

A RELAXED-INERTIAL PROXIMAL POINT ALGORITHM FOR STRONGLY QUASICONVEX EQUILIBRIUM PROBLEMS ON HADAMARD MANIFOLDS

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 a proximal point type method for approximating solutions to equilibrium problems generated by pseudomonotone and strongly quasiconvex bifunctions over Hadamard manifolds, that is complete simply connected Riemannian manifolds of nonpositive sectional curvature. Next to the usual proximal step, the method we consider also incorporates an inertia step together with a subsequent over-relaxation, the latter of which is treated in the context of Hadamard manifolds, to our knowledge, for the first time. Making use of a quantitative approach towards such proximal methods for strongly quasiconvex optimization developed by the authors in previous work, we in particular provide effective arguments for the convergence of the method, yielding explicit, fast and very uniform rates of convergence for the distance of the iterates towards the solution. These results extend previous work by Grad, Lara and Marcavillaca on such a method over finite-dimensional Euclidean spaces for the first time to a nonlinear setting, with the quantitative estimates already being novel in the Euclidean case. In particular, our effective approach allows for a fine-grained view on the assumptions on the surrounding objects, so that we are able to either weaken or even fully discharge some previous assumptions.

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

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

1. INTRODUCTION

1.1. Background and motivation. Many problems in applied mathematics, including in particular minimization problems, fixed point problems and variational inequalities, can be formulated as an equilibrium problem, that is as a problem of the form

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

for a suitable bifunction $f : K \times K \rightarrow \mathbb{R}$ and an associated constraint set $K \subseteq X$ in a space X . Also called “Ky Fan inequalities” after the pioneering work of Fan [27], these problems have been studied deeply, and we refer to [14, 22, 29, 34, 35, 49, 54] next to many other works, for such investigations.

Many prominent methods used to approximate solutions of convex equilibrium problems are based on (modifications of) the proximal point scheme. Originally developed for convex minimization problems and associated inclusion problems of maximally monotone operators by Martinet [50], Rockafellar [64] as well as Brezis and Lions [16] over Hilbert spaces $(X, \|\cdot\|)$, this

method iteratively applies the so-called proximal map

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

of a convex, lower-semicontinuous function $h : K \rightarrow \mathbb{R}$ on a closed and convex subset $K \subseteq X$ to form the sequence $\{x_k\}_k$ defined by $x_{k+1} := \text{Prox}_{\beta_k h}(x_k)$, given some fixed starting point $x_0 \in K$ and using an associated parameter sequence $\{\beta_k\}_k \subseteq (0, \infty)$. Under suitable assumptions on the parameters, $\{x_k\}_k$ weakly converges to a minimum of h (see e.g. [6]).

Many extensions of this proximal point type approach to (convex) equilibrium problems rely on utilizing a suitable analog of this proximal map defined by regularizing the whole bifunction, also called the resolvent of the bifunction (see already [51, 53] but also [19, 38, 41, 46, 52]). However, in [36] (see also [33, 37]), Iusem and Lara define an algorithm that in each iteration applies a proximal regularization only to the second (convex) argument of the bifunction, that is it utilizes $\text{Prox}_{\beta_k f(x_k, \cdot)}(x_k)$ to define a new iterate. Indeed, as illustrated in [36], this type of proximal step is particularly suitable for equilibrium problems with relaxed convexity requirements, in particular bifunctions which are strongly quasiconvex in their right argument. Here, a function $h : K \rightarrow \mathbb{R}$ is called strongly quasiconvex if

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

for all $x, y \in K$ and $\lambda \in [0, 1]$, for a given value $\gamma > 0$. This class of functions already goes back to the work of Polyak [57] and represents a restriction of the class of quasiconvex functions, that is the class of functions for which the above inequality holds with $\gamma = 0$. Nevertheless, it still accommodates various functions of practical interest, such as the Euclidean norm and its square root (see [39, 47]), and it in particular contains nonconvex functions (as e.g. witnessed by $\max\{\sqrt{\|\cdot\|}, \|\cdot\|^2 - k\}$ for $k \in \mathbb{N}$, cf. [47, Remark 18]). We refer to [32, 47] for further discussions on the relationship between strongly quasiconvex functions and various other classes of (non)convex functions.

This class of strongly quasiconvex functions has various benefits over the class of plain quasiconvex functions mentioned above. In particular, many iterative procedures known from convex settings are still reasonably well-behaved (we in particular refer to [47] for further discussions on this topic). This in particular is true for the proximal point algorithm, which was studied for these functions by Lara [47], who proved a convergence theorem over Euclidean space $\mathbb{R}^d$ (cf. [47, Theorem 10]). Contrary to the usual proximal point method for convex functions, the proximal map $\text{Prox}_{\beta h}$ is however not single-valued anymore even over Euclidean spaces (cf. [47, Remark 6]), so that the proximal step is transformed into a selection $x_{k+1} \in \text{Prox}_{\beta_k h}(x_k)$ (which is well-defined as $\text{Prox}_{\beta h}$ is still non-empty). It is also in this setting that Iusem and Lara study the method sketched above in [36], that is they consider the iteration

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

for a (pseudomonotone and suitably Lipschitz) bifunction $f : K \times K \rightarrow \mathbb{R}$ on a closed and convex domain K which is only strongly quasiconvex in its second argument, and establish its convergence over Euclidean space $\mathbb{R}^d$ (cf. [36, Theorem 3.1]). In particular, a key benefit of the above method to e.g. a method which considers a regularization of the bifunction as a whole in the style of the resolvent of a bifunction is that the above selection is immediately well-defined whereas the well-definedness of such resolvents of bifunctions is quite non-trivial in the context of strong quasiconvexity (see in particular the discussion in [24]).

The suitability of the above proximal point type method for equilibrium problems induced by strongly quasiconvex bifunctions is also further affirmed through the subsequent work of Grad, Lara and Marcavillaca [31], where they show that the method can be efficiently extended with both inertia terms and over-relaxations (extending previous work on minimization problems for strongly quasiconvex functions [30]), illustrating that more sophisticated formulations of proximal point type schemes are still viable in the context of the above class of nonconvex functions. Concretely, given a (pseudomonotone and suitably Lipschitz) bifunction $f : K \times K \rightarrow \mathbb{R}$ over an affine subspace $K \subseteq \mathbb{R}^d$ that is strongly quasiconvex in its right argument, the relaxed inertial proximal point algorithm considered in [31] iteratively computes points $\{x_k\}_k$, $\{y_k\}_k$ and $\{z_k\}_k$ via

$$\begin{cases} y_k := x_k + \alpha_k(x_k - x_{k-1}), \\ z_k \in \text{Prox}_{\beta_k f(y_k, \cdot)}(y_k), \\ x_{k+1} := (1 - \rho_k)y_k + \rho_k z_k, \end{cases}$$

from given starting points $x_{-1}, x_0 \in K$ and sequences $\{\alpha_k\}_k$, $\{\rho_k\}_k$, $\{\beta_k\}_k$ of parameters satisfying suitable conditions (discussed in detail later), and in [31] the authors establish its convergence over $\mathbb{R}^d$ (cf. [31, Theorem 11]).

While choosing $\alpha_k = 0$ and $\rho_k = 1$ for all $k \in \mathbb{N}$ allows one to recover the “bare” proximal point algorithm from [36], the strength of the above method lies of course in the use of nontrivial parameter sequences. In that context, the computation of y_k in the above scheme, i.e. the so-called “extrapolation step”, incorporates an inertia term as originally considered by Antipin [4], Alvarez [1] as well as Alvarez and Attouch [3] to improve the proximal point algorithm for monotone inclusions, motivated as a discretization of the “heavy ball with friction” system already studied by Polyak [57] (see also [5]). The computation of x_{k+1} , i.e. the over-relaxation, follows a strategy first used by Eckstein and Bertsekas [25] for the proximal point algorithm for monotone inclusions and has the particular benefit that it may indeed speed up the convergence of the method (see e.g. [13, 26]), in particular in combination with inertia terms which was first considered by Alvarez [2]. For convex equilibrium problems, such a relaxed-inertia proximal point type method was, to our knowledge, first considered in [33] (see also the related second order dynamical system studied in [67]).

1.2. The contributions of the present paper and related work. In the present paper, we study the relaxed inertial proximal point method as presented above for suitable strongly quasiconvex equilibrium problems over Hadamard manifolds, that is complete simply connected Riemannian manifolds of nonpositive sectional curvature. In many ways, Hadamard manifolds provide a very suitable but nontrivial framework for convex analysis and optimization in nonlinear contexts, and we refer to [8, 9] and the references therein for discussions in that direction.

In particular, the proximal point algorithm for convex minimization and its extension for inclusion problems for monotone vector fields have been studied deeply over Hadamard manifolds. We refer in particular [28, 48, 60] (see also [7]), next to many others. Extensions of this method for convex equilibrium problems are similarly quite well-studied and can in particular be found in [10, 11, 21] (see also [42]), all of which however use the resolvent of a bifunction in various guises, that is they approach the problem by regularizing the bifunction as a whole. For quasiconvex functions, the literature is largely limited to the setting of minimization problems, as e.g. studied in [58, 59, 60, 61, 62]. In particular, work on problems induced by strongly quasiconvex functions in these contexts is very limited so far and, to our knowledge, only comprises the recent works [56] of the second author and [24] of the authors which, respectively, extend the convergence results for the proximal point type method for strongly quasiconvex

minimization [47] and equilibrium problems [36] obtained in the linear setting to Hadamard spaces, that is complete geodesic metric spaces of nonpositive curvature (see [8, 9, 17]), a class that in particular includes Hadamard manifolds but is neither restricted to finite-dimensionality nor smoothness assumptions.

The present work can thus be seen as an extension of the results obtained in [24, 56] to the setting of [31]. In particular, we rely on the general approach initiated in [24, 56] to also study the present manifold variant of the method from [31], which concretely takes the following form: Over a Hadamard manifold X with Riemannian metric d , we iteratively compute points $\{x_k\}_k$, $\{y_k\}_k$ and $\{z_k\}_k$ via

$$\begin{cases} y_k := \exp_{x_k}(-\alpha_k \exp_{x_k}^{-1} x_{k-1}), \\ z_k \in \operatorname{argmin}_{x \in K} \left\{ f(y_k, x) + \frac{1}{2\beta_k} d^2(y_k, x) \right\}, \\ x_{k+1} := \exp_{z_k}((1 - \rho_k) \exp_{z_k}^{-1} y_k), \end{cases}$$

from given starting points $x_{-1}, x_0 \in K$ and sequences $\{\alpha_k\}_k$, $\{\rho_k\}_k$, $\{\beta_k\}_k$ of parameters as before (see Section 3 for details). Here, $\exp$ is the exponential map of the underlying manifold (see Section 2.1 for details) and, correspondingly, $K \subseteq X$ is now a metrically closed set that contains all geodesic lines determined by any two points of K , representing a suitable analog for affine subspaces in Hadamard manifolds (see Section 3 for details).

Again, for the special choices of $\alpha_k = 0$ and $\rho_k = 1$ for all $k \in \mathbb{N}$, the above method reduces to the “bare” proximal point algorithm previously studied by the authors in [24] over Hadamard spaces (extending [36]). The use of the exponential map to formulate inertia terms via

$$\exp_{x_k}(-\alpha_k \exp_{x_k}^{-1} x_{k-1})$$

as above in the context of Hadamard manifolds goes back at least to [40] (see also [20, 65]), and we crucially rely on the approach taken therein to handle the above extrapolation step in this paper. While one often finds relaxations, that is steps of the form

$$\exp_{z_k}((1 - \rho_k) \exp_{z_k}^{-1} y_k)$$

as above for the choice of $\rho_k \in [0, 1]$, in the literature on optimization methods in Hadamard manifolds (as in e.g. [40, 65], among many others), this does not seem to be true for over-relaxations, that is for the choice of $\rho_k \in [0, 2]$, where the present paper seems to provide the first such treatment, to our knowledge. Perhaps unsurprisingly, we for this rely on similar machinery as needed to handle inertia terms in this nonlinear context.

Our main result is then a convergence theorem for this method over Hadamard manifolds (see Theorems 4.12 and 4.24) under suitable conditions on the parameters and the bifunction (derived from [31], as discussed in detail in Section 3 later). In particular, this extends the convergence result obtained in [31] for the first time to a nonlinear setting. Modeled after the approach of [24, 56], our convergence proofs are quite elementary and in particular effective, contrary to [31], avoiding any kind of infinitary arguments like (weak) compactness. This was not only instrumental for obtaining our results in a nonlinear context but also allows for the formulation of explicit and fast rates of convergence (similar to [24, 56]) for the sequences $\{x_k\}$, $\{y_k\}$ and $\{z_k\}$ generated by the above method. More precisely, we in particular establish the following quantitative result:

$$\forall \varepsilon > 0 \ \forall k \geq \left(\left\lceil \frac{C}{\varepsilon^2} \right\rceil + M(E\varepsilon^2) + 2 \right) (d(x_k, x^*), d(y_k, x^*), d(z_k, x^*) < \varepsilon),$$

where x^* is the (unique) solution to the associated equilibrium problem. The corresponding rate $(\lceil C/\varepsilon^2 \rceil + M(E\varepsilon^2) + 2)$ is in particular very uniform, depending only on a rate of convergence

M for the sum $\sum_{k=0}^{\infty} \alpha_k d^2(x_k, x_{k-1}) < +\infty$ (see Section 4 later on) as well as two constants C, E which can be explicitly calculated only in terms of a few surrounding numerical parameters of the problem. To our knowledge, these are the first effective results for this method, even in the original setting of Euclidean spaces from [31]. Similar to [24], in the course of our convergence proofs, we also show how some semicontinuity assumptions on the bifunction featuring in [31] can be avoided (see Remark 4.14 later on).

While we have chosen to formulate our results over Hadamard manifolds for simplicity, it should be noted that the present approach does not at all depend on the smoothness of the space X or any other differential structure and could in principle be carried out, *mutatis mutandis*, over a suitable class Hadamard spaces that is discussed in detail in Remark 4.15 later on. In particular, this class has no dimensionality restrictions in general and in particular contains Hilbert spaces and some Hilbert manifolds. In that context, the results of the present paper in particular show that (an appropriate formulation of) the method above strongly converges even in these infinite-dimensional spaces, similar to [24, 56]. Further, the quantitative estimates remain exactly the same. A detailed discussion of the Hilbert space case can be found in the master thesis of the first author [23], on which this paper is based.

