

ON STRONGLY QUASICONVEX PSEUDOMONOTONE EQUILIBRIUM PROBLEMS IN HADAMARD SPACES

LUISA MARIE DESPRÉS^a AND NICHOLAS PISCHKE^b

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

^b Department of Computer Science, University of Bath,
Claverton Down, Bath, BA2 7AY, United Kingdom,

E-mails: luisa.despres@stud.tu-darmstadt.de, nnp39@bath.ac.uk

ABSTRACT. We study two proximal point type methods for finding equilibrium points of pseudomonotone and strongly quasiconvex bifunctions. Extending results by A. Iusem and F. Lara, we prove the strong convergence of these methods over general complete geodesic metric spaces of nonpositive curvature, so-called Hadamard spaces. Our arguments are quite elementary and in particular effective, yielding sublinear non-asymptotic guarantees for the distance of the iterates towards the solution. These quantitative results are novel even in the context of Euclidean spaces, the original setting of the work by Iusem and Lara, and the simplicity of our arguments allows us to either weaken or even fully discharge various assumptions featuring in this previous work. We also provide an existence result for solutions of equilibrium problems generated by suitably semicontinuous, pseudomonotone and strongly quasiconvex bifunctions over general Hadamard spaces, and derive from this that every lower-semicontinuous strongly quasiconvex function over a Hadamard space has a minimizer, answering a question of the second author.

Keywords: Equilibrium problems; Strong quasiconvexity; Hadamard spaces; Rates of convergence; Proof mining

MSC2020 Classification: 49J40, 47J25, 90C26, 03F10

1. INTRODUCTION

1.1. Background and motivation. Equilibrium problems, that is problems of the form

$$\text{find } x^* \in K \text{ such that } f(x^*, y) \geq 0 \text{ for all } y \in K,$$

where $f : K \times K \rightarrow \mathbb{R}$ is a suitable bifunction (e.g. satisfying appropriate continuity, convexity, and monotonicity conditions) and $K \subseteq X$ is a constraint set (commonly chosen closed and convex) over some space X , play a fundamental role in modern applied mathematics. Going back to the work of Fan [24], convex and more recently also in particular nonconvex equilibrium problems have been studied deeply. We refer e.g. to [11, 35, 40, 59, 65], and respectively to [20, 26, 36], among many others.

In the present paper, we study different approximation methods for such equilibrium problems. These methods, in one way or the other, are derived from the seminal proximal point algorithm originating in the work of Martinet [60], Rockafellar [75] as well as Brezis and Lions [13], originally devised to solve convex minimization problems (and, in a generalized formulation, inclusion problems of maximally monotone operators). Concretely, over a Hilbert space $(X, \langle \cdot, \cdot \rangle)$ with norm $\|\cdot\|$, and given a lower-semicontinuous convex function $h : K \rightarrow \mathbb{R}$ over a closed and convex set $K \subseteq X$, the proximal point algorithm iteratively compute points $\{x_k\}_k$ via

$$x_{k+1} := \text{Prox}_{\beta_k h}(x_k)$$

for a given starting point $x_0 \in K$, a sequence $\{\beta_k\}_k \subseteq (0, \infty)$ of parameters satisfying suitable conditions, and where $\text{Prox}_{\beta h}$ is the proximal map of h , defined by

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

As $h(\cdot) + \frac{1}{2\beta} \|\cdot - x\|^2$ is strongly convex whenever h is convex, this so-called proximal subproblem has a unique solution so that $\text{Prox}_{\beta h}(x)$ is a well-defined function. The sequence $\{x_k\}_k$ generated by this algorithm converges weakly to a minimizer for h if there exists one (see e.g. [4] for a modern canonical reference). If the convexity condition on h is strengthened to strong convexity, one even obtains strong convergence to the unique minimizer of h .

Proximal point type algorithms for equilibrium problems have been considered for convex bifunctions at least since the well-known work of Moudafi [61] and Moudafi and Théra [63], and have subsequently been extended in various ways (we refer in particular to [16, 42, 45, 52, 62], just to name a few).

At the core of it, the perhaps most prominent proximal point type algorithm for equilibrium problems (already going back to [61, 63]) produces an approximating sequence $\{x_k\}_k$ by using the current iterate x_k to define a regularization of the bifunction *as a whole* via

$$f_k(x, y) := f(x, y) + \frac{1}{\beta_k} \langle x - x_k, y - x \rangle,$$

and then solving the equilibrium problem associated with f_k to generate x_{k+1} . The set-valued map

$$J_{\beta}^f(x) := \left\{ z \in K \mid f(z, y) + \frac{1}{\beta} \langle z - x, y - z \rangle \geq 0 \text{ for all } y \in K \right\},$$

associating with every point x and parameter β the set of solutions of this corresponding regularized equilibrium problem, is also called the *resolvent of the bifunction f* (see e.g. [11, 19, 63]) and is single-valued whenever f is monotone as well as convex and lower-semicontinuous in its right argument (cf. [19, Lemma 2.12]), so that, using this notion, the regularized selection method sketched above reduces to the more familiar pattern $x_{k+1} := J_{\beta_k}^f(x_k)$ of the proximal point algorithm discussed above.

A second type of proximal point algorithm applies a regularization akin to the proximal subproblem only to the second (convex) argument of the bifunction, with the first argument fixed to the current iterate x_k , and so defines

$$x_{k+1} := \operatorname{argmin}_{y \in K} \left\{ f(x_k, y) + \frac{1}{2\beta_k} \|y - x_k\|^2 \right\} = \text{Prox}_{\beta_k f(x_k, \cdot)}(x_k).$$

In fact, to our knowledge, this second type of proximal point method for equilibrium problems was only recently introduced in the work of Iusem and Lara [37] (see also [34, 39] and the very recent work [12]), to solve equilibrium problems with relaxed convexity assumptions, which will also be the main concern of this paper.

Concretely, the work [37] studies the two methods above for bifunctions which exchange the convexity requirement for the right argument of f with *strong quasiconvexity*. The class of quasiconvex functions, that is functions $h : K \rightarrow \mathbb{R}$ which satisfy

$$h((1 - \lambda)x + \lambda y) \leq \max\{h(x), h(y)\}$$

for all $x, y \in K$ and $\lambda \in [0, 1]$, widely relaxes convexity and covers many practically interesting functions that fail to be convex. However, as highlighted and discussed in detail in e.g. [54], this class simultaneously suffers from a range of problems in regards to iterative procedures, in particular the proximal point algorithm. Starting with the work of Lara [54], considerable

focus has recently been placed on the class of strongly quasiconvex functions, that is functions $h : K \rightarrow \mathbb{R}$ which satisfy

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

for all $x, y \in K$ and $\lambda \in [0, 1]$, for a given modulus $\gamma > 0$. This class already goes back to the work of Polyak [69] and has various benefits over the class of quasiconvex functions. In particular, minimizers of such functions are unique if they exist and, partly as a consequence, many iterative procedures known from convex settings are still reasonably well-behaved. Indeed, in [54], Lara studies the proximal point method for strongly quasiconvex functions and, while the proximal map in this context is generally only set-valued (cf. [54, Remark 6]) so that the iterative step then takes the form of a selection $x_{k+1} \in \text{Prox}_{\beta_k h}(x_k)$, he proves a convergence theorem over Euclidean space \mathbb{R}^d (cf. [54, Theorem 10]). It is then for (pseudomonotone and suitably Lipschitz) bifunctions which are only strongly quasiconvex in their second argument that Iusem and Lara in [37] establish the convergence of the two proximal point type methods discussed above over Euclidean space \mathbb{R}^d (cf. [37, Theorem 3.1] and [37, Corollary 3.2]), building on the previous work of Lara [54].

All the while, the class is not too restrictive, containing various nonconvex functions of practical interest such as the Euclidean norm (cf. [44, Theorem 2]) which is not strongly convex on any bounded convex set, the square root of the Euclidean norm which is strongly quasiconvex on each ball (cf. [54, Theorem 17]) without being differentiable, or even functions such as $\max\{\sqrt{\|\cdot\|}, \|\cdot\|^2 - k\}$ for $k \in \mathbb{N}$ which are strongly quasiconvex without even being convex (cf. [54, Remark 18]). We refer to [54] and the more recent survey [30] for further discussions on the relationships between strongly quasiconvex functions and various other classes of (non)convex functions.

1.2. The contributions of the present paper and related work. In the present paper, we study the two proximal point type methods sketched above for strongly quasiconvex pseudomonotone equilibrium problems over Hadamard spaces, that is complete geodesic metric spaces of nonpositive curvature. Also known as (complete) CAT(0)-spaces, this general class of nonlinear spaces contains and unifies spaces such as (infinite-dimensional) Hilbert spaces, \mathbb{R} -trees and Hadamard manifolds, that is complete simply connected Riemannian manifolds of nonpositive sectional curvature, among many more.

While the fundamental concepts of Hadamard spaces can already be dated back to a paper by Wald [76], the notion reached maturity through the work of Alexandrov [1, 2] (after which such spaces are also called spaces of nonpositive curvature in the sense of Alexandrov). The moniker CAT(0) derives from the influential work of Gromov [31]. We refer to [6, 7, 14] for further historical remarks, and in general refer to [14] for an authoritative treatment of these spaces at large. Next to their traditional role in geometry, Hadamard spaces are a suitable framework for extending many concepts from convex analysis and optimization on linear spaces to nonlinear contexts, as is the topic of this paper, and we refer to [6, 7] for discussions in that direction. We provide a technical overview of the key aspects of Hadamard spaces as relevant to the present paper in Section 2 below.

The proximal point method for convex minimization was previously investigated on various subclasses of these spaces by a range of authors, such as Ferreira and Oliveira [25] on Hadamard manifolds (see also [57, 72]), which culminated in the seminal work of Bačák [5] who proved the weak convergence of the proximal point method in general Hadamard spaces. Extensions of such results to equilibrium problems have also been studied by various authors, such as Colao, López, Marino and Martín-Márquez [18] set in Hadamard manifolds or Khatibzadeh

and Mohebbi [46] as well as Iusem and Mohebbi [39] over Hadamard spaces. Yet another approach over Hadamard manifolds, utilizing Busemann functions, was provided recently by Bento, Cruz Neto, Melo, et al. [8, 9]. In particular, these works outline three distinct approaches for extending the regularization of the full bifunction to nonlinear spaces, all of which we will study in this paper. In fact, to our knowledge, in two of these cases (modeled after the works [18] and [8, 9], respectively), the corresponding regularization is considered in general Hadamard spaces for the first time.

Moving to the setting of quasiconvex functions in such nonlinear contexts, there is similarly a wide array of literature, in particular on proximal point methods. These are however, to our knowledge, largely limited to the setting of minimization problems over Hadamard manifolds, and we refer in particular to the work of Papa Quiroz and his co-authors [70, 71, 72, 73, 74]. Work on strongly quasiconvex functions in these contexts is however very limited so far and, to our knowledge, only comprises the recent work [67] of the second author where he extends Lara’s convergence results from [54] to the setting of Hadamard spaces.

In the present paper, we now follow the approach of [67] and, for appropriate metric variants of the methods from [37], establish their convergence over Hadamard spaces. Already for infinite-dimensional Hilbert spaces, these results seem to be new to the literature. The convergence proofs we provide are, similar to [67], rather different from those given in [37] and in particular quite elementary, for example being void of any (weak) compactness arguments. This elementarity of the convergence proofs given here is not only key for overcoming the obstacles of extending the results from [37] to general Hadamard spaces, but in particular allows us to simultaneously provide quantitative variants of these convergence results. Concretely, in the setting of a Hadamard space (X, d) , we provide explicit non-asymptotic guarantees for the distance of the iterates $\{x_k\}_k$ generated by any of the considered methods towards the (in the context of a strongly quasiconvex and pseudomonotone bifunction necessarily unique) solution x^* of the corresponding equilibrium problem. These take the concrete form of

$$d(x_k, x^*) < \frac{C}{\sqrt{k-2}} \text{ for all } k > 2$$

for a specific (explicitly calculated) constant C which only depends on moduli γ and η featuring in the strong quasiconvexity and a respective Lipschitz property of the bifunction (similar to [37]) and an upper bound $b^2 \geq d^2(x_0, x^*)$ on the initial displacement. Our quantitative estimates are otherwise completely independent of the space, iteration or bifunction generating the equilibrium problem. To our knowledge, these rates are novel even over Euclidean spaces, the original setting of the work of Iusem and Lara [37]. These quantitative results can be found in Theorems 4.14, 4.18 and 5.10 below.

In particular, our convergence results are thereby “strong”, i.e. we establish the convergence of the methods relative to the metric and not relative to any notion of weak convergence. This is essentially due to the fact that equilibrium problems for strongly quasiconvex and pseudomonotone bifunctions have a unique solution, and in particular sets such results apart from the “usual” proximal point method for (bi)convex functions which, as shown by Güler [32], generally only converges weakly already in Hilbert spaces.

In the course of our convergence proofs, we in particular further show how some semicontinuity assumptions on the bifunction featuring in [37] can be avoided either conditionally or even unconditionally, generalizing the respective range of bifunctions and clarifying the contributions of these continuity assumptions to the convergence of the methods (see Remarks 4.15 and 5.11 later on).

Further, in the context of the second type of method which regularizes the bifunction only in its second argument, we are able to weaken the previous parameter restriction $\{\beta_k\}_k \subseteq (\frac{1}{\gamma-8\eta}, \frac{1}{4\eta})$ featuring in [37] to $\{\beta_k\}_k \subseteq [\frac{1}{\gamma-8\eta}, \frac{1}{4\eta}]$. While perhaps only a marginal improvement at first sight, this extension of the parameter range is actually crucial for establishing our quantitative results for that method, as discussed in more detail later (see the discussion in Section 4.2 later on). In particular, this extension requires various additional technicalities compared to the proof of the original result given in [37].

Also, in the context of the first type of method which regularizes the full bifunction, we highlight some issues in the well-definedness argument of that method provided in [37] which leave it open whether this method is generally well-defined for strongly quasiconvex bifunctions already over Euclidean spaces (see Remark 5.2 later on).

As a last contribution, we provide a theorem on the existence of solutions for the equilibrium problem of suitably semicontinuous, pseudomonotone and strongly quasiconvex bifunctions over Hadamard spaces (see Theorem 3.9 later on), lifting the corresponding result obtained by Iusem and Lara (cf. [37, Proposition 3.1]) from the linear setting to this class of nonlinear spaces. Our argument here in particular relies on combining ideas from [37, 54] with the work of Khatibzadeh and Mohebbi [46], where the existence of equilibrium points over Hadamard spaces is studied for certain (in particular convex) bifunctions. As a consequence, we in particular establish that every lower-semicontinuous strongly quasiconvex function on a Hadamard space has a minimizer (see Corollary 3.10 later on), which answers a question left open in the previous work [67] of the second author and lifts a corresponding result of Lara (cf. [54, Corollary 3]) to this nonlinear setting.

1.3. Proof mining. All of the present quantitative results have been derived using methods from proof mining, a program in mathematical logic that applies results from proof theory to areas of core mathematics to classify and extract the computational content of *prima facie* “non-computational” proofs therein. We refer to the seminal monograph [49] for a comprehensive overview of this area and to the survey [50] for an overview of various applications to nonlinear analysis in particular. A detailed discussion on the concrete use of logical methods in the context of the present paper can be found in the master thesis of the first author [21]. The present paper however is (apart from a few short remarks labeled “for logicians”) presented without any explicit reference to logic and does not require any such background.

1.4. Future work. The results of Lara [54] as well as Iusem and Lara [37] have since been extended in various directions by Lara and his co-authors (see e.g. [3, 28, 29, 33, 38, 55, 56], among others), and so there is a natural range of potential future works stemming from the present paper.

In particular, we want to highlight the extensions of the proximal point algorithm which incorporate inertia terms and over-relaxations, as developed by Grad, Lara and Marcavillaca in [28] for strongly quasiconvex minimization problems and in [29] (see again also [34]) for associated equilibrium problems. A forthcoming paper by the authors (based, similarly to the present paper, on the master thesis of the first author [21]) will provide analogous (quantitative) results also for these methods, which in particular crucially rely on the approach and methods developed in the present paper.

Further, more speculative future work which we want to highlight includes extensions of the methods studied in the present paper to incorporate Bregman distances over Hadamard manifolds, which could be approached by combining the method of the present work with that of preceding work on Bregman distances and associated proximal point methods in the context of convex bifunctions [16] and strongly quasiconvex (bi)functions [3, 55] and with work on

Bregman distances over such manifolds [72] (or potentially even work of the second author [68] on Bregman distances in Hadamard spaces).

