The Annals of Probability
2002, Vol. 30, No. 3, 1195–1222

NONLINEAR MARTINGALE THEORY FOR PROCESSES WITH VALUES IN METRIC SPACES OF NONPOSITIVE CURVATURE

BY KARL-THEODOR STURM

University of Bonn

We develop a nonlinear martingale theory for time discrete processes $(Y_n)_{n\in\mathbb{N}_0}$. These processes are defined on any filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_n, \mathbb{P})_{n\in\mathbb{N}_0}$ and have values in a metric space (N, d) of nonpositive curvature (in the sense of A. D. Alexandrov). The defining martingale property for such processes is

$$\mathbb{E}(Y_{n+1}|\mathcal{F}_n) = Y_n, \qquad \mathbb{P}\text{-a.s.},$$

where the conditional expectation on the left-hand side is defined as the minimizer of the functional

$$Z \mapsto \mathbb{E}d^2(Z, Y_{n+1})$$

within the space of $\mathcal{F}_n$-measurable maps $Z:\Omega \to N$.

We give equivalent characterization of N-valued martingales (using merely the usual linear conditional expectations) and derive fundamental properties of these martingales, for example, a martingale convergence theorem.

Finally, we exploit the relation with harmonic maps. It turns out that a map $f:M \to N$ is harmonic w.r.t. a given Markov kernel p on M if and only if it maps Markov chains $(X_n)_{n\in\mathbb{N}}$ (with transition kernel p) on M onto martingales $(f(X_n))_{n\in\mathbb{N}}$ with values in N.

The nonlinear heat flow $f:\mathbb{N}_0 \times M \to N$ of a given initial map $f(0,\cdot):M \to N$ at time n is obtained as the "filtered expectation,"

$$f(n,x) := \mathbb{E}_x[f(X_n)|||(\mathcal{F}_k)_{k\geq 0}]$$

of the random map $f(X_n)$. Similarly, the unique solution to the Dirichlet problem for a given map $g:M \to N$ and a subset $D \subset M$ is obtained as

$$f(x) := \mathbb{E}_x[g(X_{\tau(D)})|||(\mathcal{F}_k)_{k\geq 0}].$$

In both cases, a crucial role is played by the notion of filtered expectation $\mathbb{E}_x[\cdot|||(\mathcal{F}_k)_{k\geq 0}]$ which will be discussed in detail.

Moreover, we prove Jensen's inequality for expectations and filtered expectations and we prove (weak and strong) laws of large numbers for sequences of i.i.d. random variables with values in N.

Our theory is an extension of the classical linear martingale theory and of the nonlinear theory of martingales with values in manifolds as developed, for example, in Emery (1989) and Kendall (1990). The goal is to extend the previous framework towards processes with values in metric spaces. This

Received June 2000; revised September 2001.

AMS 2000 subject classifications. 31C25, 58E20, 58J65, 60G42, 60J05, 60J45.

Key words and phrases. Barycenter, center of mass, center of gravity, law of large numbers, martingale, Γ-martingale, harmonic map, NPC space, Alexandrov curvature.

will lead to a stochastic approach to the theory of (generalized) harmonic maps with values in such "singular" spaces as developed by Jost (1994) and Korevaar and Schoen (1993).

1. Preliminaries on metric spaces. Throughout this paper, $(\Omega, \mathcal{F}, \mathbb{P})$ will be a probability space, $\mathcal{G} \subset \mathcal{F}$ some σ-field and $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ a filtration (where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$). (N, d) will be a global NPC space. Let us recall the definition of the latter.

DEFINITION 1.1. (N, d) is called a *global NPC space* if and only if it is a complete metric space with *nonpositive curvature* in the following sense:

$$\inf_{z\in N} \int_N d^2(x, z) q(dx) \le \tfrac{1}{2} \int_N \int_N d^2(x, y) q(dx) q(dy) \tag{1}$$

for all discrete probability measures q on N.

For examples and further details we refer to Ballmann (1995), Jost (1997) and Sturm (2001a). From the latter reference we quote the following basic properties.

PROPOSITION 1.2. *A complete metric space (N, d) is a global NPC space if and only if*:

(i) *It is a geodesic space, that is, any two points $\gamma_0, \gamma_1 \in N$ can be joined by a (continuous) curve $\gamma : [0, 1] \to N$ such that $d(\gamma_0, \gamma_1) = l_d(\gamma)$ where the length of γ is defined as*

$$l_d(\gamma) := \sup\left\{\sum_{k=1}^{n} d\bigl(\gamma_{t_k}, \gamma_{t_{k_1}}\bigr) : 0 \le t_0 \le t_1 \le \cdots \le t_n \le 1,\ n \in \mathbb{N}\right\},$$

(ii) *and it has nonpositive curvature in the sense of A. D. Alexandrov, which means that for every point z, every geodesic $t \mapsto \gamma_t$ (parametrized proportionally to arclength, as always) and every $t \in [0, 1]$,*

$$d^2(z, \gamma_t) \le (1 - t) d^2(z, \gamma_0) + t d^2(z, \gamma_1) - t(1 - t) d^2(\gamma_0, \gamma_1). \tag{2}$$

PROPOSITION 1.3. *For any quadruple $z_1, z_2, z_3, z_4 \in N$ in a global NPC space the following inequality holds true*:

$$d^2(z_1, z_3) + d^2(z_2, z_4) \le d^2(z_1, z_2) + d^2(z_3, z_4) + 2d(z_2, z_3) d(z_4, z_1). \tag{3}$$

DEFINITION 1.4. A map $Y : \Omega \to N$ is called *$\mathcal{G}$-measurable* if and only if $f^{-1}(B) \in \mathcal{G}$ for all $B \in \mathcal{N}$ where $\mathcal{N}$ denotes the Borel σ-field on N. $\mathcal{F}$-measurable maps are just called measurable maps or N-valued random variables.

PROPOSITION 1.5. *For any map $Y:\Omega \to N$ the following properties are equivalent*:

(i) *Y is $\mathcal{G}$-measurable with separable range $Y(\Omega)$*;
(ii) *Y is the uniform limit of $\mathcal{G}$-measurable maps Y_n with countable $Y_n(\Omega)$*;
(iii) *Y is the pointwise limit of $\mathcal{G}$-measurable maps Y_n with finite $Y_n(\Omega)$*;
(iv) *Y is the pointwise limit of $\mathcal{G}$-measurable maps Y_n with separable $Y_n(\Omega)$*.

For the proof, see, for example, Sturm (2001a), Lemma 4.3.

Given $r \in [1, \infty[$ and measurable maps $Y, Z:\Omega \to N$, we define their L^r-distance,

$$d_r(Y,Z) := \big(\mathbb{E}d^r(Y,Z)\big)^{1/r} := \left(\int_\Omega d^r\big(Y(\omega), Z(\omega)\big)\mathbb{P}(d\omega)\right)^{1/r}.$$

Similarly, we can define the L^∞-distance $d_\infty(Y,Z)$. We say that Y and Z are versions of each other or that they are equivalent if $d_r(Y,Z) = 0$, that is, if and only if $Y = Z$, $\mathbb{P}$-a.s.

We define the L^r-space of $\mathcal{G}$-measurable random variables

$$\begin{aligned} L^r(\mathcal{G}) := \big\{&\text{equiv. classes of } \mathcal{G}\text{-meas. } Z:\Omega \to N \\ &\text{with } d_r(y,Z) < \infty \text{ for some/all } y \in N\big\}. \end{aligned}$$

More precisely, this space should be denoted by $L^r((\Omega, \mathcal{G}, \mathbb{P}), (N, d))$, but we only specify those parameters which are not clear from the context.

PROPOSITION 1.6. *The space $L^2(\mathcal{G})$ equipped with the metric d_2 is a global NPC space*.

This is a classical result. For the proof, see, for example, Sturm (2001b), Proposition 3.3.

Given any equivalence class of measurable maps $Y:\Omega \to N$ we define its *mean conditional variance*

$$\mathbb{V}_{\mathcal{G}}(Y) := \inf\big\{\mathbb{E}d^2(Z,Y) : \mathcal{G}\text{-meas.} Z:\Omega \to N\big\}$$

and its *variance*

$$\mathbb{V}(Y) := \inf\big\{\mathbb{E}d^2(z,Y) : z \in N\big\}.$$

Note that $\mathbb{V}_{\mathcal{G}}(Y) \le \mathbb{V}(Y)$ with equality if $\mathcal{G} = \{\varnothing, \Omega\}$ and

$$\mathbb{V}_{\mathcal{G}}(Y) = 0 \quad \Longleftrightarrow \quad Y \text{ is } \mathcal{G}\text{-measurable.}$$

(More precisely, the right-hand side states that there exists a version Y' of Y which is $\mathcal{G}$-measurable.)

Occasionally, we formulate our results not only for the space $L^r(\mathcal{F})$ but for the larger space

$$L^r(\mathcal{F}, \mathcal{G}),$$

consisting of all equivalence classes of $\mathcal{F}$-measurable $Y:\Omega \to N$ with $d_r(Y, Z) < \infty$ for some $\mathcal{G}$-measurable $Z:\Omega \to N$.

Obviously, each $\mathcal{G}$-measurable, in particular, each constant map $Y \equiv y \in N$, lies in $L^r(\mathcal{F}, \mathcal{G})$. If $\mathcal{G} = \{\varnothing, \Omega\}$ then the space $L^r(\mathcal{F}, \mathcal{G})$ coincides with the space $L^r(\mathcal{F})$. Generally, $L^r(\mathcal{F}) \subset L^r(\mathcal{F}, \mathcal{G})$.

If no ambiguity is possible, we will not distinguish between random variables and the corresponding equivalence classes of random variables.

2. Expectations and conditional expectations. The crucial consequence of the uniform convexity of the function $x \mapsto d^2(z, x)$ is that it allows defining uniquely barycenters of probability measures on N or, equivalently, expectations of maps into N. This idea will be extended in a canonical way to define conditional expectations.

THEOREM 2.1. *Let $Y \in L^2(\mathcal{F})$.*

(i) *There exists a unique $Z \in L^2(\mathcal{G})$ which minimizes $Z \mapsto d_2(Z, Y)$. This Z will be denoted as*

$$Z = \mathbb{E}_{\mathcal{G}} Y = \mathbb{E}(Y|\mathcal{G})$$

and called conditional expectation of Y under $\mathcal{G}$.

Hence, $\mathbb{E}d^2(\mathbb{E}_{\mathcal{G}} Y, Y) = \mathbb{V}_{\mathcal{G}}(Y)$.

If $\mathcal{G} = \{\varnothing, \Omega\}$ then $\mathbb{E}_{\mathcal{G}} Y =: \mathbb{E}Y$ will be called the expectation of Y. It is a constant map or, *in other words*, *a point in N.*

(ii) *For all $Z \in L^2(\mathcal{G})$*,

$$\mathbb{E}d^2(Z, Y) \geq \mathbb{V}_{\mathcal{G}}(Y) + \mathbb{E}d^2(\mathbb{E}_{\mathcal{G}} Y, Z) \tag{4}$$

and $\mathbb{P}$-a.s.

$$\mathbb{E}_{\mathcal{G}} d^2(Z, Y) \geq \mathbb{E}_{\mathcal{G}} d^2(\mathbb{E}_{\mathcal{G}} Y, Y) + d^2(\mathbb{E}_{\mathcal{G}} Y, Z) \tag{5}$$

(*"conditional variance inequality"*).

REMARK. The proof will show that the conditional expectation $\mathbb{E}_{\mathcal{G}} Y$ exists uniquely for each $Y \in L^2(\mathcal{F}, \mathcal{G})$ as the minimizer of $Z \mapsto d_2(Z, Y)$ on the space of $\mathcal{G}$-measurable maps $Z:\Omega \to N$. Property (ii) of the above theorem holds true for all such Z with $d_2(Z, Y) < \infty$.

PROOF. (i) Let $Y \in L^2(\mathcal{F}, \mathcal{G})$. For $A \in \mathcal{G}$ let d_A denote the distance

$$d_A(Z, Z') := \left(\int_A d^2(Z, Z') \, d\mathbb{P} \right)^{1/2}$$

and let $L^2(A, \mathcal{G}, Y)$ denote the set of d_A-equivalence classes of $\mathcal{G}$-measurable maps $Z : \Omega \to N$ with $d_A(Z, Y) < \infty$.

