Gluing of metric measure spaces and the heat equation with homogeneous Dirichlet boundary values

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Angelo Profeta

aus
Haan

Bonn
2019

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Karl-Theodor Sturm

2. Gutachter: Prof. Dr. Patrik Ferrari

Tag der Promotion: 30.01.2020

Erscheinungsjahr: 2020

Für meine Mutter

Summary

The first part of this thesis deals with gluing together several copies of an open subset of a metric measure space along the complement. This construction results in a metric measure space. We identify the Cheeger energy and the heat flow on the glued space in terms of the corresponding objects of the underlying space. Surprisingly, the heat flow on the glued space can be expressed by using the heat flow on the underlying space and the heat flow on the open subset with *homogeneous Dirichlet boundary conditions*. This yields a possibility to deal with the Dirichlet heat flow in terms of optimal transport theory. When the glued space satisfies a lower bound on the Ricci curvature, we can infer a gradient estimate and an equivalent Bochner inequality for the Dirichlet heat flow.

As the Dirichlet heat flow does not preserve mass, we have to deal with measures of unequal masses. This makes the usual Kantorovich-Wasserstein metric useless. Instead, using a new heuristic particle interpretation for the Dirichlet heat flow that also uses *antiparticles*, we can assume the sum of particles and antiparticles to be constant and use the Kantorovich-Wasserstein metric on such sums. However, this only yields a semi-metric (i.e. the triangle inequality might not be satisfied). There is a standard way to define an induced metric from this, and we will even go a step further and define the induced length metric from it.

Another related metric is obtained by studying the one-point completion of the open subset; the added point will serve as a cemetery which makes it possible to view a subprobability measure on the open set as a probability on the one-point completion and thus using the Kantorovich-Wasserstein metric on this space.

Deriving some representation formulas in terms of other transport costs, we can compare these metrics and also clarify the relationship to weak convergence of measures. The most precise results are obtained in the case $p = 1$. Again under the assumption that the glued space has a lower bound on the Ricci curvature, we get contraction results in various of these new metrics.

Acknowledgements

I sincerely thank my supervisor Prof. Dr. Karl-Theodor Sturm for his patience, and for always coming up with a new idea once something didn't work out.

Alessandra, it is impossible to express my gratitude to you. Thank you for all the things you showed me – most of all what real friendship is. Thank you for being there in some of the darkest hours. Thank you for being silly and being serious, for cheering "KPZ!" to the Sonics (to *THE* Sonics!), for discussing the relationship of God to the ancient Hebrew people during a seminar on the stochastic heat equation, for kicking my ass, for keeping me sane. Thank you for showing me that an 11 of spades can be a useful card.

Many thanks also to Susanne, a companion without whom I wouldn't've survived ten years of math. Mein, dein, das sind doch bürgerliche Kategorien! And so we worked together all this time, shared struggles with math and life (even failed the same exam), organized anti-K.E.K.S. demonstrations, and had many coffees and Schoko-Bons.

A shout-out to Immanuel, a partner in crime and taste, for Korn, John Zorn, and a detailed study on the equivalence of an extraterrestrial sociotope with a local one. Also, whiskey.

Many more thanks are needed for many more dear friends, who helped to make my world in the last years a nice piece of space-time: Pascal (& Jolanda & Linus), Nora, Eva, Chiara R., Chiara M., Lucia, Lorenzo, Ulrike, Kathrin, Tobi, Matthäus, Luca, Lucas, Ioanna, Miri, Sonja, Melanie, Mareike, Vessi, Karsten, Sophie, Masha, Birgit and the rest of the Sturmtruppe.

A big thanks also to Patrik for being my second referee, and for trying to make me speak Italian.

There are some institutions here that also deserve credit: Thanks to Kult41 and Bonn Stomp for many many amazing concerts (and the wonderful people I met there), and to Sasà for the best pizza.

Last but not least to my Zirkus-Family, the most crazy and wonderful lot of people: Mama, Papa, Fabio & Jessica, Jerry, Tanja & Michu, Jutta & Andreas, Laura & Basti, Sara & Marten, Margrit & Franco.

On a mathematical side, I wanna thank Susanne and Lorenzo for many useful discussions. I gratefully acknowledge financial support from the CRC1060, ERC AdG RicciBounds, and BIGS.
Images in this thesis are produced with Ipe extensible drawing editor.

Contents

Chapter 1

Introduction

The theory of optimal transportation has seen an explosive growth both in content and popularity over the last 15 years. This is due to close connections with geometry, analysis, and stochastic processes. Out of these connections in classical settings like Euclidean spaces and Riemannian manifolds, a synthetic theory of analysis and geometry in metric measure spaces has evolved.

In this thesis we construct a metric measure space by gluing together several copies of a given subset of a space and identify its heat flow in terms of heat flows on the underlying space. This in turn will be used to get a description of the heat flow with homogeneous Dirichlet boundary conditions on the underlying space. Assuming a Ricci curvature bound on the glued space, we can infer gradient estimates and contraction results for this heat flow.

As the heat flow with Dirichlet boundary conditions is not mass preserving, for the latter we need to introduce a metric between measures of unequal mass. A large part of this thesis is devoted to study a number of (generalized) metrics on the space of subprobability measures.

Major results of this thesis appear in the preprint [PS18].

The rest of this chapter will give an informal overview of the involved subjects and our main results.

1.1 Analysis in Metric Measure Spaces

Among the possible approaches to analysis in metric measure spaces, we focus on the one by Ambrosio, Gigli & Savaré which was developed in the last ten years (see [AGS14a, AGS14b]) in the course of studying Ricci curvature bounds in these spaces. It is built on work by Cheeger [Che99]. Thanks to the works of Gigli [Gig15, Gig18] it is by now a very elaborate theory that encompasses a full first-order calculus in metric measure spaces, and a second-order calculus on spaces satisfying a Ricci curvature bound. We will not need this full apparatus and instead stick to the more basic notions that have been around since [AGS14a, AGS14b].

Our setting will be the one of a metric measure space $(X, d, \mathfrak{m})$ consisting of a complete, separable metric space and a Borel measure which satisfies an integrability condition so that it is in particular finite on bounded subsets. Starting from difference quotients of Lipschitz functions, a relaxation procedure yields the so-called *Cheeger energy*, which mimics the L^2-norm of the gradient of a function. Those L^2-functions whose Cheeger energy is finite will constitute the Sobolev space $W^{1,2}$. They admit an integral representation

$$\mathrm{Ch}(f) = \frac{1}{2}\int_X |\nabla f|^2 \,\mathrm{d}\mathfrak{m}$$

of the Cheeger energy with the *weak gradient* $|\nabla f|$. This weak notion of a modulus of the gradient coincides with the usual one in Euclidean space and Riemannian manifolds, and it satisfies a set of rough calculus rules. However, in general the Cheeger energy is not a quadratic form, and the Sobolev space equipped with the norm

$$\|f\|_{W^{1,2}} = \sqrt{\|f\|_{L^2}^2 + 2\,\mathrm{Ch}(f)}$$

will only be a Banach and *not* a Hilbert space. Still, this is enough to define a *Laplacian* Δ by means of convex analysis (as the element of minimal L^2-norm in the subdifferential of Ch). Subsequently, the theory of gradient flows in Hilbert spaces provides us with a heat flow P_t, given as the L^2-gradient flow of the Cheeger energy. The lack of Ch being a quadratic form now carries over to the Laplacian and heat flow *not being linear*. While for many purposes in connection with studying curvature-dimension bounds on metric measure spaces this poses no problem, for us it will be necessary to restrict to spaces whose Cheeger energies are quadratic forms. Those spaces will be called *infinitesimally Hilbertian*. In this situation, the quadratic form $\mathcal{E} = 2\,\mathrm{Ch}$ can be polarized, yielding a strongly local, quasi-regular Dirichlet form

$$\begin{cases} D(\mathcal{E}) = W^{1,2}, \\ \mathcal{E}(f,g) = \int_X \nabla f \cdot \nabla g \,\mathrm{d}\mathfrak{m} \text{ for } f,g \in W^{1,2}. \end{cases}$$

The theory of Dirichlet forms provides now close explicit connection between the form $\mathcal{E}$, the Laplacian Δ and the semigroup P_t. For instance, the Laplacian is then a self-adjoint, non-positive linear operator, connected to $\mathcal{E}$ via integration by parts

$$\mathcal{E}(f,g) = -\int_X f\Delta g \,\mathrm{d}\mathfrak{m},$$

and the heat flow is the linear semigroup $P_t = e^{t\Delta}$.

If we consider now an open subset $Y \subset X$, we can also study the *Dirichlet heat flow* P_t^0 on Y, i.e. the semigroup associated to the Dirichlet form obtained by restricting $\mathcal{E}$ to functions that vanish on $X \setminus Y$. Both heat flows enjoy nice regularity properties; through the existence of (sub-)Markovian kernels we can define dual heat flows for measures, $\mathscr{P}_t$ corresponding to P_t, and $\mathscr{P}_t^0$ corresponding to P_t^0. In the classical setting of a manifold with boundary, taking Y as the interior, P_t^0 would be the heat flow with Dirichlet boundary conditions while P_t would be the heat flow with Neumann boundary conditions.

1.2 Optimal Transport and Ricci Curvature Bounds

The other ingredient for doing geometric analysis in metric measure spaces is optimal transport. The theory of optimal transport dates back to 1781, when Gaspard Monge published the article “Mémoire sur la théorie des déblais et des remblais” [Mon81] and discussed how to optimally transport soil to a factory. In modern terms the problem is formulated in the following way: Given a complete, separable metric space (X, d) and a cost function $c\colon X \times X \to \mathbb{R}$ telling us how expensive it is to transport mass from a place $x \in X$ to a place $y \in X$, and two probability measures $\mu, \nu \in \mathcal{P}(X)$ representing the pile of soil and its destination, we want to minimize the transportation cost

$$\int_X c(x, T(x))\,\mathrm{d}\mu(x)$$

over all maps $T\colon X \to X$ transporting μ to ν, which means that the push-forward defined by $T_\#\mu(A) = \mu(T^{-1}(A))$ satisfies $T_\#\mu = \nu$. This problem is not particularly well-posed as it is for instance impossible to transport a Dirac mass $\mu = \delta_x$ to something that is not a Dirac, so for instance to $\nu = \frac{1}{2}\delta_y + \frac{1}{2}\delta_z$ with $y, z \neq x$. This is because a transport map T cannot describe how to split mass.

It took quite a while for the theory to grow up and overcome this obstruction. Monge’s problem got a satisfactory solution only in 1942, when Leonid Kantorovich relaxed the problem in a way that one easily obtains existence of minimizers (see [Kan58] for the English translation of the original Russian [Kan42]). The idea is to allow mass to *split*, i.e. the mass given at a point x is allowed to split and be transported to different points y and z. This however cannot be described by a function, which to every x associates only *one* point y. Instead, one takes a probability measure q on the product space $X \times X$ and requires it to have as *marginals* the measures μ and ν, i.e. $q(A \times X) = \mu(A)$ and $q(X \times A) = \nu(A)$ for measurable sets $A \subset X$. Then one wants to minimize

$$\int_{X\times X} c(x, y)\,\mathrm{d}q(x, y)$$

over all such *couplings* q of μ and ν. As a minimization problem it has much better properties than the original problem of Monge. First of all, the product measure $\mu \otimes \nu$ is a coupling, so the set of admissible couplings is non-empty. Furthermore the problem is linear in q with linear constraints, while the constraint on the transport map T was nonlinear. Together with Prokhorov’s theorem (i.e. compactness in the space of probability measures) this makes it easy to prove the existence of minimizers by use of the direct method of the calculus of variations. Kantorovich’s problem is a relaxation of Monge’s problem in the sense that each transport map T in the Monge problem induces a coupling $(\mathrm{id}, T)_\#\mu$ between μ and $T_\#\mu$.

Cost functions of particular interest are powers of the distance function, d^p with $p \in [1, \infty)$. They yield the so-called Kantorovich-Wasserstein distances on the set of probability measures:

$$W_p(\mu, \nu) = \inf_{\substack{q\in\mathcal{P}(X\times X)\\ q(\cdot\times X)=\mu, q(X\times\cdot)=\nu}} \left(\int_{X\times X} d^p(x, y)\,\mathrm{d}q(x, y)\right)^{\frac{1}{p}}.$$

To get a finite metric and not just an extended metric, we have to restrict to probability measures of finite p^{th} moment, the space $\mathcal{P}_p(X)$. The metric space $(\mathcal{P}_p(X), W_p)$ (known as p-Wasserstein space) is a complete, separable metric space and W_p metrizes the weak convergence in $\mathcal{P}_p(X)$, i.e. the weak convergence of the measures plus the convergence of their p^{th} moments. Furthermore, the Wasserstein space shares some properties of the underlying space; for instance it is a compact space if and only if X is. For us the basic metric geometry of the space is of importance. A metric space X is said to be a geodesic space, if for any two points $x, y \in X$ there exists a constant-speed, minimizing geodesic connecting them, i.e. a curve $\gamma\colon [0,1] \to X$ such that $\gamma_0 = x$, $\gamma_1 = y$ and

$$d(\gamma_s, \gamma_t) = |s - t| d(x, y) \tag{1.2.1}$$

for every $s, t \in [0,1]$. It turns out that the Wasserstein space is a geodesic space if and only if X is. There is a useful characterization of geodesics in the Wasserstein space which allows to express them via measures on the space of geodesics on X, which in turn lets us use (1.2.1) in integrals involving the optimal coupling.

From the point of view of geometry, the special choice $p = 2$ is the most important. In Euclidean space and Riemannian manifolds, optimal couplings for the W_2-metric are characterized quite precisely, and for measures μ, ν absolutely continuous with respect to the Lebesgue or Riemannian volume measure, respectively, such optimal couplings are indeed given by transport maps which are induced by a gradient of a potential φ. Given a Riemannian manifold (M, g, vol), viewing $\mathcal{P}_2(M)$ as a formal Riemannian manifold whose Riemannian distance is given by the Kantorovich-Wasserstein distance, a geodesic μ_t in $\mathcal{P}_2(M)$ is characterized by two equations, once the continuity equation

$$\partial_t \mu_t + \mathrm{div}(\mu_t \nabla \varphi_t) = 0$$

(understood in a distributional sense) telling us that we actually have a continuous curve of probability measures, and then a Hamilton-Jacobi equation

$$\partial_t \varphi_t + \frac{1}{2} |\nabla \varphi_t|^2 = 0$$

for the potential (or "tangent vector field") which means that the curve is a geodesic in the Wasserstein space. Doing formal Riemannian calculations while completely ignoring integrability and regularity issues, one can easily compute derivatives of functionals defined on $\mathcal{P}_2(M)$. This way one can guess for instance that certain classes of partial differential equations can be described as gradient flows on the Wasserstein space. A gradient flow is an ordinary differential equation of the form

$$\partial_t u = -\nabla E(u)$$

for some energy (or entropy) functional E, see for instance [AG13, AGS08, Ott01]. Heuristically speaking, a gradient flow curve moves in a direction that minimizes the energy (as the gradient points in the direction of steepest descent). Basic existence results for gradient flows can be obtained if the functional E is (K-)convex. In that

case one can also deduce useful contraction results.
The first (and for us most important) result in this direction was the description of the heat equation as gradient flow of the relative entropy

$$\operatorname{Ent}(f \,\mathrm{dvol}) = \int_M f \log f \,\mathrm{dvol}$$

obtained in [JKO98]. If the entropy is convex, then the heat flow exists, and in fact it is the flow $\mathscr{P}_t$ that one obtains by duality from the heat flow P_t connected to a quadratic Cheeger energy.
Convexity can be formulated in abstract geodesic metric spaces, but assuming this formal Riemannian structure of $\mathcal{P}(X)$, one can heuristically compute the Hessian of the functional by taking a second derivative along a geodesic μ_t with corresponding potential φ_t, getting

$$\operatorname{Hess}(\operatorname{Ent})(\mu_t)(\varphi_t, \varphi_t) = \frac{\mathrm{d}^2}{\mathrm{d}t^2} \operatorname{Ent}(\mu_t) = \int_M \|\operatorname{Hess}\, \varphi_t\|^2 + \operatorname{Ric}(\nabla\varphi_t, \nabla\varphi_t) \,\mathrm{dvol}\,.$$

The occurrence of the Ricci curvature shows the close connection of the relative entropy to the Ricci curvature of the underlying manifold. We see in particular that the entropy is K-convex, $K \in \mathbb{R}$, if the Ricci curvature is bounded below by K. In the seminal paper [vRS05], the authors showed that the K-convexity of the entropy is actually equivalent to a lower bound on the Ricci curvature. The big advantage is that the K-convexity can simply be formulated as

$$\operatorname{Ent}(\mu_t) \leq (1-t) \operatorname{Ent}(\mu_0) + t \operatorname{Ent}(\mu_1) - \frac{K}{2} t(1-t) W_2(\mu_0, \mu_1)^2$$

for every W_2-geodesic μ_t, not using any sort of differentiability, but only the geodesics of the Wasserstein space and the reference measure to define the entropy. One can use this now as a definition for lower Ricci curvature bounds in metric measure spaces, the so-called $\mathrm{CD}(K, \infty)$ condition ("CD" for curvature-dimension). This was done independently in [Stu06a] and [LV09]. Many geometric and analytic results valid for Riemannian manifolds whose Ricci curvature satisfies $\operatorname{Ric} \geq K$ have been shown to hold also in metric measure spaces satisfying the $\mathrm{CD}(K, \infty)$ condition. By using other entropy functionals, and a more complex notion of convexity, one can also incorporate an upper bound on the dimension of the space, resulting in more precise estimates. This was initiated in [Stu06b]. We will not need the dimensional bound in this thesis, but instead we will use a different reinforcement of the curvature-dimension condition by adding the assumption that the space is infinitesimally Hilbertian; this will be called the $\mathrm{RCD}(K, \infty)$ condition ("R" for Riemannian), and it appeared first in [AGS14b]. While the class of $\mathrm{CD}(K, \infty)$ spaces contains also Finsler manifolds (whose tangent spaces are equipped with norms instead of scalar products as for Riemannian manifolds), the latter excludes those and enforces the spaces to behave more Riemannian.

Under some weak technical assumptions, the $\mathrm{RCD}(K, \infty)$ condition is equivalent to a number of useful inequalities, namely the Bochner inequality

$$\frac{1}{2} \Delta |\nabla f|^2 - \nabla f \cdot \nabla \Delta f \geq K |\nabla f|^2,$$

which has to be understood in a weak sense (and which has been the starting point of defining synthetic Ricci curvature bounds in the setting of Dirichlet forms in the 80s, see [BÉ85]), gradient estimates

$$|\nabla P_t f|^2 \leq e^{-2K} P_t \left(|\nabla f|^2\right),$$

and Wasserstein contraction results

$$W_2(\mathscr{P}_t\mu, \mathscr{P}_t\nu) \leq e^{-K} W_2(\mu, \nu)$$

for the heat flows. One of our goals will be to get similar statements for the heat flow with Dirichlet boundary conditions.

1.3 Gluing of Metric Measure Spaces

Once these basics are settled, one can try to study the corresponding objects on related spaces, like weighted spaces (i.e. taking a weighted measure $e^{-V}\mathfrak{m}$) [AGS14a], products [AGS14b], quotients [GGKMS18], warped products [Ket13], cones [Ket15], conformal changes [Stu18, Han19]. Expecting things to be similar to the case of Riemannian manifolds, the aim is to see if curvature bounds are inherited in the sense that one can estimate a lower bound of the Ricci curvature of the resulting space in terms of the lower bound of the original space(s). For this it might be necessary to identify objects like the Cheeger energy or Laplacian in terms of the underlying space.

We will be concerned with gluing together spaces. Gluing together topological spaces along subsets is a well-known construction. Beginning with Alexandrov in the 40s, gluing has been studied in connection with curvature bounds a number of times, but mostly in Alexandrov spaces (i.e. metric spaces with a synthetic lower bound on the sectional curvature), see [Ale55, “Verheftungssatz” Kap. IX, §3], [Pog73, Chapter I, §11], [Per91, §5], [Pet97, Theorem 2.1]. When gluing together smooth Riemannian manifolds, the resulting space is no longer a manifold of the same kind, since the resulting glued metric will in general only be continuous across the gluing edge, and not smooth. One can view this space as an Alexandrov space, and indeed this idea has been exploited to deal with manifolds with boundary, see [Kos02, Theorem 1.1]. More recently, Schlichting [Sch14, Sch12] applied the method of [Kos02] to show preservation of various curvature bounds (among them Ricci curvature) on manifolds in an *approximate* sense which we will use later to give the Riemannian case as an example. See also [PV16] for a similar result. In [Pau05], metric measure spaces equipped with Dirichlet forms are glued together and the doubling property of the glued measure and the Poincaré inequality on the glued space are studied. Apart from curvature bounds, the doubling of manifolds with boundary has also been applied by other communities to produce a related manifold without boundary, see for instance [AB64].

Unlike for Alexandrov spaces and Riemannian manifolds, in metric measure spaces there is as of yet no notion of boundary. Hence, there is no natural subset at which we can glue together these spaces. Instead, given two metric measure

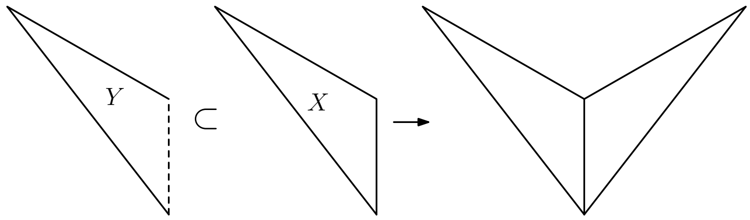

Figure 1.1: Gluing two copies of a triangle along a “bad” boundary

spaces, we have to choose isometric subsets at which to glue them together. The resulting space can easily be turned into a metric measure space. As simple examples show, it is however not possible to preserve synthetic Ricci curvature bounds when gluing together metric measure spaces. Take for instance a triangle X in $\mathbb{R}^2$ and view it as an abstract metric measure space endowed with the Euclidean distance and Lebesgue measure. As a convex subset in the Euclidean plane it has Ricci curvature bounded below by 0. Taking as an open subset Y everything but one of the sides, the gluing of two copies along $X \setminus Y$ has Ricci curvature bounded below by 0 if and only if it is convex. Thus, an example as in Figure 1.1 shows that the curvature bound is in general not preserved. Also, a recent preprint by Rizzi shows that gluing in “smooth” metric measure spaces does not preserve the dimension in the measure-contraction property [Riz18]. We will focus on the special case of gluing together two copies of the *same* space along the complement of an open subset. This allows us to identify the heat flow on the glued space in terms of heat flows on the separate copies. Let us give a few details.

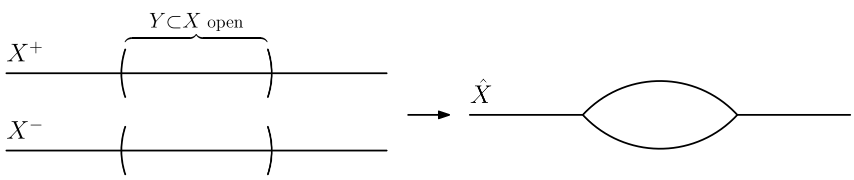

Figure 1.2: Gluing two copies of Y

It is easy to see that gluing together metric measure spaces results naturally in a metric measure space: Given a space $(X, d, \mathfrak{m})$ and an open subset $Y \subset X$, consider two exact copies of this, named X^+, Y^+ and X^-, Y^-. Then the *doubling of* Y is the space $\hat{X} = X^+ \sqcup X^- / \sim$ where we identify points in $X^+ \setminus Y^+$ with the corresponding points in $X^- \setminus Y^-$, see Figure 1.2. A distance on this space is given by

$$\hat{d}(x, y) = \begin{cases} d(x, y), & \text{if } x, y \in X^i \\ \inf_{z \in X \setminus Y} d(x, z) + d(z, y), & \text{if } x \in X^i, y \in X^j, i \neq j. \end{cases}$$

As a measure we define, for a measurable subset $A \subset \hat{X}$,

$$\hat{\mathfrak{m}}(A) = \frac{1}{2}\mathfrak{m}^+(A \cap X^+) + \frac{1}{2}\mathfrak{m}^-(A \cap X^-).$$

Being a metric measure space, it possesses a Cheeger energy $\hat{\mathrm{Ch}}$ and the related Laplacian $\hat{\Delta}$ and heat semigroup $\hat{P}_t$. Our first main result will be a characterization of the heat flow in terms of the heat flows on the single copies X. It will turn out that the Cheeger energy on $\hat{X}$ is a quadratic form because the one of X is, and for a function $u\colon \hat{X} \to \mathbb{R}$, its heat flow will be

$$\hat{P}_t u = \begin{cases} \frac{1}{2} P_t(u^+ + u^-) + \frac{1}{2} P_t^0(u^+ - u^-), & \text{on } X^+ \\ \frac{1}{2} P_t(u^+ + u^-) + \frac{1}{2} P_t^0(u^- - u^+), & \text{on } X^-. \end{cases} \tag{1.3.1}$$

Ultimately this formula will help us to study the heat flow with Dirichlet boundary conditions. The occurrence of the heat flow with Dirichlet boundary values on the glued space may be surprising at first. It is due to the fact that the mass on the separate copies does not need to be preserved, since it can move to another copy. There is the following heuristic explanation for this formula in terms of a particle interpretation. Recall that the heat equation with Neumann boundary conditions is related to Brownian particles reflected at the boundary, whereas the one with Dirichlet boundary conditions corresponds to particles killed at the boundary. In the glued space $\hat{Y}$ there is no boundary any more, so the heat equation is related to a Brownian particle in the glued space, which means for instance, if it starts on the upper half Y^+ and approaches the boundary ∂Y^+, then it can either "return" and stay on this upper half or it can change to the lower half, meaning it is killed on the upper half. On the upper copy, this behavior is captured by the terms

$$\frac{1}{2} P_t u^+ + \frac{1}{2} P_t^0 u^+.$$

But there are also particles on the lower copy which are killed there and move to the upper copy; those are represented by

$$\frac{1}{2} P_t u^- - \frac{1}{2} P_t^0 u^-.$$

1.4 Transportation Metrics for Subprobabilities

There is an equivalent way to express this intuition, namely instead of having two copies of the space we can consider two kinds of particles. We will call them particle and antiparticle. They can change their type when they hit the boundary of $Y \subset X$; half the time they continue with their type, and half the time they change to the other type. The total number of particles plus antiparticles will stay constant, and particles and antiparticles staying in the same site will annihilate. To describe such an ensemble of particles and antiparticles, we consider *charged probabilities*, couples (σ^+, σ^-) of subprobability measures σ^i that coincide when restricted to $X \setminus Y$, and the sum of which is a probability.

We will define a "Kantorovich-Wasserstein" metric $\tilde{W}_p$ on the space of charged probabilities and show that there is an isometry between this space and the Wasserstein space $(\mathcal{P}_p(\hat{X}), \hat{W}_p)$ over the doubled space. This metric will be the starting point in our journey to defining a metric on the space of subprobabilities. Along

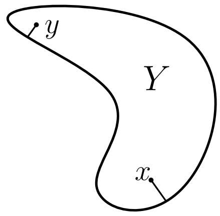

Figure 1.3: Geodesic in one-point completion Y'

the way, there will appear numerous almost-metrics. A first attempt to define such a distance is the following: Given subprobability measures μ and ν on $Y \subset X$ (not necessarily with the same mass), we consider charged probabilities σ and τ such that their *effective measures* $\sigma^0 = \sigma^+ - \sigma^-$ and $\tau^0 = \tau^+ - \tau^-$ equal μ and ν, and as distance between μ and ν we take

$$W^0(\mu, \nu) = \inf_{\sigma,\tau} \tilde{W}(\sigma, \tau)$$

where the infimum is over all such charged probabilities. Unfortunately, W^0 does not satisfy the triangle inequality. To overcome this difficulty, we introduce the biggest metric below W^0, called $W^\flat$, and further pass to its induced length metric $W^\sharp$. For all these functions, we derive various representations that make it possible to compare them. We will often focus on the case $p = 1$ since then it is possible to get more precise results. For instance, we will show that

$$\begin{aligned} W_1^0(\mu, \nu) = \inf \Big\{ W_1(\mu_1, \nu_1) + \frac{1}{2} W_1^*(\mu_0, \mu_0) + \frac{1}{2} W_1^*(\nu_0, \nu_0) \,\Big|\, & \mu = \mu_1 + \mu_0, \nu = \nu_1 + \nu_0, \\ & (\mu + \nu_0)(X) \leq 1, \\ & (\nu + \mu_0)(X) \leq 1 \Big\}, \end{aligned}$$

where $W_1^*(\mu_0, \mu_0)$ is the *annihilation cost* given by the optimal transport problem with the *reflection distance* $d^*(x, y) = \inf_{z \in X \setminus Y} d(x, z) + d(z, y)$. This auxiliary cost measures the distance that is needed to annihilate an ensemble of particles because they have to travel via the boundary to become antiparticles. Similarly we will get that

$$W_1^\sharp(\mu, \nu) = \inf \Big\{ W_1(\mu_1, \nu_1) + \frac{1}{2} W_1^*(\mu_0, \mu_0) + \frac{1}{2} W_1^*(\nu_0, \nu_0) \,\Big|\, \mu = \mu_1 + \mu_0,\ \nu = \nu_1 + \nu_0 \Big\}.$$

The proof of this requires more auxiliary costs, and a comparison to the Kantorovich-Wasserstein metric W_p' on the so-called *one-point completion* $Y' = Y \cup \{\partial\}$ of Y. Intuitively the idea is to contract the topological boundary of the open set Y to one extra point and define a metric that decides whether it's shorter to move inside Y or to move through the "boundary point", see Figure 1.3. This point will serve as a cemetery, enabling us to keep the "lost" mass and thus deal with subprobability measures on Y via probability measures on the one-point completion. In the case $p = 1$, the metric W_1' interpreted as a metric on the space of subprobability measures on Y will equal the above-mentioned metrics $W_1^\flat$ and $W_1^\sharp$, while in the case $p > 1$ we get the ordering $W_1' \leq W_p^\flat \leq W_p^\sharp \leq W_p'$. A consequence will be that for compact, geodesic spaces X, the metric $W_p^\sharp$ metrizes the vague convergence of subprobability measures on Y.

