J. Nonlinear Var. Anal. 6 (2022), No. 2, pp. 15-33
Available online at http://jnva.biemdas.com
https://doi.org/10.23952/jnva.6.2022.2.03

MULTIPLE-SETS SPLIT QUASI-CONVEX FEASIBILITY PROBLEMS: ADAPTIVE SUBGRADIENT METHODS WITH CONVERGENCE GUARANTEE

YAOHUA HU[1], GANG LI[2,*], MINGHUA LI[3], CARISA KWOK WAI YU[4]

[1]*College of Mathematics and Statistics,*
Shenzhen Key Laboratory of Advanced Machine Learning and Applications,
Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen 518060, China
[2]*Department of Mathematics, Zhejiang Sci-Tech University, Hangzhou 310018, China*
[3]*School of Mathematics and Big Data,*
Chongqing University of Arts and Sciences, Yongchuan, Chongqing 402160, China
[4]*Department of Mathematics, Statistics and Insurance,*
The Hang Seng University of Hong Kong, Shatin, Hong Kong

Abstract. In this paper, we consider a multiple-sets split quasi-convex feasibility problem (MSSQFP), which is to find a point such that itself and its image under a linear transformation fall within two families of sublevel sets of quasi-convex functions in the space and the image space, respectively. A unified framework of the adaptive subgradient methods with general control schemes is proposed to solve the MSSQFP. This paper is contributed to establish the quantitative convergence theory of adaptive subgradient methods with several general control schemes. An interesting finding is disclosed by the iteration complexity results that the stochastic control enjoys both advantages of low computational cost requirement and low iteration complexity. In addition, a notion of the Hölder-type bounded error bound property for the MSSQFP is introduced, and the linear/sublinear convergence rates for the adaptive subgradient methods to a feasible solution of the MSSQFP is established.

Keywords. Convergence rate; Iteration complexity; Quasi-convex optimization; Split feasibility problem; Subgradient method.

1. Introduction

Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a linear operator. Let C and Q be nonempty, closed, and convex sets in $\mathbb{R}^n$ and $\mathbb{R}^m$, and f and g be convex and continuous functions on $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. The split feasibility problem (SFP) is to find a point $x \in \mathbb{R}^n$ such that

$$f(x) \le 0 \quad \text{and} \quad g(Ax) \le 0. \tag{1.1}$$

Problem (1.1) includes the following widely studied SFP

$$x \in C \quad \text{and} \quad Ax \in Q. \tag{1.2}$$

*Corresponding author.
E-mail addresses: mayhhu@szu.edu.cn (Y. Hu), ligang@zstu.edu.cn (G. Li), minghuali20021848@163.com (M. Li), carisayu@hsu.edu.hk (C.K.W. Yu).

Received February 12, 2022; Accepted March 6, 2022.

as a special case by taking $f(\cdot) := \mathrm{d}_C(\cdot)$ and $g(\cdot) := \mathrm{d}_Q(\cdot)$, the distance functions to C and Q. The SFP was firstly introduced by Censor and Elfving [1] to solve the phase retrieval problems. It provides a unified framework for many inverse problems in mathematics and physical sciences, and has been extensively applied in various areas, such as signal processing [2], image reconstruction [3], intensity-modulated radiation therapy [4], and systems biology [5, 6]. Motivated by its vast applications, various numerical algorithms have been developed to solve the SFP. The most popular and practical algorithms are the subgradient methods for problem (1.1) [3, 7] and the CQ algorithms for problem (1.2) [6, 8].

To measure the separable structure involved in the SFP precisely, the multiple-sets split feasibility problem (MSSFP) was introduced by Censor et al. [4], which is a generalization of the SFP with a series of convex and continuous functions $\{f_i\}$ and $\{g_j\}$ or with a series of closed and convex sets $\{C_i\}$ and $\{Q_i\}$. Several works have been devoted to the development of projection-type algorithms, the extensions of CQ algorithm, for solving the MSSFP with different stepsize (such as constant and adaptive stepsizes) and different control schemes (such as cyclic and parallel controls); see, e.g., [4, 9–13] and the references therein.

Most literature mentioned above considered the SFP (1.1) with convex component functions; however, the convex function is 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 the convex function; e.g., Sharpe ratio in portfolio selection and Cobb-Douglas production efficiency in management. In recent decades, much attention has been drawn to quasi-convex optimization; see [14–19] and the references therein. In the scenario of the SFP (1.1), Nimana, Farajzadeh, and Petrot [20] considered the split quasi-convex feasibility problem (SQFP), where the involved functions f and g are quasi-convex and continuous. They also proposed an adaptive subgradient method to solve the SQFP and established its global convergence property.

In addition to the global convergence property, the establishment of convergence rate is important in guaranteeing the numerical performance of relevant algorithms. For the SFP (1.2), the linear convergence rate of the CQ algorithms (with varying and adaptive stepsizes) was established under an assumption of bounded linear regularity property [6]. For the quasi-convex feasibility problem, i.e., (1.1) with A being an identity matrix, the linear/sublinear convergence rates of subgradient methods (with constant and adaptive stepsizes) were explored under an assumption of Hölder-type error bound property [16]. However, to the best of our knowledge, there is limited study devoted to establishing the convergence rate of subgradient methods for solving the SQFP.

In the present paper, we consider the multiple-sets split quasi-convex feasibility problem (MSSQFP): Find a point $x \in \mathbb{R}^n$ such that

$$f_i(x) \le 0,\ \forall i \in I, \quad \text{and} \quad g_j(Ax) \le 0,\ \forall j \in J,$$

where I and J are finite index sets, and $\{f_i : i \in I\}$ and $\{g_j : j \in J\}$ be two families of quasi-convex and continuous functions. Inspired by the ideas in [16, 20], we propose the adaptive subgradient methods for solving the MSSQFP in a unified framework (see Algorithm 3.1), which covers most types of control schemes discussed in the literature. In particular, the α-most violated constraints control, the s-intermittent control, and the stochastic control are considered in this paper.

The main contribution of the present paper is to establish the quantitative convergence theory, including the global convergence, the iteration complexity and the convergence rate, of the adaptive subgradient methods with several general control schemes for solving the MSSQFP. In particular, we first establish the global convergence theorem of the adaptive subgradient methods to a feasible solution of the MSSQFP; see Theorems 4.1 and 4.4. Furthermore, we derive its (worst-case) iteration complexity to obtain a feasible solution; see Theorems 4.2 and 4.5. More importantly, we introduce a notion of the Hölder-type bounded error bound property for the MSSQFP and use it to explore the linear/sublinear convergence rates of adaptive subgradient methods; see Theorems 4.3 and 4.6. The established theorems not only extend the subgradient methods in [7, 20] to the scenario of multiple-sets context and quasi-convex system, but also improve the global convergence results to the quantitative complexity and linear convergence rate. As far as we know, the establishment of iteration complexity and convergence rates of adaptive subgradient methods are new in the literature of MSSQFP.

Moreover, the iteration complexity and convergence rates of the adaptive subgradient method with a stochastic control are presented in terms of the expectation of violation and the expectation of distance from the solution set in Theorems 4.9 and 4.10, respectively. This paper seems to be the first attempt to propose the stochastic control in subgradient methods for solving the MSSQFP, 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 4.3 for explanation.

The present paper is organized as follows. In Section 2, we present the notations and some preliminary lemmas which will be used in this paper. We provide a unified framework of adaptive subgradient methods with several general control schemes to solve the MSSQFP in Section 3, and establish the quantitative convergence theory in Section 4.

2. Notations and Preliminary Results

Notations used in this 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 Δ^m_+, that is,

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

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

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

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

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

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

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

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 $\mathrm{lev}_f^\le(x)$) is convex for each $x\in\mathbb{R}^n$.*

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 that are defined via the "normal cone" to the sublevel sets; see [17, 21, 22] and the references therein. In particular, Kiwiel [23], Censor and Segal [24], and Hu et al. [17] employed the following quasi-subgradient, defined as a normal vector to its strict sublevel set, in their concerned subgradient methods. It was reported in [17, Lemma 2.1] that the quasi-subdifferential of f is nontrivial whenever f is quasi-convex.