Consider the functional $Q_A = d_A^2(\cdot, Y)$ on $L^2(A, \mathcal{G}, Y)$ defined by

$$Q_A(Z) := \int_A d^2(Z, Y) \, d\mathbb{P}.$$

The uniform convexity (2) of $z \mapsto d^2(z, y)$ implies uniform convexity of Q_A:

$$Q_A(Z_t) \le (1 - t) Q_A(Z_0) + t Q_A(Z_1) - t(1 - t) \int_A d^2(Z_0, Z_1) \, d\mathbb{P} \tag{6}$$

for any geodesic $t \mapsto Z_t$ in $L^2(A, \mathcal{G}, Y)$.

Moreover, Q_A is continuous on $L^2(A, \mathcal{G}, Y)$. Hence, there exists a unique minimizer $Z := \mathbb{E}_{A,\mathcal{G}} Y$. Indeed, let $(Z_n)_{n \in \mathbb{N}}$ be a sequence in $L^2(A, \mathcal{G}, Y)$ which minimizes Q_A, that is,

$$\lim_{n \to \infty} Q_A(Z_n) = \alpha := \inf\{Q_A(Z) : Z \in L^2(A, \mathcal{G}, Y)\}.$$

Applying (6) to the midpoints $Z_{n,k}$ of Z_n and Z_k yields

$$\begin{aligned} d_A^2(Z_n, Z_k) &\le 2 Q_A(Z_n) + 2 Q_A(Z_k) - 4 Q_A(Z_{n,k}) \\ &\le 2 Q_A(Z_n) + 2 Q_A(Z_k) - 4\alpha \\ &\to 0 \end{aligned}$$

for $n, k \to \infty$. That is, $(Z_n)_n$ is a Cauchy sequence in the Hilbert space $L^2(A, \mathcal{G}, Y)$ and thus $Z = \lim_{n \to \infty} Z_n$ exists. Moreover, $Q_A(Z) = \alpha$ by (lower semi-)continuity of Q_A. This proves the existence of $\mathbb{E}_{A,\mathcal{G}} Y$. Uniqueness is obvious from (6). If $A = \Omega$, this is the conditional expectation $\mathbb{E}_{\mathcal{G}}(Y)$.

(ii) Let $Z \in L^2(A, \mathcal{G}, Y)$ and put $Z_1 := Z$, $Z_0 := \mathbb{E}_{A,\mathcal{G}} Y$ and let $t \mapsto Z_t$ be the joining geodesic. Then by (6),

$$\begin{aligned} d_A^2(\mathbb{E}_{A,\mathcal{G}} Y, Y) &\le d_A^2(Z_t, Y) \\ &\le (1 - t) d_A^2(\mathbb{E}_{A,\mathcal{G}} Y, Y) + t d_A^2(Z, Y) - t(1 - t) d_A^2(\mathbb{E}_{A,\mathcal{G}} Y, Z) \end{aligned}$$

for all $t \in [0, 1]$. Hence (with $t \to 0$),

$$d_A^2(\mathbb{E}_{A,\mathcal{G}} Y, Y) + d_A^2(\mathbb{E}_{A,\mathcal{G}} Y, Z) \le d_A^2(Z, Y). \tag{7}$$

Analogously, we obtain for all $Z \in L^2(\Omega \setminus A, \mathcal{G}, Y)$,

$$d_{\Omega \setminus A}^2(\mathbb{E}_{\Omega \setminus A,\mathcal{G}} Y, Y) + d_{\Omega \setminus A}^2(\mathbb{E}_{\Omega \setminus A,\mathcal{G}} Y, Z) \le d_{\Omega \setminus A}^2(Z, Y). \tag{8}$$

Obviously,

$$
\begin{aligned}
d^2(\mathbb{E}_{\mathcal{G}} Y, Y) &= \inf\{d^2_{\Omega}(Z, Y) : Z \in L^2(\Omega, \mathcal{G}, Y)\} \\
&\geq \inf\{d^2_A(Z, Y) : Z \in L^2(A, \mathcal{G}, Y)\} \\
&\quad + \inf\{d^2_{\Omega\setminus A}(Z', Y) : Z' \in L^2(\Omega \setminus A, \mathcal{G}, Y)\} \\
&= d^2_A(\mathbb{E}_{A,\mathcal{G}} Y, Y) + d^2_{\Omega\setminus A}(\mathbb{E}_{\Omega\setminus A,\mathcal{G}} Y, Y).
\end{aligned} \tag{9}
$$

Now put

$$
Y_A = \begin{cases} \mathbb{E}_{A,\mathcal{G}} Y, & \text{on } A, \\ \mathbb{E}_{\Omega\setminus A,\mathcal{G}} Y, & \text{on } \Omega \setminus A. \end{cases}
$$

Then, by (7), (8) and (9), $Y_A \in L^2(\Omega, \mathcal{G}, Y)$ is a minimizer of $\mathbb{E}d^2(\cdot, Y)$ on $L^2(\Omega, \mathcal{G}, Y)$. Hence, it coincides with $\mathbb{E}_{\mathcal{G}} Y$. That is,

$$
\mathbb{E}_{\mathcal{G}} Y = \mathbb{E}_{A,\mathcal{G}} Y, \qquad \mathbb{P}\text{-a.s. on } A.
$$

Therefore, (7) reads: $\forall\ Z \in L^2(\Omega, \mathcal{G}, Y)$, $\forall\ A \in \mathcal{G}$,

$$
\int_A d^2(Z, \mathbb{E}_{\mathcal{G}} Y)\, d\mathbb{P} + \int_A d^2(\mathbb{E}_{\mathcal{G}} Y, Y)\, d\mathbb{P} \leq \int_A d^2(Z, Y)\, d\mathbb{P}.
$$

This is equivalent to $\forall\ Z \in L^2(\Omega, \mathcal{G}, Y) : \mathbb{P}$-a.s.,

$$
d^2(Z, \mathbb{E}_{\mathcal{G}} Y) + \mathbb{E}_{\mathcal{G}} d^2(\mathbb{E}_{\mathcal{G}} Y, Y) \leq \mathbb{E}_{\mathcal{G}} d^2(Z, Y)
$$

which is the claim (5). $\square$

REMARKS. Let us mention some elementary properties of conditional expectations.

(a) If $N = \mathbb{R}$ (or, more generally, if N is a Hilbert space) then obviously our definition of expectations and conditional expectations coincides with the usual one. In particular,

$$
\mathbb{E}Y = \int_{\Omega} Y(\omega)\mathbb{P}(d\omega).
$$

Moreover, inequalities (4) and (5) are then equalities.

(b) Following the argumentation of Korevaar and Schoen (1993) one easily verifies that for each *closed convex* set $N_0 \subset N$,

$$
Y(\Omega) \subset N_0 \quad \Longrightarrow \quad \mathbb{E}_{\mathcal{G}} Y(\Omega) \subset N_0.
$$

(c) If $N = N_1 \times N_2$ is a product of global NPC spaces and $Y = (Y_1, Y_2)$ then

$$
\mathbb{E}_{\mathcal{G}} Y = (\mathbb{E}_{\mathcal{G}} Y_1, \mathbb{E}_{\mathcal{G}} Y_2).
$$

THEOREM 2.2. *Let $Y \in L^2(\mathcal{F})$ with separable range $Y(\Omega)$ and let $X \in L^2(\mathcal{G})$. Then the following assertions are equivalent*:

(i) $\mathbb{E}_{\mathcal{G}} Y = X$, $\mathbb{P}$-*a.s.*

(ii) *For all* $z \in N$, $\mathbb{E}_{\mathcal{G}} d^2(z, Y) \geq d^2(z, X) + \mathbb{E}_{\mathcal{G}} d^2(Y, X)$, $\mathbb{P}$-*a.s.*
(iii) *For all* $z \in N$, $\mathbb{E}_{\mathcal{G}} d^2(z, Y) \geq d^2(z, X) + \mathbb{E}_{\mathcal{G}} d^2(Y, \mathbb{E}_{\mathcal{G}} Y)$, $\mathbb{P}$-*a.s.*

REMARKS. (a) It suffices to verify (ii) and (iii) for a countable set of $z \in N$ which is dense in the range of Y [since $\mathbb{E}_{\mathcal{G}} d^2(z, Y)$ and $d^2(z, X)$ depend continuously on $z \in N$].

(b) Implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) hold true without the assumption of separable range $Y(\Omega)$.

PROOF. (i) $\Rightarrow$ (ii). The conditional variance inequality from Theorem 2.1 together with (i) imply $\mathbb{P}$-a.s.,

$$\begin{aligned} \mathbb{E}_{\mathcal{G}} d^2(z, Y) &\geq \mathbb{E}_{\mathcal{G}} d^2(\mathbb{E}_{\mathcal{G}} Y, Y) + d^2(z, \mathbb{E}_{\mathcal{G}} Y) \\ &= \mathbb{E}_{\mathcal{G}} d^2(X, Y) + d^2(z, X). \end{aligned}$$

(ii) $\Rightarrow$ (iii). Obvious by Theorem 2.1.

(iii) $\Rightarrow$ (i). (iii) states that $\forall z \in N$, $\forall A \in \mathcal{G}$,

$$\int_A d^2(z, Y)\, d\mathbb{P} \geq \int_A d^2(z, X)\, d\mathbb{P} + \int_A d^2(Y, \mathbb{E}_{\mathcal{G}} Y)\, d\mathbb{P}.$$

Hence, for all $\mathcal{G}$-measurable $Z : \Omega \to N$ with finite range,

$$\int d^2(Z, Y)\, d\mathbb{P} \geq \int d^2(Z, X)\, d\mathbb{P} + \int d^2(Y, \mathbb{E}_{\mathcal{G}} Y)\, d\mathbb{P}. \tag{10}$$

This immediately extends to all $\mathcal{G}$-measurable Z with countable range and, more generally, with separable range.

Since Y has separable range, also $\mathbb{E}_{\mathcal{G}} Y$ has separable range [cf. Lemma 6.4 in Sturm (2001a)]. Therefore, we may choose $Z = \mathbb{E}_{\mathcal{G}} Y$ in (10) in order to obtain

$$\mathbb{E}_{\mathcal{G}} Y = X, \qquad \mathbb{P}\text{-a.s.} \qquad \square$$

THEOREM 2.3. *For all* $Y, Z \in L^2(\mathcal{F}, \mathcal{G})$ *with* $d_2(Y, Z) < \infty$,

$$d(\mathbb{E}_{\mathcal{G}} Y, \mathbb{E}_{\mathcal{G}} Z) \leq \mathbb{E}_{\mathcal{G}} d(Y, Z)$$

and for all $r \in [1, \infty]$,

$$d_r(\mathbb{E}_{\mathcal{G}} Y, \mathbb{E}_{\mathcal{G}} Z) \leq d_r(Y, Z).$$

In particular, $d_r(\mathbb{E}Y, \mathbb{E}Z) \leq d_r(Y, Z)$ for all $Y, Z \in L^2(\mathcal{F})$ and $r \in [1, \infty]$.