The idea of adding mass at the boundary like we do with W_p' has also been used in [FG10] in the setting of open, bounded subsets of $\mathbb{R}^n$, where the authors allow to create and destroy mass at the boundary. While we add up mass to get probability measures, they more generally allow to create and destroy mass to get measures of equal mass. They obtain a gradient flow description of the heat equation with *strictly positive, constant* Dirichlet boundary conditions. However, it does not apply to the study of the heat flow with vanishing Dirichlet boundary conditions. See Remark 4.4.4 for more related to their metric. Other approaches to metrics on the space of finite Radon measures have been taken in [LMS18, PR14, KMV16, Mai11].

1.5 The Heat Flow with Dirichlet Boundary Conditions

Finally we want to use the previous results to infer some information about the heat flow with homogeneous Dirichlet boundary conditions.

For this we will from then on assume that the glued space is an $\mathrm{RCD}(K, \infty)$ space. This immediately provides us with a gradient estimate and a Wasserstein contraction for the heat flow on $\hat{X}$. Through formula (1.3.1), from this we can deduce corresponding inequalities involving the heat flow with *Dirichlet boundary conditions*, however also using the "usual" heat flow P_t. The gradient estimate for a function $f \in W^{1,2}(X)$ with $f = 0$ on $X \setminus Y$ is

$$|\nabla P_t^0 f|^2 \leq e^{-2Kt} P_t(|\nabla f|^2) \quad \mathfrak{m}\text{-a.e. in } X.$$

This gradient estimate is equivalent to the following weak Bochner inequality: for f in the domain of the Dirichlet Laplacian Δ^0 and such that $\Delta^0 f \in W^{1,2}$ with $f = 0$ on $X \setminus Y$, and for a bounded, non-negative φ in the domain of Δ:

$$\frac{1}{2} \int \Delta\varphi |\nabla f|^2 \, \mathrm{d}\mathfrak{m} - \int \varphi \nabla f \cdot \nabla \Delta^0 f \, \mathrm{d}\mathfrak{m} \geq K \int \varphi |\nabla f|^2 \, \mathrm{d}\mathfrak{m}.$$

Thanks to the self-improvement property of the Bochner inequality as shown by [Sav14], both inequalities actually hold in a p-version for every $p \in [1, \infty)$.

A related aim is to get Wasserstein-contraction-like results for the heat flow with Dirichlet boundary values. Again, the lower Ricci curvature bound on the glued space directly supplies the Wasserstein contraction for the heat flow $\tilde{\mathscr{P}}$ in the metric $\tilde{W}$. This lets us deduce contraction results with the same coefficients for the heat flow with Dirichlet boundary condition in the previously introduced semi-metric W_p^0: given subprobability measures μ and ν, their Dirichlet heat flows satisfy

$$W_p^0(\mathscr{P}_t^0 \mu, \mathscr{P}_t^0 \nu) \leq e^{-Kpt} \, W_p^0(\mu, \nu),$$

and the same with the metrics $W_p^\flat$, $W_p^\sharp$ and W_p'.

1.6 Outline of the Thesis

In Chapter 2 we collect basic definitions and well-known facts concerning analysis in metric measure spaces and optimal transport theory. Moreover, we prove basic results concerning the one-point completion Y', regularity properties of the heat semigroups, and the existence of W_1-geodesics that are supported on geodesics.

Chapter 3 discusses the gluing of $k \in \mathbb{N}$ copies of the same metric measure space, and identifies the Cheeger energy and hence also the heat flow in terms of the Dirichlet and Neumann heat flows on the underlying space.

The following Chapter 4 starts the discussion on transport metrics for subprobability measures. First, by introducing a sort of "Wasserstein" space of charged measures (which will be equivalent to the Wasserstein space of the doubled space), and then by successively going to W_p^0, $W_p^\flat$, $W_p^\sharp$, W_p' and studying in detail the connections between those functions.

Finally Chapter 5 discusses the implications of a curvature condition on the doubled space, in particular the consequences for the heat flow with Dirichlet boundary conditions.

1.7 Table of Metrics and Heat Flows

As we will encounter as much as 9 generalized "W-metrics", let us give a short overview where to find the definitions:

- W_p usual Kantorovich-Wasserstein metric on $\mathcal{P}_p(X)$, (2.5.2)
- $\tilde{W}_p$ transportation metric on $\tilde{\mathcal{P}}_p(Y|X)$, Def. 4.1.2
- W_p^0 transportation-annihilation pre-metric on $\mathcal{P}_p^{sub}(Y)$, Def. 4.2.1
- $W_p^\flat$ pseudo-metric on $\mathcal{P}_p^{sub}(Y)$, (4.3.1)
- $W_p^\sharp$ transportation-annihilation metric on $\mathcal{P}_p^{sub}(Y)$, Def. 4.3.2
- W_p' Kantorovich-Wasserstein metric on $\mathcal{P}_p(Y')$, based on shortcut metric d', Def. 4.4.1
- $W_p^\dagger$ transportation cost "over the boundary" on measures on Y of the same mass, Def. 4.4.1
- W_p^* annihilation cost; meta-metric on measures on X of the same mass, Def. 4.2.3
- $\hat{W}_p$ Kantorovich-Wasserstein metric on $\mathcal{P}_p(\hat{X})$

Since a similar number of heat flows is turning up, we give an overview; their definitions can be found in Section 2.3:

- P_t heat flow for functions on X with "Neumann boundary conditions"
- $\mathscr{P}_t$ heat flow for measures with "Neumann boundary conditions"
- P_t^0 heat flow on Y with "Dirichlet boundary conditions"
- $\mathscr{P}_t^0$ heat flow for measures "with Dirichlet boundary conditions"
- $\hat{P}_t$ heat flow for functions on the glued space
- $\hat{\mathscr{P}}_t$ heat flow for measures on the glued space
- $\tilde{\mathscr{P}}_t$ heat flow for charged measures

Chapter 2

Preliminaries

In this chapter we introduce the main objects of study, and give the main properties. Furthermore we collect some technical results.

2.1 Length Spaces

We start with the study of length and geodesic spaces. They are a natural starting point for doing geometry in an abstract setting because they allow to measure distances by the length of curves, and curves are needed to study convexity which in turn is at the basis of defining synthetic curvature bounds. After introducing some general notions connected to metric spaces we turn to length and geodesic spaces and prove some equivalent characterizations of the definitions. Anything not proven here can be found for instance in [BBI01, BH99] (note however that our definitions sometimes differ from those in [BBI01] who for instance allow the value $+\infty$ for a metric).

A metric space is a set together with a function giving distances between points. The following definition specifies this.

Definition 2.1.1. Let X be a set. A function $d\colon X \times X \to \mathbb{R}$ is a *metric* if for all $x, y, z \in X$:

$$
\begin{aligned}
\textit{Positivity:}\quad & d(x,y) \geq 0,\\
\textit{Vanishing diagonal:}\quad & d(x,x) = 0,\\
\textit{Definiteness:}\quad & d(x,y) > 0 \text{ if } x \neq y,\\
\textit{Symmetry:}\quad & d(x,y) = d(y,x),\\
\textit{Triangle inequality:}\quad & d(x,y) \leq d(x,z) + d(z,y).
\end{aligned}
$$

Then (X, d) is called a *metric space*. The metric d will also be called *distance*. Among the possible variations of this definition we will encounter:

Extended metric:	Also the value $+\infty$ may be attained.
Pseudo-metric:	May vanish also outside the diagonal.
Meta-metric:	Not necessarily vanishing on the diagonal.
Semi-metric:	Does not need to satisfy the triangle inequality.

In Chapter 4, while attempting to define a metric on the space of subprobability measures, we will encounter a function that does not satisfy the triangle inequality. There is the following elementary way of producing a metric out of it.

Lemma 2.1.2. *Consider a set X and a semi-metric $d\colon X \times X \to \mathbb{R}$ that might also vanish off the diagonal (so it is also a pseudo-metric). Then the function $d^\flat\colon X \times X \to \mathbb{R}$ defined by*

$$d^\flat(x,y) := \inf\left\{ \sum_{i=1}^n d(z_{i-1}, z_i) \;\middle|\; n \in \mathbb{N}, (z_i)_{i=1}^n \subset X,\ z_0 = x, z_n = y \right\}$$

is a pseudo-metric on X. Furthermore, it is the biggest pseudo-metric below d.

Proof. Obviously $d^\flat \geq 0$, $d^\flat(x,y) = d^\flat(y,x)$ and $d^\flat(x,x) = 0$. For the triangle inequality observe that the infimum only gets worse when restricting to paths forced to visit a third point: Given $x, y, v \in X$,

$$\begin{aligned} d^\flat(x,y) &= \inf\left\{ \sum_{i=1}^n d(z_{i-1}, z_i) \;\middle|\; n \in \mathbb{N}, (z_i)_{i=1}^n \subset X,\ z_0 = x, z_n = y \right\} \\ &\leq \inf\Bigg\{ \sum_{i=1}^j d(z_{i-1}, z_i) + \sum_{i=j+1}^n d(z_{i-1}, z_i) \;\Bigg|\; n \in \mathbb{N}, (z_i)_{i=1}^n \subset X, \\ &\qquad\qquad z_0 = x, z_n = y, z_j = v \text{ for some } j \in \{1, \dots n-1\} \Bigg\} \\ &\leq \inf\left\{ \sum_{i=1}^n d(z_{i-1}, z_i) \;\middle|\; n \in \mathbb{N}, (z_i)_{i=1}^n \subset X, z_0 = x, z_n = v \right\} \\ &\quad + \inf\left\{ \sum_{i=1}^n d(z_{i-1}, z_i) \;\middle|\; n \in \mathbb{N}, (z_i)_{i=1}^n \subset X, z_0 = v, z_n = y \right\} \\ &= d^\flat(x,v) + d^\flat(v,y). \end{aligned}$$

The maximality is a consequence of two easy facts:

1. Given a (pseudo-)metric d, the above construction yields the same (pseudo-)metric, i.e. $d^\flat = d$. Indeed, trivially $d^\flat \leq d$. The other inequality is a consequence of the triangle inequality: Given $\varepsilon > 0$, there are z_i such that $z_0 = x, z_n = y$ and $d^\flat(x,y) + \varepsilon \geq \sum_{i=1}^n d(z_{i-1}, z_i) \geq d(x,y)$.

2. This construction preserves order, i.e. if $d_1 \leq d_2$, then

$$\begin{aligned} d_1^\flat(x,y) &= \inf_{x=z_0,\dots,z_n=y} \left\{ \sum_{i=1}^n d_1(z_{i-1}, z_i) \right\} \\ &\leq \inf_{x=z_0,\dots,z_n=y} \left\{ \sum_{i=1}^n d_2(z_{i-1}, z_i) \right\} = d_2^\flat(x,y). \end{aligned}$$

Hence, for a function d lacking a triangle inequality and a metric $\tilde{d}$ with $d^\flat \le \tilde{d} \le d$, by applying the construction to these three functions, we get

$$d^\flat = (d^\flat)^\flat \le \underbrace{(\tilde{d})^\flat}_{=\tilde{d}} \le d^\flat.$$

□

Some of the most common (topological and metric) properties we will often impose on a metric space are the following:

Separability:	There exists a countable dense subset of X.
Completeness:	Every Cauchy sequence converges.
Local compactness:	Every point has a compact neighborhood.
Properness:	Every closed ball $\overline{B}_r(x)$ is compact.
total boundedness:	For every $\varepsilon > 0$ there is a finite cover of X by open balls of radius ε.

Let us now turn to more geometric notions. Given two metric spaces (X, d_X), (Z, d_Z), a function $\varphi\colon X \to Z$ is an *isometry* if $d_Z(\varphi(x), \varphi(y)) = d_X(x, y)$ for every $x, y \in X$. Unless otherwise stated, by a curve we mean a *continuous* map $\gamma\colon [a, b] \to X$, and we will usually parametrize it to be defined on $[a, b] = [0, 1]$. Sometimes we will denote by $\gamma\colon x \rightsquigarrow y$ a curve $\gamma\colon [0, 1] \to X$ with $\gamma_0 = x$ and $\gamma_1 = y$. The space $C^0([0, 1], X)$ of continuous curves equipped with the supremum-norm $d_\infty(\gamma^1, \gamma^2) := \sup_{t\in[0,1]} d(\gamma^1_t, \gamma^2_t)$ is complete and separable. A curve $\gamma\colon [0, 1] \to X$ is a *constant-speed geodesic* if for every $s, t \in [0, 1]$:

$$d(\gamma_s, \gamma_t) = |s - t| d(\gamma_0, \gamma_1).$$

The space of constant-speed geodesics in X is denoted by $\mathrm{Geo}(X)$. It is a closed subset of $C^0([0, 1], X)$. We call a metric space (X, d) *geodesic* (or strictly intrinsic), if for every two points $x, y \in X$ there is a constant-speed geodesic $\gamma\colon [0, 1] \to X$ such that $\gamma_0 = x$ and $\gamma_1 = y$. A generalization of this is given by length spaces. To introduce them, we need to define the length of curves.

Definition 2.1.3. Let $\gamma\colon [a, b] \to X$ be a curve. Its length is

$$L_d(\gamma) := \sup \left\{ \sum_{i=1}^{k} d(\gamma_{t_{i-1}}, \gamma_{t_i}) \,\middle|\, k \in \mathbb{N}, a = t_0 \le t_1 \le \ldots \le t_k = b \right\} \in [0, \infty].$$

A curve is called *rectifiable* if $L_d(\gamma)$ is finite. In case there is no possibility of confusion, we simply use $L(\gamma)$ to denote the length of γ.

The length functional $L\colon C^0([a, b], X) \to [0, \infty]$ has several properties one might intuitively expect.

Proposition 2.1.4. *Let (X, d) be a metric space and $\gamma\colon [0, 1] \to X$ a curve. Then*

i) For every $a \in [0, 1]$ we have $L(\gamma) = L(\gamma|_{[0,a]}) + L(\gamma|_{[a,1]})$.

ii) For a rectifiable curve γ, the map $[0,1] \ni t \mapsto L(\gamma|_{[0,t]})$ is continuous and non-decreasing.

iii) L is invariant under reparametrizations of the curve, i.e. given a homeomorphism $\varphi\colon [a,b] \to [0,1]$, then $L(\gamma \circ \varphi) = L(\gamma)$.

iv) $L(\gamma) \geq d(\gamma_0, \gamma_1)$.

v) L is lower semicontinuous, i.e. given a sequence of curves $\gamma^n \in C^0([0,1], X)$ converging to γ as $n \to \infty$ with respect to d_∞, then $L(\gamma) \leq \liminf_{n\to\infty} L(\gamma^n)$.

Connected to the length of a curve $\gamma\colon [0,1] \to X$ is its *metric speed*, defined by

$$|\dot{\gamma}_t| := \limsup_{h\to 0} \frac{d(\gamma_{t+h}, \gamma_t)}{|h|}.$$

This limit for instance exists almost everywhere for Lipschitz curves (which can be achieved for every rectifiable continuous curve by reparametrization). One can then compute the length of the curve also by

$$L(\gamma) = \int_0^1 |\dot{\gamma}_t| \,\mathrm{d}t.$$

An important class of metric spaces are those in which the distance between points can actually be recovered by the length of curves.

Definition 2.1.5. i) A metric space (X, d) is a *length* space if for every $x, y \in X$

$$d(x,y) = \inf \left\{ L(\gamma) \,\middle|\, \gamma \in C^0([0,1], X), \gamma_0 = x, \gamma_1 = y \right\}.$$

Given $\varepsilon > 0$, a curve is called an *ε-geodesic* (between its endpoints), if

$$|L(\gamma) - d(\gamma_0, \gamma_1)| \leq \varepsilon.$$

ii) Given a metric space (X, d), its *induced length space* is (X, d_L) with

$$d_L(x,y) := \inf \left\{ L(\gamma) \,\middle|\, \gamma \in C^0([0,1], X), \gamma_0 = x, \gamma_1 = y \right\}.$$

Remark 2.1.6. In the case of a geodesic space, minimizing curves exist and up to reparametrization they are geodesics in the sense defined above. Thus, in particular we can choose them to have constant speed, meaning that they are parametrized proportional to arc length. In fact, one can reparametrize every curve proportional to arc length by using as a homeomorphism the map

$$\varphi\colon [a,b] \to [0,1], \quad \varphi(r) := \frac{L(\gamma|_{[a,r]})}{L(\gamma)},$$

getting the constant-speed curve $\tilde{\gamma}\colon [0,1] \to X, \tilde{\gamma}_r := \gamma \circ \varphi^{-1}(r)$. Given a constant-speed curve $\gamma\colon [0,1] \to X$ and $s, t \in [0,1]$, the length of the restriction satisfies $L(\gamma|_{[s,t]}) = |s-t| L(\gamma)$.

Convention: In the following we will always assume our curves to be parametrized proportional to arc length.

The following lemma shows that constant-speed almost-geodesics are close to a geodesic also locally.

Lemma 2.1.7. *Let $\varepsilon > 0$ and $\gamma\colon [0,1] \to X$ be an ε-geodesic. Then for every $s,t \in [0,1]$, $s \le t$,*

$$|L(\gamma|_{[s,t]}) - |s-t| d(\gamma_0, \gamma_1)| \le |s-t|\varepsilon \tag{2.1.1}$$

as well as

$$|d(\gamma_s, \gamma_t) - |s-t| d(\gamma_0, \gamma_1)| \le |s-t|\varepsilon. \tag{2.1.2}$$

Proof. Knowing how to compute the length of a restriction of a constant-speed curve, and using the definition of ε-geodesic, we get

$$|L(\gamma|_{[s,t]}) - |s-t| d(\gamma_0, \gamma_1)| = |s-t| \cdot |L(\gamma) - d(\gamma_0, \gamma_1)| \le |s-t|\varepsilon.$$

This then also entails

$$d(\gamma_s, \gamma_t) \le L(\gamma|_{[s,t]}) = |s-t| L(\gamma) \le |s-t| d(\gamma_0, \gamma_1) + |s-t|\varepsilon.$$

Since in general $|s-t| d(\gamma_0, \gamma_1) \le d(\gamma_s, \gamma_t)$, we also get the other inequality, so that finally

$$|d(\gamma_s, \gamma_t) - |s-t| d(\gamma_0, \gamma_1)| \le |s-t|\varepsilon.$$

□

Remark 2.1.8. Observe that conversely a constant-speed curve satisfying either of (2.1.1) or (2.1.2) is an ε-geodesic in the sense of Def. 2.1.5.

Under the assumption of completeness there are useful characterizations of length and geodesic spaces in terms of *midpoints*.

Proposition 2.1.9. *Let (X, d) be a complete metric space.*

i) X is a geodesic space if and only if midpoints exist, i.e. if for every $x, y \in X$ there is $z \in X$ such that $d(x,z) = \frac{1}{2} d(x,y) = d(z,y)$.

ii) X is a length space if and only if for every $\varepsilon > 0$ and $x, y \in X$ there exist an ε-midpoint, i.e. a point $z \in X$ such that

$$\left| d(x,z) - \frac{1}{2} d(x,y) \right| \le \varepsilon \quad \textit{and} \quad \left| d(y,z) - \frac{1}{2} d(x,y) \right| \le \varepsilon.$$

The idea is that for one direction you can take (almost) minimizing curves and their midpoints, and for the other one you take midpoints and bisect further, getting a countable number of midpoints of midpoints and by completeness you can extend this to a curve. Another equivalent definition of (ε-)midpoints is given in the following lemma.

Lemma 2.1.10. *Let (X, d) be a complete metric space.*

i) X is a geodesic space if and only if for every couple of points $x, y \in X$ there exists a point $z \in X$ such that:

$$d(x,z) = d(y,z) \tag{2.1.3}$$

and

$$d(x,z) + d(y,z) = d(x,y). \tag{2.1.4}$$

ii) X is a length space if and only of for every $\varepsilon > 0$ and $x, y \in X$ there exists a point $z \in X$ such that:

$$|d(x,z) - d(y,z)| \leq \varepsilon \tag{2.1.5}$$

and

$$d(x,z) + d(y,z) \leq d(x,y) + \varepsilon. \tag{2.1.6}$$

Proof. We will show the equivalence to the midpoint-characterization of geodesic and length spaces.

i) Recall that a midpoint z by definition satisfies $d(x,z) = \frac{1}{2}d(x,y) = d(y,z)$. Thus we also have $d(x,z) + d(y,z) = d(x,y)$.
Conversely, the two properties in the statement imply

$$d(x,y) = d(x,z) + d(y,z) = 2d(x,z),$$

so that we recover the original definition of a midpoint.

ii) From the definition of ε-midpoints we see that

$$|d(x,z) - d(y,z)| \leq \left| d(x,z) - \frac{1}{2}d(x,y) \right| + \left| d(y,z) - \frac{1}{2}d(x,y) \right| \leq 2\varepsilon,$$

and $d(x,z) + d(y,z) \leq \frac{1}{2}d(x,y) + \varepsilon + \frac{1}{2}d(x,y) + \varepsilon \leq d(x,y) + 2\varepsilon$.
For the other direction, let us start with observing that (2.1.5) implies

$$d(x,z) \leq d(y,z) + \varepsilon \qquad \text{and} \qquad d(y,z) \leq d(x,z) + \varepsilon.$$

Inserting this in (2.1.6), we get

$$d(x,y) + \varepsilon \geq d(x,z) + d(y,z) \geq d(x,z) + d(x,z) - \varepsilon,$$

so that $d(x,z) - \frac{1}{2}d(x,y) \leq \varepsilon$. By triangle inequality and again (2.1.5), we get

$$\frac{1}{2}d(x,y) \leq \frac{1}{2}d(x,z) + \frac{1}{2}d(y,z) \leq \frac{1}{2}d(x,z) + \frac{1}{2}d(x,z) + \frac{\varepsilon}{2},$$

so that we also have $\frac{1}{2}d(x,y) - d(x,z) \leq \varepsilon$. $\square$

Remark 2.1.11. It is worth noticing that a complete, locally compact length space is a geodesic space (see [BBI01, Thm. 2.5.23]).

2.2 One-point completion

To deal with non-complete situations and for instance the loss of mass in a flow of measures it is sometimes useful to pass to the compactification of a set, introducing a *cemetery*. But instead of using the one-point compactification (which needs a locally compact space to begin with and a priori is a topological and not necessarily a metric space), we will use a *one-point completion* of an open subset of a complete space. In Chapter 4 we will compare the Kantorovich-Wasserstein metric over it with metrics on the space of subprobabilities on the open set.

Definition 2.2.1. Let (X, d) be a metric space and $Y \subset X$ be an open and non-trivial subset, where by non-trivial we mean $Y \neq \emptyset, X$. Then we define the *one-point completion* of Y as $Y' := Y \cup \{\partial\}$ with the *shortcut metric*

$$\begin{aligned} d'(\partial, \partial) &:= 0, \\ d'(x, \partial) &:= d'(\partial, x) := \inf_{z \in X \setminus Y} d(x, z), \\ d'(x, y) &:= \min\left\{d(x, y), d'(x, \partial) + d'(y, \partial)\right\} \end{aligned}$$

for $x, y \in Y$. We will further denote

$$d^{\dagger}(x, y) := d'(x, \partial) + d'(y, \partial), \tag{2.2.1}$$

so that $d' = \min\{d, d^{\dagger}\}$.

Remark 2.2.2. Observe that in the cases of trivial subsets we get: $d' = 0$ for $Y = \emptyset$, and $d' = d$ on Y and $d'(\cdot, \partial) = +\infty$ for $Y = X$.

By abuse of notation we will often call $X \setminus Y$ the boundary of Y.

Lemma 2.2.3. *Let (X, d) be a complete, separable space, and let $Y \subset X$ be open and non-trivial. Then (Y', d') is a complete, separable metric space.*

Proof. Symmetry and non-negativity are clear from the definition.

Definiteness: $d'(\partial, \partial) = 0$ by definition and $d'(x, x) \leq d(x, x) = 0$ for $x \in Y$. Let now $x, y \in Y'$ with $d'(x, y) = 0$.
Case 1: $y = \partial$: Then $0 = d'(x, y) = d'(x, \partial) = \inf_{z \in X \setminus Y} d(x, z)$. Assume $x \in Y$. Since Y is d-open in X, there is $r > 0$ such that the d-ball $B_r(x)$ is contained in Y. In particular, $B_r(x) \cap (X \setminus Y) = \emptyset$ and therefore for every $z \in X \setminus Y$ we have $d(x, z) \geq r$. This contradicts $\inf_{z \in X \setminus Y} d(x, z) = 0$. Hence $x = \partial$.
Case 2: $x, y \in Y$. Then $0 = d'(x, y) = d(x, y)$ because otherwise we would have $0 = d'(x, y) = d'(x, \partial) + d'(y, \partial)$ which would imply $x = \partial = y$ by Case 1. Hence, by the definiteness of d we conclude that $x = y$.
In order to prove the triangle inequality, first observe that $x \mapsto d'(x, \partial)$ is d-Lipschitz:

$$|d'(y, \partial) - d'(x, \partial)| \leq d(x, y).$$

Indeed, let $z_k \in X \setminus Y$ such that $d'(x, \partial) + \varepsilon > d(x, z_k)$. Then $d'(y, \partial) \leq d(y, z_k)$ and, by the triangle inequality of d,

$$d'(y, \partial) - d'(x, \partial) \leq \varepsilon + d(y, z_k) - d(x, z_k) \leq \varepsilon + d(x, y).$$

Now we deal with the different cases for the triangle inequality of d' separately. The cases where $x = y = z = \partial$, or $x = y = \partial$, or $x = z = \partial$, or $z = \partial$, are trivial. So, let $x = \partial$, $y, z \in Y$. By the Lipschitz continuity, we have

$$d'(y, \partial) - d'(z, \partial) \leq d(y, z).$$

Together with $d'(y, \partial) - d'(z, \partial) \leq d'(y, \partial) + d'(z, \partial)$ we thus have

$$d'(y, \partial) - d'(z, \partial) \leq \min\{d(y, z), d'(y, \partial) + d'(z, \partial)\} = d'(y, z).$$

The remaining case is $x, y, z \in Y$:
If $d'(x, z) = d(x, z)$ and $d'(z, y) = d(z, y)$, then $d'(x, y) \leq d(x, y) \leq d(x, z) + d(z, y)$.
If $d'(x, z) = d(x, z)$ and $d'(z, y) = d'(z, \partial) + d'(y, \partial)$, then – using the Lipschitz continuity again –

$$d'(x, y) \leq d'(x, \partial) + d'(y, \partial) \leq d'(z, \partial) + d(x, z) + d'(y, \partial).$$

The case $d'(x, z) = d'(x, \partial) + d'(z, \partial)$ and $d'(z, y) = d(z, y)$ is analogous.
Finally, if $d'(x, z) = d'(x, \partial) + d'(z, \partial)$ and $d'(z, y) = d'(z, \partial) + d'(y, \partial)$, then

$$d'(x, y) \leq d'(x, \partial) + d'(y, \partial) \leq d'(x, \partial) + d'(z, \partial) + d'(z, \partial) + d'(y, \partial).$$

Separability: This is a direct consequence of the separability of X. Let $A := \{z_i\}$ be the countable dense subset of X. Given $x \in Y$, there is a sequence $(z_{i_k})_{k \in \mathbb{N}} \subset A \cap Y$ such that $d(z_{i_k}, x) \to 0$ as $k \to \infty$. Then also $d'(z_{i_k}, x) \leq d(z_{i_k}, x) \to 0$ as $k \to \infty$. For the boundary point ∂, let $(z_{i_k})_{k \in \mathbb{N}} \subset A \cap Y$ be any sequence converging to some boundary point $z \in \partial Y$ with respect to d. But then $d'(z_{i_k}, \partial) \leq d(z_{i_k}, z) \to 0$.