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.

As a normal cone to its sublevel set, the quasi-subdifferential of a quasi-convex function contains at least a unit vector. This is a special property of the quasi-subdifferential beyond the convex subdifferential (for a convex function). Moreover, it was claimed in [17] that the quasi-subdifferential coincides with the convex cone hull of the convex subdifferential whenever f is convex.

The notion of the Hölder condition 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 and is equivalent to the bounded subgradient assumption, which is commonly assumed in the convergence study of subgradient methods for convex optimization problems; see [25] and the references therein.

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

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

The Hölder condition is a common assumption in convergence analysis of subgradient-type methods for quasi-convex optimization; see [18, 26–30] and the references therein. A fundamental property of the quasi-subgradient was provided under the Hölder condition in [31, Proposition 2.1] and in [16, Lemma 2.1], which plays an important role in the establishment of a basic inequality in convergence analysis of subgradient-type algorithms.

Lemma 2.1 ([16, 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 $S := \{x \in X : f(x) \le 0\}$. Let $0 < \beta \le 1$ and $L > 0$, and suppose that f satisfies the Hölder condition of order β with modulus L on $\mathbb{R}^n$. Then, for any $x \in S$ and $y \in X \setminus S$, it holds that*

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

We end this section by recalling the following two lemmas, which will be used in convergence rate analysis of subgradient methods.

Lemma 2.2 ([16, Lemma 2.2]). *Let $a_i \ge 0$ for $i = 1,2,\ldots,n$. Then the following assertions are true.*

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

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

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

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

Lemma 2.3 ([26, Lemma 2.2]). *Let $\sigma > 0$ and $a > 0$, and let $\{u_k\} \subseteq \mathbb{R}_+$ be a sequence of nonnegative scalars such that*

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

Then, it holds that

$$u_k \le u_0 (1 + a\sigma u_0^{\sigma} k)^{-\frac{1}{\sigma}} \quad \text{for each } k \in \mathbb{N}.$$

3. Adaptive Subgradient Methods for MSSQFP

Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a linear operator, and let $X \subseteq \mathbb{R}^n$ be a closed and convex set. Let $I := \{1,2,\ldots,M\}$ and $J := \{1,2,\ldots,N\}$ be two finite index sets, and let $\{f_i : i \in I\}$ and $\{g_j : j \in J\}$ be two families of quasi-convex and continuous functions defined on $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. In this paper, we consider the multiple-sets split quasi-convex feasibility problem (MSSQFP) that is to find a feasible point $x \in \mathbb{R}^n$ such that

$$x \in X, \quad f_i(x) \le 0, \ \forall i \in I, \quad \text{and} \quad g_j(Ax) \le 0, \ \forall j \in J. \tag{3.1}$$

Remark 3.1. (i) This type of MSSQFP includes the SFP, MSSFP, and SQFP as special cases. Particularly, it covers the SQFP [20] and the MSSFP [4] when $I = J = \{1\}$ and $\{f_i : i \in I\}$ and $\{g_j : j \in J\}$ are convex, respectively, and covers the SFP [1] when both of them are satisfied.

(ii) This type of MSSQFP includes the feasibility problem as special cases when A is an identity matrix, either convex [32] or quasi-convex [16, 33]. 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 [34], wireless sensor networks localization [35], protein conformation determination [36], and gene regulatory network inference [6].

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

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

Throughout this paper, we always assume that each component function of the MSSQFP (3.1) satisfies the Hölder condition as follows.

Assumption 3.1. There exist $\beta \in (0,1]$ and $L > 0$ such that $\{f_i\}_{i\in I}$ and $\{g_j\}_{j\in J}$ satisfy the Hölder condition of order β with modulus L on X and on AX, respectively.

Remark 3.2. Assumption 3.1 premises the unified Hölder continuity order and modulus for all component functions of the MSSQFP (3.1). This is actually equivalent to the Hölder condition of each f_i and each g_j with different orders and moduli as assumed in [18, 24] if X is bounded; one can also refer to [16, Remark 3.2]. Indeed, suppose that each f_i and each g_j in (3.1) satisfies the Hölder condition of order $\beta_{f_i} \in (0,1]$ and $\beta_{g_j} \in (0,1]$ with modulus $L_{f_i} > 0$ and $L_{g_j} > 0$ on X and AX, respectively. Then one can check that Assumption 3.1 is satisfied with

$$\beta := \min\left\{\min_{i\in I}\ \beta_{f_i},\ \min_{j\in J}\beta_{g_j}\right\}$$

and

$$L := \max\left\{\max_{i\in I}\ L_{f_i}\operatorname{diam}(X)^{\beta_{f_i}-\beta},\ \max_{j\in J}\ L_{g_j}\operatorname{diam}(AX)^{\beta_{g_j}-\beta}\right\}.$$

One of the most popular and practical numerical algorithms for solving the split (convex or quasi-convex) feasibility problem is a class of subgradient methods; see [3, 20] and references therein. Here we propose the adaptive subgradient methods for solving the MSSQFP (3.1) in a general framework (of control schemes), stated as follows. For the sake of simplicity, we write

$$\theta := \max\{1, \|A\|^2\}. \tag{3.2}$$

Algorithm 3.1. *Select an initial point $x_1 \in \mathbb{R}^n$ and a sequence of stepsizes $\{\lambda_k\} \subseteq (0,+\infty)$ satisfying*

$$0 < \underline{\lambda} \le \lambda_k \le \overline{\lambda} < \frac{1}{\theta}. \tag{3.3}$$

For each $k \in \mathbb{N}$, having $x_k \in \mathbb{R}^n$, we select nonempty index sets $I_k \subseteq I$ and $J_k \subseteq J$, weights $\{\mu_{k,i}\}_{i\in I_k} \in \Delta_+^{|I_k|}$ and $\{\nu_{k,j}\}_{j\in J_k} \in \Delta_+^{|J_k|}$, calculate quasi-subgradients $\phi_{k,i} \in \partial^* f_i(x_k) \cap \mathbb{S}$ for each $i \in I_k$ and $\psi_{k,j} \in \partial^* g_j(Ax_k) \cap \mathbb{S}$ for each $j \in J_k$, and update x_{k+1} by

$$x_{k+1} := \mathrm{P}_X\left(x_k - \lambda_k \sum_{i\in I_k}\mu_{k,i}\left(\frac{f_i^+(x_k)}{L}\right)^{\frac{1}{\beta}}\phi_{k,i} - \lambda_k\sum_{j\in J_k}\nu_{k,j}\left(\frac{g_j^+(Ax_k)}{L}\right)^{\frac{1}{\beta}}A^\top\psi_{k,j}\right). \tag{3.4}$$

It is clear by (3.4) that

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

Remark 3.3. (i) Algorithm 3.1 extends subgradient methods for solving the SQFP [20] and MSSFP [7] to the scenario of multiple-sets context and quasi-convex system, and provides a unified framework of general control schemes. In particular, when $I = J = \{1\}$ and $\{f_i : i \in I\}$, take

$$(\mu_k, \nu_k) := \begin{cases} (0,1), & \text{if } k \text{ is odd},\\ (1,0), & \text{if } k \text{ is even}, \end{cases}$$

Algorithm 3.1 is reduced to [20, Algorithm 3.3] for the SQFP.

(ii) Algorithm 3.1 extends subgradient methods for solving the CFP [32] and QFP [16] to the context of split feasibility problems. When A is an identity matrix, Algorithm 3.1 is reduced to subgradient methods for solving the QFP [16] and covers the ones in [24, 32].

To guarantee the convergence property of the adaptive subgradient methods, we shall assume the following blanket condition on parameters. It 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 [32, Remark 3.13] and [16, Assumption 2].

Assumption 3.2. There exists an $\eta > 0$ such that $\min_{i\in I_k} \mu_{k,i} \geq \eta$ and $\min_{j\in J_k} \nu_{k,j} \geq \eta$ for each $k \in \mathbb{N}$.

