Abstract. Hadamard spaces have traditionally played important roles in geometry and geometric group theory. More recently, they have additionally turned out to be a suitable framework for convex analysis, optimization and non-linear probability theory. The attractiveness of these emerging subject fields stems, inter alia, from the fact that some of the new results have already found their applications both in mathematics and outside. Most remarkably, a gradient flow theorem in Hadamard spaces was used to attack a conjecture of Donaldson in Kähler geometry. Other areas of applications include metric geometry and minimization of submodular functions on modular lattices. There have been also applications into computational phylogenetics and image processing.

We survey recent developments in Hadamard space analysis and optimization with the intention to advertise various open problems in the area. We also point out several fallacies in the existing proofs.

Keywords and phrases: Bi-Lipschitz embedding, convex function, gradient flow, Hadamard space, harmonic mapping, Mosco convergence, non-positive curvature, proximal mapping, proximal point algorithm, strongly continuous semigroup, submodular function, weak convergence

Mathematics Subject Classification (2020): Primary 46N10, 49J53; Secondary 47H20, 49M27

1. Introduction

The present paper is a follow-up to the 2014 book [16] with the aim to present new advances in the theory of Hadamard spaces and their
applications. We focus primarily on analysis and optimization, because their current development stage is, in our opinion, very favorable. On the one hand, these subject fields are very young and offer many new possibilities for further research, and on the other hand, the existing theory is already mature enough to be applied elsewhere. We will highlight the most notable applications including

- a conjecture of Donaldson on the asymptotic behavior of the Calabi flow in Kähler geometry,
- the existence of Lipschitz retractions in finite subset space,
- submodular function minimization on modular lattices,
- computing averages of trees in phylogenetics,
- computing averages of positive definite matrices in diffusion tensor imaging.

In spite of the progress described in the present survey, many intriguing questions remain unanswered and many new questions have been raised. We will gather them here, too. Last but not least, we point out that there exist a few rather problematic proofs in this area, which ought to be carefully inspected.

The history of Hadamard spaces can be traced back to a 1936 paper by Wald [216]. Their importance was recognized by Alexandrov [1] in the 1950s and that is why Hadamard spaces are sometimes referred to as *spaces of non-positive curvature in the sense of Alexandrov*. Gromov later conceived the acronym CAT(0), where C stands for Cartan, A for Alexandrov and T for Toponogov, and where 0 is the upper curvature bound. Since then, Hadamard spaces have been alternatively called complete CAT(0) spaces. For historical remarks as well as for an authoritative account on Hadamard spaces in geometry and geometric group theory, the interested reader is referred to the Bridson–Haefliger book [56].

Hadamard spaces are, by definition, geodesic metric spaces of non-positive curvature. More precisely, let \((X,d)\) be a metric space. A mapping \(\gamma: [0, 1] \to X\) is called a geodesic if \(d(\gamma(s), \gamma(t)) = |s - t|d(\gamma(0), \gamma(1))\), for every \(s, t \in [0, 1]\). If, for every \(x, y \in X\), there exists a geodesic \(\gamma: [0, 1] \to X\) such that \(\gamma(0) = x\) and \(\gamma(1) = y\), we say that \((X,d)\) is a geodesic metric space. Furthermore, if we have

\[
(1) \quad d(x, \gamma(t))^2 \leq (1 - t)d(x, \gamma(0))^2 + td(x, \gamma(1))^2 - t(1 - t)d(\gamma(0), \gamma(1))^2,
\]

for every \(x \in X\), geodesic \(\gamma: [0, 1] \to X\), and \(t \in [0, 1]\), we say that \((X,d)\) is a *CAT(0) space*. Inequality (1) expresses the non-positive curvature of the space. It also implies that geodesics in a CAT(0) space are uniquely determined by their endpoints, that is, given two points \(x, y \in X\), there exists a unique geodesic \(\gamma: [0, 1] \to X\) such that \(\gamma(0) = x\) and \(\gamma(1) = y\).
We can therefore use the linear notation $\gamma(t) = (1 - t)x + ty$. A complete CAT(0) space is called a Hadamard space and we shall denote it by $(\mathcal{H}, d)$, whereas the symbol $(X, d)$ will be used for general metric spaces. An alternative approach to define Hadamard spaces, which is not discussed herein, is via comparison triangles; see [16], [56].

The class of Hadamard spaces comprises Hilbert spaces, complete simply connected Riemannian manifolds of non-positive sectional curvature (for instance classic hyperbolic spaces and the manifold of positive definite matrices), Euclidean buildings, CAT(0) complexes, non-linear Lebesgue spaces, the Hilbert ball and many other spaces. We would like to point out that Hadamard spaces considered in the present paper do not have to be locally compact.\(^1\)

As we shall see, Hadamard spaces share many properties with Hilbert spaces. For instance, metric projections onto convex closed sets are non-expansive, Hadamard spaces have Enflo type 2, we will encounter analogs of weak convergence, the Kadec–Klee property, the Banach–Saks property, the Opial property and the finite intersection property (reflexivity). The analogy extends to more advanced topics, for instance, strongly continuous semigroups of non-expansive operators. There are however also remarkable differences between Hilbert and Hadamard spaces. A convex continuous function on a Hadamard space does not have to be locally Lipschitz, a weakly convergent sequence does not have to be bounded,\(^2\) there are non-convex Chebyshev sets, it is not known whether the closed convex hull of a compact set is compact, and the relation between weak convergence and weak topology is much more subtle.

As already alluded to above, Hadamard spaces turn out to constitute a natural framework for convexity theory. A set $C \subset \mathcal{H}$ is called convex if $(1 - t)x + ty \in C$ whenever $x, y \in C$ and $t \in (0, 1)$. A function $f : \mathcal{H} \to (-\infty, \infty]$ is called convex if $f \circ \gamma : [0, 1] \to (-\infty, \infty]$ is convex for every geodesic $\gamma : [0, 1] \to \mathcal{H}$. The trivial function $f \equiv \infty$ is of course convex, but we will (usually without explicit mentioning) exclude it from our considerations. This is to avoid repeating “and we assume that our function $f$ attains a finite value.” The domain of the function $f$ is the set $\text{dom}(f) := \{x \in \mathcal{H} : f(x) < \infty\}$. As common in convex analysis and optimization, we work with lower semicontinuous (lsc, for short) functions. The role of convexity in the whole theory of Hadamard spaces is crucial; see the authoritative monograph by Bridson and Haefliger [56]. This fact has motivated systematic study of convex analysis in Hadamard spaces summarized in [16]. See also Jost’s earlier book [112].

---

1 The only exceptions are Subsects. 6.1 and 6.4 where local compactness is required.

2 That is why we require boundedness in the definition.
Examples of convex functions include the distance function, Busemann functions, a displacement function associated with an isometry, energy functionals, the objective function in the Fermat–Weber optimal facility location problem and the objective function in the Fréchet mean problem. For a detailed list of convex functions in Hadamard spaces, see [16, Section 2.2].

A special attention will be paid to functions which are given as finite sums of convex lsc functions. Let \( f : H \rightarrow (-\infty, \infty] \) be a convex lsc function of the form

\[
(2) \quad f := \sum_{n=1}^{N} f_n,
\]

where \( f_n : H \rightarrow (-\infty, \infty] \) is a convex lsc function for each \( n = 1, \ldots, N \). As we shall see below, such functions play important roles in both theory and applications.

A basic notion in convex analysis and optimization is that of strong convexity. A function \( f : H \rightarrow (-\infty, \infty] \) is called strongly convex if there exists \( \beta > 0 \) such that

\[
f ((1-t)x + ty) \leq (1-t)f(x) + tf(y) - \beta t(1-t)d(x,y)^2,
\]

for every \( x, y \in H \) and \( t \in [0,1] \). A prime example of a strongly convex function is \( f := d(\cdot,x)^2 \), for some fixed \( x \in H \), as one can see from (1). If \( f : H \rightarrow (-\infty, \infty] \) is a strongly convex lsc function, then there exists \( z \in H \) such that

\[
(3) \quad f(z) + \beta d(z,x)^2 \leq f(x),
\]

for every \( x \in H \); see [16, Proposition 2.2.17]. In particular, each strongly convex function has a unique minimizer. Inequality (3) will appear in the sequel in several special instances (4), (14), (16) and (24).

As a first use of strong convexity, we define metric projections onto convex closed sets; see [16, p.33] or [56, p.176]. Given a convex closed set \( C \subset H \), recall that its indicator function is defined by

\[
\iota_C(x) := \begin{cases} 
0, & \text{if } x \in C, \\
\infty, & \text{if } x \notin C.
\end{cases}
\]

It is easy to see that the indicator function \( \iota_C : H \rightarrow (-\infty, \infty] \) is convex and lsc. Since the function \( f := \iota_C + d(\cdot,x)^2 \), where \( x \in H \), is strongly convex and lsc, we obtain immediately from (3) the following important result.
Theorem 1.1 (Metric projection). Let \((\mathcal{H}, d)\) be a Hadamard space. Assume that \(C \subset \mathcal{H}\) is a convex closed set and \(x \in \mathcal{H}\). Then there exists a unique point \(y \in C\) such that

\[
d(x, y) = \inf_{c \in C} d(x, c).
\]

We denote the point \(y\) by \(\text{proj}_C(x)\). Furthermore, we have

\[
d(x, \text{proj}_C(x))^2 + d(\text{proj}_C(x), c)^2 \leq d(x, c)^2,
\]

for every \(c \in C\), which can be viewed as a Pythagorean inequality.

See [56, Proposition 2.4, p.176] and [16, Theorem 2.1.12].

Definition 1.2 (Metric projection). Let \(C \subset \mathcal{H}\) be a convex closed set. The mapping \(\text{proj}_C : \mathcal{H} \to C\) from Theorem 1.1 is called the metric projection onto the set \(C\).

Many Hadamard space theorems surveyed in the present paper were later proved in some form in more general geodesic spaces, for instance, CAT(1) spaces. We will do our best to provide references to the relevant literature, but our aim here is to present Hadamard space results merely.

Let’s get started…

2. Metric and topological structure

This section is devoted to structural properties of Hadamard spaces. We first focus on various results related to the CAT(0) metric and then turn to a topology which is weaker than the topology induced by the metric.

2.1. Busemann spaces vs. CAT(0)

Apart from CAT(0), there is a weaker notion of non-positive curvature for geodesic metric spaces. It is due to Busemann [60]. We say that a geodesic metric space \((X, d)\) is a Busemann space if we have

\[
2d(\gamma(1/2), \eta(1/2)) \leq d(\gamma(1), \eta(1)),
\]

for every pair of geodesics \(\gamma : [0, 1] \to X\) and \(\eta : [0, 1] \to X\) such that \(\gamma(0) = \eta(0)\). It is easy to see that geodesics in Busemann spaces are uniquely determined by their endpoints. It also follows that a CAT(0) space is a Busemann space.

For instance, all strictly convex Banach spaces are Busemann. On the other hand, Hilbert spaces are the only Banach spaces which are CAT(0); see [56, Proposition 1.14].
It turns out that the difference between Busemann spaces and CAT(0) spaces can be neatly expressed by the Ptolemy inequality. We say that a metric space satisfies the Ptolemy inequality, or that it is Ptolemaic, if
\[ d(x_1, x_3) d(x_2, x_4) \leq d(x_1, x_2) d(x_3, x_4) + d(x_2, x_3) d(x_4, x_1) \]
for every \( x_1, \ldots, x_4 \in X \).

We can now state the promised result. It is due to Foertsch, Lytchak and Schroeder [83, Theorem 1.3].

**Theorem 2.1.** A geodesic metric space is a CAT(0) space if and only if it is Busemann and Ptolemaic.

We may now ask whether there exist other conditions which can be imposed on a Busemann space to make it CAT(0). We propose the following; see Theorem 3.1 below.

**Problem 2.2.** Let \((X, d)\) be a complete Busemann space. Assume that, given a closed convex set \( C \subset X \), the metric projection \( \text{proj}_C : X \to C \) is a well-defined single-valued non-expansive mapping. Under which conditions is then \((X, d)\) a Hadamard space?

A possible answer to Problem 2.2 in linear spaces would be: under the condition \( \dim \geq 3 \). Indeed, recall a classic result of Kakutani [114, Theorem 3], which says that the non-expansiveness of (linear) projections in Banach spaces of dimension at least 3 implies that the space is Hilbert. This was later studied also by Phelps [188, Theorem 5.2]. A deep study of non-expansiveness of projections in metric spaces and its relation to the symmetry of orthogonality has been recently carried out by Kell [117].

The following question was raised by Ivanov and Lytchak [106, Question 1].

**Problem 2.3.** Given a Busemann space, is there a CAT(0) space naturally related to it?

It is for instance known that, given a Busemann space, one cannot in general find a CAT(0) space which is affinely homeomorphic to it (even in linear spaces), or affinely isomorphic to it [106, Section 6]. One can view Problem 2.3 also like a non-linear analog of the celebrated Dvoretzky theorem [81, Theorem 6.15] from Banach space geometry.

Let us now briefly discuss the notion of (locally) Busemann space in (non Riemannian) manifolds, which is relevant in connection with Problem 2.3. Recall that a locally geodesic space is locally Busemann if each point has a neighborhood which is a Busemann space. Kristály, Varga and Kozma [139] showed that if a Finsler manifold is a Berwald space of non-positive flag curvature, then it is a locally Busemann space. It is
also known that Berwald spaces are locally Busemann spaces if and only if their flag curvature is non-positive. This theorem is due to Kristály and Kozma [137]. Ivanov and Lytchak have recently proved in [106, Theorem 2] that if a Finsler manifold is a locally Busemann space, then it is a Berwald space of non-positive flag curvature; see also [138]. Finally we are getting to a result which sheds some light on Problem 2.3. Ivanov and Lytchak showed that a Finsler manifold is a locally Busemann space if and only if the metric is affinely equivalent to a Riemannian metric of non-positive curvature; see [106, Corollary 1.3].

To close this part, we present an interesting question by Lytchak and Stadler [155, Question 1.3].

**Problem 2.4.** Given a CAT(0) space, is there a CAT(-1) metric on it defining the same topology?

### 2.2. Quadratic and non-quadratic inequalities

We will now proceed by presenting other characterizations of CAT(0). It is known that a geodesic metric space \((X,d)\) is CAT(0) if and only if it satisfies

\[
d(x_1, x_3)^2 + d(x_2, x_4)^2 \leq d(x_1, x_2)^2 + d(x_2, x_3)^2 + d(x_3, x_4)^2 + d(x_4, x_1)^2,
\]

for every \(x_1, \ldots, x_4 \in X\), or, yet equivalently if and only if it satisfies

\[
d(x_1, x_3)^2 + d(x_2, x_4)^2 \leq d(x_1, x_2)^2 + d(x_2, x_3)^2 + 2d(x_3, x_4)d(x_4, x_1),
\]

for every \(x_1, \ldots, x_4 \in X\). See [16, Theorem 1.3.3]. The fact that inequality (7)—and therefore also (6)—is satisfied in Hadamard space goes back to Reshetnyak [197].

The inverse implication, that is, the fact that (6) implies CAT(0) is a deep result due to Berg and Nikolaev [36, Theorem 6]. A simpler proof was later given by Sato [203]. This characterization of Hadamard spaces answers a question of Gromov [92, Section 1.19+], who asked for a condition implying CAT(0), which would be meaningful also for discrete spaces.

Let us also mention that inequality (6) holds for instance for the metric space \((X,d_1^\frac{1}{2})\) where \((X,d)\) is an arbitrary metric space, and for ultrametric spaces; see [203, Example 1.2].

Inequality (6) also means that Hadamard spaces have roundness 2. Recall that Enflo [76] defined that a metric space \((X,d)\) has roundness \(p\), where \(p > 0\), if \(p\) is the greatest number satisfying

\[
d(x_1, x_3)^p + d(x_2, x_4)^p \leq d(x_1, x_2)^p + d(x_2, x_3)^p + d(x_3, x_4)^p + d(x_4, x_1)^p,
\]

for every \(x_1, \ldots, x_4 \in X\).
for every \(x_1, \ldots, x_4 \in X\). Each metric space has roundness \(\geq 1\), by virtue of the triangle inequality, and geodesic metric spaces have roundness \(\leq 2\). On account of (6), \(\text{CAT}(0)\) spaces have roundness 2. It is also worth mentioning that roundness was in connection with \(\text{CAT}(0)\) spaces and groups studied by Lafont and Prassidis [142].

Enflo’s notion of roundness lead Bourgain, Milman and Wolfson to the definition of the Enflo type of a metric space [48, 3.14]; see also [77, Proposition 3].

**Definition 2.5 (Enflo type 2).** A metric space \((X, d)\) has Enflo type 2 if there exists a constant \(K \geq 1\) such that for every \(N \in \mathbb{N}\) and \(\{x_\varepsilon\}_{\varepsilon \in \{-1, 1\}^N} \subset X\) we have

\[
\sum_{\varepsilon \in \{-1, 1\}^N} d(x_\varepsilon, x_{-\varepsilon})^2 \leq K^2 \sum_{\varepsilon \sim \varepsilon'} d(x_\varepsilon, x_{\varepsilon'})^2,
\]

where \(\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)\) and \(\varepsilon \sim \varepsilon'\) stands for \(\sum_{i=1}^N |\varepsilon_i - \varepsilon_i'| = 2\). The least such a constant \(K\) will be denoted by \(E_2(X)\).

For instance, a Ptolemaic metric space \((X, d)\) has Enflo type 2 with \(E_2(X) = \sqrt{3}\); see [176, Proposition 5.3]. An iterative application of (6) yields the following result; see [77, Proposition 3] and [176, Proposition 5.2].

**Corollary 2.6.** Let \((X, d)\) be a \(\text{CAT}(0)\) space. Then it has Enflo type 2 with \(E_2(X) = 1\).

In this connection we would like to mention a recent spectacular solution to a long-standing problem in Banach space geometry by Ivanisvili, van Handel and Volberg [105], which shows that the Rademacher and Enflo types coincide.

