Convergence Rates of Subgradient Methods for Quasi-convex Optimization Problems

Yaohua Hu*,   Jiawen Li†   Carisa Kwok Wai Yu‡

Abstract   Quasi-convex optimization acts a pivotal part in many fields including economics and finance; the subgradient method is an effective iterative algorithm for solving large-scale quasi-convex optimization problems. In this paper, we investigate the quantitative convergence theory, including the iteration complexity and convergence rates, of various subgradient methods for solving quasi-convex optimization problems in a unified framework. In particular, we consider a sequence satisfying a general (inexact) basic inequality, and investigate the global convergence theorem and the iteration complexity when using the constant, diminishing or dynamic stepsize rules. More importantly, we establish the linear (or sublinear) convergence rates of the sequence under an additional assumption of weak sharp minima of Hölderian order and upper bounded noise. These convergence theorems are applied to establish the iteration complexity and convergence rates of several subgradient methods, including the standard/inexact/conditional subgradient methods, for solving quasi-convex optimization problems under the assumptions of the Hölder condition and/or the weak sharp minima of Hölderian order.

Keywords   Quasi-convex programming, subgradient method, iteration complexity, convergence rates.

AMS subject classifications   Primary, 65K05, 90C26; Secondary, 49M37.

1   Introduction

Mathematical optimization is a fundamental tool for solving decision-making problems in many disciplines. Convex optimization plays a key role in mathematical optimization, but

may not be applicable to some problems encountered in economics, finance and management science. In contrast to convex optimization, quasi-convex optimization usually provides a much more accurate representation of realities and still possesses some desirable properties of convex optimization. In recent decades, more and more attention has been drawn to quasi-convex optimization; see [1, 4, 11, 19, 32, 44, 45, 46, 47, 50] and references therein.

It is essential to develop efficient numerical algorithms for solving quasi-convex optimization problems. Subgradient methods form a class of practical and effective iterative algorithms for solving constrained (convex or quasi-convex) optimization problems. The subgradient method was originally introduced by Polyak [41] and Ermoliev [12] in the 1970s to solve nondifferentiable convex optimization problems. Over the past 40 years, various features of subgradient methods have been established for convex optimization problems and many extensions/generalizations have been devised for structured convex optimization problems; see [2, 6, 20, 27, 30, 31, 35, 37, 38, 49, 52] and references therein. Moreover, subgradient methods have also been extended and developed to solve constrained quasi-convex optimization problems; see [21, 23, 26, 29, 48, 55] and references therein. The global convergence properties of subgradient methods, in terms of function values and distances of iterates from the optimal solution set, have been well studied for either convex or quasi-convex optimization problems by using several typical stepsize rules.

In addition to the global convergence property, the establishment of convergence rates is important in analyzing the numerical performance of relevant algorithms. The convergence rates of subgradient methods for solving convex optimization problems have been investigated under the assumption of weak sharp minima [10]. For example, employing a geometrically decaying stepsize, Shor [49, Theorem 2.7] and Goffin [16, Theorem 4.4] provided a linear convergence analysis (but not necessarily converge to an optimal solution) of the subgradient method under a notion of condition number that is a stronger condition than the weak sharp minima. This work was extended by Polyak [40, Theorem 4] to the case when the subgradients are corrupted by deterministic noise. When using a dynamic stepsize rule, Brannlund [8, Theorem 2.5] established the linear convergence rate of the subgradient method under the assumption of weak sharp minima, which was generalized by Robinson [43, Theorem 3] to the inexact subgradient method. Recently, for vast applications of distributed optimization, an incremental subgradient method was proposed to solve the convex sum-minimization problem and was shown to converge linearly under the assumption of weak sharp minima (see [34, Proposition 2.11]); the linear convergence analysis has also been extended to various variant incremental subgradient methods; see [18, 33] and references therein. It is worth noting that all convergence rates of subgradient methods are measured in terms of distances of iterates from the optimal solution set. Moreover, Freund and Lu [15] and Xu et al. [54] investigated the iteration complexity (for achieving an approximate optimal value) of subgradient methods and stochastic subgradient methods under the assumption of weak sharp minima or weak sharp minima of Hölderian order, respectively. However, to the best of our knowledge, there is limited study devoted to establishing convergence rates of subgradient methods in the scenario of quasi-convex optimization.

2

Motivated by the wide applications of quasi-convex optimization, we contribute to the further development of subgradient methods for solving quasi-convex optimization problems, particularly concentrating on the iteration complexity and convergence rate analysis. In the present paper, we consider the following constrained quasi-convex optimization problem

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is a quasi-convex and continuous function, and $X \subseteq \mathbb{R}^n$ is a nonempty, closed and convex set. In 2001, Kiwiel [26] proposed a (standard) subgradient method to solve problem (1.1), which is formally stated as follows, where $\partial^* f$, $\mathbf{S}$ and $\mathbb{P}_X(\cdot)$ denote the quasi-subdifferential of f, the unit sphere centered at the origin and the Euclidean projection onto X, respectively.

Algorithm 1.1. Select an initial point $x_1 \in \mathbb{R}^n$ and a sequence of stepsizes $\{v_k\} \subseteq (0, +\infty)$. For each $k \in \mathbb{N}$, having x_k, we select $g_k \in \partial^* f(x_k) \cap \mathbf{S}$ and update x_{k+1} by

$$x_{k+1} := \mathbb{P}_X(x_k - v_k g_k).$$

Note that the basic inequality of a subgradient iteration is a key tool for the convergence analysis of subgradient methods for either convex or quasi-convex optimization. Yu et al. [55] developed a unified framework of convergence analysis for various subgradient methods, in which the global convergence theorem was established for a certain class of sequences satisfying a general basic inequality and using several typical stepsize rules. In real applications, the computation error stems from practical considerations and is inevitable in the computing process. To meet the requirements of practical applications, Hu et al. [21] introduced an inexact subgradient method to solve problem (1.1) and investigated the influence of the deterministic noise on the inexact subgradient method. In the present paper, motivated by the practical applications, we consider a more general unified framework for subgradient methods, in which a general (inexact) basic inequality is assumed to be satisfied; see conditions (H1) and (H2) in Section 3. The more general unified framework covers various subgradient methods (see [55]) and inexact subgradient methods (see Section 4.2) for either convex or quasi-convex optimization problems.

The main contribution of the present paper is to investigate the iteration complexity and convergence rates of several subgradient methods for solving quasi-convex optimization problem (1.1) via a unified framework. The stepsize rules adopted in this paper are the constant, diminishing and dynamic stepsizes. Here, we first consider a general sequence satisfying conditions (H1) and (H2), (as a by-product) establish the global convergence theorem in terms of function values and distances of iterates from the optimal solution set (see Theorem 3.1), and then derive the iteration complexity to obtain an approximate optimal solution (see Theorem 3.2 and Remark 3.2). More importantly, we explore the linear (or sublinear) convergence rates of the sequence under an additional assumption of weak sharp minima of Hölderian order and upper bounded noise; see Theorems 3.3-3.5 for details. Meanwhile, we will apply

3

the established theorems on iteration complexity and convergence rates to investigate that of several subgradient methods for solving the quasi-convex optimization problem (1.1) when using the typical stepsize rules and under the assumptions of the Hölder condition and/or the weak sharp minima of Hölderian order. As far as we know, the study of iteration complexity and convergence rates theorems of subgradient methods are new in the literature of quasi-convex optimization. Moreover, the finite convergence property of the subgradient methods are established under an interior point assumption of the solution set.

In a very recent study by Johnstone and Moulin [25], by virtue of the weak sharp minima of Hölderian order, they derived the convergence rates of the standard subgradient method for solving convex optimization problems when using the constant-type stepsize rules (including the constant and "descending stairs" stepsize rules). Our work extends [25] to more general settings with quasi-convex optimization, inexact setting, diminishing and dynamic stepsize rules.

The present paper is organized as follows. In Section 2, we present the notations and preliminary results which will be used in this paper. In Section 3, we provide a unified framework of convergence analysis for (inexact) subgradient methods, in which the global convergence theorem, iteration complexity and convergence rates are established for a certain class of sequences satisfying a general (inexact) basic inequality under the assumption of weak sharp minima of Hölderian order. In Section 4, the convergence analysis framework is applied to establish the iteration complexity and convergence rates for the standard subgradient method, the inexact subgradient method and the conditional subgradient method for solving quasi-convex optimization problems.

2 Notations and preliminary results

The notations used in the present paper are standard; see, e.g., [6]. 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 $\mathbb{B}(x, r)$ to denote the closed ball centered at x with radius r, and use $\mathbf{S}$ to denote the unit sphere centered at the origin. For $x \in \mathbb{R}^n$ and $Z \subseteq \mathbb{R}^n$, the Euclidean distance of x from Z and the Euclidean projection of x onto Z are respectively defined by

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

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

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

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

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

4

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

The convex subdifferential $\partial f(x) := \{g \in \mathbb{R}^n : f(y) \ge f(x) + \langle g, y - x\rangle, \forall y \in \mathbb{R}^n\}$ might be empty for the quasi-convex functions. Hence, the introduction of (nonempty) subdifferential of quasi-convex functions plays an important role in quasi-convex optimization. Several different types of subdifferentials of quasi-convex functions have been introduced in the literature, see [3, 17, 21, 26] and references therein. In particular, Kiwiel [26] and Hu et al. [21] introduced a quasi-subdifferential and applied this quasi-subgradient in their proposed subgradient methods; see, e.g., [21, 23, 26].

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*

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

and

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

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.

It is clear from definition that the quasi-subdifferential is a normal cone to the strict sublevel set of the quasi-convex function, and it was shown in [21, Lemma 2.1] that $\partial^* f(x) \setminus \{0\} \ne \emptyset$ whenever f is quasi-convex. Hence, the quasi-subdifferential of a quasi-convex function contains at least a unit vector. This is a special property that the convex subdifferential does not share. In particular, it was claimed in [21] that the quasi-subdifferential coincides with the convex cone hull of the convex subdifferential whenever f is convex.

The Hölder condition (restricted to the set of minima) was used in [28] to describe some properties of the quasi-subgradient, and it plays a critical role in the study of convergence analysis in quasi-convex optimization; see, e.g., [21, 55]. The notion of Hölder condition has been widely studied and applied in harmonic analysis, fractional analysis and management science; see, e.g., [3, 50]. It is worth noting that the classical Lipschitz condition (i.e., Hölder condition of order 1) is equivalent to the bounded subgradient assumption, which is always assumed in the literature of subgradient methods (see, e.g., [6, 27, 49]), whenever f is convex. f^* and X^* denote the optimal value and the solution set of problem (1.1), respectively.

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

$$f(x) - f^* \le L\mathrm{dist}^p(x, X^*) \quad \textit{for each } x \in \mathbb{R}^n.$$

We end this section by the following lemmas, which are useful to establish the unified framework of convergence analysis. In particular, Lemmas 2.1 is taken from [27, Lemma 2.1].

