COUPLING, REGULARITY AND CURVATURE

KARL-THEODOR STURM

1. Introduction

The topic of this article is the space $\mathcal{P}(M)$ of probability measures over a Euclidean or Riemannian space M. We will study the relation between dynamics on $\mathcal{P}(M)$ and stochastic processes on M. And we will try to show how curvature conditions on M determine contraction properties for the dynamics on $\mathcal{P}(M)$ and how the latter in turn determine contraction, convergence and regularity properties for the stochastic processes on M. The space $\mathcal{P}(M)$ will always be equipped with the L^θ-Wasserstein distance. Hence, contraction properties in $\mathcal{P}(M)$ mean coupling properties.

Among the most basic stochastic processes are stochastic diffusion processes and martingales. Both classes play fundamental roles in Euclidean as well as in Riemannian settings. From our point of view, in many cases it is more appropriate to investigate the dynamic on the space $\mathcal{P}(M)$ of probability measures on M than to investigate the stochastic dynamic of particles on M.

Linear diffusion equations on $\mathbb{R}^n$ or on a manifold M give rise to stochastic diffusion processes which describe the random motion of single particles or, in the same manner, the evolution of an initial distribution of particles. Nonlinear diffusion equations also lead to flows of probability meaaures. They may be interpreted as distributions of underlying systems of interacting stochastic particles. However, nonlinear diffusion equations can not be modeled by random motions of single particles.

Already half a century ago, K. Ito regarded flows of probability measures as a basic model for stochastic evolution processes. See e.g. the recent monograph [Str03] of D. Stroock on "Markov processes from K. Ito's perspective". However, at that time no appropriate tools had been available to construct and to investigate such flows. For the linear case (on Euclidean as well as on Riemannian spaces), this problem was resolved by the theory of stochastic differential equations, initiated by K. Ito at that time. This not only yields the flow of probability measures but the much more sophisticated random motion of the underlying particles.

For the nonlinear case, we will present a rather recent approach which allows to define a large class of nonlinear diffuions on M as gradient flows of probability measures w.r.t. appropriate free energy functionals on $\mathcal{P}(M)$.

An entirely different point of view is needed to understand the role of martingales on M. Whereas diffusions describe how mass spreads out in time, martingales describe the reverse procedure, namely, how one can find a center of mass for a given distribution. This leads to the concept of barycenter maps.

Let us briefly outline the role of Brownian motions and martingales in the theory of harmonic maps. For more details we refer to [Ken98]. Harmonic maps $f : M \to N$ between smooth Riemannian manifolds are critical values of the nonlinear energy functional $E(f) = \int_M ||df||^2 dm$. They may also be characterized as solutions to the corresponding Euler-Lagrange equation.

1991 *Mathematics Subject Classification.* Primary .

Key words and phrases. gradient flow, nonlinear diffusion, Bakry-Emery criterion, Ricci curvature, entropy, Wasserstein distance, barycenter, martingale, nonpositive curvature, Hadamard space, Alexandrov space, porous medium equation, fast diffusion.

Ishihara's characterization states that a map $f : M \to N$ is harmonic if and only if for each convex function $\varphi : N_0 \to \mathbb{R}$ (defined on some subset N_0 of $\varphi(M)$) the function $\varphi \circ f : M_0 \to \mathbb{R}$ (defined on $M_0 = f^{-1}(N_0)$) is subharmonic. In this sense, the minimal requirements to build up a theory of harmonic maps between singular spaces is that the domain space M is a harmonic space and that the target space is a geodesic space: on M we need to know the subharmonic functions, on N the convex functions. However, this only allows to define harmonic maps. In order to construct them one needs more structure on the domain space. One possibility is to require the domain space to be a metric measure space. This allows to construct a Dirichlet form on it. For instance, one could derive the associated process from a rescaled random walk, cf. [Stu98]. Another possibility is to assume the domain space M to be the state space of a Markov process $(X_t)_t$ and the target space N to be a metric space of nonpositive curvature (in the sense of Alexandrov). The latter allows to develop a theory of martingales on N, [Stu02], [Stu03].

However, for the following discussion let us restrict to smooth Riemannian manifolds M and N. The probabilistic characterization due to Bismut states that f is harmonic if and only if for each (stopped) Brownian motion $(X_t)_t$ on M the process $(Y_t)_t = (f(X_t))_t$ on N is a martingale. The nonlinear evolution $(t, x) \mapsto f_t(x)$ of a given map $f : M \to N$ towards a harmonic map can be described as follows: for each t and x, $f_t(x)$ is the starting point Y_0 of a martingale $(Y_s)_s$ in N with terminal value $Y_t = f(X_t)$ where $(X_s)_s$ is a Brownian motion in M with starting point $X_0 = x$. In other, more robust terms this evolution can be described as follows

$$f_t(x) = \lim_{n\to\infty} \left(P_{t/n}\right)^n f(x) \quad \text{with} \quad P_t f(x) = b(f_* p_t(x, .)).$$

Here $p_t : M \to \mathcal{P}(M)$ denotes the transition semigroup of Brownian motion on M and $b : \mathcal{P}(N) \to N$ the so-called barycenter map on N.

Regularity results for harmonic maps typically depend on lower bounds for the Ricci curvature of M and on upper bounds for the sectional curvature of N. We will see that the Ricci curvature of M is bounded from below by K if and only if

$$d_\theta^W(p_t(x, .), p_t(y, .)) \le e^{-Kt} d(x, y)$$

($\forall x, y \in M$) and that the sectional curvature of N is nonpositive if and only if

$$d(b(\mu), b(\nu)) \le d_\theta^W(\mu, \nu)$$

($\forall \mu, \nu \in \mathcal{P}_\theta(N)$), both spaces of probability measures being equipped with the L^θ-Wasserstein distance for some $\theta \in [1, \infty[$. Both properties together imply

$$\mathrm{Lip} P_t \le e^{-Kt} \mathrm{Lip} f$$

for all maps $f : M \to N$. This is a basic example of a gradient estimate for the nonlinear evolution of harmonic maps. Usually, its derivation is based on Bochner's formula (if analysts derive it) or on Bismut's formula (if probabilists do it). Here it is based on robust coupling properties for heat kernels and barycenters which allow to extend it to much more general situations (nonsymmetric, nonlocal, nonsmooth, infinite dimensional).

Let us briefly summarize the following sections. In section 2 we introduce the space $\mathcal{P}^\theta(N)$ of probability measures over a metric space (N, d), equipped with the L^θ-Wasserstein distance d_θ^W between probability measures as a natural metric.

Sections 3 and 4 are devoted to contraction properties for maps $b : \mathcal{P}^\theta(N) \to N$, called barycenter maps. The crucial ingredients will be upper curvature bounds on N in the sense of Alexandrov which generalize upper bounds for the sectional curvature.

In section 5 we discuss contraction properties for the heat semigroup on a Riemannian manifold M, regarded as a family of maps $p_t : M \to \mathcal{P}^\theta(M)$. Here lower bounds on the Ricci curvature of M will play the essential role.

Finally, in section 6 we study nonlinear diffusions on $\mathbb{R}^n$ or on a Riemannian manifold M as gradient flows of appropriate free energy functionals on $\mathcal{P}^\theta(M)$. Contraction properties for these nonlinear diffusions will be derived from convexity properties of the free energy functional. In particular, we present extensions of the Bakry-Emery criterion to nonlinear equations.