An important generalization of (6) has been obtained in [5, Lemma 27] by Andoni, Naor and Neiman and we state it as Lemma 2.7 below. Before presenting their result, we fix the notation. Let \(n \in \mathbb{N}\) and \(s_1, \ldots, s_n, t_1, \ldots, t_n \in (0, 1)\) with \(\sum_{i=1}^n s_i = \sum_{j=1}^n t_j = 1\). Let further \((a_{ij})_{i,j=1}^n\) and \((b_{ij})_{i,j=1}^n\) be matrices with entries in \([0, \infty)\) that satisfy

\[
\sum_{k=1}^n a_{ik} + \sum_{k=1}^n b_{kj} = s_i + t_j,
\]

for every \(i, j = 1, \ldots, n\).

**Lemma 2.7.** Let \((\mathcal{H}, d)\) be a Hadamard space. Then for every \(x_1, \ldots, x_n \in \mathcal{H}\) we have

\[
\sum_{i,j=1}^n \frac{a_{ij} b_{ij}}{a_{ij} + b_{ij}} d(x_i, x_j)^2 \leq \sum_{i,j=1}^n s_i t_j d(x_i, x_j)^2.
\]
Lemma 2.7 has a number of special cases. It gives various quadratic inequalities including (6) as well as non-quadratic inequalities (5) and (7). As a matter of fact, unlike the original proof of (5) from [83], the proof of Lemma 2.7 was achieved without comparisons between geodesic and Euclidean triangles.

There is however a more general phenomenon discovered in [5] which we are now going to discuss. Let us first introduce the terminology from [5]. Given $n \in \mathbb{N}$ and two matrices $A = (a_{ij})_{i,j=1}^{n}, B = (b_{ij})_{i,j=1}^{n}$ with entries in $[0, \infty)$, we say that a metric space $(X, d)$ satisfies the $(A, B)$-quadratic metric inequality if

$$\sum_{i,j=1}^{n} a_{ij} d(x_i, x_j)^2 \leq \sum_{i,j=1}^{n} b_{ij} d(x_i, x_j)^2,$$

for every $x_1, \ldots, x_n \in X$.

We say that a metric space $(X, d_X)$ admits a bi-Lipschitz embedding into a metric space $(Y, d_Y)$ if there exist $D \geq 1$ and $r > 0$ and a mapping $f : X \rightarrow Y$ such that

$$(10) \quad r \cdot d_X(x, y) \leq d_Y(f(x), f(y)) \leq D \cdot r \cdot d_X(x, y),$$

for every $x, y \in X$. Given a bi-Lipschitz embedding $f : X \rightarrow Y$, the infimum of $D$, for which (10) holds true, is called the distortion of $f$. The infimum of distortions over all bi-Lipschitz embeddings $f : X \rightarrow Y$ is denoted by $c_Y(X)$.

The following is a special case of [5, Proposition 3].

Proposition 2.8. Let $(X, d)$ be a metric space with $|X| = n$ for some $n \in \mathbb{N}$, and $D \in [1, \infty)$. Then we have

$$\inf \{ c_{\mathcal{H}}(X) : \mathcal{H} \text{ is a Hadamard space} \} \leq D$$

if and only if, given two $(n \times n)$-matrices $A = (a_{ij}), B = (b_{ij})$ with entries in $[0, \infty)$ such that every Hadamard space satisfies the $(A, B)$-quadratic metric inequality, we have that $(X, d)$ satisfies the $(A, D^2B)$-quadratic metric inequality.

This implies that subsets of Hadamard spaces are fully characterized by quadratic inequalities. In particular, each non-quadratic metric inequality can be obtained as a consequence of quadratic metric inequalities. We have seen a manifestation of this phenomenon above where Lemma 2.7 implied non-quadratic inequalities (5) and (7).

Gromov formulated the following problem in [93, §15(b)].

Problem 2.9. Characterize finite metric spaces, which have an isometric embedding into Hadamard spaces.
A very recent paper of Toyoda [213] provides solutions for some special cases; see also an alternative proof by Lebedeva and Petrunin [145]. It is natural to extend the above problem of Gromov to bi-Lipschitz embeddings; see [5, 1.4].

**Problem 2.10.** Characterize finite metric spaces, which have a bi-Lipschitz embedding into Hadamard spaces.

We refer the interested reader to the Andoni–Naor–Neiman paper [5] for detailed discussions and further open problems in this area.

Let us now take a look at bi-Lipschitz embeddings of a finite Hadamard space subset into a Hilbert space. We use the usual asymptotic notation: \( f(n) = O(g(n)) \) means there exist \( n_0 \in \mathbb{N} \) and \( C > 0 \) such that \( f(n) \leq Cg(n) \) for each \( n \geq n_0 \), and \( f(n) = \Omega(g(n)) \) means the same as \( g(n) = O(f(n)) \). Finally, \( f(n) = \Theta(g(n)) \) means that both \( f(n) = O(g(n)) \) and \( f(n) = \Omega(g(n)) \).

Bourgain [46], [47] proved that an arbitrary metric space with cardinality \( n \), where \( n \in \mathbb{N} \), can be embedded into a Hilbert space with distortion \( O(\log n) \). He also showed that each \( n \)-point subset of a metric tree admits such an embedding with distortion \( \Theta(\sqrt{\log \log n}) \). These results were later obtained also by Matoušek [157], who used different methods. See also more recent proofs due to Linial and Saks [151] and due to Kloeckner [123]. Let us now ask the following question. What is the minimal \( D \) so that each \( n \)-point subset of a Hadamard space embeds into a Hilbert space with distortion \( D \)? We can conclude from the above discussion that the answer to this question is between \( \sqrt{\log \log n} \) and \( \log n \). As a matter of fact, it is known that \( D = \Theta(\log n) \). This follows from the work of Kondo [132] on expander graphs (see also closely related results of Gromov [94]), which was later extended by Mendel and Naor [159, Theorem 1.1].

For *coarse* embeddings into Hadamard spaces, the interested reader is referred to a very recent paper by Eskenazis, Mendel and Naor [78].

We close this part by mentioning the work of Chalopin, Chepoi and Naves [62] on isometric embeddings of Busemann surfaces (that is, non-positively curved two dimensional surfaces) into \( L_1 \). For further interesting properties of Busemann surfaces, see [63] by Chepoi, Estellon and Naves.

### 2.3. Weak convergence

Since we want to avoid assuming that our Hadamard spaces are locally compact, basic to our considerations is the notion of weak convergence. While local compactness is a standard assumption in geometry and geometric group theory [56], our standpoint is that of (functional) analysis.
Jost introduced weak convergence in Hadamard spaces [109, Definition 2.7]. Given a geodesic \( \gamma: [0, 1] \to \mathcal{H} \), its range \( \gamma([0, 1]) \) is a closed convex set and we can hence consider the metric projection \( \text{proj}_{\gamma([0,1])} \); see Definition 1.2.

We say that a bounded sequence \( (x_n) \subset \mathcal{H} \) weakly converges to a point \( x \in \mathcal{H} \) if

\[
\text{proj}_{\gamma([0,1])} (x_n) \to x, \quad \text{as } n \to \infty,
\]

whenever \( \gamma: [0, 1] \to \mathcal{H} \) is a geodesic with \( x \in \gamma([0, 1]) \). Weak convergence is denoted by \( x_n \overset{w}{\to} x \). One can obviously extend this definition for trajectories \( u: (0, \infty) \to \mathcal{H} \) and write \( u(t) \overset{w}{\to} x \) as \( t \to \infty \). This will be used in gradient flow theory below.

In Hilbert spaces, the above definition of weak convergence coincides with the standard one. A well-known consequence of the Banach–Steinhaus theorem is that a weakly convergent sequence in a Banach space must be bounded. In contrast, in Hadamard spaces we have to put the boundedness assumption into the definition.

Weak convergence is related to an older notion of asymptotic centers. Given a bounded sequence \( (x_n) \subset \mathcal{H} \), define the function

\[
\psi(x, (x_n)) := \limsup_{n \to \infty} d(x, x_n)^2, \quad x \in \mathcal{H}.
\]

Since \( \psi \) is strongly convex, it has a unique minimizer, which we call the asymptotic center of the sequence \( (x_n) \). The definition, in a slightly different form, of asymptotic centers in uniformly convex Banach spaces is due to Edelstein [75]. The following characterization appeared in [79, Proposition 5.2].

**Proposition 2.11.** A bounded sequence \( (x_n) \subset \mathcal{H} \) weakly converges to a point \( x \in \mathcal{H} \) if and only if \( x \) is the asymptotic center of each subsequence of \( (x_n) \).

Like in Hilbert spaces, weak convergence in Hadamard spaces has the following three properties, which are of great importance for the theory.

(i) Each bounded sequence has a weakly convergent subsequence.

(ii) If \( C \subset \mathcal{H} \) is a convex closed set and \( (x_n) \subset C \) a sequence weakly converging to a point \( x \in \mathcal{H} \), then \( x \in C \).

(iii) If \( f: \mathcal{H} \to (-\infty, \infty] \) is a convex lsc function and \( (x_n) \subset \mathcal{H} \) a sequence weakly converging to a point \( x \in \mathcal{H} \), then \( f(x) \leq \liminf_{n \to \infty} f(x_n) \).

The fact in (i) was proved by Jost in [109, Theorem 2.1]. However, using the (easy-to-show) equivalence from Proposition 2.11, one can obtain a
shorter proof of (i); see [16, Proposition 3.1.2]. The proofs of both (ii) and (iii) are rather simple, they appeared in [24, Lemma 3.1] and [14, Lemma 3.1], respectively.

Next we look at an analog of the (weak) Banach–Sachs property. To this end we need the notion of a barycenter of a probability measure from Definition 3.15. Recall also that δₙ stands for the Dirac measure at a point x ∈ ℍ.

**Theorem 2.12 (Banach–Saks property).** Let (ℍ, d) be a Hadamard space and (xₙ) ⊂ ℍ a sequence weakly converging to a point x ∈ ℍ. Then there exists a subsequence (xₙₖ) of (xₙ) such that the barycenters

\[
\text{bar} \left( \frac{1}{k} \sum_{i=1}^{k} \delta_{x_{n_i}} \right)
\]

converge to x.

The weak Banach–Saks property is known to be enjoyed by Hilbert spaces [81]. Its counterpart in Hadamard spaces presented in Theorem 2.12 above was first stated by Jost [109, Theorem 2.2], however it is not clear to me whether his proof is correct. Yokota [223, Theorem C] later proved (a more general version of) Theorem 2.12 by different methods and we therefore know that it holds true.

The weak convergence in Hadamard spaces shares with the Hilbert space weak convergence also other properties, for instance, the uniform Kadec–Klee property, or the Opial property; see [16, Chapter 3]. Furthermore, it has been shown very recently by Lytchak and Petrunin [153, Theorem 1.2] that the weak convergence of sequences in Hadamard spaces is induced by a topology.

**Theorem 2.13 (Weak topology).** Let (ℍ, d) be a Hadamard space. Then there exists a topology τ on ℍ such that for a given bounded sequence (xₙ) ⊂ ℍ and a point x ∈ ℍ we have xₙ \(\xrightarrow{w}\) x if and only if xₙ \(\xrightarrow{τ}\) x.

By proving the above theorem, Lytchak and Petrunin resolved a well-known problem formulated by Kirk and Panyanak in [122, p. 3696]. Additionally, they showed that the existence of the weak topology in Hadamard spaces is indeed a very subtle issue in the sense that an analogous assertion does not hold for nets; see [153, Proposition 1.3 and Theorem 1.4]. This is of course in sharp contrast with the Hilbert space setting, where the weak convergence of nets is induced by the standard weak topology.

---

3 It mimicks the Banach space proof from the Goebel–Kirk book [89, Lemma 15.2].
4 I can see neither how one obtains the last estimate in the proof of [109, Theorem 2.2], nor how the right hand side thereof vanishes.
In this connection we also point out that in Hadamard spaces there exists yet another interesting topology introduced by Monod [164, Definition 13], which is defined as the weakest topology which makes every bounded convex closed set compact.

3. Convex sets and functions

In this section we concentrate upon basic concepts related to convexity. We start by recalling an important property of metric projections onto closed convex sets.

**Theorem 3.1.** Let \((\mathcal{H}, d)\) be a Hadamard space and \(C \subset \mathcal{H}\) be a closed convex set. Then the metric projection \(\text{proj}_C\) is a non-expansive (that is, 1-Lipschitz) mapping.

Theorem 3.1 is a classic result; see [56, p.176] or [16, Theorem 2.1.12]. It is easy to observe that the metric projection is even firmly non-expansive; [16, p.35]. Recall the definition.

**Definition 3.2 (Firmly non-expansive mapping).** A mapping \(F : \mathcal{H} \to \mathcal{H}\) is called firmly non-expansive if, given two points \(x, y \in \mathcal{H}\) we have that the function

\[
t \mapsto d((1-t)x + tFx, (1-t)y + tFy), \quad t \in [0,1],
\]

is non-increasing.

Another example of a firmly non-expansive mapping is the proximal mapping; see Subsect. 3.5. Yet another example is the resolvent \(R_\lambda\) associated to a non-expansive mapping, which we define below in (27); see [16, Lemma 4.2.2].

Firm non-expansiveness in a Hilbert space \(H\) is important in both theory and algorithms [32]. It has also connections to various parts of non-linear analysis and optimization. For instance, a mapping \(F : H \to H\) is firmly non-expansive if and only if there exists a maximal monotone operator \(A : H \to 2^H\) such that \(F = (I + A)^{-1}\), that is, \(F\) is the resolvent of \(A\); see [32, Corollary 23.9]. Here of course the symbol \(I\) stands for the identity operator on \(H\).

Let now \((\mathcal{H}, d)\) be a Hadamard space and \(F : \mathcal{H} \to \mathcal{H}\) a firmly non-expansive mapping which has a fixed point. It is easy to show that, given a point \(x \in \mathcal{H}\), the sequence \((F^n x)\) weakly converges to a fixed point; see [16, Exercise 6.2].
3.1. Compactness of convex hulls

It is well known that, given a compact subset $K \subset X$ of a Banach space $X$, its closed convex hull $\overline{\text{co}}
 K$ is compact; see for instance [81, Exercise 1.62]. This fact motivates the following question, which can be traced back to Gromov [91, 6.B1(f)].

Problem 3.3. Let $(\mathcal{H}, d)$ be a Hadamard space and $K \subset \mathcal{H}$ be a compact set. Is the set $\overline{\text{co}}
 K$ compact?

Kopecká and Reich showed that it would be sufficient to solve the above problem for finite sets; see [133, Theorem 2.10] and the subsequent discussion. The same fact also appears in Duchesne’s recent work [74, Lemma 2.18].

Lemma 3.4. If a Hadamard space $(\mathcal{H}, d)$ has the property that, given a finite set $\{x_1, \ldots, x_N\} \subset \mathcal{H}$, its closed convex hull $\overline{\text{co}}
 \{x_1, \ldots, x_N\}$ is compact, then, for each compact set $K \subset \mathcal{H}$, its closed convex hull $\overline{\text{co}}
 K$ is compact.

In other words, Problem 3.3 reduces to the following problem.

Problem 3.5. Let $(\mathcal{H}, d)$ be a Hadamard space and $x_1, \ldots, x_N \in \mathcal{H}$. Is the set $\overline{\text{co}}
 \{x_1, \ldots, x_N\}$ compact?

The above question is trivial for $N = 1$ and $N = 2$, but seems very difficult already for $N = 3$.

3.2. Convex hulls of extreme points

Let $A \subset \mathcal{H}$ and $x \in A$. We say that $x$ is an extreme point of $A$ if $x = \frac{1}{2}y + \frac{1}{2}z$ implies $y = z$. It is easy to see that, given a compact convex set $C \subset \mathcal{H}$, it has an extreme point and moreover, the set $C$ is the closed convex hull of its extreme points; see [74, Proposition 2.24]. A much more delicate question is whether each closed convex bounded set $C \subset \mathcal{H}$ has an extreme point [74, Question 2.23]. The answer is of course yes in Hilbert spaces—by virtue of the Krein–Milman theorem. However, the answer is in general negative due to a recently constructed counterexample of Monod [165].
3.3. Reflexivity of Hadamard spaces

We now look at an analog of reflexivity. It is a standard fact in functional analysis that a Banach space $X$ is reflexive if and only if each non-increasing family $(C_i)_{i \in I}$ of bounded closed convex sets $C_i$, indexed by an arbitrary directed set $I$, has non-empty intersection; see [81, Theorem 3.111 and Exercise 3.166]. The following Hadamard space version can be found in [144, Proposition 5.2], [143, Lemma 2.2] and [88, Lemma 2.2]; see also [16, Proposition 2.1.16] and [16, Exercise 2.11].

**Proposition 3.6.** Let $(H,d)$ be a Hadamard space. If $(C_i)_{i \in I}$ is a non-increasing family of bounded closed convex sets in $H$, where $I$ is an arbitrary directed set, then $\bigcap_{i \in I} C_i \neq \emptyset$.

Consequently, an arbitrary family $(C_i)_{i \in I}$ of bounded closed convex sets has the following property. If $\bigcap_{i \in F} C_i \neq \emptyset$ for each finite $F \subset I$, then $\bigcap_{i \in I} C_i \neq \emptyset$. The Hilbert ball case of Proposition 3.6 appears in [90, Theorem 18.1].

3.4. Ultrapowers of Hadamard spaces and convexity

Given a Hadamard space $(H,d)$ and a non-principal ultrafilter $\omega$, denote the ultrapower by $(H^\omega,d^\omega)$. The metric projection $\text{proj}_H : H^\omega \to H$ is then a 1-Lipschitz retraction and we say that Hadamard spaces are 1-complemented in their ultrapowers.

Let $f : H \to (-\infty, \infty]$ be a convex lsc function and define

$$(11) \quad f_\omega(x_\omega) := \inf \{ \omega\text{-lim}_n f(x_n) : (x_n) \in x_\omega \}, \quad x_\omega \in H^\omega.$$ 

This definition appeared in Stojkovic’s paper [206, Definition 2.26]. Note that $f_\omega$ is of course convex. Alexander Lytchak proved that the function $f_\omega$ is also lsc, which answered [16, Question 2.2.16]. I would like to thank Alexander for sending me his proof and for his kind permission to reproduce it here.

**Proposition 3.7.** Let $(H,d)$ be a Hadamard space and $f : H \to (-\infty, \infty]$ be a lsc function. Then the function $f_\omega$ defined in (11) is lsc.

