CONDITIONAL SUBGRADIENT METHODS FOR CONSTRAINED QUASI-CONVEX OPTIMIZATION PROBLEMS

YAOHUA HU, CARISA KWOK WAI YU, CHONG LI, AND XIAOQI YANG

Dedicated to Professor Tomas Dominguez-Benavides on the occasion of his 65th birthday.

ABSTRACT. Subgradient methods for solving quasi-convex optimization problems have been well studied. However, the usual subgradient method usually suffers from a zig-zagging phenomena and sustains a slow convergence in many applications. To avert the zig-zagging phenomenon and speed up the convergence behavior, we introduce a conditional subgradient method to solve a nondifferentiable constrained quasi-convex optimization problem in this paper. At each iteration, a conditional quasi-subgradient, constructed by a unit quasi-subgradient and a normal vector to the constraint set, is employed in place of the quasi-subgradient, as in the usual subgradient method. Assuming the Hölder condition of order p, we investigate convergence properties, in both objective values and iterates, of the conditional subgradient method by using the constant, diminishing and dynamic stepsize rules. We also describe the finite convergence behavior of the conditional subgradient method when the interior of optimal solution set is nonempty. Extending to the inexact setting, we further propose a conditional ϵ-subgradient method and establish its convergence results under the assumption that the computational error is deterministic and bounded.

2010 *Mathematics Subject Classification.* Primary, 65K05, 90C26; Secondary, 49M37.

Key words and phrases. Conditional subgradient method; ϵ-subgradient method; quasi-convex optimization; convergence analysis.

Corresponding author. This author's work was supported in part by the National Natural Science Foundation of China (grant 11526144, 11601343), Natural Science Foundation of Guangdong (grant 2016A030310038), and Foundation for Distinguished Young Talents in Higher Education of Guangdong (grant 2015KQNCX145).

This author's work was supported in part by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/FDS14/P03/14 and UGC/FDS14/P02/15).

This author's work was supported in part by the National Natural Science Foundation of China (grant 11571308).

This author's work was supported in part by the Research Grants Council of Hong Kong (PolyU 152167/15E).

1. Introduction

In this paper, we consider the constrained quasi-convex optimization problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in X, \end{array} \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a quasi-convex function, and $X \subseteq \mathbb{R}^n$ is a nonempty, closed and convex set. We denote the set of minima and minimum value of (1.1) by X^* and f_*, respectively. A quasi-convex function usually provides a much more accurate representation of realities than a convex function does, while it still inherits some nice properties of a convex function. This leads to a significant increase of research studies in quasi-convex optimization. Nowadays, quasi-convex optimization has been developed in various branches of applied mathematics and many application fields, such as economics, engineering and management science; see [2, 4, 7, 25] and references therein.

The development of numerical algorithms for solving constrained optimization problems, especially for large-scale optimization problems, has attracted a great amount of attention. Subgradient methods are popular iterative methods for solving constrained optimization problems. The subgradient method was originally introduced to solve a nondifferentiable convex optimization problem by Polyak [22] and Ermoliev [5] in the 1970s, which generates a sequence $\{x_k\}$ by a recursive procedure

$$x_{k+1} := P_X(x_k - v_k g_k), \tag{1.2}$$

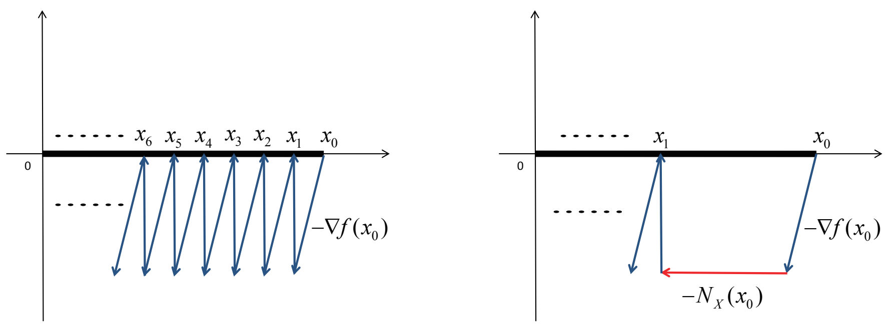
where $g_k \in \partial f(x_k) := \{g \in \mathbb{R}^n : \langle g, x - x_k \rangle \leq f(x) - f(x_k), \forall x \in \mathbb{R}^n\}$ is a subgradient of f at x_k, $v_k > 0$ is a stepsize, and P_X denotes the Euclidean projection onto the constraint set X. Over the last 40 years, various properties of subgradient methods have been discovered, many extensions and generalizations have been considered, and numerous applications have been proposed; see [3, 8, 9, 14, 19, 20, 21, 24] and references therein.

Recently, subgradient methods have been developed to solve constrained quasi-convex optimization problem (1.1), but still in its infancy. For example, Kiwiel [13] studied convergence properties of the subgradient method for solving quasi-convex optimization problems by using the diminishing stepsize rule. Extending this work, Hu et al [10] proposed a generic inexact subgradient method to solve a constrained quasi-convex optimization problem. Adopting the constant and diminishing stepsize rules, they investigated the influence of the deterministic noise to the inexact subgradient method via establishing convergence results in both objective values and iterates, and observed finite convergence to the approximate optimality. Furthermore, Hu et al [12] proposed a stochastic subgradient method, where a random noisy subgradient is adopted as the search direction, for solving quasi-convex optimization problems. They showed that the stochastic subgradient method shares the same convergence behavior as that of the exact subgradient method [13] almost surely. They also introduced a dynamic stepsize rule for the subgradient method in the category of quasi-convex optimization.

However, the subgradient method usually sustains a slow convergence, due to a zig-zagging phenomena in either convex or quasi-convex optimization problems. The zig-zagging phenomenon may stem from the fact that the subgradient may be antiparallel to the normal vector of the constraint set, and thus the Euclidean projection leads to a slow convergence of the subgradient method. In order to avoid the zig-zagging phenomenon, Larsson et al [17] proposed a conditional subgradient method for solving convex optimization problems, where a normal vector to the constraint set is taken into account in the construction of conditional subgradient direction. Convergence results of the conditional subgradient method were provided in [17] when the diminishing or dynamic stepsize rule is adopted. Extended to the inexact setting, a conditional ϵ-subgradient method, as well as its convergence theory, were studied in [18]. In particular, the following example illustrates the slow convergence of the subgradient method (1.2) and the benefit of applying the conditional subgradient method.

Example 1.1. Let $f : \mathbb{R}^2 \to \mathbb{R}$ with $f(x,y) := x+5y$, and $X := [0,10]\times\{0\}$. The (sub)gradient of f at any point is $\nabla f = (1,5)^\top$. Starting from an initial point $x_0 = (10,0)^\top$ and applying the subgradient method (1.2) to solve the associated problem (1.1), the recursive process of the subgradient method is illustrated in Figure 1(a), and the zig-zagging phenomena is observed.

