



Continuous Optimization

Inexact subgradient methods for quasi-convex optimization problems

Yaohua Hu^a, Xiaoqi Yang^{b,*}, Chee-Khian Sim^c^a Department of Mathematics, Zhejiang University, Hangzhou 310027, PR China^b Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong^c Department of Decision Sciences, National University of Singapore, 15 Kent Ridge Drive, Singapore 119245, Singapore

ARTICLE INFO

Article history:

Received 8 April 2013

Accepted 12 May 2014

Available online 20 May 2014

Keywords:

Subgradient method

Quasi-convex optimization

Noise

Weak sharp minima

ABSTRACT

In this paper, we consider a generic inexact subgradient algorithm to solve a nondifferentiable quasi-convex constrained optimization problem. The inexactness stems from computation errors and noise, which come from practical considerations and applications. Assuming that the computational errors and noise are deterministic and bounded, we study the effect of the inexactness on the subgradient method when the constraint set is compact or the objective function has a set of generalized weak sharp minima. In both cases, using the constant and diminishing stepsize rules, we describe convergence results in both objective values and iterates, and finite convergence to approximate optimality. We also investigate efficiency estimates of iterates and apply the inexact subgradient algorithm to solve the Cobb–Douglas production efficiency problem. The numerical results verify our theoretical analysis and show the high efficiency of our proposed algorithm, especially for the large-scale problems.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Subgradient methods are popular and practical techniques used to minimize a nondifferentiable convex function. Subgradient methods originated with the works of Polyak (1967) and Ermoliev (1966) and were further developed by Shor, Kiwiel, and Ruszczyński (1985). In the last 40 years, many properties of subgradient methods have been discovered, generalizations and extensions have been proposed, and various applications have been found (see Auslender & Teboulle, 2004; Bertsekas, Nedić, & Ozdaglar, 2003; Hiriart-Urruty & Lemaréchal, 1996; Larsson, Patriksson, & Strömberg, 1996; Nedić & Bertsekas, 2001; Nesterov, 2009; Patriksson, 2008; Shor et al., 1985 and references therein). Nowadays, the subgradient method still remains an important tool for nonsmooth and stochastic optimization problems, special for large-scale problems, due to its simple formulation and low storage requirement.

Motivated by practical reasons, approximate subgradient methods (also called ϵ -subgradient methods) are widely studied in Auslender and Teboulle (2004), D'Antonio and Frangioni (2009), Kiwiel (2004), Larsson, Patriksson, and Strömberg (2003), Shor et al. (1985). Kiwiel (2004) proposed a unified convergence framework for approximate subgradient methods. The author presented

convergence in objective values and convergence to a neighborhood of the optimal solution set, using both the diminishing and nonvanishing stepsize rules. Larsson et al. (2003) proposed and analyzed conditional ϵ -subgradient methods to solve convex optimization problems and convex–concave saddle-point problems. Improving conditional subgradient methods, D'Antonio and Frangioni (2009) combined the deflection and the conditional subgradient technique into one iterative process, and investigated the unified convergence analysis for the deflected conditional approximate subgradient methods, using both the Polyak-type and diminishing stepsize rules. Furthermore, Auslender and Teboulle (2004) proposed and developed an interior ϵ -subgradient method for convex constrained optimization problems over polyhedral sets, in particular \mathbb{R}_+^n , via replacing the Euclidean distance function by a logarithmic-quadratic distance-like function.

Recently, Nedić and Bertsekas (2010) investigated the effect of noise on subgradient methods for convex optimization problems. Their work was motivated by the distributed optimization in networks where the data is quantized before being transmitted between nodes (see Kashyap, Basar, & Srikant, 2007; Rabbat & Nowak, 2005 and references therein). When the constraint set is compact or the objective function has a set of weak sharp minima, the authors established convergence properties to the optimal value within some tolerance, which is expressed in terms of errors and noise, under the bounded subgradient assumption.

Quasi-convex optimization problems can be found in important applications in various areas, such as economics, engineering,

* Corresponding author. Tel.: +852 27666954; fax: +852 23629045.

E-mail addresses: hyh19840428@163.com (Y. Hu), mayangxq@polyu.edu.hk (X. Yang), bizsimc@nus.edu.sg (C.-K. Sim).

management science and various applied sciences (see Avriel, Diewert, Schaible, & Zang, 1988; Crouzeix, Martinez-Legaz, & Volle, 1998; Hadjisavvas, Komlósi, & Schaible, 2005 and references therein). The study of using subgradient methods to solve quasi-convex optimization problems has been limited. Using the diminishing stepsize rule, Kiwiel (2001) studied convergence properties and efficiency estimates of the exact subgradient method for solving a quasi-convex optimization problem under the assumption that the objective function is upper semi-continuous. On the other hand, modified dual subgradient algorithms were investigated in Gasimov (2002) and Burachik, Gasimov, Ismayilova, and Kaya (2006) for solving a general nonconvex optimization problem with equality constraints by virtue of a sharp augmented Lagrangian.

Motivated by practical and theoretical reasons, in this paper, we focus on an inexact subgradient algorithm for solving the following quasi-convex optimization problem:

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

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a quasi-convex function and the constraint set X is nonempty, closed and convex. We denote the optimal solution set and the optimal value respectively by X^* and f_* , and we assume that X^* is nonempty and compact.

Inspired by the idea in Nedić and Bertsekas (2010) and references therein, we investigate the influence of inexact terms, including both computation errors and noise, on the inexact subgradient algorithm. The computation errors, which give rise to the ϵ -subgradient, is inevitable in computing process. On the other hand, the noise may come from practical considerations and applications, and is manifested in inexact computation of subgradients. Considering a generic inexact subgradient algorithm for the quasi-convex optimization problem (1.1) and assuming that the computational errors and noise are deterministic and bounded, we establish convergence properties in both objective values and iterates within some tolerance given explicitly in terms of errors and noise. We also describe the finite convergence behavior to approximate optimality and efficiency estimates of iterates.

The quasi-convex function is more difficult to deal with, as the epigraph of a convex function is convex; while only the sublevel set of a quasi-convex function is convex. Lacking the convexity assumed in Nedić and Bertsekas (2010), the main technical challenges are defining a suitable subdifferential of a quasi-convex function, establishing the proper basic inequality, which is a key tool needed in this area of study, and applying the convexity of the sublevel set instead of that of the epigraph of a convex function, when analyzing the inexact subgradient method algorithm for the quasi-convex optimization problem. To meet these challenges, we adopt the closure of Greenberg–Pierskalla subdifferential as the quasi-convex subdifferential, introduce the Hölder condition to relate the quasi-convex subgradient with objective function values and establish the basic inequality, which is only a local property though, and then obtain the convergence property in objective values and finite convergence under the Hölder condition, instead of the upper semi-continuity of the objective function used in Kiwiel (2001). Another contribution is to describe the convergence property in iterates, which are absent in Nedić and Bertsekas (2010), by virtue of convexity of a sublevel set. When X is noncompact, we need to assume an additional generalized weak sharp minima condition. This condition extends the concept of weak sharp minima in Nedić and Bertsekas (2010) and is presented by using $\text{dist}(x, X^*)$, the distance of the decision variable x to X^* .

We also investigate the quantification of the influence of errors and noise by using both the constant and diminishing stepsize rules, while only the diminishing stepsize rule is considered in studying convergence properties and efficiency estimates of an exact subgradient method in Kiwiel (2001).

We further consider the fractional programming as an application of the quasi-convex model, describe the Cobb–Douglas production efficiency problem as an example, and perform some numerical experiments on this problem via applying the inexact subgradient method. The numerical results verify our theoretical analysis and show that the quasi-subgradient type method is highly efficient for the production efficiency problem, even when the problem is large-scale.

This paper is organized as follows. In Section 2, we present the notations used in this paper, the quasi-subdifferential theory and the inexact subgradient algorithm. In Section 3, we establish convergence properties in both objective values and iterates, and finite convergence behavior of our algorithm when the constraint set X is compact. Section 4 presents the convergence behavior when f has a set of generalized weak sharp minima over noncompact X , and Section 5 gives the efficiency estimates. Finally in Section 6, we apply our algorithm to the Cobb–Douglas production efficiency problem, and demonstrate the numerical results.

2. Preliminaries

2.1. Notation and terminology

We consider the n -dimensional Euclidean space \mathbb{R}^n . We view vector as a column vector, and denote by $\langle x, y \rangle$ the inner product of two vectors $x, y \in \mathbb{R}^n$. We use $\|x\|$ to denote the standard Euclidean norm, $\|x\| = \sqrt{\langle x, x \rangle}$. For $x \in \mathbb{R}^n$ and $\delta \in \mathbb{R}_+$, $B(x, \delta)$ denotes the closed ball of radius δ centered at x and specially B denotes the unit closed ball at the origin. For a set $Z \subseteq \mathbb{R}^n$, we denote the closure of Z by $\text{cl}Z$. We also write $\text{dist}(x, Z)$ to denote the Euclidean distance of a vector x from the set Z , i.e.,

$$\text{dist}(x, Z) = \inf_{z \in Z} \|x - z\|.$$

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

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

f is said to be upper semi-continuous (usc) on \mathbb{R}^n if $f(x) = \limsup_{y \rightarrow x} f(y)$ for all $x \in \mathbb{R}^n$. For each $\alpha \in \mathbb{R}$, we denote the (strict) sublevel sets of f by

$$\begin{aligned} S_{f, \alpha} &= \{x \in \mathbb{R}^n : f(x) < \alpha\}, \quad \bar{S}_{f, \alpha}(x) = S_{f, f(x)}, \\ \bar{S}_{f, \alpha} &= \{x \in \mathbb{R}^n : f(x) \leq \alpha\}, \quad \bar{S}_{f, \alpha}(x) = \bar{S}_{f, f(x)}. \end{aligned}$$

It is well-known that f is quasi-convex if and only if $S_{f, \alpha}(\bar{S}_{f, \alpha})$ is convex for all $\alpha \in \mathbb{R}$, and that f is usc on \mathbb{R}^n if and only if $S_{f, \alpha}$ is open for all $\alpha \in \mathbb{R}$.

2.2. Quasi-subdifferential theory

There are many different types of subdifferential, such as Clarke–Rockafellar subdifferential, Dini subdifferential, Fréchet subdifferential (see Aussel, Corvellec, & Lassonde, 1995 and references therein) and so on. They are the same for convex functions, but different for nonconvex functions. Here we introduce the Greenberg–Pierskalla subdifferential, defined by Greenberg and Pierskalla (1973), as follows.

Definition 2.1 (see Greenberg & Pierskalla, 1973). The z -quasi-conjugate of f is a function $f_z^*: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, defined by

$$f_z^*(x) = z - \inf\{f(y) : \langle x, y \rangle \geq z\}.$$

It is recalled in Greenberg and Pierskalla (1973, Theorem 1) that the z -quasi-conjugate function provides a lower bound for the corresponding convex conjugate function, and indeed, the

convex conjugate function is the supremum of the z -quasi-conjugate over z .

Definition 2.2 (see [Greenberg & Pierskalla, 1973](#)). A Greenberg–Pierskalla subgradient of f at x is a vector $g \in \mathbb{R}^n$ such that

$$f(x) + f_{(g,x)}^*(g) = \langle g, x \rangle. \quad (2.1)$$

The set of Greenberg–Pierskalla subgradients of f at x is called the Greenberg–Pierskalla subdifferential of f at x and is denoted by $\partial^*f(x)$.

The following proposition gives an equivalent formula and some important properties of the Greenberg–Pierskalla subdifferential.

Proposition 2.1 ([Greenberg & Pierskalla, 1973, Theorem 6](#)). The following statements are true:

- (i) $\partial^*f(x) = \{g : \langle g, y - x \rangle < 0, \forall y \in S_f(x)\}$,
- (ii) $\partial^*f(x)$ is a convex cone,
- (iii) $0 \in \partial^*f(x)$ if and only if $x \in \operatorname{argmin} f$.