5

Lemma 2.1. *Let $\{a_k\}$ be a scalar sequence and let $\{w_k\}$ be a sequence of nonnegative scalars. Suppose that $\lim_{k\to\infty}\sum_{i=1}^{k}w_i = \infty$. Then, it holds that*

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

Lemma 2.2. *Let $r > 0$, $a > 0$, $b \ge 0$, and let $\{u_k\}$ be a sequence of nonnegative scalars such that*

$$u_{k+1} \le u_k - au_k^{1+r} + b \quad \text{for each } k \in \mathbb{N}. \tag{2.1}$$

(i) *If $b = 0$, then*

$$u_{k+1} \le u_1 \left(1 + rau_1^r k\right)^{-\frac{1}{r}} \quad \text{for each } k \in \mathbb{N}.$$

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

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

Proof. Assertion (i) of this lemma is taken from [42, pp. 46, Lemma 6], then it remains to prove assertion (ii). For this purpose, let $u := \left(\frac{b}{a}\right)^{\frac{1}{1+r}}$. Then, (2.1) is reduced to

$$u_{k+1} - u \le u_k - u - a\left(u_k^{1+r} - u^{1+r}\right) \quad \text{for each } k \in \mathbb{N}. \tag{2.2}$$

As $r > 0$, by the convexity of $h(t) := t^{1+r}$, one has that $u_k^{1+r} - u^{1+r} \ge (1+r)u^r(u_k - u)$. Then (2.2) implies that

$$u_{k+1} - u \le \left(1 - a(1+r)u^r\right)(u_k - u) \quad \text{for each } k \in \mathbb{N}.$$

Let $\tau := 1 - a(1+r)u^r$. It follows from the assumption that $\tau \in (0,1)$, and thus,

$$u_{k+1} - u \le \tau^k(u_1 - u) \quad \text{for each } k \in \mathbb{N}.$$

The conclusion follows and the proof is complete. $\qquad\square$

Lemma 2.3. *Let $a > 0$, $b > 0$, $s \in (0,1)$ and $t \ge s$, and let $\{u_k\}$ be a sequence of nonnegative scalars such that*

$$u_{k+1} \le \left(1 - ak^{-s}\right)u_k + bk^{-t} \quad \text{for each } k \in \mathbb{N}. \tag{2.3}$$

(i) *If $t > s$, then*

$$u_{k+1} \le \frac{b}{a}k^{s-t} + o(k^{s-t}).$$

(ii) *If $t = s$, then there exists $\tau \in (0,1)$ such that*

$$u_{k+1} \le u_1 e^{\frac{as}{1-s}} \tau^k + \frac{b}{a} \quad \text{for each } k \in \mathbb{N}.$$

Proof. Assertion (i) of this lemma is taken from [42, pp. 46, Lemma 5], then it remains to prove assertion (ii). To this end, we derive by (2.3) (when $t = s$) that, for each $k \in \mathbb{N}$,

$$u_{k+1} - \frac{b}{a} \leq \left(1 - ak^{-s}\right)\left(u_k - \frac{b}{a}\right) \leq \left(u_1 - \frac{b}{a}\right)\prod_{i=1}^{k}(1 - ai^{-s}). \tag{2.4}$$

Note that

$$\prod_{i=1}^{k}(1 - ai^{-s}) = e^{\sum_{i=1}^{k}\ln(1 - ai^{-s})} < e^{\int_1^{k+1}\ln(1 - at^{-s})\mathrm{d}t} < e^{\int_1^{k+1} -at^{-s}\mathrm{d}t} \leq e^{\frac{as}{1-s}}\tau^k,$$

where $\tau := e^{-a} \in (0, 1)$. This, together with (2.4), yields the conclusion. $\qquad\square$

3 A unified framework for subgradient methods

In the present paper, we discuss subgradient methods for solving the quasi-convex optimization problem (1.1), in which the solution set and the optimal value are denoted by X^* and f^*, respectively. The class of subgradient methods is one of the most popular numerical algorithms for solving constrained optimization problems. In view of the procedure of subgradient methods, the basic inequality of a subgradient iteration is an important property and plays as a key tool for convergence analysis of subgradient methods for either convex or quasi-convex optimization problems.

This section aims to investigate the iteration complexity and convergence rates of subgradient methods via a unified framework, in which a general (inexact) basic inequality is assumed to be satisfied. In particular, we fix $\epsilon \geq 0$ and $p \in (0, 1]$, and consider a sequence $\{x_k\} \subseteq X$ that satisfies the following two conditions:

(H1) For each $x^* \in X^*$ and each $k \in \{i \in \mathbb{N} : f(x_i) > f^* + \epsilon\}$,

$$\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 \leq -\alpha_k v_k(f(x_k) - f^* - \epsilon)^{\frac{1}{p}} + \beta_k v_k^2. \tag{3.1}$$

(H2) $\{\alpha_k\}$ and $\{\beta_k\}$ are two sequences of positive scalars such that

$$\lim_{k \to \infty} \alpha_k = \alpha > 0 \quad \text{and} \quad \lim_{k \to \infty} \beta_k = \beta > 0. \tag{3.2}$$

Condition (H1) measures the difference between two distances of iterates from a possible solution by calculating the difference between the function value and the optimal value with a noise, and condition (H2) characterizes some assumptions on the parameters. In the special case when $\epsilon = 0$, conditions (H1) and (H2) are reduced to the unified framework studied in [55], where the global convergence theorem was established, but no convergence rate analysis. The nature of subgradient methods forces the generated sequence to comply with conditions (H1) and (H2) under some mild assumptions, and thus, this study provides a unified framework for various subgradient methods for either convex or quasi-convex optimization problems (one can also refer to [55] for details).

- For convex optimization problems and under a bounded subgradient assumption, condition (H1) with $p = 1$ and (H2) are satisfied for the subgradient-type methods, including the standard subgradient method [49], the approximate subgradient method [27], the primal-dual subgradient method [36], the incremental subgradient method [35], the conditional subgradient method [30] and a unified framework of subgradient methods [39].

- For quasi-convex optimization problems and under the assumption of Hölder condition of order p, conditions (H1) and (H2) are satisfied for several types of subgradient methods, such as the standard subgradient method [26], the inexact subgradient method [21], the primal-dual subgradient method [22] and the conditional subgradient method [23].

3.1   Convergence theorem

As by-products, this subsection aims to establish the convergence theorem of the sequence satisfying conditions (H1) and (H2) for some suitable stepsize rules $\{v_k\}$. The stepsize rule has a critical impact on the convergence behavior and numerical performance of subgradient methods. In the present paper, we consider three typical stepsize rules: (i) the constant stepsize rule is the most popular in applications but only guarantees the convergence to the optimal value/solution set within some tolerance; (ii) the diminishing stepsize rule guarantees the convergence to the exact optimal value/solution set but suffers a slow convergence rate; (iii) the dynamic stepsize rule enjoys the best convergence property but requires prior information of the approximate optimal value $f^* + \epsilon$; see [35, 52, 55] and references therein. Theorem 3.1 extends [55, Theorem 3.1] (considering the special case when $\epsilon = 0$) to the inexact setting, while the skeleton of the proof is similar to that of [55, Theorem 3.1]. To make this paper more self-contained, we provide a proof of the convergence theorem as follows. We write $X_\epsilon^* := X \cap \mathrm{lev}_{\leq f^*+\epsilon} f$ for the sake of simplicity, and particularly, $X_0^* = X^*$.

Theorem 3.1. *Let $\{x_k\} \subseteq X$ satisfy* (H1) *and* (H2). *Then, the following assertions are true.*

(i) *If $v_k \equiv v > 0$, then $\liminf_{k \to \infty} f(x_k) \leq f^* + \left(\frac{\beta v}{\alpha} \right)^p + \epsilon$.*

(ii) *If $\{v_k\}$ is given by*
$$v_k := ck^{-s}, \quad \text{where } c > 0,\ s \in (0,1), \tag{3.3}$$
then $\liminf_{k \to \infty} f(x_k) \leq f^* + \epsilon$.

(iii) *If $\{v_k\}$ is given by*
$$v_k := \frac{\alpha_k \lambda_k}{2\beta_k} [f(x_k) - f^* - \epsilon]_+^{\frac{1}{p}}, \quad \text{where } 0 < \underline{\lambda} \leq \lambda_k \leq \overline{\lambda} < 2, \tag{3.4}$$
then either $x_k \in X_\epsilon^*$ for some $k \in \mathbb{N}$ or $\lim_{k \to \infty} f(x_k) \leq f^* + \epsilon$.

8

Proof. Without loss of generality, we assume that $f(x_k) \le f^* + \epsilon$ only occurs for finitely many times; otherwise, assertions (i) and (ii) of this theorem hold automatically. That is, there exists $N \in \mathbb{N}$ such that $f(x_k) > f^* + \epsilon$ for each $k \ge N$; consequently, letting $x^* \in X^*$, (H1) indicates that

$$\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 \le -\alpha_k v_k (f(x_k) - f^* - \epsilon)^{\frac{1}{p}} + \beta_k v_k^2$$

for each $k \ge N$. Summing the above inequality over $k = N, \ldots, n$, we have

$$\|x_{n+1} - x^*\|^2 - \|x_N - x^*\|^2 \le -\sum_{k=N}^{n} \alpha_k v_k (f(x_k) - f^* - \epsilon)^{\frac{1}{p}} + \sum_{k=N}^{n} \beta_k v_k^2,$$

that is,

$$\frac{\sum_{k=N}^{n} \alpha_k v_k (f(x_k) - f^* - \epsilon)^{\frac{1}{p}}}{\sum_{k=N}^{n} \alpha_k v_k} \le \frac{\|x_N - x^*\|^2}{\sum_{k=N}^{n} \alpha_k v_k} + \frac{\sum_{k=N}^{n} \beta_k v_k^2}{\sum_{k=N}^{n} \alpha_k v_k}. \tag{3.5}$$

(i) By the constant stepsize rule and (3.2), one has $\lim_{n\to\infty} \sum_{k=N}^{n} \alpha_k v_k = \infty$. Then, by (3.5), Lemma 2.1 is applicable (with $(f(x_k) - f^* - \epsilon)^{\frac{1}{p}}$ and $\alpha_k v_k$ in place of a_k and w_k) to concluding that

$$\begin{aligned}
\liminf_{k\to\infty} (f(x_k) - f^* - \epsilon)^{\frac{1}{p}} &\le \liminf_{n\to\infty} \frac{\sum_{k=N}^{n} \alpha_k v_k (f(x_k) - f^* - \epsilon)^{\frac{1}{p}}}{\sum_{k=N}^{n} \alpha_k v_k} \\
&\le \liminf_{n\to\infty} \left(\frac{\|x_N - x^*\|^2}{\sum_{k=N}^{n} \alpha_k v_k} + \frac{\sum_{k=N}^{n} \beta_k v_k^2}{\sum_{k=N}^{n} \alpha_k v_k} \right). \tag{3.6}
\end{aligned}$$

Note by (3.2) and Lemma 2.1 that

$$\lim_{n\to\infty} \frac{\|x_N - x^*\|^2}{\sum_{k=N}^{n} \alpha_k v_k} = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{\sum_{k=N}^{n} \beta_k v_k^2}{\sum_{k=N}^{n} \alpha_k v_k} = \frac{\beta v}{\alpha}.$$

This, together with (3.6), shows that $\liminf_{k\to\infty} (f(x_k) - f^* - \epsilon)^{\frac{1}{p}} \le \frac{\beta v}{\alpha}$, and hence assertion (i) of this theorem is proved.

(ii) By (3.2) and (3.3), one has $\lim_{n\to\infty} \sum_{k=N}^{n} \alpha_k v_k = \infty$; consequently, (3.6) holds. Note by (3.2), (3.3) and Lemma 2.1 that

$$\lim_{n\to\infty} \frac{\|x_N - x^*\|^2}{\sum_{k=N}^{n} \alpha_k v_k} = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{\sum_{k=N}^{n} \beta_k v_k^2}{\sum_{k=N}^{n} \alpha_k v_k} = 0.$$

This, together with (3.6), yields that $\liminf_{k\to\infty} f(x_k) \le f^* + \epsilon$, as desired.