**Proof.** We need to show that, for each $\alpha \in \mathbb{R}$, the sublevel set

$$A_\alpha := f_\omega^{-1}((-\infty, \alpha]) \subset H^\omega$$

is closed. To this end let us fix $\alpha \in \mathbb{R}$ and set

$$B_{\alpha+\varepsilon} := f^{-1}((-\infty, \alpha + \varepsilon]) \subset H,$$
for every $\varepsilon > 0$. Since $f$ is lsc by assumption, the sublevel sets $B_{\alpha+\varepsilon}$ are closed and consequently their ultrapowers $B_{\omega}^{\omega}_{\alpha+\varepsilon}$ are closed in $\mathcal{H}^{\omega}$. To complete the proof it suffices to show that

$$A_{\alpha} = \bigcap_{\varepsilon > 0} B_{\omega}^{\omega}_{\alpha+\varepsilon}.$$ 

From (11) we get that a point $x_{\omega} \in \mathcal{H}^{\omega}$ lies in $A_{\alpha}$ if and only if for each $\varepsilon > 0$ there exists a sequence $(x_{n}) \in x_{\omega}$ with

$$\omega-lim_{n} f(x_{n}) \leq \alpha + \varepsilon.$$ 

However the latter condition is equivalent with $x_{\omega} \in B_{\omega}^{\omega}_{\alpha+\varepsilon}$ for every $\varepsilon > 0$, which finishes the proof.

The fact that $\mathcal{H} \subset \mathcal{H}^{\omega}$ is a $1$-retract is a crucial ingredient in the Plateau problem in Hadamard spaces [218, Theorem 1.6], [219, Theorem 1.3]. In order to state this very interesting result here, we would have to first introduce the general theory of metric currents in metric spaces due to Ambrosio and Kirchheim [4]. Instead, we refer the reader to the original papers and propose the following problem.

**Problem 3.8.** Is it possible to prove [219, Theorem 1.3] using weak convergence instead of ultrapowers?

Note that Stojkovic’s proof of the Lie–Trotter–Kato formula (see Theorem 4.9 below) from [206, Theorem 4.4] also relies on ultrapowers of Hadamard spaces. In the simplified proof [17], weak convergence was used instead of the ultrapowers. That was our motivation for raising Problem 3.8.

### 3.5. Moreau envelopes and proximal mappings

Let $f : \mathcal{H} \to (-\infty, \infty]$ be a convex lsc function. Given $\lambda > 0$, the *Moreau envelope* of $f$ is the function $f_{\lambda} : \mathcal{H} \to \mathbb{R}$ defined by

$$f_{\lambda}(x) := \inf_{y \in \mathcal{H}} \left[ f(y) + \frac{1}{2\lambda} d(x,y)^2 \right], \quad x \in \mathcal{H}.$$ 

This function is a “regularization” of $f$ and has the same set of minimizers as $f$. It is not difficult to see that Moreau envelopes are convex; see [16, Exercise 2.8]. It is also claimed in [16, p. 42] that the Moreau envelope is not in general lsc, but it is completely incorrect: it was later showed in [22, Lemma 2.1] that Moreau envelopes are locally Lipschitz.
A closely related notion is the proximal mapping. Given \( x \in \mathcal{H} \), consider the function
\[
(12) \quad f + \frac{1}{2} d(x, \cdot)^2.
\]
This function is strongly convex and hence has a unique minimizer, which we denote by \( \text{prox}_f(x) \). The mapping \( \text{prox}_f : \mathcal{H} \to \mathcal{H} \) is called the proximal mapping associated to the function \( f \). We usually work with an entire family of proximal mappings parametrized by a parameter \( \lambda > 0 \), that is, we consider the proximal mapping \( \text{prox}_{\lambda f} \) of the function \( \lambda f \) for each \( \lambda > 0 \).

Proximal mappings were introduced by Moreau in Hilbert spaces [166], [167], [168], and then used also in metric spaces (see e.g. Attouch’s book [11]). Jost [113], assuming additionally that \( f \) is non-negative, and Mayer [158] used them later in Hadamard spaces.\(^5\)

Proximal mappings were often in the Hadamard space literature (including [16]) called resolvents and denoted by \( J_\lambda \). This terminology comes from Hilbert spaces where the proximal mapping coincides with the resolvent of the convex subdifferential. Indeed, given a convex lsc function \( f : H \to (-\infty, \infty] \) on a Hilbert space \( H \) and \( \lambda > 0 \), we have
\[
\text{prox}_{\lambda f}(x) = (I + \lambda \partial f(x))^{-1},
\]
for each \( x \in H \), where \( I : H \to H \) stands for the identity mapping and \( \partial \) denotes the convex subdifferential.\(^6\)

Moreau envelopes and proximal mappings are related by the equality
\[
f_\lambda(x) = f(\text{prox}_{\lambda f}(x)) + \frac{1}{2\lambda} d(x, \text{prox}_{\lambda f}(x))^2,
\]
which holds for each \( x \in \mathcal{H} \) and \( \lambda > 0 \).

A prime example of proximal mappings is the metric projection. Indeed, let \( C \subset \mathcal{H} \) be a convex closed set and \( i_C : \mathcal{H} \to (-\infty, \infty] \) be its indicator function. Then \( \text{prox}_{i_C} = \text{proj}_C \).

Formula (13) below expresses the dependence of the proximal mapping on the parameter \( \lambda \) and is often referred to as the resolvent identity. It was established by Mayer [158, Lemma 1.10] and independently (for non-negative functions) by Jost [113, Corollary 1.3.8].

**Proposition 3.9.** Let \( f : \mathcal{H} \to (-\infty, \infty] \) be convex lsc. Then
\[
(13) \quad \text{prox}_{\lambda f}(x) = \text{prox}_{\mu f}
\left( \frac{\lambda - \mu}{\lambda} \text{prox}_{\lambda f}(x) + \frac{\mu}{\lambda} x \right), \quad x \in \mathcal{H}.
\]

\(^5\) As a matter of fact, the proof of the existence of \( \text{prox}_{\lambda f}(x) \) for a general convex lsc function \( f \) in [110, Lemma 2] does not show that the function in (12) is bounded from below and works therefore only for \( f \) bounded from below. Such an assumption is however missing in the statement of [110, Lemma 2].

\(^6\) As for the notation \( J_\lambda \) vs. \( \text{prox}_{\lambda f} \), I do not mind using either, but since some mathematicians strongly prefer the latter, I stick to it here.
for every $\lambda > \mu > 0$.

The following important result was obtained independently by Jost [110, Lemma 4] and Mayer [158, Lemma 1.12].

**Theorem 3.10.** The proximal mapping is non-expansive.

Proposition 3.9 allows to improve upon Theorem 3.10 and show that the proximal mapping is even firmly non-expansive [16, p.45].

Applying variational inequality (3) to the strongly convex function (12) leads to the following important inequality

$$f(\text{prox}_\lambda f(x)) + \frac{1}{2\lambda} d(x, \text{prox}_\lambda f(x))^2 + \frac{1}{2\lambda} d(y, \text{prox}_\lambda f(x))^2 \leq f(y) + \frac{1}{2\lambda} d(x, y)^2,$$

for each $y \in \mathcal{H}$. This inequality plays important roles in continuous- and discrete-time gradient flows; see Sects. 4 and 6, respectively.

### 3.6. Mosco convergence

We will be now concerned with sequences of convex lsc functions on a Hadamard space $(\mathcal{H}, d)$. Given $n \in \mathbb{N}$, let $f_n: \mathcal{H} \to (-\infty, \infty]$ be a convex lsc function and denote the Moreau envelope of $f_n$ by $f_{n,\lambda}$. We study the Mosco convergence of the sequence $(f_n)$ to some convex lsc function $f: \mathcal{H} \to (-\infty, \infty]$. Let us recall the definition [11], [169].

**Definition 3.11 (Mosco convergence).** We say that $(f_n)$ converges to $f$ in the sense of Mosco if, given $x \in \mathcal{H}$,

(i) for every sequence $(x_n) \subset \mathcal{H}$ such that $x_n \xrightarrow{w} x$, we have $f(x) \leq \liminf_{n \to \infty} f_n(x_n)$, and

(ii) there exists a sequence $(y_n) \subset \mathcal{H}$ such that $y_n \to x$ and $f(x) \geq \limsup_{n \to \infty} f_n(y_n)$.

We shall use the notation $f_n \xrightarrow{M} f$ for the Mosco convergence. If we replace weak convergence in the above definition by strong one, we obtain the definition of De Giorgi’s $\Gamma$-convergence [69]. It is easy to see that the Mosco convergence preserves convexity and the limit function is always lsc even if the functions $f_n$ are not.

Kuwae and Shioya showed that the Mosco convergence of non-negative convex lsc functions implies the pointwise convergence of their Moreau envelopes [141, Proposition 5.12]. The same result for general convex lsc function was proved in [18, Theorem 4.1]. The converse implication was later established in [22, Theorem 3.2]. We hence arrive at the following.
**Theorem 3.12.** Let $(H, d)$ be a Hadamard space and $f_n : H \to (-\infty, \infty]$ be a convex lsc function for each $n \in \mathbb{N}$. Then the following statements are equivalent:

(i) There exists a convex lsc function $f : H \to (-\infty, \infty]$ such that $f_n \xrightarrow{M} f$ as $n \to \infty$.

(ii) There exists a convex lsc function $f : H \to (-\infty, \infty]$ such that $f_{n,\lambda}(x) \to f_\lambda(x)$ as $n \to \infty$, for every $\lambda > 0$ and $x \in H$.

This theorem enables to introduce the Mosco topology on the set of convex lsc functions, which is a topology inducing the Mosco convergence; see [22, Section 4] for the details. Another application of Theorem 3.12 is the equivalence between the Frolík–Wijsman convergence and the Mosco convergence of sets. Let us recall a sequence of convex closed sets $C_n \subset H$, where $n \in \mathbb{N}$, converges to a convex closed set $C \subset H$ in the sense of Frolík and Wijsman if $d(x, C_n) \to d(x, C)$ for each $x \in H$. The Mosco convergence of sets is defined via indicator functions, that is, a sequence of convex closed sets $C_n \subset H$ Mosco converges to a convex closed set $C \subset H$ if the indicator functions $\iota_{C_n} \xrightarrow{\text{Mosco}} \iota_C$. A direct consequence of Theorem 3.12, observed in [22, Corollary 3.4], is that the Frolík–Wijsman and Mosco convergences are equivalent. A Banach space version of Theorem 3.12 can be found in [11, Theorem 3.33].

In [113, Definition 1.4.2] Mosco convergence on Hadamard spaces for non-negative functions is defined by the pointwise convergence of their Moreau envelopes. This is claimed to be equivalent to the pointwise convergence of the proximal mappings; this is however not true even on $\mathbb{R}$. Indeed, consider for instance the sequence of constant functions $0, 1, 0, 1, \ldots$ on $\mathbb{R}$, which does not Mosco converge, but the corresponding proximal mappings are all equal to the identity mapping. On the other hand it is known that, under an additional normalization condition, the pointwise convergence of proximal mappings in linear spaces implies the Mosco convergence; see [11, Theorem 3.26]. In Hadamard spaces, only one implication is known.

**Theorem 3.13.** Let $f_n : H \to (-\infty, \infty]$ be a convex lsc function for each $n \in \mathbb{N}$. If $f_n \xrightarrow{M} f$, then $\text{prox}_{\lambda f_n}(x) \to \text{prox}_{\lambda f}(x)$ for every $\lambda > 0$ and $x \in H$.

This theorem was proved in [18, Theorem 4.1] and an earlier version for non-negative functions is due to Kuwae and Shioya [141, Proposition 5.12]. It is therefore natural to raise the following question. See [16, Question 5.2.5].
Problem 3.14. Does the convergence
\[ \text{prox}_{\lambda f_n}(x) \to \text{prox}_{\lambda f}(x), \] for each \( x \in \mathcal{H} \) and \( \lambda > 0 \), imply, under some additional condition, that \( f_n \overset{M}{\to} f \)?

Research along these lines has already begun and first results were provided by Bördellima [35].

3.7. Barycenters of probability measures

Let \((X,d)\) be a metric space and \( p \in [1, \infty) \). Let \( \mathcal{P}^p(X) \) be the set of probability measures \( \mu \) on \( X \) such that
\[ \int_X d(x,y)^p \, d\mu(y) < \infty, \]
for some/every \( x \in X \).

Definition 3.15 (Barycenter). Let \( \mu \in \mathcal{P}^2(\mathcal{H}) \). Then the function
\[ y \mapsto \int_\mathcal{H} d(x,y)^2 \, d\mu(x), \quad y \in \mathcal{H}, \]
is strongly convex and we call its unique minimizer the barycenter of \( \mu \) and denote it by \( \text{bar}(\mu) \).

The existence and uniqueness of barycenters of \( \mathcal{P}^2(\mathcal{H}) \)-measures were first proved by Korevaar and Schoen [134, Lemma 2.5.1]. Barycenters of finitely supported measures on Hadamard manifolds were, however, studied already by Cartan; see [56, p.178]. Note also that barycenters can be defined for more general probability measures, namely \( \mathcal{P}^1(\mathcal{H}) \) measures; see Sturm’s paper [212, Proposition 4.3].

Given \( \mu \in \mathcal{P}^2(\mathcal{H}) \), we denote its variance by
\[ \text{var}(\mu) := \inf_{z \in \mathcal{H}} \int_\mathcal{H} d(x,z)^2 \, d\mu(x). \]

With this notation, inequality (3), applied to the strongly convex function in (15), reads
\[ \text{var}(\mu) + d(y, \text{bar}(\mu))^2 \leq \int_\mathcal{H} d(x,y)^2 \, d\mu(x), \quad y \in \mathcal{H}. \]

The above definitions lead naturally to the notions of the expectation and variance of Hadamard space valued random variables, which are basic building blocks for non-linear probability theory developed chiefly by
Sturm. We refer the reader to [16, Chapter 7] for a gentle introduction to this interesting subject field, and to the references therein for more advanced topics. Interestingly, Hadamard spaces can be fully characterized among complete metric spaces by probability measures [212, Theorem 4.9].

Relatedly, ergodic theorems in Hadamard spaces are due to Austin [12] and Navas [171]. In the erratum [13] to [12], the author corrects his proof of [12, Lemma 2.3], which says that the barycenter mapping $\bar{\cdot} : P^2(H) \to H$ is a $1$-Lipschitz mapping with respect to the $2$-Wasserstein distance, by proving a stronger statement with the $1$-Wasserstein distance. This fact was however proved already by Sturm [212, Theorem 6.3] and reproduced in [16, Proposition 2.3.10].

We also note that barycenters of probability measures in CAT(1) spaces have been recently studied by Yokota [223], [224].

We finish this section by remarking that some notions from convex analysis in Hilbert spaces have not been sufficiently developed in Hadamard spaces yet, for instance convex subdifferentials or the Fenchel duality theory. However research along these lines has already begun; see for instance [170] and the references therein.

4. Continuous-time gradient flows

Let $H$ be a real Hilbert space and $f : H \to (-\infty, \infty]$ be a convex lsc function. Consider the following gradient flow problem

\begin{align}
-\dot{x}(t) &\in \partial f (x(t)), \text{ for a.e. } t \in (0, \infty), \\
x(0) &= x_0,
\end{align}

where $x_0 \in H$ is a given initial value at time $t = 0$. This Cauchy problem and its numerous generalizations have been well understood for a long time. We refer the reader to Brezis’ classic book [52] and also recommend a more recent survey paper by Peypouquet and Sorin [187], which features the approach of Kobayashi [124]. We will now be interested in extending Cauchy problem (17) into Hadamard spaces.

4.1. Gradient flow semigroups

Motivated by the classic construction of Crandall and Liggett [68] in Banach spaces, Mayer showed that it is possible to define gradient flow semigroups for convex lsc functions in Hadamard spaces [158, Theorem 1.13].

---

7 even for $P^1(H)$ measures
8 Mayer’s paper was originally his thesis defended at the University of Utah in 1995.
Theorem 4.1. Let $(\mathcal{H}, d)$ be a Hadamard space and $f: \mathcal{H} \to (-\infty, \infty]$ be a convex lsc function. Then, given $t > 0$, we can define a mapping $S_t: \mathcal{H} \to \mathcal{H}$ by

$$S_t x := \lim_{k \to \infty} (\text{prox}_{\frac{1}{k} f})^k x, \quad x \in \overline{\text{dom} f},$$

where the limit is uniform in $t$ on bounded subintervals of $(0, \infty)$.

Independently of Mayer, gradient flows for convex functions on Hadamard spaces were studied by Jost [113]. Being interested in energy functionals, Jost formulates the statement of Theorem 4.1 only for non-negative convex lsc functions [113, Theorem 1.3.13] and uses (like Mayer) the Crandall–Liggett approach to construct the semigroup (18). In my opinion, however, Jost’s proof is incorrect: at a crucial step, there is a limit operation involving a quantity $l_F(x_n)$, which may tend to infinity as $n \to \infty$, and does not give the desired conclusion in [113, (1.3.17)].

We note that semigroups of operators in hyperbolic geometries were studied by Reich and Shafrir [195] and by Reich and Shoikhet [196]. Nowadays, gradient flow theory in metric spaces is a hot topic in analysis. We refer the interested reader to the authoritative book on the subject by Ambrosio, Gigli and Savaré [3]. A Hadamard space version of their proof of Theorem 4.1 appeared in [16, Theorem 5.1.6].

The following theorem summarizes the basic properties of the flow [113], [158], its proof can be found for instance in [16, Proposition 5.1.8].

Theorem 4.2. The family $(S_t)$ is a strongly continuous semigroup of non-expansive mappings, that is,

- the mapping $S_t: \mathcal{H} \to \mathcal{H}$ is non-expansive for each $t > 0$,
- $S_{s+t} = S_s \circ S_t$, for every $s, t > 0$,
- $S_t x \to x$ as $t \to 0$, for each $x \in \overline{\text{dom} f}$.

The following regularity result is due to Mayer [158, Theorem 2.9]; the proof can be found also in [16, Proposition 5.1.10].

Proposition 4.3 (Regularity of a flow). Let $f: \mathcal{H} \to (-\infty, \infty]$ be a convex lsc function and $x_0 \in \overline{\text{dom} f}$. Then the mapping $t \mapsto S_t x_0$ is locally Lipschitz on $(0, \infty)$ and Lipschitz on $[t_0, \infty)$ where $t_0$ is an arbitrary positive time.