Completeness: Let $(x_n) \subset Y'$ be a d'-Cauchy sequence. Then there is either a subsequence such that $d'(x_{n_k}, \partial) \geq c > 0$, or $d'(x_{n_k}, \partial) \to 0$. In the latter case, by definition $x_{n_k} \to \partial$ with respect to d', and hence the whole sequence converges. In the former case, there is $k^* \in \mathbb{N}$ such that for every $k, \ell > k^*$: $d'(x_{n_k}, x_{n_\ell}) = d(x_{n_k}, x_{n_\ell}) \to 0$. Since X is complete, there exists a limit in $X \setminus B_c(X \setminus Y) \subset Y$. □

Locally in Y, d' and d coincide.

Lemma 2.2.4. *Let $x \in Y$ Then there is $r > 0$ such that for every $y, z \in B_r^d(x)$*

$$d'(y, z) = d(y, z).$$

Proof. Since Y is open, there is $r^* > 0$ such that $B_{r^*}^d(x) \subset Y$. But then, given $y, z \in B_{r^*/2}(x)$, we have

$$d(y, z) \leq r^* \quad \text{and} \quad d'(y, \partial) + d'(z, \partial) \geq r^*,$$

and hence $d'(y, z) = d(y, z)$. □

In general, a subset of a geodesic space is geodesic if and only if it is convex. However, the one-point completion is intuitively speaking always geodesic.

Lemma 2.2.5. *Assume X is complete and geodesic, $Y \subset X$ open and non-trivial, and $X \setminus Y$ is proper. Then (Y', d') is geodesic.*

Proof. We will show the existence of midpoints. Depending on which expression the distance $d'(x, y)$ takes, this means that we are either taking a midpoint in X and showing that it is also one in Y', or using geodesics in X and putting them together to a curve in Y' and taking its midpoint as a candidate for a midpoint with respect to d'.

Case 1: $d'(x, y) = d(x, y)$
Let $z \in X$ be a midpoint between x and y with respect to d. If $z \in Y$, then

$$d'(x, z) \le d(x, z) \qquad \text{and} \qquad d'(y, z) \le d(y, z), \tag{2.2.2}$$

so that

$$d'(x, y) \le d'(x, z) + d'(y, z) \le d(x, z) + d(y, z) = d(x, y) = d'(x, y).$$

Hence we have equality everywhere, which together with (2.2.2) implies that

$$d'(x, z) = d(x, z) = \frac{1}{2} d'(x, y) = d(y, z) = d'(y, z).$$

Now, if $z \notin Y$, then $z \in X \setminus Y$ and in this case ∂ is a midpoint between x and y with respect to d'. Indeed, following the same strategy as before, we have that by definition

$$d'(x, \partial) \le d(x, z) \qquad \text{and} \qquad d'(y, \partial) \le d(y, z)$$

and

$$d'(x, y) \le d'(x, \partial) + d'(y, \partial) \le d(x, z) + d(y, z) = d(x, y) = d'(x, y).$$

Again, this being an equality implies

$$d'(x, \partial) = d(x, z) = \frac{1}{2} d(x, y) = \frac{1}{2} d'(x, y).$$

Case 2: $d'(x, y) = d'(x, \partial) + d'(y, \partial)$
Let $z_k, w_k \in X \setminus Y$ be minimizing sequences for $d'(x, \partial)$ and $d'(y, \partial)$ respectively. They can be chosen such that $d(x, z_k)$ and $d(y, w_k)$ are monotonically non-increasing. In particular they are bounded sequences in $X \setminus Y$, so by the properness there are converging subsequences $z_{k_\ell} \to z^*$ and $w_{k_\ell} \to w^*$. Since $X \setminus Y$ is closed, $z^*, w^* \in X \setminus Y$ and

$$d'(x, \partial) = \lim_{\ell \to \infty} d(x, z_{k_\ell}) = d(x, z^*)$$

and

$$d'(y, \partial) = \lim_{\ell \to \infty} d(y, w_{k_\ell}) = d(y, w^*).$$

Let γ^1 be a d-geodesic (in X) connecting x and z^*, and γ^2 a d-geodesic connecting w^* and y, see Figure 2.1. They have d-length $d(x, z^*)$ and $d(y, w^*)$ respectively. Let

$$M := \frac{1}{2} d'(x, y) = \frac{1}{2}(d(x, z^*) + d(y, w^*)).$$

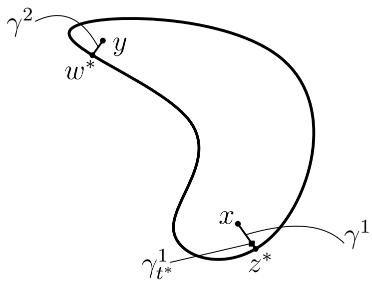

Figure 2.1: Geodesics γ^1 and γ^2 in Case 2

Without loss of generality assume that $d(x, z^*) \geq d(y, w^*)$, so that $M \leq d(x, z^*)$. Let $t^* \in [0, 1]$ be such that the d-length of $\gamma^1|_{[0,t^*]}$ is equal to M. Now we are going to show that $\gamma^1_{t^*}$ is a midpoint between x and y with respect to d'.

Claim 1: $\gamma^1((0, 1)) \subset Y$, $\gamma^2((0, 1)) \subset Y$.
If not, then there would be $s^* \in (0, 1)$ such that for instance $\gamma^1_{s^*} \in X \setminus Y$. But then, since γ^1 is a d-geodesic,

$$d(x, \gamma^1_{s^*}) = s^* d(x, z^*) < d(x, z^*) = d'(x, \partial)$$

which is in contradiction with the definition of $d'(x, \partial)$.

Claim 2: $d'(x, \gamma^1_{t^*}) = \frac{1}{2} d'(x, y)$.
Since γ^1 is a d-geodesic, and $d' \leq d$, we see that

$$\begin{aligned} d(x, \gamma^1_{t^*}) + d(\gamma^1_{t^*}, z^*) = d(x, z^*) =& d'(x, \partial) \\ \leq& d'(x, \gamma^1_{t^*}) + d'(\gamma^1_{t^*}, \partial) \\ \leq& d(x, \gamma^1_{t^*}) + d(\gamma^1_{t^*}, z^*) \end{aligned}$$

so that in fact equality holds everywhere. Thus the trivial inequalities $d'(x, \gamma^1_{t^*}) \leq d(x, \gamma^1_{t^*})$ and $d'(\gamma^1_{t^*}, z^*) \leq d(\gamma^1_{t^*}, z^*)$ are actually *equalities* and eventually

$$d'(x, \gamma^1_{t^*}) = d(x, \gamma^1_{t^*}) = L_d(\gamma^1|_{[0,t^*]}) = M = \frac{1}{2} d'(x, y).$$

Claim 3: $d'(y, \gamma^1_{t^*}) = \frac{1}{2} d'(x, y)$.
Without loss of generality assume that $d'(x, y) < d(x, y)$ (otherwise we are in Case 1). First observe that $d'(\gamma^1_{t^*}, \partial) = d(\gamma^1_{t^*}, z^*)$ because otherwise z^* would not be optimal for the distance $d'(x, \partial)$. Then

$$\begin{aligned} d'(y, \gamma^1_{t^*}) &\leq d'(y, \partial) + d'(\gamma^1_{t^*}, \partial) = d(y, w^*) + d(\gamma^1_{t^*}, z^*) \\ &= d(y, w^*) + (1 - t^*) d(x, z^*) \\ &= d(y, w^*) + d(x, z^*) - \underbrace{d(x, \gamma^1_{t^*})}_{=M} \\ &= d(y, w^*) + d(x, z^*) - \frac{1}{2} d(x, z^*) - \frac{1}{2} d(y, w^*) \end{aligned}$$

$$= \frac{1}{2}d(x, z^*) + \frac{1}{2}d(y, w^*)$$
$$= \frac{1}{2}d'(x, y).$$

In fact this is an equality. If it were not, then the inequality would be strict, so that $d'(y, \gamma^1_{t^*}) = d(y, \gamma^1_{t^*})$ by the definition of d'. Consequently, also incorporating Claim 2,

$$\begin{aligned} d'(x, y) < d(x, y) \leq d(x, \gamma^1_{t^*}) + d(\gamma^1_{t^*}, y) &= d'(x, \gamma^1_{t^*}) + d'(y, \gamma^1_{t^*}) \\ &< \frac{1}{2}d'(x, y) + \frac{1}{2}d'(x, y) = d'(x, y). \end{aligned}$$

This contradiction shows that $d'(y, \gamma^1_{t^*}) = \frac{1}{2}d'(x, y)$, which completes the proof. $\square$

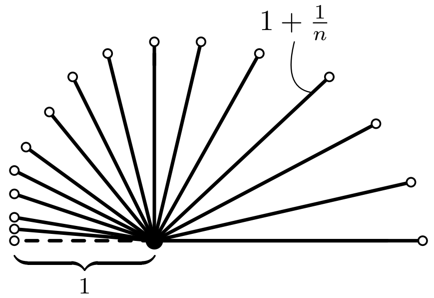

Figure 2.2: Gluing together countably many intervals. The endpoints are the boundary point in the one-point completion.

Remark 2.2.6. If X is geodesic but $X \setminus Y$ is not proper, then in general the one-point completion will not be geodesic. This can be seen by an example suggested in [BH99, Exercise 5.25(3)]: Let X be a metric graph consisting of countably many edges all of which are starting in a single vertex, the n-th having length $1 + \frac{1}{n}$. As Y we take everything but the "free" endpoints of the edges, see Figure 2.2. Then in (Y', d') the distance between the vortex and the boundary point ∂ is 1, but there is no geodesic between them.

The following elementary lemma will be used several times in the proof of the subsequent Lemma.

Lemma 2.2.7. *Let* $a, b, a', b', c' \in [0, \infty)$ *and* $\varepsilon > 0$ *with*

$$c' \leq a' + b' \leq a + b \leq c' + \varepsilon$$

and

$$a' \leq a \quad \text{and} \quad b' \leq b.$$

Then

$$a' \geq a - \varepsilon \quad \text{and} \quad b' \geq b - \varepsilon.$$

Proof. Assume for the sake of a contradiction that instead, say, $a' < a - \varepsilon$. Then

$$c' \leq a' + b' < a - \varepsilon + b \leq a + b - \varepsilon \leq c' + \varepsilon - \varepsilon = c',$$

which is a contradiction. $\square$

Considering length spaces, we do *not* need the extra assumption that the complement is proper.

Lemma 2.2.8. *Let (X, d) be a complete length space. Then (Y', d') is also a length space.*

Proof. This proof is an adaption of the one for geodesic spaces.

Case 1: $d'(x, y) = d(x, y)$
Let $z \in X$ be an ε-midpoint between x and y with respect to d. Assume first that $z \in Y$. Then

$$d'(x, y) \leq d'(x, z) + d'(y, z) \leq d(x, z) + d(y, z) \leq d(x, y) + \varepsilon = d'(x, y) + \varepsilon. \quad (2.2.3)$$

Although we cannot conclude that $d'(x, z)$ and $d(x, z)$ are equal, one can show that they actually do not differ much. Indeed, by definition of d', we have $d'(x, z) \leq d(x, z)$. And the previous Lemma 2.2.7 applied to (2.2.3) yields that $d'(x, z) \geq d(x, z) - \varepsilon$ and $d'(y, z) \geq d(y, z) - \varepsilon$. So we finally get

$$d'(x, z) - d'(y, z) \leq d(x, z) - d(y, z) + \varepsilon \leq 2\varepsilon$$

and

$$d'(y, z) - d'(x, z) \leq d(y, z) - d(x, z) + \varepsilon \leq 2\varepsilon,$$

which proves that z is a 2ε-midpoint between x and y with respect to d'.
This proof works exactly in the same way in the case $z \in X \setminus Y$, showing that ∂ is a 2ε-midpoint between x and y with respect to d'.

Case 2: $d'(x, y) = d'(x, \partial) + d'(y, \partial)$
Given $\varepsilon > 0$, let $z^*, w^* \in X \setminus Y$ such that

$$d(x, z^*) \leq d'(x, \partial) + \frac{\varepsilon}{2} \quad \text{and} \quad d(y, w^*) \leq d'(y, \partial) + \frac{\varepsilon}{2}. \quad (2.2.4)$$

Further, take two d-almost-geodesics, i.e. curves $\gamma^1, \gamma^2 \colon [0, 1] \to X$ with $\gamma^1_0 = x, \gamma^1_1 = z, \gamma^2_0 = w, \gamma^2_1 = y$ and

$$L_d(\gamma^1) \leq d(x, z^*) + \frac{\varepsilon}{2} \quad \text{and} \quad L_d(\gamma^2) \leq d(y, w^*) + \frac{\varepsilon}{2}. \quad (2.2.5)$$

Claim 1: Without loss of generality for every $\delta > 0$ it holds $\gamma^1((0, 1 - \delta)), \gamma^2((0, 1 - \delta)) \subset Y$.
Contrary to the previous proof, it is not clear if these curves stay in Y. But in fact we can assume they almost do, because in the case they don't, we take the restriction of the curves to the first time they leave Y. Let us discuss it in detail for γ^1. Let $s^* := \inf\{s > 0 \mid \gamma^1_s \in X \setminus Y\}$. Since we don't know if the infimum is attained, take

$\delta > 0$ and consider the restriction $\tilde{\gamma}^1 := \gamma^1|_{[0,s^*+\delta]}$. Then $\tilde{\gamma}^1((0,s^*)) \subset Y$, and there is $s \in [s^*, s^*+\delta)$ such that $\tilde{\gamma}^1_s \in X \setminus Y$. The d-length of $\tilde{\gamma}^1$ is at most the one of the original curve, so we possibly reach a closer boundary point.

Now let us assume without loss of generality that $d(x,z^*) \geq d(y,w^*)$. Then let $t^* \in [0,1]$ be such that $L_d(\gamma^1|_{[0,t^*]}) = \frac{1}{2}(L_d(\gamma^1) + L_d(\gamma^2))$. A candidate for being an ε-midpoint is now $\gamma^1_{t^*}$.

Claim 2: $\left|d'(x,\gamma^1_{t^*}) - \frac{1}{2}d'(x,y)\right| \leq 2\varepsilon$.
Using the additivity of the lengths of curves and (2.2.5), (2.2.4), we see that

$$\begin{aligned} d(x,\gamma^1_{t^*}) + d(\gamma^1_{t^*}, z^*) &\leq L_d(\gamma^1|_{[0,t^*]}) + L_d(\gamma^1|_{[t^*,1]}) \\ &= L_d(\gamma^1) \\ &\leq d(x,z^*) + \frac{\varepsilon}{2} \\ &\leq d'(x,\partial) + \varepsilon \\ &\leq d'(x,\gamma^1_{t^*}) + d'(\gamma^1_{t^*},\partial) + \varepsilon \\ &\leq d(x,\gamma^1_{t^*}) + d(\gamma^1_{t^*}, z^*) + \varepsilon. \end{aligned}$$

By Lemma 2.2.7 we get that

$$d'(x,\gamma^1_{t^*}) \geq d(x,\gamma^1_{t^*}) - \varepsilon. \tag{2.2.6}$$

A similar further application of that Lemma also yields that restricted almost-geodesics are still almost-geodesics between their endpoints, i.e.

$$L_d(\gamma^1|_{[0,t^*]}) \leq d(x,\gamma^1_{t^*}) + \varepsilon. \tag{2.2.7}$$

Now we can complete this step by observing that thanks to (2.2.5) and (2.2.4)

$$d'(x,\gamma^1_{t^*}) \leq d(x,\gamma^1_{t^*}) \leq L_d(\gamma^1|_{[0,t^*]}) = \frac{1}{2}(L_d(\gamma^1) + L_d(\gamma^2)) \leq \frac{1}{2}d'(x,y) + \varepsilon,$$

and by (2.2.7) and (2.2.6)

$$\frac{1}{2}d'(x,y) \leq \frac{1}{2}(L_d(\gamma^1) + L_d(\gamma^2)) = L_d(\gamma^1|_{[0,t^*]}) \leq d(x,\gamma^1_{t^*}) + \varepsilon \leq d'(x,\gamma^1_{t^*}) + 2\varepsilon.$$

Claim 3: $\left|d'(y,\gamma^1_{t^*}) - \frac{1}{2}d'(x,y)\right| \leq \varepsilon$.
We can assume that $d'(x,y) < d(x,y)$ because otherwise we are in Case 1. Observe that $d'(\gamma^1_{t^*},\partial) \geq d(\gamma^1_{t^*},z^*) - \frac{\varepsilon}{2}$ (if not, then we would again find a closer boundary point). Using that $L_d(\gamma^1|_{[t^*,1]}) = \frac{1}{2}(L_d(\gamma^1)+L_d(\gamma^2)) - L_d(\gamma^2)$, and once more (2.2.5) and (2.2.4), we obtain

$$\begin{aligned} d'(\gamma^1_{t^*}, y) &\leq d'(\gamma^1_{t^*},\partial) + d'(y,\partial) \\ &\leq d(\gamma^1_{t^*}, z^*) + d(y,w^*) \\ &\leq L_d(\gamma^1|_{[t^*,1]}) + L_d(\gamma^2) \\ &= \frac{1}{2}(L_d(\gamma^1) + L_d(\gamma^2)) - L_d(\gamma^2) + L_d(\gamma^2) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}(L_d(\gamma^1) + L_d(\gamma^2)) \\
&\le \frac{1}{2}(d'(x,\partial) + d'(y,\partial)) + \varepsilon \\
&= \frac{1}{2}d'(x,y) + \varepsilon.
\end{aligned}
$$

Finally, we have to show that also $\frac{1}{2}d'(x,y) \le d'(\gamma^1_{t^*}, y)+\varepsilon$. For the sake of contradiction, let us assume that this is not the case, so that instead $\frac{1}{2}d'(x,y) > d'(\gamma^1_{t^*}, y)+\varepsilon$. Using Claim 2, this leads to

$$
\begin{aligned}
d'(x,y) &\le d'(x,\gamma^1_{t^*}) + d'(\gamma^1_{t^*}, y) \\
&\qquad < d'(\gamma^1_{t^*}, x) + \frac{1}{2}d'(x,y) - \varepsilon \le \frac{1}{2}d'(x,y) - \varepsilon + \frac{1}{2}d'(x,y) + \varepsilon,
\end{aligned}
$$

which is a contradiction.

Thus we have found a 4ε-midpoint between x and y with respect to d'. □

Remark 2.2.9. For $X = \mathbb{R}, Y = (0, 2\pi)$, the completion Y' is isometric to the one-sphere $\mathbb{S}^1$. However, for $X = \mathbb{R}^2, Y = B_1(0)$, the resulting space is *not* isometric to a standard sphere, since locally $d' = d$, which yields that locally the curvature of (Y', d') in the sense of Alexandrov is zero, whereas the sphere has constant positive curvature. Furthermore, the completion is in general branching as can be seen from the disk example since every geodesic to the boundary point ∂ can branch at this point in any direction while staying a geodesic.

Lemma 2.2.10. *Let (X,d) be a complete metric space, and $Y \subset X$ open, non-trivial and totally bounded. Then (Y', d') is compact.*

Proof. Due to the metric version of the Heine-Borel theorem, a metric space is compact if and only if it is complete and totally bounded. □

Remark 2.2.11. A somewhat similar metric has been studied in [Man89] in connection with the one-point compactification. However, the one-point completion will in general not be compact, even if we start with a locally compact space (as is necessary for the one-point compactification). Taking for instance the half-plane, its one-point completion is still an unbounded space.

2.3 Analysis in Metric Measure Spaces and Heat Flows

This section is devoted to the analysis in metric measure spaces. We use the recent approach developed in [AGS14a, AGS14b], and for the later part on Dirichlet forms we refer to [FOT94, MR92].

Let us introduce our main object of study.

Definition 2.3.1. A *metric measure space* is a triple $(X, d, \mathfrak{m})$ consisting of a complete, separable metric space (X,d) and a σ-finite Borel measure $\mathfrak{m}$ with full support

$\operatorname{supp} \mathfrak{m} = X$, that satisfies the exponential integrability condition

$$\int_X e^{-cd(x,x_0)^2}\,\mathrm{d}\mathfrak{m}(x) < \infty \tag{2.3.1}$$

for some $c > 0, x_0 \in X$. In particular, every bounded set has finite measure, and hence $\mathfrak{m}$ is locally finite.

To do some sort of calculus, it is a good idea to start with difference quotients. We denote by $\mathrm{Lip}(X, d)$ the space of Lipschitz continuous functions $f \colon X \to \mathbb{R}$.

Definition 2.3.2. i) The *slope*, or *local Lipschitz constant*, of a function $f \colon X \to \mathbb{R}$ at $x \in X$ is

$$\mathrm{lip}(f)(x) := \limsup_{y \to x} \frac{|f(x) - f(y)|}{d(x,y)}.$$

ii) The *Cheeger energy* is the functional $\mathrm{Ch} \colon L^2(X, \mathfrak{m}) \to [0, \infty]$,

$$\mathrm{Ch}(f) := \inf \left\{ \liminf_{k \to \infty} \frac{1}{2} \int_X |\operatorname{lip}(f)|^2\,\mathrm{d}\mathfrak{m} \,\middle|\, (f_k)_{k \in \mathbb{N}} \subset \mathrm{Lip}(X,d),\ f_k \to f \text{ in } L^2 \right\}.$$

We denote its domain by $W^{1,2}(X, d, \mathfrak{m}) := D(\mathrm{Ch}) = \{f \in L^2(X, \mathfrak{m}) \mid \mathrm{Ch}(f) < \infty\}$.

Theorem 2.3.3. *Let $f \in W^{1,2}(X, d, \mathfrak{m})$.*

i) The collection of weak gradients of f,

$$\{G \in L^2(X, \mathfrak{m}) \mid \exists\, f_k \in \mathrm{Lip}(X,d) : f_k \to f \text{ in } L^2(X,\mathfrak{m}), \\ \mathrm{lip}(f_k) \rightharpoonup G \text{ in } L^2(X, \mathfrak{m})\},$$

is a non-empty, closed, convex subset of $L^2(X, \mathfrak{m})$ and thus contains a unique element of minimal L^2-norm, which we will call the minimal weak gradient of f and denote by $|\nabla f|$. The minimal weak gradient $|\nabla f|$ is also minimal in the $\mathfrak{m}$-a.e. sense.

ii) This minimal weak gradient provides an integral representation of the Cheeger energy, i.e. for $f \in W^{1,2}(X, d, \mathfrak{m})$ we can write

$$\mathrm{Ch}(f) = \frac{1}{2} \int_X |\nabla f|^2\,\mathrm{d}\mathfrak{m}.$$

iii) The Cheeger energy is a 2-homogeneous, lower semicontinuous, convex functional on $L^2(X, \mathfrak{m})$.

iv) Equipped with the norm

$$\|f\|_{W^{1,2}}^2 := \|f\|_2^2 + \||\nabla f|\|_2^2 = \|f\|_2^2 + 2\,\mathrm{Ch}(f)$$

the space $(W^{1,2}(X, d, \mathfrak{m}), \|\cdot\|_{W^{1,2}})$ is a Banach space.

Remark 2.3.4. a) The Cheeger energy will in general not be a quadratic form, and in consequence $W^{1,2}$ will in general only be a Banach and *not* a Hilbert space.

b) In the sequel, we will omit the "minimal" and just speak of the weak gradient or even just the gradient.

A useful property is the locality of weak gradients in the sense that computing Cheeger energies and weak gradients in a sub-metric measure space yields the same as restricting to the subset:

Lemma 2.3.5 ([AGS14b, Thm. 4.19])**.** *Let $(X, d, \mathfrak{m})$ be a metric measure space, and $\Omega \subset X$ an open subset with $\mathfrak{m}(\Omega) > 0$ and $\mathfrak{m}(\partial\Omega) = 0$. Let $f \in W^{1,2}(X, d, \mathfrak{m})$. Then $\tilde{f} := f|_{\overline{\Omega}} \in W^{1,2}(\overline{\Omega}, d|_{\overline{\Omega}\times\overline{\Omega}}, \mathfrak{m}|_{\overline{\Omega}})$ and*

$$|\nabla \tilde{f}|_{\overline{\Omega}} = (|\nabla f|)|_{\overline{\Omega}} \ \ \mathfrak{m}\text{-a.e. in } \overline{\Omega},$$

where $|\nabla \tilde{f}|_{\overline{\Omega}}$ is the weak gradient given in the space $(\overline{\Omega}, d|_{\overline{\Omega}\times\overline{\Omega}}, \mathfrak{m}|_{\overline{\Omega}})$.

Already in this generality one gets many properties for the weak gradient, and one could define a Laplacian and a heat flow by means of convex analysis and the theory of gradient flows in Hilbert spaces. As can be seen for example in Finsler manifolds, these operators might not be linear (see for instance [Gig15, Sec. 1.1]). For us, however, it will be necessary to restrict to those metric measure spaces in which these operators are indeed linear. In this case, one can use the theory of Dirichlet forms and give more convenient definitions.

Definition 2.3.6. A metric measure space $(X, d, \mathfrak{m})$ is called *infinitesimally Hilbertian* if the Cheeger energy is a quadratic form, i.e if it satisfies the parallelogram identity

$$\mathrm{Ch}(f+g) + \mathrm{Ch}(f-g) = 2\,\mathrm{Ch}(f) + 2\,\mathrm{Ch}(g) \quad \text{for every } \ f, g \in W^{1,2}(X, d, \mathfrak{m}).$$

Remark 2.3.7. This is equivalent to requiring the Sobolev space $W^{1,2}(X, d, \mathfrak{m})$ to be a Hilbert space.

By polarization, this makes it possible to define a symmetric bilinear map which takes the role of the scalar product between the gradients of two functions.

Definition 2.3.8. Let $(X, d, \mathfrak{m})$ be infinitesimally Hilbertian, $f, g \in W^{1,2}(X, d, \mathfrak{m})$. Then we define $\langle \nabla f, \nabla g \rangle \colon X \to \mathbb{R}$ by

$$\nabla f \cdot \nabla g := \frac{1}{4}\left(|\nabla(f+g)|^2 - |\nabla(f-g)|^2\right).$$

Weak gradients satisfy the expected calculus rules.

Theorem 2.3.9. *Let $(X, d, \mathfrak{m})$ be infinitesimally Hilbertian. Then*

$$W^{1,2}(X, d, \mathfrak{m}) \times W^{1,2}(X, d, \mathfrak{m}) \to L^1(X, \mathfrak{m}), \ \ (f, g) \mapsto \nabla f \cdot \nabla g$$

is a symmetric, bilinear, continuous map, and further satisfies:

i) Cauchy-Schwarz: For $f, g \in W^{1,2}(X, d, \mathfrak{m})$:

$$|\nabla f \cdot \nabla g| \leq |\nabla f||\nabla g| \quad \textit{and} \quad \nabla f \cdot \nabla f = |\nabla f|^2.$$

ii) Chain rule: For $f, g \in W^{1,2}(X, d, \mathfrak{m}), \varphi \colon \mathbb{R} \to \mathbb{R}$ Lipschitz it holds $\varphi \circ f \in W^{1,2}(X, d, \mathfrak{m})$ and

$$\nabla(\varphi \circ f) \cdot \nabla g = \varphi' \circ f \nabla f \cdot \nabla g$$

and in particular

$$|\nabla(\varphi \circ f)| = |\varphi' \circ f||\nabla f|.$$

(Where we set $\varphi'(f(x)) = 0$ if φ is not differentiable in $f(x)$.)

iii) Leibniz rule: Given $f, g, h \in W^{1,2}(X, d, \mathfrak{m}) \cap L^\infty(X, \mathfrak{m})$, we have that $fg \in W^{1,2}(X, d, \mathfrak{m})$ and

$$\nabla(fg) \cdot \nabla h = g \nabla f \cdot \nabla h + f \nabla g \cdot \nabla h$$

as well as

$$|\nabla(fg)| \leq |f||\nabla g| + |g||\nabla f|.$$

Theorem 2.3.10. *Let $(X, d, \mathfrak{m})$ be an infinitesimally Hilbertian metric measure space. Then*

$$\begin{cases} \mathcal{E}(f, g) := \int_X \nabla f \cdot \nabla g \, d\mathfrak{m}, & f, g \in D(\mathcal{E}) \\ D(\mathcal{E}) := W^{1,2}(X, d, \mathfrak{m}) \end{cases}$$

is a Dirichlet form on $L^2(X, \mathfrak{m})$, which means that $D(\mathcal{E})$ is dense in $L^2(X, \mathfrak{m})$, and $\mathcal{E}$ is a symmetric, bilinear form, that is additionally non-negative definite (i.e. $\mathcal{E}(f, f) \geq 0$ for every $f \in D(\mathcal{E})$), closed (i.e. $D(\mathcal{E})$ is complete when equipped with the Sobolev norm $\| \cdot \|_{\mathcal{E}} := \| \cdot \|_{W^{1,2}}$), and Markovian (i.e. for every 1-Lipschitz $\varphi \colon \mathbb{R} \to \mathbb{R}$ with $\varphi(0) = 0$ and $f \in D(\mathcal{E})$ we have $\mathcal{E}(\varphi \circ f) \leq \mathcal{E}(f)$).
Furthermore, it is strongly local, meaning that for $f, g \in D(\mathcal{E})$ such that g is constant on $\{f \neq 0\}$, we get $\mathcal{E}(f, g) = 0$.

