Mag.rer.nat. Oliver EBNER

# Stochastic aspects of refinement schemes on metric spaces

DISSERTATION

zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften

Doktoratsstudium der Naturwissenschaften im Rahmen der Doktoratsschule "Mathematik und Wissenschaftliches Rechnen"

TU Graz

Graz University of Technology

Technische Universität Graz

Betreuer:
Univ.-Prof. Dipl.-Ing. Mag.rer.nat. Dr.techn. Johannes WALLNER

Institut für Geometrie

Graz, im Mai 2012

# Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am ........................ ..........................................

(Unterschrift)

# Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitely marked all material which has been quoted either literally or by content from the used sources.

................................. ..........................................

date (signature)

# Danksagung

Ich bin vielen Menschen und der einen oder anderen Institution Dank schuldig, die zur Entstehung dieser Dissertation beigetragen haben. Zunächst möchte ich meinem Betreuer Johannes Wallner für die Bereitstellung eines fruchtbaren Themas und seine Offenheit für die fachliche Diskussion danken. Karl-Theodor Sturm danke ich für die Gastfreundschaft während meiner Aufenthalte in Regen und Bonn und manch nützliche Anregung. Für ihre wertvollen Bemerkungen bedanke ich mich bei meinen Zweitgutachtern Philipp Grohs und Kurt Jetter. Auch Wolfgang Woess, von dessen Vorlesungen an der TU Graz ich stark profitiert habe, bin ich für sein reges Interesse an meiner Arbeit Dank schuldig.

Diese Arbeit wurde im Rahmen des Doktoratskollegs "Diskrete Mathematik" (Projekt W1230 des österreichischen Fonds zur Förderung wissenschaftlicher Forschung FWF) verfasst, welches zusätzlich zu finanzieller Unterstützung ein extrem reichhaltiges wissenschaftliches Umfeld zur Verfügung stellte. Weiters möchte ich dankend die Unterstützung durch das FWF-Projekt P19870 erwähnen.

Für die angenehme menschliche und professionelle Atmosphäre am Arbeitsplatz bedanke ich mich bei meinen Kollegen vom Institut für Geometrie. Insbesondere Andreas Weinmann verdanke ich die eine oder andere sehr produktive Unterhaltung.

Abschließend gilt mein ganz besonderer Dank meiner Familie, insbesondere meinen Eltern und Aywana, für ihre fortwährende Unterstützung. Diese Arbeit ist ihnen gewidmet.

# Contents

# 1 Introduction

The convergence and smoothness analysis of refinement schemes processing data from manifolds and more generally metric spaces has been a subject of intense research over the last few years, see for example [22, 11, 27, 25]. As to convergence, complete spaces of nonpositive curvature, also known as Hadamard or global NPC spaces, have proven most accessible in terms of generalizing well-known facts from the linear theory to the nonlinear setting. An example of such a structure prominent in applications is the space of positive definite symmetric matrices, which represent measurements in diffusion tensor imaging.

While the question whether the *smoothness* properties of the linear model scheme prevail when passing to the nonlinear setting was successfully addressed in [12], the corresponding *convergence* problem remained unsolved. One aim of the present thesis is to fill this gap in the theory by presenting the author's recent results from [9, 8], augmented by yet unpublished material.

Relying on a martingale theory for discrete-time stochastic processes with values in negatively curved spaces developed in [19], we observe that the refinement processes in question actually act on bounded input data as nonlinear Markov semigroups. This fact substantially facilitates their convergence analysis.

Let us specify the general setup. Given a metric space $(X, d)$, a *refinement scheme* is a map $S : \ell^\infty(\mathbb{Z}^s, X) \to \ell^\infty(\mathbb{Z}^s, X)$. We call $S$ *convergent* if for all $x \in \ell^\infty(\mathbb{Z}^s, X)$ there exists a continuous function $S^\infty x : \mathbb{R}^s \to X$ such that

$$d_\infty(S^\infty x(\cdot/2^n), S^n x) = \sup_j d(S^\infty x(j/2^n), S^n x_j) \to 0 \quad \text{as } n \to \infty. \tag{1}$$

Visualizing $S^n x$ as a function on the refined grid $2^{-n}\mathbb{Z}^s$, convergence to $S^\infty x$ is tantamount to $d_\infty(S^n x, f|_{2^{-n}\mathbb{Z}^s}) \to 0$.

Throughout the present work we are mostly concerned with so-called *barycentric* refinement schemes associated to nonnegative real-valued $s$-variate sequences $(a_i)_{i\in\mathbb{Z}^s}$ of finite support, henceforth referred to as *masks*, which we require to fulfill the *basic sum rule*

$$\sum_{j\in\mathbb{Z}^s} a_{i-2j} = 1 \quad \text{for } i \in \mathbb{Z}^s. \tag{2}$$

Barycentric refinement schemes act on data $x \in \ell^\infty(\mathbb{Z}^s, X)$ from a complete metric space of nonpositive curvature in the sense of A. D. Alexandrov according to the following rule:

$$Sx_i = \operatorname{argmin}\left(\sum_{j\in\mathbb{Z}^s} a_{i-2j} d^2(x_j, \,\cdot\,)\right). \tag{3}$$

Much is known about the convergence of these type of refinement algorithms in the case $X = \mathbb{R}$. On complete, simply connected manifolds of nonpositive sectional curvature convergence analysis was initiated in the article [23]. The author in [9] recently proved general convergence statements for arbitrary Hadamard spaces using the principle of *contractivity*: A scheme is called contractive with respect to some nonnegative function $D : \ell^\infty(\mathbb{Z}^s, X) \to \mathbb{R}_+$ if and only if there is $\gamma < 1$ such that

$$D(Sx) < \gamma D(x), \quad \text{for all} \quad x \in \ell^\infty(\mathbb{Z}^s, X).$$

The function $D$ is referred to as a *contractivity function* for $S$. An important class of contractivity functions is associated to balanced, convex and bounded subsets $\Omega$ of $\mathbb{R}^s$:

$$D_\Omega(x) = \sup_{\rho(i-j)<2} d(x_i, x_j), \tag{4}$$

where $\rho$ denotes the Minkowski functional of $\Omega$. Contractivity functions of this type are called *admissible*, cf. [9]. The following result is taken from loc. cit.:

**Proposition 1.1.** *A barycentric refinement scheme with nonnegative mask which is contractive with respect to some admissible contractivity function also converges. This implies convergence in case the support of the mask coincides with the set of lattice points within a centered unimodular zonotope or a lattice quad with nonempty interior.*

A major result of the present thesis, taken from the author's recent article [8], is a substantial extension of this statement and describes a phenomenon which could be referred to as *linear equivalence*:

**Theorem 1.2.** *A barycentric refinement scheme converges on arbitrary Hadamard spaces if and only if it converges on the real line.*

The proof of this fact, given in Section 4.2, relies on a stochastic interpretation of the subdivision rule (3). More precisely, for each nonnegative mask $\mathbf{a} = (a_i)_{i\in\mathbb{Z}^s}$ satisfying the basic sum rule (2) one finds a so-called *characteristic* Markov chain $X_n^{\mathbf{a}}$

with state space $\mathbb{Z}^s$ and transition matrix $(a_{i-2j})_{i,j\in\mathbb{Z}^s}$ in terms of which the iterates of the refinement algorithm acting on $x \in \ell^\infty(\mathbb{Z}^s, X)$ may be written as

$$S^n x_i = E(x \circ X_n^{\mathbf{a}} |||X_0^{\mathbf{a}} = i),$$

see Theorem 4.4. Here $E(\,\cdot\,|||X_0)$ denotes the *filtered conditional expectation* introduced by K.-T. Sturm in [19]. Thus, as in the linear case,

$$\begin{cases} \mathbb{N}_0 \to \mathrm{Lip}_1(\ell^\infty(\mathbb{Z}^s, X)); \\ n \mapsto S^n \end{cases}$$

may be considered a (nonlinear) Markov semigroup. Here $\mathrm{Lip}_1(\ell^\infty(\mathbb{Z}^s, X))$ refers to the set of maps $T : \ell^\infty(\mathbb{Z}^s, X) \to \ell^\infty(\mathbb{Z}^s, X)$ satisfying the Lipschitz condition

$$d_\infty(Tx, Ty) \le d_\infty(x, y) \quad \text{for } x, y \in \ell^\infty(\mathbb{Z}^s, X). \tag{5}$$

Combining Theorem 1.2 with other recent developments in the theory of linear subdivision schemes with nonnegative masks and their barycentric counterparts on nonlinear objects, one comes up with a variety of remarkable results:

In the articles [28] and [29], X. Zhou establishes general theorems on the relation of the mask's support with its convergence properties, which, utilizing Theorem 1.2 now generalize to the following:

**Theorem 1.3.** *A barycentric subdivision scheme* $S : \ell^\infty(\mathbb{Z}^s, X) \to \ell^\infty(\mathbb{Z}^s, X)$ *with nonnegative mask converges under each of the following circumstances:*

*(i) The support of* $(a_i)_{i\in\mathbb{Z}^s}$ *coincides with the set of grid points inside a balanced zonotope.*

*(ii) The grid dimension* $s = 1$ *and, if, after a possible index translation,* $(a_i)_{i\in\mathbb{Z}} = (\dots, 0, 0, a_0, \dots, a_N, 0, 0, \dots)$*, the integers within the support are relatively prime and* $0 < a_0, a_1 < 1$*. This also constitutes a necessary condition for convergence.*

Moreover, as far as finite-dimensional Hadamard manifolds are concerned, the smoothness question is settled by a combination of Theorem 1.2 and recent work from [12]:

**Theorem 1.4.** *On a smooth Hadamard* manifold*, a barycentric subdivision scheme* $S : \ell^\infty(\mathbb{Z}^s, X) \to \ell^\infty(\mathbb{Z}^s, X)$ *with nonnegative mask converges and produces* $r$*-times differentiable limit functions if and only if the same is true for the corresponding linear scheme.*

The thesis is organized as follows: In Chapter 2 we give an account of foundational facts on the geometry and probabilistic features of metric spaces, our main references being [3, 19, 20]. Beginning with a discussion on geodesics and convexity, we proceed by introducing the central geometrical object of interest: complete, simply connected metric spaces of nonpositive curvature. These structures, introduced above as Hadamard- or global NPC-spaces, are characterized by geodesic triangles being 'slim' when compared to Euclidean triangles of the same edge lengths. Intuitively, this definition stems from the well-known fact that the defect in the angle sum of a geodesic triangle on a surface can be represented in terms of the curvature. Actually, in case the metric originates from a Riemannian structure on the underlying space, 'slimness' of triangles is tantamount to nonpositive sectional curvature. Hadamard spaces are particularly convenient in their overall tendency to 'contract' - for example, the center of mass of any probability distribution lies in the convex hull of the support of the distribution. The mere well-definedness of expectations and conditional expectations for random variables with values in Hadamard spaces addressed in Section 2.3 is already a major benefit. Above that, it turns out that nonpositive curvature also allows for efficient comparison of nonlinear and linear expectations in terms of a Jensen inequality presented in Theorem 2.35. This estimate, well known in the linear case, constitutes the pivotal feature allowing for convergence analysis of barycentric subdivision schemes in Section 4.2.

Succeeding the discussion of the geometric and stochastic fundamentals, Chapter 3 develops the very basic facts on Markov chains in the linear setting, with [26] as a main reference. As described above, the basic sum rule (2) allows for the interpretation of $(a_{i-2j})_{i,j\in\mathbb{Z}^s}$ as a row stochastic matrix. Theorem 3.6 shows how, in general, such a matrix gives rise to a time-discrete stochastic process inducing a Markov semigroup. A nonlinear version of this fact is postponed to Section 4.2.

Finally, Chapter 4 discusses the author's recent progress in the convergence analysis of barycentric refinement rules on Hadamard spaces as described above, cf. [8, 9]. After a brief introduction into the portions of the linear theory essential to our studies, it is made clear how the probabilistic and geometric concepts presented in Chapters 2 and 3 pave the way for the study of nonlinear subdivision. As mentioned above, the central observation of this chapter consists in the interpretation of a barycentric scheme as the *nonlinear* Markov semigroup associated to the transition kernel induced by the scheme's mask, see Theorem 4.4. Following the proof of Theorem 1.2, given in section 4.2, we generalize some well-known results concerning approximation order and characterization of convergence from the linear theory to the Hadamard setting. Moreover, it is shown how a strong law of large numbers leads to certain structure-preservation properties of barycentric schemes on the space of diffusion tensors, see Corollary 4.27. A concluding section addresses the relationship between the convergence properties of a scheme and its so-called characteristic Markov chain.

# 2 Probability and Stochastics on metric spaces

This chapter gives an outline of certain geometric and probabilistic properties of metric spaces. The center of interest is occupied by spaces of nonnegative curvature in the sense of Alexandrov. These structures have been intensely studied over the last and present century, see [1, 2, 3], with some recent breakthroughs on their probabilistic features, cf. [19, 20]. The material contained in this chapter is taken from the available literature, in particular [3], [19] and [20].

## 2.1 Hadamard spaces

Curvature bounds for metric spaces not necessarily endowed with the structure of a Riemannian manifold are formulated in terms of the comparison of geodesic triangles within the metric space to triangles in a comparison space of constant curvature, that is, spheres, Euclidean and hyperbolic spaces. Generally, geodesic curves are naturally defined in an object admitting a distance measure. The following paragraphs focus on the issue of existence and uniqueness of shortest paths.

### 2.1.1 Measuring length in metric spaces

Albeit the a priori absence of a sensible analogon of 'inscribed polygonal curve', the following definition comes in the spirit of the Euclidean one:

**Definition 2.1.** Suppose $(X, d)$ is a metric space, and $c : [a, b] \to X$ is a continuous curve. Then the *arc length* of $c$ is defined as

$$\ell(c) := \sup\{\sum_{i=0}^{n-1} d(c(t_i), c(t_{i+1})) \mid a \leq t_0 < \cdots < t_n \leq b\}.$$

A curve $c$ is called *rectifiable* if and only if it has finite arc length.

Having a measure of arc length at hand paves the way for an obvious definition of geodesic curves:

**Definition 2.2.** Let $(X, d)$ be a metric space. Then $X$ is called a *length space* if and only if distances in $X$ are infima of arc lengths of rectifiable curves in the sense that whenever $x_0, x_1 \in X$,

$$d(x_0, x_1) = \inf\{\ell(c) \mid c \in C([0, 1], X), \quad c(0) = x_0, \quad c(1) = x_1\}.$$

Moreover, $X$ is called *geodesic* if and only if for all $x_0, x_1 \in X$ we have

$$d(x_0, x_1) = \min\{\ell(c) \mid c \in C([0,1], X), \quad c(0) = x_0, \quad c(1) = x_1\},$$

i.e., for any two points $x_0, x_1$ there exists a *joining geodesic* $c$ satisfying $\ell(c) = d(x_0, x_1)$. Whenever two points in $X$ are joined by a *unique* geodesic, $X$ is referred to as *strongly geodesic.*

The next lemma states the easy fact that geodesic curves remain geodesic on any subinterval of their domain.

**Lemma 2.3.** *Suppose $(X, d)$ is geodesic. Choose $x_0, x_1 \in X$, and let $c : [0, b] \to X$ be a joining geodesic. Then for each $0 \leq t \leq b$,*

$$\ell(c|_{[0,t]}) = d(c(0), c(t)). \tag{6}$$

*Proof.* Assume $\ell(c|_{[0,t]}) > d(c(0), c(t))$. Choose a geodesic $\tilde{c} : [0, t] \to X$ joining $x_0$ and $c(t)$. Define

$$\bar{c}(s) = \begin{cases} \tilde{c}(s), & \text{for } s \in [0, t] \\ c(s) & \text{else.} \end{cases}$$

Then $\bar{c}$ is continuous, and

$$\begin{aligned} \ell(\bar{c}) &= \ell(\bar{c}|_{[0,t]}) + \ell(\bar{c}|_{[t,b]}) \\ &= \ell(\tilde{c}) + \ell(c|_{[t,b]}) \\ &= d(c(0), c(t)) + \ell(c|_{[t,b]}) \\ &< \ell(c|_{[0,t]}) + \ell(c|_{[t,b]}) \\ &= \ell(c), \end{aligned}$$

a contradiction. □

### 2.1.2 Arc length parametrizations

A crucial technical tool in the geometry of curves in Riemannian manifolds is the existence of an arc length parametrization in case of regularity. The following paragraph introduces an analogous construction for geodesics in metric spaces using the concept of generalized inverses.

**Definition 2.4.** Suppose $J, K \subset \mathbb{R}$ are intervals and $F : J \to K$ is a monotonously increasing, right-continuous, surjective function. Then its *generalized inverse* is defined as the map

$$F^{\leftarrow} : K \to J; y \mapsto \inf\{x \in J \mid F(x) \geq y\}.$$

The following fact is well-known and can be found in any treatise on probability, see e.g. [4].

**Lemma 2.5.** *Suppose $J, K \subset \mathbb{R}$ are intervals and $F : J \to K$ is monotonously increasing, right-continuous, and onto. Then*

$$F \circ F^{\leftarrow} = \mathrm{id}_K .$$

**Proposition 2.6.** *Suppose $\bar{c} : [0, b] \to X$ is a geodesic joining $x_0, x_1 \in X$. Then there is a reparametrization $c : [0, d(x_0, x_1)] \to X$ of $\bar{c}$ such that for all $0 \leq t \leq d(x_0, x_1)$,*

$$\ell(c|_{[0,t]}) = d(c(0), c(t)) = t.$$

*Proof.* Consider the continuous function

$$\lambda : [0, b] \to [0, d(x_0, x_1)]; \quad t \mapsto \ell(\bar{c}|_{[0,t]}).$$

and set $\gamma = \lambda^{\leftarrow} : [0, d(x_0, x_1)] \to [0, b]$. Then $\lambda \circ \gamma = \mathrm{id}_{[0,d(x_0,x_1)]}$ by Lemma 2.5. Thus, setting $c := \bar{c} \circ \gamma$, for $0 \leq s \leq t \leq d(x_0, x_1)$ we have the equalities

$$\begin{aligned} d(c(s), c(t)) &= d(\bar{c}(\gamma(s)), \bar{c}(\gamma(t))) \\ &= \ell(\bar{c}|_{[\gamma(s),\gamma(t)]}) \\ &= \ell(\bar{c}|_{[0,\gamma(t)]}) - \ell(c|_{[0,\gamma(s)]}) \\ &= \lambda(\gamma(t)) - \lambda(\gamma(s)) \\ &= t - s. \end{aligned}$$

In particular, $c$ is continuous and $\ell(c|_{[0,t]}) = d(c(0), c(t)) = t$ □

**Definition 2.7.** Let $(X, d)$ be a metric space. A geodesic $c : [0, b] \to X$ satisfying

$$d(c(0), c(t)) = t$$

is said to be *parametrized by arc length.* In case

$$d(c(0), c(t)) = \mathrm{const} \cdot t,$$

$c$ is said to be *parametrized proportional to arc length.*

**Notation.** Given two points $x_0, x_1 \in X$ in a strongly geodesic space $X$, we will write $x_t$ for the value of the geodesic joining $x_0$ and $x_1$ and parametrized proportional to arc length, at parameter value $t \in [0,1]$.

### 2.1.3 Midpoints

In general, checking existence and uniqueness of geodesics in a general metric space seems a formidable task. It appears a lot more accessible to check for every pair of points to possess a possibly unique midpoint. In the following we prove that these problems are equivalent.

**Definition 2.8.** Suppose $(X, d)$ is a metric space and $x_0, x_1 \in X$. Then $z$ is a *midpoint of $x_0$ and $x_1$* if and only if

$$d(x_0, x_{1/2}) = d(x_{1/2}, x_1) = \frac{1}{2} d(x_0, x_1).$$

The set of midpoints of $x_0, x_1$ is written $\mathrm{mpt}(x_0, x_1)$. Moreover, an *($\varepsilon$-)approximative midpoint* of $x_0, x_1 \in X$ is any $z \in X$ satisfying

$$\max(d(x_0, z), d(z, x_1)) \leq \frac{1}{2} d(x_0, x_1) + \varepsilon.$$

The set of $\varepsilon$-approximative midpoints of $x_0$ and $x_1$ is denoted by $\mathrm{mpt}_\varepsilon(x_0, x_1)$.