2. PRELIMINARIES

2.1. Hadamard spaces. Let (X, d) be a metric space. A geodesic in X is defined to be an isometry $\gamma : [0, l] \rightarrow X$, and we say that γ joins $\gamma(0)$ and $\gamma(l)$. A geodesic space X is called (uniquely) geodesic if every two points are joined by a (unique) geodesic. A geodesic ray is an isometry $r : [0, \infty) \rightarrow X$, and we say that r issues from $r(0)$. X is said to have the geodesic extension property if for all $x \neq y \in X$, there is a ray $r : [0, \infty) \rightarrow X$ issuing from x such that $r(t) = y$ for some $t > 0$. A geodesic metric space (X, d) is called a CAT(0)-space if it satisfies

$$(CN^+) \quad d^2(\gamma(\lambda l), x) \leq (1 - \lambda)d^2(\gamma(0), x) + \lambda d^2(\gamma(l), x) - \lambda(1 - \lambda)d^2(\gamma(0), \gamma(l))$$

for all $x \in X$, all $\lambda \in [0, 1]$ and all geodesics $\gamma : [0, l] \rightarrow X$, an extension of the so-called Bruhat-Tits CN-inequality [15] to geodesics. Any CAT(0)-space is uniquely geodesic and a complete CAT(0)-space is called a Hadamard space. Over a Hadamard space (X, d) , we write $(1 - \lambda)x \oplus \lambda y$ for the point $\gamma(\lambda d(x, y))$ on the unique geodesic $\gamma : [0, d(x, y)] \rightarrow X$ joining x and y . We call a subset $K \subseteq X$ convex, if with $x, y \in K$ and $\lambda \in [0, 1]$ also $(1 - \lambda)x \oplus \lambda y \in K$. We denote the convex hull of a set $A \subseteq X$ by $\text{conv}(A)$.

In this paper, we also rely on an equivalent characterization of Hadamard spaces through a metric version of the Cauchy-Schwarz inequality for the so-called quasi-inner product due to Berg and Nikolaev [10], that is the function

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

As shown in [10, Proposition 14], this function is the unique map $\langle \cdot, \cdot \rangle : X^2 \times X^2 \rightarrow \mathbb{R}$ satisfying the following conditions for any $x, v, u, y \in X$: $\langle \overrightarrow{xy}, \overrightarrow{xy} \rangle = d^2(x, y)$; $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle = \langle \overrightarrow{uv}, \overrightarrow{xy} \rangle$; $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle = -\langle \overrightarrow{yx}, \overrightarrow{uv} \rangle$; $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle + \langle \overrightarrow{xy}, \overrightarrow{vw} \rangle = \langle \overrightarrow{xy}, \overrightarrow{uw} \rangle$. Now, as established in [10, Corollary 3], a geodesic metric space (X, d) is a CAT(0)-space if, and only if,

$$(CS) \quad \langle \overrightarrow{xy}, \overrightarrow{uv} \rangle \leq d(x, y)d(u, v)$$

for all $x, y, u, v \in X$.

2.2. (Quasi)convexity over Hadamard spaces. Let (X, d) be a Hadamard space. A function $h : X \rightarrow \mathbb{R}$ is called *convex* if

$$h((1 - \lambda)x \oplus \lambda y) \leq (1 - \lambda)h(x) + \lambda h(y)$$

for all $x, y \in X$ and $\lambda \in [0, 1]$. A particular example of a convex function that we will need later is the Busemann function $b_r : X \rightarrow \mathbb{R}$ associated with a geodesic ray r , defined by

$$b_r(x) := \lim_{t \rightarrow \infty} (d(x, r(t)) - t).$$

In particular, b_r is convex and nonexpansive, i.e. 1-Lipschitz (cf. [6, Example 2.2.10]).

As motivated in the introduction, we will here be focused on the weaker notion of quasi-convexity, and its respective strengthening of strong quasiconvexity. In that vein, a function $h : X \rightarrow \mathbb{R}$ is called

(i) *quasiconvex* if

$$h((1 - \lambda)x \oplus \lambda y) \leq \max\{h(x), h(y)\}$$

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

(ii) *strongly quasiconvex* with modulus $\gamma > 0$ if

$$h((1 - \lambda)x \oplus \lambda y) \leq \max\{h(x), h(y)\} - \lambda(1 - \lambda)\frac{\gamma}{2}d^2(x, y)$$

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

Next to notions of quasiconvexity, we will also need the following notions of semicontinuity: A function $h : X \rightarrow \mathbb{R}$ is called *lower semicontinuous* (lsc) at $x \in X$ if

$$h(x) \leq \liminf_{k \rightarrow \infty} h(x_k)$$

for any sequence $\{x_k\}_k \subseteq X$ with $x_k \rightarrow x$. If instead it holds that

$$h(x) \geq \limsup_{k \rightarrow \infty} h(x_k)$$

for any sequence $\{x_k\}_k \subseteq X$ with $x_k \rightarrow x$, h is called *upper semicontinuous* (usc) at $x \in X$.

2.3. Tangent cones. We will later utilize the so-called tangent cones of a CAT(0)-space as developed in the work of Nikolaev [64] (see also [14] for further exposition). Fix a CAT(0)-space (X, d) . For two nonconstant geodesics γ, η issuing from a point $x \in X$, their Alexandrov angle $\angle_x(\gamma, \eta)$ is defined by

$$\angle_x(\gamma, \eta) := \lim_{s, t \rightarrow 0^+} \bar{\angle}_x(\gamma(s), \eta(t)),$$

where $\bar{\angle}_x(y, z)$ is the comparison angle defined through the comparison triangle $\bar{\Delta}(\bar{x}, \bar{y}, \bar{z})$ of the geodesic triangle $\Delta(x, y, z) \subseteq X$ (see [14]). Using the fact that \angle_x defines a pseudometric on the set of all nonconstant geodesics issuing from x , we define $\Sigma'_x X$ to be the set of all equivalence classes of such geodesics under the equivalence relation defined by $\angle_x(\gamma, \eta) = 0$. The completion $(\Sigma_x X, \angle_x)$ of the space $(\Sigma'_x X, \angle_x)$ is called the metric space of directions from x and the tangent cone $T_x X$ of X at x is then the Euclidean cone over $\Sigma_x X$, that is $T_x X := (\Sigma_x X \times [0, \infty)) / \sim$ where $(\gamma, t) \sim (\eta, s)$ if, and only if, $t = s = 0$ or $t = s > 0$ and $\gamma = \eta$. For brevity, we write $t\gamma$ for the equivalence class $[(\gamma, t)]_{\sim}$ and given $u = t\gamma$ and $\lambda \geq 0$, we write $\lambda u := (\lambda t)\gamma$. On $T_x X$, we define a metric

$$d_x(t\gamma, s\eta) := \sqrt{t^2 + s^2 - 2ts \cos \angle_x(\gamma, \eta)}.$$

By the results of Nikolaev [64], it follows that $(T_x X, d_x)$ is a Hadamard space. We further write $0_x := 0\gamma$ as well as $\|t\gamma\|_x := d_x(0_x, t\gamma) = t$ and define

$$g_x(t\gamma, s\eta) := \frac{1}{2} (\|t\gamma\|_x^2 + \|s\eta\|_x^2 - d_x^2(t\gamma, s\eta)) = ts \cos \angle_x(\gamma, \eta).$$

It can be easily seen that $g_x(t\gamma, s\eta) = \langle \overrightarrow{0_x t\gamma}, \overrightarrow{0_x s\eta} \rangle_x$, where $\langle \vec{\cdot}, \vec{\cdot} \rangle_x$ is the quasi-inner product on $T_x X$, and so $g_x(t\gamma, s\eta) \leq \|t\gamma\|_x \|s\eta\|_x$ by (CS) on $T_x X$. Further, one has $g_x(t\gamma, t\gamma) = \|t\gamma\|_x^2$, $g_x(t\gamma, s\eta) = g_x(s\eta, t\gamma)$ and $g_x(t\gamma, s\eta) = tg_x(\gamma, s\eta)$.

Similar as e.g. Ohta [66] we define the inverse exponential map in a general Hadamard space to be the function $\log_x : X \rightarrow T_x X$ given by $\log_x a := d(x, a)\gamma_{x,a}$ for $a \neq x$, where $\gamma_{x,a}$ is the (unique) geodesic connecting x to a , as well as $\log_x x := 0_x$. The central property of the inverse exponential map \log_x in relation to the pseudo-inner product g_x that we need is the following:

Lemma 2.1 (essentially [17, Proposition 2.16]). *For any $x, a, b \in X$:*

$$g_x(t \log_x a, s \log_x b) \geq \frac{ts}{2} (d^2(x, a) + d^2(x, b) - d^2(a, b)).$$

2.4. Weak convergence in Hadamard spaces. We now introduce the notion and basic properties of weak convergence in Hadamard spaces. First introduced by Jost [43], in [22], Espínola and Fernández-León show that the same notion was obtained by Kirk and Panyanak in [47], who brought Lim's Δ -convergence [58] into Hadamard spaces. We will thus denote weak convergence by Δ -convergence, notationally agreeing with [46], whose results we will crucially rely on in Section 3 later on for establishing the existence of equilibrium points. For a thorough discussion of weak convergence in Hadamard spaces, see e.g. [6].

Let (X, d) be a Hadamard space and $\{x_k\}_k \subseteq X$ be a bounded sequence. For $x \in X$, set

$$r(x, \{x_k\}_k) := \limsup_{n \rightarrow \infty} d(x, x_k).$$

The *asymptotic radius* $r(\{x_k\}_k)$ is given by

$$r(\{x_k\}_k) := \inf\{r(x, \{x_k\}_k) \mid x \in X\},$$

and the *asymptotic center* $A(\{x_k\}_k)$ of $\{x_k\}_k$ is the set

$$A(\{x_k\}_k) := \{x \in X \mid r(\{x_k\}_k) = r(x, \{x_k\}_k)\}.$$

It is known that $A(\{x_k\}_k)$ is a singleton in Hadamard spaces (see e.g. [47]).

A bounded sequence $\{x_k\}_k \subseteq X$ is then said to Δ -converge to $x \in X$ if x is the unique asymptotic center of every subsequence of $\{x_k\}_k$. We write $x_k \rightarrow^\Delta x$ and call x the Δ -limit of $\{x_k\}_k$. We further call a point $x \in X$ a Δ -cluster point of a sequence $\{x_k\}_k$ if there is a subsequence $\{x_{k_l}\}_l$ of $\{x_k\}_k$ such that $x_{k_l} \rightarrow^\Delta x$.

We will later rely on various results regarding this notion of weak convergence, the most essential of which we list in the following:

Lemma 2.2 (cf. [47, Proposition 3.5]). *Any bounded sequence in X has a Δ -cluster point.*

Lemma 2.3 (cf. [6, Lemma 3.2.1]). *Let $K \subseteq X$ be a closed and convex set and $\{x_k\}_k \subseteq K$. If the sequence $\{x_k\}_k$ Δ -converges to a point $x \in X$, then $x \in K$.*

Together, the above two lemmas in particular yield:

Lemma 2.4 (cf. [47, Proposition 3.6]). *Every bounded closed convex set in X is Δ -compact.*

We say that a function $h : X \rightarrow \mathbb{R}$ is Δ -lower semicontinuous (Δ -lsc) at a point $x \in X$ if

$$\liminf_{n \rightarrow \infty} h(x_k) \geq h(x)$$

for each sequence $\{x_k\}_k$ with $x_k \rightarrow^\Delta x$. The property of $h : X \rightarrow \mathbb{R}$ being Δ -upper semicontinuous (Δ -usc) is defined analogously.

Lemma 2.5 (cf. [67, Lemma 3.7], extending [6, Lemma 3.2.3]). *Let $h : X \rightarrow \mathbb{R}$ be a quasiconvex lsc function. Then h is Δ -lsc.*

3. EQUILIBRIUM PROBLEMS OVER HADAMARD SPACES AND EXISTENCE OF SOLUTIONS

We now fix the class of bifunctions and associated equilibrium problems studied in this paper: Given a Hadamard space (X, d) and a bifunction $f : X \times X \rightarrow \mathbb{R}$, the task is to

$$(EP) \quad \text{find } x^* \in X \text{ such that } f(x^*, y) \geq 0 \text{ for all } y \in X.$$

We denote the associated solution set by $S(f)$. The focus will be on a special class of pseudomonotone, strongly quasiconvex and suitably Lipschitz bifunctions. As we throughout also utilize other auxiliary properties, we enumerate the various assumptions featuring in this paper for a better overview (compare [37]):

- (A1) For every $x \in X$, the function $f(x, \cdot)$ is lsc, and for every $y \in X$, the function $f(\cdot, y)$ is usc.
 (A2) f is pseudomonotone, i.e.

$$f(x, y) \geq 0 \text{ implies } f(y, x) \leq 0$$

for all $x, y \in X$.

- (A3) f is lsc (jointly in both arguments).
 (A4) For every $x \in X$, the function $f(x, \cdot)$ is strongly quasiconvex with modulus $\gamma > 0$.
 (A5) f satisfies the following Lipschitz condition: there exists $\eta > 0$ such that

$$f(x, z) - f(x, y) - f(y, z) \leq \eta (d^2(x, y) + d^2(y, z))$$

for all $x, y, z \in X$.

- (A6) The Lipschitz constant η and the modulus of strong quasiconvexity γ are such that $12\eta < \gamma$.

While (A1) and (A2) are standard assumptions in the literature (see e.g. [46] for such assumptions in the context of equilibrium problems on Hadamard spaces), the assumptions (A4), (A5) and (A6) are metric versions of the corresponding properties used in [37] over \mathbb{R}^d , and replace the usual assumption of convexity of f . As in [37], we also note here that the standard assumption

$$(A0) \quad f(x, x) = 0 \text{ for all } x \in X.$$

is derivable from (A2) and (A5), although we will later also make use of it individually. (A3) seems to be a more uncommon assumption, and we will in particular show that it can generally be avoided in the context of the convergence and existence theorems of the present paper, improving the results of [37] (see also Remark 4.15 later on).

Before moving on to the question when and under what assumptions the above equilibrium problem (EP) has a solution in the nonlinear context of Hadamard spaces, we give a key example of a bifunction that illustrates the above assumptions (A1) – (A6). Further examples of bifunctions satisfying the above conditions simultaneously can be found in Section 4 of [37].

Example 3.1 (essentially [37, Corollary 3.1]). Let $h : X \rightarrow \mathbb{R}$ be lsc and strongly quasiconvex with modulus $\gamma > 0$. Consider the bifunction $f_h : X \times X \rightarrow \mathbb{R}$ given by

$$f_h(x, y) := h(y) - h(x) \text{ for all } x, y \in X.$$

It is then easy to see that $S(f_h) = \operatorname{argmin}_{x \in X} h(x)$, that is the solutions of the equilibrium problem for f_h correspond exactly to the solutions of the minimization problem for h . Further, f_h fulfills (A1), (A2) and (A4) – (A6). Here, (A1), (A2) and (A4) are rather immediate (see also the proof of Corollary 3.10 later on) and (A5) as well as (A6) follow from the fact that $f_h(x, z) - f_h(x, y) - f_h(y, z) = 0$.

Now, at first note that the strong quasiconvexity and pseudomonotonicity of f imply that $S(f)$ is either empty or a singleton, which generalizes parts of [37, Proposition 3.1] to our nonlinear setting (and in particular extends [67, Lemma 3.1] from minimization problems to equilibrium problems).

Lemma 3.2. *Suppose that f satisfies (A0), (A2) and (A4). Then $S(f)$ is at most a singleton.*

Proof. If $x_1, x_2 \in S(f)$ are such that $x_1 \neq x_2$, then $f(x_1, x_2) \geq 0$ and $f(x_2, x_1) \geq 0$. By pseudomonotonicity, we have $f(x_2, x_1) = 0$. Using the strong quasiconvexity of $f(x_2, \cdot)$, we get

$$f\left(x_2, \frac{1}{2}x_2 \oplus \frac{1}{2}x_1\right) \leq \max\{f(x_2, x_1), f(x_2, x_2)\} - \frac{\gamma}{8}d^2(x_1, x_2).$$

Thus, with (A0), we have $f(x_2, \frac{1}{2}x_2 \oplus \frac{1}{2}x_1) < 0$, contradicting that x_2 is an equilibrium point. \square

In the context of \mathbb{R}^d , Iusem and Lara in [37] further prove that there always exists an equilibrium point whenever the assumptions (A0), (A1), (A2) and (A4) hold (cf. [37, Proposition 3.1]). The proof given in [37] relies on local compactness arguments which are not available in general Hadamard spaces (and already not in infinite-dimensional Hilbert spaces). However, by generalizing a weak compactness argument developed in [46] for the case where f is convex in its right argument (extending the approach taken in [35, 41] over Hilbert spaces and in [18] over Hadamard manifolds), we can generalize all of [37, Proposition 3.1] to Hadamard spaces.

For this, we rely on two notions of coercivity. The first is the following usual notion for functions of one argument:

Definition 3.3. Let $o \in X$. A function $h : X \rightarrow \mathbb{R}$ is called *1-supercoercive* if

$$\liminf_{d(o,x) \rightarrow \infty} \frac{h(x)}{d(o,x)} > 0.$$

The second is a notion for bifunctions, originally introduced in a linear setting in [35].

Definition 3.4 (cf. [46, Assumption P5]). Let $o \in X$. A bifunction $f : X \times X \rightarrow \mathbb{R}$ is called *coercive as a bifunction* if for any sequence $\{x_n\}_n \subseteq X$ satisfying $\lim_{n \rightarrow \infty} d(o, x_n) = \infty$, there exists $u \in X$ and $n_0 \in \mathbb{N}$ such that $f(x_n, u) \leq 0$ for all $n \geq n_0$.