Lastly, while we focus only on equilibrium problems in this paper, our approach should immediately apply to minimization problems, not only trivially by instantiating the problem and method above, but also in the context of the generalized parameter assumptions used in that context in [30]. We however do not spell this out here.

1.3. Proof mining. The present quantitative results have been obtained using methods from proof mining, a program in mathematical logic that utilizes tools from proof theory, the foundations of mathematics, to classify and extract the computational content of *prima facie* “non-computational” proofs from various areas of core mathematics. For a comprehensive overview of this area and some of the many applications to nonlinear analysis, we refer to the seminal monograph [43] as well as to the survey [44]. As typical for many applied results of proof mining, the present paper is presented without any explicit reference to logic and does not require any such background. We instead refer to the master thesis of the first author [23] for a detailed discussion on the concrete use of logical methods in the context of the present paper.

1.4. Outline of the present paper. The paper is now organized as follows: In Section 2, we give the necessary background on Hadamard manifolds and the related objects as well as notions of convexity and quasiconvexity, together with some basic lemmas needed for the main results. Section 3 carefully introduces the method sketched above in Hadamard manifolds and discusses the central assumptions on the surrounding objects. Finally, Section 4 then provides the corresponding (quantitative) convergence theorems.

2. PRELIMINARIES

2.1. Hadamard manifolds. We now first give the (very few) required preliminaries for Riemannian and Hadamard manifolds. For further exposition beyond the results given here, we refer to [66] (as well as [8, 17]).

In the following, and for the rest of the paper unless stated otherwise, we take X to be a finite dimensional Hadamard manifold, that is a complete simply connected Riemannian manifold of nonpositive sectional curvature, with a Riemannian metric $\langle \cdot, \cdot \rangle$. We denote its tangent space at a point $x \in X$ by $T_x X$ and we denote the norm on $T_x X$ induced by the Riemannian metric by $\|\cdot\|$. The Riemannian distance on X , defined as usual, is denoted by d . It induces the original topology on X and, in particular, (X, d) is a complete metric space.

Given an interval $I \subseteq \mathbb{R}$, a geodesic in X is a map $\gamma : I \rightarrow X$ such that

$$d(\gamma(t), \gamma(s)) = c|t - s|$$

for all $t, s \in I$, where c is called the speed of γ . If $I = [0, 1]$, we say that γ is a normalized geodesic and we say that γ joins $x = \gamma(0)$ and $y = \gamma(1)$. Note that in this case $c = d(x, y)$. If $I = [0, \infty)$, we call γ a geodesic ray and if $I = \mathbb{R}$, we call γ a geodesic line. By the Cartan-Hadamard theorem (cf. [66, Theorem 4.1]), for any two points $x, y \in X$, there exists a unique normalized geodesic joining x to y . We call a subset $K \subseteq X$ convex if with $x, y \in K$ also $\gamma(t) \in K$ for all $t \in [0, 1]$ where γ is the unique normalized geodesic joining x to y .

Given a point $x \in X$ and vector $\xi \in T_x X$, the exponential map is defined by $\exp_x \xi = \gamma_{x,\xi}(1)$ where $\gamma_{x,\xi}$ is the unique normalized geodesic with $\gamma_{x,\xi}(0) = x$ and velocity $\dot{\gamma}_{x,\xi}(0) = \xi$. It can be shown that $\exp_x(t\xi) = \gamma_{x,\xi}(t)$ for all $t \in [0, 1]$. The map $\exp_x : T_x X \rightarrow X$ is a diffeomorphism and in particular has an inverse $\exp_x^{-1} : X \rightarrow T_x X$ where it holds that $d(x, y) = \|\exp_x^{-1} y\|$ for all $x, y \in X$. It follows that if $x \neq y$, then $r(t) = \exp_x(t \exp_x^{-1} y)$ is a geodesic line in X with $r(0) = x$ and $r(1) = y$.

Metrically, (X, d) is a Hadamard space, that is a complete geodesic metric space of nonpositive curvature, also known as a (complete) CAT(0) space (see in particular [8, 17]). As such, its so-called quasi-inner product as introduced by Berg and Nikolaev [12], that is the function

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

satisfies the following analogue (cf. [12, Corollary 3]) of the Cauchy-Schwarz inequality

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

for all $x, y, u, v \in X$, next to the following four further (in particular characterizing, see [12, Proposition 14]) properties for any $x, v, u, w \in X$: $\langle \overrightarrow{xy}, \overrightarrow{xy} \rangle = d^2(x, y)$; $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle = \langle \overrightarrow{uv}, \overrightarrow{xy} \rangle$; $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle = -\langle \overrightarrow{yx}, \overrightarrow{uv} \rangle$; $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle + \langle \overrightarrow{xy}, \overrightarrow{vw} \rangle = \langle \overrightarrow{xy}, \overrightarrow{uw} \rangle$.

Beyond the above, we rely on two key inequalities for the Riemannian distance. The first is an extension of the so-called Bruhat-Tits CN-inequality [18] to geodesics, which in particular characterizes Hadamard spaces:

Lemma 2.1 (cf. [8, Theorem 1.3.3]). *Let $x, y, w \in X$ and, given $t \in [0, 1]$, define $z := \exp_x(t \exp_x^{-1} y)$. Then*

$$d^2(z, w) \leq (1 - t)d^2(x, w) + td^2(y, w) - t(1 - t)d^2(x, y).$$

The next is a kind of dual to the former, formulated for negative values. In particular, it will be key for our handling of the inertia terms as well as the over-relaxations.

Lemma 2.2 (cf. [65, Proposition 9]). *Let $x, y, w \in X$ and, given $t \in [0, 1]$, define $z := \exp_x(-t \exp_x^{-1} y)$. Then*

$$d^2(z, w) \leq (1 + t)d^2(x, w) - td^2(y, w) + t(1 + t)d^2(x, y).$$

While Proposition 9 in [65] states the above as a general and abstract result, it should be noted that it already implicitly appears in the proof of Theorem 1 in [40].

2.2. Quasiconvexity over Hadamard manifolds. Over a Hadamard manifold X , a function $h : X \rightarrow \mathbb{R}$ is called *convex* if

$$(h \circ \gamma)(\lambda) \leq (1 - \lambda)h(\gamma(0)) + \lambda h(\gamma(1))$$

for any geodesic $\gamma : [0, 1] \rightarrow X$ and any $\lambda \in [0, 1]$. Going beyond convex functions, the present paper will focus on the weaker class of quasiconvex functions and their so-called strongly quasiconvex variants. In that context, a function $h : X \rightarrow \mathbb{R}$ is called

(i) *quasiconvex* if

$$(h \circ \gamma)(\lambda) \leq \max\{h(\gamma(0)), h(\gamma(1))\}$$

for any geodesic $\gamma : [0, 1] \rightarrow X$ and any $\lambda \in [0, 1]$,

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

$$(h \circ \gamma)(\lambda) \leq \max\{h(\gamma(0)), h(\gamma(1))\} - \lambda(1 - \lambda) \frac{\gamma}{2} d^2(\gamma(0), \gamma(1))$$

for any geodesic $\gamma : [0, 1] \rightarrow X$ and any $\lambda \in [0, 1]$.

Next to notions of convexity, 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 $x \in X$.

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

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

for $\beta > 0$ and $x \in X$. As remarked in the introduction, $\text{Prox}_{\beta h}$ is no longer a singleton for strongly quasiconvex functions but is non-empty under suitable assumptions (see also Remark 3.3 later). The main result for the proximal map that we utilize beyond that is the following:

Lemma 2.3 (cf. [56, 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*

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

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

The above result, which extends [47, Proposition 7], was derived in [56] already over Hadamard spaces, and hence holds over Hadamard manifolds as well.

3. RELAXED-INERTIAL PROXIMAL POINT ALGORITHMS ON HADAMARD MANIFOLDS

We now begin by defining the main method studied in this paper for solving the equilibrium problem (EP) over a Hadamard manifold X , a suitable subspace $K \subseteq X$ and a given bifunction $f : K \times K \rightarrow \mathbb{R}$. We use $S(K, f)$ to denote the associated solution set. As discussed in the introduction, we base our method on that introduced and studied in the work of Grad, Lara and Marcavillaca [31] for the associated problem in Euclidean spaces, and so first recall that method and its associated parameters here in detail.

Over $X = \mathbb{R}^d$, let $K \subseteq X$ be an affine subspace. Fix $\alpha, \rho \in [0, 1]$ and starting points $x_0, x_{-1} \in K$ as well as sequences of parameters $\{\beta_k\}_k \subseteq (0, \infty)$, $\{\alpha_k\}_k \subseteq [0, \alpha]$ and $\{\rho_k\}_k \subseteq [1 - \rho, 1 + \rho]$. Iteratively, the method introduced in [31] now calculates a new point $x_{k+1} \in K$ from previous points $x_k, x_{k-1} \in K$ as follows:

$$(\%) \quad \begin{cases} y_k := x_k + \alpha_k(x_k - x_{k-1}), \\ z_k \in \operatorname{argmin}_{x \in K} \left\{ f(y_k, x) + \frac{1}{2\beta_k} \|y_k - x\|^2 \right\}, \\ x_{k+1} := (1 - \rho_k)y_k + \rho_k z_k. \end{cases}$$

The first step y_k , commonly called the exploration step, incorporates the so-called inertia term $\alpha_k(x_k - x_{k-1})$ already discussed in the introduction. The second step z_k , commonly called the proximal step, then produces a new point from this y_k by utilizing a proximal regularization of the bifunction in its second argument, with the first argument fixed to y_k , akin to the previous proximal method studied by Iusem and Lara [36] over Euclidean spaces (and the present authors over Hadamard spaces [24]). The third step x_{k+1} , commonly called the (over)relaxation step, combines the previous two.

In [31], Grad, Lara and Marcavillaca study the above algorithm for a specific class of non-convex bifunctions, characterized by the following conditions (A1) – (A5):

(A1) For every $x \in K$, the function $f(x, \cdot)$ is lsc, and for every $y \in K$, 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 $x, y \in K$.

(A3) f is lsc (jointly in both arguments).

(A4) For every $x \in K$, 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

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

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

Further, they rely on the following crucial assumptions on the associated parameter sequences: For $\{\beta_k\}_k$, it is assumed that there exists an $e > 0$ such that

(C1) $\frac{1}{\gamma - 8\eta} < \beta_k < e \leq \frac{1}{4\eta}$ for all $k \in \mathbb{N}$,

and, relative to this e , they further require that

(C2) $0 < 1 - \rho \leq \rho_k \leq 1 + \rho$ with $0 \leq \rho \leq 1 - 4\eta e$.

In both cases, γ and η are of course as in (A4) and (A5) respectively. Similar to [31, Remark 8], we remark that if one considers $e = \frac{1}{4\eta}$, (C2) implies $\rho = 0$ and so $\rho_k = 1$ for all $k \in \mathbb{N}$. In this case, (C1) becomes $\frac{1}{\gamma - 8\eta} < \beta_k < \frac{1}{4\eta}$ for all $k \in \mathbb{N}$ as employed in the context of the proximal point method studied in [36] (as well as [24]). In any case, (C1) in this way in particular implies that $12\eta < \gamma$, which featured as assumption (A6) in [36] (as well as [24]), and we will likewise often use this property in the present paper. Further as in [36], we note that the standard assumption

(A0) $f(x, x) = 0$ for all $x \in X$

can be derived from (A2) and (A5), but we will later also use it individually.

In the setting of (A1) – (A5) and (C1) – (C2), Grad, Lara and Marcavillaca [31] show that the above method (%) is well-defined and that it converges to (the unique) solution of the associated equilibrium problem under a suitable decaying assumption on the inertia terms:

Theorem 3.1 (cf. [31, Theorem 11]). *Let $X = \mathbb{R}^d$ and let $K \subseteq X$ be an affine subspace. Suppose that $f : K \times K \rightarrow \mathbb{R}$ satisfies (Ai) for $i = 1, 2, 3, 4, 5$. Fix $\alpha, \rho \in [0, 1)$ and $x_0, x_{-1} \in K$ as well as $\{\beta_k\}_k \subseteq (0, \infty)$, $\{\alpha_k\}_k \subseteq [0, \alpha]$ and $\{\rho_k\}_k \subseteq [1 - \rho, 1 + \rho]$ such that (C1) and (C2) hold, and let $\{x_k\}_k$, $\{y_k\}_k$ and $\{z_k\}_k$ be the sequences generated by (%). If*

$$\sum_{k=0}^{\infty} \alpha_k \|x_k - x_{k-1}\|^2 < +\infty,$$

then the sequences $\{x_k\}_k$, $\{y_k\}_k$ and $\{z_k\}_k$ converge to the same limit x^ , where $\{x^*\} = S(K, f)$.*

Remark 3.2. In [31], assumption (A1) only refers to the requirement that $f(\cdot, y)$ is usc for all $y \in K$. The assumption that $f(x, \cdot)$ is lsc for every $x \in K$ is denoted by (A1') in [31] and is treated separately as it naturally follows from (A3). However, as we will in particular show in this paper that the assumption (A3) can in certain contexts be eliminated from the above convergence theorem and its generalization to Hadamard manifolds (see in particular Remark 4.14 later on), we prefer the above naming convention, which follows that of Iusem and Lara [36] (as well as [24]). Further, we want to note that a slight weakening of the assumptions on the parameters is discussed in [32, Remark 87] and while we deem it quite likely that our present results can be extended to this weakened assumption, we here focus on the original assumptions from [31] for simplicity.

Lifted to a general Hadamard manifold X , the above method (%) now takes the following form: First, to set up the problem, we fix a set $K \subseteq X$ which is metrically closed and such that the complete geodesic line determined by any two points of K lies in K , that is

$$\exp_x(t \exp_x^{-1} y) \in K \text{ for all } t \in \mathbb{R} \text{ and } x, y \in K.$$

In particular, K is a convex set and so K is a Hadamard space (see e.g. [8]). Over linear spaces X , sets K as above are exactly the affine subspaces and we will find that the above closure conditions are the right conditions to generalize the method from [31].

Over such a K , we then consider a bifunction $f : K \times K \rightarrow \mathbb{R}$ satisfying assumptions among (A1) – (A5), with (A5) now formulated using the metric via

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

for all $x, y, z \in K$, using a fixed $\eta > 0$. We still denote the set of equilibrium points of f over K , that is the set of all $x^* \in K$ such that $f(x^*, y) \geq 0$ for all $y \in K$, by $S(K, f)$. As K is a Hadamard space, [24, Theorem 3.9] guarantees that $S(K, f) \neq \emptyset$ whenever f satisfies (A0), (A1), (A2) and (A4).

