ON THE PROXIMAL POINT ALGORITHM FOR STRONGLY QUASICONVEX FUNCTIONS IN HADAMARD SPACES

NICHOLAS PISCHKE

Department of Mathematics, Technische Universität Darmstadt,
Schlossgartenstraße 7, 64289 Darmstadt, Germany,

E-mail: pischke@mathematik.tu-darmstadt.de

ABSTRACT. We prove the convergence of the proximal point algorithm for finding the unique minimizer of a strongly quasiconvex function in general nonlinear Hadamard spaces, generalizing a recent result due to F. Lara. Our argument is rather elementary and brief and relies only on a few properties of strongly quasiconvex functions and their proximal operators which are established here for the first time over these nonlinear spaces. In particular, our convergence proof is fully effective and actually yields fast (ranging up to linear) rates of convergence for the iterates towards the solution and for the function values towards the minimum. These rates are novel even in the context of Euclidean spaces.

Keywords: Proximal point algorithm; rates of convergence; strongly quasiconvex functions; Hadamard spaces; proof mining
MSC2020 Classification: 47J25, 90C26, 03F10

1. INTRODUCTION

One of the most fundamental problems in optimization is the minimization

$$\min_{x \in C} f(x)$$

of a function $f : X \to \mathbb{R}$ over a certain space X and a designated set of constraints $C \subseteq X$. Naturally, suitable restrictions have to be placed on the space and the function to make this problem feasible and the most classical formulation is the minimization of a convex and lower semicontinuous function over a Hilbert space where $C = X$. In that case, the most prominent method for approaching the corresponding minimization problem is the proximal point method originally devised in the work of Martinet [33], Rockafellar [43] as well as Brézis and Lions [6], which constructs a next iterate by solving the proximal subproblem

$$\mathrm{argmin}_{y \in X} \left\{ f(y) + \frac{1}{2c_k} \left\| y - x^k \right\|^2 \right\}$$

in terms of the current iterate x^k and a regularization parameter $c_k > 0$, which becomes tractable since $f(\cdot) + \frac{1}{2\beta} \left\| \cdot - x \right\|^2$ is strongly convex for any point x and $\beta > 0$ if the function f is convex.

When the assumption of convexity of f is relaxed, already these proximal subproblems (and hence the proximal point method in general) become considerably harder and a great amount of research has been carried out to isolate certain classes of functions and spaces where the proximal point method remains effective. The most broad concrete class of functions we want to consider here is that of quasiconvex functions. This considerably wide class of (potentially nonconvex) functions however suffers from a range of problems in regards to the proximal point

---

Date: July 10, 2025.

algorithm. As highlighted in [30], already on the real line quasiconvexity does not imply that the proximal subproblems are uniquely solvable (and in particular $f(\cdot) + \frac{1}{2\beta} \|\cdot - x\|^2$ is not strongly quasiconvex for quasiconvex functions f). Also as highlighted in [30], the proximal point method is not really well-motivated as fixed points of the proximal map are not necessarily minimizers (and so the classical stopping criteria do not apply). As a result, while the proximal point method for quasiconvex functions has been heavily investigated (see e.g. [8, 31, 34, 36, 37, 38, 39, 40, 41], among many more), essentially all convergence results rely on taking $c_k \to \infty$ which, as highlighted in [30], essentially reduces the proximal point method to a penalty method.

Starting at least with the recent work of Lara [30], an emphasis has thus been placed on strongly quasiconvex functions (which go back at least to the work of Polyak [35]), a subclass of quasiconvex functions where (some of) these issues are avoided and one nevertheless encompasses interesting nonconvex (or at least not strongly convex) functions such as the Euclidean norm, which is strongly quasiconvex (as shown by Jovanovič [23]) without being strongly convex on any bounded convex set, and even the square root of the Euclidean norm which is strongly quasiconvex on each ball (as shown by Lara [30]) without being differentiable (and which implies that e.g. $\max\{\sqrt{\|\cdot\|}, \|\cdot\|^2 - k\}$ for $k \in \mathbb{N}$ is strongly quasiconvex without even being convex as highlighted in [30]). We refer to [30] for further, very comprehensive, discussions on how this class is differentiated (or related) to various other classes of (non)convex functions. So, while some complications remain in this class, such as the proximal map not being single-valued, Lara has shown in [30] that the natural analogue of the proximal point method, obtained by choosing some minimizer of said map at each stage, allows for a convergence result where the iteration converges to the unique minimizer of the function and the function values converge to the minimum, under the weak assumption that the regularization parameters (c_k) are positively bounded from below. This method and the corresponding convergence result has since then been extended in various directions by Lara and his co-authors (see e.g. [15, 16, 20]).

In this paper, we extend the convergence result for the proximal point method as established by Lara in [30] to Hadamard spaces, that is complete geodesic metric spaces of nonpositive curvature, one of the most important classes of nonlinear hyperbolic spaces (see the next section for a more detailed introduction). On various subclasses of these spaces, the proximal point method for convex functions was previously investigated by a range of authors, culminating in the seminal work of Bačák [2] who proved the weak convergence of the proximal point method in general Hadamard spaces (under a suitable notion of weak convergence, as will also be discussed later). We refer in particular to the comprehensive discussion therein for an excellent overview of the related literature. Previous work on proximal point methods for quasiconvex functions in nonlinear contexts is also rather abundant but, to our knowledge, largely limited to the setting of certain classes of Riemannian manifolds as mainly championed in the work of Papa Quiroz and his co-authors (see e.g. [36, 37, 39, 40] among many others) and further, in these works no special subclasses of quasiconvex functions are considered. In particular, as highlighted in the recent work [36], extensions of Lara's results from [30] have so far not been developed in any kind of nonlinear context, a gap we try to fill with this paper.