PROOF OF THEOREM 2.3. The conditional variance inequality (5) implies

$$d^2(\mathbb{E}_{\mathcal{G}} Y, \mathbb{E}_{\mathcal{G}} Z) \leq \mathbb{E}_{\mathcal{G}} d^2(\mathbb{E}_{\mathcal{G}} Z, Y) - \mathbb{E}_{\mathcal{G}} d^2(\mathbb{E}_{\mathcal{G}} Y, Y)$$

as well as

$$d^2(\mathbb{E}_{\mathcal{G}}Y, \mathbb{E}_{\mathcal{G}}Z) \le \mathbb{E}_{\mathcal{G}}d^2(\mathbb{E}_{\mathcal{G}}Y, Z) - \mathbb{E}_{\mathcal{G}}d^2(\mathbb{E}_{\mathcal{G}}Z, Z).$$

Applying the quadruple inequality (3) to the points $Y(\omega)$, $Z(\omega)$, $\mathbb{E}_{\mathcal{G}}Z(\omega)$ and $\mathbb{E}_{\mathcal{G}}Y(\omega)$ yields that a.s.,

$$d^2(\mathbb{E}_{\mathcal{G}}Y, Z) + d^2(\mathbb{E}_{\mathcal{G}}Z, Y) \le d^2(\mathbb{E}_{\mathcal{G}}Y, Y) + d^2(\mathbb{E}_{\mathcal{G}}Z, Z) + 2d(\mathbb{E}_{\mathcal{G}}Y, \mathbb{E}_{\mathcal{G}}Z)d(Y, Z).$$

Taking conditional expectations and adding up these inequalities we obtain

$$2d^2(\mathbb{E}_{\mathcal{G}}Y, \mathbb{E}_{\mathcal{G}}Z) \le 2d(\mathbb{E}_{\mathcal{G}}Y, \mathbb{E}_{\mathcal{G}}Z)\mathbb{E}_{\mathcal{G}}d(Y, Z),$$

which gives the first claim. This in turn immediately implies the second one:

$$\mathbb{E}[d^r(\mathbb{E}_{\mathcal{G}}Y, \mathbb{E}_{\mathcal{G}}Z)] \le \mathbb{E}[(\mathbb{E}_{\mathcal{G}}d(Y, Z))^r] \le \mathbb{E}[d^r(Y, Z)]. \qquad \square$$

COROLLARY 2.4. (i) *The definition of conditional expectation* $\mathbb{E}_{\mathcal{G}}$ *extends continuously from* $L^2(\mathcal{F}, \mathcal{G})$ *to* $L^1(\mathcal{F}, \mathcal{G})$.

(ii) *Moreover, the assertions of Theorem* 2.3 *hold true for all* $Y, Z \in L^1(\mathcal{F}, \mathcal{G})$ *with* $d_1(Y, Z) < \infty$.

(iii) *For all* σ*-fields* $\mathcal{H} \subset \mathcal{G}$ *and all* $r \in [1, \infty]$,

$$\mathbb{E}_{\mathcal{G}} : L^r(\mathcal{F}, \mathcal{H}) \to L^r(\mathcal{G}, \mathcal{H})$$

is a contraction. In particular,

$$\mathbb{E}_{\mathcal{G}} : L^r(\mathcal{F}) \to L^r(\mathcal{G})$$

is a contraction.

(iv) *For all* $r \in [1, \infty]$ *and* $Y \in L^r(\mathcal{F}, \mathcal{G})$,

$$d_r(Y, \mathbb{E}_{\mathcal{G}}Y) < \infty.$$

PROOF. Recall that $Y \in L^1(\mathcal{F}, \mathcal{G})$ means that $Y : \Omega \to N$ is $\mathcal{F}$-measurable and that $d_1(Y, Z) < \infty$ for some $\mathcal{G}$-measurable $Z : \Omega \to N$.

Each $Y \in L^1(\mathcal{F}, \mathcal{G})$ can be approximated by $Y_n \in L^\infty(\mathcal{F}, \mathcal{G})$ with $Y_n \to Y$ in d_1. For instance, define

$$Y_n(w) = \begin{cases} Y(w), & \text{if } d\bigl(Y(w), Z(w)\bigr) \le n, \\ Z(w), & \text{else.} \end{cases}$$

Theorem 2.3 and the fact that $Y_n \to Y$ in d_1 imply that $(\mathbb{E}_{\mathcal{G}}Y_n)_n$ is a Cauchy sequence in $L^1(\Omega, \mathcal{G}, Z)$. Hence,

$$\mathbb{E}_{\mathcal{G}}Y = \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}Y_n$$

exists. Moreover, all assertions from Theorem 2.3 extend from L^2 to L^1. In particular,

$$d_1(\mathbb{E}_{\mathcal{G}} Y, \mathbb{E}_{\mathcal{G}} Y') \le d_1(Y, Y')$$

for all $Y, Y' \in L^1(\mathcal{F}, \mathcal{G})$ with $d_1(Y, Y') < \infty$ which implies that $\mathbb{E}_{\mathcal{G}}$ is a contraction from $L^1(\mathcal{F})$ to $L^1(\mathcal{G})$. $\square$

DEFINITION 2.5. Given any sequence of random variables $Y_i : \Omega \to N$ we define a sequence $(S_n)_{n\in\mathbb{N}}$ of random variables $S_n : \Omega \to N$ by induction on n as follows:

$$S_1(\omega) := Y_1(\omega)$$

and

$$S_{n+1}(\omega) := \frac{n}{n+1} S_n(\omega) + \frac{1}{n+1} Y_{n+1}(\omega),$$

where the right-hand side should denote the point on the geodesic from $S_n(\omega)$ to $Y_{n+1}(\omega)$ with distance from $S_n(\omega)$ being $\frac{1}{n+1}$ of the length of this geodesic. The map S_n will be denoted by $\frac{1}{n}\sum_{i=1}^n Y_i$.

Note, however, that in general the map $\frac{1}{n}\sum_{i=1}^n Y_i$ will strongly depend on permutations of the Y_i.

THEOREM 2.6 (Law of large numbers). *Let $(Y_i)_{i\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables $Y_i \in L^2(\mathcal{F})$. Then*

$$\frac{1}{n}\sum_{i=1}^n Y_i \to \mathbb{E}Y_1 \qquad \text{for } n \to \infty$$

in L^2 and in probability ("weak law of large numbers").

If, moreover, $Y_i \in L^\infty(\mathcal{F})$ then for $\mathbb{P}$-almost every $\omega \in \Omega$,

$$\frac{1}{n}\sum_{i=1}^n Y_i(\omega) \to \mathbb{E}Y_1 \qquad \text{for } n \to \infty$$

("strong law of large numbers").

PROOF. (a) Our first claim is that $\forall\, n \in \mathbb{N}$,

$$\mathbb{E}d^2(\mathbb{E}Y_1, S_n) \le \frac{1}{n}\mathbb{V}(Y_1).$$

This is obviously true for $n = 1$. We will prove it for all $n \in \mathbb{N}$ by induction. Assuming that it holds for n we conclude [using inequalities (2) and (4) from

Proposition 1.2 and Theorem 2.1],

$$
\begin{aligned}
&\mathbb{E}d^2(\mathbb{E}Y_1, S_{n+1})\\
&\quad = \mathbb{E}d^2\left(\mathbb{E}Y_1, \frac{n}{n+1}S_n + \frac{1}{n+1}Y_{n+1}\right)\\
&\quad \overset{(2)}{\leq} \frac{n}{n+1}\mathbb{E}d^2(\mathbb{E}Y_1, S_n) + \frac{1}{n+1}\mathbb{E}d^2(\mathbb{E}Y_1, Y_{n+1}) - \frac{n}{(n+1)^2}\mathbb{E}d^2(S_n, Y_{n+1})\\
&\quad \overset{(4)}{\leq} \frac{n}{n+1}\mathbb{E}d^2(\mathbb{E}Y_1, S_n) + \frac{1}{n+1}\mathbb{E}d^2(\mathbb{E}Y_1, Y_{n+1})\\
&\qquad - \frac{n}{(n+1)^2}\big[\mathbb{E}d^2(\mathbb{E}Y_{n+1}, S_n) + \mathbb{E}d^2(\mathbb{E}Y_{n+1}, Y_{n+1})\big]\\
&\quad = \left(\frac{n}{n+1}\right)^2 \mathbb{E}d^2(\mathbb{E}Y_1, S_n) + \frac{1}{(n+1)^2}\mathbb{V}(Y_1)\\
&\quad \leq \frac{1}{n+1}\mathbb{V}(Y_1).
\end{aligned}
$$

This proves the first claim, and of course it also proves the L^2 convergence as well as the weak law of large numbers,

$$S_n \to \mathbb{E}Y_1 \qquad \text{in probability}$$

as $n \to \infty$, that is, for all $\varepsilon > 0$,

$$\mathbb{P}\big(d(S_n, \mathbb{E}Y_1) > \varepsilon\big) \to 0$$

as $n \to \infty$.

(b) Our second claim is that

$$S_{n^2} \to \mathbb{E}Y_1$$

a.s. for $n \to \infty$. Indeed, by (a),

$$\sum_{n=1}^{\infty} \mathbb{P}\big(d(S_{n^2}, \mathbb{E}Y_1) > \varepsilon\big) \leq \sum_{n=1}^{\infty} \frac{1}{\varepsilon^2}\mathbb{E}d^2(S_{n^2}, \mathbb{E}Y_1) \leq \sum_{n=1}^{\infty} \frac{1}{\varepsilon^2 n^2}\mathbb{V}(Y_1) < \infty.$$

Due to the Borel–Cantelli lemma, this implies the second claim.

Now assume that $Y_1 \in L^\infty(\mathcal{F})$, say $d(Y_1, z) \leq R$ a.s. for some $z \in N$ and some $R \in \mathbb{R}$. Then by convexity $d(S_n, z) \leq R$ a.s. for all $n \in \mathbb{N}$ and

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

a.s. Therefore, for all $k, n \in \mathbb{N}$ with $n^2 \leq k < (n+1)^2$,

$$d(S_k, S_{n^2}) \leq \left(\frac{1}{n^2+1} + \frac{1}{n^2+2} + \cdots + \frac{1}{k}\right)2R \leq \frac{k-n^2}{n^2}2R \leq \frac{4}{n}R$$

a.s. Together with the second claim, this proves the strong law of large numbers. $\square$

REMARK 2.7. (a) Our approach to expectations and conditional expectations is based on the classical point of view of Carl Friedrich Gauss (1809). He defined the expectation of a random variable (in Euclidean space) to be the uniquely determined point which minimizes the L^2-distance ("Methode der kleinsten Quadrate").

In the context of metric spaces, this point of view was successfully used by Cartan (1928), Fréchet (1948), and many others, under the name of barycenter, center of mass or center of gravity. Iterations of barycenters (similar in spirit to our filtered expectations) on Riemannian manifolds were used by Kendall (1990) and Picard (1994). Jost (1994) applied these concepts on global NPC spaces.

(b) A different point of view is the basis for the approach of Emery and Mokobodzki (1991). They define the expectation $\mathbb{E}X$ as the *set* of all $x \in N$ such that

$$\psi(x) \leq \mathbb{E}[\psi(X)]$$

for all (continuous) convex functions $\psi : N \to \mathbb{R}$ (i.e., such that Jensen's inequality holds). Similarly, one can define conditional expectations and martingales.

A related point of view was used by Doss (1949) and Herer (1991) who define $\mathbb{E}X$ to be the *set* of all $x \in N$ such that

$$d(z, x) \leq \mathbb{E}[d(z, X)]$$

for all $z \in N$.

Note that on global NPC spaces, the functions $x \mapsto d(z, x)$ are convex. We emphasize, however, that these latter concepts to not coincide with our definition of expectations [see Remark 4.8(c)] but obviously *our expectation* $\in$ *expectation in the sense of* Emery and Mokobodzki (1991) $\subset$ *expectation in the sense of* Doss (1949) and Herer (1991).

(c) Another natural way to define "expectations" of random variables is to use (generalizations of) the law of large numbers. This requires giving a meaning to $\frac{1}{n}\sum_{i=1}^{n} Y_i$. Our definition only uses the fact that any two points in N are joined by unique geodesics. Our law of large numbers for global NPC spaces gives convergence toward the expectation defined as minimizer of the L^2 distance.

We emphasize that our definition of expectation $\mathbb{E}Y$ and our definition of $\frac{1}{n}\sum_{i=1}^{n} Y_i$ are different from the ones used by Es-Sahib and Heinich (1999). Their law of large numbers proves convergence to a point, which can be different from our expectation. For instance, let (N, d) be the tripod and let $\mathbb{P}_Y = \frac{1}{2}\delta_a + \frac{1}{4}\delta_b + \frac{1}{4}\delta_c$ where a, b, c are points on three different rays with distance 1 from the origin o (cf. Example 3.2). Then our expectation $\mathbb{E}Y$ will be the origin o, whereas an easy calculation shows that the expectation in the sense of Es-Sahib and Heinich (1999) is the point $\frac{1}{6}a$ (on the ray of a with distance $\frac{1}{6}$ from the origin o).

