J. Appl. Numer. Optim. 2 (2020), No. 2, pp. 235-247
Available online at http://jano.biemdas.com
https://doi.org/10.23952/jano.2.2020.2.08

THE EFFECT OF DETERMINISTIC NOISE ON A QUASI-SUBGRADIENT METHOD FOR QUASI-CONVEX FEASIBILITY PROBLEMS

YAOHUA HU[1], YUPING LIU[1], MINGHUA LI[2,*]

[1]*Shenzhen Key Laboratory of Advanced Machine Learning and Applications, College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China*
[2]*School of Mathematics and Big Data, Chongqing University of Arts and Sciences, Yongchuan, Chongqing, 402160, China*

Abstract. The quasi-convex feasibility problem (QFP), in which the involved functions are quasi-convex, is at the core of the modeling of many problems in various areas such as economics, finance and management science. In this paper, we consider an inexact incremental quasi-subgradient method to solve the QFP, in which an incremental control of component functions in the QFP is employed and the inexactness stems from computation error and noise arising from practical considerations and physical circumstances. Under the assumptions that the computation error and noise are deterministic and bounded and a Hölder condition on component functions in the QFP, we study the convergence property of the proposed inexact incremental quasi-subgradient method, and particularly, investigate the effect of the inexact terms on the incremental quasi-subgradient method when using the constant, diminishing and dynamic stepsize rules.

Keywords. Quasi-convex feasibility problem; Quasi-subgradient method; Inexact approach; Noise; Convergence analysis.

1. Introduction

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

$$f_i(x) \le 0 \quad \text{for each } i = 1, \ldots, m. \tag{1.1}$$

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 [18], radiation therapy treatment planning [8] and gene regulatory network inference [17, 31].

The convex feasibility problem (CFP), in which the involved functions are convex, has attracted a great deal of attention in the development of optimization algorithms and applications; see [6, 11, 33] and references therein. One of the most popular approaches for solving the CFP (1.1) is the subgradient method, which was originally introduced by Censor and Lent [9], and

*Corresponding author.
E-mail addresses: mayhhu@szu.edu.cn (Y. Hu), 1900201002@email.szu.edu.cn (Y. Liu), minghuali20021848@163.com (M. Li).
Received June 16, 2020; Accepted August 3, 2020.

until now, various variants and many features of subgradient methods have been devised and established for the CFP; one can refer to a review paper [6] and a recent book [33].

The convex function plays a central role in mathematical optimization; however, it is too restrictive for 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, while still inherits certain desirable properties of the convex function. This leads to a significant increase of studies in quasi-convex optimization; see [5, 12, 16, 30] and references therein. In particular, for feasibility problem (1.1), Goffin, Luo and Ye [15] introduced a notion of the quasi-convex feasibility problem (QFP), in which the functions involved are quasi-convex. Censor and Segal [10] and Hu, Yu and Yang [22] proposed the subgradient method (with the most violated constraint control and the almost cyclic control) and the incremental/stochastic subgradient methods (with different types of stepsize rules) to solve the QFP, respectively. The global convergence (to a feasible solution) of the subgradient-type methods was established therein.

The computation error and noise arise from practical considerations and physical circumstances, respectively, and are inevitable in applications. Motivated by practical reasons, the computation error is considered in the ε-subgradient-type methods, which were widely studied for convex optimization [1, 13, 25, 29] and quasi-convex optimization problems [19, 21, 22, 28]. Moreover, the computation error and physical noise are synthetically considered in the inexact subgradient-type methods in convex optimization [27] and quasi-convex optimization problems [20], in which the ε-subgradient with a physical noise is estimated at each iteration.

Motivated by practical considerations and theoretical requirements, in this present paper, we propose an inexact incremental quasi-subgradient method to solve the QFP and investigate its quantitative convergence property. In particular, the proposed inexact quasi-subgradient method employs an incremental updating control to follow an ordered cyclic of component functions in the QFP, and uses the ε-quasi-subgradient with a physical noise of each component function to approach the feasibility at each sub-iteration. It covers an incremental ε-quasi-subgradient method and the exact incremental quasi-subgradient method in [22] as special cases.

Inspired by the ideas in [20, 27] and references therein, we investigate the influence of inexact terms, including computation error and noise on the inexact incremental quasi-subgradient method. The computation error, which gives rise to the ε-quasi-subgradient, is inevitable in computing process; while the noise, which comes from physical circumstances, is manifested in inexact computation of quasi-subgradient. Under the assumptions that the computation error and noise are deterministic and bounded and a Hölder condition on component functions in the QFP, we establish the convergence property of the proposed inexact incremental quasi-subgradient method, and particularly, present the effect of the inexact terms on the incremental quasi-subgradient method when using the constant, diminishing and dynamic stepsize rules. The quantitative convergence result estimates the violation to the feasibility within a tolerance given explicitly in terms of error, noise and stepsize (as an additive form). This work not only extends [22, Algorithm 3] to the inexact scenario, but also improves [22, Theorem 4.1] to obtain a tighter upper bound on the total tolerance; see explanation in Remark 3.4.

The present paper is organized as follows. In Section 2, we present the notations and some preliminary lemmas that will be used in this paper. In Section 3, we propose an inexact incremental quasi-subgradient method to solve the QFP and establish its quantitative convergence results.