Beyond however merely extending said results, we even provide a quantitative convergence result and, in particular, give very fast (ranging up to linear) rates of convergence for the sequence towards the solution and for the function values towards the minimum. These rates are moreover highly uniform and depend only on very few data relating to the iteration and the

function. To our knowledge, these rates are novel even in the context of Lara's original setting of Euclidean spaces. We further want to highlight that our convergence proof is rather brief and elementary, relying only on a very small selection of key properties of strongly quasiconvex functions and their proximal maps that are generalized from the Euclidean context to the general setting of Hadamard spaces in this work as well. This elementarity of the convergence proof is key here for overcoming the obstacles of extending the results to general Hadamard spaces as Lara's proof given in [30] relies on sequential compactness arguments in Euclidean spaces which are not available in general Hadamard spaces (and this similarly differentiates us from the setting of Riemannian manifolds which also are locally compact). In particular, it should be noted that our convergence result is "strong", i.e. is phrased not with a suitable notion of weak convergence but relative to the metric. This is not true for the usual proximal point method already for convex functions in Hilbert spaces as shown by Güler [19] and is in this setting essentially due to the fact that strongly quasiconvex functions have a unique minimizer. In particular, our results thereby also extend Lara's results to infinite dimensional Hilbert spaces and show that the convergence in these spaces is strong as well. While not investigated here, it remains an interesting problem if similar, more elementary, proofs can be developed for the range of extensions of Lara's convergence result as surveyed above, and if these, or more generally the wide range of known results involving strongly quasiconvex functions (as comprehensively surveyed in the recent paper [17]), then can be likewise transferred to nonlinear contexts as broad as Hadamard spaces, if appropriate.

All of the present results have been obtained using the methodology of proof mining, a program in mathematical logic that aims to classify and extract the computational content of prima facie "non-computational" proofs. We refer to the seminal monograph [27] for a comprehensive overview of the subject and [28] for a survey of more recent developments. However, while this logical perspective of proof mining was essential for deriving the results of this paper, we want to stress that the whole paper is presented without any explicit reference to logic and does not require any such background.

The rest of the paper is now organized as follows: Section 2 introduces the main necessary (but very minimal) background on Hadamard spaces. Section 3 discusses strongly quasiconvex functions and their proximal maps in Hadamard spaces and establishes their key properties required for the main (quantitative) convergence results which are presented in Section 4.

2. Preliminaries

All our results take place over various classes of nonlinear hyperbolic spaces, the most basic of which we introduce now:

A triple (X, d, W) is called a hyperbolic space [26] if (X, d) is a metric space and $W : X \times X \times [0, 1] \to X$ is a function satisfying for all $x, y, z, w \in X$ and $\lambda, \lambda' \in [0, 1]$:

(W1) $d(W(x, y, \lambda), z) \leqslant (1 - \lambda)d(x, z) + \lambda d(y, z)$,
(W2) $d(W(x, y, \lambda), W(x, y, \lambda')) = |\lambda - \lambda'|d(x, y)$,
(W3) $W(x, y, \lambda) = W(y, x, 1 - \lambda)$,
(W4) $d(W(x, y, \lambda), W(z, w, \lambda)) \leqslant (1 - \lambda)d(x, z) + \lambda d(y, w)$.

To ease notation in the following, we write $(1 - \lambda)x \oplus \lambda y$ for the point $W(x, y, \lambda)$. We further call a set $C \subseteq X$ convex if $(1 - \lambda)x \oplus \lambda y \in C$ for all $x, y \in C$ and $\lambda \in [0, 1]$.

This class of hyperbolic spaces, which was originally introduced in [26] as a logically motivated amalgamation of various notions of convex metric and hyperbolic spaces from the literature such as Takahashi's convex metric spaces [44], the hyperbolic spaces of Reich and Shafrir [42] as well as the spaces of hyperbolic type as introduced by Goebel and Kirk [14], just to name a few, is commonly invoked as the appropriate nonlinear generalization of convexity from normed contexts and in particular allows for a nice direct approach to CAT(0) and Hadamard spaces, the main classes of spaces that we are concerned with in this paper.

Concretely, in this paper, a CAT(0) space is a hyperbolic space (X, d, W) which satisfies

$$(\text{CN}^-) \qquad d^2 \left(\frac{1}{2} x \oplus \frac{1}{2} y, z \right) \leqslant \frac{1}{2} d^2(x, z) + \frac{1}{2} d^2(y, z) - \frac{1}{4} d^2(x, y)$$

for all $x, y, z \in X$.[1]

These spaces were originally introduced by Aleksandrov [1] and were named CAT(0) spaces by Gromov [18]. A complete CAT(0) space is called a Hadamard space. Examples of such spaces include Hilbert spaces, $\mathbb{R}$-trees and complete simply connected Riemannian manifolds of nonpositive sectional curvature, among many more. We refer to the seminal monograph [7] for a comprehensive overview of CAT(0) and Hadamard spaces and further refer to [3] for a shorter treatment focused on aspects of convex analysis and optimization.

We shortly note that the characterizing relation CN^- extends to arbitrary convex combinations:

Lemma 1 (folklore, see e.g. [11]). *Let (X, d, W) be a* CAT(0) *space. Then*

$$(\text{CN}^+) \qquad d^2((1 - \lambda)x \oplus \lambda y, z) \leqslant (1 - \lambda)d^2(x, z) + \lambda d^2(y, z) - \lambda(1 - \lambda)d^2(x, y)$$

for any $x, y, z \in X$ and $\lambda \in [0, 1]$.

Crucially, we also rely on another equivalent characterization of CAT(0) spaces as developed in the work of Berg and Nikolaev [5]. Concretely, define the so-called quasi-linearization function by

$$\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle := \frac{1}{2} \left(d^2(x, v) + d^2(y, u) - d^2(x, u) - d^2(y, v) \right)$$

for all $x, y, u, v \in X$, using $\overrightarrow{xy}$ to denote pairs $(x, y) \in X^2$. As shown in [5], this function is the unique function $X^2 \times X^2 \to \mathbb{R}$ in any metric space (X, d) such that

(1) $\langle \overrightarrow{xy}, \overrightarrow{xy} \rangle = d^2(x, y)$,
(2) $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle = \langle \overrightarrow{uv}, \overrightarrow{xy} \rangle$,
(3) $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle = - \langle \overrightarrow{yx}, \overrightarrow{uv} \rangle$,
(4) $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle + \langle \overrightarrow{xy}, \overrightarrow{vw} \rangle = \langle \overrightarrow{xy}, \overrightarrow{uw} \rangle$.

for all $x, y, u, v, w \in X$. It then follows from the results in [5] that a hyperbolic space (X, d, W) is a CAT(0) space if, and only if,

$$(\text{CS}) \qquad \langle \overrightarrow{xy}, \overrightarrow{uv} \rangle \leqslant d(x, y)d(u, v)$$

---

[1]CAT(0) spaces are often introduced as geodesic metric spaces satisfying the Bruhat-Tits CN-inequality [9]. Any hyperbolic space that satisfies the CN-inequality is hence a CAT(0)-space and conversely, as every CAT(0)-space is actually uniquely geodesic, any CAT(0)-space is also a hyperbolic space by defining $W(x, y, \lambda)$ via the unique geodesic joining x and y. Hence being a CAT(0)-space in the usual sense is equivalent to the definition given here as, in the presence of (W1) – (W4), the inequality CN^- is equivalent to the CN-inequality. For more details, we refer to the discussions in [13, 26].