Under the blanket Assumptions 3.1 and 3.2, 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 and $x \in S$. Then the following assertions are true.*

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

$$\|x_{k+1}-x\|^2 \leq \|x_k - x\|^2 - 2\underline{\lambda}(1-\theta\overline{\lambda})\eta L^{-\frac{2}{\beta}} \left(\sum_{i\in I_k} f_i^+(x_k)^{\frac{2}{\beta}} + \sum_{j\in J_k} g_j^+(Ax_k)^{\frac{2}{\beta}} \right). \tag{3.5}$$

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

Proof. Assertion (ii) directly follows from the assertion (i) of this lemma and (3.3). Hence it is sufficient to prove assertion (i). To this end, fix $k \in \mathbb{N}$. It follows from (3.4) and the nonexpansive property of projection operator that

$$\begin{aligned}
&\|x_{k+1}-x\|^2 \\
\leq\, & \|x_k - \lambda_k L^{-\frac{1}{\beta}} \sum_{i\in I_k} \mu_{k,i} f_i^+(x_k)^{\frac{1}{\beta}} \phi_{k,i} - \lambda_k L^{-\frac{1}{\beta}} \sum_{j\in J_k} \nu_{k,j} g_j^+(Ax_k)^{\frac{1}{\beta}} A^\top \psi_{k,j} - x\|^2 \\
=\, & \|x_k - x\|^2 + \lambda_k^2 L^{-\frac{2}{\beta}} \| \sum_{i\in I_k} \mu_{k,i} f_i^+(x_k)^{\frac{1}{\beta}} \phi_{k,i} + \sum_{j\in J_k} \nu_{k,j} g_j^+(Ax_k)^{\frac{1}{\beta}} A^\top \psi_{k,j}\|^2 \\
& - 2\lambda_k L^{-\frac{1}{\beta}} \sum_{i\in I_k} \mu_{k,i} f_i^+(x_k)^{\frac{1}{\beta}} \langle x_k - x, \phi_{k,i}\rangle - 2\lambda_k L^{-\frac{1}{\beta}} \sum_{j\in J_k} \nu_{k,j} g_j^+(Ax_k)^{\frac{1}{\beta}} \langle Ax_k - Ax, \psi_{k,j}\rangle .
\end{aligned} \tag{3.6}$$

Note by $x \in S$ that $f_i(x) \leq 0$ for each $i \in I$. Then, for each $i \in I_k$ with $f_i(x_k) > 0$, by Assumption 3.1, Lemma 2.1 is applicable to showing that $\langle x_k - x, \phi_{k,i}\rangle \geq \left(\frac{f_i^+(x_k)}{L}\right)^{\frac{1}{\beta}}$; otherwise, $f_i^+(x_k) = 0$. Hence we conclude that

$$\sum_{i\in I_k} \mu_{k,i} f_i^+(x_k)^{\frac{1}{\beta}} \langle x_k - x, \phi_{k,i}\rangle \geq L^{-\frac{1}{\beta}} \sum_{i\in I_k} \mu_{k,i} f_i^+(x_k)^{\frac{2}{\beta}} . \tag{3.7}$$

Using the similar arguments we did for (3.7), we can obtain that

$$\sum_{j\in J_k} \nu_{k,j} g_j^+(Ax_k)^{\frac{1}{\beta}} \langle Ax_k - Ax, \psi_{k,j}\rangle \geq L^{-\frac{1}{\beta}} \sum_{j\in J_k} \nu_{k,j} g_j^+(Ax_k)^{\frac{2}{\beta}} . \tag{3.8}$$

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

$$\begin{aligned}
&\|\sum_{i\in I_k}\mu_{k,i}f_i^+(x_k)^{\frac{1}{\beta}}\phi_{k,i}+\sum_{j\in J_k}\nu_{k,j}g_j^+(Ax_k)^{\frac{1}{\beta}}A^\top\psi_{k,j}\|^2\\
=&4\|\sum_{i\in I_k}\frac{\mu_{k,i}}{2}f_i^+(x_k)^{\frac{1}{\beta}}\phi_{k,i}+\sum_{j\in J_k}\frac{\nu_{k,j}}{2}g_j^+(Ax_k)^{\frac{1}{\beta}}A^\top\psi_{k,j}\|^2\\
\le&2\sum_{i\in I_k}\mu_{k,i}f_i^+(x_k)^{\frac{2}{\beta}}+2\|A\|^2\sum_{j\in J_k}\nu_{k,j}g_j^+(Ax_k)^{\frac{2}{\beta}}.
\end{aligned}\tag{3.9}$$

(thanks to $\{\mu_{k,i}\}_{i\in I_k}\in\Delta_+^{|I_k|}$ and $\{\nu_{k,j}\}_{j\in J_k}\in\Delta_+^{|J_k|}$, $\phi_{k,i}\in\mathbb{S}$ and $\psi_{k,j}\in\mathbb{S}$ for each $i\in I_k$ and $j\in J_k$). Combining this with (3.7) and (3.8), (3.6) is reduced to

$$\begin{aligned}
\|x_{k+1}-x\|^2\le\|x_k-x\|^2&-2\lambda_k(1-\lambda_k)\sum_{i\in I_k}\mu_{k,i}\left(\frac{f_i^+(x_k)}{L}\right)^{\frac{2}{\beta}}\\
&-2\lambda_k(1-\|A\|^2\lambda_k)\sum_{j\in J_k}\nu_{k,j}\left(\frac{g_j^+(Ax_k)}{L}\right)^{\frac{2}{\beta}}.
\end{aligned}$$

This, together with (3.2)-(3.3) and Assumption 3.2, shows (3.5). The proof is complete. □

The control scheme of index sets $\{I_k\}$ and $\{J_k\}$ plays a central role in determining the active indices to be executed, and is crucial in guaranteeing the convergence property and numerical performance of subgradient methods for solving the feasibility problems. The most two extensively used control schemes are the most violated constraint control and the almost cyclic control; see, e.g., [16,24,32]. In this paper, we will consider several general control schemes as discussed in [16]. In the following definitions of the general control schemes, item (a) is an extension of the most violated constraint control and the parallel control, item (b) 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 [37–39].

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

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

(b) the s-intermittent control if

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

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

4. Quantitative Convergence Theory

In this section, we assume that Assumptions 3.1 and 3.2 are always satisfied, and establish the quantitative convergence theory, including the global convergence, 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. To explore the convergence property, the violations of the MSSQFP (3.1) are usually measured by

$$F^+(x) := \max_{i\in I} f_i^+(x) \quad \text{and} \quad G^+(Ax) := \max_{j\in J} g_j^+(Ax) \quad \text{for each } x\in\mathbb{R}^n. \tag{4.1}$$

It is clear that

$$x\in S \quad \Leftrightarrow \quad x\in X,\ F^+(x)=0,\ G^+(Ax)=0.$$

4.1. **α-most violated constraints control.** To start convergence analysis, the following lemma provides the basic inequality for Algorithm 3.1 with the α-most violated constraints control, which directly follows from Lemma 3.1 (taking $x := \mathrm{P}_S(x_k)$) and Definition 3.1(a).

Lemma 4.1. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1 with $\{I_k\}$ and $\{J_k\}$ being the α-most violated constraints controls. Then it holds for each $k\in\mathbb{N}$ that*

$$\mathrm{d}_S^2(x_{k+1}) \le \mathrm{d}_S^2(x_k) - 2\underline{\lambda}(1-\theta\overline{\lambda})\eta\left(\frac{\alpha}{L}\right)^{\frac{2}{\beta}}\left(F^+(x_k)^{\frac{2}{\beta}} + G^+(Ax_k)^{\frac{2}{\beta}}\right). \tag{4.2}$$

By virtue of Lemma 4.1, this section aims to investigate the global convergence, the iteration complexity, and the convergence rate for Algorithm 3.1 with the α-most violated constraints control.

4.1.1. *Global convergence.* We first establish the global convergence for the adaptive subgradient method with the α-most violated constraints control.