Unfortunately, different from traditional subdifferentials, the Greenberg–Pierskalla subdifferential of f is not a closed set. In order to overcome this shortcoming, in this paper, we define the following closed set, which contains the closure of $\partial^*f(x)$, instead as the quasi-subdifferential, and use it in the inexact subgradient method.

Definition 2.3. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a quasi-convex function. The quasi-subdifferential of f at x is defined by

$$\bar{\partial}^*f(x) = \{g : \langle g, y - x \rangle \leq 0, \forall y \in S_f(x)\}. \quad (2.2)$$

When f is convex, the quasi-subdifferential coincides with the convex cone hull of the convex subdifferential (i.e., $\bar{\partial}^*f(x) = \operatorname{cone}(\partial f(x))$; see [Hiriart-Urruty and Lemaréchal, \(1996\), Chapter VI, Theorem 1.3.5](#)), and the inexact subgradient method (2.4) is reduced to a normalized version of inexact subgradient method in [Nedić and Bertsekas \(2010\)](#). When f is quasi-convex, the existence and relationship between the Greenberg–Pierskalla subdifferential and the quasi-subdifferential are described in the following lemma.

Lemma 2.1. If f is quasi-convex on \mathbb{R}^n , then $\bar{\partial}^*f(x) \setminus \{0\} \neq \emptyset$. In addition, if f is usc on \mathbb{R}^n , then $\partial^*f(x) \neq \emptyset$, and $\bar{\partial}^*f(x)$ coincides with the closure of $\partial^*f(x)$, i.e., $\bar{\partial}^*f(x) = \partial^*f(x) \cup \{0\}$.

Proof. If $S_f(x) = \emptyset$, then $\bar{\partial}^*f(x) = \mathbb{R}^n$ and the conclusions hold automatically. Now suppose $S_f(x) \neq \emptyset$, since the convex sets $\{x\}$ and $S_f(x)$ are disjoint, it follows from [Bertsekas et al. \(2003, Proposition 2.4.5\)](#) that there exists a proper hyperplane separation, i.e., there exists a vector $g \neq 0$ such that

$$\sup_{y \in S_f(x)} \langle g, y \rangle \leq \langle g, x \rangle \quad \text{and} \quad \inf_{y \in S_f(x)} \langle g, y \rangle < \langle g, x \rangle.$$

Thus, the vector g is a nonzero vector in $\bar{\partial}^*f(x)$. For the second conclusion, see [Kiwiel \(2001, Lemma 3\)](#). \square

The above lemma shows that the existence of nonzero quasi-subgradient only requires the quasi-convexity. Therefore, throughout this paper, we assume that the objective function is quasi-convex. In particular, we do not assume the upper semi-continuity of the objective function as in [Kiwiel \(2001\)](#), unless otherwise specified.

Motivated by practical reasons, relaxing (2.2) by $f(x) + f_{(g,x)}^*(g) \leq \langle g, x \rangle + \epsilon$, we define the ϵ -quasi-subdifferential as follows.

Definition 2.4. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a quasi-convex function. The ϵ -quasi-subdifferential of f at x is defined by

$$\bar{\partial}_\epsilon^*f(x) = \{g : \langle g, y - x \rangle \leq 0, \forall y \in S_{f, f(x)-\epsilon}\}. \quad (2.3)$$

2.3. Inexact subgradient method

In this paper, we introduce a generic inexact subgradient method, which we also call the approximate quasi-subgradient method, to solve the quasi-convex optimization problem (1.1) as follows.

Approximate quasi-subgradient method

Select a stepsize sequence $\{\nu_k\}$, an error sequence $\{\epsilon_k\}$ and a noise sequence $\{r_k\}$. Start with an initial point $x_0 \in X$, and generate a sequence $\{x_k\} \subseteq X$ via the iteration

$$x_{k+1} = P_X(x_k - \nu_k \tilde{g}_k), \quad (2.4)$$

where $P_X(\cdot)$ denotes the Euclidean projection operator onto X and the iterative direction \tilde{g}_k is an approximate quasi-subgradient of the following form

$$\tilde{g}_k := g_k / \|g_k\| + r_k, \quad (2.5)$$

where r_k is a noise vector and $g_k \in \bar{\partial}_{\epsilon_k}^*f(x_k)$ is an arbitrary nonzero ϵ_k -quasi-subgradient of f at x_k .

Let us first consider the following example, which says that ϵ -quasi-subdifferential does not coincide with quasi-subdifferential with noise.

Example 2.1. Consider the quasi-convex function

$$f(x, y) := \begin{cases} x^2 + y^2, & x \geq 0, \\ y^2, & x < 0. \end{cases}$$

Its strict sublevel set $S_f(0, 1) = S_{f,1}$ is illustrated in [Fig. 1](#), thus it is easy to see $\bar{\partial}^*f(0, 1) = \operatorname{cone}\{(0, 1)\}$. Let the noise vector $r = (-\delta, 0)$ with $\delta > 0$. Then its quasi-subdifferential with noise and ϵ -quasi-subdifferential are respectively given by

$$\bar{\partial}^*f(0, 1) + r = \{(-\delta, \lambda) : \lambda \in \mathbb{R}_+\},$$

and

$$\bar{\partial}_\epsilon^*f(0, 1) = \begin{cases} \operatorname{cone}\{(0, 1), (\sqrt{\epsilon}, \sqrt{1-\epsilon})\}, & \epsilon < 1, \\ \mathbb{R}^2, & \epsilon \geq 1. \end{cases}$$

It is obvious that $(-\delta, 1) \notin \bar{\partial}_\epsilon^*f(0, 1)$ for all $\delta > 0$ when $\epsilon < 1$. Thus, from this example, we see that the quasi-subdifferential with noise cannot be represented by the ϵ -quasi-subdifferential.

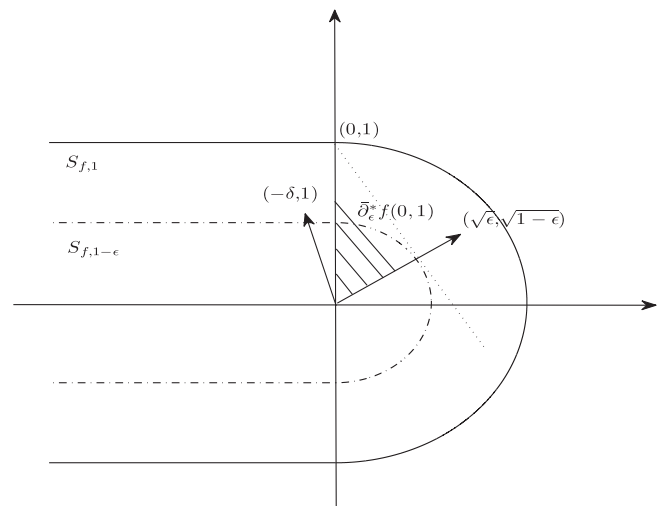


Fig. 1. Illustration of Example 2.1.

It is well-known that the stepsize rule is critical in subgradient methods. In this paper, we investigate convergence properties of the approximate quasi-subgradient method using the following stepsize rules.

- (a) *Constant stepsize rule.* The stepsize v_k is fixed to be a positive scalar v .
- (b) *Diminishing stepsize rule.* The stepsize v_k satisfies

$$v_k > 0, \quad \lim_{k \rightarrow \infty} v_k = 0, \quad \sum_{k=0}^{\infty} v_k = +\infty. \quad (2.6)$$

3. Convergence properties for a compact X

In this section, we investigate convergence properties of the approximate quasi-subgradient method when the constraint set X is compact. Throughout this section, the following three assumptions are made.

Assumption 1. The constraint set X is compact.

Assumption 2. f satisfies the Hölder condition of order $p > 0$ with modulus $\mu > 0$ on \mathbb{R}^n , that is,

$$f(x) - f_* \leq \mu(\text{dist}(x, X^*))^p, \quad \forall x \in \mathbb{R}^n. \quad (3.1)$$

Assumption 3. The noise and errors are bounded, i.e., there exist some $R, \epsilon \geq 0$ such that

$$\|r_k\| \leq R, \forall k \geq 0 \quad \text{and} \quad \limsup_{k \rightarrow \infty} \epsilon_k = \epsilon.$$

Since the constraint set X is compact, all iterates are bounded. Therefore, there exists some $d > 0$ (such as the diameter of X) such that $\|x_k - x\| \leq d$ for all $x \in X$ and $k \geq 0$. Moreover, under the bounded noise assumption, it follows from (2.5) that approximate quasi-subgradients are uniformly bounded, i.e., $\|g_k\| \leq 1 + R$ for all $k \geq 0$.

The Hölder condition of order p is used to describe some properties of quasi-subgradients in Konnov (1994). Here, we use this condition to investigate convergence properties of the approximate quasi-subgradient method. It is worth noting that the Hölder condition of order 1 is equivalent to the bounded subgradient assumption, assumed in Nedić and Bertsekas (2010), whenever f is convex.

3.1. Convergence in objective values

We now give the basic inequality and the convergence property in objective values using both the constant and diminishing stepsize rules. We start with the basic inequality, which shows a significant property of a subgradient iteration.

Lemma 3.1. Let Assumptions 1 and 3 hold and the sequence $\{x_k\}$ be generated by the approximate quasi-subgradient method. Then for all $x \in X$, we have

$$\begin{aligned} \|x_{k+1} - x\|^2 &\leq \|x_k - x\|^2 \\ &\quad - 2v_k \left(\langle g_k / \|g_k\|, x_k - x \rangle - Rd - \frac{1}{2} v_k (1 + R)^2 \right), \quad \forall k. \end{aligned} \quad (3.2)$$

Proof. By (2.4) and (2.5) and the nonexpansive property of projection operator, for all $x \in X$, we have the following basic inequality

$$\begin{aligned} \|x_{k+1} - x\|^2 &\leq \|x_k - v_k \tilde{g}_k - x\|^2 \\ &= \|x_k - x\|^2 - 2v_k \langle g_k / \|g_k\|, x_k - x \rangle \\ &\quad + v_k^2 \|g_k / \|g_k\| + r_k\|^2 \\ &\leq \|x_k - x\|^2 \\ &\quad - 2v_k \left(\langle g_k / \|g_k\|, x_k - x \rangle - Rd - \frac{1}{2} v_k (1 + R)^2 \right), \end{aligned} \quad (3.3)$$

where the last inequality follows from the compactness of X and boundedness of noise and errors. \square

The main difficulty in the study of the approximate quasi-subgradient method comes from the difference between the basic inequality (3.2) for our proposed quasi-convex subgradient method and that of convex subgradient method (cf. Nedić & Bertsekas, 2010). This difference originates from definitions and properties of subgradients: the convex subgradient directly connects with objective values and shares a global property of the objective function, while the quasi-convex subgradient is a normal direction to its sublevel set and is not directly associated with the objective function. Here, we utilize the Hölder condition to relate the quasi-convex subgradient with objective function values, which is only a local property.

Lemma 3.2 Kiwiel, 2001, Lemma 6. If $B(\bar{x}, \bar{r}) \subset \text{cl}S_{f(x_k) - \epsilon_k}$ for some $\bar{x} \in \mathbb{R}^n$ and $\bar{r} \geq 0$, then $\langle g_k / \|g_k\|, x_k - \bar{x} \rangle \geq \bar{r}$.

Lemma 3.3. If Assumption 2 holds and $f(x_k) > f_* + \mu \bar{r}^p + \epsilon_k$ holds for some $\bar{r} \geq 0$, then $\langle g_k / \|g_k\|, x_k - x^* \rangle \geq \bar{r}$ for all $x^* \in X^*$.