2. The Space of Probability Measures

Let (N,d) be a metric space and let $\mathcal{P}(N)$ denote the set of all probability measures p on N (equipped with its Borel σ-algebra $\mathcal{B}(N)$) with separable support $\text{supp}(p) \subset N$. For $1 \le \theta < \infty$, $\mathcal{P}^\theta(N)$ will denote the set of $p \in \mathcal{P}(N)$ with $\int d^\theta(x,y)p(dy) < \infty$ for some (hence all) $x \in N$, and $\mathcal{P}^\infty(N)$ will denote the set of all $p \in \mathcal{P}(N)$ with bounded support. Obviously, $\mathcal{P}^\infty(N) \subset \mathcal{P}^\theta(N) \subset \mathcal{P}^1(N)$.

Given $p,\, q \in \mathcal{P}(N)$ we say that $\mu \in \mathcal{P}(N^2)$ is a *coupling* of p and q iff its *marginals* are p and q, that is, iff $\forall A \in \mathcal{B}(N)$

$$\mu(A \times N) = p(A) \quad \text{and} \quad \mu(N \times A) = q(A). \tag{2.1}$$

One such coupling μ is the product measure $p \otimes q$. Couplings μ of p and q are also called *transportation plans* from p to q. If p is the distribution of points at which a good is produced and q is the distribution of points where it is consumed, then each coupling μ of p and q gives a plan how to transport the production to the consumer and $d^W(p,q)$ describes the smallest cost of such a transportation process.

Definition 2.1. For $\theta \in [1,\infty[$, we define the L^θ-*Wasserstein distance* distance d^W_θ on $\mathcal{P}^\theta(N)$ by

$$d^W_\theta(p,q) = \inf \left\{ \int_N \int_N d^\theta(x,y)\mu(dxdy) : \mu \in \mathcal{P}(N^2) \text{ is coupling of } p \text{ and } q \right\}^{1/\theta}.$$

In probabilistic language,

$$d^W_\theta(p,q) = \inf \Big[\mathbb{E}\, d^\theta(X,Y)\Big]^{1/\theta},$$

where the infimum is over all probability spaces $(\Omega, \mathcal{A}, \mathbb{P})$ and all measurable maps $X : \Omega \to N$ and $Y : \Omega \to N$ with separable ranges and distributions $\mathbb{P}_X = p$ and $\mathbb{P}_Y = q$.

See e.g. [Dud89], [RaRu98], [Stu01], [Vil03].

Remark 2.2. If (N,d) is complete then so is $(\mathcal{P}^\theta(N), d^W_\theta)$. If (N,d) is a geodesic space, then so is $(\mathcal{P}^\theta(N), d^W_\theta)$. If (N,d) is separable then so is $(\mathcal{P}^\theta(N), d^W_\theta)$.

The case $\theta = 2$ is of particular importance. In this case, even curvature bounds carry over from N to $\mathcal{P}^\theta(N)$.

Definition 2.3. We say that a metric space (N,d) has nonnegative curvature iff

$$\frac{1}{2k}\sum_{i,j=1}^k d^2(y_i,y_j) \le \sum_{i=1}^k d^2(z,y_i)$$

for all $k \in \mathbb{N}$ and all $z, y_1, \dots, y_k \in N$.

Obviously, the latter is equivalent to

$$\frac{1}{2}\int_N\int_N d^2(x,y)\,q(dx)\,q(dy) \le \int_N d^2(z,x)\,q(dx) \tag{2.2}$$

for all $z \in N$ and all probability measures $q \in \mathcal{P}^2(N)$.

The probabilistic interpretation of this property is as follows: N has nonnegative curvature if and only if each pair of "randomly chosen points" X, Y in N (i.e. each pair of N-valued iid random variables) is seen from each point $z \in N$ "in average" under an angle $\angle(z; X, Y) \leq 90°$. More precisely, $\mathbb{E} \cos \angle(z; X, Y) \geq \cos 90°$.

One easily verifies that (N, d) has nonnegative curvature if and only if its metric completion $(\overline{N}, \overline{d})$ has nonnegative curvature.

If (N, d) is a geodesic metric space, then our definition of nonnegative curvature can be shown to be equivalent to nonnegative curvature in the sense of A. D. Alexandrov [Stu99]. Hence, in particular, for Riemannian manifolds it is equivalent to nonnegative sectional curvature. The property "nonnegative curvature" carries over from a metric space (N, d) to the space of probability measures over this spaces [Ott01], [Stu01].

Proposition 2.4. *Let (N, d) be a metric space. The space $(\mathcal{P}^2(N), d_2^W)$ of probability measures equipped with the L^2-Wasserstein distance has nonnegative curvature if and only if the underlying space (N, d) has nonnegative curvature.*

3. Probability Measures on Metric Spaces of Nonpositive Curvature

The next two sections are devoted to two generalizations of the class of Cartan-Hadamard manifolds. The first generalization is the class of metric spaces with nonpositive curvature with nonpositive curvature in the sense of Alexandrov. The second one (to be presented in the next section) will be the class of metric spaces which admit a contracting barycenter map. We summarize some of the basic results and refer to [Stu02] and [Stu03] for more details and further references.

Definition 3.1. A metric space (N, d) is called *global NPC space* if it is complete and if for all $k \in \mathbb{N}$ and all $y_1, \ldots, y_k \in N$

$$\frac{1}{2k} \sum_{i,j=1}^{k} d^2(y_i, y_j) \geq \inf_{z \in N} \sum_{i=1}^{k} d^2(z, y_i).$$

or, equivalently, if for all probability measures $q \in \mathcal{P}^2(N)$

$$\frac{1}{2} \int_N \int_N d^2(x, y)\, q(dy)\, q(dx) \geq \inf_{z \in N} \int_N d^2(z, x)\, q(dx).$$

Here "NPC" stands for "**n**on**p**ositive **c**urvature" in the sense of A. D. Alexandrov. Global NPC spaces are also called *Hadamard spaces.* If (N, d) is a global NPC space then it is a geodesic space. Even more, for any pair of points $x_0, x_1 \in N$ there exists a unique geodesic $x : [0, 1] \to N$ connecting them. For $t \in [0, 1]$ the intermediate points x_t depend continuously on the endpoints x_0, x_1. Finally, for any $z \in N$

$$d^2(z, x_t) \leq (1-t) d^2(z, x_0) + t d^2(z, x_1) - t(1-t) d^2(x_0, x_1). \tag{3.1}$$

Our definition of nonpositive curvature is completely analogous to our definition of nonnegative curvature in Definition 2.3. Note, however, that there are two substantial differences between upper and lower curvature bounds. Firstly, if a complete metric space has globally curvature ≤ 0 then it is necessarily a geodesic space. This is not the case for complete metric spaces with (global) curvature ≥ 0. Secondly, if a complete geodesic space has "locally" curvature ≥ 0 then it has already "globally" curvature ≥ 0. The analogous statement is not true for complete geodesic spaces with local/global curvature ≤ 0.

Example 3.2. Examples of global NPC spaces are

- complete, simply connected Riemannian manifolds with nonpositive sectional curvature;
- trees and, more generally, Euclidean Bruhat-Tits buildings;
- Hilbert spaces;
- L^2-spaces of maps into such spaces;

- Finite or infinite (weighted) products of such spaces;
- Gromov-Hausdorff limits of such spaces.

See e.g. [Ale51], [BGS85], [Bal95], [BrHa99], [BBI01], [EeFu01] [Gro81/99], [Jos94], [Jos97], [KoSc93].