(iii) Without loss of generality, we assume that $f(x_k) > f^* + \epsilon$ for each $k \in \mathbb{N}$; otherwise, assertion (iii) of this theorem follows. By (3.2), there exists $N \in \mathbb{N}$ such that

$$\beta_k < 2\beta \quad \text{and} \quad \alpha_k > \frac{\alpha}{2} \quad \text{for each } k \ge N.$$

Then, for each $x^* \in X^*$, it follows from (3.1) and (3.4) that, for each $k \geq N$,

$$
\begin{aligned}
\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 &\leq -\frac{\alpha_k^2}{4\beta_k} \lambda_k (2 - \lambda_k)(f(x_k) - f^* - \epsilon)^{\frac{2}{p}} \\
&\leq -\frac{\alpha^2}{32\beta} \underline{\lambda}(2 - \overline{\lambda})(f(x_k) - f^* - \epsilon)^{\frac{2}{p}}.
\end{aligned}
\tag{3.7}
$$

This shows that $\lim_{k \to \infty} f(x_k) \leq f^* + \epsilon$; otherwise, it follows from (3.7) that there exists $\sigma > 0$ such that $\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \sigma$ for infinitely many $k \geq N$, which is impossible (as $\{\|x_k - x^*\|\}$ is nonnegative). $\qquad \square$

Remark 3.1. *It is worth noting that the conclusion of Theorem 3.1(ii) is also true for the general diminishing stepsize rule, i.e., satisfying*

$$
v_k > 0, \quad \lim_{k \to \infty} v_k = 0, \quad \sum_{k=0}^{\infty} v_k = \infty.
\tag{3.8}
$$

3.2 Iteration complexity

This subsection is devoted to the complexity issue of the sequence satisfying conditions (H1) and (H2) when using the typical stepsize rules. Given $\delta > 0$, the iteration complexity of a particular algorithm is to estimate the number of iterations K required by the algorithm to obtain an approximate solution, at which the function value is within δ of the optimal, i.e.,

$$
\min_{1 \leq k \leq K} f(x_k) \leq f^* + \delta.
$$

We write

$$
\alpha_{\inf} := \inf_{k \in \mathbb{N}} \alpha_k \quad \text{and} \quad \beta_{\sup} := \sup_{k \in \mathbb{N}} \beta_k
$$

for simplicity. It is clear that $\alpha_{\inf} \in (0, +\infty)$ and $\beta_{\sup} \in (0, +\infty)$ under the assumption (H2).

Theorem 3.2. *Let $\delta > 0$, and let $\{x_k\} \subseteq X$ satisfy (H1) and (H2).*

(i) *Let $K_1 := \frac{\mathrm{dist}^2(x_1, X^*)}{\alpha_{\inf} v \delta}$ and $v_k \equiv v > 0$. Then*

$$
\min_{1 \leq k \leq K_1} f(x_k) \leq f^* + \left(\frac{\beta_{\sup}}{\alpha_{\inf}} v + \delta \right)^p + \epsilon.
$$

(ii) *Let $K_2 := \left(\frac{(1-s)\mathrm{dist}^2(x_1, X^*)}{\alpha_{\inf} c \delta} \right)^{\frac{1}{1-s}}$ and $\{v_k\}$ be given by (3.3). Then*

$$
\min_{1 \leq k \leq K_2} f(x_k) \leq f^* + \left(\frac{\beta_{\sup}}{\alpha_{\inf}} c k^{-s} + \delta \right)^p + \epsilon.
$$

(iii) *Let $K_3 := \frac{4\beta_{\sup}\mathrm{dist}^2(x_1, X^*)}{\alpha_{\inf}^2 \underline{\lambda}(2 - \overline{\lambda})\delta^2}$ and $\{v_k\}$ be given by (3.4). Then*

$$
\min_{1 \leq k \leq K_3} f(x_k) \leq f^* + \delta^p + \epsilon.
$$

Proof. (i) We prove by contradiction, assuming for each $1 \leq k \leq K_1$ that

$$f(x_k) > f^* + \left(\frac{\beta_{\sup}}{\alpha_{\inf}} v + \delta\right)^p + \epsilon;$$

hence, by (H1), we obtain by (3.1) (with $\mathbb{P}_{X^*}(x^k)$ in place of x^*) that

$$\text{dist}^2(x_{k+1}, X^*) < \text{dist}^2(x_k, X^*) - \alpha_{\inf} v \left(\frac{\beta_{\sup}}{\alpha_{\inf}} v + \delta\right) + \beta_{\sup} v^2 = \text{dist}^2(x_k, X^*) - \alpha_{\inf} v \delta.$$

Summing the above inequality over $k = 1, \ldots, K_1$, we obtain that

$$\text{dist}^2(x_{K_1+1}, X^*) < \text{dist}^2(x_1, X^*) - K_1 \alpha_{\inf} v \delta,$$

which yields a contradiction with the definition of K_1. Assertion (i) of this theorem is proved.

(ii) Proving by contradiction, we assume that

$$f(x_k) > f^* + \left(\frac{\beta_{\sup}}{\alpha_{\inf}} ck^{-s} + \delta\right)^p + \epsilon \quad \text{for each } 1 \leq k \leq K_2.$$

Then, we obtain by (3.1) and (3.3) that

$$\text{dist}^2(x_{k+1}, X^*) < \text{dist}^2(x_k, X^*) - \alpha_{\inf} v_k \left(\frac{\beta_{\sup}}{\alpha_{\inf}} v_k + \delta\right) + \beta_{\sup} v_k^2 = \text{dist}^2(x_k, X^*) - \alpha_{\inf} c\delta k^{-s},$$

and thus,

$$\begin{aligned}
\text{dist}^2(x_{K_2+1}, X^*) \quad &< \text{dist}^2(x_1, X^*) - \alpha_{\inf} c\delta \sum_{k=1}^{K_2} k^{-s} \\
&\leq \text{dist}^2(x_1, X^*) - \alpha_{\inf} c\delta \int_1^{K_2+1} t^{-s} \mathrm{d}t \\
&= \text{dist}^2(x_1, X^*) - \alpha_{\inf} \frac{c\delta}{1-s}((K_2+1)^{1-s} - 1),
\end{aligned}$$

which is negative by the definition of K_2. This contradiction yields assertion (ii) of this theorem.

(iii) Proving by contradiction, we assume that

$$f(x_k) > f^* + \delta^p + \epsilon \quad \text{for each } 1 \leq k \leq K_3.$$

Then, it follows from (3.1) and (3.4) that

$$\begin{aligned}
\text{dist}^2(x_{k+1}, X^*) \quad &\leq \quad \text{dist}^2(x_k, X^*) - \frac{\alpha_k^2}{4\beta_k} \lambda_k (2 - \lambda_k)(f(x_k) - f^* - \epsilon)^{\frac{2}{p}} \\
&< \quad \text{dist}^2(x_k, X^*) - \frac{\alpha_{\inf}^2}{4\beta_{\sup}} \underline{\lambda}(2 - \overline{\lambda})\delta^2.
\end{aligned}$$

Summing the above inequality over $k = 1, \ldots, K_3$, we derive that

$$\text{dist}^2(x_{K_3+1}, X^*) < \text{dist}^2(x_1, X^*) - K_3 \frac{\alpha_{\inf}^2}{4\beta_{\sup}} \underline{\lambda}(2 - \overline{\lambda})\delta^2,$$

which yields a contradiction with the definition of K_3. The proof is complete. $\qquad\square$

Remark 3.2. *Theorem 3.2 shows that the sequence satisfying conditions (H1) and (H2) possesses the computational complexity of $\mathcal{O}(1/k^p)$, $\mathcal{O}(1/k^{p\min\{s,1-s\}})$ and $\mathcal{O}(1/k^{\frac{p}{2}})$ to fall within a certain region (expressed by an additive form of the stepsize and noise) of the optimal value when the constant, diminishing or dynamic stepsize rules are used, respectively. In particular, in the cases of the diminishing stepsize rule, the optimal complexity is gained when $s = \frac{1}{2}$, and thus, $v_k = ck^{-\frac{1}{2}}$ is the best choice among the type of (3.3).*

3.3 Convergence rate analysis

The establishment of convergence rates is significant in guaranteeing the numerical performance of relevant algorithms. The aim of this section is to establish the convergence rates for the sequence satisfying conditions (H1) and (H2) under the assumptions of weak sharp minima and/or some suitable assumptions on the noise.

The concept of weak sharp minima was introduced by Burke and Ferris [10], and has been extensively studied and widely used to analyze the convergence rates of many optimization algorithms; see [7, 20, 52, 56] and references therein. One natural extension of this concept is the weak sharp minima of Hölderian order; see [5, 21, 51, 56] and references therein.

Definition 3.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$ and $X^* := \arg\min\{f(x) : x \in X\}$. Let $x^* \in X^*$, $S \subseteq \mathbb{R}^n$, $\eta > 0$ and $q \geq 1$. X^* is said to be*

(a) *the set of weak sharp minima of order q for f on S over X with modulus η if*

$$f(y) - f(x^*) \geq \eta \operatorname{dist}^q(y, X^*) \quad \text{for each } y \in S \cap X;$$

(b) *the set of (global) weak sharp minima of order q for f over X with modulus η if X^* is the set of weak sharp minima of order q for f on $\mathbb{R}^n$ over X with modulus η;*

(c) *the set of boundedly weak sharp minima of order q for f over X if, for each $r > 0$ such that $X^* \cap \mathbf{B}(0, r) \neq \emptyset$, there exists $\eta_r > 0$ such that X^* is the set of weak sharp minima of order q for f on $\mathbf{B}(0, r)$ over X with modulus η_r.*

Remark 3.3. *It is clear that the global weak sharp minima of order q implies the boundedly weak sharp minima of order q. Moreover, the larger the q, the less restrictive the global (resp. boundedly) weak sharp minima of order q. In particular, when $q = 1$, this concept is reduced to the global (resp. boundedly) weak sharp minima; see, e.g., [9, 10].*

The following theorems present the linear (or sublinear) convergence rates of the sequence $\{x_k\}$ satisfying (H1) and (H2) to a certain neighborhood of the solution set when using different stepsize rules and under the assumption of boundedly weak sharp minima of Hölderian order. To this end, we further require the following condition to ensure the bounded property of $\{x_k\}$ when the constant or diminishing stepsize rule is adopted.

(H3) For each $k \in \mathbb{N}$,

$$\|x_{k+1} - x_k\| \leq \gamma_k v_k, \tag{3.9}$$

where $\{\gamma_k\}$ is a sequence of positive scalars such that $\lim_{k \to \infty} \gamma_k = \gamma > 0$.

Condition (H3) characterizes an upper bound (related to the stepsize) of the distance between the successive two iterates, which is always satisfied for various subgradient methods for either convex or quasi-convex optimization problems; see, e.g., Section 4 and [55].