Proof. Given $x^* \in X^*$, by the Hölder condition of order p and the hypotheses of this lemma, for all $x \in B(x^*, \bar{r})$, we have

$$f(x) - f_* \leq \mu(\text{dist}(x, X^*))^p \leq \mu \bar{r}^p < f(x_k) - f_* - \epsilon_k,$$

which implies $B(x^*, \bar{r}) \subset S_{f(x_k) - \epsilon_k}$. Hence, the conclusion follows from Lemma 3.2. \square

Theorem 3.1. Let Assumptions 1–3 hold. Then, for a sequence $\{x_k\}$ generated by the approximate quasi-subgradient method with the constant stepsize rule, we have

$$\liminf_{k \rightarrow \infty} f(x_k) \leq f_* + \mu \left(Rd + \frac{v}{2} (1 + R)^2 \right)^p + \epsilon.$$

Proof. We prove by contradiction, assuming that

$$\liminf_{k \rightarrow \infty} f(x_k) > f_* + \mu \left(Rd + \frac{v}{2} (1 + R)^2 \right)^p + \epsilon,$$

that is, there exists some $\delta > 0$ and positive integer k_0 such that

$$f(x_k) > f_* + \mu \left(Rd + \frac{v}{2} (1 + R)^2 + \delta \right)^p + \epsilon_k, \quad \forall k \geq k_0. \quad (3.4)$$

It follows from Lemma 3.3 that for all $x^* \in X^*$ and $k \geq k_0$ there holds

$$\langle g_k / \|g_k\|, x_k - x^* \rangle \geq Rd + \frac{v}{2} (1 + R)^2 + \delta.$$

Therefore, by using the basic inequality (3.2) with $v_k \equiv v$ and $x = x^*$, we obtain

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 \\ &\quad - 2v \left(Rd + \frac{v}{2} (1 + R)^2 + \delta - Rd - \frac{v}{2} (1 + R)^2 \right) \\ &= \|x_k - x^*\|^2 - 2v\delta \leq \dots \\ &\leq \|x_{k_0} - x^*\|^2 - 2(k - k_0 + 1)v\delta, \end{aligned}$$

which yields a contradiction for sufficiently large k . The proof is complete. \square

In [Assumption 2](#), we assume that f satisfies the Hölder condition on the whole space \mathbb{R}^n . Actually, this assumption is essential for the convergence result in [Theorem 3.1](#). Relaxing it by the assumption that f satisfies the Hölder condition on the constraint set X cannot ensure the validity of [Theorem 3.1](#) even if f is continuous on \mathbb{R}^n , as shown by the following example.

Example 3.1. Consider the objective function

$$f(u, v) := \begin{cases} M|v|, & u \leq 0, \\ u + M|v|, & u > 0, \end{cases}$$

with $M = 100$ and the constraint set $X = \{(u, v) : -1 \leq u \leq 1, v = 0\}$. Obviously, the optimal value of (1.1) is $f_* = 0$ and the optimal solution set is $X^* = \{(u, v) : -1 \leq u \leq 0, v = 0\}$. It is easy to check that f is continuous and quasi-convex on \mathbb{R}^2 and satisfies the Hölder condition (cf. (3.1)) on X with $\mu = p = 1$.

Starting from $x_0 = (1, 0)$, we use the approximate quasi-subgradient method (cf. (2.4) and (2.5)) to solve this problem. Specially, we choose the quasi-subgradient $g = (1/\sqrt{1+M^2}, M/\sqrt{1+M^2}) \in \partial^* f(x_0)$, the noise vector $r = (-1/\sqrt{1+M^2}, 0)$ and the constant stepsize rule $v = 1/2$, then we have

$$x_1 = P_X(x_0 - v(g+r)) = P_X\left((1, 0) - v\left(0, M/\sqrt{1+M^2}\right)\right) = (1, 0) = x_0.$$

Hence, a fixed sequence is generated and $\lim_{k \rightarrow \infty} f(x_k) = f(x_0) = 1$. However, $R = 0.01$, $\epsilon = 0$, $d = 2$, $v = 1/2$ and then the total error $\mu(Rd + \frac{v}{2}(1+R)^2)^p + \epsilon < 1/2$. Therefore, [Theorem 3.1](#) fails for this problem.

Using the diminishing stepsize rule, the error term involving the stepsize v in [Theorem 3.1](#) vanishes and the following theorem is obtained.

Theorem 3.2. Let Assumptions 1–3 hold. Then, for a sequence $\{x_k\}$ generated by the approximate quasi-subgradient method with the diminishing stepsize rule, we have

$$\liminf_{k \rightarrow \infty} f(x_k) \leq f_* + \mu(Rd)^p + \epsilon.$$

Proof. The proof uses properties of the diminishing stepsize rule (cf. (2.6)) and a line of analysis similar to that of [Theorem 3.1](#). We omit the details. \square

[Theorems 3.1 and 3.2](#) show convergence to the optimal value within some tolerance given in terms of errors and noise by using the constant and diminishing stepsize rules, respectively. In [Theorem 3.2](#), the total error $c := \mu(Rd)^p + \epsilon$, which is a similar formula as in [Nedić and Bertsekas \(2010\)](#), has an additive form, including the noise level R and the error level ϵ . By contrast, in [Theorem 3.1](#), the total error additionally includes a term related to the constant stepsize v .

When the noise vanishes ($R = 0$), the approximate quasi-subgradient method is reduced to the ϵ -quasi-subgradient method. In such a situation, the term $\langle r_k, x_k - x \rangle$ vanishes in the corresponding basic inequality, and [Lemma 3.1](#) holds (where $R = 0$) without the compactness hypothesis of X . Therefore, when the noise vanishes, the convergence result holds regardless of the compactness hypothesis of X . Furthermore, when the error level is precise ($\epsilon = 0$), we obtain the convergence result of the exact quasi-subgradient method, which is the main result in [Kiwiel \(2001\)](#), where upper semi-continuity of f is assumed. Here, we have

obtained the result as in [Kiwiel \(2001\)](#) without the usc assumption, but using the Hölder condition of order p instead. The following two examples show that the Hölder condition and upper semi-continuity are independent of each other.

Example 3.2 (The function satisfies the Hölder condition but is not usc). Consider the objective function

$$f(x) := \begin{cases} 0, & x \leq 0, \\ x^2, & 0 < x \leq 1, \\ 2, & x > 1, \end{cases}$$

with the constraint set $X = \{x \in \mathbb{R} : 0 \leq x \leq 10\}$. Obviously, the optimal value of problem (1.1) is $f_* = 0$ and the optimal solution set is $X^* = \{0\}$. It is easy to verify that f is quasi-convex and satisfies the Hölder condition of order 2 with modulus 1 on \mathbb{R} . However, f is not usc at $x = 1$. Thus, this example shows that the Hölder condition does not imply upper semi-continuity.

Thus, from [Kiwiel \(2001\)](#), we cannot obtain convergence of the exact quasi-subgradient method (cf. (14)–(15) in [Kiwiel \(2001\)](#)) for this example. However, the sequence generated by the exact quasi-subgradient method converges to X^* . Indeed, for any $x \in X \setminus X^*$, the strict sublevel set $S_f(x)$ is the line segment $[0, \min\{1, x\})$ and the quasi-subdifferential $\partial^* f(x) = \mathbb{R}_+$. Thus,

$$x_{k+1} = P_X(x_k - v_k g_k / \|g_k\|) = \max\{x_k - v_k, 0\},$$

and the sequence $\{x_k\}$ converges to the origin, the optimal solution, by properties of the diminishing stepsize rule. This iterative result coincides with the result in [Theorem 3.2](#) (by setting $R = 0$ and $\epsilon = 0$).

Example 3.3 (The function is usc but does not satisfy the Hölder condition). Consider the objective function

$$f(x) = e^x,$$

and the constraint set \mathbb{R}_+ . Obviously, the optimal value of problem (1.1) is $f_* = 1$ and the optimal solution set is $X^* = \{0\}$. It is easy to check that f is continuous and quasi-convex (since it is monotone) on \mathbb{R} . However, by the Taylor expansion $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$, we claim that f does not satisfy the Hölder condition on \mathbb{R} for any positive scalars p and μ . Indeed, given positive scalars p and μ , when $x \geq \exp\left(\frac{\log(\mu[p+1]!)}{[p+1]-p}\right)$, where $\exp(\cdot)$ and $[p]$ denote the exponential function and the largest integer not greater than p respectively, we have

$$f(x) - f_* = e^x - 1 > \frac{x^{[p+1]}}{[p+1]!} + \frac{x^0}{0!} - 1 = \frac{x^{[p+1]}}{[p+1]!} \geq \mu x^p,$$

which contradicts with (3.1). Thus, this example shows that upper semi-continuity does not imply the Hölder condition.

Although, from [Kiwiel \(2001\)](#), we obtain the convergence property of the exact quasi-subgradient method for this example. However, the convergence result of the approximate quasi-subgradient method (see [Theorem 3.2](#)) fails for this example. Indeed, given positive scalars p and μ , we consider the constraint set $X = \left\{x \in \mathbb{R} : 0 \leq x \leq \exp\left(\frac{\log(\mu[p+1]!)}{[p+1]-p}\right)\right\}$, noise $r_k \equiv -1$ and errors $\epsilon_k \equiv 0$. For any $x \in X \setminus X^*$, the strict sublevel set $S_f(x)$ is the line segment $[0, x)$ and the quasi-subdifferential $\partial^* f(x) = \mathbb{R}_+$. Thus, starting from $x_0 = \exp\left(\frac{\log(\mu[p+1]!)}{[p+1]-p}\right)$, we have

$$x_1 = P_X(x_0 - v_0(g_0/\|g_0\| + r_0)) = x_0.$$

Hence, the approximate quasi-subgradient method (cf. (2.4) and (2.5)) generates a fixed sequence and $\lim_{k \rightarrow \infty} f(x_k) = f(x_0) = e^{x_0}$.

However, when $R = 1$, $\epsilon = 0$ and $d = \exp\left(\frac{\log(\mu[p+1]!)}{[p+1]-p}\right)$, the total error, given in [Theorem 3.2](#), $\mu(Rd)^p + \epsilon = \mu d^p = d^{[p+1]}/[p+1]! < e^d = e^{x_0}$, where the inequality follows from the Taylor expansion. Therefore, [Theorem 3.2](#) fails for this example.

From the above two examples, we observe that the Hölder condition of order p describes some property of the objective function, which is essentially different from the upper semi-continuity, and it can be used to investigate convergence properties of the approximate quasi-subgradient method. Hence, using the mild assumptions, we have established convergence properties of the approximate quasi-subgradient method from a new perspective, which is different from that in [Kiwiel \(2001\)](#).

3.2. Finite convergence

The optimal solution set X^* has a nonempty interior in many interesting applications, such as surrogate relaxation of discrete programming problems (see [Dyer, 1980](#)). Here, we demonstrate finite convergence behavior to the approximate optimal solution set of problem (1.1) under the assumption that the optimal solution set X^* has a nonempty interior.