Theorem 4.1. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ and $\{J_k\}$ being the α-most violated constraints controls. Then $\{x_k\}$ converges to a feasible solution of the MSSQFP (3.1).*

Proof. Note by Lemma 3.1(ii) that $\{x_k\}$ is bounded, and thus has a cluster point, denoted by $\bar{x}$. It follows from the (4.2) in Lemma 4.1 that

$$\sum_{k=1}^{\infty}\left(F^+(x_k)^{\frac{2}{\beta}} + G^+(Ax_k)^{\frac{2}{\beta}}\right) \le \frac{1}{2\underline{\lambda}(1-\theta\overline{\lambda})\eta}\left(\frac{L}{\alpha}\right)^{\frac{2}{\beta}}\mathrm{d}_S^2(x_1) < \infty.$$

Consequently, $\lim_{k\to\infty} F^+(x_k) = 0$ and $\lim_{k\to\infty} G^+(Ax_k) = 0$, which implies $\bar{x}\in S$ by the continuity of $\{f_i\}_{i\in I}$ and $\{g_j\}_{j\in J}$. Recall from Lemma 3.1(ii) that $\{\|x_k-\bar{x}\|\}$ is monotonically decreasing. Hence $\{x_k\}$ converges to this $\bar{x}\in S$, and the proof is complete. □

4.1.2. *Iteration complexity.* Given $\delta>0$, 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 δ-feasible solution, that is,

$$\min_{1\le k\le K(\delta)} \max\{F^+(x_k), G^+(Ax_k)\} \le \delta.$$

As usual, we use $\lceil t\rceil$ to denote the smallest integer larger than t.

Theorem 4.2. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ and $\{J_k\}$ being the α-most violated constraints controls. Let $\delta>0$ and $K_{\mathrm{m}} := \lceil \frac{\mathrm{d}_S^2(x_1)}{2\underline{\lambda}(1-\theta\overline{\lambda})\eta}\left(\frac{L}{\alpha\delta}\right)^{\frac{2}{\beta}}\rceil$. Then*

$$\min_{1\le k\le K_{\mathrm{m}}} \max\{F^+(x_k), G^+(Ax_k)\} \le \delta.$$

Proof. Proving by contradiction, we assume that $\max\{F^+(x_k), G^+(Ax_k)\} > \delta$, consequently $F^+(x_k)^{\frac{2}{\beta}} + G^+(Ax_k)^{\frac{2}{\beta}} > \delta^{\frac{2}{\beta}}$, for each $1 \le k \le K_{\rm m}$. Then it follows from (4.2) that

$$\mathrm{d}_S^2(x_{k+1}) < \mathrm{d}_S^2(x_k) - 2\underline{\lambda}(1-\theta\overline{\lambda})\eta\left(\frac{\alpha\delta}{L}\right)^{\frac{2}{\beta}}.$$

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

$$0 \le \mathrm{d}_S^2(x_{K_{\rm m}+1}) < \mathrm{d}_S^2(x_1) - 2K_{\rm m}\underline{\lambda}(1-\theta\overline{\lambda})\eta\left(\frac{\alpha\delta}{L}\right)^{\frac{2}{\beta}},$$

which contradicts the definition of $K_{\rm m}$. The proof is complete. □

Remark 4.1. Theorem 4.2 shows that Algorithm 3.1 with the α-most violated constraints control possesses a worst-case iteration complexity of $\mathscr{O}(1/k^{\frac{\beta}{2}})$ to a feasible solution. In particular, as mentioned above, the α-most violated constraints control covers the most violated constraint control when $\alpha = \eta = 1$; also see [16, Remark 3.6]. Hence, as an application of Theorem 4.2, the iteration complexity for the most violated constraint control is

$$K_{\rm m} := \left\lceil \frac{\mathrm{d}_S^2(x_1)}{2\underline{\lambda}(1-\theta\overline{\lambda})}\left(\frac{L}{\delta}\right)^{\frac{2}{\beta}}\right\rceil.$$

4.1.3. *Convergence rate analysis.* The establishment of convergence rate is significant in guaranteeing the numerical performance of relevant algorithms. The error bound property plays an important role in convergence rate analysis of numerical algorithms. The notion of the (Lipschitz-type) error bound property was introduced by [40,41] for the linear inequality system and convex inequality system respectively, and has been extensively used in convergence rate analysis of various numerical algorithms; see [6,42,43] and references therein. As a natural extension, the Hölder-type error bound property was introduced for polynomial systems [44] and has been widely explored and applied in [45,46] and references therein. In particular, Suzuki and Kuroiwa [47] investigated the Hölder-type error bound property for the quasi-convex inequality system, and Hu et al. [16] applied the Hölder-type error bound property to establish the linear/sublinear convergence rates of subgradient methods for the quasi-convex feasibility problem. Below we introduce the Hölder-type bounded error bound for the MSSQFP (3.1).

Definition 4.1. The inequality system (3.1) is said to satisfy the Hölder-type bounded error bound property of order $q > 0$ if, for any $\gamma > 0$ such that $S \cap \mathbb{B}(0,\gamma) \neq \emptyset$, there exists $\kappa(\gamma) > 0$ such that

$$\mathrm{d}_S^q(x) \le \kappa(\gamma) \max_{i\in I,\, j\in J}\left\{f_i^+(x), g_j^+(Ax)\right\} \quad \text{for each } x \in X \cap \mathbb{B}(0,\gamma). \tag{4.3}$$

In particular, inequality system (3.1) is said to satisfy the (Lipschitz-type) bounded error bound property if (4.3) holds with $q = 1$.

Following a line of analysis similar to [16, Proposition 3.1], we can provide in the following proposition a sufficient condition for the Hölder-type bounded error bound property of the MSSQFP (3.1) in terms of the Hölder-type bounded error bound property of each component function and the Slater condition of their sublevel sets.

Proposition 4.1. *Suppose that each inequality system $\{f_i(x) \le 0\}$ and $\{g_j(Ax) \le 0\}$ satisfies the Hölder-type bounded error bound property of order q and the Slater condition is satisfied:*

$$\{x \in X : f_i(x) < 0, \forall i \in I;\ g_i(Ax) < 0, \forall j \in J\} \neq \varnothing.$$

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

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

Theorem 4.3. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ and $\{J_k\}$ being the α-most violated constraints controls. Suppose that (3.1) satisfies the Hölder-type bounded error bound property of order $q > 0$. 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*

$$\|x_k - \bar{x}\| \le c\tau^k \quad \text{for each } k \in \mathbb{N}. \tag{4.4}$$

(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*

$$\|x_k - \bar{x}\| \le ck^{-\frac{\beta}{2(q-\beta)}} \quad \text{for each } k \in \mathbb{N}. \tag{4.5}$$

Proof. Noting by Lemma 3.1(ii) that $\{x_k\}$ is bounded, we can assume that there exists $\gamma > 0$ such that $S \cap \mathbb{B}(0,\gamma) \neq \emptyset$ and $\{x_k\} \subseteq \mathbb{B}(0,\gamma)$. By the assumption that the MSSQFP (3.1) satisfies the Hölder-type bounded error bound property of order $q > 0$, there exists $\kappa > 0$ (depending on this γ) such that (4.3) holds. Since $\{x_k\} \subseteq \mathbb{B}(0,\gamma)$, one has by (4.3) and (4.1) that

$$\mathrm{d}_S^q(x_k) \le \kappa \max\left\{F^+(x_k), G^+(Ax_k)\right\} \quad \text{for each } k \in \mathbb{N}.$$

Noting that $F^+(x_k)^{\frac{2}{\beta}} + G^+(Ax_k)^{\frac{2}{\beta}} \ge \max\{F^+(x_k), G^+(Ax_k)\}^{\frac{2}{\beta}}$ and $\beta \le 1$, it follows that

$$\mathrm{d}_S^{\frac{2q}{\beta}}(x_k) \le \kappa^{\frac{2}{\beta}} \left(F^+(x_k)^{\frac{2}{\beta}} + G^+(Ax_k)^{\frac{2}{\beta}}\right) \quad \text{for each } k \in \mathbb{N}. \tag{4.6}$$