Theorem 3.3. *Let $\{x_k\} \subseteq X$ satisfy (H1)-(H3) and $v_k \equiv v > 0$. Suppose that f is coercive and that X^* is the set of boundedly weak sharp minima of order q for f over X with modulus η. Then, the following assertions are true.*

(i) *If $q = 2p$, then either $x_k \in X_\epsilon^*$ for some $k \in \mathbb{N}$ or there exist $\tau \in [0,1)$ and $N \in \mathbb{N}$ such that*

$$\mathrm{dist}^2(x_{k+N}, X^*) \leq \tau^k \mathrm{dist}^2(x_N, X^*) + 2^{\frac{1}{p}-1} \eta^{-\frac{1}{p}} \left(\epsilon^{\frac{1}{p}} + \frac{\beta v}{\alpha} \right) \quad \text{for each } k \in \mathbb{N}.$$

(ii) *If $q > 2p$ and $\epsilon^{\frac{1}{p}} + \frac{\beta v}{\alpha} < \eta^{-\frac{2}{q-2p}} (\frac{2p}{\alpha v q})^{\frac{q}{q-2p}}$, then either $x_k \in X_\epsilon^*$ for some $k \in \mathbb{N}$ or there exist $\tau \in (0,1)$ and $N \in \mathbb{N}$ such that*

$$\mathrm{dist}^2(x_{k+N}, X^*) \leq \tau^k \mathrm{dist}^2(x_N, X^*) + 2^{\frac{2}{q}-\frac{2p}{q}} \eta^{-\frac{2}{q}} \left(\epsilon^{\frac{1}{p}} + \frac{\beta v}{\alpha} \right)^{\frac{2p}{q}} \quad \text{for each } k \in \mathbb{N}.$$

Proof. Without loss of generality, we assume that $f(x_k) > f^* + \epsilon$ for each $k \in \mathbb{N}$; otherwise, this theorem holds automatically. Consequently, (H1) says that, for each $x^* \in X^*$ and $k \in \mathbb{N}$,

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \alpha_k v(f(x_k) - f^* - \epsilon)^{\frac{1}{p}} + \beta_k v^2. \tag{3.10}$$

We first claim that $\{x_k\}$ is bounded. To this end, we fix $\kappa > 1$ and define

$$\sigma := \kappa \left(\frac{\beta v}{\alpha} \right)^p + \epsilon, \quad X_\sigma^* := X \cap \mathrm{lev}_{\leq f^*+\sigma} f \quad \text{and} \quad \rho(\sigma) := \max_{x \in X_\sigma^*} \mathrm{dist}(x, X^*). \tag{3.11}$$

By the coercive assumption, it follows that its sublevel set $\mathrm{lev}_{\leq f^*+\sigma} f$ is bounded, and so is X_σ^*. Then, one has by (3.11) that $\rho(\sigma) < \infty$. By (H2) and (H3), there exists $N \in \mathbb{N}$ such that

$$\alpha \kappa^{-\frac{1}{2p}} < \alpha_k < \alpha \kappa^{\frac{1}{2p}}, \quad \beta \kappa^{-\frac{1}{2p}} < \beta_k < \beta \kappa^{\frac{1}{2p}} \quad \text{and} \quad \gamma_k < 2\gamma \quad \text{for each } k \geq N. \tag{3.12}$$

Fix $k \geq N$. Below, we show

$$\mathrm{dist}(x_{k+1}, X^*) < \max\{\mathrm{dist}(x_k, X^*), \rho(\sigma) + 2\gamma v\} \tag{3.13}$$

by claiming the following two implications:

$$[f(x_k) > f^* + \sigma] \quad \Rightarrow \quad [\mathrm{dist}(x_{k+1}, X^*) < \mathrm{dist}(x_k, X^*)]; \tag{3.14}$$

$$[f(x_k) \leq f^* + \sigma] \quad \Rightarrow \quad [\mathrm{dist}(x_{k+1}, X^*) < \rho(\sigma) + 2\gamma v]. \tag{3.15}$$

13

To prove (3.14), we suppose that $f(x_k) > f^* + \sigma$. Then, for each $x^* \in X^*$, one has by (3.10) that

$$\|x_{k+1} - x^*\|^2 < \|x_k - x^*\|^2 - \alpha_k v(\sigma - \epsilon)^{\frac{1}{p}} + \beta_k v^2 < \|x_k - x^*\|^2$$

(due to (3.11) and (3.12)). Consequently, (3.14) can be proved by selecting $x^* = \mathbb{P}_{X^*}(x_k)$. To show (3.15), we assume that $f(x_k) \leq f^* + \sigma$. Then $x_k \in X_\sigma^*$, and so (3.11) says that $\text{dist}(x_k, X^*) \leq \rho(\sigma)$. This, together with (3.9) and (3.12), shows (3.15). Therefore, (3.13) is proved, as desired.

Note by Theorem 3.1(i) that $\liminf_{k \to \infty} f(x_k) \leq f^* + (\frac{\beta v}{\alpha})^p + \epsilon$, and note by $\kappa > 1$ and (3.11) that $\sigma > (\frac{\beta v}{\alpha})^p + \epsilon$. Then, we can assume, without loss of generality, that $f(x_N) \leq f^* + \sigma$ (otherwise, we can choose a larger N). Consequently, we have by (3.15) that $\text{dist}(x_{N+1}, X^*) < \rho(\sigma) + 2\gamma v$, and inductively obtain by (3.13) that

$$\text{dist}(x_k, X^*) < \rho(\sigma) + 2\gamma v \quad \text{for each } k > N. \tag{3.16}$$

Hence, $\{x_k\}$ is proved to be bounded (since $X^* \subseteq X_\sigma^*$ is bounded), as desired. That is, there exists $r > 0$ such that $X^* \subseteq \mathbf{B}(0, r)$ and $\{x_k\} \subseteq \mathbf{B}(0, r)$ for each $k \in \mathbb{N}$. Then, by assumption of boundedly weak sharp minima property of order q, there exists $\eta > 0$ such that

$$f(x_k) - f^* \geq \eta \, \text{dist}^q(x_k, X^*) \quad \text{for each } k \in \mathbb{N}. \tag{3.17}$$

Selecting $x^* = \mathbb{P}_{X^*}(x_k)$, we deduce by (3.10) and (3.12) that, for each $k \geq N$,

$$\begin{aligned}
&\text{dist}^2(x_{k+1}, X^*) \\
&\leq \text{dist}^2(x_k, X^*) - \alpha v \kappa^{-\frac{1}{2p}} (f(x_k) - f^* - \epsilon)^{\frac{1}{p}} + \beta \kappa^{\frac{1}{2p}} v^2 \\
&\leq \text{dist}^2(x_k, X^*) - 2^{1-\frac{1}{p}} \alpha v \kappa^{-\frac{1}{2p}} (f(x_k) - f^*)^{\frac{1}{p}} + \alpha v \kappa^{-\frac{1}{2p}} \epsilon^{\frac{1}{p}} + \beta \kappa^{\frac{1}{2p}} v^2,
\end{aligned} \tag{3.18}$$

where the last inequality holds because

$$(a - b)^\gamma \geq 2^{1-\gamma} a^\gamma - b^\gamma \quad \text{whenever } a \geq b \geq 0 \text{ and } \gamma \geq 1 \tag{3.19}$$

(cf. [24, Lemma 4.1]). Below, we prove this theorem in the following two cases.

(i) Suppose that $q = 2p$. Setting $\tau := (1 - 2^{1-\frac{1}{p}} \alpha v \kappa^{-\frac{1}{2p}} \eta^{\frac{1}{p}})_+ \in [0, 1)$ and substituting (3.17) into (3.18), we achieve that

$$\text{dist}^2(x_{k+1}, X^*) \leq \tau \text{dist}^2(x_k, X^*) + \alpha v \kappa^{-\frac{1}{2p}} \epsilon^{\frac{1}{p}} + \beta \kappa^{\frac{1}{2p}} v^2 \quad \text{for each } k \geq N.$$

Then, we inductively obtain that

$$\text{dist}^2(x_{k+N}, X^*) \leq \tau^k \text{dist}^2(x_N, X^*) + \frac{1}{1 - \tau}(\alpha v \kappa^{-\frac{1}{2p}} \epsilon^{\frac{1}{p}} + \beta \kappa^{\frac{1}{2p}} v^2) \quad \text{for each } k \in \mathbb{N};$$

consequently, the conclusion follows (noting that $\kappa > 1$ is arbitrary).

(ii) Suppose that $q > 2p$ and $\alpha \epsilon^{\frac{1}{p}} + \beta v < (\alpha^p \eta)^{-\frac{2}{q-2p}} (v \frac{q}{2p})^{-\frac{q}{q-2p}}$. We obtain by (3.17) and (3.18) that, for each $k \geq N$,

$$\text{dist}^2(x_{k+1}, X^*) \leq \text{dist}^2(x_k, X^*) - 2^{1-\frac{1}{p}} \alpha v \kappa^{-\frac{1}{2p}} \eta^{\frac{1}{p}} \text{dist}^{\frac{q}{p}}(x_k, X^*) + \alpha v \kappa^{-\frac{1}{2p}} \epsilon^{\frac{1}{p}} + \beta \kappa^{\frac{1}{2p}} v^2.$$

14

Then, Lemma 2.2(ii) is applicable (with $\text{dist}^2(x_k, X^*)$, $\frac{q}{2p} - 1$, $2^{1-\frac{1}{p}}\alpha v \kappa^{-\frac{1}{2p}}\eta^{\frac{1}{p}}$, $\alpha v \kappa^{-\frac{1}{2p}}\epsilon^{\frac{1}{p}} + \beta\kappa^{\frac{1}{2p}}v^2$ in place of u_k, r, a, b) to obtaining the conclusion. $\qquad\square$

Theorem 3.4. *Let $\{x_k\} \subseteq X$ satisfy (H1) and (H2), and $\{v_k\}$ be given by (3.4). Suppose that X^* is the set of boundedly weak sharp minima of order q for f over X with modulus η. Then, the following assertions are true.*

(i) *If $q = p$ and $\epsilon \geq 0$, then either $x_k \in X_\epsilon^*$ for some $k \in \mathbb{N}$ or there exist $\tau \in [0, 1)$ and $N \in \mathbb{N}$ such that*

$$\text{dist}^2(x_{k+N}, X^*) \leq \tau^k \text{dist}^2(x_N, X^*) + 2^{\frac{2}{p}-1}\eta^{-\frac{2}{p}}\epsilon^{\frac{2}{p}} \quad \text{for each } k \in \mathbb{N}. \tag{3.20}$$

(ii) *If $q > p$ and $\epsilon = 0$, then either $x_k \in X^*$ for some $k \in \mathbb{N}$ or there exist $\gamma > 0$ and $N \in \mathbb{N}$ such that*

$$\text{dist}^2(x_{k+N}, X^*) \leq \frac{\text{dist}^2(x_N, X^*)}{(1 + \gamma k)^{\frac{p}{q-p}}} \quad \text{for each } k \in \mathbb{N}. \tag{3.21}$$

(iii) *If $q > p$ and $0 < \epsilon < \left(\frac{\alpha^2 q}{4\beta p}\underline{\lambda}(2 - \overline{\lambda})\right)^{-\frac{pq}{2(q-p)}}\eta^{-\frac{p}{q-p}}$, then either $x_k \in X_\epsilon^*$ for some $k \in \mathbb{N}$ or there exist $\tau \in (0, 1)$ and $N \in \mathbb{N}$ such that*

$$\text{dist}^2(x_{k+N}, X^*) \leq \tau^k \text{dist}^2(x_N, X^*) + 2^{\frac{2}{q}-\frac{p}{q}}\eta^{-\frac{2}{q}}\epsilon^{\frac{2}{q}} \quad \text{for each } k \in \mathbb{N}. \tag{3.22}$$

Proof. Without loss of generality, we assume that $f(x_k) > f^* + \epsilon$ for each $k \in \mathbb{N}$; otherwise, this theorem holds automatically. Combining (3.1) with (3.4), we obtain that there exists $N \in \mathbb{N}$ such that, for each $k \geq N$,

$$\begin{aligned}
&\text{dist}^2(x_{k+1}, X^*) - \text{dist}^2(x_k, X^*) \\
&\leq -\frac{\alpha_k^2}{4\beta_k}\lambda_k(2 - \lambda_k)(f(x_k) - f^* - \epsilon)^{\frac{2}{p}} \\
&\leq -2^{1-\frac{2}{p}}\kappa^{-\frac{3}{2p}}\frac{\alpha^2}{4\beta}\underline{\lambda}(2 - \overline{\lambda})(f(x_k) - f^*)^{\frac{2}{p}} + \kappa^{-\frac{3}{2p}}\frac{\alpha^2}{4\beta}\underline{\lambda}(2 - \overline{\lambda})\epsilon^{\frac{2}{p}}
\end{aligned} \tag{3.23}$$

