

STRONG LAWS, RANDOM MONOTONE VECTOR FIELDS AND GRADIENT FLOWS ON METRIC SPACES OF NONPOSITIVE CURVATURE

NICHOLAS PISCHKE

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

E-mail: nnp39@bath.ac.uk

ABSTRACT. Using a novel effective non-asymptotic concentration inequality, we establish a distribution-uniform generalization of Sturm’s strong law of large numbers for inductive means of L^1 -sequences of i.i.d. random variables on (separable) Hadamard spaces. Building on that, we establish a strong law of large numbers for integrable random monotone vector fields on (separable) Hilbert-Hadamard spaces, that is Hadamard spaces where all tangent cones isometrically embed into Hilbert spaces, extending a previous result of Salim set in Hilbert spaces. We use this latter strong law to establish a probabilistic Lie-Trotter-Kato formula for the gradient flow generated by an integral function over (separable) Hilbert-Hadamard spaces where all tangent cones are actually full Hilbert spaces. This application leverages the previous Lie-Trotter-Kato formula established for gradient flows of sums of convex functions by Stojković together with a result relating resolvent and gradient flow convergence established by Bačák, as well as a new result on the interchangeability of the subdifferential and the integral, which we establish for L^2 -Lipschitz integrands. The last ingredient provides, to our knowledge, the first nonlinear version of (a particular case of) a previous result due variously to Ioffe and Tikhomirov, Levin, Hiriart-Urruty, Thibault as well as Rockafellar and Wets. Throughout the paper, we highlight various remaining open problems.

Keywords: Strong law of large numbers; non-positive curvature; monotone vector fields; Aumann integral; subdifferential; gradient flow; Lie–Trotter–Kato formula

MSC2020 Classification: 60F15, 60B12, 47H05, 47N10, 62L20, 47H20

1. INTRODUCTION

The strong law of large numbers, originally established by Kolmogorov [64], states that for independent and identically distributed (i.i.d.) integrable random variables (X_i) on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, their sample averages converge to their common mean with probability one, i.e.

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow \mathbb{E}[X_1] \quad \mathbb{P}\text{-a.s. as } n \rightarrow \infty.$$

Due to its ubiquity both in theoretical probability and statistics as well as their applications, this result belongs arguably to the most important results in all of probability theory, and many extensions and variants of the strong law of large numbers have been studied over the years (see e.g. the surveys [101, 105], as well as [85, 84, 86] for recent quantitative studies of various related results).

In the present paper, we focus on strong laws of large numbers that take place over complete geodesic metric spaces of nonpositive curvature, so-called Hadamard spaces. Introduced by Alexandrov [1, 2] and often called CAT(0) spaces since the influential work of Gromov [48], this class unifies spaces with smooth structure such as (infinite-dimensional) Hilbert spaces

and Hadamard manifolds (i.e. complete simply connected Riemannian manifolds of nonpositive sectional curvature) with (potentially discrete) structures such as $\mathbb{R}$ -trees or the Billera-Holmes-Vogtmann tree space prominently used in phylogenetics [30], as well as many more. We generally refer to [4, 5, 21, 32] for various treatments of these spaces, which are a key common ground for many analytic, geometric and probabilistic results, with techniques from either area being used to study the others. Probability theory in particular provides various fundamental tools by which analytical or geometrical aspects of these spaces can be studied, going back to the work of Jost [53, 55] and in particular Sturm [112, 115, 114].

The present paper now aims to illustrate how such strong laws over nonpositively curved spaces can be used in combination with other results from stochastic convex analysis to derive new stochastic approximation principles for deterministic gradient flows generated by integral functions. Concretely, as is well-known, if $\mathcal{H}$ is a Hadamard space and $f : \mathcal{H} \rightarrow (-\infty, +\infty]$ is a convex and lower-semicontinuous function such that $f = \sum_{k=1}^N f_k$, then the gradient flow semigroup $S_t(x)$ associated with f (see [6]) can be characterized by a limit of products of the individual proximal mappings $\text{prox}_{\lambda f_1}, \dots, \text{prox}_{\lambda f_N}$ of $f_1, \dots, f_N$, as opposed to a limit of the proximal mapping $\text{prox}_{\lambda f}$ of f as in the usual Crandall-Liggett-type formula (see [79]). More precisely, one has the following limit for all $t \geq 0$ and all $x \in \mathcal{H}$:

$$S_t(x) = \lim_{n \rightarrow \infty} \left(\text{prox}_{\frac{t}{n} f_N} \circ \dots \circ \text{prox}_{\frac{t}{n} f_1} \right)^{(n)}(x).$$

This product formula provides a nonlinear analogue of the well-known Lie-Trotter-Kato product formula for such gradient flow semigroups in linear spaces (see [57], among many other related works) and was first established by Stojković [111] (and later simplified by Bačák [19]).¹

The ultimate goal of the present paper is to establish the following probabilistic variant of the Lie-Trotter-Kato formula (somewhat similar in spirit to [67]) for the gradient flow generated by an integral function $\int f(e, \cdot) d\mu(e)$ for a proper normal convex integrand $f : E \times \mathcal{H} \rightarrow (-\infty, +\infty]$, relating the flow of the integral to powers of products of random samples:

Theorem (informal, Theorem 7.5 in the paper). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ and $(E, \mathbf{E}, \mu)$ be suitable probability spaces and let $(\mathcal{H}, d)$ be a suitable separable Hadamard space. Let $f : E \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a proper normal convex integrand which is L^2 -Lipschitz and suitably integrable. Write $S_t(x)$ for the gradient flow associated with $\int f(e, \cdot) d\mu(e)$, and fix an i.i.d. sequence (ξ_i) of random variables $\xi_i : \Omega \rightarrow E$ with distribution μ .*

Then, there exists a set $\Omega' \subseteq \Omega$ with $\mathbb{P}(\Omega') = 1$ such that for any $\omega \in \Omega'$ as well as for any $t \geq 0$ and $x \in \mathcal{H}$:

$$S_t(x) = \lim_{n \rightarrow \infty} \left(\text{prox}_{\frac{t}{n^2} f(\xi_n(\omega), \cdot)} \circ \dots \circ \text{prox}_{\frac{t}{n^2} f(\xi_1(\omega), \cdot)} \right)^{(n)}(x).$$

In particular, the above limit holds almost surely.

The probability space $(E, \mathbf{E}, \mu)$ later has to satisfy completeness and separability restrictions, and the class of permissible Hadamard spaces is characterized by the condition that their tangent cones are isomorphic to Hilbert spaces (so-called full Hilbert-Hadamard spaces, which encompass Hilbert spaces as well as manifolds with nonpositive sectional curvature but also go beyond these classes, as discussed in more detail later). The integrability condition on f is mild, but a bit technical and so postponed to later.

¹We also mention the work of Ohta and Pálfi [94] which extended this result to CAT(1)-spaces. Generally, while not addressed explicitly throughout, a natural question for future work is whether the results of the present paper extend to more general upper curvature bounds.

The proof of the above result makes crucial use of the usual Lie-Trotter-Kato formula for sums of convex functions already mentioned previously and in particular a new uniform quantitative variant thereof tailored to Lipschitz functions (see Lemma 7.4 later), as well as of a result relating resolvent and gradient flow convergence established by Bačák [22].

However, as mentioned before, the main novel ingredients in the respective argument are strong laws of large numbers together with further results from stochastic convex analysis over these spaces. Concretely, there is the following sequence of intermediate results of independent interest leading up to the above theorem:

- (A) We derive a (distribution-uniform) strong law of large numbers for inductive means of i.i.d. sequences of integrable random variables over general Hadamard spaces.
- (B) We utilize the strong law for integrable random variables to derive a strong law of large numbers for i.i.d. selections of integrable random monotone vector fields on so-called Hilbert-Hadamard spaces (characterized by a flatness condition on the tangent cones).
- (C) To apply the strong law for integrable random monotone vector fields in the context of convex functions, we establish a result on the interchange between the subdifferential and the Aumann integral on so-called full Hilbert-Hadamard spaces (characterized by a linearity condition on the tangent cones, as mentioned before).

The use of this sequence of results can be avoided in certain cases. In the case of Hilbert spaces (see Theorem 7.8 later on), the results can be replaced by an application of a strong law for convex functions due to King and Wets [59], while in the case of a locally compact Hadamard space (see Theorem 7.10 later on, which in particular does not require any further assumptions on the tangent cones), one can utilize a related strong law due to Artstein and Wets [11]. In both cases, the L^2 -Lipschitz assumption can actually be relaxed to L^1 -Lipschitz continuity. Hence, the succession of results (A) – (B) – (C) is only required to establish the above theorem for Hadamard spaces (satisfying the respective tangent cone restrictions) which are neither linear nor locally compact, such as e.g. the infinite dimensional hyperbolic space $\mathbb{H}^\infty$ (see [49]), where it however remains the only route known to us.

Both of the strong laws in (A) and (B) are stronger than actually required in the later proof of the probabilistic Lie-Trotter-Kato formula, and both strong laws are actually established by means of a quantitative result which entails the various uniformities. On the way to each of these results and the theorem above, we prove various other related results of independent interest, and develop a considerable apparatus of stochastic nonlinear analysis for monotone vector fields over Hadamard spaces. We use this coherent and expansive picture not aimed exclusively at a single result to formulate a range of remaining open problems in the area.

The following sections of the introduction now describe each of the above ingredients and related results in more detail.

A. Strong laws for inductive means of random variables on Hadamard spaces. The strong law of large numbers for bounded random variables on Hadamard spaces goes back to the work of Sturm [114] (see Theorem 4.7 therein, and also Theorem 7.2.1 in [21]): Over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a separable Hadamard space $(\mathcal{H}, d)$, if (X_i) is an i.i.d. sequence of random variables $X_i \in L^\infty(\Omega, \mathcal{H}, \mathbb{P})$, then

$$\frac{1}{n} \overrightarrow{\sum}_{i=1, \dots, n} X_i \rightarrow \mathbb{E}[X_1] \quad \mathbb{P}\text{-a.s. as } n \rightarrow \infty.$$

Here, $\frac{1}{n} \overrightarrow{\sum}_{i=1, \dots, n} X_i$ refers to an iteratively defined notion of sample average, the so-called inductive mean introduced by Sturm (see Definition 3.1 later). These inductive means, which coincide with the usual Cesaro means over the reals, are a crucial feature of this strong law

since they allow one to actually compute the sample average without prior knowledge or ability to compute the mean of the sequence. This is in particular opposed to other notions of approximate means, such as e.g. those induced by barycenters (see e.g. [41]), in which case the strong law would be much easier to prove but also of much less interest as the iteration itself becomes already hard to compute (see Remark 4.8 in [114] for further discussions). This strong law has various applications, and can be used to develop further probability theory on these spaces (such as e.g. for deriving Jensen’s inequality [114]) or for developing algorithms to approximate Fréchet means in practice (see e.g. [20]).

Many variants of this result have been studied since then (see [37, 43, 65, 68, 74, 93, 97, 98, 118, 123], among others), but to our knowledge only the work of Yokota [123] explicitly provides an extension of Sturm’s strong law of large numbers using inductive means to L^1 -sequences of i.i.d. random variables over general Hadamard spaces, mirroring Kolmogorov’s original strong law. Indeed, all the other works above (while extending the strong law in various interesting ways in their own right) deviate from this general formulation, either regarding the notion of average by avoiding inductive means (as in [118]), still only entailing a corresponding strong law for bounded or square-integrable variables (as in [37, 43, 65, 93, 97, 98]), or focusing on different (perhaps not even Hadamard) spaces and operator-theoretic formulations (as in [43, 68, 74]).

Our first main result is set in the context of such a general strong law of large numbers on Hadamard spaces using inductive means of L^1 -sequences of i.i.d. random variables, where we establish an even stronger quantitative result by giving the following non-asymptotic concentration inequality:

Theorem (Theorem 3.4 in the paper). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(\mathcal{H}, d)$ be a separable Hadamard space. If (X_i) is an i.i.d. sequence of random variables $X_i \in L^1(\Omega, \mathcal{H}, \mathbb{P})$, then for any $\varepsilon > 0$ and any $m \in \mathbb{N}$:*

$$\mathbb{P} \left(\sup_{n \geq m} d \left(\frac{1}{n} \sum_{i=1, \dots, n}^{\rightarrow} X_i, \mathbb{E}[X_1] \right) > \varepsilon \right) \leq \frac{6}{\varepsilon} \left(\frac{\sqrt{\mathbb{E}[d(o, X_1)]}}{m^{1/2}} + U_{\mathbb{P}}(m^{1/8}) \right) + \frac{72}{\min\{\varepsilon^2, 1\} m^{1/4}},$$

where $U_{\mathbb{P}}(x) := \mathbb{E}[d(o, X_1) \mathbf{1}_{d(o, X_1) \geq x}]$ for a fixed $o \in \mathcal{H}$.

While quantitative estimates for inductive means were previously obtained by Funano [43] under substantially stronger boundedness assumptions (and under restrictions on the space), the above result gives an explicit non-asymptotic estimate (on an arbitrary separable Hadamard space) under the minimal first-moment assumption.

This inequality is derived by applying a type of geodesic truncation to the random variables which allows us to leverage the Busemann convexity of the space as well as the nonexpansiveness of barycenters w.r.t. the Wasserstein metric to reduce the result to a previous non-asymptotic concentration inequality for the usual strong law of large numbers on the reals due to Kachurovskii [56] (see Theorem 3.3 later on, and also the recent [106] for related results) together with a respective quantitative variant of Sturm’s strong law of large numbers for bounded i.i.d. sequences (see Lemma 3.2 later on).² In particular, this provides a new proof of the previous L^1 strong law of large numbers established by Yokota [123] (see Corollary 3.5 later on). Moreover, this concentration inequality is strong enough to also entail a distribution-uniform generalization of the strong law on Hadamard spaces (see Corollary 3.6 later on), extending a corresponding result of Chung [39].

²These quantitative results, as well as the remaining quantitative results in this paper, have been informed and obtained by the logic-based methodology of proof mining [62, 63] and as such are part of a recent advance of applying these methods in the context of probability theory [87, 88]. This paper however avoids any further reference to mathematical logic, as is common in proof mining.

While not explored in this paper in detail, the present approach to the (quantitative) strong law of large numbers seems moreover rather general and could perhaps be used to establish other related results in the literature, such as an extension of the L^1 -version of Birkhoff's pointwise ergodic theorem to Hadamard spaces using inductive means (see Question 3.8 later on) or establishing a strong law of large numbers for pairwise independent and identically distributed random variables over Hadamard spaces (see Question 3.9 later on), among others, as discussed in more detail later.

B. Strong laws for inductive means of random monotone vector fields on Hilbert-Hadamard spaces. Our second main result is an extension of this strong law of large numbers to random monotone vector fields. On Hilbert spaces, so-called maximal monotone operators are key objects in the modern theory underlying convex analysis as well as that of differential equations and their semigroups, among others (see e.g. [18]), and so-called monotone vector fields extend this central notion and its theory to Riemannian manifolds using their tangential structure [83, 89, 72, 73, 121]. Recently, these mappings have in particular also been studied over Hadamard spaces (see e.g. [36]), where they make use of the tangent cones of these spaces in the sense of Alexandrov, Berestovskii and Nikolaev [3]. Random monotone vector fields are, similarly to random monotone operators on Hilbert spaces, stochastically perturbed versions of these mappings which feature in stochastic convex analysis (see e.g. [27, 28, 29] as well as the recent [97] for extensions to the setting of Hadamard spaces).

A strong law of large numbers for an integrable random monotone operator on Hilbert spaces was recently established by Salim [108]. Concretely, if $\mathcal{M}(\mathcal{H})$ is the collection of all maximally monotone operators on a separable Hilbert space $\mathcal{H}$, it was shown in [108] that the averages $\frac{1}{n} \sum_{i=1}^n A(\xi_i)$ of a suitably measurable and integrable mapping $A : E \rightarrow \mathcal{M}(\mathcal{H})$ generated via an i.i.d. sequence of random variables $\xi_i : \Omega \rightarrow E$ converge a.s. to the integral operator $\int A(e) d\mu(e)$ (defined using the Aumann integral [15]), that is

$$\frac{1}{n} \sum_{i=1}^n A(\xi_i) \rightarrow^R \int A(e) d\mu(e) \quad \mathbb{P}\text{-a.s. as } n \rightarrow \infty,$$

where convergence here is in the sense of R -convergence (see e.g. [12]), that is in terms of the resolvents of the operators. This strong law of Salim is quite different to previous strong laws of large numbers for random sets [34, 109, 116, 117, 9, 10] and more closely related to strong laws for convex functions [11, 59], and was illustrated to have applications to empirical risk minimization (see the discussions in [108]).

Using an appropriate notion of inductive mean for sets in a Hadamard space and a corresponding variant of R -convergence, we extend this result of Salim (using a somewhat different approach) to integrable random monotone vector fields (generalizing the respective notion from [108]):

Theorem (informal, Theorem 5.3 in the paper). *Let $(\Omega, \mathbf{F}, \mathbb{P})$ and $(E, \mathbf{E}, \mu)$ be probability spaces and let $(\mathcal{H}, d)$ be a suitable separable Hadamard space. Further, fix an i.i.d. sequence (ξ_i) of random variables $\xi_i : \Omega \rightarrow E$ with distribution μ . Let $A : E \rightarrow \mathcal{M}(\mathcal{H})$ be a suitably measurable and integrable mapping, where $\mathcal{M}(\mathcal{H})$ refers to the collection of all monotone vector fields on $\mathcal{H}$ with total resolvents. Then*

$$\frac{1}{n} \sum_{i=1, \dots, n}^{\rightarrow} A(\xi_i) \rightarrow^R \int A(e) d\mu(e) \quad \mathbb{P}\text{-a.s.}$$

as $n \rightarrow \infty$.

The class of Hadamard spaces that this result is phrased over is that of Hilbert-Hadamard spaces as introduced by Gong, Wu and Yu [45], that is Hadamard spaces where every tangent cone is flat (which are discussed in detail later, and also featured in recent investigations into stochastic optimization [97]). In particular, the resulting strong law recovers the result from [108] and additionally holds on Hadamard manifolds and their infinite-dimensional generalizations (as well as other spaces, see Example 2.2 later on).

This result and the surrounding discussion relies on various properties of (random) monotone vector fields on (Hilbert-)Hadamard spaces, established here for the first time, which have independent use in (stochastic) convex analysis where these objects are often employed. Also this strong law of large numbers for vector fields can be made quantitative using a respective concentration inequality phrased using the resolvents (see Remark 5.4 later on).

C. An interchange result between the subdifferential and the Aumann integral on full Hilbert-Hadamard spaces. A canonical example of a (random) monotone vector field is the (random) subdifferential $\partial f(t, x)$ of a proper normal convex integrand $f : T \times \mathcal{H} \rightarrow (-\infty, +\infty]$ on a complete probability space $(T, \mathbb{T}, \tau)$ and a suitable space $\mathcal{H}$ (see [104]). Crucial for both the applicability of the associated strong law of large numbers for random monotone vector fields in the context of gradient flows as illustrated in this paper, but also for other results in (stochastic) convex analysis (see in particular [97], where the question for such a result over Hilbert-Hadamard spaces $\mathcal{H}$ is raised) are interchange results between the random subdifferential and the integral of the form

$$\int \partial f(t, x) d\tau(t) = \partial \left(\int f(t, \cdot) d\tau(t) \right) (x) \text{ for all } x \in \mathcal{H}$$

where the integral on the left-hand side is again the Aumann integral.

Results of this type were famously considered by Ioffe and Tikhomirov [52], Levin [69], Hiriart-Urruty [51], Thibault [119] as well as Rockafellar and Wets [103], among others, and have various uses in stochastic optimization and analysis next to the one mentioned above (see in particular the discussions in the above references).

As a third main result, we establish such an interchange when f is L^2 -Lipschitz and suitably integrable:

Theorem (informal, Theorem 6.15 in the paper). *Let $(T, \mathbb{T}, \tau)$ be a suitable probability space and let $(\mathcal{H}, d)$ be a suitable separable Hadamard space. Let $f : T \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a proper normal convex integrand and suppose that f is L^2 -Lipschitz and suitably integrable. Then, for any $x \in \mathcal{H}$:*

$$\int \partial f(t, x) d\tau(t) = \partial \left(\int f(t, \cdot) d\tau(t) \right) (x).$$

The integrability condition on f is quite mild, and just requires that there exists a $u_0 \in L^2(T, \mathcal{H}, \tau)$ such that $f(\cdot, u_0(\cdot)) \in L^1(T, \mathbb{R}, \tau)$. Next to being complete, the probability space $(T, \mathbb{T}, \tau)$ moreover has to be separable for this result. The class of Hadamard spaces that this result is phrased over is characterized by the assumption that all tangent cones of the Hadamard space are not only flat but are actually isometric to Hilbert spaces. This class of spaces, introduced as full Hilbert-Hadamard spaces in the present paper, is a further restriction of the Hilbert-Hadamard spaces discussed previously, also naturally covering Hilbert spaces and Hadamard manifolds and their infinite-dimensional variants, and which is still not limited to those (see again Example 2.2 later on). In particular, the present result seems to be the first such interchange result even in the manifold case. Also, this assumption of a Hilbert space structure

on the tangent cones is in a way sharp since it does not extend to plain Hilbert-Hadamard spaces, as illustrated by an example (see Example 6.2 later on).

As even full Hilbert-Hadamard spaces do not have a single global dual like linear spaces, establishing this result is rather tricky, and relies on utilizing various tools, including conjugates of convex functions on Hadamard spaces motivated by [26, 110], measurable selection theory (see [14, 35]), directional derivatives of convex functions as studied in [78, 81] (see also [75, 95, 99]), in particular also making use of recent results of Gigli and Nobile [44] on the study of the tangent cones of the Hadamard-space of square integrable functions over a Hadamard space, and classical results over Hilbert spaces like Hahn-Banach and the Riesz representation theorem.

Outline of the paper. The paper is organized as follows: Section 2 provides the various preliminaries used throughout, such as background on geodesic metric spaces, Hadamard spaces and their tangent cones as well as the special class of Hilbert-Hadamard spaces and basic probability theory over these spaces. In Section 3, we derive the (quantitative and distribution-uniform) strong law of large numbers for inductive means of integrable random variables on Hadamard spaces and discuss various related problems and questions. Section 4 introduces monotone vector fields on Hadamard spaces as well as their random variants, and derives various initial properties required later. In Section 5, we then derive the strong law of large numbers for integrable random monotone vector fields. Section 6 is concerned with the interchange result for the integral and the subdifferential and the final Section 7 combines the previous work into the probabilistic Lie-Trotter-Kato formula outlined before, and discusses special and alternative cases.

2. PRELIMINARIES

2.1. Geodesics, Hadamard spaces and tangent cones. Let $(\mathcal{H}, d)$ be a metric space. A geodesic is an isometry $\gamma : [0, l] \rightarrow \mathcal{H}$. We say that γ joins $\gamma(0)$ and $\gamma(l)$, as well as that γ issues from $\gamma(0)$. Note that necessarily $l = d(\gamma(0), \gamma(l))$. $\mathcal{H}$ is called (uniquely) geodesic if every two points $x, y \in \mathcal{H}$ are joined by a (unique) geodesic and if such a geodesic is unique, we denote it by $\gamma_{x,y}$. A CAT(0) space (also called a space of nonpositive curvature in the sense of Alexandrov) is a geodesic metric space $(\mathcal{H}, d)$ such that the (extended) Bruhat-Tits CN-inequality [33] holds (see also Proposition 2.3 in [114]):

$$d^2(\gamma(tl), x) \leq (1-t)d^2(\gamma(0), x) + td^2(\gamma(l), x) - t(1-t)d^2(\gamma(0), \gamma(l))$$

for all $x \in \mathcal{H}$, $t \in [0, 1]$ and all geodesics $\gamma : [0, l] \rightarrow \mathcal{H}$. Any CAT(0) space is uniquely geodesic and a complete CAT(0) space is called a Hadamard space. In that context, we also sometimes write $(1-t)x \oplus ty$ given $x, y \in \mathcal{H}$ and $t \in [0, 1]$ for the point $\gamma(tl)$ on the unique geodesic $\gamma : [0, l] \rightarrow \mathcal{H}$ with $\gamma(0) = x$ and $\gamma(l) = y$. It should further be noted that every CAT(0) space is a Busemann space (see e.g. Theorem 1.3.3 in [21]), that is it holds that

$$d(\gamma(tl), \eta(tk)) \leq (1-t)d(\gamma(0), \eta(0)) + td(\gamma(l), \eta(k))$$

for any two geodesics $\gamma : [0, l] \rightarrow \mathcal{H}$ and $\eta : [0, k] \rightarrow \mathcal{H}$ and any $t \in [0, 1]$. As mentioned before, we refer to [5, 21, 32] for comprehensive overviews of geodesic metric spaces, CAT(0) spaces and Hadamard spaces, including alternative definitions.

