

On the Minimization of the Sum of Nonconvex Functions with Applications to Mathematical Programming

Felipe Lara* Le Hai Yen†

June 16, 2025

Abstract We study the minimization problem of the sum of two functions in which one of them is nonconvex and nonsmooth and the other is differentiable with a Lipschitz continuous gradient (and possibly nonconvex too). By assuming that the nonconvex nonsmooth function is strongly quasiconvex in the sense of Polyak, we first provide interesting necessary optimality conditions and then we implement the proximal gradient algorithm. As a consequence, new and useful information regarding the point obtained by the stopping criteria as well as for the limit point of the generated sequence under the standard Polyak-Kurdyka-Łojasiewicz property is provided. Moreover, we apply these results to different problems from nonconvex optimization as the minimization of the sum of nonconvex functions including difference of convex programming and quadratic fractional programming problems. Finally, examples of strongly quasiconvex functions which satisfy the Polyak-Kurdyka-Łojasiewicz property as well as numerical illustrations are presented.

Keywords: Nonconvex optimization; Nonsmooth optimization; Proximal gradient algorithms; Forward-backward algorithms; DC programming; Quadratic fractional programming

1 Introduction

sec:1

Let $h, f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be two extended real-valued functions. We are interested in the following mathematical problem

$$\min_{x \in \mathbb{R}^n} F(x) := h(x) + f(x). \quad (1.1) \quad \text{min:sum}$$

Its solution set is denoted by $\text{argmin}_{\mathbb{R}^n}(h + f)$.

*Instituto de Alta Investigación (IAI), Universidad de Tarapacá, Arica, Chile. E-mail: felipelaraobregue@gmail.com; flarao@academicos.uta.cl. Web: felipelara.cl, Orcid-ID: 0000-0002-9965-0921

†Institute of Mathematics, Vietnam Academy of Sciences and Technology (VAST), Hanoi, Vietnam. Email: lhien@math.ac.vn, Orcid-ID: 0000-0002-6725-6567

Problem (1.1) encompasses different problems studied in the continuous optimization community as the convex minimization problems, difference of convex (DC) minimization problems and many other classes of mathematical programming problems as well as their applications in signal and image recovery, data sciences and biochemistry among others (see [1, 2, 7, 11] and references therein).

In order to obtain optimality conditions and design algorithms for solving problem (1.1), we can take advantage of the composite form of function F and the properties of the functions f and h in different situations such as: convexity, differentiability and/or global and locally Lipschitz continuity properties for the gradient among others (see for example [3, 4, 7, 11, 20, 24]).

Optimality conditions for problem (1.1) have been deeply developed for the case when h and f are convex and f is differentiable since, in this case, the following outstanding necessary and sufficient condition for optimality holds (see, for instance, [6, Theorem 10.7]):

$$\bar{x} \text{ is a global minimizer of } h + f \iff 0 \in \partial h(\bar{x}) + \nabla f(\bar{x}), \quad (1.2)$$

convex:case

where ∂ denotes the usual convex subdifferential (see relation (2.12) below).

Relation (1.2) is extremely useful for studying problem (1.1) in the convex case and, also, for providing numerical methods in order to find a solution as, for instance, when we use the proximal gradient method (see [6, Chapter 10]).

In the case when f and h are nonconvex, relation (1.2) is no longer useful, but another outstanding subdifferential could be used, the limiting subdifferential, which has many interesting properties (see [32] for a great account) as following optimality condition for local minimizers [32]:

$$\bar{x} \text{ is a local minimizer of } h + f \implies 0 \in \partial^{\text{lim}} h(\bar{x}) + \nabla f(\bar{x}), \quad (1.3)$$

nece:limsubd

where ∂^{lim} is the limiting subdifferential (see relation (2.13) below).

Many forward-backward algorithms for solving (1.1) were built based on this optimality condition under different assumptions on h and f . Among them, the proximal gradient method and its accelerated versions were deeply studied in [3, 4, 10, 11, 20, 24, 29, 31] under the assumption that f has global or locally Lipschitz continuous gradient and h is lsc.

When the function h is not convex, but enjoys a generalized convexity property such as: quasiconvexity or strong quasiconvexity, the optimality condition (1.2) for the convex case can not be applied while condition (1.3) still holds true, but even for global minimizers, relation (1.3) could be useless at all since the limiting subdifferential for strongly quasiconvex functions is not a good approximation in general. Indeed, let us consider the function $h(x) = \sqrt{\|x\|}$. Here the minimum point is 0, but the limiting subdifferential $\partial^{\text{lim}} h(0) = \mathbb{R}^n$ is the whole space, a completely useless information. On the other hand, the strong subdifferential [23] for $h(x) = \sqrt{\|x\|}$ at 0 is a compact convex set (see Remark 3.4 below) and includes 0. For these reasons, we consider natural to use this generalized subdifferentials for obtaining new optimality conditions for problem (1.1), and then to use those optimality conditions for studying the

convergence of forward-backward algorithms as the proximal gradient method. This is exactly what we show in this paper.

Our contribution: We deal with problem (1.1) by assuming that the non-convex nonsmooth function h is strongly quasiconvex in the sense of Polyak [35], while f is assumed to be differentiable with Lipschitz continuous gradient. Hence, in virtue of [26], we can apply the proximal operator to h , and for the differentiable function f we use the gradient step. Furthermore, by using the strong subdifferential [23], we provide necessary optimality conditions for problem (1.1) and, as a consequence, we provide new information on the limit point of the sequence generated by the proximal gradient algorithm under the usual Polyak-Kurdyka-Łojasiewicz (*PKL*) property. We apply these results in non-convex optimization problem as in the minimization of the sum of nonconvex functions including DC programming problems and quadratic fractional programming problems. Furthermore, as we will show in Remark 3.1 (below), our assumptions on the functions h and f are not restrictive since we can always encompass the usual cases of minimizing the sum of two convex functions and the DC programming problem in our context and, moreover, several nonconvex nonsmooth problems fits our assumptions while previous known results can not be applied.

The structure of the paper is as follows: In Section 2 we present preliminaries and basic definitions regarding generalized convexity, nonsmooth analysis and first-order algorithms. In Section 3, we present both new necessary optimality conditions and a proximal gradient algorithm for problem (1.1). Furthermore, we ensure the convergence of the sequence generated by the proximal gradient algorithm under standard assumptions while new information regarding its limit point is provided. In Section 4, we apply our theoretical results in the minimization of the sum of nonconvex functions which includes DC programming and quadratic fractional programming problems and we present numerical illustrations of the proposed algorithm.

2 Preliminaries and Basic Definitions

sec:2

The inner product of $\mathbb{R}^n$ and the Euclidean norm are denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively. Let K be a nonempty set in $\mathbb{R}^n$, its closure is denoted by $\text{cl } K$, its boundary by $\text{bd } K$, its topological interior by $\text{int } K$ and its convex hull by $\text{conv } K$. Given a convex and closed set K , the projection of x on K is denoted by $P_K(x)$, and the indicator function on K by ι_K . The open ball with center at x and radius $\delta > 0$ is denoted by $B(x, \delta)$.

Given a closed and convex set $K \subseteq \mathbb{R}^n$, the normal cone of K at $\bar{x} \in K$ is defined by

$$N_K(\bar{x}) := \{w \in \mathbb{R}^n : \langle w, \bar{x} - z \rangle \geq 0, \forall z \in K\}. \quad (2.1)$$

eq:NorCon

Given any extended-valued function $h : \mathbb{R}^n \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$, the effective domain of h is defined by $\text{dom } h := \{x \in \mathbb{R}^n : h(x) < +\infty\}$. It is said that h

is proper if $\text{dom } h$ is nonempty and $h(x) > -\infty$ for all $x \in \mathbb{R}^n$. The notion of properness is important when dealing with minimization problems.

It is indicated by $\text{epi } h := \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : h(x) \leq t\}$ the epigraph of h , by $S_\lambda(h) := \{x \in \mathbb{R}^n : h(x) \leq \lambda\}$ the sublevel set of h at the height $\lambda \in \mathbb{R}$ and by $\text{argmin}_{\mathbb{R}^n} h$ the set of all minimal points of h . A function h is lower semicontinuous (lsc henceforth) at $x_0 \in \mathbb{R}^n$ if for any sequence $\{x_k\}_k \in \mathbb{R}^n$ with $x_k \rightarrow x_0$, $h(x_0) \leq \liminf_{k \rightarrow +\infty} h(x_k)$. Furthermore, the current conventions $\sup_\emptyset h := -\infty$ and $\inf_\emptyset h := +\infty$ are adopted.

A function h with convex domain is said to be

- (a) convex if, given any $x, y \in \text{dom } h$, then

$$h(\lambda x + (1 - \lambda)y) \leq \lambda h(x) + (1 - \lambda)h(y), \quad \forall \lambda \in [0, 1], \quad (2.2)$$

def:convex

- (b) strongly convex on $\text{dom } h$ with modulus $\gamma_h \in]0, +\infty[$ if for all $x, y \in \text{dom } h$ and all $\lambda \in [0, 1]$, we have

$$h(\lambda y + (1 - \lambda)x) \leq \lambda h(y) + (1 - \lambda)h(x) - \lambda(1 - \lambda)\frac{\gamma_h}{2}\|x - y\|^2, \quad (2.3)$$

strong:convex

- (c) quasiconvex if, given any $x, y \in \text{dom } h$, then

$$h(\lambda x + (1 - \lambda)y) \leq \max\{h(x), h(y)\}, \quad \forall \lambda \in [0, 1], \quad (2.4)$$

def:qcx

- (d) strongly quasiconvex on $\text{dom } h$ with modulus $\gamma_h \in]0, +\infty[$ if for all $x, y \in \text{dom } h$ and all $\lambda \in [0, 1]$, we have

$$h(\lambda y + (1 - \lambda)x) \leq \max\{h(y), h(x)\} - \lambda(1 - \lambda)\frac{\gamma_h}{2}\|x - y\|^2. \quad (2.5)$$

strong:quasiconvex

It is said that h is strictly convex (resp. strictly quasiconvex) if the inequality in (2.2) (resp. (2.4)) is strict whenever $x \neq y$ and $\lambda \in]0, 1[$. Furthermore, note that convex (resp. quasiconvex) functions could be seen as strongly convex (resp. strongly quasiconvex) functions when $\gamma_h = 0$.

The relationship between all these notions is summarizing below (we denote quasiconvex by qcx):

$$\begin{array}{ccccc} \text{strongly convex} & \implies & \text{strictly convex} & \implies & \text{convex} \\ \downarrow & & \downarrow & & \downarrow \\ \text{strongly qcx} & \implies & \text{strictly qcx} & \implies & \text{qcx} \end{array} \quad (\text{SC}) \quad \text{scheme}$$

All the reverse statements do not hold in general. For instance, the Euclidean norm $h_1(x) = \|x\|$ is strongly quasiconvex without being strongly convex on any bounded convex set (see [22, Theorem 2]) and the function $h_2(x) = \frac{x}{1+|x|}$ is strictly quasiconvex without being strongly quasiconvex on $\mathbb{R}$ while the other counter examples are well-known (see [12, 16]).