The closedness follows from the lower semicontinuity of the Cheeger energy in L^2, while the Markovianity follows from the chain rule for weak gradients. Finally, the strong locality follows from a similar property of the weak gradient.

Remark 2.3.11. By abuse of notation, we will write $\mathcal{E}(f) := \mathcal{E}(f, f) = 2 \operatorname{Ch}(f)$ for the quadratic form associated to the Dirichlet form. From this one can go back to the Dirichlet form by polarization.

Using Dirichlet form theory, we can easily define the Laplacian (by integration by parts) and heat flow associated to the Cheeger energy.

Theorem 2.3.12. *Let $(X, d, \mathfrak{m})$ be an infinitesimally Hilbertian metric measure space.*

i) There is a densely defined, non-positive, self-adjoint operator $(\Delta, D(\Delta))$*, called Laplacian, connected to the Dirichlet form via integration by parts: for every* $f \in D(\Delta)$*,* $g \in W^{1,2}(X, d, \mathfrak{m})$

$$-\int_X \Delta f g \,\mathrm{d}\mathfrak{m} = \mathcal{E}(f, g).$$

ii) There is a strongly continuous contraction semigroup $(P_t)_{t>0}$ *on* $L^2(X, \mathfrak{m})$ *whose generator is the Laplacian, i.e. for every* $f \in D(\Delta)$*:*

$$\Delta f = \lim_{t \to 0} \frac{P_t f - f}{t},$$

where the limit is taken in L^2*. Conversely, by the spectral theory for self-adjoint operators it is rigorous to write* $P_t = e^{t\Delta}$*.*

iii) For $t > 0$ *and* $f, g \in L^2(X, \mathfrak{m})$ *let*

$$\mathcal{E}_t(f, g) := -\frac{1}{t} \int_X g(P_t f - f) \,\mathrm{d}\mathfrak{m}$$

be the approximate form associated to P_t*. Then we can recover the corresponding Dirichlet form by*

$$\begin{cases} D(\mathcal{E}) = \{f \in L^2(X) \,\big|\, \lim_{t \to 0} \mathcal{E}_t(f, f) < \infty\} \\ \mathcal{E}(f, g) = \lim_{t \to 0} \mathcal{E}_t(f, g), \text{ for } f, g \in D(\mathcal{E}). \end{cases} \tag{2.3.2}$$

Furthermore, for $f \in L^2(X, \mathfrak{m})$ *the map* $(0, \infty) \ni t \mapsto \mathcal{E}_t(f, f)$ *is non-increasing and non-negative.*

Remark 2.3.13. We will sometimes call P_t and Δ the heat flow and the Laplacian *with Neumann boundary conditions*. In the case where X is a closed, bounded subset of $\mathbb{R}^n$, it actually *is* the classical heat flow with Neumann conditions (meaning that the normal derivative at the boundary vanishes).

Since we are particularly interested in heat flows, let us give some more properties of P_t.

Theorem 2.3.14. *Let* $f, g \in L^2(X, \mathfrak{m})$*, and* $P_t f, P_t g$ *the corresponding heat flows starting at* f, g *respectively. Then:*

i) Maximum/Comparison principle: Let $C \in \mathbb{R}$*. If* $f \leq C$*, then* $P_t f \leq C$ *for every* $t \geq 0$*. Analogously, if* $f \geq C$*, then* $P_t f \geq C$ *for every* $t \geq 0$*. If* $f \leq g + C$*, then* $P_t f \leq P_t g + C$ *for every* $t \geq 0$*.*

ii) Mass-preservation: For every $t \geq 0$ *it holds* $\int_X P_t f \,\mathrm{d}\mathfrak{m} = \int_X f \,\mathrm{d}\mathfrak{m}$*.*

iii) P_t *is a contraction in every* $L^p(X, \mathfrak{m}), p \in [1, \infty]$*, i.e.*

$$\|P_t f\|_p \leq \|f\|_p \quad \text{for every } f \in L^2(X, \mathfrak{m}) \cap L^p(X, \mathfrak{m}).$$

iv) The heat flow P_t is an analytic semigroup; in particular $P_t\left(L^2(X,\mathfrak{m})\right) \subset D(\Delta^m)$ for every $t>0$ and $\mathfrak{m} \in \mathbb{N}$.

Remark 2.3.15. a) The mass-preservation is due to the exponential integrability of the measure.

b) Since $L^2(X,\mathfrak{m}) \cap L^p(X,\mathfrak{m})$ is dense in $L^p(X,\mathfrak{m})$, one can extend P_t to a continuous contraction semigroup in every $L^p(X,\mathfrak{m})$.

The Dirichlet form enjoys some further regularity property. Let us recall some definitions: given an open set $U \subset X$, its *capacity* is defined by

$$\operatorname{cap}(U) := \inf\{\|f\|_{\mathcal{E}}^2 \,\big|\, f \in D(\mathcal{E}), f \geq 1 \ \mathfrak{m}\text{-a.e. on } U\},$$

and for an arbitrary set $A \subset X$

$$\operatorname{cap}(A) := \inf\{\operatorname{cap}(U) \,\big|\, U \text{ open}, U \supset A\}.$$

Given a subset $F \subset X$, define $D(\mathcal{E}_F) := \{f \in D(\mathcal{E}) \,|\, f = 0 \text{ a.e. on } X \backslash F\}$. A sequence of subsets $F_k \subset X, k \in \mathbb{N}$, is an *$\mathcal{E}$-nest*, if the F_k are closed, $F_k \subset F_{k+1}$, and

$$\bigcup_{k\in\mathbb{N}} D(\mathcal{E}_{F_k}) \ \text{ is dense in } D(\mathcal{E}).$$

A set $N \subset X$ is called *$\mathcal{E}$-polar* if $\operatorname{cap}(N) = 0$. A property is said to hold *quasi-everywhere* if there is an $\mathcal{E}$-polar set $N \subset X$ such that the property holds everywhere in $X \backslash N$. Functions $f\colon X \to \mathbb{R}$ are *quasi-continuous* if there exists an $\mathcal{E}$-nest $(F_k)_{k\in\mathbb{N}}$ such that for every $k \in \mathbb{N}$, $f|_{F_k}\colon F_k \to \mathbb{R}$ is continuous.

Theorem 2.3.16. *The Dirichlet form $(\mathcal{E}, D(\mathcal{E}))$ is quasi-regular, i.e.*

i) there is an $\mathcal{E}$-nest consisting of compact sets, and

ii) there is a dense subset of $D(\mathcal{E})$ admitting quasi-continuous representatives, and

iii) there is an $\mathcal{E}$-polar set $N \subset X$ and a countable set of quasi-continuous functions in $D(\mathcal{E})$ that separates the points of $X \setminus N$.

The quasi-regularity is a consequence of the density of Lipschitz functions in $D(\mathcal{E})$ and it uses the exponential integrability of the reference measure. A proof of this can be found in [Sav14, Thm. 4.1] (which assumes a curvature bound for the theorem that is not needed in the proof of quasi-regularity).

There are many further regularity properties of the heat flow. We will not state them here right now. Instead, we will introduce the heat flow with Dirichlet boundary values (which in contrast is *not* mass-preserving) and then prove those properties for this flow in the next section, noting that everything works the same for P_t.

Theorem 2.3.17. *Let $(X,d,\mathfrak{m})$ be an infinitesimally Hilbertian metric measure space, let $Y \subset X$ be open with $\mathfrak{m}(\partial Y) = 0$. Then*

$$\begin{cases} D(\mathcal{E}^0) := \{f \in D(\mathcal{E}) \,\big|\, \tilde{f} = 0 \textit{ quasi-everywhere on } X \setminus Y\} \\ \mathcal{E}^0(f) := \mathcal{E}(f) \textit{ for } f \in D(\mathcal{E}^0), \end{cases}$$

where $\tilde{f}$ is a quasi-continuous representative of f, is a strongly local, quasi-regular Dirichlet form on $L^2(Y, \mathfrak{m}|_Y)$.
As such, it has an associated generator $(\Delta^0, D(\Delta^0))$, semigroup $(P_t^0)_{t>0}$ and approximate form

$$\mathcal{E}_t^0(f,g) := -\frac{1}{t}\int_X g(P_t^0 f - f)\,\mathrm{d}\mathfrak{m},$$

all with the same properties as the ones for $\mathcal{E}$ presented in Theorem 2.3.12.
Furthermore, the heat flow P_t^0 satisfies the same properties from Theorem 2.3.14 except the mass-preservation.

Remark 2.3.18. One can identify $L^2(Y, \mathfrak{m}|_Y)$ with $\{f \in L^2(X, \mathfrak{m}) \mid f = 0 \text{ a.e. on } X \setminus Y\}$. Thus, we will extend functions defined on Y by zero to all of X without explicitly mentioning it. One could even define the heat flow with Dirichlet boundary values for every function in $L^2(X, \mathfrak{m})$ by saying $P_t^0 f = 0$ on $X \setminus Y$. However, for those functions the heat flow will not be continuous in $t = 0$.

Thanks to the quasi-regularity of the Dirichlet form, there exists an associated stochastic process which can be used to define a Markov kernel representing the heat semigroup. Let us avoid giving details about the process and let us instead refer to [FOT94]. The result being the following:

Proposition 2.3.19. *Let $(X, d, \mathfrak{m})$ be an infinitesimally Hilbertian metric measure space. Then there exists a semigroup of sub-Markovian kernels associated to the Dirichlet form $(\mathcal{E}, D(\mathcal{E}))$, i.e. for every $t > 0$ a map $p_t \colon X \times \mathcal{B}(X) \to [0,1]$ such that*

i) $p_t(x, \cdot)$ is a Borel measure on X for every $x \in X$,

ii) $p_t(\cdot, A)$ is a Borel-measurable function for every $A \subset X$ Borel,

iii) $p_t p_s f = p_{t+s} u$ for every $s, t > 0$ and every bounded, Borel-measurable $f \colon X \to \mathbb{R}$, where we write

$$p_t f(x) := \int_X f(y) p_t(x, \mathrm{d}y). \tag{2.3.3}$$

This kernel provides a version of P_t, i.e. $P_t f = p_t f$ $\mathfrak{m}$-a.e.

Analogously we get a sub-Markovian kernel p_t^0 corresponding to $\mathcal{E}^0$.

These kernels provide us with a tool to extend the heat flow to bounded Borel-measurable functions by (2.3.3) on the one hand, and to define a dual heat flow for measures on the other hand: for $\mu \in \mathcal{P}(X)$ let

$$\mathscr{P}_t \mu(A) := \int_X p_t(x, A)\,\mathrm{d}\mu(x).$$

The heat semigroups P_t and $\mathscr{P}_t$ are dual in the following sense: For $f \colon X \to \mathbb{R}$ bounded Borel, and $\mu \in \mathcal{P}(X)$ we have

$$\int_X P_t f(x)\,\mathrm{d}\mu(x) = \int_X \int_X f(y) p_t(x, \mathrm{d}y)\,\mathrm{d}\mu(x) \tag{2.3.4}$$

$$= \int_X f(y) \int_X p_t(x, \mathrm{d}y) \,\mathrm{d}\mu(x) = \int_X f(y) \,\mathrm{d}\mathscr{P}_t\mu(y). \qquad (2.3.5)$$

In the same way we define $P_t^0 f$ for f bounded, Borel, and $\mathscr{P}_t^0\mu$ for $\mu \in \mathcal{P}^{sub}(Y)$. They also satisfy the duality relation (2.3.5).

Remark 2.3.20. With the help of the sub-Markov kernels, all of these heat flows of measures can be extended to signed, finite Borel measures.

2.4 Regularity Properties of the Heat Flows

For the proof of an equivalence of a Bochner inequality and a gradient estimate (in which both Laplacians Δ and Δ^0, or heat flows P_t and P_t^0 appear, respectively; see Proposition 5.2.1) we need some further convergence results for the Laplacians and the heat flows. We will state and prove them for the Dirichlet boundary values. **However, the statements and proofs are literally the same for the "Neumann boundary value" objects, replacing all $P_t^0, \Delta^0, \mathcal{E}^0$ by $P_t, \Delta, \mathcal{E}$, and every $(Y, \mathfrak{m}|_Y)$ and $\mathfrak{m}(Y) < \infty$ by $(X, \mathfrak{m})$ and $\mathfrak{m}(X) < \infty$, respectively.**

In this section we continue to assume that $(X, d, \mathfrak{m})$ is an infinitesimally Hilbertian metric measure space. These results are only needed for the proof of Proposition 5.2.1 and can be skipped on first reading.

Lemma 2.4.1. *i) The semigroup P_t^0 satisfies a Jensen inequality for power functions, i.e. for every $p \in [1, \infty)$ and $f \in L^p(Y, \mathfrak{m}|_Y)$ we have*

$$|P_t^0 f|^p \le P_t^0 |f|^p \quad \mathfrak{m}|_Y\text{-a.e.}$$

ii) We can extend P_t^0 to a strongly continuous contraction semigroup in every $L^p(Y, \mathfrak{m}|_Y)$, $p \in [1, \infty)$, and to a weakly-$*$-continuous semigroup in $L^\infty(Y, \mathfrak{m}|_Y)$.

iii) Given $f \in D(\Delta^0) \cap L^p(Y, \mathfrak{m}|_Y)$ with $\Delta^0 f \in L^p(Y, \mathfrak{m}|_Y)$, then for $p \in [1, \infty)$

$$\frac{P_t^0 f - f}{t} \longrightarrow \Delta^0 f \text{ in } L^p \text{ as } t \to 0,$$

whereas for $p = \infty$ we have weak-$*$-convergence in L^∞.

Proof. i) Since $p_t^0(x, \cdot)$ is a finite measure, we get a probability measure $\tilde{p}_t^0(x, A) := \frac{1}{p_t^0(x,Y)} p_t^0(x, A)$. As such, it satisfies Jensen inequality, so that for a convex function $\varphi : \mathbb{R} \to \mathbb{R}$ we have

$$\begin{aligned}
\varphi\left(P_t^0 f(x)\right) &= \varphi\left(\int_Y f(y) p_t^0(x, \mathrm{d}y)\right) = \varphi\left(p_t^0(x, Y) \int_Y f(y) \tilde{p}_t^0(x, \mathrm{d}y)\right) \\
&\le \int_Y \varphi\left(p_t^0(x, Y) f(y)\right) \tilde{p}_t^0(x, \mathrm{d}y) \\
&= \frac{1}{p_t^0(x, Y)} \int_Y \varphi\left(p_t^0(x, Y) f(y)\right) p_t^0(x, \mathrm{d}y).
\end{aligned}$$

For the power functions $\varphi(a) = |a|^p$ with $p \geq 1$ we thus get

$$\begin{aligned}\left|P_t^0 f(x)\right|^p &\leq \frac{1}{p_t^0(x,Y)} \int_Y \left|p_t^0(x,Y) f(y)\right|^p p_t^0(x,\mathrm{d}y) \\ &= p_t^0(x,Y)^{p-1} \int_Y |f(y)|^p \, p_t^0(x,\mathrm{d}y) \leq P_t^0 |f|^p(x).\end{aligned}$$

ii) The proof works as in the case of P_t. For $f \in L^2 \cap L^p(Y, \mathfrak{m}|_Y)$, we have by Jensen's inequality

$$\|P_t^0 f\|_{L^p}^p = \int_Y |P_t^0 f|^p \,\mathrm{d}\mathfrak{m} \leq \int_Y P_t^0 |f|^p \,\mathrm{d}\mathfrak{m} \leq \int_Y P_t |f|^p \,\mathrm{d}\mathfrak{m} = \int_Y |f|^p \,\mathrm{d}\mathfrak{m} = \|f\|_{L^p}.$$

By density of $L^2 \cap L^p$ in L^p, we extend P_t^0 to a contraction in $L^p(Y, \mathfrak{m}|_Y)$.

To show the strong continuity in L^p, let us first consider the case $p > 2$. Take $f \in L^p$. Further take $\varepsilon > 0$, $q > p$ and a function $g \in L^q \cap L^2$ such that $\|f - g\|_{L^p} < \varepsilon$. Then

$$\begin{aligned}\|P_t^0 f - f\|_{L^p} &\leq \|P_t^0 f - P_t^0 g\|_{L^p} + \|P_t^0 g - g\|_{L^p} + \|g - f\|_{L^p} \\ &\leq 2\varepsilon + \|P_t^0 g - g\|_{L^p},\end{aligned}$$

by the contraction in L^p. Since $2 < p < q$, there is $\lambda \in (0,1)$ such that $p = 2\lambda + (1-\lambda)q$. Using an interpolation Hölder inequality, we get

$$\|P_t^0 g - g\|_{L^p}^p \leq \|P_t^0 g - g\|_{L^2}^{2\lambda} \|P_t^0 g - g\|_{L^q}^{(1-\lambda)q}. \tag{2.4.1}$$

As P_t^0 is also a contraction in L^q, $\|P_t^0 g - g\|_{L^q}^{(1-\lambda)q}$ is bounded. Hence the strong continuity in L^2 yields the strong continuity in L^p.

For $1 < p < 2$ we use the same strategy with a $q \in (1,p)$. The case $p = 1$ is shown in [BH91, Prop.2.4.2]. For $p = \infty$, we can define $P_t^0 \colon L^\infty \to L^\infty$ via duality. Then, for $\varphi \in L^\infty$ and $f \in L^1$ we have

$$\left|\int_Y f(P_t^0 \varphi - \varphi) \,\mathrm{d}\mathfrak{m}\right| = \left|\int_Y (P_t^0 f - f)\varphi \,\mathrm{d}\mathfrak{m}\right| \leq \|\varphi\|_{L^\infty} \|P_t^0 f - f\|_{L^1} \overset{t \to 0}{\longrightarrow} 0$$

by the strong continuity of P_t^0 in L^1.

iii) Consider the case $p < \infty$ first. By strong continuity of P_t^0 in L^p, we have

$$\frac{1}{t} \int_0^t P_r^0 \Delta^0 f \,\mathrm{d}r \to \Delta^0 f \text{ in } L^p. \tag{2.4.2}$$

Then

$$\int_Y \left|\frac{P_t^0 f - f}{t} - \Delta^0 f\right|^p \mathrm{d}\mathfrak{m} = \int_Y \left|\frac{1}{t}\int_0^t P_r^0 \Delta^0 f \,\mathrm{d}r - \Delta^0 f\right|^p \mathrm{d}\mathfrak{m} \longrightarrow 0.$$

For $p = \infty$ we have weak-$*$-convergence in L^∞ in (2.4.2), so the proof works analogously. □

Thanks to the analyticity of the heat semigroups, we get continuity and other properties in the corresponding Sobolev spaces. Recall that for analytic semigroups one has $P_t^0(L^2(Y,\mathfrak{m}|_Y)) \subset D((\Delta^0)^m)$ for every $t > 0$ and $m \in \mathbb{N}$.

Lemma 2.4.2. *Let $f \in L^2(Y,\mathfrak{m}|_Y)$, $t > 0$. Then*

i) $P_{t+\varepsilon}^0 f \to P_t^0 f$ *in* $D(\mathcal{E}^0)$ *as* $\varepsilon \to 0$. *The same is true for* $t = 0$ *if* $f \in D(\mathcal{E}^0)$.

ii) $\frac{P_{t+\varepsilon}^0 f - P_t^0 f}{\varepsilon} \to \Delta^0 P_t^0 f$ *in* $D(\mathcal{E}^0)$ *as* $\varepsilon \to 0$. *The same is true for* $t = 0$ *if* $f \in D(\Delta^0)$ *with* $\Delta^0 f \in D(\mathcal{E}^0)$.

iii) If $f_n \to f$ *in* $L^2(Y,\mathfrak{m}|_Y)$, *then* $P_t^0 f_n \to P_t^0 f$ *in* $D(\mathcal{E}^0)$.

Proof. i) Let us check that the weak gradients of the semigroup are strongly continuous. For $t > 0$ the proof is easy because we can use integration by parts:

$$\begin{aligned}
\| |\nabla(P_{t+\varepsilon}^0 f - P_t^0 f)| \|_2^2 &= -\int_Y (P_{t+\varepsilon}^0 f - P_t^0 f)\Delta^0(P_{t+\varepsilon}^0 f - P_t^0 f)\,\mathrm{d}\mathfrak{m} \\
&= -\int_Y (P_{t+\varepsilon}^0 f - P_t^0 f)(P_\varepsilon^0 \Delta^0 P_t^0 f - \Delta^0 P_t^0 f)\,\mathrm{d}\mathfrak{m} \\
&\longrightarrow 0
\end{aligned}$$

by the strong continuity of P_t^0.

Now for $t = 0$ we use the proof of [FOT94, Lemma 1.3.3] which makes use of the spectral representation of the Laplacian. For completeness, we redo the argument here. Let E_λ^0 be the spectral family (also known as resolution of identity) associated to $-\Delta^0$, i.e. $(E_\lambda^0)_{\lambda\in(-\infty,\infty)}$ is a family of projection operators on $L^2(Y,\mathfrak{m}|_Y)$ which satisfy $E_\lambda^0 E_\mu^0 = E_{\min\{\lambda,\mu\}}^0$, $\lim_{\lambda'\searrow\lambda} E_{\lambda'}^0 f = E_\lambda^0 f$, $\lim_{\lambda\to-\infty} E_\lambda^0 f = 0$, $\lim_{\lambda\to\infty} E_\lambda^0 f = f$ for every $f \in L^2(Y,\mathfrak{m}|_Y)$. Since $-\Delta^0$ is non-negative definite, we have $E_\lambda^0 = 0$ for $\lambda < 0$. For $f, g \in L^2(Y,\mathfrak{m}|_Y)$, $\langle E_\lambda^0 f, g\rangle_{L^2}$ is of bounded variation in λ, so it makes sense to consider Riemann-Stieltjes integrals of the form

$$\langle \zeta(-\Delta^0) f, g\rangle_{L^2} := \int_0^\infty \zeta(\lambda)\,\mathrm{d}\langle E_\lambda^0 f, g\rangle_{L^2}$$

for continuous functions $\zeta : [0,\infty) \to \mathbb{R}$. This defines a self-adjoint operator $\zeta(-\Delta^0)$ with domain

$$D(\zeta(-\Delta^0)) := \left\{ f \in L^2(X,\mathfrak{m}) \,\middle|\, \int_0^\infty \zeta(\lambda)^2\,\mathrm{d}\langle E_\lambda^0 f, f\rangle_{L^2} < \infty \right\}.$$

For short, we will write $\zeta(-\Delta^0) = \int \zeta(\lambda)\,\mathrm{d}E_\lambda^0$. Note that $-\Delta^0 = \int \lambda\,\mathrm{d}E_\lambda^0$, $P_t^0 = e^{t\Delta^0} = \int e^{-t\lambda}\,\mathrm{d}E_\lambda^0$ and $\mathcal{E}^0(f) = \int \lambda\,\mathrm{d}\langle E_\lambda^0 f, f\rangle_{L^2} = \|(-\Delta^0)^{1/2} u\|_{L^2}^2$ for $f \in D(\mathcal{E}^0) = D((-\Delta^0)^{1/2})$. This spectral calculus for instance allows us to prove that $P_\varepsilon^0 f \to f$ in $D(\mathcal{E}^0)$ for $f \in D(\mathcal{E}^0)$:

$$\mathcal{E}^0(P_\varepsilon^0 f - f) = \int_0^\infty \lambda(e^{-\varepsilon\lambda} - 1)^2\,\mathrm{d}\langle E_\lambda^0 f, f\rangle_{L^2} \longrightarrow 0 \text{ as } \varepsilon \to 0.$$

ii) For $t > 0$, with the same calculation as above,

$$\begin{aligned}
&\left\| \left| \nabla \left(\frac{P^0_{t+\varepsilon} f - P^0_t f}{\varepsilon} - \Delta^0 P^0_t f \right) \right| \right\|_2^2 \\
&\quad = - \int_Y \left(\frac{P^0_{t+\varepsilon} f - P^0_t f}{\varepsilon} - \Delta^0 P^0_t f \right) \left(\frac{P^0_\varepsilon \Delta^0 P^0_t f - \Delta^0 P^0_t f}{\varepsilon} - \Delta^0 P^0_t f \right) \mathrm{d}\mathfrak{m} \\
&\quad \longrightarrow 0,
\end{aligned}$$

as both factors converge to 0 strongly.
For $t = 0$ we again use the spectral calculus:

$$\mathcal{E}^0 \left(\frac{P^0_\varepsilon f - f}{\varepsilon} - \Delta^0 f \right) = \int_0^\infty \lambda \left(\frac{e^{-\varepsilon\lambda} - 1}{\varepsilon} + \lambda \right)^2 \mathrm{d}\langle E^0_\lambda f, f \rangle_{L^2} \longrightarrow 0 \quad \text{as } \varepsilon \to 0.$$

iii) Thanks to the analyticity of the semigroup and the closed graph theorem, $\Delta^0 P^0_t \colon L^2 \to L^2$ is a bounded operator. Hence, for $f_n \to f$ in L^2 we also get $\Delta^0 P^0_t f_n \to \Delta^0 P^0_t f$ in L^2 and thus

$$\| \, |\nabla (P^0_t f_n - P^0_t f)| \, \|^2_{L^2} = - \int_Y (P^0_t f_n - P^0_t f) \Delta^0 P^0_t (f_n - f) \longrightarrow 0.$$

□

Without extra assumptions it is difficult to obtain the corresponding results in the L^p-Sobolev spaces, by which we mean the following spaces: For $p \in [1, \infty]$ we set

$$D_p(\mathcal{E}^0) := \{ f \in D(\mathcal{E}^0) \cap L^p(Y, \mathfrak{m}|_Y) \,\big|\, |\nabla f| \in L^p(Y, \mathfrak{m}|_Y) \}, \tag{2.4.3}$$

$$D_p(\Delta^0) := \{ f \in D(\Delta^0) \cap L^p(Y, \mathfrak{m}|_Y) \,\big|\, \Delta^0 f \in L^p(Y, \mathfrak{m}|_Y) \}. \tag{2.4.4}$$

Corollary 2.4.3. *Assume that $\mathfrak{m}(Y) < \infty$ and let $p \in [1, 2]$. Then for $f \in L^2 \cap L^p(Y, \mathfrak{m}|_Y)$ and $t > 0$:*

i) If $f \in D_p(\mathcal{E}^0)$, then $P^0_t f \in D_p(\mathcal{E}^0)$.

ii) $P^0_{t+\varepsilon} f \to P^0_t f$ with respect to $\|f\|_{W^{1,p}} := \|f\|_{L^p} + \| \, |\nabla f| \, \|_{L^p}$ as $\varepsilon \to 0$. The same is true for $t = 0$ if $f \in D_p(\mathcal{E}^0)$.

iii) $\frac{P^0_{t+\varepsilon} f - P^0_t f}{\varepsilon} \to \Delta^0 P^0_t f$ with respect to $\| \cdot \|_{W^{1,p}}$ as $\varepsilon \to 0$. The same is true for $t = 0$ if $f \in D_p(\Delta^0)$.

Proof. It is all based on the fact that the finiteness of $\mathfrak{m}$ implies that $L^2(Y, \mathfrak{m}|_Y) \subset L^p(Y, \mathfrak{m}|_Y)$ for every $p \in [1, 2]$ by Hölder's inequality:

$$\forall f \in L^2(Y, \mathfrak{m}|_Y) : \quad \|f\|^p_{L^p} \leq \mathfrak{m}(Y)^{\frac{2-p}{2}} \|f\|^p_{L^2}.$$

i) Thanks to the above, we actually have $D_p(\mathcal{E}^0) = D(\mathcal{E}^0)$. Since the analyticity of the semigroup yields $P^0_t f \in D(\mathcal{E}^0)$, we're done.

ii) By the Hölder inequality we have $\|P^0_{t+\varepsilon} f - P^0_t f\|_{W^{1,p}} \leq C \|P^0_{t+\varepsilon} f - P^0_t f\|_{W^{1,2}}$, so that the assertion follows by Lemma 2.4.2i).

iii) Same as in ii). □

Let us now introduce the semigroup mollification.