For nonconstant geodesics γ and η issuing from a point $x \in \mathcal{H}$ in a Hadamard space $(\mathcal{H}, d)$, we define their Aleksandrov angle $\angle_x(\gamma, \eta)$ by

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

with $\bar{\angle}_x(y, z)$ referring to the comparison angle, defined via the comparison triangle $\bar{\Delta}(\bar{x}, \bar{y}, \bar{z})$ of the geodesic triangle $\Delta(x, y, z) \subseteq \mathcal{H}$, as usual (see e.g. [32]). Using the Aleksandrov angle, we

can define the tangent cone $T_x\mathcal{H}$ of $\mathcal{H}$ at a point $x \in \mathcal{H}$ in the sense of Alexandrov, Berestovskii and Nikolaev [3]. For that, first note that the Aleksandrov angle $\angle_x$ is a pseudometric on the set of all nonconstant geodesics issuing from x . Writing $\Sigma'_x\mathcal{H}$ for the space of all equivalence classes of such geodesics under the equivalence relation defined by $\angle_x(\gamma, \eta) = 0$, the completion $(\Sigma_x\mathcal{H}, \angle_x)$ of the space $(\Sigma'_x\mathcal{H}, \angle_x)$ is called the space of directions from x . The tangent cone $T_x\mathcal{H}$ of $\mathcal{H}$ at x is then the Euclidean cone over $\Sigma_x\mathcal{H}$, that is $T_x\mathcal{H} := (\Sigma_x\mathcal{H} \times [0, \infty)) / \sim$ where $(\gamma, t) \sim (\eta, s)$ if, and only if, $t = s = 0$ or $t = s > 0$ and $\gamma = \eta$. For brevity, we write $t\gamma$ for the equivalence class $[(\gamma, t)]_\sim$ and given $u = t\gamma$ and $\lambda \geq 0$, we write $\lambda u := (\lambda t)\gamma$. On $T_x\mathcal{H}$, we consider the metric

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

By a result due to Nikolaev [91], $T_x\mathcal{H}$ is a Hadamard space itself. We call the set $T\mathcal{H} = \bigcup_{x \in \mathcal{H}} T_x\mathcal{H}$ the tangent bundle of $\mathcal{H}$. On $T_x\mathcal{H}$, we now write $0_x := 0\gamma$ as well as $\|t\gamma\|_x := d_x(0_x, t\gamma) = t$ and

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

Crucially, we have that $g_x(t\gamma, s\eta) \leq \|t\gamma\|_x \|s\eta\|_x$, as well as $g_x(t\gamma, t\gamma) = \|t\gamma\|_x^2$, $g_x(t\gamma, s\eta) = g_x(s\eta, t\gamma)$ and $g_x(t\gamma, s\eta) = tg_x(\gamma, s\eta)$. We refer to [71] as well as [5, 32] for further exposition and proofs of the above results.

2.2. Flatness and Hilbert-Hadamard spaces. We call a Hadamard space $\mathcal{H}$ flat (see e.g. [58]) if the CN-inequality is actually an equality, i.e. if

$$d^2(\gamma(tl), x) = (1-t)d^2(\gamma(0), x) + td^2(\gamma(l), x) - t(1-t)d^2(\gamma(0), \gamma(l))$$

for all $x \in \mathcal{H}$, $t \in [0, 1]$ and all geodesics $\gamma : [0, l] \rightarrow \mathcal{H}$. Following Gong, Wu and Yu [45], we call a Hadamard space $\mathcal{H}$ a Hilbert-Hadamard space if every tangent cone $T_x\mathcal{H}$ is flat.

Remark 2.1. This formulation diverges slightly from the definition used in [45], where instead of flatness, it is required that $T_x\mathcal{H}$ isometrically embeds into a Hilbert space (which in turn is equivalent to being isometric to a closed and convex subset of a Hilbert space). However, this condition is equivalent to flatness (see in particular Remark 2.4 in [97]).

As discussed in Example 3.2 of [45], Hilbert spaces and Hadamard manifolds are Hilbert-Hadamard spaces, as are complete simply connected Hilbert manifolds of nonpositive sectional curvature, that is generalizations of Hadamard manifolds modelled on (potentially infinite-dimensional) Hilbert spaces (such as e.g. the infinite dimensional hyperbolic space $\mathbb{H}^\infty$, see [49]). Further, Hilbert-Hadamard spaces are closed under taking closed and convex subsets. As a flat space is isometric to a closed and convex subset of a Hilbert space (recall Remark 2.1), every flat space is itself a Hilbert-Hadamard space. Therefore, tangent cones of Hilbert-Hadamard spaces are themselves Hilbert-Hadamard spaces. Further, whenever $\mathcal{H}$ is a Hilbert-Hadamard space, so is the space of all square-integrable functions $L^2(T, \mathcal{H}, \tau)$ for a finite measure space $(T, \mathbb{T}, \tau)$, discussed in detail later. This latter construction in particular allows Hilbert-Hadamard spaces to reach beyond the setting of manifolds (see the discussion in [45]). We here provide another example of a Hilbert-Hadamard space which is not a manifold:

Example 2.2. Let $\mathcal{H}_1 = \{(x, y) \in \mathbb{H} \mid x \geq 0\}$ be the right half of the usual Poincaré upper half plane $\mathbb{H}$ with constant curvature -1 (see [32]), which is a closed and convex subset of $\mathbb{H}$ and hence a Hadamard space. Let $\mathcal{H}_2$ be a scaled version of $\mathcal{H}_1$ with constant curvature -2 . As their left boundaries $\{(0, y) \mid y > 0\}$ are isometric to $\mathbb{R}$, we can glue $\mathcal{H}_1$ and $\mathcal{H}_2$ along these boundaries to produce $\mathcal{H} := \mathcal{H}_1 \sqcup_{\mathbb{R}} \mathcal{H}_2$. By Reshetnyak's gluing theorem (see e.g. Theorem 11.1 in [32]), as $\mathbb{R}$ is convex and $\mathcal{H}_1, \mathcal{H}_2$ are Hadamard spaces, we get that $\mathcal{H}$ is also a Hadamard

space. If $x \in \mathcal{H}$ is away from the seam, it is easy to see that we have $T_x \mathcal{H} \cong \mathbb{R}^2$. If $x \in \mathcal{H}$ is on the seam, then $\Sigma_x \mathcal{H}$ is the gluing of two semi-circles, that is $\Sigma_x \mathcal{H} \cong S^1$. Therefore $T_x \mathcal{H} \cong \mathbb{R}^2$ as well. In particular, $\mathcal{H}$ is a Hilbert-Hadamard space. However, $\mathcal{H}$ cannot be isometric to a smooth Riemannian manifold as the curvature is discontinuous across the seam.

Note that the tangent cones in the above Example 2.2 are not only all flat but actually all isometric to a Hilbert space, which as highlighted in the introduction is a class of Hilbert-Hadamard spaces that will be important for our interchange result between the subdifferential and the integral as well as our probabilistic Lie-Trotter-Kato formula later on.

Whenever $T_x \mathcal{H}$ is flat, we in particular have that $g_x(u, v)$ is affine, that is it is both convex and concave, in both arguments. This follows from a similar argument as in the proof of Theorem 2.4, (4) in [47] and was already explicitly noted in [97].

Lemma 2.3 (Lemma 2.6 in [97]). *Let $x \in \mathcal{H}$. If $T_x \mathcal{H}$ is flat, then $g_x(u, v)$ is affine in both arguments.*

Following [60] (see also already [3, 25]), we define a metric analog $\log_x : \mathcal{H} \rightarrow T_x \mathcal{H}$ of the inverse exponential map from Riemannian geometry by $\log_x a := d(x, a)\gamma_{x,a}$ for $a \neq x$ as well as $\log_x x := 0_x$. Crucially, note that $\log_x$ is nonexpansive:

Lemma 2.4 (see e.g. p. 224 in [45]). *For any $x, a, b \in \mathcal{H}$:*

$$d_x(\log_x a, \log_x b) \leq d(a, b).$$

The most important property of the pseudo-Riemannian metric g_x on $T_x \mathcal{H}$ is the following estimate relative to the inverse exponential map:

Lemma 2.5 (essentially Proposition 2.16 in [36]). *For any $x, a, b \in \mathcal{H}$:*

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

Further, we will need that g_x is Lipschitz continuous in both arguments.

Lemma 2.6 (Lemma 2.3 in [97]). *For $x \in \mathcal{H}$ and $u, v, w \in T_x \mathcal{H}$:*

$$|g_x(u, v) - g_x(u, w)| \leq \|u\|_x d_x(v, w).$$

2.3. Measurability and integration on CAT(0) spaces. Let $(T, \mathbb{T}, \tau)$ be a probability space and $(\mathcal{H}, d)$ be a separable Hadamard space. An $(\mathcal{H}$ -valued) random variable is a map $x : T \rightarrow \mathcal{H}$ which is $\mathbb{T}/\mathcal{B}(\mathcal{H})$ -measurable, where $\mathcal{B}(\mathcal{H})$ is the Borel σ -algebra of $\mathcal{H}$. Write $\mathcal{P}(\mathcal{H})$ for the set of all probability measures on $\mathcal{H}$ and for $p \in [1, \infty)$, write $\mathcal{P}^p(\mathcal{H})$ for the set of all measures $P \in \mathcal{P}(\mathcal{H})$ such that $\int d^p(w, z) dP(w) < \infty$ for some/any $z \in \mathcal{H}$. The push-forward measure $\tau_x(A) := \tau(x^{-1}(A))$ for $A \in \mathcal{B}(\mathcal{H})$ is called the distribution of x . Naturally, we have $\tau_x \in \mathcal{P}(\mathcal{H})$.

Fundamentally (compare e.g. Theorem 2.3.1 in [21]), we have the following result in Hadamard spaces: For any $P \in \mathcal{P}^1(\mathcal{H})$ and some/any $y \in \mathcal{H}$, there is a unique minimizer

$$b(P) := \operatorname{argmin}_{z \in \mathcal{H}} \int (d^2(z, w) - d^2(w, y)) dP(w),$$

which is independent of y , called the barycenter of P . Given $p \in [1, \infty)$, the space $\mathcal{L}^p(T, \mathcal{H}, \tau)$ is the space of all $\mathbb{T}/\mathcal{B}(\mathcal{H})$ -measurable maps $x : T \rightarrow \mathcal{H}$ with $d_p(x, z) < \infty$ for some/any $z \in \mathcal{H}$, where $d_p^p(x, y) := \int d^p(x, y) d\tau$ for $\mathbb{T}/\mathcal{B}(\mathcal{H})$ -measurable maps $x, y : T \rightarrow \mathcal{H}$. The space $L^p(T, \mathcal{H}, \tau)$ arises from this space by considering equivalence classes under the equivalence relation defined by $d_p(x, y) = 0$. Note that $x \in L^p(T, \mathcal{H}, \tau)$ if, and only if, $\tau_x \in \mathcal{P}^p(\mathcal{H})$. For

$p = \infty$, the space $L^\infty(T, \mathcal{H}, \tau)$ is the space of all almost sure equivalence classes of $\mathbb{T}/\mathbb{B}(\mathcal{H})$ -measurable maps $x : T \rightarrow \mathcal{H}$ where $d(x, z) \in L^\infty(T, \mathbb{R}, \tau)$ for some/any $z \in \mathcal{H}$. The expectation of a random variable $x \in L^1(T, \mathcal{H}, \tau)$ is now defined via

$$\int x d\tau := b(\tau_x)$$

where $b(\tau_x)$ is the barycenter of τ_x as before. Later, we will sometimes work with multiple probability spaces simultaneously. In that context, we generally use $(\Omega, \mathbb{F}, \mathbb{P})$ as notation for a ground space over which results take place, where we use $\mathbb{E}[x]$ as an alternative notation for the integral $\int x d\mathbb{P}$, and $(E, \mathbb{E}, \mu)$ as notation for a secondary space. We use the notation $(T, \mathbb{T}, \tau)$ in general sections like the present to make it easier to transfer results to either case.

Before we move on, we briefly consider the above notion in the special case of a discrete measure space: If $T = \{t_1, \dots, t_n\}$ is a finite set with full σ -algebra $\mathbb{T} = 2^T$, a measurable random variable $x : T \rightarrow \mathcal{H}$ can simply be identified with a tuple of points $(x_1, \dots, x_n)$ from $\mathcal{H}$. If τ is a measure on $(T, \mathbb{T})$ with $\tau(\{t_i\}) = w_i$ such that $\sum_{i=1}^n w_i = 1$, then the integral $\int x d\tau$ of $x = (x_1, \dots, x_n)$ reduces to

$$\operatorname{argmin}_{z \in \mathcal{H}} \int d^2(z, x(t)) d\tau(t) = \operatorname{argmin}_{z \in \mathcal{H}} \sum_{i=1}^n w_i d^2(z, x_i),$$

also called the Fréchet mean of the points $(x_1, \dots, x_n)$ (see e.g. [21]), a generalization of the arithmetic mean in Hadamard spaces which is of central importance for various applications. For this, we further refer to the discussions in [21] as well as [46], the latter of which gives various computational approaches to recognizing whether a point arises as a mean itself.

The first key theoretical result on the expectation that we rely on here is as follows:

Lemma 2.7 (Corollary 2.4 in [113]). *For any $x, y \in L^1(T, \mathcal{H}, \tau)$:*

$$d\left(\int x d\tau, \int y d\tau\right) \leq \int d(x, y) d\tau.$$

The above lemma essentially states (and can be derived from) the nonexpansivity of the barycenter relative to the Wasserstein (or Kantorovich-Rubinstein) metric

$$W_1(P, P') := \inf_{\gamma \in \Gamma(P, P')} \int d(x, y) d\gamma(x, y)$$

on $\mathcal{P}^1(\mathcal{H})$, where $\Gamma(P, P')$ is the set of all couplings

$$\Gamma(P, P') := \{\gamma \in \mathcal{P}(\mathcal{H}^2) \mid \gamma(A \times \mathcal{H}) = P(A) \text{ and } \gamma(\mathcal{H} \times A) = P'(A) \text{ for all } A \in \mathbb{B}(\mathcal{H})\}.$$

The second key result on the expectation that we rely on here is the following analogous version of Jensen's inequality:

Lemma 2.8 (Proposition 3.4 in [113]). *Let $x \in L^1(T, \mathcal{H}, \tau)$ and let $\varphi : \mathcal{H} \rightarrow \mathbb{R}$ be lower-semicontinuous and convex with $(\varphi \circ x)_+ \in L^1(T, \mathbb{R}, \tau)$. Then*

$$\int \varphi \circ x d\tau \geq \varphi\left(\int x d\tau\right).$$

We refer to the fundamental works of Sturm [113, 114] as well as the exposition of these matters given by Bačák [21] for further information on the above.

We will later use the extension of the above integral to set-valued operators. Following [97], we define this operation in an analogous way to the seminal work of Aumann [15] set in Hilbert spaces, replacing the use of the Bochner integral therein with the integral of Sturm. Concretely,

given a set-valued operator $F : T \rightarrow 2^{\mathcal{H}}$, a function $\phi : T \rightarrow \mathcal{H}$ is a measurable selection of F if it is $\mathbb{T}/\mathbb{B}(\mathcal{H})$ -measurable and $\phi(t) \in F(t)$ for all $t \in T$. The set of all measurable selections of F is denoted by $\mathcal{S}(F)$ and we write $\mathcal{S}^p(F) := \mathcal{S}(F) \cap L^p(T, \mathcal{H}, \tau)$. The Aumann-Sturm integral of F is then defined as

$$\int F d\tau = \left\{ \int \phi d\tau \mid \phi \in \mathcal{S}^1(F) \right\}.$$

Again, we briefly consider the above notion over discrete measure spaces. If $T = \{t_1, \dots, t_n\}$ is a finite set with full σ -algebra $\mathbb{T} = 2^T$, a set-valued map $F : T \rightarrow 2^{\mathcal{H}}$ can be identified with a tuple $(F_1, \dots, F_n)$ of sets $F_i \subseteq \mathcal{H}$. If now again τ is a measure on $(T, \mathbb{T})$ with $\tau(\{t_i\}) = w_i$ such that $\sum_{i=1}^n w_i = 1$, then the integral $\int F d\tau$ reduces to

$$\left\{ \operatorname{argmin}_{z \in \mathcal{H}} \sum_{i=1}^n w_i d^2(z, x_i) \mid x_i \in F_i \text{ for all } i = 1, \dots, n \right\},$$

that is the set of all Fréchet means which can be formed for tuples from $F_1 \times \dots \times F_n$.

As we will later be concerned with probabilistic notions on $T_x \mathcal{H}$, we will need to ensure that the respective concepts are well-defined thereon. In particular, this entails a separability assumption on $T_x \mathcal{H}$ which we briefly discuss in the following remark (compare Remark 2.18 in [97]).

Remark 2.9. We later crucially rely on the tangent cone $T_x \mathcal{H}$ of $\mathcal{H}$ being separable. This however directly follows if the underlying Hadamard space $\mathcal{H}$ separable as discussed in Remark 2.18 in [97]. This in particular ensures that d_x is jointly measurable and in particular also guarantees that g_x and $\|\cdot\|_x$ are (jointly) measurable as well. Further, note that $\log_x$ is measurable as it is nonexpansive and hence uniformly continuous.

It is in particular in the context of the interaction between the integral and the pseudo-Riemannian metric that we in this paper make essential use of the assumption that $T_x \mathcal{H}$ is flat later on, that is that $\mathcal{H}$ is a Hilbert-Hadamard space. Concretely, in a separable Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$, it is well-known that independent ($\mathcal{H}$ -valued) random variables $u, v : T \rightarrow \mathcal{H}$ satisfy

$$\int \langle u, v \rangle d\tau = \left\langle \int u d\tau, \int v d\tau \right\rangle,$$

where the integral in that case reduces to the Bochner integral on $\mathcal{H}$. Flatness of $T_x \mathcal{H}$ allows us to extend this property to the pseudo-Riemannian metric g_x , where we crucially rely on Lemma 2.3 above. Like that lemma, also this observation was already made and crucially employed in [97] (in fact even an extension to conditional expectations, which we will not need here).

Lemma 2.10 (Lemma 2.19 in [97]). *Fix $x \in \mathcal{H}$ and let $T_x \mathcal{H}$ be flat. For independent integrable random variables $u, v : T \rightarrow T_x \mathcal{H}$, we have*

$$\int g_x(u, v) d\tau = g_x \left(\int u d\tau, \int v d\tau \right).$$

Lastly, we discuss the space $L^2(T, \mathcal{H}, \tau)$ in more detail, which features crucially in a later section on the interchange of the subdifferential and the integral. Since at least the work of Jost (see Corollary 4.1.1 in [55] and see also e.g. Proposition 3.3 [112]), it is known that the space $L^2(T, \mathcal{H}, \tau)$ is itself a Hadamard space whenever $\mathcal{H}$ is one. As mentioned before, it was shown by Gong, Wu and Yu (see Proposition 3.13 in [45]) that if $\mathcal{H}$ is a Hilbert-Hadamard space, then also $L^2(T, \mathcal{H}, \tau)$ is a Hilbert-Hadamard space, and if $\mathcal{H}$ and T are further separable,³ then so

³A probability space $(T, \mathbb{T}, \tau)$ is called separable if the metric space $(\mathbb{T}, d_\tau)$ with $d_\tau(A, B) = \tau(A \triangle B)$ is separable, where $A \triangle B$ is the symmetric difference of $A, B \in \mathbb{T}$.

is $L^2(T, \mathcal{H}, \tau)$. A study of the tangent cones of $L^2(T, \mathcal{H}, \tau)$ can be found in Proposition 4.5 of [44], already for much more general spaces $\mathcal{H}$, by which $T_x L^2(T, \mathcal{H}, \tau)$ for $x \in L^2(T, \mathcal{H}, \tau)$ is isometric to the space of L^2 -sections of the so-called pullback $x^* T_G \mathcal{H}$ of the geometric tangent bundle $T_G \mathcal{H}$ (see [44] for precise definitions). In the case that $\mathcal{H}$ is a separable Hadamard space and that $(T, \mathbb{T}, \tau)$ is separable (as later assumed), Proposition 4.8 of [44] shows that this characterization of $T_x L^2(T, \mathcal{H}, \tau)$ reduces to the space of all equivalence classes of measurable u such that $u(t) \in T_{x(t)} \mathcal{H}$ almost surely and $\int \|u(t)\|_{x(t)}^2 d\tau(t) < +\infty$. In particular, if $x \in \mathcal{H}$, then $T_x L^2(T, \mathcal{H}, \tau)$ is isometric to $L^2(T, T_x \mathcal{H}, \tau)$. In that case, one also in particular obtains that for $u, v \in T_x L^2(T, \mathcal{H}, \tau)$ that

$$d_x^2(u, v) = \int d_x^2(u(t), v(t)) d\tau(t)$$

and so $\|u\|_x^2 = \int \|u\|_x^2 d\tau$ as well as $g_x(u, v) = \int g_x(u(t), v(t)) d\tau(t)$.

3. AN L^1 STRONG LAW OF LARGE NUMBERS

The main goal of this section is to prove the quantitative generalization of the strong law of large numbers for L^1 i.i.d. processes where crucially, as already motivated in the introduction, the notion of sample average is that of the inductive mean, following Sturm [113, 114]:

Definition 3.1 (Inductive mean). Let $(x_i) \subseteq \mathcal{H}$ be given. The sequence of inductive means $(s_n) \subseteq \mathcal{H}$ of (x_i) is defined as follows:

$$s_1 := x_1 \text{ and } s_n := \left(1 - \frac{1}{n}\right) s_{n-1} \oplus \frac{1}{n} x_n.$$

We write $\frac{1}{n} \overrightarrow{\sum}_{i=1, \dots, n} x_i$ for s_n . In general, this point will depend on the permutation of the (x_i) .

To derive the distribution-dependent non-asymptotic concentration inequality, we first require a quantitative variant of Sturm's L^∞ strong law of large numbers:

Lemma 3.2. *Let $(\Omega, \mathbb{F}, \mathbb{P})$ be a probability space and let $(\mathcal{H}, d)$ be a separable Hadamard space. If (X_i) is an i.i.d. sequence of random variables $X_i \in L^\infty(\Omega, \mathcal{H}, \mathbb{P})$ with $d(o, X_1) \leq b$ $\mathbb{P}$ -a.s. for some $b \geq 1$ and $o \in \mathcal{H}$, then for any $\varepsilon > 0$ and $m \in \mathbb{N}$:*

$$\mathbb{P} \left(\sup_{n \geq m} d \left(\frac{1}{n} \overrightarrow{\sum}_{i=1, \dots, n} X_i, \mathbb{E}[X_1] \right) > \varepsilon \right) \leq \frac{8b^2}{\min\{\varepsilon^2, 1\} m^{1/2}}.$$

Proof. Write S_n for $\frac{1}{n} \overrightarrow{\sum}_{i=1, \dots, n} X_i$. We first show that for any $\varepsilon, \lambda > 0$:

$$(*) \quad \mathbb{P} \left(\sup_{n \geq (\max\{4b^2/\varepsilon^2 \lambda, 8b/\varepsilon\})^2} d(S_n, \mathbb{E}[X_1]) > \varepsilon \right) \leq \lambda.$$

As shown in the proof of Theorem 4.7 in [114], it holds that

$$\sum_{n=1}^{\infty} \mathbb{P}(d(S_{n^2}, \mathbb{E}[X_1]) > \varepsilon) \leq \sum_{n=1}^{\infty} \frac{\mathbb{V}[X_1]}{\varepsilon^2 n^2},$$

where $\mathbb{V}[X_1] := \inf_{z \in \mathcal{H}} \mathbb{E}[d^2(z, X_1)]$ is the variance of X_1 . Note that $\mathbb{V}[X_1] \leq \mathbb{E}[d^2(o, X_1)] \leq b^2$. As the measure is sub- σ -additive, we get

$$\begin{aligned} \mathbb{P}\left(\sup_{n \geq n_0} d(S_{n^2}, \mathbb{E}[X_1]) > \varepsilon\right) &\leq \sum_{n=n_0}^{\infty} \mathbb{P}(d(S_{n^2}, \mathbb{E}[X_1]) > \varepsilon) \\ &\leq \frac{b^2}{\varepsilon^2} \sum_{n=n_0}^{\infty} \frac{1}{n^2} \leq \frac{b^2}{\varepsilon^2 n_0}. \end{aligned}$$

As further shown in the proof of Theorem 4.7 in [114], one has $d(S_{n^2}, S_k) \leq 4b/n$ for all $n, k \geq 1$ with $n^2 \leq k < (n+1)^2$. For $\varepsilon, \lambda > 0$ now fixed, write

$$n_0 := \max \left\{ \frac{4b^2}{\varepsilon^2 \lambda}, \frac{8b}{\varepsilon} \right\}.$$

For $\omega \in \Omega$ such that $d(S_{n^2}(\omega), \mathbb{E}[X_1]) \leq \varepsilon/2$ for all $n \geq n_0$ and $d(S_{n^2}(\omega), S_k(\omega)) \leq 4b/n$ for all $n, k \geq 1$ as above, if $k \geq n_0^2$, then $k \in [n^2, (n+1)^2)$ for some $n \geq n_0$. Hence, for such a k , we have

$$d(S_k(\omega), \mathbb{E}[X_1]) \leq d(S_k(\omega), S_{n^2}(\omega)) + d(S_{n^2}(\omega), \mathbb{E}[X_1]) \leq \frac{4b}{n} + \frac{\varepsilon}{2} \leq \frac{4b}{n_0} + \frac{\varepsilon}{2} \leq \varepsilon.$$

Therefore, we get

$$\mathbb{P}\left(\sup_{k \geq n_0^2} d(S_k, \mathbb{E}[X_1]) > \varepsilon\right) \leq \mathbb{P}\left(\sup_{n \geq n_0} d(S_{n^2}, \mathbb{E}[X_1]) > \varepsilon/2\right) \leq \frac{4b^2}{\varepsilon^2 n_0} \leq \lambda.$$

This shows (*). To now see the inequality claimed in the theorem, note first that (*) implies

$$\mathbb{P}\left(\sup_{n \geq n_\lambda^2} d(S_n, \mathbb{E}[X_1]) > \varepsilon\right) \leq \lambda$$

for $n_\lambda := 8b^2/(\min\{\varepsilon^2, 1\} \min\{\lambda, 1\})$. Now, if $m^{1/2} \geq 8b^2/\min\{\varepsilon^2, 1\}$, we obtain

$$\mathbb{P}\left(\sup_{n \geq m} d(S_n, \mathbb{E}[X_1]) > \varepsilon\right) \leq \frac{8b^2}{\min\{\varepsilon^2, 1\} m^{1/2}}$$

by setting $\lambda := 8b^2/(\min\{\varepsilon^2, 1\} m^{1/2})$ in the above. However, as $8b^2/(\min\{\varepsilon^2, 1\} m^{1/2}) > 1$ if $m^{1/2} < 8b^2/\min\{\varepsilon^2, 1\}$, the above inequality holds unconditionally. $\square$

We similarly require a quantitative estimate for the usual L^1 strong law of large numbers on $\mathbb{R}$, which takes the form of the following distribution-dependent non-asymptotic concentration inequality established by Kachurovskii [56]:

Theorem 3.3 (essentially Theorem 26 in [56]). *Let $(\Omega, \mathbb{F}, \mathbb{P})$ be a probability space and let (X_i) be an i.i.d. sequence of random variables $X_i \in L^1(\Omega, \mathbb{R}, \mathbb{P})$. Then, for any $\varepsilon > 0$ and $m \in \mathbb{N}$:*

$$\mathbb{P}\left(\sup_{n \geq m} \left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}[X_1] \right| > \varepsilon\right) \leq \frac{2}{\varepsilon} \left(\frac{\sqrt{\mathbb{E}[|X_1|]}}{m^{1/2}} + U_{\mathbb{P}}(m^{1/2}) \right),$$

where $U_{\mathbb{P}}(x) := \mathbb{E}[|X_1| \mathbf{1}_{|X_1| > x}]$ is the truncated first absolute moment of X_1 .