Before continuing, let us show some new examples of strongly quasiconvex functions which are not convex.

rem:exam

Remark 2.1. (i) Let $h : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by $h(x) = \sqrt{\|x\|}$. Clearly, h is nonconvex, but it is strongly quasiconvex on any $\mathbb{B}(0, r)$, $r > 0$, with modulus $\gamma = \frac{1}{5^{\frac{1}{4}} 2^{\frac{3}{4}} r^{\frac{1}{2}}}$ by [26, Theorem 17].

(ii) Let $A, B \in \mathbb{R}^{n \times n}$, $a, b \in \mathbb{R}^n$, $\alpha, \beta \in \mathbb{R}$, and $h : \mathbb{R}^n \rightarrow \mathbb{R}$ be the function given by:

$$h(x) = \frac{f(x)}{g(x)} = \frac{\frac{1}{2}\langle Ax, x \rangle + \langle a, x \rangle + \alpha}{\frac{1}{2}\langle Bx, x \rangle + \langle b, x \rangle + \beta}. \quad (2.6)$$

Take $0 < m < M$ and define:

$$K := \{x \in \mathbb{R}^n : m \leq g(x) \leq M\}.$$

If A is a positive definite matrix and at least one of the following conditions holds:

- (a) $B = 0$ (the null matrix),
- (b) f is nonnegative on K and B is negative semidefinite,
- (c) f is nonpositive on K and B is positive semidefinite,

then h is strongly quasiconvex on K with modulus $\gamma = \frac{\lambda_{\min}(A)}{M}$ by [19, Proposition 4.1], where $\lambda_{\min}(A)$ is the minimum eigenvalue of A .

(iii) Let $h_1, h_2 : \mathbb{R}^n \rightarrow \mathbb{R}$ be two strongly quasiconvex functions with modulus $\gamma_1, \gamma_2 > 0$, respectively. Then $h := \max\{h_1, h_2\}$ is strongly quasiconvex with modulus $\gamma := \min\{\gamma_1, \gamma_2\} > 0$ (straightforward).

(iv) Let $\alpha > 0$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}$ be a strongly quasiconvex function with modulus $\gamma > 0$. Then αh is strongly quasiconvex with modulus $\gamma\alpha > 0$ (straightforward).

A proper function $h : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ is said to be:

(i) 2-supercoercive, if

$$\liminf_{\|x\| \rightarrow +\infty} \frac{h(x)}{\|x\|^2} > 0,$$

(ii) coercive, if

$$\lim_{\|x\| \rightarrow +\infty} h(x) = +\infty,$$

or equivalently, if $S_\lambda(h)$ is bounded for all $\lambda \in \mathbb{R}$.

(iii) 2-weakly coercive, if

$$\liminf_{\|x\| \rightarrow +\infty} \frac{h(x)}{\|x\|^2} \geq 0. \quad (2.7)$$

2:weakly

Clearly, (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii), but the reverse statements do not hold as the functions $h(x) = \sqrt{x}$ and $h(x) = x$ shows, respectively.

The following existence result is the starting point of our research.

`strongqcx:coercive`

Lemma 2.1. ([26, Theorem 1]) *Let K be a convex set in $\mathbb{R}^n$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}$ be a strongly quasiconvex function with modulus $\gamma_h > 0$ on K . Then h is 2-supercoercive (in particular, coercive).*

As a consequence, every lsc strongly quasiconvex function has an unique minimizer on a closed and convex set K which satisfies that ([23])

$$h(\bar{x}) + \frac{\gamma_h}{8} \|y - \bar{x}\|^2 \leq h(y), \quad \forall y \in K. \quad (2.8)$$

`eq:strongmin`

If a point $\bar{x} \in K$ satisfies relation (2.8), then $\bar{x}$ is said to be a $\frac{\gamma_h}{8}$ -strong minimum point of h on K .

A function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be L -smooth on $K \subseteq \mathbb{R}^n$ if it is differentiable on K and

$$\|\nabla h(x) - \nabla h(y)\| \leq L\|x - y\|, \quad \forall x, y \in K. \quad (2.9)$$

`L:smooth`

For L -smooth functions, a fundamental result is the descent lemma, that is, if h is a L -smooth function on a convex set K with value $L \geq 0$, then for every $x, y \in K$, we have

$$h(y) \leq h(x) + \langle \nabla h(x), y - x \rangle + \frac{L}{2} \|x - y\|^2. \quad (2.10)$$

`descent:lemma`

Let $K \subseteq \mathbb{R}^n$ be a convex set and $h : K \rightarrow \mathbb{R}$ be a differentiable function. Then h is strongly convex on K with modulus γ_h if and only if

$$h(x) \geq h(y) + \langle \nabla h(y), x - y \rangle + \frac{\gamma_h}{2} \|y - x\|^2, \quad \forall x, y \in K. \quad (2.11)$$

`sconvex:diff`

Given a proper function $h : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, the convex subdifferential of h at $\bar{x} \in \text{dom } h$ is defined by

$$\partial h(\bar{x}) := \{\xi \in \mathbb{R}^n : h(y) \geq h(\bar{x}) + \langle \xi, y - \bar{x} \rangle, \quad \forall y \in \mathbb{R}^n\}, \quad (2.12)$$

`convex:subd`

and by $\partial h(x) = \emptyset$ if $x \notin \text{dom } h$.

For each $x \in \text{dom } h$, the Fréchet subdifferential of h at x , denoted by $\widehat{\partial}h(x)$, is the set defined by:

$$\widehat{\partial}h(x) := \{\xi \in \mathbb{R}^n : \liminf_{\substack{y \neq x \\ y \rightarrow x}} \frac{1}{\|y - x\|} (h(y) - h(x) - \langle \xi, y - x \rangle) \geq 0\},$$

while $\widehat{\partial}h(x) := \emptyset$ when $x \notin \text{dom } h$.

The limiting (or Mordukhovich) subdifferential (see [32]) of h at $x \in \text{dom } h$ is defined by:

$$\partial^{\text{lim}}h(x) = \{\xi \in \mathbb{R}^n : \exists x^k \rightarrow x, h(x^k) \rightarrow h(x), \xi^k \in \widehat{\partial}h(x^k) \rightarrow \xi\}. \quad (2.13)$$

`lim:subd`

We also recall that given a nonempty set $K \subseteq \mathbb{R}^n$, $\beta > 0$ and $\gamma \geq 0$. The (β, γ, K) -strong subdifferential (strong subdifferential) of h at $\bar{x} \in \text{dom } h \cap K$ is defined by (see [23])

$$\begin{aligned} \partial_{\beta, \gamma}^K h(\bar{x}) := \{ \xi \in \mathbb{R}^n : \max\{h(y), h(\bar{x})\} \geq h(\bar{x}) + \frac{\lambda}{\beta} \langle \xi, y - \bar{x} \rangle \\ + \frac{\lambda}{2} \left(\gamma - \frac{\lambda}{\beta} - \lambda \gamma \right) \|y - \bar{x}\|^2, \forall y \in K, \forall \lambda \in [0, 1] \}. \end{aligned} \quad (2.14) \quad \boxed{\text{def:subd}}$$

When $K := \mathbb{R}^n$, we write $= \partial_{\beta, \gamma}^{\mathbb{R}^n} h$.

In the particular case where $K = S_{h(\bar{x})}(h)$, we define the (β, γ) -SS (strong sublevel) subdifferential of h by (see [23])

$$\partial_{\beta, \gamma} h(\bar{x}) := \{ \xi \in \mathbb{R}^n : \langle \xi, y - \bar{x} \rangle \leq -\frac{\beta \gamma}{2} \|y - \bar{x}\|^2, \forall y \in S_{h(\bar{x})}(h) \}. \quad (2.15) \quad \boxed{\text{eq:SS}}$$

If $\gamma > 0$, then $\partial_{\beta, \gamma}^K h(\bar{x})$ is motivated for strongly quasiconvex functions, while if $\gamma = 0$, then $\partial_{\beta, 0}^K h(\bar{x})$ is motivated for quasiconvex functions. Note that when $x \in \text{dom } h$ and $S_{h(x)}(h) \subseteq K$,

$$\partial_{\beta, \gamma}^K h(x) \subseteq \partial_{\beta, \gamma} h(x), \quad \forall \beta > 0, \quad \forall \gamma \geq 0, \quad (2.16) \quad \boxed{\text{inclusion}}$$

and that if h is lsc and quasiconvex, then

$$\partial_{\beta, 0} h(x) = N_{S_{h(x)}(h)}(x).$$

We also have the following properties (see [23]):

- (P1) $\partial_{\beta, \gamma}^K h(x)$ and $\partial_{\beta, \gamma} h(x)$ are closed and convex for all $x \in \text{dom } h \cap K$, all $\beta > 0$ and all $\gamma \geq 0$ by [23, Proposition 7(a)].
- (P2) $\partial_{\beta, \gamma}^K h(\bar{x})$ is compact for all $\bar{x} \in \text{int}(\text{dom } h \cap K)$ by [23, Proposition 7(d)].
- (P3) If $\xi \in \partial_{\beta, \gamma} h(\bar{x})$, then $\lambda \xi \in \partial_{\beta, \gamma} h(\bar{x})$ for all $\lambda \geq 1$, i.e., $\partial_{\beta, \gamma} h(\bar{x})$ is unbounded for all $\bar{x} \in \text{dom } h$ (straightforward).
- (P4) If h is strongly quasiconvex with modulus $\gamma > 0$ and lsc, then $\partial_{\beta, \gamma} h(x) \neq \emptyset$ for all $\bar{x} \in \text{int dom } h$ by [23, Corollary 38(a)].
- (P5) If h is strongly quasiconvex with modulus $\gamma > 0$ and lsc, then $\partial_{\beta, \gamma}^K h(x) \neq \emptyset$ for all $x \in \text{dom } h \cap K$ by [23, Corollary 38(a)].
- (P6) If h is strongly quasiconvex with modulus $\gamma = 0$ (i.e., h is quasiconvex) and satisfies (2.7), then $\partial_{\beta, \gamma}^K h(x) \neq \emptyset$ for all $x \in \text{dom } h \cap K$ by [23, Corollary 38(b)].