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. 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

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

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

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

For any $\alpha \in \mathbb{R}$, the sublevel sets of f are denoted by

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

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. Particularly, it is well-known that f is quasi-convex if and only if $\mathrm{lev}_{<\alpha} f$ (and/or $\mathrm{lev}_{\le\alpha} f$) is convex for any $\alpha \in \mathbb{R}$.

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 a quasi-convex function (for example $f(x) := x^3$). Hence, the introduction of the (nonempty) subdifferential of quasi-convex functions plays an important role in quasi-convex optimization. Several different types of subdifferentials of quasi-convex functions have been introduced in the literature, see [2, 3, 4, 14, 20, 24] and references therein. The earliest one is the Greenberg-Pierskalla quasi-subdifferential proposed in [14]; recently, Kiwiel [24] and Hu, Yang and Sim [20] introduced a quasi-subdifferential defined as a normal cone to its sublevel set and applied this quasi-subgradient in their proposed subgradient methods, respectively.

Definition 2.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quasi-convex function, $x \in \mathbb{R}^n$ and $\varepsilon \ge 0$.

(i) The Greenberg-Pierskalla quasi-subdifferential of f at x is defined by

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

(ii) The quasi-subdifferential of f at x is defined by

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

(iii) The ε-quasi-subdifferential of f at x is defined by

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

Any vector $g \in \partial^{\mathrm{Q}} f(x)$ or $g \in \partial^{\mathrm{Q}}_{\varepsilon} f(x)$ is called a quasi-subgradient or an ε-quasi-subgradient of f at x, respectively.

It is clear from the definition that the (ε-)quasi-subdifferential is a normal cone to the sublevel set of the quasi-convex function and that $\partial^{\mathrm{GP}} f(x) \subseteq \partial^{\mathrm{Q}} f(x) \subseteq \partial^{\mathrm{Q}}_{\varepsilon} f(x)$ for any $x \in \mathbb{R}^n$. Moreover, it was shown in [20, Lemma 2.1] that $\partial^{\mathrm{Q}} f(x) \setminus \{0\} \neq \emptyset$ for any $x \in \mathbb{R}^n$ whenever f is quasi-convex. Hence, the (ε-)quasi-subdifferential of a quasi-convex function contains at least one unit vector. This is a special property of the quasi-subdifferential that the convex subdifferential does not share. In particular, it was claimed in [20] that the quasi-subdifferential coincides with the convex cone hull of the convex subdifferential whenever f is convex.

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

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

The following lemma shows that the positive part operator (partially) preserves the quasi-convexity and the (ε-)quasi-subdifferentials. The proof adopts a line of analysis similar to [22, Lemma 2.2]. Hence, the details are omitted.

Lemma 2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quasi-convex function and $z \notin \mathrm{lev}_{\leq \varepsilon} f$. Then f^+ is quasi-convex, $\partial^{\mathrm{Q}} f(z) = \partial^{\mathrm{Q}} f^+(z)$ and $\partial^{\mathrm{Q}}_{\varepsilon} f(z) = \partial^{\mathrm{Q}}_{\varepsilon} f^+(z)$.*

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, which is commonly assumed (in the form of bounded subgradient assumption) in the convergence study of subgradient methods for convex optimization problems; see, e.g., [6, 7, 27, 29].

Definition 2.2. Let $p \in (0, 1]$ and $L > 0$. $f : \mathbb{R}^n \to \mathbb{R}$ is said to satisfy the Hölder condition of order p with modulus L at $x \in \mathbb{R}^n$ if

$$|f(y) - f(x)| \leq L\|y - x\|^p \quad \text{for any } y \in \mathbb{R}^n. \tag{2.1}$$

f is said to satisfy the Hölder condition of order p with modulus L on X if (2.1) holds for any $x \in X$.

The Hölder condition was used to provide a fundamental property of the quasi-subgradient in [26, 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 [19, 20, 32] and quasi-convex feasibility problems [10, 22]. The following lemma extends the fundamental property to the ε-quasi-subgradient, which will be useful in the convergence analysis of inexact quasi-subgradient methods.

Lemma 2.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quasi-convex and continuous function, $X \subseteq \mathbb{R}^n$ be a closed and convex set, and let $S := \{x \in X : f(x) \leq 0\} \neq \emptyset$. Let $p \in (0, 1]$ and $L > 0$, and suppose that f satisfies the Hölder condition of order p with modulus L on S. Then, for any $x \in S$ and $z \in X \setminus \mathrm{lev}_{\leq \varepsilon} f$, it holds that*

$$f(z) - \varepsilon \leq L\langle g(z, \varepsilon), z - x\rangle^p \quad \textit{for any } g(z, \varepsilon) \in \partial^{\mathrm{Q}}_{\varepsilon} f(z) \cap \mathbb{S}. \tag{2.2}$$

Proof. By the assumptions of this lemma and Lemma 2.1, one can check that f^+ is quasi-convex and continuous and satisfies the Hölder condition of order p with modulus L on S, and $\partial^{\mathrm{Q}}_{\varepsilon} f^+(z) = \partial^{\mathrm{Q}}_{\varepsilon} f(z)$ for each $z \notin \mathrm{lev}_{\leq \varepsilon} f$ and $S = \arg\min_{x \in X} f^+(x)$. Then, for any $x \in S$ and