We now move on to the method. Fixing parameters $\alpha, \rho \in [0, 1)$ and $x_0, x_{-1} \in K$ as well as $\{\beta_k\}_k \subseteq (0, \infty)$, $\{\alpha_k\}_k \subseteq [0, \alpha]$ and $\{\rho_k\}_k \subseteq [1 - \rho, 1 + \rho]$ as above, the method iteratively calculates a new point $x_{k+1} \in K$ from the previous points $x_k, x_{k-1} \in K$ as follows:

$$(RIPPA) \quad \begin{cases} y_k := \exp_{x_k}(-\alpha_k \exp_{x_k}^{-1} x_{k-1}), \\ z_k \in \operatorname{argmin}_{x \in K} \left\{ f(y_k, x) + \frac{1}{2\beta_k} d^2(y_k, x) \right\}, \\ x_{k+1} := \exp_{z_k}((1 - \rho_k) \exp_{z_k}^{-1} y_k). \end{cases}$$

Again, as outlined in the introduction, the first step y_k represents the exploration step involving the inertia term in a manifold. The second step represents the proximal step as usual and the last represents an (over)relaxation in manifolds, combining the previous two. Using the fact that $\exp_x(\xi) = x + \xi$ and $\exp_x^{-1}(y) = y - x$ over Euclidean spaces, it can be easily verified that the method (RIPPA) coincides with (%) in such a setting.

The first observation we want to make is that, like (%) in [31], also the above (RIPPA) is well-defined in the context of the parameter restrictions given by (C1) and (C2). This crucially relies on the fact that

$$\operatorname{Prox}_{\beta_k f(y_k, \cdot)} := \operatorname{argmin}_{x \in K} \left\{ f(y_k, x) + \frac{1}{2\beta_k} d^2(y_k, x) \right\} \neq \emptyset,$$

which follows rather immediately from corresponding results obtained by the authors in [24] over Hadamard spaces. We collect this in the following remark.

Remark 3.3. By [24, Proposition 4.4], given a bifunction $f : Y \times Y \rightarrow \mathbb{R}$ on a Hadamard space Y which is quasiconvex and lsc in its right argument and satisfies (A5), the proximal map

$\text{Prox}_{\beta f(z, \cdot)}$ is non-empty for any $z \in X$ and $\beta \leq \frac{1}{4\eta}$ with η as in (A5). As K is in particular a Hadamard space, this result also applies here, so that (RIPPA) is well-defined whenever (A1), (A4), (A5) and (C1) hold: For $x_k, x_{k-1} \in K$, we have $y_k \in K$ since K is closed under complete geodesic lines. The above result yields that $z_k \in K$ is well-defined. We get $x_{k+1} \in K$, again using that K is closed under complete geodesic lines.

Remark 3.4. As utilized in [31], the convenient stopping criterion $y_k = z_k$ applies in the context (%) as whenever this equality holds, one already knows that $\{y_k\} = S(K, f)$. While RIPPA and (%) above were formulated without such a condition (contrary to [31]) to focus purely on the (infinite) sequences generated by these methods, we want to highlight that a similar stopping criterion also applies in the nonlinear context of (RIPPA). Indeed, under (A0) and (A4) it follows from [24, Proposition 4.6] that

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

for any $\beta > 0$, again using that K is a Hadamard space.

4. QUANTITATIVE CONVERGENCE THEOREMS

4.1. Key assumptions. We now move to the main results of the present paper, that is quantitative convergence theorems for the method (RIPPA) introduced above over general Hadamard manifolds and for suitable bifunctions. In order to derive our effective rates of convergence, we will need to consider one of two distinct *quantitative* variants of the condition (C1).

Concretely, one of the two conditions we consider on the one hand strengthens the property $\beta_k < e$ to $\beta_k \leq e - \Theta$ for some $\Theta > 0$ with $e \leq \frac{1}{4\eta}$. However, the condition we consider simultaneously weakens the lower bound from $\frac{1}{\gamma-8\eta} < \beta_k$ to $\frac{1}{\gamma-8\eta} \leq \beta_k$. Together, we arrive at the following:

$(C1)^l$ $\frac{1}{\gamma-8\eta} \leq \beta_k \leq e - \Theta$ for all $k \in \mathbb{N}$, where $e \leq \frac{1}{4\eta}$ and $\Theta > 0$ is some (fixed) constant.

The dual to this arises by instead making the lower bound $\frac{1}{\gamma-8\eta} < \beta_k$ uniformly strict by the introduction of a corresponding additional constant θ , while the upper bound $\beta_k < e \leq \frac{1}{4\eta}$ is relaxed to a non-strict variant, that is:

$(C1)^u$ $\frac{1}{\gamma-8\eta} + \theta \leq \beta_k \leq e$ for all $k \in \mathbb{N}$, where $e \leq \frac{1}{4\eta}$ and $\theta > 0$ is some (fixed) constant.

Recall that, as discussed before in the context of assumption (C1), naturally also these extended assumptions $(C1)^l$ or $(C1)^u$ imply $12\eta < \gamma$, which we utilize repeatedly throughout. Also recall that, as mentioned in Remark 3.3, the iteration (RIPPA) is well-defined in the context of (A1), (A4), (A5) and (C1) but by the argument given therein, the same holds for (C1) replaced by $(C1)^l$ or $(C1)^u$.

Our quantitative results given later depend on assuming either $(C1)^l$ or $(C1)^u$. We are not able to give similar quantitative results for the original condition (C1), nor are we able to relax any of the two assumptions $(C1)^l$ or $(C1)^u$ any further, e.g. to

$$\frac{1}{\gamma-8\eta} \leq \beta_k \leq e \leq \frac{1}{4\eta} \text{ for all } k \in \mathbb{N}.$$

This should in particular be compared to [24], where in the context of similar quantitative results and extensions to nonlinear spaces for the proximal point type method *without* inertia terms or over-relaxations featuring in [36] (which arises from (RIPPA) by setting $\alpha, \rho = 0$), it was possible to weaken the condition $\{\beta_k\}_k \subseteq (\frac{1}{\gamma-8\eta}, \frac{1}{4\eta})$ to $\{\beta_k\}_k \subseteq [\frac{1}{\gamma-8\eta}, \frac{1}{4\eta}]$. In fact, the approach taken later for deriving our (quantitative) convergence result under the condition

$(C1)^u$ relies on the machinery developed in [24] (and further extensions thereof given in this paper, see in particular Lemma 4.17 later on) for treating this generalization.

Practically, this however seems to be of rather little importance as $\{\beta_k\}_k$ is an auxiliary sequence which can be chosen freely, so that the above $(C1)^l$ or $(C1)^u$ are in a way as easy to fulfill as $(C1)$ itself.

Similar to [24], in the course of our convergence analysis, we will assume the well-definedness of (RIPPA) outright, and this, paired with the assumption of the existence of a solution to (EP), which is guaranteed under (A0), (A1), (A2) and (A4), will allow us to discharge any continuity assumptions like (A1) or (A3) (see Remark 4.14 later on). In particular, under the assumption of either $(C1)^l$ or $(C1)^u$, (A3) can be completely removed. In that way, unless stated otherwise, we from now on fix $\alpha, \rho \in [0, 1)$ and $x_0, x_{-1} \in K$ as well as $\{\beta_k\}_k \subseteq (0, \infty)$, $\{\alpha_k\}_k \subseteq [0, \alpha]$ and $\{\rho_k\}_k \subseteq [1 - \rho, 1 + \rho]$. Further, we suppose that (RIPPA) is well-defined and let $\{x_k\}_k, \{y_k\}_k$ and $\{z_k\}_k$ be its generated sequences. Further, we assume that $S(K, f) \neq \emptyset$ and denote its unique element by x^* .

4.2. Basic lemmas. In similarity to [24] (and by relation also [31, 36], although this is left rather implicit therein), a key step for our convergence analysis involves establishing the *Fejér monotonicity* of the method. However, contrary to [24, 36], we here rely on establishing the *quasi-Fejér monotonicity* of the method (RIPPA) (already featuring implicitly in [31] for the method (%)), that is we show that

$$\forall k, l \in \mathbb{N} \left(d^2(x_{k+l}, x^*) \leq a \cdot d^2(x_k, x^*) + \sum_{i=k}^{\infty} e_i \right)$$

where $a > 0$ and $\{e_i\}_i$ is a sequence in $(0, \infty)$ with $\sum_{i=0}^{\infty} e_i < +\infty$.

Towards this property, we rely on the abstract Lemma 4.1, which is a slightly corrected version (see Remark 4.2 below) of the technical [31, Lemma 2, (a)], which itself was essentially established already in [3, Theorem 2.1]. As the proof just amounts to some simple (but slightly tedious) calculations and inductions, we omit it.

Lemma 4.1 (essentially [31, Lemma 2, (a)]). *Let $\{\varphi_k\}_k, \{s_k\}_k, \{\delta_k\}_k \subseteq [0, \infty)$, let $\varphi_{-1} \in [0, \infty)$ and let $\alpha \in [0, 1)$ with $\{\alpha_k\}_k \subseteq [0, \alpha]$. Assume that*

$$(*) \quad \varphi_{k+1} - \varphi_k + s_{k+1} \leq \alpha_k(\varphi_k - \varphi_{k-1}) + \delta_k \text{ for all } k \in \mathbb{N}.$$

Then

$$\forall k \in \mathbb{N} \quad \forall n \geq k \left(\varphi_n \leq \frac{1}{1-\alpha} \varphi_k + \frac{1}{1-\alpha} \sum_{i=k}^{n-1} \delta_i - \sum_{i=k+1}^n s_i \right).$$

Remark 4.2. Note that the statement of [31, Lemma 2, (a)] is slightly wrong: Under the assumptions of Lemma 4.1, the authors of [31] claim that for all $k \geq 1$:

$$\varphi_k + \sum_{i=1}^k s_i \leq \varphi_0 + \frac{1}{1-\alpha} \sum_{i=0}^{k-1} \delta_i,$$

omitting the factor $\frac{1}{1-\alpha}$ in front of φ_0 as in the above Lemma 4.1. However, this factor can in general not be avoided. Indeed, consider e.g. $\varphi_k := 1$, $\alpha_k := 1/2$ for all $k \in \mathbb{N}$ and $s_k := \delta_k := 1/3$ if $k = 0$ and $s_k := \delta_k := 1/k^2$ otherwise. Then $\varphi_{k+1} - \varphi_k + s_k \leq \alpha_k(\varphi_k - \varphi_{k-1}) + \delta_k$ holds trivially for all $k \in \mathbb{N}$, but $\varphi_1 + s_1 \not\leq \varphi_0 + \frac{1}{1-\alpha} \delta_0$.

Rearranging the inequality obtained in Lemma 4.1, we in particular get that $\sum_{i=1}^{\infty} s_i < +\infty$, so that $\lim_{i \rightarrow \infty} s_i = 0$. The following lemma provides a quantitative variant of the corresponding (weaker) conclusion that $\liminf_{i \rightarrow \infty} s_i = 0$, which we require for the derivation of our rates later.

Lemma 4.3. *Let $\{\varphi_k\}_k, \{s_k\}_k, \{\delta_k\}_k \subseteq [0, \infty)$, let $\varphi_{-1} \in [0, \infty)$ and let $\alpha \in [0, 1)$ with $\{\alpha_k\}_k \subseteq [0, \alpha]$. Assume that $(*)$ as in Lemma 4.1 holds, and take $b \in [0, \infty)$ such that $b \geq \varphi_0$. Assume further that there exists $S \in (0, \infty)$ such that $\sum_{k=0}^{\infty} \delta_k \leq S$. Then:*

$$\forall \varepsilon > 0 \quad \forall n \in \mathbb{N} \quad \exists k \in \mathbb{N} \left(n \leq k \leq \left\lceil \frac{b+S}{(1-\alpha)\varepsilon} \right\rceil + n + 1 \text{ and } s_{k+1} < \varepsilon \right).$$

Proof. By Lemma 4.1, we get

$$\forall n \in \mathbb{N} \left(\sum_{i=1}^n s_i \leq \frac{1}{1-\alpha} \varphi_0 + \frac{1}{1-\alpha} \sum_{i=0}^{n-1} \delta_i \right)$$

so that taking $n \rightarrow \infty$ yields

$$\sum_{i=1}^{\infty} s_i \leq \frac{1}{1-\alpha} \varphi_0 + \frac{1}{1-\alpha} \sum_{i=0}^{\infty} \delta_i \leq \frac{b+S}{1-\alpha}.$$

Suppose now for a contradiction that the claim fails, i.e. there are $\varepsilon > 0$ and $n \in \mathbb{N}$ such that $s_{k+1} \geq \varepsilon$ for all $k \in \mathbb{N}$ with $n \leq k \leq \left\lceil \frac{b+S}{(1-\alpha)\varepsilon} \right\rceil + n + 1 =: l$. Then however

$$\sum_{i=1}^{\infty} s_i \geq \sum_{i=n}^l s_{i+1} \geq \left(\left\lceil \frac{b+S}{(1-\alpha)\varepsilon} \right\rceil + 1 \right) \varepsilon > \frac{b+S}{1-\alpha},$$

which is a contradiction to the above. $\square$

Before we apply Lemma 4.3 to the sequences generated by (RIPPA), we of course need to show that its conditions, in particular $(*)$ as in Lemma 4.1, will hold in our case. The next results are towards this.

The following proposition then extends [31, Proposition 6]. In difference to that result, we emphasize the slightly more general $(+)_1$ and $(+)_2$ below as we will later require varying λ 's akin to the approach taken in [24], whereas only their instantiations with $\lambda = 1/2$ are considered explicitly in [31, Proposition 6].