Generalizing parts of [54, Theorem 1] to the Hadamard setting, we now first show that any lsc, strongly quasiconvex function is 1-supercoercive.

Lemma 3.5. Let $h : X \rightarrow \mathbb{R}$ be lsc and strongly quasiconvex with modulus γ . Then h is 1-supercoercive.

Proof. Fix $o, y \in X$ and let $\{x_k\}_k \subseteq X$ with $d(o, x_k) \rightarrow \infty$. Define

$$z_k := \frac{d(o, x_k)}{1 + d(o, x_k)} y \oplus \frac{1}{1 + d(o, x_k)} x_k.$$

Notice that $\{z_k\}_k$ is bounded as d is convex. As h is strongly quasiconvex, we deduce (dividing by $d(o, x_k)$) that

$$\frac{h(z_k)}{d(o, x_k)} \leq \max \left\{ \frac{h(y)}{d(o, x_k)}, \frac{h(x_k)}{d(o, x_k)} \right\} - \frac{\gamma}{2} \frac{d^2(y, x_k)}{(1 + d(o, x_k))^2}.$$

We will now show that $\liminf_{k \rightarrow \infty} h(z_k) > -\infty$. Suppose not. Then, there exists a subsequence $\{z_{k_l}\}_l$ of $\{z_k\}_k$ such that $h(z_{k_l}) \rightarrow -\infty$. However, as $\{z_k\}_k$ and thus $\{z_{k_l}\}_l$ are bounded, by Lemma 2.2, there exists a subsequence $\{z_{k_{l_j}}\}_j$ of $\{z_{k_l}\}_l$ with $z_{k_{l_j}} \xrightarrow{\Delta} z$ for some $z \in X$. As further with Lemma 2.5, h is Δ -lsc, we obtain that $\liminf_{j \rightarrow \infty} h(z_{k_{l_j}}) \geq h(z)$ contradicting $h(z_{k_l}) \rightarrow -\infty$. Hence, $\liminf_{k \rightarrow \infty} h(z_k) > -\infty$ after all and thus $\liminf_{k \rightarrow \infty} h(z_k)/d(o, x_k) \geq 0$. Further, $h(y)/d(o, x_k) \rightarrow 0$. As however $\lim_{k \rightarrow \infty} \frac{\gamma}{2} d^2(y, x_k) / (1 + d(o, x_k))^2 = \frac{\gamma}{2}$, it must hold that $\liminf_{k \rightarrow \infty} h(x_k)/d(o, x_k) \geq \frac{\gamma}{2} > 0$. Hence h is 1-supercoercive. \square

From this it follows that, under suitable conditions, f is coercive as a bifunction:

Lemma 3.6. Let $f : X \times X \rightarrow \mathbb{R}$ be a bifunction satisfying (A1), (A2) and (A4). Then f is coercive as a bifunction.

Proof. Fix $x \in X$. From Lemma 3.5, we get that $\liminf_{d(o,y) \rightarrow \infty} f(x,y)/d(o,y) > 0$, using also (A1) and (A4). This in particular implies that for any sequence $\{y_n\}_n \subseteq X$ with $\lim_{n \rightarrow \infty} d(o,y_n) = \infty$, we find an n_0 such that $f(x,y_n) \geq 0$ for all $n \geq n_0$. Using (A2), that is pseudomonotonicity of f , implies the result. \square

Next to the above coercivity properties, the proof that $S(f)$ is nonempty under suitable assumptions on f rests on two lemmas derived from [46]. The first is a nonlinear version of the so-called KKM Lemma [48] (see also [23, Lemma 1]), derived for finite-dimensional Hadamard manifolds in [18] and then for Hadamard spaces in [46].

Lemma 3.7 (cf. [46, Lemma 1.8]). *Let $G : X \rightarrow 2^X$ be such that for each $x \in X$, $G(x)$ is Δ -closed. Suppose that*

- (a) *for all $x_1, \dots, x_m \in X$, $\text{conv}(\{x_1, \dots, x_m\}) \subseteq \bigcup_{i=1}^m G(x_i)$,*
- (b) *there exists $x_0 \in X$ such that $G(x_0)$ is Δ -compact.*

Then $\bigcap_{x \in X} G(x) \neq \emptyset$.

The second lemma states that if it is possible to find an equilibrium point x for f on some closed ball in X and one point y in the interior of that ball such that $f(x,y) \leq 0$, then x is an equilibrium point for f on the whole space X . To formulate it, we consider the following definitions akin to [46]: Fix $o \in X$. For any $n \in \mathbb{N}$, define

$$X_n := \{x \in X \mid d(o,x) \leq n\} \text{ with } X_n^\circ := \{x \in X \mid d(o,x) < n\}.$$

Further, given a bifunction $f : X \times X \rightarrow \mathbb{R}$ and $y \in X$ as well as $n \in \mathbb{N}$, define

$$L_f(n,y) := \{x \in X_n \mid f(y,x) \leq 0\}.$$

We can now adapt [46, Lemma 2.1] to our setting of strongly quasiconvex bifunctions. Here, we follow a modified version of the proof given for a preceding result of this type in linear spaces by Iusem and Sosa (cf. [41, Lemma 3.7]).

Lemma 3.8. *Let f be a bifunction satisfying (A0), (A2) and (A4). If for some $n \in \mathbb{N}$ and some $\bar{x} \in \bigcap_{y \in X_n} L_f(n,y)$ there exists $\bar{y} \in X_n^\circ$ such that $f(\bar{x},\bar{y}) \leq 0$, then $f(\bar{x},y) \geq 0$ for all $y \in X$.*

Proof. Notice that by definition

$$\bigcap_{y \in X_n} L_f(n,y) = \bigcap_{y \in X_n} \{x \in X_n \mid f(y,x) \leq 0\} = \{x \in X_n \mid f(y,x) \leq 0 \text{ for all } y \in X_n\}$$

for any $n \in \mathbb{N}$. Thus, if $\bar{x} \in \bigcap_{y \in X_n} L_f(n,y)$ for some $n \in \mathbb{N}$, we know $f(y,\bar{x}) \leq 0$ for any $y \in X_n$. We now show that $f(\bar{x},y) \geq 0$ for all $y \in X_n$. For that, let $y \in X_n$. If $y = \bar{x}$, the claim is obvious by (A0). So suppose that $y \neq \bar{x}$ and define $w_\lambda := (1-\lambda)\bar{x} \oplus \lambda y$ for $\lambda \in [0,1]$. By (A0) and (A4), we get

$$0 = f(w_\lambda, w_\lambda) \leq \max\{f(w_\lambda, y), f(w_\lambda, \bar{x})\} - \lambda(1-\lambda) \frac{\gamma}{2} d^2(y, \bar{x}).$$

As $w_\lambda \in X_n$ since balls are convex, we get $f(w_\lambda, \bar{x}) \leq 0$ from the above. Since we have $\lambda(1-\lambda) \frac{\gamma}{2} d^2(y, \bar{x}) > 0$ for $\lambda \in (0,1)$, it must hold that $f(w_\lambda, y) > 0$ for all such λ . By the continuity of geodesics, we have $w_\lambda \rightarrow \bar{x}$ for $\lambda \rightarrow 0$ and so we get

$$f(\bar{x}, y) \geq \limsup_{\lambda \rightarrow 0} f(w_\lambda, y) \geq 0$$

since f is usc in its left argument. It is left to show that $f(\bar{x}, y) \geq 0$ for all $y \in X \setminus X_n$. For that, let $y \in X \setminus X_n$ be arbitrary. Then, with $\bar{y} \in X_n^\circ$ and by the continuity of geodesics, there

exists $\lambda \in (0, 1)$ such that $z := (1 - \lambda)y \oplus \lambda\bar{y} \in X_n$. As $z \in X_n$ and by (A4), that is strong quasiconvexity, we get

$$0 \leq f(\bar{x}, z) = f(\bar{x}, (1 - \lambda)y \oplus \lambda\bar{y}) \leq \max\{f(\bar{x}, y), f(\bar{x}, \bar{y})\} - \lambda(1 - \lambda)\frac{\gamma}{2}d^2(y, \bar{y}).$$

As by assumption $f(\bar{x}, \bar{y}) \leq 0$ and as further $\lambda(1 - \lambda)\frac{\gamma}{2}d^2(y, \bar{y}) > 0$, it must hold that $f(\bar{x}, y) > 0$ and we are done. \square

We can now easily transfer [46, Theorem 2.4] to the strongly quasiconvex case, and thereby extend [37, Proposition 3.1] to the setting of Hadamard spaces.

Theorem 3.9. *Let f be a bifunction satisfying (A0), (A1), (A2) and (A4). Then $S(f)$ is nonempty.*

Proof. By [46, Lemma 2.3], f is properly quasimonotone, that is $\min_{x \in A} f(x, y) \leq 0$ for every finite set $A \subseteq X$ and every $y \in \text{conv}(A)$. Further, by Lemma 3.6, f is coercive as a bifunction. The result now follows by essentially exactly the same arguments as in [46, Theorem 2.4] but, for the sake of completeness, we present them here. Let $n \in \mathbb{N}$. We use Lemma 3.7 with X_n instead of X and $G(y) := L_f(n, y)$. To show that we are in its setting, consider $x_0, \dots, x_k \in X_n$ and $\bar{x} \in \text{conv}(\{x_0, \dots, x_k\})$. But then $\bar{x} \in \bigcup_{i=0}^k L_f(n, x_i)$ as $\bar{x} \in X_n$ by convexity of X_n and $\min_{0 \leq i \leq k} f(x_i, \bar{x}) \leq 0$ by proper quasimonotonicity of f . Further, we have that $G(y) = \{x \in X_n \mid f(y, x) \leq 0\}$ is closed as f is lsc and convex as f is (strongly) quasiconvex in its second argument. Therefore, it is Δ -closed by Lemma 2.3. $G(y)$ is also bounded as it is contained in X_n . Thus, by Lemma 2.4, it is Δ -compact for all $y \in X$. We can hence apply Lemma 3.7, deducing that $\bigcap_{y \in X_n} L_f(n, y) \neq \emptyset$ for all $n \in \mathbb{N}$. For each n , we choose $x_n \in \bigcap_{y \in X_n} L_f(n, y)$ and distinguish two cases.

Case 1: There is an $n \in \mathbb{N}$ such that $d(o, x_n) < n$. Then, $x_n \in S(f)$ by Lemma 3.8, and so we are done.

Case 2: $d(o, x_n) = n$ for all $n \in \mathbb{N}$. But then, as f is coercive as a bifunction by Lemma 3.6, there exists $u \in X$ and $n_0 \in \mathbb{N}$ such that $f(x_n, u) \leq 0$ for all $n \geq n_0$. Consider $n' \geq n_0$ such that $d(o, u) < n'$. Then $f(x_{n'}, u) \leq 0$ and $u \in X_{n'}^\circ$. Again we are done by Lemma 3.8. \square

Using the above Theorem 3.9, we can in particular show that a lsc strongly quasiconvex function $h : X \rightarrow \mathbb{R}$ over a Hadamard space has a minimizer, answering a question left open in the work [67] of the second author.

Corollary 3.10. *Let $h : X \rightarrow \mathbb{R}$ be a lsc strongly quasiconvex function. Then $\text{argmin}_{x \in X} h(x)$ is nonempty.*

Proof. As in Example 3.1, we define f_h by $f_h(x, y) := h(y) - h(x)$. This function clearly satisfies (A0) and (A1). Further, if $f_h(x, y) \geq 0$, then $h(y) - h(x) \geq 0$ and so $h(x) - h(y) \leq 0$, that is $f_h(y, x) \leq 0$. Hence f_h is pseudomonotone and so satisfies (A2). As h is strongly quasiconvex and remains strongly quasiconvex under addition of a constant, $f_h(x, \cdot)$ is strongly quasiconvex for every $x \in X$, and so satisfies (A4). We can thus apply Theorem 3.9 and obtain $x^* \in X$ such that $f_h(x^*, y) \geq 0$ for all $y \in X$. This however immediately yields that $x^* \in \text{argmin}_{x \in X} h(x)$. \square

4. A PROXIMAL POINT ALGORITHM

4.1. The method and preliminaries. We now move to the first method for solving the equilibrium problem (EP) over a Hadamard space (X, d) for a given bifunction $f : X \times X \rightarrow \mathbb{R}$ that satisfies assumptions among (A1) – (A6). As before, we write $S(f)$ for the associated solution set.

The method we consider is the following proximal point type method: For a given start point $x_0 \in X$ and a sequence $\{\beta_k\}_k \subseteq (0, \infty)$, recursively define x_{k+1} by choosing

$$(PPA) \quad x_{k+1} \in \operatorname{argmin}_{y \in X} \left\{ f(x_k, y) + \frac{1}{2\beta_k} d^2(x_k, y) \right\}.$$

This method represents a metric variant of one of the proximal point type methods studied by Iusem and Lara (cf. [37, Algorithm 1]) in the context of the above equilibrium problem (EP) over \mathbb{R}^d , corresponding to the second type of method discussed in the introduction which regularizes the bifunction only in its second argument. Concretely, as also highlighted throughout the introduction, the work [37] considers the problem over a closed and convex subset $K \subseteq \mathbb{R}^d$, and for a corresponding bifunction $f : K \times K \rightarrow \mathbb{R}$ satisfying (A1) – (A6), suitably formulated over K with a corresponding solution set $S(K, f)$. We however omit such restrictions here as any closed and convex subset of a Hadamard space (and so in particular of a Hilbert space) is naturally a Hadamard space again, so that considering the problem (EP) over X already offers the full generality here. This also applies to the existence results established previously.

In the setting described above, Iusem and Lara show in [37] that this proximal point type algorithm is well-defined and that it converges under suitable assumptions on the parameters:

Theorem 4.1 (cf. [37, Theorem 3.1]). *Let $X = \mathbb{R}^d$ and $K \subseteq X$ be closed and convex. Let $f : K \times K \rightarrow \mathbb{R}$ be a bifunction satisfying (A1) – (A6) and let $\{\beta_k\}_k \subseteq (\frac{1}{\gamma-8\eta}, \frac{1}{4\eta})$. Let $\{x_k\}_k$ be the sequence generated by (PPA) over K . Then $\{x_k\}_k$ converges to a point $x^* \in S(K, f)$.*

Notice that under (A6), it holds that $\frac{1}{\gamma-8\eta} < \frac{1}{4\eta}$ so that the interval $(\frac{1}{\gamma-8\eta}, \frac{1}{4\eta})$ is non-empty and the parameter restriction $\{\beta_k\}_k \subseteq (\frac{1}{\gamma-8\eta}, \frac{1}{4\eta})$ is satisfiable.¹ In particular, as highlighted in the introduction, we will be able to weaken this restriction to $\{\beta_k\}_k \subseteq [\frac{1}{\gamma-8\eta}, \frac{1}{4\eta}]$ later on.

Given any function $h : X \rightarrow \mathbb{R}$, we define its proximal map $\operatorname{Prox}_{\beta h}$ by

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

for $\beta > 0$ and $x \in X$. Using that definition, as already highlighted in the introduction, the method (PPA) assumes a much more compact form with

$$x_{k+1} \in \operatorname{Prox}_{\beta_k f(x_k, \cdot)}(x_k).$$

We now first show the well-definedness of the method (PPA) above, i.e. that $\operatorname{Prox}_{\beta_k f(x_k, \cdot)}(x_k) \neq \emptyset$ for any $k \in \mathbb{N}$, provided that $\beta_k \leq \frac{1}{4\eta}$. The first crucial result in this direction is the following:

Lemma 4.2 (cf. [67, Lemma 3.4]). *If $h : X \rightarrow \mathbb{R}$ is bounded from below and 1-superc coercive, then any sequence $\{y_k\}_k$ with $h(y_k) \rightarrow \inf_{x \in X} h(x)$ is bounded.*

The above result can be employed to yield the following:

Lemma 4.3 (cf. [67, Lemma 3.8]). *Let $h : X \rightarrow \mathbb{R}$ be a quasiconvex, lsc function which is bounded from below. Then, for any $x \in X$ and $\beta > 0$, $\operatorname{Prox}_{\beta h}(x)$ is nonempty.*

This can now be extended to yield the following, guaranteeing an analogous result to Lemma 4.3 above for suitable bifunctions and parameters, provided $S(f)$ is nonempty.

¹A slight weakening of assumption (A6) is discussed in [30, Remark 87] (in the context of a variant of the above method including inertia terms and over-relaxations as developed in [28, 29]), with a corresponding refined parameter restriction on $\{\beta_k\}_k$. While we deem it quite likely that our present results can be extended to this weakened assumption, we here focus on (A6) for simplicity.