Lemma 2.4.4. *Let $\eta \in C_c^\infty(0,\infty)$ be a non-negative function such that $\int_0^\infty \eta(r)\,\mathrm{d}r = 1$. Given $f \in L^p(Y,\mathfrak{m}|_Y), p \in [1,\infty]$, define for $\varepsilon > 0$*

$$\mathfrak{h}^0_\varepsilon f := \frac{1}{\varepsilon}\int_0^\infty \eta\left(\frac{r}{\varepsilon}\right) P^0_r f\,\mathrm{d}r.$$

Then:

i) $\mathfrak{h}^0_\varepsilon f \in L^p(Y,\mathfrak{m}|_Y)$ and $\mathfrak{h}^0_\varepsilon f \to f$ in $L^p(Y,\mathfrak{m}|_Y)$ as $\varepsilon \to 0$.

ii) If $f_k \in L^p(Y,\mathfrak{m}|_Y)$ with $f_k \to f$ in $L^p(Y,\mathfrak{m}|_Y)$, then $\mathfrak{h}^0_\varepsilon f_k \to \mathfrak{h}^0_\varepsilon f$ in $L^p(Y,\mathfrak{m}|_Y)$.

iii) If $f \in L^2 \cap L^p(Y,\mathfrak{m}|_Y)$, then $\mathfrak{h}^0_\varepsilon f \in D(\Delta^0) \cap L^p(Y,\mathfrak{m}|_Y)$, $\Delta^0\mathfrak{h}^0_\varepsilon f \in D(\Delta^0) \cap L^p(Y,\mathfrak{m}|_Y)$ and

$$\Delta^0\mathfrak{h}^0_\varepsilon f = -\frac{1}{\varepsilon^2}\int_0^\infty \eta'\left(\frac{r}{\varepsilon}\right) P^0_r f\,\mathrm{d}r. \tag{2.4.5}$$

iv) If $f \in D(\Delta^0)$, then $\Delta^0\mathfrak{h}^0_\varepsilon f \to \Delta^0 f$ in $L^2(Y,\mathfrak{m}|_Y)$.

Proof. i) The mollification is in L^p by

$$\left\|\mathfrak{h}^0_\varepsilon f\right\|_{L^p} \le \frac{1}{\varepsilon}\int_0^\infty \eta\left(\frac{r}{\varepsilon}\right)\|P^0_r f\|_{L^p}\,\mathrm{d}r \le \frac{1}{\varepsilon}\int_0^\infty \eta\left(\frac{r}{\varepsilon}\right)\|f\|_{L^p}\,\mathrm{d}r = \|f\|_{L^p}.$$

Similarly, convergence in L^p is obtained by

$$\begin{aligned}\left\|\mathfrak{h}^0_\varepsilon f - f\right\|_{L^p} &\le \frac{1}{\varepsilon}\int_0^\infty \eta\left(\frac{r}{\varepsilon}\right)\|P^0_r f - f\|_{L^p}\,\mathrm{d}r\\ &= \int_0^\infty \eta(s)\,\|P^0_{\varepsilon s} f - f\|_{L^p}\,\mathrm{d}s \longrightarrow 0 \text{ as } \varepsilon \to 0.\end{aligned}$$

ii) Same as i) with using the continuity of $f \mapsto P^0_t f$ instead of $t \mapsto P^0_t f$.

iii) Since the Laplacian is a closed operator, we can interchange it with the Bochner integral [EN00, Prop. C4]. By integration by parts we then have

$$\Delta^0\mathfrak{h}^0_\varepsilon f = \frac{1}{\varepsilon}\int_0^\infty \eta\left(\frac{r}{\varepsilon}\right)\Delta^0 P^0_r f\,\mathrm{d}r = \frac{1}{\varepsilon}\int_0^\infty \eta\left(\frac{r}{\varepsilon}\right)\partial_r P^0_r f\,\mathrm{d}r = -\frac{1}{\varepsilon^2}\int_0^\infty \eta'\left(\frac{r}{\varepsilon}\right) P^0_r f\,\mathrm{d}r.$$

iv) When $f \in D(\Delta^0)$, then $\Delta^0 P^0_t f = P^0_t \Delta^0 f$, so convergence in L^2 follows by i). □

Lemma 2.4.5. *i) Let $f \in D(\mathcal{E}^0)$. Then $\mathfrak{h}^0_\varepsilon f \to f$ in $D(\mathcal{E}^0)$.*

ii) If $f \in D(\Delta^0)$ with $\Delta^0 f \in D(\mathcal{E}^0)$, then $\Delta^0\mathfrak{h}^0_\varepsilon f \to \Delta^0 f$ in $D(\mathcal{E}^0)$.

iii) Let $f_k, f \in L^2(Y,\mathfrak{m}|_Y)$ with $f_k \to f$ in $L^2(Y,\mathfrak{m}|_Y)$. Then $\mathfrak{h}^0_\varepsilon f_k \to \mathfrak{h}^0_\varepsilon f$ in $D(\mathcal{E}^0)$ and $\Delta^0\mathfrak{h}^0_\varepsilon f_k \to \Delta^0\mathfrak{h}^0_\varepsilon f$ in $D(\mathcal{E}^0)$.

Proof. i) We will again make use of the spectral decomposition. This lets us express the semigroup mollification given $f, g \in L^2(Y, \mathfrak{m}|_Y)$ as

$$\langle \mathfrak{h}^0_\varepsilon f, g\rangle_{L^2} = \frac{1}{\varepsilon}\int_0^\infty \eta\left(\frac{r}{\varepsilon}\right)\langle P_r f, g\rangle_{L^2}\,\mathrm{d}r = \frac{1}{\varepsilon}\int_0^\infty\int_0^\infty \eta\left(\frac{r}{\varepsilon}\right) e^{-\lambda r}\,\mathrm{d}\langle E^0_\lambda f, g\rangle_{L^2}\,\mathrm{d}r.$$

Hence we get

$$\begin{aligned}\mathcal{E}^0(\mathfrak{h}^0_\varepsilon u - u) &= \int_0^\infty\int_0^\infty\int_0^\infty \lambda\frac{1}{\varepsilon^2}\eta\left(\frac{r}{\varepsilon}\right)\eta\left(\frac{s}{\varepsilon}\right)(e^{-\lambda r}-1)(e^{-\lambda s}-1)\,\mathrm{d}\langle E_\lambda f, g\rangle_{L^2}\,\mathrm{d}r\,\mathrm{d}s\\ &\longrightarrow 0 \text{ as } \varepsilon \to 0.\end{aligned}$$

ii) It follows directly from i) because for $f \in D(\Delta^0)$ we have $\Delta^0\mathfrak{h}^0_\varepsilon f = \mathfrak{h}^0_\varepsilon\Delta^0 f$.

iii) Using formula (2.4.5), we have

$$\begin{aligned}\mathcal{E}^0(\mathfrak{h}^0_\varepsilon(f_k - f)) &= -\int_Y \mathfrak{h}^0_\varepsilon(f_k - f)\Delta^0\mathfrak{h}^0_\varepsilon(f_k - f)\,\mathrm{d}\mathfrak{m}\\ &= -\frac{1}{\varepsilon^2}\int_0^\infty \eta'\left(\frac{r}{\varepsilon}\right)\int_Y \mathfrak{h}^0_\varepsilon(f_k - f)P_r(f_k - f)\,\mathrm{d}\mathfrak{m}\,\mathrm{d}r\end{aligned}$$

so that by the L^2-convergence of the two factors we have the desired result. The convergence of the Laplacians is shown analogously. □

Corollary 2.4.6. *Assume that $\mathfrak{m}(Y) < \infty$ and $p \in [1, 2]$. Then, if the functions are in $D_p(\mathcal{E}^0)$ or $D_p(\Delta^0)$ respectively, the convergences in Lemma 2.4.5 hold with respect to the norm $\|\cdot\|_{W^{1,p}}$.*

Proof. Follows as in Corollary 2.4.3 directly by Hölder's inequality. □

2.5 Optimal Transport and Curvature-Dimension Condition

Here we recall the basics of optimal transport theory, discuss some useful characterizations of geodesics in the Wasserstein space and introduce synthetic Ricci curvature bounds for metric measure spaces. Apart from the books [Vil03, Vil09], a good introduction can be found in [AG13].

Let (X, d) be a complete, separable metric space. Let $\mathcal{M}(X)$ be the set of finite Radon measures on X (a measure for us will always be non-negative and might take the value $+\infty$), and

$$\begin{aligned}\mathcal{M}^\alpha(X) &:= \left\{\mu \in \mathcal{M}(X) \,\middle|\, \mu(X) = \alpha\right\},\ \alpha \in (0, \infty),\\ \mathcal{P}(X) &:= \mathcal{M}^1(X),\\ \mathcal{P}^{sub}(X) &:= \left\{\mu \in \mathcal{M}(X) \,\middle|\, \mu(X) \le 1\right\}\end{aligned}$$

the sets of measures of mass α, of probability and of subprobability measures, respectively. For $p \in [1, \infty)$, let

$$\mathcal{P}_p(X) := \left\{\mu \in \mathcal{P}(X) \,\middle|\, \exists\, x_0 \in X : \int_X d(x, x_0)^p\,\mathrm{d}\mu(x) < \infty\right\}$$

be the probability measures with finite p^{th} moment. Recall that in a complete, separable space every finite Borel measure is Radon. Let $C_b^0(X)$ be the space of bounded, continuous functions, and $C_c^0(X)$ the space of continuous functions with compact support. A sequence of measures $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{M}(X)$ converges

- *weakly* to $\mu \in \mathcal{M}(X)$, if
$$\int_X f \,\mathrm{d}\mu_n \to \int_X f \,\mathrm{d}\mu \text{ as } n \to \infty \tag{2.5.1}$$
for all $f \in C_b^0(X)$
- *vaguely* to $\mu \in \mathcal{M}(X)$, if (2.5.1) holds for all $f \in C_c^0(X)$.

While for weak convergence also the total masses converge, in vague convergence mass can be lost in the limit. Observe that vague convergence only makes sense in locally compact spaces, since otherwise $C_c^0(X)$ might consist only of the zero function. If $\mu_n \to \mu$ weakly in $\mathcal{M}(X)$ and $Y \subset X$ open, then $\mu_n|_Y \to \mu|_Y$ vaguely in $\mathcal{M}(X)$ since $C_c^0(Y) \subset C_b^0(X)$.

For $\mu \in \mathcal{M}(X)$, a topological space Z, and a μ-measurable map $T\colon X \to Z$, we define the push-forward measure $T_\sharp\mu$ on Z by

$$T_\sharp\mu(A) := \mu(T^{-1}(A)) \quad \text{for every Borel set } A \subset Z.$$

We denote the natural projections of a product space to its factors by $\pi^i\colon X \times X \to X, \pi^i(x_1, x_2) = x_i$.
Though the theory of optimal transport is often presented for probability measures, it is actually the same when using finite measures *of equal mass*. Given two measures $\mu, \nu \in \mathcal{M}^\alpha(X)$, a measure $q \in \mathcal{M}^\alpha(X \times X)$ is a coupling of μ and ν if its marginals are μ and ν, respectively, i.e. if $\pi^1_\sharp q = \mu$, $\pi^2_\sharp q = \nu$. The set of all couplings between μ and ν is denoted by $\mathrm{Cpl}(\mu, \nu)$. Using the direct method of the calculus of variations, one gets the basic existence result for optimal transport problems.

Theorem 2.5.1. *Let $c\colon X \times X \to \mathbb{R} \cup \{+\infty\}$ be lower semicontinuous and bounded from below. Given $\mu, \nu \in \mathcal{M}^\alpha(X)$, the variational minimization problem*

$$\mathcal{C}(\mu, \nu) := \inf \left\{ \int_{X\times X} c(x, y) \,\mathrm{d}q(x, y) \,\middle|\, q \in \mathrm{Cpl}(\mu, \nu) \right\}$$

has a solution. Minimizers are called optimal couplings.

Given two measures μ, ν of the same mass and $\lambda > 0$, we get the scaling property $\mathcal{C}(\lambda\mu, \lambda\nu) = \lambda\mathcal{C}(\mu, \nu)$.
The most important cost functions are powers of the distance. Let $p \in [1, \infty)$ and $\mu, \nu \in \mathcal{M}^\alpha(X)$. Then the *p-Kantorovich-Wasserstein distance* between μ and ν is

$$W_p(\mu, \nu) := \inf \left\{ \int_X d(x, y)^p \,\mathrm{d}q(x, y) \,\middle|\, q \in \mathrm{Cpl}(\mu, \nu) \right\}^{\frac{1}{p}}. \tag{2.5.2}$$

These metrics are usually defined for probability measures only, but we will need them also for measures of other masses. In the case $p = 1$, there is the following "translation invariance": given a further measure ξ, one has

$$W_1(\mu + \xi, \nu + \xi) = W_1(\mu, \nu). \tag{2.5.3}$$

Let us list some of the most important properties.

Theorem 2.5.2. *Let (X, d) be a complete, separable metric space, $p \in [1, \infty)$. Then:*

i) *$(\mathcal{P}_p(X), W_p)$ is a complete and separable metric space, the so-called Wasserstein space.*

ii) *(X, d) is compact if and only $(\mathcal{P}_p(X), W_p)$ is.*

iii) *A sequence $\mu_n \in \mathcal{P}_p(X)$ converges to $\mu_* \in \mathcal{P}_p(X)$ with respect to W_p if and only if*

$$\begin{cases} \mu_n \to \mu_* \text{ weakly,} \\ \displaystyle\int_X d^p(\cdot, x_0)\,\mathrm{d}\mu_n \to \int_X d^p(\cdot, x_0)\,\mathrm{d}\mu_* \text{ for some } x_0 \in X. \end{cases}$$

If $p \neq 1$, then:

iv) *(X, d) is a length space if and only if $(\mathcal{P}_p(X), W_p)$ is.*

v) *(X, d) is a geodesic space if and only if $(\mathcal{P}_p(X), W_p)$ is.*

It is interesting to study curves in the space $\mathcal{P}_p(X)$, especially in the case it is a length or geodesic space. Recall the evaluation maps $\mathrm{e}_t : C^0([0,1], X) \to X,\ \mathrm{e}_t(\gamma) := \gamma_t$ for every $t \in [0, 1]$. Curves of measures can be constructed by taking a measure on the space of curves, and then push-forwarding it by the evaluation maps: Given $Q \in \mathcal{P}(C^0([0, 1], X))$, then $t \mapsto \mu_t := (\mathrm{e}_t)_\# Q$ is a curve in $\mathcal{P}(X)$. A crucial result now says that W_p-geodesics (for $p > 1$) are indeed given that way, and that actually in this case the measure Q is supported on the geodesics $\mathrm{Geo}(X)$. This is a quite useful feature as it allows to work with geodesics on the base space instead of the "abstract" curves of measures.

Proposition 2.5.3 ([AG13, Theorem 2.10], [Vil09, Theorem 7.21, Corollary 7.22])**.** *Let (X, d) be a complete, separable, geodesic metric space, and $p \in (1, \infty)$. Then the following are equivalent:*

(i) *The curve $(\mu_t)_{t\in[0,1]} \subset \mathcal{P}_p(X)$ is a constant-speed geodesic.*

(ii) *There is a measure $Q \in \mathcal{P}_p(\mathrm{Geo}(X))$ such that $(\mathrm{e}_0, \mathrm{e}_1)_\# Q$ is an optimal coupling of μ_0 and μ_1, and $\mu_t = (\mathrm{e}_t)_\# Q$.*

So, roughly speaking, geodesics in the space of measures are measures on the space of geodesics. Let us show some variants which we will need later in this thesis. Like the above theorem, they require to take measurable selections. Since we are going to use it again later, we will cite a useful measurable selection theorem here.

Theorem 2.5.4 ([Bog07, Theorem 6.9.13])**.** *Let $(\Omega, \mathcal{A}, \alpha)$ be a complete probability space, let W be a Souslin space, and let G be a multivalued mapping from Ω to the set of non-empty subsets of W such that its graph belongs to $\mathcal{A} \otimes \mathcal{B}(W)$. Then, there exists an $(\mathcal{A}, \mathcal{B}(W))$-measurable mapping $\Gamma \colon \Omega \to W$ such that $\Gamma(\omega) \in G(\omega)$ for all $\omega \in \Omega$.*

Remark 2.5.5.
a) As can be seen from the proof, it is not necessary to have a probability measure. One could start with a σ-finite measure since this is equivalent to a probability measure.

b) Complete, separable metric spaces are Souslin.

c) To use this theorem, the general strategy is to show that the graph is a closed set, since then it is in particular measurable.

Most of the time it will be used to get a selection of geodesics or almost-geodesics on which curves in the Wasserstein space are supported, so let us treat this situation here.

Lemma 2.5.6.
i) Let (X, d) be a complete, separable, geodesic space. Then there exists a measurable selection $\Gamma \colon X \times X \to C^0([0,1], X)$ such that for every $x, y \in X$ the curve $\Gamma(x, y)$ is a geodesic connecting x and y.

ii) Let (X, d) be a complete, separable, length space and define the set of ε-geodesics between two points as

$$G_\varepsilon(x, y) := \{\gamma \in C^0([0,1], X) \,|\, \gamma_0 = x, \gamma_1 = y, |L(\gamma) - d(x, y)| \le \varepsilon\}.$$

Then there exists a measurable selection $\Gamma_\varepsilon \colon X \times X \to C^0([0,1], X)$ with $\Gamma_\varepsilon(x, y) \in G_\varepsilon(x, y)$ for every $x, y \in X$.

Proof. i) Let $\mathrm{Geo}(x, y) := \{\gamma \in C^0([0,1], X) \,|\, \gamma \text{ geodesic}, \gamma_0 = x, \gamma_1 = y\}$ be the set of geodesics connecting x and y. Since X is a geodesic space, these sets are non-empty. Let us show that the graph of this multivalued map is closed. Let (x_n, y_n, γ^n) be a sequence such that $\gamma^n \in \mathrm{Geo}(x_n, y_n)$, $(x_n, y_n) \to (x, y)$ in $X \times X$ and γ^n converges uniformly to $\gamma^* \in C^0([0,1], X)$. Then $\gamma^*_0 = x, \gamma^*_1 = y$, and

$$d(\gamma^*_s, \gamma^*_t) = \lim_{n\to\infty} d(\gamma^n_s, \gamma^n_t) = \lim_{n\to\infty} |s - t| d(\gamma^n_0, \gamma^n_1) = |s - t| d(\gamma^*_0, \gamma^*_t).$$

Hence $\gamma^* \in \mathrm{Geo}(x, y)$. Now we can apply the above measurable selection theorem with $\Omega = X \times X$, $W = C^0([0,1], X)$ and $G = \mathrm{Geo}$, getting a measurable selection as desired.

ii) Now with G_ε instead, for a sequence (x_n, y_n, γ^n) with $(x_n, y_n) \to (x, y)$ and $\gamma^n \in G_\varepsilon(x_n, y_n)$ uniformly converging to a curve $\gamma^* \in C^0([0,1], X)$, again the end-points converge, and

$$|L(\gamma^*) - d(x, y)| \le \liminf_{n\to\infty} |L(\gamma^n) - d(x_n, y_n)| \le \varepsilon$$

thanks to the lower semicontinuity of the length. Thus, also in this case we get the desired measurable selection. □

The space $(\mathcal{P}_1(X), W_1)$ is always geodesic, since convex combinations between measures are geodesic curves in this case. However, when the underlying space X is a geodesic space, then we can actually choose geodesics in $\mathcal{P}_1(X)$ that are supported on geodesics in X. This even works in length spaces when taking almost-geodesics instead of geodesics.

Proposition 2.5.7. *i) Let (X, d) be a complete, separable, geodesic space. Then, given $\mu_0, \mu_1 \in \mathcal{P}_1(X)$, there is a W_1-geodesic connecting them which is supported on geodesics in X.*

ii) Let X be a complete, separable, length space. Then, given $\varepsilon > 0$ and $\mu_0, \mu_1 \in \mathcal{P}_1(X)$, there exists an ε-W_1-geodesic connecting them which is supported on ε-geodesics in X.

Proof. The proof works exactly as in the above Proposition 2.5.3, so let us follow the one of [AG13, Theorem 2.10].

i) Take an optimal coupling $q \in \mathrm{Cpl}(\mu_0, \mu_1)$, and a measurable selection of geodesics $\Gamma\colon X \times X \to \mathrm{Geo}(X)$ such that $\Gamma(x, y)$ is a geodesic between x and y, which exists by the above Lemma 2.5.6. Then the measure $Q := \Gamma_\# q$ is in $\mathcal{P}_1(\mathrm{Geo}(X))$ and $t \mapsto \mu_t := (\mathrm{e}_t)_\# Q$ is a geodesic in $\mathcal{P}_1(X)$ as can be seen from

$$\begin{aligned} W_1(\mu_s, \mu_t) \leq & \int_{\mathrm{Geo}(X)} d(\mathrm{e}_s(\gamma), \mathrm{e}_t(\gamma)) \,\mathrm{d}Q(\gamma) \\ =& |s-t| \int_{X \times X} d(x, y) \,\mathrm{d}q(x, y) \\ =& |s-t| W_1(\mu_0, \mu_1). \end{aligned}$$

ii) For every $\varepsilon > 0$ and $x, y \in X$ let

$$G_\varepsilon(x, y) := \{\gamma \in C^0([0,1], X) \,\big|\, \gamma_0 = x, \gamma_1 = y, |L(\gamma) - d(x, y)| \leq \varepsilon\}$$

be the set of ε-geodesics connecting x and y. By Lemma 2.5.6 we can take a measurable selection of almost-geodesics, i.e. $\Gamma_\varepsilon\colon X \times X \to C^0([0,1], X)$ such that $\Gamma_\varepsilon(x, y) \in G_\varepsilon(x, y)$. Let $q \in \mathrm{Cpl}(\mu_0, \mu_1)$ be an optimal coupling. The measure $Q_\varepsilon := (\Gamma_\varepsilon)_\# q$ is then supported on ε-geodesics. Let us show that the curve $t \mapsto \mu_t := (\mathrm{e}_t)_\# Q_\varepsilon$ is an almost-geodesic in the space $(\mathcal{P}_1(X), W_1)$:

$$\begin{aligned} W_1(\mu_s, \mu_t) \leq & \int_{C^0([0,1],X)} d(\mathrm{e}_s(\gamma), \mathrm{e}_t(\gamma)) \,\mathrm{d}Q_\varepsilon(\gamma) \\ \leq & \int_{C^0([0,1],X)} |s-t| d(\mathrm{e}_0(\gamma), \mathrm{e}_1(\gamma)) + |s-t|\varepsilon \,\mathrm{d}Q_\varepsilon(\gamma) \\ =& |s-t| \int_{X \times X} d(x, y) \,\mathrm{d}q(x, y) + |s-t|\varepsilon \\ =& |s-t| W_1(\mu_0, \mu_1) + |s-t|\varepsilon. \end{aligned}$$

where we used Lemma 2.1.7. $\square$

Let us now turn to synthetic Ricci curvature bounds. They are defined through the convexity of the relative entropy.

Definition 2.5.8. The relative entropy is the functional $\mathrm{Ent}_{\mathfrak{m}}\colon \mathcal{M}(X) \to (-\infty, \infty]$,

$$\mathrm{Ent}_{\mathfrak{m}}(\mu) := \begin{cases} \displaystyle\int_X \rho \log \rho \,\mathrm{d}\mathfrak{m}, & \text{if } \mu = \rho\mathfrak{m} \\ +\infty, & \text{otherwise.} \end{cases}$$

The domain of the entropy is denoted by $D(\mathrm{Ent}) := \{\mu \in \mathcal{M}(X) \mid \mathrm{Ent}_{\mathfrak{m}}(\mu) < \infty\}$.

By abuse of notation we will sometimes write $\mathrm{Ent}_{\mathfrak{m}}(f)$ instead of $\mathrm{Ent}_{\mathfrak{m}}(f\mathfrak{m})$ for a probability density f.

Definition 2.5.9. Let $K \in \mathbb{R}$. A geodesic metric measure space $(X, d, \mathfrak{m})$ has *Ricci curvature bounded below by* K (we also say: *is a* $\mathrm{CD}(K, \infty)$ *space*) if the relative entropy is K-convex in the Wasserstein space $(\mathcal{P}_2(X), W_2)$, i.e. if for every pair $\mu, \nu \in D(\mathrm{Ent}_{\mathfrak{m}}) \cap \mathcal{P}_2(X)$ there is a constant-speed geodesic $(\mu_t)_{t\in[0,1]} \subset D(\mathrm{Ent}_{\mathfrak{m}}) \cap \mathcal{P}_2(X)$ with $\mu_0 = \mu, \mu_1 = \nu$ such that for all $t \in [0, 1]$:

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_t) \le (1-t)\,\mathrm{Ent}_{\mathfrak{m}}(\mu) + t\,\mathrm{Ent}_{\mathfrak{m}}(\nu) - \frac{K}{2} t(1-t) W_2(\mu, \nu)^2. \tag{2.5.4}$$

If (2.5.4) holds for *every* geodesic in $D(\mathrm{Ent}_{\mathfrak{m}}) \cap \mathcal{P}_2(X)$, then $(X, d, \mathfrak{m})$ is called a *strong* $\mathrm{CD}(K, \infty)$ space.
If $(X, d, \mathfrak{m})$ is a $\mathrm{CD}(K, \infty)$ space and infinitesimally Hilbertian, then we call it an $\mathrm{RCD}(K, \infty)$ space.

There are many important geometric and analytic consequences coming from the $\mathrm{RCD}(K, \infty)$ condition, which under some additional technical assumptions are even equivalent to it.

Theorem 2.5.10. *Let $(X, d, \mathfrak{m})$ be an $\mathrm{RCD}(K, \infty)$ space. Then:*

i) The heat flow $\mathscr{P}_t$ coincides with the EVI_K flow of the entropy in the Wasserstein space, i.e. $\mathscr{P}_t\mu$ for $\mu \in \mathcal{P}_2(X)$ satisfies that for every $\alpha \in \mathcal{P}_2(X)$ and almost every $t \in (0, \infty)$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2} W_2^2(\mathscr{P}_t\mu, \alpha) \le \mathrm{Ent}_{\mathfrak{m}}(\alpha) - \mathrm{Ent}_{\mathfrak{m}}(\mathscr{P}_t\mu) - \frac{K}{2} W_2^2(\mathscr{P}_t\mu, \alpha).$$

ii) The heat flow satisfies the Wasserstein contraction result: for every $t \ge 0$ and $\mu, \nu \in \mathcal{P}_2(X)$

$$W_2(\mathscr{P}_t\mu, \mathscr{P}_t\nu) \le e^{-Kt} W_2(\mu, \nu).$$

iii) The heat flow for functions satisfies the gradient estimate: for every $t > 0$ and $f \in W^{1,2}(X, d, \mathfrak{m})$

$$|\nabla P_t f|^2 \le e^{-2Kt} P_t\left(|\nabla f|^2\right).$$

iv) The Bochner inequality holds: for every $f \in D(\Delta)$ with $\Delta f \in W^{1,2}(X, d, \mathfrak{m})$ and all $\varphi \in D_\infty(\Delta)$ with $\varphi \ge 0$

$$\frac{1}{2}\int_X \Delta\varphi |\nabla f|^2 \,\mathrm{d}\mathfrak{m} - \int_X \varphi \nabla f \cdot \nabla \Delta f \,\mathrm{d}\mathfrak{m} \ge K \int_X \varphi |\nabla f|^2 \,\mathrm{d}\mathfrak{m}.$$

Chapter 3

Gluing of Metric Measure Spaces

3.1 Gluing

Let $(X, d, \mathfrak{m})$ be an infinitesimally Hilbertian metric measure space, take an open subset $Y \subset X$ and denote $Z := X \setminus Y$. We now consider $k \in \mathbb{N}$ copies of X, denoted by $X^1, \dots, X^k$ and identify these spaces with the original one via maps $\iota_i : X \to X^i, i = 1, \dots, k$, which send points $x \in X$ to the corresponding points in X^i. Each X^i is equipped with the metric $d_i := d \circ (\iota_i^{-1}, \iota_i^{-1})$ and the measure $\mathfrak{m}^i := \iota_{i\#}\mathfrak{m}$, but we will usually suppress the indices and write d and $\mathfrak{m}$ on every X^i. Let $Y^i := \iota_i(Y)$, $Z^i := \iota_i(Z)$. We define an equivalence relation by identifying the points in the Z^i's:

$$X^i \ni x \sim y \in X^j \quad :\Leftrightarrow \quad (i = j \text{ and } x = y) \text{ or } \left(\iota_i^{-1}(x) \in Z \text{ and } \iota_i^{-1}(x) = \iota_j^{-1}(y)\right).$$

Definition 3.1.1. Given a metric measure space $(X, d, \mathfrak{m})$, the *k-gluing of X along Z* is now obtained as the quotient of the disjoint union of the X^i under this equivalence relation

$$\hat{X} := \left(\bigsqcup_{i=1}^{k} X^i\right) / \sim .$$

Define a metric $\hat{d} : \hat{X} \times \hat{X} \to \mathbb{R}$ by

$$\hat{d}(x, y) := \begin{cases} \inf_{p \in Z} \left(d_i(x, \iota_i(p)) + d_j(\iota_j(p), y)\right), & \text{if } x \in X^i, y \in X^j, i \neq j \\ d_i(x, y), & \text{if } x, y \in X^i. \end{cases}$$

For points in $p \in Z$ we will subsequently drop the isometries ι_i and for instance write $d_i(x, p)$ instead of $d_i(x, \iota_i(p))$.
As a measure we use $\hat{\mathfrak{m}} := \frac{1}{k}\sum_{i=1}^{k} \mathfrak{m}^i$, meaning that for a Borel set $A \subset \hat{X}$, we consider the restrictions to the copies and set

$$\hat{\mathfrak{m}}(A) := \frac{1}{k}\sum_{i=1}^{k} \mathfrak{m}^i(A \cap X^i).$$

For the special case of gluing together only two copies, we also call the resulting space the *doubling of Y* in X, and as indices we will use $i \in \{+, -\}$.

