathbb{N}$.*

(ii) *If $0 < b < a^{-\frac{1}{r}}(1+r)^{-\frac{1+r}{r}}$, then there exists $\tau \in (0,1)$ such that*

$$u_{k+1} \le u_1 \tau^k + \left(\frac{b}{a}\right)^{\frac{1}{1+r}} \quad \textit{for each } k \in \mathbb{N}.$$

3. Subgradient methods for quasi-convex feasibility problem

Let $m \in \mathbb{N}$ and $I := \{1, 2, \dots, m\}$ be a finite index set, and let $\{f_i : i \in I\}$ be a family of quasi-convex and continuous functions defined on $\mathbb{R}^n$ and $X \subseteq \mathbb{R}^n$ be a compact and convex set. In this paper, we consider the quasi-convex feasibility problem (QFP) that is to find a feasible point $x \in \mathbb{R}^n$ such that

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

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

$$S := \{x \in X : f_i(x) \le 0, \forall i \in I\} \ne \varnothing.$$

One of the most popular and practical algorithms for solving the (convex or quasi-convex) feasibility problem is the class of subgradient methods; see [3, 10] and references therein. In this paper, we consider subgradient methods with two typical stepsize rules and general control schemes for solving the QFP (3.1).

3.1. Subgradient methods with a constant stepsize

The constant stepsize is the most simple and practical stepsize rule in the implementation of subgradient methods; see, e.g., [4, 18, 24, 32]. Here we discuss the subgradient methods with a constant stepsize for solving the QFP (3.1) in a general framework (of control schemes), stated as follows. To proceed, for each $x \in \mathbb{R}^n$, we write $I(x)$ to denote the active index set of the system (3.1) at x, namely,

$$I(x) := \{i \in I : f_i(x) > 0\}. \tag{3.2}$$

Algorithm 3.1. *Select an initial point* $x_1 \in \mathbb{R}^n$ *and a constant stepsize* $v > 0$*. For each* $k \in \mathbb{N}$*, having* $x_k \in \mathbb{R}^n$*, we select a nonempty index set* $I_k \subseteq I$ *and weights* $\{\lambda_{k,i}\}_{i \in I_k} \in \mathbf{\Delta}_+^{|I_k|}$*, calculate* $g_{k,i} \in \partial^* f_i(x_k) \cap \mathbb{S}$ *for each* $i \in I_k \cap I(x_k)$*, and update* x_{k+1} *by*

$$x_{k+1} := \mathrm{P}_X \left(x_k - v \sum_{i \in I_k \cap I(x_k)} \lambda_{k,i} g_{k,i} \right). \tag{3.3}$$

Remark 3.1. (i) *It is clear by* (3.3) *that each sequence generated by Algorithm 3.1*

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

(ii) *It follows from* (2.1) *that, when applied to the convex feasibility problem (CFP), Algorithm 3.1 is reduced to the convex subgradient method [3, 9] with a constant stepsize.*

(iii) *Algorithm 3.1 enjoys an advantage of simple implementation of the constant stepsize. Censor and Segal [10] proposed the subgradient methods with classical control schemes to solve the QFP* (3.1)*, in which a dynamic stepsize is given in terms of the component function values and the Hölder continuity orders and moduli. However, it is not an easy task to estimate these parameters of the Hölder continuity for all component functions of* (3.1)*, which may hinder the applications of subgradient methods in [10]. In contrast, Algorithm 3.1 uses a constant stepsize and is quite practical in applications.*

(iv) *Algorithm 3.1 provides and extends a unified framework for the existing subgradient methods with general control schemes for solving the QFP. For example, Hu et al. [19] reformulated the QFP as a sum-minimization problem of quasi-convex component functions and proposed the deterministic and randomized incremental subgradient methods to solve the corresponding optimization problem. In particular, Algorithm 3.1 covers the deterministic and randomized incremental subgradient methods (i.e., [19, Algorithms 3 and 4]) when* $\{I_k\}$ *is selected as a cyclic control and a stochastic control, respectively.*

To investigate the convergence property of subgradient methods, we shall assume the following two blanket assumptions on the QFP (3.1) and parameters in Algorithm 3.1 throughout the whole paper.

Assumption 1. *There exist* $\beta \in (0,1]$ *and* $L > 0$ *such that each* f_i *satisfies the Hölder condition of order* β *with modulus* L *on* X.

Assumption 2. *There exists* $\mu > 0$ *such that* $\min_{i\in I_k} \lambda_{k,i} \geq \mu$ *for each* $k \in \mathbb{N}$.

Remark 3.2. (i) *Assumption 1 premises the unified Hölder continuity order and modulus for all component functions of the QFP* (3.1)*, which is equivalent to the Hölder condition of each* f_i *with different orders and moduli as assumed in [10, 19]. Indeed, suppose that each* f_i *in* (3.1) *satisfies the Hölder condition of order* $\beta_i \in (0,1]$ *with modulus* $L_i > 0$ *on* X*. Then one can check that Assumption 1 is satisfied with*

$$\beta := \min_{i\in I} \beta_i \quad \textit{and} \quad L := \max_{i\in I} L_i \operatorname{diam}(X)^{\beta_i - \beta}.$$

(ii) *Assumption 2 premises a unified nonzero lower bound for weights in Algorithm 3.1, which ensures each component of* (3.1) *to take sufficient contribution for seeking a feasible solution; see [3, Remark 3.13].*

Under Assumptions 1 and 2, the following lemma provides a basic inequality of Algorithm 3.1 for arbitrary type of control scheme, which shows a significant property and plays a key tool in convergence analysis of subgradient methods.

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

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 - 2v\mu L^{-\frac{1}{\beta}} \sum_{i\in I_k \cap I(x_k)} \left(f_i^+(x_k)\right)^{\frac{1}{\beta}} + v^2. \tag{3.5}$$

Proof. Fix $x \in S$ and $k \in \mathbb{N}$. We assume, without loss of generality, that $I_k \cap I(x_k) \neq \emptyset$; otherwise, one has by (3.3) that $x_{k+1} = \mathrm{P}_X(x_k)$, and thus (3.5) is satisfied automatically. Then it follows from (3.3) and the nonexpansive property of the projection operator that

$$\begin{aligned} &\|x_{k+1} - x\|^2 \\ &\leq \|x_k - v \sum_{i\in I_k\cap I(x_k)} \lambda_{k,i} g_{k,i} - x\|^2 \qquad (3.6) \\ &= \|x_k - x\|^2 - 2v \sum_{i\in I_k\cap I(x_k)} \lambda_{k,i}\langle x_k - x, g_{k,i}\rangle + v^2 \,\| \sum_{i\in I_k\cap I(x_k)} \lambda_{k,i} g_{k,i}\|^2. \end{aligned}$$

Note by (3.2) that $f_i(x_k) > 0$ for each $i \in I(x_k)$. Then by Assumption 1 and $x \in S$, it follows from (2.2) in Lemma 2.1 that $\langle x_k - x, g_{k,i}\rangle \geq L^{-\frac{1}{\beta}} f_i^{\frac{1}{\beta}}(x_k) = L^{-\frac{1}{\beta}} \left(f_i^+(x_k)\right)^{\frac{1}{\beta}}$ for each $i \in I(x_k)$. Hence we have by Assumption 2 that

$$\sum_{i\in I_k\cap I(x_k)} \lambda_{k,i}\langle x_k - x, g_{k,i}\rangle \geq \mu L^{-\frac{1}{\beta}} \sum_{i\in I_k\cap I(x_k)} \left(f_i^+(x_k)\right)^{\frac{1}{\beta}}. \tag{3.7}$$

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

$$\|\sum_{i\in I_k\cap I(x_k)}\lambda_{k,i}g_{k,i}\|^2\le 1 \tag{3.8}$$

(thanks to $\{\lambda_{k,i}\}_{i\in I_k}\in \mathbf{\Delta}_+^{|I_k|}$ and $g_{k,i}\in\mathbb{S}$). This, together with (3.6) and (3.7), implies (3.5). The proof is complete. □

The control sequence of index set $\{I_k\}$ plays a central role in determining the active indices (of component functions) to be executed, and is crucial in guaranteeing the convergence property and numerical performance of subgradient methods for solving the QFP. In this paper, we consider the following general control schemes, item (a) is an extension of the most violated constraint control and the parallel control, item (b) can be found in [3, Definition 3.18] and is an extension of the almost cyclic control and the parallel control, and item (c) takes the increasingly popular idea of the stochastic control from [31, 36].

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

(a) *the α-most violated constraints control if*

$$I_k:=\{i_k\in I: f_{i_k}^+(x_k)\ge\alpha\max_{i\in I}f_i^+(x_k)\}\quad \textit{for each } k\in\mathbb{N}.$$

(b) *the s-intermittent control if*

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

(c) *the stochastic control if $I_k=\{\omega_k\}$ that is a uniformly distributed random variable on I.*

For the remainder of this section, we assume that Assumptions 1 and 2 are always satisfied, and establish the quantitative convergence theory, including the iteration complexity and the convergence rate, of Algorithm 3.1 with the α-most violated constraints control, the s-intermittent control and the stochastic control, respectively.

3.1.1. The α-most violated constraints control

This subsection aims to investigate the iteration complexity and the convergence rate for Algorithm 3.1 with the α-most violated constraints control. To explore the convergence property, the violation of the QFP (3.1) is usually measured by

$$F^+(x):=\max_{i\in I}f_i^+(x)\quad\text{for each } x\in\mathbb{R}^n.$$

It is clear that

$$x \in S \quad \Leftrightarrow \quad F^+(x) = 0 \text{ and } x \in X.$$

By definition of the α-most violated constraints control (cf. Definition 3.1(a)), the following lemma about the basic inequality directly follows from Lemma 3.1.

Lemma 3.2. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1 with $\{I_k\}$ being the α-most violated constraints control. Then the following basic inequality holds for each $x \in S$ and $k \in \mathbb{N}$ that*

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 - 2v\mu \left(\frac{\alpha}{L}\right)^{\frac{1}{\beta}} \left(F^+(x_k)\right)^{\frac{1}{\beta}} + v^2. \tag{3.9}$$

By virtue of the basic inequality in Lemma 3.2 and following a standard line of analysis for subgradient methods, we first establish the convergence theorem and (worst-case) iteration complexity for Algorithm 3.1 with the α-most violated constraints control as a by-product. Given precision $\delta > 0$ and tolerance $T(v)$, the (worst-case) iteration complexity of a particular algorithm is to estimate the number of iterations $K(\delta)$ required by the algorithm to obtain an approximate $T(v) + \delta$-feasible solution, that is,

$$\min_{1 \leq k \leq K(\delta)} F^+(x_k) \leq T(v) + \delta.$$

As usual, we use $\lceil t \rceil$ to denote the smallest integer larger than t. The proof is given in Appendix **??** of the supplementary material.

