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

Inexact proximal point method for piecewise-star-convex function

Orizon P. Ferreira · Max L. N. Gonçalves ·
Jefferson G. Melo

the date of receipt and acceptance should be inserted later

Abstract We propose and analyze an inexact proximal point method for minimizing locally Lipschitz functions on Euclidean spaces with a piecewise star-convex structure. More precisely, the space is covered by finitely many closed convex sets, and on each set the objective function satisfies a star-convex inequality with respect to the minimizers of its restriction. This class includes, in particular, convex and star-convex functions. Under uniform bounds on the proximal parameters, we first prove that the inexact proximal scheme is well defined, the objective function values decrease and every accumulation point is Clarke stationary. We then use the piecewise geometry to show that any accumulation point in the interior of a region minimizes the objective function over that region and is therefore a local minimizer. We further prove that the iterates eventually remain in a single region and that the whole sequence converges under a summability condition on the error tolerance sequence. In particular, when the partition consists of only one region, namely, when the objective function is star-convex, the method is globally convergent. Therefore, the piecewise star-convex structure yields stronger conclusions than the accumulation-point stationarity that is typically obtained for general nonconvex functions. In addition, once the iterates remain in a fixed region, we obtain a sublinear rate of order $\mathcal{O}(1/k)$ for the objective function values. We also prove that an ε-approximate Clarke stationary point is computed in at most $\mathcal{O}(1/\varepsilon^2)$ iterations. In the piecewise strongly star-convex setting, the method achieves linear convergence. Numerical examples illustrate the behavior of the proposed algorithm.

keywords: inexact proximal point method; piecewise star-convexity; nonconvex optimization.

AMS subject classifications: 90C25, 90C60, 90C30, 65K05.

1 Introduction

Many optimization problems considered in the literature have objective functions that are defined piecewise with respect to a finite partition of Euclidean space. In such problems, the domain is decomposed into finitely many closed sets, and on each set the objective function exhibits a

IME, Universidade Federal de Goiás, Goiânia, GO 74001-970, Brazil, April 7, 2026.
orizon@ufg.br, maxlng@ufg.br and jefferson@ufg.br
The work of these authors was supported in part by CNPq [grant No. 312223/2022-6], FAPEG [grant No. 202510267001610] and FAPESC [grant No. 2024TR002238].

simpler structure than on the whole space. This piecewise description has motivated several notions of piecewise convexity and a substantial body of work on algorithms designed to exploit the underlying partition. In parallel, the proximal point method has emerged as one of the fundamental first–order schemes in optimization and monotone operator theory, and its extension to broader nonconvex classes remains an active topic of research. Developing new convergence and complexity results for this method, particularly in combination with piecewise structures, is therefore of independent interest. Before turning to our specific framework, we first relate these notions to the existing literature that motivates the present study.

In a foundational work, Louveaux [13] introduced piecewise convex programs, in which the feasible set is decomposed into finitely many closed convex subsets and, on each subset, the objective function coincides with a smooth convex function; a cutting-plane algorithm was proposed and analyzed. Subsequently, [7] discussed piecewise convex maximization problems where the objective function is given by the pointwise minimum of finitely many convex functions, and developed a method based on optimality conditions. In [16], the training and testing objective functions of neural networks with piecewise affine layers were shown to be continuous piecewise convex or piecewise multi-convex, and convergence properties were obtained for algorithms that repeatedly solve convex subproblems on the elements of the underlying partition. Altogether, these contributions illustrate that piecewise convexity covers important classes of nonconvex models while preserving convex structure on each subset.

Complementary to the developments on piecewise convexity, a substantial line of research has extended global convergence guarantees for optimization methods beyond the convex framework by imposing geometric restrictions on the objective function. A central example is the notion of star–convexity, introduced in the classical work about cubic regularization of Newton method of Nesterov and Polyak [15], and its generalization to quasar–convexity. These classes allow nonconvex level sets while enforcing a one–sided inequality along rays emanating from a reference point or from the set of global minimizers. Building on this geometry, Hinder, Sidford and Sohoni [8] constructed first–order algorithms for smooth star–convex and quasar–convex functions whose iteration complexity matches known lower bounds up to logarithmic factors. More recently, several authors have shown that suitable generalized convexity assumptions suffice to recover global linear convergence of gradient–type and proximal methods beyond the classical strongly convex setting. The notion of strong star–quasiconvexity introduced in [11] unifies several generalized strong convexity concepts and it yields linear convergence guarantees for some first order methods over closed star–shaped sets. In a complementary direction, [4] analyzes the quasar–convex case, establishing existence and uniqueness of minimizers, quadratic growth and Polyak–Łojasiewicz type properties, and deriving linear convergence for the proximal point method under strong quasar–convexity. Constrained smooth quasar–convex optimization is studied in [14], where an inexact accelerated proximal point scheme is designed and shown to achieve nearly optimal oracle complexity. The latter reference also discuss global convergence of projected gradient and Frank-Wolfe algorithms in the quasar–convex setting with general convex constraints. Star–convex functions form an intermediate class between convex and general nonconvex objective functions. Roughly speaking, they preserve convexity only along segments joining a point to a distinguished minimizer, while convexity along arbitrary segments may fail. This structure is broad enough to include relevant nonconvex models and still strong enough to support meaningful convergence guarantees, as illustrated by recent analyses of Frank–Wolfe type methods in star–convex and related settings [5,6]. These developments suggest that useful convergence properties can still be obtained under mild nonconvex geometries, and they motivate the proximal framework studied in this paper for piecewise star–convex objective functions.

Our goal is to combine this radial geometry with a piecewise description of the objective function. We consider locally Lipschitz functions on $\mathbb{R}^n$ that are piecewise star–convex with respect to a finite family of nonempty closed convex sets. On each region, the restriction of the objective function satisfies a star–convex inequality with respect to the set of minimizers of that restriction. In particular, every Clarke stationary point lying in the interior of a region is a minimizer on that region. Thus, the objective function may be globally nonconvex and may possess several distinct attraction regions, while still retaining a useful local structure on each element of the partition. For this class, we propose and analyze an unconstrained inexact proximal point method. At each iteration, we consider a quadratic regularization of the objective function and compute an approximate solution satisfying a relative residual condition and a sufficient-decrease test. Since the subproblem is generally nonconvex, it may have multiple critical points. We show that each regularized subproblem is coercive, which, in particular, allow to prove that the method is well defined.

The analysis is based on an inequality that combines the sufficient-decrease condition with the relative inexactness requirement. On the region containing the accepted iterate, this estimate controls both the decrease produced by the proximal step and the distance to the set of minimizers of the corresponding restriction. From this inequality and under uniform bounds on the proximal parameters, we derive the main convergence results. In contrast with the usual qualitative conclusion in nonsmooth nonconvex optimization, where one often obtains only stationarity of accumulation points, the piecewise star–convex structure yields stronger results. In addition to Clarke stationarity of all accumulation points, we show that any accumulation point lying in the interior of a region minimizes the objective function over that region and is therefore a local minimizer. We further prove that the iterates eventually remain in a single region and that the whole sequence converges under a summability condition on the error tolerance sequence. In particular, when the partition consists of a single region, namely when the objective function is star–convex, the method converges to a global minimizer. We also prove a sublinear rate of order $\mathcal{O}(1/k)$ for the objective function values once the iterates stay in a fixed region, an $\mathcal{O}(1/\varepsilon^2)$ complexity bound for computing an ε-approximate Clarke stationary point, and linear convergence in the piecewise strongly star–convex case. Numerical examples illustrate the practical behavior of the proposed method.

The remainder of the paper is organized as follows. In Section 2, we introduce the notation and recall basic notions from Clarke subdifferential calculus. Section 3 presents the definitions of piecewise-star-convex functions and establishes its main geometric and subdifferential properties. In Section 4, we describe the inexact proximal point algorithm, show its well definition, and derive auxiliary properties of the resulting proximal sequence. Section 4.1 contains the main convergence analysis, while Section 4.2 is devoted to the complexity analysis. In Section 5, we discuss several concrete optimization problems whose objective functions are piecewise star–convex. Section 6 presents illustrative numerical tests. Finally, Section 7 offers some concluding remarks and outlines possible extensions.

2 Preliminaries

In this section, we recall some notations, definitions and basics results used throughout the paper.

Let $\mathcal{C}$ be a convex set. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* on $\mathcal{C}$ if, for all $x, y \in \mathcal{C}$, there holds

$$f\big((1-t)x + ty\big) \le (1-t)f(x) + t\, f(y), \quad \forall t \in [0,1].$$

For $\mu > 0$, the function f is *μ-strongly convex* on $\mathcal{C}$ if, for all $x, y \in \mathcal{C}$, there holds

$$f\big((1-t)x + ty\big) \le (1-t)f(x) + t\, f(y) \;-\; \frac{\mu}{2}\, t(1-t)\, \|x-y\|^2, \qquad \forall t \in [0,1].$$

Here $\|\cdot\|$ denotes the Euclidean norm. For a comprehensive treatment of convex functions, see [9]. The function f is *locally Lipschitz* if, for all $x \in \mathbb{R}^n$, there exist a constant $L_x > 0$ and a neighborhood U_x of x such that

$$|f(z) - f(y)| \leq L_x \|z - y\|, \qquad \forall y, z \in U_x.$$

It is well known that, if $f : \mathbb{R}^n \to \mathbb{R}$ is either convex or continuously differentiable, then f is locally Lipschitz; see, for example, [3, p. 32]. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function. The *Clarke's subdifferential* of f at $x \in \mathbb{R}^n$ is the set defined by

$$\partial^c f(x) := \{v \in \mathbb{R}^n : \ f^\circ(x; d) \geq v^{\mathrm{T}} d, \ \forall d \in \mathbb{R}^n\},$$

where $f^\circ(x; d)$ is the *generalized directional derivative* of f at x in the direction d, given by

$$f^\circ(x; d) = \limsup_{\substack{u \to x \\ t \downarrow 0}} \frac{f(u + td) - f(u)}{t}.$$

For an extensive study of locally Lipschitz functions and Clarke's subdifferential see [3, p. 27]. If f is convex, then $\partial^c f(x)$ coincides with the subdifferential $\partial f(x)$ in the sense of convex analysis, and $f^\circ(x; d)$ coincides with the usual directional derivative $f'(x; d)$; see [3, p. 36]. We recall that if $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then $\partial^c f(x) = \{\nabla f(x)\}$ for any $x \in \mathbb{R}^n$; see [3, p. 33].

Theorem 2.1 ([3, p. 27]) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function. Then, $\partial^c f(x)$ is a nonempty, convex, and compact subset of $\mathbb{R}^n$, and $\|v\| \leq L_x$, for all $v \in \partial^c f(x)$, where $L_x > 0$ is the Lipschitz constant of f around x. Moreover, $f^\circ(x; d) = \max\{v^{\mathrm{T}} d : \ v \in \partial^c f(x)\}$.*

The next result follows from [3, Corollary on p. 52].

Theorem 2.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function and let $\mathcal{C} \subset \mathbb{R}^n$ be a nonempty, closed, and convex set. If $x^* \in \mathcal{C}$ is a (local) minimizer of f over $\mathcal{C}$, then there exists $v \in \partial^c f(x^*)$ such that $v^{\mathrm{T}}(x - x^*) \geq 0$, for all $x \in \mathcal{C}$.*

In view of Theorem 2.2, a point $x^* \in \mathcal{C}$ is called a Clarke stationary point of the problem $\min_{x \in \mathcal{C}} f(x)$ if there exists $v \in \partial^c f(x^*)$ such that $v^{\mathrm{T}}(x - x^*) \geq 0$, for all $x \in \mathcal{C}$. As a consequence, for unconstrained problem, i.e., $\mathcal{C} = \mathbb{R}^n$, by taking $x = x^* - v$ in the last inequality, we see that $x^* \in \mathbb{R}^n$ is a Clarke stationary point if and only if $0 \in \partial^c f(x^*)$.

Proposition 2.1 ([3, Proposition 2.1.5 on p. 29]) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function. Let $\{x_k\}_{k \in \mathbb{N}}$ and $\{u_k\}_{k \in \mathbb{N}}$ be sequences such that $u_k \in \partial^c f(x_k)$, for all $k \in \mathbb{N}$. If $\lim_{k \to +\infty} x_k = \bar{x}$ and $\lim_{k \to +\infty} u_k = \bar{u}$, then $\bar{u} \in \partial^c f(\bar{x})$.*

We conclude this section with a technical lemma on positive infinite products, which will be used in the convergence analysis of the proximal sequence. We first introduce the following notation, to be used in the next sections. Let $\{\sigma_k\}_{k \in \mathbb{N}} \subset [0, 1)$ and $k_0 \in \mathbb{N}$, and define the sequence $\{\omega_k\}_{k \geq k_0}$ by

$$\omega_{k_0} = 1, \qquad \omega_k := \prod_{t=k_0}^{k-1} \big(1 - \sigma_t\big), \qquad k = k_0 + 1, k_0 + 2, \ldots. \tag{1}$$

Lemma 2.1 *Assume that $\sum_{k=k_0}^{+\infty} \frac{\sigma_k}{1 - \sigma_k} < +\infty$. Then, the sequence $\{\omega_k\}_{k \geq k_0}$, defined in (1), converges to a positive number ω.*

Proof Since $0 \le \sigma_k < 1$, we have $1-\sigma_k \in (0,1]$, hence $-\ln(1-\sigma_k) \ge 0$, for all $k \in \mathbb{N}$. For $s \in [0,1)$, we have

$$-\ln(1-s) = \int_0^s \frac{1}{1-t}\,dt \le \int_0^s \frac{1}{1-s}\,dt = \frac{s}{1-s}, \tag{2}$$

and therefore, with $s=\sigma_k$, we obtain that $0 \le -\ln(1-\sigma_k) \le \frac{\sigma_k}{1-\sigma_k}$, for all $k \in \mathbb{N}$, which combined with $\sum_{k=k_0}^{+\infty} \frac{\sigma_k}{1-\sigma_k} < +\infty$ gives

$$\sum_{k=k_0}^{+\infty} -\ln(1-\sigma_k) < +\infty. \tag{3}$$

Taking logarithms and using definition of ω_k, we have $\ln \omega_k = \sum_{t=k_0}^{k-1} \ln(1-\sigma_t)$, hence

$$-\ln \omega_k = \sum_{t=k_0}^{k-1} -\ln(1-\sigma_t).$$

By (3), the partial sums on the right-hand side converge to some $S \in [0,+\infty)$. Therefore, $\lim_{k\to+\infty}\big(-\ln\omega_k\big) = S$ and, consequently, $\lim_{k\to+\infty}\ln\omega_k = -S$. By continuity of the exponential function, we obtain that

$$\lim_{k\to+\infty}\omega_k = \lim_{k\to+\infty} e^{\ln\omega_k} = e^{\lim_{k\to+\infty}\ln\omega_k} = e^{-S}.$$

Thus, the results follows by taking $\omega := e^{-S}$. □

3 Piecewise-star-convex functions

In this section, we recall the concept of *piecewise-star-convexity*, a relaxation of the classical star–convexity notion from [15]. Many functions that are nonconvex nevertheless satisfy this weaker property. As we will show, by identifying and exploiting piecewise-star-convexity we can significantly broaden the class of problems for which the proximal point method achieves global convergence under mild assumptions.

