Noname manuscript No.
(will be inserted by the editor)

Interior Quasi-subgradient Method with non-Euclidean Distances for Constrained Quasi-convex Optimization Problems in Hilbert Spaces

Regina S. Burachik · Yaohua Hu · Xiaoqi Yang

Received: date / Accepted: date

Abstract An interior quasi-subgradient method is proposed based on the proximal distance to solve constrained nondifferentiable quasi-convex optimization problems in Hilbert spaces. It is shown that a newly introduced generalized Gâteaux subdifferential is a subset of a quasi-subdifferential. The convergence properties, including the global convergence and iteration complexity, are investigated under the assumption of the Hölder condition of order p, when using the constant/diminishing/dynamic stepsize rules. Convergence rate results are obtained by assuming a Hölder-type weak sharp minimum condition relative to an induced proximal distance.

Keywords Quasi-convex optimization · Interior subgradient method · Proximal distance · Convergence analysis

Mathematics Subject Classification (2010) MSC 65K05 · MSC 90C26 · 49M37

1 Introduction

Mathematical optimization provides a unified framework for a wide variety of application problems in many disciplines, in which we usually consider a general constrained optimization problem

$$\begin{array}{ll} \min f(x) \\ \text{s.t.}\;\; x \in X, \end{array} \tag{1}$$

R. S. Burachik
Mathematics, UniSA STEM, University of South Australia, Mawson Lakes, SA 5095, Australia
E-mail: Regina.Burachik@unisa.edu.au

Y. Hu
Shenzhen Key Laboratory of Advanced Machine Learning and Applications, College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China
E-mail: mayhhu@szu.edu.cn

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is a real-valued function, and $X \subseteq \mathbb{R}^n$ is a nonempty, closed and convex constraint set. Convex optimization plays a fundamental role in mathematical optimization. However, convexity can become a too restrictive assumption for many real-life problems encountered in economics, finance and management science. For the latter problems, quasi-convex optimization can provide a much more accurate representation of reality, while still sharing some desirable properties enjoyed by convex optimization problems. Hence there is a significant increase of interest in quasi-convex optimization; see [2,11,27,31,36,60] and references therein. This also results in the development of numerical algorithms for solving quasi-convex optimization problems; see [20,33,34,37,40,49,59,62] and references therein.

Popular first-order algorithms for solving (convex or quasi-convex) optimization problems are the so-called *projected subgradient methods*. The classical projected subgradient method was originally introduced by Polyak [52] and Ermoliev [29] in the 1970s to solve the nondifferentiable convex optimization problem (1) (i.e., with f in (1) assumed to be convex) and has the following iterative formula:

$$x_{k+1} := \mathrm{P}_X(x_k - v_k g_k), \tag{2}$$

where $\mathrm{P}_X(\cdot)$ denotes the Euclidean projection onto X, g_k is a (convex) subgradient of f at x_k, and v_k is a positive stepsize. Over the past five decades, various features of projected subgradient methods have been established for convex optimization problems [17,38,57] and many applications/extensions have been devised for structured convex optimization problems [17,32,42].

Projected subgradient methods have also been extended and developed to solve nondifferentiable quasi-convex optimization problems (i.e., (1) with f being quasi-convex); see [33, 34,37,40,62] and references therein. In the literature of projected subgradient methods, the convergence theory, including the global convergence results, iteration complexity and convergence rates, have been well studied for the quasi-convex optimization problems by using several typical stepsize rules. In particular, Kiwiel [37] proposed a projected subgradient method with a diminishing stepsize rule and explored the global convergence of function values; in addition, the convergence of iterates was established under an additional regular condition. Konnov [40] introduced a dynamic stepsize rule for the projected subgradient method, presented the global convergence in both function values and iterates under the assumption of a Hölderian condition, and established the sublinear convergence rate under an additional assumption of weak sharp minimum of Hölderian order. Hu et al. [34] considered a generic inexact projected subgradient method with the constant or diminishing stepsize rules, and studied the influence of the deterministic noise on convergence theory (under the Hölderian condition). Moreover, Hu et al. [33] explored the iteration complexity and the linear or sublinear convergence rates (under an additional assumption of weak sharp minimum of Hölderian order) of the (inexact) projected subgradient method with the constant, diminishing and dynamic stepsize rules.

Unfortunately, the classical projected subgradient method suffers from several disadvantages arising from the Euclidean projection; see, e.g., [4,6]. In particular, the Euclidean projection operator destroys the nice descent property and might often lead to a zigzagging effect, resulting in slow convergence, despite theoretical convergence guarantee. Moreover, the Euclidean projection operator could be computationally expensive, if the constraint set X is not simple.

Note that the subgradient iteration (2) has an equivalent representation as a proximal operator applied to the linear function induced by the current subgradient. Namely,

$$x_{k+1} = \arg\min\{\langle v_k g_k, z\rangle + \frac{1}{2}\|z - x_k\|^2 : z \in X\}.$$

To avoid the disadvantages arising from the Euclidean projection operator, one common approach is to replace the Euclidean distance by a distance-like function. Namely,

$$x_{k+1} = \arg\min\{\langle v_k g_k, z\rangle + d(z, x_k) : z \in X\}, \tag{3}$$

where the distance-like function $d(\cdot,\cdot)$ ensures that the iterates always remain in the interior of X. Hence, the resulting algorithm (3) is called *interior subgradient method*. In this situation, the constraints of the original problem (1) are automatically eliminated.

The history of the interior subgradient method originates in 1983 from the mirror descent method proposed by Nemirovsky and Yudin [43], which was applied to solve efficiently convex optimization problems over the unit simplex with millions of variables [16]. It was shown in Beck and Teboulle [15] that the mirror descent method can be viewed as an interior subgradient method with $d(\cdot,\cdot)$ in (3) being a Bregman distance [19]. The convergence theory and the iteration complexity of interior subgradient methods have been well studied for constrained convex optimization, conic optimization and variational inequality problems in [6,5,4] and references therein.

It was revealed in the literature that the interior subgradient method enjoys several advantages: (i) it requires only first-order information, (ii) for particular types of constraints (such as the polyhedron and the nonnegative orthant) and suitable proximal distance $d(\cdot,\cdot)$, it generates simple iterative schemes, and (iii) it exhibits a nearly dimension independent computational complexity in terms of the problem's dimension; see, e.g., [5,6,15,16].

It is worth mentioning that the use of non-Euclidean distance-like function $d(\cdot,\cdot)$ in place of the Euclidean distance has also been extensively pursued in the literature of proximal point methods for convex optimization problems [7,8,23,24,26,41,46,50,55,56,61] and quasi-convex optimization problems [20,28,44,45,48,49,47,59]. However, to the best of our knowledge, there is still no study devoted to the interior subgradient method based on non-Euclidean distances $d(\cdot,\cdot)$ for quasi-convex optimization problems.

Our aim is twofold. First, we extend interior subgradient methods that use generalized distances $d(\cdot,\cdot)$ to the framework of constrained quasi-convex optimization problems. Second, we investigate the convergence properties of the resulting method. Our convergence analysis is developed for three prototypical types of stepsizes: constant, diminishing, and dynamic. Moreover, our types of distances include Bregman distances, φ-divergences, and second order homogeneous kernels.

In our analysis, we consider a general constrained nondifferentiable quasi-convex optimization problem in Hilbert spaces:

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t.}\ \ x \in X := \mathrm{cl}C \cap V, \end{aligned} \tag{4}$$

where $f : \mathbb{H} \to \mathbb{R}$ is a continuous and quasi-convex function, $\mathbb{H}$ is a Hilbert space, and $C \subseteq \mathbb{H}$ is a non-empty, open and convex set with closure $\mathrm{cl}C$, and $V \subseteq \mathbb{H}$ is a closed and convex set (V is considered as a linear manifold in [4,6] and the whole space in [3,5]). The optimal value of problem (4) is denoted by f^*, and the optimal solution set, assumed to be nonempty, is denoted by X^*.

Inspired by the idea of the interior subgradient method [4–6], we propose an interior quasi-subgradient method of iterative form (3) to solve problem (4). Our proposed scheme uses a unit quasi-subgradient in place of g_k, a suitable proximal distance in place of $d(\cdot,\cdot)$, and V in place of X; see Algorithm 1 for details.

The projected subgradient methods for solving convex optimization problems in Hilbert spaces have been studied. Alber et al. [1] established the convergence of function values and the weak convergence of the iterates to an optimal solution for the diminishing stepsize rule. Solodov and Zavriev [58] and Barty et al. [12] explored the weak convergence of the iterates of the inexact subgradient methods with deterministic noise or stochastic noise respectively under certain assumptions of noise. Furthermore, Barty et al. [12] also established the strong convergence of projected subgradient methods under an additional assumption of strong convexity. Kiwiel [37] studied the weak convergence of the projected subgradient method with diminishing stepsize rule under a regular condition for a quasi-convex optimization problem.

We will extend the convergence theory of quasi-subgradient methods to the non-Euclidean projection and Hilbert space setting. For the convergence analysis of the interior quasi-subgradient method, under the assumption of Hölder continuity, we establish global convergence and derive the iteration complexity for convergence towards the optimal value for the cases of constant, diminishing, and dynamic stepsizes, respectively (see Theorems 1(i),(ii)-3(i),(ii)). Under suitable assumptions, we establish weak convergence of the iterates to an optimal solution for the diminishing and dynamic stepsizes. In the finite-dimensional case, we establish convergence to a solution (see Theorems 2(iii)(b)-3(iii)(b)). Furthermore, we introduce the notion of *weak sharp minimum of Hölderian order* relative to an induced proximal distance, and use it to study the linear (or sublinear) convergence rates of the interior quasi-subgradient method; see Theorems 4-6 for details.

In particular, the present paper extends the weak convergence theory in [37] to the dynamic stepsize rule and the non-Euclidean projection; particularly, we establish the weak convergence of the interior quasi-subgradient method with the diminishing and dynamic stepsize rules under the assumptions of the Lipschitz condition and the norm compatibility/strictly convexity of Bregman kernel; see Theorems 2(iii)(a)-3(iii)(a). To the best of our knowledge, our work seems to be the first discussing the subgradient method with non-Euclidean distance in Hilbert spaces, even for convex optimization problems.

The paper is organized as follows. In section 2, we present the notations and preliminary results, including the properties of quasi-subdifferentials, used in this paper. We show that a newly introduced generalized Gâteaux subdifferential is a subset of a quasi-subdifferential. In section 3, we propose the interior quasi-subgradient method for solving the constrained quasi-convex problem (4). In section 4, we investigate convergence analysis of the interior quasi-subgradient method when using three prototypical stepsizes. The convergence properties we establish include: global convergence, iteration complexity, and convergence rates. A conclusion is given in section 5.