Like in Hilbert spaces, the gradient flow is completely determined by the evolution variational inequality; see [3, Theorem 4.0.4] or [16, Theorem 5.1.11].

Theorem 4.4 (Evolution variational inequality). Let $(\mathcal{H}, d)$ be a Hadamard space and $f: \mathcal{H} \to (-\infty, \infty]$ be a convex lsc function. Assume
\( x_0 \in \overline{\text{dom } f} \) and denote \( x_t := S_t x_0 \) for \( t \in (0, \infty) \). Then \( t \mapsto x_t \) is absolutely continuous on \((0, \infty)\) and satisfies

\[
\frac{1}{2} \frac{d}{dt} d(y, x_t)^2 + f(x_t) \leq f(y),
\]

for almost every \( t \in (0, \infty) \) and every \( y \in \text{dom } f \). Conversely, if an absolutely continuous curve \( z : (0, \infty) \to H \) with \( \lim_{t \to 0^+} z(t) = x_0 \) satisfies (19), then \( z_t = S_t x_0 \) for every \( t \in (0, \infty) \).

Next we would like to express the fact that gradient flows move in the direction of the steepest descent. To this end, we need the notion of a slope. Given a convex lsc function \( f : H \to (-\infty, \infty] \), its slope, denoted by \( |\partial f| \), is a mapping \( |\partial f| : H \to [0, \infty) \) given by

\[
|\partial f|(x) := \sup_{y \in H \setminus \{x\}} \frac{\max\{f(x) - f(y), 0\}}{d(x, y)}, \quad x \in \text{dom } f,
\]

and by \( |\partial f|(x) := \infty, \) for \( x \in H \setminus \text{dom } f \). We can now state the promised results, they are all due to Mayer [158]. Proofs can be found also in [16, Theorem 5.1.13].

**Theorem 4.5.** Let \((H, d)\) be a Hadamard space and \( f : H \to (-\infty, \infty] \) be a convex lsc function. Given \( x_0 \in \overline{\text{dom } f} \), put \( x_t := S_t x_0 \). Then,

\[
|\partial f|(x_t) = \lim_{h \to 0^+} \frac{d(x_{t+h}, x_t)}{h},
\]

as well as,

\[
|\partial f|(x_t) = \lim_{h \to 0^+} \frac{f(x_t) - f(x_{t+h})}{d(x_{t+h}, x_t)},
\]

and also,

\[
|\partial f|(x_t)^2 = \lim_{h \to 0^+} \frac{f(x_t) - f(x_{t+h})}{h},
\]

for every \( t \in (0, \infty) \).

Note also that by applying Theorem 4.1 in Hilbert spaces, we recover classic solutions to Cauchy problem (17), as pointed out already by Mayer [158, Section 2.7].

We will now take a look at gradient flows for a sequence of functions. Let \((H, d)\) be a Hadamard space and \( f_n : H \to (-\infty, \infty] \) be a convex lsc function, for each \( n \in \mathbb{N} \), and let also \( f : H \to (-\infty, \infty] \) be convex lsc. The following was established in [18, Theorem 4.6].
Theorem 4.6. If $\text{prox}_{\lambda f_n}(x) \to \text{prox}_{\lambda f}(x)$ for every $\lambda > 0$ and $x \in \mathcal{H}$, then $S_{n,t}x \to S_t x$ for every $t > 0$ and $x \in \mathcal{H}$. Here $S_{n,t}$ stands for the gradient flow semigroup associated with the function $f_n$.

Combining Theorems 3.13 and 4.6 gives that the Mosco convergence of functions implies the pointwise convergence of the corresponding gradient flow semigroups. In Hilbert spaces, one can show that the convergence in Theorem 4.6 is uniform in $t$ on each bounded time interval [52, Theorem 4.2], which leads to the following question.

Problem 4.7. Is the convergence in Theorem 4.6 uniform in $t$ on each bounded time interval?

Finally, we investigate the asymptotic behavior of gradient flows, which was in Hilbert spaces established by Bruck [59]. The following theorem was proved in [14, Theorem 1.5].

Theorem 4.8. Let $(\mathcal{H}, d)$ be a Hadamard space and $f : \mathcal{H} \to (-\infty, \infty]$ a convex lsc function which attains its minimum on $\mathcal{H}$. Given $x \in \overline{\text{dom } f}$, there exists $y \in \text{dom } f$, a minimizer of $f$, such that we have $S_t x \xrightarrow{w} y$ as $t \to \infty$.

We know that the convergence in Theorem 4.8 is not strong in general even in $\ell_2$ by an example of Baillon [25].

Before turning to applications, we note that gradient flows in Alexandrov spaces (with lower curvature bound) were studied by Perelman and Petrunin [185] and by Petrunin [186]. The theory was later improved by Lytchak [152]. In this connection we recommend a new book by Alexander, Kapovich and Petrunin [2].

4.2. Application: Donaldson’s conjecture

The asymptotic behavior of gradient flows in Hadamard spaces established in Theorem 4.8 has become an important tool for attacking a conjecture of Donaldson in Kähler geometry [72]. Let $(M, \omega)$ be a compact Kähler manifold. The set of Kähler potentials

$$\mathcal{K}_\omega = \{ \phi \in \mathcal{C}^\infty(M, \mathbb{R}) : \omega_\phi := \omega + \sqrt{-1} \partial \bar{\partial} \phi > 0 \},$$

which corresponds to the set of smooth Kähler metrics in the cohomology class $[\omega]$, can be equipped with an $L^2$-metric and completed with respect to it. We then obtain a Hadamard space $\mathcal{H}_\omega$. When we extend the Mabuchi energy functional from $\mathcal{K}_\omega$ onto $\mathcal{H}_\omega$, we get a convex lsc functional on a Hadamard space and we can study its gradient flows. It is known that gradient flows for the Mabuchi energy are the solutions to the
Calabi equation, that is, the Calabi flows. Donaldson’s conjecture states (roughly speaking) that the Calabi flow exists for all times and converges to a constant scalar curvature metric, which is a minimizer of the Mabuchi energy. The Hadamard space approach to Donaldson’s conjecture was initiated by Streets [208], who applied Theorem 4.8 to (the extension of) the Mabuchi energy on $\mathcal{H}_\omega$ and obtained a result on the asymptotic behavior of the Calabi flow. Subsequently, Berman, Darvas, and Lu [42] developed these ideas further by identifying concretely the new elements in $\mathcal{H}_\omega \setminus \mathcal{H}_\omega$, which “gives the first concrete result about the long time convergence of this flow on general Kähler manifolds, partially confirming a conjecture of Donaldson”.\(^9\) For further results along these lines, the interested reader is referred to Xia’s paper [222]. In a similar way, the convergence Theorem 4.8 impacted analogous results in Riemannian geometry; see the paper [97] by Gursky and Streets.

### 4.3. Lie–Trotter–Kato formula

Let $f : \mathcal{H} \to (-\infty, \infty]$ be a convex lsc function of the form (2). Our goal is to approximate the gradient flow semigroup $S_t$ for the function $f$ by the proximal mappings $\text{prox}_{\lambda f_1}, \ldots, \text{prox}_{\lambda f_N}$ as opposed to by the proximal mapping $\text{prox}_{\lambda f}$ like in Theorem 4.1. The Lie–Trotter–Kato formula addresses this issue.

**Theorem 4.9.** Under the above notation, we have

$$S_t x = \lim_{k \to \infty} (\text{prox}_{\lambda f_N} \circ \cdots \circ \text{prox}_{\lambda f_1})^k x, \quad x \in \text{dom } f.$$  

Theorem 4.9 is due to Stojkovic [206, Theorem 4.4]. Its proof was later simplified in [17]. A linear space version of Theorem 4.9 is due to Kato and Masuda [116]. There have been however many other closely related results in functional analysis including seminal works by Brezis and Pazy [54], [55], Chernoff [65], Kato [115], Miyadera and Ôhara [163], Reich [189], [190], [191] and Trotter [214], [215]. We would also like to mention the result of Clément and Maas [66]. Ohta and Pálfia later studied the Lie–Trotter–Kato formula in CAT(1) spaces [179].

**Problem 4.10.** Is the limit in (23) uniform in $t$ on bounded time intervals? If [206, Theorem 3.12] and [206, Theorem 3.13] are correct, then yes, as remarked in [17].

\(^9\) Quoted from the abstract of their paper [42].
4.4. Application: Retractions in finite subset space

We will now show an application of the Lie–Trotter–Kato formula into metric geometry. This application is somewhat unexpected because the original problem has nothing to do with gradient flows.

Let \((X,d)\) be a metric space. Given \(n \in \mathbb{N}\), denote by \(X(n)\) the family of subsets of \(X\) which have cardinality at most \(n\). When equipped with the Hausdorff distance \(d_H\), the metric space \((X(n), d_H)\) is called a finite subset space. Unlike Cartesian products \(X^n\) or the space of unordered \(n\)-tuples \(X^n/S_n\), finite subset spaces admit canonical isometric embeddings \(e: X(n) \to X(n+1)\), which makes them an interesting object of study. In their 1931 paper \([45]\), Borsuk and Ulam studied finite subset spaces associated to the unit interval \(I := [0,1]\), that is, sets \(I(n)\) for \(n \in \mathbb{N}\). They showed that if \(n = 1, 2, 3\), then \(I(n)\) is homeomorphic to \(I^n\), whereas if \(n \geq 4\), then \(I(n)\) is not homeomorphic to any subset of \(\ell_2^n\), the \(n\)-dimensional Euclidean space. In the same paper they also raised the following question:

**Problem 4.11.** Given \(n \in \mathbb{N}\), is \(I(n)\) homeomorphic to a subset of \(\ell_2^{n+1}\)?

The above question remains widely open. If the answer is negative, we can ask the following.

**Problem 4.12.** What is the smallest \(m \in \mathbb{N}\) such that \(I(n)\) homeomorphically embeds into \(\ell_2^m\)?

It would be also interesting to find out whether bi-Lipschitz embeddings are possible.

**Problem 4.13.** Given \(n \in \mathbb{N}\), what is the smallest \(m \in \mathbb{N}\) such that \(I(n)\) embeds into \(\ell_2^m\) in a bi-Lipschitz way?

Also determining the optimal distortion seems to be entirely unexplored.

**Problem 4.14.** Given \(n\) and \(m\), what is the distortion of an optimal bi-Lipschitz embedding of \(I(n)\) into \(\ell_2^m\)?

We want to however focus on problems directly related to Hadamard spaces, in particular, on the existence of Lipschitz retraction \(r: \mathcal{H}^n \to \mathcal{H}^{n-1}\), where \((\mathcal{H}, d)\) is a given Hadamard space and \(n \geq 2\). This problem was addressed in \([21]\), where it was shown that such a Lipschitz retraction indeed exists for each \(n \geq 2\). Interestingly, it was defined via a gradient flow for a convex functional \(F\) on \(\mathcal{H}^n\) given by

\[
F(x) := \sum_{1 \leq i < j \leq n} d(x_i, x_j), \quad x = (x_1, \ldots, x_n) \in \mathcal{H}^n,
\]

\[10 \text{ Instead of finite subset space, Borsuk and Ulam use the term symmetric product.}\]
and in order to control the gradient flow trajectory, we employed the Lie–
Trotter–Kato formula. The Lipschitz constant of the resulting retraction
is $O(n^2)$. We also note that Kovalev [136] had earlier showed that, given
a Hilbert space $H$ and $n \geq 2$, there exists a Lipschitz retraction $r: H(n) \to
H(n-1)$ with Lipschitz constant $O(n^{1.5})$. The following questions were
asked already in [21].

**Problem 4.15.** Let $(\mathcal{H}, d)$ be a Hadamard space and $n \geq 2$. Is there a
Lipschitz retraction $r: \mathcal{H}(n) \to \mathcal{H}(n-1)$ with smaller Lipschitz constant
than $O(n^2)$? In particular, is there a chance for $O(1)$?

We finish this section by mentioning a very recent and interesting con-
tribution to the theory of gradient flows in Hadamard space due to Ohta
[177], who studied the self-contracted property of gradient flow trajecto-
ries in the sense of Daniilidis, Ley and Sabourau [71, Definition 1.2]. Recall
that a curve $c: I \to X$ defined on an interval $I \subset [0, \infty)$ with values in a
metric space $(X, d)$ is self-contracted if

$$d(c(t_1), c(t_3)) \geq d(c(t_2), c(t_3)), \quad \text{for every } t_1, t_2, t_3 \in I \text{ with } t_1 \leq t_2 \leq t_3.$$ 

We can now state Ohta’s result [177, Section 4.3] here.

**Theorem 4.16.** Let $(\mathcal{H}, d)$ be a Hadamard space and $f: \mathcal{H} \to (-\infty, \infty]$ be a convex lsc function. Given $x \in \text{dom } f$, the gradient flow trajectory $t \mapsto S_t x: (0, \infty) \to \mathcal{H}$ is self-contracted.

Let us also mention that Daniilidis et al. [70, Proposition 4.16] es-
ablished the self-contracted property for discrete-time gradient flows in
Euclidean spaces, which of course extends to continuous-time gradient
flows.

5. Harmonic mappings

This section is devoted to harmonic mappings $h: M \to \mathcal{H}$, where $M$ is
a measure space and $\mathcal{H}$ is Hadamard. There are several ways to define
harmonic mappings in this setting (energy minimization, Markov opera-
tors, martingales, composition with a function) and their mutual relations
are not clear yet. To our knowledge, the study of harmonic mappings into
singular spaces was initiated by Gromov and Schoen in their celebrated
paper [95]. Since then, numerous authors have contributed including Ko-
revaar and Schoen [134], [135], Jost [109], [110], [111], Sturm [209], [210],
[211], Ohta [175] and Fuglede [84], [85]. A counterexample to Jost’s proof
of the uniqueness of harmonic mappings [112, Chapter 4] was provided by
Mese in [160, Remarks 5.3 and 5.7, Example 5.4].
Finally, a very interesting approach to harmonic mappings is due to Wang [217], which was later improved by Izeki and Nayatani [108]. In this connection we also mention the work of Izeki, Kondo and Nayatani [107] on harmonic mappings and random groups. One of the main motivations for the theory of Hadamard space harmonic mappings have been rigidity theorems; this goes back to the above mentioned paper by Gromov and Schoen [95].

We focus here on Sturm’s approach from [210] and refer the interested reader to the original papers for information on the other theories.

Let \((M, \mathcal{M}, \mu)\) be a space with \(\sigma\)-algebra \(\mathcal{M}\) and measure \(\mu\). As common in analysis, we consider mappings which differ on a null set as identical. We will now follow [210, Definition 3.1]. Given a measurable mapping \(h: M \rightarrow \mathcal{H}\), we define

\[
L^2(M, \mathcal{H}, h) := \{ f: M \rightarrow \mathcal{H} \text{ measurable: } d(h(\cdot), f(\cdot)) \in L^2(M) \},
\]

and equip this space with the \(L^2\)-metric

\[
d_2(f, g) := \int_M d(f(x), g(x))^2 \, d\mu(x), \quad f, g \in L^2(M, \mathcal{H}, h).
\]

Then \(L^2(M, \mathcal{H}, h)\) with the metric \(d_2\) becomes a Hadamard space; see [210, Proposition 3.3].

We assume that \((M, \mathcal{M}, \mu)\) is equipped with a Markov kernel \(p(x, dy)\) which is symmetric with respect to \(\mu\), that is, \(p(x, dy)\mu(dx) = p(y, dx)\mu(dy)\). Given a measurable mapping \(f: M \rightarrow \mathcal{H}\), its energy density is defined by

\[
e_f(x) := \int_M d(f(x), f(y))^2 p(x, dy), \quad x \in M.
\]

If \(e_f \in L^1(M)\), we say that \(f\) belongs to \(W^2(M, \mathcal{H})\). Furthermore, we define the energy of \(f\) by

\[
E(f) := \frac{1}{2} \int_M e_f \, d\mu.
\]

Observe that \(E(f) < \infty\) if and only if \(f \in W^2(M, \mathcal{H})\). It is not difficult to see that, given \(h \in W^2(M, \mathcal{H})\), we have \(L^2(M, \mathcal{H}, h) \subset W^2(M, \mathcal{H})\); see [210, Lemma 3.2].

Given measurable mappings \(f, g: M \rightarrow \mathcal{H}\), we define a quadratic form \(Q\) by

\[
Q(g, f) := \int_M \int_M d(g(x), f(y))^2 p(x, dy) \, d\mu(x),
\]

and the variance of \(f\) is

\[
V(f) := \inf \{ Q(h, f): h \in L^2(M, \mathcal{H}, f) \}.
\]
Observe that \( E(f) = \frac{1}{2} Q(f, f) \).

Given \( h \in W^2(M, \mathcal{H}) \), the function \( Q(\cdot, h) \) is strongly convex on \( L^2(M, \mathcal{H}, h) \) and therefore it has a unique minimizer \( Ph \in L^2(M, \mathcal{H}, h) \). The mapping \( P : W^2(M, \mathcal{H}) \to L^2(M, \mathcal{H}, h) \) is called a Markov operator. A direct consequence of (3) is the following variation inequality

\[
V(f) + d_2(Pf, g)^2 \leq Q(g, f),
\]

for each \( g \in L^2(M, \mathcal{H}, f) \); see [210, Lemma 4.1].

Another important property of the Markov operator is non-expansiveness [210, Theorem 5.2].

**Theorem 5.1.** Let \( h : M \to \mathcal{H} \) be a measurable mapping. Then \( d_2(Pf, Pg) \leq d_2(f, g) \), for every \( f, g \in L^2(M, \mathcal{H}, h) \).

It would be interesting to answer the following question.

**Problem 5.2.** Is the Markov operator \( P : W^2(M, \mathcal{H}) \to L^2(M, \mathcal{H}, h) \) firmly non-expansive?

If \( f \) has separable range, the Markov operator can be equivalently introduced by the following pointwise definition

\[
\tilde{P}f(x) := \text{bar} (f_\ast p(x, \cdot)), \quad x \in \mathcal{H},
\]

where \( f_\ast p(x, \cdot) \) is the pushforward of the probability measure \( p(x, \cdot) \) under the mapping \( f \). See [210, Corollary 4.7] and the subsequent discussion.