**Theorem 2.9.** *Let $(X, d)$ be a complete metric space. Then the following hold true:*

*(i) $X$ is a length space if and only if for any two points $x_0, x_1 \in X$ and for all $\varepsilon > 0$, the set $\mathrm{mpt}_\varepsilon(x_0, x_1)$ is nonempty.*

*(ii) The space $X$ is (strongly) geodesic if and only if each two points in $X$ possess a (unique) midpoint.*

*Proof.* We present a proof of the first statement. The second one is proven along the same lines, with some obvious modifications.

We begin by proving $\Longleftarrow$: For $\varepsilon > 0$. Set $c_\varepsilon(0) = x_0$, $c_\varepsilon(1) = x_1$, and for $n \geq 1$ recursively define

$$c_\varepsilon((2k+1)/2^n) \in \mathrm{mpt}_{2^{-2n}\varepsilon}(c_\varepsilon(k/2^{n-1}), c_\varepsilon((k+1)/2^{n-1})) \neq \emptyset$$

for $0 \leq k < 2^{n-1} - 1$.

Set $\Delta = d(x_0, x_1)$. We claim that for all $n \geq 0$ and $0 \leq i \leq 2^n$ it holds that

$$d(c_\varepsilon(i/2^n), c_\varepsilon((i+1)/2^n)) \leq \frac{\Delta + (1 - 2^{-n})\varepsilon}{2^n}.$$

Indeed, under the hypothesis that this holds, and since we may without loss of generality assume $i = 2k$, it follows that

$$\begin{aligned} d(c_\varepsilon(i/2^{n+1}), c_\varepsilon((i+1)/2^{n+1})) &= d(c_\varepsilon(k/2^n), c_\varepsilon((2k+1)/2^{n+1})) \\ &\leq \frac{1}{2} d(c_\varepsilon(k/2^n), c_\varepsilon((k+1)/2^n)) + 2^{-(2n+2)}\varepsilon \\ &\leq \frac{1}{2}\left(\frac{\Delta + (1 - 2^{-n})\varepsilon}{2^n}\right) + 2^{-(2n+2)}\varepsilon \\ &= \frac{\Delta + (1 - 2^{-n})\varepsilon + 2^{-(n+1)}\varepsilon}{2^{n+1}} \\ &= \frac{\Delta + (1 - 2^{-(n+1)})\varepsilon}{2^{n+1}}, \end{aligned}$$

whence the claim follows using induction over $n$. Consequently, for $0 \leq i, j \leq 2^n$,

$$d(c_\varepsilon(i/2^n), c_\varepsilon(j/2^n)) \leq (\Delta + \varepsilon) \cdot \frac{|i - j|}{2^n}.$$

This implies that a dyadic Lipschitz condition holds true: For $s = i/2^n$ and $t = j/2^m$, where $0 \leq i \leq 2^n$, $0 \leq j \leq 2^m$ and $m > n$, we have

$$\begin{aligned} d(c_\varepsilon(s), c_\varepsilon(t)) &= d(c_\varepsilon(i/2^n), c_\varepsilon(j/2^m)) \\ &= d(c_\varepsilon(2^{m-n}i/2^m), c_\varepsilon(j/2^m)) \\ &\leq (\Delta + \varepsilon) \cdot \frac{|i - 2^{m-n}j|}{2^n} \\ &= (\Delta + \varepsilon) \cdot |s - t|. \end{aligned} \tag{7}$$

Since $(X, d)$ is complete, this dyadic Lipschitz-continuity implies that $c_\varepsilon$ may be continuously extended from the dyadic numbers to a continuous map $[0, 1] \to X$, which we again denote by $c_\varepsilon$. Certainly, equation (7) remains valid for $s, t \in [0, 1]$, so for any partition $0 = t_0 < \cdots < t_n = 1$ of the unit interval we have

$$\begin{aligned} \sum_{i=0}^{n-1} d(c_\varepsilon(t_i), c_\varepsilon(t_{i+1})) &\leq \sum_{i=0}^{n-1} (\Delta + \varepsilon) \cdot (t_{i+1} - t_i) \\ &= \Delta + \varepsilon = d(x_0, x_1) + \varepsilon. \end{aligned}$$

Consequently $0 \leq \ell(c_\varepsilon) - d(x_0, x_1) \leq \varepsilon \to 0$ as $\varepsilon \downarrow 0$.

Now for $\Longrightarrow$: For $x_0, x_1 \in X$ choose a continuous curve $c_\varepsilon : [0,1] \to X$ joining $x_0$ and $x_1$ such that $\ell(c_\varepsilon) \leq d(x_0, x_1) + \varepsilon$. Certainly, $t \mapsto \ell(t) = \ell(c_\varepsilon|_{[0,t]})$ is a continuous map, so by the intermediate value theorem there is $0 \leq t_0 \leq 1$ such that $\ell(t_0) = \frac{1}{2}d(x_0, x_1)$. Thus, $d(x_0, c(t_0)) \leq \ell(t_0) = \frac{1}{2}d(x_0, x_1)$. On the other hand,

$$\begin{aligned} d(c(t_0), x_1) &\leq \ell(c|_{[t_0,1]}) \\ &\leq d(x_0, x_1) + \varepsilon - \ell(t_0) \\ &= \frac{1}{2}d(x_0, x_1) + \varepsilon, \end{aligned}$$

whence $c(t_0) \in \mathrm{mpt}_\varepsilon(x_0, x_1)$. □

**Corollary 2.10.** *Let $(X, d)$ be a complete metric space. Then $X$ is a length space if and only if for any $x_0, x_1 \in X$ and $\varepsilon > 0$ there exists $z \in X$ such that*

$$d^2(x_0, z) + d^2(z, x_1) \leq \frac{1}{2}d^2(x_0, x_1) + \varepsilon.$$

*Proof.* $\Longleftarrow$:In view of Theorem 2.9, it suffices to show that for any two $x_0, x_1 \in X$ and arbitrary $\varepsilon > 0$, the set $\mathrm{mpt}_\varepsilon(x_0, x_1) \neq \emptyset$. Note, however, that for $d(x_0, x_1) = u > 0$, the Taylor expansion of $f(\delta) = \sqrt{\frac{1}{2}u^2 + \delta}$ gives

$$f(\delta) = \frac{1}{\sqrt{2}}u + \frac{\delta}{\sqrt{2}u} + O(\delta^2).$$

Thus, there exists $M > 0$ such that for $\delta < \left(\frac{1-\sqrt{2}}{2}\right)u^2 = C(u)$,

$$f(\delta) \leq \frac{1}{2}u + M\delta^2.$$

In particular, given $\varepsilon > 0$, there is $\delta < C(u)$ such that $M\delta^2 < \varepsilon$. Now choose $z \in X$ such that

$$d^2(x_0, z) + d^2(z, x_1) \leq \frac{1}{2}d^2(x_0, x_1) + \delta.$$

Consequently,

$$\begin{aligned} \max(d(x_0, z), d(z, x_1)) &\leq \sqrt{d^2(x_0, z) + d^2(z, x_1)} \\ &\leq f(u) \leq \frac{1}{2}d(x_0, x_1) + \varepsilon. \end{aligned}$$

$\Longrightarrow$: Given $\varepsilon > 0$, choose $\delta > 0$ such that $\delta^2 + \delta d(x_0, x_1) < \varepsilon$. Then if $z \in \mathrm{mpt}_\delta(x_0, x_1)$, is it plain to see that

$$d^2(x_0, z) + d^2(z, x_1) \leq \frac{1}{2}d^2(x_0, x_1) + \varepsilon.$$

□

### 2.1.4 The Hadamard property

**Definition 2.11.** Suppose $(X, d)$ is a metric space. Then $X$ is called *Hadamard space* or *global NPC (Non-Positive Curvature)-space* if $X$ is complete and for all $x_0, x_1 \in X$ there is a $y \in X$ such that for all $z \in X$ the so-called *Hadamard inequality* holds true:

$$d^2(z, y) \leq \frac{1}{2} d^2(z, x_0) + \frac{1}{2} d^2(z, x_1) - \frac{1}{4} d^2(x_0, x_1). \tag{8}$$

*Remark* 2.12. As Proposition 2.14 will reveal, the point $y \in X$ in (8) is the unique midpoint $x_{\frac{1}{2}}$ of $x_0$ and $x_1$. Hence we may rewrite (8) as

$$d^2(z, x_{\frac{1}{2}}) \leq \frac{1}{2} d^2(z, x_0) + \frac{1}{2} d^2(z, x_1) - \frac{1}{4} d^2(x_0, x_1),$$

an inequality easily seen to be sharp in the Euclidean case. Thus the Hadamard inequality represents a concise way to describe 'slimness' of geodesic triangles, see Figure 2.12.

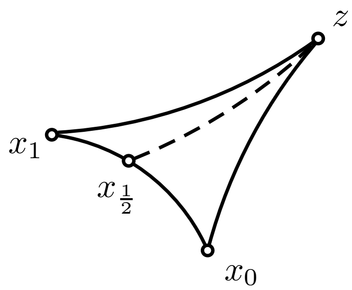


Figure 1: A 'slim' geodesic triangle.

Hadamard spaces for instance play an important role in the theory of cost- minimizing networks, see [7]. Topological examples are trees as well as Euclidean Bruhat-Tits buildings. Notably, for a measure space $M$, and $N$ Hadamard, the space of strongly measurable square-integrable functions $L^2(M, N)$ inherits the Hadamard property. It is remarkable that these spaces also occur as families of certain geometric and topological structures, such as spaces of Riemannian and Kähler metrics or spaces of connections. The latter examples actually are generically infinite-dimensional Hadamard *manifolds*, see [15]. In the smooth case the Hadamard property is equivalent to nonpositive sectional curvature and simple connectedness. An instance of a finite-dimensional Hadamard manifold significant in applications is the space of symmetric positive definite matrices, which occurs in Diffusion Tensor Imaging.

**Lemma 2.13.** *Suppose $(X,d)$ is a complete metric space. Then $X$ is Hadamard if and only if for all $x_0, x_1 \in X$ and $\varepsilon > 0$ there is $y \in X$ such that for all $z \in X$ the* approximate Hadamard inequality *holds true:*

$$d^2(z,y) \le \frac{1}{2}d^2(z,x_0) + \frac{1}{2}d^2(z,x_1) - \frac{1}{4}d^2(x_0,x_1) + \varepsilon. \tag{9}$$

*Proof.* Suppose $\varepsilon_n$ is a zero sequence. Choose $y_n \in X$ such that inequality (9) holds true for all $z \in X$, with $\varepsilon = \varepsilon_n$. Then by substituting $x_0$ and $x_1$ for $z$ in the approximate Hadamard inequality we obtain:

$$\max(d^2(y_n,x_0), d^2(y_n,x_1)) \le \frac{1}{4}d^2(x_0,x_1) + \varepsilon_n,$$

and

$$\begin{aligned} d^2(y_n,y_m) &\le \frac{1}{2}d^2(y_n,x_0) + \frac{1}{2}d^2(y_n,x_1) - \frac{1}{4}d^2(x_0,x_1) + \varepsilon_m \\ &\le \varepsilon_n + \varepsilon_m. \end{aligned}$$

Thus $y_n \to y$ as $n \to \infty$, and obviously for all $z \in X$,

$$d^2(z,y) \le \frac{1}{2}d^2(z,x_0) + \frac{1}{2}d^2(z,x_1) - \frac{1}{4}d^2(x_0,x_1).$$

□

**Proposition 2.14.** *Let $(X,d)$ be Hadamard. Then $(X,d)$ is strongly geodesic.*

*Proof.* We first show that $X$ is geodesic. Since $X$ is complete, by Theorem 2.9 it suffices to show that for each $x_0, x_1$ the set of midpoints $\mathrm{mpt}(x_0,x_1)$ is nonempty.

By the Hadamard inequality (8), there is $y \in X$ such that

$$\begin{aligned} d^2(x_0,y) &\le \frac{1}{4}d^2(x_0,x_1) \\ d^2(x_1,y) &\le \frac{1}{4}d^2(x_0,x_1). \end{aligned}$$

Hence $d(x_0,y) \le \frac{1}{2}d(x_0,x_1)$ and $d(x_1,y) \le \frac{1}{2}d(x_0,x_1)$. Thus, by the triangle inequality,

$$\begin{aligned} d(x_0,x_1) &\le d(x_0,y) + d(x_1,y) \\ &\le \frac{1}{2}d(x_0,x_1) + \frac{1}{2}d(x_0,x_1) = d(x_0,x_1). \end{aligned}$$

It follows that $y \in \mathrm{mpt}(x_0,x_1)$.

To show that $X$ is *strongly* geodesic, it suffices to show $\#\,\mathrm{mpt}(x_0, x_1) = 1$ for all $x_0, x_1 \in X$. Suppose $x_{1/2}, \tilde{x}_{1/2} \in \mathrm{mpt}(x_0, x_1)$. By the Hadamard inequality, there is $y \in X$ such that

$$\begin{aligned} d^2(y, x_{1/2}) &\leq \frac{1}{2} d^2(x_{1/2}, x_0) + \frac{1}{2} d^2(x_{1/2}, x_1) - \frac{1}{4} d^2(x_0, x_1) \\ &= \frac{1}{8} d^2(x_0, x_1) + \frac{1}{8} d^2(x_0, x_1) - \frac{1}{4} d^2(x_0, x_1) = 0, \end{aligned}$$

and the same holds true for $\tilde{x}_{1/2}$. Thus, $x_{1/2} = y = \tilde{x}_{1/2}$. □

**Definition 2.15.** Suppose $(X, d)$ is a strongly geodesic metric space. Then a function $\Phi : X \to \mathbb{R}$ is called *convex* if and only if for each geodesic $x_t$ it holds that

$$\Phi(x_t) \leq (1-t)\Phi(x_0) + t\Phi(x_1). \tag{10}$$

Moreover, $\Phi$ is called *strongly convex* if and only if

$$\Phi(x_t) \leq (1-t)\Phi(x_0) + t\Phi(x_1) - t(1-t)d^2(x_0, x_1). \tag{11}$$

**Proposition 2.16.** *Let $(X, d)$ be a strongly geodesic, complete metric space, and suppose $\Phi : X \to \mathbb{R}$ is strongly convex and continuous. Then there is a unique $x^*$ such that*

$$\Phi(x^*) = \min \Phi(X).$$

*Proof.* Choose a sequence $x_n \in X$ with $\lim_n \Phi(x_n) = \min(\Phi(X)) = \alpha$. Let $x_{nm} = \mathrm{mpt}(x_n, x_m)$. Then $\Phi(x_{nm}) \geq \alpha$ together with strong convexity implies

$$\begin{aligned} \varphi(x_{nm}) &\leq \frac{1}{2}\left(\Phi(x_n) + \Phi(x_m)\right) - \frac{1}{4} d^2(x_0, x_1) \\ \Longrightarrow \quad d^2(x_0, x_1) &= 2\left(\Phi(x_n) + \Phi(x_m)\right) - 4\Phi(x_{nm}) \\ &\leq 2\left(\Phi(x_n) + \Phi(x_m)\right) - 4\alpha, \end{aligned}$$

meaning that $x_n$ is a Cauchy sequence. Therefore, $x_n \to z_0$, and by continuity, $\Phi(z_0) = \lim_n \Phi(x_n) = \alpha$.

Assume now $\Phi(z_1) = \alpha$. Then by strong convexity again,

$$\alpha \leq \Phi(z_{1/2}) \leq \alpha - \frac{1}{4} d^2(z_0, z_1),$$

hence $d(z_0, z_1) \leq 0$.

In conclusion, $x^* = z_0$ is the unique minimizer of $\Phi$. □

**Proposition 2.17.** *Suppose* $(X, d)$ *is a strongly geodesic, complete metric space, and let* $\Phi : X \to \mathbb{R}$ *be a strongly convex function. Moreover let* $x^* = \operatorname{argmin}(\Phi)$*. Then for all* $z \in X$*,*

$$\Phi(x^*) \leq \Phi(z) - d^2(z, x^*).$$

*Proof.* Let $z_t$ denote the geodesic joining $z$ and $x^*$. Then since $\Phi(x^*) \leq \Phi(z_t)$ for $0 \leq t \leq 1$, employing the strong convexity of $\Phi$ we obtain for all $0 \leq t < 1$:

$$\begin{aligned} & \Phi(x^*) \leq \Phi(z_t) \leq (1-t)\Phi(z) + t\Phi(x^*) - t(1-t)d^2(z, x^*) \\ \Longrightarrow \quad & (1-t)\Phi(x^*) \leq (1-t)\Phi(z) - t(1-t)d^2(z, x^*) \\ \Longrightarrow \quad & \Phi(x^*) \leq \Phi(z) - td^2(z, x^*), \end{aligned}$$

hence the statement follows from taking the limit as $t \uparrow 1$. □

**Proposition 2.18** (Strong Hadamard inequality)**.** *Suppose* $(X, d)$ *is a Hadamard space, and let* $x_0, x_1, z \in X$*. Then*

$$d^2(z, x_t) \leq (1-t)d^2(z, x_0) + td^2(z, x_1) - (1-t)td^2(x_0, x_1). \tag{12}$$

*In other words, the squared distance function* $d^2(z, \cdot\,)$ *is strongly convex.*

*Proof.* Obviously, it suffices to show the statement for dyadic $t$, which is done via induction. The induction step essentially involves two computations: The first one is elementary and gives the result

$$\begin{aligned} \left(1 - \frac{2k+1}{2^n}\right)\frac{2k+1}{2^n} = \frac{1}{2}\left(\left(1 - \frac{k}{2^{n-1}}\right)\frac{k}{2^{n-1}} + \left(1 - \frac{k+1}{2^{n-1}}\right)\frac{k+1}{2^{n-1}}\right) \\ + \frac{1}{4} \cdot \frac{1}{2^{n-1}}, \end{aligned}$$

while the second one invokes an iterated application of the Hadamard inequality. Let

$t = (2k+1)/2^n$. Then

$$\begin{aligned}
d^2(x_t, z) &\le \frac{1}{2} d^2(z, x_{k/2^{n-1}}) + \frac{1}{2} d^2(z, x_{(k+1)/2^{n-1}}) \\
&- \frac{1}{4} d^2(x_{k/2^{n-1}}, x_{(k+1)/2^{n-1}}) \\
&\le \frac{1}{2}\left(\left(1 - \frac{k}{2^{n-1}}\right) + \left(1 - \frac{k+1}{2^{n-1}}\right)\right) d^2(z, x_0) \\
&+ \frac{1}{2}\left(\frac{k}{2^{n-1}} + \frac{k+1}{2^{n-1}}\right) d^2(z, x_1) \\
&- \frac{1}{2}\left(\left(1 - \frac{k}{2^{n-1}}\right)\frac{k}{2^{n-1}} + \left(1 - \frac{k+1}{2^{n-1}}\right)\frac{k+1}{2^{n-1}}\right) d^2(x_0, x_1) \\
&- \frac{1}{4} \cdot \frac{1}{2^{n-1}} d^2(x_0, x_1) \\
&= (1-t) d^2(z, x_0) + t d^2(z, x_1) - (1-t) t d^2(x_0, x_1).
\end{aligned}$$

□

## 2.2 Probability measures on Hadamard spaces

We describe how nonpositive curvature leads to convenient probabilistic features of Hadamard spaces. Most of the material appearing in this section, which could be regarded a prologue to the discussion of conditional expectations in Section 2.3, is taken from [20].

**Definition 2.19.** Suppose $(X, d)$ is a metric space, and let $\mu$ a probability measure on $X$. Then the *variance* of $\mu$ is defined as

$$\operatorname{Var}(\mu) := \inf_{z \in X} \int_X d^2(z, x) \mu(dx).$$

We call $\mu$ an *$L^2$-probability measure* if and only if $\operatorname{Var}(\mu) < \infty$.

More generally, $\mu$ is called *$L^p$-probability measure* if for one (and then all) $x_0 \in X$ it holds that

$$\int_X d^p(x_0, x) \mu(dx) < \infty.$$

The space of $L^p$-probability measures on $X$ is denoted by $\mathcal{P}^p(X)$.