for all $x, y, u, v \in X$, i.e. where a metric version of the Cauchy-Schwarz inequality holds. By these properties, this function can be seen as a nonlinear generalization of the inner product in Hilbert spaces. Beyond these properties, we will also rely on the following equality (which is immediate from the definition of the quasi-linearization function):

Lemma 2. *For any metric space (X, d) and $x, y, z \in X$:*

$$d^2(x, y) = d^2(x, z) + d^2(z, y) + 2 \langle \overrightarrow{xz}, \overrightarrow{zy} \rangle.$$

3. (Strongly) quasiconvex functions on Hadamard spaces

For the rest of this paper, unless said otherwise, we let (X, d, W) be a Hadamard space, i.e. a complete CAT(0) space. We begin with our main class of functions: A function $f : X \to \mathbb{R}$ is called

(1) *quasiconvex* if
$$f((1 - \lambda)x \oplus \lambda y) \leqslant \max\{f(x), f(y)\}$$
for all $x, y \in X$ and $\lambda \in [0, 1]$,

(2) *strongly quasiconvex* with value $\gamma > 0$ if
$$f((1 - \lambda)x \oplus \lambda y) \leqslant \max\{f(x), f(y)\} - \lambda(1 - \lambda)\frac{\gamma}{2}d^2(x, y)$$
for all $x, y \in X$ and $\lambda \in [0, 1]$.

For the rest of this paper, we are interested in solving the problem $\min_{x \in X} f(x)$ for a strongly quasiconvex function f. We write $\operatorname{argmin} f$ for the set of minimizers of f. Crucially, we first note that for such functions the minimizer is unique, similar as in Euclidean spaces (see e.g. the discussion around Corollary 3 in [30]):

Lemma 3. *Let $f : X \to \mathbb{R}$ be a strongly quasiconvex function with value $\gamma > 0$. Then $\operatorname{argmin} f$ is at most a singleton.*

Proof. If x, y are two minimizers, define $z = 1/2x \oplus 1/2y$. As f is strongly quasiconvex, we get

$$f(z) \leqslant \max\{f(x), f(y)\} - \frac{\gamma}{8}d^2(x, y) = \min f - \frac{\gamma}{8}d^2(x, y)$$

and so $d^2(x, y) \leqslant (\min f - f(z)) \cdot 8/\gamma \leqslant 0$, i.e. $x = y$. $\square$

Remark 4. While all properties and results on quasiconvex functions given in [30] are formulated relative to a closed and convex subset of the domain $\operatorname{dom} f$ of a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, we omit these restrictions in this paper as any closed and convex set $C \subseteq X$ of a Hadamard space X is naturally a Hadamard space itself and so we would not gain any further generality from considering closed and convex subsets of the domain of a function with extended real values.

Remark 5. In the context of Euclidean spaces, it is known that any lower semicontinuous strongly quasiconvex function actually has a minimizer (see e.g. Corollary 3 in [30]), which hence is unique. All such arguments (and in particular also that given by Lara [30]) however seem to rely on the local compactness of $\mathbb{R}^n$, and hence do not immediately transfer to this general setting of Hadamard spaces. Albeit unlikely, it thus remains to be seen whether an analogous result holds true in this context. When studying the proximal point algorithm later, we will thus rely on the (in Euclidean spaces superfluous) assumption that $\operatorname{argmin} f \neq \varnothing$.

The main associated object with such (strongly) quasiconvex functions f that we employ for approximating minimizers thereof is the proximal map (also sometimes called the Moreau-Yosida resolvent)

$$\mathrm{Prox}_{\beta f}(x) := \mathrm{argmin}_{y \in X} \left\{ f(y) + \frac{1}{2\beta} d^2(x, y) \right\}$$

for $\beta > 0$ and $x \in X$, used in a metric context without linear structure for the first time in [22].

If the function f is lower semincontinuous (lsc), i.e.

$$\liminf_{n \to \infty} f(x_n) \geqslant f(x)$$

for any $x \in X$ and any sequence (x_n) such that $x_n \to x$, and additionally convex, then $\mathrm{Prox}_{\beta f}(x)$ is always a singleton (see e.g. [22]). Now, in the quasiconvex case, the proximal map is in general not single-valued already in Euclidean contexts (see e.g. the discussion in [30]) but at least always non-empty (which in [30] is reduced to the lower semicontinuity and supercoercivity of the object function of the proximal map and hence relies on a compactness argument).

Before moving on to the key properties of the proximal map on which our convergence proof relies, we now first show that the same is true in Hadamard spaces for quasiconvex lsc functions f. The respective argument relies similarly on a suitable notion of supercoercivity in metric spaces together with a weak compactness argument and we first introduce the necessary ingredients for this.

A function $f : X \to \mathbb{R}$ is called supercoercive at $x \in X$ if

$$\lim_{d(x,y) \to \infty} \frac{f(y)}{d(x, y)} = \infty.$$

A similar notion on Hadamard manifolds appears in [38] under the name of 1-coercivity. Clearly, if f is bounded from below, then the function $f(y) + \frac{1}{2\beta} d^2(x, y)$ is supercoercive at x. Crucially, we have the following result:

Lemma 6 (folklore). *If $f : X \to \mathbb{R}$ is bounded from below and satisfies*

$$\liminf_{d(x,y) \to \infty} \frac{f(y)}{d(x, y)} > 0$$

for some $x \in X$, then any sequence (y_n) with $f(y_n) \to \inf f$ is bounded. In particular, the result holds if $f : X \to \mathbb{R}$ is bounded from below and supercoercive at some $x \in X$.

Proof. Suppose not, i.e. let (y_n) be an unbounded sequence such that $f(y_n) \to \inf f$. Then $d(x, y_n) \to \infty$ and so $\liminf_{n \to \infty} f(y_n)/d(x, y_n) > 0$ by assumption on f. However, we have $\lim_{n \to \infty} f(y_n)/d(x, y_n) = 0$ since $f(y_n) \to \inf f$, a contradiction. $\square$

Weak convergence in CAT(0) spaces goes back to the work of Jost [21] and is often called Δ-convergence, as it can be equivalently described via an extension of Lim's notion of Δ-convergence [32] to CAT(0) spaces developed by Kirk and Panyanak [25] (as shown by Espínola and Fernández-León [12], see also the discussion in [2]).