Theorem 3.3. *Let Assumptions 1–3 hold, $\text{int}X^* \neq \emptyset$ and the diminishing stepsize rule be chosen. Then $f(x_k) \leq f_* + \mu(Rd)^p + \epsilon$ for some k .*

Proof. By contradiction, we assume that $f(x_k) > f_* + \mu(Rd)^p + \epsilon$ for all $k \in \mathbb{N}$. Since $\text{int}X^* \neq \emptyset$, we set $B(\bar{x}, \bar{\delta}) \subset X^*$ for some $\bar{\delta} > 0$. Then for all $x \in B(\bar{x}, Rd + \frac{2}{3}\bar{\delta})$, we have

$$\begin{aligned} f(x) - f_* &\leq \mu(\text{dist}(x, X^*))^p \leq \mu\left(Rd - \frac{1}{3}\bar{\delta}\right)^p = \mu(Rd)^p - \delta' \\ &< f(x_k) - f_* - \epsilon - \delta', \end{aligned} \quad (3.5)$$

where δ' is a scalar in $\left[\frac{1}{3}\mu p \bar{\delta} (Rd - \frac{1}{3}\bar{\delta})^{p-1}, \frac{1}{3}\mu p \bar{\delta} (Rd)^{p-1}\right]$ satisfying the mean value theorem. Furthermore, since $\limsup_{k \rightarrow \infty} \epsilon_k = \epsilon$, there exists some positive integer k_0 such that $\epsilon_k \leq \epsilon + \delta'$ for all $k \geq k_0$. Therefore, (3.5) implies $f(x) < f(x_k) - \epsilon_k$ and then $B(\bar{x}, Rd + \frac{2}{3}\bar{\delta}) \subset S_{f, f(x_k) - \epsilon_k}$ for all $k \geq k_0$. Hence, it follows from [Lemma 3.2](#) that

$$\langle g_k / \|g_k\|, x_k - \bar{x} \rangle \geq Rd + \frac{2}{3}\bar{\delta}. \quad (3.6)$$

However, summing the basic inequalities (3.2) with $x = \bar{x}$ for $i = k_0, \dots, k$, we obtain

$$\begin{aligned} \min_{i=k_0, \dots, k} \left\langle \frac{g_i}{\|g_i\|}, x_i - \bar{x} \right\rangle &\leq \frac{\sum_{i=k_0}^k v_i \left\langle \frac{g_i}{\|g_i\|}, x_i - \bar{x} \right\rangle}{\sum_{i=k_0}^k v_i} \\ &\leq \frac{\|x_{k_0} - \bar{x}\|^2}{2\sum_{i=k_0}^k v_i} + Rd + \frac{\sum_{i=k_0}^k v_i^2}{2\sum_{i=k_0}^k v_i} (1+R)^2. \end{aligned} \quad (3.7)$$

By the property of the diminishing stepsize rule (cf. (2.6)), it follows from [Kiwiel \(2004, Lemma 2.1\)](#) that $\lim_{k \rightarrow \infty} (\sum_{i=k_0}^k v_i^2 / \sum_{i=k_0}^k v_i) = 0$, and thus the right hand side of (3.7) tends to Rd as k tends to infinity. Hence we arrive at a contradiction with (3.6). The proof is complete. \square

Under the same assumption of [Theorem 3.3](#), we now describe a related result for the nonvanishing stepsize rule.

Theorem 3.4. *Let Assumptions 1–3 hold. If $B(\bar{x}, \bar{\delta}) \subset X^*$ for some $\bar{\delta} > 0$ and there exists some $0 < \kappa < 1$ and $k_0 \in \mathbb{N}$ such that $v_k \in \left[\frac{\kappa^2 \bar{\delta}}{(1+R)^2}, \frac{\kappa \bar{\delta}}{(1+R)^2}\right]$ for all $k \geq k_0$, then $f(x_k) \leq f_* + \mu(Rd)^p + \epsilon$ for some k .*

Proof. By contradiction, suppose $f(x_k) > f_* + \mu(Rd)^p + \epsilon$ for all $k \in \mathbb{N}$. As in the proof of [Theorem 3.3](#) and (3.7), we have

$$\begin{aligned} Rd + \frac{2}{3}\bar{\delta} &\leq \min_{i=k_0, \dots, k} \langle g_i / \|g_i\|, x_i - \bar{x} \rangle \\ &\leq \frac{\|x_{k_0} - \bar{x}\|^2}{2\sum_{i=k_0}^k v_i} + Rd + \frac{\sum_{i=k_0}^k v_i^2}{2\sum_{i=k_0}^k v_i} (1+R)^2 \\ &\leq \frac{\|x_{k_0} - \bar{x}\|^2}{2\kappa^2 \bar{\delta} (k - k_0 + 1)} (1+R)^2 + Rd + \bar{\delta}/2, \end{aligned}$$

whose last right hand side tends to $Rd + \bar{\delta}/2$ as k tends to infinity. The contradiction happens. \square

3.3. Convergence in iterates

We have shown the convergence property in objective values in Section 3.1, and in this subsection we consider the convergence property in iterates. In [Nedić and Bertsekas \(2010\)](#), where noise in subgradient methods for convex optimization was considered, Nedić and Bertsekas did not give convergence property in iterates. In fact, convergence of $\{x_k\}$ is quite difficult to obtain. [Kiwiel \(2004\)](#) described the convergence of $\{x_k\}$ generated by ϵ -subgradient method for convex optimization. Although Kiwiel did not consider the effect of noise, his work is really helpful for our research. Following the framework of [Kiwiel \(2004\)](#), we give the convergence of $\{x_k\}$ generated by the approximate quasi-subgradient method using the diminishing stepsize rule. Besides the extension to the approximate quasi-subgradient method, another improvement of our work is to maintain the convergence property without the lower semi-continuous and coercive condition assumptions used in [Kiwiel \(2004\)](#), instead we use the usc assumption.

First, let us show a useful property of a convergent sequence, which converges in objective values as well. This result requires the additional usc assumption.

Lemma 3.4. *Suppose f is usc on \mathbb{R}^n , $\alpha > 0$, and the sequence $\{x_k\}$ converges to \bar{x} with $\lim_{k \rightarrow \infty} f(x_k) \leq f_* + \alpha$. Then $\text{dist}(\bar{x}, \bar{S}_{f, f_* + \alpha}) = 0$.*

Proof. Observe that $S_{f, f_* + \beta}$ is open and convex (as f is usc and quasi-convex) for all $\beta > f_* + \alpha$ and that $\bigcap_{\beta > f_* + \alpha} S_{f, \beta} \supset S_{f, f_* + \alpha}$, which is nonempty (as α is positive and f is usc). Since further $\{S_{f, \beta}\}$ is decreasing as $\beta \downarrow f_* + \alpha$, by [Rockafellar and Wets \(1998, Exercise 4.3\(b\)\)](#), we have

$$\lim_{\beta \downarrow f_* + \alpha} S_{f, \beta} = \bigcap_{\beta > f_* + \alpha} \text{cl} S_{f, \beta} = \text{cl} \bigcap_{\beta > f_* + \alpha} S_{f, \beta} = \text{cl} \bar{S}_{f, f_* + \alpha}, \quad (3.8)$$

where the second equality follows from [Rockafellar \(1970, Theorem 6.5\)](#). Finally, by [Rockafellar and Wets \(1998, Corollary 4.7\)](#) and (3.8), we arrive at that

$$\text{dist}(\bar{x}, \bar{S}_{f, f_* + \alpha}) = \text{dist}(\bar{x}, \text{cl}(\bar{S}_{f, f_* + \alpha})) = \lim_{\beta \downarrow f_* + \alpha} \text{dist}(\bar{x}, S_{f, \beta}) = 0,$$

where $\text{dist}(\bar{x}, S_{f, \beta}) = 0$ for all $\beta > f_* + \alpha$, since $\lim_{k \rightarrow \infty} x_k = \bar{x}$ and $\lim_{k \rightarrow \infty} f(x_k) \leq f_* + \alpha < \beta$. \square

Next, we describe the convergence of $\{x_k\}$ to some approximate optimal solution set by using the diminishing stepsize rule.

Theorem 3.5. *Let Assumptions 1–3 hold, the total error $c := \mu(Rd)^p + \epsilon > 0$, f be usc on \mathbb{R}^n and the diminishing stepsize rule be chosen. Then the following statements are true:*

- (i) $\liminf_{k \rightarrow \infty} \text{dist}(x_k, \bar{S}_{f, f_*+c} \cap X) = 0$.
(ii) $\lim_{k \rightarrow \infty} \text{dist}(x_k, X^* + \rho(c)B) = 0$, where $\rho(c)$ is defined by

$$\rho(c) := \max\{\text{dist}(x, X^*) : x \in \bar{S}_{f, f_*+c} \cap X\}.$$

Proof. First, observe that $X^* \subset \bar{S}_{f, f_*+c} \cap X \subset X^* + \rho(c)B$. Furthermore, the nonemptiness of X^* and the compactness of X imply that $\bar{S}_{f, f_*+c} \cap X$ is nonempty and bounded.

- (i) **Theorem 3.2** gives that $\liminf_{k \rightarrow \infty} f(x_k) \leq f_* + c$. The compactness of X then implies that there exists some subsequence $\{x_{k_i}\}$ that converges to some $\bar{x} \in X$ with $\lim_{i \rightarrow \infty} f(x_{k_i}) \leq f_* + c$. Thus, the conclusion follows from **Lemma 3.4**.
(ii) Given $\sigma > 0$, define

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

and

$$e_\sigma := \inf \left\{ f(x) : x \in X, \text{dist}(x, \bar{S}_{f, f_*+c} \cap X) \geq \sigma \right\} - (f_* + c). \quad (3.9)$$

We first claim that $e_\sigma > 0$. Indeed, if $e_\sigma = 0$, then there exists sequence $\{z_i\}$, in $\{x : x \in X, \text{dist}(x, \bar{S}_{f, f_*+c} \cap X) \geq \sigma\}$, converges to some $\bar{z} \in X$ with $\lim_{i \rightarrow \infty} f(z_i) = f_* + c$. It follows from **Lemma 3.4** that $\text{dist}(\bar{z}, \bar{S}_{f, f_*+c}) = 0$. Moreover, since $\bar{z} \in X$, $\text{dist}(\bar{z}, \bar{S}_{f, f_*+c} \cap X) = 0$, which is impossible as $\sigma > 0$.

For such positive e_σ , there exists some $\delta > 0$ such that

$$\mu(R\delta + \delta)^p \leq \mu(R\delta)^p + e_\sigma/2. \quad (3.10)$$

Since the stepsize v_k diminishes, there exists $k_\delta \in \mathbb{N}$ such that

$$v_k \leq \delta/(1+R)^2, \forall k \geq k_\delta. \quad (3.11)$$

Since $\limsup_{k \rightarrow \infty} \epsilon_k = \epsilon$ and $\lim_{k \rightarrow \infty} \|x_{k+1} - x_k\| = 0$ (since v_k diminishes), there exists some $k_\sigma \geq k_\delta$ such that

$$\epsilon_k < \epsilon + e_\sigma/2, \quad (3.12)$$

and

$$\|x_{k+1} - x_k\| \leq \sigma, \quad (3.13)$$

for all $k \geq k_\sigma$. Since $\liminf_{k \rightarrow \infty} \text{dist}(x_k, \bar{S}_{f, f_*+c} \cap X) = 0$ (cf. (i)), there exists some $k'_\sigma \geq k_\sigma \geq k_\delta$ such that

$$x_{k'_\sigma} \in (\bar{S}_{f, f_*+c} \cap X) + \sigma B \subset X^* + \rho(c)B + \sigma B \subset V_{2\sigma},$$

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

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

Case 1. If $\text{dist}(x_k, \bar{S}_{f, f_*+c} \cap X) \leq \sigma$, from (3.13), we have

$$\begin{aligned} x_{k+1} &\in \{x_k\} + \sigma B \subset (\bar{S}_{f, f_*+c} \cap X + \sigma B) + \sigma B \\ &\subset X^* + \rho(c)B + 2\sigma B = V_{2\sigma}. \end{aligned}$$

Case 2. Suppose $\text{dist}(x_k, \bar{S}_{f, f_*+c} \cap X) > \sigma$, from (3.9), we have

$$\begin{aligned} f(x_k) &\geq e_\sigma + f_* + c = f_* + (\mu(R\delta)^p + e_\sigma/2) + (\epsilon + e_\sigma/2) \\ &> f_* + \mu(R\delta + \delta)^p + \epsilon_k, \forall k \geq k'_\sigma, \end{aligned}$$

where the second inequality follows from (3.10) and (3.12). Hence, from **Lemmas 3.1 and 3.3**, we have