Optimality conditions for characterizing optimal solutions to the problem of minimizing a strongly quasiconvex functions in terms of both subdifferentials, that are employed in subgradient projection methods for solving such optimization problems can be found in [23] and [13, 28], respectively. Furthermore, the following result provides a straightforward method to estimate the strong subdifferential in terms of the convex subdifferential for quadratic fractional functions. This result will be useful for numerical experiments in Section 4.

prop:frac2

Lemma 2.2. ([28, Proposition 4.2]) Suppose that $h(x) = \frac{f(x)}{g(x)}$ for all $x \in \text{dom } h$, where f is strongly convex with modulus $\gamma > 0$, g is affine, positive, finite, and bounded from above by $\mu > 0$ on $\text{dom } h$, and $\text{dom } h$ is convex. Then for any $\rho > 0$ and $x_0 \in \text{dom } h$, we have

$$\frac{\rho}{\mu} \partial(f - \alpha g)(x_0) \subseteq \partial_{\rho, \frac{\gamma}{\mu}}^{\text{dom } h} h(x_0),$$

with $\alpha := h(x_0)$.

char:min

Remark 2.2. The convex, strong and strong sublevel subdifferentials characterize different notions of minimum point for any proper function, that is;

$$0 \in \partial h(\bar{x}) \iff h(y) \geq h(\bar{x}), \forall y \in \mathbb{R}^n. \quad (2.17)$$

$$0 \in \partial_{\beta, \gamma} h(\bar{x}) \iff h(y) > h(\bar{x}), \forall y \in K \setminus \{\bar{x}\}. \quad (2.18)$$

$$0 \in \partial_{\beta, \gamma}^K h(\bar{x}) \iff h(y) \geq h(\bar{x}) + \frac{\gamma^2 \beta}{8(1 + \gamma \beta)} \|y - \bar{x}\|^2, \forall y \in K, \quad (2.19)$$

where (2.17) is well-known and relations (2.18) and (2.19) where proved in [23, Theorem 24].

Furthermore, we say that a point $\bar{x} \in K$ is a ε -quasi solution of h on K (with $\varepsilon > 0$) if

$$h(\bar{x}) \leq h(y) + \varepsilon \|y - \bar{x}\|^2, \forall y \in K. \quad (2.20)$$

Let K be a closed and convex set in $\mathbb{R}^n$ and $h : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ be a proper function. Then the proximity operator on K of parameter $\beta > 0$ of h at $x \in \mathbb{R}^n$ is defined as $\text{Prox}_{\beta h} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ where

$$\text{Prox}_{\beta h}(K, x) = \text{argmin}_{y \in K} \left\{ h(y) + \frac{1}{2\beta} \|y - x\|^2 \right\}. \quad (2.21)$$

If $K = \mathbb{R}^n$, then we set $\text{Prox}_{\beta h}(\mathbb{R}^n, x) = \text{Prox}_{\beta h}(x)$. If h is proper, lsc and convex, then $\text{Prox}_{\beta, h}$ becomes a single-valued operator (see [5, Proposition 12.15]).

For strongly quasiconvex and quasiconvex functions we have the following relationship between the proximity operator and the strong subdifferential.

pro:K

Lemma 2.3. ([23, Propositions 36 and 40]) Let K be a closed and convex set in $\mathbb{R}^n$, $h : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ be a proper and lsc function such that $K \subseteq \text{dom } h$, $\beta > 0$ and $z \in K$. If h is strongly quasiconvex on K with modulus $\gamma \geq 0$, then

$$\bar{x} \in \text{Prox}_{\beta h}(K, z) \implies z - \bar{x} \in \partial_{\beta, \gamma}^K h(\bar{x}). \quad (2.22)$$

Let $h : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper and lsc function. For $x \in \text{dom } h$, we denote the limiting (Mordukhovich) subdifferential of h at x by $\partial^{\text{lim}} h(x)$.

Now, let us recall the following definition regarding the Polyak-Kurdyka-Łojasiewicz (PKL) property (see [25, 30, 34] and also [8]).

def:KL

Definition 2.1. (see, for instance, [3]) Let $h : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper and lsc function. We say that h satisfies the Polyak-Kurdyka-Lojasiewicz (PKL) property at $\bar{x} \in \text{dom } \partial^{\text{lim}} h := \{x \in \mathbb{R}^m : \partial^{\text{lim}} h(x) \neq \emptyset\}$ if there exists $\eta \in]0, +\infty]$, a neighborhood U of $\bar{x}$ and a concave function $\phi : [0, \eta[\rightarrow [0, +\infty[$ satisfied

$$(K1) \quad \phi(0) = 0;$$

$$(K2) \quad \phi \text{ is continuously differentiable on }]0, \eta[, \text{ continuous at } 0;$$

$$(K3) \quad \phi'(s) > 0 \text{ for all } s \in (0, \eta),$$

such that for all $x \in U \cap \{x \in \mathbb{R}^m : h(\bar{x}) < h(x) < h(\bar{x}) + \eta\}$, we have

$$\phi'(h(x) - h(\bar{x})) \text{dist}(0, \partial^{\text{lim}} h(x)) \geq 1. \quad (2.23)$$

KLpro

Finally, we recall a convergence result which will play a determined role in what follows. This is a classical result in the context of Fejér monotonicity techniques for proving convergence results of classical algorithms for convex optimization problems and/or for monotone inclusion problems (see [5]).

lemma:seq1

Lemma 2.4. Let $\{a_n\}_n$ and $\{b_n\}_n$ be two real sequences such that $b_n \geq 0$ for all $n \in \mathbb{N}$, $\{a_n\}_n$ is bounded from below and $a_{n+1} + b_n \leq a_n$ for all $n \in \mathbb{N}$. Then $\{a_n\}_n$ is monotonically decreasing and convergent and $\sum_{n \in \mathbb{N}} b_n < +\infty$.

For a further study on generalized convexity, nonsmooth analysis and proximal point type algorithms we refer to [3, 5, 6, 7, 10, 11, 12, 16, 17, 18, 23, 26, 29, 33, 35] and references therein.

3 Minimization of the Sum of Nonconvex Functions

sec:3

In this section, we propose to study optimality conditions for problem (1.1) where f is a real-valued function and h is a proper, extended real-valued function. To that end, let us consider the following assumptions on h and f :

(A1) h is proper, lsc and strongly quasiconvex with modulus $\gamma_h \geq 0$ on its domain.

(A2) f is real-valued, differentiable with L -Lipschitz continuous gradient (with $L > 0$).

(A3) f is 2-weakly coercive.

(A4) f is convex.

The first important comment to understand the reason behind the assumption (A1) is given in the scheme (SC), since every strongly convex function is strongly quasi-convex. As a consequence, the problem of minimizing the difference of convex (DC) functions can also be transformed to our assumptions (A1) and (A2), as we can see below.

part:cases

Remark 3.1. (Difference of convex functions). Let us consider the following DC optimization problem:

$$\min_{x \in \mathbb{R}^n} h_c(x) - f_c(x) \quad (3.1)$$

dc:problem

where $h_c, f_c : \mathbb{R}^n \rightarrow \mathbb{R}$ are convex functions and f_c is differentiable with L_c -Lipschitz continuous gradient. Then problem (3.1) can be transformed into an equivalent problem involving strongly convex functions:

$$\min_{x \in \mathbb{R}^n} h_1(x) - f_1(x), \quad (3.2)$$

sdc:problem

where $h_1(x) = h_c(x) + \frac{\rho}{2}\|x\|^2$ and $f_1(x) = f_c(x) + \frac{\rho}{2}\|x\|^2$ are strongly convex functions with modulus $\rho > 0$. Hence h_1 is strongly quasiconvex with modulus $\rho > 0$ and satisfies (A1), while $-f_1$ satisfies (A2) with $(L_c + \rho)$ -Lipschitz gradient.

We will come back to this problem in the applications of Section 4.

3.1 Necessary Optimality Conditions

subsec:3-1

We begin this subsection by ensuring the existence of solutions for problem (1.1) under assumptions (A1) and (A3) (note that (A4) implies (A3) since every convex function has an affine minorant).

exist:solutions

Proposition 3.1. Let h and f be such that f is differentiable and assumptions (Ai) with $i = 1, 3$ hold. Then $\operatorname{argmin}_{\mathbb{R}^n}(h + f) \neq \emptyset$ and compact.

Proof. Since h is strongly quasiconvex, it is 2-supercoercive by Lemma 2.1, and since f is 2-weakly coercive, its sum is supercoercive too because

$$\liminf_{\|x\| \rightarrow +\infty} \frac{h(x) + f(x)}{\|x\|^2} \geq \liminf_{\|x\| \rightarrow +\infty} \frac{h(x)}{\|x\|^2} + \liminf_{\|x\| \rightarrow +\infty} \frac{f(x)}{\|x\|^2} > 0.$$

Since $h + f$ is lsc, $\operatorname{argmin}_{\mathbb{R}^n}(h + f) \neq \emptyset$ and compact. $\square$

not:finite

Remark 3.2. The number of solutions of problem (1.1) is not necessarily finite even under assumptions (Ai), with $i = 1, 2, 4$. Indeed, let us consider the functions $h, f : \mathbb{R} \rightarrow \mathbb{R}$ given by

$$h(x) = \max\{x, x^2\}, \quad f(x) = 1 - x.$$

Here h is strongly quasiconvex with modulus $\gamma_h = \frac{1}{2}$ by Remark 2.1(iii), while f is differentiable and convex. However, $\min_{x \in \mathbb{R}}(h(x) + f(x)) = [0, 1]$.

Our first main result, which provides necessary conditions for problem (1.1) under the assumption (A1), is given below.

nece:cond

Theorem 3.1. Suppose that h satisfies assumption (A1), f is differentiable on $\mathbb{R}^n$ and $\bar{x} \in \operatorname{argmin}_{\mathbb{R}^n}(h + f)$. Then

$$0 \in \partial_{1, \gamma_h} h(\bar{x}) + \nabla f(\bar{x}). \quad (3.3)$$

nece:1

If in addition f satisfies (A2), then

$$0 \in \partial_{\frac{1}{L}, \gamma_h}^{\mathbb{R}^n} h(\bar{x}) + \frac{1}{L} \nabla f(\bar{x}). \quad (3.4) \quad \boxed{\text{nec:2}}$$

Proof. Let $\bar{x} \in \operatorname{argmin}_{\mathbb{R}^n} (h + f)$. Then, for every $y \in K$ and every $\lambda \in [0, 1]$, it follows from (A1) that

$$\begin{aligned} h(\bar{x}) + f(\bar{x}) &\leq h(\lambda y + (1 - \lambda)\bar{x}) + f(\lambda y + (1 - \lambda)\bar{x}) \\ &\leq \max\{h(y), h(\bar{x})\} - \lambda(1 - \lambda) \frac{\gamma_h}{2} \|y - \bar{x}\|^2 + f(\lambda y + (1 - \lambda)\bar{x}). \end{aligned} \quad (3.5) \quad \boxed{\text{for:nec}}$$

If $h(y) \leq h(\bar{x})$, then

$$\begin{aligned} (1 - \lambda) \frac{\gamma_h}{2} \|y - \bar{x}\|^2 &\leq \frac{f(\lambda y + (1 - \lambda)\bar{x}) - f(\bar{x})}{\lambda} \\ &\stackrel{\lambda \downarrow 0}{\implies} \frac{\gamma_h}{2} \|y - \bar{x}\|^2 \leq \langle \nabla f(\bar{x}), y - \bar{x} \rangle. \end{aligned}$$

Therefore, $0 \in \partial_{1, \gamma} h(\bar{x}) + \nabla f(\bar{x})$ and relation (3.3) holds.