**Theorem 2.20** (Barycenters)**.** *Let $(X, d)$ be a Hadamard space, and suppose $\mu$ is an $L^2$-probability measure on $X$. Then there exists a unique $b(\mu) \in X$ such that*

$$\operatorname{Var}(\mu) = \int_X d^2(b(\mu), x) \mu(dx).$$

*This point $b(\mu)$ is referred to as* barycenter *or* center of mass *of $\mu$.*

*Moreover, the so-called* variance inequality *holds true: For each* $z \in X$*,*

$$\int_X d^2(b(\mu), x)\mu(dx) \leq \int_X d^2(z, x)\mu(dx) - d^2(z, b(\mu)).$$

*Proof.* Applying the strong Hadamard inequality (12) one sees that the function

$$\Phi(z) = \int_X d^2(z, x)\mu(dx)$$

is strongly convex. Hence the statement follows from Propositions 2.16 and 2.17. □

*Remark* 2.21. Centers of mass defined as minimizers of convex functionals have been considered since the seminal article [5], and due to the influential works [14, 10] are sometimes referred to as *Karcher means* or *Fréchet means.*

**Theorem 2.22.** *Let* $(X, d)$ *be a complete metric space. Then the following are equivalent:*

*(i)* $X$ *is a Hadamard space.*

*(ii) For any* $L^2$*-probability measure* $\mu$ *there exists* $z_\mu \in X$ *such that for all* $y \in X$*,*

$$\int_X d^2(z_\mu, x)\mu(dx) \leq \int_X d^2(y, x)\mu(dx) - d^2(y, z_\mu). \tag{13}$$

*(iii) Every probability measure* $\mu$ *on* $X$ *obeys the inequality*

$$\mathrm{Var}(\mu) \leq \frac{1}{2}\int_X \int_X d^2(x, y)\mu(dx)\mu(dy).$$

*(iv)* $(X, d)$ *is a length space and for arbitrary* $x_0, x_1, x_2, x_3 \in X$ *and* $0 \leq s, t \leq 1$

$$\begin{aligned} s(1-s)d^2(x_0, x_2) + t(1-t)d^2(x_1, x_3) \leq{}& std^2(x_0, x_1) + (1-s)td^2(x_1, x_2) \\ &+ (1-s)(1-t)d^2(x_2, x_3) \\ &+ s(1-t)d^2(x_3, x_0). \end{aligned}$$

*Proof.* <u>*(i)* $\Longrightarrow$ *(ii)*</u> is Theorem 2.20.

<u>*(ii)* $\Longrightarrow$ *(iii)*</u>: It suffices to consider the case $\mathrm{Var}(\mu) < \infty$. In this case, by hypothesis there is $z_\mu \in X$ such that (13) holds true, and thus, taking integrals with respect

to $\mu$ on both sides of this inequality,

$$\begin{aligned}\mathrm{Var}(\mu) &= \int_X d^2(z_\mu, x)\mu(dx)\\ &\leq \int_X\int_X d^2(y,x)\mu(dx)\mu(dy) - \int_X d^2(y,z_\mu)\mu(dy)\\ &= \int_X\int_X d^2(y,x)\mu(dx)\mu(dy) - \mathrm{Var}(\mu).\end{aligned}$$

<u>$(iii) \implies (iv)$</u>: Note first that the elementary equality $(\alpha a-(1-\alpha)b)^2 \geq 0$ for $a,b,\alpha\in\mathbb{R}$ may be written as $\alpha(1-\alpha)(a+b)^2 \leq \alpha a^2 + (1-\alpha)b^2$. This together with the triangle inequality implies, for $x,y,z\in X$, $\alpha\in\mathbb{R}$,

$$\alpha(1-\alpha)d^2(x,y) \leq \alpha d^2(x,z) + (1-\alpha)d^2(z,y). \tag{14}$$

Now consider $\mu = s\delta_{\{x_0\}} + t\delta_{\{x_1\}} + (1-s)\delta_{\{x_2\}} + (1-t)\delta_{\{x_3\}}$. Let $\varepsilon>0$, and choose $z_\varepsilon\in X$ such that

$$\mathrm{Var}(\mu) \geq \frac{1}{2}\left(sd^2(z_\varepsilon,x_0) + td^2(z_\varepsilon,x_1) + (1-s)d^2(z_\varepsilon,x_2) + (1-t)d^2(z_\varepsilon,x_3)\right) - \varepsilon$$

Using equation (14), one deduces

$$\mathrm{Var}(\mu) \geq \frac{1}{2}\left(s(1-s)d^2(x_0,x_2) + t(1-t)d^2(x_1,x_3)\right) - \varepsilon.$$

Since $\varepsilon$ is arbitrary,

$$\mathrm{Var}(\mu) \geq \frac{1}{2}\left(s(1-s)d^2(x_0,x_2) + t(1-t)d^2(x_1,x_3)\right). \tag{15}$$

On the other hand, by $(iii)$,

$$\begin{aligned}\mathrm{Var}(\mu) &\leq \frac{1}{2}\int_X\int_X d^2(x,y)\mu(dx)\mu(dy)\\ &= \frac{1}{4}[std^2(x_0,x_1) + (1-s)td^2(x_1,x_2) + s(1-t)d^2(x_0,x_3)\\ &+ (1-s)(1-t)d^2(x_2,x_3) + s(1-s)d^2(x_0,x_2) + t(1-t)d^2(x_1,x_3)].\end{aligned}$$

Let us now prove that $(iii)$ implies that $X$ is a length space. For given $x_0,x_1\in X$ choose $\mu = \frac{1}{2}\left(\delta_{\{x_0\}} + \delta_{\{x_1\}}\right)$. Then as above, for arbitrary $\varepsilon>0$ one finds $z_\varepsilon$ such that

$$\frac{1}{2}d^2(z_\varepsilon,x_0) + \frac{1}{2}d^2(z_\varepsilon,x_1) - \varepsilon \leq \frac{1}{4}d^2(x_0,x_1).$$

In other words, $X$ admits approximative midpoints and thus is a length space by Corollary 2.10.

<u>$(iv) \Longrightarrow (i)$</u>: Given $0 < t < 1$ and $x_0, x_1 \in X$ choose $y \in X$ such that

$$\max(d^2(y,x_0), d^2(y,x_1)) \leq \frac{1}{4} d^2(x_0,x_1) + (1-t)^2.$$

Apply the inequality in $(iv)$ to the quadruple $(x_0, y, x_1, z)$ (with $s = \frac{1}{2}$) to obtain

$$t(1-t)d^2(y,z) \leq \frac{1-t}{2}(d^2(x_0,y) + d^2(y,x_1)) - \frac{1-t}{4} d^2(x_0,x_1) + (1-t)^2,$$

and use Lemma 2.13. □

The following proposition reflects the fact that any quadruple in a Hadamard space may be embedded in Euclidean space such that edge lengths are preserved while diagonals expand.

**Proposition 2.23** (Reshetnyak quadruple comparison)**.** *Suppose $(X,d)$ is a Hadamard space. Then for all $x_0, x_1, x_2, x_3 \in X$*

$$d^2(x_0,x_2) + d^2(x_1,x_3) \leq d^2(x_1,x_2) + d^2(x_3,x_0) + 2d(x_0,x_1)d(x_2,x_3). \tag{16}$$

*Proof.* Use point $(iv)$ of Theorem 2.22, with $s = t$, to obtain

$$d^2(x_0,x_2) + d^2(x_1,x_3) \leq \frac{t}{1-t} d^2(x_0,x_1) + d^2(x_1,x_2) + \frac{1-t}{t} d^2(x_2,x_3) + d^2(x_3,x_0),$$

and choose $t \in (0,1)$ such that $\frac{t}{1-t} = \frac{d(x_2,x_3)}{d(x_0,x_1)}$. □

For a proof of the following important consequence of quadruple comparison see [20].

**Corollary 2.24** (Geodesic Comparison)**.** *Suppose $X$ is a Hadamard space. Then any pair of geodesics $x_t, y_t$ obeys*

$$d(x_t,y_t) \leq (1-t)d(x_0,y_0) + td(x_1,y_1).$$

*In other words, the function $d : X \times X \to \mathbb{R}$ is convex, which in turn implies the following:*

*(i) For each $x_0 \in X$ the function $x \mapsto d(x,x_0)$ is convex. As a consequence, all geodesic balls are convex.*

*(ii) Any pair of geodesics $x_t, y_t$ obeys*

$$\sup_{0 \leq t \leq 1} d(x_t,y_t) \leq \max(d(x_0,y_0), d(x_1,y_1)).$$

*(iii) $X$ is contractible.*

## 2.3 Nonlinear conditional expectations

In this section we introduce a natural concept of conditional expectation for random variables with values in a global NPC-space first considered by K.-T. Sturm in his seminal paper [19], and describe several of its crucial properties in detail.

**Definition 2.25.** Suppose $(\Omega, \mathfrak{F}, \mu)$ is a finite measure space. Moreover let $(X, d)$ be a metric space. For $p \geq 1$ define $\mathcal{L}^p(\Omega, \mathfrak{F}, \mu; X)$ to be the space of measurable functions $Y : (\Omega, \mathfrak{F}) \to X$ satisfying $\int_\Omega d^p(Y(\omega), x)\mu(d\omega) < \infty$ for some (then all) $x \in X$. The space $\mathcal{L}^p(\Omega, \mathfrak{F}, \mu; X)$ comes with the equivalence relation of being equal almost everywhere:

$$Y \sim Z \quad \Longleftrightarrow \quad \mu(Y \neq Z) = 0.$$

The set of equivalence classes with respect to this relation is denoted by

$$L^p(\Omega, \mathfrak{F}, \mu; X) = \mathcal{L}^p(\Omega, \mathfrak{F}, \mu; X)/\sim .$$

Whenever the other parameters are clear from the context, we write $L^p(\mathfrak{F}, X)$, $L^p(\mathfrak{F})$ or $L^p(X)$.

**Theorem 2.26.** *Suppose* $(\Omega, \mathfrak{F}, \mu)$ *is a finite measure space, and let* $(X, d)$ *be a metric space. Then*

$$d_p(Y, Z) = \left( \int_\Omega d^p(Y, Z) \right)^{\frac{1}{p}},$$

*defines a metric on* $L^p(X)$. *Moreover, the following hold true:*

*(i) Completeness of* $(X, d)$ *implies completeness of* $L^p(X)$.

*(ii) If* $(X, d)$ *is a Hadamard space, then so is* $L^2(X)$.

*Proof.* We first show (i), assuming without loss of generality that $\mu(\Omega) = 1$: Suppose the sequence $Y_n : \Omega \to X$ satisfies

$$d_p(Y_n, Y_m) \to 0 \quad \text{as } n, m \to \infty.$$

Choose a subsequence $n_k$ satisfying

$$d_p(Y_{n_k}, Y_{n_{k+1}}) < \frac{1}{2^{(p+1)k}}.$$

Then the Markov inequality implies

$$\mu\left(d(Y_{n_k}, Y_{n_{k+1}}\right) > \frac{1}{2^k}) \leq \frac{1}{2^k},$$

which, according to the Borel-Cantelli lemma, provides us with the fact that almost everywhere $d(Y_{n_k}, Y_{n_{k+1}}) < \frac{1}{2^k}$ holds true, for large enough $k$. In particular, by completeness of $(X,d)$ there exists $Y : \Omega \to (X,d)$ such that $Y_{n_k} \to Y$ almost everywhere. Given $\varepsilon > 0$, choose $N_0 \in \mathbb{N}$ such that

$$n, m \geq N_0 \implies d_p(Y_n, Y_m) < \varepsilon.$$

Then for $n \geq N_0$ by the lemma of Fatou we have

$$\begin{aligned}\int_\Omega d^p(Y, Y_n) d\mu &= \int_\Omega \liminf_{n_k \geq N_0} d^p(Y_{n_k}, Y_n) d\mu \\ &\leq \liminf_{n_k \geq N_0} \int_\Omega d^p(Y_{n_k}, Y_n) d\mu < \varepsilon^p,\end{aligned}$$

which proves (i).

Claim (ii) is easy to prove. Indeed, given $Y_0, Y_1 \in L^2(X)$, set

$$Y_{1/2}(\omega) := \mathrm{mpt}(Y_0(\omega), Y_1(\omega)).$$

Then the Hadamard inequality for the metric $d$ implies

$$\begin{aligned}d_2^2(Z, Y_{1/2}) &= \int_\Omega d^2(Z(\omega), Y_{1/2}(\omega))\mu(d\omega) \\ &\leq \frac{1}{2}\int d^2(Z(\omega), Y_0(\omega))\mu(d\omega) + \frac{1}{2}\int d^2(Z(\omega), Y_1(\omega))\mu(d\omega) \\ &- \frac{1}{4}\int d^2(Y_0(\omega), Y_1(\omega))\mu(d\omega) \\ &= \frac{1}{2}d_2^2(Z, Y_0) + \frac{1}{2}d_2^2(Z, Y_1) - \frac{1}{4}d_2^2(Y_0, Y_1).\end{aligned}$$

□

*Remark* 2.27. The proof of Theorem 2.26 provides us with an explicit representation of geodesics in $L^2(X)$. Indeed, since the geodesic midpoint of $Y_0, Y_1 \in L^2(X)$ is simply the pointwise midpoint, meaning $\mathrm{mpt}(Y_0, Y_1)(\omega) = \mathrm{mpt}(Y_0(\omega), Y_1(\omega))$ for all $\omega \in \Omega$, it follows that

$$Y_t(\omega) = (Y(\omega))_t.$$

**Theorem 2.28** (Convex Projections)**.** *Suppose $(X,d)$ is a Hadamard space, and $K \subseteq X$ is closed and convex. Then there is a well-defined map*

$$\pi_K : X \to K$$

*determined by $d(\pi_K(x), x) = \min_{y \in K} d(y,x)$. This maps enjoys the following properties:*

1. *$\pi_K$ is orthogonal, i.e. for all $x \in X$ and $y \in K$,*

$$d^2(\pi_K(x), x) + d^2(\pi_K(x), y) \leq d^2(x,y). \tag{17}$$

2. *$\pi_K$ is Lipschitz-continuous in the sense that*

$$d(\pi_K(x), \pi_K(y)) \leq d(x,y).$$

*Proof.* Any closed and convex subset of $(X,d)$ is a Hadamard space of its own (with the induced metric). Moreover, for each $x \in X$, by the Hadamard inequality, $\Phi = d^2(\,\cdot\,, x)$ is a strongly convex function on the Hadamard space $K$, and hence possesses a unique minimizer $\pi_K(x)$. Orthogonality (17) is a simple consequence of applying Proposition 2.17 to $d^2(\,\cdot\,, x)$. Lipschitz continuity is slightly more subtle. Note first that (17) implies

$$d^2(z, \pi_K(w)) + d^2(w, \pi_K(z)) \geq 2d^2(\pi_K(z), \pi_K(w)) + d^2(z, \pi_K(z)) + d^2(w, \pi_K(w)).$$

On the other hand, quadruple comparison (16) gives

$$d^2(z, \pi_K(w)) + d^2(w, \pi_K(z)) \leq d^2(\pi_K(z), \pi_K(w)) + d^2(\pi_K(w), w) + d^2(w,z) + d^2(z, \pi_K(z)),$$

proving the claim. □

**Theorem 2.29.** *Let $(X,d)$ be a Hadamard space and suppose $Y, Z \in L^2(\mathfrak{F}) := L^2(\Omega, \mathfrak{F}, \mathbb{P}; X)$ are square-integrable random variables with values in $X$. Choose a subalgebra $\mathfrak{G} \subseteq \mathfrak{F}$. Then $L^2(\mathfrak{G})$ is a convex and closed subset of $L^2(\mathfrak{F})$. Define the* conditional expectation of $Y$ given $\mathfrak{G}$ *as*

$$E(Y|\mathfrak{G}) := \pi_{L^2(\mathfrak{G})}(Y), \tag{18}$$

*that is, $E(Y|\mathfrak{G})$ is the class of functions minimizing $L^2$-distance to $Y$ among all $\mathfrak{G}$-measurable classes. Then for $W \in L^2(\mathfrak{G})$ we have*

$$d(E(Y|\mathfrak{G}), E(Z|\mathfrak{G})) \leq E(d(Y,Z)|\mathfrak{G}) \tag{19}$$

$$E(d^2(E(Y|\mathfrak{G}), Y)|\mathfrak{G}) + d^2(E(Y|\mathfrak{G}), W) \leq E(d^2(Y,W)|\mathfrak{G}) \tag{20}$$

*$\mathbb{P}$-almost surely. Moreover, for all $p \in [1, \infty]$,*

$$d_p(E(Y|\mathfrak{G}), E(Z|\mathfrak{G})) \leq d_p(Y,Z). \tag{21}$$

*Proof.* Choose $Y \in L^2(\mathfrak{G})$, and consider the family of strongly convex functionals

$$\Phi_A : L^2(A, \mathfrak{G}_A, \mathbb{P}_A; X) \to \mathbb{R}; \quad Z \mapsto \int_A d^2(Z,Y) d\mathbb{P}_A.$$

parametrized by $A \in \mathfrak{G}$ such that $\mathbb{P}(A) > 0$. Let $Z_A$ denote the unique minimizer of $\Phi_A$. Then $Z_A$ admits the following interpretation: $K(A) = L^2(A, \mathfrak{G}_A, \mathbb{P}_A; X) \subseteq L^2(A, \mathfrak{F}_A, \mathbb{P}_A; X)$ is a closed and convex subset, and hence the convex projection $\pi_{K(A)}$ is well defined, cf. Theorem 2.28. Then obviously, $Z_A = \pi_{K(A)}(Y|A)$. By definition, $Z_\Omega = \pi_{K(\Omega)}(Y) = E(Y|\mathfrak{G})$.

We claim that $Z_\Omega|A = Z_A$ for each $A \in \mathfrak{G}$. Indeed, if $\Phi_A(Z_\Omega|A) > \Phi_A(Z_A)$, let

$$Z'_\Omega := \begin{cases} Z_A & \text{on } A, \\ Z_\Omega & \text{on } \Omega \setminus A. \end{cases}$$

Obviously, $Z'_\Omega \in L^2(\mathfrak{G})$, and, setting $\Phi := \Phi_\Omega$,

$$\begin{aligned} \Phi(Z'_\Omega) &= \mathbb{P}(A)\Phi_A(Z_A) + (1 - \mathbb{P}(A))\Phi_{\Omega\setminus A}(Z_\Omega|(\Omega \setminus A)) \\ &< \mathbb{P}(A)\Phi_A(Z_\Omega|A) + (1 - \mathbb{P}(A))\Phi_{\Omega\setminus A}(Z_\Omega|(\Omega \setminus A)) \\ &= \Phi(Z_\Omega), \end{aligned}$$

a contradiction.

Since it holds that $\pi_{K(A)}(\cdot) = E(\,\cdot\,|\mathfrak{G})|_A$, we deduce from Theorem 2.28 that for all $A \in \mathfrak{G}$, $Y, Z \in L^2(\mathfrak{F})$, and $W \in L^2(\mathfrak{G})$,

$$E(d(E(Y|\mathfrak{G}), E(Z|\mathfrak{G}))\mathbf{1}_A) \leq E(d(Y,Z)\mathbf{1}_A) \tag{22}$$

$$E((d^2(E(Y|\mathfrak{G}), Y) + d^2(E(Y|\mathfrak{G}), W))\mathbf{1}_A) \leq E(d^2(Y,W)\mathbf{1}_A). \tag{23}$$

Certainly, (22) respectively (23) being true for all $A \in \mathfrak{G}$ is equivalent to (19) and (20), respectively. Finally, (21) is a direct consequence of (19). □

*Remark* 2.30. The $L^p$-Lipschitz continuity (21) allows for a continuous extension of $E(\,\cdot\,|\mathfrak{G})$ to $L^1$ in the classical fashion.

*Remark* 2.31. In case $X = \mathbb{R}$, the above definition coincides with the notion of conditional expectation due to Kolmogorov. More precisely, $Z = E(Y|\mathfrak{G})$ is characterized by the following conditions:

(i) $Z$ is $\mathfrak{G}$-measurable.

(ii) For all $A \in \mathfrak{G}$ it holds that $E(\mathbf{1}_A Y) = E(\mathbf{1}_A Z)$.

The existence and uniqueness of $E(Y|\mathfrak{G})$, using $(i)$ and $(ii)$ as axioms, relies on the Radon-Nikodym theorem.