Alternatively, the conditional subgradient direction, adopted in the conditional subgradient method [17], consists of a subgradient and a normal vector in $N_X(x)$, that is, $g_k \in \nabla f(x_k) + N_X(x_k)$. Hence the iterative procedure of the conditional subgradient method is demonstrated in Figure 1(b), which averts the zig-zagging phenomena and improves the convergence behavior.



(a) The subgradient method. (b) The conditional subgradient method.

FIGURE 1. Illustrations of subgradient method vs conditional subgradient method.

Inspired by the idea in [17, 18], in this paper, we propose a conditional subgradient method to solve a nondifferentiable constrained optimization problem (1.1) in the setting of quasi-convex optimization, so as to avert the zig-zagging phenomenon and to speed-up the convergence behavior. In the proposed conditional subgradient method, the search direction of each iteration is constructed by a unit quasi-subgradient and a normal vector to the constraint set X at current iterate. Under the assumption of the Hölder condition as in [10, 12], we employ the properties of the quasi-subgradient and the normal vector to provide a proper basic inequality, which is a key tool of convergence analysis in the literature of subgradient methods. Using the constant, diminishing and dynamic stepsize rules, we establish the convergence results of the conditional subgradient method in both objective values and iterates. In particular, we show that the generated sequence converges to an optimal solution of problem (1.1) when the dynamic stepsize is adopted. We also describe the finite convergence behavior of the conditional subgradient method when the interior of optimal solution set is nonempty, which is absent in [17, 18].

Another contribution of this paper is to extend the conditional subgradient method to the inexact setting. We propose a conditional ϵ-subgradient method for solving quasi-convex optimization problem (1.1), and investigate its convergence behavior to the minimum value of (1.1) within some tolerance, which is expressed in terms of the error and the stepsize, under the assumption that the computational error is deterministic and bounded. In particular, we provide the convergence analysis of conditional ϵ-subgradient method in the case when $\{\epsilon_k\}$ is not necessary to be vanishing and the constant and diminishing stepsize rules are used; while [18] only considered the case when $\{\epsilon_k\}$ tends to zero and the diminishing stepsize is adopted.

This paper is organized as follows. In section 2, we present the notations and preliminary results that will be used in this paper. In section 3, we propose a conditional subgradient method to solve the constrained quasi-convex optimization problem (1.1), and investigate convergence properties of the conditional subgradient method by using the constant, diminishing and dynamic stepsize rules. Extended to the inexact setting, the conditional ϵ-subgradient method for solving quasi-convex optimization problem (1.1) and its convergence analysis are provided in section 4.

2. Notations and preliminary results

The notations used in this paper are standard. We consider 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 $\mathbf{B}(x,r)$ to denote the closed ball at x with radius r, and particularly use $\mathbf{B}$ and $\mathbf{S}$ to denote the unit ball and the unit sphere centered at the origin, respectively. For a set $Z \subseteq \mathbb{R}^n$, we denote the closure, boundary and interior of Z by $\mathrm{cl}Z$, $\mathrm{bd}Z$ and $\mathrm{int}Z$, respectively. For $x \in \mathbb{R}^n$ and $Z \subseteq \mathbb{R}^n$, we use $\mathrm{dist}(x,Z)$ and $P_Z(x)$ to denote the Euclidean distance of x from Z and the Euclidean projection of x onto Z, respectively, that is,

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

The normal cone to Z at x is defined by

$$N_Z(x) := \{\mu \in \mathbb{R}^n : \langle \mu, y - x \rangle \le 0, \forall y \in Z\}. \tag{2.1}$$

A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be quasi-convex if for any $x, y \in \mathbb{R}^n$ and $\alpha \in [0, 1]$, the following inequality holds

$$f((1 - \alpha)x + \alpha y) \le \max\{f(x), f(y)\}.$$

For $\alpha \in \mathbb{R}$, we denote the sublevel sets of f by

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

It is well-known that f is quasi-convex if and only if $\operatorname{lev}_{<\alpha} f$ (and/or $\operatorname{lev}_{\le\alpha} f$) is convex for any $\alpha \in \mathbb{R}$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be coercive if $\lim_{\|x\| \to \infty} f(x) = \infty$, and so the set $\operatorname{lev}_{\le\alpha} f$ is bounded for any $\alpha \in \mathbb{R}$.

The subdifferential of a quasi-convex function plays an important role in quasi-convex optimization. Several different types of subdifferentials of quasi-convex function have been introduced in the literature, see [1, 6, 10, 13] and references therein. In particular, Kiwiel [13] and Hu et al [10] introduced a quasi-subdifferential, which is a normal cone to the strict sublevel set of the quasi-convex function, and applied this quasi-subgradient in their proposed subgradient methods; see, e.g., [10, 11, 12, 13]. Here, we recall the notion of quasi-subdifferential as follows, which is taken from [10] and properties of the quasi-subdifferential have been investigated therein.

Definition 2.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quasi-convex function, and let $\epsilon > 0$. The quasi-subdifferential and ϵ-quasi-subdifferential of f at $x \in \mathbb{R}^n$ are respectively defined by

$$\partial^* f(x) := \left\{g \in \mathbb{R}^n : \langle g, y - x \rangle \le 0, \forall y \in \operatorname{lev}_{<f(x)} f\right\},$$

and

$$\partial^*_\epsilon f(x) := \left\{g \in \mathbb{R}^n : \langle g, y - x \rangle \le 0, \forall y \in \operatorname{lev}_{<f(x)-\epsilon} f\right\}.$$

Any vector $g \in \partial^* f(x)$ or $g \in \partial^*_\epsilon f(x)$ is called a quasi-subgradient or an ϵ-quasi-subgradient of f at x, respectively.

The Hölder condition of order p is used to describe some properties of the quasi-subgradient in [15], and to establish the convergence theory of subgradient methods in [10, 11, 12]. It plays an important role in the study of convergence analysis in quasi-convex optimization. It is worth noting that the Hölder condition of order 1 is equivalent to the bounded subgradient assumption, assumed in the literature of subgradient methods (e.g., [3, 14, 19, 20]), whenever f is convex.

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

$$f(x) - f_* \le L \operatorname{dist}^p(x, X^*) \quad \text{for any } x \in \mathbb{R}^n.$$

The following lemma, taken from [16, Proposition 2.1], describes an important property of a quasi-convex function that satisfies the Hölder condition of order p.