2 Notations and preliminary results

In the present paper, we consider a Hilbert space $\mathbb{H}$ with an inner product $\langle\cdot,\cdot\rangle$ and its associated norm $\|\cdot\|$, and use the norm topology, unless clearly specified. Particularly, for $\{x_k\}\subseteq\mathbb{H}$ and $x\in\mathbb{H}$, we use $x_k\to x$ and $x_k\rightharpoonup x$ to denote that $\{x_k\}$ strongly and weakly converges to x, respectively; namely, $\lim_{k\to\infty}x_k=x$ and $\text{w-}\lim_{k\to\infty}x_k=x$, respectively. For $x\in\mathbb{H}$ and $r>0$, we use $\mathbb{B}(x,r)$ to denote the closed ball centered at x with radius r, and use $\mathbb{S}$ to denote the unit sphere centered at the origin. For $x\in\mathbb{H}$ and $Z\subseteq\mathbb{H}$, we write $\mathrm{dist}(x,Z)$ and $\mathrm{P}_Z(x)$ to denote the distance of x from Z and the projection of x onto Z, respectively; namely,

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

If Z is convex, the normal cone of Z at $x\in Z$ is defined by

$$\mathrm{N}_Z(x):=\{y\in\mathbb{H}:\langle y,z-x\rangle\le 0\text{ for any }z\in Z\}.\tag{5}$$

For $h:\mathbb{H}\to\mathbb{R}$ and $x\in\mathbb{H}$, the level sets of h at x are denoted by

$$\mathrm{lev}_h^{<}(x):=\{y\in\mathbb{H}:h(y)<h(x)\}\quad\text{and}\quad \mathrm{lev}_h^{\le}(x):=\{x\in\mathbb{H}:h(y)\le h(x)\}.$$

A function $h:\mathbb{H}\to\mathbb{R}$ is said to be quasi-convex if

$$h((1-\alpha)x+\alpha y)\le\max\{h(x),h(y)\}\quad\text{for any }x,y\in\mathbb{H}\text{ and any }\alpha\in[0,1].\tag{6}$$

While a convex function can be characterized by the convexity of its epigraph, a quasi-convex function can be characterized by the convexity of its level sets. The following proposition is taken from [14, Definition 10.18 and Proposition 10.24].

Proposition 1 $h:\mathbb{H}\to\mathbb{R}$ *is quasi-convex if and only if* $\mathrm{lev}_h^{<}(x)$ *(and/or* $\mathrm{lev}_h^{\le}(x)$*) is convex for each* $x\in\mathbb{H}$*.*

The subdifferential of a quasi-convex function plays an important role in quasi-convex optimization. Several specific types of subdifferentials have been introduced and explored for quasi-convex functions that are defined via the "normal cone" to the level sets; see [9, 10,30,51] and references therein. In particular, Kiwiel [37], Censor and Segal [25], and Hu et al. [33,34] utilized a quasi-subgradient for developing and analyzing quasi-subgradient methods. In the following definition, we recall the notion of quasi-subdifferential from [51].

Definition 1 Let $h:\mathbb{H}\to\mathbb{R}$ be a quasi-convex function, and let $x\in\mathbb{H}$. The quasi-subdifferential of h at x is defined by

$$\partial^{\mathrm{Q}}h(x):=\mathrm{N}_{\mathrm{lev}_h^{<}(x)}(x)=\{g:\langle g,y-x\rangle\le 0\text{ for any }y\in\mathrm{lev}_h^{<}(x)\}.\tag{7}$$

It is an essential property that the convex subdifferential of each convex function is nonempty. The following proposition states that the quasi-subdifferential of a quasi-convex and upper semicontinuous function is nontrivial, which extends the property in Euclidean spaces [34, Lemma 2.1] to Hilbert spaces. The proof is similar to that of [34, Lemma 2.1] (and uses the separation theorem in Hilbert spaces [64, Theorem 1.1.3]), and thus is omitted.

Proposition 2 *Let $h : \mathbb{H} \to \mathbb{R}$ be a quasi-convex and upper semicontinuous function. Then $\partial^{\mathrm{Q}} h(x) \setminus \{0\} \neq \emptyset$ for each $x \in \mathbb{H}$.*

From Definition 1, the quasi-subgradient is not easy to calculate. In fact, we shall calculate the level set of the function and then estimate one of its normal vectors. To avert this difficulty, we provide an alternative approach for a quasi-subgradient via its Gâteaux derivatives. We recall the relevant definition below.

Definition 2 Let $h : \mathbb{H} \to \mathbb{R}$ and $x, u \in \mathbb{H}$. The directional derivative of h at x along u is defined by

$$h'(x,u) := \lim_{t \to 0_+} \frac{h(x+tu) - h(x)}{t}.$$

h is said to be Gâteaux differentiable at x if there exists a (necessarily unique) $h'_{\mathrm{G}}(x) \in \mathbb{H}$ (called Gâteaux derivative) such that

$$h'(x,u) = \langle h'_{\mathrm{G}}(x), u \rangle \quad \text{for each } u \in \mathbb{H}. \tag{8}$$

By [54, Theorems 4.2 and 6.2], a quasi-convex and densely continuous function in Hilbert spaces is (a.e.) Gâteaux differentiable. Hence we define the following generalized Gâteaux subdifferential by using the weak limit of the Gâteaux derivatives of h nearby x.

Definition 3 Let $f : h : \mathbb{H} \to \mathbb{R}$ be a quasi-convex and densely continuous function and $x \in \mathbb{H}$. The generalized Gâteaux subdifferential of h at x is defined by

$$\partial^{\mathrm{G}} h(x) := \left\{ g \in \mathbb{H} : \exists \; x_i \xrightarrow{h} x \text{ such that } h'_{\mathrm{G}}(x_i) \rightharpoonup g \right\}, \tag{9}$$

where $x_i \xrightarrow{h} x$ means that $x_i \to x$ with $h(x_i) \to h(x)$.

The following proposition shows that the generalized Gâteaux subdifferential of h at x is a subset of the quasi-subdifferential of h at x. Hence, we may obtain a quasi-subgradient by computing the Gâteaux derivative of a quasi-convex function (at a Gâteaux differentiable point), or by computing the weak limit of Gâteaux derivatives close to a nondifferentiable point.

Proposition 3 *Let $h : \mathbb{H} \to \mathbb{R}$ be quasi-convex and continuous. Then $\partial^{\mathrm{G}} h(\cdot) \subseteq \partial^{\mathrm{Q}} h(\cdot)$.*

Proof Fix $x \in \mathbb{H}$. We first show that

$$[h \text{ is Gâteaux differentiable at } x] \quad \Rightarrow \quad [\langle h'_{\mathrm{G}}(x), z - x \rangle \leq 0 \text{ for each } z \in \mathrm{lev}_h^{\leq}(x)]. \tag{10}$$

To this end, suppose that h is Gâteaux differentiable at x, and fix $z \in \mathrm{lev}_h^{\leq}(x)$. It follows from (8) (with $z - x$ in place of u) that for each $\epsilon > 0$, there exists $\delta(\epsilon) \in (0,1)$ such that

$$\frac{h(x + t(z-x)) - h(x)}{t} - \langle h'_{\mathrm{G}}(x), z - x \rangle \geq -\epsilon \quad \text{for each } 0 < t < \delta(\epsilon). \tag{11}$$

By the quasi-convexity of h, we can use (6) to write $h(x+t(z-x)) \leq \max\{h(x), h(z)\} = h(x)$ (thanks to $z \in \mathrm{lev}_h^{\leq}(x)$). Thus it follows from (11) that $\langle h'_{\mathrm{G}}(x), z - x \rangle \leq \epsilon$. Since $\epsilon > 0$ is arbitrary, it follows that $\langle h'_{\mathrm{G}}(x), z - x \rangle \leq 0$; consequently, (10) is proved as desired.

To complete the proof, fix $g \in \partial^{\mathrm{G}} h(x)$. By (9), there exists a sequence $\{x_i\}$, where h is Gâteaux differentiable, such that

$$x_i \xrightarrow{h} x \quad \text{and} \quad h'_{\mathrm{G}}(x_i) \rightharpoonup g. \tag{12}$$

Let $z \in \mathrm{lev}^{<}_{h}(x)$. Since $h(x_i) \to h(x)$ (by $x_i \xrightarrow{h} x$), there exists $k \in \mathbb{N}$ such that $z \in \mathrm{lev}^{<}_{h}(x_i)$ for each $i \geq k$. Then one has by (10) (with x_i in place of x) that

$$\langle h'_{\mathrm{G}}(x_i), z - x_i \rangle \leq 0 \quad \text{for each } i \geq k.$$

Using now (12), the fact that x_i converges strongly to x and the boundedness of $h'_{\mathrm{G}}(x_i)$, we arrive, after taking limits for $i \to \infty$, at $\langle g, z - x \rangle \leq 0$. Since $z \in \mathrm{lev}^{<}_{h}(x)$ is arbitrary, we can use (7) to conclude that $g \in \partial^{\mathrm{Q}} h(x)$. The proof is complete.

We end this section by recalling several lemmas, which will be useful in the convergence analysis of the interior quasi-subgradient method.

Lemma 1 ([38, Lemma 2.1]) *Let $\{a_k\}$ be a sequence of scalars, and let $\{v_k\}$ be a sequence of nonnegative scalars. Suppose that $\lim_{k\to\infty} \sum_{i=1}^{k} v_i = \infty$. Then it holds that*

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

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

Lemma 2 ([33, Lemma 2.2]) *Let $r > 0$, $a > 0$, $b \geq 0$, and let $\{u_k\}$ be a sequence of nonnegative scalars such that*

$$u_{k+1} \leq u_k - a u_k^{1+r} + b \quad \textit{for each } k \in \mathbb{N}.$$

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

$$u_k \leq u_0 \left(1 + r a u_0^r k\right)^{-\frac{1}{r}} \quad \textit{for each } k \in \mathbb{N}.$$

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

$$u_k \leq \tau^k u_0 + \left(\frac{b}{a}\right)^{\frac{1}{1+r}} \quad \textit{for each } k \in \mathbb{N}.$$

Lemma 3 ([53, pp. 46, Lemma 5]) *Let $a > 0$, $b > 0$, $s \in (0,1)$ and $t > s$, and let $\{u_k\}$ be a sequence of nonnegative scalars such that*

$$u_{k+1} \leq \left(1 - a k^{-s}\right) u_k + b k^{-t} \quad \textit{for each } k \in \mathbb{N}.$$

Then it holds that

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

Lemma 4 ([53, pp. 50, Lemma 11]) *Let $\{u_k\}$, $\{\beta_k\}$ and $\{\epsilon_k\}$ be three sequences of nonnegative scalars such that*

$$u_{k+1} \le u_k - \beta_k + \epsilon_k \quad \text{for each } k \in \mathbb{N}.$$

Suppose that $\sum_{k=1}^{\infty} \epsilon_k < +\infty$. Then $\{u_k\}$ is convergent and $\sum_{k=1}^{\infty} \beta_k < +\infty$.