$z \in X \setminus \text{lev}_{\le \varepsilon} f$ (i.e., $f^+(x) = 0$ and $f^+(z) = f(z)$), by applying [19, Lemma 4.1] (to f^+), we derive that

$$f^+(z) - f^+(x) - \varepsilon \le L \langle g(z,\varepsilon), z - x \rangle^p \quad \text{for any } g(z,\varepsilon) \in \partial^{\mathrm{Q}}_{\varepsilon} f^+(z) \cap \mathbb{S}.$$

Consequently, (2.2) is obtained. The proof is complete. □

We end this section by recalling the following two lemmas, which are useful in the convergence analysis of subgradient-type methods.

Lemma 2.3 ([23, Lemma 4.1]). *Let $\gamma \ge 1$ and $a_i \ge 0$ for $i = 1, \ldots, n$. Then it holds that*

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

Lemma 2.4 ([25, Lemma 2.1]). *Let $\{a_k\}$ be a sequence of scalars, and let $\{v_k\}$ be a sequence of nonnegative scalars. Suppose that $\lim_{k\to\infty} \sum_{i=1}^{k} v_i = \infty$. Then it holds that*

$$\liminf_{k\to\infty} a_k \le \liminf_{k\to\infty} \frac{\sum_{i=1}^{k} v_i a_i}{\sum_{i=1}^{k} v_i} \le \limsup_{k\to\infty} \frac{\sum_{i=1}^{k} v_i a_i}{\sum_{i=1}^{k} v_i} \le \limsup_{k\to\infty} a_k.$$

In particular, if $\lim_{k\to\infty} a_k = a$, then $\lim_{k\to\infty} \frac{\sum_{i=1}^{k} v_i a_i}{\sum_{i=1}^{k} v_i} = a$.

3. The inexact quasi-subgradient method for the quasi-convex feasibility problem

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

$$x \in X \quad \text{and} \quad f_i(x) \le 0 \quad \text{for each } i = 1, 2, ..., m. \tag{3.1}$$

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

$$S := \{x \in X : f_i(x) \le 0, i = 1, 2, ..., m\} \ne \varnothing.$$

Moreover, we always assume that each component function of the QFP (3.1) satisfies a Hölder condition as in the following assumption. The Hölder condition is a common assumption to develop the convergence theory of subgradient-type methods for quasi-convex optimization problems [19, 20, 21, 32] and quasi-convex feasibility problems [10, 22]. Assumption 3.1 consists of the Hölder condition for all component functions of the QFP (3.1).

Assumption 3.1. Let $p \in (0, 1]$ and $L_i > 0$ for $i = 1, \ldots, m$. For each $i = 1, \ldots, m$, f_i satisfies the Hölder condition of order p with modulus L_i on S. Furthermore, we write

$$L_{\max} := \max_{i=1,\ldots,m} L_i. \tag{3.2}$$

3.1. **Inexact incremental quasi-subgradient method.** The purpose of this subsection is to propose an inexact incremental quasi-subgradient method to solve the QFP (3.1), in which an incremental updating control is employed to follow an ordered cyclic of component functions in (3.1) and the approximate quasi-subgradient of each component function with the following form is used to approach the feasibility at each sub-iteration:

$$\hat{g}(x,\varepsilon) = g(x,\varepsilon) + r(x),$$

where $g(x,\varepsilon) \in \partial_{\varepsilon}^{\mathrm{Q}} f(x) \cap \mathbb{S}$ is an arbitrary unit ε-quasi-subgradient of f at x and $r(x)$ is a noise vector. Specifically, the inexact incremental quasi-subgradient method is formally described as follows.

Algorithm 1: Inexact incremental quasi-subgradient method - QFP.

Initialize an initial point $x_0 \in X$, a sequence of stepsizes $\{v_k\} \subseteq \mathbb{R}_+$, and let $k := 0$;
while $\max_{i=1,\ldots,m} f_i(x_k) > \varepsilon_k$ **do**
    Let $z_{k,0} := x_k$;
    for $i = 1, \ldots, m$ **do**
        if $f_i(z_{k,i-1}) \leq \varepsilon_k$ **then**
            Let $z_{k,i} := z_{k,i-1}$;
        else
            Update $z_{k,i} := \mathrm{P}_X\left(z_{k,i-1} - v_k \hat{g}_{k,i}\right)$, in which the approximate subgradient is of form $\hat{g}_{k,i} := g_{k,i} + r_{k,i}$ with $g_{k,i} \in \partial_{\varepsilon_k}^{\mathrm{Q}} f_i(z_{k,i-1}) \cap \mathbb{S}$ and $r_{k,i}$ being a noise vector;
        end
    end
    Let $x_{k+1} := z_{k,m}$ and $k := k+1$.
end

Remark 3.1. (i) In Algorithm 1, the update of the subgradient iteration is processed within an ordered cyclic sequence on each component function f_i involved in the QFP (3.1). It is worth mentioning that the convergence analysis of Algorithm 1 in the next subsection still works if any order of component functions is assumed, as long as each component is taken into account exactly once within a cycle. Hence, in applications, we could reorder the components $\{f_i\}$ by either shifting or reshuffling at the beginning of each cycle, and then proceed with the calculations until the end of this cycle.