Proposition 4.4. *Let $f : X \times X \rightarrow \mathbb{R}$ be a function which is quasiconvex and lsc in its right argument and further satisfies (A5). Assume that there exist $\bar{x} \in X$ and $t \in \mathbb{R}$ with $f(\bar{x}, y) \geq t$ for all $y \in X$. Then, for any $z \in X$ and $\beta \leq \frac{1}{4\eta}$, $\text{Prox}_{\beta f(z, \cdot)}(z)$ is nonempty.*

Proof. Fix $z \in X$ and let \bar{x}, t be as above. Then it follows from (A5) that

$$f(\bar{x}, y) - f(\bar{x}, z) - \eta(d^2(\bar{x}, z) + d^2(z, y)) \leq f(z, y) \text{ for all } y \in X.$$

As $f(\bar{x}, y) \geq t$, we can set $G := -t + f(\bar{x}, z) + \eta d^2(\bar{x}, z)$, obtaining

$$(+) \quad -G - \eta d^2(z, y) \leq f(z, y) \text{ for all } y \in X.$$

We define $h_z(y) := f(z, y) + \frac{1}{2\beta} d^2(z, y)$ and show that h_z is bounded from below. For this, notice that (+) implies

$$h_z(y) = f(z, y) + \frac{1}{2\beta} d^2(z, y) \geq \left(\frac{1}{2\beta} - \eta \right) d^2(z, y) - G.$$

But now as $\beta \leq \frac{1}{4\eta}$, we obtain that $\frac{1}{2\beta} - \eta \geq 2\eta - \eta = \eta > 0$. Thus we have $h_z(y) \geq \eta d^2(z, y) - G \geq -G$, and so h_z is bounded below. Take $\{y_k\}_k$ such that $h_z(y_k) \rightarrow \inf_{x \in X} h_z(x)$. As further

$$\liminf_{d(z, y) \rightarrow \infty} \frac{h_z(y)}{d(z, y)} \geq \liminf_{d(z, y) \rightarrow \infty} \left(\frac{\eta d^2(z, y)}{d(z, y)} - \frac{G}{d(z, y)} \right) \geq \liminf_{d(z, y) \rightarrow \infty} \frac{\eta d^2(z, y)}{d(z, y)} > 0,$$

Lemma 4.2 yields that $\{y_k\}_k$ is bounded. By Lemma 2.2, it has a Δ -cluster point, i.e. there exists a subsequence $\{y_{k_l}\}_l$ of $\{y_k\}_k$ with $y_{k_l} \rightarrow^\Delta y$. As $f(z, \cdot)$ is quasiconvex and lsc, by Lemma 2.5, $f(z, \cdot)$ is Δ -lsc. Further, as (CN⁺) implies that $y \mapsto \frac{1}{2\beta} d^2(z, y)$ is convex and as it is also lsc, Lemma 2.5 also yields that it is Δ -lsc. Hence, h_z is Δ -lsc and

$$h_z(y) \leq \liminf_{l \rightarrow \infty} h_z(y_{k_l}) = \inf_{x \in X} h_z(x).$$

Thus, y is a minimizer of h_z and hence $y \in \text{Prox}_{\beta f(z, \cdot)}(z)$. \square

With Proposition 4.4, it follows in particular that whenever (A1), (A4) and (A5) hold and whenever $S(f) \neq \emptyset$ and $\beta_k \leq \frac{1}{4\eta}$ for all $k \in \mathbb{N}$, then also $\text{Prox}_{\beta_k f(x_k, \cdot)}(x_k) \neq \emptyset$ for any $k \in \mathbb{N}$. In particular, (PPA) is well-defined under these conditions and the parameter restriction $\{\beta_k\}_k \subseteq \left[\frac{1}{\gamma - 8\eta}, \frac{1}{4\eta} \right]$.

The most essential result for strongly quasiconvex functions and their proximal maps is the following inequality, which is an extension of [54, Proposition 7] to the metric context of Hadamard spaces.

Lemma 4.5 (cf. [67, Lemma 3.9]). *Let $h : X \rightarrow \mathbb{R}$ be a strongly quasiconvex function with modulus $\gamma > 0$ and let $\beta > 0$ and $x \in X$. If $\bar{x} \in \text{Prox}_{\beta h}(x)$, then*

$$h(\bar{x}) \leq \max\{h(y), h(\bar{x})\} + \frac{\lambda}{2} \left(\frac{\lambda}{\beta} - \gamma + \lambda\gamma \right) d^2(y, \bar{x}) + \frac{\lambda}{\beta} \langle \vec{x\bar{x}}, \vec{x\bar{y}} \rangle$$

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

Akin to [37], we can also show that fixed points of this mixed proximal point map are exactly the solutions of the equilibrium problem.

Proposition 4.6. *Let $f : X \times X \rightarrow \mathbb{R}$ be a bifunction satisfying (A0) and (A4). Then*

$$S(f) = \text{Fix} \left(\underset{y \in X}{\text{argmin}} \left\{ f(\cdot, y) + \frac{1}{2\beta} d^2(\cdot, y) \right\} \right)$$

for any $\beta > 0$.

Proof. Let \bar{x} be a fixed point of $\operatorname{argmin}_{y \in X} \{f(\cdot, y) + \frac{1}{2\beta} d^2(\cdot, y)\}$, i.e. let $\bar{x} \in \operatorname{Prox}_{\beta f(\bar{x}, \cdot)}(\bar{x})$. By (A4), we get that $x \mapsto f(\bar{x}, x)$ is strongly quasiconvex with modulus γ . Thus, invoking [67, Proposition 3.10], we derive that $\bar{x} \in \operatorname{argmin}_{y \in X} f(\bar{x}, y)$. But then \bar{x} is an equilibrium point for f , as $f(\bar{x}, y) \geq f(\bar{x}, \bar{x})$ for all $y \in X$ and $f(\bar{x}, \bar{x}) = 0$ by (A0).

For the other direction, suppose that $\bar{x} \in S(f)$. Then $\bar{x} \in \operatorname{argmin}_{y \in X} f(\bar{x}, y)$ as $f(\bar{x}, y) \geq 0$ for all $y \in X$ and $f(\bar{x}, \bar{x}) = 0$ by (A0). But then again, by [67, Proposition 3.10], we get that $\bar{x} \in \operatorname{Prox}_{\beta f(\bar{x}, \cdot)}(\bar{x})$. \square

In particular, also similar to [37], we want to remark that this provides a convenient stopping criterion as whenever $x_{k+1} = x_k$, one already knows that $\{x_k\} = S(f)$.

The remainder of this section will now be concerned with establishing a variant of Theorem 4.1 in the context of Hadamard spaces, and in particular to endow it with effective information.

4.2. Quantitative convergence results. We begin by showing a result that will yield Fejér monotonicity of $\{x_k\}_k$ with respect to $S(f)$. For this, we adapt [37, Proposition 3.4] and its proof to the nonlinear setting. In particular, compared to this result, we highlight the more general inequalities $(*)_1$ and $(*)_2$ below. Only their instantiations with $\lambda = 1/2$ are considered explicitly in [37, Proposition 3.4], but we will later require varying λ 's.

Lemma 4.7. *Let f be a bifunction satisfying (A2), (A4) and (A5). Let $\{\beta_k\}_k \subseteq (0, \infty)$. Suppose that (PPA) is well-defined and let $\{x_k\}_k$ be its generated sequence. Let $x^* \in S(f)$. Then, for every $k \in \mathbb{N}$, at least one of the following inequalities holds:*

$$(*)_1 \quad \frac{\lambda(1-\lambda)(1+\gamma\beta_k)}{2\beta_k} d^2(x_{k+1}, x^*) \leq \frac{\lambda}{2\beta_k} d^2(x_k, x^*) - \frac{\lambda}{2\beta_k} d^2(x_{k+1}, x_k)$$

for all $\lambda \in [0, 1]$, or

$$(*)_2 \quad \left(\frac{\lambda(1-\lambda)(1+\gamma\beta_k)}{2\beta_k} - \eta \right) d^2(x_{k+1}, x^*) \leq \frac{\lambda}{2\beta_k} d^2(x_k, x^*) - \left(\frac{\lambda}{2\beta_k} - \eta \right) d^2(x_{k+1}, x_k)$$

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

Proof. Fix $k \in \mathbb{N}$. Using that $x_{k+1} \in \operatorname{Prox}_{\beta_k f(x_k, \cdot)}(x_k)$, applying Lemma 4.5 to the function $f(x_k, \cdot)$ with $y = x^*$ yields

$$\begin{aligned} & f(x_k, x_{k+1}) - \max\{f(x_k, x_{k+1}), f(x_k, x^*)\} \\ & \leq \frac{\lambda}{2} \left(\frac{\lambda}{\beta_k} - \gamma + \lambda\gamma \right) d^2(x_{k+1}, x^*) + \frac{\lambda}{\beta_k} \left\langle \overrightarrow{x_k x_{k+1}}, \overrightarrow{x_{k+1} x^*} \right\rangle \end{aligned}$$

for all $\lambda \in [0, 1]$. We consider two cases:

Case 1: $f(x_k, x_{k+1}) \geq f(x_k, x^*)$. Then we have

$$0 \leq \frac{\lambda}{2} \left(\frac{\lambda}{\beta_k} - \gamma + \lambda\gamma \right) d^2(x_{k+1}, x^*) + \frac{\lambda}{\beta_k} \left\langle \overrightarrow{x_k x_{k+1}}, \overrightarrow{x_{k+1} x^*} \right\rangle$$

for all $\lambda \in [0, 1]$. Using the definition of the quasi-inner product, we obtain

$$0 \leq \frac{\lambda}{2} \left(\frac{\lambda}{\beta_k} - \gamma + \lambda\gamma \right) d^2(x_{k+1}, x^*) + \frac{\lambda}{2\beta_k} (d^2(x_k, x^*) - d^2(x_{k+1}, x_k) - d^2(x_{k+1}, x^*))$$

and thus

$$\frac{\lambda(1-\lambda)(1+\gamma\beta_k)}{2\beta_k} d^2(x_{k+1}, x^*) \leq \frac{\lambda}{2\beta_k} d^2(x_k, x^*) - \frac{\lambda}{2\beta_k} d^2(x_{k+1}, x_k)$$

for all $\lambda \in [0, 1]$. This is $(*)_1$.

Case 2: $f(x_k, x_{k+1}) < f(x_k, x^*)$. Then

$$\begin{aligned} 0 &\leq \frac{\lambda}{2} \left(\frac{\lambda}{\beta_k} - \gamma + \lambda\gamma \right) d^2(x_{k+1}, x^*) + \frac{\lambda}{\beta_k} \left\langle \overrightarrow{x_k x_{k+1}}, \overrightarrow{x_{k+1} x^*} \right\rangle + f(x_k, x^*) - f(x_k, x_{k+1}) \\ &\leq \frac{\lambda}{2} \left(\frac{\lambda}{\beta_k} - \gamma + \lambda\gamma \right) d^2(x_{k+1}, x^*) + \frac{\lambda}{\beta_k} \left\langle \overrightarrow{x_k x_{k+1}}, \overrightarrow{x_{k+1} x^*} \right\rangle + f(x_{k+1}, x^*) \\ &\quad + \eta \left(d^2(x_{k+1}, x_k) + d^2(x_{k+1}, x^*) \right) \end{aligned}$$

for all $\lambda \in [0, 1]$, where the second inequality follows from the fact that we get $f(x_k, x^*) - f(x_k, x_{k+1}) - f(x_{k+1}, x^*) \leq \eta(d^2(x_{k+1}, x_k) + d^2(x_{k+1}, x^*))$ with (A5). Further, as x^* is an equilibrium point and f is pseudomonotone, we get $f(x_{k+1}, x^*) \leq 0$ and thus

$$\begin{aligned} 0 &\leq \frac{\lambda}{2} \left(\frac{\lambda}{\beta_k} - \gamma + \lambda\gamma \right) d^2(x_{k+1}, x^*) + \frac{\lambda}{\beta_k} \left\langle \overrightarrow{x_k x_{k+1}}, \overrightarrow{x_{k+1} x^*} \right\rangle \\ &\quad + \eta \left(d^2(x_{k+1}, x_k) + d^2(x_{k+1}, x^*) \right) \end{aligned}$$

for all $\lambda \in [0, 1]$. Using the definition of the quasi-inner product, similar to Case 1, we obtain

$$\left(\frac{\lambda(1-\lambda)(1+\gamma\beta_k)}{2\beta_k} - \eta \right) d^2(x_{k+1}, x^*) \leq \frac{\lambda}{2\beta_k} d^2(x_k, x^*) - \left(\frac{\lambda}{2\beta_k} - \eta \right) d^2(x_{k+1}, x_k)$$

for all $\lambda \in [0, 1]$. This is $(*)_2$. \square

Using Lemma 4.7, we obtain the following:

Lemma 4.8. *Let f be a bifunction satisfying (A2), (A4) and (A5). Let $\{\beta_k\}_k \subseteq (0, \infty)$. Suppose that (PPA) is well-defined and let $\{x_k\}_k$ be its generated sequence. Let $x^* \in S(f)$. Then, for every $k \in \mathbb{N}$ and $\lambda \in [0, 1]$:*

$$(\circ)_1 \quad \left(\frac{\lambda(1-\lambda)(1+\gamma\beta_k)}{2\beta_k} - \eta \right) d^2(x_{k+1}, x^*) \leq \frac{\lambda}{2\beta_k} d^2(x_k, x^*) - \left(\frac{\lambda}{2\beta_k} - \eta \right) d^2(x_{k+1}, x_k).$$

In particular (setting $\lambda = \frac{1}{2}$), we have

$$(\circ)_2 \quad \left(\frac{1+\gamma\beta_k}{8\beta_k} - \eta \right) d^2(x_{k+1}, x^*) \leq \frac{1}{4\beta_k} d^2(x_k, x^*) - \left(\frac{1}{4\beta_k} - \eta \right) d^2(x_{k+1}, x_k).$$

Proof. With Lemma 4.7, we obtain that for all $k \in \mathbb{N}$, inequality $(*)_1$ or inequality $(*)_2$ holds. The result then follows from the fact that $(*)_1$ actually implies $(*)_2$ as

$$\left(\frac{\lambda(1-\lambda)(1+\gamma\beta_k)}{2\beta_k} - \eta \right) < \left(\frac{\lambda(1-\lambda)(1+\gamma\beta_k)}{2\beta_k} \right) \quad \text{and} \quad \left(\frac{\lambda}{2\beta_k} - \eta \right) < \left(\frac{\lambda}{2\beta_k} \right)$$

for all $\lambda \in [0, 1]$ and $k \in \mathbb{N}$. \square

From Lemma 4.8, we now immediately obtain the Fejér monotonicity of $\{x_k\}_k$ w.r.t. $S(f)$. Over Euclidean spaces, this property is established in [37, Proposition 3.5] under the condition $\{\beta_k\}_k \subseteq (\frac{1}{\gamma-8\eta}, \frac{1}{4\eta})$. However it is easy to see that the same holds true already for $\{\beta_k\}_k \subseteq [\frac{1}{\gamma-8\eta}, \frac{1}{4\eta}]$.

Lemma 4.9. *Let f be a bifunction satisfying (Ai) with $i = 2, 4, 5, 6$. Let $\{\beta_k\}_k \subseteq [\frac{1}{\gamma-8\eta}, \frac{1}{4\eta}]$. Suppose that (PPA) is well-defined and let $\{x_k\}_k$ be its generated sequence. Let $x^* \in S(f)$. Then, for every $k \in \mathbb{N}$:*

$$d^2(x_{k+1}, x^*) \leq d^2(x_k, x^*).$$

Proof. Let $x^* \in S(f)$ and fix $k \in \mathbb{N}$. Using Lemma 4.8, we know that $(\circ)_2$ holds. Using $\beta_k \leq \frac{1}{4\eta}$, we have

$$\left(\frac{1 + \gamma\beta_k}{8\beta_k} - \eta\right) d^2(x_{k+1}, x^*) \leq \frac{1}{4\beta_k} d^2(x_k, x^*) - \left(\frac{1}{4\beta_k} - \eta\right) d^2(x_{k+1}, x_k) \leq \frac{1}{4\beta_k} d^2(x_k, x^*).$$

With $\beta_k \geq \frac{1}{\gamma - 8\eta}$, it follows that

$$\left(\frac{1 + \gamma\beta_k}{8\beta_k} - \eta\right) d^2(x_{k+1}, x^*) \geq \left(\frac{1 + \gamma\beta_k}{8\beta_k} - \frac{\gamma\beta_k - 1}{8\beta_k}\right) d^2(x_{k+1}, x^*) = \frac{1}{4\beta_k} d^2(x_{k+1}, x^*).$$

Multiplying with $4\beta_k$ yields the claim. \square

Key for our approach towards a quantitative convergence theorem for (PPA) is an effective estimate for the asymptotic behavior of $d(x_k, x_{k+1})$. This will here take the form of an explicit bound $\alpha : (0, \infty) \rightarrow \mathbb{N}$ such that for any $\varepsilon > 0$, there exists a $k \leq \alpha(\varepsilon)$ with $d(x_k, x_{k+1}) < \varepsilon$. Essential for deriving such an α is $(\circ)_1$ from Lemma 4.8, by which we obtain an inequality of the form

$$e_k d^2(x_{k+1}, x_k) \leq c_k d^2(x_k, x^*) - d_k d^2(x_{k+1}, x^*)$$

where

$$e_k := \frac{\lambda_k}{2\beta_k} - \eta, \quad c_k := \frac{\lambda_k}{2\beta_k}, \quad d_k := \frac{\lambda_k(1 - \lambda_k)(1 + \gamma\beta_k)}{2\beta_k} - \eta$$

for any choice of $\lambda_k \in [0, 1]$. In order to derive an effective asymptotic estimate on $d(x_k, x_{k+1})$ from this we, on the one hand, need to argue that $c_k d^2(x_k, x^*) - d_k d^2(x_{k+1}, x^*)$ gets arbitrarily small. Intuitively, we thus, e.g., want that $c_k \leq d_k$ and if $\{\beta_k\}_k \subseteq (\frac{1}{\gamma - 8\eta}, \frac{1}{4\eta})$, the *only* constant choice of λ_k that ensures this is $\lambda_k = \frac{1}{2}$ (as in [37]). However, to derive an estimate on $d^2(x_{k+1}, x_k)$ from this, we further, on the other hand, need that e_k can be bounded away from zero for all $k \in \mathbb{N}$, i.e. $(\frac{\lambda}{2\beta_k} - \eta) > a$ for some $a > 0$ and all $k \in \mathbb{N}$. However, having set $\lambda_k = \frac{1}{2}$, this is impossible as $(\frac{1}{4\beta_k} - \eta)$ goes to 0 if for example $\beta_k \rightarrow \frac{1}{4\eta}$, as is permitted in the condition $\{\beta_k\}_k \subseteq (\frac{1}{\gamma - 8\eta}, \frac{1}{4\eta})$.