Note that the right-hand side of the inequality from Theorem 3.3 vanishes as $m \rightarrow \infty$ since $U_{\mathbb{P}}(x) \rightarrow 0$ as $x \rightarrow \infty$ by the dominated convergence theorem.

Combined, we can now derive the following quantitative estimate on the asymptotic behavior of L^1 i.i.d. processes on $\mathcal{H}$:

Theorem 3.4. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(\mathcal{H}, d)$ be a separable Hadamard space. If (X_i) is an i.i.d. sequence of random variables $X_i \in L^1(\Omega, \mathcal{H}, \mathbb{P})$, then for any $\varepsilon > 0$ and any $m \in \mathbb{N}$:*

$$\mathbb{P} \left(\sup_{n \geq m} d \left(\frac{1}{n} \overrightarrow{\sum_{i=1, \dots, n}} X_i, \mathbb{E}[X_1] \right) > \varepsilon \right) \leq \frac{6}{\varepsilon} \left(\frac{\sqrt{\mathbb{E}[d(o, X_1)]}}{m^{1/2}} + U_{\mathbb{P}}(m^{1/8}) \right) + \frac{72}{\min\{\varepsilon^2, 1\} m^{1/4}},$$

where $U_{\mathbb{P}}(x) := \mathbb{E}[d(o, X_1) \mathbf{1}_{d(o, X_1) \geq x}]$ for a fixed $o \in \mathcal{H}$.

Also here, the right-hand side of the inequality featuring in Theorem 3.4 vanishes as $m \rightarrow \infty$.

Proof of Theorem 3.4. Let $b \geq 1$ be given. We associate with (X_i) a bounded (relative to b and o) and i.i.d. process (X_i^b) by truncating it along geodesics as follows:

$$X_i^b := \begin{cases} X_i & \text{if } d(o, X_i) \leq b, \\ \frac{b}{d(o, X_i)} X_i \oplus \frac{d(o, X_i) - b}{d(o, X_i)} o & \text{if } d(o, X_i) > b. \end{cases}$$

For brevity, write

$$S_n := \frac{1}{n} \overrightarrow{\sum_{i=1, \dots, n}} X_i \text{ and } S_n^b := \frac{1}{n} \overrightarrow{\sum_{i=1, \dots, n}} X_i^b.$$

Note that $d(o, X_i^b) = \min\{d(o, X_i), b\}$ so that (X_i^b) is i.i.d. as well as $X_i^b \in L^\infty(\Omega, \mathcal{H}, \mathbb{P})$. Our quantitative version of Sturm's L^∞ strong law of large numbers, that is Lemma 3.2, then yields

$$\mathbb{P} \left(\sup_{n \geq m} d(S_n^b, \mathbb{E}[X_1^b]) > \varepsilon \right) \leq \frac{8b^2}{\min\{\varepsilon^2, 1\} m^{1/2}}$$

for any $\varepsilon > 0$ and $m \in \mathbb{N}$. Now note that since $\mathcal{H}$ is a Hadamard space, it is a Busemann space and so we have that d is jointly convex. In particular, we inductively get

$$d \left(\frac{1}{n} \overrightarrow{\sum_{i=1, \dots, n}} x_i, \frac{1}{n} \overrightarrow{\sum_{i=1, \dots, n}} y_i \right) \leq \frac{1}{n} \sum_{i=1}^n d(x_i, y_i)$$

for any pair of sequences $(x_i), (y_i) \subseteq \mathcal{H}$. This yields

$$d(S_n, S_n^b) \leq \frac{1}{n} \sum_{i=1}^n d(X_i, X_i^b).$$

Since we have $d(X_i, X_i^b) = (d(o, X_i) - b)^+$, the real-valued process $(d(X_i, X_i^b))$ is i.i.d. and we have $d(X_i, X_i^b) \in L^1(\Omega, \mathbb{R}, \mathbb{P})$ since $X_i \in L^1(\Omega, \mathcal{H}, \mathbb{P})$. Appealing to the above quantitative variant of the real-valued L^1 strong law of large numbers, that is Theorem 3.3, now yields

$$\mathbb{P} \left(\sup_{n \geq m} \left| \frac{1}{n} \sum_{i=1}^n d(X_i, X_i^b) - \mathbb{E}[(d(o, X_1) - b)^+] \right| > \varepsilon \right) \leq \frac{2}{\varepsilon} \left(\frac{\sqrt{\mathbb{E}[d(o, X_1)]}}{m^{1/2}} + V_{\mathbb{P}}^b(m^{1/2}) \right)$$

for any $\varepsilon > 0$ and $m \in \mathbb{N}$, where $V_{\mathbb{P}}^b(x) := \mathbb{E}[(d(o, X_1) - b)^+ \mathbf{1}_{(d(o, X_1) - b)^+ > x}]$. Using the above inequality, we hence get

$$\begin{aligned} d(S_n, S_n^b) &= d(S_n, S_n^b) - \mathbb{E}[(d(o, X_1) - b)^+] + \mathbb{E}[(d(o, X_1) - b)^+] \\ &\leq \left| \frac{1}{n} \sum_{i=1}^n d(X_i, X_i^b) - \mathbb{E}[(d(o, X_1) - b)^+] \right| + \mathbb{E}[(d(o, X_1) - b)^+]. \end{aligned}$$

Lemma 2.7 now yields that

$$d(\mathbb{E}[X_1], \mathbb{E}[X_1^b]) \leq \mathbb{E}[d(X_1, X_1^b)] = \mathbb{E}[(d(o, X_1) - b)^+].$$

The triangle inequality yields

$$\begin{aligned} d(S_n, \mathbb{E}[X_1]) &\leq d(S_n, S_n^b) + d(S_n^b, \mathbb{E}[X_1^b]) + d(\mathbb{E}[X_1^b], \mathbb{E}[X_1]) \\ &\leq \left| \frac{1}{n} \sum_{i=1}^n d(X_i, X_i^b) - \mathbb{E}[(d(o, X_1) - b)^+] \right| + d(S_n^b, \mathbb{E}[X_1^b]) + 2\mathbb{E}[(d(o, X_1) - b)^+]. \end{aligned}$$

Now, note that $(d(o, X_1) - b)^+ = (d(o, X_1) - b)\mathbf{1}_{d(o, X_1) \geq b}$ and so

$$\mathbb{E}[(d(o, X_1) - b)^+] = U_{\mathbb{P}}(b) - b\mathbb{P}(d(o, X_1) \geq b) \leq U_{\mathbb{P}}(b).$$

In particular, we have $V_{\mathbb{P}}^b(x) \leq U_{\mathbb{P}}(b)$. Hence $2\mathbb{E}[(d(o, X_1) - b)^+] \leq 2U_{\mathbb{P}}(b)$. If $U_{\mathbb{P}}(b) \leq \varepsilon/6$ for $b = m^{1/8}$, then we have

$$\begin{aligned} \mathbb{P}\left(\sup_{n \geq m} d(S_n, \mathbb{E}[X_1]) > \varepsilon\right) &\leq \mathbb{P}\left(\sup_{n \geq m} \left| \frac{1}{n} \sum_{i=1}^n d(X_i, X_i^b) - \mathbb{E}[(d(o, X_1) - b)^+] \right| > \varepsilon/3\right) \\ &\quad + \mathbb{P}\left(\sup_{n \geq m} d(S_n^b, \mathbb{E}[X_1^b]) > \varepsilon/3\right) \\ &\leq \frac{6}{\varepsilon} \left(\frac{\sqrt{\mathbb{E}[d(o, X_1)]}}{m^{1/2}} + V_{\mathbb{P}}^b(m^{1/2}) \right) + \frac{72b^2}{\min\{\varepsilon^2, 1\}m^{1/2}} \\ &\leq \frac{6}{\varepsilon} \left(\frac{\sqrt{\mathbb{E}[d(o, X_1)]}}{m^{1/2}} + U_{\mathbb{P}}(m^{1/8}) \right) + \frac{72}{\min\{\varepsilon^2, 1\}m^{1/4}}. \end{aligned}$$

If $\varepsilon/6 < U_{\mathbb{P}}(m^{1/8})$, then $\frac{6}{\varepsilon}U_{\mathbb{P}}(m^{1/8}) > 1$ so that the result holds unconditionally. $\square$

In particular, the above Theorem 3.4 entails the purely qualitative strong law of large numbers for L^1 i.i.d. processes first established by Yokota [123]:

Corollary 3.5. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(\mathcal{H}, d)$ be a separable Hadamard space. If (X_i) is an i.i.d. sequence of random variables $X_i \in L^1(\Omega, \mathcal{H}, \mathbb{P})$, then

$$\frac{1}{n} \sum_{i=1, \dots, n}^{\rightarrow} X_i \rightarrow \mathbb{E}[X_1] \quad \mathbb{P}\text{-a.s.}$$

as $n \rightarrow \infty$.

However, akin to [106], also here the concentration inequality established in Theorem 3.4 is strong enough to immediately entail a distribution-uniform generalization of the strong law of large numbers in the style of Chung [39] on Hadamard spaces:

Corollary 3.6. Let $(\Omega, \mathcal{F})$ be a measurable space and let $(\mathcal{H}, d)$ be a separable Hadamard space. Let $\mathcal{P}$ be a collection of probability measures on $(\Omega, \mathcal{F})$ and (X_i) be a sequence of random variables which is i.i.d. for any $\mathbb{P} \in \mathcal{P}$. If X_1 is $\mathcal{P}$ -uniformly integrable, i.e.

$$\lim_{x \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[d(o, X_1)\mathbf{1}_{d(o, X_1) \geq x}] = 0$$

for a fixed $o \in \mathcal{H}$, then

$$\lim_{m \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(\sup_{n \geq m} d\left(\frac{1}{n} \sum_{i=1, \dots, n}^{\rightarrow} X_i, \mathbb{E}_{\mathbb{P}}[X_1]\right) > \varepsilon\right) = 0.$$

Proof. The corollary follows immediately from Theorem 3.4 once one notes that the $\mathcal{P}$ -uniform integrability implies that $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[d(o, X_1)] < +\infty$. To see this, use the $\mathcal{P}$ -uniform integrability to pick an x such that $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[d(o, X_1) \mathbf{1}_{d(o, X_1) \geq x}] \leq 1$. Then, for any $\mathbb{P} \in \mathcal{P}$:

$$\mathbb{E}_{\mathbb{P}}[d(o, X_1)] \leq \mathbb{E}_{\mathbb{P}}[d(o, X_1) \mathbf{1}_{d(o, X_1) \leq x}] + \mathbb{E}_{\mathbb{P}}[d(o, X_1) \mathbf{1}_{d(o, X_1) \geq x}] \leq x + 1$$

which yields $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[d(o, X_1)] \leq x + 1 < +\infty$. $\square$

The convergence result of Corollary 3.5 does not extend to the stochastic proximal point algorithm for strongly convex functions, one of the generalizations of the strong law of large numbers: While it is well-known that the stochastic proximal point algorithm for such functions on Hadamard spaces converges under L^∞ - or L^2 -Lipschitz assumptions (see [93] or [98], respectively), which in particular is used in [93] to re-derive Sturm's L^∞ strong law of large numbers, the following example shows that this does not extend to an L^1 -Lipschitz assumption on the function.

Example 3.7. We construct a probability space $(E, \mathbb{E}, \mu)$ and a proper Hadamard space $(\mathcal{H}, d)$ together with a function $f : E \times \mathcal{H} \rightarrow \mathbb{R}$ such that f is $\mathbb{E} \otimes \mathcal{B}(\mathcal{H})$ -measurable and $f(e, \cdot)$ is strongly convex and $L(e)$ -Lipschitz with $L \in L^1(E, (0, \infty), \mu)$, but where the stochastic proximal point iteration

$$x_{n+1} := \text{prox}_{\lambda_n f}(\xi_n, x_n) := \operatorname{argmin}_{y \in \mathcal{H}} \left\{ f(\xi_n, y) + \frac{1}{2\lambda_n} d^2(x_n, y) \right\}$$

does not converge almost surely, for a given sequence of parameters $\lambda_n \in (0, \infty)$ such that $\sum_{n \in \mathbb{N}} \lambda_n = +\infty$ and $\sum_{n \in \mathbb{N}} \lambda_n^2 < +\infty$ as well as an i.i.d. sequence (ξ_n) of random variables $\xi_n : \Omega \rightarrow E$ with distribution μ .

For that, consider the Borel space over $E := \mathbb{R}$ with a measure μ induced by the push-forward of a symmetric and integrable random variable $\xi : \Omega \rightarrow E$ with symmetric Pareto tails (see e.g. [8]), that is $\mathbb{E}[\xi] = 0$, $\mathbb{E}[|\xi|] < +\infty$ as well as

$$\mathbb{P}(\xi > t) = \frac{1}{2}t^{-r} \text{ and } \mathbb{P}(\xi < -t) = \frac{1}{2}t^{-r}$$

for all $t \geq 1$, where $r > 1$. On that space and for $\mathcal{H} := [-1, 1]$, consider $f : E \times \mathcal{H} \rightarrow \mathbb{R}$ defined by $f(e, x) := \frac{1}{2}x^2 - ex$. Then $f(e, \cdot)$ is strongly convex and

$$|f(e, x) - f(e, y)| \leq (1 + |e|)|x - y|$$

for all $x, y \in [-1, 1]$ and $e \in \mathbb{R}$, so that $f(e, \cdot)$ is Lipschitz with constant $L(e) := 1 + |e| \in L^1(E, (0, \infty), \mu)$. Further, $\underline{f}(x) := \int f(e, x) d\mu(e) = \frac{1}{2}x^2$, which has the unique minimizer 0 on $[-1, 1]$. It is easy to see that

$$\text{prox}_{\lambda f(e, \cdot)}(x) = P_{[-1, 1]} \left(\frac{x + \lambda e}{1 + \lambda} \right),$$

where $P_{[-1, 1]}$ is the projection onto the interval $[-1, 1]$. Now, for $\lambda_n = 1/n^a$ with $a \in (\frac{1}{2}, 1)$ and where $r \in (1, \frac{1}{a}]$, we have $\sum_{n \in \mathbb{N}} \lambda_n = +\infty$ and $\sum_{n \in \mathbb{N}} \lambda_n^2 < +\infty$. Let now (ξ_n) be i.i.d. with distribution μ . We then have

$$\sum_{n \in \mathbb{N}} \mathbb{P} \left(\xi_n > \frac{3}{\lambda_n} \right) = \sum_{n \in \mathbb{N}} \frac{1}{2 \cdot 3^r} \lambda_n^r = +\infty,$$

and so the Borel-Cantelli lemma yields that $\xi_n > \frac{3}{\lambda_n}$ holds infinitely often almost surely. Now if $\xi_n > \frac{3}{\lambda_n}$, we have $\frac{x + \lambda e}{1 + \lambda} > 1$ and so $x_{n+1} = 1$, which hence holds infinitely often almost surely.

Similarly, we get $\sum_{n \in \mathbb{N}} \mathbb{P} \left(\xi_n < -\frac{3}{\lambda_n} \right) = +\infty$ and so $x_{n+1} = -1$ holds infinitely often almost surely. Hence x_{n+1} cannot converge almost surely.

As already highlighted in the introduction, it still remains open whether an L^1 version of the full Birkhoff pointwise ergodic theorem holds for inductive means. Concretely, while various ergodic theorems are known on Hadamard spaces (see e.g. [7, 16, 38, 82], as well as [65]), similar as with the strong law previously, all of these results deviate in some way from a full formulation of Birkhoff's pointwise ergodic theorem on Hadamard spaces using inductive means, with [7, 38], while using inductive means, not being phrased for general dynamical systems but only Kronecker systems (which in particular exclude the Bernoulli shift otherwise used to derive the strong law of large numbers from the ergodic theorem) and with [16, 82] using other notions of averages based on (canonical) barycenters. We thus want to highlight the following question:

Question 3.8. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(\mathcal{H}, d)$ be a separable Hadamard space. Suppose $T : \Omega \rightarrow \Omega$ is a measurable and measure-preserving function and that $f \in L^1(\Omega, \mathcal{H}, \mathbb{P})$. Does it then hold that the inductive ergodic averages

$$\frac{1}{n} \sum_{i=1, \dots, n}^{\rightarrow} f(T^{(i)}(\omega)),$$

converge to $\mathbb{E}[f \mid \mathcal{I}]$ almost surely? Here, $T^{(i)}(\omega)$ denotes i -fold application of T to ω and $\mathcal{I} = \{A \in \mathcal{F} \mid T^{-1}(A) = A\}$ is the σ -algebra of invariant events.

The above is in particular known in Hilbert spaces (see e.g. [66]). A result such as in Question 3.8 would then in particular imply the strong law of large numbers from Corollary 3.5. In particular, the strategy employed above to deduce the L^1 strong law of large numbers is also viable in this context, that is using a similar bounding technique along geodesics, one can derive an L^1 variant of the pointwise ergodic theorem already from the L^∞ version of the pointwise ergodic theorem in Hadamard spaces together with the usual Birkhoff pointwise ergodic theorem on $\mathbb{R}$.

Another open problem already highlighted in the introduction is the generalization of Corollary 3.5 to weaker notions than i.i.d. random variables, such as e.g. pairwise i.i.d. random variables.

Question 3.9. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(\mathcal{H}, d)$ be a separable Hadamard space. If (X_i) is a sequence of pairwise independent and identically distributed random variables $X_i \in L^1(\Omega, \mathcal{H}, \mathbb{P})$, does it then hold that

$$\frac{1}{n} \sum_{i=1, \dots, n}^{\rightarrow} X_i \rightarrow \mathbb{E}[X_1] \quad \mathbb{P}\text{-a.s.}$$

as $n \rightarrow \infty$?

This problem is in particular open for $X_i \in L^\infty(\Omega, \mathcal{H}, \mathbb{P})$, which suffices to establish the claim for $X_i \in L^1(\Omega, \mathcal{H}, \mathbb{P})$, as also here the strategy employed above to deduce the L^1 strong law of large numbers, that is using a similar bounding technique along geodesics, is viable to derive an L^1 variant from an L^∞ version together with the L^1 strong law of large numbers for pairwise i.i.d. processes (as established by Etemadi [42], see also e.g. Theorem 5.17 in [61]). In particular, if one could establish a similar concentration inequality for an L^∞ strong law for pairwise i.i.d. processes over Hadamard spaces like that obtained in Lemma 3.2, then this result can be combined in a similar way as in the proof of Theorem 3.4 with the respective

concentration inequality for the real-valued L^1 strong law of large numbers for pairwise i.i.d. processes obtained in [90] (see Corollary 3 therein).

Next to the previous two questions, various other surrounding questions remain interesting points for future study where the above approach could be of help, like e.g. whether the strong law of large numbers extends to different classes of spaces such as e.g. Gromov's hyperbolic spaces (see the recent [92]), or whether there is an analogous result on Hadamard spaces to the strong law of large numbers by Marcinkiewicz and Zygmund [77], or even whether the associated non-asymptotic concentration inequality recently established in [106] (see Theorem 2.3 therein) can be (appropriately) extended to a more geometric case. Such a result could then perhaps be used to establish, similar to [106], related Baum-Katz results [17] (see also [85]) on Hadamard spaces.

4. RANDOM MONOTONE VECTOR FIELDS ON HADAMARD SPACES

4.1. Monotone vector fields on Hadamard spaces. Following [36], a monotone vector field over a Hadamard space $(\mathcal{H}, d)$ is a mapping $A : \mathcal{H} \rightarrow 2^{T\mathcal{H}}$ such that $A(x) \subseteq T_x\mathcal{H}$ and

$$g_x(u, \log_x y) \leq -g_y(v, \log_y x)$$

for all $(x, u), (y, v) \in A$. We denote the set of zeros of A by $\text{zer}A := \{x \in \mathcal{H} \mid 0_x \in A(x)\}$. As already discussed in the introduction, this notion simultaneously generalizes monotone operators on Hilbert spaces and monotone vector fields on Hadamard manifolds as introduced in [83, 89] (see also [72, 73, 121]).

The key derived object for a monotone vector field is its resolvent, which we defined (following [36], which in turn generalizes [73]) via

$$J_\lambda^A x := \{z \in \mathcal{H} \mid \tfrac{1}{\lambda} \log_z x \in A(z)\}.$$

Lemma 2.5 immediately yields that a $z \in \mathcal{H}$ such that $\tfrac{1}{\lambda} \log_z x \in A(z)$ is unique whenever A is monotone. In that case, we identify J_λ^A with the corresponding (potentially partial) function from $\mathcal{H}$ to $\mathcal{H}$. Following [36], we say that A satisfies the surjectivity condition if all resolvents are total functions. In the following, we write $\mathcal{M}(\mathcal{H})$ for the set of all monotone vector fields on $\mathcal{H}$ satisfying the surjectivity condition (adapting the notation used in [108]).

Remark 4.1. A monotone vector field A is called maximal if its graph $\text{gra}A := \{(x, u) \in \mathcal{H} \times T\mathcal{H} \mid u \in Ax\}$ cannot be extended properly while preserving monotonicity. By the well-known theorem of Minty [80], the maximality of a monotone operator over a Hilbert space is equivalent to the totality of the resolvent. One direction remains valid, as shown in Proposition 3.5 in [36]: if A satisfies the surjectivity condition, then it is maximal. The converse extends to the setting of Hadamard manifolds as shown in [72] under the condition that the domain $\text{dom}A := \{x \in \mathcal{H} \mid A(x) \neq \emptyset\}$ is the whole space (see Remark 4.4 in [72]).

Question 4.2. Let A be a maximally monotone vector field on a Hadamard space such that $\text{dom}A = \mathcal{H}$. Does A then satisfy the surjectivity condition? Further, can the assumption $\text{dom}A = \mathcal{H}$ be omitted? This latter problem in particular seems to be open already for Hadamard manifolds.