Theorem 3.1. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ being the α-most violated constraints control. Then the following assertions are true.*

(i) $\liminf_{k\to\infty} F^+(x_k) \leq \frac{L}{\alpha}\left(\frac{v}{2\mu}\right)^\beta$.

(ii) *Let $\delta > 0$ and $K_{\rm m}^{\rm c} := \lceil \frac{{\rm d}^2(x_1,S)}{2v\mu\delta} \rceil$. Then* $\min_{1 \leq k \leq K_{\rm m}^{\rm c}} F^+(x_k) \leq \frac{L}{\alpha}\left(\frac{v}{2\mu} + \delta\right)^\beta$.

Remark 3.3. (i) *Theorem 3.1 shows the convergence and iteration complexity of Algorithm 3.1 to a feasible solution of the QFP* (3.1) *within a tolerance when the α-most violated constraints control is adopted. The tolerance in Theorem 3.1(i) is given in terms of the stepsize v, the Hölder continuity order β and modulus L of the QFP, and parameters μ and α in Algorithm 3.1; hence it provides a stepsize adjustment rule according to the precision requirement of the solution. In particular, when applied to the most violated constraint control scheme (namely, $\alpha = 1$ and $\mu = 1$), Theorem 3.1(i) shows that the tolerance of Algorithm 3.1 is given by $T_{\rm m} = L\left(\frac{v}{2}\right)^\beta$.*

(ii) *By applying Lemma 2.2(ii) (as $\beta \leq 1$), Theorem 3.1(ii) is reduced to $\min_{1 \leq k \leq K_{\rm m}^{\rm c}} F^+(x_k) \leq \frac{L}{\alpha}\left(\frac{v}{2\mu}\right)^\beta + \frac{L}{\alpha}\delta^\beta$ with $K_{\rm m}^{\rm c} := \lceil \frac{{\rm d}^2(x_1,S)}{2v\mu\delta} \rceil$. Letting $\epsilon := \delta^\beta$, this shows that that Algorithm 3.1* *with the α-most violated constraints control possesses an iteration complexity of $\mathcal{O}(1/\epsilon^{\frac{1}{\beta}})$ to approach an approximate $\mathcal{O}(v^{\beta})+\epsilon$-feasible solution.*

The establishment of convergence rate is significant in guaranteeing the numerical performance of relevant algorithms. The error bound property plays an important role in convergence rate analysis of numerical algorithms. The notion of the (Lipschitz-type) error bound property was introduced by [15, 27] for the linear inequality system and convex inequality system respectively, and has been extensively used in convergence rate analysis of various numerical algorithms; see [35, 39] and references therein. As a natural extension, the Hölder-type error bound property was introduced for polynomial systems [23] and has been widely explored and applied in [5, 26, 37] and references therein. In particular, Suzuki and Kuroiwa [34] investigated the Hölder-type error bound property for the quasi-convex inequality system by virtue of a generator for the quasi-convex function and a constraint qualification. Below we recall the Hölder-type error bound property for the QFP (3.1).

Definition 3.2. *The inequality system* (3.1) *is said to satisfy the Hölder-type error bound property of order $q>0$ and modulus $\kappa>0$ if*

$$\mathrm{d}^q(x,S)\le \kappa \max_{i\in I} f_i^+(x) \quad \text{for each } x\in X. \tag{3.10}$$

In particular, the inequality system (3.1) *is said to satisfy the (Lipschitz-type) error bound property if* (3.10) *holds with $q=1$.*

The following proposition provides a sufficient condition for the Hölder-type error bound property of the QFP (3.1) in terms of that of each component function and the Slater condition of their sublevel sets.

Proposition 3.1. *Suppose that each inequality system $\{x\in X : f_i(x)\le 0\}$ satisfies the Hölder-type error bound property of order q and the Slater condition is satisfied:*

$$\{x\in X : f_i(x)<0, \forall i\in I\}\neq\varnothing. \tag{3.11}$$

Then the QFP (3.1) *satisfies the Hölder-type error bound property of order q.*

Proof. Write $S_i:=\{x\in\mathbb{R}^n : f_i(x)\le 0\}$ for each $i\in I$; clearly, $S=X\cap(\cap_{i\in I}S_i)$. By the continuity of f_i and the Slater condition (3.11), we have $X\cap(\cap_{i\in I}\mathrm{int}S_i)\neq\varnothing$. Hence, [3, Corollary 5.14] is applicable to concluding that there exists $\tau>0$ such that

$$\mathrm{d}(x,S)\le \tau\max_{i\in I}\mathrm{d}(x,S_i) \quad \text{for each } x\in X.$$

By the Hölder-type error bound assumption for $\{x \in X : f_i(x) \le 0\}$, there exists $\kappa_i > 0$ such that

$$\mathrm{d}^q(x, S_i) \le \mathrm{d}^q(x, S_i \cap X) \le \kappa_i f_i^+(x) \quad \text{for each } x \in X,$$

for each $i \in I$. Hence $\kappa := \tau^q \max_{i \in I} \kappa_i > 0$ is such that (3.10) is satisfied. The proof is complete. □

The main theorem of this subsection is as follows, which presents a linear convergence rate of Algorithm 3.1 with the α-most violated constraints control to a certain neighborhood of the feasible solution set under the assumption of the Hölder-type error bound property.

Theorem 3.2. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ being the α-most violated constraints control. Suppose that (3.1) satisfies the Hölder-type error bound property of order q and modulus κ. Then the following assertions are true.*

(i) *If $q = 2\beta$, then there exists $\tau \in [0, 1)$ such that, for each $k \in \mathbb{N}$,*

$$\mathrm{d}^2(x_{k+1}, S) \le \tau^k \mathrm{d}^2(x_1, S) + \frac{v}{2\mu} \left(\frac{\kappa L}{\alpha}\right)^{\frac{1}{\beta}}. \tag{3.12}$$

(ii) *If $q > 2\beta$ and $v < (2\mu)^{\frac{\beta}{\beta - q}} \left(\frac{\kappa L}{\alpha}\right)^{\frac{1}{q-\beta}} \left(\frac{2\beta}{q}\right)^{\frac{q}{2(q-\beta)}}$, then there exists $\tau \in (0, 1)$ such that, for each $k \in \mathbb{N}$,*

$$\mathrm{d}^2(x_{k+1}, S) \le \tau^k \mathrm{d}^2(x_1, S) + \left(\frac{v}{2\mu} \left(\frac{\kappa L}{\alpha}\right)^{\frac{1}{\beta}}\right)^{\frac{2\beta}{q}}. \tag{3.13}$$

Proof. Firstly, we claim that

$$\mathrm{d}^2(x_{k+1}, S) \le \mathrm{d}^2(x_k, S) - 2v\mu \left(\frac{\alpha}{\kappa L}\right)^{\frac{1}{\beta}} \mathrm{d}^{\frac{q}{\beta}}(x_k, S) + v^2 \tag{3.14}$$

for each $k \in \mathbb{N}$. To this end, recall from (3.9) (taking $x := \mathrm{P}_S(x_k)$) that

$$\mathrm{d}^2(x_{k+1}, S) \le \mathrm{d}^2(x_k, S) - 2v\mu \left(\frac{\alpha}{L}\right)^{\frac{1}{\beta}} (F^+(x_k))^{\frac{1}{\beta}} + v^2. \tag{3.15}$$

By the assumption of the Hölder-type error bound property, (3.10) holds for each x_k. This, together with (3.15), yields (3.14), as desired.

(i) Suppose that $q = 2\beta$. Setting $\tau := (1 - 2v\mu \left(\frac{\alpha}{\kappa L}\right)^{\frac{1}{\beta}})_+ \in [0, 1)$ and by (3.14), we achieve that

$$\mathrm{d}^2(x_{k+1}, S) \le \tau \mathrm{d}^2(x_k, S) + v^2 \quad \text{for each } k \in \mathbb{N}.$$

Then we inductively obtain (3.12), and the conclusion follows.

(ii) Suppose that $q > 2\beta$ and $v < (2\mu)^{\frac{\beta}{\beta - q}} \left(\frac{\kappa L}{\alpha}\right)^{\frac{1}{q-\beta}} \left(\frac{2\beta}{q}\right)^{\frac{q}{2q-2\beta}}$. Then by (3.14), Lemma 2.3(ii) is applicable (with $\mathrm{d}^2(x_k, S)$, $\frac{q}{2\beta} - 1$, $2v\mu \left(\frac{\alpha}{\kappa L}\right)^{\frac{1}{\beta}}$, v^2 in place of u_k, r, a, b) to obtaining (3.13). The proof is complete. □

3.1.2. The s-intermittent control

This subsection aims to explore the iteration complexity and the convergence rate for Algorithm 3.1 with the s-intermittent control, which is an extension of the (incremental) cyclic control as used in [19, Algorithm 3]. To proceed, we first deduce the basic inequality for Algorithm 3.1 with the s-intermittent control by virtue of Lemma 3.1.

Lemma 3.3. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1 with $\{I_k\}$ being the s-intermittent control. Then the basic inequality holds for each $x \in S$ and $k \in \mathbb{N}$ that*

$$\|x_{s(k+1)} - x\|^2 \leq \|x_{sk} - x\|^2 - 4v\mu(2L)^{-\frac{1}{\beta}} \left(F^+(x_{sk})\right)^{\frac{1}{\beta}} + sv^2(1+2\mu). \tag{3.16}$$

Proof. Fix $x \in S$ and $k \in \mathbb{N}$. By applying (3.5) in Lemma 3.1 inductively, we obtain that

$$\|x_{s(k+1)} - x\|^2 \leq \|x_{sk} - x\|^2 - 2v\mu L^{-\frac{1}{\beta}} \sum_{j=0}^{s-1} \sum_{i \in I_{sk+j}} \left(f_i^+(x_{sk+j})\right)^{\frac{1}{\beta}} + sv^2. \tag{3.17}$$

Below we estimate the second term on the right hand side of (3.17) in terms of $F^+(x_{sk})$. Firstly, let $i_k \in I$ be the most violated index of the inequality system (3.1) at x_{sk}, that is,

$$f_{i_k}^+(x_{sk}) = F^+(x_{sk}) > 0 \tag{3.18}$$

(otherwise, $\{x_i\}_{i>sk}$ terminates at $x_{sk} \in S$ due to (3.4)). By the definition of the s-intermittent control (cf. Definition 3.1(b)), there exists $j_k \in [0, s-1]$ such that $i_k \in I_{sk+j_k}$. By the Hölder condition as in Assumption 1, one has

$$\begin{aligned} f_{i_k}^{\frac{1}{\beta}}(x_{sk}) &\leq \left(f_{i_k}(x_{sk+j_k}) + L\|x_{sk+j_k} - x_{sk}\|^\beta\right)^{\frac{1}{\beta}} \\ &\leq 2^{\frac{1}{\beta}-1}\left((f_{i_k}^+(x_{sk+j_k}))^{\frac{1}{\beta}} + L^{\frac{1}{\beta}}\|x_{sk+j_k} - x_{sk}\|\right) \end{aligned} \tag{3.19}$$

