Acta Math., 196 (2006), 65–131
DOI: 10.1007/s11511-006-0002-8

On the geometry of metric measure spaces. I

by

Karl-Theodor Sturm

Universität Bonn
Bonn, Germany

1. Introduction

The notion of a 'metric space' is one of the basic concepts of mathematics. Metric spaces play a prominent role in many fields of mathematics. In particular, they constitute natural generalizations of manifolds admitting all kinds of singularities and still providing rich geometric structures.

A. D. Alexandrov [1] introduced the notion of lower curvature bounds for metric spaces in terms of comparison properties for geodesic triangles. These curvature bounds are equivalent to lower bounds for the *sectional curvature* in the case where the metric spaces are Riemannian manifolds—and they may be regarded as generalized lower bounds for the 'sectional curvature' for general metric spaces. A fundamental observation (cf. K. Grove and P. Petersen [23]) is that these lower bounds are *stable* under an appropriate notion of convergence of metric spaces, the so-called *Gromov–Hausdorff convergence*, introduced by M. Gromov [20]. The family of manifolds with sectional curvature $\geqslant K$ is, of course, not closed under Gromov–Hausdorff convergence but the family of metric spaces with curvature $\geqslant K$ in the sense of Alexandrov is closed (for each $K\in\mathbf{R}$). Even more, the family of compact metric spaces with curvature $\geqslant K$, Hausdorff dimension $\leqslant N$ and diameter $\leqslant L$ is compact (for any choice of K, N, and L), see Y. Burago, M. Gromov and G. Perelman [9].

For many fundamental results in geometric analysis, however, the crucial ingredients are not bounds for the sectional curvature but bounds for the *Ricci curvature*: estimates for heat kernels and Green functions, Harnack inequalities, eigenvalue estimates, isoperimetric inequalities, Sobolev inequalities—they all depend on lower bounds for the Ricci curvature of the underlying manifolds as pointed out by S. T. Yau and others, see e.g. [36], [10], [14] and [47]. The family of Riemannian manifolds with given lower bound for the Ricci curvature is not closed under Gromov–Hausdorff convergence (nor it is closed under any other reasonable notion of convergence). One of the great challenges thus is to establish a generalized notion of lower Ricci curvature bounds for singular spaces. For detailed investigations and a survey of the state of the art for this problem, we refer to the contributions by J. Cheeger and T. Colding [12]. Fascinating new developments have been outlined just recently by G. Perelman [43] in the context of his work on the geometrization conjecture for 3-manifolds.

Generalizations of lower Ricci curvature bounds should be formulated in the framework of metric measure spaces. These are triples (M, d, m) where (M, d) is a metric space and m is a measure on the Borel σ-algebra of M. We will always require that the metric space (M, d) is complete and separable and that the measure m is locally finite. Recall that for generalizations of sectional curvature bounds only the metric structure (M, d) is required whereas for generalizations of Ricci curvature bounds in addition a reference measure m has to be specified. In a certain sense, this phenomenon is well-known from the discussion of the curvature-dimension condition of D. Bakry and M. Émery [4] in the framework of Dirichlet forms and symmetric Markov semigroups. Of course, also the Bakry–Émery condition is a kind of generalized lower bound for the Ricci curvature (together with an upper bound for the dimension). However, it is not given in terms of the basic data (M, d, m) but in terms of the Dirichlet form (or heat semigroup) derived from the original quantities in a highly non-trivial manner.

Metric measure spaces have been studied quite intensively in recent years. Of particular interest is the study of functional inequalities, like Sobolev and Poincaré inequalities, on metric measure spaces and the construction and investigation of function spaces of various types [24], [25], [33], [26]. To some extent, doubling properties for the volume and scale invariant Poincaré inequalities on metric balls can be regarded as weak replacements of lower Ricci curvature bounds. Among others, they allow one to construct Dirichlet forms, Laplacians and heat kernels on given metric measure spaces and to derive (elliptic and parabolic) Harnack inequalities as well as (upper and lower) Gaussian estimates for heat kernels [48], [11]. On the other hand, however, even in simplest cases doubling constant and Poincaré constant do not characterize spaces with lower bounded Ricci curvature: they always allow at least also metrics which are equivalent to the given ones.

This is the first of two papers on the geometry of metric measure spaces. In the first one, we present a *dimension-independent* concept of lower Ricci curvature bounds (which in particular also applies to infinite-dimensional examples). The main results of this paper are:

- We define a *complete and separable metric* $\mathbf{D}$ on the family of all isomorphism classes of normalized metric measure spaces (Theorem 3.6). The metric $\mathbf{D}$ has a natural interpretation, based on the concept of optimal mass transportation. It is a *length metric.*

- **D**-convergence is *weaker* than measured Gromov–Hausdorff convergence. Both are equivalent on each family of compact metric measure spaces with full supports and uniform bounds for the doubling constant and the diameter. Each of these families is **D**-*compact* (Theorem 3.16).
- We introduce a notion of *lower curvature bounds* $\underline{\mathsf{Curv}}(M,\mathsf{d},m)$ for metric measure spaces (M,d,m), based on convexity properties of the relative entropy $\mathrm{Ent}(\cdot\,|m)$ with respect to the reference measure m. Here $\nu\mapsto\mathrm{Ent}(\nu|m)$ is regarded as a function on the L_2-Wasserstein space of probability measures on the metric space (M,d). For Riemannian manifolds, $\underline{\mathsf{Curv}}(M,\mathsf{d},m)\geqslant K$ if and only if $\mathrm{Ric}_M(\xi,\xi)\geqslant K|\xi|^2$ for all $\xi\in TM$ (Theorem 4.9).
- *Local* lower curvature bounds imply *global* lower curvature bounds (Theorem 4.17).
- Lower curvature bounds are *stable* under **D**-convergence (Theorem 4.20). In particular, lower curvature bounds are stable under measured Gromov–Hausdorff convergence.
- Lower curvature bounds of the form $\underline{\mathsf{Curv}}(M,\mathsf{d},m)\geqslant K$ imply *estimates for the volume growth* of concentric balls, for instance, if $K\leqslant 0$ then

$$m(B_r(x))\leqslant C(x)\exp\left(-\frac{Kr^2}{2}\right)$$

for all $r\geqslant 1$ (Theorem 4.24).

In our second paper on the geometry of metric measure spaces [53], we will treat the *finite-dimensional case.* More precisely, we will study metric measure spaces satisfying a so-called curvature-dimension condition $\mathrm{CD}(K,N)$ being more restrictive than the previous condition $\underline{\mathsf{Curv}}(M,\mathsf{d},m)\geqslant K$ (which will be obtained as the borderline case $N=\infty$). The additional parameter N plays the role of an upper bound for the dimension. In some sense, the condition $\mathrm{CD}(K,N)$ will be the geometric counterpart to the analytic curvature-dimension condition of Bakry and Émery [4]. Our main results in the second paper will be:

- For Riemannian manifolds, $\mathrm{CD}(K,N)$ is *equivalent* to the conditions $\mathrm{Ric}_M(\xi,\xi)\geqslant K|\xi|^2$ and $\dim(M)\leqslant N$.
- The curvature-dimension condition $\mathrm{CD}(K,N)$ is *stable* under **D**-convergence.
- For any triple of real numbers K, N and L the family of normalized metric measure spaces with condition $\mathrm{CD}(K,N)$ and diameter $\leqslant L$ is **D**-*compact.*
- Condition $\mathrm{CD}(K,N)$ implies a generalized version of the *Brunn–Minkowski inequality*, e.g. if $K=0$ then

$$m(A_t)^{1/N}\geqslant(1-t)m(A_0)^{1/N}+tm(A_1)^{1/N}$$

for any pair of sets $A_0, A_1 \subset M$, where A_t denotes the set of all possible points γ_t on geodesics in M with endpoints $\gamma_0 \in A_0$ and $\gamma_1 \in A_1$.

• Condition $\mathrm{CD}(K,N)$ implies the *Bishop–Gromov volume comparison theorem*:

$$\frac{m(B_r(x))}{m(B_R(x))} \geqslant \frac{\int_0^r \sin\big(t\sqrt{K/(N-1)}\,\big)^{N-1}\,dt}{\int_0^R \sin\big(t\sqrt{K/(N-1)}\,\big)^{N-1}\,dt}$$

with the usual interpretation of the right-hand side if $K \leqslant 0$, e.g. as $(r/R)^N$ if $K=0$.

• Condition $\mathrm{CD}(K,N)$ for some positive K provides a sharp upper bound on the diameter (*Bonnet–Myers theorem*):

$$L \leqslant \pi \sqrt{\frac{N-1}{K}}\,.$$

• Condition $\mathrm{CD}(K,N)$ implies the doubling property and local, scale-invariant *Poincaré inequalities* on balls. In particular, it allows one to construct canonical Dirichlet forms with *Gaussian upper and lower bounds* for the corresponding heat kernels.

The concept of optimal mass transportation plays a crucial role in our approach. It originates in the classical transportation problems of G. Monge [40] and L. V. Kantorovich [28]. The basic quantity for us is the so-called L_2-*Wasserstein distance* between two probability measures μ and ν on a given complete separable metric space (M, d) defined as

$$\mathsf{d}_W(\mu,\nu) := \inf_q \left(\int_{M\times M} \mathsf{d}^2(x,y)\,dq(x,y) \right)^{1/2},$$

where the infimum is taken over all *couplings* q of μ and ν. The latter are probability measures on the product space $M\times M$ whose marginals (i.e. image measures under the projections) are the given measures μ and ν. One choice, of course, is $q=\mu\otimes\nu$ but in most cases this will be a very bad choice if one aims for minimal transportation costs. The L_2-Wasserstein distance can be interpreted as the minimal transportation costs (measured in L_2-sense) for transporting goods from producers at locations distributed according to μ to consumers at locations distributed according to ν.

Two results may be regarded as milestones in the recent development of theory and application of mass transportation concepts; these results have raised an increasing interest in this topic of people from various fields of mathematics including partial differential equations, geometry, fluid mechanics, and probability. See, e.g., [54], [35], [5], [42], [15], [13], [3], [17] and in particular the monograph by C. Villani [55] which gives an excellent survey on the whole field.

The first of these two results is the polar factorization of Y. Brenier [6] and its extension to the Riemannian setting by R. McCann [38], [39]. The second one is F. Otto's

[27], [41] formal Riemannian calculus on the space $\mathcal{P}_2(M)$ of probability measures on M, equipped with the L_2-Wasserstein metric, and his interpretation of the heat equation (and of other non-linear dissipative evolution equations) as gradient flow(s) of the *relative entropy*

$$\mathrm{Ent}(\nu\,|\,m)=\int_M \frac{d\nu}{dm}\log\left(\frac{d\nu}{dm}\right)dm$$

(or related functionals, respectively) on $\mathcal{P}_2(M)$.

It turned out that convexity properties of the function $\nu\mapsto\mathrm{Ent}(\nu|m)$ are intimately related to curvature properties of the underlying metric measure space (M,d,m). If M is a complete Riemannian manifold with Riemannian distance d and if $m=e^{-V}dx$ then M.-K. von Renesse and the author proved (see [46] for the case $V=0$ and [51] for the general case) that the function $\nu\mapsto\mathrm{Ent}(\nu|m)$ is K-convex(1) on $\mathcal{P}_2(M)$ if and only if

$$\mathrm{Ric}_M(\xi,\xi)+\mathrm{Hess}\,V(\xi,\xi)\geqslant K|\xi|^2$$

for all $\xi\in TM$. A heuristic argument for the 'if' implication of this equivalence was presented in [42], based on the formal Riemannian calculus on $\mathcal{P}_2(M)$. In the particular case $K=0$ and $V=0$, the 'if' implication was proven in [13].

Having in mind these results, it seems quite natural to say that an arbitrary metric measure space (M,d,m) has *curvature* $\geqslant K$ if and only if for any pair $\nu_0,\nu_1\in\mathcal{P}_2(M)$ with $\mathrm{Ent}(\nu_0|m)<\infty$ and $\mathrm{Ent}(\nu_1|m)<\infty$ there exists a geodesic $\Gamma\colon[0,1]\to\mathcal{P}_2(M)$ connecting ν_0 and ν_1 with

$$\mathrm{Ent}(\Gamma(t)\,|\,m)\leqslant(1-t)\,\mathrm{Ent}(\Gamma(0)\,|\,m)+t\,\mathrm{Ent}(\Gamma(1)\,|\,m)-\frac{K}{2}t(1-t)\,\mathsf{d}_W^2(\Gamma(0),\Gamma(1))\qquad(1.1)$$

for all $t\in[0,1]$. In this case, we also briefly write $\underline{\mathsf{Curv}}(M,\mathsf{d},m)\geqslant K$.

A crucial property of this kind of curvature bound is its stability under convergence of metric measure spaces. Of course, this requires an appropriate notion of topology or distance on the family of all metric measure spaces. We define the *L_2-transportation distance* between two normalized metric measure spaces by

$$\mathbf{D}((M,\mathsf{d},m),(M',\mathsf{d}',m')):=\inf_{q,\hat{\mathsf{d}}}\left(\int_{M\times M'}\hat{\mathsf{d}}^2(x,y)\,dq(x,y)\right)^{1/2},$$

where the infimum is taken over all couplings q of m and m' and over all couplings $\hat{\mathsf{d}}$ of d and d'. The former are probability measures on $M\times M'$ with marginals m and m'. The latter are pseudo-metrics on the disjoint union $M\sqcup M'$ which extend d and d'.

(1) By definition, this means that property (1.1) below holds for each geodesic Γ in $\mathcal{P}_2(M)$.

Also, the distance $\mathbf{D}$ has an interpretation in terms of mass transportation: In order to realize the distance

$$\mathbf{D}((M,\mathsf{d},m),(M',\mathsf{d}',m'))$$

between two normalized metric measure spaces (M,d,m) and (M',d',m') one first may use isometric transformations of (M,d) and (M',d') to bring the images of m and m' in optimal position to each other. (In the sense of transportation costs, these transformations are for free.) Then one has to solve the usual mass transportation problem, trying to minimize the transportation costs in the L_2-sense.

It turns out that $\mathbf{D}$ is a complete separable length metric on the family $\mathbf{X}_1$ of all isomorphism classes of normalized metric measure spaces. For any choice of real numbers K, C and L, the family of normalized metric measure spaces with curvature $\geqslant K$, as well as the family of normalized metric measure spaces with doubling constant $\leqslant C$, are closed under $\mathbf{D}$-convergence. Moreover, the family $\mathbf{X}_1(C,L)$ of normalized metric measure spaces with doubling constant $\leqslant C$ and diameter $\leqslant L$ is compact under $\mathbf{D}$-convergence.

The $\mathbf{D}$-topology is *weaker* than the topology of *measured Gromov–Hausdorff* convergence, introduced by K. Fukaya [19]. Both topologies are equivalent on each family $\widetilde{\mathbf{X}}_1(C,L)$ of compact metric measure spaces with full supports and uniform bounds for the doubling constant and the diameter. As we will see in [53], this in particular applies to each family of compact metric measure spaces with full supports, with diameter $\leqslant L$ and satisfying a curvature-dimension condition $\mathrm{CD}(K,N)$.

For various other distances on the family $\mathbf{X}_1$, see [22, Chapter $3\frac{1}{2}$]. A completely different notion of distance between Riemannian manifolds was proposed by A. Kasue [31, 30], based on the short time asymptotics of the heat kernel. Yet another convergence concept was proposed by K. Kuwae and T. Shioya [34] extending the concept of Γ-convergence and Mosco convergence towards a notion of convergence of operators (or Dirichlet forms or heat semigroups) on varying spaces.

A major advantage of our distance $\mathbf{D}$ seems to be that it has a very natural geometric interpretation, namely, in terms of the above-mentioned mass transportation concept. We also expect that it is closely related to more analytic properties of metric measure spaces. Following [27], the heat semigroup on a metric measure space (M,d,m) should be obtained as the gradient flow on $\mathcal{P}_2(M)$ for the relative entropy $\mathrm{Ent}(\,\cdot\,|m)$. Curvature bounds of the form $\underline{\mathsf{Curv}}(M,\mathsf{d},m)\geqslant K$ should e.g. imply K-contractivity of the heat flow

$$\mathsf{d}_W(\mu p_t,\nu p_t)\leqslant e^{-Kt}\mathsf{d}_W(\mu,\nu),$$

gradient estimates for harmonic functions, isoperimetric inequalities, and volume growth estimates.

Here in the present paper, we will proceed as follows:

In Section 2 we give a brief survey on the geometry of metric spaces, recalling the concepts of length and geodesic spaces, the Gromov–Hausdorff distance and the lower curvature bounds in the sense of Alexandrov. We introduce the L_2-Wasserstein space of probability measures on a given metric space and derive some of the basic properties.

Section 3 is devoted to the metric $\mathbf{D}$. The first main result states that it indeed defines a (complete and separable) length metric on the family of isomorphism classes of normalized metric measure spaces. We collect several simple examples of $\mathbf{D}$-convergence with increasing and decreasing dimensions and we discuss closedness and compactness properties of the families of normalized metric measure spaces with the doubling property. We also present a detailed discussion of the relation between $\mathbf{D}$-convergence and the classical measured Gromov–Hausdorff convergence.

In Section 4 we study metric measure spaces with curvature bounds. First we introduce and discuss the relative entropy, then we present the definition of curvature bounds and analyze their behavior under various transformations (isomorphisms, scaling, weights, subsets and products). The main results are the globalization theorem and the convergence theorem. Finally, we deduce growth estimates for the volume of concentric balls.

After submitting this paper, the author got knowledge of related work by J. Lott and C. Villani. Their paper [37], which was finished soon after that, presents various related concepts and results. However, both papers are completely independent.

The basic concepts and main results of the present paper have also been announced in [52].

2. On the geometry of metric spaces

2.1. Length and geodesic spaces

Let us summarize some definitions and basic results on the geometry of metric spaces. For proofs and further details we refer to [7], [22], and [8].

Throughout this paper, a *pseudo-metric* on a set M will be a function $\mathsf{d}: M\times M\to [0,\infty]$ which is symmetric, vanishes on the diagonal and satisfies the triangle inequality. If it does not vanish outside the diagonal and does not take the value $+\infty$ then it is called *metric*. From now on, let (M,d) be a metric space. Open balls in M will be denoted by $B_r(x)=\{y\in M : \mathsf{d}(x,y)<r\}$, their closures by $\overline{B}_r(x)\subset\{y\in M : \mathsf{d}(x,y)\leqslant r\}$. A *curve* connecting two points $x,y\in M$ is a continuous map $\gamma:[a,b]\to M$, with $\gamma(a)=x$

and $\gamma(b)=y$. Then, obviously, $\mathrm{Length}(\gamma)\geqslant \mathsf{d}(x,y)$, with the length of γ being defined as

$$\mathrm{Length}(\gamma)=\sup \sum_{k=1}^{n} \mathsf{d}(\gamma(t_{k-1}),\gamma(t_k)),$$

where the supremum is taken over all partitions $a=t_0<t_1<...<t_n=b$. If $\mathrm{Length}(\gamma)<\infty$ then γ is called *rectifiable*. In this case we can and will henceforth always assume that (after suitable reparametrization) γ has constant speed, i.e.

$$\mathrm{Length}(\gamma|_{[s,t]})=\frac{t-s}{b-a}\mathrm{Length}(\gamma)$$

for all $a<s<t<b$. In general, we will not distinguish between curves and equivalence classes of curves which are reparametrizations of each other. The curve $\gamma\colon[a,b]\to M$ is called *geodesic* if and only if $\mathrm{Length}(\gamma)=\mathsf{d}(\gamma(a),\gamma(b))$. A geodesic in this sense is always minimizing.

A metric space (M,d) is called *length space*, or *length metric space*, if and only if for all $x,y\in M$,

$$\mathsf{d}(x,y)=\inf_{\gamma} \mathrm{Length}(\gamma),$$

where the infimum is taken over all curves γ in M which connect x and y. A metric space (M,d) is called *geodesic space*, or *geodesic metric space*, if and only if each pair of points $x,y\in M$ is connected by a geodesic. (This geodesic is not required to be unique.)

LEMMA 2.1. *A complete metric space* (M,d) *is a length space* (*or geodesic space*) *if and only if for each pair of points* $x_0,x_1\in M$ *and for each* $\varepsilon>0$ (*or for* $\varepsilon=0$, *respectively*) *there exists a point* $y\in M$ *satisfying for each* $i=0,1$,

$$\mathsf{d}(x_i,y)\leqslant \tfrac{1}{2}\mathsf{d}(x_0,x_1)+\varepsilon. \tag{2.1}$$

Any such point y *will be called* ε-midpoint *of* x_0 *and* x_1. *In the case* $\varepsilon=0$ *it will be called* midpoint *of* x_0 *and* x_1.

Remark 2.2. Given $x_0,x_1\in M$ then each ε-midpoint $y\in M$ satisfies

$$\mathsf{d}^2(x_0,y)+\mathsf{d}^2(y,x_1)\leqslant \tfrac{1}{2}\mathsf{d}^2(x_0,x_1)+\varepsilon' \tag{2.2}$$

with $\varepsilon'=2\varepsilon\,\mathsf{d}(x_0,x_1)+2\varepsilon^2$. Vice versa, each $y\in M$ which satisfies (2.2) is an ε-midpoint with $\varepsilon=\sqrt{(\mathsf{d}(x_0,x_1)/2)^2-\varepsilon'/2}-\mathsf{d}(x_0,x_1)/2-\sqrt{\varepsilon'/2}$.

Indeed, if y is an ε-midpoint then

$$\mathsf{d}^2(x_0,y)+\mathsf{d}^2(y,x_1)\leqslant 2\big[\tfrac{1}{2}\mathsf{d}(x_0,x_1)+\varepsilon\big]^2=\tfrac{1}{2}\mathsf{d}^2(x_0,x_1)+\varepsilon'$$

with ε' chosen as above. Conversely, if y satisfies (2.2) then

$$\begin{aligned}\tfrac{1}{2}\mathsf{d}^2(x_0,x_1)+\varepsilon' \geqslant \mathsf{d}^2(x_0,y)+\mathsf{d}^2(y,x_1)\\ &= \tfrac{1}{2}[\mathsf{d}(x_0,y)+\mathsf{d}(y,x_1)]^2+\tfrac{1}{2}[\mathsf{d}(x_0,y)-\mathsf{d}(y,x_1)]^2\\ &\geqslant \tfrac{1}{2}\mathsf{d}^2(x_0,x_1)+\tfrac{1}{2}[\mathsf{d}(x_0,y)-\mathsf{d}(y,x_1)]^2.\end{aligned}$$

Hence, $|\mathsf{d}(x_0,y)-\mathsf{d}(y,x_1)|\leqslant\sqrt{2\varepsilon'}$ and

$$2\mathsf{d}(x_i,y)-\sqrt{2\varepsilon'}\leqslant \mathsf{d}(x_0,y)+\mathsf{d}(y,x_1)\leqslant\sqrt{\mathsf{d}^2(x_0,x_1)+2\varepsilon'}$$

for $i=0,1$.

LEMMA 2.3. *If* (M,d) *is a complete length space then*:

(i) *The closure of* $B_r(x)$ *is* $\{y\in M:\mathsf{d}(x,y)\leqslant r\}$;

(ii) M *is locally compact if and only if each closed ball in* M *is compact*;

(iii) *if* M *is locally compact then it is a geodesic space.*

Recall that the Hausdorff distance between two subsets A_1 and A_2 of a metric space (M,d) is given by

$$\mathsf{d}^{\mathsf{H}}(A_1,A_2)=\inf\{\varepsilon>0 : A_1\subset B_\varepsilon(A_2) \text{ and } A_2\subset B_\varepsilon(A_1)\},$$

where $B_\varepsilon(A):=\{x\in M:\inf_{y\in A}\mathsf{d}(x,y)<\varepsilon\}$ denotes the ε-neighborhood of $A\subset M$. The *Gromov–Hausdorff distance* between two metric spaces (M_1,d_1) and (M_2,d_2) is defined by

$$\mathsf{D}^{\mathsf{GH}}((M_1,\mathsf{d}_1),(M_2,\mathsf{d}_2))=\inf\ \mathsf{d}^{\mathsf{H}}(j_1(M_1),j_2(M_2)),$$

where the inf is taken over all metric spaces (M,d) and over all isometric embeddings $j_1: M_1\hookrightarrow M$, $j_2: M_2\hookrightarrow M$.

A related but slightly different quantity is defined as

$$\widetilde{\mathsf{D}}^{\mathsf{GH}}((M_1,\mathsf{d}_1),(M_2,\mathsf{d}_2))=\inf\ \{\varepsilon>0 : \text{there is an } \varepsilon\text{-isometry from } (M_1,\mathsf{d}_1) \text{ to } (M_2,\mathsf{d}_2)\}.$$

Recall that a map $\psi: M_1\to M_2$ is an ε-isometry from (M_1,d_1) to (M_2,d_2) if and only if $B_\varepsilon(\psi(M_1))=M_2$ and $|\mathsf{d}_2(\psi(x),\psi(y))-\mathsf{d}_1(x,y)|\leqslant\varepsilon$ for all $x,y\in M_1$.

PROPOSITION 2.4. (i) *The Gromov–Hausdorff metric* D^{GH} *is a pseudo-metric on the family* X *of isometry classes of metric spaces. Moreover,* D^{GH} *and* $\widetilde{\mathsf{D}}^{\mathsf{GH}}$ *are equivalent*: $\tfrac{1}{2}\widetilde{\mathsf{D}}^{\mathsf{GH}}\leqslant\mathsf{D}^{\mathsf{GH}}\leqslant 2\widetilde{\mathsf{D}}^{\mathsf{GH}}$.

(ii) *If a complete metric space* (M,d) *is the GH-limit of a sequence of length spaces then* (M,d) *is a length space.*

(iii) *Let* X_c *denote the family of isometry classes of compact metric spaces. Then* $(\mathsf{X}_c,\mathsf{D}^{\mathsf{GH}})$ *is a complete separable metric space. The family* X_f *of isometry classes of metric spaces with finitely many points is GH-dense in* X_c.

2.2. Alexandrov spaces

Now let us briefly discuss metric spaces with lower curvature bounds in the sense of A. D. Alexandrov [1]. The latter are generalizations of lower bounds for the sectional curvature for Riemannian manifolds. The results of this section will not be used in the sequel. The focus in this paper is on generalizations of lower bounds for the Ricci curvature. Partly, however, there will be some analogy to Alexandrov's generalizations of lower bounds for the sectional curvature. We summarize some of the basic properties of these metric spaces and refer to [9], [22], [8] and [44] for further details.