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 - 2v_k \left(\delta - \frac{v_k}{2}(1+R)^2 \right) \\ &\leq \|x_k - x^*\|^2, \end{aligned}$$

where the second inequality follows from (3.11). Thus, $x_k \in V_{2\sigma}$ implies $x_{k+1} \in V_{2\sigma}$. Therefore, by induction, $x_k \in V_{2\sigma}$, and hence,

$\text{dist}(x_k, X^* + \rho(c)B) \leq 2\sigma$ for all $k \geq k'_\sigma$. Since $\sigma > 0$ is arbitrary, then $\text{dist}(x_k, X^* + \rho(c)B)$ vanishes as k tends to infinity. \square

4. Convergence properties for f with generalized weak sharp minima

In this section, we consider the other case when X is noncompact. Considering the similar case, **Nedić and Bertsekas (2010)** assumed that the objective function f has a set of weak sharp minima and the ϵ -subgradients are uniformly bounded on X (see **Nedić & Bertsekas, 2010, Assumptions 3.1–3.2**). The function f is said to have a set of weak sharp minima over X (see **Burke & Ferris, 1993**) if for some scalar $\eta > 0$ there holds

$$f(x) - f_* \geq \eta \text{dist}(x, X^*), \quad \forall x \in X.$$

A natural extension to generalize the weak sharp minima is the weak sharp minima of order q (see **Bonnans & Ioffe, 1995; Studniarski & Ward, 1999**), that is, there exist some scalars $\eta, q > 0$ such that

$$f(x) - f_* \geq \eta (\text{dist}(x, X^*))^q, \quad \forall x \in X. \quad (4.1)$$

However, if $p > q$, contradiction between (3.1) and (4.1) arises as $\text{dist}(x, X^*)$ tends to zero. Also, if $p < q$, contradiction arises again as $\text{dist}(x, X^*)$ tends to infinity. In order to avoid the contradiction, we weaken the assumption (4.1) as the generalized weak sharp minima, in which the constant q is replaced by a positive function $g(t)$.

Furthermore, in what follows we consider a noise sequence $\{r_k\}$ whose norm bound R is lower than $(\eta/\mu)^{1/p}$, which we refer to as a low level noise sequence (see **Nedić & Bertsekas, 2010**). In particular, we introduce the following assumptions.

Assumption 4. The function f satisfies the generalized weak sharp minima condition over X , that is, there exist some scalars $\eta > 0, q \geq p$ and a function $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, satisfying $g(\cdot) \geq p$, $\sup_{t \geq 0} g(t) = q$ and $\lim_{t \rightarrow \infty} g(t) = p$, such that

$$f(x) - f_* \geq \eta (\text{dist}(x, X^*))^{g(\text{dist}(x, X^*))}, \quad \forall x \in X, \quad (4.2)$$

where p is the order of Hölder condition used in **Assumption 2**.

Assumption 5. $\{r_k\}$ is a low level noise sequence (i.e., $R < (\eta/\mu)^{1/p}$).

When $g(t) \equiv p$, **Assumption 4** is reduced to weak sharp minima of order p , whose sufficient and necessary conditions have been described by **Studniarski and Ward (1999)** and **Bonnans and Ioffe (1995)** for specified $p = 2$. Furthermore, if $p = 1$, it is reduced to the well-known weak sharp minima introduced by **Burke and Ferris (1993)**. Note that, to arrive at the corresponding convergence results, **Assumptions 2–5** with specified $p = q = 1$ were used in **Nedić and Bertsekas (2010)**.

When

$$g(t) := \begin{cases} g(0), & 0 \leq t \leq 1, \\ p, & t > 1, \end{cases}$$

where $g(0) > p$, **Assumption 4** is reduced to

$$f(x) - f_* \geq \min\{\eta (\text{dist}(x, X^*))^{g(0)}, \eta (\text{dist}(x, X^*))^p\},$$

which is equivalent to that f has Höldrian level sets over X (see **Pang, 1997**). Another interesting example of **Assumption 4** is $g(t) = p + 1/t$.

Before we go on, we introduce an auxiliary function $H_{v,\theta}^x$ and investigate some properties of the maximum solution of $H_{v,\theta}^x(z) \geq 0$ over X , which are useful in the study of convergence

properties in objective values and iterates when X is noncompact in next two subsections.

Definition 4.1. Let μ and p be scalars given in [Assumption 2](#), R and ϵ be scalars given in [Assumption 3](#), and the function g be described in [Assumption 4](#). For each $v \geq 0$, $\theta \geq 0$ and $x \in X$, we define an auxiliary function $H_{v,\theta}^x : \mathbb{R}_+ \rightarrow \mathbb{R}$ by

$$H_{v,\theta}^x(z) := \mu \left(\frac{v}{2} (1+R)^2 + R \left(\frac{z}{\eta} \right)^{1/g(\text{dist}(x,X^*))} \right)^p + \epsilon + \theta - z, \quad (4.3)$$

We denote by $z_{v,\theta}^*$ to be the maximum solution of the inequality $H_{v,\theta}^x(z) \geq 0$ for some $x \in X$, defined by

$$z_{v,\theta}^* := \sup\{z : H_{v,\theta}^x(z) \geq 0 \text{ for some } x \in X\}. \quad (4.4)$$

[Assumption 4](#) says that $p \leq g(\text{dist}(x,X^*)) \leq q$ for all $x \in X$. Hence, by [\(4.3\)](#), we have

$$H_{v,\theta}^x(z) \leq \max\{G_{v,\theta}^p(z), G_{v,\theta}^q(z)\}, \quad \forall z \geq 0, \quad x \in X,$$

where $G_{v,\theta}^t(z)$, where $t = p, q$, is defined by

$$G_{v,\theta}^t(z) := \mu \left(\frac{v}{2} (1+R)^2 + R \left(\frac{z}{\eta} \right)^{1/t} \right)^t + \epsilon + \theta - z.$$

Thus, applying [\(4.4\)](#) and [Assumption 4](#), $z_{v,\theta}^*$ can be rewritten as

$$z_{v,\theta}^* = \{\sup\{z : G_{v,\theta}^p(z) \geq 0\}, \sup\{z : G_{v,\theta}^q(z) \geq 0\}\}.$$

For the sake of simplicity, denote

$$z_{v,\theta}^t = \sup\{z : G_{v,\theta}^t(z) \geq 0\}, \quad \text{for } t = p, q. \quad (4.5)$$

and hence,

$$z_{v,\theta}^* = \max\{z_{v,\theta}^p, z_{v,\theta}^q\}. \quad (4.6)$$

Since $H_{v,\theta}^x(0) > 0$ and $H_{v,\theta}^x(z)$ is continuous on variable z for all $x \in X$, then $z_{v,\theta}^*$ is positive. However, it might be $+\infty$. The following lemma shows that $z_{v,\theta}^*$ is finite and continuous on parameters v and θ under [Assumptions 4 and 5](#).

Lemma 4.1. Let [Assumptions 4 and 5](#) hold. Then the following statements hold:

- (i) $z_{v,\theta}^*$ is finite for all $v \geq 0$ and $\theta \geq 0$.
- (ii) $\lim_{\theta \rightarrow 0+} z_{v,\theta}^* = z_{v,0}^*$ for all $v \geq 0$.
- (iii) $\lim_{v \rightarrow 0+} z_{v,\theta}^* = z_{0,\theta}^*$ for all $\theta \geq 0$.

Proof.

- (i) By the assumptions that $R < (\eta/\mu)^{1/p}$ and $q \geq p$, we have

$$\lim_{z \rightarrow \infty} \mu \left(\frac{R}{\eta^{1/q}} z^{1/q-1/p} \right)^p < 1,$$

which is equivalent to

$$\lim_{z \rightarrow \infty} \left[\frac{\mu}{z} \left(\frac{v}{2} (1+R)^2 + R \left(\frac{z}{\eta} \right)^{1/q} \right)^p + \frac{\epsilon + \theta}{z} \right] < 1, \quad \forall v \geq 0, \quad \theta \geq 0.$$

This implies $\lim_{z \rightarrow \infty} G_{v,\theta}^q(z) < 0$. Hence, $z_{v,\theta}^q < +\infty$ for all $v \geq 0$ and $\theta \geq 0$, since $G_{v,\theta}^q(\cdot)$ is continuous. Similarly, we can prove that $z_{v,\theta}^p < +\infty$ for all $v \geq 0$ and $\theta \geq 0$. Thus, by using [\(4.6\)](#), we arrive at that $z_{v,\theta}^*$ is finite for all $v \geq 0$ and $\theta \geq 0$.

- (ii) Since $G_{v_1,\theta_1}^q(\cdot) \leq G_{v_2,\theta_2}^q(\cdot)$ for all $v_1 \leq v_2$ and $\theta_1 \leq \theta_2$, then $z_{v_1,\theta_1}^q \leq z_{v_2,\theta_2}^q$. This monotonicity immediately implies $\lim_{\theta \rightarrow 0+} z_{v,\theta}^q \geq z_{v,0}^q$.

Next, we prove the reverse inequality. By the definition of $z_{v,\theta}^q$, for given $v \geq 0$ and each positive integer n , there exists some z_n satisfying $z_n > z_{v,1/n}^q - 1/n$ and $G_{v,1/n}^q(z_n) \geq 0$. Together with the monotonicity of $z_{v,\theta}^q$, we have $-1 < z_n \leq z_{v,1/n}^q \leq z_{v,1}^q$, where the last term is finite by (i). So the sequence $\{z_n\}$ is bounded and has cluster points. Thus, for each of its cluster points \bar{z} , taking a subsequence of $\{z_n\}$ if necessary, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} G_{v,1/n}^q(z_n) &= \lim_{n \rightarrow \infty} \mu \left(\frac{v}{2} (1+R)^2 + R \left(\frac{z_n}{\eta} \right)^{1/q} \right)^p + \epsilon + \frac{1}{n} - z_n \\ &= \mu \left(\frac{v}{2} (1+R)^2 + R \left(\frac{\bar{z}}{\eta} \right)^{1/q} \right)^p + \epsilon - \bar{z} = G_{v,0}^q(\bar{z}), \end{aligned}$$

which is nonnegative, since $\{G_{v,1/n}^q(z_n)\}$ are all nonnegative. Then, by the definition of $z_{v,\theta}^q$, we have $z_{v,0}^q \geq \bar{z} \geq \lim_{\theta \rightarrow 0+} z_{v,\theta}^q$, where the second inequality holds due to $z_n > z_{v,1/n}^q - 1/n$. Therefore, we arrive at $\lim_{\theta \rightarrow 0+} z_{v,\theta}^q = z_{v,0}^q$.

Similarly, we can prove that $\lim_{\theta \rightarrow 0+} z_{v,\theta}^p = z_{v,0}^p$. Thus, from [\(4.6\)](#), we arrive at $\lim_{\theta \rightarrow 0+} z_{v,\theta}^* = z_{v,0}^*$ for all $v \geq 0$.

- (iii) The proof is similar to that of (ii). \square

4.1. Convergence in objective values

Similar to [Section 3.1](#), we obtain the following basic inequality.

Lemma 4.2. Let $\{x_k\}$ be the sequence generated by the approximate quasi-subgradient method. Then for all $x \in X$, we have

$$\begin{aligned} \|x_{k+1} - x\|^2 &\leq \|x_k - x\|^2 \\ &\quad - 2v_k \left(\langle g_k / \|g_k\|, x_k - x \rangle - R\|x_k - x\| - \frac{1}{2}v_k(1+R)^2 \right), \quad \forall k. \end{aligned}$$

Before we discuss the convergence in objective values which is the main result of this subsection, we consider the following two lemmas which show the boundedness of $\{x_k\}$, generated by the approximated quasi-subgradient method using both types of step-size rules. This interesting property is new in the literature.