Lemma 4.4 (extending [31, Proposition 6]). *Let $f : K \times K \rightarrow \mathbb{R}$ be such that (A2), (A4) and (A5) hold. Suppose that $x^* \in S(K, f)$. Then, for every $k \in \mathbb{N}$, at least one of the following is true:*

$$\begin{aligned} (+)_1 \quad d^2(x_{k+1}, x^*) &\leq d^2(y_k, x^*) - \left(\frac{2 - \rho_k}{\rho_k} \right) d^2(x_{k+1}, y_k) \\ &\quad - \rho_k ((1 - \lambda)\gamma\beta_k - \lambda) d^2(z_k, x^*) \end{aligned}$$

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

$$\begin{aligned} (+)_2 \quad d^2(x_{k+1}, x^*) &\leq d^2(y_k, x^*) - \left(\frac{2 - \rho_k - \frac{2\beta_k\eta}{\lambda}}{\rho_k} \right) d^2(x_{k+1}, y_k) \\ &\quad - \rho_k \left((1 - \lambda)\gamma\beta_k - \lambda - \frac{2\eta\beta_k}{\lambda} \right) d^2(z_k, x^*) \end{aligned}$$

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

Proof. By definition, we have $x_{k+1} := \exp_{z_k}((1 - \rho_k) \exp_{z_k}^{-1} y_k)$. Then it holds that

$$(\circ)_1 \quad d^2(x_{k+1}, x^*) \leq \rho_k d^2(z_k, x^*) + (1 - \rho_k) d^2(y^k, x^*) - \rho_k(1 - \rho_k) d^2(z_k, y_k).$$

Indeed, either $\rho_k \in [0, 1]$, where then by Lemma 2.1 we get $(\circ)_1$ immediately. Or $\rho_k \in [1, 2]$, then $\rho_k = 1 + \rho'_k$ for some $\rho'_k \in [0, 1]$ so that $x_{k+1} = \exp_{z_k} - \rho'_k \exp_{z_k}^{-1} y_k$. Hence, in that case, Lemma 2.2 yields

$$\begin{aligned} d^2(x_{k+1}, x^*) &\leq (1 + \rho'_k) d^2(z_k, x^*) - \rho'_k d^2(y_k, x^*) + \rho'_k(1 + \rho'_k) d^2(z_k, y_k) \\ &= \rho_k d^2(z_k, x^*) + (1 - \rho_k) d^2(y_k, x^*) - \rho_k(1 - \rho_k) d^2(z_k, y_k), \end{aligned}$$

that is we also get $(\circ)_1$ in this case. Rewriting $(\circ)_1$, we get

$$d^2(x_{k+1}, x^*) \leq d^2(y^k, x^*) + \rho_k (d^2(z_k, x^*) - d^2(y^k, x^*)) - \rho_k(1 - \rho_k) d^2(z_k, y_k).$$

Further, by definition of the quasi-inner product, we get

$$2\langle \overrightarrow{y_k z_k}, \overrightarrow{x^* y_k} \rangle + d^2(z_k, y_k) = d^2(z_k, x^*) - d^2(y_k, x^*)$$

so that, combining both, we obtain

$$(\circ)_2 \quad d^2(x_{k+1}, x^*) \leq d^2(y^k, x^*) + 2\rho_k \langle \overrightarrow{y_k z_k}, \overrightarrow{x^* y_k} \rangle + \rho_k^2 d^2(z_k, y_k).$$

Lemma 2.3, applied to $f(y_k, \cdot)$ with $y = x^*$, yields

$$(\circ)_3 \quad f(y_k, z_k) - \max\{f(y_k, z_k), f(y_k, x^*)\} \leq \frac{\lambda}{2} \left(\frac{\lambda}{\beta_k} - \gamma + \lambda\gamma \right) d^2(z_k, x^*) + \frac{\lambda}{\beta_k} \langle \overrightarrow{y_k z_k}, \overrightarrow{z_k x^*} \rangle$$

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

Case 1: $f(y_k, z_k) \geq f(y_k, x^*)$. From $(\circ)_3$, we obtain

$$\frac{\lambda}{\beta_k} \langle \overrightarrow{y_k z_k}, \overrightarrow{x^* z_k} \rangle \leq \frac{\lambda}{2} \left(\frac{\lambda}{\beta_k} - \gamma + \lambda\gamma \right) d^2(z_k, x^*)$$

for any $\lambda \in [0, 1]$. Then, proceed as in the next case from $(\circ)_4$, but where $\eta = 0$, to obtain $(+)_1$.

Case 2: $f(y_k, z_k) < f(y_k, x^*)$. Using (A5), we obtain from $(\circ)_3$ that

$$\begin{aligned} 0 &\leq \frac{\lambda}{2} \left(\frac{\lambda}{\beta_k} - \gamma + \lambda\gamma \right) d^2(z_k, x^*) + \frac{\lambda}{\beta_k} \langle \overrightarrow{y_k z_k}, \overrightarrow{z_k x^*} \rangle + f(y_k, x^*) - f(y_k, z_k) \\ &\leq \frac{\lambda}{2} \left(\frac{\lambda}{\beta_k} - \gamma + \lambda\gamma \right) d^2(z_k, x^*) + \frac{\lambda}{\beta_k} \langle \overrightarrow{y_k z_k}, \overrightarrow{z_k x^*} \rangle + f(z_k, x^*) + \eta (d^2(z_k, y_k) + d^2(z_k, x^*)). \end{aligned}$$

Using the pseudomonotonicity of f and since $x^* \in S(K, f)$, we get $f(z_k, x^*) \leq 0$ and so obtain

$$(\circ)_4 \quad \frac{\lambda}{\beta_k} \langle \overrightarrow{y_k z_k}, \overrightarrow{x^* z_k} \rangle \leq \frac{\lambda}{2} \left(\frac{\lambda}{\beta_k} - \gamma + \lambda\gamma + \frac{2\eta}{\lambda} \right) d^2(z_k, x^*) + \eta d^2(z_k, y_k)$$

for any $\lambda \in (0, 1]$. But then, using $\langle \overrightarrow{y_k z_k}, \overrightarrow{x^* z_k} \rangle = \langle \overrightarrow{y_k z_k}, \overrightarrow{x^* y_k} \rangle + d^2(y_k, z_k)$, we obtain

$$\langle \overrightarrow{y_k z_k}, \overrightarrow{x^* y_k} \rangle \leq \left(\frac{\eta\beta_k}{\lambda} - 1 \right) d^2(z_k, y_k) + \frac{1}{2} \left(\lambda - \gamma\beta_k + \lambda\gamma\beta_k + \frac{2\eta\beta_k}{\lambda} \right) d^2(z_k, x^*).$$

Using $(\circ)_2$, we obtain

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

As $\gamma(t) := \exp_{z_k}(t \exp_{z_k}^{-1} y_k)$ is a geodesic line, we have

$$d(x_{k+1}, y_k) = d(\gamma(1 - \rho_k), \gamma(1)) = \rho_k d(z_k, y_k),$$

which yields that $d(z_k, y_k) = \frac{1}{\rho_k} d(x_{k+1}, y_k)$. Inserting this into the above, we get

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

which is $(+)_2$. $\square$

From this, we almost immediately obtain the following lemma, which essentially is a nonlinear variant of [31, Corollary 9] with its case distinction resolved.

Lemma 4.5 (extending [31, Corollary 9]). *Let $f : K \times K \rightarrow \mathbb{R}$ be such that (A2), (A4) and (A5) hold. Suppose that either among $(C1)^l$ or $(C1)^u$ hold. Suppose further that $x^* \in S(K, f)$. Then for every $k \in \mathbb{N}$:*

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

Proof. From Lemma 4.4 we know that at least one of $(+)_1$ or $(+)_2$ holds. Taking $\lambda = \frac{1}{2}$, the result then follows with the fact that, as $\beta_k \geq (1/(\gamma - 8\eta)) \geq 0$ by either among $(C1)^l$ or $(C1)^u$,

$$\frac{\rho_k(\gamma\beta_k - 1)}{2} \geq \frac{\rho_k(\gamma\beta_k - 1 - 8\eta\beta_k)}{2} = \frac{\rho_k(\beta_k(\gamma - 8\eta) - 1)}{2} \geq 0$$

and that further

$$\frac{2 - \rho_k}{\rho_k} \geq \frac{2 - \rho_k - 4\beta_k\eta}{\rho_k}$$

for all $k \in \mathbb{N}$. $\square$

This implies the following proposition, which we will use in our main theorem to argue that we are in the setting of Lemmas 4.1 and 4.3.

Lemma 4.6 (extending [31, Proposition 10]). *Let $f : K \times K \rightarrow \mathbb{R}$ be such that (A2), (A4) and (A5) hold. Suppose that either among $(C1)^l$ or $(C1)^u$ hold. Suppose further that $x^* \in S(K, f)$. Set $\varphi_k := d^2(x_k, x^*)$ for every $k \in \mathbb{N}$. Then for every $k \in \mathbb{N}$:*

$$\varphi_{k+1} - \varphi_k + \left(\frac{2 - \rho_k - 4\eta\beta_k}{\rho_k} \right) d^2(x_{k+1}, y_k) \leq \alpha_k(\varphi_k - \varphi_{k-1}) + (\alpha_k^2 + \alpha_k) d^2(x_k, x_{k-1}).$$

Proof. By definition, we have $y_k := \exp_{x_k}(-\alpha_k \exp_{x_k}^{-1} x_{k-1})$. Using Lemma 2.2, we get

$$\begin{aligned} d^2(y_k, x^*) &\leq d^2(x_k, x^*) + \alpha_k(d^2(x_k, x^*) - d^2(x_{k-1}, x^*)) + \alpha_k(1 + \alpha_k)d^2(x_k, x_{k-1}) \\ &= \varphi_k + \alpha_k(\varphi_k - \varphi_{k-1}) + \alpha_k(1 + \alpha_k)d^2(x_k, x_{k-1}). \end{aligned}$$

Combined with Lemma 4.5, this yields the result. $\square$

As an application of Lemma 4.1, we now show that $\{x_k\}_k$ is quasi-Fejér monotone.

Lemma 4.7. *Let $f : K \times K \rightarrow \mathbb{R}$ be such that (A2), (A4) and (A5). Suppose that (C2) and either among $(C1)^l$ or $(C1)^u$ hold. Suppose further that $x^* \in S(K, f)$. Then:*

$$\forall k \in \mathbb{N} \quad \forall n \geq k \quad \left(d^2(x_n, x^*) \leq \frac{1}{1 - \alpha} d^2(x_k, x^*) + \frac{2}{1 - \alpha} \sum_{i=k}^{n-1} \alpha_i d^2(x_i, x_{i-1}) \right).$$

Proof. Define $\varphi_k := d^2(x_k, x^*)$ and $\delta_k := (\alpha_k^2 + \alpha_k)d^2(x_k, x_{k-1})$ for all $k \in \mathbb{N}$. Observe that, using Lemma 4.6, condition $(*)$ in Lemma 4.1 holds with

$$s_{k+1} := \left(\frac{2 - 4\eta\beta_k - \rho_k}{\rho_k} \right) d^2(x_{k+1}, y_k)$$

as

$$\frac{2 - 4\eta\beta_k - \rho_k}{\rho_k} \stackrel{(C2)}{\geq} \frac{1 - 4\eta\beta_k - \rho}{1 + \rho} \stackrel{\beta_k \leq e}{\geq} \frac{1 - 4\eta e - \rho}{1 + \rho} \stackrel{(C2)}{\geq} 0.$$

Hence, using that $(\alpha_k^2 + \alpha_k) \leq 2\alpha_k$, Lemma 4.1 yields the result. $\square$

4.3. Convergence under condition $(C1)^l$. We now establish the convergence of the method (RIPPA) under the condition $(C1)^l$, that is under the assumption that there exists a $\Theta > 0$ such that $\frac{1}{\gamma - 8\eta} \leq \beta_k \leq e - \Theta$ for all $k \in \mathbb{N}$, where $e \leq \frac{1}{4\eta}$. The first result we need in that vein is the following lower bound on a key quantity featuring in later arguments.

Lemma 4.8. *Suppose that $(C1)^l$ and $(C2)$ hold. Then, for any $k \in \mathbb{N}$:*

$$(2 - 4\eta\beta_k - \rho_k)\rho_k \geq \frac{16\eta^2\Theta}{(\gamma - 8\eta)} > 0.$$

Proof. Using $(C1)^l$ and $(C2)$, we have

$$\begin{aligned} (2 - 4\eta\beta_k - \rho_k)\rho_k &\geq (1 - 4\eta\beta_k - \rho)(1 - \rho) \geq (4\eta e - 4\eta\beta_k)4\eta e \\ &\geq (4\eta e - 4\eta(e - \Theta))4\eta e = 16\eta^2\Theta e. \end{aligned}$$

As we have $e \geq 1/(\gamma - 8\eta)$, and as $\eta, \Theta > 0$, this implies the result. $\square$

Essential for our quantitative convergence theorem for (RIPPA) under $(C1)^l$ is an effective estimate for the asymptotic behavior of $d(z_k, y_k)$, which will here take the form of an explicit bound $\iota : (0, \infty) \times \mathbb{N} \rightarrow \mathbb{N}$ such that for any $\varepsilon > 0$ and any $n \in \mathbb{N}$, there exists a $k \in \mathbb{N}$ with $n \leq k \leq \iota(\varepsilon, n)$ and $d(z_k, y_k) < \varepsilon$. The following result derives an explicit instantiation of such an ι , by giving a quantitative analysis of [31, Theorem 11, (a)] in the context of $(C1)^l$.