Remark 3.1.2. a) We can view X^i as a subset of $\hat{X}$, since the canonical map $\sqcup_i X^i \to \hat{X}$ restricted to X^i is injective.

b) In the following, we will also make use of the partition

$$\hat{X} = \left(\bigsqcup_{i=1}^{k} Y^i \right) \sqcup Z.$$

c) As we are gluing together copies of *the same* space, we have that for $x, y \in X^i$ and $j \neq i$

$$d_i(x, y) = d_j(\iota_j(\iota_i^{-1}(x)), \iota_j(\iota_i^{-1}(y))).$$

Proposition 3.1.3. *Let (X, d) be a complete, separable metric space. Then:*

i) $(\hat{X}, \hat{d})$ is a complete, separable metric space.

ii) If X is a geodesic space and $Z = X \setminus Y$ is proper, then $\hat{X}$ is a geodesic space.

iii) If X is a length space, then $\hat{X}$ is a length space.

Proof. i) The construction is classical and can for instance be found in [BH99, p.67f, Lemma 5.24] and [BBI01, Chapter 3]. For sake of completeness, we redo the proof here.

The function $\hat{d}$ is obviously non-negative, has a vanishing diagonal and is symmetric.

For the triangle inequality, let us start with the case that $x, y \in X^i$. Then if $z \in X^i$, we have

$$\hat{d}(x, y) = d_i(x, y) \leq d_i(x, z) + d_i(z, y) = \hat{d}(x, z) + \hat{d}(z, y).$$

If on the other hand $z \in X^j, j \neq i$, let $\varepsilon > 0$ and take $p, q \in Z$ such that

$$\hat{d}(x, z) \geq d_i(x, p) + d_j(p, z) - \varepsilon \quad \text{and} \quad \hat{d}(x, z) \geq d_i(q, y) + d_j(q, z) - \varepsilon.$$

Then, since $d_j(p, z) = d_i(p, \iota_i(\iota_j^{-1}(z)))$,

$$\begin{aligned}
\hat{d}(x, z) + \hat{d}(z, y) &\geq d_i(x, p) + d_j(p, z) + d_i(q, y) + d_j(q, z) - 2\varepsilon \\
&= d_i(x, p) + d_i(p, \iota_i(\iota_j^{-1}(z))) + d_i(q, y) + d_i(q, \iota_i(\iota_j^{-1}(z))) - 2\varepsilon \\
&\geq d_i(x, \iota_i(\iota_j^{-1}(z))) + d_i(\iota_i(\iota_j^{-1}(z)), y) - 2\varepsilon \\
&\geq d_i(x, y) - 2\varepsilon \\
&= \hat{d}(x, y) - 2\varepsilon.
\end{aligned}$$

The other cases are similar.

Let $x, y \in \hat{X}$ with $\hat{d}(x, y) = 0$. In case $x, y \in X^i$, then $x = y$ by the definiteness of d. Assume for a contradiction that $x \in Y^i, y \in Y^j, i \neq j$. Then for every $n \in \mathbb{N}$ there is $z_n \in Z$ such that $d_i(x, z_n) + d_j(z_n, y) < \frac{1}{n}$. In particular z_n converges both, to x in (X^i, d_i) and to y in (X^j, d_j). Since the copies Y^i, Y^j are open and disjoint in $\hat{X}$, this is a contradiction.

Separability is clear by taking the union of the separable sets of the different copies.

Turning to completeness, let $(x_n)_{n\in\mathbb{N}} \subset \hat{X}$ be a Cauchy sequence. Since we are gluing together only finite number of copies, there is $i \in \{1,\dots,k\}$ such that there is an infinite subsequence with $(x_{n_\ell})_{\ell\in\mathbb{N}} \subset X^i$. This is a Cauchy sequence with respect to d_i, hence there is a limit $x^* \in X^i$ which is also a limit in $(\hat{X},\hat{d})$.

ii) Let $x,y \in \hat{X}$. Then we have to find a midpoint. If x and y are in the same copy, then we can just take the midpoint we get from that copy being a geodesic space. So assume that $x \in Y^i, y \in Y^j, i \neq j$. Then there are $z_m \in Z$ such that $d_i(x,z_m) + d_j(z_m,y) \to \hat{d}(x,y)$. Hence $(z_m)_{m\in\mathbb{N}}$ is a bounded sequence in the proper space Z, so we can extract a converging subsequence $z_{m_\ell} \to z^*$ in d_i. But since $\iota_j(\iota_i^{-1}(z)) = z$ for $z \in Z$, we also get the convergence $d_j(z_{m_\ell},y) \to d_j(z^*,y)$. Thus z^* is a minimizer in the definition of $\hat{d}(x,y)$, i.e. $\hat{d}(x,y) = d_i(x,z^*) + d_j(z^*,y)$. Assume without loss of generality that $d_i(x,z^*) \geq d_j(z^*,y)$. Then we can take a geodesic $\gamma\colon [0,1] \to X^i, \gamma_0 = x, \gamma_1 = z^*$ and a time $t^* \in [0,1]$ such that

$$\hat{d}(x,\gamma_{t^*}) = d_i(x,\gamma_{t^*}) = \frac{1}{2}\hat{d}(x,y) = d_i(\gamma_{t^*},z^*) + d_j(z^*,y).$$

By a simple contradiction argument one sees that $d_i(\gamma_{t^*},z^*) + d_j(z^*,y) = \hat{d}(\gamma_{t^*},y)$, meaning that we have found a midpoint.

iii) Let $x,y \in \hat{X}$ and $\varepsilon > 0$. Now we have to find an ε-midpoint. If x and y are in the same copy, then we can just take the almost-midpoint we get from that copy being a length space. So assume that $x \in Y^i, y \in Y^j, i \neq j$. Then there is $z \in Z$ such that

$$d_i(x,z) + d_j(z,y) \leq \hat{d}(x,y) + \varepsilon.$$

Assume without loss of generality that $d_i(x,z) \geq d_j(z,y)$. Take a dyadic number $q \in (0,1)$ such that

$$\left|\frac{1}{2}\hat{d}(x,y) - qd_i(x,z)\right| \leq \frac{\varepsilon}{2} \quad \text{and} \quad \left|\frac{1}{2}\hat{d}(x,y) - [(1-q)d_i(x,z) + d_j(z,y)]\right| \leq \frac{\varepsilon}{2}.$$

By taking "midpoints of midpoints", the length property of X^i then provides us with a point $v \in X^i$ such that

$$|d_i(x,v) - qd_i(x,z)| \leq \frac{\varepsilon}{2} \quad \text{and} \quad |d_i(v,z) - (1-q)d_i(x,z)| \leq \frac{\varepsilon}{2}.$$

Finally – having in mind that $\hat{d}(x,v) = d_i(x,v)$ for $x,v \in X^i$ – this yields

$$\left|\hat{d}(x,v) - \frac{1}{2}\hat{d}(x,y)\right| \leq |d_i(x,v) - qd_i(x,z)| + \frac{\varepsilon}{2} \leq \varepsilon$$

and

$$\left|\hat{d}(v,z) - \frac{1}{2}\hat{d}(x,y)\right| \leq |d_i(v,z) + d_j(z,y) - qd_i(x,z)| + \frac{\varepsilon}{2} \leq \varepsilon.$$

□

The same counterexample as in Remark 2.2.6 shows that we cannot omit the assumption that $X \setminus Y$ is proper to conclude that the glued space is geodesic.

The metric properties directly transfer to the Wasserstein space, see for instance [Vil09].

Corollary 3.1.4. *For $p \in [1, \infty)$, the Kantorovich-Wasserstein metric $\hat{W}_p$ obtained from $\hat{d}$ is a complete, separable metric on $\mathcal{P}_p(\hat{X})$. It is a length (resp. geodesic) metric, if and only if $\hat{d}$ is.*

Lemma 3.1.5. *Given an open set $A \subset \hat{X}$, its restriction to a copy $A \cap X^i, i \in \{0, \dots, k\}$, is open in (X^i, d_i). Analogously, for a closed set $C \subset \hat{X}$, the restriction $C \cap X^i$ is closed in (X^i, d_i). In particular, for a Borel set $A \subset \hat{X}$, the restriction $A \cap X^i$ is a Borel set in (X^i, d_i).*

Proof. Let $x \in A \cap X^i$. Since A is open in $(\hat{X}, \hat{d})$, there is $\varepsilon > 0$ such that the ε-ball with respect to $\hat{d}$, $\hat{B}_\varepsilon(x)$, is contained in A. Since $\hat{d} = d_i$ on $X^i \times X^i$, we get

$$B_\varepsilon(x) = B_\varepsilon(x) \cap X^i \subset \hat{B}_\varepsilon(x) \cap X^i \subset A \cap X^i,$$

where $B_\varepsilon(x)$ is the ε-ball in (X^i, d_i) around x.
For closed sets just take complements of open sets. □

Corollary 3.1.6. *If X is compact, then also $\hat{X}$ is compact.*

Proof. Let $\{\hat{U}_\ell\}_{\ell \in \mathbb{N}}$ be an open cover of $\hat{X}$. By the previous lemma, $U^i_\ell := \hat{U}_\ell \cap X^i$ is an open cover for X^i. Hence, compactness of X gives us a finite subcover $\{U^i_{\ell_m}\}_m$ of X^i. Then $\{U^i_{\ell_m}\}_{m,i}$ is a finite subcover for $\hat{X}$. □

Corollary 3.1.7. *Let $(X, d, \mathfrak{m})$ be a metric measure space. Then $(\hat{X}, \hat{d}, \hat{\mathfrak{m}})$ is a metric measure space, i.e. the measure $\hat{\mathfrak{m}}$ is a well-defined Borel measure on $\hat{X}$ satisfying the exponential integrability condition* (2.3.1). *In case $\mathfrak{m}$ is a finite measure, $\hat{\mathfrak{m}}$ is also finite, and hence a Radon measure.*

Proof. The above lemma shows that $\hat{\mathfrak{m}}$ is a well-defined Borel measure, so we only need to show the integrability condition. Let $c > 0$ and $x_0 \in X$ be such that (2.3.1) holds for $\mathfrak{m}$. Note that once (2.3.1) holds for one $x_0 \in X$, thanks to the triangle inequality it holds for every other choice of x_0, so we can assume without loss of generality that $x_0 \in Z$. Then

$$\int_{\hat{X}} e^{-c\hat{d}^2(x,x_0)} \, \mathrm{d}\hat{\mathfrak{m}} = \frac{1}{k} \sum_{i=1}^{k} \int_{X^i} e^{-c d_i^2(x,x_0)} \, \mathrm{d}\mathfrak{m}_i < \infty.$$

□

Now we introduce some notation for dealing with functions on $\hat{X}$.

Definition 3.1.8. Let $u_i : X^i \to \mathbb{R}$, $i = 1, \dots, k$, be given by $u_i := u|_{X^i}$. Define the mean value

$$\bar{u}\colon X \to \mathbb{R}, \quad \bar{u} := \frac{1}{k}\sum_{i=1}^{k} u_i \circ \iota_i$$

and the "mean free" functions

$$\mathring{u}_i\colon X \to \mathbb{R}, \quad \mathring{u}_i := u_i \circ \iota_i - \bar{u}.$$

Observe that since the u_i all coincide on Z, the $\mathring{u}_i$ are zero *everywhere* on Z. Also, we have

$$\sum_{i=1}^{k} \mathring{u}_i = 0. \tag{3.1.1}$$

Notation: During the proof of Lemma 3.1.13 we will start to simplify notation, by mostly omitting the identification maps ι_i. Whenever a function u_i now gets an argument from X, it is understood as $u_i \circ \iota_i$ and similar for $\overline{u}, \mathring{u}_i$ with ι_i^{-1}.

Let $(\widehat{\mathrm{Ch}}, \hat{\mathcal{F}})$ denote the Cheeger energy of the space $(\hat{X}, \hat{d}, \hat{\mathfrak{m}})$.

Lemma 3.1.9. *The space $\hat{X}$ is infinitesimally Hilbertian and for every $u \in \hat{\mathcal{F}}$, the functions $u_i \circ \iota_i$ are in $\mathcal{F}$ and*

$$\widehat{\mathrm{Ch}}(u) = \frac{1}{k}\sum_{i=1}^{k} \mathrm{Ch}(u_i \circ \iota_i).$$

Proof. This follows directly from the locality property of weak gradients in Lemma 2.3.5 by applying it to the open sets Y^i and Z°. $\square$

In particular, we get a Dirichlet form $(\hat{\mathcal{E}}, D(\hat{\mathcal{E}}))$ on $\hat{X}$ by polarizing $\hat{\mathcal{E}}(u) := 2\widehat{\mathrm{Ch}}(u)$ and setting $D(\hat{\mathcal{E}}) := \hat{\mathcal{F}}$. This Dirichlet form has an associated strongly continuous, contraction semigroup which we will denote by $\hat{P}_t$. The dual heat flow on the space of measures will be denoted by $\hat{\mathscr{P}}_t$.

Lemma 3.1.10. *If $u \in D(\hat{\mathcal{E}})$, then $\bar{u} \in D(\mathcal{E})$ and $\mathring{u}_i \in D(\mathcal{E}^0)$, $i = 1, \dots, n$.*

Proof. Being in $D(\hat{\mathcal{E}})$ means $\widehat{\mathrm{Ch}}(u) < \infty$. By the previous lemma, this implies

$$\sum_{i=1}^{k} \frac{1}{k}\,\mathrm{Ch}(u_i \circ \iota_i) = \widehat{\mathrm{Ch}}(u) < \infty.$$

Since each term is non-negative, $\mathrm{Ch}(u_i \circ \iota_i) < \infty$ for every $i = 1, \dots, k$. Thus $u_i \circ \iota_i \in D(\mathcal{E})$ and also the linear combination $\bar{u} \in D(\mathcal{E})$.

The other assertion follows from the fact that all the u_i's coincide on Z. $\square$

Now we are going to define a semigroup on $\hat{X}$ and we will show that it actually is the one corresponding to $\hat{\mathcal{E}}$.

Definition 3.1.11. The *glued semigroup* $P_t^{GL} : L^2(\hat{X}, \hat{\mathfrak{m}}) \to L^2(\hat{X}, \hat{\mathfrak{m}})$ is defined by

$$P_t^{GL} u(x) := P_t \bar{u}(\iota_i^{-1}(x)) + P_t^0 \mathring{u}_i(\iota_i^{-1}(x)), \quad \text{if } x \in X^i, \ i = 1, \dots, k. \tag{3.1.2}$$

Also, define the *approximated glued Dirichlet form* $\mathcal{E}_t^{GL} : L^2(\hat{X}, \hat{\mathfrak{m}}) \times L^2(\hat{X}, \hat{\mathfrak{m}}) \to \mathbb{R}$,

$$\mathcal{E}_t^{GL}(u, v) := -\frac{1}{t} \int_{\hat{X}} v(P_t^{GL} u - u) \, \mathrm{d}\hat{\mathfrak{m}}.$$

Remark 3.1.12. Observe that P_t^{GL} is well-defined, since $u_i = u_j$ on Z for every $i, j = 1, \dots, k$.

Lemma 3.1.13. *The family of operators $(P_t^{GL})_{t>0}$ is a symmetric, strongly continuous contraction semigroup on $L^2(\hat{X}, \hat{\mathfrak{m}})$. In particular, there exists a corresponding Dirichlet form $(\mathcal{E}^{GL}, D(\mathcal{E}^{GL}))$ connected to P_t^{GL} via*

$$\begin{cases} D(\mathcal{E}^{GL}) = \left\{ u \in L^2(\hat{X}, \hat{\mathfrak{m}}) \,\middle|\, \lim_{t \to 0} \mathcal{E}_t^{GL}(u) < \infty \right\} \\ \mathcal{E}^{GL}(u, v) = \lim_{t \to 0} \mathcal{E}_t^{GL}(u, v), \ \textit{for } u, v \in D(\mathcal{E}^{GL}). \end{cases}$$

Proof. Symmetry: We use that P_t and P_t^0 are symmetric with respect to $\mathfrak{m}$:

$$\begin{aligned}
&\int_{\hat{X}} u P_t^{GL} v \, \mathrm{d}\hat{\mathfrak{m}} = \sum_{i=1}^k \frac{1}{k} \int_{X^i} u_i \left((P_t \bar{v}) \circ \iota_i^{-1} + (P_t^0 \mathring{v}_i) \circ \iota_i^{-1} \right) \mathrm{d}\mathfrak{m}^i \\
&= \sum_{i=1}^k \frac{1}{k} \int_X \bar{v} P_t(u_i \circ \iota_i) + \mathring{v}_i P_t^0 (u_i \circ \iota_i) \, \mathrm{d}\mathfrak{m} \\
&= \sum_{i,j=1}^k \frac{1}{k^2} \int_X (v_j \circ \iota_j) P_t(u_i \circ \iota_i) + (v_i \circ \iota_i) P_t^0 (u_i \circ \iota_i) - (v_j \circ \iota_j) P_t^0 (u_i \circ \iota_i) \, \mathrm{d}\mathfrak{m} \\
&= \sum_{i,j=1}^k \frac{1}{k^2} \int_X (v_j \circ \iota_j) P_t(u_i \circ \iota_i) + (v_j \circ \iota_j) P_t^0 (u_j \circ \iota_j) - (v_j \circ \iota_j) P_t^0 (u_i \circ \iota_i) \, \mathrm{d}\mathfrak{m} \\
&= \sum_{j=1}^k \frac{1}{k} \int_X (v_j \circ \iota_j) \frac{1}{k} \sum_{i=1}^k P_t(u_i \circ \iota_i) + (v_j \circ \iota_j) \left(P_t^0 (u_j \circ \iota_j) - \frac{1}{k} \sum_{i=1}^k P_t^0 (u_i \circ \iota_i) \right) \mathrm{d}\mathfrak{m} \\
&= \sum_{j=1}^k \frac{1}{k} \int_X (v_j \circ \iota_j)(P_t \bar{u} + P_t^0 \mathring{u}_j) \, \mathrm{d}\mathfrak{m} = \int_{\hat{X}} v P_t^{GL} u \, \mathrm{d}\hat{\mathfrak{m}}.
\end{aligned}$$

From now on we will apply the abuse of notation introduced before. This is in order to improve readability.

Semigroup property: First observe that on X^i we have $P_0^{GL} u = P_0 \bar{u} + P_0^0 \mathring{u}_i = \bar{u} + u_i - \bar{u} = u$. Denote $v := P_t^{GL} u$. Then $v_i = P_t \bar{u} + P_t^0 \mathring{u}_i$. Now on X^i

$$P_s^{GL} P_t^{GL} u = P_s^{GL} v = P_s \bar{v} + P_s^0 \mathring{v}_i = \frac{1}{k} \sum_{j=1}^k P_s v_j + P_s^0 v_i - \frac{1}{k} \sum_{j=1}^k P_s^0 v_j$$

$$\begin{aligned}
&=\frac{1}{k}\sum_{j=1}^{k} P_s(P_t\bar{u} + P_t^0\mathring{u}_j) + P_s^0(P_t\bar{u} + P_t^0\mathring{u}_i) - \frac{1}{k}\sum_{j=1}^{k} P_s^0(P_t\bar{u} + P_t^0\mathring{u}_j)\\
&=\frac{1}{k}\sum_{j=1}^{k} P_{s+t}\bar{u} + \underbrace{\frac{1}{k}\sum_{j=1}^{k} P_sP_t^0\mathring{u}_j}_{=0} + P_s^0P_t\bar{u} + P_{s+t}^0\mathring{u}_i - \frac{1}{k}\sum_{j=1}^{k} P_s^0P_t\bar{u} - \underbrace{\frac{1}{k}\sum_{j=1}^{k} P_{s+t}^0\mathring{u}_j}_{=0}\\
&=P_{s+t}\bar{u} + P_{s+t}^0\mathring{u}_i = P_{s+t}^{GL}u,
\end{aligned}$$

where we used (3.1.1).

Contraction: To show the contraction property in $L^2(\hat{X},\hat{\mathfrak{m}})$, we first show that P_t^{GL} is Markovian (i.e. positivity preserving and L^∞-contractive in $L^2\cap L^\infty$). By symmetry of P_t^{GL}, we also get L^1-contractivity. Using the Riesz-Thorin interpolation theorem, we finally get contractivity in L^2.

Let $u\in L^2\cap L^\infty(\hat{X},\hat{\mathfrak{m}})$ with $0\le u\le 1$. Then also $0\le u_i,\bar{u}\le 1$. Then, on X^i,

$$P_t^{GL}u = P_t\bar{u} + P_t^0\mathring{u}_i \le P_t\bar{u} + P_t\mathring{u}_i = P_tu_i \le 1.$$

For the other side, we have to show $P_t^{GL}u\ge 0$, which is equivalent to

$$P_t^0\bar{u} \le P_t\bar{u} + P_t^0u_i.$$

But this holds true because $P_t^0f\le P_tf$ for every $f\in L^2$, and $P_t^0u_i\ge 0$.

Now we use that L^1 is a subspace of the dual of L^∞. For $u\in L^1\cap L^2(\hat{X},\hat{\mathfrak{m}})$, consider the bounded, linear functional $\ell : L^\infty(\hat{X},\hat{\mathfrak{m}})\to\mathbb{R}$, $\ell(v) := \int_{\hat{X}} vP_t^{GL}u\,\mathrm{d}\hat{\mathfrak{m}}$. The dual space norm of ℓ coincides with the L^1-norm of $P_t^{GL}u$, thus

$$\begin{aligned}
\|P_t^{GL}u\|_{L^1(\hat{X})} &= \sup_{\|v\|_{L^\infty(\hat{X})}\le 1}\int_{\hat{X}} vP_t^{GL}u\,\mathrm{d}\hat{\mathfrak{m}} = \sup_{\|v\|_{L^\infty(\hat{X})}\le 1}\int_{\hat{X}} P_t^{GL}vu\,\mathrm{d}\hat{\mathfrak{m}}\\
&\le \sup_{\|v\|_{L^\infty(\hat{X})}\le 1}\int_{\hat{X}} vu\,\mathrm{d}\hat{\mathfrak{m}} = \|u\|_{L^1(\hat{X})}.
\end{aligned}$$

Here we used the symmetry of P_t^{GL} and the L^∞-contractivity.

Hence P_t^{GL} is a contraction in $L^1\cap L^2$ and also in $L^\infty\cap L^2$. By the Riesz-Thorin interpolation theorem, it is then also a contraction in L^2.

Strong continuity: This follows directly from the strong continuity of P_t and P_t^0:

$$\begin{aligned}
\|P_t^{GL}u-u\|_{L^2(\hat{X})}^2 &= \int_{\hat{X}}\left(P_t^{GL}u-u\right)^2\mathrm{d}\hat{\mathfrak{m}} = \sum_{i=1}^{k}\frac{1}{k}\int_{X^i}\left(P_t\bar{u}+P_t^0\mathring{u}_i-u_i\right)^2\mathrm{d}\mathfrak{m}^i\\
&=\sum_{i=1}^{k}\frac{1}{k}\int_X\left(P_t\bar{u}-\bar{u}+P_t^0\mathring{u}_i-\mathring{u}_i\right)^2\mathrm{d}\mathfrak{m}\\
&\le\sum_{i=1}^{k}\frac{2}{k}\int_X (P_t\bar{u}-\bar{u})^2+\left(P_t^0\mathring{u}_i-\mathring{u}_i\right)^2\mathrm{d}\mathfrak{m}
\end{aligned}$$

$$= \sum_{i=1}^{k} \frac{2}{k} \left(\|P_t \bar{u} - \bar{u}\|^2_{L^2(X)} + \|P^0_t \mathring{u}_i - \mathring{u}_i\|^2_{L^2(X)} \right) \longrightarrow 0$$

as $t \to 0$. □

Lemma 3.1.14. *For every $u, v \in L^2(\hat{X}, \hat{\mathfrak{m}})$:*

$$\mathcal{E}^{GL}_t(u, v) = \mathcal{E}_t(\bar{u}, \bar{v}) + \frac{1}{k} \sum_{i=1}^{k} \mathcal{E}^0_t(\mathring{u}_i, \mathring{v}_i). \tag{3.1.3}$$

Proof. We just compute

$$\begin{aligned}
\mathcal{E}^{GL}_t(u, v) &= -\frac{1}{t} \int_{\hat{X}} v \left(P^{GL}_t u - u \right) \mathrm{d}\hat{\mathfrak{m}} \\
&= -\sum_{i=1}^{k} \frac{1}{kt} \int_{X^i} v_i \left(P_t \bar{u} + P^0_t \mathring{u}_i - u_i \right) \mathrm{d}\mathfrak{m}^i \\
&= -\sum_{i=1}^{k} \frac{1}{kt} \int_{X} v_i \left(P_t \bar{u} - \bar{u} + P^0_t \mathring{u}_i - \mathring{u}_i \right) \mathrm{d}\mathfrak{m} \\
&= -\frac{1}{t} \int_X \bar{v} \left(P_t \bar{u} - \bar{u} \right) \mathrm{d}\mathfrak{m} - \sum_{i=1}^{k} \frac{1}{k} \int_X v_i \left(P^0_t \mathring{u}_i - \mathring{u}_i \right) \mathrm{d}\mathfrak{m} \\
&\quad + \underbrace{\sum_{i=1}^{k} \frac{1}{k} \int_X \bar{v} \left(P^0_t \mathring{u}_i - \mathring{u}_i \right) \mathrm{d}\mathfrak{m}}_{=0 \text{ by } (3.1.1)} \\
&= \mathcal{E}_t(\bar{u}, \bar{v}) + \frac{1}{k} \sum_{i=1}^{k} \mathcal{E}^0_t(\mathring{u}_i, \mathring{v}_i).
\end{aligned}$$

□

Lemma 3.1.15. *If $u \in D(\mathcal{E}^{GL})$, then $\bar{u} \in D(\mathcal{E})$ and $\mathring{u}_i \in D(\mathcal{E}^0)$, $i = 1, \dots, k$.*

Proof. By definition and (3.1.3),

$$\infty > \mathcal{E}^{GL}(u) = \lim_{t \to 0} \mathcal{E}^{GL}_t(u) = \lim_{t \to 0} \left(\mathcal{E}_t(\bar{u}, \bar{v}) + \frac{1}{k} \sum_{i=1}^{k} \mathcal{E}^0_t(\mathring{u}_i, \mathring{v}_i) \right).$$

Since the sum converges and every term is non-negative and non-decreasing as $t \to 0$, the terms converge and we can interchange sum and limit to get

$$\infty > \mathcal{E}^{GL}(u) = \lim_{t \to 0} \mathcal{E}_t(\bar{u}, \bar{v}) + \frac{1}{k} \sum_{i=1}^{k} \lim_{t \to 0} \mathcal{E}^0_t(\mathring{u}_i, \mathring{v}_i) = \mathcal{E}(\bar{u}, \bar{v}) + \frac{1}{k} \sum_{i=1}^{k} \mathcal{E}^0(\mathring{u}_i, \mathring{v}_i).$$

□

Now we come to the main theorem of this section, which identifies the semigroup P_t^{GL} with the heat semigroup $\hat{P}_t$ associated to $\hat{\mathcal{E}}$.