(by Lemma 2.2(i) as $\beta \leq 1$). Since $i_k \in I_{sk+j_k}$ and $j_k \in [0, s-1]$, we get

$$\left(f_{i_k}^+(x_{sk+j_k})\right)^{\frac{1}{\beta}} \leq \sum_{i \in I_{sk+j_k}} \left(f_i^+(x_{sk+j_k})\right)^{\frac{1}{\beta}} \leq \sum_{j=0}^{s-1} \sum_{i \in I_{sk+j}} \left(f_i^+(x_{sk+j})\right)^{\frac{1}{\beta}}. \tag{3.20}$$

On the other side, we obtain by (3.3) and the nonexpansive property of the projection operator that $\|x_{k+1} - x_k\| \leq v\|\sum_{i \in I_k \cap I(x_k)} \lambda_{k,i} g_{k,i}\| \leq v$ for each $k \in \mathbb{N}$ (thanks to (3.8)). Then we derive inductively that $\|x_{sk+j_k} - x_{sk}\| \leq sv$. This, together with (3.18)-(3.20), deduces that

$$2^{1-\frac{1}{\beta}} \left(F^+(x_{sk})\right)^{\frac{1}{\beta}} \leq \sum_{j=0}^{s-1} \sum_{i \in I_{sk+j}} \left(f_i^+(x_{sk+j})\right)^{\frac{1}{\beta}} + L^{\frac{1}{\beta}} sv.$$

This, together with (3.17), implies (3.16). The proof is complete. □

Using the basic inequality in Lemma 3.3 and a line of analysis similar to that of Theorems 3.1 and 3.2, we can obtain in the following theorems the iteration complexity and the linear convergence rate of Algorithm 3.1 with the s-intermittent control to a certain neighborhood of the feasible solution set. The detailed proofs are given in Appendixes **??** and **??** of the supplementary material.

Theorem 3.3. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ being the s-intermittent control. Then the following assertions are true.*

(i) $\liminf_{k\to\infty} F^+(x_k) \le 2L\left(\frac{sv(1+2\mu)}{4\mu}\right)^\beta$.

(ii) *Let $\delta > 0$ and $K_{\rm c}^{\rm c} := \lceil \frac{s\mathrm{d}^2(x_1,S)}{4v\mu\delta}\rceil$. Then* $\min\limits_{1\le k\le K_{\rm c}^{\rm c}} F^+(x_k) \le 2L\left(\frac{sv(1+2\mu)}{4\mu}+\delta\right)^\beta$.

Theorem 3.4. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ being the s-intermittent control. Suppose that* (3.1) *satisfies the Hölder-type error bound property of order q and modulus κ. Then the following assertions are true.*

(i) *If $q = 2\beta$, then there exists $\tau \in [0,1)$ such that, for each $k \in \mathbb{N}$,*

$$\mathrm{d}^2(x_{s(k+1)}, S) \le \tau^k \mathrm{d}^2(x_s, S) + \frac{sv(1+2\mu)}{4\mu}(2\kappa L)^{\frac{1}{\beta}}.$$

(ii) *If $q > 2\beta$ and $v < (s(1+2\mu))^{\frac{q-2\beta}{2(\beta-q)}}(4\mu)^{\frac{\beta}{\beta-q}}(2\kappa L)^{\frac{1}{q-\beta}}\left(\frac{2\beta}{q}\right)^{\frac{q}{2(q-\beta)}}$, then there exists $\tau \in (0,1)$ such that, for each $k \in \mathbb{N}$,*

$$\mathrm{d}^2(x_{s(k+1)}, S) \le \tau^k \mathrm{d}^2(x_s, S) + \left(\frac{sv(1+2\mu)}{4\mu}(2\kappa L)^{\frac{1}{\beta}}\right)^{\frac{2\beta}{q}}.$$

Remark 3.4. (i) *Theorem 3.3(i) extends the convergence result of subgradient method with an incremental control in [19] to that with the s-intermittent control, and improves [19, Theorems 4.1] to a tighter tolerance. In particular, when applied to the cyclic control scheme, Algorithm 3.1 is reduced to [19, Algorithm 3]; consequently, Theorem 3.3 (with $\mu = 1$ and $s = m$) shows that the generated sequence converges to a feasible solution within a tolerance of $T_{\rm c} = 2L\left(\frac{3mv}{4}\right)^\beta$. This result improves the one in [19, Theorem 4.1] that has a much larger tolerance of $2mL\left(\frac{mv}{4}\right)^\beta$. Moreover, Theorems 3.3(ii) and 3.4 extend the convergence theorem in [19] to the quantitative iteration complexity and linear convergence rate.*

(ii) *Theorem 3.3(ii) shows that Algorithm 3.1 with the s-intermittent control has an iteration complexity of $\mathcal{O}(1/\epsilon^{\frac{1}{\beta}})$ to an approximate $\mathcal{O}(v^\beta)+\epsilon$-feasible solution. Even though the s-intermittent control admits the same orders of iteration complexity and tolerance with the α-most violated constraints control (see Remark 3.3), but it requires a much larger number of iterations and bears a much larger tolerance than the α-most violated constraints control* *when adopting the suitable parameters μ (~ 1), s ($\sim m$) and α (~ 1). Indeed, by Theorems 3.1 and 3.3, we obtain*

$$\frac{T_{\rm c}}{T_{\rm m}} = 2^{1-\beta}(3m)^{\beta} \gg 1 \quad \textit{and} \quad \frac{K_{\rm c}^{\rm c}}{K_{\rm m}^{\rm c}} = \frac{m}{2} \gg 1.$$

This shows the benefit of the most violated constraints control and the parallel control over the (almost) cyclic control. On the other side, the (almost) cyclic control has an advantage of low computational cost requirement, especially for large-scale problems; because it only uses the information of few component functions at each iteration, while the most violated constraints control and the parallel control need to find the most violated index through all component functions or calculate the subgradients of all component functions.

3.1.3. Stochastic control

The type of deterministic control schemes always suffers from certain drawbacks. As explained in Remark 3.4(ii), the most violated constraints control and the parallel control consume expensive computational cost at each iteration when the number of component functions is large; while, the almost cyclic control bears with a higher iteration complexity than these two controls. Moreover, the order used in the almost cyclic control could significantly affect the numerical performance of subgradient methods; unfortunately, it is very difficult to determine the most favorable order in practice.

The idea of the stochastic index scheme is increasingly popular for optimization problems with a large number of component functions [4, 24] or a large number of constraints [36]. A typical example is the stochastic gradient descent algorithm in machine learning [6], in which only one component function is randomly selected to construct the descent direction at each iteration.

Inspired by the idea of the stochastic index scheme, we consider the stochastic control (cf. Definition 3.1(c)) in subgradient methods for solving the QFP (3.1) and investigate its quantitative convergence theory. To proceed convergence analysis of Algorithm 3.1 with the stochastic control, we provide a basic inequality in terms of conditional expectation.

Lemma 3.4. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ being the stochastic control, and let $\mathcal{F}_k := \{x_1, \dots, x_k\}$ for each $k \in \mathbb{N}$. Then the following basic inequality holds for each $x \in S$ and $k \in \mathbb{N}$ that*

$$\mathbb{E}\left\{\|x_{k+1} - x\|^2 \mid \mathcal{F}_k\right\} \le \|x_k - x\|^2 - \frac{2v}{m} L^{-\frac{1}{\beta}} \left(F^+(x_k)\right)^{\frac{1}{\beta}} + v^2. \tag{3.21}$$

Proof. Fix $x \in S$ and $k \in \mathbb{N}$. Since $I_k = \{\omega_k\}$ and $\mu = 1$ in the stochastic control (cf. Definition 3.1(c)), it follows from (3.5) in Lemma 3.1 that

$$\|x_{k+1} - x\|^2 \le \|x_k - x\|^2 - 2vL^{-\frac{1}{\beta}} \left(f_{\omega_k}^+(x_k)\right)^{\frac{1}{\beta}} + v^2.$$

Taking the conditional expectation with respect to $\mathcal{F}_k$, one has

$$\mathbb{E}\left\{\|x_{k+1}-x\|^2 \mid \mathcal{F}_k\right\} \le \|x_k-x\|^2 - 2vL^{-\frac{1}{\beta}}\mathbb{E}\left\{\left(f^+_{\omega_k}(x_k)\right)^{\frac{1}{\beta}} \mid \mathcal{F}_k\right\} + v^2. \tag{3.22}$$

Noting that ω_k is uniformly distributed on I, we conclude by the elementary probability theory that

$$\mathbb{E}\left\{\left(f^+_{\omega_k}(x_k)\right)^{\frac{1}{\beta}} \mid \mathcal{F}_k\right\} = \frac{1}{m}\sum_{i\in I}\left(f^+_i(x_k)\right)^{\frac{1}{\beta}} \ge \frac{1}{m}\left(F^+(x_k)\right)^{\frac{1}{\beta}}.$$

This, together with (3.22), implies (3.21). The proof is complete. □

Using the basic inequality in Lemma 3.4 and following a line of analysis similar to that of Theorem 3.1, we can establish the convergence theorem and iteration complexity for Algorithm 3.1 with the stochastic control. The detailed proof is given in Appendix **??** of the supplementary material.

Theorem 3.5. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ being the stochastic control. Then the following assertions are true.*

(i) *It holds with probability 1 that* $\liminf_{k\to\infty} F^+(x_k) \le L\left(\frac{mv}{2}\right)^\beta$.

(ii) *Let $\delta>0$ and $K^{\mathrm{c}}_{\mathrm{s}} := \lceil \frac{m\mathrm{d}^2(x_1,S)}{2v\delta}\rceil$. Then* $\min\limits_{1\le k\le K^{\mathrm{c}}_{\mathrm{s}}} \mathbb{E}\{F^+(x_k)\} \le L\left(\frac{mv}{2}+\delta\right)^\beta$.

Under the assumption of the Hölder-type error bound property, we present a linear convergence rate of Algorithm 3.1 with the stochastic control.