The following theorem, which characterizes the fixed points of the Markov operator as the minimizers of the energy functional, is due to Sturm; see [210, Theorem 6.6]. An analogous result is presented for the pointwise defined Markov operator \( \tilde{P} \) in [112, Lemma 4.1.1], but its proof is in my opinion incorrect, since it uses a pointwise operation within an integral from which the conclusion does not follow.

**Theorem 5.3.** Let \( f \in W^2(M, \mathcal{H}) \). Then \( Pf = f \) if and only if \( E(f) \leq E(g) \) for each \( g \in L^2(M, \mathcal{H}, f) \).

We will now look at the gradient flow for the energy \( E \), which was studied by Sturm in [210, Section 8]. Let \( f \in W^2(M, \mathcal{H}) \). Then \( E \) is convex and continuous on the Hadamard space \( L^2(M, \mathcal{H}, f) \) and one can hence apply Theorem 4.1 to obtain the gradient flow

\[
t \mapsto S_t f, \quad t \in (0, \infty).
\]

An alternative approach to (heat) flows relying on the Markov operator instead of on the energy functional was introduced in [23]. We first need to recall some abstract theory; see [23], [206] for the details.
Let \((\mathcal{H}, d)\) be a Hadamard space and \(F: \mathcal{H} \to \mathcal{H}\) be non-expansive. Given \(x \in \mathcal{H}\) and \(\lambda \in (0, \infty)\), the mapping

\[
G_{x, \lambda}: y \mapsto \frac{1}{1 + \lambda} x + \frac{\lambda}{1 + \lambda} F y, \quad y \in \mathcal{H},
\]

is a contraction with Lipschitz constant \(\frac{\lambda}{1 + \lambda}\) and hence has a unique fixed point, which will be denoted by \(R_\lambda x\). The mapping \(x \mapsto R_\lambda x\) will be called the resolvent of \(F\).

The following important theorem is due to Stojkovic [206, Theorem 3.10].

**Theorem 5.4.** Let \((\mathcal{H}, d)\) be a Hadamard space and \(F: \mathcal{H} \to \mathcal{H}\) be a non-expansive mapping. Then the limit

\[
T_t x := \lim_{n \to \infty} (R_t)^n x, \quad x \in \mathcal{H},
\]

exists and is uniform with respect to \(t\) on each bounded subinterval of \((0, \infty)\). Moreover, the family \((T_t)_t\) is a strongly continuous semigroup of non-expansive mappings, that is,

(i) \(\lim_{t \to 0^+} T_t x = x\),
(ii) \(T_t (T_s x) = T_{s+t} x\), for every \(s, t > 0\),
(iii) \(d(T_t x, T_t y) \leq d(x, y)\), for each \(t > 0\),

for every \(x, y \in \mathcal{H}\).

The following theorem was established in [23, Theorem 1.6]. Its Hilbert ball version is due to Reich [192].

**Theorem 5.5.** Let \(F: \mathcal{H} \to \mathcal{H}\) be a non-expansive mapping with at least one fixed point and let \(x_0 \in \mathcal{H}\). Then \(T_t x_0\) weakly converges to a fixed point of \(F\) as \(t \to \infty\).

Before returning to harmonic mappings, we make a comment.

**Remark 5.6.** One can easily notice the formal similarity between Theorems 4.1 and 5.4. As a matter of fact, these two cases are neatly unified in Hilbert spaces via maximal monotone operator theory. Indeed, let \(H\) be a Hilbert space, \(f: H \to (-\infty, \infty]\) be convex lsc and \(F: H \to H\) be non-expansive. Then the convex subdifferential \(\partial f\) and \(I - F\) are the most important instances of maximal monotone operators. For further reading we recommend the survey paper by Peypouquet and Sorin [187] as well as the authoritative monograph by Bauschke and Combettes [32]. Maximal monotone operators were in Hadamard manifolds studied in [148], [149].
Next we apply the above abstract theory to the Markov operator. Given \( f \in W^2(M, H) \), the Markov operator \( P : L^2(M, H, f) \to L^2(M, H, f) \) is non-expansive and Theorem 5.4 provides us with the heat flow semigroup\(^{(29)}\)

\[ t \mapsto T_t f, \quad t > 0.\]

We remark that the exposition of this matter in [23, Section 6] refers—by mistake—to the Markov operator \( \tilde{P} \) (defined pointwise) instead of to the Markov operator \( P \). As we have noted above, the Markov operator \( \tilde{P} \) would be appropriate only for mappings with separable ranges, but we do not want to make any such restriction.

We now arrive at the following question; see [23, Conjecture 6.3].\(^{11}\)

**Problem 5.7.** Given \( f \in W^2(M, H) \), is it true that the gradient flow in (26) coincides with the heat flow in (29) in \( L^2(M, H, f) \)?

Note that one can choose a set \( D \in M \) and consider the Dirichlet problem on \( D \) in the above discussion; see [23], [210]. However, we chose \( D = M \) here for simplicity.

We finish this section by addressing the existence of harmonic mappings. Let us again first look at an abstract theory. If \( H \) is a Hilbert space, \( F : H \to H \) is non-expansive and \( x \in H \), define

\[ m_k := \frac{1}{k+1} \sum_{n=0}^{k} F^n x, \]

for each \( k \in \mathbb{N}_0 \). Assume that the sequence \( (m_k) \) is bounded. Then each weak cluster point of \( (m_k) \) is a fixed point of \( F \). In particular, the set of fixed points of \( F \) is non-empty. See [187, Lemma 4.7]. We now aim at extending this result into Hadamard spaces. To this end define

\[ m_k := \text{bar} \left( \frac{1}{k+1} \sum_{n=0}^{k} \delta_{F^n x} \right), \]

where \( F : \mathcal{H} \to \mathcal{H} \) is a non-expansive mapping on a Hadamard space \( (\mathcal{H}, d) \) and \( x \in \mathcal{H} \).

**Problem 5.8.** If the sequence \( (m_k) \) is bounded for some/every \( x \in \mathcal{H} \), is then its set of weak cluster points (which is non-empty) contained in \( \text{Fix } F \)? In particular, has \( F \) a fixed point?

We could then apply the above claim to the Markov operator and get the existence of a harmonic mapping. However, we would also have to determine for which Markov kernels \( p(x, dy) \) the boundedness assumption is satisfied.

\(^{11}\) In [23, Conjecture 6.3] it is incorrectly referenced to [210, Theorem 8.1] but it should be [210, Theorem 8.4]
6. Discrete-time gradient flows

Continuous-time gradient flows studied in Sect. 4 are constructed from their discrete-time approximations. It turns out, however, that the discrete-time approximate solutions are of interest on their own and can be used for numerical computations. Since one defines the discrete-time approximations by applying iteratively the proximal mapping, this method is called the \textit{proximal point algorithm} and abbreviated PPA. In linear spaces, it was introduced by Martinet [156] and extended further by Rockafellar [202] and Brézis and Lions [53]. From recent literature we recommend [67], [73], [183] which witness the increasing popularity of this optimization method. On Hadamard manifolds, the PPA was studied by Ferreira and Oliveira [82], Papa Quiroz and Oliveira [182] and Li, López and Martín-Márquez [148]. The algorithm has been recently studied in metric spaces by a number of authors including Zaslavski [226], Ohta and Pálfia [178], Kimura and Kohsaka [118], [119], [120], [121], Kohsaka [131] and Espínola and Nicolae [80].

Let us now formulate a convex optimization problem in Hadamard spaces and show how the PPA is used to find a minimizer of the objective function.

Let \((\mathcal{H}, d)\) be a Hadamard space and \(f : \mathcal{H} \to (-\infty, \infty]\) be a convex lsc function attaining its minimum. We would like to find a minimizer of \(f\). To this end, we choose a point \(x_0 \in \mathcal{H}\), and define

\[
(30) \quad x_n := \text{prox}_{\lambda_n f} (x_{n-1}),
\]

for each \(n \in \mathbb{N}\). The proximal point algorithm in Hadamard spaces was first studied in [14]. The proof of the following convergence theorem can be found in [14, Theorem 1.4] and [16, Theorem 6.3.1].

**Theorem 6.1 (Basic PPA).** Let \((\mathcal{H}, d)\) be a Hadamard space and \(f : \mathcal{H} \to (-\infty, \infty]\) be a convex lsc function attaining its minimum. Then, for an arbitrary point \(x_0 \in \mathcal{H}\) and a sequence of positive reals \((\lambda_n)\) such that \(\sum_{n=1}^{\infty} \lambda_n = \infty\), the sequence \((x_n) \subset \mathcal{H}\) defined in (30) weakly converges to a minimizer of \(f\).

In general, one cannot replace weak convergence in the above theorem by strong convergence. The first counterexample in Hilbert spaces is due to Güler [96]. For further examples, see [31], [33]. A convergence analysis of the PPA for uniformly convex functions was carried out by Leuștean and Sipoș [147]. An abstract version of the PPA has been recently introduced by Leuștean, Nicolae and Sipoș [146]. We would also like to mention recent related results of Reich and Salinas [193], [194].
6.1. Splitting proximal point algorithms

In many optimization problems, the objective function is given as the sum of finitely many functions. It turns out that we can benefit from taking this structure of the objective function into account and design much more efficient algorithms for such optimization problems. To be now more concrete, we will consider objective functions of the form \( f \). While the “basic” PPA in (30) would use the proximal mapping of the function \( f \), which is typically difficult to compute, the splitting PPA relies on the evaluation of the proximal mappings associated to the individual functions \( f_n \), for \( n = 1, \ldots, N \). We obtain two variants of the splitting PPA depending on whether we apply these proximal mappings in cyclic or random order. Sadly, in this section we need to require our Hadamard space to be locally compact; see Problem 6.4. The splitting PPA in Euclidean spaces was introduced by Bertsekas in his seminal paper [43].

Let \( (\lambda_k) \) be a sequence of positive reals and let \( x_0 \in \mathcal{H} \) be an arbitrary starting point. We now introduce the cyclic order version of the splitting PPA. For each \( k \in \mathbb{N}_0 := \{0\} \cup \mathbb{N} \), we set

\[
\begin{align*}
    x_{kN+1} &= \text{prox}_{\lambda_k f_1}(x_{kN}), \\
    x_{kN+2} &= \text{prox}_{\lambda_k f_2}(x_{kN+1}), \\
    &\vdots \\
    x_{kN+N} &= \text{prox}_{\lambda_k f_N}(x_{kN+N-1}).
\end{align*}
\]  

(31)

The following convergence theorem was obtained in [15, Theorem 3.4].

**Theorem 6.2 (Splitting PPA with cyclic order).** Let \((\mathcal{H}, d)\) be a locally compact Hadamard space and let \( f : \mathcal{H} \rightarrow (-\infty, \infty] \) be a function of the form (2) which attains its minimum. Let \((\lambda_k) \in \ell_2 \setminus \ell_1\) be a sequence of positive reals. Given \( x_0 \in \mathcal{H} \), let \( (x_j) \) be the sequence defined in (31). Assume there exists \( L > 0 \) such that

\[
\begin{align*}
    f_n(x_{kN}) - f_n(x_{kN+n}) &\leq L d(x_{kN}, x_{kN+n}), \\
    f_n(x_{kN+n-1}) - f_n(x_{kN+n}) &\leq L d(x_{kN+n-1}, x_{kN+n}),
\end{align*}
\]

for each \( k \in \mathbb{N}_0 \) and \( n = 1, \ldots, N \). Then \( (x_j) \) converges to a minimizer of \( f \).

One can observe immediately from the proof\(^{12}\) of Theorem 6.2 that, given \( k \in \mathbb{N}_0 \), it is not necessary to insist on applying the proximal mappings \( \text{prox}_{\lambda_k f_n} \), where \( n = 1, \ldots, N \), in the prescribed order. Indeed, a popular approach in optimization is to shuffle randomly the indices \( 1, \ldots, N \).

\(^{12}\) See the original paper [15, Theorem 3.4] or [16, Theorem 6.3.7].
at the beginning of each cycle and apply the proximal mappings in the newly obtained order.

We now turn to the random order version of the splitting PPA. Consider the probability space $\Omega := \{1, \ldots, N\}^{\mathbb{N}_0}$ equipped with the product of the uniform probability measure on $\{1, \ldots, N\}$ and let $(r_k)$ be the sequence of random variables

$$r_k(\omega) := \omega_k, \quad \omega = (\omega_1, \omega_2, \ldots) \in \Omega.$$ 

Let again $f: \mathcal{H} \to (-\infty, \infty]$ be a function of the form (2) and let $(\lambda_k)$ be a sequence of positive reals. Given a starting point $x_0 \in \mathcal{H}$, we put

(32) $$x_{k+1} := \text{prox}_{\lambda_k f r_k}(x_k), \quad k \in \mathbb{N}_0.$$ 

The following theorem was obtained in [15, Theorem 3.7]. Its proof relies on the supermartingale convergence theorem of Robbins and Siegmund [198]. To simplify the notation, we denote by $x_{n,k+1}$ the result of the iteration (32) if $r_k(\omega) = n$.

**Theorem 6.3 (Splitting PPA with random order).** Let $(\mathcal{H}, d)$ be a locally compact Hadamard space and let $f$ be a function of the form (2) which attains its minimum. Let $(\lambda_k) \in \ell_2 \setminus \ell_1$ be a sequence of positive reals. Given a starting point $x_0 \in \mathcal{H}$, let $(x_k)$ be the sequence defined in (32). Assume there exists $L > 0$ such that

$$f_n(x_k) - f_n(x_{n,k+1}) \leq L d(x_k, x_{n,k+1}),$$

for every $k \in \mathbb{N}_0$ and $n = 1, \ldots, N$. Then $(x_k)$ converges to a minimizer of $f$ almost surely.

We raise the following natural questions.

**Problem 6.4.** Can one extend Theorems 6.2 and 6.3 into separable Hadamard spaces without the local compactness assumption? In this case—of course—we can hope for weak convergence only. This problem is open even for separable Hilbert spaces.

The splitting PPA was later extended into the 1-dimensional sphere $S_1$, which is only a locally Hadamard space, and applied successfully in image restoration problems by Bergmann, Steidl and Weinmann [38], [40], [41]; see also the follow-up paper [20] and Steidl’s survey [205]. Ohta and Pálffia [178] then extended these minimization algorithms into domains of general Alexandrov spaces. In Hadamard manifolds, Bergmann, Persch and Steidl [39] used the Douglas–Rachford algorithm in convex minimization problems in image processing. Other interesting applications of convex optimization in Hadamard manifolds into image processing are due to
Bergmann et al. [37] and due to Neumayer, Persch and Steidl [172]. The splitting PPA has been recently used in the manifold setting by Bredies et al. [50] and by Storath and Weinmann [207].

The power of the splitting PPA is showcased in the following application.

6.2. Application: Computing medians and means

Given a finite number of points $a_1, \ldots, a_N \in \mathcal{H}$ and weights $w_1, \ldots, w_N \geq 0$ with $\sum w_n = 1$, consider the function

$$f(x) := \sum_{n=1}^{N} w_n d(x, a_n)^p, \quad x \in \mathcal{H},$$

where $p \in [1, \infty)$. This function is obviously convex and continuous. In particular, we are concerned with two important cases:

(i) If $p = 1$, then $f$ becomes the objective function in the Fermat–Weber problem for optimal facility location. It has always a minimizer, which is however in general not unique. We call it a \textit{(weighted) median} of the points $a_1, \ldots, a_N$.

(ii) If $p = 2$, then $f$ is strongly convex and has therefore a unique minimizer, which is called the \textit{(weighted) mean} of the points $a_1, \ldots, a_N$. Observe that the mean can be viewed as the barycenter of the probability measure

$$\mu := \sum_{n=1}^{N} w_n \delta_{a_n}.$$ 

In Hilbert spaces, the mean coincides with the arithmetic mean.

We could in principle minimize the function $f$ in (33) by the “basic” PPA, but there is no explicit formula for the proximal mapping. On the other hand it is easy to verify that the assumptions of Theorems 6.2 and 6.3 are satisfied for the function $f$ in (33) and we can therefore minimize $f$ by the splitting PPA. The main advantage of this approach is that the proximal mappings are very easy to evaluate, see [16, Section 8.3].

Computing medians in Hadamard space is of great importance in applications. Indeed, in diffusion tensor imaging one needs to compute means of positive definite matrices (which form a Hadamard manifold as we noted above). However, since the underlying space in this case is a manifold, we can alternatively use explicit gradient-based minimization methods to compute the means [184]. Another important application, where we need to compute means in a Hadamard space, is computational phylogenetic. In
this case the underlying Hadamard space has no differentiable structure and the only known possibility to compute means is the above method based on the splitting PPA. We will present more details in Subsect. 6.3.

6.3. Application: Averaging phylogenetic trees

Given a finite number, say \( n \), of entities (for instance genes, or species), biologists represent their evolutionary history by a phylogenetic tree. It is defined as a metric tree whose terminal vertices are labeled by \( 1, \ldots, n \). In their seminal paper [44], Billera, Holmes and Vogtmann equipped the set of such trees with the structure of a CAT(0) cubical complex and the resulting locally compact Hadamard space is now referred to as the BHV tree space. We denote it by \( \mathcal{T}_n \), where \( n \geq 3 \) stands for the number of terminal vertices.

In order to use the BHV tree space in practice, one needs an (efficient) algorithm for computing distances and geodesics. By “computing geodesics” we mean: given \( x, y \in \mathcal{T}_n \) and \( t \in (0, 1) \), compute the point \( (1 - t)x + ty \). Indeed, the lack of such an algorithm hindered the use of the BHV tree space for about a decade. Based on an earlier attempt of Owen [180], who came up with an exponential time algorithm, Owen and Provan [181] finally provided the desired algorithm with polynomial runtime.

In this connection we note that algorithms for computing distances and geodesics in CAT(0) cubical complexes were studied by Chepoi and Maftuleac [64] in two dimensions, and by Ardila, Owen and Sullivant [8] in arbitrary dimensions. These works were later improved by Hayashi [101] who introduced a polynomial time algorithm in arbitrary dimensions. Such algorithms in cubical CAT(0) complexes are of interest in, for instance, robotics; see the work of Ardila, Baker and Yatchak [6] and Ardila et al. [7] and the references therein.