(due to (3.12) and (3.19)). It follows from the proof of Theorem 3.1(iii) (cf. (3.7)) that $\{x_k\}$ is bounded. Then, there exists $r > 0$ such that $X^* \cap \mathbf{B}(0, r) \neq \emptyset$ and $\{x_k\} \subseteq \mathbf{B}(0, r)$ for each $k \in \mathbb{N}$. By the assumption of boundedly weak sharp minima property of order q, there exists $\eta > 0$ such that

$$f(x_k) - f^* \geq \eta \, \text{dist}^q(x_k, X^*) \quad \text{for each } k \in \mathbb{N}. \tag{3.24}$$

(i) Suppose that $q = p$. Setting $\tau := (1 - 2^{1-\frac{2}{p}}\kappa^{-\frac{3}{2p}}\frac{\alpha^2}{4\beta}\underline{\lambda}(2 - \overline{\lambda})\eta^{\frac{2}{p}})_+ \in [0, 1)$, we obtain by substituting (3.24) into (3.23) that

$$\text{dist}^2(x_{k+1}, X^*) \leq \tau \text{dist}^2(x_k, X^*) + \kappa^{-\frac{3}{2p}}\frac{\alpha^2}{4\beta}\underline{\lambda}(2 - \overline{\lambda})\epsilon^{\frac{2}{p}} \quad \text{for each } k \geq N.$$

15

Then, we inductively obtain

$$\text{dist}^2(x_{k+N}, X^*) \leq \tau^k \text{dist}^2(x_N, X^*) + \frac{1}{1-\tau} \kappa^{-\frac{3}{2p}} \frac{\alpha^2}{4\beta} \underline{\lambda}(2 - \overline{\lambda}) \epsilon^{\frac{2}{p}} \quad \text{for each } k \in \mathbb{N};$$

hence, (3.20) holds (noting that $\kappa > 1$ is arbitrary), and assertion (i) is proved.

(ii) Suppose that $q > p$ and $\epsilon = 0$. Then, combining (3.23) and (3.24) implies that, for each $k \geq N$,

$$\text{dist}^2(x_{k+1}, X^*) \leq \text{dist}^2(x_k, X^*) - 2^{1-\frac{2}{p}} \kappa^{-\frac{3}{2p}} \frac{\alpha^2}{4\beta} \underline{\lambda}(2 - \overline{\lambda}) \eta^{\frac{2}{p}} \text{dist}^{\frac{2q}{p}}(x_k, X^*).$$

Then, Lemma 2.2(i) is applicable (with $\text{dist}^2(x_k, X^*)$, $2^{1-\frac{2}{p}} \kappa^{-\frac{3}{2p}} \frac{\alpha^2}{4\beta} \underline{\lambda}(2 - \overline{\lambda}) \eta^{\frac{2}{p}}$, $\frac{q}{p} - 1$ in place of u_k, a, r) to concluding that (3.21) holds with $\gamma := 2^{1-\frac{2}{p}} \kappa^{-\frac{3}{2p}} \frac{\alpha^2(q-p)}{4\beta p} \underline{\lambda}(2 - \overline{\lambda}) \eta^{\frac{2}{p}} \text{dist}^{\frac{2q}{p}-2}(x_N, X^*)$.

(iii) Suppose that $q > p$ and $0 < \epsilon < \left(\frac{\alpha^2 q}{4\beta p} \underline{\lambda}(2 - \overline{\lambda}) \right)^{-\frac{pq}{2(q-p)}} \eta^{-\frac{p}{q-p}}$. We obtain by (3.23) and (3.24) that, for each $k \geq N$,

$$\text{dist}^2(x_{k+1}, X^*) \leq \text{dist}^2(x_k, X^*) - 2^{1-\frac{2}{p}} \kappa^{-\frac{3}{2p}} \frac{\alpha^2}{4\beta} \underline{\lambda}(2 - \overline{\lambda}) \eta^{\frac{2}{p}} \text{dist}^{\frac{2q}{p}}(x_k, X^*) + \kappa^{-\frac{3}{2p}} \frac{\alpha^2}{4\beta} \underline{\lambda}(2 - \overline{\lambda}) \epsilon^{\frac{2}{p}}.$$

Then Lemma 2.2(ii) is applicable (with $\text{dist}^2(x_k, X^*)$, $\frac{q}{p} - 1$, $2^{1-\frac{2}{p}} \kappa^{-\frac{3}{2p}} \frac{\alpha^2}{4\beta} \underline{\lambda}(2 - \overline{\lambda}) \eta^{\frac{2}{p}}$, $\kappa^{-\frac{3}{2p}} \frac{\alpha^2}{4\beta} \underline{\lambda}(2 - \overline{\lambda}) \epsilon^{\frac{2}{p}}$ in place of u_k, r, a, b) to concluding that there exists $\tau \in (0,1)$ such that (3.22) is satisfied. $\square$

Theorem 3.5. *Let $\{x_k\} \subseteq X$ satisfy (H1)-(H3), and $\{v_k\}$ be given by (3.3). Suppose that f is coercive and that X^* is the set of boundedly weak sharp minima of order $2p$ for f over X with modulus η. Then, the following assertions are true.*

(i) *If $\epsilon = 0$, then either $x_k \in X^*$ for some $k \in \mathbb{N}$ or there exists $N \in \mathbb{N}$ such that*

$$\text{dist}^2(x_k, X^*) \leq \frac{\beta c}{\alpha} \left(\frac{2}{\eta} \right)^{\frac{1}{p}} k^{-s} \quad \text{for each } k \geq N.$$

(ii) *If $\epsilon > 0$, then either $x_k \in X_\epsilon^*$ for some $k \in \mathbb{N}$ or there exist $C > 0$ and $\tau \in (0,1)$ such that*

$$\text{dist}^2(x_k, X^*) \leq C\tau^k + \left(\frac{2\epsilon}{\eta} \right)^{\frac{1}{p}} \quad \text{for each } k \in \mathbb{N}.$$

Proof. Without loss of generality, we assume that $f(x_k) > f^* + \epsilon$ for each $k \in \mathbb{N}$; otherwise, this theorem holds automatically. Then, (H1) and (3.3) show that, for each $x^* \in X_\epsilon^*$ and $k \in \mathbb{N}$,

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \alpha_k c k^{-s} (f(x_k) - f^* - \epsilon)^{\frac{1}{p}} + \beta_k c^2 k^{-2s}. \quad (3.25)$$

Fix $\kappa > 1$ and $\sigma > \epsilon$, and define X_σ^* and $\rho(\sigma)$ by (3.11). By (H2) and (H3), there exists $N \geq (\frac{c\beta}{\alpha})^{\frac{1}{s}} (\frac{\kappa}{\sigma - \epsilon})^{\frac{1}{sp}}$ such that (3.12) is satisfied. Similar to the arguments that we did for

16

(3.16), we can derive that $\text{dist}(x_k, X^*) < \rho(\sigma) + 2\gamma c$ for each $k > N$. Hence, $\{x_k\}$ is bounded, and then, by assumption of boundedly weak sharp minima property of order $2p$, there exists $\eta > 0$ such that (3.17) is satisfied with $q = 2p$. This, together with (3.25) and (3.19), implies that, for each $k \geq N$,

$$\text{dist}^2(x_{k+1}, X^*) \leq (1 - 2^{1-\frac{1}{p}}\alpha\kappa^{-\frac{1}{2p}}\eta^{\frac{1}{p}}ck^{-s})\text{dist}^2(x_k, X^*) + \alpha\kappa^{-\frac{1}{2p}}\epsilon^{\frac{1}{p}}ck^{-s} + \beta\kappa^{\frac{1}{2p}}c^2k^{-2s}. \quad (3.26)$$

(i) Suppose that $\epsilon = 0$. Lemma 2.3(i) is applicable (with $2^{1-\frac{1}{p}}\alpha\kappa^{-\frac{1}{2p}}\eta^{\frac{1}{p}}c$, $\beta\kappa^{\frac{1}{2p}}c^2$, $2s$ in place of a, b, t) to obtaining the conclusion (as $\kappa > 1$ is arbitrary).

(ii) Suppose that $\epsilon > 0$. Letting $N \geq (\frac{c\beta}{\alpha})^{\frac{1}{s}}(\frac{\kappa}{\epsilon})^{\frac{1}{sp}}$, (3.26) is reduced to

$$\text{dist}^2(x_{k+1}, X^*) \leq (1 - 2^{1-\frac{1}{p}}\alpha\kappa^{-\frac{1}{2p}}\eta^{\frac{1}{p}}ck^{-s})\text{dist}^2(x_k, X^*) + 2\alpha\kappa^{-\frac{1}{2p}}\epsilon^{\frac{1}{p}}ck^{-s},$$

and then, Lemma 2.3(ii) is applicable (with $2^{1-\frac{1}{p}}\alpha\kappa^{-\frac{1}{2p}}\eta^{\frac{1}{p}}c$ and $2\alpha\kappa^{-\frac{1}{2p}}\epsilon^{\frac{1}{p}}c$ in place of a and b) to obtaining the conclusion. $\qquad\square$

Remark 3.4. *Theorems on convergence rates improve the results in [55], in which only the global convergence theorems were provided.*

(i) Theorems 3.3 and 3.4 show the linear convergence rates of the sequence satisfying (H1)-(H3) to a certain region (i.e., $\mathcal{O}((\epsilon^{\frac{1}{p}} + v)^{\frac{2p}{q}})$ or $\mathcal{O}(\epsilon^{\frac{2}{q}})$) of the solution set under the boundedly weak sharp minima of order q when using the constant or dynamic stepsize rules, respectively.

(ii) In the special case when $\epsilon = 0$ and using the dynamic stepsize rule, Theorem 3.4 presents the linear (or sublinear) convergence of the sequence satisfying (H1)-(H2) to the solution set under the boundedly weak sharp minima of order q.

(iii) When using the diminishing stepsize rule (3.3) and under the boundedly weak sharp minima of order $2p$, Theorem 3.5 shows the sublinear convergence rate to the solution set or the linear convergence rate to a certain region (i.e., $\mathcal{O}(\epsilon^{\frac{1}{p}})$) of the solution set for the exact or inexact framework, respectively.

4   Applications to subgradient methods

Quasi-convex optimization plays an important role in various fields such as economics, finance and engineering. The subgradient method is a popular algorithm for solving constrained quasi-convex optimization problems. It was shown in [55] that several types of subgradient methods for solving quasi-convex optimization problem satisfy conditions (H1)-(H3) assumed in the unified framework (with $\epsilon = 0$) under the Hölder condition. Hence, in this section, we directly apply the convergence theorems obtained in the preceding section to establish the convergence properties of several subgradient methods for solving quasi-convex optimization problems. The convergence theorems (resp. Theorems 4.1, 4.5 and 4.9) cover the existing results in the literature of subgradient methods (resp. [55, Theorem 4.2], [21, Theorem 3.1] and [23, Theorems 3.3 and 3.5]). To the best of our knowledge, the theorems of complexity

estimation (i.e., Theorems 4.2, 4.6 and 4.10) and convergence rates (i.e., Theorems 4.3, 4.7 and 4.11) of subgradient methods for quasi-convex optimization are new in the literature. Throughout the whole section, we make the following blanket 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 $\mathbb{R}^n$.