(ii) In the case when the noise vanishes (i.e., $R = 0$), Algorithm 1 is reduced to an incremental ε-quasi-subgradient method for solving the QFP (3.1), which is an inexact version of [22, Algorithm 3] with the ε-quasi-subgradient in place of the (exact) quasi-subgradient. Furthermore, if the computation is precise (i.e., $\varepsilon = 0$), Algorithm 1 is reduced to the exact incremental quasi-subgradient method as in [22, Algorithm 3].

Algorithm 1 extends the incremental quasi-subgradient method for solving the QFP (3.1) to the inexact scenario in terms of computation error and noise. In the rest of this paper, we aim to discuss the convergence property of Algorithm 1 and particularly investigate the effect of the inexact terms on the incremental quasi-subgradient method under the assumption that the computation error and noise are deterministic and bounded.

3.2. **Convergence analysis.** This subsection is devoted to studying the convergence property of the inexact incremental quasi-subgradient method with different types of stepsize rules for the QFP (3.1) in terms of computation error and noise. Throughout this section, the following two assumptions are made for the convergence analysis.

Assumption 3.2. X is compact with its diameter being D.

Assumption 3.3. The error and noise are bounded, i.e., there exist $\varepsilon \geq 0$ and $R \geq 0$ such that

$$\limsup_{k\to\infty} \varepsilon_k = \varepsilon \quad \text{and} \quad \|r_k\| \leq R \quad \text{for each } k \in \mathbb{N}. \tag{3.3}$$

Remark 3.2. By Assumption 3.3, it follows that the approximate quasi-subgradients involved in Algorithm 1 are uniformly bounded, i.e., $\|\hat{g}_{k,i}\| \leq \|g_{k,i}\| + \|r_{k,i}\| \leq 1+R$ for any $i = 1,\ldots,m$ and $k \in \mathbb{N}$.

We now start the convergence analysis of Algorithm 1 by providing a basic inequality, which shows a significant property of an inexact incremental quasi-subgradient iteration. To this end, we define the max-function of the feasible system $\{f_i : i = 1,\ldots,m\}$ by

$$F(x) := \max_{i=1,\ldots,m} f_i(x) \quad \text{for any } x \in \mathbb{R}^n. \tag{3.4}$$

Then, the feasible solution set of the QFP can be written as $S = \{x \in X : F(x) \leq 0\}$. The basic inequality (3.5) measures the difference of the distances of iterates from any feasible solution by the max-function value at current iterate, as well as the stepsize and inexact terms. It is worth noting that (3.5) follows from the nature of (inexact) quasi-subgradient methods and inherits the important property of the basic inequality in the literature of quasi-subgradient methods; see, e.g., [10, 19, 20, 22, 24, 32].

Lemma 3.1. *Suppose that Assumptions 3.1-3.3 are satisfied. Let $\{x_k\}$ be the sequence generated by Algorithm 1. Suppose that $F(x_k) > \varepsilon_k$. Then, for any $x \in S$, it holds that*

$$\|x_{k+1}-x\|^2 \leq \|x_k - x\|^2 - 4v_k(2L_{\max})^{-\frac{1}{p}}(F(x_k)-\varepsilon_k)^{\frac{1}{p}} + 2mv_kRD + m^2v_k^2(1+R)^2. \tag{3.5}$$

Proof. Fix $x \in S$. We first show that the following inequality holds for $i = 1,\ldots,m$:

$$\|z_{k,i}-x\|^2 \leq \|z_{k,i-1}-x\|^2 - 2v_kL_{\max}^{-\frac{1}{p}}\left(f_i(z_{k,i-1})-\varepsilon_k\right)_+^{\frac{1}{p}} + 2v_kRD + v_k^2(1+R)^2. \tag{3.6}$$

In view of Algorithm 1, if $f_i(z_{k,i-1}) \leq \varepsilon_k$, then it is updated that $z_{k,i} = z_{k,i-1}$, and so (3.6) holds automatically. Otherwise, $f_i(z_{k,i-1}) > \varepsilon_k$, then it is updated that

$$z_{k,i} = \mathrm{P}_X\left(z_{k,i-1} - v_k\hat{g}_{k,i}\right), \quad \text{where} \quad \hat{g}_{k,i} = g_{k,i} + r_{k,i}.$$