Example 4.3 (cf. Examples 2.7 and 2.11 in [97]). Let $f : \mathcal{H} \rightarrow (-\infty, +\infty]$ be a proper, convex and lower-semicontinuous (lsc) function. Then the subdifferential

$$\partial f(x) := \{u \in T_x\mathcal{H} \mid f(y) \geq f(x) + g_x(u, \log_x y) \text{ for all } y \in \mathcal{H}\},$$

as defined in the general setting of Hadamard spaces in [36], is a monotone vector field with $\text{argmin}f := \{x^* \in \mathcal{H} \mid f(x^*) \leq f(x) \text{ for all } x \in \mathcal{H}\} = \text{zer}\partial f$. As shown in [36, Proposition 3.8],

the resolvent of ∂f is given by

$$\text{prox}_{\lambda f}(x) := \operatorname{argmin}_{y \in \mathcal{H}} \left\{ f(y) + \frac{1}{2\lambda} d^2(x, y) \right\},$$

that is the proximal map of f (also called the Moreau-Yosida resolvent). Each $\text{prox}_{\lambda f}$ is total (see [54, Lemma 2]), so that ∂f satisfies the surjectivity condition (and hence is also maximal).

In the rest of this paper, we actually require very little properties of the resolvent. The first is the following lemma:

Lemma 4.4 (Proposition 4.3 in [36]). *If A is a monotone vector field, then J_λ^A is nonexpansive, that is*

$$d(J_\lambda^A(x), J_\lambda^A(y)) \leq d(x, y)$$

for all $x, y \in \operatorname{dom}(J_\lambda^A)$. Further, it holds that $\operatorname{Fix}(J_\lambda^A) = \operatorname{zer}A$ and if A also satisfies the surjectivity condition, then

$$J_\lambda^A(x) = J_\gamma^A \left(\left(1 - \frac{\gamma}{\lambda}\right) J_\lambda^A(x) \oplus \frac{\gamma}{\lambda} x \right)$$

for any $x \in \mathcal{H}$ and $0 < \gamma \leq \lambda$.

The only other property required later is the continuity of the resolvent in its real parameter. This was established as Proposition 4.12 in [36] under the condition that $\mathcal{H}$ satisfies the geodesic extension property, but we here give the following lemma which disposes of that assumption:

Lemma 4.5. *Let A be a monotone vector field that satisfies the surjectivity condition. Then for any $x \in \mathcal{H}$ and any $0 < \gamma \leq \lambda$:*

$$d(J_\lambda^A(x), J_\gamma^A(x)) \leq \left(1 - \frac{\gamma}{\lambda}\right) d(J_\lambda^A(x), x).$$

In particular, $J_\lambda^A(x)$ is continuous in $\lambda \in (0, \infty)$.

Proof. Let $x \in \mathcal{H}$ and $0 < \gamma \leq \lambda$. Using Lemma 4.4, we get

$$\begin{aligned} d(J_\lambda^A(x), J_\gamma^A(x)) &= d \left(J_\gamma^A \left(\left(1 - \frac{\gamma}{\lambda}\right) J_\lambda^A(x) \oplus \frac{\gamma}{\lambda} x \right), J_\gamma^A(x) \right) \\ &\leq d \left(\left(1 - \frac{\gamma}{\lambda}\right) J_\lambda^A(x) \oplus \frac{\gamma}{\lambda} x, x \right) \\ &\leq \left(1 - \frac{\gamma}{\lambda}\right) d(J_\lambda^A(x), x). \end{aligned}$$

To see that $J_\lambda^A(x)$ is continuous in λ , note now first that the above yields

$$\begin{aligned} d(J_\gamma^A(x), x) &\leq d(J_\gamma^A(x), J_\lambda^A(x)) + d(J_\lambda^A(x), x) \\ &\leq \left(1 - \frac{\gamma}{\lambda}\right) d(J_\lambda^A(x), x) + d(J_\lambda^A(x), x) \\ &= \left(2 - \frac{\gamma}{\lambda}\right) d(J_\lambda^A(x), x) \leq 2d(J_\lambda^A(x), x). \end{aligned}$$

If now $\gamma_n \rightarrow \lambda$ from below, then the first inequality yields

$$d(J_\lambda^A(x), J_{\gamma_n}^A(x)) \leq \left(1 - \frac{\gamma_n}{\lambda}\right) d(J_\lambda^A(x), x) \rightarrow 0$$

directly. If $\gamma_n \rightarrow \lambda$ from above, then w.l.o.g. $\gamma_n \in [\lambda, \Gamma]$ for some $\Gamma > 0$. Then, the first together with second inequality imply

$$d(J_\lambda^A(x), J_{\gamma_n}^A(x)) \leq \left(1 - \frac{\lambda}{\gamma_n}\right) d(J_{\gamma_n}^A(x), x) \leq 2 \left(1 - \frac{\lambda}{\gamma_n}\right) d(J_\Gamma^A(x), x) \rightarrow 0.$$

Combined, $J_\lambda^A(x)$ is continuous in λ . □

4.2. Random monotone vector fields and their means. Fix a probability space $(T, \mathbb{T}, \tau)$ and let $(\mathcal{H}, d)$ be a separable Hadamard space. Over these spaces, consider a map $A : T \rightarrow \mathcal{M}(\mathcal{H})$.⁴ We write $A(t, x)$ for $A(t)(x)$.

For such a perturbed vector field, we define its resolvent similar to before via

$$J_\lambda^{A(t)}(x) := \{z \in \mathcal{H} \mid \tfrac{1}{\lambda} \log_z x \in A(t, z)\}$$

for $\lambda > 0$, $t \in T$ and $x \in \mathcal{H}$. The following notion of a random monotone vector field, characterized by a measurability condition on these perturbed resolvents, was recently introduced in [97], extending the notion used in Hilbert spaces (see [27] as well as [108], among others):

Definition 4.6. An operator $A : T \rightarrow \mathcal{M}(\mathcal{H})$ is called a random monotone vector field if $J_\lambda^{A(\cdot)}(x)$ is $\mathbb{T}/\mathbb{B}(\mathcal{H})$ -measurable for any $x \in \mathcal{H}$ and any $\lambda > 0$.

As $J_\lambda^{A(t)}(\cdot)$ is single-valued, total and uniformly continuous (being nonexpansive, as discussed before) for any $t \in T$, the map $J_\lambda^{A(\cdot)}(\cdot)$ is therefore a Carathéodory map and, in particular, is $\mathbb{T} \otimes \mathbb{B}(\mathcal{H})/\mathbb{B}(\mathcal{H})$ -measurable (see e.g. Lemma 8.2.6 in [14]).

Example 4.7. Let $f : T \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a given function. Following [104], we call f a normal integrand if $f(t, \cdot)$ is lsc for all $t \in T$ and its epigraph $\text{epi} f(t, \cdot) := \{(x, \alpha) \in \mathcal{H} \times \mathbb{R} \mid f(t, x) \leq \alpha\}$ is measurable as a set-valued function $T \rightarrow 2^{\mathcal{H} \times \mathbb{R}}$. We call f a normal convex integrand if $f(t, \cdot)$ is additionally convex for all $t \in T$ and we call f proper if $f(t, \cdot)$ is proper for any $t \in T$. If $(T, \mathbb{T}, \tau)$ is complete, the measurability of the epigraph condition can equivalently be replaced with the assumption that f is $\mathbb{T} \otimes \mathbb{B}(\mathcal{H})$ -measurable.

For such a proper normal convex integrand f , consider now the associated random subdifferential

$$\partial f(t, x) := \{u \in T_x \mathcal{H} \mid f(t, y) \geq f(t, x) + g_x(u, \log_x y) \text{ for all } y \in \mathcal{H}\}$$

for $t \in T$ and $x \in \mathcal{H}$. As discussed in Example 4.3, the resolvents of $\partial f(t, \cdot)$ are given by the proximal maps $\text{prox}_{\lambda f(t, \cdot)}(x)$, which are all total. As discussed in Example 3.2 in [97], by adapting arguments from [104], it is further easy to see that $\text{prox}_{\lambda f}(t, \cdot)(x)$ is $\mathbb{T}/\mathbb{B}(\mathcal{H})$ -measurable in t , so that $\partial f : T \rightarrow \mathcal{M}(\mathcal{H})$ is a random monotone vector field.

The notion of a random monotone vector field A as defined above can be characterized by a genuine measurability condition on A defined using the so-called R -topology, extending results of Attouch [12]. This topology later gives rise to the notion of R -convergence which is the key notion of convergence for our strong law of large numbers for random monotone vector fields.

Definition 4.8. The R -topology τ_R on $\mathcal{M}(\mathcal{H})$ is the coarsest topology that makes all the functions $p_{x, \lambda}(A) := J_\lambda^A(x)$, given $x \in \mathcal{H}$ and $\lambda > 0$, continuous.

Analogously to Proposition 1.1 in [12], we also get here that this topology is metrizable and that the resulting space is separable. Before we derive this result, we however first give the following lemma, illustrating that the resolvents completely determine a monotone vector field with the surjectivity condition.

Lemma 4.9. Let $A, B \in \mathcal{M}(\mathcal{H})$ such that $J_\lambda^A(x) = J_\lambda^B(x)$ for all $\lambda > 0$ and $x \in \mathcal{H}$. Then $A = B$.

⁴Equivalently, we can regard A as a perturbed set-valued vector field $A : T \times \mathcal{H} \rightarrow 2^{T\mathcal{H}}$ with $A(t, x) \subseteq T_x \mathcal{H}$ for all $t \in T$ and $x \in \mathcal{H}$ such that $A(t, \cdot)$ is monotone and satisfies the surjectivity condition for any $t \in T$ (as studied in [97]).

Proof. We first show

$$A = \{(J_\lambda^A(x), \lambda^{-1} \log_{J_\lambda^A(x)} x) \mid x \in \mathcal{H}, \lambda > 0\}.$$

Concretely, call the right-hand side C . Then $C \subseteq A$ and hence C is a monotone vector field. Now, by definition we have $\lambda^{-1} \log_{J_\lambda^A(x)} x \in C(J_\lambda^A(x))$, and so by uniqueness of the resolvent, we have $J_\lambda^C(x) = J_\lambda^A(x)$ for any $x \in \mathcal{H}$ and $\lambda > 0$. In particular, C satisfies the surjectivity condition and hence is maximal. Therefore, we have $A = C$. Now, we similarly get the above representation for B , and so we immediately get

$$\begin{aligned} A &= \{(J_\lambda^A(x), \lambda^{-1} \log_{J_\lambda^A(x)} x) \mid x \in \mathcal{H}, \lambda > 0\} \\ &= \{(J_\lambda^B(x), \lambda^{-1} \log_{J_\lambda^B(x)} x) \mid x \in \mathcal{H}, \lambda > 0\} = B \end{aligned}$$

if $J_\lambda^A(x) = J_\lambda^B(x)$ for all $\lambda > 0$ and $x \in \mathcal{H}$, as claimed. $\square$

We can now characterize the R -topology further, using a similar approach as Attouch [12].

Lemma 4.10. *The R -topology on $\mathcal{M}(\mathcal{H})$ is metrizable with metric*

$$d_R(A, B) := \sum_{n=1}^{\infty} \frac{1}{2^n} \min\{1, d(J_{r_n}^A(x_n), J_{r_n}^B(x_n))\},$$

where $\{(x_n, r_n) \mid n \in \mathbb{N}\}$ is dense in $\mathcal{H} \times (0, \infty)$. Further, the space $(\mathcal{M}(\mathcal{H}), d_R)$ is separable.

Proof. First, note that τ_R is the topology induced by d_R . For that, note that the topology induced by d_R is the coarsest topology that makes all the functions $J_{r_n}^A(x_n)$ continuous. It hence only remains to be shown that $J_\lambda^A(x)$ is continuous in A under d_R , given $\lambda > 0$ and $x \in \mathcal{H}$. For this, suppose $d_R(A_n, A) \rightarrow 0$. Then by definition $d(J_{r_k}^{A_n}(x_k), J_{r_k}^A(x_k)) \rightarrow 0$ for any $k \in \mathbb{N}$. Take a $k \in \mathbb{N}$ such that $d(x_k, x) < \frac{\varepsilon}{3}$. Then, for large enough n , we get $d(J_{r_k}^A(x_k), J_{r_k}^{A_n}(x_k)) < \frac{\varepsilon}{3}$ for this k , and hence obtain

$$d(J_{r_k}^A(x), J_{r_k}^{A_n}(x)) \leq d(J_{r_k}^A(x), J_{r_k}^A(x_k)) + d(J_{r_k}^A(x_k), J_{r_k}^{A_n}(x_k)) + d(J_{r_k}^{A_n}(x_k), J_{r_k}^{A_n}(x)) < \varepsilon$$

for all such suitably large n . Now, using Lemma 4.5, take a k such that $r_j \geq r_k \geq \lambda$ is close enough to λ such that $d(J_\lambda^A(x), J_{r_k}^A(x)) < \frac{\varepsilon}{3}$ and

$$d(J_{r_k}^{A_n}(x), J_\lambda^{A_n}(x)) \leq 2 \left(1 - \frac{\lambda}{r_k}\right) d(J_{r_j}^{A_n}(x), x) < \frac{\varepsilon}{3},$$

using that $d(J_{r_j}^{A_n}(x), x)$ is bounded for varying n . Then, for large enough n , we get that $d(J_{r_k}^A(x), J_{r_k}^{A_n}(x)) < \frac{\varepsilon}{3}$ for this k , and hence obtain

$$d(J_\lambda^A(x), J_\lambda^{A_n}(x)) \leq d(J_\lambda^A(x), J_{r_k}^A(x)) + d(J_{r_k}^A(x), J_{r_k}^{A_n}(x)) + d(J_{r_k}^{A_n}(x), J_\lambda^{A_n}(x)) < \varepsilon.$$

To see that $(\mathcal{M}(\mathcal{H}), d_R)$ is separable, define the map $i : \mathcal{M}(\mathcal{H}) \rightarrow \mathcal{H}^\mathbb{N}$ via

$$i(A) := (J_{r_n}^A(x_n))_{n \in \mathbb{N}}.$$

The map i is now injective: Suppose that $A, B \in \mathcal{M}(\mathcal{H})$ are such that $J_{r_n}^A(x_n) = J_{r_n}^B(x_n)$ for all $n \in \mathbb{N}$. Then by Lemmas 4.4 and 4.5, we get that $J_\lambda^A(x) = J_\lambda^B(x)$ for all $\lambda > 0$ and $x \in \mathcal{H}$. Lemma 4.9 now yields that $A = B$. In particular, i is a bijection between $\mathcal{M}(\mathcal{H})$ and $i(\mathcal{M}(\mathcal{H}))$. Now, note that by definition of the R -topology, the map i is a homeomorphism between $\mathcal{M}(\mathcal{H})$ and $i(\mathcal{M}(\mathcal{H}))$. As $\mathcal{H}$ is separable, so is $\mathcal{H}^\mathbb{N}$ and hence $\mathcal{M}(\mathcal{H})$ is also separable. $\square$

Remark 4.11. Proposition 1.1 in [12] also establishes that $(\mathcal{M}(\mathcal{H}), d_R)$ is complete for a separable Hilbert space $\mathcal{H}$. Lifting this completeness to the present nonlinear setting is not completely trivial: While the approach from [12] carries over conceptually, it relies on the ability to associate a monotone vector field to a given nonexpansive map which is not fully available in this nonlinear context (see Proposition 4.7 in [36] for some partial results in the context of the geodesic extension property). We however do not need the completeness of $(\mathcal{M}(\mathcal{H}), d_R)$ here, and so simply omit results related to this.

The following result now partially extends Lemma 2.1 from [12]:

Lemma 4.12. *Let $A : T \rightarrow \mathcal{M}(\mathcal{H})$ be given. Then A is a random monotone vector field if, and only if, A is $\mathbb{T}/\mathbb{B}(\mathcal{M}(\mathcal{H}))$ -measurable.*

Proof. Suppose that A is $\mathbb{T}/\mathbb{B}(\mathcal{M}(\mathcal{H}))$ -measurable. Then $J_\lambda^{A(t)}(x) = p_{x,\lambda}(A(t))$ is $\mathbb{T}/\mathbb{B}(\mathcal{H})$ -measurable as $p_{x,\lambda}$ is continuous in $\mathcal{M}(\mathcal{H})$. Hence, A is a random monotone vector field.

Conversely, suppose that each $J_\lambda^{A(t)}(x)$ is $\mathbb{T}/\mathbb{B}(\mathcal{H})$ -measurable. Then, for an arbitrary but fixed $B \in \mathcal{M}(\mathcal{H})$, we have

$$d_R(A(t), B) := \sum_{n=1}^{\infty} \frac{1}{2^n} \min\{1, d(J_{r_n}^{A(t)}(x_n), J_{r_n}^B(x_n))\}.$$

As each summand is measurable, we get that also the series and hence $d_R(A(t), B)$ is $\mathbb{T}$ -measurable for any $B \in \mathcal{M}(\mathcal{H})$. As $(\mathcal{M}(\mathcal{H}), d_R)$ is separable by Lemma 4.10, we get that A is $\mathbb{T}/\mathbb{B}(\mathcal{M}(\mathcal{H}))$ -measurable (see e.g. Lemma III.16 in [35]). $\square$

Further, we can show that for a fixed $x \in \mathcal{H}$, the map $t \mapsto A(t, x)$ is measurable as a set-valued map $T \rightarrow 2^{T_x \mathcal{H}}$, which will be useful later. For that, we first recall the following two key notions from measurable selection theory (we generally refer to [14, 35] for further background): Over a complete separable metric space $(\mathcal{H}, d)$ and a measurable space $(T, \mathbb{T})$, a set-valued map $\varphi : T \rightarrow 2^{\mathcal{H}}$ is called graph measurable if

$$\text{gra}(\varphi) := \{(t, x) \in T \times \mathcal{H} \mid x \in \varphi(t)\} \in \mathbb{T} \otimes \mathbb{B}(\mathcal{H}).$$

Over complete σ -finite measure spaces, and if φ has nonempty closed images, this is equivalent (see e.g. Theorem 8.1.4 in [14]) to the so-called weak measurability of $\varphi : T \rightarrow 2^{\mathcal{H}}$, that is that

$$\varphi^{-1}(C) := \{t \in T \mid \varphi(t) \cap C \neq \emptyset\} \in \mathbb{T}$$

for all open sets $C \subseteq \mathcal{H}$, as well as to the so-called measurability of φ , that is that $\varphi^{-1}(C) \in \mathbb{T}$ for all closed sets $C \subseteq \mathcal{H}$.

Lemma 4.13. *Let $(T, \mathbb{T}, \tau)$ be a complete probability space. Let $A : T \rightarrow \mathcal{M}(\mathcal{H})$ be given and fix $x \in \mathcal{H}$. Then $t \mapsto A(t, x)$ is measurable as a set-valued map $T \rightarrow 2^{T_x \mathcal{H}}$.*

Proof. Similar as in the proof of Lemma 4.9, we have

$$A(t) = \{(J_\lambda^{A(t)}(y), \lambda^{-1} \log_{J_\lambda^{A(t)}(y)} y) \mid y \in \mathcal{H}, \lambda > 0\}.$$

Hence, we get

$$A(t, x) = \{\lambda^{-1} \log_{J_\lambda^{A(t)}(y)} y \mid y \in \mathcal{H}, \lambda > 0 \text{ such that } J_\lambda^{A(t)}(y) = x\}.$$

Hence, for $U \subseteq T_x \mathcal{H}$ closed, we have

$$\{t \in T \mid A(t, x) \cap U \neq \emptyset\} = \pi_T(E_U)$$

where π_T is the projection onto T and

$$E_U = \{(t, y, \lambda) \in T \times \mathcal{H} \times (0, \infty) \mid J_\lambda^{A(t)}(y) = x \text{ and } \lambda^{-1} \log_x y \in U\}.$$

Note that $J_\lambda^{A(t)}(y)$ is jointly continuous in $(y, \lambda) \in \mathcal{H} \times (0, \infty)$ since we have

$$\begin{aligned} d(J_{\lambda_n}^{A(t)}(y_n), J_\lambda^{A(t)}(y)) &\leq d(J_{\lambda_n}^{A(t)}(y_n), J_{\lambda_n}^{A(t)}(y)) + d(J_{\lambda_n}^{A(t)}(y), J_\lambda^{A(t)}(y)) \\ &\leq d(y_n, y) + d(J_{\lambda_n}^{A(t)}(y), J_\lambda^{A(t)}(y)) \rightarrow 0 \end{aligned}$$

as $(y_n, \lambda_n) \rightarrow (y, \lambda)$, using the nonexpansivity of the resolvent and Lemma 4.5. As $J_\lambda^{A(t)}(y)$ is also measurable in t , we get that $J_\lambda^{A(t)}(y)$ is a Carathéodory function and hence that it is jointly measurable in $(t, y, \lambda) \in T \times \mathcal{H} \times (0, \infty)$. Hence, by the inverse image theorem (see Theorem 8.2.9 in [14]), we have that

$$t \mapsto \{(y, \lambda) \in \mathcal{H} \times (0, \infty) \mid J_\lambda^{A(t)}(y) = x\}$$

is measurable and hence also graph-measurable. Similarly, we have that

$$t \mapsto \{(y, \lambda) \in \mathcal{H} \times (0, \infty) \mid \lambda^{-1} \log_x y \in U\}$$

is measurable and hence graph-measurable. Combined, we get that $E_U \in \mathbb{T} \otimes \mathbb{B}(\mathcal{H}) \otimes \mathbb{B}((0, \infty))$. Hence $\pi_T(E_U) \in \mathbb{T}$ by the measurable projection theorem (see Theorem III.23 in [35]). This implies that $A(\cdot, x)$ is measurable. $\square$

For a random monotone vector field $A : T \rightarrow \mathcal{M}(\mathcal{H})$, we define its integral operator via

$$\int A(t) d\tau : \mathcal{H} \rightarrow 2^{T\mathcal{H}}, \quad x \mapsto \int A(t, x) d\tau(t),$$

where the integral refers to the Aumann-Sturm integral as defined before, now on $T_x\mathcal{H}$. As already remarked in [97], if $\mathcal{H}$ is a separable Hilbert-Hadamard space and A is a random monotone vector field, then $\int A(t) d\tau(t)$ is immediately monotone, using that g_x is affine (recall Lemma 2.3).

In this paper, we now need a concept of a random monotone vector field being integrable. In [108], Salim called a random monotone operator on a Hilbert space integrable if $\int A(t) d\tau(t)$ is maximal (see also already [28, 29] for previous uses of this notion). This however is used therein exclusively to guarantee that the associated resolvents are total functions via Minty's theorem [80]. This equivalence is not available in this nonlinear context as discussed before, and we hence focus on the resolvents right away and arrive at the following condition.

Definition 4.14. A random monotone vector field $A : T \rightarrow \mathcal{M}(\mathcal{H})$ is called integrable if $\int A(t) d\tau(t)$ satisfies the surjectivity condition.

Examples of random monotone operators on Hilbert spaces which are integrable are discussed in [108] (see also [28]). Modelled on these discussions, we in the following discuss a few cases where similar results can be guaranteed in more general metric contexts.

Example 4.15. Let $(T, \mathbb{T}, \tau)$ be a complete probability space and $A : T \rightarrow \mathcal{M}(\mathcal{H})$ be a random monotone vector field.

- (1) Suppose that A is L^1 -dominated, that is there exists a $g \in L^1(T, (0, \infty), \tau)$ such that $A(t, x) \neq \emptyset$ and $\sup_{y \in A(t, x)} \|y\|_x \leq g(t)$ almost everywhere. As shown in Example 2 of [27], over $\mathbb{R}^d$, an L^1 -dominated random monotone operator is integrable. This extends to finite-dimensional Hadamard manifolds as follows: By Remark 4.4 of [72], it suffices to show that $\int A(t) d\tau(t)$ has full domain and is maximally monotone, and by Theorem 3.7 of [72], this reduces to $\int A(t) d\tau(t)$ being monotone, upper semicontinuous and having

full domain as well as having closed and convex values. Monotonicity is immediate as a Hadamard manifold is a Hilbert-Hadamard space. Further, fixing $x \in \mathcal{H}$, first recall from Lemma 4.13 that $A(\cdot, x)$ is measurable. Also, its values are closed (since A is maximal). We hence get that $\mathcal{S}^1(A(\cdot, x)) \neq \emptyset$ by the Kuratowski–Ryll–Nardzewski measurable selection theorem (see Theorem 8.1.3 in [14]), using that A is dominated to guarantee integrability. Hence, $\int A(t) d\tau(t)$ has full domain. Further, the values are convex and closed as can be immediately seen (see also [100]). As the tangent cones are just Euclidean spaces and as the values of $A(t, \cdot)$ are closed and bounded, they are compact. Hence, the results of [122] imply that $\int A(t) d\tau(t)$ is upper semicontinuous as each $A(\cdot, x)$ has measurable graph.