Together with the Owen–Provan algorithm for computing distances and geodesics in the BHV tree space, the splitting PPA allows to compute medians and means of phylogenetic trees, which are important operations in computational phylogenetics. For more details, see [16, Chapter 8] and the references therein. A real statistical model for phylogenetic inference based on these techniques was introduced in [34]. We note that computing means in the BHV tree space was independently studied by Miller, Owen and Provan [162] and we recommend their paper also for many other interesting results and observations.

Based on these first steps, statistical theory in the BHV tree space began to develop. The computation of variance was featured in the papers by Brown and Owen [58] and by Miller, Owen and Provan [162]. Barden,
Le and Owen studied the central limit theorem in [28]. Further refined analyses of means are due to Skwerer, Provan, and Marron [204] and Bar- den, Le and Owen [29]. Nye et al. [174] used the algorithm for computing means as a building block for developing more sophisticated statistical methods, namely the principal component analysis.

For further information on recent developments in tree spaces, we refer the interested reader into [27], [87], [150], [173], [220], [221]. We also recommend Miller’s interesting overview [161].

6.4. Stochastic proximal point algorithm

While we studied finite sums of convex lsc functions in the previous sections, now we turn to integrals. Such integral functionals were first considered in convex analysis by Rockafellar [199], [200], [201] and have become a classic object of study with many interesting properties and important applications, for instance, in financial mathematics [140]. Throughout this section, the underlying space will again be a locally compact Hadamard space \((\mathcal{H}, d)\).

Let \((S, \mu)\) be a probability space and assume that a function \(f: \mathcal{H} \times S \rightarrow (-\infty, \infty]\) satisfies

(i) \(f(\cdot, \xi)\) is a convex lsc function for each \(\xi \in S\),
(ii) \(f(x, \cdot)\) is a measurable function for each \(x \in \mathcal{H}\).

Then define

\[
F(x) := \int_S f(x, \xi) \, d\mu(\xi), \quad x \in \mathcal{H}.
\]

We will assume that \(F(x) > -\infty\) for every \(x \in \mathcal{H}\) and that \(F\) is lsc (which can be assured, for instance, by Fatou’s lemma).

Following [19], we now define the stochastic version of the PPA. To this end, we denote by \((\xi_k)\) the sequence of random variables

\[
\xi_k(\omega) := \omega_k, \quad \omega = (\omega_1, \omega_2, \ldots) \in \Omega,
\]

where \(\Omega := S^\mathbb{N}\). Let \((\lambda_k)\) be a sequence of positive reals. Given \(x_0 \in \mathcal{H}\), define random variables

\[
x_k := \text{prox}_{\lambda_k f(\cdot, \xi_k)}(x_{k-1}),
\]

for each \(k \in \mathbb{N}\). The convergence of the proximal sequence \((x_k)\) is guaranteed by Theorem 6.5 below. For the proof, see [19, Theorem 3.1].

**Theorem 6.5 (Stochastic PPA).** Assume that

(i) a function \(F\) is of the form (34) and has a minimizer,
(ii) there exists \( p \in \mathcal{H} \) and an \( L^2 \)-function \( L: S \to (0, \infty) \) such that
\[
f(x, \xi) - f(y, \xi) \leq L(\xi) \left[ 1 + d(x, p) \right] d(x, y),
\]
for every \( x, y \in \mathcal{H} \) and almost every \( \xi \in S \),

(iii) \( (\lambda_k) \in \ell_2 \setminus \ell_1 \) is a sequence of positive reals.

Then there exists a random variable \( x: \Omega \to \text{Min } F \) such that for almost every \( \omega \in \Omega \) the sequence \( (x_k(\omega)) \) given by (35) converges to \( x(\omega) \).

An application of Theorem 6.5 will be given in Subsect. 6.5 below. Like in the case of the splitting PPA, we raise the following question.

**Problem 6.6.** Can one prove weak convergence of the stochastic PPA in (separable) Hadamard spaces without the local compactness assumption? This is unknown even for (separable) Hilbert spaces.

To finish this section, we look at the (intuitively obvious) fact that finite sums of functions approximate integral functionals. The rigorous argument is however more involved. Artstein and Wets proved (a more general form of) the following theorem in [10, Theorem 2.3].

**Theorem 6.7 (Consistency).** Let again \( F \) be the functional from (34) and assume that the function \( f: \mathcal{H} \times S \to (-\infty, \infty] \) are measurable and for each \( x_0 \in \mathcal{H} \) there exist an open neighborhood \( U_0 \subset \mathcal{H} \) and an integrable function \( b_0: S \to \mathbb{R} \) such that
\[
f(x, \xi) \geq b_0(\xi),
\]
for every \( x \in U_0 \) and almost every \( \xi \in S \). Let \( (\xi_k) \) be an iid sequence with distribution \( \mu \). Then the sequence of functions
\[
F_k(x) := \frac{1}{k} \sum_{i=1}^{k} f(x, \xi_i), \quad x \in \mathcal{H},
\]
almost surely \( \Gamma \)-converges to \( F \).

6.5. **Application: Medians and means of general probabilities**

Consider now a probability measure \( \mu \in \mathcal{P}^2(\mathcal{H}) \). Then the assumptions of Theorem 6.5 are satisfied for the functionals
\[
\int d(\cdot, z) \, d\mu(z), \quad \text{and} \quad \int d(\cdot, z)^2 \, d\mu(z),
\]
and we can therefore employ the stochastic PPA to find their minimizers, which are called a *median* and *mean*, respectively. One can observe that
the corresponding proximal mappings are easy to evaluate. We note that
it was observed already by Miller, Owen and Provan [162] that means
can be alternatively approximated via Sturm’s law of large numbers [212,
Theorem 4.7].

With the stochastic PPA at hand, one can improve upon the statistical
model from [34]. Indeed, we are to approximate the mean of a given probabil-
ability distribution $\mu_D$ of phylogenetic trees on the BHV tree space $\mathcal{T}_n$.
The approach taken in [34] is first to generate a finite set of samples from
$\mu_D$ by a Markov Chain Monte Carlo (MCMC) simulation and then compute
their mean. However, now we can simply generate samples by MCMC
on-the-fly and average them on-the-fly by the stochastic PPA. This way
we eliminate the error coming from the approximation of the integral by
a finite sum.

6.6. Minimization of submodular functions on modular lattices

Now we start a completely new topic. It has turned out recently that the
above Hadamard space algorithms can be used for minimizing submodular
functions on modular lattices. This was discovered by Hamada and Hirai
[99]. We first recall some algebra.

A partially ordered set $(\mathcal{L}, \leq)$ is called a lattice if, given $x, y \in \mathcal{L}$,
there exist their least upper bound (denoted by $x \vee y$) and their greatest
lower bound (denoted by $x \wedge y$). Given $n \in \mathbb{N}$, we say that $x_0, \ldots, x_n \in \mathcal{L}$
form a chain if $x_0 < \cdots < x_n$. The number $n$ is called the length of
this chain. Here we consider only lattices, in which each chain has finite
length. Let $0_{\mathcal{L}}$ and $1_{\mathcal{L}}$ denote the minimum and maximum elements in $\mathcal{L}$,
respectively. Given $x \in \mathcal{L}$, the maximum length of a chain $0_{\mathcal{L}} < \cdots < x$
is called the rank of $x$ and is denoted by $\text{rank}(x)$. The rank of $\mathcal{L}$ is
deﬁned as $\text{rank}(\mathcal{L}) := \text{rank}(1_{\mathcal{L}})$.

A lattice $(\mathcal{L}, \leq)$ is modular if $x \vee (y \wedge z) = (x \vee y) \wedge z$, whenever
$x, y, z \in \mathcal{L}$ and $x \leq z$. It is known that a lattice is modular if and only if
$\text{rank}(x \vee y) + \text{rank}(x \wedge y) = \text{rank}(x) + \text{rank}(y)$,
for every $x, y \in \mathcal{L}$.

The class of modular lattices comprises distributive lattices. Recall
that a lattice $\mathcal{L}$ is distributive if $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$,
for every $x, y, z \in \mathcal{L}$.

\footnote{The law of large numbers has been recently extended into CAT(1) spaces by Yokota [225].}
A lattice $\mathcal{L}$ is complemented if, given $x \in \mathcal{L}$, there exists $y \in \mathcal{L}$ such that $x \lor y = 1_\mathcal{L}$ and $x \land y = 0_\mathcal{L}$.

A lattice is distributive and complemented if and only if it is isomorphic to a Boolean lattice.

**Example 6.8.** Let $X$ be a finite-dimensional vector space. The set of its subspaces equipped with the inclusion order is a prime example of a modular lattice. Here, given subspaces $Y, Z \subset X$, we have $Y \land Z = Y \cap Z$ and $Y \lor Z = Y + Z$. The rank of a subspace is its dimension and since

$$\dim(Y \cap Z) + \dim(Y + Z) = \dim Y + \dim Z,$$

we see that this lattice is modular.

Other examples of modular lattices include the lattice of submodules of a module over a ring (e.g. the lattice of subgroups of an abelian group), and the lattice of normal subgroups of a given group.

A function $f : \mathcal{L} \to \mathbb{R}$ is called submodular if

$$f(x \land y) + f(x \lor y) \leq f(x) + f(y),$$

for every $x, y \in \mathcal{L}$.

A key role in the further developments will be played by the orthoscheme complex introduced by Brady and McCammond [49]. We will now follow the presentation of Hamada and Hirai [99]. Given $n \in \mathbb{N}$, the $n$-dimensional orthoscheme is the simplex in $\mathbb{R}^n$ with vertices

$$0, e_1, e_1 + e_2, \ldots, e_1 + \cdots + e_n,$$

where $e_1, \ldots, e_n$ denote the canonical basis in $\mathbb{R}^n$. Let now $\mathcal{L}$ be a modular lattice of rank $n$ and $F(\mathcal{L})$ be the free $\mathbb{R}$-module over $\mathcal{L}$, that is, the set of formal linear combinations $x = \sum_{y \in \mathcal{L}} t_y y$, where $t_y \in \mathbb{R}$ for each $y \in \mathcal{L}$, and the set

$$\text{supp}(x) := \{ y \in \mathcal{L} : t_y \neq 0 \}$$

is finite. We call the above set the support of $x$. Let $K(\mathcal{L})$ be a subset of $F(\mathcal{L})$ consisting of elements $x = \sum_{y \in \mathcal{L}} t_y y$, with $t_y > 0$ and $\sum_{y \in \mathcal{L}} t_y = 1$ and supp($x$) being a chain in $\mathcal{L}$. That is, $K(\mathcal{L})$ is a geometric realization of the order complex of $\mathcal{L}$. A simplex is a subset of $K(\mathcal{L})$ consisting of all convex combinations of a given chain. Given a simplex $\sigma \subset K(\mathcal{L})$ corresponding to a maximal chain $y_0 < \cdots < y_n$, we define a mapping $\phi_\sigma$ from $\sigma$ to the $n$-dimensional orthoscheme by

$$\phi_\sigma(x) := \sum_{k=1}^{n} t_k (e_1 + \cdots + e_k), \quad x = \sum_{k=0}^{n} t_k y_k.$$
We then equip each simplex $\sigma \subset K(\mathcal{L})$ with the metric $d_\sigma$ defined by
$$d_\sigma(x, y) := \|\phi_\sigma(x) - \phi_\sigma(y)\|_2,$$
for each $x, y \in \sigma$. Finally, we equip $K(\mathcal{L})$ with the induced length metric. This results in a complete geodesic space by virtue of Bridson’s theorem \[56, \text{Theorem 7.19}\]. The following result due to Chalopin, Chepoi, Hirai and Osajda \[61\] will be instrumental for us.

**Theorem 6.9.** Let $\mathcal{L}$ be a modular lattice of finite rank. Then $K(\mathcal{L})$ is a CAT(0) space.

The above theorem settles a conjecture of Brady and McCammond \[49, \text{Conjecture 6.10}\]. We remark that an important special case of this theorem was obtained in \[98\] by Haettel, Kielak and Schwer, assuming additionally that the lattice is complemented.

We also stress that the lattice $\mathcal{L}$ in the above construction can be infinite. Only its rank is assumed to be finite.

As a matter of fact, one can associate an orthoscheme complex to a more general structure than a lattice, namely to a graded poset of finite rank. This level of generality was used in \[49, 61\]. The following result due to Hirai \[103, \text{Theorem 1.2}\] settles \[61, \text{Conjecture 7.3}\].

**Theorem 6.10.** Let $\mathcal{L}$ be a modular semilattice of finite rank. Then the orthoscheme complex $K(\mathcal{L})$ is a CAT(0) space.

We will proceed by showing that a submodular function on a modular lattice can be extended to a convex function on the orthoscheme complex. This was worked out by Hamada and Hirai in \[99\]. Let $\mathcal{L}$ be a modular lattice. The Lovász extension of a function $f : \mathcal{L} \to \mathbb{R}$ is a function $\overline{f} : K(\mathcal{L}) \to \mathbb{R}$ defined by
$$\overline{f}(x) := \sum_k t_k f(x_k),$$
where $x = \sum_k t_k x_k \in K(\mathcal{L})$. The following result is due to Hirai \[102, \text{Theorem 3.9}\].

**Theorem 6.11.** Let $f : \mathcal{L} \to \mathbb{R}$ be a function on a modular lattice $\mathcal{L}$. Then $f$ is submodular if and only if its Lovász extension $\overline{f}$ is convex.

For further information on submodular functions, we refer the interested reader to Fujishige’s book \[86\].

The above theory was applied to the maximum vanishing subspace problem, shortly MVSP, by Hamada and Hirai \[99\], \[100\]. The MVSP is an algebraic generalization of the stable-set problem for bipartite graphs in combinatorial optimization. By considering a Hadamard space relaxation
of the MVSP, where the objective function is the Lovász extension of a dimension function defined on the modular lattice of subspaces of a finite dimensional vector space, the authors designed a polynomial time algorithm for the MVSP. Their algorithm relies on the splitting PPA in Hadamard spaces. We refer the interested reader to the original papers [99], [100] for further details on this very interesting application.

Since this approach pioneered by Hamada and Hirai in [99] can be readily applied to other modular lattices and submodular functions, we believe that further applications will appear soon. In this connection we also recommend Hirai’s paper [102].

6.7. Splitting proximal point algorithm for relaxed problems

We will now look at yet another variant of the PPA. Let \((\mathcal{H}, d)\) be a Hadamard space\(^{14}\) and let \(f_1, f_2 : \mathcal{H} \to (-\infty, \infty] \) be convex lsc functions. Instead of minimizing the function \(f_1 + f_2\) on \(\mathcal{H}\), one may consider a relaxed problem: minimize the function

\[
(x, y) \mapsto f_1(x) + f_2(y) + \frac{1}{2\lambda} d(x, y)^2
\]

on the space \(\mathcal{H} \times \mathcal{H}\), where \(\lambda > 0\) is a fixed parameter. This problem was studied by Banert [26], who obtained the following convergence theorem.

**Theorem 6.12.** Given \(x_0 \in \mathcal{H}\), define \(y_n \coloneqq \text{prox}_{\lambda f_2}(x_n)\) and \(x_{n+1} \coloneqq \text{prox}_{\lambda f_1}(y_n)\), for every \(n \in \mathbb{N}_0\). Then the sequence \((x_n, y_n)\) weakly converges to a minimizer of the function in (36) provided this function attains its minimum.

Banert’s result was later obtained as a corollary of a more abstract fixed point theorem by Ariza-Ruiz, López-Acedo and Nicolae [9, Corollary 4.5].

6.8. Asymptotic regularity of the composition of projections

Our last topic is not directly related to discrete-time gradient flows, but it is an important part of convex optimization and we find the result (as well as its proof) so interesting that we decided to mention it in the present survey.

Let \(H\) be a Hilbert space and \(C_1, \ldots, C_N \subset H\) be a finite family of convex closed sets. Given \(x_0 \in H\), define

\[
x_n \coloneqq (\text{proj}_{C_N} \circ \cdots \circ \text{proj}_{C_1})(x_{n-1}),
\]

\(^{14}\) We do not assume local compactness anymore.
for $n \in \mathbb{N}$.

The sequence $(x_n)$ is used in optimization to solve \textit{convex feasibility problems}, that is, to approximate a point $x \in \bigcap_{n=1}^{N} C_n$. This approach is called the \textit{cyclic projection method}, or, if $N = 2$, we use the expression \textit{alternating projection method}. Bregman [51] proved that $(x_n)$ weakly converges to a point $x \in \bigcap_{n=1}^{N} C_n$, provided $\bigcap_{n=1}^{N} C_n \neq \emptyset$. By Hundal’s counterexample [104] we know that this cannot be improved to strong convergence even for $N = 2$.

A remarkable theorem of Bauschke [30] says that, even without the assumption $\bigcap_{n=1}^{N} C_n \neq \emptyset$, the sequence $(x_n)$ is \textit{asymptotically regular}, that is,

$$\|x_n - x_{n-1}\| \to 0,$$

as $n \to \infty$. Recall that asymptotic regularity was introduced by Browder and Petryshyn [57] and has become a basic analysis tool in optimization. For further information on this topic, we refer the interested reader to Bauschke’s paper [30] and the references therein. Here we mention only Kohlenbach’s recent paper [127] in which he obtained an explicit polynomial rate of asymptotic regularity of the above sequence by applying sophisticated methods from logic called \textit{proof mining} to Bauschke’s original proof. For further information on proof mining and its applications in Hadamard spaces, the interested reader is referred to Kohlenbach’s authoritative monograph [125] and to the papers [126], [128], [129], [130].

It is natural to ask whether one can extend the above results into Hadamard spaces.\footnote{As a matter of fact this question appeared as Problem 6.13 in the previous version of this survey paper (which is available in the arXiv).} It turns out that not. Lytchak and Petrunin [154] have constructed a (compact) Hadamard space with three subsets $C_1, C_2, C_3$, such that the corresponding sequence defined by the cyclic projections in (37) is not asymptotically regular.

We note that projection methods in Hadamard spaces had been studied also in [9], [24].

\textit{Acknowledgements.} After posting the first version of the present paper to the arXiv on July 3, 2018, I received a large number of useful comments, which helped me to improve the exposition substantially. My thanks go especially to Goulnara Arzhantseva, Victor Chepoi, Bruno Duchesne, Nicola Gigli, Koyo Hayashi, Hiroshi Hirai, Fumiaki Kohsaka, Leonid Kovalev, Alexandru Kristály, Alexander Lytchak, Uwe Mayer, Manor Mendel, Delio Mugnolo, Anton Petrunin, Gabriele Steidl and Ryokichi Tanaka.