Proposition 3.3. *Let (N, d) be a global NPC space and fix $y \in N$. For each $q \in \mathcal{P}^1(N)$ there exists a unique point $z \in N$ which minimizes the uniformly convex, continuous function $z \mapsto \int_N [d^2(z,x) - d^2(y,x)]q(dx)$. This point is independent of y; it is called* barycenter *(or, more precisely, canonical barycenter or d^2-barycenter) of q and denoted by*

$$b(q) = \operatorname*{argmin}_{z \in N} \int_N [d^2(z,x) - d^2(y,x)]\, q(dx).$$

Moreover,

$$\int_N [d^2(z,x) - d^2(b(q),x)]q(dx) \geq d^2(z, b(q)). \tag{3.2}$$

If $q \in \mathcal{P}^2(N)$ then $b(q) = \operatorname{argmin}_{z \in N} \int_N d^2(z,x)q(dx)$.

Inequality (3.2) is called *variance inequality.* If in addition to nonpositive curvature we also assume that the curvature is bounded from below, say by $-K^2$, and that the space is geodesically complete then the following *reverse variance inequality* holds [Stu03]: For each $q \in \mathcal{P}^2(N)$ and for each $z \in N$

$$\int \left[d^2(z,x) - d^2(z,b(q)) - d^2(b(q),x)\right] q(dx) \leq \frac{2K^2}{3} \int \left[d^4(z,b(q)) + d^4(b(q),x)\right] q(dx).$$

For a L^1-random variable $X : \Omega \to N$ we define its *expectation* by

$$\mathbb{E}\, X := b(\mathbb{P}_X) = \operatorname*{argmin}_{z \in N} \mathbb{E}\, [d^2(z,X) - d^2(y,X)].$$

That is, $\mathbb{E}\, X$ is the unique minimizer of the function $z \mapsto \mathbb{E}\, [d^2(z,X) - d^2(y,X)]$ on N (for each fixed $y \in N$). The variance inequality then reads as follows:

$$\mathbb{E}\left[d^2(z,X) - d^2(\mathbb{E}\, X, X)\right] \geq d^2(z, \mathbb{E}\, X)$$

for all $z \in N$. In the classical case $N = \mathbb{R}$, the corresponding *equality* should be well known after the first lessons in probability theory.

Our approach to barycenters and expectations is based on the classical point of view of [Gau1809]. 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 [Car28], [Fre48], [Kar77], and many others, under the name of barycenter, center of mass or center of gravity. Iterations of barycenters on Riemannian manifolds were used by [Ken90], [EmMo91] and [Pic94]. [Jos94] applied these concepts on global NPC spaces.

For other probabilistic approaches, see [Dos49], [Her91], [ESH99].

Another natural way to define the ”expectations” $\mathbb{E}\, Y$ of a random variable Y is to use (generalizations of) the law of large numbers. This requires to give a meaning to $\frac{1}{n}\sum_{i=1}^{n} Y_i$. Our definition below 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 towards the expectation defined as minimizer of the L^2-distance [Stu02].

Theorem 3.4 (Law of Large Numbers)**.** *Let $(Y_i)_{i\in\mathbb{N}}$ be a sequence of independent, identically distributed bounded random variables $Y_i : \Omega \to N$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in a global NPC space (N, d). Define their mean values $S_n : \Omega \to N$ by induction on $n \in \mathbb{N}$ as follows:*

$$S_1(\omega) := Y_1(\omega) \quad \text{and} \quad S_n(\omega) := \left(1 - \frac{1}{n}\right) S_{n-1}(\omega) + \frac{1}{n}\, Y_n(\omega),$$

where the RHS should denote the point $\gamma_{1/n}$ on the geodesic $\gamma : [0,1] \to N$ connecting $\gamma_0 = S_{n-1}(\omega)$ and $\gamma_1 = Y_n(\omega)$.

Then for $\mathbb{P}$-almost every $\omega \in \Omega$

$$S_n(\omega) \to \mathbb{E}\, Y_1 \qquad \text{for } n \to \infty$$

("strong law of large numbers").

Remark 3.5. (i) In strong contrast to the linear case, the mean value S_n will in general strongly depend on permutations of the iid variables $Y_i, i = 1\ldots, n$. The distribution $\mathbb{P}_{S_n}$ is of course invariant under such permutations. But even $\mathbb{E}\, S_n$ in general depends on $n \in \mathbb{N}$. The law of large numbers only yields that $\mathbb{E}\, S_1 = \lim_{n\to\infty} \mathbb{E}\, S_n$.

(ii) It might seem more natural to define the mean value of the random variables $Y_1, \ldots, Y_n$ as the barycenter of these points, more precisely, as the barycenter of the uniform distribution on these points, i.e.

$$\overline{S}_n(\omega) := b\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{Y_i(\omega)}\right).$$

In this case we also obtain a law of large numbers. Indeed, it is much easier to derive (and it holds for more or less arbitrary choices of $b(.)$), see Proposition 4.9. However, it is also of much less interest: we will obtain convergence of $\overline{S}_n(\omega)$ towards $b(\mathbb{P}_{Y_1})$, the barycenter of the distribution of Y_1, but to define $\overline{S}_n$ we already have to use $b(.)$.

(iii) Of course there are many other ways to define a mean value $\tilde{S}_n$ of the random variables $Y_1, \ldots, Y_n$ which do not depend on the a priori knowledge of $b(.)$. And indeed for many of these choices one can prove that $\tilde{S}_n$ converges almost surely to a point $\tilde{b}$ (which only depends on the distribution of Y_1). For instance, define $S_{n,1} := Y_n$ and recursively $S_{n,k+1}$ to be the midpoint of $S_{2n-1,k}$ and $S_{2n,k}$. Then $\tilde{S}_n(\omega) := S_{1,n}(\omega)$ converges for a.e. ω as $n \to \infty$ towards a point $\tilde{b} = \tilde{b}(\mathbb{P}_{Y_1})$. (Note that in the flat case, $S_{1,n} = 2^{-n}\sum_{i=1}^{2^n} Y_i$.) Another example is given by the mean value in the sense of [ESH99] which will be described as Example 4.8.

However, no choice of $\tilde{S}_n$ other than S_n is known to the author where one obtains convergence towards a point which can be characterized "extrinsically", like in our case as the minimizer of the function $z \mapsto \mathbb{E}d^2(z, Y_1)$.

The Law of Large Numbers yields a simple proof of Jensen's inequality. The key ingredient is the inequality

$$\varphi(S_n) \le \frac{1}{n}\sum_{i=1}^{n} \varphi(Y_i)$$

which holds for any convex function $\varphi : N \to \mathbb{R}$.

Theorem 3.6 (Jensen's inequality)**.** *For any global NPC space (N, d), any lower semicontinuous convex function $\varphi : N \to \mathbb{R}$ and any $q \in \mathcal{P}^1(N)$*

$$\varphi(b(q)) \le \int_N \varphi(x) q(dx),$$

provided the RHS is well-defined.

The above RHS is well-defined if either $\int \varphi^+\, dq < \infty$ or $\int \varphi^-\, dq < \infty$. In particular, it is well-defined if φ is Lipschitz continuous. Applying Jensen's inequality to the convex function $(x, y) \mapsto d(x, y)$ on the global NPC space $N \times N$ yields

Theorem 3.7 (Fundamental Contraction Property)**.** *For all $\theta \in [1,\infty[$ and all $p,q \in \mathcal{P}^\theta(N)$:*

$$d(b(p),b(q)) \le d_1^W(p,q) \le d_\theta^W(p,q). \tag{3.3}$$

4. Barycenters

In the previous section, we saw that nonpositive curvature in the sense of Alexandrov implies the existence of a canonical barycenter map. It turns out that many results require nonpositive curvature only because they rely on contraction properties of this map. The existence of such a contracting map itself may be regarded as a far reaching generalization of nonpositive curvature.

Let us fix a complete metric space (N,d) and a number $\theta \in [1,\infty[$.