*Remark* 2.32. A well-known feature of the *linear* conditional expectation is the so-called *tower property* - whenever $\mathfrak{H} \subseteq \mathfrak{G} \subseteq \mathfrak{F}$ is a nested triple of subalgebras, it holds that

$$E(E(Y|\mathfrak{G})|\mathfrak{H}) = E(Y|\mathfrak{H}). \tag{24}$$

The following example shows that the tower property is no longer valid in the nonlinear setting. Consider the tripod $T = \mathbb{R}_{\geq 0} \times \{0,1,2\}/\sim$, where $\sim$ denotes the equivalence relation generated by $(0,0) \sim (0,1) \sim (0,2)$. It is plain to show that $T$ constitutes a Hadamard space when endowed with the distance

$$d([s,i],[t,j]) = \begin{cases} |t-s|, & \text{if } i = j \\ t+s & \text{otherwise.} \end{cases}$$

Choose random variables $X_0, X_1, X_2$ possessing the following (conditional) distributions:

$$\mathbb{P}(X_0 = [1,i]) = \frac{1}{3}, \text{ where } i = 0,1,2$$
$$\mathbb{P}(X_n = [3^n, i] \mid X_1 = [3^{n-1}, j]) = \frac{4^{\delta_{ij}}}{6}, \text{ where } i,j = 0,1,2 \text{ and } n = 1,2.$$

This triple of random variables can be considered a 3-step random walk, with the probability of staying in the current branch of $T$ being twice as large as the probability of leaving it and the probabilities of ending up in each of the others being equal. Define $\mathfrak{F}_n = \sigma(X_0,\dots,X_n)$, for $n = 0,1,2$. Note that

$$\mathbb{P}(X_2 = [9,i] \mid X_0 = [1,j]) = \frac{2^{\delta_{ij}}}{4}, \text{ where } i,j = 0,1,2,$$

which in turn implies $E(X_2|\mathfrak{F}_0) = 0$. On the other hand, $E(E(X_2|\mathfrak{F}_1)|\mathfrak{F}_0) = X_0 \neq 0$.

It is for the reason of the lacking tower property that K.-T. Sturm in his article [19] defines:

**Definition 2.33** (Filtered conditional expectation)**.** Suppose $\mathfrak{F}_0 \subseteq \mathfrak{F}_1 \subseteq \cdots \mathfrak{F}_N = \mathfrak{F}$ is a sequence of subalgebras. Furthermore assume $Y \in L^2(\mathfrak{F}, X)$. Then one defines the *filtered conditional expectation* of $Y$ given $(\mathfrak{F}_n)_{0 \le n \le N}$ as

$$E(Y|||\mathfrak{F}_0) = E(\cdots E(E(Y|\mathfrak{F}_{N-1})|\mathfrak{F}_{N-2})\cdots|\mathfrak{F}_0). \tag{25}$$

**Proposition 2.34.** *Suppose $Y$ is a square-integrable random variable with values in a closed and convex subset $K$ of a Hadamard space $X$. Then $\mathbb{P}(E(Y|\mathfrak{G}) \in K) = 1$.*

*Proof.* Suppose $A = \{E(Y|\mathfrak{G}) \notin K\}$ has positive probability, and let $\pi_K$ denote the convex projection to $K$. Then obviously $Z := \pi_K(E(Y|\mathfrak{G})) \in L^2(\mathfrak{G})$, and by the conditional variance inequality (20),

$$E(d^2(Z,Y)) \le E(d^2(Y, E(Y|\mathfrak{G}))) - E(d^2(E(Y|\mathfrak{G}), Z)) < E(d^2(Y, E(Y|\mathfrak{G}))),$$

a contradiction to the minimality of $E(d^2(Y, E(Y|\mathfrak{G})))$. □

### 2.3.1 Jensen's inequality

The following theorem, addressing the comparison of nonlinear and linear conditional expectations, is essential in the convergence theory of barycentric refinement schemes developed in 4.2.

**Theorem 2.35** (Jensen's inequality, [19])**.** *Suppose $(\Omega, (\mathfrak{F}_n)_{n\in\mathbb{N}_0}, \mathfrak{F}, \mathbb{P})$ is a filtered probability space. Moreover let $Y \in L^2(\mathfrak{F}_{n_0})$ for some $n_0 \ge 0$. Then for each convex, lower semicontinuous $\psi : X \to \mathbb{R}$, it holds that*

$$\psi\big(E(Y|||\mathfrak{F}_0)\big) \le E(\psi(Y)|\mathfrak{F}_0). \tag{26}$$

*Proof.* Let $\Gamma_+(\psi)$ denote the *epigraph* of $\psi$ consisting of all pairs $(x,t) \in X \times \mathbb{R}$ such that $\psi(x) \le t$. Due to lower semicontinuity of $\psi$, $\Gamma_+(\psi)$ is closed. Indeed, given a convergent sequence $(x_n, t_n) \in \Gamma_+(\psi)$, it follows that $\psi(\lim_n x_n) \le \limsup_{n\in\mathbb{N}} \psi(x_n) \le \lim_n t_n$. Moreover, given a geodesic $(x_s, (1-s)t_0 + st_1)$ in $X \times \mathbb{R}$ joining points $(x_0,t_0), (x_1,t_1) \in \Gamma_+(\psi)$, convexity of $\psi$ implies $\psi(x_s) \le (1-s)\psi(x_0) + s\psi(x_1) \le$

$(1-s)t_0+st_1$. Therefore $\Gamma_+(\psi)$ is convex. Obviously the function $\Psi : X \to X\times\mathbb{R};\ x \mapsto (x, \psi(x))$ takes values in $\Gamma_+(\psi)$, and Proposition 2.34 implies

$$E(\Psi(Y)|\mathfrak{F}_{n_0-1}) = (E(Y|\mathfrak{F}_{n_0-1}), E(\psi(Y)|\mathfrak{F}_{n_0-1})) \in \Gamma_+(\psi).$$

In other words, $\psi(E(Y|\mathfrak{F}_{n_0-1})) \leq E(\psi(Y)|\mathfrak{F}_{n_0-1})$. Iterating this procedure and applying the tower property of the linear conditional expectation on the RHS, one obtains the desired inequality. □

Finding contractivity constants to determine the speed of convergence of a barycentric scheme is an important issue, facilitated by a Lipschitz continuity property of the barycenter map described in this paragraph.

**Definition 2.36** (Wasserstein spaces)**.** Suppose $X$ is a metric space. Let $\mathcal{P}^p(X)$ denote the space of $L^p$-probability measures on $X$. Define a *coupling* between $\mu, \nu \in \mathcal{P}^p(X)$ to be a measure $\pi \in \mathcal{P}^p(X \times X)$ whose marginals are $\mu$ and $\nu$, respectively: For all Borel sets $A \subseteq X$ it holds that

$$\pi(A \times X) = \mu(A)$$
$$\pi(X \times A) = \nu(A).$$

Then for $p \geq 1$ the *$L^p$-Wasserstein distance* of $\mu$ and $\nu$ is defined as

$$d_p^W(\mu, \nu) := \left(\inf\{\int_{X\times X} d^p(x, y)\pi(dx, dy) \mid \pi \text{ coupling of } \mu \text{ and } \nu\}\right)^{\frac{1}{p}}. \tag{27}$$

**Corollary 2.37** (Wasserstein contraction property)**.** *On a Hadamard space $X$ the following holds for any pair of probability measures $\mu$, $\nu$:*

$$d(b(\mu), b(\nu)) \leq d_p^W(\mu, \nu). \tag{28}$$

*Proof.* Consider $Y = \mathrm{id} : (X, \mu) \to X$ and $Z = \mathrm{id} : (X, \nu) \to X$. Then for any coupling $\pi$ of $\mu$ and $\nu$ obviously $U = \mathrm{id} : (X \times X, \pi) \to X \times X$ satisfies $U_1 = Y$ and $U_2 = Z$. Further on, by convexity of $d : X \times X \to \mathbb{R}$, Jensen's inequality (26) implies

$$\begin{aligned} d^p(b(\mu), b(\nu)) &= d^p(E(Y), E(Z)) \\ &= d^p(E(U)) \\ &\leq E(d^p(U)) \\ &= \int_{X\times X} d^p(x, y)\pi(dx, dy), \end{aligned}$$

which concludes the proof. □

### 2.3.2 The laws of large numbers

The topic of this paragraph is a strong law of large numbers relevant in the analysis of subdivision schemes in many respects, see further Section 4.3.

**Lemma 2.38.** *Suppose* $Y, Z : (\Omega, \mathfrak{F}, \mathbb{P}) \to (X, d)$ *are independent. Then*

$$E(d^2(Y,Z)) = \int_\Omega \int_\Omega d^2(Y(\omega), Z(\eta))\mathbb{P}(d\omega)\mathbb{P}(d\eta).$$

*Moreover, it holds that*

$$E(d^2(Y,Z)) \geq E(d^2(Y, E(Y))) + E(d^2(Z, E(Y))).$$

*Proof.* As in the linear case, one shows that

$$E(d^2(Y,Z)|Z)(\eta) = E(d^2(Y, Z(\eta))) = \int_\Omega d^2(Y(\omega), Z(\eta))\mathbb{P}(d\omega),$$

which in turn implies

$$E(d^2(Y,Z)) = E(E(d^2(Y,Z)|Z)) = \int_\Omega \int_\Omega d^2(Y(\omega), Z(\eta))\mathbb{P}(d\omega)\mathbb{P}(d\eta).$$

The second statement follows by integrating both sides of the conditional variance inequality (20). Indeed, substituting $Z$ for $W$ and $\sigma(Z)$ for $\mathfrak{G}$ in (20) yields

$$E(d^2(E(Y), Y)) + d^2(E(Y), Z) \leq E(d^2(Y,Z) \mid Z).$$ □

**Theorem 2.39** (The Laws of Large Numbers,[20])**.** *Suppose* $Y_n : (\Omega, \mathfrak{F}, \mathbb{P}) \to (X, d)$ *is an independent and identically distributed (i.i.d.) sequence of random variables with values in a Hadamard space. Define recursively*

$$\begin{aligned} S_1 &:= Y_1 \\ S_{n+1} &:= b\left(\frac{1}{n+1}\delta_{Y_{n+1}} + \frac{n}{n+1}\delta_{S_n}\right) = \frac{1}{n+1}Y_{n+1} + \frac{n}{n+1}S_n. \end{aligned}$$

*Then the following hold true:*

**Weak Law of Large Numbers:** *Suppose* $\mathrm{Var}(Y_1) < \infty$*. Then*

$$S_n \to EY_1 \quad \text{in } L^2.$$

**Strong Law of Large Numbers:** *Suppose* $Y_1$ *is bounded a.s. Then*

$$S_n \to EY_1 \quad a.s.$$

*Proof.* Simplifying notation, we set $\mu = EY_1$. To show the weak law of large numbers, we inductively prove

$$E(d^2(S_n, \mu)) \leq \frac{1}{n} \operatorname{Var}(Y_1).$$

The case $n = 1$ is trivial since $S_1 = Y_1$ implies $E(d^2(S_1, \mu)) = \operatorname{Var}(Y_1)$. Performing the induction step, we first observe that the Hadamard inequality gives

$$d^2(S_{n+1}, \mu) \leq \frac{1}{n+1} d^2(Y_{n+1}, \mu) + \frac{n}{n+1} d^2(S_n, \mu) - \frac{n}{(n+1)^2} d^2(Y_{n+1}, S_n).$$

Taking integrals on both sides of this inequality and using Lemma 2.38, one obtains

$$\begin{aligned} E(d^2(S_{n+1}, \mu)) \leq & \frac{1}{n+1} E(d^2(Y_{n+1}, \mu)) + \frac{n}{n+1} E(d^2(S_n, \mu)) \\ & - \frac{n}{(n+1)^2} E(d^2(Y_{n+1}, S_n)) \end{aligned}$$

$$\begin{aligned} & \leq \frac{1}{n+1} E(d^2(Y_{n+1}, \mu)) + \frac{n}{n+1} E(d^2(S_n, \mu)) \\ & - \frac{n}{(n+1)^2} (E(d^2(Y_{n+1}, \mu)) + E(d^2(S_n, \mu))) \\ & = \frac{1}{(n+1)^2} E(d^2(Y_1, \mu)) + \frac{n^2}{(n+1)^2} E(d^2(S_n, \mu) \\ & \leq \frac{1}{(n+1)^2} \operatorname{Var}(Y_1) + \frac{n}{(n+1)^2} \operatorname{Var}(Y_1) \\ & = \frac{1}{n+1} \operatorname{Var}(Y_1). \end{aligned}$$

This implies the weak law of large numbers.

Suppose now that there exists $z \in X$ and $R > 0$ such that $d(z, Y_1) \leq R$ almost surely. Then from the above and the Markov inequality we conclude

$$\mathbb{P}(d(S_{n^2}, \mu) > \varepsilon) \leq \frac{1}{\varepsilon^2} E(d^2(S_{n^2}, \mu)) \leq \frac{1}{n^2 \varepsilon^2} \operatorname{Var}(Y_1).$$

Thus,

$$\sum_n \mathbb{P}(d(S_{n^2}, \mu) > \varepsilon) \leq \sum_n \frac{1}{n^2 \varepsilon^2} \operatorname{Var}(Y_1) < \infty,$$

which in view of the Borel-Cantelli lemma implies that $S_{n^2} \to \mu$ almost surely. Observe that the Wasserstein contraction property implies

$$d(S_n, S_{n+1}) \leq \frac{1}{n+1} d(S_n, Y_{n+1}) \leq \frac{2R}{n+1}.$$

Hence for $n^2 \le k < (n+1)^2$ one concludes

$$d(S_{n^2}, S_k) \le \sum_{j=n^2}^{k-1} \frac{2R}{j+1} \le 2R\frac{k-n^2}{n^2} \le \frac{4R}{n},$$

from which the strong law of large numbers follows. $\square$

# 3 Markov chains

Albeit not obvious at the first glance, Markov chains turn out to be a useful tool in the convergence analysis of nonlinear refinement schemes. Markov chains are time-discrete stochastic processes with the property that, conditionally on the present, the future and the past are independent. In other words, the path such a random walk has taken up to some point in time does not provide any more information about future values than the current location. Although there is an abundance of theory on Markov chains, we, in the course of our considerations, merely make use of some foundational facts. Due to this restriction in scope, a detailed exposition of these fundamentals is given. For any of the results appearing in this section, and a wealth of additional material see the recent monograph [26].

## 3.1 The Kolmogorov Existence Theorem

**Definition 3.1** (Kolmogorov consistency conditions)**.** Let $\mathcal{X}$ be a countable set. Suppose that for distinct $i_0, \dots, i_n \in \mathbb{N}_0$

$$p_{i_0 \dots i_n} : \mathfrak{P}(\mathcal{X}) \otimes \cdots \otimes \mathfrak{P}(\mathcal{X}) \to [0,1]$$

is a probability measure on $\mathcal{X}^n$. Then the family $(p_{i_0 \dots i_n})_{n \in \mathbb{N}_0, i_j \in \mathbb{N}_0}$ fulfills the *Kolmogorov consistency conditions* if and only if

(K1) For all $B_0, \dots, B_n \subseteq \mathcal{X}$, $i_0, \dots, i_n \in \mathbb{N}_0$ and any permutation $\pi \in S_{n+1}$ it holds that

$$p_{i_{\pi(0)} \dots i_{\pi(n)}}(B_{\pi(0)} \times \cdots \times B_{\pi(n)}) = p_{i_0 \dots i_n}(B_0 \times \cdots \times B_n).$$

(K2) For all $B_0, \dots, B_{n-1} \subseteq \mathcal{X}$, $i_0, \dots, i_n \in \mathbb{N}_0$,

$$p_{i_0 \dots i_n}(B_0 \times \cdots \times B_{n-1} \times \mathcal{X}) = p_{i_0 \dots i_{n-1}}(B_0 \times \cdots \times B_{n-1}).$$

*Remark* 3.2. The marginal distributions of a stochastic process with values in a state space $\mathcal{X}$, given by

$$p_{i_0 \dots i_n}(B_0 \times \cdots \times B_n) = \mathbb{P}(X_{i_0} \in B_0, \dots, X_{i_n} \in B_n),$$

where $n, i_0, \dots, i_n \in \mathbb{N}_0$, obviously fulfill the Kolmogorov consistency conditions. Theorem 3.3 below shows that any family of probability distributions satisfying (K1) and (K2) is given as marginal distributions of a stochastic process.

For a permutation $\pi \in S_{n+1}$ consider

$$\varphi_\pi : \mathcal{X}^{n+1} \to \mathcal{X}^{n+1};\ (x_0, \ldots, x_n) \mapsto (x_{\pi^{-1}(0)}, \ldots, x_{\pi^{-1}(n)}).$$

Then under the assumption of (K2),

$$\begin{aligned}(\varphi_\pi)_* p_{i_{\pi(0)} \ldots i_{\pi(n)}}(B_0 \times \cdots \times B_n) &= p_{i_{\pi(0)} \ldots i_{\pi(n)}}(\varphi_\pi^{-1}(B_0 \times \cdots \times B_n)) \\ &= p_{i_{\pi(0)} \ldots i_{\pi(n)}}(B_{\pi(0)} \times \cdots \times B_{\pi(n)}) \\ &= p_{i_0 \ldots i_n}(B_0 \times \cdots \times B_n),\end{aligned}$$

and since product sets generate $\mathfrak{A}_n = \mathfrak{P}(\mathcal{X}) \otimes \cdots \otimes \mathfrak{P}(\mathcal{X})$, we have the following equivalent version of (K2):

$$(\varphi_\pi)_* p_{i_{\pi(0)} \ldots i_{\pi(n)}} = p_{i_0 \ldots i_n} \text{ for all } \pi \in S_{n+1}. \tag{29}$$

Following the same lines one shows that axiom (K1) is equivalent to the following: For all projections $P : \mathbb{R}^{n+1} \to \mathbb{R}^n;\ (x_0, \ldots, x_n) \mapsto (x_0, \ldots, \not{x_j}, \ldots, x_n)$ it holds that

$$P_* p_{i_0, \ldots, i_n} = p_{i_0 \ldots, \not{i_j}, \ldots, i_n}. \tag{30}$$

**Theorem 3.3.** *Suppose the family $(p_{i_0 \ldots i_n})_{n \in \mathbb{N}_0,\, i_j \in \mathbb{N}_0}$ fulfills the Kolmogorov consistency conditions. Then there exists a stochastic process $X_n : (\Omega, \mathfrak{F}, \mathbb{P}) \to \mathcal{X}$ such that for $i_0, \ldots, i_n \in \mathbb{N}_0$ the joint distribution of $(X_{i_0}, \ldots, X_{i_n})$ fulfills*

$$\mathbb{P}_{(X_{i_0}, \ldots, X_{i_n})} = p_{i_0 \ldots i_n}.$$

*Proof.* Define $\Omega = \mathcal{X}^{\mathbb{N}_0}$ (the *space of paths in* $\mathcal{X}$) and $\mathfrak{F} = \bigotimes_{n \in \mathbb{N}_0} \mathfrak{P}(\mathcal{X})$. Moreover, for $i_0, \ldots, i_n \in \mathbb{N}_0$ and $B \subseteq \mathcal{X}^{n+1}$ define the cylinder set

$$Z^B_{i_0, \ldots, i_n} = \{\omega \in \Omega \mid (\omega_{i_0}, \ldots, \omega_{i_n}) \in B\},$$

and set

$$\mathbb{P}(Z^B_{i_0, \ldots, i_n}) = p_{i_0 \ldots i_n}(B). \tag{31}$$

We prove that

$$\mathfrak{H} = \{Z^B_{i_0, \ldots, i_n} \mid n \in \mathbb{N}_0,\ i_0, \ldots, i_n \in \mathbb{N}_0,\, B \subseteq \mathcal{X}^{n+1}\}$$

is a ring, that is, $\mathfrak{H}$ is closed under forming of unions and complementation. Furthermore we show that $\mathbb{P}$ constitutes a premeasure on $\mathfrak{H}$. For this sake note that given two cylinders $Z^B_{i_0,\dots,i_n}$ and $Z^C_{j_0,\dots,j_m}$ one finds $r \in \mathbb{N}_0$, $B', C' \subseteq \mathcal{X}^{r+1}$ and $k_0, \dots, k_r$ such that $Z^{B'}_{k_0,\dots,k_r} = Z^B_{i_0,\dots,i_n}$ and $Z^{C'}_{k_0,\dots,k_r} = Z^C_{j_0,\dots,j_m}$. Thus,

$$\begin{aligned} Z^B_{i_0,\dots,i_n} \cup Z^C_{j_0,\dots,j_m} &= Z^{B'}_{k_0,\dots,k_r} \cup Z^{C'}_{k_0,\dots,k_r} \\ &= Z^{B' \cup C'}_{k_0,\dots,k_r} \in \mathfrak{H}. \end{aligned}$$

Accordingly, $\mathfrak{H}$ is closed under forming of unions. Likewise, $\Omega \setminus Z^B_{i_0,\dots,i_n} = Z^{\mathcal{X}^{n+1} \setminus B}_{i_0,\dots,i_n}$ implies closedness of $\mathfrak{H}$ under complementation. Hence $\mathfrak{H}$ constitutes a ring.