Lemma 4.3. Let [Assumptions 2–5](#) hold and $\{x_k\}$ be generated by the approximate quasi-subgradient method with the constant stepsize rule. Then $\{x_k\}$ is bounded.

Proof. Since $\limsup_{k \rightarrow \infty} \epsilon_k = \epsilon$, for any $\theta > 0$, there exists some positive integer k_0 such that

$$\epsilon_k < \epsilon + \theta, \quad \forall k \geq k_0. \quad (4.7)$$

Define the maximum solution of $t^{g(t)} \leq z_{v,\theta}^*/\eta$ by

$$T := \sup \{t \in \mathbb{R}_+ : t^{g(t)} \leq z_{v,\theta}^*/\eta\}, \quad (4.8)$$

which is finite, since $z_{v,\theta}^*$ is finite (cf. [Lemma 4.1\(i\)](#)) and $\lim_{t \rightarrow \infty} t^{g(t)} = +\infty$ (cf. [Assumption 4](#)). Next, we claim that the following inequality holds for all $i \geq k_0$:

$$\text{dist}(x_i, X^*) \leq \max\{\text{dist}(x_{k_0}, X^*), T + v(1+R)\}. \quad (4.9)$$

It is obvious that [\(4.9\)](#) holds if $i = k_0$. Proving by induction, we assume that [\(4.9\)](#) holds for some $i = k(\geq k_0)$. We consider the following two cases.

Case 1. If $f(x_k) \leq f_* + \mu \left(\frac{v}{2} (1+R)^2 + R \left(\frac{f(x_k) - f_*}{\eta} \right)^{1/g(\text{dist}(x_k, X^*))} \right)^p + \epsilon_k$, by [\(4.7\)](#), we have

$$\mu \left(\frac{v}{2} (1+R)^2 + R \left(\frac{f(x_k) - f_*}{\eta} \right)^{1/g(\text{dist}(x_k, X^*))} \right)^p + \epsilon + \theta - (f(x_k) - f_*) \geq 0,$$

that is, $H_{v,\theta}^{x_k}(f(x_k) - f_*) \geq 0$. Hence, by (4.4), we obtain $f(x_k) - f_* \leq z_{v,\theta}^*$ and then

$$\text{dist}(x_k, X^*)^{g(\text{dist}(x_k, X^*))} \leq z_{v,\theta}^*/\eta,$$

which follows from (4.2). Thus, from (4.8), we arrive at $\text{dist}(x_k, X^*) \leq T$, and thus relations (2.4) and (2.5) imply

$$\text{dist}(x_{k+1}, X^*) \leq \text{dist}(x_k, X^*) + v_k \|g_k\|/\|g_k\| + r_k < T + v(1+R).$$

That is (4.9) holds for $i = k + 1$.

Case 2. Suppose $f(x_k) > f_* + \mu \left(\frac{v}{2}(1+R)^2 + R \left(\frac{f(x_k) - f_*}{\eta} \right)^{1/g(\text{dist}(x_k, X^*))} \right)^p + \epsilon_k$, then it follows from Lemma 3.3 that

$$\begin{aligned} \langle g_k/\|g_k\|, x_k - x^* \rangle &\geq \frac{v}{2}(1+R)^2 + R \left(\frac{f(x_k) - f_*}{\eta} \right)^{1/g(\text{dist}(x_k, X^*))} \\ &\geq \frac{v}{2}(1+R)^2 + R \text{dist}(x_k, X^*), \end{aligned}$$

where the second inequality follows from (4.2). Hence, applying Lemma 4.2 with $v_k = v$ and $x^* = P_{X^*}(x_k)$, we obtain

$$\begin{aligned} (\text{dist}(x_{k+1}, X^*))^2 &\leq \|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 \\ &\quad - 2v \left(\frac{v}{2}(1+R)^2 + R \text{dist}(x_k, X^*) - R\|x^k - x^*\| - \frac{v}{2}(1+R)^2 \right) \\ &= (\text{dist}(x_k, X^*))^2. \end{aligned}$$

Hence, (4.9) holds for $i = k + 1$.

Therefore, by induction, (4.9) holds for all $i \geq k_0$. Since the right hand side of (4.9) is finite and X^* is compact, then $\{x_k\}$ is bounded. \square

When using the diminishing stepsize rule, we can also achieve the boundedness of the generated sequence as follows. The proof is omitted.

Lemma 4.4. Let Assumptions 2–5 hold and $\{x_k\}$ be generated by the approximate quasi-subgradient method with the diminishing stepsize rule. Then $\{x_k\}$ is bounded.

From Lemmas 4.3 and 4.4, one can see that $\{x_k\}$ is bounded, and hence, $\{f(x_k)\}$ is bounded from above due to the Hölder condition (cf. (3.1)), by using both types of stepsize rules. We denote by M the upper bound of $\{f(x_k)\}$ in what follows. Next, we first present the convergence property of the approximate quasi-subgradient method by using the constant stepsize rule.

Theorem 4.1. Let Assumptions 2–5 hold and $\{x_k\}$ be generated by the approximate quasi-subgradient method with the constant stepsize rule. Then, $z_{v,0}^*$ is finite and

$$\liminf_{k \rightarrow \infty} f(x_k) \leq f_* + z_{v,0}^*.$$

Proof. The finiteness of $z_{v,0}^*$ has been proved in Lemma 4.1(i). To prove the convergence property, we first show that

$$\liminf_{k \rightarrow \infty} f(x_k) < f_* + z_{v,0}^*$$

for all $\theta > 0$ by contradiction, that is, assume the following inequality holds for some $\theta > 0$,

$$\liminf_{k \rightarrow \infty} f(x_k) \geq f_* + z_{v,0}^*.$$

Thus, there exists some $\delta \in (0, \min\{\theta/2, z_{v,0}^*\})$ and positive integer k_0 such that

$$f(x_k) > f_* + z_{v,\theta}^* - \delta, \quad (4.10)$$

and

$$\epsilon_k < \epsilon + \theta/2, \quad (4.11)$$

for all $k \geq k_0$, where (4.11) holds due to $\limsup_{k \rightarrow \infty} \epsilon_k = \epsilon$.

By (4.4) and (4.10), we obtain $f(x_k) - f_* + \delta > \sup\{z : H_{v,\theta}^{x_k}(z) \geq 0\}$ and then $H_{v,\theta}^{x_k}(f(x_k) - f_* + \delta) < 0$, that is,

$$\begin{aligned} f(x_k) &> f_* + \mu \left(\frac{v}{2}(1+R)^2 + R \left(\frac{f(x_k) - f_* + \delta}{\eta} \right)^{1/g(\text{dist}(x_k, X^*))} \right)^p \\ &\quad + \epsilon + \theta - \delta > f_* + \mu \left(\frac{v}{2}(1+R)^2 \right. \\ &\quad \left. + R \left(\frac{f(x_k) - f_* + \delta}{\eta} \right)^{1/g(\text{dist}(x_k, X^*))} \right)^p + \epsilon_k \\ &\geq f_* + \mu \left(\frac{v}{2}(1+R)^2 + R \left(\frac{f(x_k) - f_*}{\eta} \right)^{1/g(\text{dist}(x_k, X^*))} + \delta' \right)^p + \epsilon_k, \end{aligned}$$

$$\forall k \geq k_0,$$

where the second inequality follows from (4.11) and $0 < \delta < \theta/2$, and the third inequality follows from the Taylor expansion with $\delta' = \min \left\{ \frac{\delta}{\eta q} \left(\frac{z_{v,\theta}^*}{\eta} \right)^{1/q-1}, \frac{\delta}{\eta p} \left(\frac{M-f_*}{\eta} \right)^{1/p-1} \right\} > 0$ (recall that M is the upper bound of $\{f(x_k)\}$). Therefore, by Lemmas 3.3 and 4.2, we obtain

$$\langle g_k/\|g_k\|, x_k - x^* \rangle \geq \frac{v}{2}(1+R)^2 + R \text{dist}(x_k, X^*) + \delta', \quad \forall k \geq k_0,$$

and thus,

$$\begin{aligned} (\text{dist}(x_{k+1}, X^*))^2 &\leq (\text{dist}(x_k, X^*))^2 - 2v\delta' \leq \dots \\ &\leq (\text{dist}(x_0, X^*))^2 - 2(k - k_0 + 1)v\delta', \end{aligned}$$

which yields a contradiction for sufficiently large k . Thus, we have

$$\liminf_{k \rightarrow \infty} f(x_k) \leq f_* + z_{v,\theta}^*, \quad \forall \theta > 0.$$

Taking the limit as $\theta \rightarrow 0$, by Lemma 4.1, we arrive at the conclusion. \square

We now give some explicit expressions for the total error in approaching f_* in Theorem 4.1 for specific cases of p and $g(t)$. By solving (4.5) and (4.6), we have the following corollaries where the total errors are given in explicit expressions.

Corollary 4.1. Let Assumptions 2–5 hold with $g(t) \equiv p$ and $p = 1$. Then, for a sequence $\{x_k\}$ generated by the approximate quasi-subgradient method with the constant stepsize rule, we have

$$\liminf_{k \rightarrow \infty} f(x_k) \leq f_* + \left(\frac{1}{2} \mu v(1+R)^2 + \epsilon \right) \frac{\eta}{\eta - R\mu}.$$

Proof. By assumptions, $g(t) \equiv p$ and $p = q = 1$, we have

$$G_{v,0}^p(z) = G_{v,0}^q(z) = \mu \left(\frac{v}{2}(1+R)^2 + R \frac{z}{\eta} \right) + \epsilon - z \quad \text{and} \quad z_{v,0}^p = z_{v,0}^q.$$

It is clear that $G_{v,0}^p(z)$ is linear and decreasing due to $R < \eta/\mu$. Thus, by (4.5), $z_{v,0}^p$ is just the solution of $G_{v,0}^p(z) = 0$. Then, by (4.6), we have $z_{v,0}^p = z_{v,0}^q = \left(\frac{1}{2} \mu v(1+R)^2 + \epsilon \right) \frac{\eta}{\eta - R\mu}$. Hence, by Theorem 4.1, we arrive at the conclusion. \square

Similar to Corollary 4.1, we obtain explicit expressions for the total error when $g(t) \equiv p$ and $p = 2$. The proof is straightforward, and thus, omitted.

Corollary 4.2. Let Assumptions 2–5 hold with $g(t) \equiv p$ and $p = 2$. Then, for a sequence $\{x_k\}$ generated by the approximate quasi-subgradient method with the constant stepsize rule, we have

$$\liminf_{k \rightarrow \infty} (x_k) \leq f_* + \eta \left(\frac{\mu v R (1+R)^2 + \sqrt{\eta \mu v^2 (1+R)^4 + 4\epsilon(\eta - \mu R^2)}}{2(\eta - \mu R^2)} \right)^2.$$

Using the diminishing stepsize rule, the total error tends to $z_{0,0}^*$ as v_k diminishes and the following theorem is obtained.

Theorem 4.2. *Let Assumptions 2–5 hold and $\{x_k\}$ be generated by the approximate quasi-subgradient method with the diminishing stepsize rule. Then $z_{0,0}^*$ is finite and*

$$\liminf_{k \rightarrow \infty} (x_k) \leq f_* + z_{0,0}^*.$$

So far, we have established the convergence property in objective values for approximate quasi-subgradient method and extended the corresponding results in Nedić and Bertsekas (2010) in Theorems 4.1 and 4.2 in the presence of generalized weak sharp minima. Specifying $g(t) \equiv p$ and $p = 1$, the generalized weak sharp minima is reduced to the weak sharp minima used in Nedić and Bertsekas (2010), and the obtained total errors (cf. Corollary 4.1) have similar formulae to that of Nedić and Bertsekas (2010, Propositions 3.1 & 3.2).