Given any $K\in\mathbf{R}$ we say that a complete length space (M,d) *locally has curvature* $\geqslant K$ if and only if each point $p\in M$ has a neighborhood $M_p\subset M$ such that for each quadruple of points $z,x_1,x_2,x_3\in M_p$,

$$\sphericalangle_K(z;x_1,x_2)+\sphericalangle_K(z;x_2,x_3)+\sphericalangle_K(z;x_3,x_1)\leqslant 2\pi. \tag{2.3}$$

We say that a complete length space (M,d) *globally has curvature* $\geqslant K$ if and only if the previous is true with $M_p:=M$. Here for any triple of points $z,x,y\in M$ we denote by $\sphericalangle_K(z;x,y)$ the angle at $\bar z$ of a triangle $\Delta(\bar z,\bar x,\bar y)$ with side lengths $\bar z\bar x=\mathsf{d}(z,x)$, $\bar z\bar y=\mathsf{d}(z,y)$ and $\bar x\bar y=\mathsf{d}(x,y)$ in the simply connected 2-dimensional space of constant curvature K, i.e.

$$\sphericalangle_K(z;x,y)=\arccos\left(\frac{\cos\big(\mathsf{d}(x,y)\sqrt{K}\,\big)-\cos\big(\mathsf{d}(z,x)\sqrt{K}\,\big)\cos\big(\mathsf{d}(z,y)\sqrt{K}\,\big)}{\sin\big(\mathsf{d}(z,x)\sqrt{K}\,\big)\sin\big(\mathsf{d}(z,y)\sqrt{K}\,\big)}\right) \tag{2.4}$$

(with appropriate interpretations/modifications if $K\leqslant 0$). This makes perfectly sense if $K[\mathsf{d}(z,x)+\mathsf{d}(x,y)+\mathsf{d}(y,z)]^2<(2\pi)^2$. Otherwise, we put $\sphericalangle_K(z;x,y):=-\infty$. There is an exceptional definition for spaces which are isometric to 1-dimensional manifolds (intervals or circles): we say that (M,d) (locally/globally) has curvature $\geqslant K$ if and only if $K\leqslant(\pi/L)^2$, where $L\leqslant\infty$ denotes the diameter. Generally, we put

$$\underline{\mathsf{curv}}(M,\mathsf{d})=\sup\{K\in\mathbf{R}:(M,\mathsf{d}) \text{ globally has curvature } \geqslant K\}.$$

Complete length spaces with curvature $\geqslant K$ and finite Hausdorff dimension are called *Alexandrov spaces* with curvature $\geqslant K$. For complete geodesic spaces there are several alternative (but equivalent) ways to define this curvature bound: via triangle comparison, angle monotonicity and convexity properties of the distance. For instance, one can interpret it as a weak formulation of

$$\operatorname{Hess}\frac{1}{K}\cos\big(\mathsf{d}(z,\cdot\,)\sqrt{K}\,\big)\geqslant-\cos\big(\mathsf{d}(z,\cdot\,)\sqrt{K}\,\big) \tag{2.5}$$

for all $z\in M$ (with appropriate modification in the case $K\leqslant 0$, e.g. $\operatorname{Hess}\mathsf{d}^2(z,\cdot\,)/2\leqslant 1$ if $K=0$).

Example 2.5. Let M be a complete Riemannian manifold with Riemannian distance d and dimension $n\geqslant 2$. Then $\underline{\mathsf{curv}}(M,\mathsf{d})$ is the greatest lower bound for the *sectional curvature* of M.

Remark 2.6. (i) Lower curvature bounds can also be defined on spaces which are not length spaces; see [49]. In this context, for instance, for each metric space (M,d) the metric space $\big(M,\sqrt{\mathsf{d}}\,\big)$ would have curvature $\geqslant 0$.

(ii) Similarly, one can define metric spaces of curvature $\leqslant K$ and a number $\overline{\mathsf{curv}}(M,\mathsf{d})$ (which coincides with the least upper bound for the sectional curvature if M is a Riemannian manifold). However, in this paper we concentrate on lower curvature bounds.

PROPOSITION 2.7. *For each complete length space* (M,d) *the following properties hold:*

(i) Scaling: $\underline{\mathsf{curv}}(M,\alpha\mathsf{d})=\alpha^{-2}\,\underline{\mathsf{curv}}(M,\mathsf{d})$ *for all* $\alpha\in\mathbf{R}_+$.

(ii) Products: *If* $(M,\mathsf{d})=\bigotimes_{i=1}^n(M_i,\mathsf{d}_i)$ *for some* $n\geqslant 2$, *with complete length spaces* $(M_1,\mathsf{d}_1),...,(M_n,\mathsf{d}_n)$ *consisting of more than one point, then*

$$\underline{\mathsf{curv}}(M,\mathsf{d})=\inf\{\,\underline{\mathsf{curv}}(M_1,\mathsf{d}_1),...,\,\underline{\mathsf{curv}}(M_n,\mathsf{d}_n),0\}.$$

(iii) Local/global: *If a complete length space locally has curvature* $\geqslant K$ *then it also globally has curvature* $\geqslant K$ (Toponogov's globalization theorem).

(iv) Convergence: *Let* $((M_n,\mathsf{d}_n))_{n\in\mathbf{N}}$ *be a sequence of complete length spaces GH-converging to a complete length space* (M,d). *Then*

$$\underline{\mathsf{curv}}(M,\mathsf{d})\geqslant\limsup_{n\to\infty}\,\underline{\mathsf{curv}}(M_n,\mathsf{d}_n).$$

In particular, for each $K\in\mathbf{R}$ *the set* $\mathsf{X}_c(K)$ *of all compact length spaces* (M,d) *with curvature* $\geqslant K$ *is a closed subset of* $(\mathsf{X}_c,\mathsf{D}^{\mathsf{GH}})$.

(v) Compactness: *For each* $K\in\mathbf{R}$, $N\in\mathbf{N}$ *and* $L\in\mathbf{R}_+$ *the set* $\mathsf{X}_c(K,N,L)$ *of all compact length spaces* (M,d) *with curvature* $\geqslant K$, *Hausdorff dimension* $\leqslant N$ *and diameter* $\leqslant L$ *is GH-compact.*

Definition 2.8. A geodesic space (M,d) is called *non-branching* if and only if for each quadruple of points z,x_0,x_1,x_2, with z being the midpoint of x_0 and x_1 as well as the midpoint of x_0 and x_2, it follows that $x_1=x_2$.

Remark 2.9. If a geodesic space has curvature $\geqslant K$ for some $K\in\mathbf{R}$ then it is non-branching.

2.3. The L_2-Wasserstein space

Probability measures on metric spaces will play an important role throughout this paper. We collect some definitions and the basic facts on the L_2-Wasserstein distance. For further reading we recommend [16], [29], [45], [55] and [56].

For the rest of this section, let (M, d) be a *complete separable* metric space. A *measure* ν on M will always mean a measure on $(M, \mathcal{B}(M))$ with $\mathcal{B}(M)$ being the Borel σ-algebra of M (generated by the open balls in M). Recall that supp$[\nu]$, the *support* of ν, is the smallest closed set $M_0 \subset M$ such that $\nu(M \setminus M_0) = 0$. The *push-forward* of ν under a measurable map $f: M \to M'$ into another metric space M' is the probability measure $f_*\nu$ on M' given by

$$(f_*\nu)(A) := \nu(f^{-1}(A))$$

for all measurable $A \subset M'$. Given two measures μ and ν on M we say that a measure q on $M \times M$ is a *coupling* of μ and ν if and only if its marginals are μ and ν, that is, if and only if

$$q(A \times M) = \mu(A) \quad \text{and} \quad q(M \times A) = \nu(A)$$

for all measurable sets $A \subset M$. (This in particular implies that the total masses coincide: $\nu(M) = q(M \times M) = \mu(M)$.) If μ and ν are probability measures then for instance one such coupling is the product measure $\mu \otimes \nu$.

The L_2-*Wasserstein distance* between μ and ν is defined as

$$\mathsf{d}_W(\mu, \nu) = \inf\left\{ \left(\int_{M \times M} \mathsf{d}^2(x, y)\, dq(x, y) \right)^{1/2} : q \text{ is a coupling of } \mu \text{ and } \nu \right\}. \tag{2.6}$$

Note that $\mathsf{d}_W(\mu, \nu) = +\infty$ whenever $\mu(M) \neq \nu(M)$. We denote by $\mathcal{P}_2(M, \mathsf{d})$ or briefly $\mathcal{P}_2(M)$ the space of all probability measures ν on M with finite second moments:

$$\int_M \mathsf{d}^2(o, x)\, d\nu(x) < \infty$$

for some (hence all) $o \in M$. The pair $(\mathcal{P}_2(M), \mathsf{d}_W)$ is called L_2-*Wasserstein space* over (M, d).

PROPOSITION 2.10. (i) *$(\mathcal{P}_2(M), \mathsf{d}_W)$ is a complete separable metric space.*

The map $x \mapsto \delta_x$ defines an isometric and totally geodesic embedding of (M, d) into $(\mathcal{P}_2(M), \mathsf{d}_W)$.

The set of all normalized configurations

$$\mu = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$$

with $n\in\mathbf{N}$ *and* $x_1,...,x_n\in M$ *is dense in* $\mathcal{P}_2(M)$.

(ii) d_W*-convergence implies weak convergence (in the sense of measures). More precisely, if* $(\mu_n)_{n\in\mathbf{N}}$ *is a sequence in* $\mathcal{P}_2(M)$ *then* $\mathsf{d}_W(\mu_n,\mu)\to 0$ *if and only if* $\mu_n\to\mu$ *weakly and*

$$\lim_{R\to\infty}\sup_n\int_{M\setminus B_R(o)}\mathsf{d}^2(o,x)\,d\mu_n(x)=0 \tag{2.7}$$

for some (hence each) point $o\in M$. *Note that obviously* (2.7) *is always satisfied if* (M,d) *is bounded.*

(iii) $(\mathcal{P}_2(M),\mathsf{d}_W)$ *is a compact space or a length space if and only if* (M,d) *is so.*

(iv) *If* M *is a length space with more than one point then*

$$\underline{\mathsf{curv}}(\mathcal{P}_2(M),\mathsf{d}_W)=0 \quad\Longleftrightarrow\quad \underline{\mathsf{curv}}(M,\mathsf{d})\geqslant 0$$

and

$$\underline{\mathsf{curv}}(\mathcal{P}_2(M),\mathsf{d}_W)=-\infty \quad\Longleftrightarrow\quad \underline{\mathsf{curv}}(M,\mathsf{d})<0.$$

Proof. (i), (ii) See [45], [55].

($\mathrm{iii_a}$) The 'only if' statements follow from the fact that M is isometrically embedded in $\mathcal{P}_2(M)$.

($\mathrm{iii_b}$) Compactness of M implies compactness of $\mathcal{P}_2(M)$ according to (ii) and Prohorov's theorem.

($\mathrm{iii_c}$) Assume that (M,d) is a length space and let $\varepsilon>0$ and $\mu,\nu\in\mathcal{P}_2(M)$ be given. We have to prove that there exists an ε-midpoint η of μ and ν. Choose $n\in\mathbf{N}$ and $x_1,...,x_n,y_1,...,y_n\in M$ such that $\mathsf{d}_W(\mu,\bar\mu)\leqslant\varepsilon/3$, $\mathsf{d}_W(\nu,\bar\nu)\leqslant\varepsilon/3$ and

$$\mathsf{d}_W^2(\bar\mu,\bar\nu)=\frac{1}{n}\sum_{i=1}^n\mathsf{d}^2(x_i,y_i),$$

where $\bar\mu:=(1/n)\sum_{i=1}^n\delta_{x_i}$ and $\bar\nu:=(1/n)\sum_{i=1}^n\delta_{y_i}$ For each $i=1,...,n$ let z_i be an $\varepsilon/3$-midpoint of x_i and y_i and put $\eta:=(1/n)\sum_{i=1}^n\delta_{z_i}$. Then

$$\begin{aligned}\mathsf{d}_W(\bar\mu,\eta)&\leqslant\left(\frac{1}{n}\sum_{i=1}^n\mathsf{d}^2(x_i,z_i)\right)^{1/2}\\&\leqslant\left(\frac{1}{n}\sum_{i=1}^n\left[\frac{1}{2}\mathsf{d}(x_i,y_i)+\frac{\varepsilon}{3}\right]^2\right)^{1/2}\\&\leqslant\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^n\mathsf{d}^2(x_i,y_i)\right)^{1/2}+\frac{\varepsilon}{3}\\&\leqslant\frac{1}{2}\mathsf{d}_W(\bar\mu,\bar\nu)+\frac{\varepsilon}{3}\end{aligned}$$

and thus $\mathsf{d}_W(\mu,\eta)\leqslant\frac{1}{2}\mathsf{d}_W(\mu,\nu)+\varepsilon$. Similarly, $\mathsf{d}_W(\nu,\eta)\leqslant\frac{1}{2}\mathsf{d}_W(\mu,\nu)+\varepsilon$. This proves the claim.

(iv$_\mathrm{a}$) Assume that (M,d) has curvature $\geqslant 0$. Then for each $n\in\mathbf{N}$ the space $M^n=M\times...\times M$ has curvature $\geqslant 0$ (see Proposition 2.7 (ii)). According to [49], the latter is equivalent to

$$\sum_{i,j=1}^{l}\lambda_i\lambda_j\mathsf{d}^2(y_i,y_j)\leqslant 2\sum_{i=1}^{l}\lambda_i\mathsf{d}^2(y_i,y_0) \tag{2.8}$$

for all $l\in\mathbf{N}$, all $\lambda_1,...,\lambda_l\in\mathbf{R}_+$ with $\sum_{i=1}^{l}\lambda_i=1$ and all $y_0,y_1,...,y_l\in M^n$. In order to prove that $(\mathcal{P}_2(M),\mathsf{d}_W)$ has curvature $\geqslant 0$, let $l\in\mathbf{N}$, $\lambda_1,...,\lambda_l\in\mathbf{R}_+$ and $\nu_0,\nu_1,...,\nu_l\in\mathcal{P}_2(M)$ be given. For $\varepsilon>0$ choose $n\in\mathbf{N}$ and $y_0=(y_{01},...,y_{0n}),...,y_l=(y_{l1},...,y_{ln})\in M^n$ such that $\mathsf{d}_W(\nu_i,\bar\nu_i)\leqslant\varepsilon$ for all $i=0,1,...,l$ and $\mathsf{d}_W^2(\bar\nu_i,\bar\nu_0)=(1/n)\sum_{k=1}^{n}\mathsf{d}^2(y_{ik},y_{0k})=(1/n)\mathsf{d}^2(y_i,y_0)$ for all $i=1,...,l$, where we put

$$\bar\nu_i=\frac{1}{n}\sum_{k=1}^{n}\delta_{y_{ik}}.$$

Then $\mathsf{d}_W^2(\bar\nu_i,\bar\nu_j)\leqslant(1/n)\sum_{k=1}^{n}\mathsf{d}^2(y_{ik},y_{jk})=(1/n)\mathsf{d}^2(y_i,y_j)$ for all $i,j=1,...,l$ and thus by (2.8)

$$\sum_{i,j=1}^{l}\lambda_i\lambda_j\mathsf{d}_W^2(\bar\nu_i,\bar\nu_j)\leqslant\frac{1}{n}\sum_{i,j=1}^{l}\lambda_i\lambda_j\mathsf{d}^2(y_i,y_j)\leqslant\frac{2}{n}\sum_{i=1}^{l}\lambda_i\mathsf{d}^2(y_i,y_0)=2\sum_{i=1}^{l}\lambda_i\mathsf{d}_W^2(\bar\nu_i,\bar\nu_0).$$

In the limit $\varepsilon\to 0$ this yields

$$\sum_{i,j=1}^{l}\lambda_i\lambda_j\mathsf{d}_W^2(\nu_i,\nu_j)\leqslant 2\sum_{i=1}^{l}\lambda_i\mathsf{d}_W^2(\nu_i,\nu_0)$$

which (again by [49]) proves the claim.

(iv$_\mathrm{b}$) It is obvious that $\underline{\mathsf{curv}}(\mathcal{P}_2(M),\mathsf{d}_W)\leqslant\underline{\mathsf{curv}}(M,\mathsf{d})$, since (M,d) is isometrically and totally geodesically embedded into $(\mathcal{P}_2(M),\,\mathsf{d}_W)$.

(iv$_\mathrm{c}$) Assume that $\underline{\mathsf{curv}}(M,\mathsf{d})<0$. Choose $K<0$ such that $\underline{\mathsf{curv}}(M,\mathsf{d})<K$. Then there exist points $x_0,x_1,x_2,x_3\in M$ with

$$\sphericalangle_K(x_0;x_1,x_2)+\sphericalangle_K(x_0;x_2,x_3)+\sphericalangle_K(x_0;x_3,x_1)>2\pi.$$

Choose a point $z\in M$ 'far away' from the x_i, say $\mathsf{d}(z,x_i)\geqslant 3\mathsf{d}(x_i,x_j)$ for all $i,j=0,1,2,3$. (This is always possible, since M is a length space and the x_i for $i=1,2,3$ can be replaced by points x_i' lying arbitrarily close to x_0 on approximate geodesics connecting x_0 and x_i.)

For $t\in]0,1]$ and $i=0,1,2,3$ define $\mu_i:=t\delta_{x_i}+(1-t)\delta_z$. Then

$$\mathsf{d}_W^2(\mu_i,\mu_j)=t\mathsf{d}^2(x_i,x_j)$$

for all $i,j=0,1,2,3$ and thus, according to formula (2.4),

$$\sphericalangle_K(x_0;x_i,x_j)=\sphericalangle_{K/t}(\mu_0;\mu_i,\mu_j).$$

Therefore,

$$\sphericalangle_{K/t}(\mu_0;\mu_1,\mu_2)+\sphericalangle_{K/t}(\mu_0;\mu_2,\mu_3)+\sphericalangle_{K/t}(\mu_0;\mu_3,\mu_1)>2\pi$$

which implies

$$\underline{\mathsf{curv}}(\mathcal{P}_2(M),\mathsf{d}_W)<\frac{K}{t}.$$

Since the latter holds for the chosen $K<0$ and all arbitrarily small $t>0$, it proves the claim.

(iv$_\mathrm{d}$) Finally, it remains to prove that $\underline{\mathsf{curv}}(\mathcal{P}_2(M),\mathsf{d}_W)\leqslant 0$ if M has more than one point. Assume that $\underline{\mathsf{curv}}(\mathcal{P}_2(M),\mathsf{d}_W)\geqslant K$ for some $K>0$ and that $x_0,x_1\in M$ with $x_0\neq x_1$. Given $\varepsilon>0$ let y be an ε-midpoint of x_0 and x_1. Put $\nu_0=\delta_{x_0}, \nu_1=\delta_{x_1}, \mu=\delta_y$ and $\eta=\frac12\delta_{x_0}+\frac12\delta_{x_1}$. Then $\mathsf{d}_W(\nu_0,\nu_1)=\mathsf{d}(x_0,x_1)$, $\mathsf{d}_W(\eta,\mu)\geqslant\frac12\mathsf{d}(x_0,x_1)$, $\mathsf{d}_W(\nu_i,\eta)=\frac{1}{\sqrt2}\mathsf{d}(x_0,x_1)$ and $\mathsf{d}_W(\nu_i,\mu)\leqslant\frac12\mathsf{d}(x_0,x_1)+\varepsilon$ for $i=0,1$. In particular, μ is an ε-midpoint of ν_0 and ν_1.

Our curvature assumption on $\mathcal{P}_2(M)$ then implies (via quadruple comparison for $(\mu;\nu_0,\nu_1,\eta)$ or via triangle comparison for (ν_0,ν_1,η)) that

$$2\cos\left(\mathsf{d}_W(\nu_0,\nu_1)\frac{\sqrt{K}}{2}\right)\cos\big(\mathsf{d}_W(\eta,\mu)\sqrt{K}\,\big)\leqslant\cos\big(\mathsf{d}_W(\eta,\nu_0)\sqrt{K}\,\big)+\cos\big(\mathsf{d}_W(\eta,\nu_1)\sqrt{K}\,\big)+\varepsilon'$$

with some $\varepsilon'\to0$ as $\varepsilon\to0$. Therefore,

$$2\cos\left(\mathsf{d}(x_0,x_1)\frac{\sqrt{K}}{2}\right)\cos\left(\mathsf{d}(x_0,x_1)\frac{\sqrt{K}}{2}\right)\leqslant\cos\left(\mathsf{d}(x_0,x_1)\sqrt{\frac{K}{2}}\right)+\cos\left(\mathsf{d}(x_0,x_1)\sqrt{\frac{K}{2}}\right).$$

Now choosing $x_0,x_1\in M$ with sufficiently small $\mathsf{d}(x_0,x_1)$ leads to a contradiction. □

Let us recall that a *Markov kernel* on M is a map $Q: M\times\mathcal{B}(M)\to[0,1]$ (where $\mathcal{B}(M)$ denotes the Borel σ-algebra of M) with the following properties:

(i) for each $x\in M$ the map $Q(x,\cdot):\mathcal{B}(M)\to[0,1]$ is a probability measure on M, usually denoted by $Q(x,dy)$;

(ii) for each $A\in\mathcal{B}(M)$ the function $Q(\,\cdot\,,A): M\to[0,1]$ is measurable.

LEMMA 2.11. (i) *For each pair* $\mu, \nu \in \mathcal{P}_2(M)$ *there exists a coupling* q (*called* optimal coupling) *such that*

$$\mathsf{d}_W^2(\mu,\nu) = \int_{M\times M} \mathsf{d}^2(x,y)\, dq(x,y)$$

and there exist Markov kernels Q *and* Q' *on* M (optimal transport kernels) *such that*

$$dq(x,y) = Q(x,dy)\, d\mu(x) = Q'(y,dx)\, d\nu(x).$$

(*In general, neither* q *nor* Q *and* Q' *are unique.*)

(ii) *For each geodesic* $\Gamma\colon [0,1]\to\mathcal{P}_2(M)$, *each* $l\in\mathbf{N}$ *and each partition*

$$0 = t_0 < t_1 < \ldots < t_l = 1$$

there exists a probability measure $\hat{q}$ *on* M^{l+1} *with the following properties*:

- *the projection on the* i*-th factor is* $\Gamma(t_i)$ (*for all* $i=0,1,\ldots,l$);
- *for* $\hat{q}$*-a.e.* $(x_0,\ldots,x_l)\in M^{l+1}$ *and every* $i,j=0,1,\ldots,l$,

$$\mathsf{d}(x_i,x_j) = |t_i - t_j|\,\mathsf{d}(x_0,x_l). \tag{2.9}$$

In particular, for every pair $i,j\in\{0,1,\ldots,l\}$ *the projection on the* i*-th and* j*-th factor is an optimal coupling of* $\Gamma(t_i)$ *and* $\Gamma(t_j)$.

In the case $l=2$ *and* $t=\frac{1}{2}$, *equation* (2.9) *states that for* $\hat{q}$*-a.e.* $(x_0,x_1,x_2)\in M^3$ *the point* x_1 *is a midpoint of* x_0 *and* x_2.

(iii) *For each geodesic* $\Gamma\colon [0,1]\to\mathcal{P}_2(M)$ *there exists a probability measure* p *on* $\mathcal{G}(M)$, *the set of geodesics* $\gamma\colon [0,1]\to M$, *such that*

$$\int_M u(x)\, d\Gamma_t(x) = \int_{\mathcal{G}(M)} u(\gamma_t)\, dp(\gamma)$$

for all $t\in[0,1]$ *and all measurable* $u\colon M\to\mathbf{R}$. *For each pair* (s,t) *the joint distribution of* (γ_s,γ_t) *under* p *is an optimal coupling of* Γ_s *and* Γ_t.

(iv) *If* M *is a non-branching geodesic space then in the situation of* (ii) *for* $\hat{q}$*-a.e.* (x_0,x_1,x_2) *and* $(y_0,y_1,y_2)\in M^3$,

$$x_1 = y_1 \quad\Longrightarrow\quad (x_0,x_2) = (y_0,y_2).$$

Proof. (i) For the existence of an optimal coupling, see [45] or [16, 11.8.2].

The existence of optimal transport kernels is a straightforward application of disintegration of measures on Polish spaces (or of the existence of regular conditional probabilities), namely, Q is the disintegration of q with respect to μ.

(ii) We assume $l=2$ and $t=\frac{1}{2}$. (The general case follows by iterated application and appropriate modifications.)

Let q_1 be an optimal coupling of $\Gamma(0)$ and $\Gamma\left(\frac{1}{2}\right)$ and let q_2 be an optimal coupling of $\Gamma\left(\frac{1}{2}\right)$ and $\Gamma(1)$. Then there exists a probability measure $\hat{q}$ on $M\times M\times M$ such that its projection on the first two factors is q_1 and the projection on last two factors is q_2 [16, Section 11.8]. Hence, for $i=1,2,3$ the projection of $\hat{q}$ on the ith factor is $\Gamma((i-1)/2)$ and for $i=1,2$,

$$\mathsf{d}_W^2\left(\Gamma\left(\frac{i-1}{2}\right),\Gamma\left(\frac{i}{2}\right)\right)=\int_{M^3}\mathsf{d}^2(x_{i-1},x_i)\,d\hat{q}(x_0,x_1,x_2).$$

Then

$$\begin{aligned}\mathsf{d}_W(\Gamma(0),\Gamma(1))&\leqslant\left[\int\mathsf{d}^2(x_0,x_2)\,d\hat{q}(x_0,x_1,x_2)\right]^{1/2}\\&\overset{(*)}{\leqslant}\left[\int[\mathsf{d}(x_0,x_1)+\mathsf{d}(x_1,x_2)]^2\,d\hat{q}(x_0,x_1,x_2)\right]^{1/2}\\&\overset{(**)}{\leqslant}\left[\int\mathsf{d}^2(x_0,x_1)\,d\hat{q}(x_0,x_1,x_2)\right]^{1/2}+\left[\int\mathsf{d}^2(x_1,x_2)\,d\hat{q}(x_0,x_1,x_2)\right]^{1/2}\\&=\mathsf{d}_W\left(\Gamma(0),\Gamma\left(\tfrac{1}{2}\right)\right)+\mathsf{d}_W\left(\Gamma\left(\tfrac{1}{2}\right),\Gamma(1)\right).\end{aligned}$$

Since $\Gamma\left(\frac{1}{2}\right)$ is a midpoint of $\Gamma(0)$ and $\Gamma(1)$, the previous inequalities $(*)$ and $(**)$ have to be equalities. From equality in $(*)$ we conclude that $\hat{q}$-almost surely the point x_1 lies on some geodesic connecting x_0 and x_2. Equality in $(**)$ implies that $\hat{q}$-almost surely the point x_1 is a midpoint of x_0 and x_2.

(iii) For each Γ, this is an immediate corollary to (ii) which yields the claim for each finite dimensional distribution (i.e. evaluation at finitely many fixed times) of Γ.

(iv) Let $\eta=\Gamma\left(\frac{1}{2}\right)$ be the distribution of the midpoints and let Q be a disintegration of $\hat{q}$ with respect to η, i.e.

$$d\hat{q}(x,z,y)=Q(z,d(x,y))\,d\eta(z).$$

We have to prove that for η-a.e. $z\in M$ the probability measure $Q(z,\cdot\,)$ is a Dirac measure (sitting on some $(x,y)\in M\times M$). Denote the marginals of $Q(z,\cdot\,)$ by $p_1(z,\cdot\,)$ and $p_2(z,\cdot\,)$. Then

$$\begin{aligned}\int\mathsf{d}^2(x,y)\,Q(z,d(x,y))&=\int[2\mathsf{d}^2(x,z)+2\mathsf{d}^2(z,y)]\,Q(z,d(x,y))\\&=\iint[2\mathsf{d}^2(x,z)+2\mathsf{d}^2(z,y)]\,p_1(z,dx)\,p_2(z,dy)\\&\overset{(***)}{\geqslant}\iint\mathsf{d}^2(x,y)\,p_1(z,dx)\,p_2(z,dy).\end{aligned}$$

The optimality of $\hat{q}$ implies that for η-a.e. $z \in M$ the measure $Q(z, \cdot\,)$ is an optimal coupling of $p_1(z, \cdot\,)$ and $p_2(z, \cdot\,)$. Hence, there has to be equality in $(***)$ which in turn implies that for $p_1(z, \cdot\,)$-a.e. $x \in M$ and $p_2(z, \cdot\,)$-a.e. $y \in M$ the point z is a midpoint of x and y. Since M is non-branching this implies that both $p_1(z, \cdot\,)$ and $p_2(z, \cdot\,)$ are Dirac measures. Thus $Q(z, \cdot\,)$ is also a Dirac measure. This proves the claim. □

Remark 2.12. (i) Couplings q of μ and ν are also called *transportation plans* from μ to ν. If μ is the distribution of locations at which a product is produced and ν is the distribution of locations where it is consumed, then each coupling q of μ and ν gives a plan how to transport the products to the consumer. More precisely, for each x the kernel $Q(x, dy)$ determines how to distribute goods produced at the location x to various consumers at location y.

(ii) The interpretation of Lemma 2.11 (ii), (iii) is that for each geodesic in $\mathcal{P}_2(M)$ the mass is transported along geodesics of the underlying space M. Lemma 2.11 (iv) states that the paths of optimal mass transportation do not cross each other halfway.