3. Filtered (conditional) expectations. Now consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}_0}, \mathbb{P})$ and put $\mathcal{F}_\infty := \bigvee_n \mathcal{F}_n$. The above nonlinear conditional expectation is not associative, that is, in general,

$$\mathbb{E}\big[\mathbb{E}[Y|\mathcal{F}_m]|\mathcal{F}_k\big] \neq \mathbb{E}[Y|\mathcal{F}_k]$$

for $m > k$. In order to overcome this disadvantage we introduce the notion of filtered conditional expectation which will play a fundamental role in the sequel. It is based on the observation that for all $m > n > k$ the conditional expectation $Y \mapsto \mathbb{E}[Y|\mathcal{F}_n]$ defines a contraction $L^1(\mathcal{F}_m, \mathcal{F}_k) \to L^1(\mathcal{F}_n, \mathcal{F}_k)$.

DEFINITION 3.1. For $m, k \in \mathbb{N}_0$ and $Y \in L^1(\mathcal{F}_m, \mathcal{F}_k)$ we define the *filtered conditional expectation* by

$$\mathbb{E}[Y \,|\!|\!|\, (\mathcal{F}_n)_{n\geq k}] := \mathbb{E}\big[\mathbb{E}\big[\cdots \mathbb{E}\big[\mathbb{E}[Y|\mathcal{F}_{m-1}] \mid \mathcal{F}_{m-2}\big] \cdots \mid \mathcal{F}_{k+1}\big] \mid \mathcal{F}_k\big],$$

provided $m > k$, and by $\mathbb{E}[Y \,|\!|\!|\, (\mathcal{F}_n)_{n\geq k}] := Y$ otherwise.

If $\mathcal{F}_0$ is trivial, then the point $\mathbb{E}[Y \,|\!|\!|\, (\mathcal{F}_n)_{n\geq 0}]$ is called *filtered expectation* of Y.

Obviously, $\mathbb{E}[Y \,|\!|\!|\, (\mathcal{F}_n)_{n\geq k}]$ is an $\mathcal{F}_k$-measurable map $\Omega \to N$. According to Theorem 2.3,

$$d_r\big(\mathbb{E}[Y \,|\!|\!|\, (\mathcal{F}_n)_{n\geq k}], \mathbb{E}[Y' \,|\!|\!|\, (\mathcal{F}_n)_{n\geq k}]\big) \leq d_r(Y, Y')$$

for all $Y, Y' \in L^r(\mathcal{F}_m)$ and $r \in [1, \infty]$. Hence, the filtered conditional expectation $\mathbb{E}[\cdot \,|\!|\!|\, (\mathcal{F}_n)_{n\geq k}]$ as well as the usual conditional expectation $\mathbb{E}[\cdot|\mathcal{F}_k]$ are nonlinear projections from $L^r(\mathcal{F}_m)$ onto the space $L^r(\mathcal{F}_k)$. *In general, however, they will not coincide!* Similarly, the filtered expectation and the expectation will not coincide in general.

EXAMPLE 3.2. Let (N, d) be a tripod consisting of three half-lines which are glued together at their base points. That is,

$$N = \{(i, t) : i \in \{1, 2, 3\},\ t \in \mathbb{R}_+\}/\sim$$

with $(1, 0) \sim (2, 0) \sim (3, 0)$. The distance is

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

Due to the negative curvature, expectations of random variables on N have a strong tendency to move towards the origin. For instance, let X be a random variable on N with $\mathbb{P}(X = (i, t_0)) = \alpha_i$ with $\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq 0$ and $\sum_{i=1}^3 \alpha_i = 1$. Then

$$\mathbb{E}[X] = \begin{cases} 0, & \text{if } \alpha_1 \leq \alpha_2 + \alpha_3, \\ \big(1, (\alpha_1 - \alpha_2 - \alpha_3)t_0\big), & \text{else.} \end{cases}$$

Now fix $n \in \mathbb{N}$, $n \geq 2$, and a discrete probability space $(\Omega, \mathbb{P})$ with $\Omega = \{1, 2, 3\}^{n+1}$ and

$$\mathbb{P}(\omega) = \prod_{k=0}^{n} p_k(\omega) \qquad \text{where } p_k(\omega) = \begin{cases} \frac{2}{3}, & \text{if } \omega_k = \omega_{k-1}, \\ \frac{1}{6}, & \text{else} \end{cases}$$

and $p_0(\omega) = 1/3$.

For $t > 0$ and $k \in \mathbb{N}_0$ let $Y_k(\omega) = (\omega_k, 3^k t)$ and $\mathcal{F}_k = \sigma\{Y_m : m \leq k\}$. Then

$$\mathbb{E}[Y_{k+1} | \mathcal{F}_k] = Y_k \quad \text{and} \quad \mathbb{E}[Y_{k+m} \| (\mathcal{F}_l)_{l \geq k}] = Y_k$$

for all $k = 0, 1, \ldots, n-1$ and all $m = 1, \ldots, n-k$. However, for all $m > 1$,

$$\mathbb{E}[Y_{k+m} | \mathcal{F}_k] = 0.$$

REMARK 3.3. (i) The above definition of the filtered conditional expectation easily extends to all Y in

$$L_0^1(\mathcal{F}_\infty) := \overline{\bigcup_n L^1(\mathcal{F}_n)} \subset L^1(\mathcal{F}_\infty).$$

Namely, for each $Y \in L_0^1(\mathcal{F}_\infty)$ there exist $Y_m \in L^1(\mathcal{F}_m)$ with $d_1(Y, Y_m) \to 0$ for $m \to \infty$. In particular, $\{Y_m\}_m$ is a Cauchy sequence in $L^1(\mathcal{F}_\infty)$. Since $d_1(\mathbb{E}[Y_m \| (\mathcal{F}_n)_{n \geq k}], \mathbb{E}[Y_{m'} \| (\mathcal{F}_n)_{n \geq k}]) \leq d_1(Y_m, Y_{m'})$ this implies that $\{\mathbb{E}[Y_m \| (\mathcal{F}_n)_{n \geq k}]\}_m$ is a Cauchy sequence in $L^1(\mathcal{F}_k)$. Hence,

$$\mathbb{E}[Y \| (\mathcal{F}_n)_{n \geq k}] := \lim_{m \to \infty} \mathbb{E}[Y_m \| (\mathcal{F}_n)_{n \geq k}]$$

exists in $L^1(\mathcal{F}_k)$. More generally, the filtered conditional expectation $\mathbb{E}[Y \| (\mathcal{F}_n)_{n \geq k}]$ is even well defined for all Y in

$$L_0^1(\mathcal{F}_\infty, \mathcal{F}_k) := \overline{\bigcup_n L^1(\mathcal{F}_n, \mathcal{F}_k)}.$$

Moreover, one easily verifies that for all $Y \in L_0^1(\mathcal{F}_\infty, \mathcal{F}_k)$,

$$\mathbb{E}[Y \| (\mathcal{F}_n)_{n \geq k}] = \lim_{m \to \infty} \mathbb{E}[Y \| (\mathcal{F}_n)_{m \geq n \geq k}],$$

where

$$\mathbb{E}[Y \| (\mathcal{F}_n)_{m \geq n \geq k}] = \mathbb{E}\big[\mathbb{E}\big[\cdots \mathbb{E}\big[\mathbb{E}[Y | \mathcal{F}_m] \mid \mathcal{F}_{m-1}\big] \cdots \mid \mathcal{F}_{k+1}\big] \mid \mathcal{F}_k\big].$$

The space $L_0^1(\mathcal{F}_\infty)$ contains all maps $L^1(\mathcal{F}_\infty)$ with separable range. Indeed, each such map Y can be approximated in d_1 by $\mathcal{F}_\infty$-measurable maps with countable range (Proposition 1.5). That is, without restriction we may assume that $\exists\, \Omega_i \in \mathcal{F}_\infty, z_i \in N : \Omega = \bigcup_{i=1}^\infty \Omega_i$ and $Y = z_i$ on Ω_i $(\forall i)$. Moreover, $\sum d(z, z_i) \times \mathbb{P}(\Omega_i) < \infty$ for each $z \in N$. Each Ω_i can be approximated from inside by suitable $\Omega_{i,n} \in \mathcal{F}_n$. Hence, we may define maps $Y_n \in L^1(\mathcal{F}_n)$ with $Y_n \to Y$ in d_1 by $Y_n := z_i$ on $\Omega_{i,n}$ and $Y_n := z_1$ on $\Omega \setminus \bigcup_i \Omega_{i,n}$.

(ii) In contrast to the conditional expectation, the filtered conditional expectation is associative; that is, for $m > k$,

$$\mathbb{E}\big[\mathbb{E}[Y \Vert\!| (\mathcal{F}_n)_{n\geq m}] \Vert\!| (\mathcal{F}_n)_{n\geq k}\big] = \mathbb{E}[Y \Vert\!| (\mathcal{F}_n)_{n\geq k}].$$

(iii) The notions of filtered (conditional) expectation strongly depend on the choice of the filtration. For instance,

$$\mathbb{E}\big[Y \Vert\!| (\mathcal{F}_{n+k})_{n\in\mathbb{N}_0}\big] \quad \text{and} \quad \mathbb{E}\big[Y \Vert\!| (\mathcal{F}_{2n+k})_{n\in\mathbb{N}_0}\big]$$

will not coincide in general. See Example 3.2.

A function $\psi : N \to \mathbb{R}$ is called *convex* if and only if for any geodesic $\gamma : [0, 1] \to N$ the function $\psi \circ \gamma : [0, 1] \to \mathbb{R}$ is convex. For instance, for every $z \in N$ the function $\psi : x \mapsto d(z, x)$ on N is convex. More generally, the function $\psi : (x, y) \mapsto d(x, y)$ on $N \times N$ is convex.

Following the proof of Jensen's inequality on global NPC spaces by Eells and Fuglede (2001) we obtain the following.

PROPOSITION 3.4. *For each $Y \in L^1(\mathcal{F})$ and each lower semicontinuous convex function $\psi : N \to \mathbb{R}$ with $(\psi \circ Y)_+ \in L^1(\mathcal{F})$,*

$$\mathbb{E}[\psi \circ Y | \mathcal{F}_k] \geq \psi(\mathbb{E}[Y | \mathcal{F}_k]).$$

If in addition $Y \in L^1_0(\mathcal{F}_\infty)$ then

$$\mathbb{E}[\psi \circ Y | \mathcal{F}_k] \geq \psi\big(\mathbb{E}[Y \Vert\!| (\mathcal{F}_n)_{n\geq k}]\big).$$

PROOF. Let ψ be convex and lower semicontinuous.

(i) Let us first assume $Y \in L^2(\mathcal{F})$ and $\psi \circ Y \in L^2(\mathcal{F})$. Consider the random variable $\hat{Y} = (Y, \psi(Y))$ with values in the global NPC space $\hat{N} = N \times R$. By assumption $\hat{Y} \in L^2(\mathcal{F})$ and

$$\mathbb{E}[\hat{Y} | \mathcal{F}_k] = \big(\mathbb{E}[Y | \mathcal{F}_k], \mathbb{E}[\psi(Y) | \mathcal{F}_k]\big).$$

The range $\hat{Y}(\Omega)$ is combined in the closed convex set $(y, t) : \psi(y) \leq t \subset \hat{N}$, hence also $\mathbb{E}[\hat{Y} | \mathcal{F}_k]$ is contained in the set. Therefore,

$$\psi(\mathbb{E}[Y | \mathcal{F}_k]) \leq \mathbb{E}[\psi(Y) | \mathcal{F}_k].$$

(ii) Now assume $Y \in L^1(\mathcal{F})$ and $(\psi \circ Y)_+ \in L^1(\mathcal{F})$. Fix any point $z \in N$ and define $Y_n \in L^\infty(\mathcal{F})$ with $\psi \circ Y_n \in L^\infty(\mathcal{F})$ by $Y_n := Y$ on $\Omega_n := \{d(Y, z) < n\} \cap |\psi(Y)| < n$ and $Y_n := z$ on $\Omega \backslash \Omega_n$. Then $Y_n \to Y$ in d_1 and thus

$$\mathbb{E}[Y_n | \mathcal{F}_k] \to \mathbb{E}[Y | \mathcal{F}_k].$$

By lower semicontinuity of ψ this implies

$$\liminf_{n\to\infty} \psi(\mathbb{E}[Y_n | \mathcal{F}_k]) \geq \psi(\mathbb{E}[Y | \mathcal{F}_k]).$$

Moreover,

$$
\begin{aligned}
\mathbb{E}[\psi \circ Y|\mathcal{F}_k] &= \lim_{n\to\infty} \mathbb{E}[\psi \circ Y \cdot 1_{\Omega_n}|\mathcal{F}_k] \\
&= \lim_{n\to\infty} \mathbb{E}[\psi \circ Y_n \cdot 1_{\Omega_n}|\mathcal{F}_k] \\
&= \lim_{n\to\infty} \mathbb{E}[\psi \circ Y_n|\mathcal{F}_k] \\
&\geq \liminf_{n\to\infty} \psi(\mathbb{E}[Y_n|\mathcal{F}_k])
\end{aligned}
$$

according to the previous part (i).