Concretely, define the asymptotic radius of a sequence $(x_n) \subseteq X$ at a point $x \in X$ via

$$r(x_n, x) := \limsup_{n \to \infty} d(x_n, x)$$

and define the general asymptotic radius of (x_n) by

$$r(x_n) := \inf_{x \in X} r(x_n, x).$$

A point $x \in X$ is called an asymptotic center of (x_n) if

$$r(x_n, x) = r(x_n)$$

and in Hadamard spaces, asymptotic centers exist and are unique (see e.g. Proposition 7 in [10]). A sequence $(x_n) \subseteq X$ is said to weakly converge to $x \in X$ if x is the asymptotic center of each subsequence of (x_n) and we write $x_n \to^w x$ in that case. A point $x \in X$ is a weak cluster point of a sequence (x_n) if there is a subsequence (x_{n_k}) of (x_n) with $x_{n_k} \to^w x$. Crucially, we rely on the following result:

Lemma 7 ([21], see also [25]). *Any bounded sequence in X has a weak cluster point.*

A function $f : X \to \mathbb{R}$ is now called weakly lower semicontinuous (weakly lsc) at a point $x \in X$ if

$$\liminf_{n \to \infty} f(x_n) \geqslant f(x)$$

whenever (x_n) is a sequence such that $x_n \to^w x$. Crucially, if $f : X \to \mathbb{R}$ is a convex lsc function on a Hadamard space, it is also weakly lsc (see [2]) and this result rather immediately extends to quasiconvex functions. For the proof, we in particular need the following result that closed convex subsets of Hadamard spaces are weakly sequentially closed:

Lemma 8 ([4]). *Let $C \subseteq X$ be a closed convex set. If $(x_n) \subseteq C$ and $x_n \to^w x$ for some $x \in X$, then $x \in C$.*

We then get the following result (which is a simple extension of Lemma 3.1 in [2]):

Lemma 9. *Let $f : X \to \mathbb{R}$ be a quasiconvex lsc function. Then f is weakly lsc.*

Proof. Suppose for a contradiction that there is a sequence $(x_n) \subseteq X$ and $x \in X$ with $x_n \to^w x$ and

$$\liminf_{n \to \infty} f(x_n) < f(x),$$

i.e. there exists a subsequence (x_{n_k}), a $k_0 \in \mathbb{N}$ and an $\varepsilon > 0$ with $f(x_{n_k}) \leqslant f(x) - \varepsilon$ for all $k \geqslant k_0$. By quasiconvexity of f, we get

$$f(y) \leqslant f(x) - \varepsilon$$

for all $y \in \mathrm{co}\{x_{n_k} \mid k \geqslant k_0\}$, where $\mathrm{co}\{x_{n_k} \mid k \geqslant k_0\}$ is the convex hull of $\{x_{n_k} \mid k \geqslant k_0\}$. By lower semicontinuity of f, we have the same for $y \in \overline{\mathrm{co}}\{x_{n_k} \mid k \geqslant k_0\}$, the closed convex hull of the respective set. But by Lemma 8, we have $x \in \overline{\mathrm{co}}\{x_{n_k} \mid k \geqslant k_0\}$ which yields a contradiction. $\square$

This allows us to show the following result:

Lemma 10. *Let $f : X \to \mathbb{R}$ be a quasiconvex lsc function which is bounded from below. Then for any $x \in X$ and $\beta > 0$, $\mathrm{Prox}_{\beta f}(x)$ is nonempty.*

Proof. As commented on before, the function $g(y) := f(y) + \frac{1}{2\beta} d^2(x, y)$ is supercoercive and by assumption f, and hence g, is bounded from below. Take a sequence (y_n) such that $g(y_n) \to \inf g$. By Lemma 6, (y_n) is bounded and so by Lemma 7, (y_n) has a weak cluster point y, i.e. there exists a subsequence (y_{n_k}) such that $y_{n_k} \to^w y$. As f is quasiconvex and lower semicontinuous, by Lemma 9 we get that f is weakly lower semicontinuous. Also, by (CN$^+$), $\frac{1}{2\beta} d^2(x, y)$ is in particular convex and clearly lower semicontinuous, and so also weakly lower semicontinuous. Therefore g is weakly lower semicontinuous and we have

$$g(y) \leqslant \liminf_{k \to \infty} g(y_{n_k}) = \inf g$$

as $g(y_{n_k}) \to \inf g$. Thus y is a minimizer of g and so clearly $y \in \mathrm{Prox}_{\beta f}(x)$. $\square$

Now, the main property of the proximal map of a strongly quasiconvex function that we rely on is given in the following lemma, which extends an analogous result due to Lara [30] (see Proposition 7 therein) from $\mathbb{R}^n$ to Hadamard spaces. It should be noted that, on $\mathbb{R}^n$, this property was used by Kabgani and Lara [24] to define the so-called strong subdifferential, which provides a novel notion of subdifferential suitable for the study of various classes of nonconvex functions.

Lemma 11. *Let $f : X \to \mathbb{R}$ be a strongly quasiconvex function with value $\gamma > 0$ and let $\beta > 0$ and $x \in X$. If $\overline{x} \in \mathrm{Prox}_{\beta f}(x)$, then*

$$f(\overline{x}) \leqslant \max\{f(y), f(\overline{x})\} + \frac{\lambda}{2}\left(\frac{\lambda}{\beta} - \gamma + \lambda\gamma\right) d^2(y, \overline{x}) + \frac{\lambda}{\beta}\left\langle \overrightarrow{x\overline{x}}, \overrightarrow{\overline{x}y} \right\rangle$$

for all $y \in X$ and $\lambda \in [0, 1]$.