For (3.4): Assume that f satisfies in addition (A2). Then, if $h(\bar{x}) < h(y)$ in (3.5), we have

$$\begin{aligned} h(\bar{x}) + f(\bar{x}) &\leq h(y) - \lambda(1 - \lambda) \frac{\gamma_h}{2} \|y - \bar{x}\|^2 + f(\lambda y + (1 - \lambda)\bar{x}) \\ &\leq h(y) - \lambda(1 - \lambda) \frac{\gamma_h}{2} \|y - \bar{x}\|^2 + f(\bar{x}) + \lambda \langle \nabla f(\bar{x}), y - \bar{x} \rangle + \frac{\lambda^2 L}{2} \|y - \bar{x}\|^2. \end{aligned} \quad (3.6) \quad \boxed{\text{res:02}}$$

We separate the proof in two cases. Let us consider $L = 1$ in (3.6). Then,

$$\begin{aligned} h(y) &\geq h(\bar{x}) + \lambda \langle -\nabla f(\bar{x}), y - \bar{x} \rangle + \lambda(1 - \lambda) \frac{\gamma_h}{2} \|y - \bar{x}\|^2 - \frac{\lambda^2}{2} \|y - \bar{x}\|^2 \\ &\geq h(\bar{x}) + \lambda \langle -\nabla f(\bar{x}), y - \bar{x} \rangle + \frac{\lambda}{2} (\gamma_h - \lambda \gamma_h - \lambda) \|y - \bar{x}\|^2, \quad \forall \lambda \in [0, 1]. \end{aligned}$$

Hence, $-\nabla f(\bar{x}) \in \partial_{1, \gamma_h}^D h(\bar{x})$ with $D := (S_{h(\bar{x})}(h))^c$ (the complement of $S_{h(\bar{x})}(h)$). Then, it follows from this and (3.3) that $-\nabla f(\bar{x}) \in \partial_{1, \gamma_h}^{\mathbb{R}^n} h(\bar{x})$.

Finally, suppose that $L \neq 1$. Let $h' = \frac{1}{L}h$ and $f' = \frac{1}{L}f$. Then, $\bar{x} \in \operatorname{argmin}_{\mathbb{R}^n} (h + f)$ if and only if $\bar{x} \in \operatorname{argmin}_{\mathbb{R}^n} (h' + f')$. In addition, h' satisfies the assumption (A1) with the modulus $\gamma'_h = \frac{1}{L}\gamma_h$ and f' satisfies the assumption (A2) with the Lipschitz constant $L' = 1$. Hence, it follows from the previous case that

$$0 \in \partial_{1, \gamma'_h}^n h'(\bar{x}) + \nabla f'(\bar{x}) \iff 0 \in \partial_{\frac{1}{L}, \gamma_h}^{\mathbb{R}^n} h(\bar{x}) + \frac{1}{L} \nabla f(\bar{x}),$$

which completes the proof. $\square$

The previous necessary condition is not sufficient for global minimum even under assumptions (Ai) with $i = 1, 2, 4$.

s:inclu

Remark 3.3. Let $h, f : \mathbb{R} \rightarrow \mathbb{R}$ be the functions given by

$$h(x) = \max\{\sqrt{|x|}, x^2\}, \quad f(x) = -\frac{3}{2}x.$$

Then h is strongly quasiconvex with modulus $\gamma_h = \frac{1}{2} > 0$ by Remark 2.1. Since $\operatorname{argmin}_{\mathbb{R}}(h + f) = \{1\}$ and $\operatorname{argminloc}_{\mathbb{R}}(h + f) = \{0, 1\}$, we take $\bar{x} = 0$. Then, by [23, Remark 20] we have

$$\partial_{1, \frac{1}{2}}^{\mathbb{R}} h(0) = \left[-\frac{3}{2}, \frac{3}{2} \right],$$

i.e., $-\nabla f(0) = \frac{3}{2} \in \partial_{1, \frac{1}{2}}^{\mathbb{R}} h(0)$, but $0 \notin \operatorname{argmin}_{\mathbb{R}}(h + f)$.

Under the assumptions in Theorem 3.1, we have the following necessary conditions for local minimizers.

nece:locmin

Corollary 3.1. Under the assumptions in Theorem 3.1: If $\bar{x}$ is a local minimizer of $h + f$, then there exist $\epsilon > 0$ such that

$$0 \in \partial_{1, \gamma_h}^{B(\bar{x}, \epsilon) \cap S_{h(\bar{x})}(h)}(\bar{x}) + \nabla f(\bar{x}).$$

In addition, if f satisfies (A2), then

$$0 \in \partial_{\frac{1}{L}, \gamma_h}^{B(\bar{x}, \epsilon)}(\bar{x}) + \frac{1}{L} \nabla f(\bar{x}).$$

Proof. Just note that if $\bar{x}$ is a local minimizer of $h + f$, then there exist $\epsilon > 0$ such that

$$h(\bar{x}) + f(\bar{x}) \leq h(y) + f(y), \quad \forall y \in B(\bar{x}, \epsilon),$$

and the proof follows exactly as in Theorem 3.1. $\square$

If the necessary condition holds, then we have the following result whenever f satisfies (A4), i.e., f is convex.

suf:e-sol

Proposition 3.2. Let h be a proper function, f be a differentiable function such that (A4) holds and $\bar{x} \in \operatorname{dom}(h + f)$. If there exists $\beta > 0$ such that

$$0 \in \partial_{\beta, \gamma_h}^{\mathbb{R}^n} h(\bar{x}) + \beta \nabla f(\bar{x}), \quad (3.7)$$

nes_beta

then the following assertions hold:

- (a) $f(\bar{x}) + \frac{\gamma_h}{2} \|y - \bar{x}\|^2 \leq f(y)$ for all $y \in S_{h(\bar{x})}(h)$.
- (b) $h(\bar{x}) + f(\bar{x}) \leq h(y) + f(y) + \frac{1}{2\beta} \|y - \bar{x}\|^2$ for all $y \notin S_{h(\bar{x})}(h)$.

Proof. If $-\beta \nabla f(\bar{x}) \in \partial_{\beta, \gamma_h}^{\mathbb{R}^n} h(\bar{x})$, then for all $y \in \mathbb{R}^n$ and all $\lambda \in [0, 1]$, we have

$$\begin{aligned}
\max\{h(y), h(\bar{x})\} &\geq h(\bar{x}) + \frac{\lambda}{\beta} \langle -\beta \nabla f(\bar{x}), y - \bar{x} \rangle + \frac{\lambda}{2} \left(\gamma_h - \frac{\lambda}{\beta} - \lambda \gamma_h \right) \|y - \bar{x}\|^2 \\
&\geq h(\bar{x}) + \lambda(f(\bar{x}) - f(y)) + \frac{\lambda}{2} \left(\gamma_h - \frac{\lambda}{\beta} - \lambda \gamma_h \right) \|y - \bar{x}\|^2, \quad (3.8)
\end{aligned}$$

for:suf

where the last inequality follows from (A4). Then we have two cases:

(i): If $h(y) \leq h(\bar{x})$, then the result follows by taking $\lambda = 0$ in (3.8).

(ii): If $h(y) > h(\bar{x})$, then the result follows by taking $\lambda = 1$ in (3.8). $\square$

We finish this section with a simple case where condition (3.7) becomes a sufficient optimality condition.

Corollary 3.2. *Let h and f be such that assumptions (Ai) with $i = 1, 2, 4$ hold. If $\operatorname{argmin}_{\mathbb{R}^n} h \subseteq \operatorname{argmin}_{\mathbb{R}^n} f$, then*

$$\operatorname{argmin}_{\mathbb{R}^n} (h + f) = \{x^*\} = \operatorname{argmin}_{\mathbb{R}^n} h. \quad (3.9)$$

If in addition there exists $\bar{x} \in \operatorname{dom}(h + f)$ which satisfies condition (3.7) for some $\beta > 0$, then $\bar{x} = x^$.*

cor1

Proof. Since h is lsc and strongly quasiconvex, $\operatorname{argmin}_{\mathbb{R}^n} h$ is a singleton. Let $x^* = \operatorname{argmin}_{\mathbb{R}^n} h \subseteq \operatorname{argmin}_{\mathbb{R}^n} f$ and suppose for the contrary that there exists $\hat{x} \in \operatorname{argmin}_{\mathbb{R}^n} (h + f)$ with $\hat{x} \neq x^*$. Then, $(h + f)(\hat{x}) \leq (h + f)(x)$ for all $x \in \mathbb{R}^n$. Thus,

$$(h + f)(\hat{x}) \leq (h + f)(x^*) \iff h(\hat{x}) - h(x^*) \leq f(x^*) - f(\hat{x}) \leq 0,$$

Hence, $h(\hat{x}) \leq h(x^*)$, a contradiction. Therefore, $\hat{x} = x^*$.

Now, let us assume that there exists $\bar{x} \in \mathbb{R}^n$ satisfies the condition (3.7) for some $\beta > 0$. Hence, $h(x^*) \leq h(\bar{x})$ and $f(x^*) \leq f(\bar{x})$, which implies that $x^* \in S_{h(\bar{x})}(h)$ and, by Proposition 3.2, we obtain

$$f(\bar{x}) + \frac{\gamma_h}{2} \|x^* - \bar{x}\|^2 \leq f(x^*).$$

Therefore, $\bar{x} = x^*$. $\square$

3.2 A Proximal Gradient Algorithm

subsec3-2

In this subsection, we study a proximal gradient algorithm for dealing with problem (1.1). To that end, we recall and introduce the following notions.

statio:point

Definition 3.1. *Let $h, f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be such that f is differentiable, $\beta > 0$ and $\gamma \geq 0$. Then $x^* \in \operatorname{dom} h$ is said to be:*

(a) *A l -critical point (see [3]) for problem (1.1) if*

$$0 \in \nabla f(x^*) + \partial^{\lim} h(x^*). \quad (3.10)$$

eq:crit

(b) *A s -critical point for problem (1.1) corresponding to β if*

$$0 \in \beta \nabla f(x^*) + \partial_{\beta, \gamma}^{\mathbb{R}^n} h(x^*). \quad (3.11)$$

eq:statio

In virtue of (2.16), if x^* satisfies (3.11), then $0 \in \beta \nabla f(x^*) + \partial_{\beta, \gamma} h(x^*)$, too.

best:info

Remark 3.4. (Importance of s -criticality instead of l -criticality) *When dealing with problem (1.1) under assumptions (Ai) with $i = 1, 2, 3$, ensures that s -criticality of a point (relation (3.11)) could be much better than l -criticality (relation (3.10)) since the strong subdifferential provide more useful information for nonsmooth strongly quasiconvex function than the limiting subdifferential even for continuous functions in the one-dimensional case as we recall below. Indeed, let $h : \mathbb{R} \rightarrow \mathbb{R}$ be given by*

$$h(x) = \begin{cases} \sqrt{|x|}, & \text{if } |x| \leq 1, \\ +\infty, & \text{otherwise.} \end{cases}$$

By Remark 2.1(i), h is strongly quasiconvex with modulus $\gamma_h = \frac{1}{2} > 0$. Take $\bar{x} = 0$ and $\beta = 1$. Then, by [23, Remark 20] we have

$$\partial_{1, \frac{1}{2}}^{\mathbb{R}} h(0) = \left[-\frac{3}{2}, \frac{3}{2} \right],$$

while $\partial^{\lim} h(0) = \mathbb{R}$.