4.2. Finite convergence and convergence in iterates

In this subsection, by the virtual of the auxiliary function $H_{v,0}^x$ and its maximum solution $z_{v,0}^*$, we describe the finite convergence behavior and convergence of $\{x_k\}$ of the approximate quasi-subgradient method when the constraint set is noncompact. The line of analysis is similar to preceding sections, and thus, we omit the details.

Theorem 4.3. *Let Assumptions 2–5 hold, $\text{int} X^* \neq \emptyset$ and the diminishing stepsize rule be chosen. Then $f(x_k) \leq f_* + z_{0,0}^*$ for some k .*

Theorem 4.4. *Let Assumptions 2–5 hold. If $B(\bar{x}, \bar{\delta}) \subset X^*$ with $\bar{\delta} > 0$ and there exists some $0 < \kappa < 1$ and $k_0 \in \mathbb{N}$ such that $v_k \in \left[\frac{\kappa^2 \bar{\delta}}{(1+R)^2}, \frac{\kappa \bar{\delta}}{(1+R)^2} \right]$ for all $k \geq k_0$, then $f(x_k) \leq f_* + z_{0,0}^*$ for some k .*

Theorem 4.5. *Let Assumptions 2–5 hold with $z_{0,0}^* > 0$ (cf. (4.4)), f be usc on \mathbb{R}^n and the diminishing stepsize rule be chosen. Then the following statements are true:*

- (i) $\liminf_{k \rightarrow \infty} \text{dist}(x_k, \bar{S}_{f, f_* + z_{0,0}^*} \cap X) = 0$.
- (ii) $\lim_{k \rightarrow \infty} \text{dist}(x_k, X^* + \rho(z_{0,0}^*)B) = 0$, where $\rho(z_{0,0}^*)$ is defined by

$$\rho(z_{0,0}^*) := \max \left\{ \text{dist}(x, X^*) : x \in \bar{S}_{f, f_* + z_{0,0}^*} \cap X \right\}.$$

5. Efficiency

In this section, under the bounded assumption (see Assumptions 1 and 3), we discuss the efficiency estimates of the approximate quasi-subgradient method. In order to quantify the efficiency, we introduce some concepts as in Kiwiel (2001).

The inradius of a set Z denotes the radius of the largest ball contained in Z , defined by

$$\dot{r}(Z) := \sup \{ r > 0 : B(x, r) \subset Z \text{ for some } x \in Z \}. \quad (5.1)$$

For any $\gamma \in (0, 1)$, the γ -solution set of problem (1.1) is defined by $X_\gamma^* := \{x \in X : \dot{r}(S_f(x)) < \gamma \dot{r}(X)\}$. (5.2)

It follows from (5.2) that x is an γ -solution of problem (1.1) if $x \in X$ and $S_f(x)$ does not contain a ball with radius $\gamma \dot{r}(X)$. Thus, the

significance of inradius is to estimate the efficiency of algorithms, inasmuch as x is an γ -solution if $\dot{r}(S_f(x)) < \gamma \dot{r}(X)$. The criterion is that the quality of iterate improves if the inradius of its strict sublevel set decreases.

At iteration $k \geq 1$, the record value $f_{\epsilon,k}^{\text{rec}}$ denotes the best approximate value found so far, and is defined by

$$f_{\epsilon,k}^{\text{rec}} := \min_{j=1,\dots,k} \{f(x_j) - \epsilon_j\}. \quad (5.3)$$

Let \dot{r}_k denote the inradius of the record strict sublevel set, defined by

$$\dot{r}_k := \dot{r}(S_{f_{\epsilon,k}^{\text{rec}}}),$$

which is nonincreasing in k .

In view of application considerations, we would like our algorithm to reach the γ -solution set as fast as possible. Since the quality of the record value/point improves if the inradius \dot{r}_k decreases (cf. Kiwiel, 2001, Lemma 13), we would like \dot{r}_k to decrease as fast as possible. For this purpose, we now give an upper bound of \dot{r}_k that depends on the stepsize rule.

Lemma 5.1. *Let Assumptions 1 and 3 hold. For a sequence $\{x_k\}$ generated by the approximate quasi-subgradient method, we have*

$$\dot{r}_k \leq Rd + \frac{d^2 + (1+R)^2 \sum_{j=1}^k v_j^2}{2 \sum_{j=1}^k v_j}, \quad \text{for } i = 1, \dots, k. \quad (5.4)$$

Proof. Suppose $\dot{r}_k > 0$. For any $\delta < \dot{r}_k$, it follows from (5.1) that there exists some \bar{x} such that $B(\bar{x}, \delta) \subset S_{f_{\epsilon,k}^{\text{rec}}}$. Then for each $j = 1, \dots, k$, from (5.3), we have $B(\bar{x}, \delta) \subset S_{f(x_j) - \epsilon_j}$. Hence, it follows from Lemma 3.2 that

$$\langle g_j / \|g_j\|, x_j - \bar{x} \rangle \geq \delta, \quad \text{for } j = 1, \dots, k.$$

Therefore, from Lemma 3.1, we have

$$\|x_{j+1} - \bar{x}\|^2 \leq \|x_j - \bar{x}\|^2 - 2v_j \delta + 2v_j R d + v_j^2 (1+R)^2.$$

Summing these inequalities over $j = 1, \dots, k$, we arrive at

$$\delta \leq Rd + \frac{d^2 + (1+R)^2 \sum_{j=1}^k v_j^2}{2 \sum_{j=1}^k v_j}, \quad \text{for } i = 1, \dots, k.$$

Since $\delta < \dot{r}_k$ is arbitrary, we arrive at the conclusion. \square

In the sense of guaranteeing that the record values/points become γ -solutions as fast as possible, the best stepsize may be found by minimizing the upper bound of \dot{r}_k in (5.4). In the following, we offer the best choice on the constant stepsize rule and estimate the rate of efficiency by using the diminishing stepsize rule.

Theorem 5.1. *Let Assumptions 1 and 3 hold. For a sequence $\{x_k\}$ generated by the approximate quasi-subgradient method, the following statements hold:*

- (i) *If a constant stepsize v is chosen, then $\dot{r}_k \leq \frac{d^2}{2kv} + Rd + \frac{v}{2}(1+R)^2$.*
- (ii) *The best constant stepsize is $v_i = \frac{d}{(1+R)\sqrt{k}}$ and $\dot{r}_k \leq \frac{d(1+R)}{\sqrt{k}} + Rd$.*
- (iii) *If the diminishing stepsize is chosen as $v_i = a/\sqrt{i}$, then*

$$\dot{r}_k \leq Rd + ck^{-1/2} \quad \text{with } c = \frac{d^2 + a^2(1+\ln 2)(1+R)^2}{a(4-2\sqrt{2})}.$$

More general, if v_k is chosen as the diminishing stepsize rule, then $\lim_{k \rightarrow \infty} \dot{r}_k \leq Rd$.

Table 1

Computation results for maximizing the Cobb–Douglas production efficiency.

Projects	Factors	QSM		AQSM-R ($R = 1$)		AQSM- ϵ ($\epsilon = 1$)	
		f_{opt}	CPU time	f_{opt}	CPU time	f_{opt}	CPU time
10	10	0.2266	0.17	0.2280	0.16	0.2271	0.17
50	50	0.0548	0.20	0.0532	0.20	0.0537	0.22
100	100	0.0349	0.26	0.0309	0.28	0.0333	0.27
500	500	0.0059	1.13	0.0047	0.81	0.0056	0.70
1000	1000	0.0027	1.68	0.0022	1.86	0.0025	1.62
2000	2000	0.0013	6.88	0.0011	6.00	0.0013	5.88

Proof.

- (i) It is (5.4) specifying $i = 1$ and $v_i \equiv v$.
(ii) Minimizing the upper bound of \hat{r}_k in (i) with respect to v , we obtain the best constant stepsize $v = \frac{d}{(1+R)\sqrt{k}}$ and the corresponding upper bound on the inradius.
(iii) It follows from Nesterov (1989, p. 157) that

$$\sum_{j=i}^k j^{-1} \leq 1 + \ln 2 \quad \text{and} \quad \sum_{j=i}^k j^{-1/2} \geq (2 - \sqrt{2})k^{1/2}, \quad \text{for } i = \left\lceil \frac{k}{2} \right\rceil.$$

Using (5.4), we obtain

$$\hat{r}_k \leq Rd + \frac{d^2 + a^2(1 + \ln 2)(1 + R)^2}{a(4 - 2\sqrt{2})k^{1/2}} = Rd + ck^{-1/2}.$$

Furthermore, the property of the diminishing stepsize rule implies $\lim_{k \rightarrow \infty} \left(\sum_{j=i}^k v_j^2 / \sum_{j=i}^k v_j \right) = 0$ (cf. Kiwiel, 2004, Lemma 2.1), and thus (5.4) implies $\lim_{k \rightarrow \infty} \hat{r}_k \leq Rd$. \square

6. Numerical experiments

Fractional programming is widely used in the modeling of practical problems arising in various areas, such as economics, information theory, management science and applied physics. In fractional programming problems, the objective is to optimize certain indicator (efficiency), characterized by a ratio of technical and economical terms, subject to the constraint imposed on the availability of goods. Examples of such situations are financial and corporate planning (debt/equity ratio), production planning (inventory/sales, output/employee), health care and hospital planning (cost/patient, nurse/patient ratio), etc. For details, one can refer to Avriel et al. (1988), Crouzeix et al. (1998), Hadjisavvas et al. (2005), Stancu-Minasian (1997) and references therein.

We consider the Cobb–Douglas production efficiency problem introduced by Bradley and Frey (1974). The problem is briefly described as follows. Consider a set of projects $i = 1, \dots, m$ and a collection of production factors $j = 1, \dots, n$, the total profit value assigned to these projects is given by the following Cobb–Douglas production function

$$\text{Profit} = a_0 \prod_{j=1}^n x_j^{a_j}, \quad \text{where} \quad \sum_{j=1}^n a_j = 1,$$

where the variables x_j designate the production factors. The Cobb–Douglas production function represents the relationship between the input variable specifying the production factors and the output variables specifying the results of the production activities. The total cost is a linear function of the levels of investment in these projects, denoted by

$$\text{Cost} = \sum_{j=1}^n c_j x_j + c_0.$$

The production efficiency problem is to maximize the profit/cost ratio, which is an efficiency indicator, i.e., the ratio between what is obtained and the expenditure, subject to a variety of constraints on funding levels. Hence, the Cobb–Douglas production efficiency model is stated as

$$\begin{aligned} \max \quad & f(x) := \frac{a_0 \prod_{j=1}^n x_j^{a_j}}{\sum_{j=1}^n c_j x_j + c_0} \\ \text{s.t.} \quad & \sum_{j=1}^n b_{ij} x_j \geq p_i, \quad i = 1, \dots, m, \\ & x \geq 0, \end{aligned} \quad (6.1)$$

where p_i represents the profit that must be obtained at project i and b_{ij} represents the contribution of the production factor j to project i to realize the profit p_i . According to the circumstance of the Cobb–Douglas production efficiency problem, all parameters on profit (a_j) and cost (c_j) are all positive. From Stancu-Minasian (1997, Theorems 2.3.3 & 2.5.1), it is clear that (6.1) is a quasi-concave maximization problem.