Lemma 4.9. *Let $f : K \times K \rightarrow \mathbb{R}$ be such that $(A2)$, $(A4)$ and $(A5)$ hold. Suppose that $(C1)^l$ and $(C2)$ hold. Suppose further that $x^* \in S(K, f)$. Let $S \in (0, \infty)$ be such that*

$$\sum_{k=0}^{\infty} \alpha_k d^2(x_k, x_{k-1}) \leq S.$$

Then

$$\forall \varepsilon > 0 \quad \forall n \in \mathbb{N} \quad \exists k \in \mathbb{N} \left(n \leq k \leq \left\lceil \frac{(b + 2S)(\gamma - 8\eta)}{16\eta^2\Theta(1 - \alpha)\varepsilon^2} \right\rceil + n + 1 \text{ and } d(z_k, y_k) < \varepsilon \right),$$

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

Proof. Define $\varphi_k := d^2(x_k, x^*)$, $\delta_k := (\alpha_k^2 + \alpha_k)d^2(x_k, x_{k-1})$ and

$$s_{k+1} := \left(\frac{2 - 4\eta\beta_k - \rho_k}{\rho_k} \right) d^2(x_{k+1}, y_k).$$

Note that it follows from $\alpha_k \in [0, 1)$ that

$$\sum_{k=0}^{\infty} \delta_k = \sum_{k=0}^{\infty} (\alpha_k^2 + \alpha_k)d^2(x_k, x_{k-1}) \leq 2 \sum_{k=0}^{\infty} \alpha_k d^2(x_k, x_{k-1}) \leq 2S.$$

Using Lemma 4.6, we get that Lemma 4.3 applies, so that for any $\varepsilon > 0$ and $n \in \mathbb{N}$, we get

$$\exists k \in \mathbb{N} \left(n \leq k \leq \left\lceil \frac{b + 2S}{(1 - \alpha)\varepsilon} \right\rceil + n + 1 \text{ and } \left(\frac{2 - 4\eta\beta_k - \rho_k}{\rho_k} \right) d^2(x_{k+1}, y_k) < \varepsilon \right).$$

As in Lemma 4.4, we have $\rho_k d(z_k, y_k) = d(x_{k+1}, y_k)$, so that Lemma 4.8 implies that

$$\left(\frac{16\eta^2\Theta}{(\gamma - 8\eta)} \right) d^2(z_k, y_k) \leq \left(\frac{2 - 4\eta\beta_k - \rho_k}{\rho_k} \right) d^2(x_{k+1}, y_k).$$

Combined with the above, this yields the result. $\square$

Remark 4.10. If in the above proposition, we impose the stronger property that

$$\sum_{k=0}^{\infty} d^2(x_k, x_{k-1}) \leq S'$$

for some $S' \in (0, \infty)$, the resulting bound becomes simpler and can be derived in a more direct way (similar to the proof of [31, Corollary 20]). To outline the corresponding construction, assume that (C2) and either among $(C1)^l$ or $(C1)^u$ hold. It is easy to see that for all $n, m \in \mathbb{N}$, there exists $k \in \mathbb{N}$ with $n \leq k \leq m + n$ and

$$d(x_{k+1}, x_k) + \alpha_k d(x_k, x_{k-1}) \leq \sqrt{\frac{8S'}{m}}.$$

Note that as $\gamma(t) := \exp_{x_k}(t \exp_{x_k}^{-1} x_{k-1})$ is a geodesic line, we have

$$d(y_k, x_k) = d(\gamma(-\alpha_k), \gamma(0)) = \alpha_k d(x_k, x_{k-1}).$$

Using that $\rho_k d(z_k, y_k) = d(x_{k+1}, y_k)$ as in Lemma 4.4, we then have

$$d(z_k, y_k) \leq \rho_k^{-1} d(x_{k+1}, y_k) \leq \rho_k^{-1} (d(x_{k+1}, x_k) + \alpha_k d(x_k, x_{k-1})).$$

Using the fact that by (C2), and either among $(C1)^l$ or $(C1)^u$, we have that $\rho_k \geq \frac{4\eta}{(\gamma - 8\eta)}$ for all $k \in \mathbb{N}$, this implies that

$$\forall \varepsilon > 0 \quad \forall n \in \mathbb{N} \quad \exists k \in \mathbb{N} \left(n \leq k \leq \left\lceil \frac{S'(\gamma - 8\eta)^2}{2\eta^2\varepsilon^2} \right\rceil + n + 1 \text{ and } d(z_k, y_k) < \varepsilon \right).$$

A particular a priori assumption under which the above stronger property holds is condition (C3) from [31], where it is assumed that there exists an $\bar{\alpha} = \xi/(2 + \xi)$ such that

$$0 \leq \alpha_k \leq \alpha_{k+1} \leq \alpha < \bar{\alpha} < 1$$

for all $k \in \mathbb{N}$, where

$$\xi := \frac{1 - 4\eta e - \rho}{1 + \rho} > 0.$$

Concretely, following the same argument as given in the proof of [31, Theorem 18], it can be shown that under condition (C3) it holds that

$$\sum_{k=0}^{\infty} d^2(x_{k+1}, x_k) \leq \frac{1}{\xi - (2 + \xi)\alpha} \cdot \frac{\mu_0}{1 - \alpha} < +\infty.$$

In particular, the above condition on the inertia parameters is a type of a priori condition which is sometimes easier to verify in practice than $\sum_{k=0}^{\infty} \alpha_k d^2(x_k, x_{k-1}) < +\infty$, similar in spirit to the a priori conditions on the inertia parameters already pioneered by Alvarez and Attouch [3] (see also the seminal work of Bot, Csetnek and Hendrich [15] on splitting methods with inertia terms as well as [65] for similar discussions in the context of Hadamard manifolds).

In our main convergence theorem, we will also utilize the following lemma which later allows us to transfer our effective convergence results for $\{x_k\}_k$ to the sequences $\{y_k\}_k$ and $\{z_k\}_k$.

Lemma 4.11. *Let $f : K \times K \rightarrow \mathbb{R}$ be such that (A2), (A4) and (A5) hold. Suppose that (C1)^l and (C2) hold. Suppose further that $x^* \in S(K, f)$. Let $M : (0, \infty) \rightarrow \mathbb{N}$ be such that*

$$\forall \varepsilon > 0 \left(\sum_{k=M(\varepsilon)}^{\infty} \alpha_k d^2(x_k, x_{k-1}) < \varepsilon \right).$$

Then

$$\forall \varepsilon > 0 \forall n \geq k \geq M \left(\frac{8\eta^2\Theta(1-\alpha)\varepsilon^2}{(\gamma-8\eta)} \right) \left(d(z_n, y_n) < \sqrt{\frac{(\gamma-8\eta)}{16\eta^2\Theta(1-\alpha)}} d(x_k, x^*) + \varepsilon \right).$$

Proof. We define φ_k, δ_k and s_{k+1} as in Lemma 4.7 and argue as therein to apply Lemma 4.1, by which we get

$$\sum_{i=k+1}^{\infty} s_i \leq \frac{1}{1-\alpha} \varphi_k + \frac{1}{1-\alpha} \sum_{i=k}^{\infty} \delta_i$$

for any $k \in \mathbb{N}$. Hence, for $\varepsilon > 0$ and $k \geq M((1-\alpha)\varepsilon^2/2)$, we obtain

$$\sum_{i=k+1}^{\infty} \left(\frac{2-4\eta\beta_i-\rho_i}{\rho_i} \right) d^2(x_{i+1}, y_i) < \frac{1}{1-\alpha} d^2(x_k, x^*) + \varepsilon.$$

We can now argue similar as at the end of Lemma 4.9 to derive

$$\left(\frac{16\eta^2\Theta}{(\gamma-8\eta)} \right) d^2(z_i, y_i) \leq \left(\frac{2-4\eta\beta_i-\rho_i}{\rho_i} \right) d^2(x_{i+1}, y_i).$$

This yields

$$d^2(z_n, y_n) < \frac{(\gamma-8\eta)}{16\eta^2\Theta(1-\alpha)} d^2(x_k, x^*) + \frac{(\gamma-8\eta)}{16\eta^2\Theta} \varepsilon$$

for any $\varepsilon > 0$ and any $n \geq k \geq M((1-\alpha)\varepsilon/2)$. This yields the claim, using sub-additivity of the root. $\square$

We now show our first main quantitative convergence result for the sequences $\{x_k\}_k, \{y_k\}_k$ and $\{z_k\}_k$ generated by (RIPPA), set in the context of the assumption (C1)^l.

Theorem 4.12. *Let X be a Hadamard manifold and let $K \subseteq X$ be metrically closed such that the complete geodesic lines determined by any two points of K lie in K . Let $f : K \times K \rightarrow \mathbb{R}$ be such that (A2), (A4) and (A5) hold. Fix $\alpha, \rho \in [0, 1)$ and $x_0, x_{-1} \in K$ as well as $\{\beta_k\}_k \subseteq (0, \infty)$, $\{\alpha_k\}_k \subseteq [0, \alpha]$ and $\{\rho_k\}_k \subseteq [1-\rho, 1+\rho]$ such that (C1)^l and (C2) hold. Assume that (RIPPA) is well-defined and let $\{x_k\}_k, \{y_k\}_k$ and $\{z_k\}_k$ be its generated sequences. Suppose further that $S(K, f) \neq \emptyset$. If*

$$\sum_{k=0}^{\infty} \alpha_k d^2(x_k, x_{k-1}) < +\infty,$$

then the sequences $\{x_k\}_k, \{y_k\}_k$ and $\{z_k\}_k$ converge to the (unique) equilibrium point x^* of f . Further, we have the following rate of convergence: If $M : (0, \infty) \rightarrow \mathbb{N}$ and $S \in (0, \infty)$ are such that M is decreasing and

$$\sum_{k=0}^{\infty} \alpha_k d^2(x_k, x_{k-1}) \leq S \text{ as well as } \forall \varepsilon > 0 \left(\sum_{k=M(\varepsilon)}^{\infty} \alpha_k d^2(x_k, x_{k-1}) < \varepsilon \right),$$

then

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

where

$$C := \max \left\{ 1, \frac{4(3\gamma - 28\eta)}{\eta^2\Theta(1 - \alpha)} \right\} \cdot \frac{256(b + 2S)(3\gamma - 16\eta)^2}{\eta^2\Theta(1 - \alpha)^2(3\gamma - 28\eta)}$$

and

$$E := \min \left\{ 1, \frac{\eta^2\Theta(1 - \alpha)}{4(\gamma - 8\eta)} \right\} \cdot \frac{(1 - \alpha)}{32}$$

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

Proof. It is enough to show the quantitative result. Here, we first provide a rate for $\{x_k\}_k$. Given $\varepsilon > 0$, using Lemma 4.9, take $k \in \mathbb{N}$ with

$$M\left(\frac{(1 - \alpha)\varepsilon^2}{32}\right) \leq k \leq \left\lceil \frac{T}{\varepsilon^2} \right\rceil + M\left(\frac{(1 - \alpha)\varepsilon^2}{32}\right) + 1$$

such that

$$d(z_k, y_k) < \frac{\sqrt{1 - \alpha}}{64} \cdot \frac{(3\gamma - 28\eta)}{(3\gamma - 16\eta)} \varepsilon < \frac{\sqrt{1 - \alpha}}{32} \varepsilon,$$

where

$$T := \frac{64^2(b + 2S)(\gamma - 8\eta)(3\gamma - 16\eta)^2}{16\eta^2\Theta(1 - \alpha)^2(3\gamma - 28\eta)^2}.$$

Note in particular that $k \leq \lceil C/\varepsilon^2 \rceil + M(E\varepsilon^2) + 1$. We now first establish the bound

$$(-) \quad d(z_k, x^*) < \frac{\sqrt{1 - \alpha}}{8} \varepsilon.$$

For that we proceed by a case distinction similar to Lemma 4.4. First, using $z_k \in \text{Prox}_{\beta_k f(y_k, \cdot)}(y_k)$ and applying Lemma 2.3 to $f(y_k, \cdot)$ with $y = x^*$ yields $(\circ)_3$ as in Lemma 4.4. Similar to Lemma 4.4, we consider two cases:

Case 1: $f(y_k, z_k) \geq f(y_k, x^*)$. From $(\circ)_3$, we obtain

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

for all $\lambda \in [0, 1]$. Using (CS) and the fact that $\beta_k \geq \frac{1}{\gamma - 8\eta}$, the above implies

$$\frac{1}{2} \left((1 - \lambda) \frac{\gamma}{\gamma - 8\eta} - \lambda \right) d^2(z_k, x^*) \leq d(z_k, y_k) d(z_k, x^*)$$

for all $\lambda \in (0, 1]$. Either $d(z_k, x^*) = 0$, in which case $(-)$ holds immediately, or $d(z_k, x^*) > 0$, where we then get

$$\frac{1}{2} \left((1 - \lambda) \frac{\gamma}{\gamma - 8\eta} - \lambda \right) d(z_k, x^*) \leq d(z_k, y_k)$$

for all $\lambda \in (0, 1]$. This in particular holds for $\lambda = \frac{1}{8}$ and we thus have

$$\frac{3\gamma + 4\eta}{8(\gamma - 8\eta)} d(z_k, x^*) \leq d(z_k, y_k) < \frac{\sqrt{1 - \alpha}}{32} \varepsilon < \frac{\sqrt{1 - \alpha}}{8} \frac{3\gamma + 4\eta}{8(\gamma - 8\eta)} \varepsilon.$$

so that we get $(-)$ in that case.

Case 2: $f(y_k, z_k) < f(y_k, x^*)$. As in Lemma 4.4, we obtain $(\circ)_4$, and so

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

for all $\lambda \in (0, 1]$. Using (CS), we get

$$\frac{1}{2} \left((1 - \lambda)\gamma\beta_k - \lambda - \frac{2\eta\beta_k}{\lambda} \right) d^2(z_k, x^*) \leq \frac{\eta\beta_k}{\lambda} d^2(z_k, y_k) + d(z_k, y_k)d(z_k, x^*).$$

We now make two further case distinctions.

Case 2a: $d(z_k, x^*) \leq d(z_k, y_k)$. As we in particular have $d(z_k, y_k) < \frac{\sqrt{1-\alpha}}{32}\varepsilon$, $(-)$ follows immediately.