(iii) If M is a complete Riemannian manifold with Riemannian volume m then for each pair $\mu, \nu \in \mathcal{P}_2(M)$ with $\mu \ll m$ there exists an *optimal transport map* $F_1: M \to M$ such that $dq(x, y) = Q(x, dy)\, d\mu(x)$, with

$$Q(x, dy) = d\delta_{F_1(x)}(y),$$

is the unique optimal coupling of μ and ν.

More precisely, there exists a function $\varphi: M \to \mathbf{R}$ such that for μ-a.e. $x \in M$ and $t \in [0, 1]$,

$$F_t(x) = \exp_x(-t \nabla \varphi(x))$$

exists and the unique geodesic Γ in $\mathcal{P}_2(M)$ connecting $\mu = \Gamma(0)$ and $\nu = \Gamma(1)$ is given by

$$\Gamma(t) := (F_t)_* \mu,$$

the push-forward of μ under F_t; see [13].

(iv) Besides the L_2-Wasserstein distance one can consider more generally the L_p-Wasserstein distance for any $p \in [1, \infty[$; see e.g. [55]. Many of the previous properties also hold in this more general case. However, all ‘curvature’ concepts like in Proposition 2.10 (iv) require $p = 2$.

3. Metric measure spaces

3.1. The metric D

Throughout this paper, a *metric measure space* will always be a triple (M, d, m), where

- (M, d) is a complete separable metric space,
- m is a measure on $(M, \mathcal{B}(M))$ which is locally finite in the sense that $m(B_r(x))<\infty$ for all $x \in M$ and all sufficiently small $r>0$.

A metric measure space (M, d, m) is called *normalized* if and only if $m(M)=1$. It is called *compact* or *locally compact* or *geodesic* if and only if the metric space (M, d) is compact or locally compact or geodesic, respectively.

Two metric measure spaces (M, d, m) and (M', d', m') are called *isomorphic* if and only if there exists an isometry $\psi: M_0 \to M_0'$ between the supports $M_0 := \mathrm{supp}[m] \subset M$ and $M_0' := \mathrm{supp}[m'] \subset M'$ such that

$$\psi_* m = m'.$$

The *variance* of a metric measure space (M, d, m) is defined as

$$\mathrm{Var}(M, \mathsf{d}, m) = \inf \int_{M'} \mathsf{d}'^2(z, x)\, dm'(x), \tag{3.1}$$

where the infimum is taken over all metric measure spaces (M', d', m') which are isomorphic to (M, d, m) and over all $z \in M'$. Note that a normalized metric measure space (M, d, m) has finite variance if and only if

$$\int_M \mathsf{d}^2(z, x)\, dm(x) < \infty \tag{3.2}$$

for some (hence all) $z \in M$. The *diameter* of a metric measure space (M, d, m) is defined as the diameter of the metric space $(\mathrm{supp}[m], d)$:

$$\mathrm{diam}(M, \mathsf{d}, m) = \sup\{d(x, y) : x, y \in \mathrm{supp}[m]\}.$$

Example 3.1. Let $M=\mathbf{R}^2$ with Euclidean distance d and $m=\frac{1}{3}(\delta_{x_1}+\delta_{x_2}+\delta_{x_3})$, where x_1, x_2 and x_3 are the vertices of an equilateral triangle of sidelength 1. Then

$$\mathrm{Var}(M, \mathsf{d}, m) = \frac{1}{4} \quad \text{whereas} \quad \inf_{z \in M} \int \mathsf{d}^2(z, x)\, dm(x) = \frac{1}{3}.$$

(*Hint*: embed $\mathrm{supp}[m]$ isometrically into a graph or into a hyperbolic space with curvature close to $-\infty$.)

The family of all isomorphism classes of metric measure spaces will be denoted by $\mathbf{X}$. For each $\lambda\in\mathbf{R}_+$, let $\mathbf{X}_\lambda$ denote the family of isomorphism classes of metric measure spaces (M,d,m) with finite variances and total mass $m(M)=\lambda$. Moreover, for $L\in\mathbf{R}_+$, let $\mathbf{X}_\lambda(L)$ denote the family of isomorphism classes of metric measure spaces (M,d,m) with diameter $\leqslant L$ and total mass $m(M)=\lambda$. If $\lambda L\neq 0$, then the map

$$(M,\mathsf{d},m)\longmapsto(M,L\mathsf{d},\lambda m)$$

defines a bijection between $\mathbf{X}_1$ and $\mathbf{X}_\lambda$ and also a bijection between $\mathbf{X}_1(1)$ and $\mathbf{X}_\lambda(L)$.

For $\lambda>0$, the family $\mathbf{X}_\lambda$ contains a unique element with $\mathrm{Var}(M,\mathsf{d},m)=0$, namely, $m=\lambda\delta_o$ for some $o\in M$. Here a priori M is an arbitrary non-empty set. But without restriction it contains just one point, say $M=\{o\}$. The family $\mathbf{X}_0$ is pathological: it contains only one element, the 'empty space'.

Definition 3.2. (i) Given two metric measure spaces (M,d,m) and (M',d',m'), we say that a measure q on the product space $M\times M'$ is a *coupling* of m and m' if and only if

$$q(A\times M')=m(A)\quad\text{and}\quad q(M\times A')=m'(A')\tag{3.3}$$

for all measurable sets $A\subset M$ and $A'\subset M'$. We say that a pseudo-metric $\hat{\mathsf{d}}$ on the disjoint union $M\sqcup M'$ is a *coupling of* d *and* d' if and only if

$$\hat{\mathsf{d}}(x,y)=\mathsf{d}(x,y)\quad\text{and}\quad\hat{\mathsf{d}}(x',y')=\mathsf{d}'(x',y')\tag{3.4}$$

for all $x,y\in\mathrm{supp}[m]\subset M$ and all $x',y'\in\mathrm{supp}[m']\subset M'$.

(ii) We define the L_2-*transportation distance* $\mathbf{D}$ between two metric measure spaces by

$$\mathbf{D}((M,\mathsf{d},m),(M',\mathsf{d}',m'))$$
$$=\inf\left\{\left(\int_{M\times M'}\hat{\mathsf{d}}^2(x,y)\,dq(x,y)\right)^{1/2}:\ \begin{array}{l}\hat{\mathsf{d}}\text{ is a coupling of }\mathsf{d}\text{ and }\mathsf{d}',\\ q\text{ is a coupling of }m\text{ and }m'\end{array}\right\}.$$

Note that the integrals involved in the definition of $\mathbf{D}$ are well-defined, since each coupling $\hat{\mathsf{d}}$ is a function on

$$(M\sqcup M')\times(M\sqcup M')=(M\times M)\sqcup(M\times M')\sqcup(M'\times M)\sqcup(M'\times M')$$

and each coupling q is a measure on $M\times M'$.

In a similar way, we can also define the L_p-transportation distance on the space of metric measure spaces for any $p\in[1,\infty[$. Many of the following properties will also hold in this more general case. However, for our purpose, the L_2-transportation distance seems to be most convenient. For simplicity, we therefore restrict the discussion to the case $p=2$.

LEMMA 3.3. (i) *For each pair of normalized metric measure spaces* (M,d,m) *and* (M',d',m') *there exists a coupling* q_* *of* m *and* m' *and a coupling* $\hat{\mathsf{d}}_*$ *of* d *and* d' *such that*

$$\mathbf{D}((M,\mathsf{d},m),(M',\mathsf{d}',m'))=\left(\int_{M\times M'}\hat{\mathsf{d}}_*^2(x,y)\,dq_*(x,y)\right)^{1/2}.$$

That is, q_* *and* $\hat{\mathsf{d}}_*$ *are* optimal couplings.

(ii) *In the definition of the distance* $\mathbf{D}$ *we may restrict ourselves to take the infimum over all* complete separable metrics $\hat{\mathsf{d}}$ *on* $M\sqcup M'$ *which are couplings of* d *and* d'.

In general, however, the optimal coupling $\hat{\mathsf{d}}_*$ *from part* (i) *will only be a pseudometric. For instance, if* (M,d,m) *and* (M',d',m') *are isomorphic then* $\hat{\mathsf{d}}_*(x,\psi(x))=0$ *for* m*-a.e.* $x\in M$.

(iii) *Moreover,*

$$\mathbf{D}((M,\mathsf{d},m),(M',\mathsf{d}',m'))=\inf\hat{\mathsf{d}}_W(\psi_*m,\psi'_*m'),\tag{3.5}$$

where the infimum is taken over all metric spaces $(\widehat{M},\hat{\mathsf{d}})$ *with isometric embeddings* $\psi:M_0\hookrightarrow\widehat{M}$ *and* $\psi':M_0'\hookrightarrow\widehat{M}$ *of the supports* M_0 *and* M_0' *of* m *and* m'*, respectively. Here* $\hat{\mathsf{d}}_W$ *denotes the* L_2*-Wasserstein distance for measures on* $\widehat{M}$ *as introduced in Section* 2. *In other words,*

$$\mathbf{D}((M,\mathsf{d},m),(M',\mathsf{d}',m'))=\inf\left(\int_{M\times M'}\hat{\mathsf{d}}^2(\psi(x),\psi'(x'))\,dq(x,x')\right)^{1/2},\tag{3.6}$$

where the infimum now is taken over all metric spaces $(\widehat{M},\hat{\mathsf{d}})$ *with isometric embeddings* $\psi:M_0\hookrightarrow\widehat{M}$ *and* $\psi':M_0'\hookrightarrow\widehat{M}$ *and over all couplings* q *of* m *and* m'.

Proof. (i) Let normalized metric measure spaces (M,d,m) and (M',d',m') be given, as well as sequences $(\hat{\mathsf{d}}_n)_{n\in\mathbf{N}}$ and $(q_n)_{n\in\mathbf{N}}$ of couplings with

$$\int_{M\times M'}\hat{\mathsf{d}}_n^2(x,x')\,dq_n(x,x')\leqslant\mathbf{D}^2((M,\mathsf{d},m),(M',\mathsf{d}',m'))+\frac{1}{n}.\tag{3.7}$$

Assume for simplicity that m and m' have full support. The q_n, $n\in\mathbf{N}$, are probability measures on the complete separable metric space $\widetilde{M}:=M\times M'$. For each $\varepsilon>0$ there exist compact subsets $K\subset M$ and $K'\subset M'$ such that $m(M\setminus K)\leqslant\varepsilon$ and $m'(M'\setminus K')\leqslant\varepsilon$. Put $\widetilde{K}:=K\times K'$. Then for all $n\in\mathbf{N}$,

$$q_n(\widetilde{M}\setminus\widetilde{K})\leqslant 2\varepsilon.$$

That is, the family $(q_n)_{n\in\mathbf{N}}$ is tight and by Prohorov's theorem there exists a subsequence converging to some probability measure q_* on $\widetilde{M}$. Of course, q_* is again a coupling of m and m'.

Now consider the family of functions $\hat{\mathsf{d}}_n$, $n\in\mathbf{N}$, on $\widetilde{M}$. By the triangle inequality

$$\left|\hat{\mathsf{d}}_n(x,x')-\hat{\mathsf{d}}_n(y,y')\right|\leqslant \mathsf{d}(x,y)+\mathsf{d}'(x',y') =: \tilde{\mathsf{d}}((x,x'),(y,y'))$$

for all $x,y\in M$ and $x',y'\in M'$. That is, the family $(\hat{\mathsf{d}}_n)_{n\in\mathbf{N}}$ is uniformly equicontinuous on $\widetilde{M}$ (equipped e.g. with the metric $\tilde{\mathsf{d}}$ or with the usual product metric).

In order to apply Ascoli's theorem we have to prove that for some point $(o,o')\in\widetilde{M}$ the sequence $(\hat{\mathsf{d}}_n(o,o'))_{n\in\mathbf{N}}$ is bounded. Choose a point (o,o') in the support of the measure q_*. Then for some constants $C,\delta,\varepsilon>0$ and all sufficiently large n,

$$\begin{aligned}
C &\geqslant \int_{M\times M'} \hat{\mathsf{d}}_n(x,x')\,dq_n(x,x')\\
&\geqslant \int_{B_\varepsilon(o)\times B_\varepsilon(o')} [\hat{\mathsf{d}}_n(o,o')-2\varepsilon]\,dq_n(x,x')\\
&\geqslant [\hat{\mathsf{d}}_n(o,o')-2\varepsilon]\cdot[q_*(B_\varepsilon(o)\times B_\varepsilon(o'))-\delta]\\
&\geqslant \frac{1}{C}\big[\hat{\mathsf{d}}_n(o,o')-2\varepsilon\big].
\end{aligned}$$

Hence, the sequence $(\hat{\mathsf{d}}_n)_{n\in\mathbf{N}}$ has a subsequence uniformly converging to some function $\hat{\mathsf{d}}_*$ on $\widetilde{M}$. Obviously, $\hat{\mathsf{d}}_*$ is again a coupling of the metrics d and d'.

Finally, uniform convergence of (an appropriate subsequence of) the $(\hat{\mathsf{d}}_n)_{n\in\mathbf{N}}$, continuity of $\hat{\mathsf{d}}_*$ and weak convergence of (an appropriate subsequence of) the $(q_n)_{n\in\mathbf{N}}$ allows one to pass to the limit in (3.7). That is,

$$\int_{M\times M'} \hat{\mathsf{d}}_*^2(x,x')\,dq_*(x,x')\leqslant \mathbf{D}^2((M,\mathsf{d},m),(M',\mathsf{d}',m')).$$

(ii) Given any (pseudo-metric) coupling $\hat{\mathsf{d}}$ of d and d' and any $\varepsilon>0$ we obtain a complete separable metric $\hat{\mathsf{d}}_\varepsilon$ which is a coupling of d and d' as follows:

$$\hat{\mathsf{d}}_\varepsilon=\begin{cases}\mathsf{d}, & \text{on } (M\times M)\sqcup(M'\times M'),\\ \mathsf{d}+\varepsilon, & \text{on } (M\times M')\sqcup(M'\times M).\end{cases}$$

(iii) The set $\widehat{M}$ can always be chosen as the disjoint union of M_0 and M_0' (the supports of m and m'), i.e.

$$\widehat{M}=M_0\sqcup M_0'$$

and ψ and ψ' can be chosen as identities. Hence,

$$\mathbf{D}((M,\mathsf{d},m),(M',\mathsf{d}',m')) \;=\; \inf_{\hat{\mathsf{d}}} \hat{\mathsf{d}}_W(m,m'),$$

where the infimum is taken over all metrics (or, equivalently, over all pseudo-metrics) on $M_0\sqcup M_0'$ whose restrictions coincide with d on M_0 and with d' on M_0'. □

Let us summarize some elementary properties of $\mathbf{D}$.

LEMMA 3.4. (i) *If* $(M,\mathsf{d})=(M',\mathsf{d}')$ *then* $\mathbf{D}((M,\mathsf{d},m),(M,\mathsf{d},m'))\leqslant\mathsf{d}_W(m,m')$. *In general, there will be no equality.*

(ii) *If* $m(M)\neq m'(M')$ *then* $\mathbf{D}((M,\mathsf{d},m),(M',\mathsf{d}',m'))=+\infty$.

(iii) *For all* $\alpha,\beta\in\mathbf{R}_+$

$$\mathbf{D}((M,\alpha\mathsf{d},\beta m),(M',\alpha\mathsf{d}',\beta m'))=\alpha\sqrt{\beta}\cdot\mathbf{D}((M,\mathsf{d},m),(M',\mathsf{d}',m')).$$

Now let us concentrate on normalized metric measure spaces.

LEMMA 3.5. (i) *If* $m(M)=1$ *and* $m'=\delta_o$ *for some* $o\in M'$ *then*

$$\mathbf{D}^2((M,\mathsf{d},m),(M',\mathsf{d}',m'))=\mathrm{Var}(M,\mathsf{d},m).$$

(ii) *The family* $\mathbf{X}_{1,*}$ *of isomorphism classes of* (M,d,m) *with finite supports* M_0, *say* $\{x_1,...,x_n\}$, *and uniform distribution* $m=(1/n)\sum_{i=1}^n\delta_{x_i}$ (normalized configurations) *is dense in* $\mathbf{X}_1$.

(iii) *For each* $(M,\mathsf{d},m)\in\mathbf{X}_1$ *with finite diameter, let* $X_1,X_2,...$ *be an independent sequence of random variables* $X_i:\Omega\to M$ (*defined on some probability space* Ω *with values in* M) *with distribution* m *and let*

$$m_n(\omega,\cdot\,):=\frac{1}{n}\sum_{i=1}^n\delta_{X_i(\omega)}$$

be their empirical distributions. Then for m*-a.e.* $\omega\in\Omega$,

$$(M,\mathsf{d},m_n(\omega,\cdot\,))\to(M,\mathsf{d},m)$$

in $(\mathbf{X}_1,\mathbf{D})$ *as* $n\to\infty$.

(iv) *If* $m=(1/n)\sum_{i=1}^n\delta_{x_i}$ *and* $m'=(1/n)\sum_{i=1}^n\delta_{x'_i}$ *then*

$$\mathbf{D}((M,\mathsf{d},m),(M',\mathsf{d}',m'))\leqslant\sup_{i,j}|\mathsf{d}_{ij}-\mathsf{d}'_{ij}|,$$

where $\mathsf{d}_{ij}:=\mathsf{d}(x_i,x_j)$ *and* $\mathsf{d}'_{ij}:=\mathsf{d}'(x'_i,x'_j)$.

Proof. (i) This is obvious.

(ii) Given $(M,\mathsf{d},m)\in\mathbf{X}_1$, we have $m\in\mathcal{P}_2(M,\mathsf{d})$ by inequality (3.2). Then, by Proposition 2.10 (i), for all $\varepsilon>0$ there exist $n\in\mathbf{N}$ and $x_1,...,x_n\in M$ such that $\mathsf{d}_W(m,\overline{m})\leqslant\varepsilon$, where $\overline{m}:=(1/n)\sum_{i=1}^n\delta_{x_i}$. Hence, $(M,\mathsf{d},\overline{m})\in\mathbf{X}_{1,*}$ and

$$\mathbf{D}((M,\mathsf{d},m),(M,\mathsf{d},\overline{m}))\leqslant\mathsf{d}_W(m,\overline{m})\leqslant\varepsilon.$$

(iii) This follows from the empirical law of large numbers or Varadarajan's theorem, e.g. [16, Theorem 11.4.1].

(iv) Assume (without restriction) that $M=\{x_1,...,x_n\}$, $M'=\{x'_1,...,x'_n\}$, and $|\mathsf{d}_{ij}-\mathsf{d}'_{ij}|\leqslant\varepsilon$ for all i,j (with d_{ij} and d'_{ij} as above). Define $\hat{\mathsf{d}}$ on $M\times M'$ by

$$\hat{\mathsf{d}}(x_i,x'_j):=\inf_{k\in\{1,...,n\}}[\mathsf{d}(x_i,x_k)+\mathsf{d}'(x'_k,x'_j)]+\varepsilon$$

and analogously on $M'\times M$. As usual, put $\hat{\mathsf{d}}:=\mathsf{d}$ on $M\times M$ and $\hat{\mathsf{d}}:=\mathsf{d}'$ on $M'\times M'$. Moreover, put

$$m=\frac{1}{n}\sum_{i=1}^{n}\delta_{(x_i,x'_i)}.$$

Then $\hat{\mathsf{d}}$ is a coupling of d and d', and q is a coupling of m and m'. Thus

$$\mathbf{D}^2((M,\mathsf{d},m),(M',\mathsf{d}',m'))\leqslant\int_{M\times M'}\hat{\mathsf{d}}^2(x,y)\,dq(x,y)=\varepsilon^2. \qquad \square$$

THEOREM 3.6. $(\mathbf{X}_1,\mathbf{D})$ *is a complete separable length metric space.*

Proof. (i) Clearly, $\mathbf{D}$ is *well-defined* and *symmetric* on $\mathbf{X}_1\times\mathbf{X}_1$ with values in $\mathbf{R}_+$.

(ii) Separability and completeness will follow from Lemma 3.7 below, at least under uniform bounds for the diameter. For sake of completeness, we give a direct proof.

(iii) According to Lemma 3.5 (ii), *separability* of $\mathbf{X}_1$ will follow from separability of $\mathbf{X}_{1,*}$. The latter is the disjoint union $\bigsqcup_{n\in\mathbf{N}}\widetilde{\mathcal{K}}(n)$, where

$$\widetilde{\mathcal{K}}(n):=\{(M,\mathsf{d},m)\in\mathbf{X}_{1,*}:\operatorname{supp}[m]\text{ has } n \text{ points}\}.$$

But $\widetilde{\mathcal{K}}(n)$ can be identified with the set $\mathcal{K}(n)$ of all $D=(D_{ij})_{i,j}\in\mathbf{R}_+^{n\times n}$ satisfying

$$D_{ij}=D_{ji},\quad D_{ij}+D_{jk}\geqslant D_{ik}\quad\text{and}\quad D_{ij}=0\iff i=j$$

for all $i,j,k\in\{1,...,n\}$. Now each of the $\mathcal{K}(n)$ is separable (as a subset of $\mathbf{R}^{n\times n}$), hence, $\widetilde{\mathcal{K}}(n)$ is separable (Lemma 3.5 (iv)) and thus finally $\mathbf{X}_{1,*}$ is separable.

(iv) In order to prove the *triangle inequality* consider three metric measure spaces $(M_i,\mathsf{d}_i,m_i)\in\mathbf{X}_1$, $i=1,2,3$. Without restriction, we may assume $M_i=\operatorname{supp}[m_i]$ for $i=1,2,3$. Then for each $\varepsilon>0$ there exist a complete separable metric d_{12} on $M_1\sqcup M_2$ and a complete separable metric d_{23} on $M_2\sqcup M_3$ such that

$$\mathbf{D}((M_1,\mathsf{d}_1,m_1),(M_2,\mathsf{d}_2,m_2))\geqslant\mathsf{d}_{12}^W(m_1,m_2)-\varepsilon,$$
$$\mathbf{D}((M_2,\mathsf{d}_2,m_2),(M_3,\mathsf{d}_3,m_3))\geqslant\mathsf{d}_{23}^W(m_2,m_3)-\varepsilon$$

and d_{ij} restricted to M_i coincides with d_i, restricted to M_j coincides with d_j for $(i,j)=(1,2)$ or $(2,3)$. (Here, for typographical reasons, we use not a lower but an upper index to indicate the Wasserstein metric derived from a given metric.) Now define d on $M\times M$ with $M:=M_1\sqcup M_2\sqcup M_3$ by

$$\mathsf{d}(x,y)=\begin{cases} \mathsf{d}_{12}(x,y), & \text{if } x,y\in M_1\sqcup M_2,\\ \mathsf{d}_{23}(x,y), & \text{if } x,y\in M_2\sqcup M_3,\\ \inf_{z\in M_2}[\mathsf{d}_{12}(x,z)+\mathsf{d}_{23}(z,y)], & \text{if } x\in M_1 \text{ and } y\in M_3,\\ \inf_{z\in M_2}[\mathsf{d}_{23}(x,z)+\mathsf{d}_{12}(z,y)], & \text{if } x\in M_3 \text{ and } y\in M_1. \end{cases}$$

Obviously, d is a complete separable metric on M and, restricted to M_i, it coincides with d_i (for each $i=1,2,3$). Then by the triangle inequality for d^W (Proposition 2.10 (i))

$$\begin{aligned}\mathbf{D}((M_1,\mathsf{d}_1,m_1),(M_3,\mathsf{d}_3,m_3)) &\leqslant \mathsf{d}^W(m_1,m_3)\\ &\leqslant \mathsf{d}^W(m_1,m_2)+\mathsf{d}^W(m_2,m_3)\\ &= \mathsf{d}_{12}^W(m_1,m_2)+\mathsf{d}_{23}^W(m_2,m_3)\\ &\leqslant \mathbf{D}((M_1,\mathsf{d}_1,m_1),(M_2,\mathsf{d}_2,m_2))\\ &\qquad +\mathbf{D}((M_2,\mathsf{d}_2,m_2),(M_3,\mathsf{d}_3,m_3))+2\varepsilon.\end{aligned}$$

This proves the claim.

(v) In order to prove *completeness* let $((M_n,\mathsf{d}_n,m_n))_{n\in\mathbf{N}}$ be a Cauchy sequence in $(\mathbf{X}_1,\mathbf{D})$. Let us choose a subsequence such that

$$\mathbf{D}((M_{n_k},\mathsf{d}_{n_k},m_{n_k}),(M_{n_{k+1}},\mathsf{d}_{n_{k+1}},m_{n_{k+1}}))\leqslant 2^{-k-1}$$

for all $k\in\mathbf{N}$. Then there exist a coupling $\hat{\mathsf{d}}_{k+1}$ of d_{n_k} and $\mathsf{d}_{n_{k+1}}$, and a coupling $\hat{q}_{k+1}$ of m_{n_k} and $m_{n_{k+1}}$ such that

$$\left(\int \hat{\mathsf{d}}^2_{k+1}(x,y)\,d\hat{q}_{k+1}(x,y)\right)^{1/2}\leqslant 2^{-k}.$$

Without restriction $\hat{\mathsf{d}}_{k+1}$ is a complete separable metric. Let us define recursively a sequence of complete separable metric spaces (M'_k,d'_k) as follows: $(M'_1,\mathsf{d}'_1):=(M_{n_1},\mathsf{d}_{n_1})$ and $M'_{k+1}=M'_k\sqcup M_{n_{k+1}}/\sim$ with $x\sim y$ if and only if $\mathsf{d}'_{k+1}(x,y)=0$, where

$$\mathsf{d}'_{k+1}(x,y)=\begin{cases} \mathsf{d}'_k(x,y), & \text{if } x,y\in M'_k,\\ \hat{\mathsf{d}}_{k+1}(x,y), & \text{if } x,y\in M_{n_k}\sqcup M_{n_{k+1}},\\ \inf_{z\in M_{n_k}}[\mathsf{d}'_k(x,z)+\hat{\mathsf{d}}_{k+1}(z,y)], & \text{if } x\in M'_k \text{ and } y\in M_{n_k}\sqcup M_{n_{k+1}}. \end{cases}$$

This way, $((M'_k, \mathsf{d}'_k))_{k\in\mathbf{N}}$ is a sequence of complete separable metric spaces with $M_{n_k}\subset M'_k$ and $M'_k\subset M'_{k+l}$ for all k and l. Hence, $M'=\bigcup_{k=1}^\infty M'_k$ is naturally equipped with a metric $\mathsf{d}'=\lim_{k\to\infty}\mathsf{d}'_k$.

Let (M,d) be the completion of (M',d'). Then $(M_{n_k},\mathsf{d}_{n_k})$ is isometrically embedded in (M,d) for each $k\in\mathbf{N}$ and the measure m_{n_k} on M_{n_k} defines a push-forward measure $\overline{m}_{n_k}$ on M. By construction

$$\mathsf{d}_W(\overline{m}_{n_k},\overline{m}_{n_{k+1}})\leqslant\left(\int\hat{\mathsf{d}}^2_{k+1}(x,y)\,d\hat{q}_{k+1}(x,y)\right)^{1/2}\leqslant 2^{-k}$$

for all $k\in\mathbf{N}$. Hence, $(\overline{m}_{n_k})_{k\in\mathbf{N}}$ is a Cauchy sequence in $(\mathcal{P}_2(M),\mathsf{d}_W)$. According to Proposition 2.10, the latter is complete. That is, there exists a probability measure m on (M,d) such that

$$\mathbf{D}((M_{n_k},\mathsf{d}_{n_k},m_{n_k}),(M,\mathsf{d},m))\leqslant\mathsf{d}_W(\overline{m}_{n_k},m)\to 0$$

as $k\to\infty$. This in turn implies that

$$\mathbf{D}((M_n,\mathsf{d}_n,m_n),(M,\mathsf{d},m))\to 0$$

as $n\to\infty$ which proves the claim.