This, together with (4.2), shows

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

with $\rho := 2\underline{\lambda}(1-\theta\overline{\lambda})\eta\left(\frac{\alpha}{\kappa L}\right)^{\frac{2}{\beta}}$. Consequently, there exists $c \ge 0$ such that

$$\mathrm{d}_S(x_k) \le \begin{cases} c\tau^k, & \text{if } q = \beta, \\ ck^{-\frac{\beta}{2(q-\beta)}}, & \text{if } q > \beta, \end{cases} \tag{4.7}$$

with $\tau := \sqrt{1-\rho}$ and by applying Lemma 2.3 (with $\mathrm{d}_S^2(x_k)$, ρ, $\frac{q}{\beta} - 1$ in place of u_k, a, σ), for each $k \in \mathbb{N}$. Fix $l > k \in \mathbb{N}$. It follows from Lemma 3.1(ii) (taking $x := \mathrm{P}_S(x_k)$) that

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

Hence, by the convergence of $\{x_l\}$ to $\bar{x} \in S$ as shown in Theorem 4.1, we obtain that

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

This, together with (4.7), implies (4.4) and (4.5). The proof is complete. □

4.2. ***s*-intermittent control.** This subsection aims to explore the global convergence, the iteration complexity and the convergence rate for the adaptive subgradient method with the s-intermittent control. To proceed, we first deduce the basic inequality for Algorithm 3.1 with the s-intermittent control by virtue of Lemma 3.1.

Lemma 4.2. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1 with $\{I_k\}$ and $\{J_k\}$ being the s-intermittent controls. Then it holds for each $k \in \mathbb{N}$ that*

$$\mathrm{d}_S^2(x_{s(k+1)}) \le \mathrm{d}_S^2(x_{sk}) - \frac{4\underline{\lambda}(1-\theta\overline{\lambda})\eta}{1+4s}(2L)^{-\frac{2}{\beta}}\left(F^+(x_{sk})^{\frac{2}{\beta}} + G^+(Ax_{sk})^{\frac{2}{\beta}}\right). \tag{4.8}$$

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

$$\|x_{s(k+1)} - x\|^2 \le \|x_{sk} - x\|^2 - 2\underline{\lambda}(1-\theta\overline{\lambda})\eta L^{-\frac{2}{\beta}}\sum_{t=0}^{s-1}\left(\sum_{i\in I_{sk+t}} f_i^+(x_{sk+t})^{\frac{2}{\beta}} + \sum_{j\in J_{sk+t}} g_j^+(Ax_{sk+t})^{\frac{2}{\beta}}\right). \tag{4.9}$$

Below we estimate the second term on the right hand side of (4.9) in terms of $F^+(x_{sk})$ and $G^+(Ax_{sk})$, respectively. Firstly, let $i_k \in I$ be the most violated index of the inequality system $\{f_i(x) \le 0\}_{i\in I}$ at x_{sk}, that is,

$$f_{i_k}^+(x_{sk}) = F^+(x_{sk}). \tag{4.10}$$

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

$$\begin{aligned} f_{i_k}^+(x_{sk})^{\frac{2}{\beta}} &\le \left(f_{i_k}^+(x_{sk+t_k}) + L\|x_{sk+t_k} - x_{sk}\|^{\beta}\right)^{\frac{2}{\beta}} \\ &\le 2^{\frac{2}{\beta}-1}\left(f_{i_k}^+(x_{sk+t_k})^{\frac{2}{\beta}} + L^{\frac{2}{\beta}}\|x_{sk+t_k} - x_{sk}\|^2\right) \end{aligned} \tag{4.11}$$

(by Lemma 2.2(ii) as $\beta \le 1$). Since $i_k \in I_{sk+t_k}$ and $t_k \in [0, s-1]$, we have

$$f_{i_k}^+(x_{sk+t_k})^{\frac{2}{\beta}} \le \sum_{i\in I_{sk+t_k}} f_i^+(x_{sk+t_k})^{\frac{2}{\beta}} \le \sum_{t=0}^{s-1}\sum_{i\in I_{sk+t}} f_i^+(x_{sk+t})^{\frac{2}{\beta}}. \tag{4.12}$$

On the other side, in view of Algorithm 3.1, we obtain by (3.4) and (3.9) that

$$\begin{aligned} \|x_{k+1} - x_k\|^2 &\le \lambda_k^2 L^{-\frac{2}{\beta}}\|\sum_{i\in I_k}\mu_{k,i} f_i^+(x_k)^{\frac{1}{\beta}}\phi_{k,i} + \sum_{j\in J_k}\nu_{k,j} g_j^+(Ax_k)^{\frac{1}{\beta}}A^\top\psi_{k,j}\|^2 \\ &\le 2\overline{\lambda}^2 L^{-\frac{2}{\beta}}\sum_{i\in I_k} f_i^+(x_k)^{\frac{2}{\beta}} + 2\overline{\lambda}^2\|A\|^2 L^{-\frac{2}{\beta}}\sum_{j\in J_k} g_j^+(Ax_k)^{\frac{2}{\beta}}. \end{aligned}$$

(thanks to (3.3) and $\{\mu_{k,i}\}_{i\in I_k} \in \Delta_+^{|I_k|}$ and $\{\nu_{k,j}\}_{j\in J_k} \in \Delta_+^{|J_k|}$) for each $k \in \mathbb{N}$. Then we derive by Lemma 2.2(ii) that

$$\begin{aligned} \|x_{sk+t_k} - x_{sk}\|^2 &\le s\sum_{t=0}^{s-1}\|x_{sk+t+1} - x_{sk+t}\|^2 \\ &\le 2s\overline{\lambda}^2 L^{-\frac{2}{\beta}}\sum_{t=0}^{s-1}\sum_{i\in I_{sk+t}} f_i^+(x_{sk+t})^{\frac{2}{\beta}} + 2s\overline{\lambda}^2\|A\|^2 L^{-\frac{2}{\beta}}\sum_{t=0}^{s-1}\sum_{j\in J_{sk+t}} g_j^+(Ax_{sk+t})^{\frac{2}{\beta}}. \end{aligned}$$

This, together with (4.10)-(4.12), deduces that

$$F^+(x_{sk})^{\frac{2}{\beta}} \le 2^{\frac{2}{\beta}-1}\left(1+2s\overline{\lambda}^2\right)\sum_{t=0}^{s-1}\sum_{i\in I_{sk+t}} f_i^+(x_{sk+t})^{\frac{2}{\beta}} + 2^{\frac{2}{\beta}} s\overline{\lambda}^2\|A\|^2\sum_{t=0}^{s-1}\sum_{j\in J_{sk+t}} g_j^+(Ax_{sk+t})^{\frac{2}{\beta}}$$

$$\le 2^{\frac{2}{\beta}-1}(1+2s)\sum_{t=0}^{s-1}\sum_{i\in I_{sk+t}} f_i^+(x_{sk+t})^{\frac{2}{\beta}} + 2^{\frac{2}{\beta}} s\sum_{t=0}^{s-1}\sum_{j\in J_{sk+t}} g_j^+(Ax_{sk+t})^{\frac{2}{\beta}}$$

(due to (3.2) and (3.3)). Using the similar arguments we did for the above inequality, we can obtain that

$$G^+(Ax_{sk})^{\frac{2}{\beta}} \le 2^{\frac{2}{\beta}-1}(1+2s)\sum_{t=0}^{s-1}\sum_{j\in J_{sk+t}} g_j^+(Ax_{sk+t})^{\frac{2}{\beta}} + 2^{\frac{2}{\beta}} s\sum_{t=0}^{s-1}\sum_{i\in I_{sk+t}} f_i^+(x_{sk+t})^{\frac{2}{\beta}}.$$

These, together with (4.9) (taking $x := \mathrm{P}_S(x_{sk})$), implies (4.8). The proof is complete. □

4.2.1. *Global convergence.*

Theorem 4.4. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ and $\{J_k\}$ being the s-intermittent controls. Then $\{x_k\}$ converges to a feasible solution of the MSSQFP (3.1).*