Lemma 2.3. *Let $p > 0$ and $L > 0$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be quasi-convex, continuous and satisfy the Hölder condition of order p with modulus L on $\mathbb{R}^n$. Let $x \in X \setminus X^*$, and let $g(x)$ be a unit quasi-subgradient of f at x, that is, $g(x) \in \partial^* f(x) \cap \mathbf{S}$. Then it holds, for any $x^* \in X^*$, that*

$$\langle g(x), x - x^* \rangle \geq \left(\frac{f(x) - f_*}{L} \right)^{\frac{1}{p}}.$$

This property locally provides a connection between the quasi-subgradient and objective values, which is a key to establish the basic inequality in the convergence analysis of conditional subgradient methods for solving the constrained quasi-convex optimization problem (1.1). To this end, throughout this paper, we make the following assumption:

- $f : \mathbb{R}^n \to \mathbb{R}$ is quasi-convex and continuous, and satisfies the Hölder condition of order p with modulus L on $\mathbb{R}^n$.

In order to make this paper self-contained, we end this section by recalling the following lemma from [14, Lemma 2.1], which is useful in the convergence analysis. For a sequence of scalars $\{a_k\}$ and a sequence of nonnegative scalars $\{v_k\}$, the accumulated sequence of $\{v_k\}$ and the averaged sequence of $\{a_k\}$ with respect to $\{v_k\}$ are respectively defined by

$$\bar{v}_k := \sum_{i=1}^{k} v_i \quad \text{and} \quad \hat{a}_k := \frac{1}{\bar{v}_k} \left(\sum_{i=1}^{k} v_i a_i \right).$$

Lemma 2.4. *Let $\{a_k\}$ be a sequence of scalars and $\{v_k\}$ be a sequence of nonnegative scalars, and let $\{\bar{v}_k\}$ be the accumulated sequence of $\{v_k\}$ and $\{\hat{a}_k\}$ be the averaged sequence of $\{a_k\}$ with respect to $\{v_k\}$. Suppose that $\lim_{k\to\infty} \bar{v}_k = \infty$. Then*

$$\liminf_{k\to\infty} a_k \leq \liminf_{k\to\infty} \hat{a}_k \leq \limsup_{k\to\infty} \hat{a}_k \leq \limsup_{k\to\infty} a_k.$$

3. Conditional subgradient method and convergence analysis

The aims of this section are to propose a conditional subgradient method to solve the constrained quasi-convex optimization problem (1.1), and to investigate its convergence properties by using some suitable types of stepsizes. The conditional subgradient method is formally stated as follows.

Algorithm 3.1. Select an initial point $x_0 \in \mathbb{R}^n$ and a sequence of stepsizes $\{v_k\} \subseteq (0, +\infty)$. Having x_k, we select

$$g_k \in \partial^* f(x_k) \cap \mathbf{S} \quad \text{and} \quad \mu_k \in \begin{cases} N_X(x_k) \cap \mathbf{S}, & \text{if } x_k \notin \text{int} X, \\ \{0\}, & \text{if } x_k \in \text{int} X, \end{cases} \tag{3.1}$$

and update x_{k+1} by

$$x_{k+1} := P_X(x_k - v_k(g_k + \mu_k)). \tag{3.2}$$

When solving quasi-convex optimization problems, the difference between the conditional subgradient method and the subgradient methods proposed in [10, 13] is that a conditional subgradient, consisting of a quasi-subgradient and a normal vector to the constraint set, is employed in Algorithm 3.1; while only the quasi-subgradient is used in the algorithms studied in [10, 13].

The stepsize rule has a critical effect on the convergence behavior and computational performance of subgradient methods. In this paper, we consider the following typical stepsize rules.

(a) *Constant stepsize rule:*

$$v_k = v(> 0) \quad \text{for any } k \in \mathbb{N}.$$

(b) *Diminishing stepsize rule:*

$$v_k > 0, \quad \lim_{k\to\infty} v_k = 0, \quad \sum_{k=0}^{\infty} v_k = +\infty. \tag{3.3}$$

(c) *Dynamic stepsize rule:*

$$v_k = \frac{\gamma_k}{4}\left(\frac{f(x_k) - f_*}{L}\right)^{\frac{1}{p}}, \quad \text{where } 0 < \underline{\gamma} \le \gamma_k \le \overline{\gamma} < 2. \tag{3.4}$$

Remark 3.1. The conditional subgradient method was introduced in [17] to solve a constrained convex optimization problem, where the subgradient in (1.2) is replaced by a conditional subgradient, which is an arbitrary vector in the conditional subdifferential

$$\partial_X f(x) := \{g \in \mathbb{R}^n : \langle g, y - x\rangle \le f(y) - f(x), \forall y \in X\}.$$

Essentially, in the context of convex optimization, the conditional subgradient method in [17] can be considered as the subgradient method for solving an associated problem

$$\begin{array}{ll} \min & f_X(x) := f(x) + \delta_X(x) \\ \text{s.t.} & x \in X. \end{array} \tag{3.5}$$

where δ_X is an indicator function of X. Indeed, it is revealed in [17] (or [23]) that

$$\partial_X f(x) = \partial f(x) + N_X(x) = \partial f_X(x) \quad \text{for any } x \in X.$$

In the context of quasi-convex optimization, we define the conditional quasi-subdifferential by

$$\partial_X^* f(x) := \{g \in \mathbb{R}^n : \langle g, y - x\rangle \le 0, \forall y \in \mathrm{lev}_{<f(x)} f \cap X\}.$$

Following from [23, Corollary 23.8.1], it also holds, for any $x \in X$, that

$$\partial_X^* f(x) = N_{\mathrm{lev}_{<f(x)} f\cap X}(x) = N_{\mathrm{lev}_{<f(x)} f}(x) + N_X(x) = \partial^* f(x) + N_X(x) = \partial^* f_X(x).$$

However, the convergence analysis of the subgradient method [10, 13] does not work when it is applied to solve quasi-convex optimization problem (3.5). This is because the objective function f_X may not be continuous, and so Lemma 2.3 and the basic inequality are not necessarily satisfied in this situation. The failure of convergence theory is also observed in a concrete example. Indeed, in Example 1.1, the quasi-subdifferential of f_X at any point in X is $\partial^* f_X = \mathbb{R}_+ \times \mathbb{R}$. When the quasi-subgradient is selected as $g = (0,1)^\top \in \partial^*_X f$, then $x_1 = P_X(x_0 - vg) = x_0$. Hence, a fixed sequence is generated by the subgradient method for solving quasi-convex optimization problem (3.5), and the convergence fails.