(vi) In order to see that $(\mathbf{X}_1,\mathbf{D})$ is a *length space* it suffices to prove that each pair of normalized configurations is connected by a geodesic. Now, let

$$(M_0,\mathsf{d}_0,m_0),(M_1,\mathsf{d}_1,m_1)\in\mathbf{X}_{1,*}$$

be given. Without restriction $M_0=\{(0,1),...,(0,n)\}$ and $M_1=\{(1,1),...,(1,n)\}$ for some $n\in\mathbf{N}$ and m_0,m_1 being uniform distributions. For each $t\in[0,1]$ we define a metric measure space $(M_t,\mathsf{d}_t,m_t))$ by $M_t:=\{(t,1),...,(t,n)\}$, $m_t:=(1/n)\sum_{i=1}^n\delta_{(t,i)}$ and

$$\mathsf{d}_t((t,i),(t,j))=(1-t)\,\mathsf{d}_0((0,i),(0,j))+t\,\mathsf{d}_1((1,i),(1,j)).$$

Let q_* and $\hat{\mathsf{d}}_*$ be a pair of optimal couplings. Again without restriction we may assume that $q=(1/n)\sum_{i=1}^n\delta_{((0,i),(1,i))}$. Then

$$\mathbf{D}((M_0,\mathsf{d}_0,m_0),(M_1,\mathsf{d}_1,m_1))=\left(\frac{1}{n}\sum_{i=1}^n\varrho_i^2\right)^{1/2}$$

with $\varrho_i:=\hat{\mathsf{d}}_*((0,i),(1,i))$. By the triangle inequality for $\hat{\mathsf{d}}_*$,

$$|\mathsf{d}_0((0,i),(0,j))-\mathsf{d}_1((1,i),(1,j))|\leqslant\varrho_i+\varrho_j$$

for all i and j. For all $s,t\in[0,1]$ we define a coupling $\hat{\mathsf{d}}_{s,t}$ of d_s and d_t by

$$\hat{\mathsf{d}}_{s,t}((s,i),(t,j))=\inf_k[\mathsf{d}_s((s,i),(s,k))+|s-t|\varrho_k+\mathsf{d}_t((t,k),(t,j))].$$

In particular $\hat{\mathsf{d}}_{s,t}((s,i)(t,i))=|s-t|\varrho_i$ for all i and thus

$$\mathbf{D}((M_s,\mathsf{d}_s,m_s),(M_t,\mathsf{d}_t,m_t))\leqslant|s-t|\mathbf{D}((M_0,\mathsf{d}_0,m_0),(M_1,\mathsf{d}_1,m_1)).$$

(vii) *Non-degeneracy* of $\mathbf{D}$ will follow from the corresponding property of Gromov's metric $\underline{\square}_1$ together with the following Lemma 3.7. □

LEMMA 3.7. *For metric measure spaces with diameter* $\leqslant L$*, the metric* $\mathbf{D}$ *is equivalent to Gromov's metric* $\underline{\square}_1$ [22, $3\frac{1}{2}$.12]:

$$\left(\tfrac{1}{2}\underline{\square}_1\right)^{3/2}\leqslant\mathbf{D}\leqslant\left(L+\tfrac{1}{4}\right)\cdot\underline{\square}_1^{1/2}.$$

The lower bound also holds if $L=\infty$.

Proof. In order to prove the lower estimate, let normalized metric measure spaces (M,d,m) and (M',d',m') be given with $\mathbf{D}((M,\mathsf{d},m),(M',\mathsf{d}',m'))<\varepsilon^{3/2}$ for some $\varepsilon>0$. Then for some metric $\hat{\mathsf{d}}$ on $M\sqcup M'$ extending d and d' and for some coupling q of m and m',

$$\int\hat{\mathsf{d}}^2(x,x')\,dq(x,x')<\varepsilon^3.$$

Hence, $q(\{(x,x')\in M\times M':\hat{\mathsf{d}}(x,x')\geqslant\varepsilon\})<\varepsilon$. Therefore there exists a measurable map $\Phi:[0,1[\to M\times M'$ such that $\Phi_*\lambda=q$, where λ denotes the Lebesgue measure on $[0,1[$ (*parametrization of* q) and there exists a measurable set $X_\varepsilon\subset[0,1[$ with $\lambda(X_\varepsilon)<\varepsilon$ such that for all $t\in[0,1[\setminus X_\varepsilon$,

$$\hat{\mathsf{d}}(\Phi(t))<\varepsilon. \tag{3.8}$$

If we write $\Phi(t)=(\varphi(t),\varphi'(t))$ with $\varphi:[0,1[\to M$ and $\varphi':[0,1[\to M'$ then $\varphi_*\lambda=m$ and $\varphi'_*\lambda=m'$. Moreover, for all $s,t\in[0,1[\setminus X_\varepsilon$,

$$|\mathsf{d}(\varphi(s),\varphi(t))-\mathsf{d}'(\varphi'(s),\varphi'(t))|\leqslant\hat{\mathsf{d}}(\varphi(s),\varphi'(s))+\hat{\mathsf{d}}(\varphi(t),\varphi'(t))=\hat{\mathsf{d}}(\Phi(s))+\hat{\mathsf{d}}(\Phi(t))<2\varepsilon,$$

according to (3.8). This yields

$$\underline{\square}_1((M,\mathsf{d},m),(M',\mathsf{d}',m'))<2\varepsilon$$

which proves the lower estimate.

For the upper estimate assume that $\square_1((M,\mathsf{d},m),(M',\mathsf{d}',m'))<\varepsilon$. Then there exists a measurable map $\Phi=(\varphi,\varphi')\colon[0,1[\to M\times M'$ with $\varphi_*\lambda=m$ and $\varphi'_*\lambda=m'$ and there exists a measurable set $X_\varepsilon\subset[0,1[$ with $\lambda(X_\varepsilon)<\varepsilon$ such that

$$|\mathsf{d}(\varphi(s),\varphi(t))-\mathsf{d}'(\varphi'(s),\varphi'(t))|<\varepsilon$$

for all $s,t\in[0,1[\setminus X_\varepsilon$. Using the map $\Phi=(\varphi,\varphi')$ and the set X_ε we define a coupling $\hat{\mathsf{d}}$ of d and d' by

$$\hat{\mathsf{d}}(x,y')=\inf_{s\in[0,1[\setminus X_\varepsilon}[\mathsf{d}(x,\varphi(s))+\varepsilon/2+\mathsf{d}(\varphi'(s),y')]$$

for $x\in M$ and $y'\in M'$. Without restriction, we may assume that $\hat{\mathsf{d}}\leqslant L$ (otherwise replace $\hat{\mathsf{d}}$ by $\hat{\mathsf{d}}\wedge L$). Moreover, we define a coupling q of m and m' by

$$q=(\varphi,\varphi')_*\lambda.$$

Then

$$\begin{aligned}\mathbf{D}^2((M,\mathsf{d},m),(M',\mathsf{d}',m'))&\leqslant\int_{M\times M'}\hat{\mathsf{d}}^2(x,x')\,dq(x,x')\\&=\int_0^1\hat{\mathsf{d}}^2(\varphi(t),\varphi'(t))\,dt\leqslant\left(\frac{\varepsilon}{2}\right)^2(1-\varepsilon)+L^2\varepsilon\leqslant\left(L+\frac{1}{4}\right)^2\varepsilon.\end{aligned}$$

This proves the upper bound. □

3.2. Examples for D-convergence

Let us demonstrate the notion of **D**-convergence with various examples.

Example 3.8. (Products) Let $(M_n,\mathsf{d}_n,m_n)\in\mathbf{X}_1$ for $n\in\mathbf{N}$. Then

$$\left(\bigotimes_{n=1}^{l}(M_n,\mathsf{d}_n,m_n)\right)_{l\in\mathbf{N}}$$

is a **D**-Cauchy sequence in $\mathbf{X}_1$ if (and only if)

$$\sum_{n=1}^{\infty}\mathrm{Var}(M_n,\mathsf{d}_n,m_n)<\infty.$$

In this case, as $l\to\infty$,

$$\bigotimes_{n=1}^{l}(M_n,\mathsf{d}_n,m_n)\xrightarrow{\mathbf{D}}\bigotimes_{n=1}^{\infty}(M_n,\mathsf{d}_n,m_n).$$

Proof. Obviously, for all k and l

$$\begin{aligned}\mathbf{D}^2\left(\bigotimes_{n=1}^{l}(M_n,\mathsf{d}_n,m_n),\bigotimes_{n=1}^{l+k}(M_n,\mathsf{d}_n,m_n)\right) &\leqslant \mathbf{D}^2\left((\{o\},0,\delta_o),\bigotimes_{n=l+1}^{l+k}(M_n,\mathsf{d}_n,m_n)\right)\\ &\leqslant \sum_{n=l+1}^{l+k}\mathbf{D}^2((\{o\},0,\delta_o),(M_n,\mathsf{d}_n,m_n))\\ &= \sum_{n=l+1}^{l+k}\mathrm{Var}(M_n,\mathsf{d}_n,m_n).\end{aligned}$$

This proves the claim. □

Example 3.9. (Dimension increasing to infinity) Let $M_n=\mathbf{R}^n$ with Euclidean distance,

$$dm_n(x)=\frac{1}{(2\pi)^{n/2}\prod_{k=1}^n\sigma_k}\exp\left(-\frac{1}{2}\sum_{k=1}^n\left(\frac{x_k}{\sigma_k}\right)^2\right)dx.$$

Then

$$(M_n,\mathsf{d}_n,m_n)\xrightarrow{\mathbf{D}}(M_\infty,\mathsf{d}_\infty,m_\infty)$$

if and only if $\sum_{k=1}^\infty\sigma_k^2<\infty$.

Example 3.10. (Increasing finite dimension) (i) Let $M_n=((1/n)\mathbf{Z}\cap[0,1])^k$ be the rescaled k-dimensional lattice, d be the Euclidean distance in $\mathbf{R}^k$ and m_n be the renormalized counting measure on M_n. Then, as $n\to\infty$,

$$(M_n,\mathsf{d},m_n)\xrightarrow{\mathbf{D}}([0,1]^k,\mathsf{d},m),$$

with m being the k-dimensional Lebesgue measure in $[0,1]^k$ (see Figure 1).

(ii) Similarly, if $\widetilde{M}_n$ denotes the graph obtained from M_n with edges between next neighbors and $\widetilde{m}_n$ being the 1-dimensional Lebesgue measure on the edges, then

$$(\widetilde{M}_n,\mathsf{d},\widetilde{m}_n)\xrightarrow{\mathbf{D}}([0,1]^k,\mathsf{d},m)$$

as $n\to\infty$ (see Figure 2).

Example 3.11. (Increasing to fractal dimension) Let $(M_n)_{n\in\mathbf{N}}$ be the usual approximation of the Sierpiński gasket $M\subset\mathbf{R}^2$ by graphs M_n with 3^n edges of sidelength 2^{1-n}, $n\in\mathbf{N}$. To be more specific, M_1 is the equilateral triangle with sidelength 1 and for each $n\in\mathbf{N}$, the graph M_n is obtained from M_{n-1} by gluing together 3 copies and rescaling the whole by the factor $\frac{1}{2}$. Let d_n be the distance from the ambient 2-dimensional Euclidean

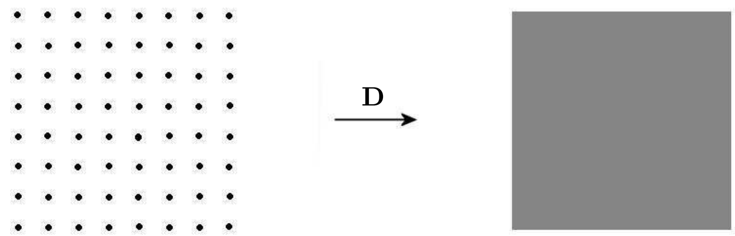

Figure 1.

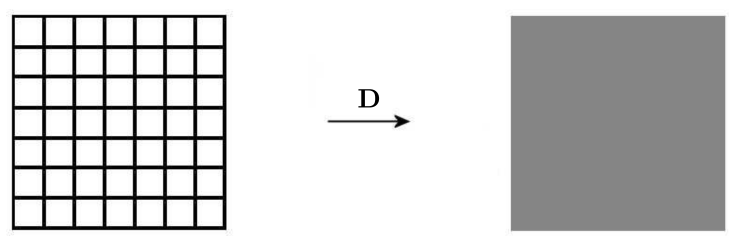

Figure 2.

space (or *alternatively* the induced length distance on M_n) and let m_n be normalized 1-dimensional Lebesgue measure on M_n. Then

$$(M_n, \mathsf{d}_n, m_n) \xrightarrow{\mathbf{D}} (M, \mathsf{d}, m),$$

where M is the *Sierpiński gasket*, d is the 2-dimensional Euclidean distance restricted to M (or the induced length distance on M, respectively) and m is the normalized $\log 3/\log 2$-dimensional Hausdorff measure on M (see Figure 3).

Similarly, we can approximate the 2-dimensional *Sierpiński carpet* $\widetilde{M}$ (equipped with Euclidean distance $\tilde{\mathsf{d}}$—or alternatively with the induced length distance—and with normalized $\log 8/\log 3$-Hausdorff measure $\widetilde{m}$) by graphs $\widetilde{M}_n$ with sidelength 3^{-n}. Here $\widetilde{M}_1$ is the square with sidelength 1 and $\widetilde{M}_n$ is obtained by gluing together 8 copies of $\widetilde{M}_{n-1}$ and rescaling the whole by the factor $\frac{1}{3}$. Then

$$(\widetilde{M}_n, \tilde{\mathsf{d}}_n, \widetilde{m}_n) \xrightarrow{\mathbf{D}} (\widetilde{M}, \tilde{\mathsf{d}}, \widetilde{m})$$

as $n \to \infty$ (Figure 4). See for instance [32].

Example 3.12. (Decreasing dimension, collapse) (i) For each metric measure space (M, d, m) and each sequence (M_n, d_n, m_n), $n \in \mathbf{N}$, with $\lim_{n\to\infty} \mathrm{Var}(M_n, \mathsf{d}_n, m_n) = 0$, one

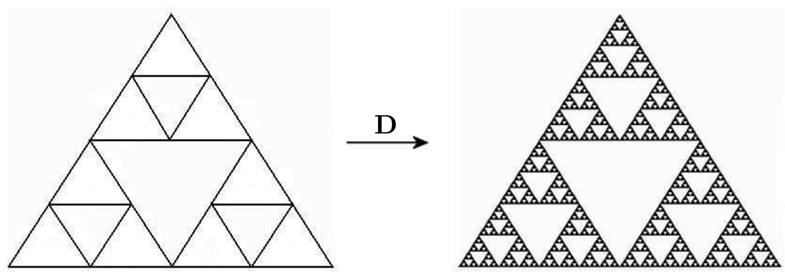

Figure 3.

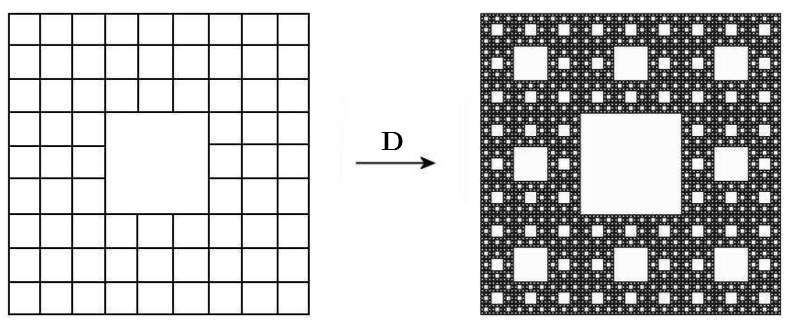

Figure 4.

has

$$(M \times M_n, \mathsf{d} \otimes \mathsf{d}_n, m \otimes m_n) \xrightarrow{\mathbf{D}} (M, \mathsf{d}, m)$$

as $n \to \infty$.

(ii) Let M be a finite graph, embedded in $\mathbf{R}^3$, let d be the graph distance and m be the 1-dimensional Lebesgue measure on M normalized to 1. Let

$$M_n := \{x \in \mathbf{R}^3 : \mathsf{d}_{\mathrm{Euclid}}(x, M) \leqslant 1/n\}$$

and

$$\widetilde{M}_n := \{x \in \mathbf{R}^3 : \mathsf{d}_{\mathrm{Euclid}}(x, M) = 1/n\}$$

be the full (and surface, respectively) tubular neighborhood of M, let d_n (and $\tilde{\mathsf{d}}_n$) be the geodesic distance on M_n (or $\widetilde{M}_n$, respectively) induced by the Euclidean distance $\mathsf{d}_{\mathrm{Euclid}}$ on the ambient space $\mathbf{R}^3$, and let m_n (and $\widetilde{m}_n$) be the 3- (or 2-, respectively) dimensional Lebesgue measure on M (or $\widetilde{M}$, respectively), normalized to 1. Then

$$(M_n, \mathsf{d}_n, m_n) \xrightarrow{\mathbf{D}} (M, \mathsf{d}, m)$$

and

$$(\widetilde{M}_n, \tilde{\mathsf{d}}_n, \widetilde{m}_n) \xrightarrow{\mathbf{D}} (M, \mathsf{d}, m)$$

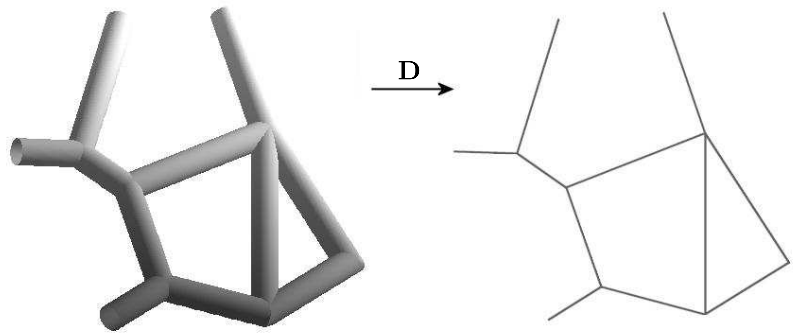

Figure 5.

as $n \to \infty$ (see Figure 5).

Being a length space is not preserved under isomorphisms of metric measure spaces. But the support $\mathrm{supp}[m]$ being a length space is preserved under isomorphisms. However, also this property is not preserved under **D**-convergence.

Example 3.13. Let $M = \mathbf{R}$ with Euclidean distance d, $dm_n(x) = \varphi_n(x)\,dx$ with

$$\varphi_n(x) = \begin{cases} 1/2n, & \text{if } x \in]-1, 1[, \\ 1/2, & \text{if } x \in [-2+1/n, -1] \cup [1, 2-1/n], \\ 0, & \text{otherwise,} \end{cases}$$

and

$$dm(x) = \tfrac{1}{2}(1_{[-2,-1]}(x) + 1_{[1,2]}(x))\,dx.$$

Then $(M, \mathsf{d}, m_n) \xrightarrow{\mathbf{D}} (M, \mathsf{d}, m)$ as $n \to \infty$ as well as $(\mathrm{supp}[m_n], \mathsf{d}, m_n) \xrightarrow{\mathbf{D}} (\mathrm{supp}[m], \mathsf{d}, m)$. However, $(\mathrm{supp}[m_n], \mathsf{d})$ does not converge with respect to D^{GH} towards $(\mathrm{supp}[m], \mathsf{d})$ as $n \to \infty$. It converges towards $([-2, 2], \mathsf{d})$. On the other hand, Example 3.10 (i) demonstrates the opposite phenomenon: non-length spaces converging to a length space.

3.3. Doubling property under D-convergence

Definition 3.14. Let $C \in \mathbf{R}_+$. We say that a metric measure space (M, d, m) has the *restricted doubling property* with *doubling constant* C if and only if for all $x \in \mathrm{supp}[m]$ and all $r \in \mathbf{R}_+$,

$$m(B_{2r}(x)) \leqslant C m(B_r(x)).$$

A metric measure space (M,d,m) has the restricted doubling property if and only if for all $x\in\operatorname{supp}[m]$ and all $r,R\in\mathbf{R}_+$,

$$m(B_R(x))\leqslant C^{\lfloor(\log R/r)/\log 2+1\rfloor}m(B_r(x)),$$

where $\lfloor a\rfloor$ denotes the greatest integer $\leqslant a$. It implies that for all x and r the sets $\overline{B}_r(x)\cap\operatorname{supp}[m]$ are compact.

Note that our definition differs from the usual definition of the doubling property: we only impose a condition on balls with center in the support of the measure. Some modification of this kind is necessary in order to obtain a property which is preserved under isomorphisms of metric measure spaces. The usual doubling property without this restriction implies that $\operatorname{supp}[m]=M$ whenever $m(M)\neq 0$.

For instance, let (M,d) be the 2-dimensional Euclidean plane and let m be the 1-dimensional Lebesgue measure on the x_1-axis. Then (M,d,m) has the restricted doubling property but not the doubling property in the usual sense.

There is a huge literature on metric measure spaces which have the doubling property; see e.g. [26] and references cited therein.

THEOREM 3.15. *The restricted doubling property is stable under* $\mathbf{D}$*-convergence.*

That is, if for all $n\in\mathbf{N}$ *the normalized metric measure spaces* (M_n,d_n,m_n) *have the restricted doubling property with a common doubling constant* C *and if* $(M_n,\mathsf{d}_n,m_n)\xrightarrow{\mathbf{D}}(M,\mathsf{d},m)$ *as* $n\to\infty$*, then also* (M,d,m) *has the restricted doubling property with the same constant* C.

Proof. Assume that the normalized metric measure spaces (M_n,d_n,m_n), $n\in\mathbf{N}$, have the restricted doubling property with a common doubling constant C and that

$$\delta_n:=\mathbf{D}((M_n,\mathsf{d}_n,m_n),(M,\mathsf{d},m))\to 0$$

as $n\to\infty$. Then for each $n\in\mathbf{N}$ the spaces $(\operatorname{supp}[m],\mathsf{d})$ and $(\operatorname{supp}[m_n],\mathsf{d}_n)$ can isometrically be embedded into some space $(\widehat{M},\hat{\mathsf{d}})$ such that

$$\hat{\mathsf{d}}_W(\widehat{m},\widehat{m}_n)\leqslant 2\delta_n,$$

where $\widehat{m}$ and $\widehat{m}_n$ denote the push-forwards of the measures m and m_n, respectively, under the embedding maps ψ and ψ_n, respectively.

Let $x\in\operatorname{supp}[m]$, $r>0$, $\varepsilon>0$ and $\alpha<1$ be given (with $2r\alpha^2-6\varepsilon>0$ for simplicity). Our first observation is that

$$\widehat{m}_n(\widehat{B}_{2\varepsilon}(\psi(x)))\geqslant\widehat{m}(\widehat{B}_\varepsilon(\psi(x)))-\frac{1}{\varepsilon^2}\hat{\mathsf{d}}_W^2(\widehat{m},\widehat{m}_n),\tag{3.9}$$

since the mass which has to be transported from the interior of the small ball to the exterior of the large ball has to be moved by a distance of at least ε. Moreover, we know that $\widehat{m}(\widehat{B}_\varepsilon(\psi(x)))=m(B_\varepsilon(x))>0$, since by assumption $x\in\operatorname{supp}[m]$. Hence, for n large enough we conclude that $\widehat{m}_n(\widehat{B}_{2\varepsilon}(\psi(x)))>0$. Therefore, there exists a point $\hat{x}_n\in\operatorname{supp}[\widehat{m}_n]\subset\widehat{M}$ with $\hat{\mathsf{d}}(\hat{x}_n,\psi(x))\leqslant 2\varepsilon$. In particular, we may apply the restricted doubling property for balls centered at x_n. This yields

$$\begin{aligned} m(B_{2\alpha^2r-6\varepsilon}(x)) &= \widehat{m}(\widehat{B}_{2\alpha^2r-6\varepsilon}(\psi(x))) \\ &\leqslant \widehat{m}(\widehat{B}_{2\alpha^2r-4\varepsilon}(\hat{x}_n)) \\ &\leqslant \widehat{m}_n(\widehat{B}_{2\alpha r-4\varepsilon}(\hat{x}_n)) + \frac{1}{(2r\alpha-2r\alpha^2)^2}\,\hat{\mathsf{d}}_W^2(\widehat{m},\widehat{m}_n) \\ &\leqslant C\widehat{m}_n(\widehat{B}_{\alpha r-2\varepsilon}(\hat{x}_n)) + \frac{1}{(2r\alpha-2r\alpha^2)^2}\,\hat{\mathsf{d}}_W^2(\widehat{m},\widehat{m}_n) \\ &\leqslant C\widehat{m}_n(\widehat{B}_{\alpha r}(\psi(x))) + \frac{1}{(2r\alpha-2r\alpha^2)^2}\,\hat{\mathsf{d}}_W^2(\widehat{m},\widehat{m}_n) \\ &\leqslant C\widehat{m}(\widehat{B}_r(\psi(x))) + \left[\frac{1}{(2r\alpha-2r\alpha^2)^2} + \frac{C}{(2r-2r\alpha)^2}\right]\hat{\mathsf{d}}_W^2(\widehat{m},\widehat{m}_n) \\ &\leqslant Cm(B_r(x)) + \frac{1+C\alpha^2}{(r\alpha(1-\alpha))^2}\,\delta_n^2. \end{aligned}$$

In the limit $n\to\infty$ we obtain

$$m(B_{2\alpha^2r-6\varepsilon}(x))\leqslant Cm(B_r(x)).$$

Since this holds for any $\alpha<1$ and any $\varepsilon>0$, we conclude that

$$m(B_{2r}(x))\leqslant Cm(B_r(x)). \qquad \square$$

3.4. D-convergence and measured Gromov–Hausdorff convergence

The L_2-transportation distance $\mathbf{D}$ on the space of metric measure spaces is closely related to the notion of measured Gromov–Hausdorff convergence (briefly: mGH-convergence) introduced by Fukaya [19].

Recall that a sequence of compact normalized metric measure spaces

$$((M_n,\mathsf{d}_n,m_n))_{n\in\mathbf{N}}$$

converges in *measured Gromov–Hausdorff sense* to a compact normalized metric measure space (M,d,m) if and only if there exist numbers $\varepsilon_n\searrow 0$ and ε_n-isometries $\psi_n\colon M_n\to M$ such that $(\psi_n)_*m_n\to m$ weakly on M for $n\to\infty$.