Case 2b: $d(z_k, x^*) > d(z_k, y_k)$. This yields

$$\frac{1}{2} \left((1 - \lambda)\gamma\beta_k - \lambda - \frac{2\eta\beta_k}{\lambda} \right) d(z_k, x^*) \leq \left(1 + \frac{\eta\beta_k}{\lambda} \right) d(z_k, y_k).$$

for all $\lambda \in (0, 1]$. We multiply both sides by $\frac{\lambda}{\lambda + \beta_k\eta}$ and rewrite the expression slightly to get

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

for all $\lambda \in (0, 1]$. Setting $\lambda = \frac{3}{8}$ and using $\beta_k \geq 1/(\gamma - 8\eta)$, we obtain

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

since $12\eta < \gamma$. Then further, again as $\beta_k \geq 1/(\gamma - 8\eta)$, 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 implies

$$\frac{(3\gamma - 28\eta)}{8(3\gamma - 16\eta)} d(z_k, x^*) \leq d(z_k, y_k) < \frac{\sqrt{1-\alpha}}{8} \cdot \frac{(3\gamma - 28\eta)}{8(3\gamma - 16\eta)} \varepsilon,$$

so that we also get $(-)$ in this case. As in Lemma 4.4, we have $\rho_k d(z_k, y_k) = d(x_{k+1}, y_k)$. Since $\rho_k \leq 1 + \rho \leq 2$ by (C2), using the triangle inequality we get

$$\begin{aligned} d(x_{k+1}, x^*) &\leq \rho_k d(z_k, y_k) + d(y_k, z_k) + d(z_k, x^*) \\ &\leq 3d(z_k, y_k) + d(z_k, x^*) \end{aligned}$$

With the bound on $d(z_k, y_k)$ and $(-)$, we thus obtain

$$d(x_{k+1}, x^*) < \frac{3}{32}\sqrt{1-\alpha}\varepsilon + \frac{4}{32}\sqrt{1-\alpha}\varepsilon = \frac{7}{32}\sqrt{1-\alpha}\varepsilon.$$

But now, as in particular $k \geq M((1 - \alpha)\varepsilon^2/32)$, Lemma 4.7 yields

$$d(x_n, x^*) \leq \sqrt{\frac{1}{1-\alpha}} d(x_{k+1}, x^*) + \frac{\varepsilon}{4} < \frac{7}{32}\varepsilon + \frac{1}{4}\varepsilon = \frac{15}{32}\varepsilon < \varepsilon$$

for all $n \geq k + 1$. This implies the rate for $\{x_k\}_k$ to x^* . We now turn to the rates for $\{z_k\}_k$ and $\{y_k\}_k$. Using triangle inequality, we get

$$d(z_n, x^*) \leq d(z_n, y_n) + d(x_{n+1}, y_n) + d(x_{n+1}, x^*)$$

for all $n \in \mathbb{N}$. Using $\rho_n d(z_n, y_n) = d(x_{n+1}, y_n)$ and $\rho_n \leq 2$, we obtain

$$d(z_n, x^*) \leq 3d(z_n, y_n) + d(x_{n+1}, x^*).$$

Using Lemma 4.11, we obtain

$$\forall j \in \mathbb{N} \quad \forall n \geq j \geq \left(M \left(\frac{\eta^2 \Theta(1-\alpha) \varepsilon^2}{32(\gamma-8\eta)} \right) + 1 \right) \left(d(z_n, y_n) < \sqrt{\frac{(\gamma-8\eta)}{16\eta^2 \Theta(1-\alpha)}} d(x_j, x^*) + \frac{\varepsilon}{16} \right).$$

As M is decreasing, the above inequality holds for any $n \geq j$ such that

$$j \geq j_0 + 1 := \left\lceil \frac{T}{\varepsilon^2} \frac{8^2(\gamma-8\eta)}{16\eta^2 \Theta(1-\alpha)} \right\rceil + M \left(\frac{\eta^2 \Theta(1-\alpha)^2 \varepsilon^2}{2 \cdot 8^2(\gamma-8\eta)} \right) + 1.$$

For a $j \geq j_0 + 1$ as above, the above rate for $\{x_k\}_k$ to x^* then in particular implies

$$d(x_j, x^*) < \frac{\varepsilon}{16 \sqrt{\frac{\gamma-8\eta}{16\eta^2 \Theta(1-\alpha)}}}$$

Combined, the above yields $d(z_n, y_n) < \varepsilon/8$ for all $n \geq j_0 + 2$. Finally, this implies

$$d(z_n, x^*) \leq 3d(z_n, y_n) + d(x_{n+1}, x^*) < \frac{3}{8}\varepsilon + \frac{4}{8}\varepsilon = \frac{7}{8}\varepsilon$$

for any $n \in \mathbb{N}$ such that

$$n \geq \max \left\{ \left\lceil \frac{T}{\varepsilon^2} \right\rceil + M \left(\frac{(1-\alpha)\varepsilon^2}{32} \right), j_0 \right\} + 2.$$

Further, for an n as above, a simple application of the triangle inequality yields

$$d(y_n, x^*) \leq d(z_n, y_n) + d(z_n, x^*) < \frac{1}{8}\varepsilon + \frac{7}{8}\varepsilon = \varepsilon.$$

This yields the rates for $\{z_k\}_k$ and $\{y_k\}_k$ after some slightly tedious but obvious simplifications. $\square$

Remark 4.13. Notice that the bound $S \in (0, \infty)$ on $\sum_{k=0}^{\infty} \alpha_k d^2(x_k, x_{k-1})$ featuring in Theorem 4.12 above can in fact be computed from the rate $M : (0, \infty) \rightarrow \mathbb{N}$ assumed for the same series. Concretely, note that $\sum_{k=M(1)}^{\infty} \alpha_k d^2(x_k, x_{k-1}) < 1$ by the defining property of M , so that

$$\sum_{k=0}^{\infty} \alpha_k d^2(x_k, x_{k-1}) \leq M(1) \cdot \max\{\alpha_k d^2(x_k, x_{k-1}) \mid k < M(1)\} + 1.$$

Further, while a rate of convergence M for the series $\sum_{k=0}^{\infty} \alpha_k d^2(x_k, x_{k-1}) < +\infty$ as in Theorem 4.12 above always exists, based on results from computability theory, it is generally not possible to obtain a computable such M (see e.g. [43]). However, it is often possible to obtain a computable so-called rate of metastability, that is a bound $M'(\varepsilon, g)$ such that

$$\forall \varepsilon > 0 \quad \forall g : \mathbb{N} \rightarrow \mathbb{N} \quad \exists N \leq M'(\varepsilon, g) \quad \forall m, n \in [N; N + g(N)] \quad \left(\sum_{k=m}^n \alpha_k d^2(x_k, x_{k-1}) < \varepsilon \right).$$

Indeed, general logical metatheorems from the proof mining program [43] (recall also the references in the introduction) guarantee that in an appropriate logical framework one can extract an effective M' from a corresponding proof of $\sum_{k=0}^{\infty} \alpha_k d^2(x_k, x_{k-1}) < +\infty$. In fact, a construction thereof is already immediate from an upper bound $\sum_{k=0}^{\infty} \alpha_k d^2(x_k, x_{k-1}) \leq S$, where it follows from [43, Proposition 2.27] (see also [43, Remark 2.29]) that $M'(\varepsilon, g) := \tilde{g}^{(\lceil S \cdot \varepsilon^{-1} \rceil)}(0)$ with $\tilde{g}(n) := n + g(n)$ is a rate of metastability. In particular, such a rate can hence immediately be given in the context of the parameter conditions from Remark 4.10.

While the previous arguments can be adapted to this computationally weaker assumption of a rate of metastability for the series, it would then naturally restrict us to only be able

to obtain rates of metastability instead of full rates of convergence in Theorem 4.12 above (as well as Theorem 4.24 below), which is not the focus of this paper. While computable rates of convergence for M are generally difficult to obtain, there are common online choices for α_k (already pioneered by Alvarez and Attouch [3]) where an explicit and effective rate M for $\sum_{k=0}^{\infty} \alpha_k d^2(x_k, x_{k-1}) < +\infty$ can be directly given. Concretely, consider the online choice discussed in [31, Remark 12]:

$$\alpha_k := \min \left\{ \alpha, \frac{\theta^k}{d^2(x_k, x_{k-1})} \right\}$$

for some $\theta \in (0, 1)$, supposing (w.l.o.g.) that $x_k \neq x_{k-1}$ for all $k \in \mathbb{N}$. Then one has

$$\sum_{k=n}^{\infty} \alpha_k d^2(x_k, x_{k-1}) \leq \sum_{k=n}^{\infty} \theta^k = \frac{1}{1-\theta} - \sum_{k=0}^{n-1} \theta^k = \frac{\theta^n}{1-\theta}$$

so that $M(\varepsilon) := \frac{\log(\varepsilon(1-\theta))}{\log(\theta)} + 1$ is a corresponding rate of convergence.

Remark 4.14. In the context of assumption $(C1)^l$, the quantitative result contained in Theorem 4.12 is, to our knowledge, already novel in the Euclidean case. In terms of assumptions on the bifunction, we find that compared to [31], (A3) can be dropped completely. Further, if (A1) is added as an assumption (which, compared to [31], here crucially contains the lower semicontinuity of the bifunction in its right argument in the absence of (A3)), then the assumptions that $S(K, f) \neq \emptyset$ and that (RIPPA) is well-defined can both be dropped from Theorem 4.12, using [24, Theorem 3.9] and [24, Proposition 4.4], respectively.

Remark 4.15. As highlighted in the introduction, since we do not rely on the finite dimensionality of the manifold in any way, the result extends mutatis mutandis to suitable Hadamard spaces which satisfy all the required metric properties outlined in Section 2.1. Concretely, the results given here already hold over Hadamard spaces (X, d) which satisfy the following two properties:

- (1) For each pair of distinct points $x, y \in X$, there is a unique metric line $r : \mathbb{R} \rightarrow X$ which passes through x and y .
- (2) For any $x, y, w \in X$ and $t \in [0, 1]$, for the unique point z on the metric line passing through x and y such that $d(z, x) = td(x, y)$ and $d(z, y) = (1+t)d(x, y)$, it holds that

$$d^2(z, w) \leq (1-t)d^2(x, w) + td^2(y, w) - t(1-t)d^2(x, y).$$

Spaces closed under metric lines as in item (1) above are also key for the notion of hyperbolic space studied by Reich and Shafrir [63]. Writing $(1+t)x \ominus ty$ for the unique point z on the metric line passing through x and y such that $d(z, x) = td(x, y)$ and $d(z, y) = (1+t)d(x, y)$ and $(1-t)x \oplus ty$ for the unique point z on the metric line passing through x and y such that $d(z, x) = td(x, y)$ and $d(z, y) = (1-t)d(x, y)$, we can see that

$$\exp_x(t \exp_x^{-1} y) = (1-t)x \oplus ty \text{ and } (1+t)x \ominus ty = \exp_x(-t \exp_x^{-1} y)$$

for all $x, y \in X$ and $t \in [0, 1]$. Lemma 2.2 thereby shows that Hadamard manifolds have the property from item (2) above. Lemma 2.1 holds just as well for general Hadamard spaces if we take $z = (1-t)x \oplus ty$. Next to Hadamard manifolds, another class of spaces that satisfies the above properties are Hilbert spaces. Now, with that notation the method (RIPPA) can be

defined over such Hadamard spaces via

$$\begin{cases} y_k := (1 + \alpha_k)x_k \ominus \alpha_k x_{k-1}, \\ z_k \in \operatorname{argmin}_{x \in K} \left\{ f(y_k, x) + \frac{1}{2\beta_k} d^2(y_k, x) \right\}, \\ x_{k+1} := \begin{cases} \rho_k z_k \oplus (1 - \rho_k)y_k, & \text{if } \rho_k \in [1 - \rho, 1], \\ \rho_k z_k \ominus (\rho_k - 1)y_k, & \text{if } \rho_k \in [1, 1 + \rho]. \end{cases} \end{cases}$$

for similar parameter restrictions as before, and the above results extend suitably to this purely metric variant, and hence in particular to Hilbert spaces (a detailed derivation tailored to that context can be found in the master thesis of the first author [23]). In that case, the rates remain exactly the same. Further, the convergence is then still strong, that is w.r.t. the norm and not relative to any notion of weak convergence. Akin to [24, 56], this is also here essentially due to the fact that equilibrium problems for strongly quasiconvex and pseudomonotone bifunctions have a unique solution.

Remark 4.16. The quantitative result contained in Theorem 4.12 above could also have been obtained by an application of the general results for quasi-Fejér monotone sequences under general metric regularity assumptions as developed by the second author in [55] (extending the work of Kohlenbach, López-Acedo and Nicolae [45]). We refer to the master thesis of the first author [23] for further discussions in that vein.

4.4. Convergence under condition $(C1)^u$. We now establish the convergence of the method (RIPPA) under the condition $(C1)^u$, that is under the assumption that there exists a $\theta > 0$ such that $\frac{1}{\gamma - 8\eta} + \theta \leq \beta_k \leq e$ for all $k \in \mathbb{N}$, where $e \leq \frac{1}{4\eta}$.

Similar to before, an essential ingredient for our quantitative convergence theorem for (RIPPA) under this assumption is an effective estimate for the asymptotic behavior of $d(z_k, y_k)$, again taking the form of an explicit bound $\iota : (0, \infty) \times \mathbb{N} \rightarrow \mathbb{N}$ such that for any $\varepsilon > 0$ and any $n \in \mathbb{N}$, there exists a $k \in \mathbb{N}$ with $n \leq k \leq \iota(\varepsilon, n)$ and $d(z_k, y_k) < \varepsilon$.

To motivate how this is achieved under the assumption $(C1)^u$, first recall how we have proceeded in the analogous result for $(C1)^l$, that is Lemma 4.9: Using $(+)_2$ from Lemma 4.4 with $\lambda = \frac{1}{2}$, we suitably instantiated $(*)$ from Lemma 4.1, after which we were able to infer the existence of a point $k \geq n$ (together with a respective bound) where

$$\left(\frac{2 - 4\eta\beta_k - \rho_k}{\rho_k} \right) d^2(x_{k+1}, y_k)$$

is suitably small. While $(C1)$ and $(C2)$ as in [31] allow for $(2 - 4\eta\beta_k - \rho_k)/\rho_k \rightarrow 0$ (setting $\rho_k = 2 - 4\eta e$ for all $k \in \mathbb{N}$ and considering $\beta_k \rightarrow e$), the upper bound $\Theta > 0$ such that $\beta_k \leq e - \Theta$ for all $k \in \mathbb{N}$ contained in $(C1)^l$ prohibits this behavior, so that in that context $(2 - 4\eta\beta_k - \rho_k)\rho_k$ can be bounded uniformly away from zero so that it can be ensured that also $\rho_k^{-2} d^2(x_{k+1}, y_k) = d^2(z_k, y_k)$ is suitably small at this point. In that context, one can then even generalize the lower bound on $\{\beta_k\}_k$ to $\beta_k \geq \frac{1}{\gamma - 8\eta}$ for all $k \in \mathbb{N}$.

Now, we can actually treat the non-strict upper bound $\beta_k \leq \frac{1}{4\eta}$ by using $(+)_2$ from Lemma 4.4 with λ chosen adaptively as λ_k with k (or rather, the values of β_k) instead of setting $\lambda = \frac{1}{2}$. Using an appropriate λ_k , we will be able to infer the existence of a point $k \geq n$ (together with a respective bound) where

$$\left(\frac{2 - \rho_k - \frac{2\beta_k\eta}{\lambda_k}}{\rho_k} \right) d^2(x_{k+1}, y_k)$$

is suitably small. But to now simultaneously ensure that we are able to uniformly bound $(2 - \rho_k - 2\beta_k\eta/\lambda_k) \rho_k$ away from zero for our adaptive $\lambda_k, \beta_k \rightarrow \frac{1}{\gamma-8\eta}$ must be excluded, leading to the assumption $(C1)^u$ which introduces a strict lower bound $\theta > 0$ such that $\frac{1}{\gamma-8\eta} + \theta \leq \beta_k$ for all $k \in \mathbb{N}$. Once this is guaranteed, we are able to infer that also $\rho_k^{-2}d^2(x_{k+1}, y_k) = d^2(z_k, y_k)$ is suitably small, similar to before.