Fortunately, Algorithm 3.1 is only a special case of the subgradient method for solving the associated quasi-convex optimization problem (3.5), and the convergence analysis of Algorithm 3.1 will work under the use of some suitable stepsize rules, as shown in the remainder of this section. Indeed, the search direction adopted in Algorithm 3.1 is $g_k + \mu_k$, where $\mu_k \in N_X(x_k)$ but g_k is a unit vector in $\partial^* f(x_k)$. This excludes the extreme case when the search direction falls in $\{0\} + N_X(x)$. A nonzero vector in $\partial^* f(x_k)$, involved in the search direction, maintains the descent property of Algorithm 3.1, which plays an important role in the convergence analysis of the conditional subgradient method for solving quasi-convex optimization problem (1.1).

We now start the convergence analysis of the conditional subgradient method by providing the following basic inequality, which shows a significant property of a conditional subgradient iteration.

Lemma 3.2. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1. Fix $k \in \mathbb{N}$. If $x_k \notin X^*$, then it holds, for any $x^* \in X^*$, that*

$$\|x_{k+1} - x^*\|^2 \le \|x_k - x^*\|^2 - 2v_k \left(\frac{f(x_k) - f_*}{L}\right)^{\frac{1}{p}} + 4v_k^2. \tag{3.6}$$

Proof. Fix $x^* \in X^*$. In view of Algorithm 3.1 (cf. (3.2)), it follows from the nonexpansive property of projection operator that

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\le \|x_k - v_k(g_k + \mu_k) - x^*\|^2 \\ &= \|x_k - x^*\|^2 - 2v_k \langle g_k + \mu_k, x_k - x^* \rangle + v_k^2 \|g_k + \mu_k\|^2. \end{aligned} \tag{3.7}$$

By the assumption that $x_k \notin X^*$, Lemma 2.3 is applicable (to x_k, g_k in place of x, $g(x)$) to concluding that

$$\langle g_k, x_k - x^* \rangle \ge \left(\frac{f(x_k) - f_*}{L}\right)^{\frac{1}{p}}.$$

By the second inclusion of (3.1) and by (2.1), one has that $\langle \mu_k, x^* - x_k \rangle \le 0$. Note also by (3.1) that $\|g_k + \mu_k\|^2 \le (\|g_k\| + \|\mu_k\|)^2 \le 4$. These, together with (3.7), imply (3.6). The proof is complete. $\square$

When the constant stepsize is adopted, we establish the convergence of the conditional subgradient method to the minimum value of (1.1) within some tolerance, given in terms of the stepsize.

Theorem 3.3. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1 with the constant stepsize rule. Then*

$$\liminf_{k\to\infty} f(x_k) \le f_* + L(2v)^p. \tag{3.8}$$

Proof. We prove by contradiction, assuming that

$$\liminf_{k\to\infty} f(x_k) > f_* + L(2v)^p.$$

Consequently, there exist some $\sigma > 0$ and $k_0 \in \mathbb{N}$ such that

$$f(x_k) > f_* + L(2v+\sigma)^p \quad \text{for any } k \ge k_0.$$

Hence $x_k \notin X^*$ for any $k \ge k_0$, and then Lemma 3.2 is applicable (to v in place of v_k) to concluding that

$$\|x_{k+1} - x^*\|^2 \le \|x_k - x^*\|^2 - 2v\left(\frac{f(x_k)-f_*}{L}\right)^{\frac{1}{p}} + 4v^2 < \|x_k - x^*\|^2 - 2v\sigma.$$

Summing the above inequality over $k = k_0, \ldots, n$, we obtain that

$$\|x_{n+1} - x^*\|^2 \le \|x_{k_0} - x^*\|^2 - 2(n-k_0+1)v\sigma,$$

which yields a contradiction for a sufficiently large n. The proof is complete. □

If the stepsize tends to zero, then the tolerance in (3.8) vanishes. Below, we present convergence results of the conditional subgradient method, in objective values and distances from the set of minima of (1.1), when the diminishing stepsize is adopted.

Theorem 3.4. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1 with the diminishing stepsize rule* (3.3). *Then the following assertions are true:*

(i) $\liminf_{k\to\infty} f(x_k) = f_*$.

(ii) *If f is coercive, then*

$$\lim_{k\to\infty} \operatorname{dist}(x_k, X^*) = 0 \quad \text{and} \quad \lim_{k\to\infty} f(x_k) = f_*.$$

(iii) *If $\sum_{k=0}^{\infty} v_k^2 < \infty$, then $\{x_k\}$ converges to an optimal solution of* (1.1).

Proof. The proof utilizes the basic inequality (3.6) of Algorithm 3.1 and the property of the diminishing stepsize rule (3.3).

(i) Summing (3.6) over $k = 0, 1, \ldots, n-1$, one has

$$\|x_n - x^*\|^2 \le \|x_0 - x^*\|^2 - 2L^{-\frac{1}{p}} \sum_{k=0}^{n-1} v_k (f(x_k) - f_*)^{\frac{1}{p}} + 4\sum_{k=0}^{n-1} v_k^2, \tag{3.9}$$

which implies that

$$\sum_{k=0}^{n-1} \frac{v_k(f(x_k)-f_*)^{\frac{1}{p}}}{\sum_{k=0}^{n-1} v_k} \le \frac{L^{\frac{1}{p}}\|x_0-x^*\|^2}{2\sum_{k=0}^{n-1} v_k} + \frac{2L^{\frac{1}{p}}\sum_{k=0}^{n-1} v_k^2}{\sum_{k=0}^{n-1} v_k}.$$

Then, by Lemma 2.4 (with $(f(x_k)-f_*)^{\frac{1}{p}}$ in place of a_k), it follows that

$$\begin{aligned} \liminf_{k\to\infty}\,(f(x_k)-f_*)^{\frac{1}{p}} &\le \liminf_{n\to\infty} \sum_{k=0}^{n-1} \frac{v_k(f(x_k)-f_*)^{\frac{1}{p}}}{\sum_{k=0}^{n-1} v_k} \\ &\le \liminf_{n\to\infty} \left(L^{\frac{1}{p}} \frac{\|x_0-x^*\|^2}{2\sum_{k=0}^{n-1} v_k} + 2L^{\frac{1}{p}} \frac{\sum_{k=0}^{n-1} v_k^2}{\sum_{k=0}^{n-1} v_k} \right). \end{aligned} \tag{3.10}$$

By (3.3) (in particular, $\sum_{k=0}^{\infty} v_k = +\infty$), one has that

$$\lim_{n\to\infty} \frac{\|x_0-x^*\|^2}{2\sum_{k=0}^{n-1} v_k} = 0. \tag{3.11}$$

Also by (3.3) (in particular, $\lim_{n\to\infty} v_n = 0$), we obtain from Lemma 2.4 (with v_k^2 in place of a_k) that

$$\liminf_{n\to\infty} \frac{\sum_{k=0}^{n-1} v_k^2}{\sum_{k=0}^{n-1} v_k} \le \lim_{k\to\infty} v_k = 0. \tag{3.12}$$

These, together with (3.10), conclude that $\liminf_{k\to\infty} f(x_k) = f_*$.