We will overcome this issue by showing that if λ_k is chosen adaptively with k (or rather, the values of β_k), we can ensure both properties, that is both $c_k - d_k \leq 0$ and the existence of a uniform positive lower bound on e_k , at the same time. This not only enables our quantitative results, but also allows us to strengthen the original convergence result of Iusem and Lara (recall Theorem 4.1 above) to the assumption that $\beta_k \in [\frac{1}{\gamma - 8\eta}, \frac{1}{4\eta}]$ for all $k \in \mathbb{N}$.²

Before we give these quantitative results in the following part of this section, we need to develop some further machinery. Once this is established, the quantitative results for the asymptotic behavior of $d(x_k, x_{k+1})$ will follow from the following abstract result derived from [67, Lemma 4.2]:

Lemma 4.10. *Let $\{a_k\}_k$ be a nonincreasing sequence of nonnegative reals and let $b \geq a_0$. Let further $\{c_k\}_k$ and $\{d_k\}_k$ be sequences of nonnegative reals with $c_k \leq d_k$ for all $k \in \mathbb{N}$ and such that there exists $u \in \mathbb{R}$ with $c_k \leq u$ for all $k \in \mathbb{N}$. Then, for any $\varepsilon > 0$:*

$$\exists k \leq \left\lceil \frac{bu}{\varepsilon} \right\rceil (c_k a_k - d_k a_{k+1} < \varepsilon).$$

²We further refer to the extensive discussions in the master thesis of the first author [21] for why this extension is also crucial from the logical perspective of the proof mining program, through which the present results have been obtained, to allow for the extraction of rates as in Theorem 4.14 later on.

Proof. For any $k \in \mathbb{N}$, we have

$$c_k a_k - d_k a_{k+1} \leq c_k (a_k - a_{k+1}) \leq u(a_k - a_{k+1}).$$

The result now follows from [67, Lemma 4.2] applied to $\{ua_k\}_k$. \square

For the rest of this section, we now suppose that $S(f) \neq \emptyset$ and denote the unique element of $S(f)$ by x^* . Similarly, we will presume that, given a suitable sequence $\{\beta_k\}_k$, (PPA) is well-defined and generates a sequence $\{x_k\}_k$.

By Theorem 3.9, the former is guaranteed under the conditions (A0), (A1), (A2) and (A4) and by Proposition 4.4, the latter is guaranteed under the conditions (A1), (A4) and (A5), and whenever $S(f) \neq \emptyset$. However, as we will see, once these two properties are assumed, the remainder of the convergence proof does not rely on (A1) (see in particular Remark 4.15 later on).

As outlined before, key for establishing this result over the broader set of parameters $[\frac{1}{\gamma-8\eta}, \frac{1}{4\eta}]$ is to choose λ in $(o)_1$ of Lemma 4.8 adaptively as λ_k , depending on the values of β_k . The key function that later allows us to compute λ_k from β_k is now the following: For $x \geq \frac{1}{\gamma-8\eta}$, set

$$(†) \quad N(x) := \frac{\gamma x}{2(1+\gamma x)} + \frac{\sqrt{\gamma^2 x^2 - 8x\eta(1+\gamma x)}}{2(1+\gamma x)},$$

where γ and η are as in (A4) – (A6). Indeed, notice that N is well-defined on $[\frac{1}{\gamma-8\eta}, \infty)$, as for $x \geq \frac{1}{\gamma-8\eta}$ the value $\gamma^2 x^2 - 8x\eta(1+\gamma x)$ is positive. The key properties of N are now as follows:

Lemma 4.11. *Consider N as defined in (†). Then, we have the following:*

- (i) $N(\frac{1}{\gamma-8\eta}) = \frac{1}{2}$,
- (ii) $N(\frac{1}{4\eta}) > \frac{4\gamma}{3(8\eta+2\gamma)} > \frac{1}{2}$,
- (iii) $N : [\frac{1}{\gamma-8\eta}, \infty) \rightarrow [0, 1]$ is increasing,
- (iv) $\frac{N(x)}{2x}$ is decreasing on $[\frac{1}{\gamma-8\eta}, \frac{1}{4\eta}]$.

Proof. (i) Notice that $\sqrt{\gamma^2 x^2 - 8x\eta(1+\gamma x)} = x \cdot \sqrt{\gamma^2 - 8\eta/x - 8\gamma\eta}$. Hence

$$N\left(\frac{1}{\gamma-8\eta}\right) = \frac{\gamma}{2(\gamma-8\eta)+2\gamma} + \frac{\sqrt{\gamma^2 - 8\eta(\gamma-8\eta) - 8\gamma\eta}}{2(\gamma-8\eta)+2\gamma} = \frac{\gamma + \sqrt{(\gamma-8\eta)^2}}{4\gamma-16\eta} = \frac{1}{2}.$$

(ii) Using (A6), we obtain

$$\begin{aligned} N\left(\frac{1}{4\eta}\right) &= \frac{\gamma}{8\eta+2\gamma} + \frac{\sqrt{\gamma^2 - 32\eta^2 - 8\eta\gamma}}{8\eta+2\gamma} \stackrel{12\eta < \gamma}{>} \frac{\gamma}{8\eta+2\gamma} + \frac{\sqrt{\gamma^2 - \frac{2}{9}\gamma^2 - \frac{6}{9}\gamma^2}}{8\eta+2\gamma} \\ &= \frac{\gamma + \frac{1}{3}\gamma}{8\eta+2\gamma} \stackrel{12\eta < \gamma}{>} \frac{\gamma + 4\eta}{8\eta+2\gamma} = \frac{1}{2}. \end{aligned}$$

(iii) Notice that

$$N(x) \leq \frac{\gamma x}{2(1+\gamma x)} + \frac{\sqrt{\gamma^2 x^2}}{2(1+\gamma x)} = \frac{2\gamma x}{2+2\gamma x} < 1$$

for all $x \geq 1/(\gamma-8\eta)$. Together with (i), this implies that $N : [\frac{1}{\gamma-8\eta}, \infty) \rightarrow [0, 1]$. Notice further that

$$(–) \quad N(x) = \frac{\gamma}{\frac{2}{x} + 2\gamma} + \frac{\sqrt{\gamma^2 - \frac{8\eta}{x} - 8\gamma\eta}}{\frac{2}{x} + 2\gamma}.$$

Form this, it is clear that $N : [\frac{1}{\gamma-8\eta}, \infty) \rightarrow [0, 1]$ is increasing.

(iv) We show that the derivative of $\frac{N(x)}{2x}$ is negative for $x \in [\frac{1}{\gamma-8\eta}, \frac{1}{4\eta}]$. With $(-)$, we obtain

$$\left(\frac{N(x)}{2x}\right)'(x) = -\frac{\gamma^2}{4(1+\gamma x)^2} + \frac{4\eta}{\sqrt{\gamma(\gamma-8\eta) - \frac{8\eta}{x}} \cdot 4(1+\gamma x)x^2} - \frac{\sqrt{\gamma(\gamma-8\eta) - \frac{8\eta}{x}} \cdot \gamma}{4(1+\gamma x)^2}.$$

As $x \geq \frac{1}{\gamma-8\eta}$, we know $\sqrt{\gamma(\gamma-8\eta) - \frac{8\eta}{x}} \geq (\gamma-8\eta)$ and so get

$$\left(\frac{N(x)}{2x}\right)'(x) \leq -\frac{\gamma^2}{4(1+\gamma x)^2} + \frac{4\eta(\gamma-8\eta)(1+\gamma x)}{4(1+\gamma x)^2} - \frac{(\gamma-8\eta)\gamma}{4(1+\gamma x)^2}.$$

With $x \leq \frac{1}{4\eta}$, this yields

$$\left(\frac{N(x)}{2x}\right)'(x) \leq -\frac{\gamma^2}{4(1+\gamma x)^2} + \frac{4\eta(\gamma-8\eta)(1+\frac{\gamma}{4\eta})}{4(1+\gamma x)^2} - \frac{(\gamma-8\eta)\gamma}{4(1+\gamma x)^2} = \frac{-\gamma^2 + 4\eta(\gamma-8\eta)}{4(1+\gamma x)^2}.$$

Finally, using $12\eta < \gamma$, we obtain

$$\left(\frac{N(x)}{2x}\right)'(x) \leq \frac{-2/3\gamma^2 - 32\eta^2}{4(1+\gamma x)^2}$$

which is (using that $x \geq 0$) negative by (A6). □

Using N , we can now give an explicit definition for the adaptive choice of λ_k . The following lemma provides this definition, together with the essential properties of the corresponding quantities e_k , c_k and d_k as derived from Lemma 4.11.

Lemma 4.12. *Suppose that (A6) holds. For any $k \in \mathbb{N}$, let $\beta_k \in [\frac{1}{\gamma-8\eta}, \frac{1}{4\eta}]$ and consider*

$$e_k := \frac{\lambda_k}{2\beta_k} - \eta, \quad c_k := \frac{\lambda_k}{2\beta_k}, \quad d_k := \frac{\lambda_k(1-\lambda_k)(1+\gamma\beta_k)}{2\beta_k} - \eta$$

for

$$\lambda_k := \min\left\{N(\beta_k), \frac{1}{2} + \frac{1}{\sqrt{8}}\right\},$$

where N is defined as in (\dagger) . Then, for any $k \in \mathbb{N}$:

- (i) $c_k - d_k \leq 0$,
- (ii) $\lambda_k \in [\frac{1}{2}, \frac{1}{2} + \frac{1}{\sqrt{8}}]$,
- (iii) $e_k \geq (2\gamma\eta - 24\eta^2)/(3(8\eta + 2\gamma)) > 0$,
- (iv) $c_k \leq (\gamma - 8\eta)/4$.

Proof. (i) We have

$$c_k - d_k = \frac{\lambda_k - \lambda_k(1-\lambda_k)(1+\gamma\beta_k)}{2\beta_k} + \eta$$

and argue that $c_k - d_k = 0$ if $\lambda_k = N(\beta_k)$. To see this, notice that

$$c_k - d_k = 0 \Leftrightarrow \lambda_k^2 - \lambda_k \frac{\gamma\beta_k}{(1+\gamma\beta_k)} + \frac{2\beta_k\eta}{(1+\gamma\beta_k)} = 0.$$

By definition of N , the right hand side now holds by the quadratic formula. Hence, the claim holds if $\lambda_k = N(\beta_k)$. It is left to show the result for cases where $\lambda_k = \frac{1}{2} + \frac{1}{\sqrt{8}} < N(\beta_k)$. For this, notice that if we regard $c_k - d_k$ as a function in (the values of) λ_k ,

it is increasing on the interval $[\frac{1}{2}, 1]$. Thus, $c_k - d_k \leq 0$ if $\lambda_k = \frac{1}{2} + \frac{1}{\sqrt{8}} < N(\beta_k)$, as $N(\beta_k) < 1$ by Lemma 4.11, (iii).

- (ii) From Lemma 4.11, (i), we know that $N(\frac{1}{\gamma-8\eta}) = \frac{1}{2}$. As by Lemma 4.11, (iii), we know further that N is increasing, the result follows from the definition of λ_k .
- (iii) Suppose $\lambda_k = N(\beta_k)$. As $\frac{N(x)}{2x}$ decreases on the interval $[\frac{1}{\gamma-8\eta}, \frac{1}{4\eta}]$ by Lemma 4.11, (iv), together with Lemma 4.11, (ii) we obtain

$$\begin{aligned} e_k &= \frac{N(\beta_k)}{2\beta_k} - \eta \geq \frac{N\left(\frac{1}{4\eta}\right)}{2} 4\eta - \eta \\ &> \frac{8\gamma\eta}{3(8\eta + 2\gamma)} - \eta = \frac{8\gamma\eta - 6\gamma\eta - 24\eta^2}{3(8\eta + 2\gamma)} = \frac{2\gamma\eta - 24\eta^2}{3(8\eta + 2\gamma)}. \end{aligned}$$

It follows that $e_k > 0$ as $12\eta < \gamma$ by (A6). Suppose that $\lambda_k = \frac{1}{2} + \frac{1}{\sqrt{8}}$. But then, we obtain

$$e_k = \frac{\lambda_k}{2\beta_k} - \eta \geq \frac{1}{2} \left(\frac{1}{2} + \frac{1}{\sqrt{8}} \right) 4\eta - \eta = \frac{2}{\sqrt{8}}\eta > \frac{2}{3}\eta > \frac{2\gamma\eta - 24\eta^2}{3(8\eta + 2\gamma)}.$$

As before, we have $e_k > 0$ since $12\eta < \gamma$ by (A6). This implies the result.

- (iv) With Lemma 4.11, (iv) and Lemma 4.11, (i) we obtain

$$c_k = \frac{\lambda_k}{2\beta_k} \leq \frac{N(\beta_k)}{2\beta_k} \leq \frac{1}{2} N\left(\frac{1}{\gamma-8\eta}\right) (\gamma - 8\eta) = \frac{\gamma - 8\eta}{4}. \quad \square$$

Our main quantitative result for the asymptotic behavior of $d(x_k, x_{k+1})$ is now the following:

Lemma 4.13. *Let f be a bifunction satisfying (Ai) with $i = 2, 4, 5, 6$. Let $\{\beta_k\}_k \subseteq [\frac{1}{\gamma-8\eta}, \frac{1}{4\eta}]$. Suppose that (PPA) is well-defined and let $\{x_k\}_k$ be its generated sequence. Let $x^* \in S(f)$. Then, for any $\varepsilon > 0$:*

$$\exists k \leq \left\lceil \frac{3b^2(\gamma - 8\eta)(2\gamma + 8\eta)}{8\eta(\gamma - 12\eta)\varepsilon} \right\rceil (d^2(x_k, x_{k+1}) < \varepsilon),$$

where $b^2 \geq d^2(x_0, x^*)$.

Proof. With Lemma 4.8, we know that $(\circ)_1$ holds and in particular obtain

$$e_k d^2(x_{k+1}, x_k) \leq c_k d^2(x_k, x^*) - d_k d^2(x_{k+1}, x^*)$$

for all $k \in \mathbb{N}$, where c_k, e_k and d_k are defined as in Lemma 4.12. By Lemma 4.12, (i) and (ii), it holds that $0 < c_k \leq d_k$ for all $k \in \mathbb{N}$. Further, Lemma 4.12, (iv) implies that $c_k \leq \frac{\gamma-8\eta}{4}$. We can thus apply Lemma 4.10 to obtain that for any $\varepsilon > 0$,

$$\exists k \leq \left\lceil \frac{b^2(\gamma - 8\eta)}{4\varepsilon} \right\rceil (e_k d^2(x_{k+1}, x_k) < \varepsilon).$$

With Lemma 4.12, (iii), we now obtain

$$\exists k \leq \left\lceil \frac{b^2(\gamma - 8\eta)}{4\varepsilon} \right\rceil \left(\frac{2\gamma\eta - 24\eta^2}{3(8\eta + 2\gamma)} d^2(x_{k+1}, x_k) < \varepsilon \right)$$

for any $\varepsilon > 0$. This implies the result. \square

With this, we obtain a quantitative convergence theorem that generalizes the assumptions of Theorem 4.1. In particular, the quantitative result we provide comes in the form of a sublinear non-asymptotic guarantee which only depends on the parameters γ, η as in (A4) – (A6), as well as a bound $b > 0$ on the distance of the starting point of (PPA) to the equilibrium point x^* .