Lemma 5 ([1, Proposition 2]) *Let $a > 0$, $\{u_k\}$ and $\{\beta_k\}$ be two sequences of nonnegative scalars satisfying*

$$\sum_{k=1}^{\infty} \beta_k = +\infty, \quad \sum_{k=1}^{\infty} \beta_k u_k < +\infty \quad \text{and} \quad u_{k+1} \le u_k + a\beta_k \quad \text{for each } k \in \mathbb{N}.$$

Then $\lim_{k\to\infty} u_k = 0$.

3 Interior quasi-subgradient method

In this section, we propose an interior quasi-subgradient method to solve the quasi-convex optimization problem (4). The principle of the interior quasi-subgradient method is to use a non-Euclidean distance-like function (satisfying certain desirable properties) in place of the Euclidean distance in the classical quasi-subgradient method [37]. More precisely, the projection is performed w.r.t. a suitably chosen distance. This will force the generated iterates to stay in the interior of the constraint set, and thus, automatically eliminate the constraints and also produce interior trajectories; see, e.g., [5,6].

To introduce the method, we define first a proximal distance that replaces the usual Euclidean distance and allows to handle the problem constraints as desired. In the definition below, item (i) is a distance-like property, item (ii) forces the generated iterates to stay in the open set C, and item (iii) constitutes a useful property for the convergence analysis. For a function $d : X \times Y \to \mathbb{R}$, and a fixed $y \in Y$, we denote by $\nabla_1 d(\cdot, y)$ and $\partial_1 d(\cdot, y)$ the derivative and the subdifferential of $d(\cdot, y)$ w.r.t. the first variable, respectively.

Definition 4 Let $C \subseteq \mathbb{H}$ be open and convex, and $V \subseteq \mathbb{H}$ be closed and convex. The mapping $d : \mathbb{H} \times \mathbb{H} \to \mathbb{R}_+ \cup \{+\infty\}$ is said to be a *proximal distance* if, for each $y \in C \cap V$, the following properties are satisfied:

(i) $d(\cdot, y)$ is proper, lower semicontinuous (lsc), convex and continuously differentiable on $C \cap V$ with $d(y, y) = 0$ and $\nabla_1 d(y, y) = 0$.
(ii) $\operatorname{dom} d(\cdot, y) \subseteq \operatorname{cl}C$ and $\operatorname{dom} \partial_1 d(\cdot, y) = C$.
(iii) $d(\cdot, y)$ is σ-strongly convex over $C \cap V$, i.e.,

$$\langle \nabla_1 d(x_1, y) - \nabla_1 d(x_2, y), x_1 - x_2 \rangle \ge \sigma \|x_1 - x_2\|^2 \quad \text{for each } x_1, x_2 \in C \cap V.$$

We use $\mathcal{D}(C, V)$ to denote the family of all functions $d(\cdot, \cdot)$ satisfying the premises above.

The types of proximal distances include Bregman distances, φ-divergences, and second order homogeneous kernels; see [5–7,22] and references therein. Given $d \in \mathcal{D}(C, V)$, we can define a projection-like mapping $\mathcal{P} : \mathbb{H} \times \mathbb{H} \to \mathbb{H}$

$$\mathcal{P}(g, x) := \arg\min\{\langle g, z\rangle + d(z, x) : z \in V\} \quad \text{for each } g \in \mathbb{H}, x \in C \cap V. \tag{13}$$

Remark 1 For $d \in \mathcal{D}(C, V)$, and thanks to Definition 4(i)-(iii), we have that $\mathcal{P}(\cdot,\cdot)$ is well-defined and is a single-valued map with images in $C \cap V$ (see also [6]). Since $\mathcal{P}(g, x) \in C$ and C is open, $\mathcal{P}$ is called an *interior projection mapping*. Moreover, for a given proximal distance $d(\cdot,\cdot)$, the interior projection mapping (13) performs a proximal-like step associated with a linear function without the need of considering the constraint C. On the other hand, the Euclidean projection operator executes a proximal step with a linear function, and the constraint $\mathrm{cl}C \cap V$ still needs to be imposed separately. This iteration is likely to be more complicated than (13).

The interior quasi-subgradient method for solving the quasi-convex optimization problem (4) is formally stated as follows. Recall that $\mathbb{S}$ denotes the unit sphere centered at the origin. In the special case when the proximal distance is chosen as the Euclidean distance $d(x, y) = \frac{1}{2}\|x - y\|^2$, the interior quasi-subgradient method is reduced to the classical projected quasi-subgradient method [37] for solving the quasi-convex optimization problem.

Algorithm 1 (Interior quasi-subgradient method) Select an initial point $x_1 \in \mathbb{H}$ and a sequence of stepsizes $\{v_k\} \subseteq (0, +\infty)$. For each $k \in \mathbb{N}$, given x_k, select $g_k \in \partial^{\mathrm{Q}} f(x_k) \cap \mathbb{S}$ and update x_{k+1} by

$$x_{k+1} := \mathcal{P}(v_k g_k, x_k). \tag{14}$$

The main computational task in Algorithm 1 is the interior projection mapping (13). For some choices of d, C, and V, the projection can be computed via a closed formula, and thus the resulting Algorithm 1 is particularly attractive; one can refer to [4–6, 22] for the detailed examples.

To measure the convergence properties of the interior quasi-subgradient method, we recall from [5, 6, 22] the definition of an induced proximal distance $\mathcal{H}(\cdot,\cdot)$ to $d(\cdot,\cdot)$.

Definition 5 Let $C \subseteq \mathbb{H}$ be open and convex, $V \subseteq \mathbb{H}$ be closed and convex, and $d \in \mathcal{D}(C, V)$. The function $\mathcal{H} : \mathbb{H} \times \mathbb{H} \to \mathbb{R}_+ \cup \{+\infty\}$ is said to be the *induced proximal distance to* $d(\cdot,\cdot)$, if

(a) $\mathcal{H}(\cdot,\cdot)$ is finite-valued on $\mathrm{cl}C \times C$, and $\mathcal{H}(x, x) = 0$ for each $x \in C$,
(b) $\mathcal{H}(z, x) - \mathcal{H}(z, y) \geq \langle z - y, \nabla_1 d(y, x)\rangle$ for each $x, y \in C$ and $z \in \mathrm{cl}C$.

In addition to (a) and (b), $\mathcal{H}$ verifies the following properties.

(i) For each $y \in \mathrm{cl}C$ and each bounded sequence $\{y_k\} \subseteq C$ with $\lim_{k\to\infty} \mathcal{H}(y, y_k) = 0$, then $y_k \rightharpoonup y$.
(ii) For each $y \in \mathrm{cl}C$ and $\{y_k\} \subseteq C$ with $y_k \to y$, then $\lim_{k\to\infty} \mathcal{H}(y, y_k) = 0$.
(iii) For each $y \in \mathrm{cl}C$ and $\{y_k\} \subseteq C$ with $\lim_{k\to\infty} \|y_k\| = \infty$, then $\lim_{k\to\infty} \mathcal{H}(y, y_k) = \infty$.

We write $(d, \mathcal{H}) \in \mathcal{F}_+(C, V)$ to express the fact that the quaternity $[C, V, d, \mathcal{H}]$ satisfies (a)-(b) and (i)-(iii) above.

Given an induced proximal distance $\mathcal{H}$ to d, we define the $\mathcal{H}$-distance from a point $x \in \mathbb{H}$ to a set $Z \subseteq \mathbb{H}$ as follows

$$\mathrm{dist}_{\mathcal{H}}(Z, x) := \inf_{z \in Z} \mathcal{H}(z, x).$$

The proximal distance is said to be self-proximal if $d = \mathcal{H}$. The readers can refer to [4, 5, 7, 22, 61] for examples of pair (d, H) in Euclidean spaces, such as Bregman distances, φ-divergences and second order homogeneous kernels. The following is an example of self-proximal distance, that uses Bregman distances [19]. This example defines a distance in an infinite dimensional Hilbert space following an idea in [21, Section 5.6]. The idea is to use a finite intersection of open half-spaces as the constraint set C, and then use the sum of two distances, one of them the square of the norm, to ensure strong convexity, and the other a distance d that takes care of the constraint C. In the following example, any φ-divergence, second order homogeneous kernel, or Bregman distance can be used as a first term in equation (18).

Example 1 We first recall the Bregman distance in Euclidean spaces; see, e.g., [26,22]. Let $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a Legendre function satisfying the following conditions:

(A_1) φ is proper, lsc, and convex with $\operatorname{dom}\varphi \subseteq \mathbb{R}^n_+$ and $\operatorname{dom}\nabla\varphi = \mathbb{R}^n_{++}$.
(A_2) φ is strictly convex and continuous on $\operatorname{dom}\varphi$, and continuously differentiable on $\mathbb{R}^n_{++}$.

Associated to this φ, the Bregman distance $\mathcal{D}_\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ is defined by

$$\mathcal{D}_\varphi(x,y) := \begin{cases} \varphi(x) - \varphi(y) - \langle \nabla\varphi(y), x-y\rangle, \ \forall x \in \mathbb{R}^n_+, y \in \mathbb{R}^n_{++}, \\ +\infty, \qquad\qquad\qquad\qquad\qquad \text{otherwise.} \end{cases} \tag{15}$$

The Bregman distance $\mathcal{D}_\varphi$ enjoys a remarkable three point identity [26, Lemma 3.1] that

$$\mathcal{D}_\varphi(z,x) = \mathcal{D}_\varphi(z,y) + \mathcal{D}_\varphi(y,x) + \langle \nabla_1 \mathcal{D}_\varphi(y,x), z-y\rangle. \tag{16}$$

By (16) and the convexity of φ, the Bregman distance is self-proximal, that is, $d = \mathcal{H} = \mathcal{D}_\varphi$, or equivalently, $(\mathcal{D}_\varphi, \mathcal{D}_\varphi) \in \mathcal{F}_+(\mathbb{R}^n_{++}, \mathbb{R}^n)$.

Now we introduce the Bregman distance in infinite-dimensional Hilbert spaces. Let

$$C := \{x \in \mathbb{H} : \langle A_i, x\rangle > b_i, i = 1,\ldots,n\}, \tag{17}$$

where $A_i \in \mathbb{H}$ and $b_i \in \mathbb{R}$ for $i = 1,\ldots,n$, and write $A := [A_1,\ldots,A_n]$ and $b := [b_1,\ldots,b_n]$. Inspired by the idea in [21, Section 5.6], we define a regularized Bregman kernel associated to the set C as

$$\psi(x) := \varphi(Ax - b) + \frac{\sigma}{2}\|x\|^2 \quad \text{for each } x \in \mathbb{H}, \tag{18}$$

where $\sigma > 0$. By assumptions (A_1) and (A_2) of φ and by (18), one can check that (see [21, Proposition 5.2]) $\psi : \mathbb{H} \to \mathbb{R} \cup \{+\infty\}$ satisfies the following conditions:

(B_1) ψ is proper, weakly lsc, and convex with $\operatorname{dom}\psi \subseteq \operatorname{cl}C$ and $\operatorname{dom}\nabla\psi = C$.
(B_2) ψ is strictly convex and continuous on $\operatorname{dom}\psi$, and continuously differentiable on C.