4.1 Subgradient method

It was claimed in [55] the standard subgradient method (i.e., Algorithm 1.1) satisfies the following basic inequality under the blanket assumption:

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

whenever $f(x_k) > f^*$, and $\|x_{k+1} - x_k\| \leq v_k$. That is, conditions (H1)-(H3) are satisfied with

$$\epsilon = 0, \quad \alpha_k \equiv 2L^{-\frac{1}{p}}, \quad \beta_k \equiv 1, \quad \gamma_k \equiv 1.$$

Therefore, the convergence theorems established in the preceding section can be directly applied to establish the convergence properties of the standard subgradient method as follows.

Theorem 4.1. *Let $\{x_k\}$ be a sequence generated by Algorithm 1.1.*

(i) *If $v_k \equiv v > 0$, then $\liminf_{k\to\infty} f(x_k) \leq f^* + L \left(\frac{1}{2}v \right)^p$.*

(ii) *If $\{v_k\}$ satisfies (3.8), then $\liminf_{k\to\infty} f(x_k) \leq f^*$.*

(iii) *If $\{v_k\}$ is given by*

$$v_k := \lambda_k \left(\frac{f(x_k) - f^*}{L} \right)^{\frac{1}{p}}, \quad \text{where } 0 < \underline{\lambda} \leq \lambda_k \leq \overline{\lambda} < 2, \tag{4.1}$$

then $\lim_{k\to\infty} f(x_k) = f^*$.

Remark 4.1. *For the dynamic stepsize rule (4.1) ((4.8) or (4.9) in the sequel), once x_k enters X^* (or X^*_ϵ), the stepsize will be zero, the iterates will stay at the solution (or the approximate solution), and thus, the conclusions of global convergence and convergence rate follow automatically.*

Theorem 4.2. *Let $\delta > 0$, and let $\{x_k\}$ be a sequence generated by Algorithm 1.1.*

(i) *Let $K_1 := \frac{L^{\frac{1}{p}} \operatorname{dist}^2(x_1, X^*)}{2v\delta}$ and $v_k \equiv v > 0$. Then $\min_{1 \leq k \leq K_1} f(x_k) \leq f^* + \left(\frac{1}{2} L^{\frac{1}{p}} v + \delta \right)^p$.*

(ii) *Let $K_2 := \left(\frac{(1-s)L^{\frac{1}{p}} \operatorname{dist}^2(x_1, X^*)}{2c\delta} \right)^{\frac{1}{1-s}}$ and $\{v_k\}$ be given by (3.3). Then $\min_{1 \leq k \leq K_2} f(x_k) \leq f^* + \left(\frac{1}{2} L^{\frac{1}{p}} ck^{-s} + \delta \right)^p$.*

(iii) *Let* $K_3 := \frac{L^{\frac{2}{p}}\mathrm{dist}^2(x_1,X^*)}{\underline{\lambda}(2-\overline{\lambda})\delta^2}$ *and* $\{v_k\}$ *be given by* (4.1). *Then* $\min\limits_{1\le k\le K_3} f(x_k) \le f^* + \delta^p$.

Theorem 4.3. *Let* $\{x_k\}$ *be a sequence generated by Algorithm 1.1. Suppose that* X^* *is the set of boundedly weak sharp minima of order* q *for* f *over* X *with modulus* η.

(I) *Suppose that* $v_k \equiv v > 0$ *and that* f *is coercive.*

(i) *If* $q = 2p$, *then either* $x_k \in X^*$ *for some* $k \in \mathbb{N}$ *or there exist* $\tau \in [0,1)$ *and* $N \in \mathbb{N}$ *such that*

$$\mathrm{dist}^2(x_{k+N}, X^*) \le \tau^k \mathrm{dist}^2(x_N, X^*) + 2^{\frac{1}{p}-2}\left(\frac{L}{\eta}\right)^{\frac{1}{p}} v \quad \text{for each } k \in \mathbb{N}.$$

(ii) *If* $q > 2p$ *and* $v < \left(2^{-p}\frac{L}{\eta}(\frac{2p}{q})^{\frac{q}{2}}\right)^{\frac{1}{q-p}}$, *then either* $x_k \in X^*$ *for some* $k \in \mathbb{N}$ *or there exist* $\tau \in (0,1)$ *and* $N \in \mathbb{N}$ *such that*

$$\mathrm{dist}^2(x_{k+N}, X^*) \le \tau^k \mathrm{dist}^2(x_N, X^*) + 2^{\frac{2}{q}-\frac{4p}{q}}\left(\frac{L}{\eta}\right)^{\frac{2}{q}} v^{\frac{2p}{q}} \quad \text{for each } k \in \mathbb{N}.$$

(II) *Suppose that* $\{v_k\}$ *is given by* (3.3), $q = 2p$ *and that* f *is coercive. Then, either* $x_k \in X^*$ *for some* $k \in \mathbb{N}$ *or there exists* $N \in \mathbb{N}$ *such that*

$$\mathrm{dist}^2(x_k, X^*) \le \frac{c}{2}\left(\frac{2L}{\eta}\right)^{\frac{1}{p}} k^{-s} \quad \text{for each } k \ge N.$$

(III) *Suppose that* $\{v_k\}$ *is given by* (4.1).

(i) *If* $q = p$, *then there exist* $\tau \in [0,1)$ *and* $N \in \mathbb{N}$ *such that*

$$\mathrm{dist}^2(x_{k+N}, X^*) \le \tau^k \mathrm{dist}^2(x_N, X^*) \quad \text{for each } k \in \mathbb{N}.$$

(ii) *If* $q > p$, *then there exist* $\gamma > 0$ *and* $N \in \mathbb{N}$ *such that*

$$\mathrm{dist}^2(x_{k+N}, X^*) \le \frac{\mathrm{dist}^2(x_N, X^*)}{(1+\gamma k)^{\frac{p}{q-p}}} \quad \text{for each } k \in \mathbb{N}.$$

Remark 4.2. *(i) When using the dynamic stepsize rule, the linear convergence rate of the subgradient method for solving convex optimization problems was shown in [8, Theorem 2.5] under the assumption of weak sharp minima. Theorem 4.3 remarkably extends the result in [8] to quasi-convex optimization, using several typical stepsize rules and under the weaker assumption of weak sharp minima of Hölderian order.*

(ii) When using the constant stepsize rule, [25, Theorem 1] proved the linear convergence rate of the subgradient method for solving convex optimization problems under the assumption of weak sharp minima of Hölderian order. Theorem 4.3 extends the result in [8] to quasi-convex optimization and using several typical stepsize rules.

The finite convergence is a strong convergence property of optimization algorithms. For example, the proximal point algorithm has the finite convergence property under the assumption of weak sharp minima; see, e.g., [13, 53]. However, the subgradient method may not converge within the finite iterations to an optimal solution under the assumption of weak sharp minima, even for the convex optimization problem; see the following counter-example.

Example 4.1. *Consider the optimization problem*

$$\min_{x \in \mathbb{R}} f(x) := |x|.$$

Clearly, its solution set is $X^* = \{0\}$, and f is a (quasi-)convex function and satisfies the weak sharp minima and the Lipschitz condition. Note that its convex subdifferential $\partial f(x) = \{1\}$ and $\{-1\}$ (resp. quasi-subdifferential $\partial^* f(x) = \mathbb{R}_+$ and $\mathbb{R}_-$) if $x > 0$ and $x < 0$, respectively. Selecting the initial point $x_1 = \sqrt{2}$ and the constant stepsize $v = 1$ or the diminishing stepsize $v_k = \frac{1}{k}$ or the dynamic stepsize (4.1) (with $\lambda_k \equiv \lambda \neq 1$), one can check that the sequence generated by the convex subgradient method (resp. Algorithm 1.1) cannot achieve the solution within finite iterations.

In the following theorem, we establish the finite convergence property of the subgradient method to the solution set X^* of problem (1.1) under the assumption that X^* has a nonempty interior.

Theorem 4.4. *Let $\{x_k\}$ be a sequence generated by Algorithm 1.1. 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) $v_k = v \in (0, 2\sigma)$ *for each $k \in \mathbb{N}$.*

(ii) $\{v_k\}$ *satisfies (3.8).*

Proof. Proving by contradiction, we assume that $f(x_k) > f_*$ for each $k \in \mathbb{N}$. By the process of Algorithm 1.1, it follows from the nonexpansive property of projection operator that

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - v_k g_k - x^*\|^2 = \|x_k - x^*\|^2 - 2v_k \langle g_k, x_k - x^* \rangle + v_k^2$$

for each $k \in \mathbb{N}$. Summing the above inequality over $k = 0, \ldots, n$, one has

$$\frac{\sum_{k=0}^{n} v_k \langle g_k, x_k - x^* \rangle}{\sum_{k=0}^{n} v_k} \leq \frac{\|x_0 - x^*\|^2}{2 \sum_{k=0}^{n} v_k} + \frac{\sum_{k=0}^{n} v_k^2}{2 \sum_{k=0}^{n} v_k}. \tag{4.2}$$

By the assumption that $\mathbb{B}(x^*, \sigma) \subseteq X^*$ and $\|g_k\| = 1$, one has $x^* + \sigma g_k \in X^* \subseteq X \cap \text{lev}_{< f(x_k)} f$. Then, it follows from Definition 2.1 that $\langle g_k, x^* + \sigma g_k - x_k \rangle \leq 0$, that is, $\langle g_k, x_k - x^* \rangle \geq \sigma$. This, together with (4.2), shows that

$$\frac{\|x_0 - x^*\|^2}{2 \sum_{k=0}^{n} v_k} + \frac{\sum_{k=0}^{n} v_k^2}{2 \sum_{k=0}^{n} v_k} \geq \sigma. \tag{4.3}$$

On the other hand, under either assumption (i) or (ii), one can check by Lemma 2.1 that

$$\liminf_{n \to \infty} \frac{\|x_0 - x^*\|^2}{2 \sum_{k=0}^{n} v_k} + \frac{\sum_{k=0}^{n} v_k^2}{2 \sum_{k=0}^{n} v_k} < \sigma,$$

which arrives at a contradiction with (4.3). The proof is complete. $\qquad \square$

4.2 Inexact subgradient method

In many applications, the computation error stems from practical considerations and is inevitable in the computing process. To meet the requirement of practical applications, an inexact subgradient method was proposed in [21] to solve a constrained quasi-convex optimization problem (1.1), in which an ϵ-quasi-subgradient is employed (with an additional noise), and the global convergence theorem is established there. This section is devoted to establishing the iteration complexity and convergence rates of the inexact subgradient method for solving problem (1.1), which is formally described as follows.

Algorithm 4.1. Let $\epsilon > 0$. Select an initial point $x_1 \in \mathbb{R}^n$ and a sequence of stepsizes $\{v_k\} \subseteq (0, +\infty)$. For each $k \in \mathbb{N}$, having x_k, we select $g_k \in \partial_\epsilon^* f(x_k) \cap \mathbf{S}$ and update x_{k+1} by

$$x_{k+1} := \mathbb{P}_X(x_k - v_k g_k). \tag{4.4}$$

We claim that the inexact subgradient method satisfies conditions (H1)-(H3) under the blanket assumption. To this end, we provide in the following lemma an important property of a quasi-convex function, which is inspired by [29, Proposition 2.1].