THEOREM 3.16. (Compactness) *For each pair* $(C,L)\in\mathbf{R}_+\times\mathbf{R}_+$ *the family* $\mathbf{X}_1(C,L)$ *of all isomorphism classes of normalized metric measure spaces with ('restricted') doubling constant* $\leqslant C$ *and diameter* $\leqslant L$ *is* $\mathbf{D}$*-compact (and thus also* $\square_1$*-compact).*

Moreover, the family $\widetilde{\mathbf{X}}_1(C,L)$ *of normalized metric measure spaces* (M,d,m) *with full support, doubling constant* $\leqslant C$ *and diameter* $\leqslant L$ *is mGH-compact.*

Proof. Let C and L be given and consider a space $(M,\mathsf{d},m)\in\mathbf{X}_1(C,L)$. Without restriction, we may assume that $\mathrm{supp}[m]=M$. Then each ball $B_\varepsilon(x)\subset M=B_L(x)$ has volume

$$m(B_\varepsilon(x))\geqslant\left(\frac{\varepsilon}{2L}\right)^N$$

with 'doubling dimension' $N:=\log C/\log 2$. Hence, each family of disjoint balls of radius ε in M contains at most $(2L/\varepsilon)^N$ elements which in turn implies that M can be covered by $(2L/\varepsilon)^N$ balls of radius 2ε. Firstly, this implies that M is compact. Secondly, since this covering property holds uniformly in $(M,\mathsf{d},m)\in\widetilde{\mathbf{X}}_1(C,L)$, according to Gromov's precompactness theorem ([20] or [7, Theorem 5.41]) it implies compactness of the family $\widetilde{\mathbf{X}}_1(C,L)$ under Gromov–Hausdorff convergence. Moreover, due to Theorem 3.15 (together with Lemma 3.18 (i)) the family $\widetilde{\mathbf{X}}_1(C,L)$ is closed under $\mathbf{D}$-convergence (as well as under mGH-convergence). Due to the following Lemma 3.17 this in turn implies compactness under mGH-convergence. Due to Lemma 3.18 (i) below, the mGH-compactness implies $\mathbf{D}$-compactness. □

LEMMA 3.17. *Let* $\{(M_i,\mathsf{d}_i,m_i):i\in I\}$ *be an arbitrary family of normalized compact metric measure spaces which is closed under mGH-convergence. If the family* $\mathsf{X}'=\{(M_i,\mathsf{d}_i):i\in I\}$ *is compact with respect to Gromov–Hausdorff convergence then the family* $\mathbf{X}'=\{(M_i,\mathsf{d}_i,m_i):i\in I\}$ *is compact with respect to measured Gromov–Hausdorff convergence.*

Proof. (Cf. [19, Proposition 2.10].) Let a sequence $((M_n,\mathsf{d}_n,m_n))_{n\in\mathbf{N}}$ in $\mathbf{X}'$ be given. Then (by the assumption of compactness of X') there exists a subsequence

$$((M_{n_k},\mathsf{d}_{n_k},m_{n_k}))_{k\in\mathbf{N}}$$

and a compact metric space (M,d) such that $(M_{n_k},\mathsf{d}_{n_k})\to(M,\mathsf{d})$ in D^{GH}. To simplify notation, assume that the whole sequence is GH-convergent. Then there exist sequences of numbers $\varepsilon_n\searrow 0$ and ε_n-isometries $\psi_n: M_n\to M$. Now consider the sequence of probability measures $m'_n=(\psi_n)_*m_n$ on M. Since M is compact, Prohorov's theorem implies that $m'_{n_l}\to m$ as $l\to\infty$ weakly on M for a suitable subsequence $(m'_{n_l})_{l\in\mathbf{N}}$ and for some probability measure m on M. In other words,

$$(M_{n_l},\mathsf{d}_{n_l},m_{n_l})\to(M,\mathsf{d},m)$$

in mGH-sense. This proves the compactness. □

Lemma 3.18. (i) *For any sequence of normalized compact metric measure spaces, mGH-convergence implies* $\mathbf{D}$*-convergence.*

(ii) *For any sequence of normalized compact metric measure spaces with full supports and uniform bounds for the doubling constants and the diameters, mGH-convergence is equivalent to* $\mathbf{D}$*-convergence* (*as well as to* $\underline{\Box}_1$*-convergence*).

Proof. (i) Assume that a sequence of compact normalized metric measure spaces $((M_n,\mathsf{d}_n,m_n))_{n\in\mathbf{N}}$ mGH-converges to a compact normalized metric measure space (M,d,m). That is, there exist numbers $\delta_n,\varepsilon_n\searrow 0$ and ε_n-isometries $\psi_n\colon M_n\to M$ such that $\mathsf{d}_W(m_n',m)\leqslant\delta_n$, where $m_n':=(\psi_n)_*m_n$. We claim that this implies that

$$\mathbf{D}((M,\mathsf{d},m),(M_n,\mathsf{d}_n,m_n))\leqslant\delta_n+\frac{\varepsilon_n}{2}.$$

Obviously,

$$\mathbf{D}((M,\mathsf{d},m),(M,\mathsf{d},m_n'))\leqslant\delta_n.$$

It remains to prove that

$$\mathbf{D}((M,\mathsf{d},m_n'),(M_n,\mathsf{d}_n,m_n))\leqslant\frac{\varepsilon_n}{2}.$$

For this purpose, we define couplings q_n of m_n' and m_n by $q_n=(\psi_n,\mathrm{Id})_*m_n$, and couplings $\hat{\mathsf{d}}_n$ of d and d_n by

$$\hat{\mathsf{d}}_n(x,y)=\frac{\varepsilon_n}{2}+\inf_{z\in M_n}[\mathsf{d}(x,\psi_n(z))+\mathsf{d}_n(z,y)]$$

for $x\in M$ and $y\in M_n$. Indeed, q_n and $\hat{\mathsf{d}}_n$ are couplings and

$$\int_{M\times M_n}\hat{\mathsf{d}}_n^2(x,y)\,dq_n(x,y)=\int_{M_n}\hat{\mathsf{d}}_n^2(\psi_n(y),y)\,dm_n(y)=\Big(\frac{\varepsilon_n}{2}\Big)^2.$$

This proves the claim.

(ii) By the homeomorphism theorem, part (i) together with the mGH-compactness (Theorem 3.16) imply the equivalence of the topologies. □

In general, there is no converse to the implication (i) of the previous Lemma 3.18.

Example 3.19. Let $M_n=[-n,n]$ and $m_n(dx)=c_n\cdot 1_{[-n,n]}(x)\exp(-n|x|)\,dx$, with $c_n=(n/2)(1-\exp(-n^2))^{-1}$. Then there exists no mGH-limit of the sequence

$$((M_n,\mathsf{d}_{\mathrm{Euclid}},m_n))_{n\in\mathbf{N}}.$$

It also has no mGH-converging subsequence.

However, the above sequence $\mathbf{D}$-converges to $(\mathbf{R},\mathsf{d}_{\mathrm{Euclid}},\delta_0)$ (or equivalently, to $(0,\mathsf{d}_{\mathrm{Euclid}},\delta_0)$).

Example 3.20. Let $(\sigma_k)_{k\in\mathbf{N}}$ be a sequence of positive numbers with $\sum_{k=1}^{\infty}\sigma_k^2<\infty$ and consider for each $n\in\mathbf{N}$ the metric measure space (M_n, d_n, m_n) from Example 3.9, where (M_n, d_n) is the n-dimensional Euclidean space and m_n is some Gaussian measure with covariance matrix given by the $\sigma_1, ..., \sigma_n$.

Let $M'_n=(\mathbf{R}/\mathbf{Z})^n$ be the n-dimensional torus with induced length metric d'_n and let $\pi_n: M_n\to M'_n$ be the projection map. Put $m'_n=(\pi_n)_*m_n$. Then the sequence

$$((M'_n, \mathsf{d}'_n, m'_n))_{n\in\mathbf{N}}$$

will be $\mathbf{D}$-convergent to some metric measure space $(M'_\infty, \mathsf{d}'_\infty, m'_\infty)$, where M'_∞ is the infinite-dimensional torus and m'_∞ is some probability measure on it.

However, neither the given sequence $((M'_n, \mathsf{d}'_n, m'_n))_{n\in\mathbf{N}}$ nor any subsequence of it will converge in mGH-sense to some metric measure space.

Remark 3.21. (i) The concept of mGH-convergence is not appropriate to study *isomorphism classes* of metric measure spaces.

In particular, the mGH-convergence of a sequence (M_n, d_n, m_n) towards (M, d, m) *does not* imply the mGH-convergence of the sequence $(\operatorname{supp}[m_n], \mathsf{d}_n, m_n)$ towards $(\operatorname{supp}[m], \mathsf{d}, m)$.

Similarly, the mGH-convergence of $(\operatorname{supp}[m_n], \mathsf{d}_n, m_n)$ towards $(\operatorname{supp}[m], \mathsf{d}, m)$ does not imply the mGH-convergence of the sequence (M_n, d_n, m_n) towards (M, d, m).

(ii) According to part (v) of the proof of Theorem 3.6 (cf. Gromov's union lemma [22]), for each sequence of metric measure spaces with $\mathbf{D}((M, \mathsf{d}, m), (M_n, \mathsf{d}_n, m_n))\to 0$ sufficiently fast there exists a (huge) metric space $(\widehat{M}, \hat{\mathsf{d}})$ such that all the spaces M_n, as well as the space M, can be isometrically embedded into $\widehat{M}$ and the push-forward measures $\widehat{m}_n=(\psi_n)_*m_n$ converge weakly to $\widehat{m}=\psi_*m$ on $\widehat{M}$. Hence (at least in some extended sense), the sequence of metric measure spaces $(\widehat{M}, \hat{\mathsf{d}}, \widehat{m}_n)$ mGH-converges to $(\widehat{M}, \hat{\mathsf{d}}, \widehat{m})$. This convergence, however, has no meaning in the sense of 'convergence of the original spaces'. The space $\widehat{M}$ has no geometric meaning. In general, it will be not compact.

4. Curvature bounds for metric measure spaces

4.1. The relative entropy

Recall that a metric measure space always means a triple (M, d, m), where (M, d) is a complete separable metric space and m is a locally finite measure on M equipped with its Borel σ-algebra. To avoid pathologies, in the sequel we always exclude the case $m(M)=0$.

Given a metric measure space (M,d,m) we denote by $\mathcal{P}_2(M,\mathsf{d},m)$ the subspace of all $\nu\in\mathcal{P}_2(M,\mathsf{d})$ which are absolutely continuous with respect to m, that is, which can be written as $\nu=\varrho m$, with Radon–Nikodym density ϱ. In other words, $\mathcal{P}_2(M,\mathsf{d},m)$ can be identified with the set of all m-equivalence classes of non-negative Borel-measurable functions $\varrho\colon M\to\mathbf{R}$ satisfying $\int\varrho(x)\,dm(x)=1$ and $\int\mathsf{d}^2(o,x)\varrho(x)\,dm(x)<\infty$ for some $o\in M$.

For $\nu=\varrho m\in\mathcal{P}_2(M,\mathsf{d},m)$ we define the *relative entropy* of ν with respect to m by

$$\mathrm{Ent}(\nu\,|\,m):=\lim_{\varepsilon\searrow 0}\int_{\{\varrho>\varepsilon\}}\varrho\log\varrho\,dm. \tag{4.1}$$

This coincides with

$$\int_{\{\varrho>0\}}\varrho\log\varrho\,dm$$

provided $\int_{\{\varrho>1\}}\varrho\log\varrho\,dm<\infty$. Otherwise $\mathrm{Ent}(\nu|m):=+\infty$. We also define $\mathrm{Ent}(\nu|m):=+\infty$ for $\nu\in\mathcal{P}_2(M,\mathsf{d})\setminus\mathcal{P}_2(M,\mathsf{d},m)$. Finally, we put

$$\mathcal{P}_2^*(M,\mathsf{d},m):=\{\nu\in\mathcal{P}_2(M,\mathsf{d}):\mathrm{Ent}(\nu\,|\,m)<\infty\}.$$

LEMMA 4.1. *If m has finite mass, then the relative entropy* $\mathrm{Ent}(\,\cdot\,|m)$ *is lower semicontinuous and $\neq-\infty$ on $\mathcal{P}_2(M,\mathsf{d})$. More precisely, for all $\nu\in\mathcal{P}_2(M,\mathsf{d})$,*

$$\mathrm{Ent}(\nu\,|\,m)\geqslant-\log m(M). \tag{4.2}$$

Proof. The lower estimate for the relative entropy is a simple application of Jensen's inequality. The lower semicontinuity is more subtle. For $N>1$ define $U_N(r)=-Nr^{1-1/N}$. Then $r\mapsto U_N(r)$ is convex on $\mathbf{R}_+$, hence

$$U_N(r)\geqslant U_N(r_0)-(N-1)r_0^{-1/N}(r-r_0) \tag{4.3}$$

for all r and r_0. Moreover,

$$\lim_{N\to\infty}[Nr+U_N(r)]=\sup_N[Nr+U_N(r)]=r\log r.$$

Now consider $\widetilde{S}_N\colon\mathcal{P}_2(M,\mathsf{d})\to\mathbf{R}$ with

$$\widetilde{S}_N(\nu):=\int U_N(\varrho)\,dm+N$$

for $\nu=\nu_0+\nu_*\in\mathcal{P}_2(M,\mathsf{d})$, with $\nu_*\perp m$ and $\nu_0=\varrho m$. Note that

$$N-N\nu_0(M)^{1-1/N}m(M)^{1/N}\leqslant\widetilde{S}_N(\nu)\leqslant N$$

for all ν. Therefore, for all $\nu\in\mathcal{P}_2(M,\mathsf{d})$

$$\mathrm{Ent}(\nu\,|\,m)=\lim_{N\to\infty}\widetilde{S}_N(\nu)=\sup_N \widetilde{S}_N(\nu). \tag{4.4}$$

Hence, lower semicontinuity of $\mathrm{Ent}(\,\cdot\,|m)$ will follow from lower semicontinuity of $\widetilde{S}_N$. In order to verify the latter, let $\nu_n=\varrho_n m+\nu_n^*$ be any sequence in $\mathcal{P}_2(M,\mathsf{d})$ which converges to some $\nu=\varrho m+\nu^*\in\mathcal{P}_2(M,\mathsf{d})$. It implies that $\nu_n\to\nu$ weakly in the sense of measures. Since ϱ is a sub-probability density, it lies in $L_{1-1/N}(M,m)$ and it can be approximated in the metric $D_{1-1/N}$ by non-negative, bounded, continuous $\varrho^{(i)}\in L_{1-1/N}(M,m)$. Here $D_{1-1/N}(u,v):=\int|u-v|^{1-1/N}\,dm$. Put $\varrho_n^{(i)}:=\varrho_n-\varrho+\varrho^{(i)}$. Then

$$|\widetilde{S}_N(\varrho^{(i)}m)-\widetilde{S}_N(\varrho m)|\leqslant N D_{1-1/N}(\varrho^{(i)},\varrho)\to 0$$

as well as

$$|\widetilde{S}_N(\varrho_n^{(i)}m)-\widetilde{S}_N(\varrho_n m)|\leqslant N D_{1-1/N}(\varrho^{(i)},\varrho)\to 0$$

as $i\to\infty$, uniformly in n. According to (4.3),

$$\begin{aligned}\widetilde{S}_N(\varrho_n^{(i)}m)-\widetilde{S}_N(\varrho^{(i)}m)&\geqslant-(N-1)\int[\varrho^{(i)}]^{-1/N}(\varrho_n^{(i)}-\varrho^{(i)})\,dm\\&=-(N-1)\int\big[\varrho^{(i)}\big]^{-1/N}(\varrho_n-\varrho)\,dm,\end{aligned}$$

which tends to 0 as $n\to\infty$ due to the weak convergence of $\varrho_n m$ to ϱm and since $\varrho^{(i)}$ is continuous and bounded. Summing up, we obtain

$$\liminf_{n\to\infty}\widetilde{S}_N(\varrho_n m)\geqslant\widetilde{S}_N(\varrho m)$$

and thus finally, as $N\to\infty$,

$$\liminf_{n\to\infty}\mathrm{Ent}(\varrho_n m\,|\,m)\geqslant\mathrm{Ent}(\varrho m\,|\,m). \qquad\square$$

Remark 4.2. The relative entropy can in the same manner also be defined for finite, non-normalized measures ν (and m) on M. Then for all $\alpha,\beta>0$,

$$\mathrm{Ent}(\alpha\nu\,|\,\beta m)=\alpha\,\mathrm{Ent}(\nu\,|\,m)+(\log\alpha-\log\beta)\alpha\nu(M). \tag{4.5}$$

Moreover, for all finite or countable sets I and all finite measures ν_i, $i\in I$,

$$\mathrm{Ent}\bigg(\sum_{i\in I}\nu_i\,\bigg|\,m\bigg)\geqslant\sum_{i\in I}\mathrm{Ent}(\nu_i\,|\,m) \tag{4.6}$$

with equality if and only if the ν_i, $i\in I$, are mutually singular, and

$$\mathrm{Ent}\Big(\sum_{i\in I}\nu_i\,\Big|\,m\Big)\leqslant\sum_{i\in I}\mathrm{Ent}(\nu_i\,|\,m)-\sum_{i\in I}\nu_i(M)\log\nu_i(M). \tag{4.7}$$

For convex combinations of probability measures ν_i, inequalities (4.6) and (4.7) become

$$\sum_{i\in I}\alpha_i\,\mathrm{Ent}(\nu_i\,|\,m)+\sum_{i\in I}\alpha_i\log\alpha_i\leqslant\mathrm{Ent}\Big(\sum_{i\in I}\alpha_i\nu_i\,\Big|\,m\Big)\leqslant\sum_{i\in I}\alpha_i\,\mathrm{Ent}(\nu_i\,|\,m). \tag{4.8}$$

Proof. Indeed,

$$\mathrm{Ent}\Big(\sum_{i\in I}\nu_i\,\Big|\,m\Big)=\sum_{i\in I}\int\varrho_i\log\Big(\sum_{k\in I}\varrho_k\Big)\,dm\overset{(*)}{\geqslant}\sum_{i\in I}\int\varrho_i\log\varrho_i\,dm=\sum_{i\in I}\mathrm{Ent}(\nu_i\,|\,m),$$

with equality in $(*)$ if and only if $\varrho_k\varrho_i=0$ m-a.e. on M for all $k\neq i$. On the other hand, according to Jensen's inequality (applied to the convex function $\varphi(r)=r\log r$)

$$\begin{aligned}\mathrm{Ent}\Big(\sum_{i\in I}\nu_i\,\Big|\,m\Big)&=\int\Big(\sum_{i\in I}\alpha_i\bar\varrho_i\Big)\log\Big(\sum_{i\in I}\alpha_i\bar\varrho_i\Big)\,dm\\&\leqslant\int\sum_{i\in I}\alpha_i\bar\varrho_i\log\bar\varrho_i\,dm\\&=\sum_{i\in I}\int\varrho_i\log\varrho_i\,dm-\sum_{i\in I}\alpha_i\log\alpha_i,\end{aligned}$$

with $\alpha_i=\nu_i(M)$, $\bar\varrho_i=(1/\alpha_i)\varrho_i$ and $\varrho_i=d\nu_i/dm$. □

Remark 4.3. (i) If m has infinite mass then $\mathrm{Ent}(\cdot\,|m)$ may exhibit strange behavior. In particular, it can attain the value $-\infty$ and also lower semicontinuity may fail. See the example below.

(ii) If m is finite on all balls and if $\mathrm{Ent}(\nu|m)<\infty$ then

$$\mathrm{Ent}(\nu\,|\,m)=\lim_{R\to\infty}\int_{B_R(o)}\varrho\log\varrho\,dm \tag{4.9}$$

for each $\nu=\varrho m$ (with any $o\in M$). Indeed, due to the finiteness of m on $B_R(o)$ the integral on the right-hand side exists for all R and as $R\to\infty$ by monotone convergence

$$\int_{B_R(o)\cap\{\varrho>1\}}\varrho\log\varrho\,dm\to\int_{\{\varrho>1\}}\varrho\log\varrho\,dm<\infty,$$

whereas

$$\int_{B_R(o)\cap\{\varrho<1\}}\varrho\log\varrho\,dm\to\int_{\{\varrho<1\}}\varrho\log\varrho\,dm\leqslant\infty.$$

Example 4.4. Let $M=\mathbf{R}$ with Euclidean distance d and $dm(x)=\exp(\exp(x^2))\,dx$, $d\mu_\alpha(x)=(1/2\alpha)\exp(-|x|/\alpha)\,dx$.

(i) Then for all $\alpha>0$ we have $\mu_\alpha\in\mathcal{P}_2(M)$ with $\mathsf{d}_W(\mu_\alpha,\delta_0)=\alpha\sqrt{2}$ and

$$\mathrm{Ent}(\mu_\alpha\,|\,m)=-\infty.$$

(ii) For each $\eta\in\mathcal{P}_2(M,d)$ with $\mathrm{Ent}(\eta|m)>-\infty$ the relative entropy $\mathrm{Ent}(\,\cdot\,|m)$ is *not* lower semicontinuous at η, since

$$\eta_\alpha := (1-\alpha)\eta+\alpha\mu_\alpha\to\eta$$

in $(\mathcal{P}_2(M),\mathsf{d}_W)$ as $\alpha\to0$ and

$$-\infty=\lim_{\alpha\to0}\mathrm{Ent}(\eta_\alpha\,|\,m)<\mathrm{Ent}(\eta\,|\,m).$$

(iii) Moreover, given any $\nu_0,\nu_1\in\mathcal{P}_2(M,\mathsf{d})$ there exists a midpoint η of them. If $\mathrm{Ent}(\eta|m)<\infty$ then for each $\varepsilon>0$ there exists an $\alpha>0$ such that η_α (defined as before) is an ε-midpoint of ν_0 and ν_1 and

$$-\infty=\mathrm{Ent}(\eta_\alpha\,|\,m)\leqslant\frac{1}{2}\mathrm{Ent}(\nu_0\,|\,m)+\frac{1}{2}\mathrm{Ent}(\nu_1\,|\,m)-\frac{K}{8}\mathsf{d}_W^2(\nu_0,\nu_1)$$

for each $K\in\mathbf{R}$.

4.2. Curvature bounds

Definition 4.5. (i) We say that a metric measure space (M,d,m) *has curvature* $\geqslant K$ for some number $K\in\mathbf{R}$ if and only if the relative entropy $\mathrm{Ent}(\,\cdot\,|m)$ is weakly K-convex on $\mathcal{P}_2^*(M,\mathsf{d},m)$ in the following sense: for each pair $\nu_0,\nu_1\in\mathcal{P}_2^*(M,\mathsf{d},m)$ there exists a geodesic $\Gamma\colon[0,1]\to\mathcal{P}_2^*(M,\mathsf{d},m)$ connecting ν_0 and ν_1 with

$$\mathrm{Ent}(\Gamma(t)\,|\,m)\leqslant(1-t)\,\mathrm{Ent}(\Gamma(0)\,|\,m)+t\,\mathrm{Ent}(\Gamma(1)\,|\,m)-\frac{K}{2}t(1-t)\,\mathsf{d}_W^2(\Gamma(0),\Gamma(1)) \quad (4.10)$$

for all $t\in[0,1]$. To be more specific, we say that in the previous case the metric measure space (M,d,m) *globally has curvature* $\geqslant K$. Moreover, we put

$$\underline{\mathsf{Curv}}(M,\mathsf{d},m):=\sup\{K\in\mathbf{R}:(M,\mathsf{d},m)\text{ has curvature }\geqslant K\}$$

(with $\sup\varnothing:=-\infty$ as usual). Note that then (M,d,m) has curvature $\geqslant\underline{\mathsf{Curv}}(M,\mathsf{d},m)$. Occasionally, we use slightly modified concepts:

(ii) We say that a metric measure space (M,d,m) *(globally) has curvature* $\geqslant K$ *in the lax sense* if and only if for each $\varepsilon>0$ and for each pair $\nu_0,\nu_1\in\mathcal{P}_2^*(M,\mathsf{d},m)$ there exists an ε-midpoint $\eta\in\mathcal{P}_2^*(M,\mathsf{d},m)$ of ν_0 and ν_1 with

$$\mathrm{Ent}(\eta\,|\,m)\leqslant\frac{1}{2}\,\mathrm{Ent}(\nu_0\,|\,m)+\frac{1}{2}\,\mathrm{Ent}(\nu_1\,|\,m)-\frac{K}{8}\,\mathsf{d}_W^2(\nu_0,\nu_1)+\varepsilon. \tag{4.11}$$

We denote the maximal K with this property by $\underline{\mathsf{Curv}}_{\mathrm{lax}}(M,\mathsf{d},m)$.

(iii) We say that a metric measure space (M,d,m) *locally has curvature* $\geqslant K$ if each point of M has a neighborhood M' such that for each pair $\nu_0,\nu_1\in\mathcal{P}_2^*(M,\mathsf{d},m)$ supported in M' there exists a geodesic $\Gamma\colon[0,1]\to\mathcal{P}_2^*(M,\mathsf{d},m)$ connecting ν_0 and ν_1 and satisfying (4.10). (Note that we do not require that the $\Gamma(t)$ are supported in M'.)

The maximal K with this property will be denoted by $\underline{\mathsf{Curv}}_{\mathrm{loc}}(M,\mathsf{d},m)$.

Remark 4.6. Let (M,d,m) be a metric measure space of finite mass.

(i) Then $\underline{\mathsf{Curv}}(M,\mathsf{d},m)\geqslant K$ if and only if for each pair $\nu_0,\nu_1\in\mathcal{P}_2^*(M,\mathsf{d},m)$ there exists a midpoint $\eta\in\mathcal{P}_2^*(M,\mathsf{d},m)$ of ν_0 and ν_1 with

$$\mathrm{Ent}(\eta\,|\,m)\leqslant\frac{1}{2}\,\mathrm{Ent}(\nu_0\,|\,m)+\frac{1}{2}\,\mathrm{Ent}(\nu_1\,|\,m)-\frac{K}{8}\,\mathsf{d}_W^2(\nu_0,\nu_1). \tag{4.12}$$

(ii) Similarly, we have that $\underline{\mathsf{Curv}}_{\mathrm{lax}}(M,\mathsf{d},m)\geqslant K$ if and only if for all $\varepsilon>0$ and all $\nu_0,\nu_1\in\mathcal{P}_2^*(M,\mathsf{d},m)$ there exists a curve $\Gamma\colon[0,1]\to\mathcal{P}_2^*(M,\mathsf{d},m)$ connecting ν_0 and ν_1 with

$$\mathrm{Length}(\Gamma)\leqslant\mathsf{d}_W(\nu_0,\nu_1)+\varepsilon \tag{4.13}$$

and

$$\mathrm{Ent}(\Gamma(t)\,|\,m)\leqslant(1-t)\,\mathrm{Ent}(\nu_0\,|\,m)+t\,\mathrm{Ent}(\nu_1\,|\,m)-\frac{K}{2}t(1-t)\,\mathsf{d}_W^2(\nu_0,\nu_1)+\varepsilon \tag{4.14}$$

for all $t\in[0,1]$.

(iii) The fact that $\underline{\mathsf{Curv}}_{\mathrm{lax}}(M,\mathsf{d},m)>-\infty$ implies that $\mathcal{P}_2^*(M,\mathsf{d},m)$ is a length space (with metric d_W) and that $M_0=\mathrm{supp}[m]\subset M$ is a length space (with metric d).

(iv) Obviously, $\underline{\mathsf{Curv}}_{\mathrm{loc}}(M,\mathsf{d},m)\geqslant K$ provided each point of M has a neighborhood M' such that (M',d,m) globally has curvature $\geqslant K$. Due to the previous remark (iii) this requires M' to be convex, at least in some weak sense.

Proof. (i), (ii) We have to prove that the existence of (approximate) midpoints with property (4.12) (or (4.11)) implies the existence of (approximate) geodesics with property (4.10) (or (4.14), respectively).

Given $\varepsilon=0$ (or $\varepsilon>0$, respectively) define $\Gamma\big(\frac{1}{2}\big)$ as an ε-midpoint of $\Gamma(0):=\nu_0$ and $\Gamma(1):=\nu_1$ satisfying (4.11). Then define $\Gamma\big(\frac{1}{4}\big)$ as an $\varepsilon/2$-midpoint of $\Gamma(0)$ and $\Gamma\big(\frac{1}{2}\big)$

satisfying (4.11) with $\varepsilon/2$ and define $\Gamma\left(\frac{3}{4}\right)$ as an $\varepsilon/2$-midpoint of $\Gamma\left(\frac{1}{2}\right)$ and $\Gamma(1)$ satisfying (4.11) with $\varepsilon/2$. By iteration, we obtain $\Gamma(t)$ for all dyadic $t\in[0,1]$. The continuous extension yields the required curve. (See, for instance, the proof of [50, Proposition 2.3], for a similar argument.) Lower semicontinuity of the relative entropy then proves the claim for all $t\in[0,1]$.