Therefore, our goal is to ensure s -criticality of the generated sequence of the Proximal gradient algorithm in its classical version:

Algorithm 1 Proximal gradient Algorithm for Nonconvex Nonsmooth Minimization (PGANNM)

Prox:gradient

Step 0. Take $x^1 \in \text{dom}(h + f)$, $k = 1$ and sequences $\{\beta_k\}_k \subseteq \mathbb{R}_{++}$.

Step 1. Compute

$$y^k = x^k - \beta_k \nabla f(x^k). \quad (3.12)$$

$$x^{k+1} \in \text{Prox}_{\beta_k h}(y^k). \quad (3.13)$$

step:y

step:x

Step 2. If $x^{k+1} = x^k$, then STOP, x^k is a s -critical point of problem (1.1). Otherwise, take $k = k + 1$ and go to Step 1.

Algorithm 1 has been studied deeply in the literature (see, for instance, [3, 5, 6, 7, 11, 10, 20, 24, 29, 31, 33]). In particular, in [3], a general version of the proximal gradient algorithm for minimizing the sum of a nonconvex extended real-valued function h and a L -smooth function f was studied under the (PKL) property. If the restriction of h on its domain is continuous and the sequence generated by this algorithm is bounded, then it converges to a l -critical point $\bar{x}$ of $f + h$ (see [3, Theorem 5.1]), i.e., the sequence $\{x^k\}_k$ converges to a point $\bar{x}$ such that

$$0 \in \partial^{\lim}(h + f)(\bar{x}).$$

In this subsection, we will study Algorithm 1 when h is strongly quasiconvex and nonsmooth and f is L -smooth and possible nonconvex too. Thanks to the

strong quasiconvexity of h and the optimality conditions developed in Subsection 3.1, we will show that the generated sequence $\{x^k\}_k$ converges also to a s -critical point as defined in Definition 3.1(b) which is, at least, a quasi solution for problem (1.1) (see relation (2.20)).

First of all, let us present here important results of the sequence generated by the proximal gradient method in our setting. The properties in part (a), (c), (d) also presented in [3, Theorem 5.1] under the assumptions that $h + f$ is a (PKL) function and the sequence $\{x^k\}$ is bounded.

rel:k:k+1

Proposition 3.3. *Let h and f be such that assumption (Ai) with $i = 1, 2$ hold and $\{\beta_k\}_k$ be a sequence of positive numbers. Then for every $k \in \mathbb{N}$, we have*

$$(h + f)(x^{k+1}) + \frac{(1 - \beta_k L)}{2\beta_k} \|x^{k+1} - x^k\|^2 \leq (h + f)(x^k). \quad (3.14)$$

eq:decrease_1

If in addition (A3) holds and $0 < \beta_k \leq \frac{1}{L}$ for all $k \in \mathbb{N}$, then

- (a) $(f + h)(x^{k+1}) \leq (f + h)(x^k)$ for all $k \in \mathbb{N}$.
- (b) The sequence $\{x^k\}_k$ is bounded.
- (c) The sequence $\{(f + h)(x^k)\}_k$ is convergent.

- (d) The sequence $\{x^k\}_k$ has a finite length, i.e., $\sum_{k=k_0}^{\infty} \|x^{k+1} - x^k\|^2 < \infty$.

Proof. (a): By [3], relation (3.14) holds, and since $\beta_k \leq \frac{1}{L}$ for all $k \in \mathbb{N}$, we have

$$(h + f)(x^{k+1}) \leq (h + f)(x^k), \quad \forall k \in \mathbb{N}. \quad (3.15)$$

(b): Since $h + f$ is coercive by assumptions (A1) and (A3), $S_{(h+f)(x^1)}(h + f)$ is bounded, thus the sequence $\{x^k\}_k \subseteq S_{(h+f)(x^1)}(h + f)$ is bounded too.

(c)-(d): In virtue of (3.14), parts (a) and (b) proved above, it follows from Lemma 2.4 that the sequence $\{(f + h)(x^k)\}_k$ is convergent and that $\sum_{k=k_0}^{\infty} \|x^k - x^{k+1}\|^2 < \infty$, which completes the proof. $\square$

Basic results are given below. For the first one, note that there is no assumption on h and that a sufficient condition for the existence of iterates x^{k+1} is that h is lsc and 2-weakly coercive (see [14, Proposition 3.1]).

prop:start-sol

Proposition 3.4. *Let h and f be such that f is differentiable and $\beta > 0$. Then for any $x \in \mathbb{R}^n$ and $y = x - \beta \nabla f(x)$, the following assertions hold:*

- (a) If f satisfies (A2) and $\beta \leq \frac{1}{L}$, then

$$x \in \operatorname{argmin}_{\mathbb{R}^n} (h + f) \implies x \in \operatorname{Prox}_{\beta h}(y). \quad (3.16)$$

- (b) Suppose in addition that f is strongly convex with modulus $\gamma_f > 0$. If $\beta \geq \frac{1}{\gamma_f}$ for all $k \in \mathbb{N}$, then

$$x \in \operatorname{Prox}_{\beta h}(y) \implies x \in \operatorname{argmin}_{\mathbb{R}^n} (h + f). \quad (3.17)$$

Proof. It is straightforward. $\square$

k=k+1

Proposition 3.5. *Let h be such that assumption (A1) holds, f be differentiable and $\{\beta_k\}_k$ be a sequence of positive numbers. Then the following assertions hold:*

(a) *If $x^{k+1} = x^k$, then*

$$0 \in \partial_{\beta_k, \gamma_h}^{\mathbb{R}^n} h(x^k) + \beta_k \nabla f(x^k), \quad (3.18) \quad \text{nece:cond1}$$

(b) *If f satisfies (A4) and $x^{k+1} = x^k$, then*

$$x^k \in \text{Fix}(\text{Prox}_{\beta_k F}(\cdot)). \quad (3.19) \quad \text{suffcond1}$$

Proof. (b): It is straightforward. (a): Since $x^{k+1} = x^k$ and $x^{k+1} \in \text{Prox}_{\beta_k h}(y^k)$, $y^k - x^k = -\beta_k \nabla f(x^k) \in \partial_{\beta_k, \gamma_h}^{\mathbb{R}^n} h(x^k)$ by Lemma 2.3 and Step (3.12). $\square$

Remark 3.5. (i) *If in addition to the assumptions in Proposition 3.5 we assume that one of the following condition holds*

(a1) *F is convex;*

(a2) *F is strongly quasiconvex with modulus $\gamma_F > 0$;*

then x^k is an optimal solution of (1.1) (see [26, Theorem 10] for (a2)).

(ii) *Condition (3.19) becomes necessary for the optimality of $F = f + h$ when $\beta_k = 1/L$. Furthermore, for $\alpha \geq \beta > 0$ and $\xi \in \mathbb{R}^n$, we have*

$$-\alpha \xi \in \partial_{\alpha, \gamma_h}^{\mathbb{R}^n} h(x) \implies -\beta \xi \in \partial_{\beta, \gamma_h}^{\mathbb{R}^n} h(x).$$

Our second main result, which shows that every cluster point of the sequence $\{x^k\}_k$, generated by Algorithm 1, is a s -critical point of problem (1.1), is given below under a continuity assumption on h .

cluster:point1

Theorem 3.2. *Let $h, f : \mathbb{R}^n \rightarrow \mathbb{R}$ be two real-valued functions such that assumptions (Ai) with $i = 1, 2, 3$ hold and $\{\beta_k\}_k$ be a sequence of positive numbers. If h is continuous and*

$$0 < \lim_{k \rightarrow +\infty} \beta_k = \beta' < \frac{1}{L}, \quad (3.20) \quad \text{cluster:cond}$$

then every cluster point $\hat{x}$ of the sequence $\{x^k\}_k$ is a s -critical point of problem (1.1), that is,

$$0 \in \partial_{\beta', \gamma_h}^{\mathbb{R}^n} h(\hat{x}) + \beta' \nabla f(\hat{x}). \quad (3.21) \quad \text{station}$$

Proof. Since $\{x^k\}_k$ is bounded, it has cluster points. Let $\hat{x}$ be a cluster point of $\{x^k\}_k$. Then there exists a subsequence $\{x^{k_\ell}\}_\ell \subset \{x^k\}_k$ such that $x^{k_\ell} \rightarrow \hat{x}$ as $\ell \rightarrow +\infty$. Combining this with Proposition 3.3(d), we obtain

$$\lim_{\ell \rightarrow \infty} x^{k_\ell+1} = \hat{x}.$$

By relation (3.13) and Lemma 2.3, $y^k - x^{k+1} \in \partial_{\beta_k, \gamma_h}^{\mathbb{R}^n} h(x^{k+1})$. Then for every $y \in \mathbb{R}^n$ and every $\lambda \in [0, 1]$, we have

$$\begin{aligned}
& \max\{h(y), h(x^{k+1})\} \\
& \geq h(x^{k+1}) + \frac{\lambda}{\beta_k} \langle y^k - x^{k+1}, y - x^{k+1} \rangle + \frac{\lambda}{2} \left(\gamma_h - \frac{\lambda}{\beta_k} - \lambda \gamma_h \right) \|y - x^{k+1}\|^2 \\
& = h(x^{k+1}) + \frac{\lambda}{\beta_k} \langle x^k - \beta_k \nabla f(x^k) - x^{k+1}, y - x^{k+1} \rangle \\
& + \frac{\lambda}{2} \left(\gamma_h - \frac{\lambda}{\beta_k} - \lambda \gamma_h \right) \|y - x^{k+1}\|^2.
\end{aligned} \tag{3.22}$$

strongine

By replacing k by k_ℓ and taking the limit when $\ell \rightarrow +\infty$, we obtain from the continuity of h and ∇f , and assumption (3.20) that

$$\begin{aligned}
\max\{h(y), h(\hat{x})\} & \geq h(\hat{x}) + \lambda \langle \nabla f(\hat{x}), y - \hat{x} \rangle + \frac{\lambda}{2} \left(\gamma_h - \frac{\lambda}{\beta'} - \lambda \gamma_h \right) \|y - \hat{x}\|^2, \\
& \forall y \in \mathbb{R}^n, \forall \lambda \in [0, 1].
\end{aligned} \tag{3.23}$$

Therefore, $-\beta' \nabla f(\hat{x}) \in \partial_{\beta', \gamma_h}^{\mathbb{R}^n} h(\hat{x})$, and the proof is complete. $\square$

If in addition we assume (A4) instead of (A3), i.e., f is convex, we have:

cluster:point2

Proposition 3.6. *Let $h, f : \mathbb{R}^n \rightarrow \mathbb{R}$ be two real-valued functions such that assumptions (Ai) with $i = 1, 2, 4$ hold and $\{\beta_k\}_k$ be a sequence of positive numbers such that condition (3.20) holds. Then every cluster point $\hat{x}$ of the sequence $\{x^k\}_k$ is a $\frac{1}{2\beta'}$ -quasi solution of problem (1.1), that is,*

$$(h + f)(\hat{x}) \leq (h + f)(x) + \frac{1}{2\beta'} \|\hat{x} - x\|^2, \quad \forall x \in \mathbb{R}^n. \tag{3.24}$$

beta:quasi:sol

Proof. Since (A4) $\Rightarrow$ (A3), it follows from Proposition 3.3(b) that the sequence $\{x^k\}_k$ is bounded, it has cluster points. Let $\hat{x} \in \mathbb{R}^n$ be a cluster point, then there exists a subsequence $\{x^{k_\ell}\}_\ell$ such that $x^{k_\ell} \rightarrow \hat{x}$ as $\ell \rightarrow +\infty$.