Proof. Note by Lemma 3.1(ii) that $\{x_{sk}\}$ is bounded, and thus has a cluster point, denoted by $\bar{x}$. It follows from (4.8) in Lemma 4.2 that

$$\sum_{k=1}^{\infty}\left(F^+(x_{sk})^{\frac{2}{\beta}} + G^+(Ax_{sk})^{\frac{2}{\beta}}\right) \le \frac{1+4s}{4\underline{\lambda}(1-\theta\overline{\lambda})\eta}(2L)^{\frac{2}{\beta}}\mathrm{d}_S^2(x_s) < \infty.$$

This shows that $\lim_{k\to\infty} F^+(x_{sk}) = 0$ and $\lim_{k\to\infty} G^+(Ax_{sk}) = 0$, and thus $\bar{x}\in S$ by the continuity of $\{f_i\}_{i\in I}$ and $\{g_j\}_{j\in J}$. This, together with Lemma 3.1(ii), shows that $\{x_k\}$ converges to $\bar{x}\in S$. The proof is complete. □

4.2.2. *Iteration complexity.*

Theorem 4.5. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ and $\{J_k\}$ being the s-intermittent controls. Let $\delta > 0$ and $K_c := \lceil \frac{s(1+4s)\mathrm{d}_S^2(x_1)}{4\underline{\lambda}(1-\theta\overline{\lambda})\eta}\left(\frac{2L}{\delta}\right)^{\frac{2}{\beta}}\rceil$. Then*

$$\min_{1\le k\le K_c}\max\{F^+(x_k), G^+(Ax_k)\} \le \delta.$$

Proof. Proving by contradiction, we assume that $\max\{F^+(x_k), G^+(Ax_k)\} > \delta$, consequently $F^+(x_k)^{\frac{2}{\beta}} + G^+(Ax_k)^{\frac{2}{\beta}} > \delta^{\frac{2}{\beta}}$, for each $1\le k\le K_c$. Then it follows from (4.8) that

$$\mathrm{d}_S^2(x_{s(k+1)}) < \mathrm{d}_S^2(x_{sk}) - \frac{4\underline{\lambda}(1-\theta\overline{\lambda})\eta}{1+4s}\left(\frac{\delta}{2L}\right)^{\frac{2}{\beta}}.$$

Summing the above inequality over $k = 1,\ldots,\frac{K_c}{s}$, we derive that

$$0 \le \mathrm{d}_S^2(x_{K_c+s}) < \mathrm{d}_S^2(x_s) - \frac{K_c}{s}\frac{4\underline{\lambda}(1-\theta\overline{\lambda})\eta}{1+4s}\left(\frac{\delta}{2L}\right)^{\frac{2}{\beta}},$$

which contradicts with the definition of K_c. The proof is complete. □

Remark 4.2. Theorem 4.5 shows that Algorithm 3.1 with the s-intermittent control has a worst-case iteration complexity of $\mathscr{O}(1/k^{\frac{\beta}{2}})$ to a feasible solution. Particularly, as mentioned above, the s-intermittent control coves the almost cyclic control when $s = \max\{M,N\}$ and $\eta = 1$; also see [16, Remark 3.6]. Hence, as a direct application of Theorem 4.5, the iteration complexity for the almost cyclic control is

$$K_{\rm c} := \left\lceil \frac{\max\{M,N\}(1+4\max\{M,N\})\mathrm{d}_S^2(x_1)}{4\underline{\lambda}(1-\theta\overline{\lambda})}\left(\frac{2L}{\delta}\right)^{\frac{2}{\beta}}\right\rceil.$$

In contrast to Remark 4.1, we observe that the almost cyclic control requires a much larger number of iterations than the most violated constraint control:

$$\frac{K_{\rm c}}{K_{\rm m}} = \max\{M,N\}(1+4\max\{M,N\})2^{\frac{2}{\beta}-1} \gg 1.$$

This shows a benefit of the most violated constraints control over the almost cyclic control. Nevertheless, 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 constraint control and the parallel control need to find the most violated index through all component functions or calculate the subgradients of all component functions, respectively.

4.2.3. *Convergence rate analysis.*

Theorem 4.6. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ and $\{J_k\}$ being the s-intermittent controls. Suppose that (3.1) satisfies the Hölder-type bounded error bound property of order $q > 0$. 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.*
(ii) *If $q > \beta$, then there exist $\bar{x} \in S$ and $c \geq 0$ such that*

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

Proof. Similar to the beginning of proof of Theorem 4.3, there exists $\kappa > 0$ such that (4.6) is satisfied. By (4.6) and (4.8), we obtain that

$$\mathrm{d}_S^2(x_{s(k+1)}) \leq \mathrm{d}_S^2(x_{sk}) - \rho\, \mathrm{d}_S^{\frac{2q}{\beta}}(x_{sk}) \quad \text{for each } k \in \mathbb{N},$$

where $\rho := \frac{4\underline{\lambda}(1-\theta\overline{\lambda})\eta}{1+4s}\left(\frac{1}{2\kappa L}\right)^{\frac{2}{\beta}}$. Consequently, there exists $c \geq 0$ such that

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

with $\tau := \sqrt{1-\rho}$ and by applying Lemma 2.3, for each $k \in \mathbb{N}$. Similar to the end of proof of Theorem 4.3, we can obtain the conclusions, and the proof is complete. □

4.3. **Stochastic control.** Deterministic control schemes always suffer from certain drawbacks. Specifically, the most violated constraints control and the parallel control consume expensive computational cost to find the most violated index through all component functions or calculate the subgradients of all component functions 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 types of control schemes; see, e.g., [16].

The idea of the stochastic index scheme is increasingly popular in optimization algorithms and applications; e.g., the first-order algorithms with random projections in large-scale network optimization problems [39], and the incremental subgradient methods with random component selections in distributed optimization problems [18, 25]. A typical example is the stochastic gradient descent algorithms in machine learning [37], 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, this subsection aims to consider the stochastic control in the adaptive subgradient method for solving the MSSQFP (3.1) and investigate its quantitative convergence theory. An interesting finding is disclosed by Theorem 4.9 that the stochastic control has a significant favorable effect on the performance of the adaptive subgradient method; concretely, the stochastic control enjoys both advantages of the low computational cost requirement and the low iteration complexity; see Remark 4.3 for details.

To proceed the convergence analysis of the adaptive subgradient method with the stochastic control, we provide below a basic inequality in terms of conditional expectation.

Lemma 4.3. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ and $\{J_k\}$ being the stochastic controls, and let $\mathscr{F}_k := \{x_1, \ldots, x_k\}$ for each $k \in \mathbb{N}$. Then it holds for each $x \in S$ and $k \in \mathbb{N}$ that*

$$\mathbb{E}\left\{\mathrm{d}_S^2(x_{k+1}) \mid \mathscr{F}_k\right\} \le \mathrm{d}_S^2(x_k) - \frac{2\underline{\lambda}(1-\theta\overline{\lambda})}{\max\{M,N\}} L^{-\frac{2}{\beta}} \left(F^+(x_k)^{\frac{2}{\beta}} + G^+(Ax_k)^{\frac{2}{\beta}}\right). \tag{4.13}$$

Proof. Fix $x \in S$ and $k \in \mathbb{N}$. Since $I_k = \{\omega_k\}$, $J_k = \{\pi_k\}$ and $\eta = 1$ in the stochastic control (cf. Definition 3.1(c)), it follows from Lemma 3.1 (take $x := \mathrm{P}_S(x_k)$) that

$$\mathrm{d}_S^2(x_{k+1}) \le \mathrm{d}_S^2(x_k) - 2\underline{\lambda}(1-\theta\overline{\lambda}) L^{-\frac{2}{\beta}} \left(f_{\omega_k}^+(x_k)^{\frac{2}{\beta}} + g_{\pi_k}^+(Ax_k)^{\frac{2}{\beta}}\right).$$

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

$$\mathbb{E}\left\{\mathrm{d}_S^2(x_{k+1}) \mid \mathscr{F}_k\right\} \le \mathrm{d}_S^2(x_k) - 2\underline{\lambda}(1-\theta\overline{\lambda}) L^{-\frac{2}{\beta}} \left(\mathbb{E}\left\{f_{\omega_k}^+(x_k)^{\frac{2}{\beta}} \mid \mathscr{F}_k\right\} + \mathbb{E}\left\{g_{\pi_k}^+(Ax_k)^{\frac{2}{\beta}} \mid \mathscr{F}_k\right\}\right). \tag{4.14}$$

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

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

Similarly, we have

$$\mathbb{E}\left\{g_{\pi_k}^+(Ax_k)^{\frac{2}{\beta}} \mid \mathscr{F}_k\right\} = \frac{1}{N}\sum_{j\in J} g_j^+(Ax_k)^{\frac{2}{\beta}} \ge \frac{1}{N}G^+(Ax_k)^{\frac{2}{\beta}}.$$

These, together with (4.14), imply (4.13). The proof is complete. □

The supermartingale convergence theorem is taken from [48, p. 148], which is useful in convergence analysis of the subgradient method with the stochastic control.