(iii) According to part (ii), it only remains to prove that M_0 is a length space. Given $x_0, x_1\in M_0$ let ν_i, for $i=0,1$, be the normalized volume in $B_\varepsilon(x_i)$, i.e.

$$\nu_i = \frac{1}{m(B_\varepsilon(x_i))} 1_{B_\varepsilon(x_i)} m,$$

with $\varepsilon>0$ to be chosen later. (Note that $m(B_\varepsilon(x_i))>0$ for all $\varepsilon>0$, since $x_i\in M_0$, and $m(B_\varepsilon(x_i))<\infty$ for all sufficiently small $\varepsilon>0$, since m is locally finite.) Then we have $\nu_i\in\mathcal{P}_2^*(M,\mathsf{d},m)$. Hence, there exists $\eta\in\mathcal{P}_2^*(M,\mathsf{d},m)$ with

$$\mathsf{d}_W(\nu_i,\eta) \leqslant \tfrac{1}{2}\mathsf{d}_W(\nu_0,\nu_1)+\varepsilon$$

for $i=0,1$. Therefore

$$\begin{aligned}\int [\mathsf{d}^2(x_0,y)+\mathsf{d}^2(x_1,y)]\,d\eta(y) &= \mathsf{d}_W^2(\delta_{x_0},\eta)+\mathsf{d}_W^2(\delta_{x_1},\eta)\\ &\leqslant [\mathsf{d}_W(\nu_0,\eta)+\varepsilon]^2+[\mathsf{d}_W(\nu_1,\eta)+\varepsilon]^2\\ &\leqslant 2\left[\tfrac{1}{2}\mathsf{d}_W(\nu_0,\nu_1)+2\varepsilon\right]^2\\ &\leqslant 2\left[\tfrac{1}{2}\mathsf{d}(x_0,x_1)+3\varepsilon\right]^2\\ &= \tfrac{1}{2}\mathsf{d}^2(x_0,x_1)+\varepsilon'\end{aligned}$$

for arbitrarily small $\varepsilon'>0$. It implies that there exists a point $y\in\operatorname{supp}[\eta]$ with

$$\mathsf{d}^2(x_0,y)+\mathsf{d}^2(x_1,y) \leqslant \tfrac{1}{2}\mathsf{d}^2(x_0,x_1)+\varepsilon'.$$

In other words, y is an approximate midpoint and thus M_0 is a length space. □

LEMMA 4.7. *If M is compact then curvature bounds in the usual sense and in the lax sense coincide*:

$$\underline{\mathsf{Curv}}(M,\mathsf{d},m) = \underline{\mathsf{Curv}}_{\text{lax}}(M,\mathsf{d},m).$$

Proof. Given $\nu_0,\nu_1\in\mathcal{P}_2^*(M,\mathsf{d},m)$ let $\eta^{(i)}$ be a family of ε-midpoints of ν_0 and ν_1 satisfying (4.11) with $\varepsilon=1/i$. Consider the family of probability measures $Q:=\{\eta^{(i)}:i\in\mathbf{N}\}$. This family is tight. Indeed, we may assume without restriction that M is a compact length space (otherwise, replace M by $M_0=\operatorname{supp}[m]$; see Remark 4.6 (iii)). Hence, there exists a suitable subsequence $(\eta^{(i_j)})_{j\in\mathbf{N}}$ which converges to some $\eta\in\mathcal{P}_2(M,\mathsf{d})$. Continuity of the distance d_W and lower semicontinuity of the relative entropy $\mathrm{Ent}(\,\cdot\,|m)$ imply that η is a midpoint of ν_0 and ν_1 and (4.11) holds with $\varepsilon=0$. Iterating this procedure yields a geodesic connecting ν_0 and ν_1 and satisfying (4.10). □

The usual definition of K-convexity for the relative entropy would require that (4.10) holds for *each* geodesic connecting ν_0 and ν_1. This leads to the following definition which, however, will be not used in this paper.

We say that a metric measure space (M,d,m) *(globally) has curvature* $\geqslant K$ *in the restricted sense* if and only if $\mathcal{P}_2^*(M,\mathsf{d},m)$ is a geodesic space and if and only if each geodesic Γ in $\mathcal{P}_2^*(M,\mathsf{d},m)$ satisfies (4.10).

Remark 4.8. Assume that M is a compact non-branching geodesic space where each pair of points in M is connected by a unique geodesic which depends continuously on the endpoints. Then curvature bounds in the restricted sense and curvature bounds in the usual sense coincide.

Proof. The (uniformly) continuous dependence of the geodesics on the endpoints implies (and actually is equivalent to the fact) that for each $\varepsilon>0$ there exists $\delta>0$ such that the midpoint z' of $x'\in B_\delta(x)$ and $y'\in B_\delta(y)$ lies in $B_\varepsilon(z)$ whenever z is the midpoint of x and y. Now let the probability measures q on $M\times M$ and η on M be given, which are an optimal coupling and a midpoint, respectively, of some ν_0 and ν_1. Decompose q into a sum $q=\sum_{i\in\mathbf{N}} q_i$ of mutually singular q_i, $i\in\mathbf{N}$, with $\mathrm{supp}[q_i]\subset B_\delta(x_i)\times B_\delta(y_i)$ for suitable $x_i,y_i\in M$, $i\in\mathbf{N}$. Let $\nu_{0,i}$ and $\nu_{1,i}$ denote the marginals of q_i. Assuming that (M,d,m) has curvature $\geqslant K$ *in the usual sense* then implies that for each $i\in\mathbf{N}$ there exists a midpoint $\tilde\eta_i$ of $\nu_{0,i}$ and $\nu_{1,i}$ satisfying

$$\mathrm{Ent}(\tilde\eta_i\,|\,m)\leqslant\frac{1}{2}\mathrm{Ent}(\nu_{0,i}\,|\,m)+\frac{1}{2}\mathrm{Ent}(\nu_{1,i}\,|\,m)-\frac{K}{8}\mathsf{d}_W^2(\nu_{0,i},\nu_{1,i}).$$

The $\tilde\eta_i$ for $i\in\mathbf{N}$ are mutually singular, since M is non-branching and since the q_i are mutually singular (Lemma 2.11 (iii)). Hence, $\tilde\eta=\sum_{i\in\mathbf{N}}\tilde\eta_i$ satisfies

$$\mathrm{Ent}(\tilde\eta\,|\,m)\leqslant\frac{1}{2}\mathrm{Ent}(\nu_0\,|\,m)+\frac{1}{2}\mathrm{Ent}(\nu_1\,|\,m)-\frac{K}{8}\mathsf{d}_W^2(\nu_0,\nu_1).$$

Moreover, $\mathsf{d}_W(\eta,\tilde\eta)\leqslant 2\varepsilon$ since for each $i\in\mathbf{N}$, $\mathrm{supp}[\eta_i]\subset B_\varepsilon(z_i)$ as well as $\mathrm{supp}[\tilde\eta_i]\subset B_\varepsilon(z_i)$, with z_i being the midpoint of x_i and y_i. By lower semicontinuity of $\mathrm{Ent}(\,\cdot\,|m)$ this implies

$$\mathrm{Ent}(\eta\,|\,m)\leqslant\frac{1}{2}\mathrm{Ent}(\nu_0\,|\,m)+\frac{1}{2}\mathrm{Ent}(\nu_1\,|\,m)-\frac{K}{8}\mathsf{d}_W^2(\nu_0,\nu_1).$$ □

THEOREM 4.9. (Riemannian spaces) *Let M be a complete Riemannian manifold with Riemannian distance* d *and Riemannian volume m and put $m'=e^{-V}m$, with a $\mathcal{C}^2$ function $V: M\to\mathbf{R}$. Then*

$$\underline{\mathsf{Curv}}(M,\mathsf{d},m')=\inf\{\mathrm{Ric}_M(\xi,\xi)+\mathrm{Hess}\,V(\xi,\xi):\xi\in TM \text{ and } |\xi|=1\}. \tag{4.15}$$

In particular, (M,d,m) has curvature $\geqslant K$ if and only if the Ricci curvature of M is $\geqslant K$.

Note that in the above Riemannian setting for each pair of points $\nu_0,\nu_1\in\mathcal{P}_2(M,\mathsf{d},m)$ there exists a *unique* geodesic connecting them. Hence, curvature bounds in the usual sense coincide with curvature bounds in the restricted sense. Moreover, note that in this setting, local curvature bounds always coincide with global curvature bounds.

Proof. Let us briefly sketch the main ideas of the proof, ignoring smoothness and regularity questions. For details, see [46] for the case $V=0$ and [51] for the general case.

Let $\nu_0=\varrho_0 m$ and $\nu_1=\varrho_1 m$ be given. According to Remark 2.12 (iii), there exists a function $\varphi: M\to\mathbf{R}$ such that

$$\nu_t=(F_t)_*\nu_0,$$

with

$$F_t(x)=\exp_x(-t\nabla\varphi(x)),$$

defines the unique geodesic $t\mapsto\nu_t$ in $\mathcal{P}_2(M,\mathsf{d})$ connecting ν_0 and ν_1. The change of variable formula then gives

$$\mathrm{Ent}(\nu_t\,|\,e^{-V}m)=\int\varrho_0\log\varrho_0\,dm-\int y_t\varrho_0\,dm+\int V(F_t)\varrho_0\,dm, \tag{4.16}$$

with $y_t=\log\det dF_t$ being the logarithm of the determinant of the Jacobian of F_t (in some weak sense). Now for ν_0-a.e. $x\in M$ the function $t\mapsto y_t(x)$ satisfies the differential inequality

$$\ddot{y}_t(x)\leqslant-\frac{1}{n}(\dot{y}_t)^2(x)-\mathrm{Ric}(\dot{F}_t(x),\dot{F}_t(x)). \tag{4.17}$$

Together with (4.16), this yields

$$\frac{\partial^2}{\partial t^2}\mathrm{Ent}(\nu_t\,|\,e^{-V}m)\geqslant\int[\mathrm{Ric}(\dot{F}_t,\dot{F}_t)+\mathrm{Hess}\,V(\dot{F}_t,\dot{F}_t)]\varrho_0\,dx\geqslant K\mathsf{d}_W^2(\nu_0,\nu_1),$$

provided $\mathrm{Ric}(\xi,\xi)+\mathrm{Hess}\,V(\xi,\xi)\geqslant K|\xi|^2$ for all $\xi\in TM$. This 'proves' the K-convexity of $\mathrm{Ent}(\,\cdot\,|e^{-V}m)$. □

Some of the most simple examples are the following.

Example 4.10. (i) If M is an n-dimensional Riemannian manifold of constant sectional curvature $\varkappa$ then

$$\underline{\mathsf{Curv}}(M,\mathsf{d},m)=(n-1)\varkappa.$$

(ii) If M is the Euclidean space $\mathbf{R}^n$ with the weighted measure

$$dm(x)=\exp\left(\frac{-K\|x\|^2}{2}\right)dx$$

then

$$\underline{\mathsf{Curv}}(M,\mathsf{d},m)=K.$$

(iii) If supp$[m]$ consists of one point then $\underline{\mathsf{Curv}}(M,\mathsf{d},m)=+\infty$.

Remark 4.11. If m is finite on all balls then it suffices to verify (4.11) for all $\nu_i=\varrho_i m$ with bounded density ϱ_i and with bounded support supp$[\nu_i]$, $i=0,1$.

Proof. Let $\nu_0,\nu_1\in\mathcal{P}_2^*(M,\mathsf{d},m)$ be given, say $\nu_0=\varrho_0 m$ and $\nu_1=\varrho_1 m$. Fix $o\in M$ and define for $i=0,1$,

$$\varrho_{i,R}=\frac{1}{\alpha_{i,R}}1_{B_R(o)}[\varrho_i\wedge R],\quad\text{with}\quad \alpha_{i,R}=\int_{B_R(o)}(\varrho_i\wedge R)\,dm.$$

Then, according to Remark 4.3, we have $\alpha_{i,R}\to 1$ and $\mathrm{Ent}(\varrho_{i,R}m|m)\to\mathrm{Ent}(\nu_i|m)$ as $R\to\infty$. Moreover, $\mathsf{d}_W(\varrho_{i,R}m,\nu_i)\to 0$ and thus for sufficiently large R, each $\varepsilon/2$-midpoint of $\varrho_{0,R}m$ and $\varrho_{1,R}m$ will be an ε-midpoint of ν_0 and ν_1. □

4.3. Basic transformations

Proposition 4.12. (Isomorphism) *If* (M,d,m) *and* (M',d',m') *are isomorphic metric measure spaces then*

$$\underline{\mathsf{Curv}}(M,\mathsf{d},m)=\underline{\mathsf{Curv}}(M',\mathsf{d}',m').$$

Thus $\underline{\mathsf{Curv}}(\,\cdot\,)$ *extends to a function on* $\mathbf{X}$*, the family of isomorphism classes of metric measure spaces.*

Analogous statements hold for $\underline{\mathsf{Curv}}_{\mathrm{lax}}(\,\cdot\,)$ and $\underline{\mathsf{Curv}}_{\mathrm{loc}}(\,\cdot\,)$.

Proof. Let $\Psi\colon M_0\to M_0'$ be an isometry between $M_0:=\mathrm{supp}[m]$ and $M_0':=\mathrm{supp}[m']$ such that $\Psi_*m=m'$. Then for all $\nu=\varrho m\in\mathcal{P}_2(M,\mathsf{d},m)$ the push-forward measure $\Psi_*\nu$ is absolutely continuous with respect to m' with density $\varrho(\Psi^{-1})$. Thus Ψ induces an isometry $\nu\mapsto\Psi_*\nu$ between $\mathcal{P}_2(M,\mathsf{d})$ and $\mathcal{P}_2(M',\mathsf{d}')$ which maps $\mathcal{P}_2(M,\mathsf{d},m)$ onto $\mathcal{P}_2(M',\mathsf{d}',m')$. Moreover,

$$\mathrm{Ent}(\Psi_*\nu\,|\,m')=\mathrm{Ent}(\nu\,|\,m).$$

This proves the claim. □

PROPOSITION 4.13. (Scaled spaces) *For each metric measure space* (M,d,m) *and all* $\alpha,\beta>0$, *we have*

$$\underline{\mathsf{Curv}}(M,\alpha\mathsf{d},\beta m)=\alpha^{-2}\,\underline{\mathsf{Curv}}(M,\mathsf{d},m).$$

Analogous statements hold for $\underline{\mathsf{Curv}}_{\mathrm{lax}}(\,\cdot\,)$ and $\underline{\mathsf{Curv}}_{\mathrm{loc}}(\,\cdot\,)$.

Proof. Clearly,

$$\mathrm{Ent}(\nu\,|\,\beta m)=\mathrm{Ent}(\nu\,|\,m)-\log\beta \quad\text{and}\quad (\alpha\cdot\mathsf{d})_W(\nu_0,\nu_1)=\alpha\cdot\mathsf{d}_W(\nu_0,\nu_1). \qquad \square$$

PROPOSITION 4.14. (Weighted spaces) *For each metric measure space* (M,d,m) *and each lower bounded, measurable function* $V\colon M\to\mathbf{R}$, *we have*

$$\underline{\mathsf{Curv}}(M,\mathsf{d},e^{-V}m)\geqslant\underline{\mathsf{Curv}}(M,\mathsf{d},m)+\underline{\mathrm{Hess}}\,V,$$

where $\underline{\mathrm{Hess}}\,V:=\sup\{K\in\mathbf{R}:\ V$ *is K-convex on* $\mathrm{supp}[m]\}$. *If V is locally bounded from below then an analogous statement holds for* $\underline{\mathsf{Curv}}_{\mathrm{loc}}(\,\cdot\,)$.

Recall that a function $V\colon M\to\,]-\infty,+\infty]$, defined on a geodesic space M, is called K-*convex*, for some $K\in\mathbf{R}$, if for each geodesic $\gamma\colon[0,1]\to M$ and for each $t\in[0,1]$ the following holds:

$$V(\gamma(t))\leqslant(1-t)V(\gamma(0))+tV(\gamma(1))-\frac{K}{2}t(1-t)\mathsf{d}^2(\gamma(0),\gamma(1)). \tag{4.18}$$

Proof. A simple calculation yields

$$\mathrm{Ent}\big(\nu\,|\,e^{-V}m\big)=\mathrm{Ent}(\nu\,|\,m)+\int V\,d\nu.$$

Moreover, $\mathcal{P}_2^*(M,\mathsf{d},e^{-V}m)\subset\mathcal{P}_2^*(M,\mathsf{d},m)$ by the lower boundedness of V. Now put $K_0:=\underline{\mathsf{Curv}}(M,\mathsf{d},m)$ and $K_1:=\underline{\mathrm{Hess}}\,V$. Given any geodesic $\Gamma\in\mathcal{P}_2^*(M,\mathsf{d},m)$ and any $t\in[0,1]$, choose an optimal coupling $\hat{q}$ on M^3 with marginals $\Gamma(0)$, $\Gamma(t)$ and $\Gamma(1)$ in the sense of Lemma 2.11 (ii). Then

$$\begin{aligned}
&\mathrm{Ent}(\Gamma(t)\,|\,e^{-V}m)-(1-t)\,\mathrm{Ent}(\Gamma(0)\,|\,e^{-V}m)-t\,\mathrm{Ent}(\Gamma(1)\,|\,e^{-V}m)\\
&\quad=\mathrm{Ent}(\Gamma(t)\,|\,m)-(1-t)\,\mathrm{Ent}(\Gamma(0)\,|\,m)-t\,\mathrm{Ent}(\Gamma(1)\,|\,m)\\
&\qquad+\int_M V\,d\Gamma(t)-(1-t)\int_M V\,d\Gamma(0)-t\int_M V\,d\Gamma(1)\\
&\quad=\mathrm{Ent}(\Gamma(t)\,|\,m)-(1-t)\,\mathrm{Ent}(\Gamma(0)\,|\,m)-t\,\mathrm{Ent}(\Gamma(1)\,|\,m)\\
&\qquad+\int_{M^3}\big[V(x_t)-(1-t)V(x_0)-V(x_1)\big]\,d\hat{q}(x_0,x_t,x_1)\\
&\quad\overset{(*)}{\leqslant}-\frac{K_0}{2}t(1-t)\mathsf{d}_W^2(\Gamma(0),\Gamma(1))-\int_{M^3}\frac{K_1}{2}t(1-t)\mathsf{d}^2(x_0,x_1)\,d\hat{q}(x_0,x_t,x_1)\\
&\quad=-\frac{K_0+K_1}{2}t(1-t)\mathsf{d}_W^2(\Gamma(0),\Gamma(1)).
\end{aligned}$$

The inequality $(*)$ follows from K_1-convexity of V, since for $\hat{q}$-a.e. (x_0, x_t, x_1) the point x_t lies on a geodesic connecting x_0 and x_1 (Lemma 2.11 (ii)). □

Proposition 4.15. (Subsets) *Let (M, d, m) be a metric measure space and let M' be a convex subset of M. Then*

$$\underline{\mathsf{Curv}}(M', \mathsf{d}, m) \geqslant \underline{\mathsf{Curv}}(M, \mathsf{d}, m).$$

Here a set $M' \subset M$ is called *convex* if and only if $\gamma_t \in M'$ for all $t \in [0,1]$ and all geodesics $\gamma: [0,1] \to M$ with endpoints $\gamma_0, \gamma_1 \in M'$.

Proof. Let ν_0 and ν_1 be probability measures in $\mathcal{P}_2^*(M', \mathsf{d}, m)$. Regard them as probability measures on M. Let Γ be a geodesic in $\mathcal{P}_2^*(M, \mathsf{d}, m)$ connecting them and satisfying (4.10). It remains to prove that for each $t \in [0,1]$ the measure $\Gamma(t)$ is supported by M', i.e.

$$\Gamma(t)(M \setminus M') = 0.$$

According to Lemma 2.11 (ii), there exists an optimal coupling $\hat{q}$ of $\Gamma(0) = \nu_0$, $\Gamma(t)$ and $\Gamma(1) = \nu_1$ such that for $\hat{q}$-a.e. $(x, z, y) \in M^3$ the point z lies on some geodesic connecting the points x and y. But then $\hat{q}$-almost surely z has to lie in M', since x and y lie in M' and the latter is assumed to be convex. This proves that $\Gamma(t)(M \setminus M') = 0$ and thus yields the claim for $\underline{\mathsf{Curv}}(M', \mathsf{d}, m)$. □

Proposition 4.16. (Products) *Let (M_i, d_i, m_i) for $i = 1, ..., l$ be metric measure spaces and*

$$(M, \mathsf{d}, m) = \bigotimes_{i=1}^{l} (M_i, \mathsf{d}_i, m_i).$$

Assume that M is non-branching and compact. Then

$$\underline{\mathsf{Curv}}(M, \mathsf{d}, m) = \inf_{i \in \{1, \ldots, l\}} \underline{\mathsf{Curv}}(M_i, \mathsf{d}_i, m_i). \tag{4.19}$$

Proof. (i) Let us first prove the inequality

$$\underline{\mathsf{Curv}}(M, \mathsf{d}, m) \leqslant \inf_{i \in \{1, \ldots, l\}} \underline{\mathsf{Curv}}(M_i, \mathsf{d}_i, m_i).$$

Assume that this is not true. Then for some $K \in \mathbf{R}$ and $i \in \{1, ..., l\}$,

$$\underline{\mathsf{Curv}}(M, \mathsf{d}, m) \geqslant K > \underline{\mathsf{Curv}}(M_i, \mathsf{d}_i, m_i). \tag{4.20}$$

Without restriction, we may assume that $i = 1$. Then the last inequality implies that there exist $\nu_0^{(1)}, \nu_1^{(1)} \in \mathcal{P}_2^*(M_1, \mathsf{d}_1, m_1)$ such that for each midpoint $\eta^{(1)}$ in $\mathcal{P}_2^*(M_1, \mathsf{d}_1, m_1)$ between $\nu_0^{(1)}$ and $\nu_1^{(1)}$, the inequality (4.12) is violated, i.e.

$$\operatorname{Ent}(\eta^{(1)} \,|\, m_1) > \frac{1}{2} \operatorname{Ent}(\nu_0^{(1)} \,|\, m_1) + \frac{1}{2} \operatorname{Ent}(\nu_1^{(1)} \,|\, m_1) - \frac{K}{8} \mathsf{d}_W^2(\nu_0^{(1)}, \nu_1^{(1)}). \tag{4.21}$$

Now for $j=0,1$ put $\nu_j := \nu_j^{(1)} \otimes \overline{m}_2 \otimes ... \otimes \overline{m}_l$ with normalized measures $\overline{m}_i = (1/m_i(M_i))m_i$ for $i=2, ..., l$. Then obviously

$$\mathrm{Ent}(\nu_j \,|\, m) = \mathrm{Ent}(\nu_j^{(1)}) - \sum_{i=2}^{l} \log m_i(M_i) \tag{4.22}$$

and $\nu_j \in \mathcal{P}_2^*(M, \mathsf{d}, m)$. Moreover

$$\mathsf{d}^W(\nu_0, \nu_1) = \mathsf{d}_1^W(\nu_0^{(1)}, \nu_1^{(1)})$$

(where for typographical reasons, we replace the lower index 'W' by an upper index, again denoting the L_2-Wasserstein distances derived from d and d_1, respectively).

Now the first inequality in (4.20) implies that there exists a midpoint η of ν_0 and ν_1 satisfying (4.12). According to Remark 2.2, it implies that

$$\mathsf{d}^W(\nu_0, \eta)^2 + \mathsf{d}^W(\eta, \nu_1)^2 \leqslant \tfrac{1}{2} \mathsf{d}^W(\nu_0, \nu_1)^2,$$

which in turn implies, for the first marginal $\eta^{(1)}$ of η, that

$$\mathsf{d}_1^W(\nu_0^{(1)}, \eta^{(1)})^2 + \mathsf{d}_1^W(\eta^{(1)}, \nu_1^{(1)})^2 \leqslant \tfrac{1}{2} \mathsf{d}_1^W(\nu_0^{(1)}, \nu_1^{(1)})^2.$$

Again, according to Remark 2.2, this yields that $\eta^{(1)}$ is a midpoint of $\nu_0^{(1)}$ and $\nu_1^{(1)}$. But (4.21) and (4.22) imply that (4.12) is violated, contradicting our previous assertion. Thus $\underline{\mathsf{Curv}}(M, \mathsf{d}, m) < K$, which proves our first claim.

(ii) To prove the reverse implication, we start with treating the particular case $\nu_0 = \nu_0^{(1)} \otimes ... \otimes \nu_0^{(l)}$ and $\nu_1 = \nu_1^{(1)} \otimes ... \otimes \nu_1^{(l)}$. Assume that $\underline{\mathsf{Curv}}(M_i, d_i, m_i) \geqslant K$ for each $i = 1, ..., l$. Then for each i there exists a midpoint $\eta^{(i)}$ of $\nu_0^{(i)}$ and $\nu_1^{(i)}$ with

$$\mathrm{Ent}(\eta^{(i)} \,|\, m_i) \leqslant \frac{1}{2} \mathrm{Ent}(\nu_0^{(i)} \,|\, m) + \frac{1}{2} \mathrm{Ent}(\nu_1^{(i)} \,|\, m) - \frac{K}{8} \mathsf{d}_i^W(\nu_0^{(i)}, \nu_1^{(i)})^2.$$

Put $\eta := \eta^{(1)} \otimes ... \otimes \eta^{(l)}$. Then η is a midpoint of ν_0 and ν_1, since

$$\mathsf{d}^W(\eta, \nu_0)^2 = \sum_{i=1}^{l} \mathsf{d}_i^W(\eta^{(i)}, \nu_0^{(i)})^2 \leqslant \sum_{i=1}^{l} \left[\frac{1}{2} \mathsf{d}_i^W(\nu_0^{(i)}, \nu_1^{(i)})\right]^2 = \left[\frac{1}{2} \mathsf{d}^W(\nu_0, \nu_1)\right]^2.$$

Moreover,

$$\mathrm{Ent}(\nu_0 \,|\, m) = \sum_{i=1}^{l} \mathrm{Ent}(\nu_0^{(i)} \,|\, m_i), \quad \mathrm{Ent}(\nu_1 \,|\, m) = \sum_{i=1}^{l} \mathrm{Ent}(\nu_1^{(i)} \,|\, m_i)$$

and

$$\mathrm{Ent}(\eta\,|\,m)=\sum_{i=1}^{l}\mathrm{Ent}(\eta^{(i)}\,|\,m_i).$$

Hence,

$$\begin{aligned}\mathrm{Ent}(\eta\,|\,m)&\leqslant\sum_{i=1}^{l}\left[\frac{1}{2}\mathrm{Ent}(\nu_0^{(i)}\,|\,m)+\frac{1}{2}\mathrm{Ent}(\nu_1^{(i)}\,|\,m)-\frac{K}{8}\mathsf{d}_i^W(\nu_0^{(i)},\nu_1^{(i)})^2\right]\\&\leqslant\frac{1}{2}\mathrm{Ent}(\nu_0\,|\,m)+\frac{1}{2}\mathrm{Ent}(\nu_1\,|\,m)-\frac{K}{8}\mathsf{d}^W(\nu_0,\nu_1)^2.\end{aligned}$$

This proves the claim in the particular case.