In this case, it follows from the nonexpansive property of the projection operator that

$$\begin{aligned}\|z_{k,i}-x\|^2 &\leq \left\|z_{k,i-1} - v_k\hat{g}_{k,i} - x\right\|^2 \\ &= \|z_{k,i-1}-x\|^2 - 2v_k\left\langle g_{k,i}+r_{k,i}, z_{k,i-1}-x\right\rangle + v_k^2\|g_{k,i}+r_{k,i}\|^2 \\ &\leq \|z_{k,i-1}-x\|^2 - 2v_k\left\langle g_{k,i}, z_{k,i-1}-x\right\rangle + 2v_kRD + v_k^2(1+R)^2,\end{aligned} \tag{3.7}$$

where the last inequality follows from Assumptions 3.2-3.3 and Remark 3.2. Noting by $x \in S$ that $f_i(x) \leq F(x) \leq 0$, Lemma 2.2 is applicable (with f_i, $z_{k,i-1}$, ε_k, $g_{k,i}$ in place of f, z, ε, $g(z,\varepsilon)$) to concluding that

$$f_i(z_{k,i-1}) - \varepsilon_k \leq L_i\langle g_{k,i}, z_{k,i-1}-x\rangle^p \leq L_{\max}\langle g_{k,i}, z_{k,i-1}-x\rangle^p$$

(due to (3.2)). Then, (3.7) is reduced to (3.6) in this case. Hence, (3.6) is proved as desired.

Next, we estimate the second term on the right hand side of (3.6) in terms of $(f_i(x_k)-\varepsilon_k)_+$. Note by the subadditivity of a_+ that

$$(f_i(x_k)-\varepsilon_k)_+ \le (f_i(x_k)-f_i(z_{k,i-1}))_+ + (f_i(z_{k,i-1})-\varepsilon_k)_+.$$

Since $p\in(0,1]$, one has by Lemma 2.3 that

$$2^{1-\frac{1}{p}}(f_i(x_k)-\varepsilon_k)_+^{\frac{1}{p}} \le (f_i(x_k)-f_i(z_{k,i-1}))_+^{\frac{1}{p}} + (f_i(z_{k,i-1})-\varepsilon_k)_+^{\frac{1}{p}}. \tag{3.8}$$

By Assumption 3.1 (cf. (2.1)) and in view of Algorithm 1, it follows that

$$\begin{aligned}(f_i(x_k)-f_i(z_{k,i-1}))_+ &\le |f_i(x_k)-f_i(z_{k,i-1})| \\ &\le L_i\|z_{k,i-1}-x_k\|^p \\ &\le L_{\max}\left(\sum_{j=1}^{i-1}\|z_{k,j}-z_{k,j-1}\|\right)^p \\ &\le L_{\max}\left(v_k(i-1)(1+R)\right)^p\end{aligned}$$

(cf. Remark 3.2). Hence, (3.8) is reduced to

$$(f_i(z_{k,i-1})-\varepsilon_k)_+^{\frac{1}{p}} \ge 2^{1-\frac{1}{p}}(f_i(x_k)-\varepsilon_k)_+^{\frac{1}{p}} - L_{\max}^{\frac{1}{p}}v_k(i-1)(1+R).$$

So, (3.6) yields that

$$\|z_{k,i}-x\|^2 \le \|z_{k,i-1}-x\|^2 - 4v_k(2L_{\max})^{-\frac{1}{p}}(f_i(x_k)-\varepsilon_k)_+^{\frac{1}{p}} + 2v_kRD + v_k^2(2i-1)(1+R)^2.$$

Finally, summing the above inequality over $i=1,\ldots,m$, we derive that

$$\|x_{k+1}-x\|^2 \le \|x_k-x\|^2 - 4v_k(2L_{\max})^{-\frac{1}{p}}\sum_{i=1}^{m}(f_i(x_k)-\varepsilon_k)_+^{\frac{1}{p}} + 2mv_kRD + m^2v_k^2(1+R)^2.$$

From the assumption $F(x_k)>\varepsilon_k$ and definition in (3.4), we have that

$$\sum_{i=1}^{m}(f_i(x_k)-\varepsilon_k)_+^{\frac{1}{p}} \ge \max_{i=1,\ldots,m}(f_i(x_k)-\varepsilon_k)_+^{\frac{1}{p}}, = (F(x_k)-\varepsilon_k)^{\frac{1}{p}},$$

whihc achieves (3.5). The proof is complete. □

Remark 3.3. In the case when the noise vanishes (i.e., $R=0$), the term $\langle r_{k,i}, z_{k,i-1}-x\rangle$ vanishes on (3.7), and correspondingly, the third term $2mv_kRD$ on the right hand side of (3.6) vanishes in the basic inequality. Hence, Lemma 3.1 is satisfied without the compactness hypothesis of X (i.e., Assumption 3.2). Therefore, the convergence theorems established below are true for the incremental ε-quasi-subgradient method for the QFP (3.1) regardless of Assumption 3.2.

The stepsize rule has a critical effect on the convergence property and computational capacity of subgradient methods. In this section, by virtue of the basic inequality provided in Lemma 3.1, we establish in Theorems 3.1-3.3 the convergence properties of the inexact incremental quasi-subgradient method for the QFP (3.1) by using the typical constant, diminishing and dynamic stepsize rules, respectively.