(iii) Next assume $Y \in L^1(\mathcal{F}_m)$ and $(\psi \circ Y)_+ \in L^1(\mathcal{F}_m)$. For $k \leq m$ define

$$Y_k = \mathbb{E}[Y \| (\mathcal{F}_n)_{n\geq k}].$$

Using (ii) we deduce iteratively from $Y_k \in L^1(\mathcal{F})$ and $(\psi \circ Y_k)_+ \in L^1(\mathcal{F})$ that $Y_{k-1} \in L^1(\mathcal{F})$ and

$$\psi \circ Y_{k-1} \leq \mathbb{E}[\psi \circ Y_k|\mathcal{F}_{k-1}].$$

Thus, in particular, $(\psi \circ Y_{k-1})_+ \in L^1(\mathcal{F})$. Iterating the previous inequality yields the claim.

(iv) Finally, assume $Y \in L^1_0(\mathcal{F}_\infty)$. Then by lower semicontinuity of ψ,

$$
\begin{aligned}
\psi\big(\mathbb{E}[Y|(\mathcal{F}_n)_{n\geq k}]\big) &\leq \liminf_{m\to\infty} \psi\big(\mathbb{E}[Y|(\mathcal{F}_n)_{m\geq n\geq k}]\big) \\
&\leq \mathbb{E}[\psi \circ Y|\mathcal{F}_k]. \qquad \square
\end{aligned}
$$

COROLLARY 3.5. *Let $Y \in L^1(\mathcal{F})$ and $\psi : N \to \mathbb{R}$ be a lower semicontinuous, convex function with $(\psi(Y))_+ \in L^1(\mathcal{F})$. Then*

$$\mathbb{E}[\psi(Y)] \geq \psi(\mathbb{E}[Y]).$$

This is, of course, part of the previous proposition. However, we present an alternative proof based on the law of large numbers (Theorem 2.6).

PROOF OF COROLLARY 3.5. As in the previous proof, we may assume without restriction that $Y \in L^\infty(\mathcal{F})$ and $\psi(Y) \in L^\infty(\mathcal{F})$. Choose an i.i.d. sequence $(Y_i)_i$ with the same distribution as Y and put $Z_i := \psi(Y_i)$. Moreover, put $S_n := \frac{1}{n}\sum_{i=1}^n Y_i$ and $T_n := \frac{1}{n}\sum_{i=1}^n Z_i$.

Then by the strong law of large numbers (for N-valued and for $\mathbb{R}$-valued random variables, resp.),

$$S_n \to \mathbb{E}Y, \qquad T_n \to \mathbb{E}\psi(Y).$$

Moreover, we claim that

$$\psi(S_n) \leq T_n.$$

Indeed, this is true for $n = 1$ and follows for general n by induction:

$$\begin{aligned}\psi(S_{n+1}) &= \psi\left(\frac{n}{n+1}S_n + \frac{1}{n+1}Y_{n+1}\right)\\ &\le \frac{n}{n+1}\psi(S_n) + \frac{1}{n+1}\psi(Y_{n+1})\\ &\le \frac{n}{n+1}T_n + \frac{1}{n+1}Z_{n+1} = T_{n+1},\end{aligned}$$

where we only used the convexity of ψ along geodesics. □

4. Martingales.

DEFINITION 4.1. (a) Let $L^r(\mathcal{F}.)$ denote the set of equivalence classes of (discrete time, N-valued) processes $Y. = (Y_n)_{n\in\mathbb{N}_0}$ on the filtered probability space $(\Omega, \mathcal{F},$
$(\mathcal{F}_n)_{n\in\mathbb{N}_0}, \mathbb{P})$ with $Y_n \in L^r(\mathcal{F}_n)$ for each $n \in \mathbb{N}_0$.

(b) A process $Y. = (Y_n)_{n\in\mathbb{N}_0} \in L^1(\mathcal{F}.)$ is called (discrete time, N-valued) *martingale* if and only if $\forall\, k \in \mathbb{N}_0$,

$$\mathbb{E}[Y_{k+1}|\mathcal{F}_k] = Y_k. \tag{11}$$

An immediate consequence of the definitions of martingale and filtered conditional expectation and of Jensen's inequality is the following.

PROPOSITION 4.2. (i) *A process $Y. = (Y_n)_{n\in\mathbb{N}_0} \in L^1(\mathcal{F}.)$ is a martingale if and only if for all $k, m \in \mathbb{N}_0$,*

$$\mathbb{E}[Y_{k+m}\|\!|(\mathcal{F}_n)_{n\ge k}] = Y_k. \tag{12}$$

(ii) *Let $Y. \in L^1(\mathcal{F}.)$ be an N-valued martingale and $\psi : N \to \mathbb{R}$ be a lower semicontinuous, convex function with $\psi \circ Y. \in L^1(\mathcal{F}.)$. Then the real-valued process $Z. := \psi(Y.)$ is a submartingale.*

(iii) *Let $X., Y. \in L^1(\mathcal{F}.)$ be two N-valued martingales (w.r.t. the same filtration) then the distance process $d(X., Y.)$ is a real-valued submartingale. In particular, $d_1(X_n, Y_n)$ is increasing in $n \in \mathbb{N}_0$.*

THEOREM 4.3. (i) *For each $Z \in L^1(\mathcal{F}_m)$ there exists a unique N-valued martingale $(Y_k)_{0\le k\le m}$ with $Y_m = Z$; namely,*

$$Y_k = \mathbb{E}[Z\|\!|(\mathcal{F}_n)_{n\ge k}].$$

(ii) *For each* $Z \in L^1_0(\mathcal{F}_\infty)$ *there exists a unique* N*-valued martingale* $(Y_k)_{k\in\mathbb{N}_0}$ *with* $Y_k \to Z$ *in* $L^1(\mathcal{F})$, *namely,*

$$Y_k = \mathbb{E}[Z \,|||\, (\mathcal{F}_n)_{n\geq k}].$$

PROOF. (i) The existence is obvious from the definition of martingales and filtered conditional expectations. The uniqueness follows from Proposition 4.2(iii).

(ii) Again the uniqueness is clear by Proposition 4.2(iii). Namely, if $(Y_n)_n$ and $(Y'_n)_n$ are two such martingales then $d_1(Y_k, Y'_k) \leq d_1(Y_n, Y'_n)$ for all $k \leq n$ and $d_1(Y_n, Y'_n) \to 0$ for all $n \to \infty$.

Existence. Since $Z \in L^1(\mathcal{F}_\infty)$ there exists $Z_m \in L^1(\mathcal{F}_m)$ with $Z_m \to Z$ in $L^1(\mathcal{F})$. Let

$$Y_{m,k} = \mathbb{E}[Z_m \,|||\, (\mathcal{F}_n)_{n\geq k}]$$

and

$$Y_k = \mathbb{E}[Z \,|||\, (\mathcal{F}_n)_{n\geq k}].$$

Then $d_1(Y_{m,k}, Y_k) \leq d_1(Z_m, Z)$ for all m, k and $Y_{m,k} = Z_m$ for $k \geq m$. Hence,

$$d_1(Z, Y_m) \leq 2d_1(Z, Z_m) \to 0$$

for $m \to \infty$. $\square$

REMARK 4.4. (a) We emphasize that for general target spaces N the martingale property (11) does *not imply* that

$$\mathbb{E}[Y_{k+m}|\mathcal{F}_k] = Y_k$$

for $m > 1$. (See Example 3.2.) For a related phenomenon, see Example 6.6 in Sturm (2001a). In particular, if $(Y_n)_{n\in\mathbb{N}_0}$ is a martingale on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}_0}, \mathbb{P})$ then $(Y_{2n})_{n\in\mathbb{N}_0}$ is *not* necessarily a martingale on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_{2n})_{n\in\mathbb{N}_0}, \mathbb{P})$.

(b) The integrability assumption in the definition of martingales can be weakened; it suffices to require that $Y. = (Y_n)_{n\in\mathbb{N}_0}$ is an N-valued process with $Y_n \in L^1(\mathcal{F}_n, \mathcal{F}_{n-1})$ for all $n \in \mathbb{N}$.

(c) The notion of martingales with time parameter $n \in \mathbb{N}_0 \cup \{\infty\}$ can be defined based on Proposition 4.2(i) and Theorem 4.3.

DEFINITION 4.5. For a process $Y. = (Y_n)_{n\in\mathbb{N}_0} \in L^2(\mathcal{F}.)$ we define the *bracket* $[Y].$ by

$$[Y]_n := \sum_{k=1}^{n} d^2(Y_k, Y_{k-1}),$$

the *sharp bracket* $\langle Y \rangle_.$ by

$$\langle Y \rangle_n := \sum_{k=1}^{n} \mathbb{E}[d^2(Y_k, Y_{k-1})|\mathcal{F}_{k-1}]$$

and the *reduced bracket* $[[Y]]_.$ by

$$[[Y]]_n := \sum_{k=1}^{n} d^2(Y_k, \mathbb{E}[Y_k|\mathcal{F}_{k-1}]).$$

PROPOSITION 4.6. *For a process $Y_. \in L^2(\mathcal{F}_.)$ the following are equivalent*:

(i) *$Y_.$ is a martingale*;
(ii) $[Y]_. = [[Y]]_.$;
(iii) $\mathbb{E}[Y]_. = \mathbb{E}[[Y]]_.$.

PROOF. (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) is trivial.

(iii) $\Rightarrow$ (i). The variance inequality from Theorem 2.1 implies

$$\mathbb{E}d^2(Y_{k-1}, Y_k) \geq \mathbb{E}d^2(\mathbb{E}_{\mathcal{F}_{k-1}} Y_k, Y_k) + \mathbb{E}d^2(\mathbb{E}_{\mathcal{F}_{k-1}} Y_k, Y_{k-1});$$

that is,

$$\mathbb{E}[Y]_n \geq \mathbb{E}[[Y]]_n + \sum_{k=1}^{n} \mathbb{E}d^2(\mathbb{E}_{\mathcal{F}_{k-1}} Y_k, Y_{k-1}).$$

Assuming now $\mathbb{E}[Y]_n = \mathbb{E}[[Y]]_n$, implies

$$\mathbb{E}_{\mathcal{F}_{k-1}} Y_k = Y_{k-1}, \qquad \mathbb{P}\text{-a.s.}$$

for all $k \leq n$. That is, (i). $\square$

THEOREM 4.7. *Let $Y_. \in L^2(\mathcal{F}_.)$ with separable range. Then the following assertions are equivalent*:

(i) *$Y_.$ is a martingale*;
(ii) *for each $z \in N$, the $\mathbb{R}$-valued process $Z_. = (Z_n)_{n \in \mathbb{N}_0}$ defined by $Z_n := d^2(z, Y_n) - [Y]_n$ is a submartingale*;
(iii) *for each $z \in N$, the $\mathbb{R}$-valued process $Z_. = (Z_n)_{n \in \mathbb{N}_0}$ defined by $Z_n := d^2(z, Y_n) - \langle Y \rangle_n$ is a submartingale*;
(iv) *for each $z \in N$, the $\mathbb{R}$-valued process $Z_. = (Z_n)_{n \in \mathbb{N}_0}$ defined by $Z_n := d^2(z, Y_n) - [[Y]]_n$ is a submartingale*.