I am also very grateful to the referees for suggesting various improvements and pointing out many misprints. Finally, I would like to express my sincere gratitude to numerous colleagues who encouraged and supported me during the review process.
References

[1] A.D. Aleksandrov, A theorem on triangles in a metric space and some of its applications, Trudy Mat. Inst. Steklov., 38 (1951), 5–23.
[2] S. Alexander, V. Kapovitch and A. Petrunin, An Invitation to Alexandrov Geometry: CAT(0) Spaces, SpringerBriefs Math., Springer-Verlag, 2019.
[3] L. Ambrosio, N. Gigli and G. Savaré, Gradient flows in metric spaces and in the space of probability measures. Second ed., Lectures Math. ETH Zürich, Birkhäuser Verlag, Basel, 2008.
[4] L. Ambrosio and B. Kirchheim, Currents in metric spaces, Acta Math., 185 (2000), 1–80.
[5] A. Andoni, A. Naor and O. Neiman, Snowflake universality of Wasserstein spaces, Ann. Sci. Éc. Norm. Supér. (4), 51 (2018), 657–700.
[6] F. Ardila, T. Baker and R. Yatchak, Moving robots efficiently using the combinatorics of CAT(0) cubical complexes, SIAM J. Discrete Math., 28 (2014), 986–1007.
[7] F. Ardila, H. Bastidas, C. Ceballos and J. Guo, The configuration space of a robotic arm in a tunnel, SIAM J. Discrete Math., 31 (2017), 2675–2702.
[8] F. Ardila, M. Owen and S. Sullivant, Geodesics in CAT(0) cubical complexes, Adv. in Appl. Math., 48 (2012), 142–163.
[9] D. Ariza-Ruiz, G. López-Acedo and A. Nicolae, The asymptotic behavior of the composition of firmly nonexpansive mappings, J. Optim. Theory Appl., 167 (2015), 409–429.
[10] Z. Artstein and R.J.-B. Wets, Consistency of minimizers and the SLLN for stochastic programs, J. Convex Anal., 2 (1995), 1–17.
[11] H. Attouch, Variational Convergence for Functions and Operators, Applicable Mathematics Series, Pitman, Advanced Publ. Program, Boston, MA, 1984.
[12] T. Austin, A CAT(0)-valued pointwise ergodic theorem, J. Topol. Anal., 3 (2011), 145–152.
[13] T. Austin, Erratum: A CAT(0)-valued pointwise ergodic theorem, J. Topol. Anal., 8 (2016), 373–374.
[14] M. Bačák, The proximal point algorithm in metric spaces, Israel J. Math., 194 (2013), 689–701.
[15] M. Bačák, Computing medians and means in Hadamard spaces, SIAM J. Optim., 24 (2014), 1542–1566.
[16] M. Bačák, Convex Analysis and Optimization in Hadamard Spaces, De Gruyter Ser. Nonlinear Anal. Appl., 22, De Gruyter, Berlin, 2014.
[17] M. Bačák, A new proof of the Lie–Trotter–Kato formula in Hadamard spaces, Commun. Contemp. Math., 16 (2014), no. 6, 1350044.
[18] M. Bačák, Convergence of nonlinear semigroups under nonpositive curvature, Trans. Amer. Math. Soc., 367 (2015), 3929–3953.
[19] M. Bačák, A variational approach to stochastic minimization of convex functionals, Pure Appl. Funct. Anal., 3 (2018), 287–295.
[20] M. Bačák, R. Bergmann, G. Steidl and A. Weinmann, A second order non-smooth variational model for restoring manifold-valued images, SIAM J. Sci. Comput., 38 (2016), A567–A597.
[21] M. Bačák and L.V. Kovalev, Lipschitz retractions in Hadamard spaces via gradient flow semigroups, Canad. Math. Bull., 59 (2016), 673–681.
[22] M. Bačák, M. Montag and G. Steidl, Convergence of functions and their Moreau envelopes on Hadamard spaces, J. Approx. Theory, 224 (2017), 1–12.
[23] M. Bačák and S. Reich, The asymptotic behavior of a class of nonlinear semi-
groups in Hadamard spaces, J. Fixed Point Theory Appl., 16 (2014), 189–202.
[24] M. Bačák, I. Searston and B. Sims, Alternating projections in CAT(0) spaces,
J. Math. Anal. Appl., 385 (2012), 599–607.
[25] J.-B. Baillon, Un exemple concernant le comportement asymptotique de la solution
du problème \( du/dt + \partial \varphi(u) \ni 0 \), J. Functional Analysis, 28 (1978), 369–376.
[26] S. Banert, Backward-backward splitting in Hadamard spaces, J. Math. Anal.
Appl., 414 (2014), 656–665.
[27] D. Barden and H. Le, The logarithm map, its limits and Fréchet means in
orthant spaces, Proc. Lond. Math. Soc. (3), 117 (2018), 751–789.
[28] D. Barden, H. Le and M. Owen, Central limit theorems for Fréchet means in
the space of phylogenetic trees, Electron. J. Probab., 18 (2013), no. 25.
[29] D. Barden, H. Le and M. Owen, Limiting behaviour of Fréchet means in the
space of phylogenetic trees, Ann. Inst. Statist. Math., 70 (2018), 99–129.
[30] H.H. Bauschke, The composition of projections onto closed convex sets in
Hilbert space is asymptotically regular, Proc. Amer. Math. Soc., 131 (2003),
141–146.
[31] H.H. Bauschke, J.V. Burke, F.R. Deutsch, H.S. Hundal and J.D. Vanderwerff,
A new proximal point iteration that converges weakly but not in norm, Proc.
Amer. Math. Soc., 133 (2005), 1829–1835.
[32] H.H. Bauschke and P.L. Combettes, Convex analysis and monotone operator
theory in Hilbert spaces. Second ed., With a foreword by Hédy Attouch, CMS
Books Math./Ouvrages Math. SMC, Springer-Verlag, 2017.
[33] H.H. Bauschke, E. Matoušková and S. Reich, Projection and proximal point
methods: convergence results and counterexamples, Nonlinear Anal., 56 (2004),
715–738.
[34] P. Benner, M. Bačák and P.-Y. Bourguignon, Point estimates in phylogenetic
reconstructions, Bioinformatics, 30 (2014), i534–i540.
[35] A. Bërdëllima, A note on Mosco convergence in CAT(0) spaces, Canad. Math.
Bull., 65 (2022), 994–1003.
[36] I.D. Berg and I.G. Nikolaev, Quasilinearization and curvature of Aleksandrov
spaces, Geom. Dedicata, 133 (2008), 195–218.
[37] R. Bergmann, R.H. Chan, R. Hielscher, J. Persch and G. Steidl, Restoration of
manifold-valued images by half-quadratic minimization, Inverse Probl. Imaging,
10 (2016), 281–304.
[38] R. Bergmann, F. Laus, G. Steidl and A. Weinmann, Second order differences
of cyclic data and applications in variational denoising, SIAM J. Imaging Sci.,
7 (2014), 2916–2953.
[39] R. Bergmann, J. Persch and G. Steidl, A parallel Douglas–Rachford algorithm
for minimizing ROF-like functionals on images with values in symmetric Hadam-
ard manifolds, SIAM J. Imaging Sci., 9 (2016), 901–937.
[40] R. Bergmann and A. Weinmann, Inpainting of cyclic data using first and second
order differences, preprint, arXiv:1410.1998v1; Accepted conference paper at
EMMCVPR’15.
[41] R. Bergmann and A. Weinmann, A second-order TV-type approach for inpaint-
ing and denoising higher dimensional combined cyclic and vector space data, J.
Math. Imaging Vision, 55 (2016), 401–427.
[42] R.J. Berman, T. Darvas and C.H. Lu, Convexity of the extended K-energy
and the large time behavior of the weak Calabi flow, Geom. Topol., 21 (2017),
2945–2988.
D.P. Bertsekas, Incremental proximal methods for large scale convex optimization, Math. Program., 129 (2011), 163–195.

L.J. Billera, S.P. Holmes and K. Vogtmann, Geometry of the space of phylogenetic trees, Adv. in Appl. Math., 27 (2001), 733–767.

K. Borsuk and S. Ulam, On symmetric products of topological spaces, Bull. Amer. Math. Soc., 37 (1931), 875–882.

J. Bourgain, On Lipschitz embedding of finite metric spaces in Hilbert space, Israel J. Math., 52 (1985), 46–52.

J. Bourgain, The metrical interpretation of superreflexivity in Banach spaces, Israel J. Math., 56 (1986), 222–230.

J. Bourgain, V. Milman and H. Wolfson, On type of metric spaces, Trans. Amer. Math. Soc., 294 (1986), 295–317.

T. Brady and J. McCammond, Braids, posets and orthoschemes, Algebr. Geom. Topol., 10 (2010), 2277–2314.

K. Bredies, M. Holler, M. Storath and A. Weinmann, Total generalized variation for manifold-valued data, SIAM J. Imaging Sci., 11 (2018), 1785–1848.

L.M. Brégman, Finding the common point of convex sets by the method of successive projection, Dokl. Akad. Nauk SSSR, 162 (1965), 487–490.

H. Brézis, Opérateurs Maximaux Monotones Et Semi-Groupes De Contractions Dans Les Espaces De Hilbert, North-Holland Math. Stud., 5, North-Holland Publ. Co., Amsterdam, 1973.

H. Brézis and P.-L. Lions, Produits infinis de résolvantes, Israel J. Math., 29 (1978), 329–345.

H. Brézis and A. Pazy, Semigroups of nonlinear contractions on convex sets, J. Functional Analysis, 6 (1970), 237–281.

H. Brézis and A. Pazy, Convergence and approximation of semigroups of nonlinear operators in Banach spaces, J. Functional Analysis, 9 (1972), 63–74.

M.R. Bridson and A. Haefliger, Metric Spaces of Non-Positive Curvature, Grundlehren Math. Wiss., 319, Springer-Verlag, 1999.

F.E. Browder and W.V. Petryshyn, The solution by iteration of nonlinear functional equations in Banach spaces, Bull. Amer. Math. Soc., 72 (1966), 571–575.

D.G. Brown and M. Owen, Mean and variance of phylogenetic trees, Systematic Biology, 69 (2019), 139–154.

R.E. Bruck, Jr., Asymptotic convergence of nonlinear contraction semigroups in Hilbert space, J. Functional Analysis, 18 (1975), 15–26.

H. Busemann, Spaces with non-positive curvature, Acta Math., 80 (1948), 259–310.

J. Chalopin, V. Chepoi, H. Hirai and D. Osajda, Weakly Modular Graphs and Nonpositive Curvature, Mem. Amer. Math. Soc., 268, no. 1309, Amer. Math. Soc., Providence, RI, 2020.

J. Chalopin, V. Chepoi and G. Naves, Isometric embedding of Busemann surfaces into $L_1$, Discrete Comput. Geom., 53 (2015), 16–37.

V. Chepoi, B. Estellon and G. Naves, Packing and covering with balls on Busemann surfaces, Discrete Comput. Geom., 57 (2017), 985–1011.

V. Chepoi and D. Maftuleac, Shortest path problem in rectangular complexes of global nonpositive curvature, Comput. Geom., 46 (2013), 51–64.

P.R. Chernoff, Note on product formulas for operator semigroups, J. Functional Analysis, 2 (1968), 238–242.

P. Clément and J. Maas, A Trotter product formula for gradient flows in metric spaces, J. Evol. Equ., 11 (2011), 405–427.
[67] P.L. Combettes and L.E. Glaudin, Proximal activation of smooth functions in splitting algorithms for convex image recovery, SIAM J. Imaging Sci., 12 (2019), 1905–1935.

[68] M.G. Crandall and T.M. Liggett, Generation of semi-groups of nonlinear transformations on general Banach spaces, Amer. J. Math., 93 (1971), 265–298.

[69] G. Dal Maso, An Introduction to Γ-Convergence, Progr. Nonlinear Differential Equations Appl., 8, Birkhäuser Boston Inc., Boston, MA, 1993.

[70] A. Daniilidis, G. David, E. Durand-Cartagena and A. Lemenant, Rectifiability of self-contracted curves in the Euclidean space and applications, J. Geom. Anal., 25 (2015), 1211–1239.

[71] A. Daniilidis, O. Ley and S. Sabourau, Asymptotic behaviour of self-contracted planar curves and gradient orbits of convex functions, J. Math. Pures Appl. (9), 94 (2010), 183–199.

[72] S.K. Donaldson, The construction of an asymptotic center with a fixed-point property, Bull. Amer. Math. Soc., 78 (1972), 206–208.

[73] P. Enflo, On the nonexistence of uniform homeomorphisms between $L_p$-spaces, Ark. Mat., 8 (1969), 103–105.

[74] B. Duchesne, Groups acting on spaces of non-positive curvature, In: Handbook of Group Actions. Vol. III, Adv. Lect. Math. (ALM), 40, Int. Press, Somerville, MA, 2018, pp. 101–141.

[75] M. Fabian, P. Habala, P. Hájek, V. Montesinos and V. Zizler, Banach Space Theory. The Basis for Linear and Nonlinear Analysis, CMS Books Math./Ouvrages Math. SMC, Springer-Verlag, 2011.

[76] T. Foertsch, A. Lytchak and V. Schroeder, Nonpositive curvature and the Ptolemy inequality, Int. Math. Res. Not. IMRN, 2007 (2007), rnm100.

[77] B. Fuglede, Harmonic maps from Riemannian polyhedra to geodesic spaces with curvature bounded from above, Calc. Var. Partial Differential Equations, 31 (2008), 99–136.