Proof. If $\overline{x} \in \mathrm{Prox}_{\beta f}(x)$, then by definition

$$f(\overline{x}) + \frac{1}{2\beta}d^2(\overline{x}, x) \leqslant f(z) + \frac{1}{2\beta}d^2(z, x)$$

for all $z \in X$. Now, given $y \in X$ and $\lambda \in [0, 1]$, we set $z = (1 - \lambda)\overline{x} \oplus \lambda y$ in the above to obtain

$$f(\overline{x}) + \frac{1}{2\beta}d^2(\overline{x}, x) \leqslant f((1 - \lambda)\overline{x} \oplus \lambda y) + \frac{1}{2\beta}d^2((1 - \lambda)\overline{x} \oplus \lambda y, x)$$

$$\leqslant \max\{f(y), f(\overline{x})\} - \lambda(1 - \lambda)\frac{\gamma}{2}d^2(y, \overline{x}) + \frac{1}{2\beta}d^2((1 - \lambda)\overline{x} \oplus \lambda y, x)$$

using that f is strongly quasiconvex with value γ. Now, using (CN^+), we get

$$d^2((1 - \lambda)\overline{x} \oplus \lambda y, x) \leqslant (1 - \lambda)d^2(\overline{x}, x) + \lambda d^2(y, x) - \lambda(1 - \lambda)d^2(y, \overline{x})$$

and so

$$f(\overline{x}) \leqslant \max\{f(y), f(\overline{x})\} - \frac{\lambda}{2}(\gamma - \lambda\gamma)d^2(y, \overline{x}) + \frac{\lambda}{2}\frac{\lambda}{\beta}d^2(y, \overline{x})$$

$$+ \frac{1}{2\beta}d^2(\overline{x}, x) - \frac{\lambda}{2\beta}d^2(\overline{x}, x) + \frac{\lambda}{2\beta}d^2(y, x) - \frac{\lambda}{2\beta}d^2(y, \overline{x}) - \frac{1}{2\beta}d^2(\overline{x}, x)$$

$$\leqslant \max\{f(y), f(\overline{x})\} + \frac{\lambda}{2}\left(\frac{\lambda}{\beta} - \gamma + \lambda\gamma\right) d^2(y, \overline{x}) + \frac{\lambda}{\beta} \cdot \frac{1}{2}\left(d^2(y, x) - d^2(\overline{x}, x) - d^2(y, \overline{x})\right)$$

$$\leqslant \max\{f(y), f(\overline{x})\} + \frac{\lambda}{2}\left(\frac{\lambda}{\beta} - \gamma + \lambda\gamma\right) d^2(y, \overline{x}) + \frac{\lambda}{\beta}\left\langle \overrightarrow{x\overline{x}}, \overrightarrow{\overline{x}y} \right\rangle$$

which is what we wanted to show. $\square$

Before moving on to the proximal point algorithm, we want to highlight that, also in this nonlinear context, fixed points of the proximal map are exactly the minimizers of the strongly quasiconvex function. The proof given here is a simple nonlinear version of Lara's proof of an analogous result given in [30] (see Proposition 9 therein).

Proposition 12. *Let $f : X \to \mathbb{R}$ be a strongly quasiconvex function with value $\gamma > 0$. Then*

$$\mathrm{Fix}(\mathrm{Prox}_{\beta f}) = \mathrm{argmin} f$$

for any $\beta > 0$.

Proof. Let $\overline{x} \in \text{Prox}_{\beta f}(\overline{x})$. Then

$$f(\overline{x}) = f(\overline{x}) + \frac{1}{2\beta}d^2(\overline{x}, \overline{x}) \leqslant f(x) + \frac{1}{2\beta}d^2(x, \overline{x})$$

for all $x \in X$. Taking $x = (1 - \lambda)\overline{x} \oplus \lambda y$ for some arbitrary $y \neq \overline{x}$ (if X consists of a single point, the result is trivial) and $\lambda \in [0, 1]$ in the above then yields

$$f(\overline{x}) \leqslant f((1 - \lambda)\overline{x} \oplus \lambda y) + \frac{1}{2\beta}d^2((1 - \lambda)\overline{x} \oplus \lambda y, \overline{x})$$

$$\leqslant \max\{f(\overline{x}), f(y)\} - \lambda(1 - \lambda)\frac{\gamma}{2}d^2(\overline{x}, y) + \frac{1}{2\beta}d^2((1 - \lambda)\overline{x} \oplus \lambda y, \overline{x})$$

$$= \max\{f(\overline{x}), f(y)\} - \lambda(1 - \lambda)\frac{\gamma}{2}d^2(\overline{x}, y) + \frac{1}{2\beta}\lambda^2 d^2(\overline{x}, y)$$

$$= \max\{f(\overline{x}), f(y)\} + \frac{\lambda}{2}\left(\frac{\lambda}{\beta} - \gamma + \lambda\gamma\right)d^2(\overline{x}, y)$$

where the third line follows from the fact that

$$d((1 - \lambda)\overline{x} \oplus \lambda y, \overline{x}) = \lambda d(\overline{x}, y)$$

holds in the hyperbolic space X. Taking $\lambda \in (0, \beta\gamma/(1 + \beta\gamma))$ yields $\lambda/\beta - \gamma + \lambda\gamma < 0$ and so

$$f(\overline{x}) < \max\{f(\overline{x}), f(y)\}$$

which implies $f(\overline{x}) < f(y)$. As $y \in X\backslash\{\overline{x}\}$ was arbitrary, we get $\overline{x} \in \text{argmin}f$.

Conversely, let $\overline{x} \in \text{argmin}f$. Then

$$f(\overline{x}) + \frac{1}{2\beta}d^2(\overline{x}, \overline{x}) = f(\overline{x}) \leqslant f(y) \leqslant f(y) + \frac{1}{2\beta}d^2(\overline{x}, y)$$

for any $y \in X$ and so $\overline{x} \in \text{Prox}_{\beta f}(\overline{x})$. $\qquad\square$

4. THE PROXIMAL POINT ALGORITHM FOR MINIMIZING STRONGLY QUASICONVEX FUNCTIONS

We now are concerned with the proximal point algorithm for minimizing a strongly quasi-convex lsc function f and its convergence to the minimizer, if existent.

Concretely, by the proximal point algorithm, we shall in the following understand the following method: Given a sequence (c_k) of nonnegative reals with $\inf_{k\in\mathbb{N}} c_k \geqslant c > 0$ and an arbitrary $x^0 \in X$, define an iteration (x^k) by choosing

(PPA) $$x^{k+1} \in \text{Prox}_{c_k f}(x^k).$$

We assume throughout that $\text{argmin}f \neq \varnothing$. By Lemma 3, we have that $\text{argmin}f$ is a singleton and we denote the unique minimizer by x^* in the following. By Lemma 10, the sequence (x^k) is well-defined.