PROOF. (i) $\Rightarrow$ (ii). Let $Y_.$ be a martingale and let $Z_.$ be the process defined in (ii). Then for all $n \in \mathbb{N}_0$, by the conditional variance inequality,

$$\begin{aligned}\mathbb{E}[Z_{n+1}|\mathcal{F}_n] &= \mathbb{E}\big[d^2(z, Y_{n+1}) - [Y]_{n+1}|\mathcal{F}_n\big] \\ &\geq d^2(z, Y_n) - [Y]_n = Z_n.\end{aligned}$$

(ii) $\Rightarrow$ (iii). According to (ii),

$$\mathbb{E}\big[d^2(z, Y_{n+1}) - [Y]_{n+1}|\mathcal{F}_n\big] \geq d^2(z, Y_n) - [Y]_n$$

for all $n \in \mathbb{N}_0$ and all $z \in N$. This is equivalent to

$$\mathbb{E}[d^2(z, Y_{n+1})|\mathcal{F}_n] \geq d^2(z, Y_n) + \mathbb{E}[d^2(Y_n, Y_{n+1})|\mathcal{F}_n],$$

which in turn is equivalent to (iii).

(iii) $\Rightarrow$ (iv). Now assume (iii) and let $Z.$ be the process in (iv). Again, the conditional variance inequality implies

$$\begin{aligned}\mathbb{E}[Z_{n+1}|\mathcal{F}_n] &= \mathbb{E}[d^2(z, Y_{n+1})|\mathcal{F}_n] - \mathbb{E}\big[d^2(Y_{n+1}, \mathbb{E}[Y_{n+1}|\mathcal{F}_n])|\mathcal{F}_n\big] - [[Y]]_n \\ &\geq \mathbb{E}[d^2(z, Y_{n+1})|\mathcal{F}_n] - \mathbb{E}[d^2(Y_{n+1}, Y_n)|\mathcal{F}_n] - [[Y]]_n \\ &\geq d^2(z, Y_n) - [[Y]]_n = Z_n.\end{aligned}$$

(iv) $\Rightarrow$ (i). By assumption $\forall\, z \in N$,

$$\mathbb{E}[d^2(z, Y_{n+1})|\mathcal{F}_n] \geq d^2(z, Y_n) + \mathbb{E}\big[d^2(Y_{n+1}, \mathbb{E}[Y_{n+1}|\mathcal{F}_n])|\mathcal{F}_n\big].$$

According to Theorem 2.2 this is equivalent to

$$\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = Y_n. \qquad \square$$

REMARK 4.8. (a) The implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) in Theorem 4.7 are valid without the separability of $Y_n(\Omega)$. In particular, they imply that for each N-valued martingale $Y.$ and for each $z \in N$ the $\mathbb{R}$-valued process $d^2(z, Y.)$ is a submartingale.

(b) The crucial point in Theorem 4.7 is that properties (ii) and (iii) characterize N-valued martingales in terms of $\mathbb{R}$-valued processes, without using any kind of nonlinear expectation or nonlinear conditional expectation.

(c) In the ("classical") case of continuous time processes with values in Riemannian manifolds, martingales can be characterized by means of property (ii) of Proposition 4.2. Namely, $Y.$ is a martingale if and only if for each lower semicontinuous convex function $\psi : N \to \mathbb{R}$ the process $Z. = \psi(Y.)$ is a submartingale. [See Emery (1989) or Kendall (1990).] This characterization does not hold in our framework!

Actually, such a characterization would immediately imply the optional sampling theorem. Indeed, if for all lower semicontinuous convex function $\psi : N \to \mathbb{R}$ the processes $(\psi(Y_n))_{n\in\mathbb{N}_0}$ are submartingales w.r.t. the filtration $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$, then so are the processes $(\psi(Y_{T_n}))_{n\in N_0}$ w.r.t. the filtration $(\mathcal{F}_{T_n})_{n\in\mathbb{N}_0}$ for each increasing sequence of bounded stopping times T_n, $n \in \mathbb{N}_0$. The above martingale characterization would then imply that $(Y_{T_n})_{n\in\mathbb{N}_0}$ is an N-valued martingale w.r.t. the filtration $(\mathcal{F}_{T_n})_{n\in\mathbb{N}_0}$. However, this implication does not hold! See Example 3.2 (with $T_n = 2n$).

COROLLARY 4.9 (Optional stopping theorem). *Let $Y_\cdot \in L^2(\mathcal{F}_\cdot)$ be a martingale with separable range and let T be a stopping time. Then the stopped process $Y_\cdot^T = (Y_{T\wedge n})_{n\in\mathbb{N}_0}$ is a martingale* [*w.r.t. the original filtration $\mathcal{F}_\cdot = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$ as well as w.r.t. the filtration $\mathcal{F}^T_\cdot = (\mathcal{F}_{T\wedge n})_{n\in\mathbb{N}_0}$*].

PROOF. Since $Y_\cdot$ is a martingale, for each $z \in N$ the process

$$Z_n = d^2(z, Y_n) - [Y]_n$$

is a submartingale. By the classical optional stopping theorem for submartingales, the stopped process

$$Z_n^T = d^2(z, Y_{T\wedge n}) - [Y]_{T\wedge n} = d^2(z, Y_n^T) - [Y^T]_n$$

is a submartingale. More precisely, it is a submartingale w.r.t. the filtration $\mathcal{F}_\cdot = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$ as well as w.r.t. the filtration $\mathcal{F}^T_\cdot = (\mathcal{F}_{T\wedge n})_{n\in\mathbb{N}_0}$. Since this holds for each $z \in N$, Theorem 4.6 implies that $Y_\cdot^T$ is a martingale (w.r.t. both filtrations). □

REMARK. The optional sampling theorem in the usual form will not be true in the general nonlinear framework. In particular, given a martingale Y and bounded stopping times S, T, in general,

$$\mathbb{E}[Y_S] \neq \mathbb{E}[Y_T].$$

See Example 3.2. However, an appropriate version holds true for the filtered expectations.

COROLLARY 4.10. *Let $Y_\cdot \in L^2(\mathcal{F}_\cdot)$ be a martingale with separable range and let S, T be bounded stopping times. Then*

$$\mathbb{E}\big[Y_S \,\|\!|\, (\mathcal{F}_n)_{n\in\mathbb{N}_0}\big] = \mathbb{E}\big[Y_T \,\|\!|\, (\mathcal{F}_n)_{n\in\mathbb{N}_0}\big].$$

PROOF. Since the stopped process Y^T is a martingale, we get for $m \geq k$,

$$\mathbb{E}[Y_{T\wedge m} \,\|\!|\, (\mathcal{F}_n)_{n\geq k}] = Y_{T\wedge k}.$$

Choosing $m \geq T$ and $k = 0$ yields $\mathbb{E}[Y_T \,\|\!|\, (\mathcal{F}_n)_{n\geq k}] = Y_0$. Since the same result holds with S in the place of T, the claim is proven. □

THEOREM 4.11 (Martingale convergence theorem). *Let $Y_\cdot \in L^r(\mathcal{F}_\cdot)$ be a martingale with locally compact, separable range (i.e., there exists a locally compact, separable subset $N_0 \subset N$ such that $Y_n \in N_0$ $\mathbb{P}$-a.s. for all $n \in \mathbb{N}_0$) and let $r \in [1, \infty[$. Recall that $\mathcal{F}_\infty = \bigvee_{n\in\mathbb{N}_0} \mathcal{F}_n$. Assume that $Y_\cdot$ is uniformly L^r-bounded, i.e., $\sup_n d_r(z, Y_n) < \infty$ for some/all $z \in N$.*

(i) *Then there exists an $\mathcal{F}_\infty$-measurable map $Y_\infty : \Omega \to N$ such that*

$$Y_n \to Y_\infty \qquad \mathbb{P}\text{-}a.s.$$

(ii) *If $r > 1$ then also $Y_n \to Y_\infty$ in $L^r(\mathcal{F})$.*
(iii) *If the latter holds true for some $r \in [1, \infty]$ then for each $n \in \mathbb{N}_0$,*

$$Y_n = \mathbb{E}[Y_\infty \| (\mathcal{F}_k)_{k \geq n}].$$

REMARK. In the case $r = 2$, the process $(Y_n)_{n \in \mathbb{N}_0 \cup \{\infty\}}$ satisfies the martingale characterizations (ii) and (iii) of Theorem 4.7 on the filtered probability space $(\Omega, \mathcal{F}_n, \mathbb{P})_{n \in \mathbb{N}_0 \cup \{\infty\}}$.

PROOF. (i) Fix $z \in N$ and put $R_{z,n} := d(z, Y_n) : \Omega \to \mathbb{R}$. Then by Theorem 3.4 $(R_{z,n})_{n \in \mathbb{N}_0}$ is a submartingale with

$$\sup_n \mathbb{E} R_{z,n}^r < \infty.$$

Hence, by the classical submartingale convergence theorem there exists an $\mathcal{F}_\infty$-measurable random variable $R_{z,\infty} : \Omega \to \mathbb{R}$ such that

$$\lim_{n \to \infty} R_{z,n} = R_{z,\infty}, \qquad \mathbb{P}\text{-a.s.}$$

Now choose a countable set N_1 which is dense in the compact set $N_0 \subset N$. Then there exists a set $\Omega_1 \subset \Omega$ with $\mathbb{P}(\Omega_1) = 1$ such that

$$\lim_{n \to \infty} R_{z,n}(\omega) = R_{z,\infty}(\omega) \qquad \forall\, z \in N_1,\ \forall\, \omega \in \Omega_1.$$

By density of N_1 in N_0 and by uniform continuity of $z \mapsto R_{z,n}(\omega)$ it follows that

$$\lim_{n \to \infty} R_{z,n}(\omega) = R_{z,\infty}(\omega) \qquad \forall\, z \in N_0,\ \forall\, \omega \in \Omega_1. \tag{13}$$

Without restriction we may assume

$$Y_n(\omega) \in N_0 \qquad \forall\, \omega \in \Omega_1,\ \forall\, n \in \mathbb{N}_0.$$

Note that in a global NPC space all closed, locally compact balls are compact. Hence, for each $\omega \in \Omega_1$ by compactness there exists a point $Y_\infty(\omega) \in N_0$ and a subsequence $(Y_{n_k}(\omega))_{k \in \mathbb{N}}$ such that

$$\lim_{k \to \infty} Y_{n_k}(\omega) = Y_\infty(\omega).$$

Now for each $\omega \in \Omega_1$ choose $z = Y_\infty(\omega)$ in (13) in order to obtain that

$$\lim_{n \to \infty} R_{z,n}(\omega) = \lim_{n \to \infty} d\Big(\lim_{k \to \infty} Y_{n_k}(\omega), Y_n(\omega)\Big)$$

exists (and then of course vanishes). In other words, $\forall\, \omega \in \Omega_1$,

$$\lim_{n \to \infty} Y_n(\omega) = Y_\infty(\omega).$$

This obviously implies that Y_∞ is $\mathcal{F}_\infty$-measurable.

(ii) Finally, the a.s.-convergence together with the uniform L^r-boundedness yields the L^r-convergence of the Y_n for $n \to \infty$. [This in particular implies $Y_\infty \in L_0^r(\mathcal{F}_\infty)$.]