[78] A. Gavryushkin and A.J. Drummond, The space of ultrametric phylogenetic trees, J. Theoret. Biol., 403 (2016), 197–208.
[88] T. Gelander, A. Karlsson and G.A. Margulis, Superrigidity, generalized harmonic maps and uniformly convex spaces, Geom. Funct. Anal., 17 (2008), 1524–1550.
[89] K. Goebel and W.A. Kirk, Topics in Metric Fixed Point Theory, Cambridge Stud. Adv. Math., 28, Cambridge Univ. Press, Cambridge, 1990.
[90] K. Goebel and S. Reich, Uniform convexity, hyperbolic geometry, and nonexpansive mappings, Monogr. Textbooks Pure Appl. Math., 83, Marcel Dekker Inc., New York, 1984.
[91] M. Gromov, Asymptotic invariants of infinite groups, In: Geometric Group Theory. Vol. 2, London Math. Soc. Lecture Note Ser., 182, Cambridge Univ. Press, Cambridge, 1993, pp. 1–295.
[92] M. Gromov, Metric Structures for Riemannian and Non-Riemannian Spaces, Progr. Math., 152, Birkhäuser Boston Inc., Boston, MA, 1999. Based on the 1981 French original, with appendices by M. Katz, P. Pansu and S. Semmes, translated from the French by Sean Michael Bates.
[93] M. Gromov, CAT(κ)-spaces: construction and concentration, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI), 280 (2001), 101–140.
[94] M. Gromov, Random walk in random groups, Geom. Funct. Anal., 13 (2003), 73–146.
[95] M. Gromov and R. Schoen, Harmonic maps into singular spaces and p-adic superrigidity for lattices in groups of rank one, Inst. Hautes Études Sci. Publ. Math., 76 (1992), 165–246.
[96] O. Güler, On the convergence of the proximal point algorithm for convex minimization, SIAM J. Control Optim., 29 (1991), 403–419.
[97] M.J. Gursky and J. Streets, A formal Riemannian structure on conformal classes and the inverse Gauss curvature flow, Geom. Flows, 4 (2019), 30–50.
[98] T. Haettel, D. Kielak and P. Schwer, The 6-strand braid group is CAT(0), Geom. Dedicata, 182 (2016), 263–286.
[99] M. Hamada and H. Hirai, Maximum vanishing subspace problem, CAT(0)-space relaxation, and block-triangularization of partitioned matrix, preprint, arXiv:1705.02060.
[100] M. Hamada and H. Hirai, Computing the nc-rank via discrete convex optimization on CAT(0) spaces, SIAM J. Appl. Algebra Geom., 5 (2021), 455–478.
[101] K. Hayashi, A polynomial time algorithm to compute geodesics in CAT(0) cubical complexes, In: 45th International Colloquium on Automata, Languages, and Programming, LIPIcs. Leibniz Int. Proc. Inform., 107, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2018, no. 78.
[102] H. Hirai, L-convexity on graph structures, J. Oper. Res. Soc. Japan, 61 (2018), 71–109.
[103] H. Hirai, A nonpositive curvature property of modular semilattices, Geom. Dedicata, 214 (2021), 427–463.
[104] H.S. Hundal, An alternating projection that does not converge in norm, Nonlinear Anal., 57 (2004), 35–61.
[105] P. Ivanisvili, R. van Handel and A. Volberg, Rademacher type and Enflo type coincide, Ann. of Math. (2), 192 (2020), 665–678.
[106] S. Ivanov and A. Lytchak, Rigidity of Busemann convex Finsler metrics, Comment. Math. Helv., 94 (2019), 855–868.
[107] H. Izeki, T. Kondo and S. Nayatani, N-step energy of maps and the fixed-point property of random groups, Groups Geom. Dyn., 6 (2012), 701–736.
[108] H. Izeki and S. Nayatani, Combinatorial harmonic maps and discrete-group actions on Hadamard spaces, Geom. Dedicata, 114 (2005), 147–188.
[109] J. Jost, Equilibrium maps between metric spaces, Calc. Var. Partial Differential Equations, 2 (1994), 173–204.
[110] J. Jost, Convex functionals and generalized harmonic maps into spaces of non-positive curvature, Comment. Math. Helv., 70 (1995), 659–673.
[111] J. Jost, Generalized Dirichlet forms and harmonic maps, Calc. Var. Partial Differential Equations, 5 (1997), 1–19.
[112] J. Jost, Nonpositive Curvature: Geometric and Analytic Aspects, Lectures Math. ETH Zürich, Birkhäuser Verlag, Basel, 1997.
[113] J. Jost, Nonlinear Dirichlet forms, In: New Directions in Dirichlet Forms, AMS/IP Stud. Adv. Math., 8, Amer. Math. Soc., Providence, RI, 1998, pp. 1–47.
[114] S. Kakutani, Some characterizations of Euclidean space, Jpn. J. Math., 16 (1939), 93–97.
[115] T. Kato, Nonlinear semigroups and evolution equations, J. Math. Soc. Japan, 19 (1967), 508–520.
[116] T. Kato and K. Masuda, Trotter’s product formula for nonlinear semigroups generated by the subdifferentials of convex functionals, J. Math. Soc. Japan, 30 (1978), 169–178.
[117] M. Kell, Symmetric orthogonality and non-expansive projections in metric spaces, Manuscripta Math., 161 (2020), 141–159.
[118] Y. Kimura and F. Kohsaka, Spherical nonspreadingness of resolvents of convex functions in geodesic spaces, J. Fixed Point Theory Appl., 18 (2016), 93–115.
[119] Y. Kimura and F. Kohsaka, Two modified proximal point algorithms for convex functions in Hadamard spaces, Linear Nonlinear Anal., 2 (2016), 69–86.
[120] Y. Kimura and F. Kohsaka, The proximal point algorithm in geodesic spaces with curvature bounded above, Linear Nonlinear Anal., 3 (2017), 133–148.
[121] Y. Kimura and F. Kohsaka, Two modified proximal point algorithms in geodesic spaces with curvature bounded above, Rend. Circ. Mat. Palermo (2), 68 (2019), 83–104.
[122] W.A. Kirk and B. Panyanak, A concept of convergence in geodesic spaces, Nonlinear Anal., 68 (2008), 3689–3696.
[123] B.R. Kloeckner, Yet another short proof of Bourgain’s distortion estimate for embedding of trees into uniformly convex Banach spaces, Israel J. Math., 200 (2014), 419–422.
[124] Y. Kobayashi, Difference approximation of Cauchy problems for quasi-dissipative operators and generation of nonlinear semigroups, J. Math. Soc. Japan, 27 (1975), 640–665.
[125] U. Kohlenbach, Applied Proof Theory: Proof Interpretations and their Use in Mathematics, Springer Monogr. Math., Springer-Verlag, 2008.
[126] U. Kohlenbach, Recent progress in proof mining in nonlinear analysis, IFCoLog J. Log. Appl., 4 (2017), 3357–3406.
[127] U. Kohlenbach, A polynomial rate of asymptotic regularity for compositions of projections in Hilbert space, Found. Comput. Math., 19 (2019), 83–99.
[128] U. Kohlenbach and L. Leuștean, Effective metastability of Halpern iterates in CAT(0) spaces, Adv. Math., 231 (2012), 2526–2556.
[129] U. Kohlenbach and L. Leuștean, Addendum to “Effective metastability of Halpern iterates in CAT(0) spaces”, Adv. Math., 250 (2014), 650–651.
[130] U. Kohlenbach, L. Leuștean and A. Nicolae, Quantitative results on Fejér monotone sequences, Commun. Contemp. Math., 20 (2018), 1750015.
[131] F. Kohsaka, Existence and approximation of fixed points of vicinal mappings in geodesic spaces, Pure Appl. Funct. Anal., 3 (2018), 91–106.
[132] T. Kondo, CAT(0) spaces and expanders, Math. Z., 271 (2012), 343–355.
[133] E. Kopecká and S. Reich, Nonexpansive retracts in Banach spaces, In: Fixed Point Theory and Its Applications, Banach Center Publ., 77, Polish Acad. Sci. Inst. Math., Warsaw, 2007, pp. 161–174.
[134] N.J. Korevaar and R.M. Schoen, Sobolev spaces and harmonic maps for metric space targets, Comm. Anal. Geom., 1 (1993), 561–659.
[135] N.J. Korevaar and R.M. Schoen, Global existence theorems for harmonic maps to non-locally compact spaces, Comm. Anal. Geom., 5 (1997), 333–387.
[136] L.V. Kovalev, Lipschitz retraction of finite subsets of Hilbert spaces, Bull. Aust. Math. Soc., 93 (2016), 146–151.
[137] A. Kristály and L. Kozma, Metric characterization of Berwald spaces of non-positive flag curvature, J. Geom. Phys., 56 (2006), 1257–1270.
[138] A. Kristály and A. Róth, Testing metric relations on Finsler manifolds via a geodesic detecting algorithm, In: 2014 IEEE 9th IEEE International Symposium on Applied Computational Intelligence and Informatics, 2014, pp. 331–336.
[139] A. Kristály, C. Varga and L. Kozma, The dispersing of geodesics in Berwald spaces of non-positive flag curvature, Houston J. Math., 30 (2004), 413–420.
[140] H.J. Kushner and G.G. Yin, Stochastic Approximation and Recursive Algorithms and Applications. Second ed., Appl. Math. (N. Y.), 35, Stoc. Mod. Appl. Probab., Springer-Verlag, 2003.
[141] K. Kuwae and T. Shioya, Variational convergence over metric spaces, Trans. Amer. Math. Soc., 360 (2008), 35–75.
[142] J.-F. Lafont and S. Prassidis, Roundness properties of groups, Geom. Dedicata, 117 (2006), 137–160.
[143] U. Lang, B. Pavlović and V. Schroeder, Extensions of Lipschitz maps into Hadamard spaces, Geom. Funct. Anal., 10 (2000), 1527–1553.
[144] U. Lang and V. Schroeder, Kirszbraun’s theorem and metric spaces of bounded curvature, Geom. Funct. Anal., 7 (1997), 535–560.
[145] N. Lebedeva and A. Petrunin, 5-point CAT(0) spaces after Tetsu Toyoda, Anal. Geom. Metr. Spaces, 9 (2021), 160–166.
[146] L. Leuştean, A. Nicolae and A. Sipoş, An abstract proximal point algorithm, J. Global Optim., 72 (2018), 553–577.
[147] L. Leuştean and A. Sipoş, An application of proof mining to the proximal point algorithm in CAT(0) spaces, In: Mathematics Almost Everywhere, World Sci. Publ., Hackensack, NJ, 2018, pp. 153–167.
[148] C. Li, G. López and V. Martín-Márquez, Monotone vector fields and the proximal point algorithm on Hadamard manifolds, J. Lond. Math. Soc. (2), 79 (2009), 663–683.
[149] C. Li, G. López, V. Martín-Márquez and J.-H. Wang, Resolvents of set-valued monotone vector fields in Hadamard manifolds, Set-Valued Var. Anal., 19 (2011), 361–383.
[150] B. Lin, B. Sturmfels, X. Tang and R. Yoshida, Convexity in tree spaces, SIAM J. Discrete Math., 31 (2017), 2015–2038.
[151] N. Linial and M. Saks, The Euclidean distortion of complete binary trees, Discrete Comput. Geom., 29 (2003), 19–21.
[152] A. Lytchak, Open map theorem for metric spaces, Algebra i Analiz, 17 (2005), 139–159.
[153] A. Lytchak and A. Petrunin, Weak topology on CAT(0) spaces, preprint, arXiv:2107.09295.
[154] A. Lytchak and A. Petrunin, Cyclic projections in Hadamard spaces, J. Optim. Theory Appl., 194 (2022), 636–642.
[155] A. Lytchak and S. Stadler, Improvements of upper curvature bounds, Trans. Amer. Math. Soc., 373 (2020), 7153–7166.
[156] B. Martinet, Régularisation d’inéquations variationnelles par approximations successives, Rev. Française Informat. Recherche Opérationnelle, 4 (1970), 154–158.
[157] J. Matoušek, On embedding trees into uniformly convex Banach spaces, Israel J. Math., 114 (1999), 221–237.
[158] U.F. Mayer, Gradient flows on nonpositively curved metric spaces and harmonic maps, Comm. Anal. Geom., 6 (1998), 199–253.
[159] M. Mendel and A. Naor, Expanders with respect to Hadamard spaces and random graphs, Duke Math. J., 164 (2015), 1471–1548.
[160] C. Mese, Uniqueness theorems for harmonic maps into metric spaces, Commun. Contemp. Math., 4 (2002), 725–750.
[161] E. Miller, Fruit flies and moduli: interactions between biology and mathematics, Notices Amer. Math. Soc., 62 (2015), 1178–1184.
[162] E. Miller, M. Owen and J.S. Provan, Polyhedral computational geometry for averaging metric phylogenetic trees, Adv. in Appl. Math., 68 (2015), 51–91.
[163] I. Miyadera and S. Ōharu, Approximation of semi-groups of nonlinear operators, Tohoku Math. J. (2), 22 (1970), 24–47.
[164] N. Monod, Superrigidity for irreducible lattices and geometric splitting, J. Amer. Math. Soc., 19 (2006), 781–814.
[165] N. Monod, Extreme points in non-positive curvature, Studia Math., 234 (2016), 265–270.
[166] J.-J. Moreau, Fonctions convexes duales et points proximaux dans un espace hilbertien, C. R. Acad. Sci. Paris, 255 (1962), 2897–2899.
[167] J.-J. Moreau, Propriétés des applications “prox”, C. R. Acad. Sci. Paris, 256 (1963), 1069–1071.
[168] J.-J. Moreau, Proximité et dualité dans un espace hilbertien, Bull. Soc. Math. France, 93 (1965), 273–299.
[169] U. Mosco, Convergence of convex sets and of solutions of variational inequalities, Advances in Math., 3 (1969), 510–585.
[170] M. Movahedi, D. Behnardi and S. Hosseini, On the density theorem for the subdifferential of convex functions on Hadamard spaces, Pacific J. Math., 276 (2015), 437–447.
[171] A. Navas, An $L^1$ ergodic theorem with values in a non-positively curved space via a canonical barycenter map, Ergodic Theory Dynam. Systems, 33 (2013), 609–623.
[172] S. Neumayer, J. Persch and G. Steidl, Morphing of manifold-valued images inspired by discrete geodesics in image spaces, SIAM J. Imaging Sci., 11 (2018), 1898–1930.
[173] T.M.W. Nye, Convergence of random walks to Brownian motion on cubical complexes, preprint, arXiv:1508.02906.
[174] T.M.W. Nye, X. Tang, G. Weyenberg and R. Yoshida, Principal component analysis and the locus of the Fréchet mean in the space of phylogenetic trees, Biometrika, 104 (2017), 901–922.
[175] S. Ohta, Harmonicity of totally geodesic maps into nonpositively curved metric spaces, Manuscripta Math., 114 (2004), 127–138.
[176] S. Ohta, Markov type of Alexandrov spaces of non-negative curvature, Mathematika, 55 (2009), 177–189.
[177] S. Ohta, Self-contracted curves in CAT(0)-spaces and their rectifiability, J. Geom. Anal., 30 (2020), 936–967.
[178] S. Ohta and M. Pálfia, Discrete-time gradient flows and law of large numbers in Alexandrov spaces, Calc. Var. Partial Differential Equations, 54 (2015), 1591–1610.

[179] S. Ohta and M. Pálfia, Gradient flows and a Trotter–Kato formula of semiconvex functions on CAT(1)-spaces, Amer. J. Math., 139 (2017), 937–965.

[180] M. Owen, Computing geodesic distances in tree space, SIAM J. Discrete Math., 25 (2011), 1506–1529.

[181] M. Owen and J.S. Provan, A fast algorithm for computing geodesic distances in tree space, IEEE/ACM Trans. Comput. Biol. Bioinform., 8 (2011), 2–13.

[182] E.A. Papa Quiroz and P.R. Oliveira, Proximal point methods for quasiconvex and convex functions with Bregman distances on Hadamard manifolds, J. Convex Anal., 16 (2009), 49–69.

[183] N. Parikh and S. Boyd, Proximal algorithms, Found. Trends Optim., 1 (2013), 123–231.

[184] X. Pennec, P. Fillard and N. Ayache, A Riemannian framework for tensor computing, Int. J. Comput. Vis., 66 (2006), 41–66.

[185] G. Perelman and A. Petrunin, Quasigeodesics and gradient curves in Alexandrov spaces, preprint, 1994.

[186] A. Petrunin, Applications of quasigeodesics and gradient curves, In: Comparison Geometry, Math. Sci. Res. Inst. Publ., 30, Cambridge Univ. Press, Cambridge, 1997, pp. 203–219.

[187] J. Peypouquet and S. Sorin, Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time, J. Convex Anal., 17 (2010), 1113–1163.

[188] R.R. Phelps, Convex sets and nearest points, Proc. Amer. Math. Soc., 8 (1957), 790–797.

[189] S. Reich, Product formulas, nonlinear semigroups, and accretive operators, J. Functional Analysis, 36 (1980), 147–168.

[190] S. Reich, A complement to Trotter’s product formula for nonlinear semigroups generated by the subdifferentials of convex functionals, Proc. Japan Acad. Ser. A Math. Sci., 58 (1982), 193–195.

[191] S. Reich, Solutions of two problems of H. Brézis, J. Math. Anal. Appl., 95 (1983), 243–250.

[192] S. Reich, The asymptotic behavior of a class of nonlinear semigroups in the Hilbert ball, J. Math. Anal. Appl., 157 (1991), 237–242.

[193] S. Reich and Z. Salinas, Weak convergence of infinite products of operators in Hadamard spaces, Rend. Circ. Mat. Palermo (2), 65 (2016), 55–71.

[194] S. Reich and Z. Salinas, Metric convergence of infinite products of operators in Hadamard spaces, J. Nonlinear Convex Anal., 18 (2017), 331–345.

[195] S. Reich and I. Shafrir, Nonexpansive iterations in hyperbolic spaces, Nonlinear Anal., 15 (1990), 537–558.

[196] S. Reich and D. Shoikhet, Semigroups and generators on convex domains with the hyperbolic metric, Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl., 8 (1997), 231–250.

[197] J.G. Rešetnjak, Non-expansive maps in a space of curvature no greater than $\kappa$, Sibirsk. Mat. Z., 9 (1968), 918–927.

[198] H. Robbins and D. Siegmund, A convergence theorem for non negative almost supermartingales and some applications, In: Optimizing Methods in Statistics, Proc. Sympos., Ohio State Univ., Columbus, Ohio, 1971, Academic Press, New York, 1971, pp. 233–257.
[199] R.T. Rockafellar, Integrals which are convex functionals, Pacific J. Math., 24 (1968), 525–539.
[200] R.T. Rockafellar, Convex integral functionals and duality, In: Contributions to Nonlinear Functional Analysis, Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, 1971, Academic Press, New York, 1971, pp. 215–236.
[201] R.T. Rockafellar, Integrals which are convex functionals. II, Pacific J. Math., 39 (1971), 439–469.
[202] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control Optim., 14 (1976), 877–898.
[203] T. Sato, An alternative proof of Berg and Nikolaev’s characterization of CAT(0)-spaces via quadrilateral inequality, Arch. Math. (Basel), 93 (2009), 487–490.
[204] S. Skwerer, S. Provan and J.S. Marron, Relative optimality conditions and algorithms for treespace Fréchet means, SIAM J. Optim., 28 (2018), 959–988.
[205] G. Steidl, Combined first and second order variational approaches for image processing, Jahresber. Dtsch. Math.-Ver., 117 (2015), 133–160.
[206] I. Stojkovic, Approximation for convex functionals on non-positively curved spaces and the Trotter–Kato product formula, Adv. Calc. Var., 5 (2012), 77–126.
[207] M. Storath and A. Weinmann, Variational regularization of inverse problems for manifold-valued data, Inf. Inference, 10 (2021), 195–230.
[208] J. Streets, The consistency and convergence of K-energy minimizing movements, Trans. Amer. Math. Soc., 368 (2016), 5075–5091.
[209] K.-T. Sturm, Monotone approximation of energy functionals for mappings into metric spaces. I, J. Reine Angew. Math., 486 (1997), 129–151.
[210] K.-T. Sturm, Nonlinear Markov operators associated with symmetric Markov kernels and energy minimizing maps between singular spaces, Calc. Var. Partial Differential Equations, 12 (2001), 317–357.
[211] K.-T. Sturm, Nonlinear Markov operators, discrete heat flow, and harmonic maps between singular spaces, Potential Anal., 16 (2002), 305–340.
[212] K.-T. Sturm, Probability measures on metric spaces of nonpositive curvature, In: Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces, Paris, 2002, Contemp. Math., 338, Amer. Math. Soc., Providence, RI, 2003, pp. 357–390.
[213] T. Toyoda, An intrinsic characterization of five points in a CAT(0) space, Anal. Geom. Metr. Spaces, 8 (2020), 114–165.
[214] H.F. Trotter, Approximation of semi-groups of operators, Pacific J. Math., 8 (1958), 887–919.
[215] H.F. Trotter, On the product of semi-groups of operators, Proc. Amer. Math. Soc., 10 (1959), 545–551.
[216] A. Wald, Begründung einer koordinatenlosen Differentialgeometrie der Flächen, Erg. Math. Kolloqu., 7 (1936), 24–46.
[217] M.-T. Wang, Generalized harmonic maps and representations of discrete groups, Comm. Anal. Geom., 8 (2000), 545–563.
[218] S. Wenger, Isoperimetric inequalities of Euclidean type in metric spaces, Geom. Funct. Anal., 15 (2005), 534–554.
[219] S. Wenger, Plateau’s problem for integral currents in locally non-compact metric spaces, Adv. Calc. Var., 7 (2014), 227–240.
[220] A. Willis, Confidence sets for phylogenetic trees, J. Amer. Statist. Assoc., 114 (2019), 235–244.
[221] A. Willis and R. Bell, Uncertainty in phylogenetic tree estimates, J. Comput. Graph. Statist., 27 (2018), 542–552.
[222] M. Xia, On sharp lower bounds for Calabi-type functionals and destabilizing properties of gradient flows, Anal. PDE, 14 (2021), 1951–1976.

[223] T. Yokota, Convex functions and barycenter on CAT(1)-spaces of small radii, J. Math. Soc. Japan, 68 (2016), 1297–1323.

[224] T. Yokota, Convex functions and $p$-barycenter on CAT(1)-spaces of small radii, Tsukuba J. Math., 41 (2017), 43–80.

[225] T. Yokota, Law of large numbers in CAT(1)-spaces of small radii, Calc. Var. Partial Differential Equations, 57 (2018), no. 35.

[226] A.J. Zaslavski, Inexact proximal point methods in metric spaces, Set-Valued Var. Anal., 19 (2011), 589–608.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.