We begin by recalling the definition of star–convexity.

Definition 3.1 A function $f : \mathbb{R}^n \to \mathbb{R}$ is *star-convex* on $\mathbb{R}^n$ if its set of global minimizers $X^* := \arg\min_{\mathbb{R}^n} f$ is nonempty and, for any $x^* \in X^*$, there holds

$$f(tx^* + (1-t)x) \le tf(x^*) + (1-t)f(x), \qquad \forall\, x \in \mathbb{R}^n,\ \forall\, t \in [0,1]. \tag{4}$$

Every convex function with a nonempty set of global minimizers is star–convex, but the converse does not hold. We present two star-convex functions that are not convex; see [15] (see also [6] for additional examples).

Example 3.1 The functions $\phi(t) = |t|(1-e^{-|t|})$ and $f(s,t) = s^2t^2 + s^2 + t^2$ are star–convex but are not convex.

We next recall the piecewise-star-convexity on the whole space $\mathbb{R}^n$.

Definition 3.2 Let I be a finite index set and let $\mathcal{P} := \{\mathcal{C}_i\}_{i\in I}$ be a family of nonempty closed convex subsets $\mathcal{C}_i \subset \mathbb{R}^n$ such that $\bigcup_{i\in I}\mathcal{C}_i = \mathbb{R}^n$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *piecewise-star-convex (PWSC) with respect to* $\mathcal{P}$ if, for every $i \in I$, the following conditions hold:

(i) the set of minimizers of f on $\mathcal{C}_i$, denoted by X_i^*, is nonempty;
(ii) for any $x_i^* \in X_i^*$, there holds

$$f\big(tx_i^* + (1-t)x\big) \leq tf(x_i^*) + (1-t)f(x), \qquad \forall x \in \mathcal{C}_i, \quad \forall t \in [0,1]. \tag{5}$$

Remark 3.1 Note that if the family $\mathcal{P}$ is trivial, i.e., if $\mathcal{P} = \{\mathbb{R}^n\}$ and $X_1^* = X^*$, then the class of PWSC functions reduces to the class of star-convex functions in Definition 3.1. In practice, piecewise-star-convexity often arises when f is star-convex on each region $\mathcal{C}_i$ but not across different regions. This structure ensures that star–type inequalities hold on each set, with respect to its minimizers X_i^*, even if such inequalities fail when crossing set boundaries. In this setting, the minimizers $X_i^* = \arg\min_{\mathcal{C}_i} f$ are *global* minimizers over their corresponding sets $\mathcal{C}_i$, although they may be only *local* and possibly non–global minimizers of f on the whole domain $\mathbb{R}^n$. Moreover, an element of X_i^* may fail to be even a local minimizer of f on $\mathbb{R}^n$. To illustrate this fact, consider the function $f(x) = \min\{x^2, (x-1)^2 - 1\}$ and the family of convex, and closed sets $\mathcal{P} := \{(-\infty, 0], [0, +\infty)\}$. It can be easily seen that f is PWSC with respect to $\mathcal{P}$, and the point $\bar{x} := 0$ is a minimizer of f on the set $\mathcal{C}_1 := (-\infty, 0]$, but it is not a local minimizer of f on the whole space $\mathbb{R}$. In fact, $\bar{x}$ is a local maximizer of f on the set $\mathcal{C}_2 := [0, +\infty)$.

We next present an example of a PWSC function that fails to be convex on the subsets.

Example 3.2 Let $n = 2$ and consider the two closed convex sets

$$\mathcal{C}_1 := \{x = (x_1, x_2) \in \mathbb{R}^2 : \ x_1 \leq 0\}, \qquad \mathcal{C}_2 := \{x = (x_1, x_2) \in \mathbb{R}^2 : \ x_1 \geq 0\},$$

so that $\mathcal{P} = \{\mathcal{C}_1, \mathcal{C}_2\}$ covers $\mathbb{R}^2$. Define the polynomial function $\psi : \mathbb{R}^2 \to \mathbb{R}$ by

$$\psi(u_1, u_2) := u_1^4 + u_2^4 - u_1^2 u_2^2.$$

Let $a_1 := (-1, 0) \in \operatorname{int} \mathcal{C}_1$ and $a_2 := (1, 0) \in \operatorname{int} \mathcal{C}_2$, and define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x) := \begin{cases} \psi(x - a_1), & x \in \mathcal{C}_1, \\ \psi(x - a_2), & x \in \mathcal{C}_2. \end{cases}$$

Then f is locally Lipschitz on $\mathbb{R}^2$, is PWSC with respect to $\mathcal{P}$, is not convex on $\mathcal{C}_1$ nor on $\mathcal{C}_2$, and has two distinct strict global minimizers, namely a_1 and a_2. Indeed, first note the identity

$$\psi(u_1, u_2) = (u_1^2 - u_2^2)^2 + u_1^2 u_2^2 \ \geq \ 0.$$

Moreover, $\psi(u_1, u_2) = 0$ holds only at the origin $(u_1, u_2) = (0, 0)$. Moreover, ψ is 4–homogeneous: $\psi(\alpha u) = \alpha^4 \psi(u)$ for all $\alpha \geq 0$.

On the boundary $x_1 = 0$ we have $x - a_1 = (1, x_2)$ and $x - a_2 = (-1, x_2)$. It can be easily seen that

$$\psi(1, x_2) = \psi(-1, x_2) = 1 + x_2^4 - x_2^2.$$

Thus, the two expressions defining f agree on $x_1 = 0$. Therefore f is continuous and piecewise polynomial, hence locally Lipschitz on $\mathbb{R}^2$.

Fix $i \in \{1, 2\}$. Since $\psi \geq 0$ and it vanishes only at the origin, we have $f(x) = \psi(x - a_i) \geq 0$ for all $x \in \mathcal{C}_i$, with equality at $x = a_i$. Moreover, if $x \in \mathcal{C}_i$ satisfies $f(x) = 0$, then $\psi(x - a_i) = 0$, hence $x = a_i$. Therefore a_i is the unique global minimizer of f on $\mathcal{C}_i$, and

$$X_i^* = \arg\min_{x \in \mathcal{C}_i} f(x) = \{a_i\},$$

which verifies Definition 3.2(i). Now take $x_i^* = a_i \in X_i^*$, $x \in \mathcal{C}_i$, and $t \in [0,1]$. Since $\mathcal{C}_i$ is convex, the point $tx_i^* + (1-t)x$ belongs to $\mathcal{C}_i$. Using $f(x_i^*) = 0$ and the 4–homogeneity of ψ, we have

$$f\big(tx_i^* + (1-t)x\big) = \psi\big((tx_i^* + (1-t)x) - a_i\big) = \psi\big((1-t)(x-a_i)\big) = (1-t)^4\,\psi(x-a_i).$$

Because $\psi(x-a_i) \geq 0$ and $(1-t)^4 \leq (1-t)$ for $t \in [0,1]$, we obtain

$$f\big(tx_i^* + (1-t)x\big) \leq (1-t)\,\psi(x-a_i) = (1-t)f(x) = tf(x_i^*) + (1-t)f(x),$$

which proves (5) in Definition 3.2(ii).

To see that f is not convex on either set, fix $i \in \{1,2\}$ and consider the line $x_i(s) = a_i + (1,s) \in \mathcal{C}_i$. Along this line we have $x_i(s) - a_i = (1,s)$, and therefore

$$f\big(x_i(s)\big) = \psi(1,s) = 1 + s^4 - s^2.$$

Since $\frac{d^2}{ds^2} f\big(x_i(s)\big)\big|_{s=0} = -2 < 0$, the restriction $s \mapsto f(x_i(s))$ fails to be convex in a neighborhood of 0. Hence f is not convex on $\mathcal{C}_1$ nor on $\mathcal{C}_2$. Since $\psi(u) \geq 0$ with equality only at $u = 0$, each point a_i satisfies $f(a_i) = 0$ and $f(x) > 0$ for all $x \neq a_i$ in a neighborhood contained in $\mathcal{C}_i$. Thus a_1 and a_2 are strict global minimizers of f.

The previous example illustrates that PWSC functions may be far from being convex on the subsets. Nevertheless, the star–convex inequality (5) still enforces a useful first–order behavior on the interior of each region. In particular, it yields a one–sided support inequality for Clarke subgradients, which will be the key tool in our convergence analysis. This is made precise in the following proposition. The next definition will be needed for the result.

We recall that f is said to be *μ_i-strongly star-convex on $\mathcal{C}_i$* for $\mu_i > 0$ if its set minimizers X_i^* is nonempty and, for any $x^* \in X_i^*$, there holds

$$f\big(tx^* + (1-t)x\big) \leq tf(x^*) + (1-t)f(x) - \frac{\mu_i}{2}t(1-t)\|x - x^*\|^2, \qquad \forall x \in \mathcal{C}_i,\ \forall t \in [0,1]. \quad (6)$$

Proposition 3.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function and PWSC with respect to the finite family $\mathcal{P} = \{\mathcal{C}_i\}_{i\in I}$. Assume that $\operatorname{int}\mathcal{C}_i \neq \varnothing$, and let $x_i^* \in X_i^*$. Then,*

$$v^\top(x_i^* - x) \leq f^\circ(x; x_i^* - x) \leq f(x_i^*) - f(x), \qquad \forall x \in \operatorname{int}\mathcal{C}_i, \quad \forall\, v \in \partial^c f(x). \quad (7)$$

If, in addition, f is μ_i-strongly star-convex on $\mathcal{C}_i$, then the upper bound in (7) *improves to*

$$f^\circ(x; x_i^* - x) \leq f(x_i^*) - f(x) - \frac{\mu_i}{2}\|x - x_i^*\|^2, \qquad \forall x \in \operatorname{int}\mathcal{C}_i. \quad (8)$$

Proof Let $i \in I$, $x_i^* \in X_i^*$, and $x \in \operatorname{int}\mathcal{C}_i$. Thus, by applying Theorem 2.1, we have

$$f^\circ(x;\, x_i^* - x) = \max_{w \in \partial^c f(x)} w^\top(x_i^* - x).$$

Therefore, for every $v \in \partial^c f(x)$, we obtain that $v^\top(x_i^* - x) \leq f^\circ(x;\, x_i^* - x)$, which proves the first inequality in (7). We now prove inequality (8). The proof of the second inequality in (7) follows by a similar argument with $\mu_i = 0$. Let $L_x > 0$ be a Lipschitz constant of f around x. Since $x \in \operatorname{int}\mathcal{C}_i$, choose $r > 0$ such that $B(x,r) \subset \mathcal{C}_i$ and f is L_x–Lipschitz on $B(x,r)$. Let $d := x_i^* - x$. Consider arbitrary $u \in B(x, r/2)$ and $t \in (0, 1/2]$, and define

$$y := \frac{u - tx}{1-t}.$$

Then, we have

$$u = (1-t)y + tx, \qquad u + td = (1-t)y + tx_i^*, \qquad \|y - x\| = \frac{1}{1-t}\|u - x\| \le 2\|u - x\| < r,$$

which, in particular, implies that $y \in B(x, r) \subset \mathcal{C}_i$. Since f is μ_i-strongly star-convex on $\mathcal{C}_i$, we obtain that

$$f(u + td) = f\big((1-t)y + tx_i^*\big) \le (1-t)f(y) + tf(x_i^*) - \frac{t(1-t)\mu_i}{2}\|y - x_i^*\|^2.$$

Subtracting $f(u)$ from both sides of the above inequality and dividing by $t > 0$, we conclude that

$$\frac{f(u+td) - f(u)}{t} \le f(x_i^*) - f(u) + \frac{1-t}{t}\big(f(y) - f(u)\big) - \frac{(1-t)\mu_i}{2}\|y - x_i^*\|^2. \tag{9}$$

Using the L_x–Lipschitz continuity of f on $B(x, r)$ and the fact that $\|y - u\| = [t/(1-t)]\|u - x\|$, we obtain

$$|f(y) - f(u)| \le L_x\|y - u\| \le \frac{tL_x}{1-t}\|u - x\|.$$

Hence, it follows from (9) that

$$\frac{f(u+td) - f(u)}{t} \le f(x_i^*) - f(u) + L_x\|u - x\| - \frac{(1-t)\mu_i}{2}\|y - x_i^*\|^2.$$

It follows from the above inequality, the continuity of f, and the fact that y approaches x as u approaches x that

$$f^\circ(x; d) = \limsup_{\substack{u \to x \\ t \downarrow 0}} \frac{f(u+td) - f(u)}{t} \le f(x_i^*) - f(x) - \frac{\mu_i}{2}\|x - x_i^*\|^2,$$

which proves the inequality in (8), concluding the proof of the proposition. □

Remark 3.2 The interior assumption in the proof of Proposition 3.1 is essential. Indeed, for instance, let $\mathcal{C}_1 = [0, \infty)$, $\mathcal{C}_2 = (-\infty, 0]$, and define the function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} (x-1)^2, & x \ge 0, \\ (x+1)^2, & x \le 0. \end{cases}$$

Then f is locally Lipschitz and PWSC on each subset, and $X_1^* = \{1\}$. At the boundary point $x = 0 \in \mathcal{C}_1$, take $x_i^* = 1 \in X_1^*$ and set $d := x_i^* - x = 1$. It is easy to see that $f^\circ(0; d) = 2$. On the other hand, $f(x_i^*) - f(x) = f(1) - f(0) = 0 - 1 = -1$. Hence, we see that

$$f^\circ(x; x_i^* - x) = f^\circ(0; d) > f(x_i^*) - f(x).$$

Therefore, the second inequality in (7) fails at $x = 0$, showing that one cannot drop the requirement $x \in \operatorname{int} \mathcal{C}_i$.