- (2) Suppose $A(t, x) = \partial f(t, x)$ is the subdifferential of a proper normal convex integrand $f : T \times \mathcal{H} \rightarrow (-\infty, +\infty]$, which is a random monotone vector field as discussed in Example 4.7. Assume that $\int f(t, \cdot) d\tau(t)$ is proper and lsc (it is immediately convex). If we can interchange the integral and the subdifferential, that is if

$$\int \partial f(t, x) d\tau(t) = \partial \left(\int f(t, \cdot) d\tau(t) \right) (x)$$

for all $x \in \mathcal{H}$, then $\int \partial f(t, \cdot) d\tau(t)$ satisfies the surjectivity condition as $\partial \left(\int f(t, \cdot) d\tau(t) \right)$ does (recall Example 4.3). Suitable conditions which guarantee this interchange of the integral and the subdifferential on Hilbert spaces can be found in [103]. We will later extend this result to separable Hilbert-Hadamard spaces where every tangent cone is isometric to a Hilbert space, covering in particular Hadamard manifolds, provided that $(T, \mathbb{T}, \tau)$ is additionally separable and f is L^2 -Lipschitz and suitably integrable.

If $(T, \mathbb{T}, \tau)$ is a finite discrete measure space with $T = \{t_1, \dots, t_n\}$ as well as $\tau(\{t_i\}) = w_i$, and $\mathcal{H}$ is such that each $T_x \mathcal{H}$ is isometric to a Hilbert space, then $\int A(t, x) d\tau(t)$ reduces to an arithmetic mean of the image sets $\sum_{i=1}^n w_i A(t_i, x)$, where the sum refers to the Minkowski sum. In particular, if $\mathcal{H}$ is a finite dimensional Hadamard manifold and all $A(t_i, \cdot)$ have full domain, $\int A(t) d\tau(t)$ satisfies the surjectivity condition as can be seen similarly to item (1) above: Since the sum also has full domain, by Remark 4.4 of [72] and Theorem 3.7 of [72], $\int A(t) d\tau(t)$ satisfying the surjectivity condition reduces to $\int A(t) d\tau(t)$ being monotone, upper semicontinuous and having closed and convex values which we can establish in this case. First, in the present context, $\int A(t) d\tau(t)$ is clearly monotone and its values are convex as the Minkowski sum preserves convexity. By Lemma 3.6 in [72], the A_i are locally bounded and so, as they are also closed, we get that $\underline{A}$ has closed values since the tangent cones of $\mathcal{H}$ are finite dimensional Hilbert spaces. Also, by the same argument as given in item (1) above, we get that $\underline{A}$ is upper semicontinuous.

5. A STRONG LAW OF LARGE NUMBERS FOR RANDOM MONOTONE VECTOR FIELDS

We now turn to the second main result of this paper, an extension of the previous L^1 strong law of large numbers for random variables on Hadamard spaces to integrable random monotone vector fields. For this, unless said otherwise, let $(\mathcal{H}, d)$ now be a separable Hilbert-Hadamard space (and recall Remark 2.9).

Again, we require a notion of finite mean, and for this we extend the previously discussed inductive mean of Sturm from points to sets.

Definition 5.1 (Inductive mean of sets). The inductive mean of $A_1, \dots, A_n \subseteq \mathcal{H}$ is defined as follows:

$$\frac{1}{n} \overrightarrow{\sum}_{i=1, \dots, n} A_i := \left\{ \frac{1}{n} \overrightarrow{\sum}_{i=1, \dots, n} x_i \mid x_i \in A_i \text{ for } i = 1, \dots, n \right\}.$$

If $A_i : \mathcal{H} \rightarrow 2^{T\mathcal{H}}$ are monotone vector fields, we define inductive mean vector field via

$$\frac{1}{n} \overrightarrow{\sum}_{i=1, \dots, n} A_i : \mathcal{H} \rightarrow 2^{T\mathcal{H}}, \quad x \mapsto \frac{1}{n} \overrightarrow{\sum}_{i=1, \dots, n} A_i(x).$$

In particular, this vector field is monotone in a separable Hilbert-Hadamard space $\mathcal{H}$, using that g_x is affine (recall Lemma 2.3) similar as with $\int A(t) d\tau$.

Before we can derive this second strong law of large numbers, we require an intermediate step. In a Hadamard manifold $\mathcal{H}$, using that $\nabla_x \frac{1}{2} d^2(x, y) = -\log_x y$ (as follows from the first variation formula, see [107]) and that $\frac{1}{2} d^2(x, y)$ is strongly convex, it can be shown that $-\log_x$ is strongly monotone, that is

$$g_z(\log_z x, \log_z y) \geq -g_y(\log_y x, \log_y z) + d^2(z, y)$$

for all $x, y, z \in \mathcal{H}$. Our proof of the strong law of large numbers for integrable random monotone vector fields relies on this property and while the conclusion of strong monotonicity as formulated above does make sense in the context of Hadamard spaces, there are two immediate issues with the above approach: First, a Hadamard space does not carry differential structure as is, and second, the expression $-\log_x$ is not well-defined as the space $T_x \mathcal{H}$ is not linear but just a cone.

We hence rely on a different approach here. Concretely, we follow the argument used in Proposition 2.9 from [97] to show that ∂f is strongly monotone whenever f is strongly convex, but utilize $-\frac{1}{2} d^2(x, y)$ in place of f .

Lemma 5.2. *Let $(\mathcal{H}, d)$ be a Hadamard space. For any $x, y, z \in \mathcal{H}$:*

$$g_z(\log_z x, \log_z y) \geq -g_y(\log_y x, \log_y z) + d^2(z, y)$$

Proof. For $f(a) := -\frac{1}{2} d^2(a, x)$, we get for any $t \in [0, 1]$ and any $a \in \mathcal{H}$ that

$$(1-t)f(a) + tf(b) + t(1-t)\frac{1}{2}d^2(a, b) \leq f((1-t)a \oplus tb).$$

Lemma 2.5 yields

$$\begin{aligned} 2g_a(\log_a x, \log_a y) &\geq d^2(a, x) + d^2(a, y) - d^2(x, y) \\ &\geq d^2(a, x) - d^2(x, y) \end{aligned}$$

and so

$$f(y) \leq f(a) + g_a(\log_a x, \log_a y)$$

for all $y \in \mathcal{H}$. In particular, we have

$$f((1-t)a \oplus tb) \leq f(a) + g_a(\log_a x, \log_a((1-t)a \oplus tb)),$$

so that combined, we have

$$(1-t)f(a) + tf(b) + t(1-t)\frac{1}{2}d^2(a, b) \leq f(a) + g_a(\log_a x, \log_a((1-t)a \oplus tb)).$$

Hence

$$f(b) - f(a) \leq \frac{g_a(\log_a x, \log_a((1-t)a \oplus tb))}{t} - (1-t)\frac{1}{2}d^2(a, b)$$

for all $t \in (0, 1]$. Similarly, we get

$$f(a) - f(b) \leq \frac{g_b(\log_b x, \log_b((1-t)b \oplus ta))}{t} - (1-t)\frac{1}{2}d^2(b, a).$$

Using $g_a(\log_a x, \log_a((1-t)a \oplus tb))/t = g_a(\log_a x, \log_a b)$ (and similarly for b), as established in the proof of Proposition 2.9 in [97], we have

$$g_a(\log_a x, \log_a b) - (1-t)d^2(a, b) \geq -g_b(\log_b x, \log_b a).$$

Taking $t \rightarrow 0$, we get

$$g_a(\log_a x, \log_a b) \geq -g_b(\log_b x, \log_b a) + d^2(a, b)$$

as claimed. $\square$

We can now establish our strong law of large numbers for integrable random monotone vector fields over Hilbert-Hadamard spaces. Similar to [108], the mode of convergence for the random monotone vector fields will be that of R -convergence, as already mentioned before. Concretely, the R -topology discussed previously in Definition 4.8 gives rise to the following notion of R -convergence: Given a sequence $(A_n) \subseteq \mathcal{M}(\mathcal{H})$ together with $A \in \mathcal{M}(\mathcal{H})$, we say that (A_n) R -converges to A , written $A_n \rightarrow^R A$, if

$$J_\lambda^{A_n}(x) \rightarrow J_\lambda^A(x)$$

for all $x \in \mathcal{H}$ and $\lambda > 0$. In particular, if $A_n, A : \Omega \rightarrow \mathcal{M}(\mathcal{H})$ are random monotone vector fields, then $A_n \rightarrow^R A$ holds $\mathbb{P}$ -a.s. if, and only if, there exists a set $\Omega' \subseteq \Omega$ of measure one such that $J_\lambda^{A_n(\omega)}(x) \rightarrow J_\lambda^{A(\omega)}(x)$ for all $\omega \in \Omega'$, $x \in \mathcal{H}$ and $\lambda > 0$.

We now turn to the second main result of this paper:

Theorem 5.3. *Let $(\Omega, \mathbb{F}, \mathbb{P})$ and $(E, \mathbb{E}, \mu)$ be probability spaces and let $(\mathcal{H}, d)$ be a separable Hilbert-Hadamard space. Further, fix an i.i.d. sequence (ξ_i) of random variables $\xi_i : \Omega \rightarrow E$ with distribution μ . Let $A : E \rightarrow \mathcal{M}(\mathcal{H})$ be an integrable random monotone vector field such that $\frac{1}{n} \sum_{i=1, \dots, n}^{\rightarrow} A(\xi_i)$ satisfies the surjectivity condition $\mathbb{P}$ -a.s. Then*

$$\frac{1}{n} \sum_{i=1, \dots, n}^{\rightarrow} A(\xi_i) \rightarrow^R \int A(e) d\mu(e) \quad \mathbb{P}\text{-a.s.}$$

as $n \rightarrow \infty$.

Proof. For brevity, write

$$\underline{A}_n := \frac{1}{n} \sum_{i=1, \dots, n}^{\rightarrow} A(\xi_i) \text{ and } \underline{A} := \int A(e) d\mu(e).$$

Given any $x \in \mathcal{H}$ and $\lambda > 0$, write $y := J_\lambda^{\underline{A}}(x)$ and $z := \frac{1}{\lambda} \log_y x \in \underline{A}(y)$. As $z \in \underline{A}(y)$, there is an $\alpha \in \mathcal{S}^1(A(\cdot, y))$ such that $z = \int \alpha d\mu$. Define

$$z_n := \frac{1}{n} \sum_{i=1, \dots, n}^{\rightarrow} \alpha(\xi_i)$$

and note that $z_n(\omega) \in \underline{A}_n(\omega, y)$. Then by Corollary 3.5 on $T_y \mathcal{H}$, we have

$$z_n \rightarrow z \quad \mathbb{P}\text{-a.s.}$$

as $n \rightarrow \infty$. Now, define $y_n(\omega) := J_\lambda^{\underline{A}_n(\omega)}(x)$ and $u_n(\omega) := \frac{1}{\lambda} \log_{y_n(\omega)} x \in \underline{A}_n(\omega, y_n(\omega))$. As $\underline{A}_n(\omega, \cdot)$ is monotone for any ω , we get

$$g_{y_n}(u_n, \log_{y_n} y) \leq -g_y(z_n, \log_y y_n)$$

almost surely. Lemma 5.2 yields that

$$g_{y_n}(\log_{y_n} x, \log_{y_n} y) \geq -g_y(\log_y x, \log_y y_n) + d^2(y_n, y),$$

and so

$$g_{y_n}(u_n, \log_{y_n} y) \geq -g_y(z, \log_y y_n) + \frac{1}{\lambda} d^2(y_n, y).$$

Combined, we have

$$\begin{aligned} d^2(y_n, y) &\leq \lambda (g_{y_n}(u_n, \log_{y_n} y) + g_y(z, \log_y y_n)) \\ &\leq \lambda (g_y(z, \log_y y_n) - g_y(z_n, \log_y y_n)) \\ &\leq \lambda \|\log_y y_n\| d_y(z, z_n) \\ &= \lambda d(y, y_n) d_y(z, z_n), \end{aligned}$$

where the third inequality follows from Lemma 2.6. Hence, we have

$$d(J_\lambda^{A_n(\omega)}(x), J_\lambda^A(x)) = d(y_n(\omega), y) \leq \lambda d_y(z, z_n(\omega)) \rightarrow 0$$

for almost all ω as $n \rightarrow \infty$. This shows that for any $x \in \mathcal{H}$ and $\lambda > 0$, there exists a set $\Omega_{x,\lambda}$ of measure one such that

$$J_\lambda^{A_n(\omega)}(x) \rightarrow J_\lambda^A(x)$$

for all $\omega \in \Omega_{x,\lambda}$. Using that $\mathcal{H}$ is separable, let $Z \subseteq \mathcal{H}$ be a countable dense subset. Define

$$\Omega' := \bigcap_{z \in Z, r \in \mathbb{Q}_{>0}} \Omega_{z,r},$$

which has measure one as well. Now, let $\omega \in \Omega'$ as well as $x \in \mathcal{H}$ and $\lambda > 0$ be arbitrary. Let $\varepsilon \in (0, 1]$ be arbitrary and let $z \in Z$ be such that $d(x, z) < \frac{\varepsilon}{6}$ and let $r \in \mathbb{Q}_{>0}$ be such that $\lambda \leq r \leq \min\{\lambda/(1 - \varepsilon/24b), \lceil 2\lambda \rceil\}$ for $b \geq 1$ such that $b \geq d(J_{\lceil 2\lambda \rceil}^A(z), z)$. Then, using the nonexpansivity of the resolvents J_λ^A and $J_\lambda^{A_n}$, we get

$$\begin{aligned} d(J_\lambda^{A_n(\omega)}(x), J_\lambda^A(x)) &\leq d(J_\lambda^{A_n(\omega)}(x), J_\lambda^{A_n(\omega)}(z)) + d(J_\lambda^{A_n(\omega)}(z), J_\lambda^A(z)) + d(J_\lambda^A(x), J_\lambda^A(z)) \\ &\leq 2d(x, z) + d(J_\lambda^{A_n(\omega)}(z), J_\lambda^A(z)) \\ &\leq 2d(x, z) + d(J_\lambda^{A_n(\omega)}(z), J_r^{A_n(\omega)}(z)) + d(J_r^{A_n(\omega)}(z), J_r^A(z)) + d(J_r^A(z), J_\lambda^A(z)). \end{aligned}$$

Now, note that

$$d(J_r^A(z), J_\lambda^A(z)) \leq 2 \left(1 - \frac{\lambda}{r}\right) d(J_{\lceil 2\lambda \rceil}^A(z), z) \leq 2 \left(1 - \frac{\lambda}{r}\right) b$$

using (the proof of) Lemma 4.5. Similarly, we have

$$\begin{aligned} d(J_\lambda^{A_n(\omega)}(z), J_r^{A_n(\omega)}(z)) &\leq 2 \left(1 - \frac{\lambda}{r}\right) d(J_{\lceil 2\lambda \rceil}^{A_n(\omega)}(z), z) \\ &\leq 2 \left(1 - \frac{\lambda}{r}\right) d(J_{\lceil 2\lambda \rceil}^{A_n(\omega)}(z), J_{\lceil 2\lambda \rceil}^A(z)) + 2 \left(1 - \frac{\lambda}{r}\right) d(J_{\lceil 2\lambda \rceil}^A(z), z) \\ &\leq 2d(J_{\lceil 2\lambda \rceil}^{A_n(\omega)}(z), J_{\lceil 2\lambda \rceil}^A(z)) + 2 \left(1 - \frac{\lambda}{r}\right) b. \end{aligned}$$

Combined, we have

$$\begin{aligned} d(J_\lambda^{A_n(\omega)}(x), J_\lambda^A(x)) &\leq 2d(x, z) + 2d(J_{[2\lambda]}^{A_n(\omega)}(z), J_{[2\lambda]}^A(z)) + 4\left(1 - \frac{\lambda}{r}\right)b + d(J_r^{A_n(\omega)}(z), J_r^A(z)) \\ &\leq \frac{3\varepsilon}{6} + 2d(J_{[2\lambda]}^{A_n(\omega)}(z), J_{[2\lambda]}^A(z)) + d(J_r^{A_n(\omega)}(z), J_r^A(z)). \end{aligned}$$

Using $\omega \in \Omega' \subseteq \Omega_{z,r} \cap \Omega_{z,[2\lambda]}$, there is an m such that

$$d(J_r^{A_n(\omega)}(z), J_r^A(z)), d(J_{[2\lambda]}^{A_n(\omega)}(z), J_{[2\lambda]}^A(z)) \leq \frac{\varepsilon}{6}$$

for all $n \geq m$. Hence

$$d(J_\lambda^{A_n(\omega)}(x), J_\lambda^A(x)) \leq \varepsilon \text{ for all } n \geq m.$$

As $\varepsilon \in (0, 1]$ was arbitrary, we get that $J_\lambda^{A_n(\omega)}(x)$ converges to $J_\lambda^A(x)$. This yields the result. $\square$

Remark 5.4. Also the strong law for random monotone vector fields in the above Theorem 5.3 can be made quantitative. At first, note that by following the proof of Theorem 5.3, we get

$$d(J_\lambda^{A_n(\omega)}(x), J_\lambda^A(x)) \leq \lambda d_y(z, z_n(\omega))$$

for almost all ω . Theorem 3.4 can then be employed to provide a quantitative rendering of $z_n \rightarrow z$ $\mathbb{P}$ -a.s.: For any $\varepsilon > 0$ and any $m \in \mathbb{N}$:

$$\mathbb{P}\left(\sup_{n \geq m} d(z_n, z) > \varepsilon\right) \leq \frac{6}{\varepsilon} \left(\frac{\sqrt{\mathbb{E}[\|\alpha(\xi_1)\|_y]}}{m^{1/2}} + U_{\mathbb{P}}(m^{1/8}) \right) + \frac{72}{\min\{\varepsilon^2, 1\}m^{1/4}}$$

where $U_{\mathbb{P}}(a) := \mathbb{E}[\|\alpha(\xi_1)\|_y \mathbf{1}_{\|\alpha(\xi_1)\|_y \geq a}]$ where $\alpha \in \mathcal{S}^1(A(\cdot, y))$ such that $z = \int \alpha d\mu$ and $y = J_\lambda^A(x)$. Hence, we in particular have

$$\mathbb{P}\left(\sup_{n \geq m} d(J_\lambda^{A_n}(x), J_\lambda^A(x)) > \varepsilon\right) \leq \frac{6}{\varepsilon/\lambda} \left(\frac{\sqrt{\mathbb{E}[\|\alpha(\xi_1)\|_y]}}{m^{1/2}} + U_{\mathbb{P}}(m^{1/8}) \right) + \frac{72}{\min\{\varepsilon^2/\lambda^2, 1\}m^{1/4}}$$

for all $\varepsilon > 0$ and $m \in \mathbb{N}$. If A is L^1 -uniformly majorized (in the spirit of [96]), that is $\sup_{y \in A(e, x)} \|y\|_x \leq g(e, b)$ almost everywhere for all $b > 0$ and all $x \in \mathcal{H}$ with $d(o, x) \leq b$, where $g(\cdot, b) \in L^1(E, (0, \infty), \mu)$ for all $b > 0$, then we can estimate $\sqrt{\mathbb{E}[\|\alpha(\xi_1)\|_y]} \leq \sqrt{\mathbb{E}[g(\cdot, c)]}$ as well as

$$U_{\mathbb{P}}(a) = \mathbb{E}[\|\alpha(\xi_1)\|_y \mathbf{1}_{\|\alpha(\xi_1)\|_y \geq a}] \leq \mathbb{E}[g(\cdot, c) \mathbf{1}_{g(\cdot, c) \geq a}],$$

where $c \geq d(o, y)$. In particular, if $d \geq d(o, J_\lambda^A(o))$ and $b \geq d(o, x)$, then

$$d(o, y) \leq d(o, J_\lambda^A(o)) + d(J_\lambda^A(o), J_\lambda^A(x)) \leq d + b$$

so that we can take $c = d + b$ and so that the above bound holds uniformly for all $x \in \mathcal{H}$ in terms of such d and b as well as $\lambda > 0$. In particular, if A is L^1 -dominated, that is $\sup_{y \in A(e, x)} \|y\|_x \leq g(e)$ almost everywhere and $g(\cdot) \in L^1(E, (0, \infty), \mu)$ for all $x \in \mathcal{H}$, then the above holds uniformly for all $x \in \mathcal{H}$. We do not spell this out here further.

6. ON THE INTERCHANGE OF INTEGRATION AND SUBDIFFERENTIATION

Let $(T, \mathbb{T}, \tau)$ be a complete probability space and $(\mathcal{H}, d)$ be a separable Hadamard space, and let $f : T \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a proper normal convex integrand (recall [104] and Example 4.7). The goal of this section is to investigate in what way the interchange between subdifferentiation and integration in the form of

$$\int \partial f(t, x) d\tau(t) = \partial \left(\int f(t, \cdot) d\tau(t) \right) (x) \text{ for all } x \in \mathcal{H}$$

holds true in separable Hilbert-Hadamard spaces. Results of this type have a longer history. One of the strongest instances was obtained by Rockafellar and Wets [103], who actually proved a stronger relation on the interchange of the conditional expectation and the subdifferential of $F(x) := \int f(t, x(t)) d\tau(t)$ defined for $x \in L^2(T, \mathcal{H}, \tau)$, over $\mathcal{H} = \mathbb{R}^d$ (see also [31] for similar results). A result such as the above for suitable functions f over $\mathbb{R}^d$ was first derived by Ioffe and Tikhomirov [52], and later extended and modified by Levin [69], Hiriart-Urruty [51] and Thibault [119], among others.

Throughout, we now write $f(x) := \int f(t, x) d\tau(t)$ as well as $\underline{\partial} f(x) := \int \partial f(t, x) d\tau(t)$. As a first observation, we note that over separable Hilbert-Hadamard spaces, one direction always holds true (as already observed in Example 3.2 of [97]):

Lemma 6.1. *Let $(\mathcal{H}, d)$ be a separable Hilbert-Hadamard space and let $f : T \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a proper normal convex integrand. For any $x \in \mathcal{H}$: $\underline{\partial} f(x) \subseteq \partial f(x)$.*

Proof. Let $u \in L^1(T, T_x \mathcal{H}, \tau)$ be a selection of ∂f , that is for all $t \in T$ and all $y \in \mathcal{H}$:

$$f(t, y) \geq f(t, x) + g_x(u(t), \log_x y).$$

Integrating and using Lemma 2.10 yields

$$\underline{f}(y) \geq \underline{f}(x) + g_x \left(\int u d\tau, \log_x y \right)$$

for all $y \in \mathcal{H}$, so that $\int u d\tau \in \underline{\partial} f(x)$. □

Unfortunately, the converse does not hold true over general Hilbert-Hadamard spaces, as the following construction shows:

Example 6.2. Consider $\mathcal{H} := [0, \infty)$. This is a non-empty, closed and convex subset of a Hilbert-Hadamard space, and so itself a Hilbert-Hadamard space. In particular, we have $T_0 \mathcal{H} = [0, \infty)$ and $\log_0 x = x$ for all $x \in \mathcal{H} := [0, \infty)$. Consider the complete probability space on $T := \{1, 2\}$ with measure τ induced by $\tau(\{1\}) := \tau(\{2\}) := \frac{1}{2}$, and define the function $f : T \times \mathcal{H} \rightarrow \mathbb{R}$ by

$$f(1, x) := -x \text{ and } f(2, x) := 2x \text{ for all } x \in \mathcal{H} := [0, \infty).$$

Clearly $f(1, \cdot)$ and $f(2, \cdot)$ are proper, convex and continuous. Further, f is clearly measurable. We now have

$$\underline{f}(x) = \frac{1}{2}f(1, x) + \frac{1}{2}f(2, x) = \frac{x}{2}.$$

It can be easily seen that $\partial f(0) = [0, \frac{1}{2}]$. Now, it holds that $u \in \partial f(1, 0)$ if, and only if, $u \in T_0 \mathcal{H}$ and $u \leq -1$, but since $u \in T_0 \mathcal{H}$ implies $u \geq 0$, we get $\partial f(1, 0) = \emptyset$. Hence, we have

$$\underline{\partial} f(0) = \frac{1}{2}\partial f(1, 0) + \frac{1}{2}\partial f(2, 0) = \emptyset$$

which shows that $\underline{\partial} f(0) \not\subseteq \partial f(0)$.

A key problem in the above example was that $T_0\mathcal{H}$, while flat, is not a linear space, so that $u \in T_0\mathcal{H}$ contradicted $u \leq -1$.

Indeed, once the linearity of the tangent cones is imposed, we are actually able to establish the converse to the above result (if f is additionally Lipschitz). We first give these spaces a name:

Definition 6.3. We call a Hadamard space $\mathcal{H}$ a full Hilbert-Hadamard space if each tangent cone $T_x\mathcal{H}$ is isometric to a Hilbert space.