Using the regularized Bregman kernel ψ, we now define the Bregman distance $\mathcal{D}_\psi : \mathbb{H}\times\mathbb{H} \to \mathbb{R}_+ \cup \{+\infty\}$ via (15), with ψ in place of φ. By (18), (15) and (16), one can check that $\mathcal{D}_\psi \in \mathcal{D}(C,\mathbb{H})$ and $\mathcal{D}_\psi$ is self-proximal; consequently, $(\mathcal{D}_\psi, \mathcal{D}_\psi) \in \mathcal{F}_+(C,\mathbb{H})$.

Separable Bregman distances are the most commonly used in the literature. In detail, the Legendre function φ is written as the summation of one-dimensional functions

$$\varphi(x) := \sum_{i=1}^n \theta(x_i),$$

where $\theta : \mathbb{R} \to \mathbb{R}_+ \cup \{+\infty\}$ satisfies conditions (A_1) and (A_2) and is twice differentiable on $\mathbb{R}_{++}$. By the separable structure and (15), one has $\mathcal{D}_\varphi(x, y) = \sum_{i=1}^n \mathcal{D}_\theta(x_i, y_i)$; moreover, one can check by (18) that

$$\psi(x) = \sum_{i=1}^n \theta(\langle A_i, x\rangle - b_i) + \frac{\sigma}{2}\|x\|^2$$

and

$$\mathcal{D}_\psi(x, y) = \sum_{i=1}^n \mathcal{D}_\theta(\langle A_i, x\rangle - b_i, \langle A_i, y\rangle - b_i) + \frac{\sigma}{2}\|x - y\|^2.$$

One can refer to [5,13] for several popular examples of separable Bregman kernels and their associated Bregman distances; e.g., the Euclidean distance, the Kullback-Leibler divergence and the Itakura-Saito divergence are Bregman distances with the separable kernel $\theta(\cdot)$ given by the energy, the Boltzmann-Shannon entropy and the Burg entropy, respectively.

In the rest of this paper, we focus on the convergence properties and convergence rates of the interior quasi-subgradient method, i.e., Algorithm 1.

4 Convergence analysis

In this section we establish the convergence properties of the interior quasi-subgradient method for three types of stepsize strategies.

To this end, we first present some useful properties of the interior projection mapping in the following proposition, which extends [5, Proposition 4.1] to Hilbert spaces.

Proposition 4 *Let $x \in C \cap V$, $g \in \mathbb{H}$ and $v > 0$. Suppose that $d \in \mathcal{D}(C, V)$. Then the following assertions are true.*

(i) $\mathcal{P}(0, x) = x$.
(ii) $\sigma\|x - \mathcal{P}(vg, x)\|^2 \le v\langle g, x - \mathcal{P}(vg, x)\rangle$.
(iii) $\sigma\|x - \mathcal{P}(vg, x)\| \le v\|g\|$.

Proof From the strong convexity of $d(\cdot, y)$ we have that $\mathcal{P}(\cdot,\cdot)$ is a single-valued mapping (see Remark 1). Assertion (i) directly follows from (13) and Definition 4(i); assertion (iii) immediately follows from assertion (ii) and the Cauchy-Schwarz inequality. Hence, it only remains to show assertion (ii). Writing the optimality conditions for (13) with vg in place of g become

$$0 \in vg + \nabla_1 d(\mathcal{P}(vg, x), x) + \mathrm{N}_V(\mathcal{P}(vg, x)). \tag{19}$$

Recalling that V is a closed and convex set, we deduce by (5) that (19) is equivalent to

$$\langle vg + \nabla_1 d(\mathcal{P}(vg, x), x), z - \mathcal{P}(vg, x)\rangle \ge 0 \quad \text{for each } z \in V. \tag{20}$$

Then, by the strong convexity of $d(\cdot, y)$ (i.e., Definition 4(iii) with $\mathcal{P}(vg, x)$, x, x in place of x_1, x_2, y) and Definition 4(i), we obtain that

$$\sigma\|x - \mathcal{P}(vg, x)\|^2 \le \langle \nabla_1 d(\mathcal{P}(vg, x), x), \mathcal{P}(vg, x) - x\rangle \le v\langle g, x - \mathcal{P}(vg, x)\rangle,$$

where we also used (20) for $z := x$. The above expression is (ii). The proof is complete.

The Hölder condition of order p was used in [39] to describe some basic properties of quasi-subgradient, and plays an important role in the convergence study of quasi-subgradient methods for quasi-convex optimization problems; see [33–35, 40].

Definition 6 Let $0 < p \leq 1$ and $L > 0$. The function $f : \mathbb{H} \to \mathbb{R}$ is said to satisfy the Hölder condition (restricted to the set of minima X^*) of order p with modulus L on $\mathbb{H}$ if

$$f(x) - f^* \leq L \operatorname{dist}^p(x, X^*) \quad \text{for each } x \in \mathbb{H}.$$

The Hölder condition of order p reduces to the Lipschitz condition (restricted to the set of minima X^*) when $p = 1$, and this property holds for very broad classes of functions with various values of $p > 0$. The following lemma extends [62, Lemma 2.5] to Hilbert spaces and describes an important property of a quasi-convex function that satisfies the Hölder condition of order p. The line of analysis is similar to that of [62, Lemma 2.5], and thus is omitted.

Lemma 6 *Let $f : \mathbb{H} \to \mathbb{R}$ be a quasi-convex and continuous function and $x^* \in X^*$. Let $0 < p \leq 1$ and $L > 0$, and suppose that f satisfies the Hölder condition of order p with modulus L on $\mathbb{H}$. Then, for each $x \in X \setminus X^*$, it holds that*

$$f(x) - f^* \leq L \langle g, x - x^* \rangle^p \quad \textit{for each } g \in \partial^{\mathrm{Q}} f(x) \cap \mathbb{S}.$$

This property is a key to establish a basic inequality in the convergence analysis of quasi-subgradient methods; see, e.g., [34, 35, 62]. Hence, throughout this section, we make the following assumption to investigate the convergence properties of the interior quasi-subgradient method.

Assumption 1 *$f : \mathbb{H} \to \mathbb{R}$ is quasi-convex and continuous, and satisfies the Hölder condition of order $p \in (0, 1]$ with modulus L on $\mathbb{H}$.*

Under this assumption, the following lemma provides the basic inequality of the interior quasi-subgradient method, which is the foundation of its convergence analysis.

Lemma 7 *Let $(d, \mathcal{H}) \in \mathcal{F}_+(C, V)$ and let $\{x_k\}$ be a sequence generated by Algorithm 1. Then the following assertions are true.*

(i) *Let $z \in \mathrm{cl}C$, then we have*

$$\mathcal{H}(z, x_{k+1}) - \mathcal{H}(z, x_k) \leq v_k \langle g_k, z - x_{k+1} \rangle.$$

(ii) *Suppose that Assumption 1 is satisfied and fix $k \in \mathbb{N}$. If $x_k \notin X^*$, then it holds that*

$$\mathcal{H}(x^*, x_{k+1}) \leq \mathcal{H}(x^*, x_k) - v_k \left(\frac{f(x_k) - f^*}{L} \right)^{\frac{1}{p}} + \frac{v_k^2}{\sigma} \quad \textit{for each } x^* \in X^*. \tag{21}$$

Moreover, if $x_k \notin X^*$ for $k = 1, \ldots, n$, then

$$\mathcal{H}(x^*, x_{n+1}) \leq \mathcal{H}(x^*, x_1) - L^{-\frac{1}{p}} \sum_{k=1}^{n} v_k (f(x_k) - f^*)^{\frac{1}{p}} + \sum_{k=1}^{n} \frac{v_k^2}{\sigma} \quad \textit{for each } x^* \in X^*. \tag{22}$$

Proof (i) Fix $z \in \mathrm{cl}C$. From Definition 5(b) we obtain

$$\begin{aligned}\mathcal{H}(z, x_{k+1}) - \mathcal{H}(z, x_k) &\leq -\langle z - x_{k+1}, \nabla_1 d(x_{k+1}, x_k)\rangle \\ &\leq v_k \langle g_k, z - x_{k+1}\rangle,\end{aligned}$$

where (20) and (14) are used in the last inequality.
(ii) From part (i) for $z = x^*$ we deduce

$$\mathcal{H}(x^*, x_{k+1}) - \mathcal{H}(x^*, x_k) \leq v_k \langle g_k, x_k - x_{k+1}\rangle - v_k \langle g_k, x_k - x^*\rangle. \tag{23}$$

In view of Algorithm 1 and Assumption 1, we obtain by Lemma 6 and Proposition 4(iii) the following inequalities

$$\langle g_k, x_k - x^*\rangle \geq \left(\frac{f(x_k) - f^*}{L}\right)^{\frac{1}{p}} \quad \text{and} \quad \langle g_k, x_k - x_{k+1}\rangle \leq \|x_k - x_{k+1}\| \leq \frac{v_k}{\sigma},$$

respectively. Combining these inequalities with (23), (21) follows. Suppose further that $x_k \notin X^*$ for $k = 1, \ldots, n$. Then (21) holds for $k = 1, \ldots, n$; adding these yields (22). The proof is complete.

To end this subsection, we provide a framework for establishing the weak convergence of the interior quasi-subgradient method under the following additional assumption on the proximal distance. Assumption 2 reduces to the so-called *norm compatibility* introduced in [23, pp. 200 (B_6)] for the case in which $d(\cdot,\cdot)$ is a Bregman distance, and in [21, Definition 5.3] for infinite-dimensional case. The Boltzmann-Shannon entropy is an example satisfying the norm compatibility; see [21, Proposition 5.3]. In particular, consider the constraint set C defined in Example 1 (see (17)), and the separable Boltzmann-Shannon kennel given by $\theta(t) := t\log(t)$. Define $\eta : \mathbb{H} \to \mathbb{R}^n$ with $\eta_i(z) := \langle A_i, z\rangle - b_i$ for any $z \in \mathbb{H}$ and $i = 1, \ldots, n$; so the vector $\eta(z) \in \mathbb{R}^n_{++}$ for all $z \in C$. Hence the regularized Boltzmann-Shannon kennel is

$$\psi(x) := \sum_{i=1}^{n} \eta_i(x)\log(\eta_i(x)) + \frac{\sigma}{2}\|x\|^2,$$

and the regularized Boltzmann-Shannon entropy is

$$\mathcal{D}_\psi(x, y) := d_{\mathrm{KL}}(\eta(x), \eta(y)) + \frac{\sigma}{2}\|x - y\|^2,$$

where $d_{\mathrm{KL}}(\cdot,\cdot)$ is the Kullback-Liebler divergence defined by

$$d_{\mathrm{KL}}(\mu, \nu) := \sum_{i=1}^{n} \mu_i \log\frac{\mu_i}{\nu_i} - \mu_i + \nu_i \quad \text{for each } \mu \in \mathbb{R}^n_+, \nu \in \mathbb{R}^n_{++}.$$

Lemma 8 will be useful in Theorems 2 and 3 for establishing the weak convergence of the interior quasi-subgradient method with the diminishing or dynamic stepsize rules.