Corollary 3.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function that is PWSC with respect to the finite family $\mathcal{P} = \{\mathcal{C}_i\}_{i \in I}$. Let $\bar{x} \in \mathbb{R}^n$ be a Clarke stationary point of f, i.e., $0 \in \partial^c f(\bar{x})$. Then, for every index $i \in I$ such that $\bar{x} \in \operatorname{int} \mathcal{C}_i$, the following statements hold*

a) we have $f(\bar{x}) = \min_{x \in \mathcal{C}_i} f(x)$, equivalently $\bar{x} \in X_i^*$.
b) the element $\bar{x}$ is a local minimizer of f in $\mathbb{R}^n$;

c) If f is μ_i-strongly star-convex on $\mathcal{C}_i$, then the following inequality holds

$$f(x) \geq f(\bar{x}) + \frac{\mu_i}{2}\|x - \bar{x}\|^2, \qquad \forall x \in \mathcal{C}_i. \tag{10}$$

As a consequence, $X_i^* = \{\bar{x}\}$.

Proof Fix $i \in I$ such that $\bar{x} \in \text{int}\mathcal{C}_i$. Since $0 \in \partial^c f(\bar{x})$, Proposition 3.1 applied with $x = \bar{x}$ and $v = 0$ yields, for every $x_i^* \in X_i^*$,

$$0 = v^\top (x_i^* - \bar{x}) \leq f^\circ(\bar{x}; x_i^* - \bar{x}) \leq f(x_i^*) - f(\bar{x}).$$

Hence, we have

$$f(\bar{x}) \leq f(x_i^*). \tag{11}$$

Since $x_i^* \in X_i^*$ is a minimizer of f over $\mathcal{C}_i$, we obtain $f(\bar{x}) = \min_{x \in \mathcal{C}_i} f(x)$, which proves the statement in (a). Now note that $\bar{x} \in \text{int}\mathcal{C}_i$ implies that there exists $r > 0$ such that $B(\bar{x}, r) \subset \mathcal{C}_i$. Thus, for every $y \in B(\bar{x}, r)$, we have $f(y) \geq \min_{x \in \mathcal{C}_i} f(x) = f(\bar{x})$, which proves the statement in (b). Let us now prove (c). Since f is μ_i-strongly star-convex on $\mathcal{C}_i$, from inequality (6) and statement (a), we have

$$f(\bar{x}) \leq f\big(t\bar{x} + (1-t)x\big) \leq tf(\bar{x}) + (1-t)f(x) - \frac{\mu_i}{2}t(1-t)\|x - \bar{x}\|^2, \qquad \forall x \in \mathcal{C}_i, \quad \forall\, t \in [0,1],$$

Hence, from the above inequailties we obtain

$$\frac{\mu_i}{2}t(1-t)\|x - \bar{x}\|^2 \leq (1-t)(f(x) - f(\bar{x})), \qquad \forall x \in \mathcal{C}_i, \quad \forall\, t \in [0,1].$$

Considering $t \in (0,1)$ and dividing both sides of the above inequality by $1-t$, we arrive at

$$\frac{\mu_i}{2}t\|x - \bar{x}\|^2 \leq f(x) - f(\bar{x}), \qquad \forall x \in \mathcal{C}_i, \quad \forall\, t \in [0,1).$$

Therefore, (10) follows by taking limit as t goes to 1 on both sides of the above inequality. The last statement of (c) follows trivially from (10). □

Remark 3.3 The interior assumption $\bar{x} \in \text{int}\,\mathcal{C}_i$ in Corollary 3.1 is essential. Indeed, a Clarke stationary point located on the boundary of the subsets may fail to minimize f on any adjacent subset and may even be a local maximizer; see Remarks 3.1 and 3.2.

We now turn to points located on the boundary between subsets. While Corollary 3.1 provides a strong conclusion for Clarke stationary points lying in the interior of a subset, the same type of statement does not extend to boundary points. The next remark illustrates that, even under PWSC, a Clarke stationary point on the boundary may exhibit different local behaviors.

Remark 3.4 Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz and PWSC with respect to a *finite* family $\mathcal{P} = \{\mathcal{C}_i\}_{i \in I}$ of nonempty closed convex sets with $\bigcup_{i \in I} \mathcal{C}_i = \mathbb{R}^n$. For a point $\bar{x} \in \mathbb{R}^n$, define the set of *incident subsets* $I(\bar{x}) := \{\, i \in I : \ \bar{x} \in \mathcal{C}_i \,\}$. Clarke stationarity on the boundary does not force a single behavior. Indeed, even if $0 \in \partial^c f(\bar{x})$ and $\bar{x}$ lies on the common boundary of at least two subsets, $\bar{x}$ may be a global minimizer, a strict local maximizer, or a saddle point. For instance, in $\mathbb{R}$ with $\mathcal{C}_1 = [0, +\infty)$ and $\mathcal{C}_2 = (-\infty, 0]$:

(i) For global minimizer, take $f(x) = |x|$. Then, f is convex on each subset, $0 \in \partial^c f(0)$, and 0 is a global minimizer of f.
(ii) For strict local maximizer, the construction in Remark 3.2 yields a PWSC function for which $0 \in \partial^c f(0)$ and 0 is a strict local maximizer of f.

(iii) For saddle point, define

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

Each piece is (strongly) convex and therefore f is PWSC at $x = 0$. The one-sided derivatives satisfy $f'_+(0) = 0$ and $f'_-(0) = 1$, and hence $\partial^c f(0) = \text{co}\{0, 1\} = [0, 1]$; in particular, $0 \in \partial^c f(0)$. On the other hand, for $\varepsilon > 0$ small, we have $f(-\varepsilon) = \varepsilon^2 - \varepsilon < 0 = f(0)$ while $f(\varepsilon) = \varepsilon^2 > 0 = f(0)$, thus 0 is neither a local minimizer nor a local maximizer of f.

Therefore, in view of Remark 3.4, the inclusion $0 \in \partial^c f(\bar{x})$ does not determine the local nature of a point $\bar{x}$ lying on the boundary of a subset, even for PWSC functions. Nevertheless, local minimizers admit a precise characterization in terms of the incident subsets $I(\bar{x})$, as stated in the next corollary.

Corollary 3.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function that is PWSC with respect to a finite family $\mathcal{P} = \{\mathcal{C}_i\}_{i\in I}$ of nonempty closed convex sets covering $\mathbb{R}^n$. For $\bar{x} \in \mathbb{R}^n$, let $I(\bar{x})$ be as in Remark 3.4. Then, $\bar{x}$ is a local minimizer of f on $\mathbb{R}^n$ if and only if*

$$\bar{x} \in X_i^* \quad \textit{for every } i \in I(\bar{x}), \qquad \textit{or, equivalently,} \qquad f(\bar{x}) = \min_{x\in\mathcal{C}_i} f(x) \quad \forall i \in I(\bar{x}). \tag{12}$$

Proof First, assume that $\bar{x}$ is a local minimizer of f in $\mathbb{R}^n$. Fix $i \in I(\bar{x})$ and take any $x_i^* \in X_i^*$. Suppose, by contradiction, that $f(x_i^*) < f(\bar{x})$. For $t \in (0, 1)$ define $y_t := tx_i^* + (1-t)\bar{x}$. Since $\mathcal{C}_i$ is convex and $x_i^*, \bar{x} \in \mathcal{C}_i$, we have $y_t \in \mathcal{C}_i$, and clearly $\lim_{t\downarrow 0} y_t = \bar{x}$. Because $\bar{x}$ is a local minimizer, there exists $t_0 \in (0, 1)$ such that

$$f(y_t) \;\geq\; f(\bar{x}), \qquad \forall t \in (0, t_0].$$

On the other hand, since f is PWSC, applying (5) with $x = \bar{x}$ and this x_i^* yields

$$f(y_t) \;=\; f\big(tx_i^* + (1-t)\bar{x}\big) \;\leq\; tf(x_i^*) + (1-t)f(\bar{x}), \qquad \forall t \in [0, 1].$$

Since $f(x_i^*) < f(\bar{x})$, the right-hand side is *strictly smaller* than $f(\bar{x})$ for every $t \in (0, 1)$, hence

$$f(y_t) \;<\; f(\bar{x}), \qquad \forall t \in (0, 1),$$

which contradicts $f(y_t) \geq f(\bar{x})$ for small t. Therefore $f(x_i^*) \geq f(\bar{x})$. But $x_i^* \in X_i^*$ is a minimizer on $\mathcal{C}_i$, so $f(x_i^*) \leq f(\bar{x})$, because $\bar{x} \in \mathcal{C}_i$. Hence $f(x_i^*) = f(\bar{x})$, i.e., $\bar{x} \in X_i^*$. As $i \in I(\bar{x})$ was arbitrary, (12) follows.

For the converse, assume (12). Set $\mathcal{U} := \bigcup_{i\notin I(\bar{x})} \mathcal{C}_i$. Because $\mathcal{P}$ is finite and each $\mathcal{C}_i$ is closed, $\mathcal{U}$ is closed. Moreover $\bar{x} \notin \mathcal{U}$ by definition of $I(\bar{x})$. Hence $\text{dist}(\bar{x}, \mathcal{U}) =: r > 0$. Therefore, we have $B(\bar{x}, r) \cap \mathcal{U} = \varnothing$, which implies that $B(\bar{x}, r) \subset \mathbb{R}^n \setminus \mathcal{U}$. Take any $y \in B(\bar{x}, r)$. Then $y \in \mathcal{C}_i$ for some $i \in I(\bar{x})$, and so $f(y) \;\geq\; \min_{x\in\mathcal{C}_i} f(x) \;=\; f(\bar{x})$, where we used (12). This proves that $\bar{x}$ is a local minimizer of f. $\square$

We conclude this section by noting that Section 5 presents further concrete optimization problems whose objective functions are piecewise–star–convex in the sense of Definition 3.2.

4 Inexact proximal point method

This section presents an inexact proximal point method for minimizing a locally Lipschitz PWSC function, proves that the method is well defined, and establishes its asymptotic convergence properties and iteration-complexity bounds. At each iteration, an unconstrained local solver may be used to compute an approximate stationary solution of the proximal subproblem, together with an associated residual. The solution satisfies a sufficient-decrease condition and an inexactness criterion, quantified through a residual-based relative measure consistent with the notion of criticality returned by typical numerical solvers. When the objective function is globally convex, the proposed scheme reduces to an inexact version of the classical proximal point method.

Henceforth, we consider the problem of minimizing a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ that is PWSC with respect to a family $\mathcal{P} := \{\mathcal{C}_i\}_{i\in I}$ of nonempty, closed, and convex sets, where I is a finite index set.

In the following, we state the proposed method.

Algorithm 1 Inexact proximal point method

Input: An initial point $x_0 \in \mathbb{R}^n$*, and two scalar sequences* $\{\lambda_k\}_{k\in\mathbb{N}} \subset (0, \infty)$ *and* $\{\sigma_k\}_{k\in\mathbb{N}} \subset [0, 1)$*.*

Step 0. Initialization: Set $k \leftarrow 0$*.*

Step 1. Proximal subproblem: Define the regularized function

$$\varphi_k(x) := f(x) + \tfrac{\lambda_k}{2}\,\|x - x_k\|^2, \tag{13}$$

and compute a pair $(x_{k+1}, v_{k+1}) \in \mathbb{R}^n \times \mathbb{R}^n$ *such that*

$$v_{k+1} \in \partial^c \varphi_k(x_{k+1}), \tag{14}$$

$$\|v_{k+1}\| \le \sigma_k\,\lambda_k\,\|x_{k+1} - x_k\|, \qquad \varphi_k(x_{k+1}) \le \varphi_k(x_k). \tag{15}$$

Step 2. Iteration update: Set $k \leftarrow k + 1$ *and return to* **Step 1**.

We now make some remarks about Algorithm 1. For fixed $x_k \in \mathbb{R}^n$ and $\lambda_k > 0$, the regularized function φ_k defined in (13) need not be convex and may admit multiple critical points. The proximal subproblems are not constrained to any set $\mathcal{C}_i$. Indeed, to compute a pair (x_{k+1}, v_{k+1}) as in Step 1, one may apply an unconstrained local solver to minimize φ_k, typically initialized at x_k, and accepts its output when the stationarity inclusion in (14), together with the two inequalities in (15), are satisfied. This conditions can be seen as a residual-based *relative* stationarity criterion and a sufficient-decrease condition for φ_k. For instance, one may employ the nonsmooth descent method proposed in [2]. As will be shown below, the function φ_k is coercive, and since it is continuous, its sublevel set $S_k := \{x \in \mathbb{R}^n : \ \varphi_k(x) \le \varphi(x_k)\}$ is compact. Hence, noting that $\arg\min_{\mathbb{R}^n} \varphi_k \subseteq S_k$, we may prove that a Clarke stationary point x_{k+1} of this problem with residue $v_{k+1} := 0$ satisfy the inexact criterions (14)–(15). The last inequality in (15) together with the definition of φ_k in (13) imply that

$$\varphi_k(x_{k+1}) = f(x_{k+1}) + \tfrac{\lambda_k}{2}\,\|x_{k+1} - x_k\|^2 \le \varphi_k(x_k) = f(x_k). \tag{16}$$

Hence, the sequence of functional values $\{f(x_k)\}_{k\ge 0}$ is nonincreasing and for every iteration k such that $x_{k+1} \neq x_k$, we obtain, in particular, the strict decreasing inequality $f(x_{k+1}) < f(x_k)$. Finally, if $\sigma_k = 0$, then $v_{k+1} = 0$, in view of the first inequality in (15). Additionally, if f is convex on $\mathbb{R}^n$, then φ_k is strongly convex for every k, and there exists a unique point x_{k+1} satisfying the approximate stationarity conditions (14)–(15) (with $\sigma_k = 0$), which coincides with the unique minimizer of φ_k. Hence, Algorithm 1 can be interpreted as an inexact version of the classical proximal point method and reduces to its exact counterpart when $\sigma_k = 0$ for all $k \in \mathbb{N}$.