To now choose these λ_k adaptively with β_k , we will rely on some technical machinery introduced for a similar purpose in [24].¹ For $x \geq \frac{1}{\gamma-8\eta}$, set

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

where γ and η are as in (A4) and (A5). 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 which we require are collected in the following lemma. All items apart from item (v) already appear with corresponding proofs in [24], so that we only provide details for the last item here.

Lemma 4.17 (extending [24, Lemma 4.11]). *Consider N as defined in $(\dagger)$. 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) $x \mapsto \frac{N(x)}{2x}$ is decreasing for $x \in [\frac{1}{\gamma-8\eta}, \frac{1}{4\eta}]$,
- (v) For any $\theta > 0$:

$$N\left(\frac{1}{\gamma-8\eta} + \theta\right) \geq \frac{1}{2 - \frac{2\theta(\gamma-8\eta)^2}{(1+\theta(\gamma-8\eta))(2\gamma-8\eta)}}.$$

Proof. We only prove (v), the proofs for the remaining items can be found in the proof of [24, Lemma 4.11]. Note first that a simple calculation yields

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

Further, we have $\sqrt{\gamma(\gamma-8\eta) - \frac{8\eta}{x}} \geq (\gamma-8\eta)$ for all $x \geq \frac{1}{\gamma-8\eta}$. Together, this yields

$$\begin{aligned} N\left(\frac{1}{\gamma-8\eta} + \theta\right) &\geq \frac{\gamma + (\gamma-8\eta)}{\frac{2}{\frac{1}{\gamma-8\eta} + \theta} + 2\gamma} = \frac{2\gamma-8\eta}{\frac{2(\gamma-8\eta)}{1+\theta(\gamma-8\eta)} + 2\gamma} \\ &= \frac{(2\gamma-8\eta)(1+\theta(\gamma-8\eta))}{2(2\gamma-8\eta) + 2\gamma\theta(\gamma-8\eta)} = \frac{1+\theta(\gamma-8\eta)}{2 + \frac{2\gamma\theta(\gamma-8\eta)}{2\gamma-8\eta}}. \end{aligned}$$

¹The strategy of choosing λ_k adaptively with β_k used here was first used in [24] to extend the parameter restrictions of the proximal point type method from [36] (and its nonlinear version studied in [24]) from the open interval $(\frac{1}{\gamma-8\eta}, \frac{1}{4\eta})$ to the closed interval $[\frac{1}{\gamma-8\eta}, \frac{1}{4\eta}]$.

For shorter notation, set $t := \theta(\gamma - 8\eta)$. We continue to observe that

$$\begin{aligned}
\frac{2}{1+t} + \frac{2\gamma t}{(1+t)(2\gamma - 8\eta)} &= 2 - \left(2 - \frac{2}{1+t} - \frac{2\gamma t}{(1+t)(2\gamma - 8\eta)} \right) \\
&= 2 - \left(\frac{2(1+t)(2\gamma - 8\eta) - 2(2\gamma - 8\eta) - 2\gamma t}{(1+t)(2\gamma - 8\eta)} \right) \\
&= 2 - \left(\frac{2t(2\gamma - 8\eta) - 2\gamma t}{(1+t)(2\gamma - 8\eta)} \right) \\
&= 2 - \left(\frac{2t(\gamma - 8\eta)}{(1+t)(2\gamma - 8\eta)} \right).
\end{aligned}$$

Finally, combining the above, we obtain the result:

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

□

The function N can now be used to introduce our adaptive parameter choices λ_k and the following lemma provides both the definition and essential property related to this. A variant of this result, featuring a slightly refined λ_k , already appears in [24, Lemma 4.12], and while the proof is essentially the same, we provide the details here for self-containedness.

Lemma 4.18. *Suppose that $12\eta < \gamma$ holds. For any $k \in \mathbb{N}$, let $\beta_k \in [\frac{1}{\gamma - 8\eta}, \frac{1}{4\eta}]$ and consider*

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

for $\lambda_k := N(\beta_k)$, where N is defined as in $(\dagger)$. Then for any $k \in \mathbb{N}$: $c_k - d_k = 0$.

Proof. We have

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

To see that $c_k - d_k = 0$, 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. □

Having introduced our adaptive convex combination parameters λ_k , we also here rely on a specific lower bound on a key quantity featuring in later arguments, similar to Lemma 4.8.

Lemma 4.19. *Suppose that $(C1)^u$ and $(C2)$ hold. Then, for all $k \in \mathbb{N}$:*

$$\left(2 - \rho_k - \frac{2\beta_k}{N(\beta_k)}\eta \right) \rho_k \geq \frac{16\eta^2\theta}{(1 + \theta(\gamma - 8\eta))(2\gamma - 8\eta)} > 0.$$

Proof. Using the fact that N is increasing by Lemma 4.17, (iii) yields

$$\left(2 - \rho_k - \frac{2\beta_k}{N(\beta_k)}\eta \right) \rho_k \stackrel{(C2)}{\geq} \left(4\eta e - \frac{2\beta_k}{N(\beta_k)}\eta \right) 4\eta e \stackrel{(C1)^u}{\geq} \left(4\eta e - \frac{2\eta}{N(1/(\gamma - 8\eta) + \theta)} \right) 4\eta e.$$

Now using Lemma 4.17, (v), and since $e \geq 1/(\gamma - 8\eta)$ by $(C1)^u$, we obtain

$$\begin{aligned} \left(2 - \rho_k - \frac{2\beta_k}{N(\beta_k)}\eta\right) \rho_k &\geq \left(4\eta e - 2\eta e \left(2 - \frac{2\theta(\gamma - 8\eta)^2}{(1 + \theta(\gamma - 8\eta))(2\gamma - 8\eta)}\right)\right) 4\eta e \\ &= 16\eta^2 e^2 \frac{\theta(\gamma - 8\eta)^2}{(1 + \theta(\gamma - 8\eta))(2\gamma - 8\eta)} \\ &\geq \frac{16\eta^2 \theta}{(1 + \theta(\gamma - 8\eta))(2\gamma - 8\eta)}. \end{aligned}$$

Notice that the last term is greater than 0 by $(C1)^u$. $\square$

We now provide the following counterpart to Lemma 4.5.

Lemma 4.20. *Let $f : K \times K \rightarrow \mathbb{R}$ be such that (A2), (A4) and (A5) hold. Suppose that either among $(C1)^l$ or $(C1)^u$ hold. Suppose further that $x^* \in S(K, f)$. Then, for every $k \in \mathbb{N}$:*

$$d^2(x_{k+1}, x^*) \leq d^2(y_k, x^*) - \left(\frac{2 - \rho_k - \frac{2\beta_k\eta}{N(\beta_k)}}{\rho_k}\right) d^2(x_{k+1}, y_k).$$

where N is defined as in $(\dagger)$.

Proof. From Lemma 4.4 we know that for all $k \in \mathbb{N}$ that at least one of $(+)_1$ or $(+)_2$ holds, that is (using Lemma 4.17, (iii)), that either

$$d^2(x_{k+1}, x^*) \leq d^2(y_k, x^*) - \left(\frac{2 - \rho_k}{\rho_k}\right) d^2(x_{k+1}, y_k) - \rho_k ((1 - N(\beta_k))\gamma\beta_k - N(\beta_k)) d^2(z_k, x^*)$$

or

$$\begin{aligned} d^2(x_{k+1}, x^*) &\leq d^2(y_k, x^*) - \left(\frac{2 - \rho_k - \frac{2\beta_k\eta}{N(\beta_k)}}{\rho_k}\right) d^2(x_{k+1}, y_k) \\ &\quad - \rho_k \left((1 - N(\beta_k))\gamma\beta_k - N(\beta_k) - \frac{2\eta\beta_k}{N(\beta_k)} \right) d^2(z_k, x^*). \end{aligned}$$

We begin by arguing that, in both of the above inequalities, we can discard the last term. As $\rho_k \geq 0$, we do this by showing that

$$((1 - N(\beta_k))\gamma\beta_k - N(\beta_k)) \geq \left((1 - N(\beta_k))\gamma\beta_k - N(\beta_k) - \frac{2\eta\beta_k}{N(\beta_k)} \right) \geq 0.$$

The first inequality is obvious. The second is (by multiplying with $\frac{N(\beta_k)}{2\beta_k}$) equivalent to showing

$$\left(\frac{N(\beta_k)(1 - N(\beta_k))(1 + \gamma\beta_k)}{2\beta_k} - \frac{N(\beta_k)}{2\beta_k} - \eta \right) \geq 0.$$

But now

$$\left(\frac{N(\beta_k)(1 - N(\beta_k))(1 + \gamma\beta_k)}{2\beta_k} - \frac{N(\beta_k)}{2\beta_k} - \eta \right) = d_k - c_k,$$

where c_k and d_k are as in Lemma 4.18, which implies that $d_k - c_k = 0$ holds. The inequality now follows from the fact that

$$\left(\frac{2 - \rho_k - \frac{2\beta_k\eta}{N(\beta_k)}}{\rho_k} \right) \leq \frac{2 - \rho_k}{\rho_k}. \quad \square$$

In the same way as Lemma 4.6, we also obtain the following result:

Lemma 4.21 (extending [31, Proposition 10]). *Let $f : K \times K \rightarrow \mathbb{R}$ be such that (A2), (A4) and (A5) hold. Suppose that either among $(C1)^l$ or $(C1)^u$ hold. Suppose further that $x^* \in S(K, f)$. Set $\varphi_k := d^2(x_k, x^*)$ for every $k \in \mathbb{N}$. Then, for every $k \in \mathbb{N}$:*

$$\varphi_{k+1} - \varphi_k - \alpha_k(\varphi_k - \varphi_{k-1}) + \left(\frac{2 - \rho_k - \frac{2\beta_k\eta}{N(\beta_k)}}{\rho_k} \right) d^2(x_{k+1}, y_k) \leq (\alpha_k^2 + \alpha_k) d^2(x_k, x_{k-1})$$

where N is defined as in $(\dagger)$.

We continue with the analogue of Lemma 4.9.

Lemma 4.22. *Let $f : K \times K \rightarrow \mathbb{R}$ be such that (A2), (A4) and (A5) hold. Suppose that $(C1)^u$ and (C2) hold. Suppose further that $x^* \in S(K, f)$. Let $S \in (0, \infty)$ be such that*

$$\sum_{k=0}^{\infty} \alpha_k d^2(x_k, x_{k-1}) \leq S.$$

Then

$$\forall \varepsilon > 0 \forall n \in \mathbb{N} \exists k \in \mathbb{N} \left(n \leq k \leq \left\lceil \frac{(b+2S)((1+\theta(\gamma-8\eta))(2\gamma-8\eta))}{16\eta^2(1-\alpha)\theta\varepsilon^2} \right\rceil + n + 1 \text{ and } d(z_k, y_k) < \varepsilon \right),$$

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

Proof. Proceeding as in the proof of Lemma 4.9, define $\varphi_k := d^2(x_k, x^*)$ and $\delta_k := (\alpha_k^2 + \alpha_k) d^2(x_k, x_{k-1})$ for all $k \geq 0$. We have

$$s_{k+1} := \left(\frac{2 - \rho_k - \frac{2\beta_k\eta}{N(\beta_k)}}{\rho_k} \right) d^2(x_{k+1}, y_k) \geq 0$$

by Lemma 4.19 and $\sum_{k=0}^{\infty} \delta_k \leq 2S$. Using Lemma 4.21, we get that Lemma 4.3 applies so that for $\varepsilon > 0$ and $n \in \mathbb{N}$ as before, we have

$$\exists k \in \mathbb{N} \left(n \leq k \leq \left\lceil \frac{b+2S}{(1-\alpha)\varepsilon} \right\rceil + n + 1 \text{ and } \left(\frac{2 - \rho_k - \frac{2\beta_k\eta}{N(\beta_k)}}{\rho_k} \right) d^2(x_{k+1}, y_k) < \varepsilon \right).$$

Using that $\rho_k d(z_k, y_k) = d(x_{k+1}, y_k)$, this implies that

$$\left(2 - \rho_k - \frac{2\beta_k\eta}{N(\beta_k)} \right) \rho_k d^2(z_k, y_k) < \varepsilon.$$

With Lemma 4.19, we obtain that

$$\frac{16\eta^2\theta}{(1+\theta(\gamma-8\eta))(2\gamma-8\eta)} d^2(z_k, y_k) < \varepsilon$$

for such a k . Put together, this yields the claim. $\square$

Towards showing a quantitative convergence result not only for $\{x_k\}_k$, but also for $\{y_k\}_k$ and $\{z_k\}_k$, we establish an analogous result to Lemma 4.11.

Lemma 4.23. *Let $f : K \times K \rightarrow \mathbb{R}$ be such that (A2), (A4) and (A5) hold. Suppose that $(C1)^u$ and (C2) hold. Suppose further that $x^* \in S(K, f)$. Let $M : (0, \infty) \rightarrow \mathbb{N}$ be such that*

$$\forall \varepsilon > 0 \left(\sum_{k=M(\varepsilon)}^{\infty} \alpha_k d^2(x_k, x_{k-1}) < \varepsilon \right).$$

Then:

$$\forall \varepsilon > 0 \quad \forall k \in \mathbb{N} \quad \forall n > k \geq M \left(\frac{8\eta^2\theta(1-\alpha)\varepsilon^2}{(1+\theta(\gamma-8\eta))(2\gamma-8\eta)} \right) \\ \left(d(z_n, y_n) < \sqrt{\frac{(1+\theta(\gamma-8\eta))(2\gamma-8\eta)}{16\eta^2\theta(1-\alpha)}} d(x_k, x^*) + \varepsilon \right).$$

Proof. We argue as in Lemma 4.11 to obtain that

$$\sum_{i=k+1}^{\infty} \left(\frac{2 - \rho_i - \frac{2\beta_i\eta}{N(\beta_i)}}{\rho_i} \right) d^2(x_{i+1}, y_i) < \frac{1}{1-\alpha} d^2(x_k, x^*) + \varepsilon.$$

for any $\varepsilon > 0$ and any $k \geq M((1-\alpha)\varepsilon/2)$. Proceeding as in the proof of Lemma 4.11, using $\rho_k d(z_k, y_k) = d(x_{k+1}, y_k)$ and Lemma 4.19, yields the result. $\square$

The following is now our second main quantitative convergence result for the sequences $\{x_k\}_k$, $\{y_k\}_k$ and $\{z_k\}_k$ generated by (RIPPA), set in the context of the assumption $(C1)^u$.