Definition 4.1. A *L^θ-barycenter contraction* is a map $b : \mathcal{P}^\theta(N) \to N$ such that

- $b(\delta_x) = x$ for all $x \in N$;
- $d(b(p),b(q)) \le d_\theta^W(p,q)$ for all $p,q \in \mathcal{P}^\theta(N)$.

Obviously, a L^θ-barycenter contraction is a $L^{\theta'}$-barycenter contraction for each $\theta' \ge \theta$,

Example 4.2. For each global NPC space the canonical barycenter yields a L^θ-barycenter contraction ($\forall\theta$).

Actually, also partly the converse holds. If there exists a L^θ-barycenter contraction on (N,d) then (N,d) is a geodesic space: For each pair of points $x_0,x_1 \in N$ we can define a geodesic $t \mapsto x_t$ connecting x_0 and x_1 by $x_t := b((1-t)\delta_{x_0}+t\delta_{x_1})$. Given any four points $x_0,x_1,y_0,y_1 \in N$, the function $t \mapsto d(x_t,y_t)$ is convex. In particular, the geodesic $t \mapsto x_t$ depends continuously on x_0 and x_1. However, it is not necessarily the only geodesic connecting x_0 and x_1.

If geodesics in N are unique then the existence of a L^θ-barycenter contraction implies that $d : N\times N \to \mathbb{R}$ is convex. Thus N has globally "nonpositive curvature" in the sense of Busemann.

Corollary 4.3. *Let N be a complete, simply connected Riemannian manifold and let d be a Riemannian distance. Then (N,d) admits a L^θ-barycenter contraction b if and only if N has nonpositive sectional curvature.*

Indeed, if (N,d) admits a L^θ-barycenter contraction then so does (N_0,d) for each closed convex $N_0 \subset N$. Hence, geodesics in N_0 are unique and thus $t \mapsto d(\gamma_t,\zeta_t)$ is convex for any pair of geodesics γ and ζ in N_0. This implies that N has nonpositive curvature.

Example 4.4. Let $(N,\|.\|)$ be a (real or complex) Banach space and put $d(x,y) := \|x-y\|$. Then $\mathcal{P}^\theta(N)$ is the set of Radon measures p on N satisfying $\int_N \|x\|^\theta\, p(dx) < \infty$. For each $p \in \mathcal{P}^\theta(N)$, the identity $x \mapsto x$ on N is *Bochner integrable* and

$$b(p) := \int_N x\, p(dx)$$

defines a barycenter contraction on (N,d).

Example 4.5. Let I be a countable set and for each $i \in I$, let (N_i,d_i) be a complete metric space with L^1-barycenter contraction b_i and "base" point $o_i \in N_i$. Given $\eta \in [1,\infty]$, define a complete metric space (N,d) with base point $o = (o_i)_{i\in I}$ by

$$N := \left\{ x = (x_i)_{i\in I} \in \bigotimes_{i\in I} N_i :\ d(x,o) < \infty \right\}, \quad d(x,y) := \left[\sum_{i\in I} d_i^\eta(x_i,y_i) \right]^{\frac{1}{\eta}}$$

provided $\eta < \infty$ or by $d(x,y) = \sup_{i\in I} d_i(x_i,y_i)$ if $\eta = \infty$. One can define a barycenter contraction b on $\mathcal{P}^1(N)$ by

$$b(p) := (b_i(p_i))_{i\in I}$$

where $p_i \in \mathcal{P}^1(N_i)$ with $p_i : A \mapsto p(\{x = (x_j)_{j \in I} \in N : x_i \in A\})$ denotes the projection of $p \in \mathcal{P}^1(N)$ onto the i-th factor of N.

For instance, this applies to $N = \mathbb{R}^n, n \geq 2$ with the usual notion of barycenter but with "unusual" metric $d(x, y) = \sup\{|x_i - y_i| : i = 1, .., n\}$. In this case, geodesics are not unique, e.g. each curve $t \mapsto (t, \varphi_2(t), ..., \varphi_n(t))$ with $\varphi \in \mathcal{C}^1(\mathbb{R}), \varphi_i(0) = \varphi_i(1) = 0$ and $|\varphi_i'| \leq 1$ is a geodesic connecting $(0, 0, ..., 0)$ and $(1, 0, ..., 0)$.

Remark 4.6. Each barycenter map b on a complete metric space (N, d) gives rise to a whole family of barycenter maps b_n, $n \in \mathbb{N}$ (which in general do not coincide with b, see Example below). Let b be a L^θ-barycenter contraction and $\Phi : N \times N \to N$ be the "midpoint map" induced by b, i.e. $\Phi(x, y) = b(\frac{1}{2}\delta_x + \frac{1}{2}\delta_y)$. Define a map $\Xi : \mathcal{P}^1(N) \to \mathcal{P}^1(N)$ by

$$\Xi(q) := \Phi_*(q \otimes q).$$

Then Ξ is a contraction with respect to d_θ^W. Thus for each $n \in \mathbb{N}$

$$b_n(q) := b(\Xi^n(q))$$

defines a contracting barycenter map $b_n : \mathcal{P}^\theta(N) \to N$.

Example 4.7. Define the *tripod* by gluing together 3 copies of $\mathbb{R}_+$ at their origins, i.e.

$$N = \{(i, r) : i \in \{1, 2, 3\}, r \in \mathbb{R}_+\}/\sim \quad \text{where } (i, r) \sim (j, s) :\Leftrightarrow r = s = 0.$$

It can be realized as the subset $\{r \cdot \exp(\frac{l}{3}2\pi i) \in \mathbb{C} : r \in \mathbb{R}_+, l \in \{1, 2, 3\}\}$ of the complex plane, however, equipped with the (non-Euclidean!) intrinsic metric

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

Then (N, d) is a complete metric space of globally nonpositive curvature and according to Example 3.2 there exists a canonical barycenter map b. Derive from that the barycenter map $b_1 = b(\Xi(.))$ as above. Then the maps b and b_1 do not coincide. Indeed, choose $q = \frac{1}{2}\delta_{(1,1)} + \frac{1}{4}\delta_{(2,1)} + \frac{1}{4}\delta_{(3,1)}$. Then $\Xi(q) = \frac{1}{4}\delta_{(1,1)} + \frac{1}{16}\delta_{(2,1)} + \frac{1}{16}\delta_{(3,1)} + \frac{5}{8}\delta_o$. Hence, $b(q) = (1, 0)$ and $b_1(q) = b(\Xi(q)) = (1, \frac{1}{8})$.

Example 4.8 (Barycenter Map of Es-Sahib & Heinich [ESH99])**.** Let (N, d) be a locally compact complete separable metric space with *negative curvature in the sense of Busemann* or a global NPC space in the sense of Alexandrov (see next section). Then one can define recursively for each $n \in \mathbb{N}$ a unique map $\beta_n : N^n \to N$ satisfying

- $\beta_n(x_1, \ldots, x_1) = x_1$
- $d(\beta_n(x_1, \ldots, x_n), \beta_n(y_1, \ldots, y_n)) \leq \frac{1}{n}\sum_{i=1}^n d(x_i, y_i)$
- $\beta_n(x_1, \ldots, x_n) = \beta_n(\check{x}_1, \ldots, \check{x}_n)$ where $\check{x}_i := \beta_{n-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.