In order to show that (31) defines a premeasure on $\mathfrak{H}$, we first prove well-definedness. Observe that, using the above notation, $Z^B_{i_0,\dots,i_n} = Z^C_{j_0,\dots,j_m}$ implies $B' = C'$. Moreover, the representation $Z^{B'}_{k_0,\dots,k_r}$ is constructed by subjecting $Z^B_{i_0,\dots,i_n}$ to permutations of indices and pullbacks with respect to projections. The Kolmogorov consistency conditions in the shape of (29) and (30) thus imply well-definedness of (31). Since the cylinders $Z^{B'}_{k_0,\dots,k_r}$ and $Z^{C'}_{k_0,\dots,k_r}$ are disjoint if and only if $C' \cap B' = \emptyset$, it follows that in this case,

$$\begin{aligned} \mathbb{P}(Z^{B'}_{k_0,\dots,k_r} \cup Z^{C'}_{k_0,\dots,k_r}) &= \mathbb{P}(Z^{B' \cup C'}_{k_0,\dots,k_r}) \\ &= p_{k_0,\dots,k_r}(B' \cup C') \\ &= p_{k_0,\dots,k_r}(B') + p_{k_0,\dots,k_r}(C') \\ &= \mathbb{P}(Z^{B'}_{k_0,\dots,k_r}) + \mathbb{P}(Z^{C'}_{k_0,\dots,k_r}), \end{aligned}$$

which provides finite additivity of $\mathbb{P}$. To prove countable additivity it suffices to show that for cylinder sets $U_n$ with $U_n \downarrow \emptyset$ it holds that $\mathbb{P}(U_n) \to 0$. Assume, on the contrary, that there exists $\varepsilon >$ such that $\mathbb{P}(U_n) > \varepsilon$ for all $n \in \mathbb{N}_0$. Since $U_{n+1} \subseteq U_n$ we may assume without loss of generality that

$$U_n = Z^{B_n}_{0,\dots,n}.$$

By regularity there exists a compact (hence finite) sets $K_n \subseteq B_n$ such that $p_{0,\dots,n}(B_n \setminus K_n) < \frac{\varepsilon}{2^{n+1}}$. Define $V_n = Z^{K_n}_{0,\dots,n}$, and set $W_n = \cap_{j=0}^n V_j$. Then

$$\begin{aligned} \mathbb{P}(U_n \setminus W_n) &\leq \sum_{j=0}^n \mathbb{P}(U_j \setminus V_j) \\ &= \sum_{j=0}^n p_{0,\dots,j}(B_j \setminus K_j) < \frac{\varepsilon}{2}. \end{aligned}$$

It follows that $\mathbb{P}(W_n) > \frac{\varepsilon}{2}$ for all $n \in \mathbb{N}_0$, implying that $W_n$ is nonempty, with $W_{n+1} \subseteq W_n$. Choose a point $x^{(n)} \in W_n$. Then obviously for all $n \geq k$, $(x_0^{(n)}, \ldots, x_k^{(n)}) \in K_k$. Due to compactness of the $K_n$ one may apply a standard diagonalization technique to find a subsequence $n_j$ such that $x_k^{(n_j)}$ converges for all $k \in \mathbb{N}_0$. It thus follows that $x = \lim_{j\to\infty} x^{(n_j)} \in \bigcap_{n\in\mathbb{N}_0} W_n \subseteq \bigcap_{n\in\mathbb{N}_0} U_n$, a contradiction. □

## 3.2 Markov chains and Markov semigroups

In this section we introduce the notion of a Markov transition kernel and show how such an object gives rise to a time-discrete stochastic process with special properties on the one hand and an operator semigroup on the other.

**Definition 3.4.** Suppose $\mathcal{X}$ is a countable set of *states*. A *Markov chain* with state space $\mathcal{X}$ is a discrete-time stochastic process $X_n : (\Omega, \mathfrak{F}, \mathbb{P}) \to \mathcal{X}$ adapted to a filtration $(\mathfrak{F}_n)_{n\in\mathbb{N}_0}$, such that the so-called *Markov property* holds true: for each bounded $f : \mathcal{X} \to \mathbb{R}$ and $n < m$ we have

$$E(f(X_m)|\mathfrak{F}_n) = E(f(X_m)|X_n). \tag{32}$$

**Definition 3.5.** A *Markov transition kernel* is a bivariate family of nonnegative real numbers

$$P = (p_{xy})_{x,y\in\mathcal{X}}$$

such that $\sum_{y\in\mathcal{X}} p_{xy} = 1$ for all $y \in \mathcal{X}$.

**Theorem 3.6.** *For each Markov transition kernel and an initial distribution $\alpha \in \mathcal{P}(\mathcal{X})$ there exists a filtered probability space $(\Omega, \mathfrak{F}, (\mathfrak{F}_n)_{n\in\mathbb{N}_0}, \mathbb{P}_\alpha)$ and a Markov chain $X_n : (\Omega, \mathfrak{F}, \mathbb{P}_\alpha) \to \mathcal{X}$ satisfying $\mathbb{P}_\alpha(X_0 = x) = \alpha(x)$ and*

$$\mathbb{P}_\alpha(X_{n+1} = x|X_n = y) = p_{xy}$$

*for $x, y \in \mathcal{X}$ and $n \in \mathbb{N}_0$.*

*Proof.* Define a probability measure on $\mathcal{X}^{n+1}$ via

$$p_{0,\ldots,n}((x_0, \ldots, x_n)) = \alpha(x_0)p(x_0, x_1)\cdots p(x_{n-1}, x_n).$$

It is easy to check that the induced family of probability distributions fulfills the Kolmogorov consistency criteria and thus gives rise to a stochastic process $X_n$ with marginals given by $(\mathbb{P}_\alpha)_{(X_0,\ldots,X_n)} = p_{0,\ldots,n}$. In particular,

$$\begin{aligned}\mathbb{P}_\alpha(X_{n+1} = x_{n+1}, X_n = x_n, \ldots, X_0 = x_0) &= \alpha(x_0)p(x_0,x_1)\cdots p(x_{n-1},x_n)p(x_n,x_{n+1}) \\ &= \mathbb{P}_\alpha(X_n = x_n, \ldots, X_0 = x_0)p(x_n,x_{n+1}),\end{aligned}$$

implying $p(x_n, x_{n+1}) = \mathbb{P}_\alpha(X_{n+1} = x_{n+1} \mid X_n = x_n, \ldots, X_0 = x_0) = \mathbb{P}_\alpha(X_{n+1} = x_{n+1} \mid X_n = x_n)$. It follows for nonnegative $f : \mathcal{X} \to \mathbb{R}$ that

$$\begin{aligned}E_\alpha(f(X_{n+1})|X_n = x_n, \ldots, X_0 = x_0) &= \sum_{x\in\mathcal{X}} f(x)\mathbb{P}_\alpha(X_{n+1} = x|X_n = x_n, \ldots, X_0 = x_0) \\ &= \sum_{x\in\mathcal{X}} \mathbb{P}_\alpha(X_{n+1} = x|X_n = x_n)f(x) \\ &= \sum_{x\in\mathcal{X}} p(x_n, x)f(x) \\ &= E_\alpha(f(X_{n+1})|X_n = x_n).\end{aligned}$$

Thus, setting $\mathfrak{F}_n = \sigma(X_0, \ldots, X_n)$, we obtain $E_\alpha(f(X_{n+1})|\mathfrak{F}_n) = E_\alpha(f(X_{n+1})|X_n)$ as desired. □

*Remark* 3.7. Notice that although $\mathbb{P}_\alpha$ strongly depends on the initial distribution $\alpha$, the conditional distribution of $X_m$ given $X_n$, where $n \leq m$, does not. Thus, in the following we will specify the initial distribution only when speaking about absolute probabilities.

**Definition 3.8.** Let $P$ be a Markov transition kernel on $\mathcal{X}$ and let $X_n$ denote the associated Markov chain. The *n-step transition kernel* is defined as

$$p^{(n)}(x, y) := \mathbb{P}(X_n = y|X_0 = x) = P^n(x, y),$$

where the last expression denotes the $n$-fold power of the possibly infinite matrix $P$. Moreover, the *Markov semigroup* associated to $P$ acts on positive functions $f : \mathcal{X} \to \mathbb{R}$ as

$$T_n f(x) = E(f(X_n)|X_0 = x) = \sum_{y\in\mathcal{X}} p^{(n)}(x, y)f(y).$$

*Remark* 3.9. Suppose $P$ is a Markov transition kernel. Then the obvious identities

$$p^{(n)}(x, y) = \sum_{z\in\mathcal{X}} p^{(k)}(x, z)p^{(n-k)}(z, y), \quad \text{for } x, y \in \mathcal{X} \tag{33}$$

are referred to as the *Chapman-Kolmogorov equations.*

# 4 Refinement schemes on metric spaces

This chapter describes recent progress in the convergence theory of refinement algorithms on metric spaces to be found in the author's articles [9, 8]. Using the probabilistic preliminaries presented in the previous chapters, we develop a stochastic viewpoint on multivariate barycentric subdivision schemes with nonnegative masks on Hadamard spaces. In particular, we establish a link between these types of refinement algorithms and the theory of Markov chains by characterizing barycentric subdivision schemes as nonlinear Markov semigroups. Exploiting this connection, we subsequently prove the Linear Equivalence Theorem 1.2, whose statement we want to recall at this point:

**Theorem** ([8]). *A barycentric refinement scheme converges on arbitrary Hadamard spaces if and only if it converges on the real line.*

Moreover, we generalize a characterization of convergence from the linear theory, and consider approximation qualities of barycentric subdivision schemes. Subsequently, it is shown how the strong law of large numbers leads to certain structure-preserving properties of refinement schemes on the space of diffusion tensors. A concluding section addresses the relationship between the convergence properties of a scheme and its so-called characteristic Markov chain.

## 4.1 Refinement schemes as Markov semigroups

This section is devoted to a stochastic interpretation of the subdivision rule (3) that first appeared in the author's recent work [8]. More precisely, we view barycentric subdivision as the semigroup acting on $\ell^\infty(\mathbb{Z}^s, X)$ associated to the so-called characteristic Markov chain of $(a_i)_{i\in\mathbb{Z}^s}$. This result requires some more prerequisites about conditional expectations of random variables with values in Hadamard spaces.

In view of the convergence analysis of barycentric subdivision schemes, it is of particular interest to gain a deeper understanding of principle of conditioning in case the filtration stems from a Markov chain. A *nonlinear Markov property* analogous to (32), see [19, Theorem 5.2], leads to a representation of the conditional expectation explicit enough for our purposes. We provide a short proof adapted to our setting, beginning with an auxiliary result which can be found e.g. in [19]:

**Lemma 4.1.** *Suppose $(X_k)_{k\in\mathbb{N}_0}$ is a Markov chain in $\mathbb{Z}^s$ associated to the transition kernel $P$. Choose an initial distribution $\alpha$. Furthermore assume $Y:\Omega\to X$ is $\mathfrak{F}_n$-measurable with separable range, and let $x\in\ell^\infty(\mathbb{Z}^s,X)$. Then for nonnegative and measurable $f:X\times X\to\mathbb{R}$ and $m\geq n$ we have*

$$\int_\Omega f(x\circ X_m(\omega),Y(\omega))\mathbb{P}_\alpha(d\omega)=\int_\Omega\sum_j p_{n,m}(X_n(\omega),j)f(x_j,Y(\omega))\mathbb{P}_\alpha(d\omega).$$

**Proposition 4.2** (Nonlinear Markov property)**.** *Let $(X_k)_{k\in\mathbb{N}_0}$ be a Markov chain as in Lemma 4.1, and suppose $x\in\ell^\infty(\mathbb{Z}^s,X)$, with $(X,d)$ Hadamard. Choose $n,m\in\mathbb{N}_0$ with $n<m$. Then*

$$\begin{aligned}E_\alpha(x(X_m)|\mathfrak{F}_n)(\omega)&=\operatorname{argmin}\sum_{j\in\mathbb{Z}^s}p^{(m-n)}(X_n(\omega),j)d^2(x_j,\,\cdot\,)\\&=E_{X_n(\omega)}(x(X_m)).\end{aligned}\tag{34}$$

*Proof.* By the linear Markov property (32),

$$\begin{aligned}Y(\omega)&:=\operatorname{argmin}\big(E_\alpha(d^2(x\circ X_m,\,\cdot\,)\mid\mathfrak{F}_n)(\omega)\big)\\&=\operatorname{argmin}\sum_{j\in\mathbb{Z}^s}p^{(m-n)}(X_n(\omega),j)d^2(x_j,\,\cdot\,).\end{aligned}$$

Clearly $Y$, as a measurable function of $X_n$, is $\mathfrak{F}_n$-measurable. Thus, in order to verify that $Y$ is indeed the conditional expectation of $x(X_m)$ given $\mathfrak{F}_n$, it remains to show that for each $\mathfrak{F}_n$-measurable function $Z$ with separable range the inequality $E_\alpha(d^2(X_m,Y))\leq E_\alpha(d^2(X_m,Z))$ holds true, cf. Definition (18). For this sake, define

$$\begin{cases}\psi:\mathbb{Z}^s\times X\to[0,\infty];\\(i,z)\mapsto\sum_j p^{(m-n)}(i,j)d(x_j,z).\end{cases}$$

By construction of $Y$ we have $\psi(X_m,Y)\leq\psi(X_m,Z)$. Thus, Lemma 4.1 implies

$$\begin{aligned}E_\alpha(d^2(x\circ X_m,Y))&=E_\alpha(\psi(X_m,Y))\\&\leq E_\alpha(\psi(X_m,Z))\\&=E_\alpha(d^2(x\circ X_m,Z)).\end{aligned}$$

□

*Remark* 4.3. Proposition 4.2 implies that the expression $E_\alpha(x(X_m)|\mathfrak{F}_n)$ actually is independent of the initial distribution $\alpha$. Therefore we omit $\alpha$ in the following and simply write $E(x(X_m)|\mathfrak{F}_n)$.

We are now in a position to establish a link between nonlinear Markov semigroups and barycentric refinement processes. Suppose $\mathbf{a} = (a_i)_{i\in\mathbb{Z}^s}$ is a nonnegative compactly supported $s$-variate sequence such that $\sum_j a_{i-2j} = 1$ for all $i \in \mathbb{Z}^s$. Then clearly

$$p^{\mathbf{a}}(i,j) = a_{i-2j} \tag{35}$$

defines a Markov transition kernel. Define recursively $a_i^0 = \delta(i)$, where $\delta$ denotes the Dirac delta on the origin, and

$$a_i^{(n+1)} = \sum_{j\in\mathbb{Z}^s} a_{i-2j} a_j^{(n)}. \tag{36}$$

Then obviously $\left(p^{\mathbf{a}}\right)^{(n)}(i,j) = a^{(n)}_{i-2^n j}$. We write $P^{\mathbf{a}} = (p^{\mathbf{a}}(i,j))_{i,j\in\mathbb{Z}^s}$, denote the associated Markov chain by $X_n^{\mathbf{a}}$, and refer to $X_n^{\mathbf{a}}$ as the *characteristic Markov chain* of $(a_i)_{i\in\mathbb{Z}^s}$. The central observation of this chapter is the following consequence of the nonlinear Markov property (34):

**Theorem 4.4.** *Suppose $x : \mathbb{Z}^s \to X$ is bounded, where $(X, d)$ is a Hadamard space. Let $S$ be a barycentric refinement scheme acting on data from $X$ according to the subdivision rule* (3)*. Let $X_n^{\mathbf{a}}$ denote the characteristic Markov chain of $(a_i)_{i\in\mathbb{Z}^s}$. Then*

$$S^n x \circ X_0^{\mathbf{a}} = E(x \circ X_n^{\mathbf{a}} ||| \mathfrak{F}_0).$$

*Proof.* This statement is proven by induction over $n$ using the following computation based on Proposition 4.2:

$$\begin{aligned} E(x \circ X_n^{\mathbf{a}} \mid \mathfrak{F}_{n-1}) &= \operatorname{argmin}\Big(\sum_{j\in\mathbb{Z}^s} p^{\mathbf{a}}(X_{n-1}^{\mathbf{a}}, j) d^2(x_j, \cdot)\Big) \\ &= \operatorname{argmin}\Big(\sum_{j\in\mathbb{Z}^s} a_{X_{n-1}^{\mathbf{a}}-2j} d^2(x_j, \cdot)\Big) \\ &= Sx \circ X_{n-1}^{\mathbf{a}}. \end{aligned}$$ □

The rest of this chapter is devoted to analyzing the effects of this representation of the iterates of $S$ on the convergence properties of barycentric schemes with nonnegative masks.

## 4.2 The convergence problem

### 4.2.1 A primer on linear subdivision schemes

We begin this section by summarizing some well-known facts about the convergence of barycentric schemes acting on real-valued input data. As a standard reference we mention [6]. Other classical resources on this topic are [16] and, in the irregular grid case, [17].

**Theorem 4.5.** *Suppose* $\mathbf{a} = (a_i)_{i\in\mathbb{Z}^s}$ *is an s-variate compactly supported sequence of nonnegative reals. Define a refinement scheme* $\widetilde{S} : \ell^\infty(\mathbb{Z}^s, \mathbb{R}) \to \ell^\infty(\mathbb{Z}^s, \mathbb{R})$ *via*

$$\widetilde{S}x_i = \sum_{j\in\mathbb{Z}^s} a_{i-2j}x_j, \quad \textit{where } x \in \ell^\infty(\mathbb{Z}^s).$$

*We call* $\widetilde{S}$ convergent *if, in addition to* (1)*, there is at least one nonvanishing limit function, that is,* $\widetilde{S}^\infty \neq 0$*. Then a necessary condition for the convergence of* $\widetilde{S}$ *on* $\mathbb{R}$ *is the basic sum rule* (2)*. In case the mask* $(a_i)_{i\in\mathbb{Z}^s}$ *obeys this rule, we conclude*

$$\widetilde{S}x_i = \operatorname{argmin}\left(\sum_{j\in\mathbb{Z}^s} a_{i-2j}|x_j - \cdot\,|^2\right) = \operatorname{argmin}\left(\sum_{j\in\mathbb{Z}^s} a_{i-2j}d_{|\cdot|}(x_j,\,\cdot\,)^2\right).$$

*Moreover,* $\widetilde{S}$ *converges if and only if there exists a continuous* $\varphi : \mathbb{R} \to \mathbb{R}$ *subject to the functional equations*

$$\varphi(t) = \sum_j a_j\varphi(2t - j) \tag{37}$$

$$\sum_j \varphi(t - j) = 1. \tag{38}$$

*Due to Equation* (37)*,* $\varphi$ *is referred to as an* **a***-refinable function. Given bounded, real-valued input data* $(x_i)_{i\in\mathbb{Z}^s}$*, the limit function may be written as*

$$\widetilde{S}^\infty x(t) = \sum_{j\in\mathbb{Z}^s} \varphi(t - j)x_j.$$

*In particular,* $\varphi = \widetilde{S}^\infty\delta$*, where* $\delta$ *denotes the Dirac distribution on the origin.*

We decompose the proof of Theorem 4.5 into a series of lemmas, the first of which relates the convergence of $\widetilde{S}$ to the convergence of the so-called *cascade algorithm.* The idea behind this algorithm designed to construct an **a**-refinable function $\varphi$ is elegant and simple - define the operator $T : C(\mathbb{R}^s, \mathbb{R}) \to C(\mathbb{R}^s, \mathbb{R})$ via

$$Tf(t) := \sum_{i\in\mathbb{Z}^s} a_i f(2t - i),$$

for any continuous $f : \mathbb{R}^s \to \mathbb{R}$. Then obviously if the iterates of $T$ applied to some $f \in C(\mathbb{R}^s, \mathbb{R})$ would converge to some continuous function

$$\varphi = \lim_{n\to\infty} T^n f$$

in the supremum norm, $\varphi$, as a fixed point of $T$, would satisfy the refinement equation (37).