Theorem 4.14. *Let X be a Hadamard space and let $f : X \times X \rightarrow \mathbb{R}$ be a bifunction satisfying (Ai) with $i = 2, 4, 5, 6$. Let $\{\beta_k\}_k \subseteq [\frac{1}{\gamma-8\eta}, \frac{1}{4\eta}]$. Suppose that (PPA) is well-defined and let $\{x_k\}_k$ be its generated sequence. Suppose that $S(f) \neq \emptyset$. Then $\{x_k\}_k$ converges to the (unique) equilibrium point x^* of f . Further, we have the non-asymptotic guarantee*

$$\forall k > 2 \left(d(x_k, x^*) < \frac{C}{\sqrt{k-2}} \right)$$

where

$$C := \frac{12\sqrt{3}b\gamma^2}{\sqrt{\eta}(\gamma-12\eta)^{3/2}}$$

with $b^2 \geq d^2(x_0, x^*)$.

Proof. It is enough to show the quantitative result. To that end, we first show

$$(Q) \quad \forall \varepsilon > 0 \quad \forall k \geq \left(\left\lceil \frac{C^2}{\varepsilon^2} \right\rceil + 1 \right) \left(d(x_k, x^*) < \varepsilon \right).$$

Given $\varepsilon > 0$, using Lemma 4.13, take

$$k \leq \left\lceil \frac{3b^2(\gamma-8\eta)(2\gamma+8\eta)}{8\eta(\gamma-12\eta)} \cdot \frac{64(3\gamma-16\eta)^2}{(3\gamma-28\eta)^2\varepsilon^2} \right\rceil \leq \left\lceil \frac{C^2}{\varepsilon^2} \right\rceil$$

such that

$$d(x_{k+1}, x_k) < \frac{3\gamma-28\eta}{8(3\gamma-16\eta)}\varepsilon.$$

Notice that as $12\gamma-112\eta \leq 24\gamma-128\eta$ by (A6), we in particular obtain that $d(x_{k+1}, x_k) < \frac{1}{4}\varepsilon$. Using that $x_{k+1} \in \text{Prox}_{\beta_k f(x_k, \cdot)}(x_k)$, applying Lemma 4.5 to the function $f(x_k, \cdot)$ with $y = x^*$ yields

$$\begin{aligned} & f(x_k, x_{k+1}) - \max\{f(x_k, x_{k+1}), f(x_k, x^*)\} \\ & \leq \frac{\lambda}{2} \left(\frac{\lambda}{\beta_k} - \gamma + \lambda\gamma \right) d^2(x_{k+1}, x^*) + \frac{\lambda}{\beta_k} \left\langle \overrightarrow{x_k x_{k+1}}, \overrightarrow{x_{k+1} x^*} \right\rangle \end{aligned}$$

for all $\lambda \in [0, 1]$. We consider two cases:

Case 1: $f(x_k, x_{k+1}) \geq f(x_k, x^*)$. We get

$$0 \leq \frac{\lambda}{2} \left(\frac{\lambda}{\beta_k} - \gamma + \lambda\gamma \right) d^2(x_{k+1}, x^*) + \frac{\lambda}{\beta_k} \left\langle \overrightarrow{x_k x_{k+1}}, \overrightarrow{x_{k+1} x^*} \right\rangle.$$

for all $\lambda \in [0, 1]$. Multiplying by $\frac{\beta_k}{\lambda}$, using (CS) and the fact that $\beta_k \geq \frac{1}{\gamma-8\eta}$, we obtain

$$\begin{aligned} & \frac{1}{2} \left((1-\lambda) \frac{\gamma}{\gamma-8\eta} - \lambda \right) d^2(x_{k+1}, x^*) \\ & \leq \left\langle \overrightarrow{x_k x_{k+1}}, \overrightarrow{x_{k+1} x^*} \right\rangle \leq d(x_{k+1}, x_k) d(x_{k+1}, x^*) \end{aligned}$$

for all $\lambda \in (0, 1]$. Now either $d(x_{k+1}, x^*) = 0$, and we are done by the Fejér monotonicity of $\{x_k\}_k$ w.r.t. $S(f) = \{x^*\}$ (recall Lemma 4.9), or $d(x_{k+1}, x^*) > 0$, and we obtain

$$\frac{1}{2} \left((1-\lambda) \frac{\gamma}{\gamma-8\eta} - \lambda \right) d(x_{k+1}, x^*) \leq d(x_{k+1}, x_k)$$

for all $\lambda \in (0, 1]$. This in particular holds for $\lambda = \frac{1}{8}$. By the above, we now have that

$$d(x_{k+1}, x_k) < \frac{1}{4}\varepsilon = \frac{3\gamma+4\eta}{12\gamma+16\eta}\varepsilon < \frac{3\gamma+4\eta}{8(\gamma-8\eta)}\varepsilon.$$

As $(3\gamma + 4\eta)/(8(\gamma - 8\eta)) > 0$ by (A6), we get $d(x_{k+1}, x^*) < \varepsilon$ and thus, with Lemma 4.9, we obtain $d(x_j, x^*) < \varepsilon$ for all $j \geq k + 1$.

Case 2: $f(x_k, x_{k+1}) < f(x_k, x^*)$. Then, just as in the proof of Lemma 4.7 in that case, we obtain

$$0 \leq \frac{\lambda}{2} \left(\frac{\lambda}{\beta_k} - \gamma + \lambda\gamma \right) d^2(x_{k+1}, x^*) + \frac{\lambda}{\beta_k} \left\langle \overrightarrow{x_k x_{k+1}}, \overrightarrow{x_{k+1} x^*} \right\rangle + \eta \left(d^2(x_{k+1}, x_k) + d^2(x_{k+1}, x^*) \right)$$

for all $\lambda \in [0, 1]$. Similar to Case 1, multiplying by $\frac{\beta_k}{\lambda}$ and using (CS), we obtain

$$\begin{aligned} & \frac{1}{2} \left((1 - \lambda)\gamma\beta_k - \lambda - \frac{2\eta\beta_k}{\lambda} \right) d^2(x_{k+1}, x^*) \\ & \leq d(x_{k+1}, x_k) d(x_{k+1}, x^*) + \frac{\eta\beta_k}{\lambda} d^2(x_{k+1}, x_k) \end{aligned}$$

for all $\lambda \in (0, 1]$. We make two further case distinctions:

Case 2a: $d(x_{k+1}, x^*) \leq d(x_{k+1}, x_k)$. As with $d(x_{k+1}, x_k) < \frac{1}{4}\varepsilon$, we in particular have that $d(x_{k+1}, x_k) < \varepsilon$, the result follows with Lemma 4.9.

Case 2b: $d(x_{k+1}, x^*) > d(x_{k+1}, x_k)$. This yields

$$\frac{1}{2} \left((1 - \lambda)\gamma\beta_k - \lambda - \frac{2\eta\beta_k}{\lambda} \right) d(x_{k+1}, x^*) < d(x_{k+1}, x_k) + \frac{\eta\beta_k}{\lambda} d(x_{k+1}, x_k).$$

for all $\lambda \in (0, 1]$. Multiplying both sides by $\frac{\lambda}{\lambda + \beta_k \eta}$ and rewriting the expression slightly, we thus get

$$\frac{1}{2\left(\frac{\lambda}{\beta_k} + \eta\right)} \left((\lambda(1 - \lambda)\gamma - 2\eta) - \frac{\lambda^2}{\beta_k} \right) d(x_{k+1}, x^*) < d(x_{k+1}, x_k)$$

for all $\lambda \in (0, 1]$. We now want to bound the factor in front of $d(x_{k+1}, x^*)$ from below. To this end, we set $\lambda = \frac{3}{8}$ and, using $\beta_k \geq \frac{1}{\gamma - 8\eta}$, obtain

$$(\lambda(1 - \lambda)\gamma - 2\eta) - \frac{\lambda^2}{\beta_k} \geq (\lambda(1 - \lambda)\gamma - 2\eta) - \lambda^2(\gamma - 8\eta) = \frac{6\gamma - 56\eta}{64} > 0$$

with (A6). Then further, again as $\beta_k \geq \frac{1}{\gamma - 8\eta}$, we get

$$\frac{1}{2\left(\frac{\lambda}{\beta_k} + \eta\right)} \left((\lambda(1 - \lambda)\gamma - 2\eta) - \frac{\lambda^2}{\beta_k} \right) \geq \frac{1}{2\left(\frac{3(\gamma - 8\eta)}{8} + \eta\right)} \left(\frac{6\gamma - 56\eta}{64} \right) = \frac{3\gamma - 28\eta}{8(3\gamma - 16\eta)}.$$

Together, the above yields

$$\frac{3\gamma - 28\eta}{8(3\gamma - 16\eta)} d(x_{k+1}, x^*) < d(x_{k+1}, x_k) < \frac{3\gamma - 28\eta}{8(3\gamma - 16\eta)} \varepsilon$$

and since $(3\gamma - 28\eta)/8(3\gamma - 16\eta) > 0$, we get $d(x_{k+1}, x^*) < \varepsilon$. Using Lemma 4.9, we in particular have $d(x_j, x^*) < \varepsilon$ for any $j \geq k + 1$.

This completes the proof of (Q). To obtain the non-asymptotic variant of that effective convergence result, set $\varepsilon := C/\sqrt{k - 2}$ given a $k > 2$, and rearrange. \square

Remark 4.15. Already over \mathbb{R}^d , the above result strengthens the convergence result of Iusem and Lara (recall Theorem 4.1) in three ways: First, the above result provides quantitative estimates that (to our knowledge) are already novel in the Euclidean case. Furthermore, the assumption that the parameter sequence $\{\beta_k\}_k$ is contained in the open interval $(\frac{1}{\gamma - 8\eta}, \frac{1}{4\eta})$

was weakened to the closed interval $[\frac{1}{\gamma-8\eta}, \frac{1}{4\eta}]$. Lastly, the above result allows for a careful examination of the main assumptions on the bifunction. In that context, we find that (A3) can be dropped completely and, if the existence of an equilibrium point and the well-definedness of the algorithm are assumed beforehand, also (A1) can be dropped. If, on the other hand, (A1) is added as an assumption to Theorem 4.14 above, then the assumptions that $S(f) \neq \emptyset$ and that the algorithm is well-defined can both be dropped, using Theorem 3.9 and Proposition 4.4 respectively.

Remark 4.16. The quantitative estimate contained in Theorem 4.14 above 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 the work of Kohlenbach, López-Acedo and Nicolae [51]. We refer to the master thesis of the first author [21] for a detailed discussion in that vein.

Remark 4.17 (For logicians). As the assumption (A2), that is pseudomonotonicity, is a (generalized) Π_2 -statement, the logical methodology of proof mining underlying the quantitative result contained in Theorem 4.14 would a priori suggest a dependence of the respective rate on a (uniform) quantitative rendering of this pseudomonotonicity of f , say in the form of a *modulus of uniform pseudomonotonicity* for f , that is for a fixed $o \in X$ a function $p : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ such that

$$f(x, y) \geq -\frac{1}{p(b, n) + 1} \rightarrow f(y, x) \leq \frac{1}{n + 1}$$

for all $b, n \in \mathbb{N}$ and all $x, y \in X$ with $d(o, x), d(o, y) \leq b$. By the general logical metatheorems of proof mining for metric and CAT(0)-spaces [27], an effective modulus of uniform pseudomonotonicity can be extracted from a corresponding (suitable) proof of pseudomonotonicity. Indeed, as an example for this, consider $f_h(x, y) := h(y) - h(x)$ as defined in Example 3.1 for some lsc and strongly quasiconvex $h : X \rightarrow \mathbb{R}$, where it can be rather immediately seen that a modulus of uniform pseudomonotonicity is given by $p(b, n) := n$.

However, the extracted rate of convergence from Theorem 4.14 *does not* depend on such a modulus and in fact this independence can be logically explained and guaranteed a priori. Indeed, in the context of Theorem 4.14, the assumption (A2) is only ever used to deduce $\forall x \in X (f(x, x^*) \leq 0)$ from the characterizing property $\forall x \in X (f(x^*, x) \geq 0)$ of a given (but fixed) solution x^* . Instead of (A2), it hence suffices to work only with the two properties above, formulated relative to such a fixed solution x^* . As these are universal properties, they do not contribute to the extracted bounds, which guarantees the independence of the resulting rate from any such modulus of uniform pseudomonotonicity a priori. We refer to the master thesis of the first author [21] for a more in-depth discussion of the surrounding logical aspects.

We end this section by mentioning two further quantitative results that allow for a simplification of the constant C featuring in Theorem 4.14 under additional assumptions on the parameters. Concretely, as $\{\beta_k\}_k$ can be freely chosen, and since $(\frac{1}{\gamma-8\eta}, \frac{1}{4\eta})$ is nonempty, a permissible restriction for applying (PPA) would be to choose $\{\beta_k\}_k$ from a compact subinterval of $(\frac{1}{\gamma-8\eta}, \frac{1}{4\eta})$. If one follows this line, both the constant C and the resulting arguments for the convergence of the method simplify drastically, to the point where the adaptive choice of λ_k depending on β_k , as facilitated above by the function N from (†), is no longer necessary. We collect the resulting quantitative estimates in the following theorem, but omit the (simplified) proofs for them, which can be found in the master thesis of the first author [21].

Theorem 4.18. *Let X be a Hadamard space and let $f : X \times X \rightarrow \mathbb{R}$ be a bifunction satisfying (Ai) with $i = 2, 4, 5, 6$. Let $\Theta > 0$ with $\frac{1}{\gamma-8\eta} < \Theta < \frac{1}{4\eta}$ and let $\{\beta_k\}_k \subseteq [\frac{1}{\gamma-8\eta}, \Theta]$. Suppose that*

(PPA) is well-defined and let $\{x_k\}_k$ be its generated sequence. Suppose that $S(f) \neq \emptyset$. Then $\{x_k\}_k$ converges to the (unique) equilibrium point x^* of f . Further, we have the non-asymptotic guarantee

$$\forall k > 2 \left(d(x_k, x^*) < \frac{C'}{\sqrt{k-2}} \right)$$

where

$$C' := \frac{24b\gamma}{\sqrt{\Theta^*}(3\gamma - 8\eta)}$$

with $b^2 \geq d^2(x_0, x^*)$ and $\Theta^* := 1 - 4\eta\Theta$. If we further have a $\theta > 0$ such that $\{\beta_k\}_k \subseteq [\theta, \Theta] \subseteq (\frac{1}{\gamma-8\eta}, \frac{1}{4\eta})$, then we can set

$$C' := \max \left\{ 1, \frac{3}{2\theta^*} \right\} \frac{b}{\sqrt{\Theta^*}}$$

with b and Θ^* as before as well as $\theta^* := (\gamma - 8\eta)\theta - 1$.

5. REGULARIZED SELECTION METHODS

5.1. The method and preliminaries. We now move to the second method for solving the equilibrium problem (EP) over a Hadamard space (X, d) for a given bifunction $f : X \times X \rightarrow \mathbb{R}$, now following the first type of method discussed in the introduction which regularizes the bifunction as a whole (yielding the notion of a resolvent of the bifunction). We recall it here first over a Hilbert space $(X, \langle \cdot, \cdot \rangle)$ and a given closed and convex set $K \subseteq X$. Given a starting point $x_0 \in K$ and a sequence $\{\beta_k\}_k \subseteq (0, \infty)$, recursively define x_{k+1} by choosing

$$(\%) \quad x_{k+1} \in S(K, f_k) \text{ for } f_k(x, y) := f(x, y) + \frac{1}{\beta_k} \langle x - x_k, y - x \rangle,$$

where $S(K, f_k)$ refers to the solution set of the equilibrium problem defined by f_k over K , similar to before. Over Euclidean spaces, Iusem and Lara show that the sequence $\{x_k\}_k$ generated by this algorithm converges to an equilibrium point in the context of a suitable strongly quasiconvex and pseudomonotone bifunction:

Theorem 5.1 (cf. [37, Corollary 3.2]). *Let $X = \mathbb{R}^d$ and $K \subseteq X$ be closed and convex. Let $f : K \times K \rightarrow \mathbb{R}$ be a bifunction satisfying (A1) – (A5). Take $\beta > 0$ such that $\beta_k \geq \beta$ for all $k \in \mathbb{N}$ and let $\{x_k\}_k$ be the sequence generated by (%) over K . Then $\{x_k\}_k$ converges to a point $x^* \in S(K, f)$.*

In fact, in [37, Proposition 3.6], Iusem and Lara further “show” that (%) is well-defined, i.e. that $S(K, f_k) \neq \emptyset$ for all $k \in \mathbb{N}$. However, we cannot confirm their result. The following remark provides a detailed discussion in that vein, and in particular provides a counterexample to their main line of reasoning.