Theorem 3.6. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ being the stochastic control. Suppose that* (3.1) *satisfies the Hölder-type error bound property of order q and modulus κ. Then the following assertions are true.*

(i) *If $q=2\beta$, then there exists $\tau\in[0,1)$ such that, for each $k\in\mathbb{N}$,*

$$\mathbb{E}\left\{\mathrm{d}^2(x_k,S)\right\} \le \tau^k\mathrm{d}^2(x_1,S) + \frac{mv}{2}(\kappa L)^{\frac{1}{\beta}}. \tag{3.23}$$

(ii) *If $q>2\beta$ and $v<\left(\frac{m}{2}\right)^{\frac{\beta}{q-\beta}}(\kappa L)^{\frac{1}{q-\beta}}\left(\frac{2\beta}{q}\right)^{\frac{q}{2(q-\beta)}}$, then there exists $\tau\in(0,1)$ such that, for each $k\in\mathbb{N}$,*

$$\mathbb{E}\left\{\mathrm{d}^2(x_k,S)\right\} \le \tau^k\mathrm{d}^2(x_1,S) + \left(\frac{mv}{2}(\kappa L)^{\frac{1}{\beta}}\right)^{\frac{2\beta}{q}}. \tag{3.24}$$

Proof. Fix $k\ge N$ and $x:=\mathrm{P}_S(x_k)$. Then (3.21) in Lemma 3.4 is reduced to

$$\mathbb{E}\left\{\mathrm{d}^2(x_{k+1},S) \mid \mathcal{F}_k\right\} \le \mathrm{d}^2(x_k,S) - \frac{2v}{m}L^{-\frac{1}{\beta}}\left(F^+(x_k)\right)^{\frac{1}{\beta}} + v^2.$$

By the assumption of the Hölder-type error bound property, (3.10) holds for each x_k. Then the above inequality is reduced to

$$\mathbb{E}\left\{\mathrm{d}^2(x_{k+1},S)\mid \mathcal{F}_k\right\}\le \mathrm{d}^2(x_k,S)-\frac{2v}{m}(\kappa L)^{-\frac{1}{\beta}}\mathrm{d}^{\frac{q}{\beta}}(x_k,S)+v^2.$$

Taking its expectation, we have by the convexity of $t^{\frac{q}{2\beta}}$ on $\mathbb{R}_+$ (as $q\ge 2\beta$) that

$$\mathbb{E}\left\{\mathrm{d}^2(x_{k+1},S)\right\}\le \mathbb{E}\left\{\mathrm{d}^2(x_k,S)\right\}-\frac{2v}{m}(\kappa L)^{-\frac{1}{\beta}}\left(\mathbb{E}\left\{\mathrm{d}^2(x_k,S)\right\}\right)^{\frac{q}{2\beta}}+v^2. \tag{3.25}$$

(i) Suppose that $q=2\beta$. Setting $\tau:=(1-\frac{2v}{m}(\kappa L)^{-\frac{1}{\beta}})_+\in[0,1)$ and by (3.25), we obtain

$$\mathbb{E}\left\{\mathrm{d}^2(x_{k+1},S)\right\}\le \tau\mathbb{E}\left\{\mathrm{d}^2(x_k,S)\right\}+v^2\quad\text{for each } k\in\mathbb{N}.$$

Then we inductively obtain (3.23), and the conclusion follows.

(ii) Suppose that $q>2\beta$ and $v<\left(\frac{m}{2}\right)^{\frac{\beta}{q-\beta}}(\kappa L)^{\frac{1}{q-\beta}}\left(\frac{2\beta}{q}\right)^{\frac{q}{2(q-\beta)}}$. Then by (3.25), Lemma 2.3(ii) is applicable (with $\mathbb{E}\{\mathrm{d}^2(x_k,S)\}$, $\frac{q}{2\beta}-1$, $\frac{2v}{m}(\kappa L)^{-\frac{1}{\beta}}$, v^2 in place of u_k, r, a, b) to obtaining (3.24). The proof is complete. □

Remark 3.5. *Note that Algorithm 3.1 with the stochastic control is reduced to [19, Algorithm 4]. Theorem 3.3(i) shows its convergence to a feasible solution of the QFP* (3.1) *within a tolerance of* $T_{\mathrm{s}}=L\left(\frac{mv}{2}\right)^{\beta}$. *This result improves the one in [19, Theorem 4.2] that has a larger tolerance of* $mL\left(\frac{v}{2}\right)^{\beta}$. *Moreover, Theorems 3.5(ii) and 3.6 extend the convergence theorem in [19] to the quantitative iteration complexity and linear convergence rate.*

3.2. Subgradient methods with a dynamic stepsize

Although the constant stepsize is easy to implement in practical applications; however, as shown in the preceding section, the convergence theory is only guaranteed to converge to an approximate feasible solution within a tolerance relevant to the stepsize. If the Hölder continuity parameters (including the order β and modulus L) are known, a dynamic stepsize rule (depending on the function value and parameters β and L) is quite popular and widely used in subgradient methods for solving the CFP [3] and the QFP [10]. The dynamic stepsize rule inherits the comparable computational cost as the constant stepsize rule, but enjoys the benefit of converging to an exact feasible solution. Here we discuss subgradient methods with a dynamic stepsize for solving the QFP (3.1) in a unified framework (of control schemes).

Algorithm 3.2. *Select an initial point* $x_1\in\mathbb{R}^n$ *and a sequence of stepsizes* $\{v_k\}\subseteq(0,+\infty)$ *satisfying*

$$0<\underline{v}\le v_k\le \overline{v}<2. \tag{3.26}$$

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

$$x_{k+1} := \mathrm{P}_X \left(x_k - v_k \sum_{i\in I_k} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L} \right)^{\frac{1}{\beta}} g_{k,i} \right). \tag{3.27}$$

Remark 3.6. (i) *The sequence generated by Algorithm 3.2 satisfies* (3.4).

(ii) *It follows from* (2.1) *that, when applied to the CFP, Algorithm 3.2 covers the convex subgradient method [3] and extends to the general control schemes.*

(iii) *Focused on the QFP, Algorithm 3.2 covers the subgradient methods presented in [10, 19]. In particular, Algorithm 3.2 covers [10, Algorithms 13, 15 and 17] when the control scheme is selected as the most violated constraint control, the almost cyclic control and the parallel control, respectively. Algorithm 3.2 covers [19, Algorithms 3 and 4] when the control scheme is selected as the cyclic control and the stochastic control, respectively.*

(iv) *If the Hölder continuity order β_i and modulus L_i for each f_i are known, the iteration* (3.27) *in Algorithm 3.2 can be revised to*

$$x_{k+1} := \mathrm{P}_X \left(x_k - v_k \sum_{i\in I_k} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L_i} \right)^{\frac{1}{\beta_i}} g_{k,i} \right). \tag{3.28}$$

Its quantitative convergence theory can also be established with $\beta := \min_{i\in I} \beta_i$ and $L := \max_{i\in I} L_i$ by following a line of analysis similar to that in the sequel. To simplify the notations, we adopt Assumption 1 and investigate the convergence theory of Algorithm 3.2 in the rest of this section.

In the remainder of this section, we aim to establish the quantitative convergence theory of Algorithm 3.2 with the α-most violated constraints control, the s-intermittent control and the stochastic control, respectively, under Assumptions 1 and 2. To this end, we first provide a basic inequality and the descent property of Algorithm 3.2 for arbitrary type of control scheme.

Lemma 3.5. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.2, and let $x \in S$. Then the following assertions are true.*

(i) *The following basic inequality holds for each $k \in \mathbb{N}$ that*

$$\|x_{k+1} - x\|^2 \le \|x_k - x\|^2 - \underline{v}(2 - \overline{v})\mu L^{-\frac{2}{\beta}} \sum_{i\in I_k} \left(f_i^+(x_k)\right)^{\frac{2}{\beta}}. \tag{3.29}$$

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

Proof. Assertion (ii) directly follows from assertion (i) of this lemma and (3.26). Hence it suffices to prove assertion (i). To this end, fix $x \in S$ and $k \in \mathbb{N}$. We assume, without loss of generality, that $x_k \notin S$; otherwise, $f_i^+(x_k) = 0$ for each $i \in I$, and thus (3.29) is satisfied automatically by (3.4) (see Remark 3.6(i)). It follows from (3.27) and the nonexpansive property of projection operator that

$$\begin{aligned}
\|x_{k+1} - x\|^2 &\leq \|x_k - v_k \sum_{i \in I_k} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L} \right)^{\frac{1}{\beta}} g_{k,i} - x\|^2 \\
&= \|x_k - x\|^2 - 2v_k \sum_{i \in I_k} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L} \right)^{\frac{1}{\beta}} \langle x_k - x, g_{k,i} \rangle \\
&\quad + v_k^2 \| \sum_{i \in I_k} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L} \right)^{\frac{1}{\beta}} g_{k,i}\|^2.
\end{aligned} \tag{3.30}$$

Note by $x \in S$ that $f_i(x) \leq 0$ for each $i \in I$. Then, for each $i \in I_k$ such that $f_i(x_k) > 0$, it follows from (2.2) in Lemma 2.1 that $\langle x_k - x, g_{k,i} \rangle \geq \left(\frac{f_i^+(x_k)}{L} \right)^{\frac{1}{\beta}}$; otherwise, $f_i^+(x_k) = 0$. Hence we conclude that

$$\sum_{i \in I_k} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L} \right)^{\frac{1}{\beta}} \langle x_k - x, g_{k,i} \rangle \geq \sum_{i \in I_k} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L} \right)^{\frac{2}{\beta}}. \tag{3.31}$$

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

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

(thanks to $\{\lambda_{k,i}\}_{i \in I_k} \in \mathbf{\Delta}_+^{|I_k|}$ and $g_{k,i} \in \mathbb{S}$). This, together with (3.30) and (3.31), implies

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

By (3.26) and Assumption 2, it is reduced to (3.29). The proof is complete. □

3.2.1. The α-most violated constraints control

This subsection aims to investigate the iteration complexity and the convergence rate for Algorithm 3.2 with the α-most violated constraints control. To proceed, by definition of the α-most violated constraints control (cf. Definition 3.1(a)), the following lemma about the basic inequality directly follows from Lemma 3.5.

Lemma 3.6. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1 with $\{I_k\}$ being the α-most violated constraints control. Then the following basic inequality holds for each $k \in \mathbb{N}$ that*

$$\mathrm{d}^2(x_{k+1}, S) \leq \mathrm{d}^2(x_k, S) - \underline{v}(2 - \overline{v})\mu L^{-\frac{2}{\beta}} \left(\alpha F^+(x_k) \right)^{\frac{2}{\beta}}. \tag{3.32}$$

By virtue of the basic inequality in Lemma 3.6 and following a line of analysis similar to [10], we first establish the convergence theorem and iteration complexity for Algorithm 3.2 with the α-most violated constraints control as a by-product, which extends that for the most violated constraint control and parallel control in [10, Theorems 14 and 18]. The proof is given in Appendix **??** of the supplementary material.