Theorem 3.1.16. *The semigroups P_t^{GL} and $\hat{P}_t$ coincide on $L^2(\hat{X}, \hat{\mathfrak{m}})$.*

Proof. We will proof that the Dirichlet forms $(\mathcal{E}^{GL}, D(\mathcal{E}^{GL}))$ and $(\hat{\mathcal{E}}, D(\hat{\mathcal{E}}))$ coincide. Let $u, v \in D(\hat{\mathcal{E}})$. By Lemma 3.1.14,

$$\mathcal{E}_t^{GL}(u,v) = \mathcal{E}_t(\bar{u}, \bar{v}) + \frac{1}{k}\sum_{i=1}^{k} \mathcal{E}_t^0(\mathring{u}_i, \mathring{v}_i).$$

By Lemma 3.1.10, $\bar{u}, \bar{v} \in D(\mathcal{E})$ and $\mathring{u}_i, \mathring{v}_i \in D(\mathcal{E}^0)$, so that we can take the limit $t \to 0$. This yields

$$\begin{aligned}
\mathcal{E}^{GL}(u,v) =& \lim_{t\to 0} \mathcal{E}_t^{GL}(u,v) = \lim_{t\to 0}\left(\mathcal{E}_t(\bar{u}, \bar{v}) + \frac{1}{k}\sum_{i=1}^{k} \mathcal{E}_t^0(\mathring{u}_i, \mathring{v}_i)\right) \\
=& \mathcal{E}(\bar{u}, \bar{v}) + \frac{1}{k}\sum_{i=1}^{k} \mathcal{E}^0(\mathring{u}_i, \mathring{v}_i) = \mathcal{E}(\bar{u}, \bar{v}) + \frac{1}{k}\sum_{i=1}^{k} \mathcal{E}(\mathring{u}_i, \mathring{v}_i) \\
=& \mathcal{E}(\bar{u}, \bar{v}) + \frac{1}{k}\sum_{i=1}^{k} \mathcal{E}(u_i - \bar{u}, v_i - \bar{v}) = \frac{1}{k}\sum_{i=1}^{k} \mathcal{E}(u_i, v_i) = \hat{\mathcal{E}}(u,v),
\end{aligned}$$

where we used that $\mathcal{E}$ is an extension of $\mathcal{E}^0$. This also shows that $D(\hat{\mathcal{E}}) \subset D(\mathcal{E}^{GL})$. The other direction works with the same argument but using Lemma 3.1.15 instead. □

3.2 The Case of Riemannian Manifolds

When interested in curvature properties, gluing together Riemannian manifolds is a delicate issue, since in general the glued Riemannian metric will only be continuous and so one cannot define the curvature tensors. Schlichting [Sch12, Sch14] showed that for the Ricci curvature (and various other curvature operators) a lower bound is preserved under gluing in an approximate sense. We will now use this result to show that the doubling is an $\mathrm{RCD}(K, \infty)$ space.

Theorem 3.2.1. *Let (M, g) be a complete, n-dimensional Riemannian manifold with Ricci curvature bounded below by $K \in \mathbb{R}$. Let $Y \subset M$ be an open, bounded, convex subset with a smooth, compact boundary. We set $X := \overline{Y}$ and equip it with the Riemannian distance d and volume measure $\mathfrak{m}$. Then the 2-gluing of $(X, d, \mathfrak{m})$ along ∂Y, denoted by $(\hat{X}, \hat{d}, \hat{\mathfrak{m}})$, is an $\mathrm{RCD}(K, n)$ space.*

Proof. First observe that the gluing of Riemannian manifolds yields a continuous Riemannian metric

$$\hat{g}(p) = \begin{cases} g_+(p), & \text{if } p \in Y^+ \\ g_-(p), & \text{if } p \in Y^-, \end{cases}$$

whose Riemannian distance and volume measure are $d_{\hat{g}} = \hat{d}$ and $\mathfrak{m}_{\hat{g}} = 2\hat{\mathfrak{m}}$ in terms of our metric gluing.

By convexity, the submanifold Y satisfies the same lower bound on the Ricci curvature and the boundary ∂Y has non-negative second fundamental form. The result in [Sch12, Sch14] now ensures that there is a sequence of *smooth* Riemannian metrics $\hat{g}_\varepsilon$ on the glued manifold $\hat{X}$ converging to $\hat{g}$ uniformly as $\varepsilon \to 0$ and such that

$$\mathrm{Ric}_{\hat{g}_\varepsilon} \geq (K - \varepsilon).$$

Thus we get a sequence of smooth, compact metric measure spaces $(\hat{X}, d_{\hat{g}_\varepsilon}, \mathfrak{m}_{\hat{g}_\varepsilon})$ which satisfy the $\mathrm{RCD}(K-\varepsilon, n)$ condition. The stability of the RCD-condition under measured Gromov-Hausdorff convergence together with the convergence result in the following lemma completes the proof. □

Lemma 3.2.2. *Let $(g_\varepsilon)_{\varepsilon>0}$ be a sequence of smooth Riemannian metrics and g a continuous Riemannian metric on a compact, smooth manifold $\mathfrak{M}$. If $g_\varepsilon \to g$ uniformly as $\varepsilon \to 0$, then $(\mathfrak{M}, d_\varepsilon, \mathfrak{m}_\varepsilon) \to (\mathfrak{M}, d, \mathfrak{m})$ in the measured Gromov-Hausdorff sense, where $d_\varepsilon, \mathfrak{m}_\varepsilon$ and $d, \mathfrak{m}$ are the distance functions and volume measures obtained by g_ε and g, respectively.*

Proof. First, we have to show that for every $\delta > 0$ there is $\varepsilon^* = \varepsilon^*(\delta) > 0$ such that for every $\varepsilon < \varepsilon^*$ the identity $\mathrm{id} : (X, d_\varepsilon) \to (X, d)$ is a δ-isometry. Being a δ-isometry in this case means that for every $x, y \in X$ we have $|d_\varepsilon(x, y) - d(x, y)| < \delta$ (i.e. uniform convergence).

By the uniform convergence of g_ε and the uniform continuity of the square root, we have uniform convergence of $\sqrt{g_\varepsilon(v, v)} =: |v|_\varepsilon \to |v| := \sqrt{g(v, v)}$, which means that given $\delta > 0$, there is $\varepsilon^* > 0$ such that for $\varepsilon < \varepsilon^*$ and *all* $p \in \mathfrak{M}, v \in T_p\mathfrak{M}$ we have

$$|v| - \delta < |v|_\varepsilon < |v| + \delta.$$

Integrating over the speed of curves $\gamma : [0, 1] \to M$ yields

$$\int_0^1 |\dot{\gamma}| \,\mathrm{d}s - \delta < \int_0^1 |\dot{\gamma}|_\varepsilon \,\mathrm{d}s < \int_0^1 |\dot{\gamma}| \,\mathrm{d}s + \delta.$$

Now taking the infimum over all curves connecting x and y, we finally have

$$d(x, y) - \delta < d_\varepsilon(x, y) < d(x, y) + \delta.$$

By the uniform convergence of g_ε, this reasoning is independent of the points x and y, thus we have uniform convergence $d_\varepsilon \to d$ which means we found δ-isometries.

Now we have to show that the push-forward measures $\mathrm{id}_\# \mathfrak{m}_\varepsilon$ converge weakly to $\mathfrak{m}$. But this follows by observing that in coordinates the volume measure has the density $\sqrt{|\det g_\varepsilon|}$, which still converges uniformly. □

Chapter 4

Transportation Distances for Subprobability Measures

This chapter aims to introduce a metric on the space of subprobability measures. Our approach is based on a heuristic particle interpretation that involves also antiparticles. Particles can change to antiparticles when they hit the boundary. In contrast to the classical particle interpretation of the heat equation with Dirichlet boundary conditions, the particles do not get killed at the boundary, but reflected and thereby possibly changing to antiparticles. However, using this intuitive idea, we only get a function that does not satisfy the triangle inequality. Instead, we will study the induced metric and the further induced length metric in detail.

In all of this chapter we will assume that (X, d) is a complete, separable metric space and $Y \subset X$ a non-trivial, open subset, i.e. such that $\emptyset \neq Y \neq X$. Additional assumptions are given in the beginnings of the sections, if needed.

4.1 Charged Probability Measures and Identification with the Doubled Space

By abuse of notation we will often call $Z := X \setminus Y$ the *boundary* of Y despite it being different from the topological boundary ∂Y in general. While the distance between two particles – as well as between two antiparticles – at locations $x, y \in X$ is $d(x, y)$,

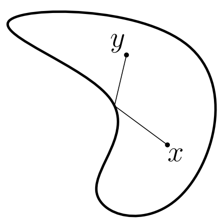

Figure 4.1: Distance between a particle and an antiparticle.

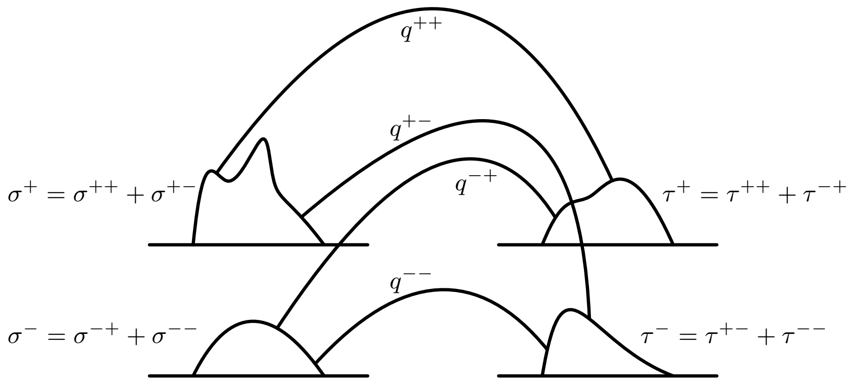

Figure 4.2: Decomposition of charged measures.

the distance between a particle at $x \in X$ and an antiparticle at $y \in X$ is given by

$$d^*(x,y) := \inf_{z \in X \setminus Y} \big[d(x,z) + d(z,y) \big],$$

see Figure 4.1. This expresses our heuristic idea that particles can change to antiparticles when they hit the boundary. To use the idea of particles and antiparticles, we use pairs of measures:

Definition 4.1.1. The space of *charged probability measures* is

$$\tilde{\mathcal{P}}(Y|X) := \Big\{ \sigma = (\sigma^+, \sigma^-) \Big| \sigma^\pm \in \mathcal{P}^{sub}(X),\, \sigma^+|_{X\setminus Y} = \sigma^-|_{X\setminus Y}, \\ \sigma^+(X) + \sigma^-(X) = 1 \Big\}.$$

The subprobability σ^+ represents a distribution of particles whereas σ^- represents a distribution of antiparticles. When at the same place, they annihilate, so what is left is the *effective* measure $\sigma^+ - \sigma^-$, supported on Y. Denote by $\sigma^0 := \sigma^+ - \sigma^-$ the *effective* measure and by $\overline{\sigma} := \sigma^+ + \sigma^-$ the *total* measure. Observe that σ^0 is in general a *signed* measure. However, we will mostly use charged measures with $\sigma^0 \geq 0$ since we usually start with a given subprobability μ and choose an appropriate σ with $\sigma^0 = \mu$.

A problem in defining a "transport" metric for subprobabilities is that it does not make sense to look for couplings between measures of unequal mass. To overcome this difficulty we interpret given measures $\mu, \nu \in \mathcal{P}^{sub}(Y)$ as effective measures of some $\sigma, \tau \in \tilde{\mathcal{P}}(Y|X)$. For charged measures, we will now define the L^p-transportation distance. We have to distinguish between transports from particles to antiparticles (and vice versa), and transports between particles and particles (or antiparticles and antiparticles), because the former use the metric d^* whereas the latter use d. To do so, given $\sigma, \tau \in \tilde{\mathcal{P}}(Y|X)$, we take a coupling $q \in \mathrm{Cpl}(\overline{\sigma}, \overline{\tau})$ between the total measures and decompose it in the following way. Since $\sigma^i \leq \overline{\sigma}$, there are densities such that $\sigma^i = u^i \overline{\sigma}, i \in \{+,-\}$, and analogously there are functions $v_j, j \in \{+,-\}$ such that $\tau^j = v^j \overline{\tau}$. Setting

$$\mathrm{d}q^{ij}(x,y) := u^i(x)v^j(y)\,\mathrm{d}q(x,y)$$

and

$$\sigma^{ij}(\cdot) := q(\cdot, X), \quad \tau^{ij}(\cdot) := q(X, \cdot),$$

we obtain a decomposition

$$q = q^{++} + q^{+-} + q^{-+} + q^{--}, \quad \sigma^i = \sigma^{i+} + \sigma^{i-}, \quad \tau^j = \tau^{+j} + \tau^{-j}$$

such that $q^{ij} \in \mathrm{Cpl}(\sigma^{ij}, \tau^{ij})$, $i, j \in \{+, -\}$, see Figure 4.2. Given this decomposition, we can now give the following definition.

Definition 4.1.2. For $p \in [1, \infty)$, we define the L^p-transportation cost between charged probability measures

$$\begin{aligned}\tilde{W}_p(\sigma, \tau) := \inf_{q \in \mathrm{Cpl}(\overline{\sigma}, \overline{\tau})} \Bigg\{ \int_{X \times X} d(x,y)^p \,\mathrm{d}q^{++}(x,y) + \int_{X \times X} d^*(x,y)^p \,\mathrm{d}q^{+-}(x,y) \\ + \int_{X \times X} d^*(x,y)^p \,\mathrm{d}q^{-+}(x,y) + \int_{X \times X} d(x,y)^p \,\mathrm{d}q^{--}(x,y) \Bigg\}^{1/p}.\end{aligned} \tag{4.1.1}$$

We further define

$$\tilde{\mathcal{P}}_p(Y|X) := \left\{ \sigma \in \tilde{\mathcal{P}}(Y|X) \,\middle|\, \tilde{W}_p\left(\sigma, \left(\frac{1}{2}\delta_x, \frac{1}{2}\delta_x\right)\right) < \infty \text{ for some/all } x \in X \right\}.$$

Lemma 4.1.3. *The map $\mu \mapsto (\frac{1}{2}\mu, \frac{1}{2}\mu)$ defines an isometric embedding of $\mathcal{P}_p(X)$ into $\tilde{\mathcal{P}}_p(Y|X)$.*

Proof. Using that $d(x,y) \le d^*(x,y)$, we have for every admissible coupling $q \in \mathrm{Cpl}(\frac{1}{2}\mu + \frac{1}{2}\mu, \frac{1}{2}\nu + \frac{1}{2}\nu)$:

$$\begin{aligned}&\int_{X \times X} d(x,y)^p \,\mathrm{d}q^{++}(x,y) + \int_{X \times X} d^*(x,y)^p \,\mathrm{d}q^{+-}(x,y) \\ &\qquad + \int_{X \times X} d^*(x,y)^p \,\mathrm{d}q^{-+}(x,y) + \int_{X \times X} d(x,y)^p \,\mathrm{d}q^{--}(x,y) \\ &\ge \int_{X \times X} d(x,y)^p \,\mathrm{d}q^{++}(x,y) + \int_{X \times X} d(x,y)^p \,\mathrm{d}q^{+-}(x,y) \\ &\qquad + \int_{X \times X} d(x,y)^p \,\mathrm{d}q^{-+}(x,y) + \int_{X \times X} d(x,y)^p \,\mathrm{d}q^{--}(x,y) \\ &\ge \int_{X \times X} d(x,y)^p \,\mathrm{d}q(x,y) \\ &\ge W_p(\mu, \nu)^p.\end{aligned}$$

On the other hand, every coupling $q \in \mathrm{Cpl}(\mu, \nu)$ in the definition of $W_p(\mu, \nu)$ is also an admissible coupling for $\tilde{W}_p$ with decomposition $q^{++} = q^{--} = \frac{1}{2}q$ and $q^{+-} = q^{-+} \equiv 0$, so we also have $W_p \ge \tilde{W}_p$. □

We will now show that $(\tilde{\mathcal{P}}_p(Y|X), \tilde{W}_p)$ can be isometrically identified with the Wasserstein space $(\mathcal{P}_p(\hat{X}), \hat{W}_p)$ over $\hat{X}$. This identification is very useful and for instance immediately tells us that $(\tilde{\mathcal{P}}_p(Y|X), \tilde{W}_p)$ is a complete separable metric space. Since we only look at two copies of $Y \subset X$, we index the different copies in the glued space by Y^+ and Y^- instead of the numerical indices in Subsection 3.1. Still, $Z := X \setminus Y$ and $\hat{X} = (X^+ \sqcup X^-) / \sim$. As we are dealing now with measures which are not equal on the different copies of X, in this section we do keep track of the identification maps $\iota_i, i \in \{+,-\}$. Every subset used in this section is assumed to be a Borel-measurable set in the space it is taken from.

Lemma 4.1.4. *The maps $\Phi : \tilde{\mathcal{P}}(Y|X) \to \mathcal{P}(\hat{X})$ and $\Psi : \mathcal{P}(\hat{X}) \to \tilde{\mathcal{P}}(Y|X)$, given by*

$$\begin{aligned}\Phi((\sigma^+, \sigma^-))(A) := \sigma^+(\iota_+^{-1}(A \cap Y^+)) + \sigma^-(\iota_-^{-1}(A \cap Y^-))\\ + \sigma^+(\iota_+^{-1}(A \cap Z)) + \sigma^-(\iota_-^{-1}(A \cap Z))\end{aligned}$$

for $A \subset \hat{X}$ and

$$\Psi(\hat{\sigma})^i(B) := \hat{\sigma}(\iota_i(B) \cap Y^i) + \frac{1}{2}\hat{\sigma}(\iota_i(B) \cap Z), \;\; i \in \{+,-\},$$

for $B \subset X$, respectively, are inverse to each other.

Proof. Let us first check that the maps are well-defined. For Φ this simply is

$$\Phi((\sigma^+, \sigma^-))(\hat{X}) = \sigma^+(Y^+) + \sigma^-(Y^-) + \sigma^+(Z) + \sigma^-(Z) = \sigma^+(X) + \sigma^-(X) = 1$$

For Ψ we first observe that

$$\Psi(\hat{\sigma})^+(X) + \Psi(\hat{\sigma})^-(X) := \hat{\sigma}(Y^+) + \frac{1}{2}\hat{\sigma}(Z) + \hat{\sigma}(Y^-) + \frac{1}{2}\hat{\sigma}(Z) = \hat{\sigma}(\hat{X}) = 1.$$

By definition it is clear that $\Psi(\hat{\sigma})^+\big|_Z = \Psi(\hat{\sigma})^-\big|_Z$. Hence both maps are well-defined. Now let us check that $\Phi \circ \Psi = \mathrm{id}_{\mathcal{P}(\hat{X})}$. Let $\hat{\sigma} \in \mathcal{P}(\hat{X})$ and $A \subset \hat{X}$ be measurable. Then

$$\begin{aligned}\Phi \circ \Psi(\hat{\sigma})(A) &= \Phi((\Psi(\hat{\sigma})^+, \Psi(\hat{\sigma})^-))(A)\\ &= \Psi(\hat{\sigma})^+(\iota_+^{-1}(A \cap Y^+)) + \Psi(\hat{\sigma})^-(\iota_-^{-1}(A \cap Y^-)) + \Psi(\hat{\sigma})^+(\iota_+^{-1}(A \cap Z))\\ &\quad + \Psi(\hat{\sigma})^-(\iota_-^{-1}(A \cap Z))\\ &= \hat{\sigma}(\iota_+(\iota_+^{-1}(A \cap Y^+)) \cap Y^+) + \frac{1}{2}\hat{\sigma}(\iota_+(\iota_+^{-1}(A \cap Y^+)) \cap Z)\\ &\quad + \hat{\sigma}(\iota_-(\iota_-^{-1}(A \cap Y^-)) \cap Y^-) + \frac{1}{2}\hat{\sigma}(\iota_-(\iota_-^{-1}(A \cap Y^-)) \cap Z)\\ &\quad + \hat{\sigma}(\iota_+(\iota_+^{-1}(A \cap Z) \cap Y^+) + \frac{1}{2}\hat{\sigma}(\iota_+(\iota_+^{-1}(A \cap Z)) \cap Z)\\ &\quad + \hat{\sigma}(\iota_-(\iota_-^{-1}(A \cap Z)) \cap Y^-) + \frac{1}{2}\hat{\sigma}(\iota_-(\iota_-^{-1}(A \cap Z)) \cap Z)\\ &= \hat{\sigma}(A \cap Y^+) + \hat{\sigma}(A \cap Y^-) + \frac{1}{2}\hat{\sigma}(A \cap Z) + \frac{1}{2}\hat{\sigma}(A \cap Z)\end{aligned}$$

$$= \hat{\sigma}(A).$$

We are left with showing that $\Psi \circ \Phi = \mathrm{id}_{\tilde{\mathcal{P}}(Y|X)}$. Let $\sigma = (\sigma^+, \sigma^-) \in \tilde{\mathcal{P}}(Y|X)$ and $B \subset X$ be measurable. Then

$$\begin{aligned}
\Psi \circ \Phi(\sigma)^+(B) &= \Phi(\sigma)(\iota_+(B) \cap Y^+) + \frac{1}{2}\Phi(\sigma)(\iota_+(B) \cap Z) \\
&= \sigma^+(\iota_+^{-1}(\iota_+(B) \cap Y^+ \cap Y^+)) + \sigma^-(\iota_-^{-1}(\iota_+(B) \cap Y^+ \cap Y^-)) \\
&\quad + \sigma^+(\iota_+^{-1}(\iota_+(B) \cap Y^+ \cap Z)) + \sigma^-(\iota_-^{-1}(\iota_+(B) \cap Y^+ \cap Z)) \\
&\quad + \frac{1}{2}\sigma^+(\iota_+^{-1}(\iota_+(B) \cap Z \cap Y^+)) + \frac{1}{2}\sigma^-(\iota_-^{-1}(\iota_+(B) \cap Z \cap Y^-)) \\
&\quad + \frac{1}{2}\sigma^+(\iota_+^{-1}(\iota_+(B) \cap Z \cap Z)) + \frac{1}{2}\sigma^-(\iota_-^{-1}(\iota_+(B) \cap Z \cap Z)) \\
&= \sigma^+(B \cap Y) + \frac{1}{2}\sigma^+(B \cap Z) + \frac{1}{2}\sigma^-(B \cap Z) \\
&= \sigma^+(B),
\end{aligned}$$

since $\sigma^+\big|_Z = \sigma^-\big|_Z$, and analogously for $\Psi \circ \Phi(\sigma)^-$. $\square$

Lemma 4.1.5. *Φ and Ψ are isometries between $(\tilde{\mathcal{P}}_p(Y|X), \tilde{W}_p)$ and $(\mathcal{P}_p(\hat{X}), \hat{W}_p)$ for every $p \in [1, \infty)$.*

Proof. We will show that

$$\hat{W}_p(\hat{\sigma}, \hat{\tau}) = \tilde{W}_p(\Psi(\hat{\sigma}), \Psi(\hat{\tau})).$$

Let $\hat{q} \in \mathrm{Cpl}(\hat{\sigma}, \hat{\tau})$. Then we have to construct a coupling $q \in \mathrm{Cpl}(\overline{\Psi(\hat{\sigma})}, \overline{\Psi(\hat{\tau})})$ giving the same cost. Given a set $\mathcal{A} \subset X \times X$, we will define q by defining the decomposition q^{ij} by embedding $\mathcal{A}$ into the "$X^i \times X^j$".

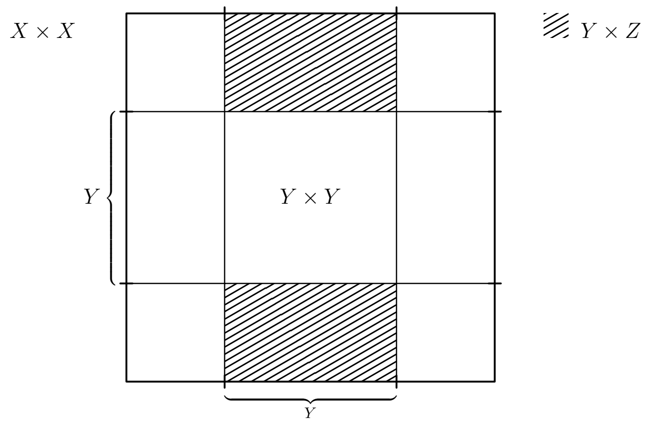

Figure 4.3: Decomposition of $X \times X$ to define the coupling.

Let $\mathcal{A}^{ij} := \iota_{ij}(\mathcal{A}) := \{(\iota_i(x_1), \iota_j(x_2)) \,\big|\, (x_1, x_2) \in \mathcal{A}\}$. Then, for $i, j \in \{+, -\}$, we define

$$q^{ij}(\mathcal{A}) := \tilde{q}^{ij}(\mathcal{A} \cap (Y \times Y)) + \frac{1}{2}\tilde{q}^{ij}(\mathcal{A} \cap (Y \times Z)) + \frac{1}{2}\tilde{q}^{ij}(\mathcal{A} \cap (Z \times Y))$$

$$
\begin{aligned}
&+ \frac{1}{4}\tilde{q}^{ij}(\mathcal{A} \cap (Z \times Z))\\
:= {}& \hat{q}(\mathcal{A}^{ij} \cap (Y^i \times Y^j)) + \frac{1}{2}\hat{q}(\mathcal{A}^{ij} \cap (Y^i \times Z)) + \frac{1}{2}\hat{q}(\mathcal{A}^{ij} \cap (Z \times Y^j))\\
&+ \frac{1}{4}\hat{q}(\mathcal{A}^{ij} \cap (Z \times Z)),
\end{aligned}
$$

see Figure 4.3. Observe that $\tilde{q}^{ij}(\mathcal{A}\cap(Y\times Y)) = \hat{q}(\iota_{ij}(\mathcal{A}\cap(Y\times Y))) = (\iota_{ij}^{-1})_{\#}\hat{q}(\mathcal{A}\cap(Y\times Y))$ and similarly for the other terms. This seemingly complicated decomposition into 4 terms instead of just taking "$\hat{q}(\mathcal{A}^{ij})$" is necessary because otherwise we would count the parts on the boundary wrong. As a candidate for a coupling between $\overline{\Psi(\hat{\sigma})}$ and $\overline{\Psi(\hat{\tau})}$ we define

$$
q(\mathcal{A}) := q^{++}(\mathcal{A}) + q^{+-}(\mathcal{A}) + q^{-+}(\mathcal{A}) + q^{--}(\mathcal{A}). \tag{4.1.2}
$$

Let us first show that it is a probability measure on $X \times X$:

$$
\begin{aligned}
\sum_{i,j\in\{+,-\}} q^{ij}(X \times X) &= \sum_{i,j\in\{+,-\}} \Big[\hat{q}(Y^i \times Y^j) + \frac{1}{2}\hat{q}(Y^i \times Z)\\
&\qquad + \frac{1}{2}\hat{q}(Z \times Y^j) + \frac{1}{4}\hat{q}(Z \times Z)\Big]\\
&= \hat{q}(\hat{X} \times \hat{X}) = 1,
\end{aligned}
$$

since the sum is a disjoint partition of $\hat{X} \times \hat{X} = (Y^+ \sqcup Y^- \sqcup Z) \times (Y^+ \sqcup Y^- \sqcup Z)$.

Next we show that it is indeed a coupling. Taking a subset $A \subset X$ and defining $A^i := \iota_i(A)$, we evaluate (4.1.2) at $\mathcal{A} := A \times X$:

$$
\begin{aligned}
q(A \times X) = {}& \hat{q}((A^+ \cap Y^+) \times Y^+) + \frac{1}{2}\hat{q}((A^+ \cap Y^+) \times Z) + \frac{1}{2}\hat{q}((A^+ \cap Z) \times Y^+)\\
&+ \frac{1}{4}\hat{q}((A^+ \cap Z) \times Z) + \hat{q}((A^+ \cap Y^+) \times Y^-) + \frac{1}{2}\hat{q}((A^+ \cap Y^+) \times Z)\\
&+ \frac{1}{2}\hat{q}((A^+ \cap Z) \times Y^-) + \frac{1}{4}\hat{q}((A^+ \cap Z) \times Z) + \hat{q}((A^- \cap Y^-) \times Y^+)\\
&+ \frac{1}{2}\hat{q}((A^- \cap Y^-) \times Z) + \frac{1}{2}\hat{q}((A^- \cap Z) \times Y^+) + \frac{1}{4}\hat{q}((A^- \cap Z) \times Z)\\
&+ \hat{q}((A^- \cap Y^-) \times Y^-) + \frac{1}{2}\hat{q}((A^- \cap Y^-) \times Z) + \frac{1}{2}\hat{q}((A^- \cap Z) \times Y^-)\\
&+ \frac{1}{4}\hat{q}((A^- \cap Z) \times Z)\\
= {}& \hat{q}((A^+ \cap Y^+) \times \hat{X}) + \hat{q}((A^- \cap Y^-) \times \hat{X}) + \frac{1}{2}\hat{q}((A^+ \cap Z) \times \hat{X})\\
&+ \frac{1}{2}\hat{q}((A^- \cap Z) \times \hat{X})\\
= {}& \hat{\sigma}(A^+ \cap Y^+) + \hat{\sigma}(A^- \cap Y^-) + \frac{1}{2}\hat{\sigma}(A^+ \cap Z) + \frac{1}{2}\hat{\sigma}(A^- \cap Z)\\
= {}& \Psi(\hat{\sigma})^+(A) + \Psi(\hat{\sigma})^-(A) = \overline{\Psi(\hat{\sigma})}(A).
\end{aligned}
$$

This works analogously for $\mathcal{A} = X \times A$ and $\overline{\Psi(\hat{\tau})}$. We are left to show that this coupling used in the definition of $\tilde{W}_2$ gives the same value as $\hat{W}_2(\hat{\sigma}, \hat{\tau})$. Let us only discuss one term in order to simplify the exposition.

$$\begin{aligned}
&\int_{X\times X} d(x,y)^p \, \mathrm{d}q^{++}(x,y) \\
&\qquad = \int_{Y\times Y} d(x,y)^p \, \mathrm{d}\tilde{q}^{++}(x,y) + \frac{1}{2}\int_{Y\times Z} d(x,y)^p \, \mathrm{d}\tilde{q}^{++}(x,y) \\
&\qquad\quad + \frac{1}{2}\int_{Z\times Y} d(x,y)^p \, \mathrm{d}\tilde{q}^{++}(x,y) + \frac{1}{4}\int_{Z\times Z} d(x,y)^p \, \mathrm{d}\tilde{q}^{++}(x,y) \\
&\qquad = \int_{Y^+\times Y^+} d(x,y)^p \, \mathrm{d}\hat{q}(x,y) + \frac{1}{2}\int_{Y^+\times Z} d(x,y)^p \, \mathrm{d}\hat{q}(x,y) \\
&\qquad\quad + \frac{1}{2}\int_{Z\times Y^+} d(x,y)^p \, \mathrm{d}\hat{q}(x,y) + \frac{1}{4}\int_{Z\times Z} d(x,y)^p \, \mathrm{d}\hat{q}(x,y)
\end{aligned}$$

where we abused notation once more by omitting the identification maps ι_i. Taking into account that $d(x,y) = d^*(x,y)$ in case at least one of the two points is lying in Z, and adding all the terms up, we indeed see that with this choice of q we have $\hat{W}_p(\hat{\sigma}, \hat{\tau}) \geq \tilde{W}_p(\Psi(\hat{\sigma}), \Psi(\hat{\tau}))$.