Remark 5.2. In the proof of the well-definedness of the method generated by (%) under the assumptions of Theorem 5.1 as given in [37, Proposition 3.6], it is argued that the nonemptiness of $S(K, f_{z,\beta})$ for the function $f_{z,\beta}(x, \cdot) := f(x, \cdot) + \frac{1}{\beta} \langle x - z, \cdot - x \rangle$ (and hence the nonemptiness of $S(K, f_k)$) follows from the fact that $f_{z,\beta}(x, \cdot)$ attains its minimum on K for any $z, x \in K$ and $\beta > 0$. However it seems like this implication should not hold. Towards a counterexample, we utilize [37, Example 4.2]. Take $\alpha > 0$ and $K = [0, 2]$. Define the bifunction $f : K \times K \rightarrow \mathbb{R}$ by

$$f(x, y) := \sqrt{y} - \sqrt{x} + \langle \alpha x, y - x \rangle.$$

Note that $\langle a, b \rangle := ab$ for all $a, b \in \mathbb{R}$. It follows by [37, Example 4.2] that the assumptions (A1) – (A6) hold for this function. Also, the function

$$f_{z,\beta}(x, \cdot) := \sqrt{\cdot} - \sqrt{x} + \langle (\alpha + 1/\beta)x - z/\beta, \cdot - x \rangle$$

clearly attains a minimum on $K = [0, 2]$ for any z, β and x . Now, consider $\beta = 1$ and $z = 1$. We will show that $S(K, f_{1,1}) = \emptyset$. For a contradiction, assume that $x^* \in S(K, f_{1,1})$. Then by assumption

$$f_{1,1}(x^*, y) = \sqrt{y} - \sqrt{x^*} + \langle (1 + \alpha)x^* - 1, y - x^* \rangle \geq 0$$

for all $y \in K$. Suppose that $x^* = 0$. Then, for $y = 2$, we obtain

$$f_{1,1}(0, 2) = \sqrt{2} + \langle -1, 2 \rangle = \sqrt{2} - 2 < 0,$$

contradicting that $x^* \in S(K, f_{1,1})$. However if $x^* > 0$, we obtain for $y = 0$ that

$$f_{1,1}(x^*, 0) = -\sqrt{x^*} + ((1 + \alpha)x^* - 1)(-x^*) = -\sqrt{x^*} - (1 + \alpha)(x^*)^2 + x^*.$$

But now, if $x^* \leq 1$ we have $\sqrt{x^*} \geq x^*$ and if $x^* > 1$, we have $(x^*)^2 > x^*$. Thus, the above expression is strictly negative for all choices of $x^* > 0$. We conclude that $S(K, f_{1,1})$ must be empty.

Nevertheless, we now move on towards generalizing the method (%) and its convergence result contained in Theorem 5.1 to the nonlinear setting of Hadamard spaces, and outfitting it with quantitative information. So, fix a Hadamard space (X, d) and a bifunction $f : X \times X \rightarrow \mathbb{R}$. Similar to before, as closed and convex subsets of Hadamard spaces are again Hadamard spaces, we will (without loss of generality) not consider a relativization onto such sets.

There are now multiple ways to generalize the method (%), based on different ways of regularizing f in a Hadamard space: Concretely, two immediate regularizations corresponding to the above definition in Hilbert spaces are

$$f_k(x, y) := f(x, y) + \frac{1}{\beta_k} \langle \overrightarrow{x_k x}, \overrightarrow{x y} \rangle$$

on the one hand, using the quasi-inner product on X , as well as

$$f_k(x, y) := f(x, y) - \frac{1}{\beta_k} g_x(\log_x x_k, \log_x y)$$

on the other, using the pseudo-inner product on $T_x X$. Yet another possible regularization utilizes the Busemann function of an associated geodesic ray: Supposing that X has the geodesic extension property, one can consider

$$f_k(x, y) := f(x, y) - \frac{1}{\beta_k} d(x_k, x)(b_{r_{x_k, x}}(y) - b_{r_{x_k, x}}(x)),$$

where $b_{r_{x_k, x}}$ is the Busemann function corresponding to the (unique) geodesic ray $r_{x_k, x}$ such that $r_{x_k, x}(0) = x_k$ and $r_{x_k, x}(d(x_k, x)) = x$ (which exists since X has the geodesic extension property).³

Indeed, all regularizations coincide with (%) in Hilbert spaces. The first is employed in [46] in the context of Hadamard spaces. The second was previously taken as a point of departure in work of Colao, López, Marino and Martín-Márquez [18] for equilibrium problems over

³Since we rely on X having the geodesic extension property for the last type of regularization based on Busemann functions, there would be a gain in generality if we would formulate our equilibrium problem and resulting method over a fixed closed and convex set $K \subseteq X$. However, since the approach we outline here can be trivially extended to such situations, and to not create notational mismatch between the methods, we refrain from introducing such a relativization in the following.

Hadamard manifolds and is considered here (to our knowledge) for the first time in general Hadamard spaces. The last represents a regularization of a bifunction based on Busemann functions motivated by the recent work of Bento, Cruz Neto, Melo, et al. [8, 9], also set over Hadamard manifolds, modified here to both adequately extend the resulting object to Hadamard spaces and to allow applications in the context of strongly quasiconvex functions. All have previously however only been considered for convex bifunctions f .

Based on this plurality, and to provide a uniform study of (at least) these three methods above, we in the following consider a more abstract regularized selection method via

$$(RS) \quad x_{k+1} \in S(f_k) \text{ for } f_k(x, y) := f(x, y) + \frac{1}{\beta_k} r(x_k, x, y),$$

where $r : X^3 \rightarrow \mathbb{R}$ is an abstract regularization function for which we will assume two key properties for any $u, v, w \in X$:

$$(R1) \quad r(u, v, w) \leq \frac{1}{2}(d^2(w, u) - d^2(w, v) - d^2(u, v)),$$

$$(R2) \quad r(u, v, w) \leq d(u, v)d(v, w).$$

Before moving on, the next two lemmas quickly note that (R1) and (R2) hold for the two regularization functions used above.

Lemma 5.3. *Define $r(u, v, w) := \langle \overrightarrow{uv}, \overrightarrow{vw} \rangle$. Then r satisfies (R1) and (R2).*

Proof. By definition, we have

$$\langle \overrightarrow{uv}, \overrightarrow{vw} \rangle = \frac{1}{2} (d^2(w, u) - d^2(w, v) - d^2(u, v))$$

which yields (R1). Further, by (CS), we get $\langle \overrightarrow{uv}, \overrightarrow{vw} \rangle \leq d(u, v)d(v, w)$ which is (R2). \square

Lemma 5.4. *Define $r(u, v, w) := -g_v(\log_v u, \log_v w)$. Then r satisfies (R1) and (R2).*

Proof. Lemma 2.1 yields

$$-g_v(\log_v u, \log_v w) \leq \frac{1}{2} (d^2(w, u) - d^2(w, v) - d^2(u, v))$$

which is (R1). Further, recall from Section 2.3 that

$$-g_v(\log_v u, \log_v w) \leq \|\log_v u\|_v \|\log_v w\|_v = d(u, v)d(v, w)$$

by (CS) on $T_x X$, which is (R2). \square

To deal with the last regularization, we first require the following result about Busemann functions.

Lemma 5.5. *Let X have the geodesic extension property. Then, for any $u, v, w \in X$:*

$$d(u, v)b_{r_{u,v}}(w) \geq \frac{1}{2} (d^2(w, v) - d^2(w, u) - d^2(u, v)).$$

Proof. To simplify notation, write r for $r_{u,v}$ and a for $d(u, v)$. Given $t > a$, note that $v = (1 - \frac{a}{t})u \oplus \frac{a}{t}r(t)$. Hence, using (CN⁺), we get

$$d^2(w, v) \leq \left(1 - \frac{a}{t}\right) d^2(w, u) + \frac{a}{t} d^2(w, r(t)) - \frac{a}{t} \left(1 - \frac{a}{t}\right) d^2(u, r(t)).$$

As r is a geodesic ray issuing at u , we have $d^2(u, r(t)) = t^2$ and so $\frac{a}{t}(1 - \frac{a}{t})t^2 = a(t - a)$. Rearranging the above and multiplying by $\frac{t}{a}$, we get

$$\begin{aligned} d^2(w, r(t)) &\geq \frac{t}{a} (d^2(w, v) - d^2(w, u)) + d^2(w, u) + t(t - a) \\ &\geq t^2 + tc + d^2(w, u) \end{aligned}$$

where $c = \frac{1}{a}(d^2(w, v) - d^2(w, u) - a^2)$. Hence, we get

$$\begin{aligned} d(w, r(t)) - t &\geq \sqrt{t^2 + tc + d^2(w, u)} - t \\ &= \frac{tc + d^2(w, u)}{\sqrt{t^2 + tc + d^2(w, u)} + t} \geq \frac{tc + d^2(w, u)}{2t + \sqrt{tc + d^2(w, u)}} \end{aligned}$$

so that we get $b_{r_{u,v}}(w) = \lim_{t \rightarrow \infty} (d(w, r(t)) - t) \geq \frac{c}{2}$. Multiplying by a gives the result. \square

Now, we can recognize the last type of regularization as an instance of our abstract schema.

Lemma 5.6. *Let X have the geodesic extension property. Define $r(u, v, w) := -d(u, v)(b_{r_{u,v}}(w) - b_{r_{u,v}}(v))$. Then r satisfies (R1) and (R2).*

Proof. It is easy to see that $b_{r_{u,v}}(v) = -d(u, v)$. Hence, using Lemma 5.5, we get

$$\begin{aligned} d(u, v)(b_{r_{u,v}}(w) - b_{r_{u,v}}(v)) &= d(u, v)b_{r_{u,v}}(w) + d^2(u, v) \\ &\geq \frac{1}{2}(d^2(w, v) - d^2(w, u) - d^2(u, v)) + d^2(u, v) \\ &\geq \frac{1}{2}(d^2(w, v) - d^2(w, u) + d^2(u, v)) \end{aligned}$$

so that

$$r(u, v, w) \leq \frac{1}{2}(d^2(w, u) - d^2(w, v) - d^2(u, v))$$

which is (R1). For (R2), note that we have

$$-d(u, v)(b_{r_{u,v}}(w) - b_{r_{u,v}}(v)) = d(u, v)(b_{r_{u,v}}(v) - b_{r_{u,v}}(w)) \leq d(u, v)d(v, w)$$

since $b_{r_{u,v}}$ is nonexpansive (recall Section 2). \square

In comparison to (9%), where the well-definedness of the algorithm was already unclear even in the linear case for strongly quasiconvex functions, this abstract phrasing via (RS) makes the well-definedness of the method even more precarious, especially in Hadamard spaces and already for convex bifunctions. We will hence in the following generally only assume (and not show) that $S(f_k) \neq \emptyset$ for all $k \in \mathbb{N}$ in the Hadamard space setting, so that the method (RS) is well-defined. However, we briefly comment on some partial results in that vein in Remark 5.7 below.

In that way, although the proof of the main convergence theorem and its corresponding quantitative estimates given below (see Theorem 5.10) are simpler than their counterparts for (PPA) from the previous section (recall Theorems 4.14 and 4.18), the great disadvantage of (RS) is that it is less clear whether the algorithm is well-defined, and it further does not seem immediate in what respect the corresponding subproblems are easy to solve, as they lack important structure (e.g. w.r.t. quasiconvexity).

Remark 5.7. In [46], Khatibzadeh and Mohebbi consider well-definedness of (RS) in Hadamard spaces for $r(u, v, w) := \langle \vec{u}\vec{v}, \vec{v}\vec{w} \rangle$, but where f is convex in its second argument. They encounter a different difficulty compared to the above Remark 5.2 in this nonlinear context: They can show that $S(f)$ is nonempty if f is convex in its right argument, is θ -undermonotone (see [46, p. 233]) and satisfies (A0) and (A1). However, as the function $y \mapsto \langle \vec{z}\vec{x}, \vec{x}\vec{y} \rangle$ is not necessarily convex in general Hadamard spaces, neither is $f_{z,\beta}(x, \cdot) := f(x, \cdot) + \frac{1}{\beta}r(z, x, \cdot)$ for the above r . Thus, the previous arguments, reliant on the convexity property of f , cannot be applied to show nonemptiness of $S(f_{z,\beta})$. Khatibzadeh and Mohebbi leave it as an open question whether the algorithm is well-defined only assuming (A0), (A1), convexity and θ -undermonotonicity of f and only show nonemptiness of $S(f_{z,\beta})$ under additional assumptions on f (e.g. if f is

cyclically monotone). It was claimed in [18, Proposition 2.9] (as well as [72, Proposition 3.4, (ii)]) that the issue of the convexity of the regularization function r is mitigated for $r(u, v, w) := -g_v(\log_v u, \log_v w)$ over Hadamard manifolds, which was used in [18, Theorem 4.9] to establish that $S(f_{z,\beta})$ is nonempty whenever f convex in its right argument, monotone and satisfies some further auxiliary conditions (including a compactness condition). However, as shown in [53], these results from [18, 72] are actually false in general and the function r is only convex over linear spaces. The last type of regularization $r(u, v, w) := -d(u, v)(b_{r_{u,v}}(w) - b_{r_{u,v}}(v))$ is actually concave in w since the Busemann function is convex (recall Section 2). This is contrary to the regularization $d(u, v)b_{r_{v,u}}(w)$ introduced in [8], which served as a motivation for the above regularization introduced in this paper. Indeed, the convexity of this regularization is crucially used therein to establish the well-definedness of the method (cf. [8, Theorem 1.1]). In any case, such convexity results for r would not immediately solve the associated problem in the strongly quasiconvex case as the addition of a strongly quasiconvex and a convex function need not be strongly quasiconvex, so that not even Theorem 3.9 can be applied without additional considerations. Naturally, if $f_{z,\beta}$ happens to be strongly quasiconvex for all $\beta > 0$ and $z \in X$ in its right argument and further satisfies (A0),(A1) and (A2), then Theorem 3.9 of course guarantees $S(f_{z,\beta}) \neq \emptyset$.

5.2. Quantitative convergence results. We now turn to the main focus of this section, the generalized and effective variant of Theorem 5.1. In [37], Iusem and Lara refer⁴ to [42, Theorem 1] where, in a Hilbert space setting, weak convergence of the sequence generated by (%) is shown under the assumptions that $S(f) \neq \emptyset$ and that f is suitably semicontinuous, monotone and convex. They then argue that this can easily be transferred to the strongly quasiconvex setting of Theorem 5.1. Inspired by [67], we follow a slightly different argument which in particular will allow us to show a strong convergence result together with a rate of convergence in the general setting of strongly quasiconvex and pseudomonotone bifunctions on Hadamard spaces.

We begin by adapting [42, Proposition 4] to our setting, which will allow us to show Fejér monotonicity of $\{x_k\}_k$ and derive a quantitative result on the asymptotic behavior of $d(x_{k+1}, x_k)$. The proof is an easy adaptation of the proof of [42, Proposition 4].

Lemma 5.8. *Let f be a bifunction satisfying (A2) and let r satisfy (R1). Let $\{\beta_k\}_k \subseteq (0, \infty)$. Suppose that (RS) is well-defined and let $\{x_k\}_k$ be its generated sequence. Let $x^* \in S(f)$. Then, for every $k \in \mathbb{N}$:*

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

Proof. Fix $k \in \mathbb{N}$. As $x_{k+1} \in S(f_k)$, we obtain

$$0 \leq f_k(x_{k+1}, x^*) = f(x_{k+1}, x^*) + \frac{1}{\beta_k} r(x_k, x_{k+1}, x^*)$$

and thus $-f(x_{k+1}, x^*) \leq \frac{1}{\beta_k} r(x_k, x_{k+1}, x^*)$. As x^* is an equilibrium point for f , we have $f(x^*, y) \geq 0$ for all $y \in X$, and by pseudomonotonicity of f it follows that $f(y, x^*) \leq 0$ for all $y \in X$. In particular, we have $f(x_{k+1}, x^*) \leq 0$, and hence

$$0 \leq \frac{1}{\beta_k} r(x_k, x_{k+1}, x^*) \leq \frac{1}{2\beta_k} (d^2(x^*, x_k) - d^2(x^*, x_{k+1}) - d^2(x_{k+1}, x_k)),$$

using (R1). As $\beta_k > 0$ for all $k \in \mathbb{N}$, this implies the result. \square

The approach towards our quantitative convergence theorem for (RS) is now similar as with (PPA). First, we derive an effective estimate for the asymptotic behavior of $d(x_k, x_{k+1})$ from the above.

⁴They actually reference [40], but we believe this might be a typo.

Lemma 5.9. *Let f be a bifunction satisfying (A2) and let r satisfy (R1). Let $\{\beta_k\}_k \subseteq (0, \infty)$. Suppose that (RS) is well-defined and let $\{x_k\}_k$ be its generated sequence. Let $x^* \in S(f)$. Then, for any $\varepsilon > 0$:*

$$\exists k \leq \left\lceil \frac{b^2}{\varepsilon} \right\rceil \left(d^2(x_k, x_{k+1}) < \varepsilon \right),$$

where $b^2 \geq d^2(x_0, x^*)$.

Proof. Immediately follows with Lemma 4.10 and Lemma 5.8. \square

The above asymptotic estimate for $d(x_k, x_{k+1})$ can now already be used to show the main quantitative convergence theorem.

Theorem 5.10. *Let X be a Hadamard space and let f be a bifunction satisfying (A2), (A4) and (A5). Further, let r satisfy (R1) and (R2). Let $\{\beta_k\}_k \subseteq (0, \infty)$ be a sequence such that $\beta \leq \beta_k$ for all $k \in \mathbb{N}$. Suppose that (RS) is well-defined and let $\{x_k\}_k$ be its generated sequence. Suppose that $S(f) \neq \emptyset$. Then $\{x_k\}_k$ converges to the (unique) equilibrium point x^* of f . Further, we have the non-asymptotic guarantee*

$$\forall k > 2 \left(d(x_k, x^*) < \frac{D}{\sqrt{k-2}} \right)$$

where $D := 4b/\beta\gamma$ with $b^2 \geq d^2(x_0, x^*)$.