Theorem 3.7. *Let $\{x_k\}$ be generated by Algorithm 3.2 with $\{I_k\}$ being the α-most violated constraints control. Then the following assertions are true.*

(i) *$\{x_k\}$ converges to a feasible solution of the QFP* (3.1)*.*

(ii) *Let $\delta > 0$ and $K_{\rm m}^{\rm d} := \lceil \frac{{\rm d}^2(x_1,S)}{\underline{v}(2-\overline{v})\mu} \left(\frac{L}{\alpha\delta}\right)^{\frac{2}{\beta}} \rceil$. Then $\min_{1\le k\le K_{\rm m}^{\rm d}} F^+(x_k) \le \delta$.*

Remark 3.7. *Theorem 3.7 indicates an advantage of the dynamic stepsize that Algorithm 3.2 converges to an exact feasible solution, while Algorithm 3.1 is only guaranteed to converge to a neighbourhood (depending on the stepsize) of a feasible solution; see Theorem 3.1. Moreover, Theorem 3.7(ii) shows that Algorithm 3.2 with the α-most violated constraints control possesses an iteration complexity of $\mathcal{O}(1/\delta^{\frac{2}{\beta}})$ to an approximate δ-feasible solution.*

The main theorem of this subsection is as follows, which presents the convergence rates of Algorithm 3.2 with the α-most violated constraints control under the assumption of the Hölder-type error bound property.

Theorem 3.8. *Let $\{x_k\}$ be generated by Algorithm 3.2 with $\{I_k\}$ being the α-most violated constraints control. Suppose that* (3.1) *satisfies the Hölder-type error bound property of order q. Then the following assertions are true.*

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

$$\|x_k - \bar{x}\| \le c\tau^k \quad \text{for each } k \in \mathbb{N}. \tag{3.33}$$

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

$$\|x_k - \bar{x}\| \le ck^{-\frac{\beta}{2(q-\beta)}} \quad \text{for each } k \in \mathbb{N}. \tag{3.34}$$

Proof. By the assumption of the Hölder-type error bound property, there exists $\kappa > 0$ such that (3.10) holds for each x_k. This, together with (3.32), yields

$$ {\rm d}^2(x_{k+1}, S) \le {\rm d}^2(x_k, S) - \rho\, {\rm d}^{\frac{2q}{\beta}}(x_k, S) \quad \text{for each } k \in \mathbb{N}, $$

with $\rho := \underline{v}(2-\overline{v})\mu\left(\frac{\alpha}{\kappa L}\right)^{\frac{2}{\beta}}$. Consequently, there exists $c \geq 0$ such that

$$\mathrm{d}(x_k, S) \leq \begin{cases} c\tau^k, & \text{if } q = \beta, \\ ck^{-\frac{\beta}{2(q-\beta)}}, & \text{if } q > \beta, \end{cases} \tag{3.35}$$

with $\tau := \sqrt{1-\rho}$ and by applying Lemma 2.3(i) (with $\mathrm{d}^2(x_k, S)$, ρ, $\frac{q}{\beta} - 1$ in place of u_k, a, r), for each $k \in \mathbb{N}$.

Fix $l > k \in \mathbb{N}$. It follows from Lemma 3.5(ii) (taking $x := \mathrm{P}_S(x_k)$) that

$$\|x_l - x_k\| \leq \|x_l - \mathrm{P}_S(x_k)\| + \|x_k - \mathrm{P}_S(x_k)\| \leq 2\|x_k - \mathrm{P}_S(x_k)\| = 2\mathrm{d}(x_k, S).$$

Hence, by the convergence of $\{x_l\}$ to $\bar{x} \in S$ as shown in Theorem 3.7, we obtain

$$\|x_k - \bar{x}\| = \lim_{l\to\infty} \|x_l - x_k\| \leq 2\mathrm{d}(x_k, S).$$

This, together with (3.35), implies (3.33) and (3.34). The proof is complete. $\square$

As mentioned in Remark 3.6(iii), [10, Algorithm 13] can be covered by Algorithm 3.2 with the most violated constraint control and using $\{i(k)\}$ and 1 in place of I_k and $\lambda_{k,i}$; [10, Algorithm 17] can be covered by Algorithm 3.2 with the parallel control and using I and λ_i in place of I_k and $\lambda_{k,i}$. As direct applications of Theorems 3.7-3.8, we present the quantitative convergence theory of [10, Algorithms 13 and 17] as in the following corollaries, in which assertions (i) cover [10, Theorems 14 and 18], and more importantly, assertions (ii) and (iii) improve [10] to provide the quantitative iteration complexity and convergence rates. Moreover, we observe from Corollary 3.2(ii) that the best weighting strategy (in sense of the iteration complexity) for the parallel control scheme is equi-weights (that is, $\mu = \frac{1}{m}$), if we do not have any prior ordering information on component functions.

Corollary 3.1. *Let $\{x_k\}$ be generated by [10, Algorithm 13]. Then the following assertions are true.*

(i) *$\{x_k\}$ converges to a feasible solution of the QFP (3.1).*

(ii) *Let $\delta > 0$ and $K_{\mathrm{m}}^{\mathrm{d}} := \lceil \frac{\mathrm{d}^2(x_1,S)}{\underline{v}(2-\overline{v})} \left(\frac{L}{\delta}\right)^{\frac{2}{\beta}} \rceil$. Then $\min_{1\leq k\leq K_{\mathrm{m}}^{\mathrm{d}}} F^+(x_k) \leq \delta$.*

(iii) *Suppose that (3.1) satisfies the Hölder-type error bound of order q.*

(iii-a) *If $q = \beta$, then $\{x_k\}$ converges to a feasible solution at a linear rate.*

(iii-b) *If $q > \beta$, then there exist $\bar{x} \in S$ and $c \geq 0$ such that*

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

Corollary 3.2. *Let $\{x_k\}$ be generated by [10, Algorithm 17]. Then the following assertions are true.*

(i) *$\{x_k\}$ converges to a feasible solution of the QFP* (3.1)*.*

(ii) *Let $\delta > 0$ and $K_{\rm p}^{\rm d} := \lceil \frac{{\rm d}^2(x_1,S)}{\underline{v}(2-\overline{v})\mu} \left(\frac{L}{\delta}\right)^{\frac{2}{\beta}} \rceil$. Then $\min_{1 \le k \le K_{\rm p}^{\rm d}} F^+(x_k) \le \delta$.*

(iii) *Suppose that* (3.1) *satisfies the Hölder-type error bound of order q.*

(iii-a) *If $q = \beta$, then $\{x_k\}$ converges to a feasible solution at a linear rate.*

(iii-b) *If $q > \beta$, then there exist $\bar{x} \in S$ and $c \ge 0$ such that*

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

Remark 3.8. *Corollaries 3.1(ii) and 3.2(ii) can be also obtained from [10, (26) and (44)], respectively.*

3.2.2. The s-intermittent control

This subsection aims to provide the iteration complexity and the convergence rate of Algorithm 3.2 with the s-intermittent control, which is an extension of the almost cyclic control as used in [10, Algorithm 15] and the parallel control as used in [10, Algorithm 17]. To proceed, we first deduce its basic inequality by virtue of Lemma 3.5.

Lemma 3.7. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1 with $\{I_k\}$ being the s-intermittent control. Then the basic inequality holds for each $x \in S$ and $k \in \mathbb{N}$ that*

$$\mathrm{d}^2(x_{s(k+1)}, S) \le \mathrm{d}^2(x_{sk}, S) - \frac{2\underline{v}(2-\overline{v})\mu}{1+4s} \left(\frac{F^+(x_{sk})}{2L}\right)^{\frac{2}{\beta}}. \tag{3.36}$$

Proof. Fix $x \in S$ and $k \in \mathbb{N}$. By applying (3.29) in Lemma 3.5 inductively, we obtain that

$$\|x_{s(k+1)} - x\|^2 \le \|x_{sk} - x\|^2 - \underline{v}(2-\overline{v})\mu L^{-\frac{2}{\beta}} \sum_{j=0}^{s-1} \sum_{i \in I_{sk+j}} \left(f_i^+(x_{sk+j})\right)^{\frac{2}{\beta}}. \tag{3.37}$$

Below we estimate the second term on the right hand side of (3.37) in terms of $F^+(x_{sk})$. Firstly, let $i_k \in I$ be the most violated index of the inequality system (3.1) at x_{sk}, that is,

$$f_{i_k}^+(x_{sk}) = F^+(x_{sk}) > 0 \tag{3.38}$$

(otherwise, $\{x_i\}_{i>sk}$ terminates at $x_{sk} \in S$ due to (3.4)). By the definition of the s-intermittent control (cf. Definition 3.1(b)), there exists $j_k \in [0, s-1]$ such that $i_k \in I_{sk+j_k}$. Then one has by Assumption 1 that

$$\begin{aligned} f_{i_k}^{\frac{2}{\beta}}(x_{sk}) &\le \left(f_{i_k}(x_{sk+j_k}) + L\|x_{sk+j_k} - x_{sk}\|^\beta\right)^{\frac{2}{\beta}} \\ &\le 2^{\frac{2}{\beta}-1}\left((f_{i_k}^+(x_{sk+j_k}))^{\frac{2}{\beta}} + L^{\frac{2}{\beta}}\|x_{sk+j_k} - x_{sk}\|^2\right) \end{aligned} \tag{3.39}$$

(by Lemma 2.2(i) as $\beta \leq 1$). Since $i_k \in I_{sk+j_k}$ and $j_k \in [0, s-1]$, we get

$$\left(f_{i_k}^+(x_{sk+j_k})\right)^{\frac{2}{\beta}} \leq \sum_{i \in I_{sk+j_k}} \left(f_i^+(x_{sk+j_k})\right)^{\frac{2}{\beta}} \leq \sum_{j=0}^{s-1} \sum_{i \in I_{sk+j}} \left(f_i^+(x_{sk+j})\right)^{\frac{2}{\beta}}. \tag{3.40}$$

On the other side, in view of Algorithm 3.2, we obtain by (3.27) and the nonexpansive property of the projection operator that

$$\|x_{k+1} - x_k\|^2 \leq v_k^2 \|\sum_{i \in I_k} \lambda_{k,i} \left(\frac{f_i^+(x_k)}{L}\right)^{\frac{1}{\beta}} g_{k,i}\|^2 \leq 4 \sum_{i \in I_k} \left(\frac{f_i^+(x_k)}{L}\right)^{\frac{2}{\beta}}$$

(thanks to (3.26), $\{\lambda_{k,i}\}_{i \in I_k} \in \mathbf{\Delta}_+^{|I_k|}$ and $g_{k,i} \in \mathbb{S}$) for each $k \in \mathbb{N}$. Then we derive by Lemma 2.2(i) that

$$\|x_{sk+j_k} - x_{sk}\|^2 \leq s \sum_{j=0}^{s-1} \|x_{sk+j+1} - x_{sk+j}\|^2 \leq 4s \sum_{j=0}^{s-1} \sum_{i \in I_{sk+j}} \left(\frac{f_i^+(x_{sk+j})}{L}\right)^{\frac{2}{\beta}}.$$

This, together with (3.38)-(3.40), deduces that

$$\left(F^+(x_{sk})\right)^{\frac{2}{\beta}} \leq (1+4s)\, 2^{\frac{2}{\beta}-1} \sum_{j=0}^{s-1} \sum_{i \in I_{sk+j}} \left(f_i^+(x_{sk+j})\right)^{\frac{2}{\beta}}.$$

This, together with (3.37) (taking $x := \mathrm{P}_S(x_{sk})$), implies (3.36). The proof is complete. □

Using the basic inequality in Lemma 3.7 and a line of analysis similar to that of Theorems 3.7 and 3.8, we can obtain in the following theorems the iteration complexity and convergence rates of Algorithm 3.2 with the s-intermittent control to a feasible solution. The detailed proofs are given in Appendixes **??** and **??** of the supplementary material.