Hilbert spaces and Hadamard manifolds as well as their infinite-dimensional generalizations are not only Hilbert-Hadamard spaces but also full Hilbert-Hadamard spaces. Further, Example 2.2 shows that also full Hilbert-Hadamard spaces are not restricted to the settings of manifolds. Before moving on, we note the following result relating Hilbert-Hadamard spaces to full Hilbert-Hadamard spaces. Say that $\mathcal{H}$ has the geodesic extension property if every geodesic in $\mathcal{H}$ can be extended to an isometry $\mathbb{R} \rightarrow \mathcal{H}$.

Lemma 6.4. *If $\mathcal{H}$ is a Hilbert-Hadamard space and it satisfies the geodesic extension property, then $\mathcal{H}$ is a full Hilbert-Hadamard space.*

Proof. Fix $x \in \mathcal{H}$. By assumption (recall Remark 2.1), $T_x\mathcal{H}$ isometrically embeds into a Hilbert space H , say via a mapping $\iota : T_x\mathcal{H} \rightarrow H$. Using a shift, we may assume that $\iota(0_x) = 0$. We now show that $\iota(T_x\mathcal{H})$ is closed under additive inverses. For this, let $\gamma \in \Sigma'_x\mathcal{H}$ and let $-\gamma \in \Sigma'_x\mathcal{H}$ be an opposite geodesic. Note that $\|\gamma\|_x = \|-\gamma\|_x$ as well as $\angle_x(\gamma, -\gamma) = \pi$ so that

$$d_x(\gamma, -\gamma) = \sqrt{2 - 2 \cos \angle_x(\gamma, -\gamma)} = \sqrt{4} = 2 \|\gamma\|_x.$$

As ι is an isometry, we get $\|\iota(\gamma)\|_H = \|\iota(-\gamma)\|_H$ and $\|\iota(\gamma) - \iota(-\gamma)\|_H = 2 \|\iota(\gamma)\|_H$, where $\|\cdot\|_H$ is the norm on H . As we hence have an equality in the triangle inequality in H , we get that $\iota(-\gamma) = -\iota(\gamma)$. As $\Sigma'_x\mathcal{H}$ is dense in $\Sigma_x\mathcal{H}$, the result extends to $\Sigma_x\mathcal{H}$ by continuity and then it further extends to $T_x\mathcal{H}$, the cone over $\Sigma_x\mathcal{H}$, where we in particular use that $\iota(tv) = t\iota(v)$, which holds since $\|\iota(v) - \iota(tv)\|_H = d_x(v, tv) = \|v\|_x \sqrt{(t-1)^2} = \|v\|_x |t-1| = \|\iota(v)\|_H |t-1|$. Combined, we get that $\iota(T_x\mathcal{H})$ is a nonempty closed and convex subset of H containing 0 and which is closed under non-negative scalar multiplication and additive inverses. Hence, $\iota(T_x\mathcal{H})$ is linear. $\square$

To now establish the converse of the above Lemma 6.1 over full Hilbert-Hadamard spaces (and suitable Lipschitz functions), we broadly follow a similar approach as used in [51] (which in turn relies on the exactness result from Valadier [120]), together with suitably adapted variants of a variety of results due to Castaing and Valadier [35] highlighted throughout.

In particular, we rely on the conjugate of F , which we introduce following the work [26, 110] (see also [24, 50] for related works using Busemann functions). Generally, for a Hadamard space $\mathcal{H}$ and $h : \mathcal{H} \rightarrow (-\infty, +\infty]$ and $p \in \mathcal{H}$, define $h_p^* : T_p\mathcal{H} \rightarrow (-\infty, +\infty]$ via

$$h_p^*(u) := \sup_{q \in \mathcal{H}} \{g_p(u, \log_p q) - h(q)\}.$$

We begin with some immediate lemmas.

Lemma 6.5. *Let $\mathcal{H}$ be a Hadamard space and let $h : \mathcal{H} \rightarrow (-\infty, +\infty]$ be given. For any $p \in \mathcal{H}$ and $u \in T_p\mathcal{H}$:*

- (1) $h_p^*(u) + h(p) \geq 0$,
- (2) $u \in \partial h(p)$ if, and only if, $h_p^*(u) + h(p) = 0$.

Proof. (1) By definition, we have that

$$h_p^*(u) + h(p) \geq g_p(u, \log_p p) - h(p) + h(p) = 0.$$

(2) Suppose $u \in \partial h(p)$, that is

$$h(q) \geq h(p) + g_p(u, \log_p q) \text{ for all } q \in \mathcal{H}.$$

Re-arranged, we get

$$-h(p) \geq g_p(u, \log_p q) - h(q) \text{ for all } q \in \mathcal{H}.$$

Taking the supremum over q , we get $-h(p) \geq h_p^*(u)$, that is $h_p^*(u) + h(p) \leq 0$. By (1), we get $h_p^*(u) + h(p) = 0$.

Conversely, if $h_p^*(u) + h(p) = 0$, then $h_p^*(u) + h(p) \leq 0$ and so

$$-h(p) \geq h_p^*(u) \geq g_p(u, \log_p q) - h(q) \text{ for all } q \in \mathcal{H}.$$

Re-arranging yields that $u \in \partial h(p)$. □

As discussed in Section 2, $L^2(T, \mathcal{H}, \tau)$ is a separable Hilbert-Hadamard space whenever $\mathcal{H}$ is so and T is separable, and the tangent cone $T_p L^2(T, \mathcal{H}, \tau)$ of $L^2(T, \mathcal{H}, \tau)$ at a point $p \in \mathcal{H}$ (identified with its constant map) is (isometric to) $L^2(T, T_p \mathcal{H}, \tau)$. In particular, the pseudo-Riemannian metric takes the form of

$$g_p(u, v) = \int g_p(u(t), v(t)) d\tau(t).$$

Hence, we have

$$F_p^*(v) := \sup_{q \in L^2(T, \mathcal{H}, \tau)} \{g_p(v, \log_p q) - F(q)\}$$

for $p \in \mathcal{H}$ and $v \in L^2(T, T_p \mathcal{H}, \tau)$.

We now show that F_p^* can be represented as the integral over the conjugate f_p^* . For that, we follow the approach taken in [35] for a similar result. At first, we show that the latter is a normal convex integrand over Hilbert-Hadamard spaces, which is an adaptation of Corollary VII-2 in [35].

Lemma 6.6. *Let $\mathcal{H}$ be a separable Hilbert-Hadamard space and let $f : T \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a proper normal integrand. For any $p \in \mathcal{H}$, the conjugate $f_p^* : T \times T_p \mathcal{H} \rightarrow (-\infty, +\infty]$ defined by*

$$f_p^*(t, u) := \sup_{q \in \mathcal{H}} \{g_p(u, \log_p q) - f(t, q)\}$$

is a normal convex integrand.

Proof. First, note that by Lemma 2.3, $f_p^*(t, \cdot)$ is a supremum of convex and lsc functions and hence is convex and lsc. As T is complete, it suffices to show that f_p^* is $\mathbb{T} \otimes \mathbb{B}(T_p \mathcal{H})$ -measurable. To see this, note that

$$f_p^*(t, u) = \sup_{(q, r) \in \text{epi} f(t, \cdot)} \{g_p(u, \log_p q) - r\}.$$

As $\text{epi} f(t, \cdot)$ is measurable, there is a sequence (q_n, r_n) of measurable selections such that, for any $t \in T$, the sequence $(q_n(t), r_n(t))$ is dense in $\text{epi} f(t, \cdot)$. Hence, we have

$$f_p^*(t, u) = \sup_{n \in \mathbb{N}} \{g_p(u, \log_p q_n(t)) - r_n(t)\},$$

using that g_p and $\log_p$ are continuous. Now, $g_p(u, \log_p q_n(t))$ is a Carathéodory function and hence is $\mathbb{T} \otimes \mathbb{B}(T_p \mathcal{H})$ -measurable. Therefore, f_p^* is a supremum of $\mathbb{T} \otimes \mathbb{B}(T_p \mathcal{H})$ -measurable functions and hence is $\mathbb{T} \otimes \mathbb{B}(T_p \mathcal{H})$ -measurable. □

Further, we need the following lemma, which is an adaptation of Proposition VII-4 in [35].

Lemma 6.7. *Let $\mathcal{H}$ be a complete separable metric space and $s : T \rightarrow \mathcal{H}$ be measurable. Then there exists an increasing sequence of measurable sets T_n such that $\tau(T \setminus \bigcup_{n \in \mathbb{N}} T_n) = 0$ and $\overline{s(T_n)}$ is compact for all $n \in \mathbb{N}$.*

Proof. Consider the push-forward measure τ_s on $\mathcal{B}(\mathcal{H})$. As $\mathcal{H}$ is a Polish space, τ_s is a Radon measure (as follows from Prohorov's theorem, see Theorem 13.29 in [61]). Therefore, there exists an increasing sequence of compact sets $K_n \subseteq \mathcal{H}$ such that $\tau_s(\bigcup_{n \in \mathbb{N}} K_n) = \tau_s(\mathcal{H})$. Define $T_n := s^{-1}(K_n)$. This gives the claim. $\square$

Lastly, we require another result from measurable selection theory, stating the graph-measurability of set-valued maps induced by inequalities (extending Proposition 2I in [102]):

Lemma 6.8. *Let $\mathcal{H}$ be a complete separable metric space and let $f : T \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a normal integrand and α be measurable. Then*

$$\Gamma(t) := \{x \in \mathcal{H} \mid f(t, x) \leq \alpha(t)\}$$

is closed-valued and graph measurable.

Proof. That Γ is closed-valued follows immediately as f is lsc in its right argument. We have

$$\begin{aligned} \text{gra}\Gamma &= \{(t, x) \in T \times \mathcal{H} \mid f(t, x) \leq \alpha(t)\} \\ &= \{(t, x) \in T \times \mathcal{H} \mid (x, \alpha(t)) \in \text{epi}f(t, \cdot)\} \\ &= \varphi^{-1}(\text{gra}(\text{epi}f)) \end{aligned}$$

for $\varphi(t, x) = (t, x, \alpha(t))$, which is clearly measurable. Hence $\text{gra}\Gamma \in \mathbb{T} \otimes \mathcal{B}(\mathcal{H})$ since $\text{gra}(\text{epi}f) \in \mathbb{T} \otimes \mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathbb{R})$. $\square$

We now can give the respective result on the conjugate F_p^* , where we follow the approach taken in the proof of Theorem VII-7 in [35].

Lemma 6.9. *Let T be separable and let $\mathcal{H}$ be a separable Hilbert-Hadamard space, let $f : T \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a proper normal convex integrand and assume that F is proper and lsc. For any $p \in \mathcal{H}$ and any $v \in L^2(T, T_p\mathcal{H}, \tau)$:*

$$F_p^*(v) = \int f_p^*(t, v(t)) d\tau(t).$$

Proof. Let $q \in L^2(T, \mathcal{H}, \tau)$. Then

$$g_p(v(t), \log_p q(t)) - f(t, q(t)) \leq f_p^*(t, v(t))$$

for any $t \in T$, and hence

$$g_p(v, \log_p q) - F(q) \leq \int f_p^*(t, v(t)) d\tau(t).$$

Now, taking the supremum over all such q , we get

$$\int f_p^*(t, v(t)) d\tau(t) \geq \sup_{q \in L^2(T, \mathcal{H}, \tau)} \{g_p(v, \log_p q) - F(q)\}.$$

For the converse, let $\beta < \int f_p^*(t, v(t)) d\tau(t)$. Now, we define a real-valued integrable function γ such that $\beta < \int \gamma(t) d\tau(t) - \varepsilon$ for some $\varepsilon > 0$ and $\gamma(t) < f_p^*(t, v(t))$ for all $t \in T$. If $\int f_p^*(t, v(t)) d\tau(t) < +\infty$, then we set $\gamma(t) := f_p^*(t, v(t)) - \varepsilon$ for $\varepsilon > 0$ such that

$$\beta < \int f_p^*(t, v(t)) d\tau(t) - \varepsilon.$$

If $\int f_p^*(t, v(t)) d\tau(t) = +\infty$, then let N and $\varepsilon > 0$ be such that

$$\beta < \int \min\{N, f_p^*(t, v(t))\} d\tau(t) - \varepsilon$$

and set $\gamma(t) := \min\{N, f_p^*(t, v(t))\} - \varepsilon$. Define

$$\Gamma(t) := \{x \in \mathcal{H} \mid g_p(v(t), \log_p x) - f(t, x) \geq \gamma(t)\}.$$

Then Γ is graph-measurable by Lemma 6.8 and so, by completeness, measurable. Further, the values of Γ are closed and non-empty, the latter since $\gamma(t) < f_p^*(t, v(t))$. Hence, by the Kuratowski–Ryll–Nardzewski measurable selection theorem (see Theorem 8.1.3 in [14]), there exists a measurable $s : T \rightarrow \mathcal{H}$ such that

$$g_p(v(t), \log_p s(t)) - f(t, s(t)) \geq \gamma(t) \text{ for all } t \in T.$$

Now, take $u_0 \in \text{dom} F$, which exists since F is proper. Then, there exists an integrable α such that

$$g_p(v(t), \log_p u_0(t)) - f(t, u_0(t)) \geq \alpha(t) \text{ for all } t \in T.$$

Using Lemma 6.7, take now an increasing sequence T_n such that $\tau(T \setminus \bigcup_n T_n) = 0$ and such that $\overline{s(T_n)}$ is compact for all $n \in \mathbb{N}$. Define

$$u_n(t) := \begin{cases} s(t) & \text{if } t \in T_n, \\ u_0(t) & \text{if } t \in T \setminus T_n. \end{cases}$$

Clearly we have $u_n \in L^2(T, \mathcal{H}, \tau)$. In particular, we have

$$\int g_p(v(t), \log_p u_n(t)) d\tau(t) - \int f(t, u_n(t)) d\tau(t) \geq \int_{T_n} \gamma(t) d\tau(t) + \int_{T \setminus T_n} \alpha(t) d\tau(t).$$

As $\int_{T \setminus T_n} \alpha(t) d\tau(t) \rightarrow 0$ as well as $\int_{T_n} \gamma(t) d\tau(t) \rightarrow \int_T \gamma(t) d\tau(t)$ when $n \rightarrow \infty$, there is an n such that

$$\int_{T \setminus T_n} \alpha(t) d\tau(t) \geq -\varepsilon/2 \text{ and } \int_{T_n} \gamma(t) d\tau(t) \geq \int_T \gamma(t) d\tau(t) - \varepsilon/2.$$

Hence, for such an n , we get

$$g_p(v, \log_p u_n) - F(u_n) = \int g_p(v(t), \log_p u_n(t)) d\tau(t) - \int f(t, u_n(t)) d\tau(t) \geq \int_T \gamma(t) d\tau(t) - \varepsilon > \beta.$$

This completes the proof. $\square$

The above can now be used rather immediately to show the following characterization of the subdifferential of F :

Lemma 6.10. *Let T be separable and let $\mathcal{H}$ be a separable Hilbert-Hadamard space, let $f : T \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a proper normal convex integrand and suppose that F is proper and lsc. Then, for any $x \in \mathcal{H}$:*

$$\partial F(x) = \{v \in L^2(T, T_x \mathcal{H}, \tau) \mid v(t) \in \partial f(t, x) \text{ almost surely}\}.$$

Proof. By Lemma 6.5, we have $v \in \partial F(x)$ if, and only if, $F_x^*(v) + F(x) = 0$. By Lemma 6.9, the latter is equivalent to

$$\begin{aligned} 0 &= \int f_x^*(t, v(t)) d\tau(t) + \int f(t, x) d\tau(t) \\ &= \int (f_x^*(t, v(t)) + f(t, x)) d\tau(t). \end{aligned}$$

As $f_x^*(t, v(t)) + f(t, x) \geq 0$ by Lemma 6.5, the above is hence equivalent to $f_x^*(t, v(t)) + f(t, x) = 0$ almost surely, which by Lemma 6.5 again is equivalent to $v(t) \in \partial f(t, x)$ almost surely. $\square$

The following key lemma now requires the linearity of the tangent cones as well as the Lipschitz continuity of the integrand.

Lemma 6.11. *Let T be separable and let $\mathcal{H}$ be a separable full Hilbert-Hadamard space. Let $f : T \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a proper normal convex integrand and suppose that f is L^2 -Lipschitz, that is*

$$|f(t, x) - f(t, y)| \leq L(t)d(x, y) \text{ for all } x, y \in \mathcal{H} \text{ and } t \in T,$$

where $L \in L^2(T, (0, \infty), \tau)$. Assume that there exists a $u_0 \in L^2(T, \mathcal{H}, \tau)$ such that $f(\cdot, u_0(\cdot)) \in L^1(T, \mathbb{R}, \tau)$. Then, for any $x \in \mathcal{H}$:

$$\partial \underline{f}(x) = \left\{ \int v \, d\tau \mid v \in \partial F(x) \right\}.$$

Conceptually, the above result is similar to (a special case of) Theorem I-29 in [35] on the representation of the subdifferential $\partial(h \circ A)(x)$ of a composition $h \circ A$ of a convex function h with a linear operator via its conjugate A^* as $A^*(\partial h(A(x)))$.

To prove Lemma 6.11, we rely on the directional derivative $d_x h(v)$ at a point $x \in \mathcal{H}$ and $v \in T_x \mathcal{H}$ of a convex and Lipschitz continuous function $h : \mathcal{H} \rightarrow \mathbb{R}$. To define such an object, we follow the approach of [78] and first define the function $\sigma_x h : \Sigma'_x \mathcal{H} \rightarrow \overline{\mathbb{R}}$ via

$$\sigma_x h(\gamma) := \lim_{t \rightarrow 0} \frac{h(\gamma(t)) - h(x)}{t}$$

for $x \in \mathcal{H}$ and $\gamma \in \Sigma'_x \mathcal{H}$. Note that by convexity of h , the above fraction is actually nondecreasing in t . A related object was studied already in [81] as well as [75, 95, 99] (the latter in different, partly more general settings). As shown in [78], using that h is Lipschitz, this function can be extended to $T_x \mathcal{H}$ and this extension retains various properties of h .

Lemma 6.12 (Proposition 2.16 in [78]). *Let $\mathcal{H}$ be a Hadamard space and let $h : \mathcal{H} \rightarrow \mathbb{R}$ be convex and Lipschitz-continuous with constant $L > 0$. Then, for each $x \in \mathcal{H}$, there exists a unique continuous map $d_x h : T_x \mathcal{H} \rightarrow \mathbb{R}$ such that*

$$d_x h(\gamma) = \sigma_x h(\gamma) \text{ for all } \gamma \in \Sigma'_x \mathcal{H}.$$

Further, $d_x h$ is convex, Lipschitz-continuous with constant L and positively 1-homogeneous, that is $d_x h(\lambda v) = \lambda d_x h(v)$ for any $v \in T_x \mathcal{H}$ and $\lambda \geq 0$.

As every convex and positively 1-homogeneous function is sub-additive over linear spaces, we in particular get the following result for linear tangent cones:

Corollary 6.13. Let $\mathcal{H}$ be a Hadamard space and let $h : \mathcal{H} \rightarrow \mathbb{R}$ be convex and Lipschitz-continuous with constant $L > 0$. Fix $x \in \mathcal{H}$. If $T_x \mathcal{H}$ is a Hilbert space, then $d_x h$ is sub-additive.

The last property we need of the directional derivative is that it characterizes the subgradient:

Lemma 6.14 (extending Theorem 2.7 in [81]). *Let $\mathcal{H}$ be a Hadamard space and let $h : \mathcal{H} \rightarrow \mathbb{R}$ be convex and Lipschitz-continuous with constant $L > 0$. Then, for any $x \in \mathcal{H}$ and $v \in T_x \mathcal{H}$, $v \in \partial h(x)$ if, and only if, $g_x(v, w) \leq d_x h(w)$ for all $w \in T_x \mathcal{H}$.*

Proof. Suppose that $v \in \partial h(x)$, that is

$$h(y) \geq h(x) + g_x(v, \log_x y) \text{ for all } y \in \mathcal{H}.$$

Hence, for $\gamma \in \Sigma'_x \mathcal{H}$ with length l , we have

$$h(\gamma(t)) \geq h(x) + g_x(v, \log_x \gamma(t)) \text{ for all } t \in [0, l].$$

Now, note that

$$g_x(v, \log_x \gamma(t)) = t g_x(v, \gamma).$$

Dividing by t and taking $t \rightarrow 0$ yields $\sigma_x h(\gamma) \geq g_x(v, \gamma)$. Using continuity of $d_x h$, this now extends to $\gamma \in \Sigma_x \mathcal{H}$. For $w = \lambda \gamma \in T_x \mathcal{H}$, we then get

$$d_x h(\lambda \gamma) = \lambda d_x h(\gamma) \geq \lambda g_x(v, \gamma) = g_x(v, \lambda \gamma)$$

using positive 1-homogeneity of $d_x h$ and g_x .

For the converse, suppose $g_x(v, w) \leq d_x h(w)$ for all $w \in T_x \mathcal{H}$, that is in particular

$$g_x(v, \log_x y) \leq d_x h(\log_x y) = d(x, y) d_x h(\gamma_{x,y}) \leq d(x, y) \frac{h(\gamma_{x,y}(t)) - h(x)}{t}$$

for $y \neq x$ and for all $t \in (0, d(x, y)]$. Taking $t = d(x, y)$, we get

$$g_x(v, \log_x y) \leq h(y) - h(x) \text{ for all } y \in \mathcal{H}$$

and so $v \in \partial h(x)$. □

We can now give the proof of Lemma 6.11. Note for this that if f is L^2 -Lipschitz with $L \in L^2(T, (0, \infty), \tau)$, integrating the respective Lipschitz condition

$$|f(t, x(t)) - f(t, y(t))| \leq L(t) d(x(t), y(t)) \text{ for all } t \in T$$

for $x, y \in L^2(T, \mathcal{H}, \tau)$ and using the Cauchy-Schwarz inequality yields that

$$|F(x) - F(y)| \leq \|L\|_2 d_2(x, y),$$

where $\|L\|_2$ is the L^2 -norm of L , so that F is $\|L\|_2$ -Lipschitz on $L^2(T, \mathcal{H}, \tau)$, and in particular $\text{dom} F = L^2(T, \mathcal{H}, \tau)$ if F is proper, i.e. if there exists a $u_0 \in L^2(T, \mathcal{H}, \tau)$ such that $f(\cdot, u_0(\cdot)) \in L^1(T, \mathbb{R}, \tau)$. In particular, $\underline{f}$ is $\|L\|_2$ -Lipschitz on $\mathcal{H}$. Note also that for $x \in \mathcal{H}$ and $v \in T_x \mathcal{H}$, we have $d_x F(v) = d_x \underline{f}(v)$. To see the last statement, note first that for $\gamma \in \Sigma'_x \mathcal{H}$, we have $\sigma_x \underline{f}(\gamma) = \sigma_x F(\gamma)$ simply because $\underline{f}$ and F coincide on $\mathcal{H}$. Then, for the unique extensions $d_x F$ of $\sigma_x F$ on $T_x L^2(T, \mathcal{H}, \tau)$ and $d_x \underline{f}$ of $\sigma_x \underline{f}$ on $T_x \mathcal{H}$ and $v = \lambda \gamma \in T_x \mathcal{H}$, we have

$$d_x F(v) = \lambda d_x F(\gamma) = \lambda \sigma_x F(\gamma) = \lambda \sigma_x \underline{f}(\gamma) = \lambda d_x \underline{f}(\gamma) = d_x \underline{f}(v),$$

using Lemma 6.12.

Proof of Lemma 6.11. If $v \in \partial F(x)$, then by Lemma 6.10, we have $v(t) \in \partial f(t, x)$ almost surely, that is

$$f(t, y) \geq f(t, x) + g_x(v(t), \log_x y) \text{ for all } y \in \mathcal{H}.$$

Integrating and using Lemma 2.10 yields $\int v d\tau \in \partial \underline{f}(x)$. For the converse, let $u \in \partial \underline{f}(x)$ and define the function $l_0 : T_x \mathcal{H} \rightarrow \mathbb{R}$ by $l_0(w) := g_x(u, w)$. Then Lemma 6.14 yields

$$l_0(w) = g_x(u, w) \leq d_x \underline{f}(w) = d_x F(w)$$

for any $w \in T_x \mathcal{H}$. Since $T_x \mathcal{H}$ is a Hilbert space, so is $T_x L^2(T, \mathcal{H}, \tau) = L^2(T, T_x \mathcal{H}, \tau)$ of which $T_x \mathcal{H}$ is a subspace of. By Lemma 6.12 and Corollary 6.13, we hence have that l_0 is a continuous linear functional on $T_x \mathcal{H}$ which is dominated by the continuous and sub-linear functional $d_x F : T_x L^2(T, \mathcal{H}, \tau) \rightarrow \mathbb{R}$. By the Hahn-Banach theorem, l_0 extends to continuous linear function $l : T_x L^2(T, \mathcal{H}, \tau) \rightarrow \mathbb{R}$ such that $l(w) \leq d_x F(w)$ for all $w \in T_x L^2(T, \mathcal{H}, \tau)$. By the Riesz representation theorem, there is a $v \in T_x L^2(T, \mathcal{H}, \tau)$ such that

$$l(w) = g_x(v, w) \text{ for all } w \in T_x L^2(T, \mathcal{H}, \tau).$$

Since we thus have $g_x(v, w) \leq d_x F(w)$ for all $w \in T_x L^2(T, \mathcal{H}, \tau)$, we get $v \in \partial F(x)$. Further, for any $w \in T_x \mathcal{H}$, we have

$$g_x(u, w) = l_0(w) = l(w) = g_x(v, w) = \int g_x(v(t), w) d\tau(t) = g_x\left(\int v d\tau, w\right)$$

so that $\int v d\tau = u$. $\square$

We now turn to the main result of this section, the converse to the above Lemma 6.1 over full Hilbert-Hadamard spaces.