This map is symmetric and satisfies $d(z, \beta_n(x_1, \ldots, x_n)) \leq \frac{1}{n}\sum_{i=1}^n d(z, x_i)$ for all $z \in N$. Given any $p \in \mathcal{P}^1(N)$ let $(Y_i)_i$ be an independent sequence of maps $Y_i : \Omega \to N$ on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with distribution $\mathbb{P}_{Y_i} = p$ and define $\tilde{S}_n(\omega) := b_n((Y_1(\omega), \ldots, Y_n(\omega))$. Then there exists a point $\beta(p) \in N$ such that $\tilde{S}_n(\omega) \to \beta(p)$ for $\mathbb{P}$-a.e. ω and $n \to \infty$. The map $\beta : \mathcal{P}^1 \to N$ is easily seen to be a L^1-contracting barycenter map. Note, however, that in general, $\beta\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right) \neq \beta_n(x_1, \ldots, x_n)$.

Moreover, we emphasize that on non-flat Riemannian manifolds as well as on trees this barycenter map β is different from the canonical one, defined via minimizing the L^2-distance (see previous section). For instance, let (N, d) be the tripod and let $p = \frac{1}{2}\delta_{(1,1)} + \frac{1}{4}\delta_{(2,1)} + \frac{1}{4}\delta_{(3,1)}$. Then with the latter choice $b(p) = (1, 0)$, whereas an easy calculation shows that the previous choice yields $\beta(p) = (1, 1/6)$.

Proposition 4.9 (Empirical Law of Large Numbers)**.** *Let (N,d) be a complete metric space with a contracting barycenter map $b:\mathcal{P}^\theta(N)\to N$ and fix $p\in\mathcal{P}^\infty(N)$. Moreover, let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space and $(X_i)_{i\in\mathbb{N}}$ be an independent sequence of measurable maps $X_i:\Omega\to N$ with identical distribution $\mathbb{P}_{X_i}=p$. Define the "barycentric mean value" $s_n:\Omega\to N$ by $s_n(\omega):=b\left(\frac{1}{n}\sum_{i=1}^n\delta_{X_i(\omega)}\right)$. Then for $\mathbb{P}$-almost every $\omega\in\Omega$*

$$s_n(\omega)\longrightarrow\beta(p)\qquad \text{as } n\to\infty.$$

5. Transport Inequalities and Gradient Estimates

In previous sections, we studied contraction properties of the canonical barycenter maps $b:\mathcal{P}(N)\to N$ on spaces N with some kind of upper curvature bounds (e.g. nonpositive curvature in the sense of Alexandrov). Now we will be concerned with the reverse situation. We will study contraction properties of "canonical" maps $p_t:M\to\mathcal{P}(M)$ on spaces M with appropriate lower curvature bounds. To be more specific, we will restrict ourselves (mostly) to Riemannian manifolds and we will choose $(p_t)_t$ to be the heat semigroup. The appropriate curvature bounds will turn out to be lower bounds for the Ricci curvature.

We will present various equivalences between transportation inequalities (for volume measures, heat kernels, Brownian motions), gradient estimates for the heat semigroup and lower bounds for the Ricci curvature, These are joint results with Max-K. von Renesse. Details can be found in [vRSt03].

In the sequel, (M,g) always is assumed to be a complete smooth Riemannian manifold with dimension n, Riemannian distance $d(x,y)$ and Riemannian volume $m(dx)$. Here and henceforth, $p_t(x,y)$ always denotes the heat kernel on M, i.e. the minimal positive fundamental solution to the heat equation $(\Delta-\frac{\partial}{\partial t})p_t(x,y)=0$. It is smooth in (t,x,y), symmetric in (x,y) and satisfies $\int_M p_t(x,y)m(dy)\le 1$. Hence, it defines a subprobability measure $p_t(x,dy):=p_t(x,y)m(dy)$ as well as operators $p_t:\mathcal{C}_c^\infty(M)\to\mathcal{C}^\infty(M)$ and $p_t:L^2(M)\to L^2(M)$ which are all denoted by the same symbol. Given $\mu\in\mathcal{P}^\theta(M)$ and $t>0$ we define a new measure $\mu p_t\in\mathcal{P}^\theta(M)$ by $\mu p_t(A)=\int_A\int_M p_t(x,y)\mu(dx)m(dy)$.

Brownian motion on M is by definition the Markov process with generator $\frac{1}{2}\Delta$. Thus its transition (sub-)probabilities are given by $p_{t/2}$.

If the Ricci curvature of the underlying manifold M is bounded from below then all the $p_t(x,.)$ are probability measures. If the latter holds true we say that the heat kernel and the associated Brownian motion are conservative. It means that the Brownian motion has infinite lifetime.

Our first main result in thsi section deals with robust versions of gradient estimates.

Theorem 5.1. *For any complete smooth Riemannian manifold M and any $K\in\mathbb{R}$ the following properties are equivalent:*

(i): $\mathrm{Ric}(M)\ge K$,
which always should be read as: $\mathrm{Ric}_x(v,v)\ge K|v|^2$ *for all* $x\in M, v\in T_xM$.

(ii): *For all $f\in C_c^\infty(M)$, all $x\in M$ and all $t>0$*

$$|\nabla p_tf|(x)\le e^{-Kt}\,p_t|\nabla f|(x).$$

(iii): *For all $f\in\mathcal{C}_c^\infty(M)$ and all $t>0$*

$$\|\nabla p_tf\|_\infty\le e^{-Kt}\,\|\nabla f\|_\infty\,.$$

(iv): *For all bounded $f\in\mathcal{C}^{\mathrm{Lip}}(M)$ and all $t>0$*

$$\mathrm{Lip}(p_tf)\le e^{-Kt}\,\mathrm{Lip}(f).$$

The equivalence of (i) and (ii), perhaps, is one of the most famous general results which relate heat kernels with Ricci curvature. It is due to D. Bakry & M. Emery [BaEm84], see also [ABC00] and references therein. Property (ii) is successfully used in various applications as a replacement (or definition) of lower Ricci curvature bounds for symmetric Markov semigroups on general state spaces. Our result states that (ii) can be weakened in two respects:
– one can replace the pointwise estimate by an estimate between L^∞-norms;
– one can drop the p_t on the RHS.
Besides being formally weaker than (ii) one other advantage of (iii) is that it is an explicit statement on the smoothing effect of p_t whereas (ii) is implicit (since p_t appears on both sides).

As an easy corollary to the equivalence of the statements (ii) and (iii) one may deduce the well known fact that (ii) is equivalent to the assertion that for all f, x and t as above

$$|\nabla p_t f|(x) \le e^{-Kt} \left[p_t(|\nabla f|^2)(x)\right]^{1/2}.$$

Property (iv) may be considered as a replacement (or as one possible definition) for lower Ricci curvature bounds for Markov semigroups on metric spaces. For several non-classical examples (including nonlocal generators as well as infinite dimensional or singular finite dimensional state spaces) we refer to [Stu03], [DaRo02] and [vRe03]. This property turned out to be the key ingredient to prove Lipschitz continuity for harmonic maps between metric spaces in [Stu03].

According to the Kantorovich-Rubinstein duality, property (iv) is equivalent to a contraction property for the heat kernels in terms of the L^1-Wasserstein distance d_1^W. Actually, however, much more can be proven:
– one obtains contraction in d_θ^W for each $\theta \in [1,\infty]$ and for any initial data;
– one obtains pathwise contraction for Brownian trajectories.

Corollary 5.2. *For any smooth complete Riemannian manifold M and any $K \in \mathbb{R}$ the following properties are equivalent:*

(i): $\mathrm{Ric}(M) \ge K$.