Assumption 2 *Let $\{x_k\} \subseteq \mathbb{H}$, and let $\bar{x}_1$ and $\bar{x}_2$ be the arbitrary weak cluster points of $\{x_k\}$. If $\lim_{k\to\infty}\langle \nabla_1 d(x_{k+1}, x_k), \bar{x}_1 - \bar{x}_2\rangle = 0$, then we must have $\bar{x}_1 = \bar{x}_2$.*

Lemma 8 *Let $(d, \mathcal{H}) \in \mathcal{F}_+(C, V)$ and let $\{x_k\}$ be a sequence generated by Algorithm 1. Suppose that the following assumptions are satisfied:*

(a) $\lim_{k\to\infty} v_k = 0$, $\{x_k\}$ *is bounded*, $\{\mathcal{H}(x^*, x_k)\}$ *is convergent for each* $x^* \in X^*$.
(b) *Each weak cluster point is an optimal solution of* (4).
(c) *Assumption 2 is satisfied.*

Then $\{x_k\}$ *weakly converges to an optimal solution of* (4).

Proof Let $\bar{x}_1$ and $\bar{x}_2$ be two weak cluster points of $\{x_k\}$. Then one has by assumption (b) that $\bar{x}_1, \bar{x}_2 \in X^*$. By Definition 5(b) (with x_k, x_{k+1}, $\bar{x}_1$ in place of x, y, z), one has

$$\langle \nabla_1 d(x_{k+1}, x_k), \bar{x}_1 - x_{k+1} \rangle \leq \mathcal{H}(\bar{x}_1, x_k) - \mathcal{H}(\bar{x}_1, x_{k+1}).$$

Moreover, (20) (with x_k, x_{k+1}, $\bar{x}_2$, g_k in place of x, $\mathcal{P}(vg, x)$, z, g) is reduced to

$$\langle \nabla_1 d(x_{k+1}, x_k), x_{k+1} - \bar{x}_2 \rangle \leq \langle v_k g_k, \bar{x}_2 - x_{k+1} \rangle.$$

Combining the above two inequalities, we obtain that

$$\langle \nabla_1 d(x_{k+1}, x_k), \bar{x}_1 - \bar{x}_2 \rangle \leq \mathcal{H}(\bar{x}_1, x_k) - \mathcal{H}(\bar{x}_1, x_{k+1}) + \langle v_k g_k, \bar{x}_2 - x_{k+1} \rangle. \tag{24}$$

One can see that $\lim_{k\to\infty} \mathcal{H}(\bar{x}_1, x_k) - \mathcal{H}(\bar{x}_1, x_{k+1}) = 0$ by the assumption in (a) that $\{\mathcal{H}(\bar{x}_1, x_k)\}$ is convergent. Since $\|g_k\| = 1$, we obtain by Cauchy-Schwartz that

$$\limsup_{k\to\infty} \langle v_k g_k, \bar{x}_2 - x_{k+1} \rangle \leq \limsup_{k\to\infty} v_k \|\bar{x}_2 - x_{k+1}\| = 0.$$

The above expression, combined with (24) gives

$$\limsup_{k\to\infty} \langle \nabla_1 d(x_{k+1}, x_k), \bar{x}_1 - \bar{x}_2 \rangle \leq 0.$$

By exchanging $\bar{x}_2$ with $\bar{x}_1$, we obtain that $\limsup_{k\to\infty} \langle \nabla_1 d(x_{k+1}, x_k), \bar{x}_2 - \bar{x}_1 \rangle \leq 0$. Consequently, we arrive at $\lim_{k\to\infty} \langle \nabla_1 d(x_{k+1}, x_k), \bar{x}_1 - \bar{x}_2 \rangle = 0$, and then by Assumption 2 $\bar{x}_1 = \bar{x}_2$. Therefore, $\{x_k\}$ is shown to weakly converge to an optimal solution of (4).

By virtue of the notion of induced proximal distance and Lemmas 7-8, we establish the convergence theorems of the interior quasi-subgradient method when using three typical stepsize rules: the constant, diminishing and dynamic stepsize rules (see [33, 42, 62]).

4.1 Global convergence and iteration complexity

In this subsection, we will investigate the convergence properties, including the global convergence and iteration complexity, under the assumption of the Hölder condition of order p, when using the constant/diminishing/dynamic stepsize rules respectively.

4.1.1 Constant stepsize

Theorem 1 *Let* $(d,\mathcal{H}) \in \mathcal{F}_+(C,V)$ *and let* $\{x_k\}$ *be a sequence generated by Algorithm 1. Suppose that Assumption 1 is satisfied and* $v_k \equiv v > 0$ *for any* $k \in \mathbb{N}$*. Then the following assertions are true.*

(i) $\min_{i=1,\dots,k} f(x_i) \le f^* + L\left(\frac{v}{\sigma} + \frac{1}{kv}\mathrm{dist}_{\mathcal{H}}(X^*,x_1)\right)^p$.

(ii) $\liminf_{k\to\infty} f(x_k) \le f^* + L\left(\frac{v}{\sigma}\right)^p$.

Proof (i) Without loss of generality, we assume that $x_i \notin X^*$ for $i = 1,\dots,k$; otherwise, assertion (i) of this theorem follows automatically. Let $x^* \in X^*$. Then, following Lemma 7(ii), we obtain by (22) that

$$0 \le \mathcal{H}(x^*,x_{k+1}) \le \mathcal{H}(x^*,x_1) - kv\left(\frac{\min_{i=1,\dots,k} f(x_i) - f^*}{L}\right)^{\frac{1}{p}} + k\frac{v^2}{\sigma}.$$

Since $x^* \in X^*$ is arbitrary, assertion (i) of this theorem now follows from the above expression by taking the infimum over X^* in the right-hand side and re-arranging the resulting expression.

(ii) It is enough to assume that $x_k \in X^*$ only occurs for finitely many times; otherwise, the left-hand side of assertion (ii) reduces to f^*, and hence it automatically holds. We can further assume that $x_k \notin X^*$ for each $k \in \mathbb{N}$. Indeed, since there exists $N \in \mathbb{N}$ such that $x_k \notin X^*$ for all $k \ge N$, we can restrict the analysis to the tail $\{x_k\}_{k\ge N}$ instead. Hence, we can assume that $x_k \notin X^*$ for each $k \in \mathbb{N}$, and we are in conditions of Lemma 7. Let $x^* \in X^*$. By Lemma 7(ii), (22) holds with v in place of v_k. Re-arranging (22) gives

$$\sum_{k=1}^n \frac{(f(x_k) - f^*)^{\frac{1}{p}}}{n} \le \frac{L^{\frac{1}{p}}}{nv}\mathcal{H}(x^*,x_1) + \frac{v}{\sigma}L^{\frac{1}{p}}.$$

Then, by applying Lemma 1 with $(f(x_k) - f^*)^{\frac{1}{p}}$ and 1 in place of a_k and v_k, we derive

$$\liminf_{n\to\infty}(f(x_n) - f^*)^{\frac{1}{p}} \le \liminf_{n\to\infty}\sum_{k=1}^n \frac{(f(x_k) - f^*)^{\frac{1}{p}}}{n} \le \liminf_{n\to\infty}\frac{L^{\frac{1}{p}}}{nv}\mathcal{H}(x^*,x_1) + \frac{v}{\sigma}L^{\frac{1}{p}} = \frac{v}{\sigma}L^{\frac{1}{p}}.$$

Consequently, assertion (ii) of this theorem follows, and the proof is complete.

4.1.2 Diminishing stepsize

Theorem 2 *Let* $(d,\mathcal{H}) \in \mathcal{F}_+(C,V)$ *and let* $\{x_k\}$ *be a sequence generated by Algorithm 1. Suppose that Assumption 1 is satisfied and* $\{v_k\}$ *satisfies*

$$v_k > 0, \quad \lim_{k\to\infty} v_k = 0 \quad \textit{and} \quad \sum_{k=1}^\infty v_k = \infty. \tag{25}$$

Then the following assertions are true.

(i) $\min_{i=1,\ldots,k} f(x_i) \le f^* + L \left(\frac{\mathrm{dist}_{\mathcal{H}}(X^*, x_1) + \frac{1}{\sigma} \sum_{i=1}^k v_i^2}{\sum_{i=1}^k v_i} \right)^p$.

(ii) $\liminf_{k\to\infty} f(x_k) = f^*$.

(iii) *If* $\sum_{k=0}^\infty v_k^2 < \infty$, *then* $\{x_k\}$ *weakly converges to an optimal solution of* (4) *provided one of the following assumptions:*

(a) $p = 1$, f *is weakly lsc on* $\mathrm{cl}C \cap V$, *and Assumption 2 is satisfied.*

(b) $\mathbb{H}$ *is finite-dimensional.*

Proof (i) As in the previous theorem, we may assume that $x_i \notin X^*$ for $i = 1, \ldots, k$; otherwise, assertion (i) follows automatically. Let $x^* \in X^*$. Then, following Lemma 7(ii), we obtain by (22) that

$$0 \le \mathcal{H}(x^*, x_{k+1}) \le \mathcal{H}(x^*, x_1) - \left(\frac{\min_{i=1,\ldots,k} f(x_i) - f^*}{L} \right)^{\frac{1}{p}} \sum_{i=1}^k v_i + \frac{1}{\sigma} \sum_{i=1}^k v_i^2$$

Since $x^* \in X^*$ is arbitrary, we can take infimum in the right-hand side of the expression above, and part (i) follows by re-arranging the resulting expression.

(ii) Let $x^* \in X^*$. As in the proof of Theorem 1(ii), we can assume, without loss of generality, that $x_k \notin X^*$ for each $k \in \mathbb{N}$, and then Lemma 7 follows. Re-arranging (22) gives

$$\frac{\sum_{k=1}^n v_k (f(x_k) - f^*)^{\frac{1}{p}}}{\sum_{k=1}^n v_k} \le \frac{L^{\frac{1}{p}}}{\sum_{k=1}^n v_k} \mathcal{H}(x^*, x_1) + \frac{1}{\sigma} \frac{\sum_{k=1}^n v_k^2}{\sum_{k=1}^n v_k}.$$

By our assumption, $\sum_{k=1}^\infty v_k = \infty$. Hence, we can apply the second statement of Lemma 1 with $a_k := v_k$ and $a := 0$, to conclude that $\lim_{n\to\infty} \frac{\sum_{k=1}^n v_k^2}{\sum_{k=1}^n v_k} = 0$ (as $\lim_{k\to\infty} v_k = 0$). By taking limits for $n \to \infty$ in the expression above, and recalling that $\sum_{k=1}^\infty v_k = \infty$, we deduce that the left-hand side tends to 0. Now we apply Lemma 1 again, with $a_k := (f(x_k) - f^*)^{\frac{1}{p}}$, v_k as in the Lemma, and use the fact that the left-hand side in the expression above tends to 0, to deduce part (ii).