(iii) Fix $n \in \mathbb{N}_0$. Since $Y_m \to Y_\infty$ in $L^r(\mathcal{F}_\infty)$ for $m \to \infty$, we get

$$\mathbb{E}[Y_m \| (\mathcal{F}_k)_{k \geq n}] \to \mathbb{E}[Y_\infty \| (\mathcal{F}_k)_{k \geq n}] \qquad \text{in } L^r(\mathcal{F}_n)$$

for $m \to \infty$. But since $Y.$ is a martingale, $\mathbb{E}[Y_m \| (\mathcal{F}_k)_{k \geq n}] = Y_n$ for all sufficiently large m. □

5. The nonlinear Markov property. From now on, let $(M, \mathcal{M})$ be a measurable space and let $(\mathbb{P}_x, X_n)_{x \in M, n \in \mathbb{N}_0}$ be a Markov chain with values in M and being defined on some filtered measurable space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \in \mathbb{N}_0})$ with shifts Θ_n, $n \in \mathbb{N}_0$.

In the sequel, Y will always denote a measurable map $Y : \Omega \to N$ with separable range $Y(\Omega)$ and such that all the maps $Y \circ \Theta_k$ have finite variances w.r.t. all the probability measures $\mathbb{P}_x$, that is, $Y \circ \Theta_k \in L^2((\Omega, \mathcal{F}, \mathbb{P}_x), (N, d))$ for all $x \in M$, $k \in \mathbb{N}_0$. For instance, each bounded map satisfies this condition.

LEMMA 5.1. *The map $x \mapsto \mathbb{E}_x[Y]$ is $\mathcal{M}$-measurable.*

PROOF. In the case $Y = f(X_1)$ the result is proven in Sturm (2001a), Lemma 6.4. Replacing in that proof the kernel $p(x, dy) = \mathbb{P}_x(X_1 \in dy)$ (defined on $M \times \mathcal{M}$) by the kernel $\mathbb{P}_x(d\omega)$ (defined on $M \times \mathcal{F}$), the same arguments also apply to the general situation. □

THEOREM 5.2 (Nonlinear Markov property). *For all $k \in \mathbb{N}_0$ and all $x \in M$,*

$$\mathbb{E}_x[Y \circ \Theta_k | \mathcal{F}_k] = \mathbb{E}_x[Y \circ \Theta_k | X_k] = \mathbb{E}_{X_k}[Y], \qquad \mathbb{P}_x\text{-a.s.}$$

PROOF. (a) In order to make the arguments clear, we plug in all the integration variables. Let us first derive the following form of the *linear* Markov property: given measurable maps $Y, Y' : \Omega \to N$ and $u : N \times N \to \mathbb{R}_+$ then

$$\iint u\big(Y'(\omega), Y(\omega')\big)\mathbb{P}_{X_k(\omega)}(d\omega')\mathbb{P}_x(d\omega) = \int u\big(Y'(\omega), (Y \circ \Theta_k)(\omega)\big)\mathbb{P}_x(d\omega)$$

provided Y' is $\mathcal{F}_k$-measurable. By a monotone class argument it suffices to prove this for functions u of the form $u(y', y) = v(y') \cdot w(y)$ with measurable $v, w : N \to \mathbb{R}_+$. In this case, the usual Markov property obviously yields

$$\begin{aligned}
&\iint v\big(Y'(\omega)\big) \cdot w\big(Y(\omega')\big)\mathbb{P}_{X_k(\omega)}(d\omega')\mathbb{P}_x(d\omega) \\
&\quad = \mathbb{E}_x\big[v \circ Y' \cdot \mathbb{E}_{X_k}[w \circ Y]\big] = \mathbb{E}_x\big[v \circ Y' \cdot \mathbb{E}_x[w \circ Y \circ \Theta_k | \mathcal{F}_k]\big] \\
&\quad = \mathbb{E}_x\big[\mathbb{E}_x[(v \circ Y') \cdot (w \circ Y \circ \Theta_k) | \mathcal{F}_k]\big] = \mathbb{E}_x[(v \circ Y') \cdot (w \circ Y \circ \Theta_k)],
\end{aligned}$$

which is the claim.

(b) The variance inequality (applied to the map Y) yields for all $x \in M$ and $z \in N$,

$$d^2\left(z, \int Y(\omega')\mathbb{P}_x(d\omega')\right) \leq \int d^2\big(z, Y(\omega')\big)\mathbb{P}_x(d\omega') - \int d^2\left(\int Y(\omega'')\mathbb{P}_x(d\omega''), Y(\omega')\right)\mathbb{P}_x(d\omega').$$

Choosing $x = X_k(\omega)$ and $z = Z(\omega)$ for some $\mathcal{F}_k$-measurable map $Z : \Omega \to N$ yields

$$d^2\big(Z(\omega), Z'(\omega)\big) \leq \int d^2\big(Z(\omega), Y(\omega')\big)\mathbb{P}_{X_k(\omega)}(d\omega') - \int d^2\big(Z'(\omega), Y(\omega')\big)\mathbb{P}_{X_k(\omega)}(d\omega'),$$

where we have put $Z'(\omega) = \int Y(\omega')\mathbb{P}_{X_k(\omega)}(d\omega')$. Integrating w.r.t. $\mathbb{P}_x(d\omega)$ and applying the *linear* Markov property yields (since Z and Z' are $\mathcal{F}_k$-measurable)

$$\begin{aligned}\int d^2\big(Z(\omega), Z'(\omega)\big)\mathbb{P}_x(d\omega) &\leq \iint d^2\big(Z(\omega), Y(\omega')\big)\mathbb{P}_{X_k(\omega)}(d\omega')\mathbb{P}_x(d\omega) \\ &\quad - \iint d^2\big(Z'(\omega), Y(\omega')\big)\mathbb{P}_{X_k(\omega)}(d\omega')\mathbb{P}_x(d\omega) \\ &= \int d^2\big(Z(\omega), (Y \circ \Theta_k)(\omega)\big)\mathbb{P}_x(d\omega) \\ &\quad - \int d^2\big(Z'(\omega), (Y \circ \Theta_k)(\omega)\big)\mathbb{P}_x(d\omega),\end{aligned}$$

provided the integrals are finite. That is, we end up with

$$\mathbb{E}_x d^2(Z, Z') \leq \mathbb{E}_x d^2(Z, Y \circ \Theta_k) - \mathbb{E}_x d^2(Z', Y \circ \Theta_k) = \text{RHS}$$

for $Z' := \mathbb{E}_{X_k}[Y]$ and for any $\mathcal{F}_k$-measurable Z for which $\mathbb{E}_x d^2(Z, Y \circ \Theta_k) < \infty$.

(c) Choosing now $Z = \mathbb{E}_x[Y \circ \Theta_k|\mathcal{F}_k]$, the assumptions on Y imply that $\mathbb{E}_x d^2(Z, Y \circ \Theta_k) < \infty$, hence, also $\mathbb{E}_x d^2(Z', Y \circ \Theta_k) < \infty$. Applying the conditional variance inequality to the map $Y \circ \Theta_k$ yields that

$$\text{RHS} \leq -\mathbb{E}_x d^2(Z, Z').$$

That is, $Z = Z'$ or, in other words,

$$\mathbb{E}_x[Y \circ \Theta_k|\mathcal{F}_k] = \mathbb{E}_{X_k}[Y].$$

Of course, this implies that $\mathbb{E}_x[Y \circ \Theta_k|\mathcal{F}_k]$ is $\sigma(X_k)$-measurable, hence,

$$\mathbb{E}_x[Y \circ \Theta_k|\mathcal{F}_k] = \mathbb{E}_x[Y \circ \Theta_k|X_k]$$

by the conditional variance inequality. $\square$

Mostly, the nonlinear Markov property will be applied to maps of the form $Y = f(X_n)$. In this particular case, it reads as follows.

COROLLARY 5.3. *Let $f : M \to N$ be a measurable map with separable range and such that the maps $f(X_n)$, $n \in \mathbb{N}_0$, have finite variances w.r.t. all the probability measures $\mathbb{P}_x$, $x \in M$. Then for all $k, n \in \mathbb{N}_0$ and all $x \in M$,*

$$\mathbb{E}_x[f(X_{k+n})|\mathcal{F}_k] = \mathbb{E}_x[f(X_{k+n})|X_k] = \mathbb{E}_{X_k} f(X_n), \qquad \mathbb{P}_x\text{-a.s.}$$

Note that [in contrast to the martingale property (11)] this also holds true for $n > 1$. Indeed, the nonlinear Markov property for conditional expectations will be used in the sequel only with $k = 1$. For $k > 1$, one needs a nonlinear Markov property for filtered conditional expectations.

THEOREM 5.4. *For all $k \in \mathbb{N}_0$ and all $x \in M$,*

$$\mathbb{E}_x[Y \circ \Theta_k \| (\mathcal{F}_n)_{n \geq k}] = \mathbb{E}_{X_k}[Y \| (\mathcal{F}_n)_{n \geq 0}], \qquad \mathbb{P}_x\text{-a.s.}$$

PROOF. A straightforward generalization of the nonlinear Markov property of Theorem 5.2 yields

$$\begin{aligned}\mathbb{E}_x[Y \circ \Theta_k|\mathcal{F}_{k+m}] &= \mathbb{E}_x[Y \circ \Theta_k|X_k, \ldots, X_{k+m}] = \mathbb{E}_{X_0}[Y|X_0, \ldots, X_m] \circ \Theta_k \\ &= \mathbb{E}_{X_0}[Y|\mathcal{F}_m] \circ \Theta_k.\end{aligned}$$

Hence, for $Y \in \mathcal{F}_m$ (which implies $Y \circ \Theta_k \in \mathcal{F}_{k+m}$),

$$\begin{aligned}&\mathbb{E}_x[Y \circ \Theta_k \| (\mathcal{F}_n)_{n \geq k}] \\ &\quad = \mathbb{E}_x\big[\mathbb{E}_x\big[\cdots \mathbb{E}_x\big[\mathbb{E}_x[Y \circ \Theta_k|\mathcal{F}_{k+m-1}]|\mathcal{F}_{k+m-2}\big] \cdots |\mathcal{F}_{k+1}\big]|\mathcal{F}_k\big] \\ &\quad = \mathbb{E}_x\big[\mathbb{E}_x\big[\cdots \mathbb{E}_x\big[\mathbb{E}_{X_0}[Y|\mathcal{F}_{m-1}] \circ \Theta_k|\mathcal{F}_{k+m-2}\big] \cdots |\mathcal{F}_{k+1}\big]|\mathcal{F}_k\big] \\ &\quad = \mathbb{E}_x\big[\mathbb{E}_x\big[\cdots \mathbb{E}_{X_0}\big[\mathbb{E}_{X_0}[Y|\mathcal{F}_{m-1}]|\mathcal{F}_{m-2}\big] \circ \Theta_k \cdots |\mathcal{F}_{k+1}\big]|\mathcal{F}_k\big] \\ &\quad = \cdots \\ &\quad = \mathbb{E}_{X_0}\big[\mathbb{E}_{X_0}\big[\cdots \mathbb{E}_{X_0}\big[\mathbb{E}_{X_0}[Y|\mathcal{F}_{m-1}]|\mathcal{F}_{m-2}\big] \cdots |\mathcal{F}_1\big]|\mathcal{F}_0\big] \circ \Theta_k \\ &\quad = \mathbb{E}_{X_0}[Y \| (\mathcal{F}_n)_{n \geq k}] \circ \Theta_k = \mathbb{E}_{X_k}[Y \| (\mathcal{F}_n)_{n \geq k}].\end{aligned}$$

The extension from $Y \in \bigcup_m L^2(\mathcal{F}_m)$ to $Y \in L^2_0(\mathcal{F}_\infty)$ is obvious. $\square$

6. Harmonic maps and martingales. In this section, we want to establish the relation between N-valued martingales and harmonic maps $f : M \to N$ as introduced in Sturm (2001a).

We denote by $\mathcal{L}(M)$ the set of all measurable maps $f : M \to N$ with separable range $f(M)$ and such that the random variables $f(X_k) : \Omega \to N$ have finite variances w.r.t. all the probability measures $\mathbb{P}_x$ for all $x \in M$, $k \in \mathbb{N}_0$.

Similarly, we denote by $\mathcal{L}(\mathbb{N}_0 \times M)$ the set of all measurable maps $f : \mathbb{N}_0 \times M \to N$ with separable range $f(\mathbb{N}_0 \times M)$ and such that the random variables $f(n, X_k) : \Omega \to N$ have finite variances w.r.t. all the probability measures $\mathbb{P}_x$; that is,

$$\inf_{z \in N} \mathbb{E}_x d^2\big(z, f(n, X_k)\big) < \infty,$$

for all $x \in M$ and $k, n \in \mathbb{N}_0$ with $k \leq n$.