Now, by step (3.13), we have for every k that

$$\begin{aligned}
& h(x^{k+1}) + \frac{1}{2\beta_k} \|x^{k+1} - y^k\|^2 \leq h(x) + \frac{1}{2\beta_k} \|x - y^k\|^2 \\
& \iff h(x^{k+1}) \leq h(x) - \frac{1}{\beta_k} \langle x - x^{k+1}, y^k - x \rangle - \frac{1}{2\beta_k} \|x^{k+1} - x\|^2.
\end{aligned}$$

Then, by step (3.12), we have

$$h(x^{k+1}) \leq h(x) - \frac{1}{\beta_k} \langle x - x^{k+1}, x^k - x \rangle + \langle \nabla f(x^k), x - x^{k+1} \rangle - \frac{1}{2\beta_k} \|x^{k+1} - x\|^2.$$

By replacing k by k_ℓ , and since f and h are continuous and assumption (3.20) holds, then by taking the limit when $\ell \rightarrow +\infty$, we have

$$\begin{aligned} h(\hat{x}) &\leq h(x) - \frac{1}{\beta'} \langle x - \hat{x}, \hat{x} - x \rangle + \langle \nabla f(\hat{x}), x - \hat{x} \rangle - \frac{1}{2\beta'} \|\hat{x} - x\|^2 \\ &= h(x) + \langle \nabla f(\hat{x}), x - \hat{x} \rangle + \frac{1}{2\beta'} \|\hat{x} - x\|^2 \\ &\leq h(x) + f(x) - f(\hat{x}) + \frac{1}{2\beta'} \|\hat{x} - x\|^2, \quad \forall x \in \mathbb{R}^n, \quad (\text{by (A4)}). \end{aligned}$$

Therefore, relation (3.24) holds and the proof is complete. $\square$

Under the extra usual assumption that the function $f+h$ satisfies the (PKL) property, we can prove that the whole sequence $\{x^k\}_k$ is convergent and that the limit point has the previous interesting properties.

main:theo

Theorem 3.3. *Let $h, f : \mathbb{R}^n \rightarrow \mathbb{R}$ be two real-valued functions such that assumptions (Ai) with $i = 1, 2, 3$ hold, h continuous and $\{\beta_k\}_k$ be a sequence of positive numbers. Suppose that condition (3.20) holds. If the function $f + h$ satisfies the (PKL) property, then the sequence $\{x^k\}_k$ converges to a point x^* which satisfies the following properties:*

- (a) x^* is a l -critical point of problem (1.1).
- (b) x^* is a s -critical point of problem (1.1) corresponding to β' .
- (c) If in addition f satisfies (A4), then x^* is also a $\frac{1}{2\beta'}$ -quasi solution of problem (1.1), i.e., satisfies relation (3.24).

Proof. (a): Since $f + h$ satisfies the (PKL) property and the sequence $\{x^k\}_k$ is bounded by Proposition 3.3, it follows from [3, Theorem 5.1] that the sequence $\{x^k\}_k$ is convergent to a point x^* and that

$$0 \in \partial^{\lim}(f + h)(x^*). \quad (3.25)$$

cric:point

Since f is continuously differentiable, (3.25) implies $0 \in \nabla f(x^*) + \partial^{\lim}(h)(x^*)$, i.e., x^* is a l -critical point for problem (1.1).

(b)-(c): By part (a) we have $x^k \rightarrow x^*$ and by Propositions 3.2 and 3.6, every cluster point is a s -critical point and a $\frac{1}{2\beta'}$ -quasi solution for problem (1.1), respectively, which completes the proof. $\square$

Note that in Theorem 3.3, the continuity assumption on h is needed for (b) and (c), but it is not necessary for (a) as proved in [11] under lower semicontinuity. On the other hand, the rate of convergence of the proximal gradient method under local Lipschitz assumption together with (PKL) property was provided in case that the desingularization function Φ has the form $\phi(s) = cs^k$ for $k \in]0, 1]$ in [20, Theorem 4.6].

In virtue of Theorem 3.3 and Proposition 3.3, we have the following complexity analysis.

comp:anal

Proposition 3.7. *Let $h, f : \mathbb{R}^n \rightarrow \mathbb{R}$ be two real-valued functions such that assumptions (Ai) with $i = 1, 2, 3$ hold and suppose that $0 < \beta_k \leq \frac{1}{L}$ for all k . Then for every $\varepsilon > 0$, we need at most*

$$N \geq \frac{2}{\varepsilon L} ((h + f)(x^1) - \min_{x \in \mathbb{R}^n} (h + f)(x)), \quad (3.26)$$

comp:rate

iterations to satisfy the stopping criteria $\|x^{k+1} - x^k\|^2 < \varepsilon$.

Proof. By relation (3.14), we have:

$$\frac{1}{2\beta_k} \|x^{k+1} - x^k\|^2 \leq (h + f)(x^k) - (h + f)(x^{k+1}), \quad \forall k \in \mathbb{N}.$$

By telescoping this over the first N iterations and since $0 < \beta_k \leq \frac{1}{L}$, we have

$$\begin{aligned} \sum_{k=1}^N \|x^{k+1} - x^k\|^2 &\leq \frac{2}{L} \sum_{k=1}^N ((h + f)(x^k) - (h + f)(x^{k+1})) \\ &= \frac{2}{L} (h + f)(x^1) - (h + f)(x^{N+1}) \\ &\leq \frac{2}{L} \left((h + f)(x^1) - \min_{x \in \mathbb{R}^n} (h + f)(x) \right). \end{aligned}$$

Hence,

$$N \left(\min_{1 \leq i \leq N} \|x^{i+1} - x^i\|^2 \right) \leq \frac{2}{L} \left((h + f)(x^1) - \min_{x \in \mathbb{R}^n} (h + f)(x) \right),$$

and the result follows. $\square$

We finish this section with the following examples of strongly quasiconvex functions (which are not convex) that satisfy the (PKL) property.

Example 3.1. *Let us consider the function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ given by $h(x) = \sqrt{\|x\|}$, which is strongly quasiconvex on $\mathbb{B}(0, r)$, $r > 0$, with modulus $\gamma = \frac{1}{5^{\frac{1}{4}} 2^{\frac{5}{4}} r^{\frac{1}{2}}}$ by Remark 2.1(i) and its minimum point is $\bar{x} = 0$. Hence, it follows from relation (2.8) with $\bar{x} = 0$ that*

$$h(y) \geq h(0) + \frac{1}{85^{\frac{1}{4}} 2^{\frac{5}{4}} r^{\frac{1}{2}}} \|y\|^2, \quad \forall y \in \mathbb{B}(0, r). \quad (3.27)$$

Following [4, Subsection 4.2.2], we claim that h satisfies the (PKL) property with $\phi(s) = 4r^{\frac{3}{4}} s^{\frac{1}{2}}$ on $\mathbb{B}(0, r)$ for all $r > 0$.

Clearly, ϕ satisfies conditions (K1), (K2) and (K3) in Definition 2.1. Now, let us prove that relation (2.23) holds. Indeed, since h is differentiable on

$\mathbb{B}(0, r) \setminus \{0\}$, relation (2.23) for $\bar{x} = 0$ becomes

$$\begin{aligned}\phi'(h(x) - 0) \operatorname{dist}(0, \nabla h(x)) &= 4r^{\frac{3}{4}} \frac{1}{2} \left(\frac{1}{\sqrt{\|x\|}} \right)^{\frac{1}{2}} \left\| 0 - \frac{1}{2} \frac{1}{\|x\|^{\frac{3}{2}}} x \right\| \\ &= \left(\frac{r}{\|x\|} \right)^{\frac{3}{4}} \\ &\geq 1,\end{aligned}$$

i.e., h satisfies the (PKL) property.

Another interesting case of strongly quasiconvex functions which satisfies the (PKL) property is given below.

Example 3.2. Let $A, B \in \mathbb{R}^{n \times n}$, $a, b \in \mathbb{R}^n$, $\alpha, \beta \in \mathbb{R}$. Let us consider the functions $f(x) = \frac{1}{2} \langle Ax, x \rangle + \langle a, x \rangle + \alpha$ and $g(x) = \langle b, x \rangle + \beta$. Take $0 < m < M$ and $K = \{x \in \mathbb{R}^n : m \leq \langle b, x \rangle + \beta \leq M\}$. We consider the function:

$$h(x) = \begin{cases} \frac{\frac{1}{2} \langle Ax, x \rangle + \langle a, x \rangle + \alpha}{\langle b, x \rangle + \beta}, & \text{if } x \in K, \\ +\infty, & \text{otherwise.} \end{cases} \quad (3.28)$$

If A is definite positive, then h is strongly quasiconvex on K with modulus $\gamma = \frac{\lambda_{\min}(A)}{M} > 0$ by Remark 2.1(iii).

This function is semi-algebraic since its graph is a semi-algebraic subset of $\mathbb{R}^{n+1}$. Indeed,

$$\begin{aligned}& \{(x, t) \in \mathbb{R}^{n+1} : h(x) = t\} \\ &= \{(x, t) \in \mathbb{R}^{n+1} : f(x) = tg(x)\} \cap \{(x, t) \in \mathbb{R}^{n+1} : x \in K\}\end{aligned}$$

Therefore, by [9, Theorem 3], the function h satisfies the (PKL) property at any point of $\operatorname{dom} h$.

Fractional functions have been deeply studied in continuous optimization due to its interesting applications in economics, finance and management among others (see [12, 36, 37, 38] and references therein).

4 Applications and Numerical Experiments

sec:4

In this section, we apply our theoretical results in the nonconvex nonsmooth optimization problems described below.