Theorem 3.9. *Let $\{x_k\}$ be generated by Algorithm 3.2 with $\{I_k\}$ being the s-intermittent control. Then the following assertions are true.*

(i) *$\{x_k\}$ converges to a feasible solution of the QFP (3.1).*

(ii) *Let $\delta > 0$ and $K_{\mathrm{c}}^{\mathrm{d}} := \lceil \frac{s(1+4s)\mathrm{d}^2(x_1,S)}{2\underline{v}(2-\overline{v})\mu} \left(\frac{2L}{\delta}\right)^{\frac{2}{\beta}} \rceil$. Then $\min_{1 \leq k \leq K_{\mathrm{c}}^{\mathrm{d}}} F^+(x_k) \leq \delta$.*

Theorem 3.10. *Let $\{x_k\}$ be generated by Algorithm 3.2 with $\{I_k\}$ being the s-intermittent control. Suppose that (3.1) satisfies the Hölder-type error bound property of order q. Then the following assertions are true.*

(i) *If $q = \beta$, then $\{x_k\}$ converges to a feasible solution at a linear rate.*

(ii) *If $q > \beta$, then there exist $\bar{x} \in S$ and $c \geq 0$ such that*

$$\|x_k - \bar{x}\| \leq c k^{-\frac{\beta}{2(q-\beta)}} \quad \text{for each } k \in \mathbb{N}.$$

Remark 3.9. *Theorem 3.9(ii) shows that Algorithm 3.2 with the s-intermittent control has an iteration complexity of* $\mathcal{O}(1/\delta^{\frac{2}{\beta}})$ *to an approximate* δ*-feasible solution, which admits the same order with the* α*-most violated constraints control (see Remark 3.7), but requires a much larger number of iterations when adopting the suitable parameters* s $(\sim m)$ *and* α (~ 1)*. Indeed, by Theorems 3.7 and 3.9, we obtain*

$$\frac{K_{\rm c}^{\rm d}}{K_{\rm m}^{\rm d}} = s(1+4s)2^{\frac{\beta}{2}-1} \gg 1.$$

As mentioned in Remark 3.6, [10, Algorithm 15] can be covered by Algorithm 3.2 with the almost cyclic control and using $\{i(k)\}$ and 1 in place of I_k and $\lambda_{k,i}$. As a direct application of Theorems 3.9 and 3.10, we present the quantitative convergence theory of [10, Algorithm 15] as follows, in which assertion (i) covers [10, Theorem 16], and more importantly, assertions (ii) and (iii) improve [10] to provide the iteration complexity and convergence rates.

Corollary 3.3. *Let* $\{x_k\}$ *be generated by [10, Algorithm 15]. Then the following assertions are true.*

(i) $\{x_k\}$ *converges to a feasible solution of the QFP* (3.1)*.*

(ii) *Let* $\delta > 0$ *and* $K_{\rm c}^{\rm d} := \lceil \frac{s(1+4s){\rm d}^2(x_1,S)}{2\underline{v}(2-\overline{v})} \left(\frac{2L}{\delta}\right)^{\frac{2}{\beta}} \rceil$*. Then* $\min_{1\le k\le K_{\rm c}^{\rm d}} F^+(x_k) \le \delta$*.*

(iii) *Suppose that* (3.1) *satisfies the Hölder-type error bound of order* q*.*

(iii-a) *If* $q = \beta$*, then* $\{x_k\}$ *converges to a feasible solution at a linear rate.*

(iii-b) *If* $q > \beta$*, then there exist* $\bar{x} \in S$ *and* $c \ge 0$ *such that*

$$\|x_k - \bar{x}\| \le c k^{-\frac{\beta}{2(q-\beta)}} \quad \textit{for each } k \in \mathbb{N}.$$

3.2.3. Stochastic control

This subsection aims to provide the iteration complexity and the convergence rate of Algorithm 3.1 the stochastic control. An interesting finding is disclosed by Theorem 3.11 that the stochastic control has a significant favorable effect on the performance of the subgradient method; concretely, the stochastic control enjoys both advantages of the low computational cost requirement of the (almost) cyclic control and the low (worst-case) iteration complexity of the parallel control; see Remark 3.10 for details. To proceed convergence analysis, we provide below a basic inequality in terms of conditional expectation.

Lemma 3.8. *Let* $\{x_k\}$ *be generated by Algorithm 3.2 with* $\{I_k\}$ *being the stochastic control, and let* $\mathcal{F}_k := \{x_1, \ldots, x_k\}$ *for each* $k \in \mathbb{N}$*. Then the following basic inequality holds for each* $x \in S$ *and* $k \in \mathbb{N}$ *that*

$$\mathbb{E}\left\{\|x_{k+1} - x\|^2 \mid \mathcal{F}_k\right\} \le \|x_k - x\|^2 - \frac{\underline{v}(2-\overline{v})}{m} L^{-\frac{2}{\beta}} \left(F^+(x_k)\right)^{\frac{2}{\beta}}. \tag{3.41}$$

Proof. Fix $x \in S$ and $k \in \mathbb{N}$. Since $I_k = \{\omega_k\}$ and $\mu = 1$ in the stochastic control (cf. Definition 3.1(c)), it follows from (3.29) in Lemma 3.5 that

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 - \underline{v}(2 - \overline{v})L^{-\frac{2}{\beta}} \left(f_{\omega_k}^+(x_k)\right)^{\frac{2}{\beta}}.$$

Taking the conditional expectation with respect to $\mathcal{F}_k$, one has

$$\mathbb{E}\left\{\|x_{k+1} - x\|^2 \mid \mathcal{F}_k\right\} \leq \|x_k - x\|^2 - \underline{v}(2 - \overline{v})L^{-\frac{2}{\beta}}\mathbb{E}\left\{\left(f_{\omega_k}^+(x_k)\right)^{\frac{2}{\beta}} \mid \mathcal{F}_k\right\}. \tag{3.42}$$

Noting by definition of the stochastic control (cf. Definition 3.1(c)) that ω_k is uniformly distributed on I, we have by the elementary probability theory that

$$\mathbb{E}\left\{\left(f_{\omega_k}^+(x_k)\right)^{\frac{2}{\beta}} \mid \mathcal{F}_k\right\} = \frac{1}{m}\sum_{i\in I}\left(f_i^+(x_k)\right)^{\frac{2}{\beta}} \geq \frac{1}{m}\left(F^+(x_k)\right)^{\frac{2}{\beta}}.$$

This, together with (3.42), implies (3.41). The proof is complete. □

Using the basic inequality in Lemma 3.8 and following a line of analysis similar to that of Theorem 3.7, we can establish the convergence theorem and iteration complexity for Algorithm 3.2 with the stochastic control. The detailed proof is given in Appendix **??** of the supplementary material.

Theorem 3.11. *Let $\{x_k\}$ be generated by Algorithm 3.2 with $\{I_k\}$ being the stochastic control. Then the following assertions are true.*

(i) *$\{x_k\}$ converges to a feasible solution of the QFP* (3.1) *with probability 1.*

(ii) *Let $\delta > 0$ and $K_{\mathrm{s}}^{\mathrm{d}} := \lceil \frac{m\mathrm{d}^2(x_1,S)}{\underline{v}(2-\overline{v})}\left(\frac{L}{\delta}\right)^{\frac{2}{\beta}}\rceil$. Then $\min\limits_{1\leq k\leq K_{\mathrm{s}}^{\mathrm{d}}} \mathbb{E}\{F^+(x_k)\} \leq \delta$.*

Remark 3.10. *Theorem 3.11(ii) provides a theoretical evidence for the benefit of the s-tochastic control in the sense of the iteration complexity. In particular, the stochastic control enjoys both advantages of the low computational cost requirement of the (almost) cyclic control (much less than the parallel control) and the low (worst-case) iteration complexity of the parallel control (much less than the cyclic control). Indeed, by Theorem 3.11 and Corollaries 3.2 and 3.3, we derive*

$$\frac{K_{\mathrm{s}}^{\mathrm{d}}}{K_{\mathrm{p}}^{\mathrm{d}}} = m\mu \sim 1 \quad \text{and} \quad \frac{K_{\mathrm{s}}^{\mathrm{d}}}{K_{\mathrm{c}}^{\mathrm{d}}} = \frac{m}{s(1+4s)}2^{1-\frac{2}{\beta}} \ll 1,$$

where the best weighting strategy for the parallel control is such that $\mu = \frac{1}{m}$ and the cyclic control is such that $s = m$.

Under the assumption of the Hölder-type error bound property, we present the convergence rates of Algorithm 3.2 with the stochastic control.