(v): *For all $x, y \in M$ and all $t > 0$ there exists $\theta \in [1,\infty]$ with*

$$d_\theta^W(p_t(x,.), p_t(y,.)) \le e^{-Kt} \cdot d(x,y).$$

(vi): *For all $\theta \in [1,\infty]$, all $\mu, \nu \in \mathcal{P}^\theta(M)$ and all $t > 0$:*

$$d_\theta^W(\mu p_t, \nu p_t) \le e^{-Kt} \cdot d_\theta^W(\mu,\nu).$$

(vii): *For all x_1, $x_2 \in M$ there exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and two conservative Brownian motions $(X_1(t))_{t\ge 0}$ and $(X_2(t))_{t\ge 0}$ defined on it, with values in M and starting in x_1 and x_2, respectively, such that for all $t > 0$*

$$\mathbb{E}\left[d(X_1(t), X_2(t))\right] \le e^{-Kt/2} \cdot d(x_1, x_2).$$

(viii): *There exists a conservative Markov process $(\Omega, \mathcal{A}, \mathbb{P}^x, X(t))_{x\in M\times M, t\ge 0}$ with values in $M \times M$ such that the coordinate processes $(X_1(t))_{t\ge 0}$ and $(X_2(t))_{t\ge 0}$ are Brownian motions on M and such that for all $x = (x_1, x_2) \in M \times M$ and all $t > 0$*

$$d(X_1(t), X_2(t)) \le e^{-Kt/2} \cdot d(x_1, x_2) \qquad \mathbb{P}^x\text{-}a.s.$$

Note that each of the statements (v) and (vi) implicitly includes the conservativity of the heat kernel. Indeed, the finiteness of the Wasserstein distance implies that the measures under consideration must have the same total mass. Thus $p_t(x, M)$ is constant in x, hence also constant in t and therefore equal to 1.

The interpretation of these results is as follows: if we put mass distributions μ and ν on M and if they spread out according to the heat equation then the lower bound for the Ricci curvature of M controls how fast the distances between these distributions may expand (or have to decay) in time.

The second main result in this section deals with transportation inequalities for uniform distributions on spheres and analogous inequalities for uniform distributions on balls. Here the lower Ricci bound is characterized as a control for the increase of the distances if we replace Dirac masses δ_x and δ_y by uniform distributions $\sigma_{r,x}$ and $\sigma_{r,y}$ on spheres around x and y, resp. or if we replace them by uniform distributions $m_{r,x}$ and $m_{r,y}$ on balls around x and y, resp.

Theorem 5.3. *For any smooth compact Riemannian manifold M and any $K \in \mathbb{R}$ the following properties are equivalent:*

(i): $\mathrm{Ric}(M) \geq K$.

(ix): *The normalized surface measure on spheres of radius $\sqrt{2n}\,r$*

$$\sigma_{r,x}(A) := \frac{\mathcal{H}^{n-1}(A \cap \partial B_{\sqrt{2n}r}(x))}{\mathcal{H}^{n-1}(\partial B_{\sqrt{2n}r}(x))}, \qquad A \in \mathcal{B}(M)$$

satisfies the asymptotic estimate

$$d_1^W(\sigma_{r,x}, \sigma_{r,y}) \leq \Big(1 - Kr^2 + o(r^2)\Big) \cdot d(x,y) \tag{5.1}$$

where the error term is uniform w.r.t. $x, y \in M$.

(x): *The normalized Riemannian uniform distribution on balls of radius $\sqrt{2(n+2)}\,r$*

$$m_{r,x}(A) := \frac{m(A \cap B_{\sqrt{2(n+2)}r}(x))}{m(B_{\sqrt{2(n+2)}r}(x))}, \qquad A \in \mathcal{B}(M)$$

satisfies the asymptotic estimate

$$d_1^W(m_{r,x}, m_{r,y}) \leq \Big(1 - Kr^2 + o(r^2)\Big) \cdot d(x,y) \tag{5.2}$$

where the error term is uniform w.r.t. $x, y \in M$.

The heat kernel on a Riemannian manifold is a fundamental object for analysis, geometry and stochastics. Many properties and precise estimates are known. In most of these results, lower bounds on the Ricci curvature of the underlying manifold play a crucial role. However, for more general spaces, like e.g. metric measure spaces, there is neither a notion of Ricci curvature nor a common notion of bounds for the Ricci curvature (comparable for instance to Alexandrov's notion of bounds for the sectional curvature for metric spaces).

The advantage of the above characterization of Ricci curvature is that it depends only on the basic, robust data: measure and metric. It does not require any heat kernel, any Laplacian or any Brownian motion. It might be used as a guideline in much more general situations.

For instance, let (M, d) be an arbitrary separable metric space equipped with a measure m on its Borel σ-field and assume that (5.2) holds true (with some number $K \in \mathbb{R}$). Define an operator m_r acting on bounded measurable functions by $m_r f(x) = \int_M f(y)\, m_{r,x}(dy)$. Then by the Arzela-Ascoli theorem there exists a sequence $(l_j)_j \subset \mathbb{N}$ such that

$$p_t f := \lim_{j\to\infty} \Big(m_{\sqrt{t/l_j}}\Big)^{l_j} f$$

exists (as a uniform limit) for all bounded $f \in \mathcal{C}^{\mathrm{Lip}}(M)$ and it defines a Markov semigroup on M satisfying

$$\mathrm{Lip}(p_t f) \leq e^{-Kt} \mathrm{Lip}(f).$$

For Riemannian manifolds, the invariance principle for Brownian motions implies that this semigroup $(p_t)_t$ is just the usual heat semigroup and thus (x) obviously implies (iv). Analogously, (ix)$\Longrightarrow$ (iv). The converse implications are rather involved and required detailed estimates for the transportation costs. The basic ingredient, however, is an elementary quadrilateral estimate for geodesic parallel transports (and actually this easily explains the final asymptotic formula).

6. Gradient Flows on Metric Spaces and Nonlinear Diffusions

This section is devoted to contraction properties of nonlinear diffusions on $\mathbb{R}^n$ or on a Riemannian manifold M. Following [Ott01] and [OtVi00], we regard them as gradient flows of appropriate free energy functionals S on $\mathcal{P}^\theta(M)$. Contraction properties for these nonlinear diffusions will be derived from convexity properties of the free energy functional. In particular, we present extensions of the Bakry-Emery criterion to nonlinear equations. For details, proofs and further references we reer to [Stu04].

Given an arbitrary geodesic space (N, d_N), a number $K \in \mathbb{R}$ and a function $S : N \to [-\infty, +\infty]$ we say that S is K*-convex* iff for each (constant speed, as usual) geodesic $\gamma : [0,1] \to N$ with $S(\gamma_0) < \infty$ and $S(\gamma_1) < \infty$) and for each $t \in [0,1]$:

$$S(\gamma_t) \leq (1-t)\, S(\gamma_0) + t\, S(\gamma_1) - \frac{K}{2} t(1-t)\, d_N^2(\gamma_0, \gamma_1). \tag{6.1}$$

If S is lower semicontinuous, then it suffices to verify this for all geodesics γ and $t = \frac{1}{2}$.

K-convexity is a local property. The above inequality (6.1) holds for a given function S and a given geodesic $\gamma : [0,1] \to N$ provided there exists a partition $0 = t_0 < t_1 < \ldots < t_{n+1} = 1$ such that for each $i = 1, \ldots, n$ the geodesic $\gamma : [t_{i-1}, t_{i+1}] \to N$ satisfies (after suitable reparametrization) inequality (6.1).