4.1 DC Programming

subsec:4-1

Let us recall the DC optimization problem:

$$\min_{x \in \mathbb{R}^n} (h_c(x) - f_c(x)), \quad (4.1) \quad \text{dc:problem1}$$

where $h_c, f_c : \mathbb{R}^n \rightarrow \mathbb{R}$ are convex functions and f_c is differentiable with L_c -Lipschitz continuous gradient.

As stated in Remark 3.1, for any $\rho > 0$, problem (4.1) can be equivalently written with strongly convex functions, that is,

$$\min_{x \in \mathbb{R}^n} \{h_1(x) - f_1(x)\}, \quad (4.2) \quad \boxed{\text{sdc:problemequiv}}$$

where $h_1(x) = h_c(x) + \frac{\rho}{2}\|x\|^2$ and $f_1(x) = f_c(x) + \frac{\rho}{2}\|x\|^2$ are strongly convex functions with modulus $\rho > 0$.

The following DC algorithm was considered in [39], and then improved in [1, 2] (Boosted DC Algorithm) for constrained DC programming can be applied for solving (4.2) as follows: Given $x^0 \in \mathbb{R}^n$, we compute

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \{h_1(x) - \langle \nabla f_1(x^k), x \rangle\}, \quad \forall k \geq 0. \quad (4.3) \quad \boxed{\text{alg:LDM}}$$

We can also transform problem (4.1) to

$$\min_{x \in \mathbb{R}^n} \{h(x) + f(x)\}, \quad (4.4) \quad \boxed{\text{sdc:problemequiv1}}$$

where $h(x) = h_c(x) + \frac{\alpha}{2}\|x\|^2$ and $f(x) = -f_c(x) - \frac{\alpha}{2}\|x\|^2$ for any $\alpha > 0$.

In the following, we show that Algorithm (4.3) can be considered as a particular case of our Algorithm 1 for solving (4.4). Indeed, let us assume that $0 < \alpha < \rho$. Then the update rule (4.3) is equivalent with

$$\begin{aligned} x^{k+1} &= \operatorname{argmin}_{x \in \mathbb{R}^n} \{h_c(x) + \frac{\rho}{2}\|x\|^2 - \langle \nabla f_c(x^k) + \rho x^k, x \rangle\} \\ &= \operatorname{argmin}_{x \in \mathbb{R}^n} \{h(x) + \frac{\rho - \alpha}{2}\|x\|^2 + \langle \nabla f(x^k) - (\rho - \alpha)x^k, x \rangle\} \\ &= \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ h(x) + \frac{\rho - \alpha}{2} \left\| x - \left(x^k - \frac{1}{\rho - \alpha} \nabla f(x^k) \right) \right\|^2 \right\}. \end{aligned} \quad (4.5) \quad \boxed{\text{updateruledc}}$$

Clearly, (4.5) is the update rule of Algorithm 1 for solving (4.4) with $\beta_k = \frac{1}{\rho - \alpha}$ for all k .

Therefore, the following result provides sufficient conditions for the convergence of the generated sequence (4.5) to a l -critical and s -critical point.

main:theo2

Theorem 4.1. *Let h_c and f_c be two convex functions where h_c is continuous and f_c is continuously differentiable. Let us suppose the following:*

(B1) f_c has L_{f_c} -Lipschitz continuous gradient with $L_{f_c} > 0$.

(B2) The function $-f_c$ is 2-weakly coercive.

(B3) The function $h_c - f_c$ satisfies the (PKL) property.

Then by taking $\rho > L_{f_c}$ and $0 < \alpha < \frac{\rho - L_{f_c}}{2}$, we obtain that the sequence $\{x^k\}_k$, generated by the update rule (4.5), satisfies the following:

- (a) If $x^{k+1} = x^k$, then stop, x^k is a s -critical point corresponding to $\frac{1}{\rho-\alpha}$ for problem (4.4).
- (b) The sequence $\{x^k\}_k$ converges to a point x^* which is l -critical and s -critical for problem (4.4).

Proof. Since f_c has a L_{f_c} -Lipschitz continuous gradient with $L_{f_c} > 0$, we choose $\rho > L_{f_c}$ and $0 < \alpha < \frac{\rho - L_{f_c}}{2}$.

Let us define $h(x) = h_c(x) + \frac{\alpha}{2}\|x\|^2$ and $f(x) = -f_c(x) - \frac{\alpha}{2}\|x\|^2$. Since h_c is convex and $\alpha > 0$, h is strongly convex with modulus $\alpha > 0$, i.e., it is strongly quasiconvex with the same modulus. Furthermore, since f_c has a L_{f_c} -Lipschitz continuous gradient, then f has a $(L_{f_c} + \alpha)$ -Lipschitz continuous gradient. Furthermore, since $-f_c$ is 2-weakly coercive, $h_c - f_c$ is 2-supercoercive by Proposition 3.1, thus $h + f = h_c - f_c$ is 2-supercoercive too.

Then, h is strongly quasiconvex with modulus $\rho > 0$, f has a $(L_{f_c} + \rho)$ -Lipschitz continuous gradient, $h + f$ is coercive, $h + f = h_c - f_c$ satisfies the (PKL) property by assumption (P3) and

$$\frac{1}{\rho - \alpha} \leq \frac{1}{L_{f_c} + \alpha},$$

i.e., the sequence $\beta_k \equiv \frac{1}{\rho - \alpha}$ satisfies condition (3.20), thus part (a) follows from Proposition 3.5(a) while (b) follows from Theorem 3.3. $\square$

subsec:4-3

4.2 Numerical Experiments

In the following examples, we construct the function h as the maximum of a finite number of strongly convex and strongly quasiconvex functions in virtue of Remark 2.1, i.e., the resulting function is strongly quasiconvex and nonsmooth.

The algorithm was implemented and executed in Python on a ASUS Laptop with Windows 11 and an AMD Ryzen 7 5800H CPU with 16GB RAM. In the following numerical experiments, there is no closed form or any available solver for finding the global solution of the proximal operator of a nonconvex function h . The proximal step was computed by using function *minimize* of the SciPy.optimize package.

ex1 **Example 4.1.** ([15]) Let $q \in \mathbb{N}$ and $h_1, h_2 : \mathbb{R}^n \rightarrow \mathbb{R}$ given by $h_1(x) = \sqrt{\|x\|}$ and $h_2(x) = \|x\|^2 - q$. Both h_1 and h_2 are continuous functions. It is known that h_2 is strongly convex, hence also strongly quasiconvex, while according to Remark 2.1 the function h_1 is strongly quasiconvex on any convex and bounded set $K \subseteq \mathbb{R}^n$. Then the function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ defined by $h(x) := \max\{h_1(x), h_2(x)\}$ is continuous and strongly quasiconvex (as the maximum of two strongly quasiconvex functions) without being convex and nonsmooth.

Then we consider problem (1.1) with

$$h(x) := \max\{h_1(x), h_2(x)\} \quad \text{and} \quad f(x) = \|x - b\|^2,$$

where b is a given vector in $\mathbb{R}^n$. Clearly, f is a convex and differentiable function with Lipschitz continuous gradient, i.e., assumptions (Ai) with $i = 1, 2, 3$ hold.

In the first experiment, we test the proximal gradient algorithm for this problem with $p = 1$, $n = 5$ and $b = (0, 0, 0, 0, 0)^T$ with starting point $x^0 = (1, 1, 1, 1, 1)^T$ and stopping criteria $\|x^k - x^{k+1}\| < 10^{-5}$ or the number of iterations exceeds 1000. Under these assumptions, the unique solution of (1.1) is $x^* = (0, 0, 0, 0, 0)^T$ and the optimal value is 0. Table 1 reports the CPU time(s) and the error

$$err = \frac{\|x^k - x^*\|}{n},$$

with different choices of β_k .

Table 1: CPU time(s) and Error of the Proximal Gradient Algorithm in the first experiment with different choices of β_k in Example 4.1

	$\beta_k \equiv \frac{1}{5}$	$\beta_k \equiv \frac{1}{3}$	$\beta_k \equiv \frac{1}{2}$
CPU time(s)	0.03750324249267578	0.1652665138244629	0.018276453018188477
Error	1.0064413352965294e-06	8.680309102298729e-06	0

tab1

In the second experiment, the proximal gradient algorithm was tested for this problem with $p = 1$, $n = 10$ and $b = 1_{10}$ with starting point $x^0 = 0_{10}$ and stopping criteria $\|x^k - x^{k+1}\| < 10^{-5}$ or the number of iterations exceeds 1000. Under these assumptions, the unique solution of (1.1) is $x^* = \frac{1}{2} \times 1_{10}$ and the optimal value is 4. Table 1 reports the CPU time(s) and the error

$$err = \frac{\|x^k - x^*\|}{n},$$

with different choices of β_k .

Table 2: CPU time(s) and Error of the Proximal Gradient Algorithm in the second experiment with different choices of β_k in Example 4.1

	$\beta_k \equiv \frac{1}{5}$	$\beta_k \equiv \frac{1}{3}$	$\beta_k \equiv \frac{1}{2}$
CPU time(s)	0.08352446556091309	0.05029726028442383	0.004403352737426758
Error	8.213949607813013e-06	5.827785724233184e-06	2.216049577025723e-07

tab2

We emphasize that in the next example, we compute a family of 100 hundred problems randomly generated for each dimension $n = 20, 50$ and 100.

ex2 **Example 4.2.** Let us consider problem (1.1) when the function h is the maximum of a finite number of quadratic fractional functions:

$$h(x) = \max_{j \in J} \left\{ \frac{\frac{1}{2} \langle A_j x, x \rangle + \langle a_j, x \rangle + \alpha_j}{\langle b_j, x \rangle + \beta_j} \right\}, \quad (4.6) \quad \text{eq:const}$$

with $A_j \in \mathbb{R}^{n \times n}$ symmetric and definite positive, $a_j, b_j \in \mathbb{R}^n$ and $\alpha_j, \beta_j \in \mathbb{R}$ for all $j \in J$ and $2 \leq |J| < +\infty$. Take fixed $M > m > 0$. We choose the feasible

set K such that for any $x \in K$, we have

$$m \leq \langle b_j, x \rangle + \beta_j \leq M, \quad \forall j \in \{1, \dots, T\},$$

and K is convex and compact.

By Remark 2.1(ii), h is the maximum of $|T| < +\infty$ strongly quasiconvex functions, thus is strongly quasiconvex by Remark 2.1(iii) with modulus $\gamma_i = \frac{\min_{j \in J} \{\lambda_{\min}(A_j^i)\}}{M} > 0$ and it is not necessarily differentiable because $|J| \geq 2$.