Theorem 6.15. *Let $\mathcal{H}$ be a separable full Hilbert-Hadamard space and T be a complete and separable probability space. Let $f : T \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a proper normal convex integrand and suppose that f is L^2 -Lipschitz and that there exists a $u_0 \in L^2(T, \mathcal{H}, \tau)$ such that $f(\cdot, u_0(\cdot)) \in L^1(T, \mathbb{R}, \tau)$. Then, for any $x \in \mathcal{H}$:*

$$\int \partial f(t, x) d\tau(t) = \partial \left(\int f(t, \cdot) d\tau(t) \right) (x).$$

Proof. By Lemma 6.1, we only have to show the inclusion $\partial \underline{f}(x) \subseteq \underline{\partial f}(x)$. Let $u \in \partial \underline{f}(x)$. Hence, by Lemma 6.11, we have that there exists a $v \in \partial F(x)$ such that $u = \int v d\tau$. As $v \in \partial F(x)$, Lemma 6.10 yields that $v(t) \in \partial f(t, x)$ almost surely. Hence, we have $u \in \underline{\partial f}(x)$. $\square$

For the above Theorem 6.15, we in particular highlight the following natural question:

Question 6.16. Are the assumptions of Theorem 6.15 that f is L^2 -Lipschitz or that T is separable necessary?

7. A PROBABILISTIC LIE-TROTTER-KATO FORMULA

Combining Theorem 6.15 with Theorem 5.3, we in particular obtain the following result on full Hilbert-Hadamard spaces, in analogy to Theorem 1 in [108].

Theorem 7.1. *Let $(\Omega, \mathbb{F}, \mathbb{P})$ and $(E, \mathbb{E}, \mu)$ be probability spaces and let $\mathcal{H}$ be a separable full Hilbert-Hadamard space. Further, fix an i.i.d. sequence (ξ_i) of random variables $\xi_i : \Omega \rightarrow E$ with distribution μ . Let $A : E \rightarrow \mathcal{M}(\mathcal{H})$ be an integrable random monotone vector field such that $\frac{1}{n} \sum_{i=1}^n A(\xi_i)$ satisfies the surjectivity condition $\mathbb{P}$ -a.s. Then*

$$\frac{1}{n} \sum_{i=1}^n A(\xi_i) \xrightarrow{R} \int A(e) d\mu(e) \quad \mathbb{P}\text{-a.s.}$$

as $n \rightarrow \infty$.

In particular, suppose that $(E, \mathbb{E}, \mu)$ is complete and separable and $A = \partial f$ for a proper normal convex integrand $f : E \times \mathcal{H} \rightarrow (-\infty, +\infty]$ which is L^2 -Lipschitz and where there exists a $u_0 \in L^2(E, \mathcal{H}, \mu)$ such that $f(\cdot, u_0(\cdot)) \in L^1(E, \mathbb{R}, \mu)$. Then

$$\partial \left(\frac{1}{n} \sum_{i=1}^n f(\xi_i, \cdot) \right) \xrightarrow{R} \partial \left(\int f(e, \cdot) d\mu(e) \right) \quad \mathbb{P}\text{-a.s.}$$

as $n \rightarrow \infty$.

This latter conclusion can now be used to derive a probabilistic representation of the gradient flow associated with $\int f(e, \cdot) d\mu(e)$ which resembles the well-known Lie-Trotter-Kato formula.

Concretely, let $\mathcal{H}$ be a Hadamard space and let $f : \mathcal{H} \rightarrow (-\infty, +\infty]$ be a convex lsc function. Then, given $t > 0$, one can define the mapping $S_t : \text{dom } f \rightarrow \mathcal{H}$ by

$$S_t(x) := \lim_{k \rightarrow \infty} \left(\text{prox}_{\frac{t}{k} f} \right)^{(k)}(x),$$

the so-called exponential formula. The limit exists and is uniform in t on bounded subintervals of $(0, \infty)$ (see Theorem 1.13 in [79]), and the above map can be continuously extended to $t \in [0, \infty)$ by $S_0(x) := x$. Moreover, the collection of maps (S_t) is a nonexpansive semigroup, that is every S_t is nonexpansive and $S_{t+s} = S_t \circ S_s$. As such, the above definition and existence result provides a nonlinear extension of a similar result of Crandall and Liggett [40] set over Hilbert spaces, where this semigroup can also be characterized as the solution semigroup of the parabolic inclusion problem $-\dot{x} \in \partial f(x)$ which generalizes the gradient flow equation $\dot{x} = \nabla f(x)$ to non-smooth convex functions f (in fact, these results extend to general maximally monotone operators in place of ∂f). In particular, the nonlinear variant inherits many aspects of the associated theory of this inclusion, such as regularity of the flow or characterizations via an evolution variational inequality, and further extends to various other classes of spaces, making it a seminal tool in geometric analysis. We refer to [6, 21] for further background.

The first key result we need is that resolvent convergence actually implies the convergence of the gradient flow semigroup, as established by Bačák [22]:

Lemma 7.2 (Theorem 4.4 in [22]). *Let $\mathcal{H}$ be a Hadamard space and let $h_n : \mathcal{H} \rightarrow \mathbb{R}$ as well as $h : \mathcal{H} \rightarrow \mathbb{R}$ be convex and lsc. Let $S_t(x)$ be the gradient flow semigroup associated with h and $S_{n,t}(x)$ be the gradient flow semigroup associated with h_n . If $\text{prox}_{\lambda h_n}(x) \rightarrow \text{prox}_{\lambda h}(x)$ as $n \rightarrow \infty$ for all $\lambda > 0$ and $x \in \mathcal{H}$, then $S_{n,t}(x) \rightarrow S_t(x)$ as $n \rightarrow \infty$ for all $t \geq 0$ and $x \in \mathcal{H}$.*

Fundamentally, there is the following theorem of Stojković [111], of which an easier proof is given by Bačák in [19].

Lemma 7.3 (Theorem 4.4 in [111]). *Let $\mathcal{H}$ be a Hadamard space and let $h : \mathcal{H} \rightarrow \mathbb{R}$ be convex and lsc such that $h = \sum_{k=1}^N h_k$. Let $S_t(x)$ be the gradient flow semigroup associated with h . Then, for every $t \geq 0$ and all $x \in \mathcal{H}$:*

$$S_t(x) = \lim_{n \rightarrow \infty} \left(\text{prox}_{\frac{t}{n} h_N} \circ \cdots \circ \text{prox}_{\frac{t}{n} h_1} \right)^{(n)}(x).$$

If the functions $h_1, \dots, h_N$ are suitably Lipschitz, we can improve the above to the following uniform quantitative version, essentially by following (and at points quantitatively improving) the argument provided in [111]:

Lemma 7.4. *Let $\mathcal{H}$ be a Hadamard space and let $h : \mathcal{H} \rightarrow \mathbb{R}$ be convex such that $h = \sum_{k=1}^N h_k$ where h_i is L_i -Lipschitz. Let $S_t(x)$ be the gradient flow semigroup associated with h . Then, for every $t \geq 0$ and all $x \in \mathcal{H}$ as well as any $n \in \mathbb{N}$:*

$$d \left(S_t(x), \left(\text{prox}_{\frac{t}{n} h_N} \circ \cdots \circ \text{prox}_{\frac{t}{n} h_1} \right)^{(n)}(x) \right) \leq 8Ltn^{-1/4},$$

where $L := \sum_{k=1}^N L_k$.

Proof. Given $r > 0$, write $Q_r := \text{prox}_{rh_N} \circ \cdots \circ \text{prox}_{rh_1}$. Note that Q_r is nonexpansive. Consider the (λ, r) resolvent $J_{\lambda, r}$ associated with Q_r , defined as in Definition 3.2 in [111]. For each $\lambda, r > 0$, the map $J_{\lambda, r}$ is also nonexpansive (see Lemma 3.3 in [111]). Let $S_t^r(x)$ be the nonexpansive semigroup generated by $J_{\lambda, r}^r$ as in Theorem 3.10 in [111], and note that one has

$$d(S_t^r(x), Q_r^{(m)}(x)) \leq 2\sqrt{\left(m - \frac{t}{r}\right)^2 + \frac{t}{r}} d(x, Q_r(x))$$

for any $t, r > 0$, $x \in \mathcal{H}$ and $m \in \mathbb{N}$ (see Lemma 3.11 in [111]). Setting $r := t/m$, we get

$$d(S_t^{t/m}(x), Q_{t/m}^{(m)}(x)) \leq 2\sqrt{m}d(x, Q_{t/m}(x))$$

As in e.g. eq. (3.32) in [111] (see also [6]), we further have

$$d\left((\text{prox}_{\frac{t}{m}h})^{(m)}(x), S_t(x)\right) \leq \frac{t}{\sqrt{2m}}|\partial h|(x),$$

where, again, $|\partial h|(x)$ is the metric slope of h at x . As h is L -Lipschitz, we have $|\partial h|(x) \leq L$. Using Theorem 3.1.6 in [6], we get

$$d(x, \text{prox}_{rh_i}x) \leq r|\partial h_i|(x) \leq rL_i$$

and hence

$$d(x, Q_rx) \leq \sum_{k=1}^N d(\text{prox}_{rh_k} \circ \cdots \circ \text{prox}_{rh_1}(x), \text{prox}_{rh_{k-1}} \circ \cdots \circ \text{prox}_{rh_1}(x)) \leq \sum_{k=1}^N rL_k = rL.$$

Combined, we have

$$d(S_t^{t/m}x, Q_{t/m}^{(m)}x) \leq 2tL/\sqrt{m}.$$

Eq. (4.13) in [111] gives

$$\frac{1}{2\lambda}d^2(v, J_{\lambda,t}x) + \frac{1}{2\lambda}d^2(x, J_{\lambda,t}x) + \sum_{i=1}^N h_i(x_i) \leq h(v) + \frac{1}{2\lambda}d^2(x, v)$$

for $x_i := \text{prox}_{th_i}(x_{i-1})$ with $x_0 := J_{\lambda,t}x$. We get $d(x_i, x_{i-1}) \leq tL_i$ again using Theorem 3.1.6 in [6], so that $d(x_i, x_0) \leq t \sum_{k=1}^i L_k$ and hence

$$h(x_0) - \sum_{i=1}^N h_i(x_i) \leq \sum_{i=1}^N L_i d(x_i, x_0) \leq t \sum_{i=1}^N L_i \sum_{k=1}^i L_k \leq tL^2,$$

using that h_i is L_i -Lipschitz. Hence

$$\frac{1}{2\lambda}d^2(v, J_{\lambda,t}x) - tL^2 \leq \left(h(v) + \frac{1}{2\lambda}d^2(x, v)\right) - \left(h(J_{\lambda,t}x) + \frac{1}{2\lambda}d^2(x, J_{\lambda,t}x)\right).$$

For $v = \text{prox}_{\lambda h}x$, the right-hand side is nonpositive, so that

$$\frac{1}{2\lambda}d^2(\text{prox}_{\lambda h}x, J_{\lambda,t}x) \leq tL^2$$

and hence $d(J_{\lambda,t}x, \text{prox}_{\lambda h}x) \leq L\sqrt{2\lambda t}$. Using that both maps are nonexpansive, we get

$$d(J_{\lambda,t}^{(n)}x, (\text{prox}_{\lambda h})^{(n)}x) \leq nL\sqrt{2\lambda t}.$$

Lastly, note that by Theorem 3.10 in [111], we have

$$d(S_t^r(x), J_{t/n,r}^{(n)}x) \leq 2t \frac{1}{\sqrt{n}} \frac{d(x, Q_rx)}{r} \leq \frac{2tL}{\sqrt{n}}.$$

Combined, we get

$$\begin{aligned} d(S_t^r(x), S_t(x)) &\leq d(S_t^r(x), J_{t/k,r}^{(k)}x) + d(J_{t/k,r}^{(k)}x, \text{prox}_{t/kh}^{(k)}x) + d(\text{prox}_{t/kh}^{(k)}x, S_t(x)) \\ &\leq \frac{2tL}{\sqrt{k}} + L\sqrt{2trk} + \frac{t}{\sqrt{2k}}L. \end{aligned}$$

For $r = t/n$ and $k = \lceil \sqrt{n} \rceil$, we get

$$d(S_t^{t/n}(x), S_t(x)) \leq Lt \left(\frac{2}{n^{1/4}} + \frac{2\sqrt{2}}{\sqrt{n}} + \frac{1}{\sqrt{2}\sqrt{n}} \right) \leq 6Ltn^{-1/4}.$$

All together, we finally obtain

$$\begin{aligned} d(Q_{t/n}^{(n)}x, S_t(x)) &\leq d(Q_{t/n}^{(n)}x, S_t^{t/n}x) + d(S_t^{t/n}x, S_t(x)) \\ &\leq 2tLn^{-1/2} + 6Ltn^{-1/4} \leq 8Ltn^{-1/4}, \end{aligned}$$

as claimed. $\square$

Note that by the above Lemma 7.4, the convergence in the Lie-Trotter-Kato formula is in particular uniform in t on each bounded time interval in this case (compare Problem 4.10 in [23]).

Combined, we can now show the following:

Theorem 7.5. *Let $(\Omega, \mathbb{F}, \mathbb{P})$ and $(E, \mathbb{E}, \mu)$ be probability spaces, with the latter complete and separable, and let $\mathcal{H}$ be a separable full Hilbert-Hadamard space. Further, fix an i.i.d. sequence (ξ_i) of random variables $\xi_i : \Omega \rightarrow E$ with distribution μ . Let $f : E \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a proper normal convex integrand which is L^2 -Lipschitz and where there exists a $u_0 \in L^2(E, \mathcal{H}, \mu)$ such that $f(\cdot, u_0(\cdot)) \in L^1(E, \mathbb{R}, \mu)$. Write $S_t(x)$ for the gradient flow associated with $\int f(e, \cdot) d\mu(e)$. Then, there exists a set $\Omega' \subseteq \Omega$ with $\mathbb{P}(\Omega') = 1$ such that for any $\omega \in \Omega'$ as well as for any $t \geq 0$ and $x \in \mathcal{H}$:*

$$S_t(x) = \lim_{n \rightarrow \infty} \left(\text{prox}_{\frac{t}{n^2} f(\xi_n(\omega), \cdot)} \circ \cdots \circ \text{prox}_{\frac{t}{n^2} f(\xi_1(\omega), \cdot)} \right)^{(n)}(x).$$

In particular, the above limit holds almost surely.

Proof. Using Theorem 7.1, we get that there exists a single set Ω_0 of measure one such that

$$\text{prox}_{\lambda \frac{1}{n} \sum_{i=1}^n f(\xi_i(\omega), \cdot)}(x) \rightarrow \text{prox}_{\lambda \int f(e, \cdot) d\mu(e)}(x)$$

as $n \rightarrow \infty$, for all $\lambda > 0$ and $x \in \mathcal{H}$ as well as $\omega \in \Omega_0$. Let $S_{n,t}^\omega(x)$ be the gradient flow semigroup associated with $\frac{1}{n} \sum_{i=1}^n f(\xi_i(\omega), \cdot)$. Using Lemma 7.2, we hence get that

$$S_{n,t}^\omega(x) \rightarrow S_t(x)$$

as $n \rightarrow \infty$, for all $t \geq 0$ and $x \in \mathcal{H}$ as well as $\omega \in \Omega_0$. Now, using Lemma 7.4, we get

$$d \left(S_{n,t}^\omega(x), \left(\text{prox}_{\frac{t}{jn} f(\xi_n(\omega), \cdot)} \circ \cdots \circ \text{prox}_{\frac{t}{jn} f(\xi_1(\omega), \cdot)} \right)^{(j)}(x) \right) \leq 8tj^{-1/4} \left(\frac{1}{n} \sum_{i=1}^n L(\xi_i(\omega)) \right)$$

for any $j, n \in \mathbb{N}$, $t \geq 0$ and $x \in \mathcal{H}$ as well as $\omega \in \Omega_0$. As L is square-integrable and (ξ_i) is i.i.d., we get $\frac{1}{n} \sum_{i=1}^n L(\xi_i) \rightarrow \int L(e) d\mu(e)$ $\mathbb{P}$ -a.s. as $n \rightarrow \infty$ by the strong law of large numbers, say on a set Ω_1 of measure one. In particular, $L^s(\omega) := \sup_{n \in \mathbb{N}} \left(\frac{1}{n} \sum_{i=1}^n L(\xi_i(\omega)) \right) < +\infty$ for all $\omega \in \Omega_1$. In particular, we have

$$d \left(S_t(x), \left(\text{prox}_{\frac{t}{n^2} f(\xi_n(\omega), \cdot)} \circ \cdots \circ \text{prox}_{\frac{t}{n^2} f(\xi_1(\omega), \cdot)} \right)^{(n)}(x) \right) \leq d(S_t(x), S_{n,t}^\omega(x)) + 8tn^{-1/4} L^s(\omega)$$

for all $\omega \in \Omega_0 \cap \Omega_1$ as well as $t \geq 0$, $x \in \mathcal{H}$ and $n \in \mathbb{N}$, and so we get the result by letting $n \rightarrow \infty$. $\square$

As mentioned in the introduction, the use of Theorem 7.1 can be avoided in certain cases which moreover allow for the simplification of the Lipschitz assumption in the above Theorem 7.5.

At first, in the special case where $\mathcal{H}$ is a separable Hilbert space, we can employ a (consequence of a) strong law of large numbers for convex functions established by King and Wets [59].

Lemma 7.6 (consequence of Theorem 2.4 in [59]). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ and $(E, \mathcal{E}, \mu)$ be probability spaces, with the latter complete, and let $\mathcal{H}$ be a separable Hilbert space. Further, fix an i.i.d. sequence (ξ_i) of random variables $\xi_i : \Omega \rightarrow E$ with distribution μ . Let $f : E \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a normal convex integrand which is L^1 -Lipschitz and where there exists a $u_0 \in \mathcal{H}$ such that $f(\cdot, u_0) \in L^1(E, \mathbb{R}, \mu)$. Then*

$$\frac{1}{n} \sum_{i=1}^n f(\xi_i, \cdot) \xrightarrow{M} \int f(e, \cdot) d\mu(e) \quad \mathbb{P}\text{-a.s.}$$

as $n \rightarrow \infty$, with $\rightarrow^M$ meaning convergence in the sense of Mosco (see e.g. [21]).

Over Hilbert spaces, by a result of Attouch [13], Mosco convergence of convex functions is equivalent to the pointwise convergence of proximal maps. While it is not known whether this equivalence fully transfers to Hadamard spaces (see Problem 3.14 in [23]), one direction remains valid which will be the crucial one for the present paper.

Lemma 7.7 (Theorem 4.1 in [22]). *Let $(\mathcal{H}, d)$ be a Hadamard space and let $h_n : \mathcal{H} \rightarrow (-\infty, +\infty]$ be convex and lsc for each $n \in \mathbb{N}$. If $h_n \xrightarrow{M} h$ as $n \rightarrow \infty$, then $\text{prox}_{\lambda h_n}(x) \rightarrow \text{prox}_{\lambda h}(x)$ as $n \rightarrow \infty$ for all $\lambda > 0$ and $x \in \mathcal{H}$.*

This now allows us to establish the following variant of Theorem 7.5:

Theorem 7.8. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ and $(E, \mathcal{E}, \mu)$ be probability spaces, with the latter complete, and let $\mathcal{H}$ be a separable Hilbert space. Further, fix an i.i.d. sequence (ξ_i) of random variables $\xi_i : \Omega \rightarrow E$ with distribution μ . Let $f : E \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a proper normal convex integrand which is L^1 -Lipschitz and where there exists a $u_0 \in \mathcal{H}$ such that $f(\cdot, u_0) \in L^1(E, \mathbb{R}, \mu)$. Write $S_t(x)$ for the gradient flow associated with $\int f(e, \cdot) d\mu(e)$. Then, there exists a set $\Omega' \subseteq \Omega$ with $\mathbb{P}(\Omega') = 1$ such that for any $\omega \in \Omega'$ as well as for any $t \geq 0$ and $x \in \mathcal{H}$:*

$$S_t(x) = \lim_{n \rightarrow \infty} \left(\text{prox}_{\frac{t}{n^2} f(\xi_n(\omega), \cdot)} \circ \cdots \circ \text{prox}_{\frac{t}{n^2} f(\xi_1(\omega), \cdot)} \right)^{(n)}(x).$$

In particular, the above limit holds almost surely.

Proof. Proceed as in the proof of Theorem 7.5, using Lemmas 7.6 and 7.7 in place of Theorem 7.1 to derive that there exists a single set Ω_0 of measure one such that

$$\text{prox}_{\lambda \frac{1}{n} \sum_{i=1}^n f(\xi_i(\omega), \cdot)}(x) \rightarrow \text{prox}_{\lambda \int f(e, \cdot) d\mu(e)}(x)$$

as $n \rightarrow \infty$, for all $\lambda > 0$ and $x \in \mathcal{H}$ as well as $\omega \in \Omega_0$. From that point onwards, the proof proceeds similar, noting that it suffices that L is integrable to apply the strong law of large numbers. $\square$

The second case we consider is that $\mathcal{H}$ is a locally compact Hadamard space. In that case, we cannot only weaken the L^2 -Lipschitz assumption to L^1 -Lipschitz continuity but also do not have to require any additional assumptions on the tangent cones of $\mathcal{H}$. The result that makes this possible is the following (consequence of a) strong law of large numbers for convex functions due to Artstein and Wets [11]:

Lemma 7.9 (consequence of Theorem 2.3 in [11]). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ and $(E, \mathcal{E}, \mu)$ be probability spaces, with the latter complete, and let $\mathcal{H}$ be a separable Hadamard space. Further, fix an i.i.d. sequence (ξ_i) of random variables $\xi_i : \Omega \rightarrow E$ with distribution μ . Let $f : E \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a normal convex integrand which is L^1 -Lipschitz and where there exists a $u_0 \in \mathcal{H}$ such that*

$f(\cdot, u_0) \in L^1(E, \mathbb{R}, \mu)$. Then

$$\frac{1}{n} \sum_{i=1}^n f(\xi_i, \cdot) \rightarrow^\Gamma \int f(e, \cdot) d\mu(e) \quad \mathbb{P}\text{-a.s.}$$

as $n \rightarrow \infty$, with $\rightarrow^\Gamma$ meaning Γ -convergence in the sense of De Giorgi (see e.g. [21]).

The reason why we are restricted to locally compact Hadamard spaces when applying the above Lemma 7.9 is that it only establishes the weaker Γ -convergence, which is generally not enough to guarantee the convergence of the proximal maps (this already fails in infinite-dimensional Hilbert spaces, see e.g. Example 3.2 in [70]). However, over locally compact Hadamard spaces, weak and strong convergence coincide (see e.g. Example 3.2 in [76]), and so Γ - and Mosco-convergence also coincide. It is hence exactly outside of Hilbert spaces and locally compact Hadamard spaces where our Theorem 7.1 has its full use.

We now end this paper with the corresponding variant of Theorem 7.5 in the case of locally compact Hadamard spaces:

Theorem 7.10. *Let $(\Omega, \mathbb{F}, \mathbb{P})$ and $(E, \mathbb{E}, \mu)$ be probability spaces, with the latter complete, and let $\mathcal{H}$ be a locally compact Hadamard space. Further, fix an i.i.d. sequence (ξ_i) of random variables $\xi_i : \Omega \rightarrow E$ with distribution μ . Let $f : E \times \mathcal{H} \rightarrow (-\infty, +\infty]$ be a proper normal convex integrand which is L^1 -Lipschitz and where there exists a $u_0 \in \mathcal{H}$ such that $f(\cdot, u_0) \in L^1(E, \mathbb{R}, \mu)$. Write $S_t(x)$ for the gradient flow associated with $\int f(e, \cdot) d\mu(e)$. Then, there exists a set $\Omega' \subseteq \Omega$ with $\mathbb{P}(\Omega') = 1$ such that for any $\omega \in \Omega'$ as well as for any $t \geq 0$ and $x \in \mathcal{H}$:*

$$S_t(x) = \lim_{n \rightarrow \infty} \left(\text{prox}_{\frac{t}{n^2} f(\xi_n(\omega), \cdot)} \circ \cdots \circ \text{prox}_{\frac{t}{n^2} f(\xi_1(\omega), \cdot)} \right)^{(n)}(x).$$

In particular, the above limit holds almost surely.