Theorem 3.12. *Let* $\{x_k\}$ *be generated by Algorithm 3.2 with* $\{I_k\}$ *being the stochastic control. Suppose that* (3.1) *satisfies the Hölder-type error bound property of order* q. *Then the following assertions are true.*

(i) *If* $q = \beta$, *then there exist* $c \geq 0$ *and* $\tau \in (0,1)$ *such that*

$$\mathbb{E}\{\mathrm{d}(x_k, S)\} \leq c\tau^k \quad \textit{for each } k \in \mathbb{N}. \tag{3.43}$$

(ii) *If* $q > \beta$, *then there exists* $c \geq 0$ *such that*

$$\mathbb{E}\{\mathrm{d}(x_k, S)\} \leq ck^{-\frac{\beta}{2(q-\beta)}} \quad \textit{for each } k \in \mathbb{N}. \tag{3.44}$$

Proof. Fix $k \geq N$ and $x := \mathrm{P}_S(x_k)$. Then (3.41) in Lemma 3.8 is reduced to

$$\mathbb{E}\left\{\mathrm{d}^2(x_{k+1}, S) \mid \mathcal{F}_k\right\} \leq \mathrm{d}^2(x_k, S) - \frac{\underline{v}(2-\overline{v})}{m} L^{-\frac{2}{\beta}} \left(F^+(x_k)\right)^{\frac{2}{\beta}}.$$

By the assumption of the Hölder-type error bound property, there exists $\tau > 0$ such that (3.10) holds for each x_k. Then the above inequality is reduced to

$$\mathbb{E}\left\{\mathrm{d}^2(x_{k+1}, S) \mid \mathcal{F}_k\right\} \leq \mathrm{d}^2(x_k, S) - \rho\, \mathrm{d}^{\frac{2q}{\beta}}(x_k, S),$$

where $\rho := \frac{\underline{v}(2-\overline{v})}{m}\left(\frac{1}{\kappa L}\right)^{\frac{2}{\beta}}$. Taking the expectation on the above inequality, we derive by the convexity of $t^{\frac{q}{\beta}}$ on $\mathbb{R}_+$ (as $q \geq \beta$) that

$$\mathbb{E}\left\{\mathrm{d}^2(x_{k+1}, S)\right\} \leq \mathbb{E}\left\{\mathrm{d}^2(x_k, S)\right\} - \rho\left(\mathbb{E}\left\{\mathrm{d}^2(x_k, S)\right\}\right)^{\frac{q}{\beta}}.$$

This shows that there exists $c \geq 0$ such that

$$\mathbb{E}\left\{\mathrm{d}^2(x_k, S)\right\} \leq \begin{cases} c\tau^{2k}, & \text{if } q = \beta, \\ ck^{-\frac{\beta}{(q-\beta)}}, & \text{if } q > \beta, \end{cases}$$

with $\tau := \sqrt{1-\rho}$ and by applying Lemma 2.3(i) (with $\mathbb{E}\left\{\mathrm{d}^2(x_k, S)\right\}$, ρ, $\frac{q}{\beta} - 1$ in place of u_k, a, r), for each $k \in \mathbb{N}$. These, together with that $\left(\mathbb{E}\{\mathrm{d}(x_k, S)\}\right)^2 \leq \mathbb{E}\left\{\mathrm{d}^2(x_k, S)\right\}$, imply (3.43) and (3.44), respectively. The proof is complete. □

Remark 3.11. *As shown in Theorems 3.7, 3.9 and 3.11, the sequence generated by Algorithm 3.2 with general control schemes converges to a feasible solution of the QFP* (3.1). *Hence the (linear or sublinear) convergence rates of Algorithm 3.2 can be established under a weaker assumption of the Hölder-type* local *error bound property at* $\bar{x} \in S$, *namely,*

for any $r > 0$, *there is* $\kappa_r > 0$ *such that* (3.10) *holds for each* $x \in X \cap \mathbb{B}(\bar{x}, r)$.

4. Numerical experiments

This section aims to consider the application of the feasibility problem to the multiple Cobb-Douglas (CD) production efficiency problem (MCDPE) [7] and to carry out numerical experiments to illustrate the performance of subgradient methods for solving the corresponding feasibility problem. Numerical results indicate the advantage on the modeling capability of the QFP over the CFP and show the high efficiency and stability of subgradient methods for solving the QFP.

In this numerical study, we consider the MCDPE that has a variety of important applications in economics and management science. The classical CD production efficiency problem aims to find an optimal strategy of factors such that the CD efficiency (profit/cost) of one production is maximized. Extended to the case of multiple productions, the MCDPE aims to find a feasible strategy such that the CD efficiency of these productions are larger than certain targets, respectively. Formally, consider a set of m productions with s projects and n factors. Let $x \in \mathbb{R}^n$ denote the amounts of n factors. For $i = 1, \dots, m$, the profit function of production i can be expressed as a CD function

$$\text{Profit}_i(x) := w_i \prod_{j=1}^{n} x_j^{a_{i,j}},$$

where $w_i \geq 0$ and $a_{i,\cdot} \in \mathbf{\Delta}_+^n$; and the cost function of production i is formulated as a linear function

$$\text{Cost}_i(x) := u_i + \langle c_{i,\cdot}, x \rangle,$$

where $u_i \geq 0$ and $c_{i,\cdot} \in \mathbb{R}_+^n$. Due to the constraints of daily profit or operating cost, the investment amounts of factors should fall in the following closed and convex constraint set

$$X := \{x \in \mathbb{R}_+^n : \|x\|_\infty \leq D, \langle b_{t,\cdot}, x \rangle \geq p_t, t = 1, \dots, s\},$$

where $D \geq 0$, $p_t \geq 0$ and $b_{t,\cdot} \in \mathbb{R}_+^n$. Given a family of targets $\{r_i \in \mathbb{R}_+ : i = 1, \dots, m\}$ on the production efficiency of these m productions, the MCDPE aims to find a feasible solution such that the CD efficiency of these productions are larger than given targets. Thus it can be formulated as a feasibility problem

$$x \in X \quad \text{and} \quad f_i(x) := r_i - \frac{\text{Profit}_i(x)}{\text{Cost}_i(x)} \leq 0 \quad \text{for } i = 1, \dots, m. \tag{4.1}$$

This is a QFP because each CD efficiency function is quasi-concave on $\mathbb{R}_+^n$ due to [33, Theorems 2.3.3 and 2.5.1]. Furthermore, as $X \subseteq \mathbb{R}_+^n$, the QFP (4.1) is equivalent to the following CFP

$$x \in X \quad \text{and} \quad g_i(x) := r_i \text{Cost}_i(x) - \text{Profit}_i(x) \leq 0 \quad \text{for } i = 1, \dots, m, \tag{4.2}$$

which could also be solved by the subgradient methods for the CFP [3].

In the numerical experiments, we will apply subgradient methods with two stepsize rules and several types of control schemes to solve the QFP (4.1) associated to the MCDPE, and compare with subgradient methods for solving the CFP (4.2). The simulation data of the MCDPE are randomly generated as follows. Firstly, the parameters of the MCDPE are randomly generated by a uniform distribution from different intervals:

$$w_i \in [0,10], \quad a_{i,j}, b_{t,j}, u_i, c_{i,j} \in [0,1], \quad \text{and} \quad p_t \in [0, n/2];$$

then a feasible point $\bar{x} \in X$ is generated by a uniform distribution from $[0, D]$ with $D := 100$, and the targets of CD efficiency (with an additive noise) are calculated by

$$r_i := \frac{\text{Profit}_i(\bar{x})}{\text{Cost}_i(\bar{x})} + 10^{-6}\epsilon_i,$$

where ϵ_i is a uniformly distributed random variable from $[0,1]$. In the implementation of Algorithm 3.2, the Hölder continuity order is set as $\beta_i := \min_{j=1,\ldots,n} a_{i,j}$ as in (3.28). The Hölder continuity modulus L is a constant that is not easy to estimate (see Remark 3.1(iii)); alternatively we involve this modulus L in the setting of the stepsize v, that is to set $v \leftarrow vL^{-\frac{1}{\beta}}$ as different values. In the implementation of Algorithms 3.1 and 3.2 (as well as subgradient methods for the CFP), the stepsize is set to be $v = 1$ as default.

All numerical experiments are implemented in MATLAB R2014a and executed on a personal desktop (Intel Core Duo i7-8550, 1.80 GHz, 8.00 GB of RAM). The performance of each algorithm is evaluated by:

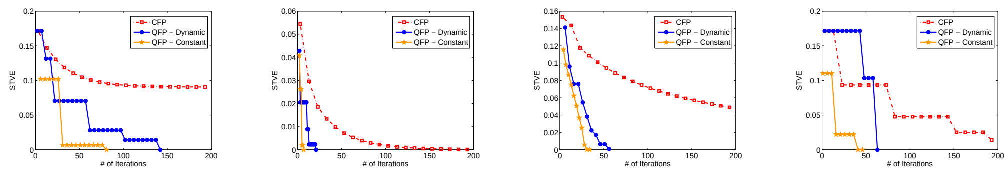
- (Accuracy) The sum of total violation error (STVE):

$$\text{STVE} := \sum_{i=1}^{m} \left(r_i - \frac{\text{Profit}_i(x)}{\text{Cost}_i(x)} \right)^{+}.$$

- (Speed) The CPU time cost from the algorithm begins to it stops, in which the stopping criterion is set as STVE $\leq 1e$-6 or the number of iterations is larger than 200.

- (Stability) The ratio of successful estimating, in which STVE $< 1e$-5.

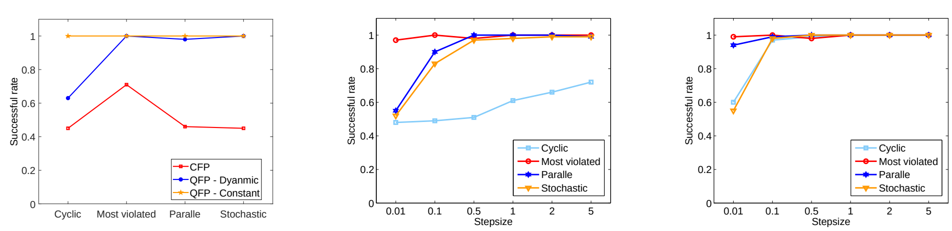
The first experiment is to compare the convergence behavior of subgradient methods for solving the QFP (4.1) and the CFP (4.2) with two stepsize rules (including the constant stepsize and the dynamic stepsize) and different types of control schemes (including the cyclic control, the most violated constraint control, the parallel control and the stochastic control). Figure 1 plots the variation of the STVE along the number of iterations for the MCDPE problem with $(m, n, s) = (50, 10, 10)$. It is illustrated in Figure 1 that subgradient methods (with both the constant and dynamic stepsize rules) for solving the QFP converge much faster than that for the CFP, and they approach a feasible solution of the MCDPE within a few iterations for these typical control schemes.



(a) Cyclic (b) Most violated (c) Parallel (d) Stochastic

Figure 1: Convergence behavior of subgradient methods.

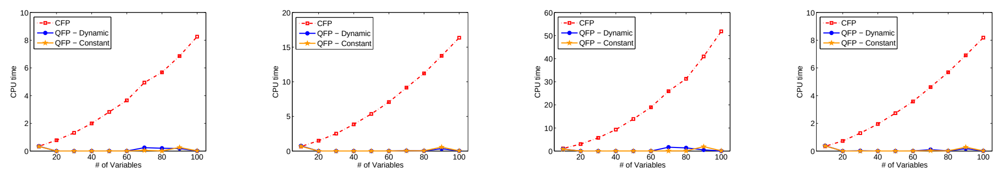
The second experiment is to show the stability of subgradient methods for solving the MCDPE problem with $(m, n, s) = (50, 10, 10)$. Conducting 100 random simulations, Figure 2(a) plots the successful rates of subgradient methods for the QFP (4.1) and the CFP (4.2) with different control schemes, and Figures 2(b) and (c) draws the successful rates of subgradient methods for the QFP with the dynamic stepsize and the constant stepsize, respectively, in which the stepsize v varies from [0.01, 5]. It is illustrated from Figure 2(a) subgradient methods with both the constant and dynamic stepsize rules for solving the QFP are more stable than that for solving the CFP, and from Figures 2(b) and (c) that subgradient methods for the QFP are quite stable on the stepsize and have a high successful rate, except a too small stepsize leads to a quite slow convergence.