Lemma 4.1. *Let $x \in X$ be such that $f(x) > f^* + \epsilon$ and let $g \in \partial_\epsilon^* f(x) \cap \mathbf{S}$. Then, it holds that*

$$\langle g, x - x^* \rangle \geq \left(\frac{f(x) - f^* - \epsilon}{L} \right)^{\frac{1}{p}} \quad \textit{for each } x^* \in X^*.$$

Proof. Fix $x \in X$ be such that $f(x) > f^* + \epsilon$. Then, the level set $\mathrm{lev}_{<f(x)-\epsilon} f$ is nonempty open and convex by the blanket assumption that f is quasi-convex and continuous on $\mathbb{R}^n$. Given $x^* \in X^*$, we define

$$r := \inf \left\{ \|y - x^*\| : y \in \mathrm{bd}\left(\mathrm{lev}_{<f(x)-\epsilon} f \right) \right\}, \tag{4.5}$$

where $\mathrm{bd}(Z)$ denotes the boundary of the set Z. It follows that $r > 0$ by the fact that $f(x) - \epsilon > f(x^*)$ and the Hölder condition. Furthermore, we have by Definition 2.2 that

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

Taking the infimun over $\mathrm{bd}\left(\mathrm{lev}_{<f(x)-\epsilon} f \right)$, we can show that

$$f(x) - f^* - \epsilon \leq L \inf \left\{ \mathrm{dist}^p(y, X^*) : y \in \mathrm{bd}\left(\mathrm{lev}_{<f(x)-\epsilon} f \right) \right\} \leq L r^p. \tag{4.6}$$

Let $\delta \in (0, 1)$. Since $g \in \partial_\epsilon^* f(x) \cap \mathbf{S}$, we obtain by (4.5) that $x^* + \delta r g \in \text{lev}_{<f(x)-\epsilon} f$. Hence, it follows by definition that

$$\langle g, x - x^* \rangle = \langle g, x - (x^* + \delta r g) \rangle + \delta r \geq \delta r.$$

Since $\delta \in (0, 1)$ is arbitrary, one has $\langle g, x - x^* \rangle \geq r$. This, together with (4.6), implies the conclusion. $\qquad\square$

Lemma 4.2. *Let $\{x_k\}$ be a sequence generated by Algorithm 4.1.*

(i) $\|x_{k+1} - x_k\| \leq v_k$.

(ii) *If $f(x_k) > f^* + \epsilon$, then it holds for each $x^* \in X^*$ that*

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

Proof. Assertion (i) of this theorem directly follows from Algorithm 4.1 (cf. (4.4)). To show assertion (ii), we suppose that $f(x_k) > f^* + \epsilon$ and fix $x^* \in X^*$. In view of Algorithm 4.1, it follows from the nonexpansive property of projection operator that

$$
\begin{aligned}
\|x_{k+1} - x^*\|^2 &\leq \|x_k - v_k g_k - x^*\|^2 \\
&= \|x_k - x^*\|^2 - 2v_k \langle g_k, x_k - x^* \rangle + v_k^2.
\end{aligned}
\tag{4.7}
$$

By the assumption that $f(x_k) > f^* + \epsilon$, Lemma 4.1 is applicable to concluding that

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

This, together with (4.7), implies the conclusion. $\qquad\square$

Lemma 4.2 shows that conditions (H1)-(H3) are satisfied for Algorithm 4.1 with

$$\epsilon > 0, \quad \alpha_k \equiv 2L^{-\frac{1}{p}}, \quad \beta_k \equiv 1, \quad \gamma_k \equiv 1.$$

Hence, the convergence theorems established in the preceding section can be directly applied to establish the convergence properties of the inexact subgradient method as follows.

Theorem 4.5. *Let $\{x_k\}$ be a sequence generated by Algorithm 4.1.*

(i) *If $v_k \equiv v > 0$, then $\liminf_{k\to\infty} f(x_k) \leq f^* + L \left(\frac{1}{2} v\right)^p + \epsilon$.*

(ii) *If $\{v_k\}$ satisfies (3.8), then $\liminf_{k\to\infty} f(x_k) \leq f^* + \epsilon$.*

(iii) *If $\{v_k\}$ is given by*

$$v_k := \lambda_k \left(\frac{f(x_k) - f^* - \epsilon}{L} \right)_+^{\frac{1}{p}}, \quad \text{where } 0 < \underline{\lambda} \leq \lambda_k \leq \overline{\lambda} < 2,
\tag{4.8}$$

then $\lim_{k\to\infty} f(x_k) \leq f^* + \epsilon$.

Theorem 4.6. *Let $\delta > 0$, and let $\{x_k\}$ be a sequence generated by Algorithm 4.1.*

(i) *Let $K_1 := \frac{L^{\frac{1}{p}} \operatorname{dist}^2(x_1, X^*)}{2v\delta}$ and $v_k \equiv v > 0$. Then $\min\limits_{1 \leq k \leq K_1} f(x_k) \leq f^* + \left(\frac{1}{2} L^{\frac{1}{p}} v + \delta\right)^p + \epsilon$.*

(ii) *Let $K_2 := \left(\frac{(1-s)L^{\frac{1}{p}} \operatorname{dist}^2(x_1, X^*)}{2c\delta}\right)^{\frac{1}{1-s}}$ and $\{v_k\}$ be given by (3.3). Then $\min\limits_{1 \leq k \leq K_2} f(x_k) \leq f^* + \left(\frac{1}{2} L^{\frac{1}{p}} c k^{-s} + \delta\right)^p + \epsilon$.*

(iii) *Let $K_3 := \frac{L^{\frac{2}{p}} \operatorname{dist}^2(x_1, X^*)}{\underline{\lambda}(2-\overline{\lambda})\delta^2}$ and $\{v_k\}$ be given by (4.8). Then $\min\limits_{1 \leq k \leq K_3} f(x_k) \leq f^* + \delta^p + \epsilon$.*

Theorem 4.7. *Let $\{x_k\}$ be a sequence generated by Algorithm 4.1. Suppose that X^* is the set of boundedly weak sharp minima of order q for f over X with modulus η.*

(I) *Suppose that $v_k \equiv v > 0$ and that f is coercive.*

    (i) *If $q = 2p$, then either $x_k \in X_\epsilon^*$ for some $k \in \mathbb{N}$ or there exist $\tau \in [0, 1)$ and $N \in \mathbb{N}$ such that, for each $k \in \mathbb{N}$,*

$$\operatorname{dist}^2(x_{k+N}, X^*) \leq \tau^k \operatorname{dist}^2(x_N, X^*) + 2^{\frac{1}{p}-1} \eta^{-\frac{1}{p}} \left(\epsilon^{\frac{1}{p}} + \frac{1}{2} L^{\frac{1}{p}} v\right).$$

    (ii) *If $q > 2p$ and $\epsilon^{\frac{1}{p}} + \frac{1}{2} L^{\frac{1}{p}} v < \eta^{-\frac{2}{q-2p}} \left(\frac{p}{vq} L^{\frac{1}{p}}\right)^{\frac{q}{q-2p}}$, then either $x_k \in X_\epsilon^*$ for some $k \in \mathbb{N}$ or there exist $\tau \in (0, 1)$ and $N \in \mathbb{N}$ such that, for each $k \in \mathbb{N}$,*

$$\operatorname{dist}^2(x_{k+N}, X^*) \leq \tau^k \operatorname{dist}^2(x_N, X^*) + 2^{\frac{2}{q}-\frac{2p}{q}} \eta^{-\frac{2}{q}} \left(\epsilon^{\frac{1}{p}} + \frac{1}{2} L^{\frac{1}{p}} v\right)^{\frac{2p}{q}}.$$

(II) *Suppose that $\{v_k\}$ is given by (3.3), $q = 2p$ and that f is coercive. Then, either $x_k \in X_\epsilon^*$ for some $k \in \mathbb{N}$ or there exist $C > 0$ and $\tau \in (0, 1)$ such that*

$$\operatorname{dist}^2(x_k, X^*) \leq C\tau^k + \left(\frac{2\epsilon}{\eta}\right)^{\frac{1}{p}} \quad \text{for each } k \in \mathbb{N}.$$

(III) *Suppose that $\{v_k\}$ is given by (4.8).*

    (i) *If $q = p$, then either $x_k \in X_\epsilon^*$ for some $k \in \mathbb{N}$ or there exist $\tau \in [0, 1)$ and $N \in \mathbb{N}$ such that*

$$\operatorname{dist}^2(x_{k+N}, X^*) \leq \tau^k \operatorname{dist}^2(x_N, X^*) + 2^{\frac{2}{p}-1} \eta^{-\frac{2}{p}} \epsilon^{\frac{2}{p}} \quad \text{for each } k \in \mathbb{N}.$$

    (ii) *If $q > p$ and $\epsilon < \left(\frac{L}{\eta}\right)^{\frac{q}{q-p}} \left(\frac{q}{p}\underline{\lambda}(2-\overline{\lambda})\right)^{-\frac{pq}{2(q-p)}}$, then either $x_k \in X_\epsilon^*$ for some $k \in \mathbb{N}$ or there exist $\tau \in (0, 1)$ and $N \in \mathbb{N}$ such that*

$$\operatorname{dist}^2(x_{k+N}, X^*) \leq \tau^k \operatorname{dist}^2(x_N, X^*) + 2^{\frac{2}{q}-\frac{p}{q}} \eta^{-\frac{2}{q}} \epsilon^{\frac{2}{q}} \quad \text{for each } k \in \mathbb{N}.$$

23

Following a line of analysis similar to that of Theorem 4.4, we can obtain the finite convergence property of the inexact subgradient method to the approximate solution set X_ϵ^* under the interior point assumption of X^*. The proof is omitted.

Theorem 4.8. *Let $\{x_k\}$ be a sequence generated by Algorithm 4.1. Let $x^* \in X^*$ and $\sigma > 0$, and suppose that $\mathbf{B}(x^*, \sigma) \subseteq X^*$. Then $x_k \in X_\epsilon^*$ for some $k \in \mathbb{N}$, provided one of the following conditions:*

(i) *$v_k = v \in (0, 2\sigma)$ for each $k \in \mathbb{N}$.*

(ii) *$\{v_k\}$ satisfies (3.8).*

4.3 Conditional subgradient method

The standard subgradient method (i.e., Algorithm 1.1) usually suffers from a zig-zagging phenomenon and sustains a slow convergence in practical applications. To avoid the zig-zagging phenomenon and speed up the convergence behavior, an idea of conditional subgradient method was proposed for either convex optimization [30] or quasi-convex optimization problems [23], which is stated as follows. It is worth mentioning that the algorithmic procedure of the conditional subgradient method is totally different from the conditional gradient method (also named the Franke-Wolfe method) [14], although they share a similar name.

Algorithm 4.2. Select an initial point $x_1 \in \mathbb{R}^n$ and a sequence of stepsizes $\{v_k\} \subseteq (0, +\infty)$. For each $k \in \mathbb{N}$, given x_k, we calculate

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

and update x_{k+1} by

$$x_{k+1} := \mathbb{P}_X(x_k - v_k(g_k + \mu_k)).$$

It was proved in [23, Lemma 3.2] the condition subgradient method satisfies the basic inequality as follow under the blanket assumption:

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

whenever $f(x_k) > f^*$, and $\|x_{k+1} - x_k\| \leq 2v_k$. This shows that conditions (H1)-(H3) are satisfied with

$$\epsilon = 0, \quad \alpha_k \equiv 2L^{-\frac{1}{p}}, \quad \beta_k \equiv 4, \quad \gamma_k \equiv 2.$$

Hence, the convergence theorems established in the preceding section can be directly applied to the conditional subgradient method as follows.