(ii) Fix $\sigma > 0$. Note that $\lim_{n\to\infty} v_n = 0$. We can let $k_0 \in \mathbb{N}$ be such that

$$v_k \le \tfrac{1}{4}\sigma^{\frac{1}{p}} \quad \text{for any } k \ge k_0. \tag{3.13}$$

We consider the estimation of x_k ($k \ge k_0$) in the following two cases:

Case 1: $f(x_k) > f_* + L\sigma$. Then $x_k \notin X^*$, and so Lemma 3.2 is applicable to concluding, for any $x^* \in X^*$, that

$$\|x_{k+1}-x^*\|^2 \le \|x_k-x^*\|^2 - 2v_k\sigma^{\frac{1}{p}} + 4v_k^2 \le \|x_k-x^*\|^2 - 4v_k^2$$

(due to (3.13)), and hence

$$\operatorname{dist}(x_{k+1}, X^*) \le \operatorname{dist}(x_k, X^*) - 4v_k^2. \tag{3.14}$$

That is, $\operatorname{dist}(x_k, X^*)$ is decreasing in this case.

Case 2: $f(x_k) \le f_* + L\sigma$. It follows from assertion (i) that this case must occur for infinitely many k. Define

$$X_\sigma := X \cap \operatorname{lev}_{\le f_*+L\sigma} f \quad \text{and} \quad \rho(\sigma) := \max_{x\in X_\sigma} \operatorname{dist}(x, X^*).$$

By the assumption that f is coercive, it follows that its sublevel set $\operatorname{lev}_{\le f_*+L\sigma} f$ is bounded, and so $\rho(\sigma) < \infty$. Since $f(x_k) \le f_* + L\sigma$, one has that $\operatorname{dist}(x_k, X^*) \le \rho(\sigma)$. In view of Algorithm 3.1, for any $x^* \in X^*$, we obtain that

$$\|x_{k+1}-x^*\| \le \|x_k - v_k(g_k+\mu_k) - x^*\| \le \|x_k-x^*\| + 2v_k,$$

and thus

$$\operatorname{dist}(x_{k+1}, X^*) \le \operatorname{dist}(x_k, X^*) + 2v_k \le \rho(\sigma) + 2v_k. \tag{3.15}$$

By (3.14) and (3.15) (in both cases), we conclude that

$$\operatorname{dist}(x_k, X^*) \le \rho(\sigma) + 2v_k \quad \text{for any } k \ge k_0. \tag{3.16}$$

Since f is continuous and coercive, its sublevel sets are compact, and so, it is trivial to see that $\lim_{\sigma\to 0} \rho(\sigma) = 0$. Hence, it follows from (3.16) that $\lim_{k\to\infty} \operatorname{dist}(x_k, X^*) = 0$, and so $\lim_{k\to\infty} f(x_k) = f_*$ (by the continuity of f).

(iii) When $\sum_{k=0}^{\infty} v_k^2 < \infty$, one sees from (3.9) that $\{\|x_k - x^*\|^2\}$ is bounded, and hence $\{x_k\}$ is bounded. Since further $\liminf_{k\to\infty} f(x_k) = f_*$ (proved in (i)), it follows that $\{x_k\}$ must have a cluster point $\bar{x} \in X^*$. Noting that $\lim_{n\to\infty} \sum_{k=n}^{\infty} v_k^2 = 0$, we conclude by (3.6) that $\{\|x_k - \bar{x}\|^2\}$ is a Cauchy sequence, and thus it converges to 0. Therefore $\{x_k\}$ converges to such $\bar{x}$.

The proof is complete. □

Given the prior information of f_*, the dynamic stepsize rule is usually considered in the literature of subgradient methods; see, e.g., [12, 17, 20, 24]. Below, we show the convergence of the conditional subgradient method to an optimal solution of (1.1), when the dynamic stepsize is adopted.

Theorem 3.5. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1 with the dynamic stepsize rule* (3.4)*. Then $\{x_k\}$ converges to an optimal solution of* (1.1)*.*

Proof. Fix $x^* \in X^*$. By using the dynamic stepsize rule (cf. (3.4)), it follows from Lemma 3.2 that

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\le \|x_k - x^*\|^2 - 2v_k \left(\tfrac{f(x_k)-f_*}{L}\right)^{\frac{1}{p}} + 4v_k^2 \\ &= \|x_k - x^*\|^2 - \tfrac{1}{4}\gamma_k(2-\gamma_k)\left(\tfrac{f(x_k)-f_*}{L}\right)^{\frac{2}{p}} \\ &\le \|x_k - x^*\|^2 - \tfrac{1}{4}\underline{\gamma}(2-\overline{\gamma})\left(\tfrac{f(x_k)-f_*}{L}\right)^{\frac{2}{p}} \end{aligned} \tag{3.17}$$

(which automatically holds when $x_k \in X^*$, since $f(x_k) = f_*$, $v_k = 0$ and $x_{k+1} = x_k$ in this situation). The sequence $\{\|x_k - x^*\|\}$ is decreasing, and hence $\{x_k\}$ is bounded. It also follows that $\lim_{k\to\infty} f(x_k) = f_*$; otherwise, it follows from (3.17) that there exists some $\sigma > 0$ such that $\|x_{k+1} - x^*\|^2 \le \|x_k - x^*\|^2 - \sigma$ occurs for infinitely many k, which is impossible. Hence, any cluster point $\bar{x}$ of $\{x_k\}$ is an optimal solution of (1.1), that is, $\bar{x} \in X^*$.

Since further the sequence $\{\|x_k - x^*\|\}$ is decreasing, it must converge to $\|\bar{x} - x^*\|$. If there are two distinct cluster points of $\{x_k\}$, namely $\bar{x}$ and $\tilde{x}$, we have $\bar{x} \in X^*$, $\tilde{x} \in X^*$, and $\|\bar{x} - x^*\| = \|\tilde{x} - x^*\|$ for any $x^* \in X^*$; consequently, we conclude that $\bar{x} = \tilde{x}$. Therefore, $\{x_k\}$ converges to an optimal solution $\bar{x} \in X^*$. The proof is complete. □

At the end of this section, we present the finite convergence behavior of the conditional subgradient method to the set of minima X^* of problem (1.1) under the assumption that X^* has a nonempty interior.