(a) CFP - QFP (b) QFP - dynamic (c) QFP - constant

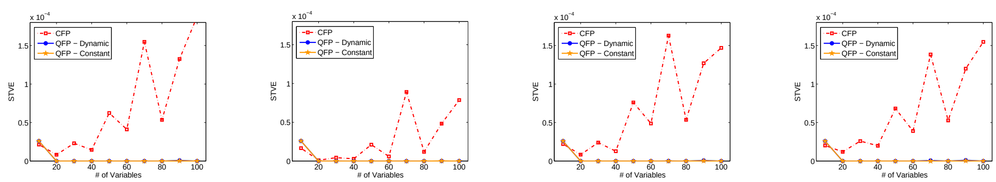
Figure 2: Successful rate of subgradient methods.

The third experiment is to illustrate the numerical performance (in terms of the STVE and the CPU time) of subgradient methods for different dimensional MCDPE. In this experiment, we set $(m, n, s) = (5n, n, n)$ with n varying from [10,100]. Figure 3 denotes the CPU time (seconds) cost by subgradient methods for solving the QFP and the CFP, and Figure 4 denotes the STVE of solutions obtained by subgradient methods along with the variation of n, with different stepsize rules and control schemes. Figures 3 and 4 demonstrate that subgradient methods for the QFP obtain a more accurate feasible solution of the MCDPE within much less CPU time than that for the CFP, especially for large-scale problems. Numerical results of these three experiments reveal the advantage of the QFP on the modeling capability over the CFP and show the high efficiency and stability of subgradient methods (with the constant/dynamic stepsize rules and typical control schemes) for solving the QFP.



(a) Cyclic (b) Most Violated (c) Parallel (d) Stochastic

Figure 3: CPU time of subgradient methods.



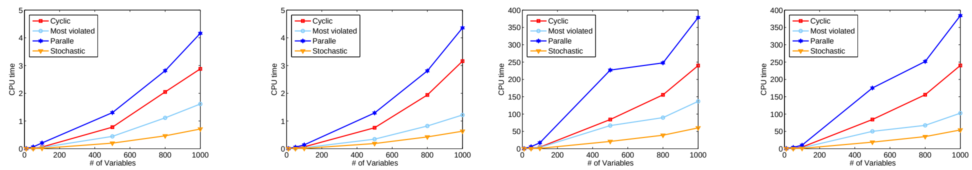
(a) Cyclic (b) Most Violated (c) Parallel (d) Stochastic

Figure 4: Accuracy of subgradient methods.

The last experiment aims to compare the numerical capability of these four types of control schemes in subgradient methods for solving the large-scale QFP, e.g., $(m, n, s) = (5n, n, n)$ or $(m, n, s) = (500n, n, n)$ with n varying from [10,1000]. Subgradient methods with these four types of control schemes can all obtain a feasible solution of the MCDPE within a short CUP time. Figure 5 shows the CPU time (seconds) cost by subgradient methods with different stepsize rules and control schemes to obtain a feasible solution of the QFP when $m = 5n$ and $m = 500n$, respectively. It is observed that the stochastic control converges faster than the most violated constraint control, both of which cost less CPU time than the cyclic control and the parallel control. This result is consistent with Remarks 3.4, 3.9 and 3.10.

5. Conclusion

In this paper, we proposed a unified framework of subgradient methods for solving the QFP, in which the constant/dynamic stepsize rules and several general control schemes were discussed. The quantitative convergence results, including the iteration complexity



(a) $m = 5n$, Constant stepsize

(b) $m = 5n$, Dynamic stepsize

(c) $m = 500n$, Constant stepsize

(d) $m = 500n$, Dynamic stepsize

Figure 5: CPU time of subgradient methods for large-scale QFP.

and the convergence rates, of subgradient methods were explored under the assumptions of the Hölder condition and/or the Hölder-type error bound property. Particularly, the iteration complexity results validated the benefit of the stochastic control that it enjoys both advantages of low computational cost requirement and low (worst-case) iteration complexity; the linear (or sublinear) convergence rates of subgradient methods to a feasible solution were established via an error bound-based analysis.

A difficulty on estimating the Hölder continuity modulus remains in the implementation of subgradient methods with the dynamic stepsize. To avert this obstacle, Nesterov [25] proposed a universal gradient method with an adaptive adjustment strategy of Hölder continuity modulus for its gradient. However, due to the difference between the analysis of gradient methods and that of subgradient methods, it still remains an open question how to establish the convergence theory of subgradient methods with an adaptive adjustment strategy of Hölder continuity modulus for the objective function.

Acknowledgement

The authors are grateful to the editor and the anonymous reviewers for their valuable comments and suggestions toward the improvement of this paper. This research was supported by the National Natural Science Foundation of China (12071306, 11871347), Natural Science Foundation of Guangdong (2019A1515011917, 2020B1515310008), Project of Educational Commission of Guangdong Province of China (2019KZDZX1007), Natural Science Foundation of Shenzhen (JCYJ20190808173603590), Hong Kong Polytechnic University (A-PL05, B-Q31F), Post-doctoral Fellowship of Department of Logistics and Maritime Studies and the Research Grants Council of Hong Kong (UGC/FDS14/P02/17).

References

[1] Aussel, D., & Pištěk, M. (2015). Limiting normal operator in quasiconvex analysis. *Set-Valued and Variational Analysis*, *23*, 669–685.

[2] Avriel, M., Diewert, W. E., Schaible, S., & Zang, I. (1988). *Generalized Concavity*. New York: Plenum Press.

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

[4] Bertsekas, D. P. (2015). *Convex Optimization and Algorithms*. Massachusetts: Athena Scientific.

[5] Borwein, J. M., Li, G., & Yao, L. (2014). Analysis of the convergence rate for the cyclic projection algorithm applied to basic semialgebraic convex sets. *SIAM Journal on Optimization*, *24*, 498–527.

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

[7] Bradley, S. P., & Frey, S. C. (1974). Fractional programming with homogeneous functions. *Operations Research*, *22*, 350–357.

[8] Byrne, C. (2014). *Iterative Optimization in Inverse Problems*. New York: CRC Press.

[9] Censor, Y., & Lent, A. (1982). Cyclic subgradient projections. *Mathematical Programming*, *24*, 233–235.

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

[11] Combettes, P. L. (1996). The convex feasibility problem in image recovery. (pp. 155–270). Elsevier volume 95 of *Advances in Imaging and Electron Physics*.

[12] Greenberg, H. J., & Pierskalla, W. P. (1973). Quasiconjugate functions and surrogate duality. *Cahiers Centre Études Recherche Opertionnelle*, *15*, 437–448.

[13] Hadjisavvas, N., Komlósi, S., & Schaible, S. (2005). *Handbook of Generalized Convexity and Generalized Monotonicity*. New York: Springer-Verlag.

[14] Hishinuma, K., & Iidukab, H. (2020). Fixed point quasiconvex subgradient method. *European Journal of Operational Research*, *282*, 428–437.

[15] Hoffman, A. J. (1952). On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, *49*, 263–265.

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

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

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

[19] Hu, Y. H., Yu, C. K. W., & Yang, X. Q. (2019). Incremental quasi-subgradient methods for minimizing the sum of quasi-convex functions. *Journal of Global Optimization*, *75*, 1573–2916.

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

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

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

[23] Luo, X.-D., & Luo, Z.-Q. (1994). Extension of Hoffman's error bound to polynomial systems. *SIAM Journal on Optimization*, *4*, 383–392.

[24] Nedić, A., & Bertsekas, D. P. (2001). Incremental subgradient methods for nondifferentiable optimization. *SIAM Journal on Optimization*, *12*, 109–138.

[25] Nesterov, Y. (2015). Universal gradient methods for convex optimization problems. *Mathematical Programming*, *152*, 381–404.

[26] Ng, K. F., & Zheng, X. Y. (2000). Global error bounds with fractional exponents. *Mathematical Programming*, *88*, 357–370.

[27] Pang, J.-S. (1997). Error bounds in mathematical programming. *Mathematical Programming*, *79*, 299–332.

[28] Papa Quiroz, E. A., Apolinario, H. C. F., Villacorta, K. D., & Oliveira, P. R. (2019). A linear scalarization proximal point method for quasiconvex multiobjective minimization. *Journal of Optimization Theory and Applications*, *183*, 1028–1052.

[29] Papa Quiroz, E. A., & Oliveira, P. R. (2012). An extension of proximal methods for quasiconvex minimization on the nonnegative orthant. *European Journal of Operational Research*, *216*, 26–32.

[30] Papa Quiroz, E. A., Ramirez, L. M., & Oliveira, P. R. (2015). An inexact proximal method for quasiconvex minimization. *European Journal of Operational Research*, *246*, 721–729.

[31] Polyak, B. (2001). Random algorithms for solving convex inequalities. In D. Butnariu, Y. Censor, & S. Reich (Eds.), *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications* (pp. 409–422). Elsevier volume 8 of *Studies in Computational Mathematics*.

[32] Shor, N. Z. (1985). *Minimization Methods for Non-differentiable Functions*. New York: Springer-Verlag.

[33] Stancu-Minasian, I. M. (1997). *Fractional Programming*. Dordrecht: Kluwer Academic Publishers.

[34] Suzuki, S., & Kuroiwa, D. (2017). Nonlinear error bounds for quasiconvex inequality systems. *Optimization Letters*, *11*, 107–120.

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

[36] Wang, M., & Bertsekas, D. P. (2016). Stochastic first-order methods with random constraint projection. *SIAM Journal on Optimization*, *26*, 681–717.

[37] Wang, X. M., Hu, Y. H., Li, C., Guu, S.-M., & Yao, J.-C. (2021). Convergence rates of subgradient algorithms for convex feasibility problems . *Preprint*.

[38] Yang, K., & Murty, K. G. (1992). New iterative methods for linear inequalities. *Journal of Optimization Theory and Applications*, *72*, 163–185.

[39] Yue, M.-C., Zhou, Z., & So, A. M.-C. (2019). 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*, 327–358.

[40] Zaslavski, A. J. (2016). *Approximate Solutions of Common Fixed-Point Problems*. Switzerland: Springer.