The first crucial result is that the sequence (x^k) defined by (PPA) satisfies the following recursive inequality (modeled, together with its proof, after Eq. (4.5) in [30]), which in particular implies that (x^k) is Fejér monotone w.r.t. $\text{argmin}f$:

Lemma 13. *Let (x^k) be defined by* (PPA). *Then for any $k \in \mathbb{N}$:*

$$d^2(x^{k+1}, x^*) \leqslant d^2(x^k, x^*) - d^2(x^k, x^{k+1}).$$

Proof. Fix $k \in \mathbb{N}$. Using that $x^{k+1} \in \operatorname{Prox}_{c_k f}(x^k)$, Lemma 11 implies

$$f(x^{k+1}) \leqslant \max\{f(y), f(x^{k+1})\} + \frac{\lambda}{2}\left(\frac{\lambda}{c_k} - \gamma + \lambda\gamma\right) d^2(y, x^{k+1}) + \frac{\lambda}{c_k}\left\langle \overrightarrow{x^k x^{k+1}}, \overrightarrow{x^{k+1} y}\right\rangle$$

for all $y \in X$ and $\lambda \in [0, 1]$. Take $y = x^*$. Then we get

$$0 = f(x^{k+1}) - \max\{f(x^*), f(x^{k+1})\}$$
$$\leqslant \frac{\lambda}{2}\left(\frac{\lambda}{c_k} - \gamma + \lambda\gamma\right) d^2(x^{k+1}, x^*) + \frac{\lambda}{c_k}\left\langle \overrightarrow{x^k x^{k+1}}, \overrightarrow{x^{k+1} x^*}\right\rangle$$

for all $\lambda \in [0, 1]$ and so (using the characterizing properties of the quasi-linearization function), we have

$$\left\langle \overrightarrow{x^k x^{k+1}}, \overrightarrow{x^* x^{k+1}}\right\rangle \leqslant \frac{c_k}{2}\left(\frac{\lambda}{c_k} - \gamma + \lambda\gamma\right) d^2(x^{k+1}, x^*)$$

whenever $\lambda > 0$. Now, Lemma 2 together with the characterizing properties of the quasi-linearization function imply

$$d^2(x^{k+1}, x^*) = d^2(x^{k+1}, x^k) + d^2(x^k, x^*) + 2\left\langle \overrightarrow{x^{k+1} x^k}, \overrightarrow{x^k x^*}\right\rangle$$
$$= d^2(x^{k+1}, x^k) + d^2(x^k, x^*) + 2\left\langle \overrightarrow{x^{k+1} x^k}, \overrightarrow{x^k x^{k+1}}\right\rangle + 2\left\langle \overrightarrow{x^{k+1} x^k}, \overrightarrow{x^{k+1} x^*}\right\rangle$$
$$= d^2(x^{k+1}, x^k) + d^2(x^k, x^*) - 2d^2(x^{k+1}, x^k) + 2\left\langle \overrightarrow{x^{k+1} x^k}, \overrightarrow{x^{k+1} x^*}\right\rangle$$
$$\leqslant d^2(x^k, x^*) - d^2(x^{k+1}, x^k) + c_k\left(\frac{\lambda}{c_k} - \gamma + \lambda\gamma\right) d^2(x^{k+1}, x^*)$$

for any $\lambda \in (0, 1]$. Taking $\lambda \in (0, \gamma c_k/(1 + \gamma c_k))$, we get $c_k\left(\frac{\lambda}{c_k} - \gamma + \lambda\gamma\right) < 0$ so that

$$d^2(x^{k+1}, x^*) \leqslant d^2(x^k, x^*) - d^2(x^k, x^{k+1}). \qquad \square$$

The above lemma not only yields that the sequence (x^k) defined by (PPA) is Fejér monotone but also that $\lim_{k\to\infty} d^2(x^k, x^{k+1}) = 0$. For the construction of our rate of convergence, we need a quantitative version of the special case that $\liminf_{k\to\infty} d^2(x^k, x^{k+1}) = 0$ which we construct in the following. For that, we first need the following abstract quantitative result on sequences of real numbers.

Lemma 14 (folklore)**.** *Let (a_k) be a sequence of nonnegative reals and let $b \geqslant a_0$. Then for any $\varepsilon > 0$:*

$$\exists k \leqslant \left\lceil \frac{b}{\varepsilon} \right\rceil \ (a_k - a_{k+1} < \varepsilon).$$

Proof. Suppose the claim is false, i.e. there is an $\varepsilon > 0$ such that

$$a_k - a_{k+1} \geqslant \varepsilon$$

for all $k \leqslant \left\lceil \frac{b}{\varepsilon} \right\rceil$. Then

$$a_0 \geqslant a_0 - a_{\left\lceil \frac{b}{\varepsilon} \right\rceil + 1} = \sum_{k=0}^{\left\lceil \frac{b}{\varepsilon} \right\rceil}(a_k - a_{k+1}) \geqslant \left(\left\lceil \frac{b}{\varepsilon} \right\rceil + 1\right)\varepsilon \geqslant b + \varepsilon > b$$

which is a contradiction. $\qquad \square$

Lemma 15. *Let $b \geqslant d^2(x^0, x^*)$. Then for any $\varepsilon > 0$:*

$$\exists k \leqslant \left\lceil \frac{b}{\varepsilon} \right\rceil \ \left(d^2(x^k, x^{k+1}) < \varepsilon \right).$$

Proof. Using Lemma 13, we get

$$d^2(x^k, x^{k+1}) \leqslant d^2(x^k, x^*) - d^2(x^{k+1}, x^*).$$

The result now follows from Lemma 14 using $a_k = d^2(x^k, x^*)$.   □

The preceding results allow us to give the following elementary proof of a rate of convergence for the iteration (x^k) and hence in particular also for the first time derive the convergence of it towards the minimizer in the general setting of Hadamard spaces.

Theorem 16. *Let X be a Hadamard space and let $f : X \to \mathbb{R}$ be a strongly quasiconvex lsc function with value $\gamma > 0$ and where $\mathrm{argmin} f \neq \varnothing$. Let (c_k) be a sequence of nonnegative reals with $\inf_{k \in \mathbb{N}} c_k \geqslant c > 0$. Then the sequence (x^k) defined by (PPA) converges to the (unique) minimizer x^* of f and $f(x^k)$ converges to $\min f$. Further:*

$$\forall \varepsilon > 0 \ \forall k \geqslant \varphi(\varepsilon) \ \left(d(x^k, x^*) < \varepsilon \right)$$

as well as

$$\forall \varepsilon > 0 \ \forall k \geqslant \varphi(\sqrt{2c\varepsilon}) + 1 \ \left(f(x^k) - \min f < \varepsilon \right)$$

where

$$\varphi(\varepsilon) := \left\lceil \frac{4b}{\varepsilon^2((1-\lambda_0)\gamma c - \lambda_0)^2} \right\rceil + 1$$

with $\lambda_0 = \frac{1}{2} \frac{\gamma c}{1+\gamma c}$ and $b \geqslant d^2(x^0, x^*)$.