A function S is K-convex if and only if for each geodesic $\gamma : [0,1] \to N$ with $S(\gamma_0) < \infty$ and $S(\gamma_1) < \infty$ one has $S(\gamma_t) < \infty$ (for all $t \in [0,1]$) and

$$\liminf_{t \to 0} \frac{1}{t^2} \cdot [S(\gamma_{2t}) - 2S(\gamma_t) + S(\gamma_0)] \geq K \cdot d_N^2(\gamma_0, \gamma_1).$$

Example 6.1. A smooth function S on a Riemannian manifold (N, d) is K-convex if and only if

$$\mathrm{Hess}\, S \geq K.$$

Given a function S on a geodesic space (N, d), we say that a map $\sigma : \mathbb{R}_+ \times N \to N$, $(t, x) \mapsto \sigma_t(x)$ is a *gradient flow for S* iff for each $x \in N$, $t \mapsto \sigma(t, x)$ is a curve in N starting in x with $|\partial_t \sigma(t, x)| = -\partial_\sigma S(\sigma)(t, x)$ and $\partial_\sigma S(\sigma)(t_0, x) \leq \partial_\eta S(\eta)(t_0)$ for any t_0 and any other curve η in N with $\sigma(t_0, x) = \eta_{t_0}$.

Proposition 6.2. *Assume that S is K-convex. Then there exists a unique gradient flow σ for S and it satisfies:*

- **(i):** *$d(\sigma(t,x), \sigma(t,y)) \leq e^{-Kt} d(x,y)$ for all $x, y \in N$ and all $t \geq 0$.*
- **(ii):** *If in addition $K > 0$ and $S > -\infty$ then there exists a unique "ground state" $x_0 \in N$ satisfying $S(x) \geq S(x_0) + \frac{K}{2} d_N^2(x, x_0)$ for all $x \in N$.*
- **(iii):** *If $K > 0$ and $S(x_0) = 0$ then $-\partial_t S(\sigma(t,x)) \geq 2K \cdot S(\sigma(t,x))$ and thus $S(\sigma(t,x)) \leq e^{-2Kt} S(x)$.*

Property (i) is deduced in unpublished papers by Perelman and Petrunin as well as by A. Lytchak (private communication). Properties (ii) and (iii) may be regarded as generalized versions of Talagrand's inequality and Gross' logarithmic Sobolev inequality, resp. This may be seen in Example 6.4 below where we choose the space N and the function S more specifically.

From now on, let N be the space $\mathcal{P}^2(M)$ of probability measures on a smooth complete Riemannian manifold M and let d_N be the L^2-Wasserstein distance d_2^W distance derived from the Riemannian distance $d = d_M$. Following [McC01], K-convex functions on $\mathcal{P}^2(M)$ are also called *displacement K-convex* (to emphasize that it means K-convexity along the geodesics $t \mapsto \gamma_t$ w.r.t. d_2^W and not along the geodesics $t \mapsto (1-t)\gamma_0 + t\gamma_1$ in the linear space of signed measures).

Given an increasing function $U : \mathbb{R} \to \mathbb{R}$ and a lower semicontinuous function $V : M \to \mathbb{R}$ we define the *free energy* $S : \mathcal{P}^2(M) \to [-\infty, \infty]$ by

$$S(\nu) := \int_M U\left(\log \frac{d\nu}{dm}\right) d\nu + \int_M V\, d\nu \tag{6.2}$$

provided ν is absolutely continuous w.r.t. the Riemannian volume measure m and $\int U_+(\log \frac{d\nu}{dm})\, d\nu + \int V_+\, d\nu < \infty$. Otherwise, we define $S(\nu) := +\infty$.

Remark 6.3. Under minimal assumptions on M, U and V the gradient flow σ for S as above is given by $\sigma(t,\nu)(dx) = \rho(t,x)\, m(dx)$ where the densities ρ solve the nonlinear PDE

$$\partial_t \rho(t,x) = \Delta(\rho U'(\log \rho))(t,x) + \nabla(\rho \cdot \nabla V)(t,x)$$

on $\mathbb{R}_+ \times M$, [OtVi00], [Vil03].

If we can verify K-convexity of S for some $K > 0$ then this nonlinear diffusion equation has a unique stationary solution and any other solution converges exponentially fast to the stationary solution.

Example 6.4. The main examples are:

- $U(r) = r, V = 0$ yields the relative entropy $S(\nu) = \int_M \log \frac{d\nu}{dm}\, d\nu$. Its gradient flow is the usual heat equation $\partial_t \rho = \Delta \rho$. More precisely, the densities of the gradient flow are solutions of the heat equation.
- $U(r) = r$ leads to the Fokker-Planck equation
$$\partial_t \rho = \Delta \rho + \nabla(\rho \cdot \nabla V).$$
In this case, an easy calculation shows $S(\sigma(t,\nu)) = \int u^2 \log(u^2) e^{-V} dm$ and $-\partial_t S(\sigma(t,\nu)) = \frac{1}{4} \int |\nabla u|^2 e^{-V} dm$ provided we write $\frac{d\sigma}{dm} = u^2 e^{-V}$. Hence, here we indeed obtain the usual version of the logarithmic Sobolev inequality.
- $U(r) = \frac{1}{a} \exp(ar)$ for a constant $a \neq 0$ and $V = 0$ yields $S(\nu) = \frac{1}{a} \int_M \left(\frac{d\nu}{dm}\right)^a d\nu$. The associated gradient flow is given by the porous medium equation (if $a > 0$) or fast diffusion equation (if $a < 0$)
$$\partial_t \rho = \Delta(\rho^{1+a}).$$

Our main result from [Stu04] yields K-convexity for large classes of energy functionals associated to nonlinear diffusions on Euclidean and Riemannian spaces. As a consequence it yields exponential convergence to equilibrium for the solutions to these equations together with explicit bounds for the rate of convergence.

Theorem 6.5. *The free energy S from (6.2) is K-convex if and only if $U''(r) + \frac{1}{n} U'(r) \geq 0$ and*

$$U'(r) \cdot Ric_x(\xi,\xi) + Hess_x V(\xi,\xi) \geq K \cdot |\xi|^2$$

for all $r \in \mathbb{R}$, $x \in M$ and $\xi \in T_x M$.

Applications of this result to heat equation, Fokker-Planck equation and porous medium equation are straightforward.

Corollary 6.6. *The free energy $S(\nu) = \int_M \log \frac{d\nu}{dm}\, d\nu + \int_M V d\nu$ associated with the Fokker-Planck equation is K-convex if and only if the Bakry-Emery criterion*

$$Ric_x(\xi,\xi) + Hess_x V(\xi,\xi) \geq K \cdot |\xi|^2$$

is satisfied ($\forall x \in M, \forall \xi \in T_x M$).

In particular, the relative entropy $S(\nu) = \int_M \log \frac{d\nu}{dm}\, d\nu$ is a K-convex function on the metric space $\mathcal{P}^2(M)$ if and only if the Ricci curvature of the underlying Riemannian manifold M is bounded from below by K.

Corollary 6.7. *For any $N > 0$ the free energy $S(\nu) = -N \cdot \int_M \left(\frac{d\nu}{dm}\right)^{-1/N} d\nu$ associated with the fast diffusion equation $\partial_t \rho = \Delta(\rho^{1-1/N})$ is a 0-convex function on the metric space $\mathcal{P}^2(M)$ if and only if the underlying Riemannian manifold M has nonnegative Ricci curvature and dimension $\leq N$.*

Parts of the above corollaries had been obtained in [OtVi00], [CMS01] and [vRSt03]. The previous results yields a characterization of the curvature-dimension conditions $CD(K, \infty)$ as well as $CD(0, N)$ of Bakry-Emery in terms of contraction properties of nonlinear diffusions. The general condition $CD(K, N)$ may be characterized in a similar manner:

Theorem 6.8. *i) For $K > 0$ and $N > 0$ consider the free energy functional*

$$S(\nu) = \int_M \left[\log\left(\frac{d\nu}{dm}\right) - N\left(\frac{d\nu}{dm}\right)^{-1/N}\right] d\nu$$

associated with the nonlinear diffusion equation

$$\partial_t \rho = \Delta(\rho(1 + \rho^{-1/N})).$$

Then S is K-convex if and only if the dimension of the manifold is bounded from above by N and its Ricci curvature is bounded from below by K.

iii) For $K < 0$ and $N > 0$ consider the free energy functional

$$S(\nu) = -N \int_M \log\left[1 + \left(\frac{d\nu}{dm}\right)^{-1/N}\right] d\nu$$

associated with the nonlinear diffusion equation

$$\partial_t \rho = \Delta(\rho(1 + \rho^{1/N})^{-1}).$$

Then S is K-convex if and only if the dimension of the manifold is bounded from above by N and its Ricci curvature is bounded from below by K.