Theorem 4.7. *Let $\{Y_k\}$, $\{Z_k\}$, and $\{W_k\}$ be three sequences of random variables, and let $\{\mathscr{F}_k\}$ be a sequence of sets of random variables such that $\mathscr{F}_k \subseteq \mathscr{F}_{k+1}$ for each $k \in \mathbb{N}$. Suppose for any $k \in \mathbb{N}$ that*

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

4.3.1. *Global convergence.*

Theorem 4.8. *Let* $\{x_k\}$ *be generated by Algorithm 3.1 with* $\{I_k\}$ *and* $\{J_k\}$ *being the stochastic controls. Then* $\{x_k\}$ *converges to a feasible solution of the MSSQFP (3.1) with probability 1.*

Proof. Note by Lemma 3.1(ii) that $\{x_k\}$ is bounded, and thus has a cluster point, denoted by $\bar{x}$. By Lemma 4.3, Theorem 4.7 is applicable to showing that $\sum_{k=1}^{\infty} F^+(x_k)^{\frac{2}{\beta}} + G^+(Ax_k)^{\frac{2}{\beta}} < \infty$ with probability 1. Hence $\lim_{k\to\infty} F^+(x_k)^{\frac{2}{\beta}} = 0$ and $\lim_{k\to\infty} G^+(Ax_k)^{\frac{2}{\beta}} = 0$, and consequently $\bar{x} \in S$ (by the continuity of $\{f_i\}_{i\in I}$ and $\{g_j\}_{j\in J}$), with probability 1. This, together with Lemma 3.1(ii), shows that $\{x_k\}$ converges to this $x \in S$ with probability 1. The proof is complete. □

4.3.2. *Iteration complexity.*

Theorem 4.9. *Let* $\{x_k\}$ *be generated by Algorithm 3.1 with* $\{I_k\}$ *and* $\{J_k\}$ *being the stochastic controls. Let* $\delta > 0$ *and* $K_{\rm s} := \frac{\max\{M,N\}\mathrm{d}_S^2(x_1)}{2\underline{\lambda}(1-\theta\overline{\lambda})}\left(\frac{L}{\delta}\right)^{\frac{2}{\beta}}$*. Then*

$$\min_{1\le k\le K_{\rm s}} \max\{\mathbb{E}\{F^+(x_k)\},\mathbb{E}\{G^+(Ax_k)\}\} \le \delta.$$

Proof. Proving by contradiction, we assume that $\max\{\mathbb{E}\{F^+(x_k)\},\mathbb{E}\{G^+(Ax_k)\}\} > \delta$ for each $1 \le k \le K_{\rm s}$. Consequently, $\max\left\{\mathbb{E}\{F^+(x_k)^{\frac{2}{\beta}}\},\mathbb{E}\{G^+(Ax_k)^{\frac{2}{\beta}}\}\right\} > \delta^{\frac{2}{\beta}}$ by the convexity of $t^{\frac{2}{\beta}}$ on $\mathbb{R}_+$ (as $\beta \le 1$), and thus $\mathbb{E}\left\{F^+(x_k)^{\frac{2}{\beta}} + G^+(Ax_k)^{\frac{2}{\beta}}\right\} > \delta^{\frac{2}{\beta}}$. Taking the expectation on (4.13), one has for each $1 \le k \le K_{\rm s}$ that

$$\begin{aligned}\mathbb{E}\left\{\mathrm{d}_S^2(x_{k+1})\right\} &\le \mathbb{E}\left\{\mathrm{d}_S^2(x_k)\right\} - \frac{2\underline{\lambda}(1-\theta\overline{\lambda})}{\max\{M,N\}}L^{-\frac{2}{\beta}}\mathbb{E}\left\{F^+(x_k)^{\frac{2}{\beta}} + G^+(Ax_k)^{\frac{2}{\beta}}\right\}\\ &< \mathbb{E}\left\{\mathrm{d}_S^2(x_k)\right\} - \frac{2\underline{\lambda}(1-\theta\overline{\lambda})}{\max\{M,N\}}\left(\frac{\delta}{L}\right)^{\frac{2}{\beta}}.\end{aligned}$$

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

$$0 \le \mathbb{E}\left\{\mathrm{d}_S^2(x_{k+1})\right\} < \mathrm{d}_S^2(x_1) - K_{\rm s}\frac{2\underline{\lambda}(1-\theta\overline{\lambda})}{\max\{M,N\}}\left(\frac{\delta}{L}\right)^{\frac{2}{\beta}},$$

which contradicts with the definition of $K_{\rm s}$. The proof is complete. □

Remark 4.3. Theorem 4.9 provides a theoretical evidence for the benefit of the stochastic control in the sense of the worst-case iteration complexity. In particular, the stochastic control not only enjoys the significant advantage of the low computational cost requirement as the almost cyclic control (much less than the most violated control and the parallel control), but also owns a much lower iteration complexity than the almost cyclic control. Indeed, by Theorem 4.9 and Remark 4.2, we derive

$$\frac{K_{\rm s}}{K_{\rm c}} = \frac{2^{1-\frac{2}{\beta}}}{(1+4\max\{M,N\})} \ll 1.$$

4.3.3. *Convergence rate analysis.*

Theorem 4.10. *Let $\{x_k\}$ be generated by Algorithm 3.1 with $\{I_k\}$ and $\{J_k\}$ being the stochastic controls. Suppose that (3.1) satisfies the Hölder-type bounded error bound property of order $q > 0$. Then the following assertions are true.*

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

$$\mathbb{E}\{\mathrm{d}_S(x_k)\} \leq c\tau^k \quad \textit{for each } k \in \mathbb{N}. \tag{4.15}$$

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

$$\mathbb{E}\{\mathrm{d}_S(x_k)\} \leq ck^{-\frac{\beta}{2(q-\beta)}} \quad \textit{for each } k \in \mathbb{N}. \tag{4.16}$$

Proof. Similar to the beginning of proof of Theorem 4.3, there exists $\kappa > 0$ such that (4.6) is satisfied. By (4.13) and (4.6), we obtain that

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

where $\rho := \frac{2\underline{\lambda}(1-\theta\overline{\lambda})}{\max\{M,N\}}\left(\frac{1}{\kappa L}\right)^{\frac{2}{\beta}}$. 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

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

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

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

with $\tau := \sqrt{1-\rho}$ and by applying Lemma 2.3, for each $k \in \mathbb{N}$. These, together with that $(\mathbb{E}\{\mathrm{d}_S(x_k)\})^2 \leq \mathbb{E}\left\{\mathrm{d}_S^2(x_k)\right\}$, imply (4.15) and (4.16), respectively. The proof is complete. □

Acknowledgments

The authors are grateful to the editor and the anonymous reviewers for their valuable comments and suggestions toward the improvement of this paper. The first author was supported by the National Natural Science Foundation of China (12071306, 32170655, 11871347), Natural Science Foundation of Guangdong Province (2019A1515011917, 2020B1515310008), Project of Educational Commission of Guangdong Province (2021KTSCX103, 2019KZDZX1007), Natural Science Foundation of Shenzhen (JCYJ20190808173603590) and Interdisciplinary Innovation Team of Shenzhen University. The second author was supported by the Science Foundation of Zhejiang Sci-Tech University (19062150-Y). The third author was supported by the Foundation for High-level Talents of Chongqing University of Art and Sciences (P2017SC01), Chongqing Key Laboratory of Group and Graph Theories and Applications, and Key Laboratory of Complex Data Analysis and Artificial Intelligence of Chongqing Municipal Science and Technology Commission. The forth author was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/FDS14/P03/20).