**Lemma 4.6.** *Let $a^{(n)} = \widetilde{S}^n\delta$ as in (36). Then it holds that*

$$T^n f(t) = \sum_{i\in\mathbb{Z}^s} a_i^{(n)} f(2^n t - i). \tag{39}$$

*Proof.* To simplify notation set $p = p^{\mathbf{a}}$, where $p^{\mathbf{a}}$ is given by (35). Recall that $p^{(n)}(i,j) = a^{(n)}_{i-2^n j}$. The Chapman-Kolmogorov equations (33) imply

$$\begin{aligned}\sum_j a^{(n)}_{i-2^n j} a_j &= \sum_j p^{(n)}(i,j)p(j,0)\\ &= p^{(n+1)}(i,0) = a_i^{(n+1)}.\end{aligned}$$

Obviously (39) is true in case $n = 0$, so the general case follows from induction:

$$\begin{aligned}T^{n+1}f(t) &= \sum_{j\in\mathbb{Z}^s} a_j T^n f(2t-j)\\ &= \sum_{j\in\mathbb{Z}^s} a_j \big(\sum_{i\in\mathbb{Z}^s} a_i^{(n)} f(2^{n+1}t - 2^n j - i)\big)\\ &= \sum_{i\in\mathbb{Z}^s}\big(\sum_{j\in\mathbb{Z}^s} a^{(n)}_{i-2^n j} a_j\big) f(2^{n+1}t - i)\\ &= \sum_{i\in\mathbb{Z}^s} a_i^{(n+1)} f(2^{n+1}t - i).\end{aligned}$$ □

**Lemma 4.7.** *Suppose $\widetilde{S}$ converges, and assume* supp($\mathbf{a}$) *is bounded. Then $\widetilde{S}^\infty\delta$ has compact support.*

*Proof.* Define $\Omega := \operatorname{supp}(\mathbf{a})$. Observe that $a_i^{(2)} = \sum_{j\in\mathbb{Z}^s} a_{i-2j}a_j \neq 0$ implies that there is $j \in \Omega$ such that $i - 2j \in \Omega$. In particular, $\operatorname{supp}(a^{(2)}) \subseteq \Omega + 2\Omega$. Following the same lines and using induction, one shows that

$$\operatorname{supp}(a^{(n)}) \subseteq \Omega + 2\Omega + \cdots + 2^{n-1}\Omega.$$

Now if $t = i/2^n$ is a dyadic real number such that $\widetilde{S}^\infty \delta(t) \neq 0$, we may assume $n$ large enough for $a_i^{(n)} \neq 0$ to hold true. In particular, this implies $i \in \Omega + 2\Omega + \cdots + 2^{n-1}\Omega$ and hence

$$t \in \frac{1}{2}\Omega + \cdots + \frac{1}{2^n}\Omega.$$

Moreover, in case $\Omega$ is bounded, say $\|x\| \leq R$ for $x \in \Omega$, it follows that

$$\|t\| \leq R,$$

proving the claim. □

**Lemma 4.8.** *Suppose $\widetilde{S}$ converges. Then the cascade algorithm converges in the sense that for any continuous, compactly supported $f : \mathbb{R}^s \to \mathbb{R}$ satisfying*

$$\sum_{i \in \mathbb{Z}^s} f(i) = 1, \tag{40}$$

*the limit $\lim_{n\to\infty} T^n f$ in the supremum norm exists and is compactly supported. Moreover, for any $f$ as specified above, the limit of the cascade algorithm applied to $f$ satisfies*

$$\lim_{n\to\infty} T^n f = \widetilde{S}^\infty \delta.$$

*Proof.* For $n \in \mathbb{N}$ and any $i \in \mathbb{Z}^s$ we observe

$$|\widetilde{S}^\infty \delta(i/2^n) - T^n f(i/2^n)| \leq |\widetilde{S}^\infty \delta(i/2^n) - \widetilde{S}^n \delta_i| + |\widetilde{S}^n \delta_i - T^n f(i/2^n)|$$

The first term on the right hand side of this inequality converges to 0 uniformly in $i$ as $n \to \infty$, so all we have to take care of is the second summand: By (39), it holds that

$$\begin{aligned} T^n f(t) &= \sum_{j \in \mathbb{Z}^s} a_j^{(n)} f(2^n t - j) \\ &= \sum_{j \in \mathbb{Z}^s} \widetilde{S}^n \delta_j f(2^n t - j). \end{aligned}$$

Using this basic fact as well as $\sum_{j \in \mathbb{Z}^s} f(j) = 1$, one concludes

$$\begin{aligned} |\widetilde{S}^n \delta_i - T^n f(i/2^n)| &= |\sum_{j \in \mathbb{Z}^s} \left(\widetilde{S}^n \delta_i - \widetilde{S}^n \delta_j\right) f(i-j)| \\ &\leq M \sup_{j\,:\,(i-j) \in \mathrm{supp}(f)} |\widetilde{S}^n \delta_i - \widetilde{S}^n \delta_j|, \end{aligned}$$

where $M = \sum_{j\in\mathbb{Z}^s} |f(j)|$. The last essential estimate is

$$|\widetilde{S}^n\delta_i - \widetilde{S}^n\delta_j| \le |\widetilde{S}^n\delta_i - \widetilde{S}^\infty\delta(i/2^n)| + |\widetilde{S}^\infty\delta(i/2^n) - \widetilde{S}^\infty\delta(j/2^n)| + |\widetilde{S}^n\delta_j - \widetilde{S}^\infty\delta(j/2^n)|,$$

which, together with the uniform continuity of $\widetilde{S}^\infty\delta$ proves the claim. □

**Lemma 4.9.** *Suppose $\widetilde{S}$ converges. Then the basic sum rule* (2) *holds true, meaning that for all $i \in \mathbb{Z}^s$ one has*

$$\sum_{j\in\mathbb{Z}^s} a_{i-2j} = 1.$$

*Proof.* Choose initial data $x \in \ell^\infty(\mathbb{Z}^s, \mathbb{R})$ mapped to a nonvanishing function by the limit operator, and, given $i_0 \in \mathbb{Z}^s$, select a dyadic vector $i/2^{n+1}$ such that $\widetilde{S}^\infty x(i/2^{n+1}) \neq 0$ and $i - i_0 \equiv 0 \mod 2$. The equation

$$\widetilde{S}^{n+1}x_i = \sum_{j\in\mathbb{Z}^s} a_{i-2j}\widetilde{S}^n x_j \tag{41}$$

points at the further proof strategy–given $\varepsilon > 0$, choose $n \in \mathbb{N}$ large enough for $|\widetilde{S}^k x_j - \widetilde{S}^\infty x(j/2^k)| < \varepsilon$ to hold, for arbitrary $j \in \mathbb{Z}^s$ and $k \ge n$. Since the input data sequence $x$ without loss of generality has compact support, $x_j \neq 0 \implies \|j\| \le R$, we may modify the choice of $n \in \mathbb{N}$ in a way that guarantees

$$\|t - s\| \le \frac{R}{2^{n+1}} \quad \implies \quad |\widetilde{S}^\infty x(t) - \widetilde{S}^\infty x(s)| < \varepsilon.$$

Omitting the details, it is clear from the above considerations, in addition to $a_{i-2j} \neq 0 \implies \|\frac{i}{2^{n+1}} - \frac{j}{2^n}\| \le \frac{R}{2^{n+1}}$, that, heuristically, (41) implies

$$\begin{aligned}\widetilde{S}^\infty x(i/2^{n+1}) \cdot \sum_j a_{i-2j} &= \sum_j a_{i-2j}\widetilde{S}^\infty x(i/2^{n+1}) \approx \sum_j a_{i-2j}\widetilde{S}^\infty x(j/2^n)\\ &\approx \sum_j a_{i-2j}\widetilde{S}^n x_j = \widetilde{S}^{n+1}x_i\\ &\approx \widetilde{S}^\infty x(i/2^{n+1}),\end{aligned}$$

and thus $\sum_{j\in\mathbb{Z}^s} a_{i-2j} = \sum_{j\in\mathbb{Z}^s} a_{i_0-2j} = 1$ as desired. □

**Lemma 4.10.** *Suppose $\widetilde{S}$ converges, and set $\varphi := \widetilde{S}^\infty\delta$. Then*

$$\sum_{i\in\mathbb{Z}^s} \varphi(t - j) = 1 \quad \textit{for } t \in \mathbb{R}^s.$$

*Moreover, for each bounded input data sequence $x \in \ell^\infty(\mathbb{Z}^s, X)$ the limit function $\widetilde{S}^\infty x$ takes the following form:*

$$\widetilde{S}^\infty x(t) = \sum_{i \in \mathbb{Z}^s} x_i \varphi(t-i).$$

*Proof.* The property of $(\varphi(\cdot - i))_{i\in\mathbb{Z}^s}$ as a partition of unity follows readily from the basic sum rule. Following the same reasoning that led to equation (39) and using the refinement equation $\sum_{i\in\mathbb{Z}^s} a_i\varphi(2t-i) = \varphi(t)$, one shows that

$$\sum_{i\in\mathbb{Z}^s} x_i \varphi(t-i) = \sum_{i\in\mathbb{Z}^s} \widetilde{S}^n x_i \varphi(2^n t - i).$$

This, together with $\sum_{i\in\mathbb{Z}^s}\varphi(t-i) = 1$ easily leads to the desired form of $\widetilde{S}^\infty x$. □

### 4.2.2 The impact of nonlinearity

Assuming that conditional expectations of bounded random variables mapping to the metric space $X$ are well-defined in the sense of equation (18), and in addition satisfy the tower property (24), we could deduce from Theorem 4.4

$$\begin{aligned} S^n x \circ X_0 &= E(x \circ X_n^{\mathbf{a}} ||| \mathfrak{F}_0) \\ &= E(x \circ X_n^{\mathbf{a}} | \mathfrak{F}_0) \\ &= \operatorname{argmin}\Big(\sum_j \big(p^{\mathbf{a}}\big)^{(n)}(X_0, j) d^2(x_j, \cdot)\Big). \end{aligned}$$

Recall that $\big(p^{\mathbf{a}}\big)^{(n)}(i,j)$, the $n$-step transition probabilities of $(X_k^{\mathbf{a}})_{k\in\mathbb{N}_0}$, can be viewed as $(\widetilde{S}^n\delta)_{i-2^n j}$, where $\widetilde{S}$ denotes the linear counterpart to $S$, and $\delta$ the Dirac delta on the origin, cf. Theorem 4.5. Thus, the assumption of the tower property would immediately imply that every scheme converging for linear input data would converge on $X$ as well. Indeed, the limit functions for given input data $x \in \ell^\infty(\mathbb{Z}^s, X)$ would satisfy

$$S^\infty x(t) = \operatorname{argmin}(\sum_j \varphi(t-j) d^2(x_j, \cdot)),$$

where, as above, $\varphi = \widetilde{S}^\infty \delta$, leading to a complete analogy to the linear case. However, nonlinear conditioning does not obey the tower rule. The above observations demonstrate that this lack of property (24) constitutes the need for a further discussion of convergence.

### 4.2.3 Proximity and Contractivity

We describe how classical convergence arguments involving contractivity and proximity properties, which constitute the backbone of the analysis of subdivision schemes, carry over to the NPC setting. The first result of this section is Theorem 1 of [9].

**Theorem 4.11.** *Let $S, T$ be refinement schemes acting on data from a Hadamard space $X$. Then $S$ converges under the following assumptions:*

*(i) There is a function $D : \ell^\infty(\mathbb{Z}^s, X) \to \mathbb{R}_{\geq 0}$, a nonnegative real number $\gamma < 1$ and a positive integer $n_0$ such that*

$$D(S^n x) \leq \gamma^{[n/n_0]} D(x) \tag{42}$$

*for $x \in \ell^\infty(\mathbb{Z}^s, X)$ and $n \in \mathbb{N}$. Here $[\,\cdot\,]$ denotes the floor function.*

*(ii) $T$ is convergent and satisfies*

$$d_\infty(Tx, Ty) \leq d_\infty(x, y)$$

*for $x, y \in \ell^\infty(\mathbb{Z}^s, X)$.*

*(iii) There is $C \geq 0$ such that*

$$d_\infty(Sx, Tx) \leq C \cdot D(x)$$

*for $x \in \ell^\infty(\mathbb{Z}^s, X)$.*

*Proof.* We set $f_n(y) := T^\infty(S^n x)(2^n y)$ and claim that this defines a Cauchy sequence in $(C(\mathbb{R}^s, X), d_\infty)$. Note first that given $n \in \mathbb{N}$ and $y \in \mathbb{R}^s$, by continuity of $f_n$ respectively $f_{n+1}$, we find $j \in \mathbb{Z}^s$ and $m \in \mathbb{N}$ such that

$$d(f_r(y), f_r(2^{-m} j)) < C \cdot D(x) \gamma^{[n/n_0]} \quad \text{for } r = n, n+1. \tag{43}$$

Moreover, due to convergence of $T$, by multiplying both the numerator and the denominator of the number $j/2^m$ with a power of two if necessary we may assume $m$ to be sufficiently large for

$$d(f_r(2^{-m} j), T^{m-r}(S^r x)_j) = d(T^\infty S^r x(2^{r-m} j), T^{m-r}(S^r x)_j) < C \cdot D(x) \gamma^{[n/n_0]}$$

to hold for $r = n, n+1$, in addition to (43). This together with (i) and (ii) implies

$$\begin{aligned} d(f_n(y), f_{n+1}(y)) &\leq d(f_n(y), f_n(2^{-m}j)) \\ &+ d(f_n(2^{-m}j), T^{m-n}S^n x_j) \\ &+ d(T^{m-n}S^n x_j, T^{m-n-1}S^{n+1}x_j) \\ &+ d(T^{m-n-1}S^{n+1}x_j, f_{n+1}(2^{-m}j)) \\ &+ d(f_{n+1}(2^{-m}j), f_{n+1}(y)) \\ &< 5C \cdot D(x)\gamma^{[n/n_0]}, \end{aligned}$$

showing that $f_n$ is a Cauchy sequence. Since $X$ is complete, we find a continuous $f : \mathbb{R}^s \to X$ with $f_n \to f$ uniformly. We claim that $S^n x$ converges to $f$ in the sense of (1). For $m \geq n$ and $j \in \mathbb{Z}^s$, we obtain the inequality

$$\begin{aligned} d(T^{m-n}S^n x_j, S^m x_j) &\leq \sum_{k=n}^{m-1} d(T^{m-k}S^k x_j, T^{m-k-1}S^{k+1}x_j) \\ &\leq \sum_{k=n}^{m-1} \gamma^{[k/n_0]} \cdot D(x)C \leq \gamma^{[n/n_0]} \left(\frac{n_0 D(x)C}{1-\gamma}\right), \end{aligned}$$

which together with

$$d(f_n(2^{-m}j), S^m x_j) \leq d(f_n(2^{-m}j), T^{m-n}S^n x_j) + d(T^{m-n}S^n x_j, S^m x_j)$$

establishes the claim. □

**Definition 4.12.** In accordance with [9], we call a scheme $S$ satisfying (42) *weakly contractive*. Thus, a weakly contractive scheme is contractive if and only if $n_0 = 1$.

In the following we rely on the nonlinear version of Jensen's inequality introduced in Chapter 2:

**Theorem** (Conditional Jensen's inequality, [19])**.** *Suppose $\psi : X \to \mathbb{R}$ is a convex, lower semicontinuous function on a Hadamard space $(X, d)$, and $(\Omega, \mathfrak{F}, \mathbb{P})$ is a probability space. Moreover suppose $(\mathfrak{F}_k)_{k\in\mathbb{N}_0}$ is a filtration in $\mathfrak{F}$. Then for each bounded, $\mathfrak{F}_N$-measurable random variable $Y : \Omega \to X$ the following holds true:*

$$\psi(E(Y|||(\mathfrak{F}_k)_{k\geq n})) \leq E(\psi(Y)|\mathfrak{F}_n). \tag{44}$$

The following lemma shows that barycentric schemes with convergent linear counterpart are always weakly contractive.

**Lemma 4.13** ([8]). *Suppose the linear scheme associated to $(a_i)_{i\in\mathbb{Z}^s}$ converges, and* $\operatorname{supp}(\mathbf{a}) \subseteq \Omega$, *with $\Omega$ bounded, convex and balanced. Denote by $\rho : \mathbb{R}^s \to \mathbb{R}_{\geq 0}$ the Minkowski functional of $\Omega$. Furthermore define $D : \ell^\infty(\mathbb{Z}^s, X) \to \mathbb{R}_{\geq 0}$ via*

$$D(x) = \sup_{\rho(i-j)<2} d(x_i, x_j).$$

*Then the barycentric scheme $S$ associated to $(a_i)_{i\in\mathbb{Z}^s}$ is weakly contractive with respect to $D$.*

*Proof.* The Hadamard property implies that for each $z_0 \in X$ the function

$$X \to \mathbb{R}_{\geq 0}; \quad z \mapsto d(z, z_0),$$

which clearly is continuous, is convex as well. Thus, by Jensen's inequality (44) and Theorem 4.4,

$$d(S^n x \circ X_0, z_0) = d(E(x \circ X_n^{\mathbf{a}} || \mathfrak{F}_0), z_0) \leq E(d(x \circ X_n^{\mathbf{a}}, z_0)|\mathfrak{F}_0). \tag{45}$$

Recall that the transition kernel of $X_n^{\mathbf{a}}$ takes the form

$$(P^{\mathbf{a}})^k = \left(a^{(k)}_{i-2^k j}\right)_{i,j\in\mathbb{Z}^s}.$$

Thus Proposition 4.2 implies $E(d(X_n^{\mathbf{a}}, z_0)|\mathfrak{F}_0) = \sum_{k\in\mathbb{Z}^s} a^{(n)}_{X_0^{\mathbf{a}}-2^n k} d(x_k, z_0)$. Together with (45) this gives

$$d(S^n x_i, z_0) \leq \sum_{k\in\mathbb{Z}^s} a^{(n)}_{i-2^n k} d(x_k, z_0) \quad \text{for all } i \in \mathbb{Z}^s.$$

Substituting $S^n x_j$ for $z_0$, we deduce

$$\begin{aligned} d(S^n x_i, S^n x_j) &\leq \sum_{k\in\mathbb{Z}^s} a^{(n)}_{i-2^n k} d(x_k, S^n x_j) \\ &\leq \sum_{k\in\mathbb{Z}^s, \ell\in\mathbb{Z}^s} a^{(n)}_{i-2^n k} a^{(n)}_{j-2^n \ell} d(x_k, x_\ell). \end{aligned}$$

The fact that the support of $(a_i)_{i\in\mathbb{Z}^s}$ is contained in the balanced, convex and bounded set $\Omega$ together with the recursion $a_i^{(n)} = \sum_j a_{i-2j} a_j^{(n-1)}$ implies $\operatorname{supp}(a^{(n)}) \subseteq (2^n - 1)\Omega$, see [6] and Lemma 4.7.