Proof. It is enough to show the quantitative result, and to that end we show

$$(Q') \quad \forall \varepsilon > 0 \quad \forall k \geq \left(\left\lceil \frac{D^2}{\varepsilon^2} \right\rceil + 1 \right) \left(d(x_k, x^*) < \varepsilon \right).$$

The non-asymptotic guarantee above can then be derived similar as in Theorem 4.14. Given $\varepsilon > 0$, using Lemma 5.9, choose

$$k \leq \left\lceil \frac{16b^2}{\beta^2\gamma^2\varepsilon^2} \right\rceil \leq \left\lceil \frac{D^2}{\varepsilon^2} \right\rceil \text{ such that } d(x_{k+1}, x_k) < \frac{\beta\gamma}{4}\varepsilon.$$

Using that $x_{k+1} \in S(f_k)$, by definition we obtain

$$0 \leq f(x_{k+1}, y) + \frac{1}{\beta_k} r(x_k, x_{k+1}, y)$$

for all $y \in X$. Consider $y := (1 - \lambda)x_{k+1} \oplus \lambda x^*$ where $\lambda \in (0, 1)$. Then

$$0 \leq f(x_{k+1}, (1 - \lambda)x_{k+1} \oplus \lambda x^*) + \frac{1}{\beta_k} r(x_k, x_{k+1}, (1 - \lambda)x_{k+1} \oplus \lambda x^*)$$

By strong quasiconvexity of f and (R2), we obtain

$$\begin{aligned} 0 &\leq \max\left\{ f(x_{k+1}, x^*), f(x_{k+1}, x_{k+1}) \right\} - \lambda(1 - \lambda) \frac{\gamma}{2} d^2(x^*, x_{k+1}) \\ &\quad + \frac{1}{\beta_k} d(x_{k+1}, x_k) d((1 - \lambda)x_{k+1} \oplus \lambda x^*, x_{k+1}). \end{aligned}$$

As x^* is an equilibrium point for f , we know $f(x^*, y) \geq 0$ for all $y \in X$ and thus by pseudomonotonicity of f , we in particular have $f(x_{k+1}, x^*) \leq 0$. Further, $f(x_{k+1}, x_{k+1}) = 0$ by (A0) (which follows from (A2) and (A5)). Thus, using convexity of d in its first argument, we obtain

$$\lambda(1 - \lambda) \frac{\gamma}{2} d^2(x^*, x_{k+1}) \leq \frac{\lambda}{\beta_k} d(x_{k+1}, x_k) d(x^*, x_{k+1}).$$

Choosing $\lambda = 1/2$ and using $\beta \leq \beta_k$, this now finally yields

$$\frac{\beta\gamma}{4}d^2(x^*, x_{k+1}) \leq d(x_{k+1}, x_k)d(x^*, x_{k+1}).$$

If $d(x^*, x_{k+1}) = 0$, we are done by Lemma 5.8. Otherwise, we obtain

$$\frac{\beta\gamma}{4}d(x^*, x_{k+1}) \leq d(x_{k+1}, x_k) < \frac{\beta\gamma}{4}\varepsilon.$$

As $\frac{\beta\gamma}{4} > 0$, we get $d(x^*, x_{k+1}) < \varepsilon$ and again we are done with Lemma 5.8. This completes the proof of (Q'). \square

Remark 5.11. Already over \mathbb{R}^d , the above result strengthens the convergence result of Iusem and Lara (recall Theorem 5.1) by providing quantitative estimates, which (to our knowledge) are already novel in the Euclidean case. Again, the above approach allows for a careful examination of the main assumptions on the bifunction, whereby (A1) and (A3) are not needed to show the convergence result if the existence of an equilibrium point and well-definedness of the algorithm are assumed.

Further, both Remark 4.16 and Remark 4.17 extend from (PPA) to this setting of (RS), and we again refer to the master thesis of the first author [21] for additional discussions in these regards.

Acknowledgments: This paper is a revised version of parts of the master thesis [21] of the first author, written under the supervision of Prof. Dr. Ulrich Kohlenbach at TU Darmstadt and co-supervised by the second author. Both authors want to thank Prof. Kohlenbach for many insightful comments on the topic of the thesis [21]. The authors also want to thank Sorin-Mihai Grad and Felipe Lara for helpful conversations on and around the topic of this paper.

REFERENCES

- [1] A.D. Aleksandrov. A theorem on triangles in a metric space and some of its applications. *Proceedings of the Steklov Institute of Mathematics*, 38:5–23, 1951.
- [2] A.D. Aleksandrov. Über eine Verallgemeinerung der Riemannschen Geometrie. *Schriftenreihe des Forschungsinstituts für Mathematik*, 1:33–84, 1957.
- [3] Q.H. Ansari, F. Babu, and M.S. Raju. Proximal point method with Bregman distance for quasiconvex pseudomonotone equilibrium problems. *Optimization*, 74(9):2035–2055, 2025.
- [4] H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer, 2nd edition, 2017.
- [5] M. Bačák. The proximal point algorithm in metric spaces. *Israel Journal of Mathematics*, 194:689–701, 2013.
- [6] M. Bačák. *Convex Analysis and Optimization in Hadamard Spaces*. Series in Nonlinear Analysis and Applications. De Gruyter, Berlin, 2014.
- [7] M. Bačák. Old and new challenges in Hadamard spaces. *Japanese Journal of Mathematics*, 18(2):115–168, 2023.
- [8] G.d.C. Bento, J.X. Cruz Neto, J.O. Lopes, Í.D.L. Melo, and P.S. Filho. A New Approach About Equilibrium Problems via Busemann Functions. *Journal of Optimization Theory and Applications*, 200(1):428–436, 2024.
- [9] G.d.C. Bento, J.X. Cruz Neto, and Í.D.L. Melo. Combinatorial Convexity in Hadamard Manifolds: Existence for Equilibrium Problems. *Journal of Optimization Theory and Applications*, 195(3):1087–1105, 2022.
- [10] I.D. Berg and I.G. Nikolaev. Quasilinearization and curvature of Aleksandrov spaces. *Geometriae Dedicata*, 133(1):195–218, 2008.

- [11] E. Blum and W. Oettli. From optimization and variational inequalities to equilibrium problems. *Mathematics Student*, 63(1):123–145, 1994.
- [12] K. Boufi, A. Fadil, and A. Roubi. New formulation with proximal point algorithm and proximal bundle methods for equilibrium problems. *Optimization*, 2026. In press.
- [13] H. Brezis and P.L. Lions. Produits infinis de resolvantes. *Israel Journal of Mathematics*, 29(4):329–345, 1978.
- [14] M. Bridson and A. Haefliger. *Metric Spaces of Non-Positive Curvature*, volume 319 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 1999.
- [15] F. Bruhat and J. Tits. Groupes Réductifs Sur Un Corps Local. *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, 41(1):5–251, 1972.
- [16] R. Burachik and G. Kassay. On a generalized proximal point method for solving equilibrium problems in Banach spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 75(18):6456–6464, 2012.
- [17] P. Chaipunya, F. Kohsaka, and P. Kumam. Monotone vector fields and generation of nonexpansive semi-groups in complete CAT(0) spaces. *Numerical Functional Analysis and Optimization*, 42(9):989–1018, 2021.
- [18] V. Colao, G. López, G. Marino, and V. Martín-Márquez. Equilibrium problems in Hadamard manifolds. *Journal of Mathematical Analysis and Applications*, 388(1):61–77, 2012.
- [19] P.L. Combettes and S.A. Hirstoaga. Equilibrium programming in Hilbert spaces. *Journal of Nonlinear and Convex Analysis*, 1(6):117–136, 2005.
- [20] J. Cotrina and Y. García. Equilibrium Problems: Existence Results and Applications. *Set-Valued and Variational Analysis*, 26(1):159–177, 2018.
- [21] L.M. Després. On the proximal point algorithm for strongly quasiconvex pseudomonotone equilibrium problems in Hadamard spaces. Master's thesis, TU Darmstadt, 2026.
- [22] R. Espínola and A. Fernández-León. CAT(k)-spaces, weak convergence and fixed points. *Journal of Mathematical Analysis and Applications*, 353(1):410–427, 2009.
- [23] K. Fan. A generalization of Tychonoff's fixed point theorem. *Mathematische Annalen*, 142(3):305–310, 1961.
- [24] K. Fan. A Minimax Inequality and Applications. In O. Sisha, editor, *Inequalities*, volume 3, pages 103–113. Academic Press, New York, 1972.
- [25] O.P. Ferreira and P.R. Oliveira. Proximal point algorithm on Riemannian manifolds. *Optimization*, 51:257–270, 2002.
- [26] F. Flores-Bazán. Existence Theorems for Generalized Noncoercive Equilibrium Problems: The Quasi-Convex Case. *SIAM Journal on Optimization*, 11(3):675–690, 2001.
- [27] P. Gerhardy and U. Kohlenbach. General Logical Metatheorems for Functional Analysis. *Transactions of The American Mathematical Society*, 360:2615–2661, 2008.
- [28] 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.
- [29] 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.
- [30] S.-M. Grad, F. Lara, and R.T. Marcavillaca. Strongly Quasiconvex Functions: What We Know (So Far). *Journal of Optimization Theory and Applications*, 205(2), 2025. Article no. 38.
- [31] 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.
- [32] O. Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM Journal on Control and Optimization*, 29:403–419, 1991.
- [33] N. Hadjisavvas, F. Lara, R.T. Marcavillaca, and P.T. Vuong. Heavy ball and Nesterov accelerations with Hessian-driven damping for nonconvex optimization. *Applied Mathematics & Optimization*, 93(3), 2026. Article no. 63.
- [34] D. Hieu, H.N. Duong, and B.H. Thai. Convergence of relaxed inertial methods for equilibrium problems. *Journal of Applied and Numerical Optimization*, 3(1):215–229, 2021.
- [35] A. Iusem, G. Kassay, and W. Sosa. On certain conditions for the existence of solutions of equilibrium problems. *Mathematical Programming*, 116:259–273, 2009.
- [36] A. Iusem and F. Lara. Optimality Conditions for Vector Equilibrium Problems with Applications. *Journal of Optimization Theory and Applications*, 180(1):187–206, 2019.
- [37] A. Iusem and F. Lara. Proximal Point Algorithms for Quasiconvex Pseudomonotone Equilibrium Problems. *Journal of Optimization Theory and Applications*, 193(1–3):443–461, 2022.

- [38] A. Iusem, F. Lara, R.T. Marcavillaca, and L.H. Yen. A two-step proximal point algorithm for nonconvex equilibrium problems with applications to fractional programming. *Journal of Global Optimization*, 90(3):755–779, 2024.
- [39] A. Iusem and V. Mohebbi. Convergence analysis of the extragradient method for equilibrium problems in Hadamard spaces. *Computational and Applied Mathematics*, 39, 2020. Article no. 44.
- [40] A. Iusem and W. Sosa. Iterative Algorithms for Equilibrium Problems. *Optimization*, 52(3):301–316, 2003.
- [41] A. Iusem and W. Sosa. New existence results for equilibrium problems. *Nonlinear Analysis: Theory, Methods & Applications*, 52(2):621–635, 2003.
- [42] A. Iusem and W. Sosa. On the proximal point method for equilibrium problems in Hilbert spaces. *Optimization*, 59(8):1259–1274, 2010.
- [43] J. Jost. Equilibrium maps between metric spaces. *Calculus of Variations and Partial Differential Equations*, 2(2):173–204, 1994.
- [44] M. Jovanović. A note on strongly convex and quasiconvex functions. *Mathematical Notes*, 60:584–585, 1996.
- [45] H. Khatibzadeh and V. Mohebbi. Proximal point algorithm for infinite pseudo-monotone bifunctions. *Optimization*, 65(8):1629–1639, 2016.
- [46] H. Khatibzadeh and V. Mohebbi. Monotone and pseudo-monotone equilibrium problems in Hadamard spaces. *Journal of the Australian Mathematical Society*, 110(2):220–242, 2021.
- [47] W.A. Kirk and B. Panyanak. A concept of convergence in geodesic spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 68(12):3689–3696, 2008.
- [48] B. Knaster, C. Kuratowski, and S. Mazurkiewicz. Ein Beweis des Fixpunktsatzes für n-dimensionale Simplexes. *Fundamenta Mathematicae*, 14:132–137, 1929.
- [49] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer, Berlin, Heidelberg, 2008.
- [50] U. Kohlenbach. Proof-theoretic Methods in Nonlinear Analysis. In B. Sirakov, P. Ney de Souza, and M. Viana, editors, *Proceedings of the International Congress of Mathematicians (ICM 2018)*, volume 2, pages 61–82. World Scientific, 2019.
- [51] U. Kohlenbach, G. López-Acedo, and A. Nicolae. Moduli of regularity and rates of convergence for Fejér monotone sequences. *Israel Journal of Mathematics*, 232(1):261–297, 2019.
- [52] I.V. Konnov. Application of the proximal point method to nonmonotone equilibrium problems. *Journal of Optimization Theory and Applications*, 119:317–333, 2003.
- [53] A. Kristály, C. Li, G. López-Acedo, and A. Nicolae. What do ‘convexities’ imply on Hadamard manifolds? *Journal of Optimization Theory and Applications*, 170(3):1068–1074, 2016.
- [54] F. Lara. On Strongly Quasiconvex Functions: Existence Results and Proximal Point Algorithms. *Journal of Optimization Theory and Applications*, 192(3):891–911, 2022.
- [55] F. Lara and R.T. Marcavillaca. Bregman proximal point type algorithms for quasiconvex minimization. *Optimization*, 73(3):497–515, 2024.
- [56] F. Lara, R.T. Marcavillaca, and P.T. Vuong. Characterizations, dynamical systems and gradient methods for strongly quasiconvex functions. *Journal of Optimization Theory and Applications*, 206(3), 2025. Article no. 60.
- [57] C. Li, G. López, and V. Martín-Márquez. Monotone vector fields and the proximal point algorithm on Hadamard manifolds. *Journal of the London Mathematical Society*, 79:663–683, 2009.
- [58] T.-C. Lim. Remarks on some fixed point theorems. *Proceedings of the American Mathematical Society*, 60:179–182, 1976.
- [59] R. López. Approximations of equilibrium problems. *SIAM Journal on Control and Optimization*, 50(2):1038–1070, 2012.
- [60] B. Martinet. Brève communication. Régularisation d’inéquations variationnelles par approximations successives. *Revue française d’informatique et de recherche opérationnelle. Série rouge*, 4(R3):154–158, 1970.
- [61] A. Moudafi. Proximal point algorithm extended to equilibrium problems. *Journal of Natural Geometry*, 15(1–2):91–100, 1999.
- [62] A. Moudafi. Second-order differential proximal methods for equilibrium problems. *Journal of Inequalities in Pure and Applied Mathematics*, 4, 2003. Article no. 18.
- [63] A. Moudafi and M. Théra. Proximal and Dynamical Approaches to Equilibrium Problems. In M. Théra and R. Tichatschke, editors, *Ill-posed Variational Problems and Regularization Techniques*, pages 187–201. Springer, Berlin, Heidelberg, 1999.
- [64] I.G. Nikolaev. The tangent cone of an Aleksandrov space of curvature $\leq K$. *Manuscripta Mathematica*, 86(1):137–147, 1995.

- [65] W. Oettli. A remark on vector-valued equilibria and generalized monotonicity. *Acta Mathematica Vietnamica*, 22:213–221, 1997.
- [66] S. Ohta. Barycenters in Alexandrov spaces of curvature bounded below. *Advances in Geometry*, 12(4):571–587, 2012.
- [67] N. Pischke. On the proximal point algorithm for strongly quasiconvex functions in Hadamard spaces. *Optimization Methods and Software*, 40(6):1438–1453, 2025.
- [68] N. Pischke. Duality, Fréchet differentiability and Bregman distances in hyperbolic spaces. *Israel Journal of Mathematics*, 2026. In press.
- [69] B.T. Polyak. Existence Theorems and Convergence of Minimizing Sequences in Extremum Problems with Restrictions. *Soviet Mathematics. Doklady*, 7:72–75, 1966.
- [70] E.A. Papa Quiroz. Proximal Point Method for Quasiconvex Functions in Riemannian Manifolds. *Journal of Optimization Theory and Applications*, 202(3):1268–1285, 2024.
- [71] E.A. Papa Quiroz, N. Baygorrea CusiHuallpa, and N. Maculan. Inexact Proximal Point Methods for Multiobjective Quasiconvex Minimization on Hadamard Manifolds. *Journal of Optimization Theory and Applications*, 186(3):879–898, 2020.
- [72] 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):46–69, 2009.
- [73] 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.
- [74] 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.
- [75] R.T. Rockafellar. Monotone Operators and the Proximal Point Algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976.
- [76] A. Wald. Begründung einer koordinatenlosen Differentialgeometrie der Flächen. *Ergebnisse eines Mathematischen Kolloquiums*, 7:24–46, 1936.