Theorem 4.9. *Let $\{x_k\}$ be a sequence generated by Algorithm 4.2.*

(i) *If $v_k \equiv v > 0$, then $\liminf_{k\to\infty} f(x_k) \leq f^* + L(2v)^p$.*

(ii) *If $\{v_k\}$ satisfies (3.8), then $\liminf_{k\to\infty} f(x_k) \le f^*$.*

(iii) *If $\{v_k\}$ is given by*

$$v_k := \frac{\lambda_k}{4}\left(\frac{f(x_k) - f^*}{L}\right)^{\frac{1}{p}}, \quad \text{where } 0 < \underline{\lambda} \le \lambda_k \le \overline{\lambda} < 2, \tag{4.9}$$

then $\lim_{k\to\infty} f(x_k) = f^*$.

Theorem 4.10. *Let $\delta > 0$, and let $\{x_k\}$ be a sequence generated by Algorithm 4.2.*

(i) *Let $K_1 := \frac{L^{\frac{1}{p}}\operatorname{dist}^2(x_1, X^*)}{2v\delta}$ and $v_k \equiv v > 0$. Then $\min_{1 \le k \le K_1} f(x_k) \le f^* + \left(2L^{\frac{1}{p}}v + \delta\right)^p$.*

(ii) *Let $K_2 := \left(\frac{(1-s)L^{\frac{1}{p}}\operatorname{dist}^2(x_1, X^*)}{2c\delta}\right)^{\frac{1}{1-s}}$ and $\{v_k\}$ be given by (3.3). Then $\min_{1 \le k \le K_2} f(x_k) \le f^* + \left(2L^{\frac{1}{p}}ck^{-s} + \delta\right)^p$.*

(iii) *Let $K_3 := \frac{4L^{\frac{2}{p}}\operatorname{dist}^2(x_1, X^*)}{\underline{\lambda}(2-\overline{\lambda})\delta^2}$ and $\{v_k\}$ be given by (4.9). Then $\min_{1 \le k \le K_3} f(x_k) \le f^* + \delta^p$.*

Theorem 4.11. *Let $\{x_k\}$ be a sequence generated by Algorithm 4.2. Suppose that X^* is the set of boundedly weak sharp minima of order q for f over X with modulus η.*

(I) *Suppose that $v_k \equiv v > 0$ and that f is coercive.*

(i) *If $q = 2p$, then either $x_k \in X^*$ for some $k \in \mathbb{N}$ or there exist $\tau \in [0, 1)$ and $N \in \mathbb{N}$ such that*

$$\operatorname{dist}^2(x_{k+N}, X^*) \le \tau^k \operatorname{dist}^2(x_N, X^*) + 2^{\frac{1}{p}}\left(\frac{L}{\eta}\right)^{\frac{1}{p}}v \quad \text{for each } k \in \mathbb{N}.$$

(ii) *If $q > 2p$ and $v < \left(2^p\frac{L}{\eta}(\frac{p}{2q})^{\frac{q}{2}}\right)^{\frac{1}{q-p}}$, then either $x_k \in X^*$ for some $k \in \mathbb{N}$ or there exist $\tau \in (0, 1)$, $c > 0$ and $N \in \mathbb{N}$ such that*

$$\operatorname{dist}^2(x_{k+N}, X^*) \le \tau^k \operatorname{dist}^2(x_N, X^*) + 2^{\frac{2}{q}}\left(\frac{L}{\eta}\right)^{\frac{2}{q}}v^{\frac{2p}{q}} \quad \text{for each } k \in \mathbb{N}.$$

(II) *Suppose that $\{v_k\}$ is given by (3.3), $q = 2p$ and that f is coercive. Then, either $x_k \in X^*$ for some $k \in \mathbb{N}$ or there exists $N \in \mathbb{N}$ such that*

$$\operatorname{dist}^2(x_k, X^*) \le 2c\left(\frac{2L}{\eta}\right)^{\frac{1}{p}}k^{-s} \quad \text{for each } k \ge N.$$

(III) *Suppose that $\{v_k\}$ is given by (4.9).*

(i) *If $q = p$, then there exist $\tau \in [0, 1)$ and $N \in \mathbb{N}$ such that*

$$\text{dist}^2(x_{k+N}, X^*) \leq \tau^k \text{dist}^2(x_N, X^*) \quad \text{for each } k \in \mathbb{N}.$$

(ii) *If $q > p$, then there exist $\gamma > 0$ and $N \in \mathbb{N}$ such that*

$$\text{dist}^2(x_{k+N}, X^*) \leq \frac{\text{dist}^2(x_N, X^*)}{(1 + \gamma k)^{\frac{p}{q-p}}} \quad \text{for each } k \in \mathbb{N}.$$

The following finite convergence property of the conditional subgradient method is recalled from [23, Theorem 3.6].

Theorem 4.12. *Let $\{x_k\}$ be a sequence generated by Algorithm 4.2. 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 each $k \in \mathbb{N}$.*

(ii) *$\{v_k\}$ satisfies (3.8).*

References

[1] Apolinario H.C.F., Papa Quiroz, E.A., Oliveira, P.R.: A scalarization proximal point method for quasiconvex multiobjective minimization. *Journal of Global Optimization* 64, 79–96 (2016)

[2] Auslender, A., Teboulle, M.: Projected subgradient methods with non-Euclidean distances for non-differentiable convex minimization and variational inequalities. *Mathematical Programming* 120, 27–48 (2009)

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

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

[5] Bao, J., Yu, C.K.W., Wang, J., Hu, Y., Yao, J.C.: Modified inexact Levenberg-Marquardt methods for solving nonlinear least squares problems. *Computational Optimization and Applications* (2019).

[6] Bertsekas, D.P.: *Nonlinear Programming.* Athena Scientific, Cambridge (1999)

[7] Bolte, J., Nguyen, T.P., Peypouquet, J., Suter, B.W.: From error bounds to the complexity of first-order descent methods for convex functions. *Mathematical Programming* 165, 471–507 (2017)

[8] Brännlund, U.: A generalized subgradient method with relaxation step. *Mathematical Programming* 71, 207–219 (1995)

[9] Burke, J.V., Deng, S.: Weak sharp minima revisited, part I: Basic theory. *Control and Cybernetics* 31, 439–469 (2002)

[10] Burke, J.V., Ferris, M.C.: Weak sharp minima in mathematical programming. *SIAM Journal on Control and Optimization* 31, 1340–1359 (1993)

[11] Crouzeix, J.P., Martinez-Legaz, J.E., Volle, M.: *Generalized Convexity, Generalized Monotonicity*. Kluwer Academic Publishers, Dordrecht (1998)

[12] Ermoliev, Y.M.: Methods of solution of nonlinear extremal problems. *Cybernetics and Systems Analysis* 2, 1–14 (1966)

[13] Ferris, M.C.: Finite termination of the proximal point algorithm. *Mathematical Programming* 50, 359–366 (1991)

[14] Frank, M., Wolfe, P.: An algorithm for quadratic programming. *Naval Research Logistics Quarterly* 3, 95–110 (1956)

[15] Freund, R.M., Lu, H.: New computational guarantees for solving convex optimization problems with first order methods, via a function growth condition measure. *Mathematical Programming* 170, 445–477 (2018)

[16] Goffin, J.L.: On convergence rates of subgradient optimization methods. *Mathematical Programming* 13, 329–347 (1977)

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

[18] Gürbüzbalaban, M., Ozdaglar, A., Ribeiro, P.A.: On the convergence rate of incremental aggregated gradient algorithms. *SIAM Journal on Optimization* 27, 1035–1048 (2017)

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

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

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

[22] Hu, Y., Yang, X., Yu, C.K.W.: Subgradient methods for saddle point problems of quasiconvex optimization. *Pure and Applied Functional Analysis* 2, 83–97 (2017)

[23] Hu, Y., Yu, C.K.W., Li, C., Yang, X.: Conditional subgradient methods for constrained quasi-convex optimization problems. *Journal of Nonlinear and Convex Analysis* 17, 2143–2158 (2016)

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

[25] Johnstone, P. R., Moulin, P.: Faster subgradient methods for functions with Hölderian growth. *Mathematical Programming* 180, 417–450 (2020).

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

[27] Kiwiel, K.C.: Convergence of approximate and incremental subgradient methods for convex optimization. *SIAM Journal on Optimization* 14, 807–840 (2004)

[28] Konnov, I.V.: On properties of supporting and quasi-supporting vectors. *Journal of Mathematical Sciences* 71, 2760–2763 (1994)

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

[30] Larsson, T., Patriksson, M., Strömberg, A.B.: Conditional subgradient optimization–theory and applications. *European Journal of Operational Research* 88, 382–403 (1996)

[31] Lu, Z., Zhang, Y., Lu, J.: ℓ_p Regularized low-rank approximation via iterative reweighted singular value minimization. *Computational Optimization and Applications* 68, 619-642 (2017)

[32] Mastrogiacomo, M., Gianin, E.R.: Portfolio optimization with quasiconvex risk measures. *Mathematics of Operations Research* 40, 1042–1059 (2015)

[33] Mokhtari, A., Gürbüzbalaban, M., Ribeiro, A. Surpassing gradient descent provably: A cyclic incremental method with linear convergence rate. *SIAM Journal on Optimization* 28, 1420–1447 (2018)

[34] Nedić, A., Bertsekas, D.P.: Convergence rate of incremental subgradient algorithms. *In: Stochastic Optimization: Algorithms and Applications*, pp. 223–264. Springer (2001)

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

[36] Nedić, A., Ozdaglar, A.: Subgradient methods for saddle-point problems. *Journal of Optimization Theory and Applications* 142, 205–228 (2009)

[37] Nesterov, Y.: Primal-dual subgradient methods for convex problems. *Mathematical Programming* 120, 221–259 (2009)

[38] Nesterov, Y., Shikhman, V.: Dual subgradient method with averaging for optimal resource allocation. *European Journal of Operational Research* 270, 907–916 (2018)

[39] Neto, E.S.H., Pierro, A.R.D.: Incremental subgradients for constrained convex optimization: A unified framework and new methods. *SIAM Journal on Optimization* 20, 1547–1572 (2009)

[40] Poljak, B.T.: Nonlinear programming methods in the presence of noise. *Mathematical Programming* 14, 87–97 (1978)

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

[42] Polyak, B.T.: *Introduction to Optimization.* Optimization Software, New York (1987)

[43] Robinson, S.M.: Linear convergence of epsilon-subgradient descent methods for a class of convex functions. *Mathematical Programming* 86, 41–50 (1999)

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

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

[46] Papa Quiroz, E.A., Oliveira, P.R.: Proximal point methods for quasiconvex and convex functions with Bregman distances on Hadamard manifolds. *Journal of Convex Analysis* 16, 49–69 (2009)

[47] Papa Quiroz, E.A., Quispe Cardenas, E.M., Oliveira P.R.: Steepest descent method with a generalized Armijo search for quasiconvex functions on Riemannian manifolds. *Journal of Mathematical Analysis and Applications* 341, 467–477 (2008)

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

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

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

[51] Studniarski, M., Ward, D.E.: Weak sharp minima: Characterizations and sufficient conditions. *SIAM Journal on Control and Optimization* 38, 219–236 (1999)

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

[53] Wang, J., Li, C., Lopez, G., Yao, J.-C.: Proximal point algorithms on Hadamard manifolds: Linear convergence and finite termination. *SIAM Journal on Optimization* 26, 2696–2729 (2016).

[54] Xu, Y., Lin, Q., Yang, T.: Accelerate stochastic subgradient method by leveraging local error bound. *arXiv:1607.01027* (2018)

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

[56] Zhou, Z., So, A.M.C.: A unified approach to error bounds for structured convex optimization problems. *Mathematical Programming* 165, 689–728 (2017)