References

[Ale51] A. D. Alexandrov (1951): *A theorem on triangles in a metric space and some applications* (Russian), Trudy Math. Inst. Steklov **38**, 5-23.

[ABC00] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, G. Scheffer (2000): *Sur les inégalités de Sobolev logarithmiques.* Société Mathématique de France, Paris.

[Bal95] W. Ballmann (1995): *Lectures on spaces of nonpositive curvature.* DMV Seminar Band **25**, Birkhäuser Verlag, Basel.

[BGS85] W. Ballmann, M. Gromov, V. Schroeder (1985): *Manifolds of nonpositive curvature.* Progress in Mathematics **61**. Birkhäuser Boston Inc., Boston, MA.

[BaEm84] D. Bakry, M. Emery (1984): Hypercontractivite de semi-groupes de diffusion. *C. R. Acad. Sci., Paris, Ser. I*, **299**, 775–778.

[BrHa99] M.R. Bridson, A. Haefliger (1999): *Metric spaces of non-positive curvature.* Grundlehren der Mathematischen Wissenschaften, **319**. Springer-Verlag, Berlin.

[BBI01] D. Burago, Y. Burago, S. Ivanov (2001): *A course in metric geometry.* Graduate Studies in Mathematics, **33**. American Mathematical Society, Providence, RI.

[Car28] H. Cartan (1928): *Lecons sur lá geometrie des espaces de Riemann.* Gauthiers-Villars, Paris.

[CMS01] D. Cordero-Erausquin, R. McCann, M. Schnuckenschläger (2001): A Riemannian interpolation inequality à la Borell, Brascamb and Lieb. *Invent. Math.*, **146**, 219-257.

[DaRo02] G. Da Prato, M. Röckner (2002): Singular dissipative stochastic equations in Hilbert spaces. *Probab. Theory Relat. Fields*, **124**, 261-303.

[Dos49] S. Doss (1949): Sur la moyenne d'un élément aléatoire dans un espace distancié. *Bull. Sci. Math.* **73**, 48–72.

[Dud89] R. M. Dudley (1989): *Real analysis and probability.* The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.

[EeFu01] J. Eells, B. Fuglede (2001): *Harmonic maps between Riemannian polyhedra.* Cambridge Tracts in Mathematics, **142.** Cambridge University Press, Cambridge.

[EmMo91] D. Emery, G. Mokobodzki (1991): Sur le barycentre d'une probabilité dans une variété. *Séminaire de Probabilités XXV*, 220–233. Lecture Notes in Math. **1485**, Springer, Berlin.

[ESH99] A. Es-Sahib, H. Heinich (1999): Barycentre canonique pour un espace métrique à courbure négative. *Séminaire de Probabilités, XXXIII,* 355–370. Lecture Notes in Math., **1709**. Springer, Berlin.

[Fre48] M. Fréchet (1948): Les éléments alétoires de nature quelconque dans un espace distancié, *Ann. Inst. H. Poincaré* **10**, 215–310.

[Gau1809] Carl Friedrich Gauss(1809): *Theoria Motus Corporum Celestium.*

[Gro81/99] M. Gromov (1981), (1999): *Structures métriques pour les variétés Riemanniennes.* Rédigé par J. Lafontaine et P. Pansu, Cedic/Fernand Nathan.

[Her91] W. Herer (1991): Espérance mathématique au sens Doss d'une variable aléatoire dans un espace métrique. *C. R. Acad. Sci. Paris Sér. I*, **302**, 131–134.

[Jos94] J. Jost (1994): Equilibrium maps between metric spaces, *Calc. Var. Partial Differential Equations* **2**, 173-204.

[Jos97] J. Jost (1997): *Nonpositive curvature: geometric and analytic aspects.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel.

[Kar77] H. Karcher (1977): Riamannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.* **30**, 509-541.

[Ken90] W.S. Kendall (1990): Probability, convexity, and harmonic maps with small images. I. Uniqueness and fine existence. *Proc. London Math. Soc.* **61**, 371-406.

[Ken98] W.S. Kendall (1998): From stochastic parallel transport to harmonic maps. In: *New directions in Dirichlet forms.* Edt. J.Jost, W.Kendall, U. Mosco, M.Röckner, K.T.Sturm. pp 49-115. AMS and International Press.

[KoSc93] N. Korevaar, R. Schoen (1993): Sobolev spaces and harmonic maps for metric space targets. *Comm. Anal. Geom.* **1**, 561-569.

[McC01] R. McCann (2001): Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.* **11**, 589-608.

[Ott01] F. Otto (2001): The geometry of dissipative evolution equation: the porous medium equation. *Comm. PDE* **26**, 101-174.

[OtVi00] F. Otto, C. Villani (2000): Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, **173**, 361–400.

[Pic94] J. Picard (1994): Barycentres et martingales sur une variéte. *Ann. Inst. Henri Poincaré, Prob. Stat.* **30**, 647–702.

[RaRu98] S. T. Rachev, L. Rüschendorf (1998): *Mass transportation problems. Vol. I. Theory.* Probability and its Applications. Springer, New York.

[vRe03] M.-K. von Renesse (2003): Intrinsic Coupling on Riemannian Manifolds and Polyhedra. Preprint Univ. Bonn, 2003.

[vRSt03] M.-K. von Renesse, K.T. Sturm (2003): Transport inequalities, gradient estimates, entropy and Ricci curvature. *SFB 611 Preprint* **80**. Universität Bonn.

[Str03] D.W. Stroock (2003): *Markov processes from K. Ito's perspective.* Annals of Mathematics Studies. Princeton University Press.

[Stu98] K.T. Sturm (1998): Diffusion processes and heat kernels on metric spaces. *Ann. Probab.*, **26**, 1–55.

[Stu99] K.T. Sturm (1999): Metric spaces of lower bounded curvature. *Expo. Math.* **17**, 35-47.

[Stu01] K.T. Sturm (2001): *Stochastics and analysis on metric spaces.* Lecture Notes, Universität Bonn.

[Stu02] K.T. Sturm (2002): Nonlinear martingale theory for processes with values in metric spaces of nonpositive curvature. *Ann. Probab.* **30**, 1195-1222.

[Stu03] K.T. Sturm (2003): A Semigroup Approach to Harmonic Maps. *SFB 611 Preprint* **39**. Universität Bonn.

[Stu04] K.T. Sturm (2004): Nonlinear diffusions as gradient flows: convexity and contraction properties. Preprint Universität Bonn.

[Vil03] C. Villani (2003) *Topics in Mass Transportation.* Graduate Studies in Mathematics. American Mathematical Society.

Institut für Angewandte Mathematik
Universität Bonn
Wegelerstrasse 6
53115 Bonn
Germany
E-mail address: sturm@uni-bonn.de