Theorem 4.24. *Let X be a Hadamard manifold and let $K \subseteq X$ be metrically closed such that the complete geodesic lines determined by any two points of K lie in K . Let $f : K \times K \rightarrow \mathbb{R}$ be such that (A2), (A4) and (A5) hold. Fix $\alpha, \rho \in [0, 1)$ and $x_0, x_{-1} \in K$ as well as $\{\beta_k\}_k \subseteq (0, \infty)$, $\{\alpha_k\}_k \subseteq [0, \alpha]$ and $\{\rho_k\}_k \subseteq [1-\rho, 1+\rho]$ such that $(C1)^u$ and $(C2)$ hold. Assume that (RIPPA) is well-defined and let $\{x_k\}_k$, $\{y_k\}_k$ and $\{z_k\}_k$ be its generated sequences. Suppose further that $S(K, f) \neq \emptyset$. If*

$$\sum_{k=0}^{\infty} \alpha_k d^2(x_k, x_{k-1}) < +\infty,$$

then the sequences $\{x_k\}_k$, $\{y_k\}_k$ and $\{z_k\}_k$ converge to the (unique) equilibrium point x^ of f . Further, we have the following rate of convergence: If $M : (0, \infty) \rightarrow \mathbb{N}$ and $S \in (0, \infty)$ are such that M is decreasing and*

$$\sum_{k=0}^{\infty} \alpha_k d^2(x_k, x_{k-1}) \leq S \text{ as well as } \forall \varepsilon > 0 \left(\sum_{k=M(\varepsilon)}^{\infty} \alpha_k d^2(x_k, x_{k-1}) < \varepsilon \right),$$

then

$$\forall \varepsilon > 0 \quad \forall k \geq \left(\left\lceil \frac{C'}{\varepsilon^2} \right\rceil + M(E'\varepsilon^2) + 2 \right) (d(x_k, x^*), d(y_k, x^*), d(z_k, x^*) < \varepsilon),$$

where

$$C' := \max \left\{ 1, \frac{(2\gamma-8\eta)^2}{\eta^3\theta(1-\alpha)} \right\} \cdot \frac{64(b+2S)(2\gamma-8\eta)(3\gamma-16\eta)^2}{\eta^3\theta(1-\alpha)^2(3\gamma-28\eta)}$$

and

$$E' := \min \left\{ 1, \frac{\eta^3\theta(1-\alpha)}{32(2\gamma-8\eta)^2} \right\} \cdot \frac{(1-\alpha)}{32}$$

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

Proof. The proof proceeds in exactly the same way as the proof of Theorem 4.12, with the only difference that we refer to Lemma 4.22 instead of Lemma 4.9 and Lemma 4.23 instead of Lemma 4.11. We in that way only provide a sketch. Write

$$T' := \frac{64^2(b+2S)((1+\theta(\gamma-8\eta))(2\gamma-8\eta)(3\gamma-16\eta)^2}{16\eta^2(1-\alpha)^2(3\gamma-28\eta)^2\theta}.$$

Given $\varepsilon > 0$, using Lemma 4.22, take

$$M((1 - \alpha)\varepsilon^2/32) \leq k \leq \left\lceil \frac{T'}{\varepsilon^2} \right\rceil + M((1 - \alpha)\varepsilon^2/32) + 1$$

such that

$$d(z_k, y_k) < \frac{\sqrt{1 - \alpha}}{64} \frac{3\gamma - 28\eta}{(3\gamma - 16\eta)} \varepsilon < \frac{\sqrt{1 - \alpha}}{32} \varepsilon.$$

Proceeding similarly as in the proof of Theorem 4.12, one obtains

$$\forall n \geq \left\lceil \frac{T'}{\varepsilon^2} \right\rceil + M((1 - \alpha)\varepsilon^2/32) + 2 \left(d(x_n, x^*) < \frac{\varepsilon}{2} \right).$$

Again, as $M : (0, \infty) \rightarrow \mathbb{N}$ is decreasing, Lemma 4.23 combined with the above yields

$$d(z_n, y_n) < \sqrt{\frac{(1 + \theta(\gamma - 8\eta))(2\gamma - 8\eta)}{16\eta^2\theta(1 - \alpha)}} d(x_j, x^*) + \frac{\varepsilon}{16}.$$

for any $n \geq j$ with

$$j \geq \left\lceil \frac{T' 8^2(1 + \theta(\gamma - 8\eta))(2\gamma - 8\eta)}{\varepsilon^2 16\eta^2\theta(1 - \alpha)} \right\rceil + M\left(\frac{\eta^2\theta(1 - \alpha)^2\varepsilon^2}{128(1 + \theta(\gamma - 8\eta))(2\gamma - 8\eta)}\right) + 2.$$

For such an n , similar to before, we get

$$d(x_n, x^*) < \frac{\varepsilon^2}{16\sqrt{\frac{(1 + \theta(\gamma - 8\eta))(2\gamma - 8\eta)}{16\eta^2\theta(1 - \alpha)}}}$$

Combining the above yields $d(z_n, y_n) < \frac{\varepsilon}{8}$ for such an n . One now obtains the result in the same way as in the proof of Theorem 4.12, using various (tedious, but trivial) simplifications along the way. $\square$

Remarks 4.13 – 4.16 also apply in this context of $(C1)^u$, that is to Theorem 4.24.

Acknowledgments: This paper is a revised version of parts of the master thesis [23] 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 [23]. 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] F. Alvarez. On the minimizing property of a second order dissipative system in Hilbert spaces. *SIAM Journal on Control and Optimization*, 38:1102–1119, 2000.
- [2] F. Alvarez. Weak Convergence of a Relaxed and Inertial Hybrid Projection-Proximal Point Algorithm for Maximal Monotone Operators in Hilbert Space. *SIAM Journal on Optimization*, 14(3):773–782, 2004.
- [3] F. Alvarez and H. Attouch. An Inertial Proximal Method for Maximal Monotone Operators via Discretization of a Nonlinear Oscillator with Damping. *Set-Valued Analysis*, 9(1):3–11, 2001.
- [4] A.S. Antipin. Minimization of convex functions on convex sets by means of differential equations. *Differential Equations*, 30(9):1365–1375, 1994.
- [5] H. Attouch, X. Goudou, and P. Redont. The heavy ball with friction method. I. The continuous dynamical system. *Communications in Contemporary Mathematics*, 2(1):1–34, 2000.
- [6] H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer Cham, 2nd edition, 2017.
- [7] M. Bačák. The proximal point algorithm in metric spaces. *Israel Journal of Mathematics*, 194:689–701, 2013.

- [8] M. Bačák. *Convex Analysis and Optimization in Hadamard Spaces*. Series in Nonlinear Analysis and Applications. De Gruyter, Berlin, 2014.
- [9] M. Bačák. Old and new challenges in Hadamard spaces. *Japanese Journal of Mathematics*, 18(2):115–168, 2023.
- [10] 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*, 200:428–436, 2024.
- [11] 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*, 195:1087–1105, 2022.
- [12] I.D. Berg and I.G. Nikolaev. Quasilinearization and curvature of Aleksandrov spaces. *Geometriae Dedicata*, 133(1):195–218, 2008.
- [13] D.P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York, 1982.
- [14] E. Blum and W. Oettli. From optimization and variational inequalities to equilibrium problems. *Mathematics Student*, 63(1):123–145, 1994.
- [15] R.I. Boş, E.R. Csetnek, and C. Hendrich. Inertial Douglas–Rachford splitting for monotone inclusion problems. *Applied Mathematics and Computation*, 256:472–487, 2015.
- [16] H. Brézis and P.L. Lions. Produits infinis de resolvantes. *Israel Journal of Mathematics*, 29(4):329–345, 1978.
- [17] M. Bridson and A. Haefliger. *Metric Spaces of Non-Positive Curvature*, volume 319 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 1 edition, 1999.
- [18] 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.
- [19] 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.
- [20] S.S. Chang, J.C. Yao, M. Liu, and L.C. Zhao. Inertial proximal point algorithm for variational inclusion in Hadamard manifolds. *Applicable Analysis*, 102(7):2055–2066, 2023.
- [21] 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.
- [22] J. Cotrina and Y. García. Equilibrium Problems: Existence Results and Applications. *Set-Valued and Variational Analysis*, 26(1):159–177, 2018.
- [23] L.M. Després. On the proximal point algorithm for strongly quasiconvex pseudomonotone equilibrium problems in Hadamard spaces. Master’s thesis, TU Darmstadt, 2026.
- [24] L.M. Després and N. Pischke. On strongly quasiconvex pseudomonotone equilibrium problems in Hadamard spaces, 2026. Preprint, <https://arxiv.org/abs/2606.27563>.
- [25] J. Eckstein and D.P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992.
- [26] J. Eckstein and M.C. Ferris. Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control. *INFORMS Journal on Computing*, 10:218–235, 1998.
- [27] K. Fan. A Minimax Inequality and Applications. In O. Sisha, editor, *Inequalities*, volume 3, pages 103–113. Academic Press, New York, 1972.
- [28] O.P. Ferreira and P.R. Oliveira. Proximal point algorithm on Riemannian manifolds. *Optimization*, 51:257–270, 2002.
- [29] F. Flores-Bazán. Existence Theorems for Generalized Noncoercive Equilibrium Problems: The Quasi-Convex Case. *SIAM Journal on Optimization*, 11(3):675–690, 2001.
- [30] S.-M. Grad, F. Lara, and R. T. Marcavillaca. Relaxed-inertial proximal point type algorithms for quasi-convex minimization. *Journal of Global Optimization*, 85(3):615–635, 2023.
- [31] 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.
- [32] S.-M. Grad, F. Lara, and R.T. Marcavillaca. Strongly Quasiconvex Functions: What We Know (So Far). *Journal of Optimization Theory and Applications*, 205, 2025. Article no. 38.
- [33] D. Hieu, H.N. Duong, and B.H. Thai. Convergence of relaxed inertial methods for equilibrium problems. *Journal of Applied and Numerical Optimization*, 3:215–229, 2021.
- [34] A. Iusem, G. Kassay, and W. Sosa. On certain conditions for the existence of solutions of equilibrium problems. *Mathematical Programming*, 116:259–273, 2009.

- [35] A. Iusem and F. Lara. Optimality Conditions for Vector Equilibrium Problems with Applications. *Journal of Optimization Theory and Applications*, 180(1):187–206, 2019.
- [36] 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.
- [37] 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.
- [38] A. Iusem and W. Sosa. On the proximal point method for equilibrium problems in Hilbert spaces. *Optimization*, 59(8):1259–1274, 2010.
- [39] M. Jovanović. A note on strongly convex and quasiconvex functions. *Mathematical Notes*, 60:584–585, 1996.
- [40] K. Khammahawong, P. Chaipunya, and P. Kumam. An inertial Mann algorithm for nonexpansive mappings on Hadamard manifolds. *AIMS Mathematics*, 8(1):2093–2116, 2023.
- [41] H. Khatibzadeh and V. Mohebbi. Proximal point algorithm for infinite pseudo-monotone bifunctions. *Optimization*, 65(8):1629–1639, 2016.
- [42] 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.
- [43] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer, Berlin, Heidelberg, 2008.
- [44] 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.
- [45] 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.
- [46] I.V. Konnov. Application of the proximal point method to nonmonotone equilibrium problems. *Journal of Optimization Theory and Applications*, 119:317–333, 2003.
- [47] F. Lara. On Strongly Quasiconvex Functions: Existence Results and Proximal Point Algorithms. *Journal of Optimization Theory and Applications*, 192(3):891–911, 2022.
- [48] 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.
- [49] R. López. Approximations of equilibrium problems. *SIAM Journal on Control and Optimization*, 50(2):1038–1070, 2012.
- [50] 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.
- [51] A. Moudafi. Proximal point algorithm extended to equilibrium problems. *Journal of Natural Geometry*, 15(1–2):91–100, 1999.
- [52] A. Moudafi. Second-order differential proximal methods for equilibrium problems. *Journal of Inequalities in Pure and Applied Mathematics*, 4, 2003. Article no. 18.
- [53] 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.
- [54] W. Oettli. A remark on vector-valued equilibria and generalized monotonicity. *Acta Mathematica Vietnamica*, 22:213–221, 1997.
- [55] N. Pischke. Quantitative Results on Algorithms for Zeros of Differences of Monotone Operators in Hilbert Space. *Journal of Convex Analysis*, 30(1):295–315, 2023.
- [56] N. Pischke. On the proximal point algorithm for strongly quasiconvex functions in Hadamard spaces. *Optimization Methods and Software*, 40(6):1438–1453, 2025.
- [57] B.T. Polyak. Existence Theorems and Convergence of Minimizing Sequences in Extremum Problems with Restrictions. *Soviet Mathematics. Doklady*, 7:72–75, 1966.
- [58] E.A. Papa Quiroz. Proximal Point Method for Quasiconvex Functions in Riemannian Manifolds. *Journal of Optimization Theory and Applications*, 202(3):1268–1285, 2024.
- [59] 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.
- [60] 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.
- [61] 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.

- [62] 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.
- [63] S. Reich and I. Shafrir. Nonexpansive iterations in hyperbolic spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 15(6):537–558, 1990.
- [64] R.T. Rockafellar. Monotone Operators and the Proximal Point Algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976.
- [65] D.R. Sahu, S. Sharma, and P. Gautam. Efficient Douglas–Rachford Methods on Hadamard Manifolds with Applications to the Heron Problems. *Applied Mathematics & Optimization*, 93, 2026. Article no. 78.
- [66] T. Sakai. *Riemannian Geometry*, volume 149 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, 1996.
- [67] L. Van Vinh, V.N. Tran, and P.T. Vuong. A second-order dynamical system for equilibrium problems. *Numerical Algorithms*, 91(1):327–351, 2022.