Since f is only PWSC and locally Lipschitz, the function φ_k may be nonconvex. In the following, our goal is to show that, for every $k \in \mathbb{N}$, the function φ_k is coercive and hence admits global minimizers. Consequently, the system (14)–(15) admits at least one solution pair (x_{k+1}, v_{k+1}), without additional assumptions on the sets $\{\mathcal{C}_i\}_{i\in I}$. Before proceeding to the proof of this result, we show that PWSC functions are bounded from below.

Lemma 4.1 *There exists $f^* \in \mathbb{R}$ such that $f(x) \geq f^*$ for every $x \in \mathbb{R}^n$.*

Proof By Definition 3.2, for each $i \in I$, the set $X_i^* = \arg\min_{x\in\mathcal{C}_i} f(x)$ is nonempty. Hence the minimum value of f on $\mathcal{C}_i$, $f_i^* := \min_{x\in\mathcal{C}_i} f(x)$, is finite for every $i \in I$. Since the family $\{\mathcal{C}_i\}_{i\in I}$ is finite, we can define $f^* := \min_{i\in I} f_i^* > -\infty$. Moreover, in view of the fact that $\bigcup_{i\in I} \mathcal{C}_i = \mathbb{R}^n$, every point $x \in \mathbb{R}^n$ belongs to at least one $\mathcal{C}_i$, and hence, for such an index i, we have $f(x) \geq f_i^* \geq f^*$, which proves the lemma. □

Next, we present some properties of φ_k which ensures, in particular, that Algorithm 1 is well defined.

Lemma 4.2 *For every $k \in \mathbb{N}$, the function φ_k, defined in* (13)*, is locally Lipschitz and coercive. In particular, the sublevel set $S_k = \{\, x \in \mathbb{R}^n : \varphi_k(x) \leq \varphi_k(x_k)\,\}$ is compact and $\arg\min_{x\in\mathbb{R}^n} \varphi_k(x) \neq \emptyset$. Consequently, there exists at least one approximate proximal stationary solution pair (x_{k+1}, v_{k+1}) satisfying* (14)–(15).

Proof The locally Lipschitz continuity of φ_k follows from the fact that it is the sum of two locally Lipschitz functions, namely, the function f, which is locally Lipschitz by assumption, and the function $(\lambda_k/2)\|\cdot - x_k\|^2 : \mathbb{R}^n \to \mathbb{R}$, which is smooth and therefore locally Lipschitz. Now, it follows from Lemma 4.1 that there exists $f^* \in \mathbb{R}$ such that $f(x) \geq f^*$ for all $x \in \mathbb{R}^n$. Hence, the definition of φ_k in (13) yields

$$\varphi_k(x) \geq f^* + \frac{\lambda_k}{2}\,\|x - x_k\|^2, \qquad \forall x \in \mathbb{R}^n.$$

Since $\lambda_k > 0$, the above inequality implies that $\lim_{\|x\|\to\infty} \varphi_k(x) = +\infty$, and hence the function φ_k is coercive. Thus the first statement of the lemma is proved. Since φ_k is locally Lipschitz, it is continuous. In particular, every sublevel set S_k is closed. Since the coercivity of φ_k implies that S_k is bounded, we conclude that S_k is compact, which together with the continuity of φ_k, implies that φ_k has a global minimizer on S_k, and hence on $\mathbb{R}^n$. Let $\bar{x}_k \in \operatorname{argmin}_{x\in\mathbb{R}^n} \varphi_k(x)$, and define $(x_{k+1}, v_{k+1}) := (\bar{x}_k, 0)$. Hence, from the optimality of $\bar{x}_k$, we have $v_{k+1} \in \partial\varphi_k^c(x_{k+1})$ and $\varphi_k(x_{k+1}) = \varphi(\bar{x}_k) \leq \varphi_k(x_k)$. Moreover, for any $\sigma_k \geq 0$, we have $\|v_{k+1}\| = 0 \leq \sigma_k\lambda_k\|x_{k+1} - x_k\|$. Therefore, we conclude that the pair (x_{k+1}, v_{k+1}) satisfies conditions (14)–(15), which in turn implies that Step 1 of Algorithm 1 is well defined for every iteration k. □

Note that if the local solver used to compute an approximate stationary solution for the k-th proximal supbroblem returns a point $x_{k+1} = x_k$, then the first inequality in (15) forces v_{k+1} to be null, which, in view of (14), implies that $0 \in \partial^c\varphi_k(x_k)$. Hence, the iterate x_k turns out to be a Clarke stationary point of the regularized function φ_k. Moreover, since

$$\partial^c\varphi_k(x) = \partial^c f(x) + \lambda_k(x - x_k), \tag{17}$$

it follows that $\partial^c\varphi_k(x_k) = \partial^c f(x_k)$, and hence we conclude that if $x_{k+1} = x_k$, then x_k is also a Clarke stationary point of f. If, in addition, x_k lies in the interior of some set $\mathcal{C}_i$, Corollary 3.1(b) implies that x_k is a local minimizer of f. Therefore, the condition $x_{k+1} = x_k$ is a simple and useful stopping criterion for Algorithm 1. To analyze the asymptotic convergence properties of Algorithm 1, we henceforth assume that it generates an infinite sequence $\{x_k\}_{k\in\mathbb{N}}$ such that $x_{k+1} \neq x_k$ for all k.

4.1 Asymptotic convergence analysis

In this section, we establish the asymptotic convergence properties of the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by Algorithm 1. We first show that the acceptance conditions in (15) enforce a sufficient decrease of the objective, which implies convergence of the values $\{f(x_k)\}_{k\in\mathbb{N}}$ and, under uniform bounds on $\{\lambda_k\}_{k\in\mathbb{N}}$, asymptotic regularity and Clarke stationarity of any accumulation point. We then exploit the PWSC structure on the subsets $\{\mathcal{C}_i\}_{i\in I}$ to derive a three–point inequality on each region, which yields refined consequences, namely, accumulation points are minimizers of the restrictions $f|_{\mathcal{C}_i}$ on the subsets containing them and, under a mild summability condition, the method eventually remains in a single subset and the whole sequence converges. We close the section with a simple level–boundedness assumption ensuring the existence of accumulation points.

We start with the basic properties of Algorithm 1 without assuming any stringent conditions.

Theorem 4.1 *Let $\{x_k\}_{k\in\mathbb{N}}$ be a sequence generated by Algorithm 1. Then, the following statements hold:*

a) The sequence $\{f(x_k)\}_{k\in\mathbb{N}}$ is monotonically decreasing and converges to some scalar $\bar{f}$. Consequently, if $\bar{x}$ is an accumulation point of $\{x_k\}_{k\in\mathbb{N}}$, then $f(\bar{x}) = \bar{f}$.

b) for every $k \in \mathbb{N}$, we have

$$u_{k+1} := v_{k+1} + \lambda_k(x_k - x_{k+1}) \in \partial^c f(x_{k+1}), \quad \|u_{k+1}\| \leq (\sigma_k + 1)\lambda_k\|x_{k+1} - x_k\|.$$

c) If there exists $\bar{\lambda} > 0$ such that $\lambda_k \geq \bar{\lambda}$ for all $k \in \mathbb{N}$, then

$$\sum_{k=0}^{+\infty} \|x_{k+1} - x_k\|^2 < \frac{2}{\bar{\lambda}}\left(f(x_0) - \bar{f}\right).$$

In particular, the sequence $\{\|x_{k+1} - x_k\|\}_{k\in\mathbb{N}}$ converges to zero.

d) If, in addition to the assumption in (c), there exists $\hat{\lambda} > 0$ such that $\lambda_k \leq \hat{\lambda}$ for all $k \in \mathbb{N}$, then every accumulation point of $\{x_k\}_{k\in\mathbb{N}}$, if any, is a Clarke stationary point of f.

Proof (a) It follows from (16), the assumption that $x_{k+1} \neq x_k$, and Lemma 4.1 that

$$f^* \leq f(x_{k+1}) < f(x_{k+1}) + \frac{\lambda_k}{2}\|x_{k+1} - x_k\|^2 \leq f(x_k), \qquad \forall k \in \mathbb{N}. \tag{18}$$

Hence $\{f(x_k)\}_{k\in\mathbb{N}}$ is monotonically decreasing and bounded below, which implies that it converges to some $\bar{f} \in \mathbb{R}$. Thus, the first statement of (a) follows. The last statement of (a) follows immediately from the first one and the continuity of f.

(b) It follows immediately from (14) and (17) that

$$u_{k+1} := v_{k+1} + \lambda_k(x_k - x_{k+1}) \in \partial^c f(x_{k+1}), \qquad \forall k \in \mathbb{N}.$$

Using he first inequality in (15) and the Cauchy-Schwarz inequality, we obtain

$$\|u_{k+1}\| \leq \|v_{k+1}\| + \lambda_k\|x_{k+1} - x_k\| \leq (\sigma_k + 1)\lambda_k\|x_{k+1} - x_k\|,$$

and hence the proof of (b) follows.

(c) Since $\lambda_k \geq \bar{\lambda} > 0$ and $\{f(x_k)\}_{k\in\mathbb{N}}$ is monotonically decreasing and converges to $\bar{f}$, it follows from the last inequality in (18) that, for every $j \in \mathbb{N}$,

$$\frac{\bar{\lambda}}{2}\sum_{k=0}^{j}\|x_{k+1}-x_k\|^2 \leq \sum_{k=0}^{j}\frac{\lambda_k}{2}\|x_{k+1}-x_k\|^2 \leq \sum_{k=0}^{j}(f(x_k)-f(x_{k+1})) \leq f(x_0)-\bar{f}.$$

Therefore, the first statement in (c) follows from the preceding inequalities, and the second follows immediately.

(d) Le $\bar{x}$ be an accumulation point of $\{x_k\}_{k\in\mathbb{N}}$, and let $\{x_{k_\ell}\}_{\ell\in\mathbb{N}}$ be a subsequence converging to $\bar{x}$. From the Cauchy-Schwarz inequality, we have

$$\|x_{k_\ell+1}-\bar{x}\| \leq \|x_{k_\ell+1}-x_{k_\ell}\| + \|x_{k_\ell}-\bar{x}\|.$$

Hence, in view of the last statement in (c), the above inequality, and the convergence of $\{x_{k_\ell}\}_{\ell\in\mathbb{N}}$ to $\bar{x}$, we conclude that $\{x_{k_\ell+1}\}_{\ell\in\mathbb{N}}$ also converges to $\bar{x}$. Since $\lim_{k\to\infty}\|x_{k+1}-x_k\| = 0$, it follows from the inequality in (b) and the bounds $\lambda_k \leq \hat{\lambda}$ and $\sigma_k \leq 1$ that $\lim_{k\to\infty} u_{k+1} = 0$. Thus, as $\lim_{\ell\to\infty} x_{k_\ell+1} = \bar{x}$, and $u_{k+1} \in \partial^c f(x_{k+1})$ for all $k \in \mathbb{N}$, by the inclusion in (b), the closedness of the graph of $\partial^c f$ (see Proposition 2.1) yields $0 \in \partial^c f(\bar{x})$. Therefore, $\bar{x}$ is a Clarke stationary point of f, completing the proof of (d). □

In the next corollary, we recall a simple condition that ensures the existence of an accumulation point for the proximal sequence.

Corollary 4.1 *Let x_0, $\{\lambda_k\}_{k\in\mathbb{N}}$, and $\{\sigma_k\}_{k\in\mathbb{N}}$ be the inputs of Algorithm 1, and assume that the following conditions hold: i) $\mathcal{L}_{f(x_0)} := \{x \in \mathbb{R}^n : \ f(x) \leq f(x_0)\}$ is bounded; ii) $\{\lambda_k\}_{k\in\mathbb{N}} \subset [\bar{\lambda}, \hat{\lambda}]$ for some $\hat{\lambda} \geq \bar{\lambda} > 0$. Then, the sequence $\{x_k\}_{k\in\mathbb{N}}$ is bounded, and every accumulation point is a Clarke stationary point of f.*

Proof Since the sequence $\{f(x_k)\}_{k\in\mathbb{N}}$ is nonincreasing, we have $f(x_k) \leq f(x_0)$ for all $k \in \mathbb{N}$. Consequently, $x_k \in \mathcal{L}_{f(x_0)}$ for every $k \in \mathbb{N}$. By assumption, the level set $\mathcal{L}_{f(x_0)}$ is bounded; hence, the sequence $\{x_k\}_{k\in\mathbb{N}}$ is bounded and thus admits at least one accumulation point. Therefore, the desired result follows from Theorem 4.1(d), in view of the latter conclusion and the assumptions of the lemma. □

Next, we incorporate the piecewise star–convex geometry by establishing a three-point inequality on each $\mathcal{C}_i$. This estimate is the main tool for the remainder of the section, as it allows us to relate the decrease mechanism of the inexact proximal steps to the setwise minimizers X_i^* and to derive identification and full convergence results.