(iii) By (21) and the assumption that $\sum_{k=0}^\infty v_k^2 < \infty$, one sees that $\{\mathcal{H}(x^*, x_k)\}$ is a quasi-Fejér sequence; thus Lemma 4 is applicable to concluding that

$$\{\mathcal{H}(x^*, x_k)\} \text{ is convergent for each } x^* \in X^* \tag{26}$$

and

$$\sum_{k=1}^\infty v_k (f(x_k) - f^*)^{\frac{1}{p}} < \infty. \tag{27}$$

Then by Definition 5(iii), $\{x_k\}$ is bounded and must have weak cluster points. Below we show the convergence of $\{x_k\}$ in the following two situations.

(a) Note that (25), (26) and the boundedness of $\{x_k\}$ validate assumption (a) in Lemma 8. On the other hand, by the assumption that $p = 1$, we obtain by the Lipschitz continuity and Proposition 4(iii) (with x_k, x_{k+1}, g_k in place of x, $\mathcal{P}(vg, x)$, g) that

$$f(x_{k+1}) - f(x_k) \le L\|x_{k+1} - x_k\| \le \frac{L}{\sigma} v_k$$

(thanks to $\|g_k\| = 1$). This, together with (25) and (27), verifies the assumptions in Lemma 5 (with $f(x_k) - f^*$, v_k, $\frac{L}{\sigma}$ in place of u_k, β_k, a). Hence it follows from Lemma 5 that $\lim_{k\to\infty} f(x_k) = f^*$. Then, by the assumption that f is weakly lsc on $\mathrm{cl}C \cap V$, one has that each weak cluster point of $\{x_k\}$ is an optimal solution of (4); this validates assumption (b) in Lemma 8. These, together with Assumption 2 make Lemma 8 applicable to showing the weak convergence of $\{x_k\}$ to an optimal solution of (4).

(b) Suppose that $\mathbb{H}$ is finite-dimensional. Since $\{x_k\}$ is bounded, and by assertion (ii) and by the continuity of f, we have that $\{x_k\}$ has a cluster point $\bar{x} \in X^*$. Noting by assumption that $\lim_{n\to\infty} \sum_{k=n}^{\infty} v_k^2 = 0$, we derive by (21) that $\{\mathcal{H}(\bar{x}, x_k)\}$ is a Cauchy sequence, and thus it converges to 0 (by Definition 5(ii)). Therefore, by Definition 5(i), we conclude that $\{x_k\}$ converges to this $\bar{x}$. The proof is complete.

4.1.3 Dynamic stepsize

Theorem 3 *Let $(d, \mathcal{H}) \in \mathcal{F}_+(C, V)$ and let $\{x_k\}$ be a sequence generated by Algorithm 1. Suppose that Assumption 1 is satisfied and $\{v_k\}$ is given by*

$$v_k = \frac{\sigma}{2}\lambda_k \left(\frac{f(x_k) - f^*}{L}\right)^{\frac{1}{p}} \quad \text{with } 0 < \underline{\lambda} \le \lambda_k \le \overline{\lambda} < 2. \tag{28}$$

Then the following assertions are true.

(i) $\min_{i=1,\dots,k} f(x_i) \le f^* + L\left(\frac{4}{k\sigma\underline{\lambda}(2-\overline{\lambda})}\mathrm{dist}_{\mathcal{H}}(X^*, x_1)\right)^{\frac{p}{2}}$.

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

(iii) $\{x_k\}$ *weakly converges to an optimal solution of* (4) *provided one of the following assumptions:*
  (a) f *is weakly lsc on* $\mathrm{cl}C \cap V$*, and Assumption 2 is satisfied.*
  (b) $\mathbb{H}$ *is finite-dimensional.*

Proof Without loss of generality, we assume that $x_k \notin X^*$ for each $k \in \mathbb{N}$; otherwise, one checks by (28) that $v_k = 0$ whenever $x_k \in X^*$, and so the generated sequence stays at this optimal solution, that is, assertions of this theorem follow automatically.

(i) Let $x^* \in X^*$. Following Lemma 7(ii), we obtain by (21) and (28) that

$$\begin{aligned}\mathcal{H}(x^*, x_{k+1}) &\le \mathcal{H}(x^*, x_k) - \tfrac{\sigma}{4}\lambda_k(2-\lambda_k)\left(\tfrac{f(x_k)-f^*}{L}\right)^{\frac{2}{p}}\\ &\le \mathcal{H}(x^*, x_k) - \tfrac{\sigma}{4}\underline{\lambda}(2-\overline{\lambda})\left(\tfrac{f(x_k)-f^*}{L}\right)^{\frac{2}{p}}.\end{aligned} \tag{29}$$

Noting that $f(x_k) \ge f^*$ (as $x_k \in X$) for each $k \in \mathbb{N}$, we have inductively that

$$\begin{aligned}0 \le \mathcal{H}(x^*, x_{k+1}) &\le \mathcal{H}(x^*, x_1) - \textstyle\sum_{i=1}^{k} \tfrac{\sigma}{4}\underline{\lambda}(2-\overline{\lambda})\left(\tfrac{f(x_i)-f^*}{L}\right)^{\frac{2}{p}}\\ &\le \mathcal{H}(x^*, x_1) - k\tfrac{\sigma}{4}\underline{\lambda}(2-\overline{\lambda})\left(\tfrac{\min_{i=1,\dots,k} f(x_i)-f^*}{L}\right)^{\frac{2}{p}}.\end{aligned}$$

Assertion (i) of this theorem follows as $x^* \in X^*$ is arbitrary.

(ii) Note that $f(x_k) \geq f^*$ (as $x_k \in X$) for each $k \in \mathbb{N}$ and by (29) that $\{\mathcal{H}(x^*, x_k)\}$ is decreasing and thus convergent (as $\mathcal{H}(\cdot,\cdot)$ is nonnegative) for each $x^* \in X^*$. Hence, we obtain by (29) that $\lim_{k\to\infty} f(x_k) = f^*$.

(iii) Recalling that $\{\mathcal{H}(x^*, x_k)\}$ is decreasing, one sees by Definition 5(iii) that $\{x_k\}$ is bounded and must have weak cluster points. Below we show the convergence of $\{x_k\}$ under the assumptions indicated either by (a) or (b).

(a) Note by (28) and assertion (ii) of this theorem that $\lim_{k\to\infty} v_k = 0$. This, together with the bounded property of $\{x_k\}$ and the convergence property of $\{\mathcal{H}(x^*, x_k)\}$ mentioned above, confirms assumption (a) in Lemma 8. The weakly lsc assumption of f and assertion (ii) of this theorem validate assumption (b) in Lemma 8. These, together with Assumption 2, make Lemma 8 applicable to showing the weak convergence of $\{x_k\}$ to an optimal solution of (4).

(b) Suppose that $\mathbb{H}$ is finite-dimensional. Assertion (ii) of this theorem and the continuity of f show that each cluster point of $\{x_k\}$ is an optimal solution of (4), i.e., there exists a subsequence $\{x_{k_i}\}$ such that $\lim_{i\to\infty} x_{k_i} = \bar{x} \in X^*$. Then one has by Definition 5(ii) that $\lim_{i\to\infty} \mathcal{H}(\bar{x}, x_{k_i}) = 0$, and then, by the decreasing property of $\{\mathcal{H}(\bar{x}, x_k)\}$ that $\lim_{k\to\infty} \mathcal{H}(\bar{x}, x_k) = 0$. Hence, Definition 5(i) reveals that $\{x_k\}$ converges to this $\bar{x}$. The proof is complete.

Remark 2 (a) Assertions (i) of the above theorems show the iteration complexity of the interior quasi-subgradient method when using different types of stepsizes. In particular, Theorems 1(i) and 3(i) show that the interior quasi-subgradient method possesses the computational complexity of $\mathcal{O}(1/k^p)$ and $\mathcal{O}(1/k^{\frac{p}{2}})$ to approach (a certain region of) the optimal value when the constant and dynamic stepsize rules are used, respectively. When the diminishing stepsize is given by

$$v_k := ck^{-s} \quad \text{with } c > 0 \text{ and } s \in (0, 1), \tag{30}$$

Theorem 2(i) exhibits the computational complexity of $\mathcal{O}(1/k^{p\min\{s,1-s\}})$ for the interior quasi-subgradient method, and particularly, the optimal complexity is gained when $s = \frac{1}{2}$.

(b) Theorem 1(ii) presents the convergence of the function values to the optimal value within a certain region, proportional to v^p, for the constant stepsize rule, and assertions (ii) of Theorems 2 and 3 reveal the convergence to the optimal value when using the diminishing and dynamic stepsize rules, respectively.

(c) Assertions (iii) of Theorems 2 and 3 demonstrate the weak convergence of the iterates to the optimal solution when using the diminishing and dynamic stepsize rules. To the best of our knowledge, this paper seems the first work discussing the weak convergence of subgradient method with non-Euclidean distance in Hilbert spaces.

(d) When $d(x, y) = \mathcal{H}(x, y) = \frac{1}{2}\|x-y\|^2$, the resulting interior quasi-subgradient method is reduced to the (Euclidean) projected quasi-subgradient method [37], and Theorems 1-3 cover the global convergence results for the (Euclidean) projected quasi-subgradient method in Euclidean spaces in [33, Theorems 4.1 and 4.2].

4.2 Convergence rates

The establishment of convergence rates is important in analyzing the numerical performance of relevant algorithms. The concept of weak sharp minimum is a common assumption for analyzing the convergence rates of many optimization algorithms; see [18,32,33,63] and references therein. Particularly, the convergence rates of (Euclidean) projected quasi-subgradient methods in Euclidean spaces were established in [33] under the assumption of weak sharp minimum of Hölderian order. To establish the convergence rate for the interior quasi-subgradient method in the sequel, we introduce a notion of weak sharp minimum property relative to the induced proximal distance.

Definition 7 Let $\eta > 0$ and $q > 0$. f is said to satisfy the weak sharp minimum property of order q relative to the induced proximal distant $\mathcal{H}$ on X with modulus η if

$$f(x) - f^* \geq \eta \, \mathrm{dist}^q_{\mathcal{H}}(X^*, x) \quad \text{for each } x \in X. \tag{31}$$

In the case when the proximal distance is the Euclidean distance (see Example 1(i)) or based on the second order homogeneous kernels [7], the induced proximal distance $\mathcal{H}$ is proportional to the Euclidean distance, and thus, the weak sharp minimum property of order q relative to $\mathcal{H}$ is reduced to the classical weak sharp minimum property of order $2q$ with a different modulus. In the rest of this paper, we make the following assumption to establish the convergence rates for the interior quasi-subgradient method.