Since the linear subdivision scheme with mask $(a_i)_{i\in\mathbb{Z}^s}$ converges, we find a refinable function $\varphi : \mathbb{R}^s \to \mathbb{R}$ satisfying (37) and (38), cf. Theorem 4.5. Recall that one obtains this refinable function as the limit of the linear scheme acting on the input data $y_j = \delta_{j0}$, cf. Theorem 4.5 :

$$\sup_i |a_i^{(n)} - \varphi(i/2^n)| = \varepsilon_n \underset{n\to\infty}{\longrightarrow} 0. \tag{46}$$

Accordingly, setting $A = \{(k,\ell) \in \mathbb{Z}^s \times \mathbb{Z}^s \mid \max(\rho(i-2^n k), \rho(j-2^n\ell)) \le 2^n - 1\}$, we obtain

$$\begin{aligned} d(S^n x_i, S^n x_j) &\le \sum_{(k,\ell)\in A} a^{(n)}_{i-2^nk} a^{(n)}_{j-2^n\ell} d(x_k, x_\ell) \\ &\le \sum_{(k,\ell)\in A} \varphi(i/2^n - k)\varphi(j/2^n - \ell) d(x_k, x_\ell) \\ &\quad + \varepsilon_n \left( \sum_{(k,\ell)\in A} (a^{(n)}_{i-2^nk} + a^{(n)}_{j-2^n\ell}) d(x_k, x_\ell) \right) \\ &\quad + \varepsilon_n^2 \left( \sum_{(k,\ell)\in A} d(x_k, x_\ell) \right). \end{aligned} \tag{47}$$

Now, if $i, j, k, \ell \in \mathbb{Z}^s$ are such that $\rho(i-j) < 2$, $\rho(i-2^nk) \le 2^n - 1$ and $\rho(j-2^n\ell) \le 2^n - 1$, one concludes

$$\begin{aligned} \rho(k-\ell) &\le \frac{1}{2^n}(\rho(i-2^nk) + \rho(i-j) + \rho(j-2^n\ell) \\ &< \frac{1}{2^n}(2(2^n-1)+2) = 2. \end{aligned} \tag{48}$$

Define

$$\psi(s,t) = \sum_{i\in\mathbb{Z}^s} \varphi(t-i)\varphi(s-i).$$

Then, since the refinable function is uniformly continuous, the property (38) implies that for $n$ large enough,

$$\alpha_n = \inf_{\rho(t-s)<2^{-n+1}} \psi(s,t) > \varepsilon > 0. \tag{49}$$

By boundedness of $\Omega$ we also obtain

$$M = \sup_{t\in\mathbb{R}^s} |\mathbb{Z}^s \cap (t+\Omega)| < \infty. \tag{50}$$

Combining (38) with (47) through (50) further gives

$$D(S^n x) = \sup_{\rho(i-j)<2} d(S^n x_i, S^n x_j) \leq \gamma_n D(x), \tag{51}$$

where $\gamma_n = (1-\alpha_n+2\varepsilon_n+M^2\varepsilon_n^2)$. Clearly, for $n_0$ large enough, $\gamma = \gamma_{n_0} < 1$. Moreover, the estimate (51) is uniform in $x$ (and even $d$). The same argument leading to the first inequality in (47) together with (48) provides

$$D(S^m x) \leq D(S^k x) \quad \text{for } m \geq k.$$

Thus, for $n \in \mathbb{N}$ one concludes:

$$\begin{aligned} D(S^n x) &\leq \gamma D(S^{n-n_0} x) \\ &\leq \gamma^{[n/n_0]} D(S^{n-n_0[n/n_0]}) \\ &\leq \gamma^{[n/n_0]} D(x), \end{aligned}$$

which completes the proof. □

In Lemma 4.13 it was proven that every convergent linear scheme gives rise to a weakly contractive barycentric scheme. We now show how strong contractivity in some instances follows from special properties of the mask's support and how the mask influences the shape of the contractivity constant.

**Lemma 4.14.** *Let $\alpha_1, \dots \alpha_n \geq 0$ and $\beta_1, \dots, \beta_n \geq 0$ be probability distribution functions on $\Omega = \{0, \dots, n\}$. Then there exists a coupling $\pi : \Omega \times \Omega \to [0,1]$ of $\alpha$ and $\beta$ such that $\pi_{ii} = \min(\alpha_i, \beta_i)$.*

*Proof.* Without restriction of generality suppose $\alpha_i = \min(\alpha_i, \beta_i)$ for $i = 1, \dots, k$ and $\beta_j = \min(\alpha_j, \beta_j)$ for $j = k+1 \dots, n$. Then obviously

$$c = \sum_{i=1}^{k} (\beta_i - \alpha_i) = \sum_{i=k+1}^{n} (\alpha_i - \beta_i).$$

Excluding the trivial case $k = n$, we may assume $c > 0$. Define

$$\pi_{ij} = \begin{cases} \frac{(\beta_i-\alpha_i)(\alpha_j-\beta_j)}{c} & \text{for } i = 1, \dots, k \text{ and } j = k+1, \dots, n \\ \min(\alpha_i, \beta_i) & \text{for } 1 \leq i = j \leq n \\ 0 & \text{otherwise.} \end{cases}$$

One readily verifies that $\pi$ constitutes a coupling of $\alpha$ and $\beta$, and by definition satisfies $\pi_{ii} = \min(\alpha_i, \beta_i)$. □

The next proposition establishes a class of strongly contractive schemes on Hadamard spaces, including the ones generating splines of arbitrary degree. This result and its linear counterpart, see Proposition 3.1 in [6], are equally powerful in identifying contractivity.

**Proposition 4.15** ([9])**.** *Suppose $S$ and the corresponding $D$ are as in Lemma 4.13. Then we have*

$$D(Sx) \leq \gamma D(x),$$

*where*

$$\gamma = 1 - \min_{\rho(i-j)<2} \left( \sum_{k\in\mathbb{Z}^s} \min(a_{i-2k}, a_{j-2k}) \right).$$

*In particular, if for each $i, j \in \mathbb{Z}^s$ with $\rho(i-j) < 2$ one finds $k \in \mathbb{Z}^s$ such that $i - 2k \in \operatorname{supp}(a)$ and $j - 2k \in \operatorname{supp}(a)$, then $\gamma < 1$ and hence $S$ is contractive w.r.t. $D$.*

*Proof.* Introducing probability distributions $\alpha^i$ by letting $\alpha^i_k = a_{i-2k}$, we have $Sx_i = b(x_*\alpha^i)$ for $i \in \mathbb{Z}^s$. Thus, with the notation

$$\eta_{ij} = \sum_{k\in\mathbb{Z}^s} \min(a_{i-2k}, a_{j-2k}),$$

Theorem (28) implies

$$\begin{aligned} D(Sx) &= \sup_{\rho(i-j)<2} d(b(x_*\alpha^i), b(x_*\alpha^j)) \\ &\leq \sup_{\rho(i-j)<2} [(1-\eta_{ij}) \max_{\rho(i-2k),\rho(i-2\ell)\leq 1} d(x_k, x_\ell)]. \end{aligned} \tag{52}$$

Certainly $\rho(i-j) < 2$ together with $\rho(i-2\ell) \leq 1$ and $\rho(j-2k) \leq 1$ implies

$$\rho(k-\ell) \leq \rho(k - \frac{j}{2}) + \rho(\frac{\ell}{2} - i) + \rho(\frac{1}{2}(i-j)) < 2.$$

Combining this with (52), we obtain $D(Sx) \leq \gamma D(x)$ as required. □

Proposition 4.15 provides us with a contractivity criterion that solely depends on the structure of the mask's support. Thus every linear scheme seen to be contractive using the linear version of the above proposition possesses a contractive barycentric analogue. In particular, this applies to the class of schemes identified in chapter 3 of [6], see Corollary 4.16 below. Recall that a *centered zonotope* is defined as $Z(X) = \{Xu \mid u \in \mathbb{R}^n,\ \|u\|_\infty \leq 1\}$ with $X \in \mathbb{Z}^{s\times n}$, $n > s$. $Z(X)$ is called *unimodular* if and only if each $s \times s$ – minor of $X$ has determinant $-1$, 0, or 1, and $\operatorname{rank}(X) = s$.

**Corollary 4.16** ([9]). *Suppose the barycentric scheme $S$ possesses a mask whose support is an integer quad with edges of length at least 2, or supp$(a) = Z(X) \cap \mathbb{Z}^s$ with $Z(X)$ unimodular. Then $S$ is contractive w.r.t. some contractivity function $D$ admissible in the sense of* (4).

*Proof.* This is a direct consequence of the proofs of Theorems 3.3 and 3.4 of [6], and Proposition 4.15. □

Recall that the *tensor product* $(a \otimes b)_{i \in \mathbb{Z}^{s+t}}$of two masks $(a_i)_{i \in \mathbb{Z}^s}$ and $(b_j)_{j \in \mathbb{Z}^t}$ is defined by

$$(a \otimes b)_{(i,j)} = a_i \cdot b_j.$$

The next lemma identifies linear B-spline subdivision as a model scheme suitable for our convergence analysis.

**Lemma 4.17** ([9]). *Suppose $S$ and the corresponding $D$ are as in Lemma 4.13. Define $(b_i)_{i \in \mathbb{Z}}$ via $b_0 = 1$, $b_{-1} = b_1 = \frac{1}{2}$, and $b_i = 0$ for $|i| > 1$. Let $T$ denote the barycentric scheme associated to the $s$-fold tensor product $b \otimes \cdots \otimes b$. Then $T$ is Lipschitz in the sense of* (5) *and converges on any Hadamard space. Moreover, there is $C > 0$ such that $d_\infty(Sx, Tx) \leq C \cdot D(x)$ for all $x \in \ell^\infty(\mathbb{Z}^s, X)$.*

*Proof.* We begin by proving convergence. Define $D_\infty(x) = \sup_{\|i-j\| \leq 1} d(x_i, x_j)$. By Corollary 4.16, $D_\infty(Tx) \leq \gamma D_\infty(x)$, with $\gamma < 1$. For $n \in \mathbb{N}_0$, define $f_n : \mathbb{R}^s \to X$ as follows:

1. For $t \in \mathbb{R}$, set $\varphi_0(t) = \max\{1 - |t|, 0\}$ and define $\varphi(t_1, \ldots, t_s) = \prod_i \varphi_0(t_i)$.
2. Set $f_n(\zeta) = \operatorname{argmin}(\sum_k \varphi(2^{n-1}\zeta - k) d(\,\cdot\,, T^{n-1}x_k)^2)$.

This function is continuous since the center of mass depends continuously on the weights. Moreover, $f_r(j/2^r) = T^r x_j$ for each $j \in \mathbb{Z}^s$ by construction of $\varphi$. Suppose $\zeta \in \mathbb{R}^s$ is contained in some dyadic cube $Q = \prod [k_i 2^{-r+1}, (k_i + 1)2^{-r+1}]$, where $k_i \in \mathbb{Z}$. Clearly $\varphi \equiv 0$ outside $\{\xi \in \mathbb{R}^s \mid \|\xi\|_\infty < 1\}$, from which we conclude that

$$f_r(\zeta) = \operatorname{argmin} \sum_{v \in V(Q)} \varphi(2^{r-1}(\zeta - v)) d(\cdot, T^{r-1}x(v))^2,$$

where $V(Q)$ denotes the vertex set of $Q$. Applying the inequality (28), we obtain

$$\begin{aligned}\max_{v\in V(Q)} d(T^{r-1}x(v), f_r(\zeta)) &\le \max_{v,w\in V(Q)} d(T^{r-1}x(v), T^{r-1}x(w)) \\ &\le D_\infty(T^{r-1}x) \le \gamma^{r-1} D_\infty(x).\end{aligned} \tag{53}$$

Certainly, every dyadic cube of edge length $2^{-n}$ shares a vertex with a dyadic cube of edge length $2^{-n+1}$. Together with (53) applied to $r = n, n+1$, this implies $d(f_n(\zeta), f_{n+1}(\zeta)) < 2\gamma^{n-1}D_\infty(x)$. It is straightforward to show that $f := \lim_n f_n$ is a uniform limit of $T^n x$.

Lipschitz continuity is an easy consequence of inequality (28). Indeed, denoting the mask of $T$ by $b$, for $i \in \mathbb{Z}^s$ we have

$$d(Tx_i, Ty_i) \le \sum_{j\in\mathbb{Z}^s} b_{i-2j} d(x_j, y_j) \le d_\infty(x, y),$$

since $T$ is affine invariant, i.e. $\sum_j b_{i-2j} = 1$.

The proximity inequality $d_\infty(Sx, Tx) \le C \cdot D(x)$ is proven along the same lines, for details see [9].ity $d_\infty(Sx, Tx) \le C \cdot D(x)$ is proven along the same lines, for details see [9]. □

### 4.2.4 Proof of Theorem 1.2

Suppose $S$ denotes the barycentric scheme associated to the nonnegative mask $(a_i)_{i\in\mathbb{Z}^s}$. Under the assumption that the linear counterpart of $S$ converges, combining Lemmas 4.13 and 4.17, we find a function $D : \ell^\infty(\mathbb{Z}^s, X) \to \mathbb{R}_{\ge 0}$, a convergent scheme $T : \ell^\infty(\mathbb{Z}^s, X) \to \ell^\infty(\mathbb{Z}^s, X)$, and constants $\gamma < 1$ and $C \ge 0$ such that

(i) There is a positive integer $n_0$ such that $D(S^n x) \le \gamma^{[n/n_0]} D(x)$ for $x \in \ell^\infty(\mathbb{Z}^s, X)$ and $n \in \mathbb{N}$.

(ii) $T \in \mathrm{Lip}_1(\ell^\infty(\mathbb{Z}^s, X))$ is convergent.

(iii) $d_\infty(Sx, Tx) \le C \cdot D(x)$ for $x \in \ell^\infty(\mathbb{Z}^s, X)$.

Thus, by Theorem 4.11, the scheme $S$ converges. □

*Remark* 4.18. Statements relating the convergence of a nonlinear subdivision scheme to the convergence of its linear counterpart have been obtained in the smooth setting in [11, 25], although with severe restrictions on the density of the input data. As for convergence theorems applying to arbitrary bounded input data, initial univariate results from [23] have been substantially extended in the articles [9, 8] presented within this thesis. To the best of our knowledge the latter articles also first consider subdivision algorithms in metric spaces without differentiable structure. Another recent result on the convergence of a small class of interpolatory schemes on smooth manifolds without restriction on the input data can be found in [21]. Independent of the author, Jetter and Li in their article [13] recently reproved and extended results on *linear* subdivision schemes using notions from the theory of Markov chains.

### 4.2.5 A characterization of convergence

A well-known result from the linear theory is the following:

**Proposition 4.19.** *The* univariate *and* linear *scheme $\widetilde{S}$ associated to the mask $(a_i)_{i\in\mathbb{Z}}$ converges if and only if there is $\gamma < 1$ and $C \geq 0$ such that*

$$\sup_{i\in\mathbb{Z}} |\widetilde{S}^n x_i - \widetilde{S}^n x_{i+1}| \leq C \cdot \gamma^n \sup_{i\in\mathbb{Z}} |x_i - x_{i+1}| \quad \textit{for all } n \in \mathbb{N}_0.$$

Theorems 1.2 and 4.11 along with Lemma 4.13 put us in a position to generalize this statement to the setting of Hadamard spaces. Still we need an easy auxiliary result.

**Lemma 4.20.** *Suppose $(X, d)$ is a metric space, and let*

$$D_\infty(x) = \sup_{\|i-j\|_\infty \leq 1} d(x_i, x_j).$$

*Then a refinement scheme $S$ is weakly contractive with respect to an admissible contractivity function if and only if there is $\gamma < 1$ and $C \geq 0$ such that*

$$D_\infty(S^n x) \leq C\gamma^n D_\infty(x).$$

*Proof.* Suppose $S$ is weakly contractive with respect to $D_\Omega$, meaning there is $n_0 \in \mathbb{N}$ and $\widetilde{\gamma} < 1$ such that $D \circ S^n \leq \widetilde{\gamma}^{(n/n_0)} D$. It is not difficult to see (cf. [9]) that there are $r, R > 0$ such that

$$rD_\Omega \leq D_\infty \leq RD_\Omega.$$

Observe that, since $[n/n_0] \geq n/n_0 - 1$ one has $\widetilde{\gamma}^{[n/n_0]} \leq \widetilde{\gamma}^{n/n_0 - 1} = \widetilde{C}\gamma^n$, where $\widetilde{C} = \widetilde{\gamma}^{-1}$ and $\gamma = \widetilde{\gamma}^{1/n_0} < 1$. Moreover define $C = \frac{R\widetilde{C}}{r}$. Then

$$\begin{aligned} D_\infty \circ S^n &\leq RD_\Omega \circ S^n \leq R\widetilde{\gamma}^{[n/n_0]} D_\Omega \\ &\leq \frac{R}{r}\widetilde{C}\gamma^n D_\infty = C\gamma^n D_\infty. \end{aligned}$$

Now assume there is $\gamma < 1$ and $C \geq 0$ such that

$$D_\infty \circ S^n \leq C\gamma^n D_\infty.$$

Choose $n_0 \in \mathbb{N}$ such that $\frac{CR}{r}\gamma^{n_0} \leq 1$. From (44) it follows that $D_\Omega \circ S^n \leq D_\Omega$. On the other hand, for $n \geq n_0$ we have $[n/n_0] \leq n - n_0$ and thus

$$\begin{aligned} D_\Omega \circ S^n &\leq \frac{1}{r} D_\infty \circ S^n \leq \frac{C}{r}\gamma^n D_\infty \\ &\leq \left(\frac{RC}{r}\gamma^{n_0}\right)\gamma^{n-n_0} D_\Omega \leq \gamma^{[n/n_0]} D_\Omega. \end{aligned}$$

□

We are now able to generalize Proposition 4.19:

**Theorem 4.21** ([8])**.** *The refinement scheme associated to $(a_i)_{i\in\mathbb{Z}^s}$ converges on arbitrary Hadamard spaces if and only if there is $C \geq 0$ and $\gamma < 1$ such that for all $(X, d)$ Hadamard*

$$D_\infty(S^n x) \leq C \cdot \gamma^n D_\infty(x) \quad \textit{for all } x \in \ell^\infty(\mathbb{Z}^s, X),$$

*where, as above, $D_\infty(x) = \sup_{\|i-j\|_\infty \leq 1} d(x_i, x_j)$.*

*Proof.* This follows from combining Lemma 4.20 with Lemma 4.13 and Theorem 4.11.

□

### 4.2.6 Approximation power of barycentric subdivision schemes

We present an approximation result for Lipschitz functions and a statement highlighting the effect of the contractivity constant on the quality of convergence.

**Theorem 4.22** ([9])**.** *Suppose $f : (\mathbb{R}^s, \|\cdot\|) \to (X, d)$ is Lipschitz-continuous with constant $C > 0$, and $S$ is a convergent barycentric scheme whose mask is supported on $\{x \in \mathbb{R}^s \mid \|x\| \leq r\}$. Sample $f$ on the grid $h\mathbb{Z}^s$, $h > 0$, via $x_i = f(hi)$. Then*

$$d_\infty(S^\infty x(h^{-1}\cdot), f(\cdot)) \leq rC \cdot h.$$

*Proof.* Suppose $n \in \mathbb{N}_0$ and $i \in \mathbb{Z}^s$. Then by Theorems 2.35 and 4.4 we obtain

$$\begin{aligned} d(S^n x_{2^{n-k}i}, f(hi/2^k)) &\leq \sum_{\|2^{n-k}i-2^n j\| \leq (2^n-1)r} a^{(n)}_{2^{n-k}i-2^n j} d(x_j, f(hi/2^k)) \\ &\leq \sup_{\|i/2^k - j\| \leq (1-2^{-n})r} d(f(hj), f(hi/2^k)) \\ &\leq rC \cdot h, \end{aligned}$$

from which the claim follows. □

**Theorem 4.23** ([9])**.** *Suppose the barycentric subdivision scheme $S$ converges. Then there is $C > 0$ and $\gamma < 1$ such that*

$$d(S^\infty x(j/2^n), S^n x_j) \leq C \cdot \gamma^n D_\infty(x).$$

*Proof.* Suppose $\Omega$ is a balanced convex set with nonempty interior such that $\mathrm{supp}(a) \subseteq 4\Omega^\circ$. Let $D_\Omega$ denote the admissible contractivity function associated to $\Omega$, cf. (4). Then by (28),

$$d(Sx_{2j}, x_j) \leq \sum_{\rho_\Omega(2(j-k))<4} a_{2(j-k)} d(x_k, x_j) \leq D_\Omega(x).$$

By Theorem 4.21, there is $C_1 > 0$ and $\gamma < 1$ such that

$$D_\infty(S^n x) \leq C_1 \gamma^n D_\infty(x).$$

Choose $C_2$ such that $D_\Omega \leq C_2 D_\infty$. Then

$$d((S^m x)_{2^{m-n}j}, S^n x_j) \leq C_1 C_2 \gamma^n \left( \frac{D_\infty(x)}{1-\gamma} \right). \tag{54}$$

Thus the statement follows by taking the limit in $m$ on the left hand side of equation (54), and setting $C := \frac{C_1 C_2}{1-\gamma}$. □

## 4.3 A note on diffusion tensor subdivision

In this section we address the impact of the strong law of large numbers, cf. Theorem 2.39, on the properties of barycentric subdivision algorithms acting on diffusion tensor valued data. For a survey of mathematical and algorithmic methods in tensor field processing see e.g [24]. Recall that a possible realization of the space of diffusion tensors is the set of positive definite symmetric matrices $\mathcal{P}(n)$. Introducing the metric

$$d(x, y) = \left\| \log(x^{-1/2} y x^{-1/2}) \right\|_F,$$

where $\| \cdot \|_F$ denotes the Frobenius norm, renders the space of diffusion tensors a symmetric space of noncompact type and thus a Hadamard space. Consequently, it is possible to perform barycentric subdivision, with the linear equivalence theorem (Theorem 1.2) at hand. Even more, the limits of convergent schemes enjoy structure preserving features such as invariance under inversion. The central result of the present section is Corollary 4.27, which extends work from [18].