Proof. It suffices to show the quantitative results. Given $\varepsilon > 0$, take

$$k \leqslant \left\lceil \frac{4b}{\varepsilon^2((1-\lambda_0)\gamma c - \lambda_0)^2} \right\rceil$$

such that

$$d(x^{k+1}, x^k) < \frac{1}{2}((1-\lambda_0)\gamma c - \lambda_0)\varepsilon$$

using Lemma 15. As $x^{k+1} \in \mathrm{Prox}_{c_k f}(x^k)$ and as $c_k \geqslant c > 0$, we get

$$f(x^{k+1}) \leqslant \max\{f(y), f(x^{k+1})\} + \frac{\lambda}{2}\left(\frac{\lambda}{c_k} - \gamma + \lambda\gamma\right) d^2(y, x^{k+1}) + \frac{\lambda}{c_k}\left\langle \overrightarrow{x^k x^{k+1}}, \overrightarrow{x^{k+1} y} \right\rangle$$

for all $y \in X$ and $\lambda \in [0, 1]$ from Lemma 11. Taking $y = x^*$, we in particular have

$$\frac{\lambda}{2}\left((1-\lambda)\gamma - \frac{\lambda}{c_k}\right) d^2(x^{k+1}, x^*) \leqslant \frac{\lambda}{c_k}\left\langle \overrightarrow{x^k x^{k+1}}, \overrightarrow{x^{k+1} x^*} \right\rangle$$

and hence for $\lambda > 0$ we get

$$\frac{1}{2}\left((1-\lambda)\gamma c - \lambda\right) d^2(x^{k+1}, x^*) \leqslant \left\langle \overrightarrow{x^k x^{k+1}}, \overrightarrow{x^{k+1} x^*} \right\rangle \leqslant d(x^{k+1}, x^k)d(x^{k+1}, x^*).$$

using (CS). Now, either $d(x^{k+1}, x^*) = 0$, in which case the claim follows immediately from Lemma 13, or $d(x^{k+1}, x^*) > 0$, in which case we have

$$\frac{1}{2}((1-\lambda)\gamma c - \lambda)d(x^{k+1}, x^*) \leqslant d(x^{k+1}, x^k)$$

for any $\lambda \in (0, 1]$. In particular, this holds for $\lambda = \lambda_0$ and we hence have

$$\frac{1}{2}((1 - \lambda_0)\gamma c - \lambda_0)d(x^{k+1}, x^*) \leqslant d(x^{k+1}, x^k) < \frac{1}{2}((1 - \lambda_0)\gamma c - \lambda_0)\varepsilon.$$

Since $(1 - \lambda_0)\gamma c - \lambda_0 > 0$ holds by definition, we get

$$d(x^{k+1}, x^*) < \varepsilon,$$

and so, using Lemma 13, we in particular have $d(x^j, x^*) < \varepsilon$ for any $j \geqslant k + 1$.

For the second claim, note that again as $x^{k+1} \in \mathrm{Prox}_{c_k f}(x^k)$, we have

$$f(x^{k+1}) \leqslant f(x) + \frac{1}{2c_k}d^2(x, x^k)$$

for any $k \in \mathbb{N}$ and any $x \in X$. In particular, for $x = x^*$, we get

$$f(x^{k+1}) - \min f \leqslant \frac{1}{2c}d^2(x^*, x^k).$$

Now, using the first claim, we get that $d(x^*, x^k) < \sqrt{2c\varepsilon}$ for any $k \geqslant \varphi(\sqrt{2c\varepsilon})$ and so

$$\frac{1}{2c}d^2(x^*, x^k) < \varepsilon$$

from which the second claim follows immediately. $\square$

Remark 17. The rate in the above Theorem 16 could also have been obtained by an application of the general results on rates of convergence for Fejér monotone sequences under general metric regularity assumptions as developed in [29]. Concretely, by Lemma 13, the iteration (x^k) is Fejér monotone w.r.t. $\mathrm{argmin} f$ and the assumption that f is strongly quasiconvex, which yields that the minimizer of f is unique, can be used to construct a so-called modulus of uniqueness, which is a particular instance of a modulus of regularity as introduced in [29] (see in particular the comprehensive discussions in [29] on such issues). In fact, the strategy of the proof of Theorem 16 is closely modeled after the proof of Theorem 4.1 in [29] but the above presentation allows us simultaneously to be more self-contained and to incorporate various slight optimizations into the rate and the argument.

Remark 18. The above Theorem 16 in particular implies Theorem 10 in [30] since any (closed and convex subset of a) Euclidean space is a Hadamard space and since a strongly quasiconvex lsc function in Euclidean space has a minimizer which follows from Theorem 1 in [30]. Naturally however, the above theorem is much broader as it not only applies in infinite dimensional Hilbert spaces but also in general arbitrary Hadamard spaces (recall also the discussion in the introduction). In particular, the rate of convergence from Theorem 16 applies to the context of Lara's work [30] which is not only, to our knowledge, the first general rate of convergence for that iteration but is even of very low complexity (being linear in the squared metric and in the convergence of the function values towards the minimizer).

Remark 19. The proof of Theorem 16 is still valid if the function is only strongly quasiconvex (i.e. not necessarily lsc) and if X is only a CAT(0) space (i.e. not necessarily complete) but in that case, we do not know if the iteration (x^k) is well-defined.

References

[1] A.D. Aleksandrov. A theorem on triangles in a metric space and some of its applications. *Trudy Matematicheskogo Instituta imeni V.A. Steklova*, 38:5–23, 1951.

[2] M. Bačák. The proximal point algorithm in metric spaces. *Israel Journal of Mathematics*, 194:689–701, 2013.

[3] M. Bačák. *Convex Analysis and Optimization in Hadamard Spaces*, volume 22 of *Series in Nonlinear Analysis and Applications*. Berlin, München, Boston: De Gruyter, 2014.

[4] M. Bačák, I. Searston, and B. Sims. Alternating projections in CAT(0) spaces. *Journal of Mathematical Analysis and Applications*, 385:599–607, 2012.

[5] I.D. Berg and I.G. Nikolaev. Quasilinearization and curvature of Aleksandrov spaces. *Geometriae Dedicata*, 133:195–218, 2008.

[6] H. Brézis and P.L. Lions. Produits infinis de resolvantes. *Israel Journal of Mathematics*, 29(4):329–345, 1978.

[7] M.R. Bridson and A. Haefliger. *Metric Spaces of Non-Positive Curvature*, volume 319 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin, Heidelberg, 1999.