(iii) Now let arbitrary $\nu_0,\nu_1\in\mathcal{P}_2^*(M,\mathsf{d},m)$ and $\varepsilon>0$ be given. Then there exist

$$\tilde\nu_0=\frac{1}{n}\sum_{j=1}^{n}\nu_{0,j},$$

with mutually singular product measures $\nu_{0,j}$, $j=1,...,n$, and

$$\tilde\nu_1=\frac{1}{n}\sum_{j=1}^{n}\nu_{1,j},$$

with mutually singular product measures $\nu_{1,j}$, $j=1,...,n$, such that

$$\begin{aligned}\mathrm{Ent}(\tilde\nu_0\,|\,m)&\leqslant\mathrm{Ent}(\nu_0\,|\,m)+\varepsilon,\quad \mathsf{d}^W(\nu_0,\tilde\nu_0)\leqslant\varepsilon,\\ \mathrm{Ent}(\tilde\nu_1\,|\,m)&\leqslant\mathrm{Ent}(\nu_1\,|\,m)+\varepsilon,\quad \mathsf{d}^W(\nu_1,\tilde\nu_1)\leqslant\varepsilon\end{aligned}$$

and

$$\mathsf{d}^W(\tilde\nu_0,\tilde\nu_1)\geqslant\left[\frac{1}{n}\sum_{j=1}^{n}\mathsf{d}^W(\nu_{0,j},\nu_{1,j})^2\right]^{1/2}-\varepsilon.$$

Furthermore, since ν_0 is the sum of mutually singular $\nu_{0,j}$,

$$\mathrm{Ent}(\tilde\nu_0\,|\,m)=\frac{1}{n}\sum_{j=1}^{n}\mathrm{Ent}(\nu_{0,j}\,|\,m)-\log n$$

and, similarly,

$$\mathrm{Ent}(\tilde\nu_1\,|\,m)=\frac{1}{n}\sum_{j=1}^{n}\mathrm{Ent}(\nu_{1,j}\,|\,m)-\log n.$$

According to part (ii), for each $j=1,...,n$ there exists a midpoint η_j of $\nu_{0,j}$ and $\nu_{1,j}$ satisfying

$$\mathrm{Ent}(\eta_j\,|\,m)\leqslant\frac{1}{2}\mathrm{Ent}(\nu_{0,j}\,|\,m)+\frac{1}{2}\mathrm{Ent}(\nu_{1,j}\,|\,m)-\frac{K}{8}\mathsf{d}^W(\nu_{0,j},\nu_{1,j})^2.$$

According to Lemma 2.11 (iii), since M is non-branching and since the $\nu_{0,j}$ for $j=1, ..., n$ are mutually singular, also the η_j for $j=1, ..., n$ must be mutually singular. Hence,

$$\eta := \frac{1}{n}\sum_{j=1}^{n}\eta_j$$

satisfies

$$\mathrm{Ent}(\eta\,|\,m) = \frac{1}{n}\sum_{j=1}^{n}\mathrm{Ent}(\eta_j\,|\,m) - \log n$$

and thus

$$\begin{aligned}\mathrm{Ent}(\eta\,|\,m) &\leqslant \frac{1}{2}\mathrm{Ent}(\tilde\nu_0\,|\,m) + \frac{1}{2}\mathrm{Ent}(\tilde\nu_1\,|\,m) - \frac{K}{8}\frac{1}{n}\sum_{j=1}^{n}\mathsf{d}^W(\nu_{0,j},\nu_{1,j})^2 \\ &\leqslant \frac{1}{2}\mathrm{Ent}(\nu_0\,|\,m) + \frac{1}{2}\mathrm{Ent}(\nu_1\,|\,m) - \frac{K}{8}[\mathsf{d}^W(\nu_0,\nu_1) \mp 3\varepsilon]^2 + \varepsilon\end{aligned}$$

(with $\mp$ to be chosen according to the sign of K). Moreover, η is an approximate midpoint of ν_0 and ν_1:

$$\begin{aligned}2\mathsf{d}^W(\nu_0,\eta) &\leqslant 2\left[\frac{1}{n}\sum_{j=1}^{n}\mathsf{d}^W(\nu_{0,j},\eta_j)^2\right]^{1/2} \\ &\leqslant \left[\frac{1}{n}\sum_{j=1}^{n}\mathsf{d}^W(\nu_{0,j},\nu_{1,j})^2\right]^{1/2} \leqslant \mathsf{d}^W(\tilde\nu_0,\tilde\nu_1) + \varepsilon \leqslant \mathsf{d}^W(\nu_0,\nu_1) + 3\varepsilon\end{aligned}$$

and similarly for $\mathsf{d}^W(\nu_1,\eta)$. This proves that $\underline{\mathsf{Curv}}_{\mathrm{lax}}(M,\mathsf{d},m) \geqslant K$. Together with compactness of M this finally yields the claim. □

4.4. From local to global

A crucial implication of our definition of lower curvature bounds for metric measure spaces is the following globalization theorem which states that local curvature bounds imply global curvature bounds. This is in the spirit of the globalization theorem of Toponogov for lower curvature bounds (in the sense of Alexandrov) for metric spaces.

THEOREM 4.17. (Globalization) *Let (M,d,m) be a compact, non-branching metric measure space and assume that $\mathcal{P}_2^*(M,\mathsf{d},m)$ is a geodesic space. Then*

$$\underline{\mathsf{Curv}}(M,\mathsf{d},m) = \underline{\mathsf{Curv}}_{\mathrm{loc}}(M,\mathsf{d},m). \tag{4.23}$$

Proof. Put $K=\underline{\mathsf{Curv}}_{\mathrm{loc}}(M,\mathsf{d},m)$ and consider for each number $k\in\mathbf{N}\cup\{0\}$ the following property:

C(k): For each geodesic $\Gamma\colon[0,1]\to\mathcal{P}_2^*(M,\mathsf{d},m)$ and for each pair $s,t\in[0,1]$ with $0\leqslant t-s\leqslant 2^{-k}$ there exists a midpoint $\eta(s,t)$ of $\Gamma(s)$ and $\Gamma(t)$ such that

$$\mathrm{Ent}(\eta(s,t)\,|\,m)\leqslant\frac{1}{2}\mathrm{Ent}(\Gamma(s)\,|\,m)+\frac{1}{2}\mathrm{Ent}(\Gamma(t)\,|\,m)-\frac{K}{8}\mathsf{d}_W^2(\Gamma(s),\Gamma(t)). \tag{4.24}$$

Our first claim is:

- For each $k\in\mathbf{N}$, C(k) implies C($k-1$).

In order to prove this claim, let $k\in\mathbf{N}$ be given with property C(k), and let a geodesic Γ and numbers $s,t\in[0,1]$ be given with $0\leqslant t-s\leqslant 2^{1-k}$. Define iteratively a sequence $(\Gamma^{(i)})_{i\in\mathbf{N}}$ of geodesics in $\mathcal{P}_2^*(M,\mathsf{d},m)$ which coincide with Γ on $[0,s]\cup[t,1]$ as follows: start with $\Gamma^{(0)}:=\Gamma$; assuming that $\Gamma^{(2i)}$ is already given, let $\Gamma^{(2i+1)}\colon[0,1]\to\mathcal{P}_2^*(M,\mathsf{d},m)$ be any geodesic which coincides with Γ on $[0,s]\cup[t,1]$ and for which $\Gamma^{(2i+1)}(s+(t-s)/4)$ is a midpoint of $\Gamma(s)=\Gamma^{(2i)}(s)$ and $\Gamma^{(2i)}(s+(t-s)/2)$, and for which $\Gamma^{(2i+1)}(s+3(t-s)/4)$ is a midpoint of $\Gamma^{(2i)}(s+(t-s)/2)$ and $\Gamma(t)=\Gamma^{(2i)}(t)$ satisfying

$$\begin{aligned}\mathrm{Ent}\left(\Gamma^{(2i+1)}\left(s+\frac{t-s}{4}\right)\,\middle|\,m\right)\leqslant{}&\frac{1}{2}\mathrm{Ent}(\Gamma(s)\,|\,m)\\&+\frac{1}{2}\mathrm{Ent}\left(\Gamma^{(2i)}\left(s+\frac{t-s}{2}\right)\,\middle|\,m\right)-\frac{K}{32}\,\mathsf{d}_W^2(\Gamma(s),\Gamma(t))\end{aligned} \tag{4.25}$$

and

$$\begin{aligned}\mathrm{Ent}\left(\Gamma^{(2i+1)}\left(s+3\frac{t-s}{4}\right)\,\middle|\,m\right)\leqslant{}&\frac{1}{2}\mathrm{Ent}(\Gamma(t)\,|\,m)\\&+\frac{1}{2}\mathrm{Ent}\left(\Gamma^{(2i)}\left(s+\frac{t-s}{2}\right)\,\middle|\,m\right)-\frac{K}{32}\,\mathsf{d}_W^2(\Gamma(s),\Gamma(t)).\end{aligned} \tag{4.26}$$

Such midpoints exist by the assumption C(k).

Then let $\Gamma^{(2i+2)}\colon[0,1]\to\mathcal{P}_2^*(M,\mathsf{d},m)$ be any geodesic which coincides with Γ on $[0,s]\cup[t,1]$ and for which $\Gamma^{(2i+2)}(s+(t-s)/2)$ is a midpoint of $\Gamma^{(2i+1)}(s+(t-s)/4)$ and $\Gamma^{(2i+1)}(s+3(t-s)/4)$ satisfying

$$\begin{aligned}\mathrm{Ent}\left(\Gamma^{(2i+2)}\left(s+\frac{t-s}{2}\right)\,\middle|\,m\right)\leqslant{}&\frac{1}{2}\mathrm{Ent}\left(\Gamma^{(2i+1)}\left(s+\frac{t-s}{4}\right)\,\middle|\,m\right)\\&+\frac{1}{2}\mathrm{Ent}\left(\Gamma^{(2i+1)}\left(s+3\frac{t-s}{4}\right)\,\middle|\,m\right)\\&-\frac{K}{32}\,\mathsf{d}_W^2(\Gamma(s),\Gamma(t)).\end{aligned} \tag{4.27}$$

Again, such a midpoint exists by the assumption C(k). This yields a sequence of geodesics $\Gamma^{(i)}$, $i\in\mathbf{N}$.

Combining (4.25)–(4.27) gives

$$\operatorname{Ent}\left(\Gamma^{(2i+2)}\left(s+\frac{t-s}{2}\right)\,\middle|\, m\right) \leqslant \frac{1}{2}\operatorname{Ent}\left(\Gamma^{(2i)}\left(s+\frac{t-s}{2}\right)\,\middle|\, m\right) + \frac{1}{4}\operatorname{Ent}(\Gamma(s)\,|\,m)+\frac{1}{4}\operatorname{Ent}(\Gamma(t)\,|\,m)-\frac{K}{16}\mathsf{d}_W^2(\Gamma(s),\Gamma(t)).$$

By iteration, it yields

$$\operatorname{Ent}\left(\Gamma^{(2i)}\left(s+\frac{t-s}{2}\right)\,\middle|\, m\right) \leqslant 2^{-i}\operatorname{Ent}\left(\Gamma\left(s+\frac{t-s}{2}\right)\,\middle|\, m\right) +(1-2^{-i})\left[\frac{1}{2}\operatorname{Ent}(\Gamma(s)\,|\,m)+\frac{1}{2}\operatorname{Ent}(\Gamma(t)\,|\,m)-\frac{K}{8}\mathsf{d}_W^2(\Gamma(s),\Gamma(t))\right].$$

By compactness of $\mathcal{P}_2(M,\mathsf{d})$, there exists a subsequence of $(\Gamma^{(2i)}(s+(t-s)/2))_{i\in\mathbf{N}}$ converging to some $\eta\in\mathcal{P}_2(M,\mathsf{d})$. Continuity of the distance implies that η is a midpoint of $\Gamma(s)$ and $\Gamma(t)$ (since each of the $\Gamma^{(2i)}(s+(t-s)/2)$ is a midpoint) and lower semicontinuity of the relative entropy implies

$$\operatorname{Ent}(\eta\,|\,m)\leqslant\frac{1}{2}\operatorname{Ent}(\Gamma(s)\,|\,m)+\frac{1}{2}\operatorname{Ent}(\Gamma(t)\,|\,m)-\frac{K}{8}\mathsf{d}_W^2(\Gamma(s),\Gamma(t)).$$

This proves property C($k-1$).

Now, according to our curvature assumption, each point $x\in M$ has a neighborhood $M(x)$ such that probability measures in $\mathcal{P}_2^*(M,\mathsf{d},m)$ which are supported on $M(x)$ can be joined by geodesics in $\mathcal{P}_2^*(M,\mathsf{d},m)$ satisfying (4.10). By compactness of M, there exist $\lambda>0$, finitely many disjoint sets $L_1,...,L_n$ which cover M, and closed sets $M_j\supset B_\lambda(L_j)$, for $j=1,...,n$, such that probability measures in $\mathcal{P}_2^*(M,\mathsf{d},m)$ which are supported on M_j can be joined by geodesics in $\mathcal{P}_2^*(M,\mathsf{d},m)$ satisfying (4.10). Choose $k'\in\mathbf{N}$ such that

$$2^{-k'}\operatorname{diam}(M,\mathsf{d},m)\leqslant\lambda. \tag{4.28}$$

Our next claim is:

- Property C(k') is satisfied.

In order to prove this claim, fix Γ, s and t, and let $\hat{q}$ be a coupling of $\Gamma(0)$, $\Gamma(s)$, $\Gamma(t)$ and $\Gamma(1)$ on M^4. Then, according to Lemma 2.11, for $\hat{q}$-a.e. $(x_0,x_s,x_t,x_1)\in M^4$ the points x_s and x_t lie on some geodesic connecting x_0 and x_1 with

$$d(x_s,x_t)=|t-s|d(x_0,x_1)\leqslant 2^{-k'}\operatorname{diam}(M,\mathsf{d},m)\leqslant\lambda, \tag{4.29}$$

by (4.28). Define probability measures $\Gamma_j(s)$ and $\Gamma_j(t)$, for $j=1,...,n$, by

$$\Gamma_j(s)(A) := \frac{1}{\alpha_j}\Gamma(s)(A\cap L_j) = \frac{1}{\alpha_j}\hat{q}(M\times(A\cap L_j)\times M\times M)$$

and

$$\Gamma_j(t)(A) := \frac{1}{\alpha_j}\hat{q}(M\times L_j\times A\times M),$$

provided $\alpha_j:=\Gamma_s(L_j)\neq 0$. (Otherwise, define $\Gamma_j(s)$ and $\Gamma_j(t)$ arbitrarily.) Then

$$\operatorname{supp}[\Gamma_j(s)]\subset \bar{L}_j$$

which, together with (4.29), implies that

$$\operatorname{supp}[\Gamma_j(s)]\cup\operatorname{supp}[\Gamma_j(t)]\subset \overline{B}_\lambda(L_j)\subset M_j. \tag{4.30}$$

Therefore, for each $j\in\{1,...,n\}$, the assumption $\underline{\mathsf{Curv}}_{\mathrm{loc}}(M,\mathsf{d},m)\geqslant K$ can be applied to the probability measures $\Gamma_j(s),\Gamma_j(t)\in\mathcal{P}_2^*(M,\mathsf{d},m)$, both supported on M_j. It yields the existence of a midpoint $\eta_j(s,t)$ of them with the property

$$\operatorname{Ent}(\eta_j(s,t)\,|\,m)\leqslant\frac{1}{2}\operatorname{Ent}(\Gamma_j(s)\,|\,m)+\frac{1}{2}\operatorname{Ent}(\Gamma_j(t)\,|\,m)-\frac{K}{8}\mathsf{d}_W^2(\Gamma_j(s),\Gamma_j(t)). \tag{4.31}$$

Define

$$\eta(s,t) := \sum_{j=1}^n \alpha_j\eta_j(s,t).$$

Then $\eta(s,t)$ is a midpoint of $\Gamma(s)=\sum_{j=1}^n\alpha_j\Gamma_j(s)$ and $\Gamma(t)=\sum_{j=1}^n\alpha_j\Gamma_j(t)$. Moreover, since the $\Gamma_j(s)$, for $j=1,...,n$, are mutually singular and since M is non-branching, also the $\eta_j(s,t)$, for $j=1,...,n$, are mutually singular (see Lemma 2.11 (iii)). Hence, by (4.6),

$$\operatorname{Ent}(\eta(s,t)\,|\,m)=\sum_{j=1}^n\alpha_j\operatorname{Ent}(\eta_j(s,t)\,|\,m)+\sum_{j=1}^n\alpha_j\log\alpha_j \tag{4.32}$$

and

$$\operatorname{Ent}(\Gamma(s)\,|\,m)=\sum_{j=1}^n\alpha_j\operatorname{Ent}(\Gamma_j(s)\,|\,m)+\sum_{j=1}^n\alpha_j\log\alpha_j, \tag{4.33}$$

whereas

$$\operatorname{Ent}(\Gamma(t)\,|\,m)\geqslant\sum_{j=1}^n\alpha_j\operatorname{Ent}(\Gamma_j(t)\,|\,m)+\sum_{j=1}^n\alpha_j\log\alpha_j, \tag{4.34}$$

since the $\Gamma_j(t)$, for $j=1,...,n$, are not necessarily mutually singular. Summing up (4.31) over $j=1,...,n$ and using (4.32)–(4.34) yields (4.24). This proves property $\mathrm{C}(k')$.

In order to finish the proof of the theorem, let two probability measures $\nu_0, \nu_1 \in \mathcal{P}_2^*(M, \mathsf{d}, m)$ be given. By assumption, there exists a geodesic Γ in $\mathcal{P}_2^*(M, \mathsf{d}, m)$ connecting them. According to our second claim, property $\mathrm{C}(k')$ is satisfied and according to our first claim, this implies property $\mathrm{C}(k)$ for all $k=k'-1, k'-2, ..., 0$. Property $\mathrm{C}(0)$ finally states that there exists a midpoint η of $\Gamma(0)$ and $\Gamma(1)$ with

$$\mathrm{Ent}(\eta \,|\, m) \leqslant \frac{1}{2}\,\mathrm{Ent}(\Gamma(0) \,|\, m) + \frac{1}{2}\,\mathrm{Ent}(\Gamma(1) \,|\, m) - \frac{K}{8}\mathsf{d}_W^2(\Gamma(0), \Gamma(1)). \tag{4.35}$$

This proves the theorem. □

Remark 4.18. Let M be a compact space.

(i) The condition on $\mathcal{P}_2^*(M, \mathsf{d}, m)$ to be a geodesic space is always satisfied if $\underline{\mathsf{Curv}}(M, \mathsf{d}, m) > -\infty$ (Remark 4.6 (iii)).

(ii) If $\mathcal{P}_2^*(M, \mathsf{d}, m)$ is a geodesic space then $\mathrm{supp}[m]$ is a geodesic space. The converse is not true in general; however, we conjecture that it is true under the additional assumption $\underline{\mathsf{Curv}}_{\mathrm{loc}}(M, \mathsf{d}, m) > -\infty$.

(iii) If $M_0 := \mathrm{supp}[m]$ is a geodesic space then $\mathcal{P}_2(M_0, \mathsf{d})$ is a geodesic space . Moreover, the space $\mathcal{P}_2^*(M_0, \mathsf{d}, m)$ is dense in $\mathcal{P}_2(M_0, \mathsf{d})$. Indeed, given any $\mu \in \mathcal{P}_2(M_0, \mathsf{d})$ and any $\varepsilon > 0$ there exist $n \in \mathbf{N}$ and $x_1, ..., x_n \in M_0$ such that $\mathsf{d}_W(\mu, \mu') \leqslant \varepsilon$, where

$$\mu' := \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}.$$

Moreover, $\mathsf{d}_W(\mu', \mu'') \leqslant \varepsilon$, where

$$\mu'' := \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m(B_\varepsilon(x_i))} \cdot 1_{B_\varepsilon(x_i)}\, m$$

and

$$\mathrm{Ent}(\mu'' \,|\, m) \leqslant \sup_{x \in M_0} \log \left[\frac{1}{n} \sum_{i=1}^{n} \frac{1}{m(B_\varepsilon(x_i))} \cdot 1_{B_\varepsilon(x_i)}(x) \right] \leqslant - \inf_{i \in \{1, ..., n\}} \log\, m(B_\varepsilon(x_i)) < \infty.$$

That is, $\mu'' \in \mathcal{P}_2^*(M_0, \mathsf{d}, m)$ and $\mathsf{d}_W(\mu, \mu'') \leqslant 2\varepsilon$, which proves the density.

4.5. Stability under convergence

One of the most important results in this paper is that our curvature bounds for metric measure spaces are stable under convergence. The key to this result is the fact that we are able to construct a transformation Q' from the L_2-Wasserstein space over one metric measure space (M, d, m) to the L_2-Wasserstein space over any other metric measure space

(M',d',m') which reduces the relative entropy and which is almost an isometry between these spaces, provided the underlying spaces are close to each other in the metric $\mathbf{D}$.

Actually, this is quite easy in the particular case where m is the push-forward of m' under a map $\psi'\colon M'\to M$. In this case we can define (similarly to the construction in the proof of Proposition 4.12)

$$\begin{aligned} Q'\colon \mathcal{P}_2(M,\mathsf{d},m) &\longrightarrow \mathcal{P}_2(M',\mathsf{d}',m') \\ \varrho m &\longmapsto (\varrho\circ\psi')m'. \end{aligned}$$

The general case is more subtle, since we may not restrict ourselves to transformations derived from push-forward maps. For instance, if m' is the Riemannian measure of a collapsed space, then the Riemannian measure m of the initial space cannot be represented as a push-forward measure.

Given two normalized metric measure spaces (M,d,m) and (M',d',m') we will define a canonical map

$$Q'\colon \mathcal{P}_2(M,\mathsf{d},m) \longrightarrow \mathcal{P}_2(M',\mathsf{d}',m')$$

as follows: Let q be a coupling of m and m' and $\hat{\mathsf{d}}$ be a coupling of d and d' such that

$$\int \hat{\mathsf{d}}^2(x,x')\,dq(x,x') \leqslant 2\mathbf{D}^2((M,\mathsf{d},m),(M',\mathsf{d}',m')).$$

Let Q' and Q be the disintegrations of q with respect to m' and m, respectively, that is,

$$dq(x,x') = Q'(x',dx)\,dm'(x') = Q(x,dx')\,dm(x)$$

and let $\hat{L}$ denote the m-essential supremum of the map

$$x \longmapsto \left[\int_{M'} \hat{\mathsf{d}}^2(x,x')\,Q(x,dx')\right]^{1/2}.$$

In general, of course, $\hat{L}$ may attain the value ∞. However, if both metric measure spaces have finite diameter we have

$$\hat{L} \leqslant \operatorname{diam}(M,\mathsf{d},m) + \operatorname{diam}(M',\mathsf{d}',m') < \infty.$$

For $\nu=\varrho m\in\mathcal{P}_2(M,\mathsf{d},m)$ define $Q'(\nu)\in\mathcal{P}_2(M',\mathsf{d}',m')$ by $Q'(\nu):=\varrho' m'$, where

$$\varrho'(x') := \int_M \varrho(x) Q'(x',dx). \tag{4.36}$$

In other words, for all measurable $A\subset M'$,

$$\begin{aligned}Q'(\nu)(A) &= \int_{M'} 1_A(x')\varrho'(x')\,dm'(x')\\ &= \int_{M'}\int_M 1_A(x')\varrho(x)Q'(x',dx)\,dm'(x')\\ &= \int_{M\times M'} 1_A(x')\varrho(x)\,dq(x,x').\end{aligned}$$

LEMMA 4.19. *The map Q' defined as above satisfies $Q'(m)=m'$ and, for all $\nu=\varrho m$,*

$$\mathrm{Ent}(Q'(\nu)\,|\,m')\leqslant \mathrm{Ent}(\nu\,|\,m) \tag{4.37}$$

and

$$\mathsf{d}_W^2(\nu,Q'(\nu))\leqslant \frac{2+\hat{L}^2\,\mathrm{Ent}(\nu\,|\,m)}{-\log \mathbf{D}((M,\mathsf{d},m),(M',\mathsf{d}',m'))} \tag{4.38}$$

provided $\mathbf{D}((M,\mathsf{d},m),(M',\mathsf{d}',m'))<1$.

Proof. Inequality (4.37) is a consequence of Jensen's inequality, applied to the convex function $r\mapsto r\log r$, as follows

$$\begin{aligned}\mathrm{Ent}(Q'(\nu)\,|\,m') &= \int \varrho'\log\varrho'\,dm'\\ &= \int\biggl[\int \varrho(x)Q'(x',dx)\biggr]\log\biggl[\int \varrho(x)Q'(x',dx)\biggr]\,dm'(x')\\ &\leqslant \iint \varrho(x)\log\varrho(x)Q'(x',dx)\,dm'(x')\\ &= \int \varrho(x)\log\varrho(x)\,dm(x)\\ &= \mathrm{Ent}(\nu\,|\,m).\end{aligned}$$

Inequality (4.38) follows from the fact that the measure

$$\varrho(x)\,dq(x,x') = \varrho(x)Q(x,dx')\,dm(x)$$

is a coupling of $\varrho(x)\,dm(x)$ and

$$\varrho'(x')\,dm'(x') = \int_M \varrho(x)Q'(x',dx)\,dm'(x') = \int_M \varrho(x)\,dq(x,x')$$

and thus

$$\mathsf{d}_W^2(\varrho m,\varrho' m')\leqslant \iint \hat{\mathsf{d}}^2(x,x')Q(x,dx')\varrho(x)\,dm(x) =: \Phi(\varrho).$$

Now, again with Jensen's inequality applied to the convex function $\psi(r) := r\log r$,

$$\psi\left(\frac{\Phi(\varrho)}{\Phi(1)}\right) \leqslant \frac{1}{\Phi(1)} \iint \hat{\mathsf{d}}^2(x,x')Q(x,dx')\psi(\varrho(x))\,dm(x) \leqslant \frac{\hat{L}^2}{\Phi(1)}\operatorname{Ent}(\nu\,|\,m).$$

Hence, since by assumption $\Phi(1)\leqslant 2\mathbf{D}((M,\mathsf{d},m),(M',\mathsf{d}',m'))<2$,

$$\begin{aligned}
\mathsf{d}_W^2(\varrho m, \varrho' m') &\leqslant \Phi(\varrho)\\
&\leqslant \Phi(1)\psi^{-1}(\hat{L}^2\operatorname{Ent}(\nu\,|\,m)/\Phi(1))\\
&\leqslant \Phi(1)\psi^{-1}([2+\hat{L}^2\operatorname{Ent}(\nu\,|\,m)]/\Phi(1))\\
&= \frac{2+\hat{L}^2\operatorname{Ent}(\nu\,|\,m)}{\log\psi^{-1}([2+\hat{L}^2\operatorname{Ent}(\nu\,|\,m)]/\Phi(1))}\\
&\leqslant 2\frac{2+\hat{L}^2\operatorname{Ent}(\nu\,|\,m)}{\log([2+\hat{L}^2\operatorname{Ent}(\nu\,|\,m)]/\Phi(1))}\\
&\leqslant 2\frac{2+\hat{L}^2\operatorname{Ent}(\nu\,|\,m)}{\log([2+\hat{L}^2\operatorname{Ent}(\nu\,|\,m)]/2\mathbf{D}^2)}\\
&\leqslant \frac{2+\hat{L}^2\operatorname{Ent}(\nu\,|\,m)}{-\log\mathbf{D}},
\end{aligned}$$

where we have used the abbreviation $\mathbf{D} := \mathbf{D}((M,\mathsf{d},m),(M',\mathsf{d}',m'))$. □

THEOREM 4.20. (Convergence) *Let $((M_n,\mathsf{d}_n,m_n))_{n\in\mathbf{N}}$ be a sequence of normalized metric measure spaces with uniformly bounded diameter. If*

$$(M_n,\mathsf{d}_n,m_n) \xrightarrow{\mathbf{D}} (M,\mathsf{d},m)$$

as $n\to\infty$, then

$$\limsup_{n\to\infty} \underline{\mathsf{Curv}}_{\mathrm{lax}}(M_n,\mathsf{d}_n,m_n) \leqslant \underline{\mathsf{Curv}}_{\mathrm{lax}}(M,\mathsf{d},m).$$

In particular, for each $K\in\mathbf{R}$ and $L\in\mathbf{R}_+$ the family $\mathbf{X}_1(K,L)$ of isomorphism classes of normalized metric measure spaces with curvature $\geqslant K$ in the lax sense and diameter $\leqslant L$ is closed with respect to $\mathbf{D}$.

Remarks. (i) As an obvious corollary to Theorem 4.20 and Lemma 3.18 we obtain that our curvature bounds are also preserved under measured Gromov–Hausdorff convergence. As usual, of course, then one has to restrict to compact spaces.

(ii) Another corollary (to Theorems 4.20 and 3.16) is that for each triple of real numbers K, C and L the family of all isomorphism classes of normalized metric measure spaces with curvature $\geqslant K$, doubling constant $\leqslant C$ and diameter $\leqslant L$ is $\mathbf{D}$-compact.