Theorem 3.6. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1. Let $x^* \in X^*$ and $\sigma > 0$, and suppose that $\mathbf{B}(x^*, \sigma) \subseteq X^*$. Then $x_k \in X^*$ for some $k \in \mathbb{N}$, provided one of the following conditions:*

(i) *$v_k = v \in (0, \frac{\sigma}{2})$ for any $k \in \mathbb{N}$.*
(ii) *$\{v_k\}$ satisfies the diminishing stepsize rule* (3.3)*.*

Proof. We prove by contradiction, assuming that $f(x_k) > f_*$ for any $k \in \mathbb{N}$. Fix $k \in \mathbb{N}$. Since $\mathbf{B}(x^*, \sigma) \subseteq X^*$, one sees that $\mathbf{B}(x^*, \sigma) \subseteq X \cap \mathrm{lev}_{<f(x_k)} f$. Let $\theta \in (0, 1)$. Since $\|g_k\| = 1$, we obtain that $x^* + \theta\sigma g_k \in \mathrm{lev}_{<f(x_k)} f$, and then it follows from Definition 2.1 that $\langle g_k, x^* + \theta\sigma g_k - x_k\rangle \le 0$, that is, $\langle g_k, x_k - x^*\rangle \ge \theta\sigma$. Since $\theta \in (0, 1)$ is arbitrary, it follows that

$$\langle g_k, x_k - x^*\rangle \ge \sigma. \tag{3.18}$$

Note by definition that $\langle \mu_k, x^* - x_k\rangle \le 0$. This, together with (3.18), implies that

$$\langle g_k + \mu_k, x_k - x^*\rangle \ge \sigma. \tag{3.19}$$

On the other hand, summing (3.7) over $k = 0, \ldots, n$, one has

$$\frac{\sum_{k=0}^n v_k \langle g_k + \mu_k, x_k - x^*\rangle}{\sum_{k=0}^n v_k} \le \frac{\|x_0 - x^*\|^2}{2\sum_{k=0}^n v_k} + \frac{2\sum_{k=0}^n v_k^2}{\sum_{k=0}^n v_k}. \tag{3.20}$$

We now claim, under the assumption of (i) or (ii), that

$$\liminf_{n\to\infty} \langle g_k + \mu_k, x_k - x^*\rangle < \sigma. \tag{3.21}$$

(i) When a constant stepsize $v \in (0, \frac{\sigma}{2})$ is used, (3.20) is reduced to

$$\frac{\sum_{k=0}^n v_k \langle g_k + \mu_k, x_k - x^*\rangle}{\sum_{k=0}^n v_k} \le \frac{\|x_0 - x^*\|^2}{2nv} + 2v,$$

and thus, by Lemma 2.4, we obtain that

$$\liminf_{n\to\infty} \langle g_k + \mu_k, x_k - x^*\rangle \le \liminf_{n\to\infty} \frac{\sum_{k=0}^n v_k \langle g_k + \mu_k, x_k - x^*\rangle}{\sum_{k=0}^n v_k} \le 2v < \sigma.$$

(ii) When a diminishing stepsize is used, by (3.11) and (3.12), it follows from Lemma 2.4 and (3.20) that

$$\liminf_{n\to\infty} \langle g_k + \mu_k, x_k - x^*\rangle \le 0 < \sigma.$$

Hence we have proved (3.21) under the assumption of (i) or (ii), which arrives at a contradiction with (3.18). The proof is complete. □

4. Conditional ϵ-subgradient method and convergence analysis

In many applications, the computation error stems from practical considerations, and is inevitable in the computing process. Usually, the computation error gives rise to the calculation of the ϵ-subgradient. To meet the requirement of applications, this section is devoted to the study of the conditional ϵ-subgradient method for solving constrained quasi-convex optimization problem (1.1), where an ϵ-quasi-subgradient is employed in place of the quasi-subgradient as in Algorithm 3.1. Hence the conditional ϵ-subgradient method is formally described as follows.

Algorithm 4.1. Select an initial point $x_0 \in \mathbb{R}^n$, a sequence of stepsizes $\{v_k\} \subseteq (0, +\infty)$ and a sequence of errors $\{\epsilon_k\} \subseteq (0, +\infty)$. Having x_k, we calculate

$$g_k \in \partial^*_{\epsilon_k} f(x_k) \cap \mathbf{S} \quad \text{and} \quad \mu_k \in \begin{cases} N_X(x_k) \cap \mathbf{S}, & \text{if } x_k \notin \text{int}X, \\ \{0\}, & \text{if } x_k \in \text{int}X, \end{cases} \tag{4.1}$$

and update x_{k+1} by

$$x_{k+1} := P_X(x_k - v_k(g_k + \mu_k)). \tag{4.2}$$

Assuming that the computational error is deterministic and bounded, we investigate the influence of the computation error on the proposed conditional ϵ-subgradient method. For this purpose, we first present an important property in the following lemma, which is a key to establish the basic inequality in the convergence analysis.

Lemma 4.1. *Let $x \in X$ be such that $f(x) > f_* + \epsilon$, and let $g(x, \epsilon)$ be a unit ϵ-quasi-subgradient of f at x, i.e., $g(x, \epsilon) \in \partial^*_\epsilon f(x) \cap \mathbf{S}$. Then it holds, for any $x^* \in X^*$, that*

$$\langle g(x, \epsilon), x - x^* \rangle \geq \left(\frac{f(x) - f_* - \epsilon}{L} \right)^{\frac{1}{p}}.$$

Proof. By assumptions that f is quasi-convex and continuous and that $f(x) > f_* + \epsilon$, it follows that its strict sublevel set $\text{lev}_{<f(x)-\epsilon} f$ is nonempty, open and convex. Given $x^* \in X^*$, we define

$$r := \inf \left\{ \|y - x^*\| : y \in \text{bd}\left(\text{lev}_{<f(x)-\epsilon} f\right) \right\}. \tag{4.3}$$

One has by the Hölder condition that

$$f(y) - f_* \leq L\text{dist}^p(y, X^*) \quad \text{for any } y \in \mathbb{R}^n.$$

This, by taking the infimun over $\text{bd}\left(\text{lev}_{<f(x)-\epsilon} f\right)$, implies that

$$f(x) - \epsilon - f_* \leq L \inf \left\{ \text{dist}^p(y, X^*) : y \in \text{bd}\left(\text{lev}_{<f(x)-\epsilon} f\right) \right\} \leq Lr^p. \tag{4.4}$$

Let $\theta \in (0, 1)$. Since $\|g(x, \epsilon)\| = 1$, one has by (4.3) that $x^* + \theta r g(x, \epsilon) \in \text{lev}_{<f(x)-\epsilon} f$, and we obtain by Definition 2.1 that

$$\langle g(x, \epsilon), x^* + \theta r g(x, \epsilon) - x \rangle \leq 0,$$

that is, $\langle g(x, \epsilon), x - x^* \rangle \geq \theta r$. Since $\theta \in (0, 1)$ is arbitrary, it follows that $\langle g(x, \epsilon), x - x^* \rangle \geq r$. This, together with (4.4), implies that

$$\langle g(x, \epsilon), x - x^* \rangle \geq \left(\frac{f(x) - f_* - \epsilon}{L} \right)^{\frac{1}{p}}.$$