[8] A.S. Brito, J.X. da Cruz Neto, J.O. Lopes, and P.R. Oliveira. Interior proximal algorithm for quasiconvex programming problems and variational inequalities with linear constraints. *Journal of Optimization Theory and Applications*, 154:217–234, 2012.

[9] F. Bruhat and J. Tits. Groupes réductifs sur un corps local. I. Données radicielles valuées. *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, 41:5–251, 1972.

[10] S. Dhompongsa, W.A. Kirk, and B. Sims. Fixed points of uniformly Lipschitzian mappings. *Nonlinear Analysis. Theory, Methods & Applications*, 65:762–772, 2006.

[11] S. Dhompongsa and B. Panyanak. On Δ-convergence theorems in CAT(0) spaces. *Computers & Mathematics with Applications*, 56(10):2572–2579, 2008.

[12] R. Espínola and A. Fernández-León. CAT(k)-spaces, weak convergence and fixed points. *Journal of Mathematical Analysis and Applications*, 353:410–427, 2009.

[13] P. Gerhardy and U. Kohlenbach. General logical metatheorems for functional analysis. *Transactions of the American Mathematical Society*, 360(5):2615–2660, 2008.

[14] K. Goebel and W.A. Kirk. Iteration processes for nonexpansive mappings. In S.P. Singh, S. Thomeier, and B. Watson, editors, *Topological methods in nonlinear functional analysis*, volume 21 of *Contemporary Mathematics*, pages 115–123. American Mathematical Society, Providence, RI, 1983.

[15] S.-M. Grad, F. Lara, and R.T. Marcavillaca. Relaxed-inertial proximal point type algorithms for quasiconvex minimization. *Journal of Global Optimization*, 85(3):615–635, 2023.

[16] S.-M. Grad, F. Lara, and R.T. Marcavillaca. Relaxed-Inertial Proximal Point Algorithms for Nonconvex Equilibrium Problems with Applications. *Journal of Optimization Theory and Applications*, 203(3):2233–2262, 2024.

[17] S.-M. Grad, F. Lara, and R.T. Marcavillaca. Strongly Quasiconvex Functions: What We Know (So Far). *Journal of Optimization Theory and Applications*, 205, 2025. 38, 41pp.

[18] M. Gromov. Hyperbolic groups. In S.M. Gersten, editor, *Essays in group theory*, volume 8 of *Mathematical Sciences Research Institute Publications*, pages 75–263. Springer, New York, 1987.

[19] O. Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM Journal on Control and Optimization*, 29:403–419, 1991.

[20] A. Iusem and F. Lara. Proximal point algorithms for quasiconvex pseudomonotone equilibrium problems. *Journal of Optimization Theory and Applications*, 193(1):443–461, 2022.

[21] J. Jost. Equilibrium maps between metric spaces. *Calculus of Variations and Partial Differential Equations*, 2:173–204, 1994.

[22] J. Jost. Convex functionals and generalized harmonic maps into spaces of nonpositive curvature. *Commentarii Mathematici Helvetici*, 70:659–673, 1995.

[23] M. Jovanovič. A note on strongly convex and quasiconvex functions. *Mathematical Notes*, 60:584–585, 1996.

[24] A. Kabgani and F. Lara. Strong subdifferentials: theory and applications in nonconvex optimization. *Journal of Global Optimization*, 84(2):349–368, 2022.

[25] W.A. Kirk and B. Panyanak. A concept of convergence in geodesic spaces. *Nonlinear Analysis. Theory, Methods & Applications*, 68:3689–3696, 2008.

[26] U. Kohlenbach. Some logical metatheorems with applications in functional analysis. *Transactions of the American Mathematical Society*, 357(1):89–128, 2005.

[27] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer-Verlag Berlin Heidelberg, 2008.

[28] U. Kohlenbach. Proof-theoretic Methods in Nonlinear Analysis. In B. Sirakov, P. Ney de Souza, and M. Viana, editors, *Proc. ICM 2018*, volume 2, pages 61–82. World Scientific, 2019.

[29] U. Kohlenbach, G. López-Acedo, and A. Nicolae. Moduli of regularity and rates of convergence for Fejér monotone sequences. *Israel Journal of Mathematics*, 232:261–297, 2019.

[30] F. Lara. On Strongly Quasiconvex Functions: Existence Results and Proximal Point Algorithms. *Journal of Optimization Theory and Applications*, 192:891–911, 2022.

[31] F. Lara and R.T. Marcavillaca. Bregman type proximal point algorithms for quasiconvex minimization. *Optimization*, 73(3):497–515, 2024.

[32] T.C. Lim. Remarks on some fixed point theorems. *Proceedings of the American Mathematical Society*, 60:179–182, 1976.

[33] B. Martinet. Régularisation dinéquations variationnelles par approximations successives. *Revue française d'informatique et de recherche opérationnelle*, 4:154–159, 1970.

[34] S. Pan and J.-S. Chen. Entropy-like proximal algorithms based on a second-order homogeneous distance function for quasi-convex programming. *Journal of Global Optimization*, 39:555–575, 2007.

[35] B.T. Polyak. Existence Theorems and Convergence of Minimizing Sequences in Extremum Problems with Restrictions. *Soviet Mathematics. Doklady*, 7:72–75, 1966.

[36] E.A. Papa Quiroz. Proximal Point Method for Quasiconvex Functions in Riemannian Manifolds. *Journal of Optimization Theory and Applications*, 202:1268–1285, 2024.

[37] E.A. Papa Quiroz, N. Baygorrea Cusihuallpa, and N. Maculan. Inexact proximal point methods for multi-objective quasiconvex minimization on Hadamard manifolds. *Journal of Optimization Theory and Applications*, 186:879–898, 2020.

[38] E.A. Papa Quiroz and P.R. Oliveira. Proximal point methods for quasiconvex and convex functions with Bregman distances on Hadamard manifolds. *Journal of Convex Analysis*, 16(1):49–69, 2009.

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

[40] E.A. Papa Quiroz and P.R. Oliveira. Full convergence of the proximal point method for quasiconvex functions on Hadamard manifolds. *ESAIM: Control, Optimisation and Calculus of Variations*, 18(2):483–500, 2012.

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

[42] S. Reich and I. Shafrir. Nonexpansive iterations in hyperbolic spaces. *Nonlinear Analysis. Theory, Methods & Applications*, 15(6):537–558, 1990.

[43] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal of Control and Optimization*, 14:877–898, 1976.

[44] W. Takahashi. A convexity in metric space and nonexpansive mappings. I. *Kodai Mathematical Seminar Reports*, 22:142–149, 1970.