REFERENCES

[1] Y. Censor, T. Elfving, A multiprojection algorithm using Bregman projections in a product space, Numer. Algo. 8 (1994), 221-239.
[2] C. Byrne, A unified treatment of some iterative algorithms in signal processing and image reconstruction, Inverse Probl. 20 (2004), 103–120.
[3] D.A. Lorenz, F. Schöpfer, S. Wenger, The linearized Bregman method via split feasibility problems: Analysis and generalizations, SIAM J. Imaging Sci. 7 (2014), 1237-1262.
[4] Y. Censor, T. Elfving, N. Kopf, T. Bortfeld, The multiple-sets split feasibility problem and its applications for inverse problems, Inverse Probl. 21 (2005), 2071-2084.
[5] Y. Hu, C. Li, K. Meng, J. Qin, X. Yang, Group sparse optimization via $\ell_{p,q}$ regularization, J. Machine Learn. Res. 18 (2017), 1-52.
[6] J. Wang, Y. Hu, C. Li, J.-C. Yao, Linear convergence of CQ algorithms and applications in gene regulatory network inference, Inverse Probl. 33 (2017), 055017.
[7] F. Wang, Polyak's gradient method for split feasibility problem constrained by level sets, Numer. Algo. 77 (2018), 925-938.
[8] C. Byrne, Iterative oblique projection onto convex sets and the split feasibility problem, Inverse Prob. 18 (2002), 441-453.
[9] J. Wang, Y. Hu, C.K.W. Yu, X. Zhuang, A family of projection gradient methods for solving the multiple-sets split feasibility problem, J. Optim. Theory Appl. 183 (2019), 520-534.
[10] M. Wen, J. Peng, Y. Tang, A cyclic and simultaneous iterative method for solving the multiple-sets split feasibility problem, J. Optim. Theory Appl. 166 (2015), 844-860.
[11] H.-K. Xu, A variable Krasnosel'skií-Mann algorithm and the multiple-set split feasibility problem, Inverse Probl. 22 (2006), 2021-2034.
[12] W. Zhang, D. Han, Z. Li, A self-adaptive projection method for solving the multiple-sets split feasibility problem, Inverse Probl. 25 (2009), 115001.
[13] J. Zhao, Q. Yang, Self-adaptive projection methods for the multiple-sets split feasibility problem, Inverse Probl. 27 (2011), 035009.
[14] J.-P. Crouzeix, J.-E. Martinez-Legaz, M. Volle, Generalized Convexity, Generalized Monotonicity, Kluwer Academic Publishers, Dordrecht, 1998.
[15] N. Hadjisavvas, S. Komlósi, S. Schaible, Handbook of Generalized Convexity and Generalized Monotonicity, Springer-Verlag, New York, 2005.
[16] Y. Hu, G. Li, C.K.W. Yu, T.L. Yip, Quasi-convex feasibility problems: Subgradient methods and convergence rates, Eur. J. Oper. Res. 298 (2022), 45-58.
[17] Y. Hu, X. Yang, C.-K. Sim, Inexact subgradient methods for quasi-convex optimization problems, Eur. J. Oper. Res. 240 (2015), 315-327.
[18] Y. Hu, C.K.W. Yu, X. Yang, Incremental quasi-subgradient methods for minimizing the sum of quasi-convex functions, J. Global Optim. 75 (2019), 1003-1028.
[19] I.M. Stancu-Minasian, Fractional Programming, Kluwer Academic Publishers, Dordrecht, 1997.
[20] N. Nimana, A.P. Farajzadeh, N. Petrot, Adaptive subgradient method for the split quasi-convex feasibility problems, Optimization 65 (2016), 1885-1898.
[21] D. Aussel, M. Pištěk, Limiting normal operator in quasiconvex analysis, Set-Valued Var. Anal. 23 (2015), 669-685.
[22] H.J. Greenberg, W.P. Pierskalla, Quasiconjugate functions and surrogate duality, Cahiers Centre Études Rech Oper. 15 (1973), 437-448.
[23] K.C. Kiwiel, Convergence and efficiency of subgradient methods for quasiconvex minimization, Math. Program. 90 (2001), 1-25.
[24] Y. Censor, A. Segal, Algorithms for the quasiconvex feasibility problem, J. Comput. Appl. Math. 185 (2006), 34-50.
[25] D.P. Bertsekas, Convex Optimization ang Algorithms, Athena Scientific, Massachusetts, 2015.
[26] Y. Hu, J. Li, C.K.W. Yu, Convergenece rates of subgradient methods for quasi-convex optimization problems, Comput. Optim. Appl. 77 (2020), 183-212.

[27] Y. Hu, X. Yang, C.K.W. Yu, Subgradient methods for saddle point problems of quasiconvex optimization, Pure Appl. Funct. Anal. 2 (2017), 83-97.

[28] Y. Hu, C.K.W. Yu, C. Li, Stochastic subgradient method for quasi-convex optimization problems, J. Nonlinear Convex Anal. 17 (2016), 711-724.

[29] Y. Hu, C.K.W. Yu, C. Li, X. Yang, Conditional subgradient methods for constrained quasi-convex optimization problems, J. Nonlinear Convex Anal. 17 (2016), 2143-2158.

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

[31] I.V. Konnov, On convergence properties of a subgradient method, Optim. Meth. Softw. 18 (2003), 53-62.

[32] H.H. Bauschke, J.M. Borwein, On projection algorithms for solving convex feasibility problems, SIAM Rev. 38 (1996), 367-426.

[33] Y. Hu, Y. Liu, M. Li, The effect of deterministic noise on quasi-subgradient method for quasi-convex feasibility problems, J. Appl. Numer. Optim. 2 (2020), 235-247.

[34] P.L. Combettes, The convex feasibility problem in image recovery, In: Advances in Imaging and Electron Physics, vol. 95, pp. 155-270, Elsevier, 1996.

[35] Y. Hu, C. Li, X. Yang, On convergence rates of linearized proximal algorithms for convex composite optimization with applications, SIAM J. Optim. 26 (2016), 1207-1235.

[36] J.M. Borwein, M.K. Tam, Reflection methods for inverse problems with application to protein conformation determination, In: Generalized Nash Equilibrium Problems, Bilevel Programming and MPEC, pp. 83-100, Springer, Singapore, 2018.

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

[38] A. Nedić, Random algorithms for convex minimization problems, Math. Program. 129 (2011), 225-253.

[39] M. Wang, D.P. Bertsekas, Stochastic first-order methods with random constraint projection, SIAM J. Optimi. 26 (2016), 681-717.

[40] A.J. Hoffman, On approximate solutions of systems of linear inequalities, J. Res. Nat. Bur. Stand. 49 (1952), 263-265.

[41] J.S. Pang, Error bounds in mathematical programming, Math. Program. 79 (1997), 299-332.

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

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

[44] X.-D. Luo, Z.-Q. Luo, Extension of Hoffman's error bound to polynomial systems, SIAM J. Optim. 4 (1994), 383-392.

[45] J.M. Borwein, G. Li, L. Yao, Analysis of the convergence rate for the cyclic projection algorithm applied to basic semialgebraic convex sets, SIAM J. Optim. 24 (2014), 498-527.

[46] K.F. Ng, X.Y. Zheng, Global error bounds with fractional exponents, Math. Program. 88 (2000), 357-370.

[47] S. Suzuki, D. Kuroiwa, Nonlinear error bounds for quasiconvex inequality systems, Optim. Lett. 11 (2017), 107-120.

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