Recall the laws of large numbers 2.39 from Chapter 2:

**Theorem** (The Laws of Large Numbers, [20])**.** *Suppose* $Y_n : (\Omega, \mathfrak{F}, \mathbb{P}) \to (X, d)$ *is an i.i.d. sequence of random variables with values in a Hadamard space. Define recursively*

$$\begin{aligned} S_1 &:= Y_1 \\ S_{n+1} &:= b\left(\frac{1}{n+1}\delta_{Y_{n+1}} + \frac{n}{n+1}\delta_{S_n}\right) = \frac{1}{n+1}Y_{n+1} + \frac{n}{n+1}S_n. \end{aligned}$$

*Then the following hold true:*

**Weak Law of Large Numbers:** *Suppose* $\mathrm{Var}(Y_1) < \infty$*. Then*

$$S_n \to EY_1 \quad \textit{in } L^2.$$

**Strong Law of Large Numbers:** *Suppose* $Y_1$ *is bounded a.s. Then*

$$S_n \to EY_1 \quad \textit{a.s.}$$

*Remark* 4.24. The laws of large numbers as stated above are of great significance in the sense that they provide a way to compute the expected value of a random variable using repeated binary averaging. The impact of this observation is two-fold: First, it is of great value in generalizing facts on binary averages to expected values. Second, it provides a Monte Carlo method to compute expected values of random variables with values in Hadamard spaces whose geodesics are well-understood.

We will also make use of the following straightforward Lemma:

**Proposition 4.25** (Isometries)**.** *Let* $\psi : (X, d) \to (X', d')$ *be an isometry of Hadamard spaces. Then for each random variable* $Y$ *with values in* $X$ *one has*

$$E(\psi(Y)) = \psi(E(Y)).$$

**Theorem 4.26** ([8])**.** *Suppose* $Y : (\Omega, \mathfrak{F}, \mathbb{P}) \to (\mathcal{P}(n), d)$ *is a bounded random variable with values in the space of positive definite symmetric matrices. Moreover, let* $\alpha : (\Omega, \mathfrak{F}, \mathbb{P}) \to \mathcal{P}(1) = \mathbb{R}_{>0}$ *be bounded, and* $Q \in SO(n)$*. Then*

$$E(Y^{-1}) = E(Y)^{-1} \tag{55}$$

$$E(\alpha Y) = E(\alpha)E(Y) \tag{56}$$

$$E(Q^T Y Q) = Q^T E(Y) Q. \tag{57}$$

*Moreover, if* $\det(Y) = 1$, *then* $\det(E(Y)) = 1$.

*Proof.* Equations (55) and (57) follow from the fact that both the inverse function and conjugation with an orthogonal matrix constitute isometries on $(\mathcal{P}(n), d)$. Since $\{x \in \mathcal{P}(n) \mid \det(x) = 1\}$ is a convex and closed subset of $\mathcal{P}(n)$, the last statement follows from the fact that the expected value of a random variable lies within the convex hull of its range. The conformity property (56) is trivial to show in case $\#Y(\Omega) = 2$, for the geodesic joining $x_0$ and $x_1$ is given explicitly as

$$x_t = x_0^{1/2}(x_0^{-1/2} x_1 x_0^{-1/2})^t x_0^{1/2}.$$

Passing to the general case, we invoke the strong law of large numbers (Theorem 2.39). Suppose $Y_n$ is an i.i.d. sequence of random variables with the same law as $Y$, and $\alpha_n$ are i.i.d. versions of $\alpha$. Define $S_n = \frac{1}{n}\sum_{i=1}^n Y_i$ and $\sigma_n = \frac{1}{n}\sum_{i=1}^n \alpha_i$ as in Theorem 2.39. Inductively one shows

$$\frac{1}{n}\sum_{i=1}^{n} \alpha_i Y_i = \sigma_n \cdot S_n. \tag{58}$$

Indeed, the case $n = 1$ being trivial, the induction step reduces to the case of a binary average as above:

$$\begin{aligned}
\frac{1}{n+1}\sum_{i=1}^{n+1} \alpha_i Y_i &= \frac{1}{n+1}\alpha_{n+1}Y_{n+1} + \frac{n}{n+1}\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_i Y_i\right) \\
&= \frac{1}{n+1}\alpha_{n+1}Y_{n+1} + \frac{n}{n+1}\sigma_n \cdot S_n \\
&= \left(\frac{1}{n+1}\alpha_{n+1} + \frac{n}{n+1}\sigma_n\right)\left(\frac{1}{n+1}Y_{n+1} + \frac{n}{n+1}S_n\right) \\
&= \sigma_{n+1}S_{n+1}
\end{aligned}$$

Choosing $\omega \in \Omega$ such that $\sigma_n(\omega) \to E(\alpha)$ and $S_n(\omega) \to E(Y)$, evaluating both sides of equation (58) at $\omega$, and letting $n \to \infty$ in the same equation concludes the proof. □

**Corollary 4.27** ([8]). *Suppose* $S : \ell^\infty(\mathbb{Z}^s, \mathcal{P}(n)) \to \ell^\infty(\mathbb{Z}^s, \mathcal{P}(n))$ *is a convergent barycentric subdivision scheme, and let*

$$S^\infty : \ell^\infty(\mathbb{Z}^s, \mathcal{P}(n)) \to C(\mathbb{R}^s, \mathcal{P}(n))$$

*denote the limit operator. Moreover suppose* $\alpha : \mathbb{Z}^s \to \mathcal{P}(1) = \mathbb{R}_{>0}$ *is bounded, and*

$Q \in SO(n)$. *Then for all* $x \in \ell^\infty(\mathbb{Z}^s, \mathcal{P}(n))$

$$\begin{aligned} S^\infty(x^{-1}) &= (S^\infty x)^{-1} \\ S^\infty(\alpha x) &= S^\infty \alpha \cdot S^\infty x \\ S^\infty(Q^T x Q) &= Q^T (S^\infty x) Q. \end{aligned}$$

*In addition, if* $\det(x_i) = 1$ *for all* $i \in \mathbb{Z}^s$, *then* $\det(S^\infty x) \equiv 1$.

*Remark* 4.28. In case of subdivision algorithms acting on input data from $\mathcal{P}(n)$ via a *finite* number of repeated binary averages in each refinement step, an analogous version of Corollary 4.27 was shown in [18]. In this particular situation the aforementioned properties of $S^\infty$ already follow from a restriction of 4.26 to random variables with $\#Y(\Omega) = 2$.

## 4.4 $L^p$-convergence of the characteristic Markov chain

This short section clarifies the relationship between the stochastic convergence of the Markov chain associated to $(a_i)_{i \in \mathbb{Z}^s}$ and its counterpart in the theory of barycentric subdivision schemes.

**Lemma 4.29.** *Suppose* $\operatorname{supp}(\mathbf{a}) \subseteq C \cap \mathbb{Z}^s$, *where* $C$ *is a convex, balanced, and compact set. Let* $\rho : \mathbb{R}^s \to \mathbb{R}_{\geq 0}$ *denote the Minkowski functional of* $C$. *Recall the notation* $\mathbb{P}_i$ *for the probability measure on* $(\mathbb{Z}^s)^{\mathbb{N}_0}$ *induced by the transition kernel* $P^{\mathbf{a}}$ *and the initial distribution* $\delta_{\{i\}}$. *Then*

$$\rho(i) \leq 2^n \quad \Longrightarrow \quad \mathbb{P}_i(X_n^{\mathbf{a}} \in 2C) = 1.$$

*In other words, the Markov chain with deterministic initial condition* $X_0^{\mathbf{a}} = i$ *reaches* $2C$ *within* $[\log_2(\rho(i))] + 1$ *steps and remains in this set thereafter.*

*Proof.* Recall that since $\operatorname{supp}(\mathbf{a}) \subseteq C$, for any $j \in \mathbb{Z}^s$ we obtain

$$a_{i-2^n j}^{(n)} \neq 0 \quad \Longrightarrow \quad \rho(i - 2^n j) \leq 2^n.$$

Thus the fact that $\rho(i)/2^n \leq 1$ renders the right hand side of

$$\mathbb{P}_i(X_n^{\mathbf{a}} \in \mathbb{Z}^s \setminus 2C) = \sum_{\rho(j) > 2} a_{i-2^n j}^{(n)}$$

an empty sum, since $\rho(i - 2^n j) \leq 2^n$ implies

$$\rho(j) \leq \rho(i - 2^n j)/2^n + \rho(i)/2^n \leq 2. \qquad \square$$

**Theorem 4.30** ([8]). *Let $p \in [1, \infty)$. Suppose the characteristic Markov chain $X_n^{\mathbf{a}}$ of $(a_i)_{i\in\mathbb{Z}^s}$ with deterministic initial condition $\ell \in \mathbb{Z}^s$ possesses a stationary distribution $\pi$ in the sense that for all $j \in \mathbb{Z}^s$, $|a^{(n)}_{\ell-2^n j} - \pi_j| \to 0$ as $n \to \infty$. Then $X_n^{\mathbf{a}}$ converges in $L^p(\Omega, \mathbb{P}_\ell; \mathbb{R}^s)$ if and only if there is $k \in \mathbb{Z}^s$ such that $\pi = \delta_k$. In this case,*

$$E_\ell(\|X_n^{\mathbf{a}} - k\|^p) \to 0 \quad \textit{as } n \to \infty.$$

*Proof.* Let $\rho$ denote the Minkowski functional of a balanced, closed and convex set containing $\operatorname{supp}(\mathbf{a})$. Moreover, for $n \in \mathbb{N}$ define

$$A_n = \{(i,j) \in \mathbb{Z}^s \times \mathbb{Z}^s \mid \max(\rho(\ell - 2^n j), \rho(j - 2^n i)) \le 2^n - 1\}.$$

Note that $(i,j) \in A_n$ implies that $\rho(j) \le 1 + \frac{\rho(\ell)-1}{2^n}$ as well as $\rho(i) \le 1 + \frac{\rho(\ell)-1}{2^{2n}}$. Thus there is a bounded set $B$ such that

$$\bigcup_{n\in\mathbb{N}} A_n \subseteq B.$$

Moreover we have

$$\begin{aligned}
\int_\Omega \|X_{2n}^{\mathbf{a}}(\omega) - X_n^{\mathbf{a}}(\omega)\|^p \mathbb{P}_\ell(d\omega) &= \sum_{i,j\in\mathbb{Z}^s} \|i-j\|^p \mathbb{P}_\ell(X_{2n}^{\mathbf{a}} = i \wedge X_n^{\mathbf{a}} = j) \\
&= \sum_{i,j\in\mathbb{Z}^s} \|i-j\|^p \mathbb{P}_\ell(X_{2n}^{\mathbf{a}} = i | X_n^{\mathbf{a}} = j)\mathbb{P}_\ell(X_n^{\mathbf{a}} = j) \\
&= \sum_{(i,j)\in A_n} \|i-j\|^p a^{(n)}_{j-2^n i} a^{(n)}_{\ell-2^n j}.
\end{aligned} \tag{59}$$

Certainly, since $B$ is bounded, the sequence

$$\varepsilon_n := \sup_{(i,j)\in B} (|\pi_i - a^{(n)}_{j-2^n i}| + |\pi_j - a^{(n)}_{\ell-2^n j}|)$$

converges to zero as $n \to \infty$. Consequently, we obtain $E_\ell(\|X_{2n}^{\mathbf{a}} - X_n^{\mathbf{a}}\|^p) \ge c_n$, where

$$\begin{aligned}
c_n = &\sum_{(i,j)\in A_n} \|i-j\|^p \pi_i \pi_j \\
&- \varepsilon_n \sum_{(i,j)\in A_n} \|i-j\|^p (a^{(n)}_{j-2^n i} + a^{(n)}_{\ell-2^n j}) \\
&- \varepsilon_n^2 \sum_{(i,j)\in A_n} \|i-j\|^p.
\end{aligned} \tag{60}$$

Thus, whenever there are integers $i \ne j$ such that $\pi_i > 0$ and $\pi_j > 0$, Equation (60) implies that $E_\ell(\|X_{2n}^{\mathbf{a}} - X_n^{\mathbf{a}}\|)$ is bounded away from zero asymptotically. Hence for $L^p$-convergence of $X_n^{\mathbf{a}}$ we need the existence of some $k \in \mathbb{Z}^s$ with $\pi_i = \delta_{ki}$.

Conversely, assume that $\pi = \delta_{\{k\}}$. Then since $X_n^{\mathbf{a}} \to k$ in distribution, $X_n^{\mathbf{a}} \to k$ in probability. From Lemma 4.29 we conclude that there is $M > 0$ such that $\|X_n^{\mathbf{a}}\| \leq M$ holds $\mathbb{P}_\ell$-almost surely for $n \in \mathbb{N}$. Hence for $\delta > 0$,

$$\begin{aligned} E_\ell(\|X_n^{\mathbf{a}} - k\|^p) &\leq \int_{\{\|X_n^{\mathbf{a}} - k\| \geq \delta\}} \|X_n^{\mathbf{a}} - k\|^p d\mathbb{P}_\ell + \int_{\{\|X_n^{\mathbf{a}} - k\| < \delta\}} \|X_n^{\mathbf{a}} - k\|^p d\mathbb{P}_\ell \\ &\leq (M + k)^p \mathbb{P}_\ell(\|X_n^{\mathbf{a}} - k\| \geq \delta) + \delta^p, \end{aligned}$$

showing that $E_\ell(\|X_n^{\mathbf{a}} - k\|)$ converges to zero. □

Suppose now that the subdivision scheme associated to $(a_i)_{i \in \mathbb{Z}^s}$ converges. Then a refinable function $\varphi$ satisfying (37) and (38) exists. Substituting $i \in \mathbb{Z}^s$ for $t$ in

$$\varphi(t) = \sum_j a_j \varphi(2t - j)$$

and exploiting the fact that $\sum_i \varphi(i) = 1$, we observe that $\pi_i = \varphi(-i)$ is a stationary distribution for $X_n^{\mathbf{a}}$. Moreover recall that a convergent scheme is called *interpolatory* if and only if there is $k \in \mathbb{Z}^s$ such that for $j \in \mathbb{Z}^s$, $\varphi(j) = \delta_{kj}$. Now Theorem 4.30 translates to the language of refinement schemes as follows:

**Corollary 4.31.** *Suppose the linear scheme associated to $(a_i)_{i \in \mathbb{Z}^s}$ converges, and $p \in [1, \infty)$. Then the characteristic Markov chain of $X_n^{\mathbf{a}}$ with deterministic initial condition $\ell \in \mathbb{Z}^s$ converges in $L^p(\Omega, \mathbb{P}_\ell; \mathbb{R}^s)$ if and only if the scheme is interpolatory. In this case the $L^p$-limit is a constant lattice point.*

# References

[1] A. D. Alexandrov. A theorem on triangles in a metric space and some of its applications. *Trudy Mat. Inst. Steklov.*, 38:5–23, 1951. [Russian].

[2] A.D. Alexandrov. Über eine Verallgemeinerung der Riemannschen Geometrie. *Schriftenreihe des Forschungsinstituts für Mathematik bei der Deutschen Akademie der Wissenschaften zu Berlin*, 1:33–84, 1957.

[3] W. Ballmann. *Lectures on spaces of nonpositive curvature.* Birkhäuser, 1995.

[4] P. Billingsley. *Probability and measure.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995.

[5] E. Cartan. Groupes simples clos et ouverts et géométrie riemannienne. *Journal de mathématiques pures et appliquées 9e série*, 8:1–34, 1929.

[6] A. S. Cavaretta, W. Dahmen, and C. A. Micchelli. *Stationary Subdivision.* American Mathematical Society, 1991.

[7] J. Dahl. Steiner problems in optimal transport. *Transactions of the American Mathematical Society*, 363:1805–1819, 2011.

[8] O. Ebner. Stochastic aspects of refinement schemes on metric spaces. Technical report, TU Graz, 2012.

[9] O. Ebner. Convergence of refinement schemes on metric spaces. *Proceedings of the American Mathematical Society*, to appear.

[10] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, 10:215–310, 1948.

[11] P. Grohs. Smoothness analysis of subdivision schemes on regular grids by proximity. *SIAM Journal on Numerical Analysis*, 46:2169–2182, 2008.

[12] P. Grohs. A general proximity analysis of nonlinear subdivision schemes. *SIAM Journal on Mathematical Analysis*, 42(2):729–750, 2010.

[13] K. Jetter and X.-J. Li. SIA matrices and non-negative subdivision. *Results in Mathematics*, to appear.

[14] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, 30(5):509–541, 1977.

[15] S. Lang. *Fundamentals of Differential Geometry*, volume 191 of *Graduate Texts in Mathematics*. Springer, 1999.

[16] C. A. Micchelli and H. Prautzsch. Uniform refinement of curves. *Linear Algebra and Applications*, 114/115:841–870, 1989.

[17] U. Reif and J. Peters. *Subdivision Surfaces*, volume 3 of *Geometry and Computing*. Springer, 2008.

[18] N. Sharon and U. Itai. Subdivision schemes for positive definite matrices. *Technical Report*, University of Tel Aviv, 2011.

[19] 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.

[20] K.-T. Sturm. Probability measures on metric spaces of nonpositive curvature. In *Heat kernels and analysis on manifolds, graphs, and metric spaces*, volume 338 of *Contemporary Mathematics*, pages 357–390. American Mathematical Society, 2003.

[21] J. Wallner. On convergent interpolatory subdivision schemes in Riemannian geometry. Geometry Preprint 2012/02, TU Graz, April 2012.

[22] J. Wallner and N. Dyn. Convergence and $C^1$ analysis of subdivision schemes on manifolds by proximity. *Computer Aided Geometric Design*, 22:593–622, 2005.

[23] J. Wallner, E. Nava Yazdani, and A. Weinmann. Convergence and smoothness analysis of subdivision rules in Riemannian and symmetric spaces. *Advances in Computational Mathematics*, 34(2):201–218, 2011.

[24] J. Weickert and H. Hagen, editors. *Visualization and processing of tensor fields.* Mathematics and Visualization. Springer-Verlag, Berlin, 2006.

[25] A. Weinmann. Nonlinear subdivision schemes on irregular meshes. *Constructive Approximation*, 31:395–415, 2010.

[26] W. Woess. *Denumerable Markov Chains - Generating Functions, Boundary Theory, Random Walks on Trees.* EMS Textbooks in Mathematics. European Mathematical Society Publishing House, 2009.

[27] G. Xie and T. P.-Y. Yu. Smoothness equivalence properties of general manifold-valued data subdivision schemes. *Multiscale Modeling and Simulation*, 7(3):1073–1100, 2008.

[28] X. Zhou. Subdivision schemes with nonnegative masks. *Mathematics of Computation*, 74:819–839, 2005.

[29] X. Zhou. On multivariate subdivision schemes with nonnegative finite masks. *Proceedings of the American Mathematical Society*, 134:859–869, 2006.