Proof. Proceed as in the proof of Theorem 7.5 (as well as Theorem 7.8), now using Lemmas 7.9 and 7.7 in place of Theorem 7.1. $\square$

Acknowledgments: I want to thank Morenikeji Neri and Thomas Powell for helpful comments on a previous draft of this paper. The author interacted with OpenAI's GPT-5.6 while working on this paper, as detailed in the following:

- (1) Examples 3.7 and 6.2 were directly provided by GPT-5.6.
- (2) Example 2.2 as well as Lemmas 4.5, 4.13, 6.4, 6.8 and 6.14 were first obtained by the author alone, partially in weaker form, and then have been extended and streamlined (together with their proofs) after interacting with GPT-5.6. GPT-5.6 made a similar contribution to Lemma 4.10, where it in particular suggested a much easier proof of Lemma 4.9 which does not require the surjectivity of $\log$, as was used in a first version of the author. However, the author then found another proof alone, the present one, which is quite different to this proof suggested by GPT-5.6.
- (3) The approach to Corollary 3.5 and Theorem 5.3 was first developed by the author alone over Hadamard manifolds without a focus on quantitative aspects, and then later substantially simplified after interactions with GPT-5.6, not yet being aware of the work by Yokota [123]. The lift to Hadamard spaces was then carried out by the author alone, as was the quantitative analysis of this first purely qualitative proof which resulted in the non-asymptotic concentration inequalities provided in Theorem 3.4 and Remark 5.4. After learning of [123], the resulting exposition was restructured.
- (4) The greatest mathematical contribution of GPT-5.6 is the formulation and proof of Lemma 6.11. This result was first formulated (in a slightly different way) and proven

over Hadamard manifolds by GPT-5.6 after various interactions on the content of Section 6. In particular, GPT-5.6 suggested the approach to use Hahn-Banach to extend a linear functional on the tangent cone. The author then developed the lift of this result to full Hilbert-Hadamard spaces, together with GPT-5.6.

- (5) The outline of the approach to Lemma 7.4 was sketched first by GPT-5.6 after the author asked an adjacent question of removing a double limit in a first version of Theorem 7.5, obtained by the author alone. It was later refined by the author alone.

GPT-5.6 has also been used to find relevant related literature and typos. The author takes full accountability for the work. No parts of this paper were written or worded by an LLM.

REFERENCES

- [1] A.D. Aleksandrov. A theorem on triangles in a metric space and some of its applications. *Proceedings of the Steklov Institute of Mathematics*, 38:5–23, 1951.
- [2] A.D. Aleksandrov. Über eine Verallgemeinerung der Riemannschen Geometrie. *Schriftenreihe des Forschungsinstituts für Mathematik*, 1:33–84, 1957.
- [3] A.D. Aleksandrov, V.N. Berestovskii, and I.G. Nikolaev. Generalized Riemannian spaces. *Russian Mathematical Surveys*, 41(3):1–54, 1986.
- [4] S. Alexander, V. Kapovitch, and A. Petrunin. *An Invitation to Alexandrov Geometry: CAT(0) Spaces*. Springer Briefs in Mathematics. Springer Cham, 2019.
- [5] S. Alexander, V. Kapovitch, and A. Petrunin. *Alexandrov Geometry: Foundations*, volume 236 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2024.
- [6] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics. ETH Zürich. Birkhäuser Basel, 2008.
- [7] J. Antezana, E. Ghigloni, and D. Stojanoff. Ergodic theorem in CAT(0) spaces in terms of inductive means. *Ergodic Theory and Dynamical Systems*, 43:1433–1454, 2023.
- [8] B.C. Arnold. *Pareto Distributions*, volume 140 of *Monographs on Statistics and Applied Probability*. Chapman and Hall/CRC, New York, 2nd edition, 2015.
- [9] Z. Artstein and S. Hart. Law of large numbers for random sets and allocation processes. *Mathematics of Operations Research*, 6(4):485–492, 1981.
- [10] Z. Artstein and R.A. Vitale. A strong law of large numbers for random compact sets. *The Annals of Probability*, 3(5):879–882, 1975.
- [11] Z. Artstein and R.J.-B. Wets. Consistency of minimizers and the SLLN for stochastic programs. *Journal of Convex Analysis*, 2(1–2):1–17, 1995.
- [12] H. Attouch. Familles d’opérateurs maximaux monotones et mesurabilité. *Annali di Matematica Pura ed Applicata*, 120:35–111, 1979.
- [13] H. Attouch. *Variational Convergence for Functions and Operators*. Applicable Mathematics Series. Pitman Publishing Ltd, Boston, MA, 1984.
- [14] J.-P. Aubin and H. Frankowska. *Set-Valued Analysis*. Springer, New York, 2009.
- [15] R.J. Aumann. Integrals of set-valued functions. *Journal of Mathematical Analysis and Applications*, 12(1):1–12, 1965.
- [16] T. Austin. A CAT(0)-valued pointwise ergodic theorem. *Journal of Topology and Analysis*, 3(2):145–152, 2011.
- [17] L.E. Baum and M. Katz. Convergence rates in the law of large numbers. *Transactions of the American Mathematical Society*, 120(1):108–123, 1965.
- [18] H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer Cham, 2nd edition, 2017.
- [19] M. Bačák. A new proof of the Lie–Trotter–Kato formula in Hadamard spaces. *Communications in Contemporary Mathematics*, 16(6), 2014. Article no. 1350044, 15pp.
- [20] M. Bačák. Computing medians and means in Hadamard spaces. *SIAM Journal of Optimization*, 24(3):1542–1566, 2014.
- [21] M. Bačák. *Convex analysis and optimization in Hadamard spaces*, volume 22 of *De Gruyter Series in Nonlinear Analysis and Applications*. Walter de Gruyter GmbH, Berlin/Boston, 2014.
- [22] M. Bačák. Convergence of nonlinear semigroups under nonpositive curvature. *Transactions of the American Mathematical Society*, 367:3929–3953, 2015.

- [23] M. Bačák. Old and new challenges in Hadamard spaces. *Japanese Journal of Mathematics*, 18(2):115–168, 2023.
- [24] G.D.C. Bento, J.C. Neto, and Í.D.L. Melo. Fenchel conjugate via Busemann function on Hadamard manifolds. *Applied Mathematics & Optimization*, 88(3), 2023. Article no. 83, 29pp.
- [25] V.N. Berestovskii. Introduction of a Riemannian structure in certain metric spaces. *Siberian Mathematical Journal*, 16:499–507, 1975.
- [26] R. Bergmann, R. Herzog, M. Silva Louzeiro, D. Tenbrinck, and J. Vidal-Núñez. Fenchel duality theory and a primal-dual algorithm on Riemannian manifolds. *Foundations of Computational Mathematics*, 21(6):1465–1504, 2021.
- [27] P. Bianchi. Ergodic convergence of a stochastic proximal point algorithm. *SIAM Journal on Optimization*, 26(4):2235–2260, 2016.
- [28] P. Bianchi and W. Hachem. Dynamical behavior of a stochastic forward–backward algorithm using random monotone operators. *Journal of Optimization Theory and Applications*, 171(1):90–120, 2016.
- [29] P. Bianchi, W. Hachem, and A. Salim. A Constant Step Forward-Backward Algorithm Involving Random Maximal Monotone Operators. *Journal of Convex Analysis*, 26(2):397–436, 2019.
- [30] L.J. Billera, S.P. Holmes, and K. Vogtmann. Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics*, 27(4):733–767, 2001.
- [31] J.M. Bismut. Intégrales convexes et probabilités. *Journal of Mathematical Analysis and Applications*, 42(3):639–673, 1973.
- [32] M.R. Bridson and A. Haefliger. *Metric Spaces of Non-Positive Curvature*, volume 319 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin, Heidelberg, 1999.
- [33] F. Bruhat and J. Tits. Groupes réductifs sur un corps local. I. Données radicielles valuées. *Publications Mathématiques de l’Institut des Hautes Études Scientifiques*, 41:5–251, 1972.
- [34] C. Castaing and P. Raynaud de Fitte. Law of large numbers and ergodic theorem for convex weak star compact valued Gelfand-integrable mappings. In S. Kusuoka and T. Maruyama, editors, *Advances in Mathematical Economics*, volume 17, pages 1–37. Springer, Berlin, 2013.
- [35] C. Castaing and M. Valadier. *Convex Analysis and Measurable Multifunctions*. Lecture Notes in Mathematics. Springer Berlin, Heidelberg, 1977.
- [36] P. Chaipunya, F. Kohsaka, and P. Kumam. Monotone vector fields and generation of nonexpansive semi-groups in complete CAT(0) spaces. *Numerical Functional Analysis and Optimization*, 42(9):989–1018, 2021.
- [37] B.J. Choi, J. Heo, and U.C. Ji. Law of large numbers for weighted inductive means in a Hadamard space. *Mathematische Nachrichten*, 290(17–18):2775–2787, 2017.
- [38] B.J. Choi and J.H. Kim. Birkhoff’s ergodic theorems in terms of weighted inductive means. *Journal of the Korean Mathematical Society*, 59(3):549–570, 2022.
- [39] K.L. Chung. The strong law of large numbers. In J. Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 341–353. University of California Press, 1951.
- [40] M.G. Crandall and T.M. Liggett. Generation of semi-groups of nonlinear transformations on general Banach spaces. *American Journal of Mathematics*, 93(2):265–298, 1971.
- [41] A. Es-Sahib and H. Heinich. Barycentre canonique pour un espace métrique à courbure négative. *Séminaire de Probabilités*, 33:355–370, 1999.
- [42] N. Etemadi. An elementary proof of the strong law of large numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 55(1):119–122, 1981.
- [43] K. Funano. Rate of convergence of stochastic processes with values in $\mathbb{R}$ -trees and Hadamard manifolds. *Osaka Journal of Mathematics*, 47(4):911–920, 2010.
- [44] N. Gigli and F. Nobile. A differential perspective on gradient flows on CAT(κ)-spaces and applications. *The Journal of Geometric Analysis*, 31(12):11780–11818, 2021.
- [45] S. Gong, J. Wu, and G. Yu. The Novikov conjecture, the group of volume preserving diffeomorphisms and Hilbert-Hadamard spaces. *Geometric and Functional Analysis*, 31:206–267, 2021.
- [46] A. Goodwin, A.S. Lewis, G. López-Acedo, and A. Nicolae. Recognizing Weighted Means in Geodesic Spaces. *Foundations of Computational Mathematics*, 2025. In press.
- [47] T. Le Gouic, Q. Paris, P. Rigollet, and A.J. Stromme. Fast convergence of empirical barycenters in Alexandrov spaces and the Wasserstein space. *Journal of the European Mathematical Society*, 25(6):2229–2250, 2022.

- [48] M. Gromov. Hyperbolic groups. In S.M. Gersten, editor, *Essays in group theory*, volume 8 of *Mathematical Sciences Research Institute Publications*, pages 75–263. Springer, New York, 1987.
- [49] M. Gromov. *Geometric Group Theory. Volume 2: Asymptotic Invariants of Infinite Groups*, volume 182 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1993.
- [50] H. Hirai. Convex analysis on Hadamard spaces and scaling problems. *Foundations of Computational Mathematics*, 24(6):1979–2016, 2024.
- [51] J.-B. Hiriart-Urruty. About properties of the mean value functional and of the continuous infimal convolution in stochastic convex analysis. In J. Cea, editor, *Optimization Techniques Modeling and Optimization in the Service of Man Part 2*, volume 41 of *Lecture Notes in Computer Science*, pages 763–789. Springer-Verlag, Berlin, 1976.
- [52] A.D. Ioffe and V.M. Tikhomirov. On minimization of integral functionals. *Functional Analysis and Its Applications*, 3(3):218–227, 1969.
- [53] J. Jost. Equilibrium maps between metric spaces. *Calculus of Variations and Partial Differential*, 2:173–204, 1994.
- [54] J. Jost. Convex functionals and generalized harmonic maps into spaces of nonpositive curvature. *Commentarii Mathematici Helvetici*, 70:659–673, 1995.
- [55] J. Jost. *Nonpositive curvature: geometric and analytic aspects*. Lectures in Mathematics. ETH Zürich. Birkhäuser Basel, 1997.
- [56] A.G. Kachurovskii. The rate of convergence in ergodic theorems. *Russian Mathematical Surveys*, 51(4):653–703, 1996.
- [57] T. Kato and K. Masuda. Trotter’s product formula for nonlinear semigroups generated by the subdifferentials of convex functionals. *Journal of the Mathematical Society of Japan*, 30(1):169–178, 1978.
- [58] H. Khatibzadeh and S. Ranjbar. Monotone operators and the proximal point algorithm in complete CAT(0) metric spaces. *Journal of the Australian Mathematical Society*, 103(1):70–90, 2017.
- [59] A.J. King and R.J.-B. Wets. Epi-consistency of convex stochastic programs. *Stochastics and Stochastic Reports*, 34(1–2):83–92, 1991.
- [60] B. Kleiner and B. Leeb. Rigidity of quasi-isometries for symmetric spaces and Euclidean buildings. *Publications Mathématiques de l’Institut des Hautes Études Scientifiques*, 86:115–197, 1997.
- [61] A. Klenke. *Probability Theory: A Comprehensive Course*. Universitext. Springer Cham, 3rd edition, 2020.
- [62] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer-Verlag Berlin Heidelberg, 2008.
- [63] U. Kohlenbach. Proof-theoretic Methods in Nonlinear Analysis. In B. Sirakov, P. Ney de Souza, and M. Viana, editors, *Proceedings ICM 2018*, volume 2, pages 61–82. World Scientific, Singapore, 2019.
- [64] A.N. Kolmogorov. Sur la loi forte des grands nombres. *Comptes rendus de l’Académie des Sciences*, 191:910–912, 1930.
- [65] G. Köstenberger and T. Stark. Robust signal recovery in Hadamard spaces. *Journal of Multivariate Analysis*, 210, 2025. Article no. 105469, 18pp.
- [66] U. Krengel. *Ergodic theorems*, volume 6 of *De Gruyter Studies in Mathematics*. Walter de Gruyter GmbH, Berlin/Boston, 2011.
- [67] T.G. Kurtz. A random Trotter product formula. *Proceedings of the American Mathematical Society*, 35(1):147–154, 1972.
- [68] Z. Léka and M. Pálfi. Strong law of large numbers for generalized operator means. *Advances in Mathematics*, 457, 2024. Article no. 109933, 38pp.
- [69] V.L. Levin. Subdifferentials of convex functionals. *Uspekhi Matematicheskikh Nauk*, 25(4(154)):183–184, 1970.
- [70] A.B. Levy, R.A. Poliquin, and L. Thibault. Partial extensions of Attouch’s theorem with applications to proto-derivatives of subgradient mappings. *Transactions of the American Mathematical Society*, 347(4):1269–1294, 1995.
- [71] A.S. Lewis, G. López-Acedo, and A. Nicolae. Basic convex analysis in metric spaces with bounded curvature. *SIAM Journal on Optimization*, 34(1):366–388, 2024.
- [72] C. Li, G. López, and V. Martín-Márquez. Monotone vector fields and the proximal point algorithm on Hadamard manifolds. *Journal of the London Mathematical Society*, 79(3):663–683, 2009.
- [73] C. Li, G. López, V. Martín-Márquez, and J.H. Wang. Resolvents of set-valued monotone vector fields in Hadamard manifolds. *Set-Valued and Variational Analysis*, 19(3):361–383, 2011.
- [74] Y. Lim and M. Pálfi. Strong law of large numbers for the L^1 -Karcher mean. *Journal of Functional Analysis*, 279(7), 2020. Article no. 108672, 37pp.

- [75] A. Lytchak. Differentiation in metric spaces. *Algebra i Analiz*, 16(6):128–161, 2004.
- [76] A. Lytchak and A. Petrunin. Weak topology on $\text{CAT}(0)$ spaces. *Israel Journal of Mathematics*, 255(2):763–781, 2023.
- [77] J. Marcinkiewicz and A. Zygmund. Sur les fonctions indépendantes. *Fundamenta Mathematicae*, 29:60–90, 1937.
- [78] S. Di Marino, N. Gigli, E. Pasqualetto, and E. Soultanis. Infinitesimal Hilbertianity of locally $\text{CAT}(\kappa)$ -spaces. *The Journal of Geometric Analysis*, 31(8):7621–7685, 2021.
- [79] U. Mayer. Gradient flows on nonpositively curved metric spaces and harmonic maps. *Communications in Analysis and Geometry*, 6:199–253, 1998.
- [80] G.J. Minty. Monotone (nonlinear) operators in Hilbert spaces. *Duke Mathematical Journal*, 29:341–346, 1962.
- [81] M. Movahedi, D. Behmardi, and S. Hosseini. On the density theorem for the subdifferential of convex functions on Hadamard spaces. *Pacific Journal of Mathematics*, 276(2):437–447, 2015.
- [82] A. Navas. An L^1 ergodic theorem with values in a non-positively curved space via a canonical barycenter map. *Ergodic Theory and Dynamical Systems*, 33:609–623, 2013.
- [83] S.Z. Németh. Monotone vector fields. *Publicationes Mathematicae Debrecen*, 54:437–449, 1999.
- [84] M. Neri. A finitary Kronecker’s lemma and large deviations in the Strong Law of Large numbers on Banach spaces. *Annals of Pure and Applied Logic*, 176(6), 2025. Article no. 103569, 31pp.
- [85] M. Neri. Quantitative Strong Laws of Large Numbers. *Electronic Journal of Probability*, 30, 2025. Article no. 20, 22pp.
- [86] M. Neri. Oscillations in the Strong Law of Large Numbers. *Proceedings of the American Mathematical Society*, 2026. In press.
- [87] M. Neri, P. Oliva, and N. Pischke. A systematic way of analysing proofs in probability theory, 2026. Preprint, <https://arxiv.org/abs/2604.08078>.
- [88] M. Neri and N. Pischke. Proof mining and probability theory. *Forum of Mathematics, Sigma*, 13, 2025. Article no. e187, 47pp.
- [89] J.X. Da Cruz Neto, O.P. Ferreira, and L.R. Lucambio Pérez. Monotone point-to-set vector fields. *Balkan Journal of Geometry and Its Applications*, 5:69–79, 2000.
- [90] T. Nguyen and H. Pham. A maximal theorem of Hardy-Littlewood type for pairwise i.i.d. random variables and the law of large numbers. *Theory of Probability & Its Applications*, 66(3):445–454, 2021.
- [91] I.G. Nikolaev. The tangent cone of an Aleksandrov space of curvature $\leq K$. *Manuscripta Mathematica*, 86(1):137–147, 1995.
- [92] S.I. Ohta. Barycenters and a law of large numbers in Gromov hyperbolic spaces. *Revista Matemática Iberoamericana*, 40(3):1185–1206, 2024.
- [93] S.I. Ohta and M. Pálfi. Discrete-time gradient flows and law of large numbers in Alexandrov spaces. *Calculus of Variations and Partial Differential Equations*, 54(2):1591–1610, 2015.
- [94] S.I. Ohta and M. Pálfi. Gradient flows and a Trotter-Kato formula of semi-convex functions on $\text{CAT}(1)$ -spaces. *American Journal of Mathematics*, 139(4):937–965, 2017.
- [95] A. Petrunin. Semiconcave functions in Alexandrov’s geometry. In J. Cheeger and K. Grove, editors, *Metric and comparison geometry*, volume 11 of *Surveys in Differential Geometry*, pages 137–201. International Press, Somerville, 2007.
- [96] N. Pischke. Logical Metatheorems for Accretive and (Generalized) Monotone Set-Valued Operators. *Journal of Mathematical Logic*, 24(2), 2024. Article no. 2350008, 59pp.
- [97] N. Pischke. Mean-square and sublinear convergence of a stochastic proximal point algorithm in metric spaces of nonpositive curvature, 2026. Preprint, <https://arxiv.org/abs/2510.10697>.
- [98] N. Pischke and T. Powell. Convergence guarantees for stochastic algorithms solving non-unique problems in metric spaces, 2026. Preprint, <https://arxiv.org/abs/2605.06129>.
- [99] C. Plaut. Metric spaces of curvature $\geq k$. In R.J. Daverman and R.B. Sher, editors, *Handbook of Geometric Topology*, pages 819–898. North-Holland, Amsterdam, 2001.
- [100] K. Podczek. On the convexity and compactness of the integral of a Banach space valued correspondence. *Journal of Mathematical Economics*, 44(7–8):836–852, 2008.
- [101] P. Révész. *The laws of large numbers*, volume 4 of *Probability and Mathematical Statistics*. Academic Press, New York, London, 1968.
- [102] R.T. Rockafellar. Integral functionals, normal integrands and measurable selections. In J.P. Gossez, E.J. Lami Dozo, J. Mawhin, and L. Waelbroeck, editors, *Nonlinear Operators and the Calculus of Variations*, volume 543 of *Lecture Notes in Mathematics*, pages 157–207. Springer, Berlin, Heidelberg, 1976.

- [103] R.T. Rockafellar and R.J.-B. Wets. On the interchange of subdifferentiation and conditional expectation for convex functionals. *Stochastics*, 7(3):173–182, 1982.
- [104] R.T. Rockafellar and R.J.B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin, Heidelberg, 1998.
- [105] A. Rosalsky and L.V. Thành. The Laws of Large Numbers. In M. Lovric, editor, *International Encyclopedia of Statistical Science*, pages 1323–1328. Springer, Berlin, Heidelberg, 2025.
- [106] J. Ruf and I. Waudby-Smith. Concentration inequalities for strong laws and laws of the iterated logarithm, 2025. Preprint, <https://arxiv.org/abs/2511.00175>.
- [107] T. Sakai. *Riemannian Geometry*, volume 149 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, 1996.
- [108] A. Salim. A Strong Law of Large Numbers for Random Monotone Operators. *Set-Valued and Variational Analysis*, 31, 2023. 38, 13pp.
- [109] A. Shapiro and H. Xu. Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions. *Journal of Mathematical Analysis and Applications*, 325(2):1390–1399, 2007.
- [110] M. Silva Louzeiro, R. Bergmann, and R. Herzog. Fenchel duality and a separation theorem on Hadamard manifolds. *SIAM Journal on Optimization*, 32(2):854–873, 2022.
- [111] I. Stojković. Approximation for convex functionals on non-positively curved spaces and the Trotter–Kato product formula. *Advances in Calculus of Variations*, 5:77–126, 2012.
- [112] K.-T. Sturm. Nonlinear Markov operators associated with symmetric Markov kernels and energy minimizing maps between singular spaces. *Calculus of Variations and Partial Differential Equations*, 12:317–357, 2001.
- [113] K.-T. Sturm. Nonlinear martingale theory for processes with values in metric spaces of nonpositive curvature. *The Annals of Probability*, 30(3):1195–1222, 2002.
- [114] K.-T. Sturm. Probability measures on metric spaces of nonpositive curvature. In P. Auscher, T. Coulhon, and A. Grigoryan, editors, *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*, volume 338 of *Contemporary Mathematics*, pages 357–390. American Mathematical Society, Providence, RI, 2003.
- [115] K.T. Sturm. Nonlinear Markov operators, discrete heat flow, and harmonic maps between singular spaces. *Potential Analysis*, 16:305–340, 2002.
- [116] R.L. Taylor and H. Inoue. Laws of Large Numbers for Random Sets. In J. Goutsias, R.P.S. Mahler, and H.T. Nguyen, editors, *Random Sets. Theory and Applications*, volume 97 of *The IMA Volumes in Mathematics and its Applications*, pages 347–360. Springer, New York, 1997.
- [117] P. Terán. On a uniform law of large numbers for random sets and subdifferentials of random functions. *Statistics & Probability Letters*, 78(1):42–49, 2008.
- [118] P. Terán and I. Molchanov. The law of large numbers in a metric space with a convex combination operation. *Journal of Theoretical Probability*, 19(4):875–898, 2006.
- [119] L. Thibault. Espérances conditionnelles d’intégrandes semi-continues. *Annales de l’institut Henri Poincaré. Section B. Calcul des probabilités et statistiques*, 17(4):337–350, 1981.
- [120] M. Valadier. Intégration de convexes fermés notamment d’épigraphe inf-convolution continue. *Revue française d’informatique et de recherche opérationnelle. Série rouge*, 4(R2):57–73, 1970.
- [121] J.H. Wang, G. López, V. Martín-Márquez, and C. Li. Monotone and accretive vector fields on Riemannian manifolds. *Journal of Optimization Theory and Applications*, 146(3):691–708, 2010.
- [122] N.C. Yannelis. On the upper and lower semicontinuity of the Aumann integral. *Journal of Mathematical Economics*, 19(4):373–389, 1990.
- [123] T. Yokota. Law of large numbers in CAT(1)-spaces of small radii. *Calculus of Variations and Partial Differential Equations*, 57(2), 2018. Article no. 35, 28pp.