Lemma 4.3 *Let (x_{k+1}, v_{k+1}) be a pair satisfying the inexact proximal criteria* (14)–(15)*, and assume that there exists $i \in I$ such that $x_{k+1} \in \operatorname{int}\mathcal{C}_i$. Then, for every $x_i^* \in X_i^*$, the following inequality holds:*

$$\big(1-\sigma_k\big)\|x_i^*-x_{k+1}\|^2 \leq \|x_i^*-x_k\|^2 - (1-\sigma_k)\,\|x_{k+1}-x_k\|^2 + \tfrac{2}{\lambda_k}\big(f(x_i^*)-f(x_{k+1})\big). \qquad (19)$$

If additionally, f is μ_i-strongly star-convex on $\mathcal{C}_i$, then the term on the left-hand side of (19) *becomes $\big(1-\sigma_k+\mu_i/\lambda_k\big)\|x_i^*-x_{k+1}\|^2$.*

Proof Let $i \in I$ be such that $x_{k+1} \in \operatorname{int} \mathcal{C}_i$ and fix $x_i^* \in X_i^*$. We prove the inequality in the strongly star-convex case. The other case follows by the same argument with $\mu_i = 0$. Since $x_{k+1} \in \operatorname{int} \mathcal{C}_i$ and f is μ_i-strongly star-convex on $\mathcal{C}_i$, by Proposition 3.1 we have

$$u_{k+1}^T(x_i^* - x_{k+1}) \leq f(x_i^*) - f(x_{k+1}) - \frac{\mu_i}{2}\|x_{k+1} - x_i^*\|^2, \qquad \forall\, u_{k+1} \in \partial^c f(x_{k+1}).$$

In view of inclusions (14) and (17), we have that $u_{k+1} := v_{k+1} + \lambda_k(x_k - x_{k+1}) \in \partial^c f(x_{k+1})$. Thus, substituting this element into the above inequality, we obtain

$$\lambda_k(x_k - x_{k+1})^T(x_i^* - x_{k+1}) \leq f(x_i^*) - f(x_{k+1}) - v_{k+1}^T(x_i^* - x_{k+1}) - \tfrac{\mu_i}{2}\|x_{k+1} - x_i^*\|^2.$$

Using the identity $2(x_k - x_{k+1})^T(x_i^* - x_{k+1}) = \|x_i^* - x_{k+1}\|^2 + \|x_{k+1} - x_k\|^2 - \|x_i^* - x_k\|^2$, and the previous inequality, a simple rearrangement yields that

$$\big(1+\tfrac{\mu_i}{\lambda_k}\big)\|x_i^*-x_{k+1}\|^2 \leq \|x_i^*-x_k\|^2-\|x_{k+1}-x_k\|^2+\tfrac{2}{\lambda_k}\big(f(x_i^*)-f(x_{k+1})\big)-\tfrac{2}{\lambda_k}v_{k+1}^T(x_i^*-x_{k+1}). \quad (20)$$

On the other hand, from the Cauchy–Schwarz inequality and the first inequality in (15), we have

$$-\tfrac{2}{\lambda_k}v_{k+1}^T(x_i^* - x_{k+1}) \leq \tfrac{2}{\lambda_k}\|v_{k+1}\|\,\|x_i^* - x_{k+1}\| \leq 2\sigma_k\,\|x_{k+1} - x_k\|\,\|x_i^* - x_{k+1}\|.$$

Using the inequality $2ab \leq a^2 + b^2$ with $a = \|x_i^* - x_{k+1}\|$ and $b = \|x_{k+1} - x_k\|$, we obtain

$$-\tfrac{2}{\lambda_k}v_{k+1}^T(x_i^* - x_{k+1}) \leq \sigma_k\big(\|x_i^* - x_{k+1}\|^2 + \|x_{k+1} - x_k\|^2\big).$$

Therefore, the version of (19) in the strong case follows by substituting the above estimate into (20) and rearranging terms. □

As a first consequence of Lemma 4.3, we show that under a mild condition on the initial parameters of Algorithm 1, the entire sequence converges to a stationary solution.

Theorem 4.2 *Let $\bar{x}$ be an accumulation point of $\{x_k\}_{k\in\mathbb{N}}$ and assume that $\{\lambda_k\}_{k\in\mathbb{N}} \subset [\bar{\lambda}, \hat{\lambda}]$ for some $\hat{\lambda} \geq \bar{\lambda} > 0$. Then the following statements hold:*

a) If $\bar{x} \in \operatorname{int} \mathcal{C}_i$ for some $i \in I$, then $\bar{x} \in X_i^*$. Moreover, $\bar{x}$ is a local minimizer of f in $\mathbb{R}^n$.
b) Assume that $\bar{x} \in \operatorname{int} \mathcal{C}_{i_0}$ for some $i_0 \in I$, and that

$$\sum_{k=0}^{\infty} \frac{\sigma_k}{1 - \sigma_k} < +\infty.$$

Then, there exists $k_0 \in \mathbb{N}$ such that $x_k \in \operatorname{int} \mathcal{C}_{i_0}$ for all $k \geq k_0$. Moreover, the entire sequence $\{x_k\}_{k\in\mathbb{N}}$ converges to $\bar{x}$.

Proof It follows from Theorem 4.1(d) that $\bar{x}$ is a Clarke stationary point of f. Hence, for every $i \in I$ such that $\bar{x} \in \operatorname{int} \mathcal{C}_i$, Corollary 3.1(a) and (b) imply that $\bar{x} \in X_i^*$ and that $\bar{x}$ is a local minimizer of f in $\mathbb{R}^n$. This proves item (a).

(b) Let $i_0 \in I$ such that $\bar{x} \in \operatorname{int} \mathcal{C}_{i_0}$. Define

$$\delta := \operatorname{dist}\big(\bar{x},\ \mathbb{R}^n \setminus \mathcal{C}_{i_0}\big) \ \in\ (0, +\infty], \quad (21)$$

and fix an arbitrary $\varepsilon \in (0, \delta/4)$. In view of Theorem 4.1(c), we have $\lim_{k\to+\infty} \|x_{k+1} - x_k\| = 0$, and then there exists $\hat{k} \in \mathbb{N}$ such that

$$\|x_{k+1} - x_k\| \leq \varepsilon, \qquad \forall k \geq \hat{k}. \quad (22)$$

Now note that, for every $j \in \mathbb{N}$, we have

$$\ln\Big(\prod_{k=0}^{j} \frac{1}{1-\sigma_k}\Big) = \sum_{k=0}^{j} \ln\Big(\frac{1}{1-\sigma_k}\Big) = \sum_{k=0}^{j} \ln\Big(1+\frac{\sigma_k}{1-\sigma_k}\Big) \le \sum_{k=0}^{+\infty} \ln\Big(1+\frac{\sigma_k}{1-\sigma_k}\Big) < +\infty,$$

where the last inequality is due to the assumption $\sum_{k=0}^{+\infty} \sigma_k(1-\sigma_k)^{-1} < +\infty$. Hence, we conclude that the product $\prod_{k=0}^{+\infty}(1-\sigma_k)^{-1}$ is finite. Thus, there exists $\bar{k} \in \mathbb{N}$ such that

$$\prod_{k=\bar{k}}^{+\infty} \frac{1}{1-\sigma_k} \le 4. \tag{23}$$

Since $\lim_{\ell\to\infty} x_{k_\ell} = \bar{x}$, choose ℓ_0 sufficiently large such that $k_{\ell_0} \ge \max\{\hat{k}, \bar{k}\}$ and $\|x_{k_{\ell_0}} - \bar{x}\| \le \varepsilon/2$. We will prove by induction that

$$\|\bar{x} - x_{k_{\ell_0}+n}\| \le \varepsilon \quad \text{and} \quad x_{k_{\ell_0}+n} \in \operatorname{int} \mathcal{C}_{i_0}, \qquad \forall n \in \mathbb{N}. \tag{24}$$

This is true for $n = 0$ by construction. Suppose it holds for some $n \ge 0$ and to simplify the notations set $s := k_{\ell_0} + n$. Then

$$\|\bar{x} - x_{s+1}\| \le \|\bar{x} - x_s\| + \|x_{s+1} - x_s\| \le \varepsilon + \varepsilon = 2\varepsilon < \delta,$$

By (21), this implies that $x_{s+1} \in \operatorname{int} \mathcal{C}_{i_0}$. Thus, we conclude by induction that $x_k \in \operatorname{int} \mathcal{C}_{i_0}$ for every $k \ge k_{\ell_0}$, which proves the first part of item (b) by taking $k_0 := k_{\ell_0}$. On the other hand, since $\bar{x} \in \operatorname{int} \mathcal{C}_{i_0}$, it follows from (a) that $\bar{x} \in X^*_{i_0}$. Hence Lemma 4.3 with $x^*_i = \bar{x}$, $i = i_0$ and $k \ge k_{\ell_0}$, combined with the fact that $\mu_{i_0} \ge 0$, imply that

$$(1-\sigma_k)\,\|\bar{x} - x_{k+1}\|^2 \le \|\bar{x} - x_k\|^2 - (1-\sigma_k)\,\|x_{k+1} - x_k\|^2 + \tfrac{2}{\lambda_k}\big(f(\bar{x}) - f(x_{k+1})\big). \tag{25}$$

By (a) of Theorem 4.1 and the monotonicity $\{f(x_k)\}_{k\in\mathbb{N}}$, we obtain that $f(\bar{x}) \le f(x_{k+1})$. Since $\lambda_k > 0$ and $\sigma_k < 1$, it follows from (25) that

$$\|\bar{x} - x_{k+1}\|^2 \le \frac{1}{1-\sigma_k}\,\|\bar{x} - x_k\|^2, \qquad \forall k \ge k_{\ell_0}.$$

Using the above inequality recursively for $k = k_{\ell_0}, \ldots, s$, we obtain

$$\|\bar{x} - x_{s+1}\|^2 \le \|\bar{x} - x_{k_{\ell_0}}\|^2 \prod_{t=k_{\ell_0}}^{s} \frac{1}{1-\sigma_t} \le \frac{\varepsilon^2}{4} \prod_{t=\bar{k}}^{+\infty} \frac{1}{1-\sigma_t} \le \varepsilon^2,$$

where the last two inequalities follow from $\|x_{k_{\ell_0}} - \bar{x}\| \le \varepsilon/2$ and (23), respectively. Since $s = k_{\ell_0} + n$, we obtain $\|\bar{x} - x_{k_{\ell_0}+1}\| \le \varepsilon$, which establishes the inequality in (24) for $n+1$. This completes the induction argument. Moreover, since $\varepsilon \in (0, \delta/4)$ is arbitrary, the inequality in (24) implies that $\lim_{k\to\infty} x_k = \bar{x}$, concluding the proof of the theorem. □

4.2 Iteration complexity analysis

In this section, we analyze the iteration complexity and convergence rate of Algorithm 1 under some mild assumptions. First, we show that once the sequence $\{x_k\}_{k\in\mathbb{N}}$ enters a set $\mathcal{C}_i$, associate to its limit point $\bar{x}$, then the sequence of functional values $\{f(x_k)\}_{k\in\mathbb{N}}$ converges sublinearly. This result is then used to derive an iteration-complexity bound to obtain a point whose functional value approximates the optimal value of f in the set $\mathcal{C}_i$. Second, we introduce a concept of approximate stationary solution and establish the iteration complexity of Algorithm 1 to compute such a solution. Finally, we end the section by showing that $\{x_k\}_{k\in\mathbb{N}}$ converges linearly if the objective function is strongly PWSC (a notion defined below) and the initial parameters pair (σ_k, λ_k) of Algorithm 1 is suitably chosen depending on the strong convexity parameter.

Theorem 4.3 *Let $\bar{x}$ be an accumulation point of the sequence $\{x_k\}_{k\in\mathbb{N}}$. Assume that $\lambda_k \leq \hat{\lambda}$ for all $k \in \mathbb{N}$, and that there exist $k_0 \in \mathbb{N}$ and $i \in I$ such that $x_k \in \operatorname{int}\mathcal{C}_i$ for all $k \geq k_0$. Then, for every integer $N > k_0$,*

$$f(x_N) - f(\bar{x}) \;\leq\; \frac{\hat{\lambda}}{2\,(N-k_0)\omega_N}\,\|\bar{x} - x_{k_0}\|^2, \tag{26}$$

where ω_N is defined by (1). Moreover, if $\sum_{k=k_0}^{+\infty} \frac{\sigma_k}{1-\sigma_k} < +\infty$, then

$$f(x_N) - f(\bar{x}) \;\leq\; \frac{\hat{\lambda}}{2\omega(N-k_0)}\,\|\bar{x} - x_{k_0}\|^2, \qquad \forall\, N > k_0, \tag{27}$$

where $0 < \omega = \lim_{k\to\infty}\omega_k$. In particular, one has $f(x_N) - f(\bar{x}) = \mathcal{O}(1/N)$ as N goes to $+\infty$.

Proof Since $x_{k+1} \in \operatorname{int}\mathcal{C}_i$ for all $k \geq k_0$, it follows from Lemma 4.3 that, for every $k \geq k_0$,

$$(1-\sigma_k)\,\|\bar{x} - x_{k+1}\|^2 \leq \|\bar{x} - x_k\|^2 - (1-\sigma_k)\,\|x_{k+1} - x_k\|^2 + \frac{2}{\lambda_k}\big(f(\bar{x}) - f(x_{k+1})\big).$$

Since $\sigma_k \in [0,1)$, the last inequality yields

$$\frac{2}{\lambda_k}\big(f(x_{k+1}) - f(\bar{x})\big) \leq \|\bar{x} - x_k\|^2 - (1-\sigma_k)\,\|\bar{x} - x_{k+1}\|^2, \qquad \forall k \geq k_0.$$

Define $\omega_{k_0} = 1$ and $\omega_{k+1} = \omega_k(1-\sigma_k)$ for all $k > k_0$. Thus, multiplying the last inequality by ω_k gives

$$\frac{2\,\omega_k}{\lambda_k}\big(f(x_{k+1}) - f(\bar{x})\big) \leq \omega_k\|\bar{x} - x_k\|^2 - \omega_{k+1}\|\bar{x} - x_{k+1}\|^2, \qquad \forall k \geq k_0.$$

Summing both sides of the above inequality from $k = k_0$ to $k = N-1$ and using that $\omega_{k_0} = 1$ and $\omega_N > 0$, we obtain

$$\sum_{k=k_0}^{N-1} \frac{2\,\omega_k}{\lambda_k}\big(f(x_{k+1}) - f(\bar{x})\big) \leq \|\bar{x} - x_{k_0}\|^2 - \omega_N\|\bar{x} - x_N\|^2 \leq \|\bar{x} - x_{k_0}\|^2. \tag{28}$$

Since $f(x_{k+1}) \leq f(x_k)$ for all k, we conclude that

$$f(x_{k+1}) - f(\bar{x}) \geq f(x_N) - f(\bar{x}), \qquad \forall\, k_0 \leq k \leq N-1.$$

Using $\lambda_k \leq \hat{\lambda}$ and (28), we obtain

$$\frac{2}{\hat{\lambda}}\big(f(x_N) - f(\bar{x})\big) \sum_{k=k_0}^{N-1} \omega_k \leq \sum_{k=k_0}^{N-1} \frac{2\,\omega_k}{\lambda_k}\big(f(x_{k+1}) - f(\bar{x})\big) \leq \|\bar{x} - x_{k_0}\|^2.$$

Since $\omega_{k+1} = \omega_k(1-\sigma_k) \leq \omega_k$, the sequence $\{\omega_k\}_{k\geq k_0}$ is nonincreasing, and thus $\omega_k \geq \omega_N$ for all $k_0 \leq k \leq N-1$. Therefore $\sum_{k=k_0}^{N-1} \omega_k \geq (N-k_0)\omega_N$, and consequently

$$f(x_N) - f(\bar{x}) \leq \frac{\hat{\lambda}}{2\,(N-k_0)\,\omega_N}\,\|\bar{x} - x_{k_0}\|^2,$$

which proves (26).