Assumption 3 *Let $\eta > 0$ and $q > 0$, and suppose that f satisfies the weak sharp minimum property of order q relative to the induced proximal distant $\mathcal{H}$ on X with modulus η.*

4.2.1 Constant stepsize

Theorem 4 *Let $(d, \mathcal{H}) \in \mathcal{F}_+(C, V)$ and let $\{x_k\}$ be a sequence generated by Algorithm 1. Suppose that Assumptions 1 and 3 are satisfied and $v_k \equiv v > 0$ for each $k \in \mathbb{N}$. Then the following assertions are true.*

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

$$\mathrm{dist}_{\mathcal{H}}(X^*, x_{k+1}) \leq \tau^k \mathrm{dist}_{\mathcal{H}}(X^*, x_1) + \max\left\{ \left(\frac{L}{\eta}\right)^{\frac{1}{p}}, v \right\} \frac{v}{\sigma} \quad \text{for each } k \in \mathbb{N}.$$

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

$$\mathrm{dist}_{\mathcal{H}}(X^*, x_{k+1}) \leq \tau^k \mathrm{dist}_{\mathcal{H}}(X^*, x_1) + \left(\frac{L}{\eta}\right)^{\frac{1}{q}} \left(\frac{v}{\sigma}\right)^{\frac{p}{q}} \quad \text{for each } k \in \mathbb{N}.$$

Proof Without loss of generality, we assume that $x_k \notin X^*$ for each $k \in \mathbb{N}$; otherwise, this theorem holds automatically. By Lemma 7(ii) and Assumption 3, we obtain by (21) and (31) that, for each $k \in \mathbb{N}$

$$\mathrm{dist}_{\mathcal{H}}(X^*, x_{k+1}) \leq \mathrm{dist}_{\mathcal{H}}(X^*, x_k) - v \left(\frac{\eta}{L}\right)^{\frac{1}{p}} \mathrm{dist}^{\frac{q}{p}}_{\mathcal{H}}(X^*, x_k) + \frac{v^2}{\sigma}. \tag{32}$$

Below, we prove this theorem in the following two cases.

(i) Suppose that $q = p$. Setting $\tau := (1 - v\left(\frac{\eta}{L}\right)^{\frac{1}{p}})_+ \in [0, 1)$, we achieve by (32) that

$$\mathrm{dist}_{\mathcal{H}}(X^*, x_{k+1}) \le \tau \mathrm{dist}_{\mathcal{H}}(X^*, x_k) + \frac{v^2}{\sigma} \quad \text{for each } k \in \mathbb{N}.$$

Then we inductively obtain that, for each $k \in \mathbb{N}$

$$\begin{aligned}\mathrm{dist}_{\mathcal{H}}(X^*, x_{k+1}) &\le \tau^k \mathrm{dist}_{\mathcal{H}}(X^*, x_1) + \frac{v^2}{(1-\tau)\sigma} \\ &= \tau^k \mathrm{dist}_{\mathcal{H}}(X^*, x_1) + \max\left\{\left(\frac{L}{\eta}\right)^{\frac{1}{p}}, v\right\}\frac{v}{\sigma}.\end{aligned}$$

(ii) Suppose that $q > p$. By the assumptions and (32), Lemma 2(ii) is applicable (with $\mathrm{dist}_{\mathcal{H}}(X^*, x_k)$, $\frac{q}{p} - 1$, $v\left(\frac{\eta}{L}\right)^{\frac{1}{p}}$ and $\frac{v^2}{\sigma}$ in place of u_k, r, a and b) to concluding that there exists $\tau \in (0, 1)$ such that the conclusion follows. The proof is complete.

4.2.2 Diminishing stepsize

Theorem 5 *Let $(d, \mathcal{H}) \in \mathcal{F}_+(C, V)$ and let $\{x_k\}$ be a sequence generated by Algorithm 1. Suppose that Assumptions 1 and 3 are satisfied with $q = p$ and v_k is given by (30). Then, either $x_k \in X^*$ for some $k \in \mathbb{N}$ or there exists $N \in \mathbb{N}$ such that*

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

Proof Without loss of generality, we assume that $x_k \notin X^*$ for each $k \in \mathbb{N}$; otherwise, this theorem holds automatically. Then, we obtain by (21) and (31) that

$$\mathrm{dist}_{\mathcal{H}}(X^*, x_{k+1}) \le \left(1 - ck^{-s}\left(\frac{\eta}{L}\right)^{\frac{1}{p}}\right)\mathrm{dist}_{\mathcal{H}}(X^*, x_k) + \frac{c^2}{\sigma}k^{-2s} \quad \text{for each } k \in \mathbb{N}.$$

Lemma 3(i) is applicable (with $c\left(\frac{\eta}{L}\right)^{\frac{1}{p}}$, $\frac{c^2}{\sigma}$ and $2s$ in place of a, b and t) to obtaining the conclusion. The proof is complete.

4.2.3 Dynamic stepsize

Theorem 6 *Let $(d, \mathcal{H}) \in \mathcal{F}_+(C, V)$ and let $\{x_k\}$ be a sequence generated by Algorithm 1. Suppose that Assumptions 1 and 3 are satisfied and v_k is given by (28). Then the following assertions are true.*

(i) *If $2q = p$, then there exists $\tau \in [0, 1)$ such that*

$$\mathrm{dist}_{\mathcal{H}}(X^*, x_{k+1}) \le \tau^k \mathrm{dist}_{\mathcal{H}}(X^*, x_1) \quad \textit{for each } k \in \mathbb{N}.$$

(ii) *If $2q > p$, then there exists $\gamma > 0$ such that*

$$\mathrm{dist}_{\mathcal{H}}(X^*, x_{k+1}) \le \frac{\mathrm{dist}_{\mathcal{H}}(X^*, x_1)}{(1+\gamma k)^{\frac{p}{2q-p}}} \quad \textit{for each } k \in \mathbb{N}. \tag{33}$$

Proof Without loss of generality, we assume that $x_k \notin X^*$ for each $k \in \mathbb{N}$; otherwise, one checks by (28) that $v_k = 0$ whenever $x_k \in X^*$, and so the generated sequence stays at this optimal solution; consequently, assertions of this theorem follow automatically. Then we obtain by (29) and (31) that

$$\mathrm{dist}_{\mathcal{H}}(X^*, x_{k+1}) \le \mathrm{dist}_{\mathcal{H}}(X^*, x_k) - \frac{\sigma}{4}\underline{\lambda}(2-\overline{\lambda})\left(\frac{\eta}{L}\right)^{\frac{2}{p}} \mathrm{dist}_{\mathcal{H}}^{\frac{2q}{p}}(X^*, x_k) \quad \text{for each } k \in \mathbb{N}. \tag{34}$$

(i) Suppose that $2q = p$. Setting $\tau := (1 - \frac{\sigma}{4}\underline{\lambda}(2-\overline{\lambda})\left(\frac{\eta}{L}\right)^{\frac{2}{p}})_+ \in [0, 1)$, we obtain by (34) that

$$\mathrm{dist}_{\mathcal{H}}(X^*, x_{k+1}) \le \tau \mathrm{dist}_{\mathcal{H}}(X^*, x_k) \le \tau^k \mathrm{dist}_{\mathcal{H}}(X^*, x_1) \quad \text{for each } k \in \mathbb{N}.$$

(ii) Suppose that $2q > p$. Then, by (34), Lemma 2(i) is applicable (with $\mathrm{dist}_{\mathcal{H}}(X^*, x_k)$, $\frac{\sigma}{4}\underline{\lambda}(2-\overline{\lambda})(\frac{\eta}{L})^{\frac{2}{p}}$, $\frac{2q}{p}-1$ in place of u_k, a, r) to concluding that (33) holds with $\gamma := \frac{(2q-p)\sigma}{4p}\underline{\lambda}(2-\overline{\lambda})\left(\frac{\eta}{L}\right)^{\frac{2}{p}} \mathrm{dist}_{\mathcal{H}}^{\frac{2q-p}{p}}(X^*, x_1)$.

Remark 3 (a) Theorem 6 (resp., 4) shows the linear (or sublinear) convergence rate (measured by the induced proximal distance) of the interior quasi-subgradient method with the dynamic (resp., constant) stepsize to the solution set (resp., a certain region of the solution set) under the assumption of weak sharp minimum of Hölderian order relative to the induced proximal distance.

(b) When using the diminishing stepsize rule (30) and under the weak sharp minimum property of order p relative to the induced proximal distance, Theorem 5 shows the convergence of the interior quasi-subgradient method to an optimal solution at a sublinear rate $\mathcal{O}(k^{-s})$. This is faster than the one established in Theorem 2 without the assumption of the weak sharp minimum property, namely $\mathcal{O}(1/k^{p\min\{s,1-s\}})$ in Remark 2(a), because of the assumed $p \le 1$.

(c) To the best of our knowledge, the established convergence rates of subgradient method with non-Euclidean distance in Hilbert spaces are new in the literature. When $d(x,y) = \mathcal{H}(x,y) = \frac{1}{2}\|x-y\|^2$, the resulting interior quasi-subgradient method is reduced to the (Euclidean) projected quasi-subgradient method [37], and Theorems 4-6 cover the convergence rate results for the (Euclidean) projected quasi-subgradient method in Euclidean spaces in [33, Theorem 4.3].

5 Conclusion

In this paper, we proposed an interior quasi-subgradient method based on non-Euclidean distances to solve constrained and nondifferentiable quasi-convex optimization problems. In the proposed algorithm, a proximal distance was adopted in the projection-like mapping. The types of proximal distances include Bregman distances, φ-divergences, and second order homogeneous kernels. Moreover, a new approach for calculating a quasi-subgradient was provided via the weak limit of Gâteaux derivatives.

Our convergence analysis of the interior quasi-subgradient method was developed for three prototypical types of stepsizes: constant, diminishing, and dynamic. Convergence results of objective values, including the global convergence and iteration complexity, were established under the assumption of the Hölder condition of order p. Weak convergence results of iterates were obtained under the assumptions of the Lipschitz condition and the norm compatibility/strictly convexity of Bregman kernel. Moreover, convergence rates were estimated by assuming a Hölder-type weak sharp minimum condition relative to an induced proximal distance.

It is an open question whether the weak convergence can be obtained under the Hölder condition with general $p < 1$.

Acknowledgements The authors are grateful to the editor and the anonymous reviewers for their valuable comments and suggestions toward the improvement of this paper. Regina S. Burachik's and Xiaoqi Yang's work was supported in part by the Research Grants Council of Hong Kong (PolyU 152342/16E). Yaohua Hu's work was supported in part by the National Natural Science Foundation of China (12071306, 11871347, 32170655), Natural Science Foundation of Guangdong Province of China (2019A1515011917, 2020B1515310008), Natural Science Foundation of Shenzhen (JCYJ20190808173603590).