Theorem 3.1. *Suppose that Assumptions 3.1-3.3 are satisfied. Let $\{x_k\}$ be the sequence generated by Algorithm 1 with $v_k\equiv v>0$. Then*

$$\liminf_{k\to\infty} F(x_k) \le \varepsilon + 2L_{\max}\left(\frac{1}{2}mRD + \frac{1}{4}m^2(1+R)^2v\right)^p. \tag{3.9}$$

Proof. Without loss of generality, we assume that

$$F(x_k) \le \varepsilon_k + 2L_{\max}\left(\frac{1}{2}mRD + \frac{1}{4}m^2(1+R)^2 v\right)^p$$

only occurs for finitely many times; otherwise, (3.9) holds automatically. We further assume that

$$F(x_k) > \varepsilon_k + 2L_{\max}\left(\frac{1}{2}mRD + \frac{1}{4}m^2(1+R)^2 v\right)^p \tag{3.10}$$

for each $k \in \mathbb{N}$ (otherwise, one can choose some $N \in \mathbb{N}$ such that (3.10) is satisfied for each $k \ge N$ and focus on the subsequence $\{x_k\}_{k \ge N}$ instead). Hence, Lemma 3.1 is applicable to ensuring (3.5) for each $k \in \mathbb{N}$. Summing (3.5) with $v_k \equiv v$ over $k = 0, \ldots, n-1$, we deduce that

$$\|x_n - x\|^2 - \|x_0 - x\|^2 \le -4(2L_{\max})^{-\frac{1}{p}} v \sum_{k=0}^{n-1} (F(x_k) - \varepsilon_k)^{\frac{1}{p}} + 2mRDnv + m^2(1+R)^2 nv^2.$$

Consequently,

$$4(2L_{\max})^{-\frac{1}{p}} \frac{\sum_{k=0}^{n-1}(F(x_k)-\varepsilon_k)^{\frac{1}{p}}}{n} \le \frac{\|x_0 - x\|^2}{nv} + 2mRD + m^2(1+R)^2 v.$$

Then, we obtain by (3.3) and Lemma 2.4 that

$$\begin{aligned}
\liminf_{k\to\infty} (F(x_k) - \varepsilon)^{\frac{1}{p}} &= \liminf_{k\to\infty} (F(x_k) - \varepsilon_k)^{\frac{1}{p}} \\
&\le \liminf_{n\to\infty} \frac{\sum_{k=0}^{n-1}(F(x_k)-\varepsilon_k)^{\frac{1}{p}}}{n} \\
&\le (2L_{\max})^{\frac{1}{p}} \left(\frac{1}{2}mRD + \frac{1}{4}m^2(1+R)^2 v\right).
\end{aligned}$$

Consequently, (3.9) is achieved, and the proof is complete. □

Remark 3.4. (i) Theorem 3.1 shows the convergence of the violation of the sequence generated by Algorithm 1 (with the constant stepsize rule) to the feasibility within some tolerance given in terms of error and noise. As shown in Theorem 3.1, the total tolerance

$$T_v := \varepsilon + 2L_{\max}\left(\frac{1}{2}mRD + \frac{1}{4}m^2(1+R)^2 v\right)^p \tag{3.11}$$

has an additive form of the error level ε and the noise level R and plus a term related to the stepsize v.

(ii) As mentioned in Remarks 3.1 and 3.3, in the special case when $R = 0$ and $\varepsilon = 0$, Theorem 3.1 is applicable to concluding the convergence property of the (exact) incremental quasi-subgradient method (i.e., [22, Algorithm 3]) with the constant stepsize rule for solving the QFP (3.1) as

$$\liminf_{k\to\infty} F(x_k) \le 2L_{\max}\left(\frac{1}{4}m^2 v\right)^p$$

under Assumption 3.1. While, the convergence result in [22, Theorem 4.1(a)] is

$$\liminf_{k\to\infty} F(x_k) \le 2m^{1-p} L_{\max}\left(\frac{1}{4}m^2 v\right)^p$$

under the same assumption. From these above results, we can see that Theorem 3.1 not only extends [22, Algorithm 3] to the inexact scenario, but also improves [22, Theorem 4.1(a)] to obtain a tighter upper bound on the total tolerance.

Theorem 3.2. *Suppose that Assumptions 3.1-3.3 are satisfied. Let* $\{x_k\}$ *be a sequence generated by Algorithm 1 with the stepsize satisfying*

$$\lim_{k\to\infty} v_k = 0 \quad \textit{and} \quad \sum_{k=0}^{\infty} v_k = \infty. \tag{3.12}$$

Then

$$\liminf_{k\to\infty} F(x_k) \le \varepsilon + 2L_{\max}\left(\frac{1}{2}mRD\right)^p. \tag{3.13}$$

Proof. Similar to the beginning of the proof of Theorem 3.1, we can assume, without loss of generality, that

$$F(x_k) > \varepsilon_k + 2L_{\max}\left(\frac{1}{2}mRD\right)^p \quad \text{for each } k\in\mathbb{N}.$$

Then, Lemma 3.1 is applicable to ensuring (3.5) for each $k\in\mathbb{N}$. Summing (3.5) over $k = 0,\ldots,n-1$, we obtain that

$$\|x_n - x\|^2 - \|x_0 - x\|^2 \le -4(2L_{\max})^{-\frac{1}{p}}\sum_{k=0}^{n-1} v_k(F(x_k)-\varepsilon_k)^{\frac{1}{p}} + 2mRD\sum_{k=0}^{n-1} v_k + m^2(1+R)^2\sum_{k=0}^{n-1} v_k^2.$$