Assume now that $\sum_{k=k_0}^{+\infty} \frac{\sigma_k}{1-\sigma_k} < +\infty$. By Lemma 2.1, the sequence $\{\omega_k\}_{k\geq k_0}$ converges to a finite limit ω satisfying $\omega > 0$. Since $\{\omega_k\}_{k\geq k_0}$ is nonincreasing, one has $\omega_N \geq \omega$ for all $N > k_0$. Substituting this bound into (26) yields (27). □

As a consequence of (27), we obtain an explicit bound of $\mathcal{O}(1/\varepsilon)$ on the number of iterations needed to reach a prescribed accuracy level on the region $\mathcal{C}_{i^*}$.

Corollary 4.2 *Under the assumptions of Theorem 4.3, let $\varepsilon > 0$ and $R := \|\bar{x} - x_{k_0}\|$. Then, for any integer $N \geq k_0 + \frac{\hat{\lambda}R^2}{2\omega\varepsilon}$, we have $f(x_N) - f(\bar{x}) \leq \varepsilon$.*

Now, let us introduce a concept of approximate stationary solution and analyze the iteration complexity of Algorithm 1 for computing such a solution. First, recall that $\bar{x}$ is a stationary point of f if and only if $0 \in \partial^c f(\bar{x})$. Hence, given a tolerance $\varepsilon > 0$, a point x is said to be an ε-approximate stationary solution of f if there exists $u \in \partial^c f(x)$ such that $|u| \leq \varepsilon$.

In the following, we show that Algorithm 1 computes an ε-approximate stationary solution in at most $\mathcal{O}(1/\varepsilon^2)$ iterations.

Proposition 4.1 *Let $\{x_k\}_{k\in\mathbb{N}}$ be generated by Algorithm 1 and assume that $\{\lambda_k\}_{k\in\mathbb{N}} \subset [\bar{\lambda}, \hat{\lambda}]$ for some scalars $\hat{\lambda} \geq \bar{\lambda} > 0$. Then, given a tolerance $\varepsilon > 0$, Algorithm 1 provides an ε-approximate stationary solution $x := x_{N+1}$ in at most $N = \mathcal{O}(1/\varepsilon^2)$ iterations.*

Proof From Theorem 4.1(b) and the assumptions $\lambda_k \leq \hat{\lambda}$ and $\sigma_k \leq 1$, we obtain

$$u_{k+1} = v_{k+1} + \lambda_k(x_k - x_{k+1}) \in \partial^c f(x_{k+1}), \qquad \|u_{k+1}\| \leq 2\hat{\lambda}\|x_k - x_{k+1}\|. \tag{29}$$

From Theorem 4.1(c), we have $\sum_{k=0}^{j} \|x_k - x_{k+1}\|^2 \leq 2(f(x_0) - \bar{f})/\bar{\lambda}$, for all $j \in \mathbb{N}$. Combining the latter inequality with the one in (29), we deduce that $\sum_{k=0}^{j} \|u_{k+1}\|^2 \leq 8\hat{\lambda}^2(f(x_0) - \bar{f})/\bar{\lambda}$, for all $j \in \mathbb{N}$. Thus, if $\{u_{k+1}\}_{k=0,\cdots,K-1}$ is not ε-approximate stacionary, i,e, $\|u_{k+1}\| > \varepsilon$, then

$$K\varepsilon^2 < \sum_{k=0}^{K-1} \|u_{k+1}\|^2 \leq \frac{8\hat{\lambda}^2}{\bar{\lambda}}(f(x_0) - \bar{f}),$$

which implies that $K < 8\hat{\lambda}^2(f(x_0) - \bar{f})/(\bar{\lambda}\varepsilon^2)$. Therefore, if $N = \lceil 8\hat{\lambda}^2(f(x_0) - \bar{f})/(\bar{\lambda}\varepsilon^2)\rceil$, then we must necessarily have $\|u_{N+1}\| \leq \varepsilon$. Since $u_{N+1} \in \partial^c f(x_{N+1})$, in view of (29), the proof of the proposition follows. □

To conclude this section, we show that if the objective function f is strongly PWSC, then the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by Algorithm 1, with suitably chosen initial parameters, converges linearly once it enters the interior of a set $\mathcal{C}_i$.

Proposition 4.2 *Assume that, for every $i \in I$, f is μ_i-strongly star-convex on $\mathcal{C}_i$, and define $\bar{\mu} := \min_{i\in I} \mu_i > 0$. Let $\{x_k\}_{k\in\mathbb{N}}$ be generated by Algorithm 1, where the sequence of initial parameters $\{(\lambda_k, \sigma_k)\}_{k\in\mathbb{N}}$ satisfies $\lambda_k(\sigma_k + \bar{\sigma}) \leq \bar{\mu}$ for some $\bar{\sigma} > 0$ and for all $k \in \mathbb{N}$. Assume also that $\{x_k\}_{k\in\mathbb{N}}$ admits an accumulation point $\bar{x}$ belonging to $\operatorname{int}\mathcal{C}_i$ for some $i \in I$. Let $k_0 \in \mathbb{N}$ be such that $x_k \in \operatorname{int}\mathcal{C}_i$ for all $k \geq k_0$. Then the following inequalities hold*

$$\|x_{k+1} - \bar{x}\|^2 \leq \frac{1}{1+\bar{\sigma}}\|x_k - \bar{x}\|^2 \tag{30}$$

$$f(x_{k+1}) - f(\bar{x}) \leq \frac{1}{\sigma_k + \bar{\sigma}}\left[f(x_k) - f(\bar{x})\right], \qquad \forall k \geq k_0. \tag{31}$$

Proof Fix $i \in I$ and $k_0 \in \mathbb{N}$ as in the statement, and let $N > k_0 + 1$. Since $x_{k+1} \in \operatorname{int}\mathcal{C}_i$ for all $k \geq k_0$, it follows from Lemma 4.3 and the fact that, for every $k \geq k_0$,

$$\left(1 - \sigma_k + \frac{\mu_i}{\lambda_k}\right)\|\bar{x} - x_{k+1}\|^2 \leq \|\bar{x} - x_k\|^2 - (1-\sigma_k)\|x_{k+1} - x_k\|^2 + \frac{2}{\lambda_k}\big(f(\bar{x}) - f(x_{k+1})\big).$$

Since $\sigma_k \in [0, 1)$ and $f(\bar{x}) \leq f(x_{k+1})$ the last inequality yields

$$\left(1 - \sigma_k + \frac{\mu_i}{\lambda_k}\right)\|\bar{x} - x_{k+1}\|^2 \leq \|\bar{x} - x_k\|^2, \qquad \forall k \geq k_0. \tag{32}$$

Now, note that the assumption $\lambda_k(\sigma_k + \bar{\sigma}) \leq \bar{\mu}$ together with the fact that $\bar{\mu} = \min_{i\in I}\mu_i$ imply that $1+\bar{\sigma} \leq 1-\sigma_k+\mu_i/\lambda_k$. Therefore, (32) yields $(1+\bar{\sigma})\|\bar{x}-x_{k+1}\|^2 \leq \|\bar{x}-x_k\|^2$, for all $k \geq k_0$, which proves (30).

Now, Since $0 < 1 + \bar{\sigma} \leq 1 - \sigma_k + \mu_i/\lambda_k$, it follows from Lemma 4.3 that

$$f(x_{k+1}) - f(\bar{x}) \leq \frac{\lambda_k}{2}\|x_k - \bar{x}\|^2.$$

On the other hand, from (10) with $x = x_k$ and the fact that $\mu_i \geq \bar{\mu}$, we have

$$\|x_k - \bar{x}\|^2 \leq \frac{2}{\bar{\mu}}\left(f(x_k) - f(\bar{x})\right).$$

Thus, combining both inequalities, we arrive at

$$f(x_{k+1}) - f(\bar{x}) \leq \frac{\lambda_k}{\bar{\mu}}\left(f(x_k) - f(\bar{x})\right),$$

which proves (31), in view of the assumption $\lambda_k(\sigma_k + \bar{\sigma}) \leq \bar{\mu}$. □

Note that, under the assumptions of Proposition 4.2, if the scalar $\bar{\sigma}$ is larger than 1, then (31) implies that the sequence of functional values $\{f(x_k)\}_{k\geq k_0}$ is linearly convergent to $f(\bar{x})$.

5 Aplications

In this section we discuss several concrete optimization problems whose objective functions are piecewise star–convex in the sense of Definition 3.2. For each model we explicitly construct a convex partition of $\mathbb{R}^n$ and identify the corresponding family of local minimizers, thereby verifying the PWSC structure and the well–posedness of the proximal subproblems. This shows how the inexact proximal point method can be applied to nonconvex problems whose objective is locally convex on each region, while still enjoying the convergence and complexity properties established in Section 4. The constructions presented below are mathematically based on the

PWSC examples studied in [5, Section 3] and are reformulated here in a self–contained manner so as to highlight the features that are relevant for the proximal point framework developed in this paper.

We begin with a contrastive learning model. In many contrastive and metric learning formulations, one seeks a representation $x \in \mathbb{R}^n$ whose images under suitable feature maps are close to "positive" reference examples and far from "negative" ones. Typical loss functions combine a squared Euclidean term that attracts x toward positive samples with nonsmooth or robust penalties that discourage proximity to negative samples; see, for instance, the survey on contrastive self–supervised learning in [10] and robust distance–metric learning formulations with ℓ_1–type distances such as [17]. The next example captures this structure in a finite–dimensional quadratic/ℓ_1 setting that fits into our PWSC framework. For later use, and to simplify notation, we set

$$\{-1,1\}^\ell := \big\{\, s = (s_1,\ldots,s_\ell) \in \mathbb{R}^\ell :\ s_i \in \{-1,1\} \text{ for all } i \in \{1,\ldots,\ell\} \,\big\}. \tag{33}$$

Example 5.1 Let $m, r \in \mathbb{N}$ and, for each $i = 1,\ldots,m$, let $B_i \in \mathbb{R}^{d_i \times n}$, $\bar{p}_i \in \mathbb{R}^{d_i}$ and $\lambda_i^+ > 0$ be given. For each $j = 1,\ldots,r$, let $A_j \in \mathbb{R}^{m_j \times n}$, $\bar{q}_j \in \mathbb{R}^{m_j}$ and $\lambda_j^- > 0$ be given. Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by

$$f(x) := \sum_{i=1}^{m} \lambda_i^+ \big\|B_i x - \bar{p}_i\big\|_2^2 - \sum_{j=1}^{r} \lambda_j^- \big\|A_j x - \bar{q}_j\big\|_1, \qquad x \in \mathbb{R}^n.$$

Note that f is the difference of two convex functions on $\mathbb{R}^n$. Moreover, each convex function is locally Lipschitz on $\mathbb{R}^n$, and the difference of two locally Lipschitz functions is locally Lipschitz. Therefore, f is locally Lipschitz on $\mathbb{R}^n$.