The proof is complete. □

Lemma 4.2. *Let $\{x_k\}$ be a sequence generated by Algorithm 4.1. Fix $k \in \mathbb{N}$. If x_k is such that $f(x_k) > f_* + \epsilon_k$, then it holds, for any $x^* \in X^*$, that*

$$\|x_{k+1} - x^*\|^2 \le \|x_k - x^*\|^2 - 2v_k \left(\frac{f(x_k) - f_* - \epsilon_k}{L}\right)^{\frac{1}{p}} + 4v_k^2. \tag{4.5}$$

Proof. The proof adopts a line of analysis similar to that of Lemma 3.2. Since $f(x_k) > f_* + \epsilon_k$, Lemma 4.1 is applicable (to x_k, g_k, ϵ_k in place of x, $g(x, \epsilon)$, ϵ) to concluding that

$$\langle g_k, x_k - x^* \rangle \ge \left(\frac{f(x_k) - f_* - \epsilon_k}{L}\right)^p.$$

This, together with (3.7), deduces (4.5). The proof is complete. □

By virtue of Lemma 4.2, we establish the convergence of the conditional ϵ-subgradient method to the minimum value of (1.1) within some tolerance, which is expressed in terms of the error and the stepsize, under the assumption that the computational error is deterministic and bounded.

Theorem 4.3. *Let $\{x_k\}$ be a sequence generated by Algorithm 4.1 with the constant stepsize rule and $\limsup_{k\to\infty} \epsilon_k = \epsilon$. Then*

$$\liminf_{k\to\infty} f(x_k) \le f_* + L(2v)^p + \epsilon.$$

Proof. We prove by contradiction, assuming that

$$\liminf_{k\to\infty} f(x_k) > f_* + (2v)^p + \epsilon.$$

Since $\limsup_{k\to\infty} \epsilon_k = \epsilon$, there exist some $\sigma > 0$ and $k_0 \in \mathbb{N}$ such that

$$f(x_k) > f_* + L(2v + \sigma)^p + \epsilon_k \quad \text{for any } k \ge k_0.$$

Hence, by Lemma 4.2 (with v in place of v_k), it follows for any $k \ge k_0$ that

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\le \|x_k - x^*\|^2 - 2v \left(\tfrac{f(x_k) - f_* - \epsilon_k}{L}\right)^{\frac{1}{p}} + 4v^2 \\ &< \|x_k - x^*\|^2 - 2v\sigma. \end{aligned}$$

Summing the above inequality over $k = k_0, \ldots, n$, we obtain

$$\|x_{n+1} - x^*\|^2 \le \|x_{k_0} - x^*\|^2 - 2(n - k_0 + 1)v\sigma,$$

which yields a contradiction for a sufficiently large n. The proof is complete. □

Theorem 4.4. *Let $\{x_k\}$ be a sequence generated by Algorithm 4.1 with the diminishing stepsize rule* (3.3) *and $\limsup_{k\to\infty} \epsilon_k = \epsilon$. Then the following assertions are true:*

(i) $\liminf_{k\to\infty} f(x_k) \le f_* + \epsilon$.
(ii) *If f is coercive, then*

$$\liminf_{k\to\infty} \operatorname{dist}(x_k, \operatorname{lev}_{\le f_*+\epsilon} \cap X) = 0. \tag{4.6}$$

Let $\rho(\epsilon) := \max\{\mathrm{dist}(x, X^*) : x \in \mathrm{lev}_{\le f_*+\epsilon} \cap X\}$. *Then*

$$\lim_{k\to\infty} \mathrm{dist}(x_k, X^* + \rho(\epsilon)\mathbf{B}) = 0. \tag{4.7}$$

Proof. Taking a line of analysis similar to that of Theorem 4.3 and the property of diminishing stepsize rule (3.3), we can prove assertion (i). Then it remains the proof of assertion (ii).

Assertion (i) says that there exists a subsequence $\{x_{k_i}\}$ such that

$$\lim_{i\to\infty} f(x_{k_i}) \le f_* + \epsilon. \tag{4.8}$$

By the assumption that f is coercive, it follows that $\{x_{k_i}\}$ is bounded. Without loss of generality, we can assume that $\{x_{k_i}\}$ converges to some $\bar{x} \in X$; otherwise, one can select a subsequence of $\{x_{k_i}\}$ such that it converges to some $\bar{x} \in X$ and satisfies (4.8). Hence it follows from [10, Lemma 3.4] that $\mathrm{dist}(\bar{x}, \mathrm{lev}_{\le f_*+\epsilon}) = 0$. Since further $\bar{x} \in X$, then $\mathrm{dist}(\bar{x}, \mathrm{lev}_{\le f_*+\epsilon} \cap X) = 0$, and so (4.6) is proved.

To prove (4.7), we fix $\sigma > 0$ and define

$$V_{2\sigma} := X^* + \rho(\epsilon)\mathbf{B} + 2\sigma\mathbf{B},$$

and

$$e_\sigma := \inf\{f(x) : x \in X, \mathrm{dist}(x, \mathrm{lev}_{\le f_*+\epsilon} f \cap X) \ge \sigma\} - (f_* + \epsilon). \tag{4.9}$$

Then $e_\sigma > 0$. Indeed, if $e_\sigma = 0$, then there exists a sequence $\{z_i\}$, in $\{x \in X : \mathrm{dist}(x, \mathrm{lev}_{\le f_*+\epsilon} f \cap X) \ge \sigma\}$, such that $\lim_{i\to\infty} f(z_i) = f_* + \epsilon$. Thus, by the arguments as we did for (4.6), we conclude that some cluster point $\bar{z}$ of $\{z_i\}$ satisfies that $\mathrm{dist}(\bar{z}, \mathrm{lev}_{\le f_*+\epsilon} f \cap X) = 0$, which is impossible as $\sigma > 0$.