References

1. Y. I. Alber, A. N. Iusem, and M. V. Solodov. On the projected subgradient method for nonsmooth convex optimization in a Hilbert space. *Mathematical Programming*, 81:23–35, 1998.
2. A. Agrawal and S. Boyd. Disciplined quasiconvex programming. *Optimization Letters*, 14:1643–1657, 2020.
3. A. Auslender and M. Teboulle. Interior gradient and ϵ-subgradient descent methods for constrained convex minimization. *Mathematics of Operations Research*, 29(1):1–26, 2004.
4. A. Auslender and M. Teboulle. Interior projection-like methods for monotone variational inequalities. *Mathematical Programming*, 104:39–68, 2005.
5. A. Auslender and M. Teboulle. Interior gradient and proximal methods for convex and conic optimization. *SIAM Journal on Optimization*, 16(3):697–725, 2006.
6. A. Auslender and M. Teboulle. Projected subgradient methods with non-Euclidean distances for non-differentiable convex minimization and variational inequalities. *Mathematical Programming*, 120:27–48, 2009.
7. A. Auslender, M. Teboulle, and S. Ben-Tiba. Interior proximal and multiplier methods based on second order homogeneous kernels. *Mathematics of Operations Research*, 24(3):645–668, 1999.
8. A. Auslender, M. Teboulle, and S. Ben-Tiba. A logarithmic-quadratic proximal method for variational inequalities. *Computational Optimization and Applications*, 12:31–40, 1999.
9. D. Aussel and N. Hadjisavvas. Adjusted sublevel sets, normal operator, and quasi-convex programming. *SIAM Journal on Optimization*, 16(2):358–367, 2005.
10. D. Aussel and M. Pištěk. Limiting normal operator in quasiconvex analysis. *Set-Valued and Variational Analysis*, 23(4):669–685, 2015.
11. M. Avriel, W. E. Diewert, S. Schaible, and I. Zang. *Generalized Concavity*. Plenum Press, New York, 1988.
12. K. Barty, J.-S. Roy, and C. Strugarek. Hilbert-valued perturbed subgradient algorithms. *Mathematics of Operations Research*, 32(3):551–562, 2007.
13. H. H. Bauschke, J. Bolte, and M. Teboulle. A descent lemma beyond Lipschitz gradient continuity: First-order methods revisited and applications. *Mathematics of Operations Research*, 42(2):330–348, 2017.
14. H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, 2011.
15. A. Beck and M. Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.
16. A. Ben-tal, T. Margalit, and A. Nemirovski. The ordered subsets mirror descent optimization method with applications to tomography. *SIAM Journal on Optimization*, 12:79–108, 2001.
17. D. P. Bertsekas. *Convex Optimization ang Algorithms*. Athena Scientific, Massachusetts, 2015.
18. J. Bolte, T. P. Nguyen, J. Peypouquet, and B. W. Suter. From error bounds to the complexity of first-order descent methods for convex functions. *Mathematical Programming*, 165(2):471–507, 2017.

19. L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967.
20. A. S. Brito, J. X. da Cruz Neto, J. O. Lopes, and P. R. Oliveira. Interior proximal algorithm for quasiconvex programming problems and variational inequalities with linear constraints. *Journal of Optimization Theory and Applications*, 154(1):217–234, 2012.
21. R. S. Burachik. *Generalized Proximal Point Method for the Variational Inequality Problem*. PhD Thesis, 1995.
22. R. S. Burachik and J. Dutta. Inexact proximal point methods for variational inequality problems. *SIAM Journal on Optimization*, 20(5):2653–2678, 2010.
23. R. S. Burachik and A. N. Iusem. A generalized proximal point algorithm for the variational inequality problem in a Hilbert space. *SIAM Journal on Optimization*, 8(1):197–216, 1998.
24. R. S. Burachik and S. Scheimberg. A proximal point method for the variational inequality problem in Banach spaces. *SIAM Journal on Control and Optimization*, 39(5):1633–1649, 2001.
25. Y. Censor and A. Segal. Algorithms for the quasiconvex feasibility problem. *Journal of Computational and Applied Mathematics*, 185(1):34–50, 2006.
26. G. Chen and M. Teboulle. Convergence analysis of a proximal-like minimization algorithm using bregman functions. *SIAM Journal on Optimization*, 3(3):538–543, 1993.
27. J.-P. Crouzeix, J.-E. Martinez-Legaz, and M. Volle. *Generalized Convexity, Generalized Monotonicity*. Kluwer Academic Publishers, Dordrecht, 1998.
28. F. G. M. Cunha, J. X. da Cruz Neto, and P. R. Oliveira. A proximal point algorithm with a ϕ-divergence for quasiconvex programming. *Optimization*, 59(5-6):777–792, 2010.
29. Y. M. Ermoliev. Methods of solution of nonlinear extremal problems. *Cybernetics and Systems Analysis*, 2:1–14, 1966.
30. H. J. Greenberg and W. P. Pierskalla. Quasiconjugate functions and surrogate duality. *Cahiers Centre Études Recherche Opertionnelle*, 15:437–448, 1973.
31. N. Hadjisavvas, S. Komlósi, and S. Schaible. *Handbook of Generalized Convexity and Generalized Monotonicity*. Springer-Verlag, New York, 2005.
32. Y. Hu, C. Li, and X. Yang. On convergence rates of linearized proximal algorithms for convex composite optimization with applications. *SIAM Journal on Optimization*, 26(2):1207–1235, 2016.
33. Y. Hu, J. Li, and C. K. W. Yu. Convergence rates of subgradient methods for quasi-convex optimization problems. *Computational Optimization and Applications*, 77(1):183-212, 2020.
34. Y. Hu, X. Yang, and C.-K. Sim. Inexact subgradient methods for quasi-convex optimization problems. *European Journal of Operational Research*, 240(2):315–327, 2015.
35. Y. Hu, C. K. W. Yu, and X. Yang. Incremental quasi-subgradient methods for minimizing the sum of quasi-convex functions. *Journal of Global Optimization*, 75(4):1003–1028, 2019.
36. Q. Ke and T. Kanade. Quasiconvex optimization for robust geometric reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(10):1834–1847, 2007.
37. K. C. Kiwiel. Convergence and efficiency of subgradient methods for quasiconvex minimization. *Mathematical Programming*, 90:1–25, 2001.
38. K. C. Kiwiel. Convergence of approximate and incremental subgradient methods for convex optimization. *SIAM Journal on Optimization*, 14(3):807–840, 2004.
39. I. V. Konnov. On properties of supporting and quasi-supporting vectors. *Journal of Mathematical Sciences*, 71:2760–2763, 1994.
40. I. V. Konnov. On convergence properties of a subgradient method. *Optimization Methods and Software*, 18(1):53–62, 2003.
41. J. López and E. A. Papa Quiroz. Construction of proximal distances over symmetric cones. *Optimization*, 66(8):1301–1321, 2017.
42. A. Nedić and D. P. Bertsekas. Incremental subgradient methods for nondifferentiable optimization. *SIAM Journal on Optimization*, 12(1):109–138, 2001.
43. A. S. Nemirovsky and D. B. Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley, New York, 1983.
44. S. Pan and J.-S. Chen. Entropy-like proximal algorithms based on a second-order homogeneous distance function for quasi-convex programming. *Journal of Global Optimization*, 39:555–575, 2007.
45. E. A. Papa Quiroz, J. López D. Borda, and F. Collantes. A proximal method for multiobjective quasiconvex minimization on the nonnegative orthant and its application to demand theory in microeconomy. *In Proceedings of 3rd International Conference on Robot Systems and Applications, ACM International Conference Proceeding Series*, 75–81, 2020.

46. E. A. Papa Quiroz, J. López Luis, and M. A. Cano Lengua. A proximal multiplier method for convex separable symmetric cone optimization. *In Proceeding of 5th International Conference on Multimedia Systems and Signal Processing, ACM International Conference Proceeding Series*, 92–97, 2020.
47. E. A. Papa Quiroz and P. R. Oliveira. An extension of proximal methods for quasiconvex minimization on the nonnegative orthant. *European Journal of Operational Research*, 216(1):26–32, 2012.
48. E. A. Papa Quiroz, L. M. Ramirez, and P. R. Oliveira. An inexact algorithm with proximal distances for variational inequalities. *RAIRO Operations Research*, 52(1):159–176, 2018.
49. E. A. Papa Quiroz, L. M. Ramirez, and P. R. Oliveira. An inexact proximal method for quasiconvex minimization. *European Journal of Operational Research*, 246(3):721–729, 2015.
50. E. A. Papa Quiroz, O. Sarmiento, and P. R. Oliveira. An inexact proximal decomposition method for variational inequalities with separable structure. *RAIRO Operations Research*, 55:S873–S884, 2021.
51. J.-P. Penot. Are generalized derivatives useful for generalized convex functions? In M. V. J.-P. Crouzeix, J.E. Martinez-Legaz, editor, *Generalized Convexity, Generalized Monotonicity: Recent Results*, pages 3–59. Kluwer Academic Publishers, 1998.
52. B. T. Polyak. A general method for solving extremum problems. *Soviet Mathematics Doklady*, 8:593–597, 1967.
53. B. T. Polyak. *Introduction to Optimization*. Optimization Software, New York, 1987.
54. P. J. Rabier. Differentiability of quasiconvex functions on separable Banach spaces. *Israel Journal of Mathematics*, 207:11–51, 2015.
55. O. Sarmiento, E. A. Papa Quiroz, and P. R. Oliveira. A proximal multiplier method for separable convex minimization. *Optimization*, 65(2):501–537, 2016.
56. O. Sarmiento, E. A. Papa Quiroz, and P. R. Oliveira. Convergence rate of a proximal multiplier algorithm for separable convex minimization. *Optimization*, 66(2):251–270, 2017.
57. N. Z. Shor. *Minimization Methods for Non-differentiable Functions*. Springer-Verlag, New York, 1985.
58. M. V. Solodov and S. K. Zavriev. Error stability properties of generalized gradient-type algorithms. *Journal of Optimization Theory and Applications*, 98:663–680, 1998.
59. S. d. S. Souza, P. R. Oliveira, J. X. da Cruz Neto, and A. Soubeyran. A proximal method with separable Bregman distances for quasiconvex minimization over the nonnegative orthant. *European Journal of Operational Research*, 201(2):365–376, 2010.
60. I. M. Stancu-Minasian. *Fractional Programming*. Kluwer Academic Publishers, Dordrecht, 1997.
61. M. Teboulle. Convergence of proximal-like algorithms. *SIAM Journal on Optimization*, 7(4):1069–1083, 1997.
62. C. K. W. Yu, Y. Hu, X. Yang, and S. K. Choy. Abstract convergence theorem for quasi-convex optimization problems with applications. *Optimization*, 68(7):1289–1304, 2019.
63. Z. Zhou and A. M.-C. So. A unified approach to error bounds for structured convex optimization problems. *Mathematical Programming*, 165(2):689–728, 2017.
64. C. Zălinescu. *Convex Analysis in General Vector Spaces*. World Scientific, Singapore, 2002.