Define the matrix $H := \sum_{i=1}^m \lambda_i^+ B_i^\top B_i$. We assume that H is positive definite. Then the first term is a smooth quadratic with Hessian $2H \succ 0$, and the second term is convex and piecewise linear. To describe the PWSC structure, we first make the notation explicit. For each $j \in \{1,\ldots,r\}$ and $x \in \mathbb{R}^n$, we write

$$A_j x - \bar{q}_j \;=\; \big((A_j x - \bar{q}_j)_k\big)_{k=1}^{m_j} \in \mathbb{R}^{m_j},$$

where, for each $k = 1,\ldots,m_j$, the scalar $(A_j x - \bar{q}_j)_k$ denotes the k-th component of the vector $A_j x - \bar{q}_j$. For each multi–sign pattern

$$S = (s_{j,k})_{1 \le j \le r,\, 1 \le k \le m_j} \in \{-1,1\}^{\sum_{j=1}^r m_j},$$

define the closed convex set

$$\mathcal{C}_S := \Big\{\, x \in \mathbb{R}^n : s_{j,k}\big((A_j x - \bar{q}_j)_k\big) \ge 0,\ j = 1,\ldots,r,\ k = 1,\ldots,m_j \Big\}.$$

Since S ranges over the finite set $\{-1,1\}^{\sum_{j=1}^r m_j}$, the family $\{\mathcal{C}_S\}$ is finite. Each $\mathcal{C}_S$ is an intersection of closed halfspaces of the form $s_{j,k}((A_j x - \bar{q}_j)_k) \ge 0$. Hence $\mathcal{C}_S$ is closed and convex. Moreover, if $S \ne S'$, there exists at least one pair (j,k) such that $s_{j,k} \ne s'_{j,k}$. For any x in the interior of $\mathcal{C}_S$ we have the strict inequalities $s_{j,k}(A_j x - \bar{q}_j)_k > 0$ for all (j,k), while for any y in the interior of $\mathcal{C}_{S'}$ we have $s'_{j,k}(A_j y - \bar{q}_j)_k > 0$ for all (j,k). When $s'_{j,k} = -s_{j,k}$, these two strict inequalities are incompatible. Thus, we have

$$\operatorname{int}(\mathcal{C}_S) \cap \operatorname{int}(\mathcal{C}_{S'}) = \emptyset \quad \text{whenever } S \ne S'.$$

Finally, given any $x \in \mathbb{R}^n$, define a sign pattern $S(x)$ by

$$s_{j,k}(x) := \begin{cases} 1, & \text{if } (A_j x - \bar{q}_j)_k \ge 0, \\ -1, & \text{if } (A_j x - \bar{q}_j)_k < 0, \end{cases}$$

for all $j = 1, \dots, r$ and $k = 1, \dots, m_j$. By construction, x satisfies $s_{j,k}(x)(A_j x - \bar{q}_j)_k \geq 0$ for every (j,k). Hence $x \in \mathcal{C}_{S(x)}$. Thus $\bigcup_S \mathcal{C}_S = \mathbb{R}^n$, and $\mathcal{P} := \{\mathcal{C}_S\}$ is a finite convex partition of $\mathbb{R}^n$ with pairwise disjoint interiors. Hence $\mathcal{P} := \{\mathcal{C}_S\}$ is a convex partition of $\mathbb{R}^n$. On each set $\mathcal{C}_S$ we have

$$\left\|A_j x - \bar{q}_j\right\|_1 = \sum_{k=1}^{m_j} \left|(A_j x - \bar{q}_j)_k\right| = \sum_{k=1}^{m_j} s_{j,k}\,(A_j x - \bar{q}_j)_k, \qquad x \in \mathcal{C}_S,$$

so that the restriction of f to $\mathcal{C}_S$ can be written as

$$f_S(x) := f(x) = \sum_{i=1}^{m} \lambda_i^+ \left\|B_i x - \bar{p}_i\right\|_2^2 \;-\; \sum_{j=1}^{r} \lambda_j^- \sum_{k=1}^{m_j} s_{j,k}\,(A_j x - \bar{q}_j)_k, \qquad x \in \mathcal{C}_S.$$

Since $H \succ 0$, each f_S is a strongly convex quadratic on the closed convex set $\mathcal{C}_S$, and therefore admits a unique minimizer $x_S^* \in \mathcal{C}_S$. Let $X_S^* := \{x_S^*\}$. Then, for any $x \in \mathcal{C}_S$ and $t \in [0,1]$, the convexity of f_S gives

$$f\big(tx_S^* + (1-t)x\big) = f_S\big(tx_S^* + (1-t)x\big) \leq tf_S(x_S^*) + (1-t)f_S(x) = tf(x_S^*) + (1-t)f(x),$$

so that condition (5) holds on $\mathcal{C}_S$ with respect to X_S^*. Since this is valid for every set $\mathcal{C}_S$ in the partition $\mathcal{P}$, the function f is PWSC on $\mathbb{R}^n$ with respect to $\mathcal{P}$. Finally, the positive definiteness of H implies that the quadratic part of f has quadratic growth and dominates the linear growth of the second term at infinity. Thus, f is coercive and attains at least one global minimizer.

The next example shows how the minimum of strongly convex functions sharing a common strongly convex base function induces a polyhedral partition of the space, a construction closely related to (additively weighted) Voronoi diagrams with applications in clustering, facility location, distance geometry and constrained optimization; see, for example, [1,12] for comprehensive surveys on this subject.

Example 5.2 Let $\psi : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable, μ–strongly convex function with $\mu > 0$. For $i = 1, \dots, m$, let $b_i \in \mathbb{R}^n$ and $c_i \in \mathbb{R}$ be given, and define

$$\phi_i(x) \;:=\; \psi(x) + b_i^\top x + c_i, \qquad x \in \mathbb{R}^n, \quad i = 1, \dots, m,$$

and

$$f(x) \;:=\; \min_{1 \leq i \leq m} \phi_i(x), \qquad x \in \mathbb{R}^n.$$

Since ψ is continuously differentiable on $\mathbb{R}^n$, it is locally Lipschitz on $\mathbb{R}^n$. For each $i \in \{1, \dots, m\}$, the function ϕ_i is the sum of a locally Lipschitz function and an affine function, hence ϕ_i is locally Lipschitz on $\mathbb{R}^n$. Finally, the pointwise minimum of finitely many locally Lipschitz functions is locally Lipschitz. Therefore, f is locally Lipschitz on $\mathbb{R}^n$.

Each ϕ_i is μ–strongly convex on $\mathbb{R}^n$, hence in particular convex and coercive. For each $i \in \{1, \dots, m\}$, consider the set

$$\mathcal{C}_i := \left\{\, x \in \mathbb{R}^n : \phi_i(x) \leq \phi_j(x) \text{ for all } j = 1, \dots, m \,\right\}.$$

Using the fact that all the functions ϕ_i share the same strongly convex base function ψ, we can rewrite the inequalities defining the set $\mathcal{C}_i$ as follows. For any $x \in \mathbb{R}^n$, $\phi_i(x) \leq \phi_j(x)$, if and only if $\psi(x) + b_i^\top x + c_i \leq \psi(x) + b_j^\top x + c_j$, and, after can subset ing $\psi(x)$ on both sides, this is equivalent to $(b_i - b_j)^\top x \leq c_j - c_i$. Hence

$$\mathcal{C}_i = \bigcap_{j=1}^{m} \left\{\, x \in \mathbb{R}^n : (b_i - b_j)^\top x \leq c_j - c_i \,\right\},$$

so each $\mathcal{C}_i$ is a closed convex polyhedron (intersection of affine halfspaces). As in Voronoi-type constructions, these sets have pairwise disjoint interiors and satisfy $\bigcup_{i=1}^m \mathcal{C}_i = \mathbb{R}^n$, so that

$$\mathcal{P} := \{\mathcal{C}_i : i = 1, \ldots, m\}$$

is a finite convex polyhedral partition of $\mathbb{R}^n$. By construction, $f(x) = \phi_i(x)$ for all $x \in \mathcal{C}_i$. Restricting to $\mathcal{C}_i$, the function f coincides with ϕ_i, which is μ–strongly convex on $\mathcal{C}_i$ and therefore admits a unique minimizer $x_i^* \in \mathcal{C}_i$ of f on that set. Let $X_i^* := \{x_i^*\}$. Then, for any $x \in \mathcal{C}_i$ and $t \in [0,1]$, the convexity of ϕ_i yields

$$f\big(tx_i^* + (1-t)x\big) = \phi_i\big(tx_i^* + (1-t)x\big) \leq t\,\phi_i(x_i^*) + (1-t)\,\phi_i(x) = t\,f(x_i^*) + (1-t)\,f(x),$$

so that (5) holds on $\mathcal{C}_i$ with respect to X_i^*. Hence f is PWSC on $\mathbb{R}^n$ with respect to the finite partition $\mathcal{P} = \{\mathcal{C}_i\}_{i=1}^m$. Finally, since ψ is coercive and $\phi_i(x) = \psi(x) + b_i^\top x + c_i$, each ϕ_i is coercive and so is their pointwise minimum f. In particular, f attains at least one global minimizer.

Generalized Fermat–Weber problems provide a classical family of location models in which piecewise structures naturally arise. In our setting, a particularly simple and important instance is obtained when the strongly convex base function and the sites are chosen in a Euclidean way, as described in the following remark.

Remark 5.1 Let $m \in \mathbb{N}$ and, for each $i = 1, \ldots, m$, let $P_i := \{a_{i1}, \ldots, a_{i\ell_i}\} \subset \mathbb{R}^n$ be a finite nonempty set of reference points, and let $\omega_i \geq 0$ be weights such that $\sum_{i=1}^m \omega_i = 1$. The squared distance to P_i is

$$d_{P_i}^2(x) := \min_{1 \leq \ell \leq \ell_i} \|x - a_{i\ell}\|^2, \qquad x \in \mathbb{R}^n,$$

and we consider the generalized Fermat–Weber objective given by

$$f(x) := \sum_{i=1}^m \omega_i\, d_{P_i}^2(x), \qquad x \in \mathbb{R}^n.$$

This model is a particular case of Example 5.2. Indeed, take $\psi(x) := \|x\|^2$, for $x \in \mathbb{R}^n$, as the common strongly convex base function in Example 5.2, and use the index set

$$\Lambda := \{1, \ldots, \ell_1\} \times \cdots \times \{1, \ldots, \ell_m\}.$$

For each multi-index $\ell = (\ell_1, \ldots, \ell_m) \in \Lambda$, introduce the strongly convex quadratic

$$\phi_\ell(x) := \sum_{i=1}^m \omega_i \|x - a_{i\ell_i}\|^2, \qquad x \in \mathbb{R}^n.$$

By expanding the squared norms, $\phi_\ell : \mathbb{R}^n \to \mathbb{R}$ can be written in the affine–perturbation form of Example 5.2, namely

$$\phi_\ell(x) = \psi(x) + b_\ell^\top x + c_\ell, \qquad b_\ell := -2\sum_{i=1}^m \omega_i a_{i\ell_i}, \quad c_\ell := \sum_{i=1}^m \omega_i \|a_{i\ell_i}\|^2.$$

For a fixed $x \in \mathbb{R}^n$, each term $d_{P_i}^2(x)$ is a minimum over the finite set P_i. Thus, for every $i \in \{1, \ldots, m\}$ there exists an index $\ell_i(x) \in \{1, \ldots, \ell_i\}$ such that

$$d_{P_i}^2(x) = \|x - a_{i\ell_i(x)}\|^2.$$

Define the multi-index $\ell(x) := \big(\ell_1(x), \ldots, \ell_m(x)\big) \in \Lambda$. Then, we conclude that

$$f(x) = \sum_{i=1}^{m} \omega_i d_{P_i}^2(x) = \sum_{i=1}^{m} \omega_i \|x - a_{i\ell_i(x)}\|^2 = \phi_{\ell(x)}(x) \;\geq\; \min_{\ell \in \Lambda} \phi_\ell(x).$$

On the other hand, for any fixed $\ell \in \Lambda$ and any $x \in \mathbb{R}^n$ we have $d_{P_i}^2(x) \leq \|x - a_{i\ell_i}\|^2$, for $i = 1, \ldots, m$, so that

$$f(x) = \sum_{i=1}^{m} \omega_i d_{P_i}^2(x) \leq \sum_{i=1}^{m} \omega_i \|x - a_{i\ell_i}\|^2 = \phi_\ell(x).$$

Therefore, taking the minimum over $\ell \in \Lambda$ in the last inequality yields $f(x) \leq \min_{\ell \in \Lambda} \phi_\ell(x)$, and combining the two inequalities we obtain

$$f(x) \;=\; \min_{\ell \in \Lambda} \phi_\ell(x), \qquad x \in \mathbb{R}^n.$$

Consequently, the same polyhedral partition $\mathcal{P} = \{\mathcal{C}_\ell : \ell \in \Lambda\}$ as in Example 5.2 induces a PWSC structure for f. As a special case, when each P_i is a singleton, $P_i = \{a_i\}$, one recovers the classical squared Fermat–Weber problem with point sites.

These above examples show that PWSC objectives arise naturally in sparse regularization, contrastive learning models, and minimum–of–quadratics constructions related to Voronoi–type partitions. In all cases, the PWSC structure is induced by a finite partition of the space into convex regions on which the objective coincides with a strongly convex function. This allows the proximal point method to exploit the resulting locally convex geometry on each subset while still treating a globally nonconvex objective.

6 Illustrative numerical experiments

In this section, we present a few illustrative numerical experiments for the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a PWSC function. We emphasize that the primary contribution of this work is theoretical in nature. Accordingly, the numerical experiments reported here are not intended to provide an extensive computational assessment or a comparison with existing methods, but rather to illustrate the practical behavior of the proposed algorithm.

All experiments were implemented in Python and executed on a machine equipped with a 3.5 GHz Dual-Core Intel Core i5 processor and 16 GB of memory. The pair (x_{k+1}, v_{k+1}), satisfying (14) and (15), was computed using the robust gradient sampling algorithm of [2], with a maximum of 30 iterations. Moreover, Algorithm 1 was terminated when

$$f(x_k) \leq f^* + 10^{-6},$$

where f^* denotes the known global optimal value of each test problem.

We first consider a two–dimensional PWSC example, as in Example 5.2, in order to illustrate the influence of the choices of the sequences $\{\lambda_k\}$ and $\{\sigma_k\}$ on the performance of the proposed method.

Example 6.1 Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x) = \min \left\{ \|x - a\|^2, \ \|x + a\|^2 \right\},$$

where $a \in \mathbb{R}^2$ is a randomly generated vector with components uniformly distributed in the box $[0, 20] \times [0, 20]$. The function f is piecewise smooth and convex, being the pointwise minimum of two strongly convex quadratic functions. The corresponding partition of $\mathbb{R}^2$ is given by

$$\mathcal{P} := \left\{ \{x \in \mathbb{R}^2 : \langle x, a\rangle \geq 0\}, \ \{x \in \mathbb{R}^2 : \langle x, a\rangle \leq 0\} \right\}.$$

The set of global minimizers consists of the two points $\{-a, a\}$, both attaining the optimal value $f^* = 0$.

We first investigate the influence of the proximal parameter sequence $\{\lambda_k\}$ on the computational performance of the algorithm. To this end, we fix the inexactness sequence as $\sigma_k = 1/(k+1)^2$ for all $k \geq 0$ and consider a randomly generated initial point $x_0 \in [-20, 20] \times [-20, 20]$. The algorithm was executed for constant proximal parameters of the form $\lambda_k := \lambda$, with

$$\lambda \in \{0.1, 0.2, 0.3, \dots, 0.9, 1.0, 1.5, 2, 4\}.$$

Performance was assessed in terms of computational cost, defined as the total number of function evaluations and gradient evaluations at points where the objective function is differentiable, the latter arising within the robust gradient sampling algorithm used to solve the subproblems:

$$\text{Cost} := \text{fe} + \text{ge}.$$

Figure 1 shows that the computational cost of Algorithm 1 is strongly influenced by the choice of the proximal parameter λ. The cost exhibits a clearly nonmonotonic behavior, with intermediate values of λ providing the most efficient performance. In contrast, larger values of λ lead to a significant increase in computational cost, indicating that overly aggressive proximal steps adversely affect efficiency.