Proof of Theorem 4.20. Let $((M_n,\mathsf{d}_n,m_n))_{n\in\mathbf{N}}$ be a sequence in $(\mathbf{X}_1,\mathbf{D})$ with

$$(M_n,\mathsf{d}_n,m_n)\longrightarrow(M,\mathsf{d},m)$$

and assume that $\operatorname{diam}(M,\mathsf{d},m)\leqslant L$ and $\underline{\mathsf{Curv}}_{\mathrm{lax}}(M_n,\mathsf{d}_n,m_n)\geqslant K$ for some $L,K\in\mathbf{R}$ and all sufficiently large $n\in\mathbf{N}$. Now let $\varepsilon>0$ and $\nu_0=\varrho_0 m,\nu_1=\varrho_1 m\in\mathcal{P}_2^*(M,\mathsf{d},m)$ be given. Choose R with

$$\sup_{i=0,1}\operatorname{Ent}(\nu_i\,|\,m)+\frac{|K|}{8}[\mathsf{d}_W(\nu_0,\nu_1)+2\varepsilon]^2+\varepsilon\leqslant R. \tag{4.39}$$

We have to deduce the existence of an ε-midpoint η which satisfies inequality (4.11). Choose $n\in\mathbf{N}$ with

$$\mathbf{D}((M_n,\mathsf{d}_n,m_n),(M,\mathsf{d},m))\leqslant\exp\biggl(-\frac{2+(L+L')^2R}{\varepsilon^2}\biggr). \tag{4.40}$$

Define the map $Q_n'\colon\mathcal{P}_2(M,\mathsf{d},m)\to\mathcal{P}_2(M_n,\mathsf{d}_n,m_n)$ as in the previous lemma, now with m_n in the place of m', and analogously the map $Q_n\colon\mathcal{P}_2(M_n,\mathsf{d}_n,m_n)\to\mathcal{P}_2(M,\mathsf{d},m)$. Put

$$\nu_{i,n}:=Q_n'(\nu_i)=\varrho_{i,n}m_n,$$

with $\varrho_{i,n}(y)=\int\varrho_i(x)Q_n'(y,dx)$ for $i=0,1$, and let η_n be an ε-midpoint of $\nu_{0,n}$ and $\nu_{1,n}$, with

$$\operatorname{Ent}(\eta_n\,|\,m_n)\leqslant\frac12\operatorname{Ent}(\nu_{0,n}\,|\,m_n)+\frac12\operatorname{Ent}(\nu_{1,n}\,|\,m_n)-\frac{K}{8}\mathsf{d}_W^2(\nu_{0,n},\nu_{1,n})+\varepsilon. \tag{4.41}$$

From (4.38)–(4.40) we conclude that

$$\begin{aligned}\mathsf{d}_W^2(\nu_0,\nu_{0,n})&\leqslant\frac{2+\hat{L}^2\operatorname{Ent}(\nu_0\,|\,m)}{-\log\mathbf{D}((M,\mathsf{d},m),(M_n,\mathsf{d}_n,m_n))}\\&\leqslant\frac{2+(L+L')^2R}{-\log\mathbf{D}((M,\mathsf{d},m),(M_n,\mathsf{d}_n,m_n))}\leqslant\varepsilon^2\end{aligned}$$

and analogously $\mathsf{d}_W^2(\nu_1,\nu_{1,n})\leqslant\varepsilon^2$. Moreover, (4.37) and (4.41) imply

$$\begin{aligned}\operatorname{Ent}(\eta_n\,|\,m_n)&\leqslant\frac12\operatorname{Ent}(\nu_{0,n}\,|\,m_n)+\frac12\operatorname{Ent}(\nu_{1,n}\,|\,m_n)-\frac{K}{8}\mathsf{d}_W^2(\nu_{0,n},\nu_{1,n})+\varepsilon\\&\leqslant\frac12\operatorname{Ent}(\nu_0\,|\,m)+\frac12\operatorname{Ent}(\nu_1\,|\,m)-\frac{K}{8}\mathsf{d}_W^2(\nu_0,\nu_1)+\varepsilon',\end{aligned}$$

with $\varepsilon'=[1+(|K|/2)(\mathsf{d}_W(\nu_0,\nu_1)+\varepsilon)]\varepsilon$. Finally, put

$$\eta=Q_n(\eta_n).$$

Then again by (4.38)–(4.40) and by the previous estimate for $\mathrm{Ent}(\eta_n|m)$,

$$\mathsf{d}_W^2(\eta_n,\eta) \leqslant \frac{2+\hat{L}^2\,\mathrm{Ent}(\eta_n\,|\,m)}{-\log \mathbf{D}((M,\mathsf{d},m),(M_n,\mathsf{d}_n,m_n))} \leqslant \frac{2+(L+L')^2 R}{-\log \mathbf{D}((M,\mathsf{d},m),(M_n,\mathsf{d}_n,m_n))} \leqslant \varepsilon^2.$$

Hence,

$$\sup_{i=0,1} \mathsf{d}_W(\eta,\nu_i) \leqslant \frac{1}{2}\mathsf{d}_W(\nu_0,\nu_1)+4\varepsilon,$$

i.e. η is a 4ε-midpoint of ν_0 and ν_1. Furthermore, by (4.37),

$$\mathrm{Ent}(\eta\,|\,m) \leqslant \mathrm{Ent}(\eta_n\,|\,m_n) \leqslant \frac{1}{2}\mathrm{Ent}(\nu_0\,|\,m)+\frac{1}{2}\mathrm{Ent}(\nu_1\,|\,m)-\frac{K}{8}\mathsf{d}_W^2(\nu_0,\nu_1)+\varepsilon',$$

with ε' as above. This proves that $\underline{\mathsf{Curv}}_{\mathrm{lax}}(M,\mathsf{d},m)\geqslant K$. □

As an immediate consequence of Theorem 4.20, together with Proposition 4.16, we obtain the following result.

Corollary 4.21. (Infinite products) *Let*

$$(M,\mathsf{d},m) = \bigotimes_{n\in\mathbf{N}} (M_n,\mathsf{d}_n,m_n),$$

where (M_n,d_n,m_n) for $n\in\mathbf{N}$ are normalized metric measure spaces with compact non-branching M_n. Assume that $\sum_{n\in\mathbf{N}} \mathrm{Var}(M_n,\mathsf{d}_n,m_n)<\infty$. Then

$$\underline{\mathsf{Curv}}(M,\mathsf{d},m) = \inf_{n\in\mathbf{N}} \underline{\mathsf{Curv}}(M_n,\mathsf{d}_n,m_n).$$

Important infinite-dimensional examples are given by abstract Wiener spaces. Let (M,H,m) be an abstract Wiener space, that is, M is a separable Banach space, m is a Gaussian measure on M, and H is a separable Hilbert space that is continuously and densely embedded in M, such that

$$\int_M \exp(i\langle x,y\rangle)\,dm(x) = \exp\bigl(-\tfrac{1}{2}\|y\|_H^2\bigr)$$

for any $y\in M^*\subset H$ (where we identify H with its dual). For the classical Wiener space, $M=\mathcal{C}(\mathbf{R}_+,\mathbf{R})$ is the path space of 1-dimensional Brownian motion,

$$H = \{u\in M : u \text{ is absolutely continuous with } \textstyle\int_{\mathbf{R}_+} |\dot u(t)|^2\,dt<\infty\}$$

is the Cameron–Martin space, and m is the Wiener measure.

Given any abstract Wiener space (M,H,m), define a pseudo-metric on M by

$$\mathsf{d}(x,y):=\begin{cases}\|x-y\|_H, & \text{if } x-y\in H,\\ \infty, & \text{otherwise,}\end{cases}$$

and consider the *pseudo-metric measure space* (M,d,m). Of course, formally this does not fit in our framework. Nevertheless, the definition of the L_2-Wasserstein distance d_W derived from this pseudo-metric d perfectly makes sense. It is a pseudo-metric on the space of probability measures on M and a metric on the subspace of all those probability measures which have finite d_W-distance from m. Also, the relative entropy $\mathrm{Ent}(\cdot\,|m)$ and the curvature bound $\underline{\mathsf{Curv}}(M,\mathsf{d},m)$ are well defined. Particular attention, however, has to be paid to the fact that d_W is not continuous with respect to the weak topology of measures on M; it behaves very singular. For a detailed analysis, we refer to [17] and [18]. Here we restrict ourselves to the following result.

Proposition 4.22. (Wiener space)

$$\underline{\mathsf{Curv}}_{\mathrm{lax}}(M,\mathsf{d},m)\geqslant 1.$$

Proof. Let $\nu_0,\nu_1\in\mathcal{P}_2^*(M,\mathsf{d},m)$ and $\varepsilon>0$ be given. Choose an increasing (total) sequence $(M_n)_{n\in\mathbf{N}}$ of regular finite-dimensional subspaces with $\bigcup_{n\in\mathbf{N}}M_n$ being dense in M and H. Let m_n, $\nu_{0,n}$ and $\nu_{1,n}$ be the image measures of m, ν_0 and ν_1, respectively, under the projections $\pi_n: M\to M_n$. Then

$$\mathsf{d}_W(\nu_{0,n},\nu_0)\to 0,\quad \mathsf{d}_W(\nu_{1,n},\nu_1)\to 0$$

and

$$\mathrm{Ent}(\nu_{0,n}\,|\,m_n)\to\mathrm{Ent}(\nu_0\,|\,m),\quad \mathrm{Ent}(\nu_{1,n}\,|\,m_n)\to\mathrm{Ent}(\nu_1\,|\,m)$$

as $n\to\infty$. The space (M_n,M_n,m_n) is a finite-dimensional abstract Wiener space. Thus, it is isomorphic to $(\mathbf{R}^N,\mathbf{R}^N,\exp(-\|x\|^2/2)\,dx)$ for some $N=N(n)\in\mathbf{N}$ (where dx denotes the Lebesgue measure in $\mathbf{R}^N$). Hence, according to Theorem 4.9,

$$\underline{\mathsf{Curv}}(M_n,\mathsf{d},m_n)=1.$$

Thus, for each $n\in\mathbf{N}$, there exists a midpoint η_n of $\nu_{0,n}$ and $\nu_{1,n}$ with

$$\begin{aligned}\mathrm{Ent}(\eta_n\,|\,m)&=\mathrm{Ent}(\eta_n\,|\,m_n)\\ &\leqslant\frac{1}{2}\mathrm{Ent}(\nu_{0,n}\,|\,m_n)+\frac{1}{2}\mathrm{Ent}(\nu_{1,n}\,|\,m_n)-\frac{K}{8}\mathsf{d}_W^2(\nu_{0,n},\nu_{1,n})\\ &\leqslant\frac{1}{2}\mathrm{Ent}(\nu_0\,|\,m)+\frac{1}{2}\mathrm{Ent}(\nu_1\,|\,m)-\frac{K}{8}\mathsf{d}_W^2(\nu_0,\nu_1)+\varepsilon\end{aligned}$$

for n large enough. This proves the claim, since η_n is an ε-midpoint of ν_0 and ν_1 (again, for large enough n). □

4.6. Volume growth estimates

In the Riemannian setting, it is well known that lower bounds for the Ricci curvature of the underlying space imply upper bounds for the growth

$$R \longmapsto m(\overline{B}_R(x))$$

of the volume of concentric balls. In particular, this growth is at most exponential in R. This is the content of the well-known Bishop volume estimate.

Also, for general metric measure spaces, lower bounds for the curvature will imply upper estimates for the volume growth of concentric balls. These estimates, however, have to take into account that in the general case (without any dimensional restriction) the volume can grow much faster than exponentially. For instance, already in the following standard example we observe squared exponential volume growth.

Example 4.23. Let (M, d) be the 1-dimensional Euclidean space equipped with the measure $dm(x) = \exp(-(K/2)x^2)\,dx$ for some $K \in \mathbf{R}$. Then $\underline{\mathsf{Curv}}(M, \mathsf{d}, m) = K$ and, if $K < 0$,

$$m(\overline{B}_R(x)) \geqslant \exp\left(\frac{|K|}{2}\left(R - \frac{1}{2}\right)^2\right)$$

for each $x \in M$ and $R \geqslant \frac{1}{2}$.

THEOREM 4.24. *Let* (M, d, m) *be an arbitrary metric measure space satisfying* $\underline{\mathsf{Curv}}(M, \mathsf{d}, m) \geqslant K$ *for some* $K \leqslant 0$. *For fixed* $x \in \operatorname{supp}[m] \subset M$ *consider the volume growth*

$$v_R := m(\overline{B}_R(x))$$

of closed balls centered at x. *Then for all* $R \geqslant 2\varepsilon > 0$,

$$v_R \leqslant v_{2\varepsilon}\left(\frac{v_{2\varepsilon}}{v_\varepsilon}\right)^{R/\varepsilon} \exp\left(\frac{|K|}{2}\left(R + \frac{\varepsilon}{2}\right)^2\right). \tag{4.42}$$

In particular, each ball in M *has finite volume.*

Proof. Apply the following lemma with $r = \varepsilon$. □

LEMMA 4.25. *Let* (M, d, m), K *and* x *be as in the above theorem. Then for all* $\varepsilon, R > 0$ *and all* $t \in]0, 1]$,

$$\log v_R \leqslant \frac{1}{t} \log v_{\varepsilon + t(R+\varepsilon)} + \left(1 - \frac{1}{t}\right) \log v_\varepsilon + \frac{|K|}{2}(1-t)(R+\varepsilon)^2. \tag{4.43}$$

In other words, for all $\varepsilon, r > 0$ *and all* $R > \varepsilon + r$,

$$v_R \leqslant v_\varepsilon \left(\frac{v_{\varepsilon + r}}{v_\varepsilon}\right)^{(R+\varepsilon)/r} \exp\left(\frac{|K|}{2}(R + \varepsilon - r)(R + \varepsilon)\right). \tag{4.44}$$

Proof. Fix $x\in\mathrm{supp}[m]$ and $\varepsilon, R>0$. Let ν_0 and ν_1 denote the uniform distributions in $\overline{B}_\varepsilon(x)$ and $\overline{B}_R(x)$, respectively. That is,

$$d\nu_0(x)=\frac{1}{v_\varepsilon}\cdot 1_{\overline{B}_\varepsilon(x)}\,dm(x) \quad\text{and}\quad d\nu_1(x)=\frac{1}{v_R}\cdot 1_{\overline{B}_R(x)}\,dm(x).$$

Then, obviously, $\nu_0,\nu_1\in\mathcal{P}_2^*(M,\mathsf{d},m)$, with

$$\mathrm{Ent}(\nu_0\,|\,m)=-\log v_\varepsilon \quad\text{and}\quad \mathrm{Ent}(\nu_1\,|\,m)=-\log v_R.$$

Let ν_t, $t\in[0,1]$, be a geodesic in $\mathcal{P}_2^*(M,\mathsf{d},m)$ connecting ν_0 and ν_1 such that

$$\mathrm{Ent}(\nu_t\,|\,m)\leqslant(1-t)\,\mathrm{Ent}(\nu_0\,|\,m)+t\,\mathrm{Ent}(\nu_1\,|\,m)-\frac{K}{2}t(1-t)\mathsf{d}_W^2(\nu_0,\nu_1).$$

Such a geodesic exists according to our curvature assumption. Since $\mathsf{d}_W(\nu_0,\delta_x)\leqslant\varepsilon$ and $\mathsf{d}_W(\nu_1,\delta_x)\leqslant R$ it follows that

$$\mathsf{d}_W(\nu_0,\nu_1)\leqslant R+\varepsilon. \tag{4.45}$$

Moreover, if $\hat{q}$ is an optimal coupling of ν_0, ν_t and ν_1, then for $\hat{q}$-a.e. (y_0,y_t,y_1) the point y_t lies on a geodesic connecting y_0 and y_1, with $\mathsf{d}(y_0,y_t)=t\mathsf{d}(y_0,y_1)$. Together with inequality (4.45), the latter implies

$$\mathrm{supp}[\nu_t]\subset\overline{B}_{\varepsilon+t(R+\varepsilon)}(x). \tag{4.46}$$

Now, according to Jensen's inequality, for all ν_t satisfying (4.46),

$$\mathrm{Ent}(\nu_t\,|\,m)\geqslant\mathrm{Ent}(m_t\,|\,m),$$

where

$$m_t:=\frac{1}{v_{\varepsilon+t(R+\varepsilon)}}1_{\overline{B}_{\varepsilon+t(R+\varepsilon)}(x)}m$$

denotes uniform distribution in the closed ball $\overline{B}_{\varepsilon+t(R+\varepsilon)}(x)$. Hence,

$$\begin{aligned}
-\log v_{\varepsilon+t(R+\varepsilon)} &= \mathrm{Ent}(m_t\,|\,m)\\
&\leqslant \mathrm{Ent}(\nu_t\,|\,m)\\
&\leqslant (1-t)\,\mathrm{Ent}(\nu_0\,|\,m)+t\,\mathrm{Ent}(\nu_1\,|\,m)-\frac{K}{2}t(1-t)\mathsf{d}_W^2(\nu_0,\nu_1)\\
&\leqslant -(1-t)\log v_\varepsilon-t\log v_R+\frac{|K|}{2}t(1-t)(R+\varepsilon)^2.
\end{aligned}$$

This proves the first claim. For the second claim, choose $t=r/(R+\varepsilon)$ and apply the first claim. □

Slightly modifying the previous arguments also yields estimates for the volume of spherical shells

$$v_{R,\delta} := m(\overline{B}_R(x)\setminus B_{R-\delta}(x)).$$

Let ν_1 denote uniform distribution in the shell $\overline{B}_R(x)\setminus B_{R-\delta}(x)$ and let ν_0 (as before) be uniform distribution in $\overline{B}_\varepsilon(x)$. Then we now obtain

$$R-\varepsilon-\delta \leqslant \mathsf{d}_W(\nu_0,\nu_1) \leqslant R+\varepsilon$$

and

$$\mathrm{supp}[\nu_t] \subset \overline{B}_{\varepsilon+t(R+\varepsilon)}(x)\setminus B_{R-\delta-(1-t)(R+\varepsilon)} \tag{4.47}$$

for the probability measures on the geodesic connecting ν_0 and ν_1. Hence, arguing similarly as before, we deduce the following result.

THEOREM 4.26. *Let (M,d,m) be an arbitrary metric measure space with*

$$\underline{\mathsf{Curv}}(M,\mathsf{d},m) \geqslant K$$

for some $K\in\mathbf{R}$. For fixed $x\in\mathrm{supp}[m]$ consider $v_{R,\delta}:=m(\overline{B}_R(x)\setminus B_{R-\delta}(x))$. Then, for all $\varepsilon,\delta,r>0$ and all $R>r>2\varepsilon+\delta$,

$$v_{R,\delta} \leqslant v_\varepsilon \left(\frac{v_{\varepsilon+r,2\varepsilon+\delta}}{v_\varepsilon}\right)^{(R+\varepsilon)/r} \exp\left(-\frac{K}{2}\left(1-\frac{r}{R+\varepsilon}\right)\left(R-\frac{\delta}{2}\pm\frac{2\varepsilon+\delta}{2}\right)^2\right), \tag{4.48}$$

where $\pm$ has to be chosen as $+$ if $K\leqslant 0$ and as $-$ if $K>0$.

Choosing $\varepsilon=\delta=r/2$, this yields, in the case $K\geqslant 0$, for all $R\geqslant 3\varepsilon>0$,

$$v_{R,\varepsilon} \leqslant v_{3\varepsilon}\left(\frac{v_{3\varepsilon}}{v_\varepsilon}\right)^{R/2\varepsilon} \exp\left(-\frac{K}{2}[(R-3\varepsilon)^2-\varepsilon^2]\right). \tag{4.49}$$

In particular, $K>0$ implies that m has finite mass and finite variance.

In general, estimating the volume of concentric balls in terms of squared exponential growing functions is best possible, as demonstrated in the previous example. In the accompanying paper [53], we discuss metric measure spaces satisfying a so-called curvature-dimension condition (K,N) (replacing the condition that the curvature is $\geqslant K$), with some additional number $N\in\mathbf{R}_+$ playing the role of a dimension. We will prove that under this condition the volume of balls grows at most exponentially.

References

[1] Alexandrov, A. D., A theorem on triangles in a metric space and some applications. *Trudy Mat. Inst. Steklov*, 38 (1951), 5–23 (Russian; translated into German and combined with more material in [2]).

[2] — Über eine Verallgemeinerung der Riemannschen Geometrie. *Schr. Forschungsinst. Math. Berlin*, 1 (1957), 33–84.

[3] Ambrosio, L. & Tilli, P., *Topics on Analysis in Metric Spaces*. Oxford Lecture Series in Mathematics and its Applications, 25. Oxford University Press, Oxford, 2004.

[4] Bakry, D. & Émery, M., Diffusions hypercontractives, in *Séminaire de Probabilités,* XIX, 1983/84, pp. 177–206. Lecture Notes in Math., 1123. Springer, Berlin, 1985.

[5] Bobkov, S. G. & Götze, F., Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163 (1999), 1–28.

[6] Brenier, Y., Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44 (1991), 375–417.

[7] Bridson, M. R. & Haefliger, A., *Metric Spaces of Non-positive Curvature*. Grundlehren der Mathematischen Wissenschaften, 319. Springer, Berlin, 1999.

[8] Burago, D., Burago, Y. & Ivanov, S., *A Course in Metric Geometry*. Graduate Studies in Mathematics, 33. Amer. Math. Soc., Providence, RI, 2001.

[9] Burago, Y., Gromov, M. & Perelman, G., A. D. Aleksandrov spaces with curvatures bounded below. *Uspekhi Mat. Nauk*, 47:2 (1992), 3–51, 222 (Russian); English translation in *Russian Math. Surveys*, 47:2 (1992), 1–58.

[10] Chavel, I., *Riemannian Geometry—a Modern Introduction*. Cambridge Tracts in Mathematics, 108. Cambridge University Press, Cambridge, 1993.

[11] Cheeger, J., Differentiability of Lipschitz functions on metric measure spaces. *Geom. Funct. Anal.*, 9 (1999), 428–517.

[12] Cheeger, J. & Colding, T. H., On the structure of spaces with Ricci curvature bounded below I, II, III. *J. Differential Geom.*, 46 (1997), 406–480; *Ibid*, 54 (2000), 13–35; *Ibid*, 54 (2000), 37–74.

[13] Cordero-Erausquin, D., McCann, R. J. & Schmuckenschläger, M., A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.*, 146 (2001), 219–257.

[14] Davies, E. B., *Heat Kernels and Spectral Theory*. Cambridge Tracts in Mathematics, 92. Cambridge University Press, Cambridge, 1989.

[15] Del Pino, M. & Dolbeault, J., Best constants for Gagliardo–Nirenberg inequalities and applications to nonlinear diffusions. *J. Math. Pures Appl.*, 81 (2002), 847–875.

[16] Dudley, R. M., *Real Analysis and Probability*. Wadsworth & Brooks/Cole, Pacific Grove, CA, 1989.

[17] Feyel, D. & Üstünel, A. S., Monge–Kantorovitch measure transportation and Monge–Ampère equation on Wiener space. *Probab. Theory Related Fields*, 128 (2004), 347–385.

[18] — The strong solution of the Monge–Ampère equation on the Wiener space for log-concave densities. *C. R. Math. Acad. Sci. Paris*, 339 (2004), 49–53.

[19] Fukaya, K., Collapsing of Riemannian manifolds and eigenvalues of Laplace operator. *Invent. Math.*, 87 (1987), 517–547.

[20] Gromov, M., Groups of polynomial growth and expanding maps. *Inst. Hautes Études Sci. Publ. Math.*, 53 (1981), 53–73.

[21] — *Structures Métriques pour les Variétés Riemanniennes*. Textes Mathématiques, 1. CEDIC, Paris, 1981.

[22] — *Metric Structures for Riemannian and non-Riemannian Spaces*. Progress in Mathematics, 152. Birkhäuser Boston, Boston, MA, 1999. Based on [21].

[23] Grove, K. & Petersen, P., Manifolds near the boundary of existence. *J. Differential Geom.*, 33 (1991), 379–394.
[24] Hajłasz, P. & Koskela, P., Sobolev meets Poincaré. *C. R. Acad. Sci. Paris Sér. I Math.*, 320 (1995), 1211–1215.
[25] — Sobolev met Poincaré. *Mem. Amer. Math. Soc.*, 145 (2000).
[26] Heinonen, J., *Lectures on Analysis on Metric Spaces*. Universitext. Springer, New York, 2001.
[27] Jordan, R., Kinderlehrer, D. & Otto, F., The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.*, 29 (1998), 1–17.
[28] Kantorovich, L. V., On the translocation of masses. *C. R.* (*Doklady*) *Acad. Sci. URSS*, 37 (1942), 199–201.
[29] Kantorovich, L. V. & Rubinshteĭn, G. S., On a functional space and certain extremum problems. *Dokl. Akad. Nauk SSSR*, 115 (1957), 1058–1061.
[30] Kasue, A., Convergence of Riemannian manifolds and Laplace operators II. Preprint, 2004.
[31] Kasue, A. & Kumura, H., Spectral convergence of Riemannian manifolds. *Tohoku Math. J.*, 46 (1994), 147–179.
[32] Kigami, J., *Analysis on Fractals*. Cambridge Tracts in Mathematics, 143. Cambridge University Press, Cambridge, 2001.
[33] Koskela, P., Upper gradients and Poincaré inequalities, in *Lecture Notes on Analysis in Metric Spaces* (Trento, 1999), pp. 55–69. Appunti Corsi Tenuti Docenti Sc. Scuola Norm. Sup., Pisa, 2000.
[34] Kuwae, K. & Shioya, T., Convergence of spectral structures: a functional analytic theory and its applications to spectral geometry. *Comm. Anal. Geom.*, 11 (2003), 599–673.
[35] Ledoux, M., *The Concentration of Measure Phenomenon*. Mathematical Surveys and Monographs, 89. Amer. Math. Soc., Providence, RI, 2001.
[36] Li, P. & Yau, S.-T., On the parabolic kernel of the Schrödinger operator. *Acta Math.*, 156 (1986), 153–201.
[37] Lott, J. & Villani, C., Ricci curvature for metric-measure spaces via optimal transport. Preprint, 2005.
[38] McCann, R. J., Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80 (1995), 309–323.
[39] — A convexity principle for interacting gases. *Adv. Math.*, 128 (1997), 153–179.
[40] Monge, G., Mémoire sur la théorie des déblais et des remblais, in *Histoire de l'Académie Royale des Sciences*. Paris, 1781.
[41] Otto, F., The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26 (2001), 101–174.
[42] Otto, F. & Villani, C., Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173 (2000), 361–400.
[43] Perelman, G., The entropy formula for the Ricci flow and its geometric applications. Preprint, 2002.
[44] Plaut, C., Metric spaces of curvature $\geqslant k$, in *Handbook of Geometric Topology*, pp. 819–898. North-Holland, Amsterdam, 2002.
[45] Rachev, S. T. & Rüschendorf, L., *Mass Transportation Problems*. Vol. I. Probability and its Applications (New York). Springer, New York, 1998.
[46] von Renesse, M.-K. & Sturm, K.-T., Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58 (2005), 923–940.
[47] Saloff-Coste, L., *Aspects of Sobolev-type Inequalities*. London Math. Soc. Lecture Note Series, 289. Cambridge University Press, Cambridge, 2002.

[48] Sturm, K.-T., Diffusion processes and heat kernels on metric spaces. *Ann. Probab.*, 26 (1998), 1–55.
[49] — Metric spaces of lower bounded curvature. *Exposition. Math.*, 17 (1999), 35–47.
[50] — Probability measures on metric spaces of nonpositive curvature, in *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces* (Paris, 2002), pp. 357–390. Contemp. Math., 338. Amer. Math. Soc., Providence, RI, 2003.
[51] — Convex functionals of probability measures and nonlinear diffusions on manifolds. *J. Math. Pures Appl.*, 84 (2005), 149–168.
[52] — Generalized Ricci bounds and convergence of metric measure spaces. *C. R. Math. Acad. Sci. Paris*, 340 (2005), 235–238.
[53] — On the geometry of metric measure spaces. II. *Acta Math.*, 196 (2006), 133–177.
[54] Talagrand, M., Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, (1995), 73–205.
[55] Villani, C., *Topics in Optimal Transportation.* Graduate Studies in Mathematics, 58. Amer. Math. Soc., Providence, RI, 2003.
[56] Wasserstein [Vasershtein], L. N., Markov processes over denumerable products of spaces describing large system of automata. *Problemy Peredači Informacii*, 5:3 (1969), 64–72 (Russian). English translation in *Problems of Information Transmission*, 5:3 (1969), 47–52.

Karl-Theodor Sturm
Institut für Angewandte Mathematik
Universität Bonn
Wegelerstrasse 6
DE-53115 Bonn
Germany
sturm@uni-bonn.de

Received October 29, 2004
Received in revised form January 11, 2006