Note that each *bounded* measurable map $f : M \to N$ (or $f : \mathbb{N}_0 \times M \to N$) with separable range lies in $\mathcal{L}(M)$ [or $\mathcal{L}(\mathbb{N}_0 \times M)$, resp.].

We define the *nonlinear Markov operator* $P = P_{M,N}$ acting on $\mathcal{L}(M)$ by

$$Pf(x) := \mathbb{E}_x f(X_1) \qquad \forall\, x \in M,\ f \in \mathcal{L}(M).$$

DEFINITION 6.1. (i) A map $f \in \mathcal{L}(M)$ is called *harmonic* on M if and only if $Pf = f$ on M.

(ii) A map $f \in \mathcal{L}(\mathbb{N} \times M)$ is called *space–time harmonic* or *solution of the nonlinear heat equation* if and only if

$$f(n+1, \cdot) = Pf(n, \cdot) \qquad \text{for all } n \in \mathbb{N}_0.$$

THEOREM 6.2. *For a map $f \in \mathcal{L}(\mathbb{N}_0 \times M)$ the following properties are equivalent*:

(i) *f is a solution of the nonlinear heat equation*.

(ii) *For all $n \in \mathbb{N}_0$ and all $x \in M$ the process $(f(n-k, X_k))_{k=0,1,\dots,n}$ is a martingale w.r.t. $\mathbb{P}_x$*.

PROOF. Denote the map $f(k, \cdot)$ by f_k. Fix $n \in \mathbb{N}_0$ and put $Y_k := f_{n-k}(X_k)$ for $k = 0, 1, \dots, n$. Then according to the nonlinear Markov property,

$$\mathbb{E}_x[Y_{k+1}|\mathcal{F}_k] = \mathbb{E}_x[f_{n-k-1}(X_{k+1})|\mathcal{F}_k] = \mathbb{E}_{X_k} f_{n-k-1}(X_1) = Pf_{n-k-1}(X_k).$$

Now assume (i). Then

$$\text{RHS} = f_{n-k}(X_k) = Y_k.$$

Hence, $\mathbb{E}_x[Y_{k+1}|\mathcal{F}_k] = Y_k$ which is (ii).

Conversely, assume (ii). Then

$$\text{LHS} = Y_k = f_{n-k}(X_k).$$

For $k = 0$ this yields $f_n(X_0) = Pf_{n-1}(X_0)$. Of course, all the above equalities hold a.s. w.r.t. $\mathbb{P}_x$ for all x. Hence,

$$f_n(x) = Pf_{n-1}(x)$$

for all $x \in M$ which is (i). $\square$

COROLLARY 6.3. *Let $f \in \mathcal{L}(\mathbb{N}_0 \times M)$ be a solution of the nonlinear heat equation. Then f is uniquely determined by $f(0, \cdot)$ in the following way*:

$$f(n, x) = \mathbb{E}_x\big[f(0, X_n) \| (\mathcal{F}_k)_{k \in \mathbb{N}_0}\big],$$

that is, it is the filtered expectation of the map $f(0, X_n)$ w.r.t. the probability measure $\mathbb{P}_x$. Or, in other words, it is the starting point Y_0 of a martingale $(Y_k)_{k=0,1,\dots,n}$ with terminal value $Y_n = f(0, X_n)$.

Of course, in general, the map $\tilde{f}$ defined by $\tilde{f}(n, x) = \mathbb{E}_x[f(0, X_n)]$ will be no solution of the nonlinear heat equation.

COROLLARY 6.4. *A map $f \in \mathcal{L}(M)$ is harmonic on M if and only if the process $(f(X_k))_{k \in \mathbb{N}_0}$ is a martingale w.r.t. all probability measure $\mathbb{P}_x$, $x \in M$.*

7. The Dirichlet problem. The previous results easily extend to the following more general situation. Let D be any measurable subset of M. Denote by $\mathcal{L}(D)$ the set of all measurable maps $f : M \to N$ with separable range $f(M)$ and such that the random variables $f(X_k) : \Omega \to N$ have finite variances w.r.t. all the probability measures $\mathbb{P}_x$ for all $x \in D$, $k \in \mathbb{N}_0$. That is,

$$v(x) := \inf_{z \in N} \mathbb{E}_x d^2\big(z, f(X_k)\big) < \infty$$

for all $x \in D$, $k \in \mathbb{N}_0$.

Note that the map f has to be defined on the whole space M whereas the condition $v(x) < \infty$ is required only for $x \in D$. However, even for $x \in D$ the number $v(x)$ depends also on the value of f on $M \setminus D$.

We say that a map $f \in \mathcal{L}(D)$ is *harmonic on D* if and only if $Pf = f$ on D.

THEOREM 7.1. *The map $f \in \mathcal{L}(D)$ is harmonic on D if and only if the stopped process $(f(X_{k \wedge \tau})_{k \in \mathbb{N}_0}$ is a martingale w.r.t. all probability measure $\mathbb{P}_x$, $x \in M$, where $\tau = \tau(D) = \inf\{n \in \mathbb{N}_0 : X_n \notin D\}$ denotes the first exit time of D.*

PROOF. Let $X'_k := X_{k \wedge \tau}$ denote the stopped Markov chain. Note that $\mathbb{P}_x$-a.s.,

$$X'_1 = \begin{cases} X_1, & \text{if } x \in D, \\ x, & \text{if } x \notin D. \end{cases}$$

Hence, the associated nonlinear Markov operator is given by

$$P' f(x) = \mathbb{E}_x f(X'_1) = \begin{cases} Pf(x), & \text{if } x \in D, \\ f(x), & \text{if } x \notin D. \end{cases}$$

Therefore,

$$\begin{aligned} f \text{ is harmonic on } D \quad &\iff \quad Pf = f \text{ on } D \\ &\iff \quad P'f = f \text{ on } M \\ &\iff \quad \big(f(X'_k)\big)_{k\in\mathbb{N}_0} \text{ is a martingale} \\ &\iff \quad \big(f(X_{k\wedge\tau})\big)_{k\in\mathbb{N}_0} \text{ is a martingale.} \qquad \square \end{aligned}$$

DEFINITION 7.2. Given a map $g \in \mathcal{L}(D)$ we say that a map $f \in \mathcal{L}(D)$ is a solution of the *Dirichlet problem* for the map g if and only if f is harmonic on D and coincides with g on $M \setminus D$.

THEOREM 7.3. *Assume that $\tau = \tau(D) < \infty$ $\mathbb{P}_x$-a.s. for all $x \in M$. Then for each bounded map $g \in \mathcal{L}(D)$ there exists a unique bounded map $f \in \mathcal{L}(D)$ which solves the Dirichlet problem for the map g. It is given as the filtered expectation of the random map $g(X_\tau)$; that is, for each $x \in M$,*

$$f(x) = \mathbb{E}_x\big[g(X_\tau) \| \!|(\mathcal{F}_n)_{n\in\mathbb{N}_0}\big].$$

PROOF. If f is harmonic in D then $(f(X_{\tau\wedge k}))_k$ is a martingale. Hence,

$$f(x) = \mathbb{E}_x\big[f(X_{\tau\wedge k}) \| \!|(\mathcal{F}_n)_{n\in\mathbb{N}_0}\big]$$

for all $k \in \mathbb{N}_0$. Since $\tau < \infty$ and f is bounded, we get that $f(X_{\tau\wedge k}) \to f(X_\tau)$ for $k \to \infty$ in $L^2(\mathcal{F}_\infty)$. Hence,

$$f(x) = \mathbb{E}_x\big[f(X_\tau) \| \!|(\mathcal{F}_n)_{n\in\mathbb{N}_0}\big] = \mathbb{E}_x\big[g(X_\tau) \| \!|(\mathcal{F}_n)_{n\in\mathbb{N}_0}\big].$$

Conversely, define f by $f(x) = \mathbb{E}_x[g(X_\tau) \| \!|(\mathcal{F}_n)_{n\in\mathbb{N}_0}]$. Then of course $f(x) = g(x)$ for $x \in M \setminus D$. For $x \in D$ the nonlinear Markov property for filtered expectations yields

$$\begin{aligned} Pf(x) &= \mathbb{E}_x\big[\mathbb{E}_{X_1}[g(X_\tau) \| \!|(\mathcal{F}_n)_{n\geq 0}]\big] = \mathbb{E}_x\big[\mathbb{E}_x[g(X_\tau) \circ \Theta_1 \| \!|(\mathcal{F}_n)_{n\geq 1}]\big] \\ &= \mathbb{E}_x[g(X_\tau) \circ \Theta_1 \| \!|(\mathcal{F}_n)_{n\geq 0}] = \mathbb{E}_x[g(X_\tau) \| \!|(\mathcal{F}_n)_{n\geq 0}] = f(x). \end{aligned}$$

This proves the claim. $\square$

It should be clear from the previous discussions that, in general, the map

$$\tilde{f}(x) = \mathbb{E}_x[g(X_\tau)|]$$

does *not solve* the Dirichlet problem for g.

Acknowledgments. The author is grateful to James Eells, Bent Fuglede, Jürgen Jost and Anton Thalmaier for stimulating discussions, and to Daniel Speicher for contributing to simplify various proofs (e.g., of Theorems 2.1 and 5.2).

REFERENCES

BALLMANN, W. (1995). *Lectures on Spaces of Nonpositive Curvature.* Birkhäuser, Berlin.

CARTAN, H. (1928). *Leçons sur la géometrie des espaces de Riemann.* Gauthiers-Villars, Paris.

DOSS, S. (1949). Sur la moyenne d'un élément aléatoire dans un espace distancié. *Bull. Sci. Math.* **73** 48–72.

EELLS, J. and FUGLEDE, B. (2001). *Harmonic Maps between Riemannian Polyhedra.* Cambridge Univ. Press.

EMERY, D. (1989). *Stochastic Calculus on Manifolds.* Springer, New York.

EMERY, D. and MOKOBODZKI, G. (1991). Sur le barycentre d'une probabilité dans une variété. *Séminaire de Probabilités XXV. Lecture Notes in Math.* **1485** 220–233. Springer, Berlin.

ES-SAHIB, A. and HEINICH, H. (1999). Barycentres canoniques pour un espace métrique à courbure négative. *Seminaire de Probabilités XXXIII. Lecture Notes in Math.* **1709** 355–370. Springer, Berlin.

FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré* **10** 215–310.

GAUSS, C. F. (1809). *Theoria Motus Corporum Celestium*.

HERER, W. (1991). Espérance mathématique au sens de Doss d'une variable aléatoire dans un espace métrique. *C. R. Acad. Sci. Paris Sér. I* **302** 131–134.

JOST, J. (1994). Equilibrium maps between metric spaces. *Calc. Var. Partial Differential Equations* **2** 173–204.

JOST, J. (1997). Nonpositive curvature: geometric and analytic aspects. *Lectures Math. ETH Zürich.* Birkhäuser, Basel.

KENDALL, W. (1990). Probability, convexity, and harmonic maps with small images I. Uniqueness and fine existence. *Proc. London Math. Soc.* **61** 371–406.

KOREVAAR, N. and SCHOEN, R. (1993). Sobolev spaces and harmonic maps for metric space targets. *Comm. Anal. Geom.* **1** 561–569.

PICARD, J. (1994). Barycentres et martingales sur une variété. *Ann. Inst. H. Poincaré* **30** 647–702.

STURM, K. T. (2001a). Nonlinear Markov operators, discrete heat flow, and harmonic maps between singular spaces. *Potential Anal.* To appear.

STURM, K. T. (2001b) Nonlinear Markov operators associated with symmetric Markov kernels and energy minimizing maps between singular spaces. *Calc. Var. Partial Differential Equations* **12** 317–357.

INSTITUT FÜR ANGEWANDTE MATHEMATIK
UNIVERSITÄT BONN
WEGELERSTRASSE 6
53115 BONN
GERMANY
E-MAIL: sturm@uni-bonn.de