Consequently,

$$4(2L_{\max})^{-\frac{1}{p}}\frac{\sum_{k=0}^{n-1} v_k(F(x_k)-\varepsilon_k)^{\frac{1}{p}}}{\sum_{k=0}^{n-1} v_k} \le \frac{\|x_0-x\|^2}{\sum_{k=0}^{n-1} v_k} + 2mRD + m^2(1+R)^2\frac{\sum_{k=0}^{n-1} v_k^2}{\sum_{k=0}^{n-1} v_k}.$$

Then, we obtain by (3.3) and Lemma 2.4 that

$$\begin{aligned}\liminf_{k\to\infty}(F(x_k)-\varepsilon)^{\frac{1}{p}} &= \liminf_{k\to\infty}(F(x_k)-\varepsilon_k)^{\frac{1}{p}} \\ &\le \liminf_{n\to\infty}\frac{\sum_{k=0}^{n-1} v_k(F(x_k)-\varepsilon_k)^{\frac{1}{p}}}{\sum_{k=0}^{n-1} v_k} \\ &\le \frac{1}{4}(2L_{\max})^{\frac{1}{p}}\liminf_{n\to\infty}\left(\frac{\|x_0-x\|^2}{\sum_{k=0}^{n-1} v_k} + 2mRD + m^2(1+R)^2\frac{\sum_{k=0}^{n-1} v_k^2}{\sum_{k=0}^{n-1} v_k}\right).\end{aligned} \tag{3.14}$$

Note by (3.12) and Lemma 2.4 that

$$\lim_{n\to\infty}\frac{\|x_0-x\|^2}{\sum_{k=0}^{n-1} v_k} = 0 \quad \text{and} \quad \lim_{n\to\infty}\frac{\sum_{k=0}^{n-1} v_k^2}{\sum_{k=0}^{n-1} v_k} = 0.$$

Then, (3.14) is reduced to

$$\liminf_{k\to\infty}(F(x_k)-\varepsilon)^{\frac{1}{p}} \le \frac{1}{2}mRD(2L_{\max})^{\frac{1}{p}}.$$

Thus, (3.13) is achieved. The proof is complete. □

Remark 3.5. (i) Theorem 3.2 shows the convergence of the violation of Algorithm 1 (by using the diminishing stepsize rule) to the feasibility within a total tolerance given in terms of error and noise:

$$T := \varepsilon + 2L_{\max}\left(\frac{1}{2}mRD\right)^{p}.$$

This total tolerance has an additive form of the error level ε and the noise level R, and can be understood as an asymptotic result of the tolerance of the constant stepsize rule (3.11) as $T = \lim_{v\to 0} T_v$.

(ii) In the special case when $R = 0$ and $\varepsilon = 0$, Theorem 3.2 is applicable to concluding the convergence property of the (exact) incremental quasi-subgradient method (i.e., [22, Algorithm 3]) with the diminishing stepsize rule for the QFP (3.1) as

$$\liminf_{k\to\infty} F(x_k) = 0.$$

This is indeed the same as [22, Theorem 4.1(b)]. Hence, Theorem 3.2 extends [22, Algorithm 3] to the inexact scenario and covers [22, Theorem 4.1(b)] as a special case.

Theorem 3.3. *Suppose that Assumptions 3.1-3.3 are satisfied. Let $\{x_k\}$ be the sequence generated by Algorithm 1 with the stepsize given by*

$$v_k := \frac{2\gamma_k}{m^2(1+R)^2}\left(\left(\frac{F(x_k)-\varepsilon_k}{2L_{\max}}\right)^{\frac{1}{p}} - \frac{1}{2}mRD\right)_+ \quad \textit{for any } k \in \mathbb{N}, \tag{3.15}$$

where $0 < \underline{\gamma} \le \gamma_k \le \overline{\gamma} < 2$. Then, either $F(x_k) \le \varepsilon_k$ for some $k \in \mathbb{N}$ or

$$\limsup_{k\to\infty} F(x_k) \le \varepsilon + 2L_{\max}\left(\frac{1}{2}mRD\right)^{p}. \tag{3.16}$$

Proof. Without loss of generality, we assume that $F(x_k) > \varepsilon_k$ for each $k \in \mathbb{N}$; otherwise, the conclusion of this theorem follows. Then, Lemma 3.1 is applicable to ensuring (3.5) for each $k \in \mathbb{N}$. By the definition of (3.15), (3.5) is reduced to

$$\begin{aligned}
\|x_{k+1} - x\|^2 &\le \|x_k - x\|^2 - \frac{4\gamma_k(2-\gamma_k)}{m^2(1+R)^2}\left(\left(\frac{F(x_k)-\varepsilon_k}{2L_{\max}}\right)^{\frac{1}{p}} - \frac{1}{2}mRD\right)_+^2 \\
&\le \|x_k - x\|^2 - \frac{4\underline{\gamma}(2-\bar{\gamma})}{m^2(1+R)^2}\left(\left(\frac{F(x_k)-\varepsilon_k}{2L_{\max}}\right)^{\frac{1}{p}} - \frac{1}{2}mRD\right)_+^2.
\end{aligned}$$