We conduct all numerical experiments in a personal laptop (Intel Core i7, 2.00 GHz, 8.00 GB of RAM) using MATLAB R2009a. In the numerical experiments, the parameters of the problem (6.1) are randomly chosen from different intervals,

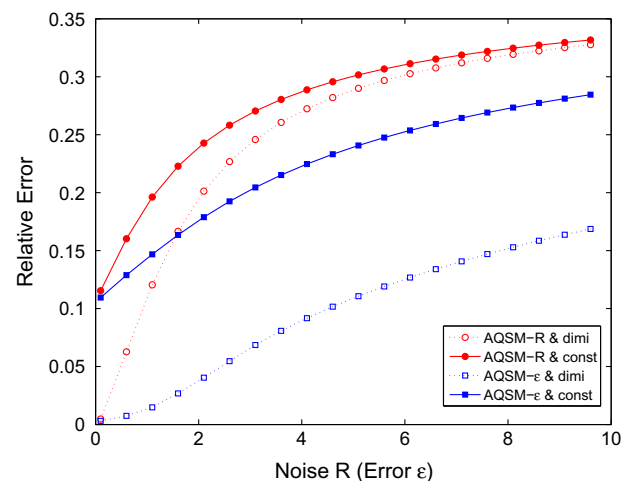
$$a_j, b_{ij} \in [0, 1], \quad a_0, c_0, c_j \in [0, 10], \quad \text{and} \quad p_i \in [0, n/2].$$

The diminishing stepsize rule is chosen as

$$v_k = v / (1 + 0.1k),$$

where v is always chosen between [2, 5], while the constant stepsize is selected between [0.2, 0.5]. The larger the problem size, the larger the stepsize.

We first show the performance (in both optimal value and CPU time) of the approximate quasi-subgradient algorithm using the diminishing stepsize rule for different dimensions. The

**Fig. 2.** Sensitivity analysis on noise and error, respectively.

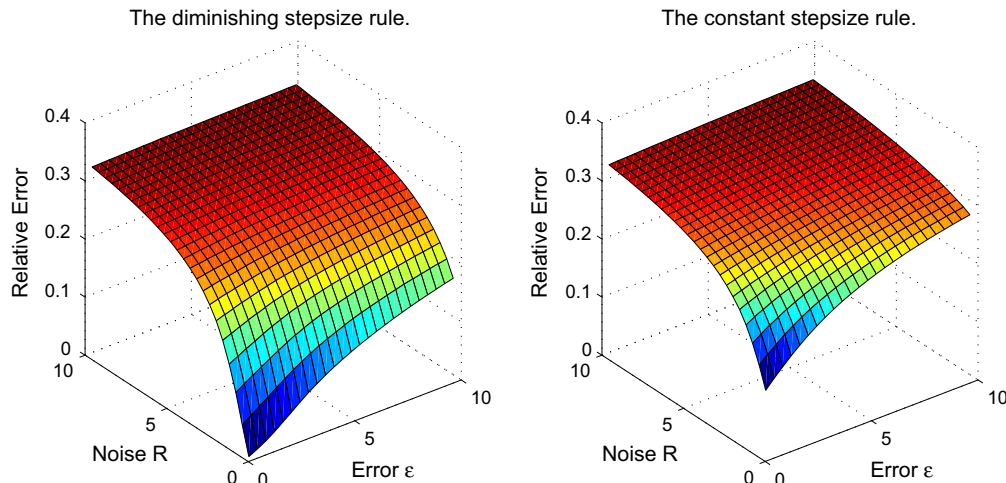


Fig. 3. Sensitivity analysis on noise and error simultaneously.

computation results are displayed in Table 1. In this table, QSM (resp. AQSM-R, AQSM- ϵ) denotes the exact quasi-subgradient method (resp. the approximate quasi-subgradient method with noise only, the approximate quasi-subgradient method with error only), the columns of Projects and Factors represent the numbers of projects and production factors of the problem (6.1) respectively, f_{opt} and CPU time denote the obtained optimal value and the CPU time (seconds) cost to reach f_{opt} by each algorithm, respectively.

From the results in Table 1, we can see that the quasi-subgradient type methods are highly efficient for the Cobb–Douglas production efficiency problem, even when the problem is large-scale. In the presence of persistent noise ($R = 1$) or error ($\epsilon = 1$), there are some tolerances from the optimal value of the QSM, which is consistent with the theoretical analysis in the preceding section. We can also note that the AQSM- ϵ achieves the better optimal value than the AQSM-R.

The second experiment is performed to study the sensitivity analysis on noise and error, by using both the constant and diminishing stepsize rules. In this experiment, we fix the problem size 100×100 , generate the noise and error series on $[0, 10]$, respectively. We characterize the performance by the relative error of optima value $(f^* - f_{opt})/f^*$, where f^* is the optimal value obtained by the QSM.

The numerical results, plotted in Fig. 2, are consistent with the theoretical analysis in Section 3. Although the constraint set of the problem (6.1) may be noncompact, the optimal solution and the iterates are always placed in some bounded area. Recall that Theorems 3.1 and 3.2 provide tolerances away from the optimal value of the forms

$$\mu \left(Rd + \frac{\nu}{2} (1 + R)^2 \right)^p + \epsilon \quad \text{and} \quad \mu (Rd)^p + \epsilon,$$

respectively by using the constant and diminishing stepsize rules, where $p < 1$ as $a_j < 1$ in the problem (6.1). In absence of the error ϵ , the curves (plotted by \circ) of AQSM-R basically fit the exponential form of tolerance. When the noise R vanishes, the curves (plotted by \square) of AQSM- ϵ verify the linear dependence of tolerance on ϵ .

We further analyze the sensitivity behavior on noise and error simultaneously. The results are plotted in Fig. 3, where the left one is for the diminishing stepsize rule and the right one is for the constant stepsize rule. These results are also consistent with the theoretical analysis in Theorems 3.1 and 3.2.

We also test the global convergence property of the QSM by randomly selecting initial starting points. We adopt the same

diminishing stepsize rule as the one used in Table 1, that is $\nu_k = 3/(1 + 0.1k)$, and start from several different initial points, either feasible or infeasible. As long as the iteration number is taken large enough, the sequence of the function values always converges to the same value. Also, the QSM starting from feasible points significantly outperforms when starting from infeasible points.

Acknowledgments

The authors are grateful to the anonymous reviewer for his/her valuable suggestions and remarks which helped to improve the quality of the paper. The work was supported by the Research Grants Council of Hong Kong [PolyU 5306/11E].

References

- Auslender, A., & Teboulle, M. (2004). Interior gradient and ϵ -subgradient descent methods for constrained convex minimization. *Mathematics of Operations Research*, 29, 1–26.
- Aussel, D., Corvellec, J.-N., & Lassonde, M. (1995). Mean value property and subdifferential criteria for lower semicontinuous functions. *Transactions of the American Mathematical Society*, 347, 4147–4161.
- Avriel, M., Diewert, W. E., Schaible, S., & Zang, I. (1988). *Generalized concavity*. New York: Plenum Press.
- Bertsekas, D. P., Nedić, A., & Ozdaglar, A. (2003). *Convex analysis and optimization*. Cambridge: Athena Scientific.
- Bonnans, J. F., & Ioffe, A. (1995). Second-order sufficiency and quadratic growth for nonisolated minima. *Mathematics of Operations Research*, 20, 801–817.
- Bradley, S. P., & Frey, S. C. (1974). Fractional programming with homogeneous functions. *Operations Research*, 22, 350–357.
- Burachik, R. S., Gasimov, R. N., Ismayilova, N. A., & Kaya, C. Y. (2006). On a modified subgradient algorithm for dual problems via sharp augmented Lagrangian. *Journal of Global Optimization*, 34, 55–78.
- Burke, J. V., & Ferris, M. C. (1993). Weak sharp minima in mathematical programming. *SIAM Journal on Control and Optimization*, 31, 1340–1359.
- Crouzeix, J.-P., Martinez-Legaz, J.-E., & Volle, M. (1998). *Generalized convexity, generalized monotonicity*. Dordrecht: Kluwer Academic Publishers.
- D'Antonio, G., & Frangioni, A. (2009). Convergence analysis of deflected conditional approximate subgradient methods. *SIAM Journal on Optimization*, 20, 357–386.
- Dyer, M. E. (1980). Calculating surrogate constraints. *Mathematical Programming*, 19, 255–278.
- Ermoliev, Y. M. (1966). Methods of solution of nonlinear extremal problems. *Cybernetics and Systems Analysis*, 2, 1–14.
- Gasimov, R. N. (2002). Augmented Lagrangian duality and nondifferentiable optimization methods in nonconvex programming. *Journal of Global Optimization*, 24, 187–203.
- Greenberg, H. J., & Pierskalla, W. P. (1973). Quasi-conjugate functions and surrogate duality. *Cahiers Centre Études Recherche Opérationnelle*, 15, 437–448.
- Hadjisavvas, N., Komlósi, S., & Schaible, S. (2005). *Handbook of generalized convexity and generalized monotonicity*. New York: Springer-Verlag.
- Hiriart-Urruty, J.-B., & Lemaréchal, C. (1996). *Convex analysis and minimization algorithms*. Berlin: Springer-Verlag.

- Kashyap, A., Basar, T., & Srikant, R. (2007). Quantized consensus. *Automatica*, 43, 1192–1203.
- Kiwiel, K. C. (2001). Convergence and efficiency of subgradient methods for quasiconvex minimization. *Mathematical Programming*, 90, 1–25.
- Kiwiel, K. C. (2004). Convergence of approximate and incremental subgradient methods for convex optimization. *SIAM Journal on Optimization*, 14, 807–840.
- Konnov, I. V. (1994). On properties of supporting and quasi-supporting vectors. *Journal of Mathematical Sciences*, 71, 2760–2763.
- Larsson, T., Patriksson, M., & Strömberg, A.-B. (1996). Conditional subgradient optimization—theory and applications. *European Journal of Operational Research*, 88, 382–403.
- Larsson, T., Patriksson, M., & Strömberg, A.-B. (2003). On the convergence of conditional ϵ -subgradient methods for convex programs and convex–concave saddle-point problems. *European Journal of Operational Research*, 151, 461–473.
- Nedić, A., & Bertsekas, D. P. (2001). Incremental subgradient methods for nondifferentiable optimization. *SIAM Journal on Optimization*, 12, 109–138.
- Nedić, A., & Bertsekas, D. P. (2010). The effect of deterministic noise in subgradient methods. *Mathematical Programming*, 125, 75–99.
- Nesterov, Y. (1989). *Effective methods in nonlinear programming*. Moscow: Radio i Svyaz.
- Nesterov, Y. (2009). Primal-dual subgradient methods for convex problems. *Mathematical Programming*, 120, 221–259.
- Pang, J.-S. (1997). Error bounds in mathematical programming. *Mathematical Programming*, 79, 299–332.
- Patriksson, M. (2008). A survey on the continuous nonlinear resource allocation problem. *European Journal of Operational Research*, 185, 1–46.
- Polyak, B. T. (1967). A general method for solving extremum problems. *Soviet Mathematics Doklady*, 8, 593–597.
- Rabbat, M. G., & Nowak, R. D. (2005). Quantized incremental algorithms for distributed optimization. *IEEE Journal on Selected Areas in Communications*, 23, 798–808.
- Rockafellar, R. T. (1970). *Convex analysis*. Princeton: Princeton University Press.
- Rockafellar, R. T., & Wets, R. J. B. (1998). *Variational analysis*. Berlin: Springer-Verlag.
- Shor, N. Z., Kiwiel, K. C., & Ruszczyński, A. (1985). *Minimization methods for non-differentiable functions*. New York: Springer-Verlag.
- Stancu-Minasian, I. M. (1997). *Fractional programming*. Dordrecht: Kluwer Academic Publishers.
- Studniarski, M., & Ward, D. E. (1999). Weak sharp minima: Characterizations and sufficient conditions. *SIAM Journal on Control and Optimization*, 38, 219–236.