In the following experiment, let us take $f(x) = \|Ax - b\|^2$, $K = [0, 5]^n$, $|J| = 2$ and

$$\begin{aligned} A_1 &= 2 \times I_n, & a_1 &= -1_n, & \alpha_1 &= n+1, & b_1 &= 1_n, & \beta_1 &= 1, \\ A_2 &= 0_{n \times n}, & a_2 &= 1_n, & \alpha_2 &= -n+1, & b_2 &= 0_n, & \beta_2 &= 1. \end{aligned}$$

In virtue of Lemma 2.2, we can estimate the strong subdifferential by using the convex subdifferential, and using the structure of the problem, we have:

$$h(x) = \max \left\{ \frac{\|x\|^2 - 1_n^T x + n + 1}{1_n^T x + 1}, 1_n^T x - n + 1 \right\}.$$

Clearly, $h(x) \geq 1$ for every $x \in [0, 5]^n$ and $h(1_n) = 1$. Every entry of A was independently generated using a standard normal distribution in the interval $[0, 1]$ and b was chosen as $b = A \times 1_n$. Under these assumptions, the unique solution of (1.1) is $x^* = 1_n$ and the optimal value is 1. In this setting, $x^* \in \operatorname{argmin}_K h \cap \operatorname{argmin}_K f$, therefore, it follows from Corollary 3.2 that the sequence generated by the proximal gradient method converges to x^* .

Table 3 reports the average CPU time(s) and the average error

$$\text{err} := \frac{\|x^k - x^*\|}{n},$$

over 100 problems for each dimension $n = 20, 50, 100$ with $\beta_k \equiv \frac{1}{2\lambda_{\max}(A^T A)}$, $x^0 = 0_n$ and we stop the algorithm if $\|x^k - x^{k+1}\| < 10^{-3}$ or the number of iterations exceeds 1000.

Table 3: Average CPU time and error for different values of n in Example 4.2

n	20	50	100
CPU time(s)	0.4206161332130433	1.0433911538124083	4.349398927688599
Error	0.01346800752509246	0.007167582429400066	0.004199154120685113

tab3

5 Conclusions

sec:06

We contributed to the discussion on the minimization of the sum of nonconvex functions by providing both optimality conditions and proximal gradient algorithms when one of the involved functions is nonsmooth and nonconvex and the

other is differentiable (and possibly nonconvex too). In particular, we assume that the nonsmooth nonconvex function is strongly quasiconvex (in the sense of Polyak [35]) and, by using adequate tools for this class, we established new information regarding the point provided by the stopping criteria as well as for the limit point of the sequence generated by the algorithm. Furthermore, we provided applications in nonconvex optimization as DC programming and quadratic fractional programming problems as well as numerical illustrations.

We hope that this new information will be very useful for the optimization community for dealing with the composite optimization problem involving nonconvex functions and its applications. In particular, by developing accelerations/flexibilizations/generalization of the proximal gradient method with its inertial, relaxed and adaptive versions as well as with Bregman distances (see [10, 31] and also [15, 27] for the minimization of strongly quasiconvex functions). This will be the subject for a subsequent work.

6 Declarations

6.1 Availability of supporting data

No data sets were generated during the current study. The used PYTHON codes are available from all authors on reasonable request.

6.2 Author Contributions

All authors contributed equally to the study conception, design and implementation and wrote and corrected the manuscript.

6.3 Competing Interests

There are no conflicts of interest or competing interests related to this manuscript.

6.4 Funding

The authors wish to thank the reviewers for their suggestions that helped to improve the paper. This research was partially supported by ANID-Chile through Fondecyt Regular 1241040 (Lara) and by Vietnam Academy of Science and Technology (VAST) through project number CTTH00.02/24-25 (Yen).

References

- AFV

[1] F.J. ARAGÓN-ARTACHO, R.M.T. FLEMING, P.T. VUONG, Accelerating the DC algorithm for smooth functions, *Math. Program.*, **30**, 980–1006, (2018).
- AV

[2] F.J. ARAGÓN-ARTACHO, P.T. VUONG, The boosted DC algorithm for nonsmooth functions, *SIAM, J. Optim.*, **169**, 95–118, (2020).

- [3] H. ATTOUCH, J. BOLTE, B.F. SVAITER, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods. *Math. Program.* **137**, 91–129, (2013).
- [4] H. ATTOUCH, J. BOLTE, P. REDONT, A. SOUBEYRAN, Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality, *Math. Oper. Res.*, **35**, 438–457, (2010).
- [5] H.H. BAUSCHKE, P.L. COMBETTES. “Convex Analysis and Monotone Operators Theory in Hilbert Spaces”. *CMS Books in Mathematics*. Springer-Verlag, second edition, (2017).
- [6] A. BECK. “First Order Methods in Optimization”. *MOS-SIAM, Series on Optimization*. SIAM, Philadelphia, (2017).
- [7] A. BECK, M. TEBoulLE, Gradient-based algorithms with applications to signal recovery, in: Y. Eldar and D. Palomar (EDS.) “Convex Optimization in Signal Processing and Communications”, pp. 3–50, (2009).
- [8] J. BOLTE, A. DANILIDIS, A. LEWIS, The Lojasiewicz inequality for non-smooth subanalytic functions with applications to subgradient dynamical systems, *SIAM J. Optim.*, **17**(4), 1205–1223, (2006).
- [9] J. BOLTE, S. SABACH, M. TEBoulLE, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, *Math. Programm.*, **146**, 459–494, (2014).
- [10] R.I. BOŦ, E.R. CSETNEK, An inertial Tseng’s type proximal algorithm for nonsmooth and nonconvex optimization problems, *J. Optim. Theory Appl.*, **171**, 600–616, (2016).
- [11] R.I. BOŦ, E.R. CSETNEK, C.S. LÁSZLÓ, An inertial forward–backward algorithm for the minimization of the sum of two nonconvex functions, *EURO, J. Comput. Optim.*, **4**, 3–25, (2016).
- [12] A. CAMBINI, L. MARTEIN. “Generalized Convexity and Optimization”. Springer-Verlag, Berlin-Heidelberg, (2009).
- [13] J. CHOQUE, F. LARA, R.T. MARCAVILLACA, A subgradient projection method for quasiconvex minimization, *Positivity*, Vol. 28, Issue 5, Paper 64, (2024).
- [14] S.-M. GRAD, F. LARA, An extension of proximal point algorithms beyond convexity, *J. Global Optim.*, **82**, 313–329, (2022).
- [15] S.-M. GRAD, F. LARA, R.T. MARCAVILLACA, Relaxed-inertial proximal point type algorithms for quasiconvex minimization, *J. Global Optim.*, **85**, 615–635, (2023).

- [HKS] [16] N. HADJISAVVAS, S. KOMLOSI, S. SCHAIBLE. “Handbook of Generalized Convexity and Generalized Monotonicity”. Springer-Verlag, Boston, (2005).
- [HU-L] [17] J.-B. HIRIART-URRUTY, C. LEMARÉCHAL. “Fundamentals of Convex Analysis”. Springer-Verlag, Berlin, Second Edition, (2004).
- [IL1] [18] A. IUSEM, F. LARA, Second order asymptotic functions and applications to quadratic programming, *J. of Convex Anal.*, **25**, 271–291, (2018).
- [ILMY] [19] A. IUSEM, F. LARA, R.T. MARCAVILLACA, L.H. YEN, A two-step PPA for nonconvex equilibrium problems with applications to fractional programming, *J. Global Optim.*, Vol. 90, 755–779, (2024).
- [JKM] [20] X. JIA, C. KANZOW, P. MEHLITZ, Convergence analysis of the proximal gradient method in the presence of the Kurdyka-Lojasiewicz property without global Lipschitz assumptions, *SIAM J. Optim.*, **33**, 3038–3056, (2023).
- [J-1] [21] M. JOVANOVIĆ, On strong quasiconvex functions and boundedness of level sets, *Optimization*, **20**, 163–165, (1989).
- [J-2] [22] M. JOVANOVIĆ, A note on strongly convex and quasiconvex functions, *Math. Notes*, **60**, 584–585, (1996).
- [KL-2] [23] A. KABGANI, F. LARA, Strong subdifferentials: theory and applications in nonconvex optimization, *J. Global Optim.*, **84**, 349–368, (2022).
- [KM] [24] C. KANZOW, P. MEHLITZ, Convergence properties of monotone and non-monotone proximal gradient methods revisited, *J. Optim. Theory Appl.*, **195**, 624646, (2022).
- [K] [25] K. KURDYKA, On gradients of functions definable in $\mathcal{o}$ -minimal structures, *Ann. Inst. Fourier*, **48**, 769–783, (1998).
- [Lara-9] [26] F. LARA, On strongly quasiconvex functions: existence results and proximal point algorithms, *J. Optim. Theory Appl.*, **192**, 891–911, (2022).
- [LM] [27] F. LARA, R.T. MARCAVILLACA, Bregman type proximal point algorithms for quasiconvex minimization, *Optimization*, **73**, 497–515, (2024).
- [LMY] [28] F. LARA, R.T. MARCAVILLACA, L.H. YEN, An extragradient method for nonconvex pseudomonotone equilibrium problems with applications, *Comp. Appl. Math.*, **43**, 128, (2024).
- [L] [29] S.C. LÁSZLÓ, A forward-backward algorithm with different inertial terms for structured non-convex minimization problems, *J. Optim. Theory Appl.*, **198**, 387–427, (2023).
- [Lo] [30] S. LOJASIEWICZ, Une propriété topologique des sous-ensembles analytiques réels, in “Les Équations aux Dérivées Partielles”. Éditions du Centre National de la Recherche Scientifique, Paris, 87–89, (1963).

- [MT] [31] Y. MALITSKY, M.K. TAM, A forward-backward splitting method for monotone inclusions without cocoercivity, *SIAM J. Optim.*, **30**, 1451–1472, (2020).
- [M-VAA] [32] B. MORDUKHOVICH. “Variational Analysis and Applications”. Springer Monographs in Mathematics, Springer Cham, Switzerland, (2018).
- [OCBP] [33] P. OCHS, Y. CHEN, T. BROX, T. POCK, iPiano: Inertial proximal algorithm for non-convex optimization, *SIAM J. Imag. Sci.*, **7**, 1388–1419, (2014).
- [P1] [34] B.T. POLYAK, Gradient methods for minimizing functionals, *Zh. Vychisl. Math. Mat. Fiz.*, **3**, 643–653, (1963).
- [P] [35] B.T. POLYAK, Existence theorems and convergence of minimizing sequences in extremum problems with restrictions, *Soviet Math.*, **7**, 72–75, (1966).
- [SCH] [36] S. SCHAIBLE, Fractional programming, In: R. Horst and P. Pardalos (eds.), “Handbook of Global Optimization”, pp. 495–608. Kluwer Academic Publishers, Dordrecht, (1995).
- [SZI] [37] S. SCHAIBLE, W.T. ZIEMBA. “Generalized Concavity in Optimization and Economics”. Academic Press, New York, (1981).
- [Stancu] [38] I.M. STANCU-MINASIAN. “Fractional Programming: Theory, Methods and Applications”. Kluwer Academic Publishers, (1997).
- [LDM] [39] H.A.L. THI, T.P. DINH, L.D. MUU, Numerical solution for optimization over the efficient set by D.C. optimization algorithms, *Oper. Res. Lett.*, **19**, 117–128, (1996).