Then, it follows that

$$\sum_{k=1}^{\infty}\left(\left(\frac{F(x_k)-\varepsilon_k}{2L_{\max}}\right)^{\frac{1}{p}} - \frac{1}{2}mRD\right)_+^2 \le \frac{m^2(1+R)^2}{4\underline{\gamma}(2-\overline{\gamma})}\|x_0 - x^*\|^2,$$

which is finite. Consequently, one has $\lim_{k\to\infty}\left(\left(\frac{F(x_k)-\varepsilon_k}{2L_{\max}}\right)^{\frac{1}{p}} - \frac{1}{2}mRD\right)_+ = 0$, and thus, (3.16) is achieved. The proof is complete. ☐

Acknowledgements

The first author was supported in part by the National Natural Science Foundation of China (11871347), Natural Science Foundation of Guangdong Province of China (2019A1515011917, 2020B1515310008), the Project of Educational Commission of Guangdong Province of China (2019KZDZX1007), Natural Science Foundation of Shenzhen (JCYJ20190808173603590, JCYJ 20170817100950436) and Interdisciplinary Innovation Team of Shenzhen University. The third author was supported in part by the Natural Science Foundation of Chongqing Municipal Science and Technology Commission (cstc2018jcyjAX0226), Basic Science and Frontier Technology Research of Yongchuan (Ycstc, 2018nb1401), Foundation for High-level Talents of Chongqing University of Art and Sciences (R2016SC13, 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.

REFERENCES

[1] A. Auslender, M. Teboulle. Interior gradient and ε-subgradient descent methods for constrained convex minimization, Math. Oper. Res. 29 (2004), 1-26.

[2] D. Aussel, New developments in quasiconvex optimization. In Q. H. A. Saleh Abdullah R. Al-Mezel, Falleh Rajallah M. Al-Solamy, editor, Fixed Point Theory, Variational Analysis, and Optimization, pp. 171-205. Taylor & Francis Group, 2014.

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

[4] D. Aussel, M. Pištěk, Limiting normal operator in quasiconvex analysis, Set-Valued Var. Anal. 23 (2015), 669-685.

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

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

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

[8] 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.

[9] Y. Censor, A. Lent, Cyclic subgradient projections, Math. Program. 24 (1982), 233-235.

[10] Y. Censor, A. Segal, Algorithms for the quasiconvex feasibility problem, J. Comput. Appl. Math. 185 (2006), 34-50.

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

[12] J.-P. Crouzeix, J.-E. Martinez-Legaz, M. Volle, Generalized Convexity, Generalized Monotonicity, Kluwer Academic Publishers, Dordrecht, 1998.

[13] G. D'Antonio, A. Frangioni, Convergence analysis of deflected conditional approximate subgradient methods, SIAM J. Optim. 20 (2009), 357-386.

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

[15] J.-L. Goffin, Z. Q. Luo, Y. Ye, On the complexity of a column generation algorithm for convex or quasiconvex feasibility problems. In W. W. Hager, D. W. Hearn, P. M. Pardalos, ed. Large Scale Optimization: State of the Art, pp. 182–191, Kluwer Academic Publishers, 1994.

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

[17] X. Hu, Y. Hu, F. Wu, R. W. T. Leung, J. Qin, Integration of single-cell multi-omics for gene regulatory network inference, Comput. Structural Biotech. J. 18 (2020), 1925-1938.

[18] 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.
[19] Y. Hu, J. Li, C. K. W. Yu, Convergence rates of subgradient methods for quasi-convex optimization problems, Comput. Optim. Appl. 77 (2020), 183-212.
[20] Y. Hu, X. Yang, C.-K. Sim, Inexact subgradient methods for quasi-convex optimization problems, Euro. J. Oper. Res. 240 (2015), 315-327.
[21] Y. Hu, C. K. W. Yu, C. Li, Stochastic subgradient method for quasi-convex optimization problems, J. Nonlinear Convex Anal. 17 (2016), 711-724.
[22] 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.
[23] X. Huang, X. Yang, A unified augmented Lagrangian approach to duality and exact penalization, Math. Oper. Res. 28 (2003), 533-552.
[24] K. C. Kiwiel, Convergence and efficiency of subgradient methods for quasiconvex minimization, Math. Program. 90 (2001), 1-25.
[25] K. C. Kiwiel, Convergence of approximate and incremental subgradient methods for convex optimization, SIAM J. Optim. 14 (2004), 807-840.
[26] I. V. Konnov, On convergence properties of a subgradient method, Optim. Methods Softw. 18 (2003), 53-62.
[27] A. Nedić, D. P. Bertsekas, The effect of deterministic noise in subgradient methods, Math. Program. 125 (2010), 75-99.
[28] E. A. Papa Quiroz, L. Mallma Ramirez, P. R. Oliveira, An inexact proximal method for quasiconvex minimization, Euro. J. Oper. Res. 246 (2015), 721-729.
[29] N. Z. Shor, Minimization Methods for Non-differentiable Functions, Springer-Verlag, New York, 1985.
[30] I. M. Stancu-Minasian, Fractional Programming, Kluwer Academic Publishers, Dordrecht, 1997.
[31] 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.
[32] 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.
[33] A. J. Zaslavski, Approximate Solutions of Common Fixed-Point Problems, Springer, Switzerland, 2016.