Note that $\limsup_{k\to\infty} \epsilon_k = \epsilon$, $\lim_{k\to\infty} v_k = 0$ and $\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0$. Then there exists some $k_\sigma \in \mathbb{N}$ such that

$$\epsilon_k < \epsilon + \frac{e_\sigma}{2}, \quad v_k \le \frac{1}{2}\left(\frac{e_\sigma}{2L}\right)^{\frac{1}{p}}, \quad \text{and} \quad \|x_{k+1} - x_k\| \le \sigma, \tag{4.10}$$

for any $k \ge k_\sigma$. Note by definition that $X^* \subseteq \mathrm{lev}_{\le f_*+\epsilon} f \cap X \subseteq X^* + \rho(\epsilon)\mathbf{B}$. Then, by (4.6), one sees that there exists some $k'_\sigma \ge k_\sigma$ such that

$$x_{k'_\sigma} \in (\mathrm{lev}_{\le f_*+\epsilon} f \cap X) + \sigma\mathbf{B} \subseteq X^* + \rho(\epsilon)\mathbf{B} + \sigma\mathbf{B} \subseteq V_{2\sigma},$$

that is, $x_{k'_\sigma} \in V_{2\sigma}$.

Next, we claim that $x_k \in V_{2\sigma}$ for any $k \ge k'_\sigma$. Proving by induction, we assume that $x_k \in V_{2\sigma}$ for some $k \ge k'_\sigma$ and consider the following two cases.

Case 1. If $\mathrm{dist}(x_k, \mathrm{lev}_{\le f_*+\epsilon} f \cap X) \le \sigma$, by the third inequality of (4.10), we have

$$x_{k+1} \in \{x_k\} + \sigma\mathbf{B} \subseteq (\mathrm{lev}_{\le f_*+\epsilon} f \cap X + \sigma\mathbf{B}) + \sigma\mathbf{B} \subseteq X^* + \rho(\epsilon)\mathbf{B} + 2\sigma\mathbf{B} = V_{2\sigma}.$$

Case 2. If $\mathrm{dist}(x_k, \mathrm{lev}_{\le f_*+\epsilon} f \cap X) > \sigma$, it follows from (4.9) and (4.10) that

$$f(x_k) \ge e_\sigma + f_* + \epsilon > f_* + \epsilon_k + \frac{e_\sigma}{2}.$$

Then Lemma 4.2 is applicable to concluding that

$$\|x_{k+1} - x^*\|^2 < \|x_k - x^*\|^2 - 2v_k\left(\left(\frac{e_\sigma}{2L}\right)^{\frac{1}{p}} - 2v_k\right) \leq \|x_k - x^*\|^2$$

(thanks to the second inequality of (4.10)). Thus, in both cases, $x_k \in V_{2\sigma}$ implies that $x_{k+1} \in V_{2\sigma}$. Hence, by induction, $x_k \in V_{2\sigma}$, and so $\mathrm{dist}(x_k, X^* + \rho(\epsilon)\mathbf{B}) \leq 2\sigma$, for any $k \geq k'_\sigma$. Since $\sigma > 0$ is arbitrary, then (4.7) is achieved, and the proof is complete. □

Remark 4.5. (i) Theorems 4.3 and 4.4 show the convergence behavior to the minimum value of (1.1) within some tolerance by using the constant and diminishing stepsize rules, respectively. In particular, it is exhibited by Theorem 4.3 that the tolerance has an additive form, including the computation error ϵ and the constant stepsize v, which is of a similar formula as in [10] (where a noise is additionally considered). When the diminishing stepsize is adopted, the stepsize tends to zero, and so the term involving the stepsize vanishes in the tolerance, which is indicated by Theorem 4.4.

(ii) In the convergence analysis of conditional ϵ-subgradient method, we assume that the computation error satisfies $\limsup_{k\to\infty} \epsilon_k = \epsilon$, which is not necessarily to be vanishing. While [18] only considered the case when $\lim_{k\to\infty} \epsilon_k = 0$ and only the diminishing stepsize is adopted in the category of convex optimization. In particular, in the case when $\epsilon = 0$, Theorems 4.3 and 4.4 are applicable to concluding the exact convergence results of the conditional ϵ-subgradient method in the category of quasi-convex optimization, which cover and extend the results obtained in [18].

References

[1] D. Aussel and A. Daniilidis, *Normal characterization of the main classes of quasiconvex functions*, Set-Valued Anal. **8** (2000), 219–236.

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

[3] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Cambridge, 1999.

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

[5] Y. M. Ermoliev, *Methods of solution of nonlinear extremal problems*, Cybern. Syst. Anal. **2** (1996), 1–14.

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

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

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

[9] Y. H. Hu, C.-K. Sim and X. Q. Yang, *A subgradient method based on gradient sampling for solving convex optimization problems*, Numer. Func. Anal. Opt. **36** (2015), 1559–1584.

[10] Y. H. Hu, X. Q. Yang and C.-K. Sim, *Inexact subgradient methods for quasi-convex optimization problems*, European J. Oper. Res. **240** (2015), 315–327.

[11] Y. H. Hu, X. Q. Yang and C. K. W. Yu, *Subgradient methods for saddle point problems of quasiconvex optimization*, Pure and Applied Functional Analysis, to appear.

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

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

[14] K. C. Kiwiel, *Convergence of approximate and incremental subgradient methods for convex optimization*, SIAM J. Optim. **14** (2004), 807–840.

[15] I. V. Konnov, *On properties of supporting and quasi-supporting vectors*, J. Math. Sci. **71** (1994), 2760–2763.

[16] I. V. Konnov, *On convergence properties of a subgradient method*, Optim. Method. Softw. **18** (2003), 53–62.

[17] T. Larsson, M. Patriksson and A.-B. Strömberg, *Conditional subgradient optimization – Theory and applications*, European J. Oper. Res. **88** (1996), 382–403.

[18] T. Larsson, M. Patriksson and A.-B. Strömberg, *On the convergence of conditional ϵ-subgradient methods for convex programs and convex-concave saddle-point problems*, European J. Oper. Res. **151** (2003), 461–473.

[19] A. Nedić and D. P. Bertsekas, *Incremental subgradient methods for nondifferentiable optimization*, SIAM J. Optim. **12** (2001), 109–138.

[20] A. Nedić and D. P. Bertsekas, *The effect of deterministic noise in subgradient methods*, Math. Program. **125** (2010), 75–99.

[21] Y. Nesterov, *Primal-dual subgradient methods for convex problems*, Math. Program. **120** (2009), 221–259.

[22] B. T. Polyak, *A general method for solving extremum problems*, Soviet Mathematics Doklady **8** (1967), 593–597.

[23] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

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

[25] I. M. Stancu-Minasian, *Fractional Programming*, Kluwer Academic Publishers, Dordrecht, 1997.

(Y. H. Hu) College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, P. R. China
E-mail address: mayhhu@szu.edu.cn

(Carisa K. W. Yu) Department of Mathematics and Statistics, Hang Seng Management College, Hong Kong
E-mail address: carisayu@hsmc.edu.hk

(C. Li) Department of Mathematics, Zhejiang University, Hangzhou 310027, P. R. China
E-mail address: cli@zju.edu.cn

(X. Q. Yang) Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong
E-mail address: mayangxq@polyu.edu.hk