We next investigate the influence of the inexactness sequence $\{\sigma_k\}$ on the computational performance of the algorithm. Recall that setting $\sigma_k = 0$ for all $k \geq 0$ corresponds to requiring exact solutions of the proximal subproblems. In this round, we fix the proximal parameter as $\lambda_k = 0.3$ and consider inexactness sequences of the form $\sigma_k = 1/(k + 1)^p$. For a randomly generated initial point $x_0 \in [-20, 20] \times [-20, 20]$, the algorithm was executed for different values of the exponent p,

$$p \in \{1.1, \ 1.2, \ 1.3, \dots, \ 1.9, \ 2.0, \ 2.5, \ 3, \ 4\}.$$

Figure 2 reports the corresponding computational cost as a function of p. The results indicate that allowing inexact subproblem solutions can substantially reduce the computational cost, provided that the error sequence decays sufficiently fast. In particular, moderate values of p yield the lowest cost, while larger values of p increase the computational burden and approach the behavior of the exact variant of the algorithm.

We conclude this section with a nonsmooth and star-convex test problem.

Example 6.2 Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x) = h_1(x)\, h_2(x),$$

where

$$h_1(x) := \max\{\|x\|, \ 2\|x\| - 2\},$$

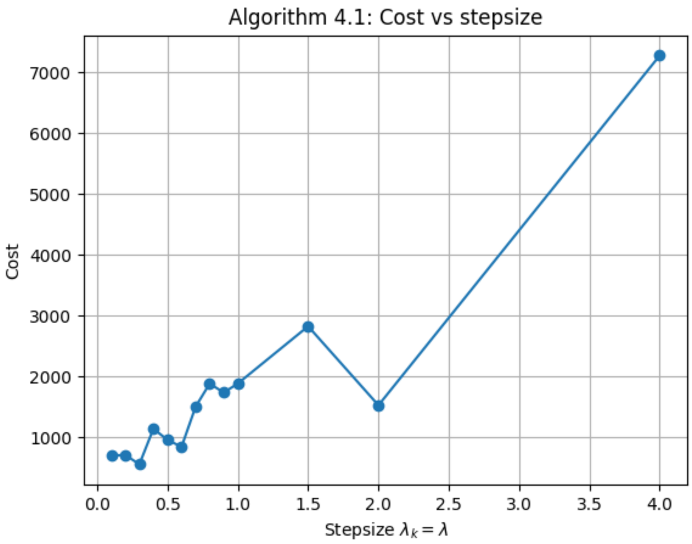

Fig. 1 Computational cost (function plus gradient evaluations) of Algorithm 1 for the two–dimensional PWSC function $f(x) = \min\{\|x - a\|^2, \|x + a\|^2\}$ as a function of the proximal parameter $\lambda_k = \lambda$.

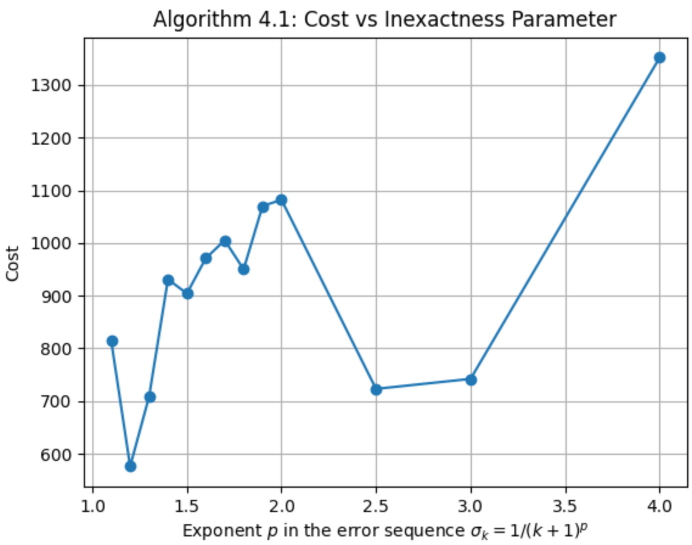

Fig. 2 Computational cost of Algorithm 1 for the two–dimensional PWSC function $f(x) = \min\{\|x-a\|^2, \|x+a\|^2\}$ as a function of the inexactness parameter p in the error sequence $\sigma_k = 1/(k+1)^p$.

and

$$h_2(x) := \begin{cases} 1, & \text{if } x = 0, \\ \dfrac{1}{4N} \displaystyle\sum_{i=1}^{N} \left(a_i \sin^2\left(b_i \frac{x_1}{\|x\|}\right) + c_i \cos^2\left(d_i \frac{x_2}{\|x\|}\right)\right) + 1, & \text{otherwise,} \end{cases}$$

with $a_i, b_i, c_i, d_i \in \mathbb{R}$ for $i = 1, \ldots, N$. The function h_1 is convex, while h_2 is smooth and bounded away from zero. In particular, it can be shown that f is a nonsmooth and star-convex function on $\mathbb{R}^2$. The unique global minimizer is $x^* = (0, 0)$, with optimal value $f^* = 0$.

The purpose of this experiment is to assess the ability of Algorithm 1 to efficiently handle functions with mixed nonsmooth radial and smooth angular structures. Numerical experiments were conducted as follows. The coefficients $\{a_i, c_i\}_{i=1}^N$ were generated independently from the uniform distribution on $[0, 20]$, while $\{b_i, d_i\}_{i=1}^N$ were generated independently from the uniform distribution on $[-25, 25]$. The algorithm parameters were fixed as $\lambda_k = 0.3$ and $\sigma_k = 1/(k+1)^{1.2}$ for all $k \geq 0$. For each value of N, the initial point $[0.5, 0.5]$ was used, and the algorithm was tested for $N \in \{2,\ 5,\ 10,\ 20\}$.

The numerical results reported in Table 1 and illustrated in Figure 3 show that Algorithm 1 converges reliably for all tested values of N. In all instances, the algorithm terminated successfully and reached a stationary point of f, with a consistent decay of the objective value as the computational cost increases. These results demonstrate that the proposed algorithm performs stably on this nonsmooth star-convex problem, in full agreement with the theoretical results established in Section 4.

Table 1 Performance of Algorithm 1 for the star-convex function $f(x) = h_1(x)h_2(x)$ with different values of N.

N	Iterations	fe	ge	Cost (fe + ge)
2	45	27.901	8.100	36.001
5	59	37.069	10.620	47.689
10	66	41.451	11.880	53.331
20	98	61.566	17.640	79.206

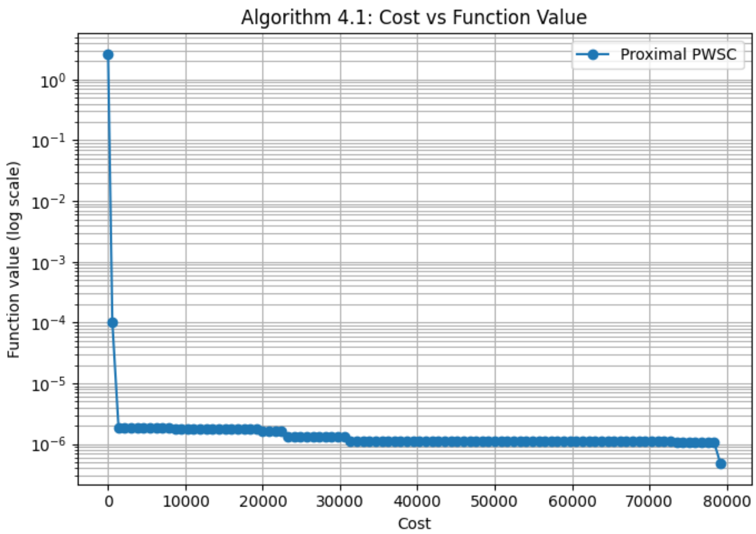

Fig. 3 Function value versus computational cost (function plus gradient evaluations) for Algorithm 1 applied to the star-convex function $f(x) = h_1(x)h_2(x)$ with $N = 20$. A logarithmic scale is used for the function value.

7 Final remark

In this work, we proposed and analyzed an inexact proximal point framework for a class of nonsmooth and nonconvex problems characterized by a piecewise star-convex structure. By exploiting this geometry, we established well-posedness of the method, monotonic decrease of the objective function values, and convergence of the generated sequence under mild conditions on the inexactness. In contrast to the standard guarantees available for general nonsmooth nonconvex problems, our analysis shows that the piecewise structure leads to stronger conclusions, convergence of the whole sequence, and local optimality of accumulation points lying in the interior of a region. Furthermore, we derived complexity bounds for computing approximate stationary points and obtained sublinear and linear convergence rates under additional assumptions. These results highlight the relevance of piecewise star-convexity as a meaningful extension of convexity that still allows convergence guarantees of the proximal methods.

A natural direction for future research is to extend the present piecewise structure from star–convexity to (strongly) quasar–convexity, a class that has recently received considerable attention in nonconvex optimization; see, for instance, [4, 8, 11, 18]. Quasar–convexity and strong quasar–convexity were introduced in [8] as relaxations of convexity and star–convexity, and have since been used to derive first–order methods and proximal schemes with global convergence guarantees; see in particular the proximal point analysis in [4] and the related developments in [11, 18]. Motivated by these notions, we propose the following piecewise variants, in the spirit of Definition 3.2.

Definition 7.1 Let $\mathcal{P} = \{\, \mathcal{C}_i \,:\, i \in I\}$ be a family of nonempty closed convex subsets of $\mathbb{R}^n$ with pairwise disjoint interiors and $\bigcup_{i\in I} \mathcal{C}_i = \mathbb{R}^n$. We say that $f : \mathbb{R}^n \to \mathbb{R}$ is *piecewise quasar–convex with respect to* $\mathcal{P}$ if, for every $i \in I$,

(i) the set of global minimizers of f on $\mathcal{C}_i$, denoted X_i^*, is nonempty; and
(ii) there exists a parameter $\gamma_i \in (0, 1]$ such that, for any $x_i^* \in X_i^*$, the restriction $f|_{\mathcal{C}_i}$ is γ_i–quasar–convex on $\mathcal{C}_i$ with respect to x_i^* in the sense of [8, Definition 1] or [4].

Similarly, we say that f is *piecewise strongly quasar–convex with respect to* $\mathcal{P}$ if, for every $i \in I$,

(i) the set X_i^* is nonempty; and
(ii) there exist parameters $\gamma_i \in (0, 1]$ and $\mu_i > 0$ such that, for any $x_i^* \in X_i^*$, the restriction $f|_{\mathcal{C}_i}$ is (γ_i, μ_i)–strongly quasar–convex on $\mathcal{C}_i$ with respect to x_i^* in the sense of [8, Eq. (2)] or [4].

It would be interesting to develop a convergence and complexity analysis of the inexact proximal point method for such piecewise (strongly) quasar–convex functions, clarifying how the local parameters (γ_i, μ_i) on each region influence the global behavior and convergence rates of the proximal sequence, and to investigate whether mechanisms inspired by [4, 8, 18] can be incorporated into this piecewise setting.

References

1. F. Aurenhammer. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Comput. Surv.*, 23(3):345–405, 1991.
2. J. V. Burke, A. S. Lewis, and M. L. Overton. A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM J. Optim.*, 15(3):751–779, 2005.
3. F. Clarke. *Optimization and Nonsmooth Analysis.* Canadian Mathematical Society series of monographs and advanced texts. Wiley, 1983.
4. J. de Brito, F. Lara, and D. Liu. Extending linear convergence of the proximal point algorithm: The quasar-convex case. *arXiv preprint arXiv:2509.04375*, 2025.

5. R. Díaz Millán, O. P. Ferreira, and J. Ugon. Frank–Wolfe algorithm for DC optimization problem. *arXiv e-prints*, 2023.
6. R. D. Díaz Millán, O. P. Ferreira, and J. Ugon. Frank–Wolfe algorithm for star-convex functions. *arXiv e-prints*, 2025.
7. D. Fortin and I. Tsevendorj. Piecewise-convex maximization problems: algorithm and computational experiments. *J. Glob. Optim.*, 24(1):61–77, 2002.
8. O. Hinder, A. Sidford, and N. Sohoni. Near-optimal methods for minimizing star-convex functions and beyond. *Conference on learning theory*, pages 1894–1938, 2020.
9. J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms. I*, volume 305 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993. Fundamentals.
10. A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon. A survey on contrastive self-supervised learning. *Technologies*, 9(1):2, 2021. Article 2.
11. P. Q. Khanh and F. Lara. Star quasiconvexity: an unified approach for linear convergence of first-order methods beyond convexity. *arXiv preprint arXiv:2510.24981*, 2025.
12. L. Liberti, C. Lavor, N. Maculan, and A. Mucherino. Euclidean distance geometry and applications. *SIAM Rev.*, 56(1):3–69, 2014.
13. F. V. Louveaux. Piecewise convex programs. *Math. Program.*, 15(1):53–62, 1978.
14. D. Martínez-Rubio. Smooth quasar-convex optimization with constraints. *arXiv preprint arXiv:2510.01943*, 2025.
15. Y. Nesterov and B. T. Polyak. Cubic regularization of Newton method and its global performance. *Math. Program.*, 108(1, Ser. A):177–205, 2006.
16. B. Rister and D. L. Rubin. Piecewise convexity of artificial neural networks. *Neural Networks*, 94:34–45, 2017.
17. H. Wang, F. Nie, and H. Huang. Robust distance metric learning via simultaneous L1-norm minimization and maximization. In E. P. Xing and T. Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 1836–1844, Beijing, China, 2014. PMLR.
18. J.-K. Wang and A. Wibisono. Continuized acceleration for quasar convex functions in non-convex optimization. *arXiv preprint arXiv:2302.07851*, 2023.