For the other direction, given $q \in \mathrm{Cpl}(\overline{\Psi(\hat{\sigma})}, \overline{\Psi(\hat{\tau})})$ we have to construct a suitable $\hat{q} \in \mathrm{Cpl}(\hat{\sigma}, \hat{\tau})$. Given a set $\mathcal{B} \subset \hat{X} \times \hat{X}$, define $\mathcal{B}^{ij} := \mathcal{B} \cap (X^i \times X^j)$ and

$$\hat{q}(\mathcal{B}) := \sum_{i,j\in\{+,-\}} q^{ij}(\iota_{ij}^{-1}(\mathcal{B}^{ij})).$$

With similar arguments as above one sees that indeed $\hat{q}$ is a coupling of $\hat{\sigma}$ and $\hat{\tau}$ with the same cost. Hence we also have $\hat{W}_p(\hat{\sigma}, \hat{\tau}) \leq \tilde{W}_p(\Psi(\hat{\sigma}), \Psi(\hat{\tau}))$. □

Lemma 4.1.6. *For each $p \in [1,\infty)$, $\tilde{W}_p$ is a complete separable metric on $\tilde{\mathcal{P}}_p(Y|X)$. It is a length metric if d is a length metric; $\tilde{\mathcal{P}}_p(Y|X)$ is compact if X is compact.*

Proof. This is an immediate consequence of the isometry between $\tilde{\mathcal{P}}_p(Y|X)$ and $\mathcal{P}_p(\hat{X})$, together with Lemma 3.1.3. □

4.2 Transportation-Annihilation Pre-Distance

From now on, we assume our space X to be a length space, so that in particular also $(\tilde{\mathcal{P}}_p(Y|X), \tilde{W}_p)$ is a length space.

We will now use the metric for charged probability measures to define a semi-metric between subprobabilities. The idea is that we interpret a subprobability measure as a distribution of particles and complete them to a charged probability by adding the same amount of particles and antiparticles which in the effective measure annihilate, see Figure 4.4.

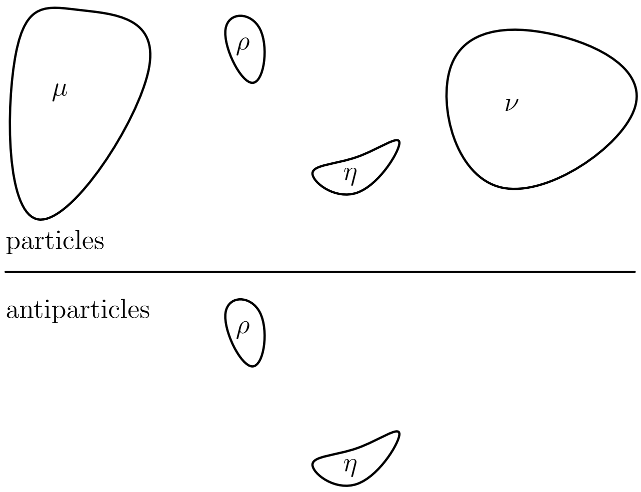

Figure 4.4: Charged measures in the definition of the transportation-annihilation pre-distance.

Definition 4.2.1. For $\mu, \nu \in \mathcal{P}^{sub}(Y)$ and $p \in [1, \infty)$ we define

$$\begin{aligned} W_p^0(\mu, \nu) :=& \inf \left\{ \tilde{W}_p(\sigma, \tau) \,\middle|\, \sigma, \tau \in \tilde{\mathcal{P}}(Y|X), \sigma^0 = \mu, \tau^0 = \nu \right\} \\ =& \inf \Big\{ \tilde{W}_p\big((\mu + \rho, \rho), (\nu + \eta, \eta)\big) \,\Big|\, \rho, \eta \in \mathcal{P}^{sub}(X), (\mu + 2\rho)(X) = 1, \\ & \qquad (\nu + 2\eta)(X) = 1 \Big\}, \end{aligned}$$

called the *transportation-annihilation pre-distance.* Moreover, we put

$$\mathcal{P}_p^{sub}(Y) := \left\{ \mu \in \mathcal{P}^{sub}(Y) \,\middle|\, W_p^0(\mu, \delta_y) < \infty \text{ for some/all } y \in Y \right\}.$$

Remark 4.2.2.
a) The infima in the previous Definition will be attained if X is compact. Observe that without compactness this is not clear because we don't know if minimizing sequences $(\sigma_n)_n, (\tau_n)_n$ are tight.

b) If μ and ν are *probability* measures, then $W_p^0(\mu, \nu)$ coincides with the usual L^p-Kantorovich-Wasserstein metric $W_p(\mu, \nu)$.

c) In general, W_p^0 will not satisfy the triangle inequality. For instance, let $X = \mathbb{R}, Y = (-3, 3), \mu = \delta_{-2}, \nu = \delta_2, \xi = 0$. Then

$$W_p^0(\mu, \nu) = W_p(\delta_{-2}, \delta_2) = 4,$$

but

$$W_p^0(\mu, \xi) = \inf_{\tau = (\tau_1, \tau_1)} \tilde{W}_p((\delta_{-2}, 0), (\tau_1, \tau_1)) \le \tilde{W}_p\left((\delta_{-2}, 0), \left(\frac{1}{2}\delta_{-3}, \frac{1}{2}\delta_{-3} \right) \right) = 1,$$

i.e.

$$4 = W_p^0(\mu, \nu) \not\le W_p^0(\mu, \xi) + W_p^0(\xi, \nu) = 2.$$

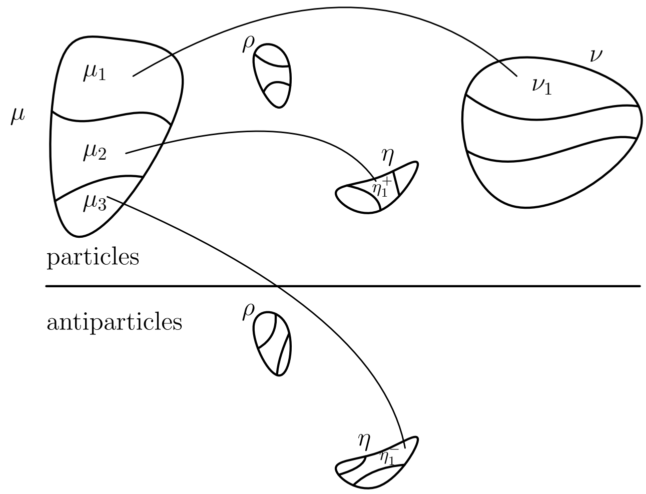

Figure 4.5: Decomposition of $W_p^0(\mu, \nu)$ in Lemma 4.2.5.

This definition is impractical for another reason than just the lack of a triangle inequality: given a sequence $(\mu_n)_n$ and a measure μ, to study $W_p^0(\mu_n, \mu)$ we get sequences $(\sigma_n)_n, (\tau_n)_n \in \tilde{\mathcal{P}}_p(Y|X)$ where $\tau_n^0 = \mu$. This means we cannot choose a *fixed* charged measure representing μ, but it also depends on the element in the sequence we are comparing it with. This makes it hard to extract converging subsequences in the case that the base space is not compact because in principle the added masses in τ_n could wander off to infinity. The rest of this section will be devoted to derive more useful characterizations of W_p^0 through more conventional terms. In these descriptions, a related transportation cost appears:

Definition 4.2.3. Given subprobability measures $\mu, \nu \in \mathcal{P}^{sub}(X)$ with equal mass $\mu(X) = \nu(X)$, we define the transport cost with respect to d^*:

$$W_p^*(\mu, \nu)^p := \inf_{q \in \mathrm{Cpl}(\mu, \nu)} \int_{X \times X} d^*(x, y)^p \, \mathrm{d}q(x, y).$$

Further, we introduce

$$W_p^*(\mu) := \frac{1}{2} W_p^*(\mu, \mu).$$

Both functions will be referred to as *annihilation costs.*

Remark 4.2.4. W_p^* is symmetric in its arguments and satisfies the triangle inequality but typically $W_p^*(\mu, \mu) \neq 0$, so it is a meta-metric.

A first, easy step consists in decomposing the transport between $(\mu + \rho, \rho)$ and $(\nu + \eta, \eta)$ into nine transports, see Figure 4.5.

Lemma 4.2.5. *Let $\mu, \nu \in \mathcal{P}_p^{sub}(Y)$. Then*

$$\begin{aligned} W_p^0(\mu, \nu)^p = \inf \Big\{ & W_p(\mu_1, \nu_1)^p + W_p(\mu_2, \eta_1^+)^p + W_p^*(\mu_3, \eta_1^-)^p \\ & + W_p(\rho_1^+, \nu_2)^p + W_p(\rho_2^+, \eta_2^+)^p + W_p^*(\rho_3^+, \eta_2^-)^p \end{aligned}$$

$$+ W_p^*(\rho_1^-, \nu_3)^p + W_p^*(\rho_2^-, \eta_3^+)^p + W_p(\rho_3^-, \eta_3^-)^p \Big| \tag{4.2.1}$$
$$\mu = \mu_1 + \mu_2 + \mu_3, \rho = \rho_1^+ + \rho_2^+ + \rho_3^+ = \rho_1^- + \rho_2^- + \rho_3^-,$$
$$\nu = \nu_1 + \nu_2 + \nu_3,\ \eta = \eta_1^+ + \eta_2^+ + \eta_3^+ = \eta_1^- + \eta_2^- + \eta_3^-,$$
$$(\mu + 2\rho)(X) = 1, (\nu + 2\eta)(X) = 1\Big\}.$$

The decompositions implicitly require the coupled measures to have the same mass, so for instance $\mu_1(X) = \nu_1(X)$ etc.

Proof. Given $\varepsilon > 0$, let $\rho, \eta \in \mathcal{P}^{sub}(X)$ such that $W_p^0(\mu, \nu) + \varepsilon > \tilde{W}_p((\mu, \rho, \rho), (\nu + \eta, \eta))$. We will switch to the setting of the glued space for convenience. Thus, we will now consider the measures $\Phi((\mu + \rho, \rho))$ and $\Phi((\nu + \eta, \eta))$ on $\mathcal{P}(\hat{X})$. By abuse of notation we will stick to the names of the measures and add pluses and minuses depending on whether they are measures on the upper or lower part of the glued space. Let $q \in \mathrm{Cpl}(\mu^+ + \rho^+ + \rho^-, \nu^+ + \eta^+ + \eta^-)$ be an optimal coupling for $\hat{W}_p(\mu^+ + \rho^+ + \rho^-, \nu^+ + \eta^+ + \eta^-)$. Now we disintegrate q with respect to $\mu^+ + \rho^+ + \rho^-$, getting a family of measures $(q_x)_{x \in \hat{X}}$. For $(\mu^+ + \rho^+ + \rho^-)$-almost every $x \in \hat{X}$, q_x is absolutely continuous with respect to $\nu^+ + \eta^+ + \eta^-$: Indeed, given a set $B \subset \hat{X}$ with $(\nu^+ + \eta^+ + \eta^-)(B) = 0$, we have

$$0 = (\nu^+ + \eta^+ + \eta^-)(B) = q(\hat{X} \times B) = \int_{\hat{X}} q_x(B) \, \mathrm{d}(\mu^+ + \rho^+ + \rho^-)(x),$$

thus $q_x(B) = 0$ for $(\mu^+ + \rho^+ + \rho^-)$-almost every $x \in \hat{X}$. Denote the density by

$$\varphi(x, y) := \frac{\mathrm{d}q_x}{\mathrm{d}(\nu^+ + \eta^+ + \eta^-)}(y).$$

If we now for instance define

$$\mathrm{d}\mu_1(x) := \left(\int_{\hat{X}} \varphi(x, y) \, \mathrm{d}\nu^+(y) \right) \mathrm{d}\mu^+(x), \quad \mathrm{d}\nu_1(y) := \left(\int_{\hat{X}} \varphi(x, y) \, \mathrm{d}\mu^+(x) \right) \mathrm{d}\nu^+(y),$$

then $\varphi(x, y) \, \mathrm{d}\mu^+(x) \, \mathrm{d}\nu^+(y)$ is an optimal coupling for μ_1 and ν_1. Analogously defining the remaining 14 measures in (4.2.1), we get 9 couplings, the sum of which is the original coupling q. Optimality of these "partial" couplings is inherited because if there were better ones for the 9 terms, then the sum of those 9 couplings would be again a coupling for $\mu^+ + \rho^+ + \rho^-$ and $\nu^+ + \eta^+ + \eta^-$, but with a lower cost, which is a contradiction to the optimality of q. Thus, for every $\varepsilon > 0$ we found a decomposition into 9 terms that is ε-close to $W_p^0(\mu, \nu)$. □

For $p = 1$ many of the a priori possible ways of transporting mass in this formula are not necessary and we can simplify it to a more convenient representation which does not need the additional measures ρ and η. However, for $p > 1$ we only get an upper bound.

Lemma 4.2.6. *i) For $p \in [1, \infty)$ and all $\mu, \nu \in \mathcal{P}_p^{sub}(Y)$*

$$W_p^0(\mu, \nu)^p \leq \inf \Big\{ W_p(\mu_1, \nu_1)^p + W_p^*(\mu_0)^p + W_p^*(\nu_0)^p \Big|$$
$$\mu = \mu_1 + \mu_0, \nu = \nu_1 + \nu_0, (\mu + \nu_0)(X) \leq 1, (\nu + \mu_0)(X) \leq 1 \Big\}.$$

ii) For $\mu, \nu \in \mathcal{P}_1^{sub}(Y)$

$$W_1^0(\mu,\nu) = \inf \Big\{ W_1(\mu_1,\nu_1) + W_1^*(\mu_0) + W_1^*(\nu_0) \,\Big|\, \mu = \mu_1 + \mu_0, \nu = \nu_1 + \nu_0, (\mu+\nu_0)(X) \le 1, (\nu+\mu_0)(X) \le 1 \Big\}.$$

Remark 4.2.7. As one can see from the proof, part ii) is actually true without assuming X to be a length space.

Proof of Lemma 4.2.6. i) In the previous Lemma 4.2.5 choose the decomposition $\rho_3^+ = \eta_2^- = \rho_2^- = \eta_3^+ = 0$ and $\rho_2^+ = \eta_2^+ = \rho_3^- = \eta_3^-$, so that

$$W_p^0(\mu,\nu)^p \le \inf \big\{ W_p(\mu_1,\nu_1)^p + W_p(\mu_2,\eta_1^+)^p + W_p^*(\mu_3,\eta_1^-)^p + W_p(\rho_1^+,\nu_2)^p + W_p^*(\rho_1^-,\nu_3)^p \,\big|\, (\mu+2\nu_2)(X) \le 1, (\nu+2\mu_2)(X) \le 1 \big\}.$$

Let us first discuss the case $p = 1$. Then

$$\inf_{\eta_1^+,\, \mu_2+\mu_3=\mu_0} \Big[W_1(\mu_2,\eta_1^+) + W_1^*(\eta_1^+,\mu_3) \Big] \le \frac{1}{2} W_1^*(\mu_0,\mu_0) = W_1^*(\mu_0)$$

by choosing $\eta_1^+ = \mu_2 = \mu_3 = \frac{1}{2}\mu_0$. Together with the fact that $\rho_1^+ = \rho_1^-, \eta_1^+ = \eta_1^-$ we thus get

$$W_1^0(\mu,\nu) \le \inf \big\{ W_1(\mu_1,\nu_1) + W_1^*(\mu_0) + W_1^*(\nu_0) \,\big|\, (\mu+\nu_0)(X) \le 1, (\nu+\mu_0)(X) \le 1 \big\}.$$

For the case $p > 1$ we are working with optimal transport in the glued space $\hat{X}$, using the identification of $(\tilde{\mathcal{P}}_p(Y|X), \tilde{W}_p)$ and $(\mathcal{P}_p(\hat{X}), \hat{W}_p)$.
Given an ε-$\tilde{W}_p$-geodesic $(\sigma_t)_{t\in[0,1]}$ connecting $\sigma_0 := (\mu_0, 0)$ and $\sigma_1 := (0, \mu_0)$, we decompose it into two ε-$\tilde{W}_p$-geodesics $(\sigma'_t)_{t\in[0,1]}$ and $(\sigma''_t)_{t\in[0,1]}$ such that

$$\tilde{W}_p(\sigma'_0,\sigma'_1) = \tilde{W}_p(\sigma''_0,\sigma''_1) = \frac{1}{2}\tilde{W}_p(\sigma_0,\sigma_1) \quad \text{and} \quad \sigma'_{1/2}(Y^-) = \sigma''_{1/2}(Y^+) = 0.$$

Choosing $\mu_2 = (\sigma'_0)^+$, $\mu_3 = (\sigma'_1)^-$, and $\eta_1^+ = (\sigma'_{1/2})^+$ and using that $\sigma'_{1/2}$ is an ε-midpoint then yields

$$\begin{aligned}
\inf_{\eta_1^+,\, \mu_2+\mu_3=\mu_0} \Big[W_p(\mu_2,\eta_1^+)^p + W_p^*(\eta_1^+,\mu_3)^p \Big]
&\le W_p\big((\sigma'_0)^+, (\sigma'_{1/2})^+\big)^p + W_p^*\big((\sigma'_{1/2})^+, (\sigma'_1)^+\big)^p \\
&= \tilde{W}_p\big(\sigma'_0, \sigma'_{1/2}\big)^p + \tilde{W}_p\big(\sigma'_{1/2}, \sigma'_1\big)^p \\
&\le 2\left(\frac{1}{2}\tilde{W}_p(\sigma'_0,\sigma'_1) + \varepsilon\right)^p \\
&= 2\left(\frac{1}{4}\tilde{W}_p(\sigma_0,\sigma_1) + \varepsilon\right)^p \\
&\le \left(\frac{1}{2}\tilde{W}_p(\sigma_0,\sigma_1) + 2\varepsilon\right)^p
\end{aligned}$$

$$= \left(W_p^*(\mu_0) + 2\varepsilon\right)^p.$$

By this we can continue as in the case $p = 1$.

ii) To prove the "$\geq$"- inequality, we assume for simplicity that minimizers in the definition of W_1^0 exist. This is for instance the case when X is compact. For the general case one has to work with almost-minimizers.

Let subprobabilities μ and ν be given as well as ρ and η with $(\mu + 2\rho)(X) = 1, (\nu + 2\eta)(X) = 1$ such that

$$\begin{aligned} W_1^0(\mu, \nu) &= \tilde{W}_1\big((\mu + \rho, \rho), (\nu + \eta, \eta)\big) \\ &= \hat{W}_1\big(\mu + \rho + \rho^*, \nu + \eta + \eta^*\big) \end{aligned}$$

where for the last identity we switched to the picture of the glued space $\hat{X}$ with subprobabilities μ, ν, ρ, η on the "upper" sheet X^+ and their copies ρ^*, η^* on the "lower" sheet X^-. We further assume for the moment that all masses are rational numbers. This is to approximate the measures in a convenient way by sums of Dirac measures:
Given $\varepsilon > 0$, choose $n, n_1, n_2 \in \mathbb{N}$ and $x_i, y_i, u_i, v_i \in X^+$ for $i = 1, \ldots, n$ such that the measures

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n-2n_1} \delta_{x_i}, \quad \nu_n = \frac{1}{n}\sum_{i=1}^{n-2n_2} \delta_{y_i}, \quad \rho_n = \frac{1}{n}\sum_{i=1}^{n_1} \delta_{u_i}, \quad \eta_n = \frac{1}{n}\sum_{i=1}^{n_2} \delta_{v_i}$$

satisfy

$$W_1(\mu, \mu_n) \leq \varepsilon, \quad W_1(\nu, \nu_n) \leq \varepsilon, \quad W_1(\rho, \rho_n) \leq \varepsilon, \quad W_1(\eta, \eta_n) \leq \varepsilon.$$

To avoid ambiguity, we may assume that the sets $\{x_i\}$ and $\{y_i\}$ are disjoint form each other. Such an approximation is possible as shown for instance in [Vil09, Theorem 6.18].
Denoting *conjugate points* by

$$u^* := \begin{cases} \iota_- \circ \iota_+^{-1}(u), & \text{if } u \in X^+ \\ \iota_+ \circ \iota_-^{-1}(u), & \text{if } u \in X^-, \end{cases}$$

(so that in particular $(u^*)^* = u$) we also have

$$W_1(\rho^*, \rho_n^*) \leq \varepsilon, \quad W_1(\eta^*, \eta_n^*) \leq \varepsilon$$

for

$$\rho_n^* = \frac{1}{n}\sum_{i=1}^{n_1} \delta_{u_i^*}, \quad \eta_n^* = \frac{1}{n}\sum_{i=1}^{n_2} \delta_{v_i^*}.$$

In particular we have $\frac{n_1}{n} = \rho(X)$ and so on.

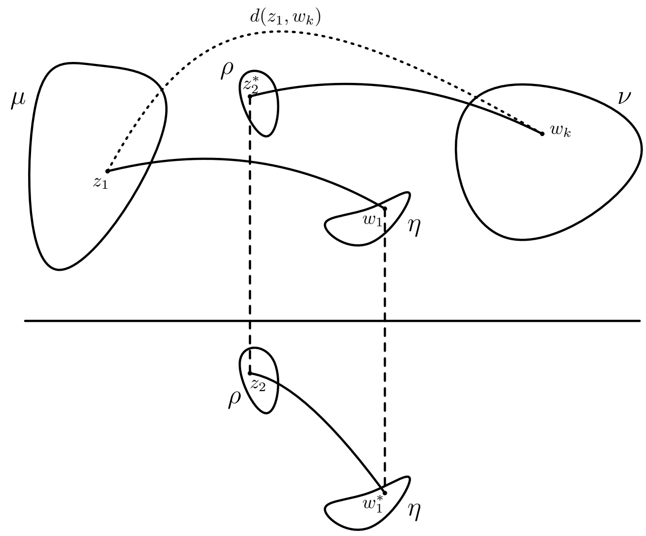

Figure 4.6: Chains in Case 1.

Now fix a $\hat{W}_1$-optimal coupling q_n of $\mu_n + \rho_n + \rho_n^*$ and $\nu_n + \eta_n + \eta_n^*$ on $\hat{X}$. As shown in [EH15, Lemma 2.6], we can choose this coupling q_n as a matching (i.e. it does not split mass), that is,

$$q_n = \frac{1}{n} \sum_{\xi \in Q_n} \delta_\xi$$

with suitable $Q_n \subset Z \times W$ where $Z := \{x_i\} \cup \{u_i\} \cup \{u_i^*\}$ and $W := \{y_i\} \cup \{v_i\} \cup \{v_i^*\}$. We consider now chains consisting of a sequence of coupled pairs $(z_1, w_1), \ldots, (z_k, w_k) \in Q_n$, $k \in \mathbb{N}$, with $z_{i+1} = z_i^*$ or $w_{i+1} = w_i^*$. This means that whenever we have a pair that contains an element of $\{u_i\} \cup \{u_i^*\} \cup \{v_i\} \cup \{v_i^*\}$, we also look at the pair containing the conjugated point. Extending these sequences this way until no further pair can be added satisfying the constraint, we arrive at three classes of *maximal* chains.

Case 1: Chains such that $z_1 \in \{x_i\}$ and $w_k \in \{y_i\}$.
See Figure 4.6 for a generic chain in this case. Observe that the constraint on consecutive pairs tells us that $w_{2i} = w_{2i-1}^*$ and $z_{2i+1} = z_{2i}^*$ for $i = 1, \ldots, \frac{k-1}{2}$. Thanks to the general inequality $\hat{d}(x, y) \geq d(\iota_i^{-1}(x), \iota_j^{-1}(y))$ for $x \in X^i$, $y \in X^j$, $i, j \in \{+, -\}$, we can give a lower bound on the transportation cost of this sequence of pairs:

$$\begin{aligned}
\hat{d}(z_1, w_1) + \hat{d}(z_2, w_2) + &\hat{d}(z_3, w_3) + \cdots + \hat{d}(z_k, w_k) \\
&= \hat{d}(z_1, w_1) + \hat{d}(z_2, w_1^*) + \hat{d}(z_2^*, w_3) + \cdots + \hat{d}(z_{k-1}^*, w_k) \\
&\geq d(z_1, w_1) + d(z_2, w_1) + d(z_2, w_3) + \cdots + d(z_{k-1}, w_k) \\
&\geq d(z_1, w_k).
\end{aligned}$$

Here, by abuse of notation, we omitted the identification maps to project the points in the same copy. Collecting all the initial points z_1 of chains of this form in $X_1 \subset \{x_i\}$

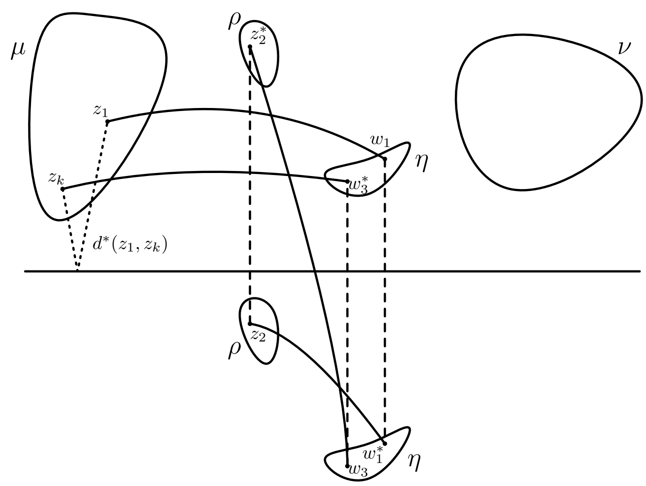

Figure 4.7: Chains in Case 2a.

and the endpoints w_k in $Y_1 \subset \{y_i\}$, the sum of the costs of chains of this type is bounded below by the cost $W_1(\mu_n^1, \nu_n^1)$ with measures

$$\mu_n^1 := \frac{1}{n} \sum_{x \in X_1} \delta_x, \qquad \nu_n^1 := \frac{1}{n} \sum_{y \in Y_1} \delta_y.$$

Case 2a: Chains such that $z_1 \in \{x_i\}$ and $z_k \in \{x_i\}$.
Chains in this case look like Figure 4.7. In this case there exists a pair in the chain that couples points on the different copies of the doubled space, i.e. there is $j_* \in \{1, \dots, k\}$ such that $z_{j_*} \in X^\alpha$ and $w_{j_*} \in X^\beta$ with $\alpha, \beta \in \{+,-\}$, $\alpha \neq \beta$ and hence

$$\hat{d}(z_{j_*}, w_{j_*}) = d^*(z_{j_*}, w_{j_*}) = d^*(z_{j_*}^*, w_{j_*}^*).$$

Without loss of generality we assume that $z_{j_*} = z_{j_*-1}^*$ and $w_{j_*+1} = w_{j_*}^*$. Then, as in Case 1,

$$\begin{aligned}
\hat{d}(z_1, w_1) &+ \hat{d}(z_2, w_2) + \hat{d}(z_3, w_3) + \cdots + \hat{d}(z_{j_*}, w_{j_*}) + \cdots + \hat{d}(z_k, w_k) \\
&\geq d(z_1, w_1) + d(z_2, w_1) + d(z_2, w_3) + \cdots + d^*(z_{j_*}, w_{j_*}) + \cdots + d(z_{k-1}, w_k) \\
&= d(z_1, z_{j_*-1}) + d^*(z_{j_*}^*, w_{j_*}^*) + d(w_{j_*+1}, z_k) \\
&= d(z_1, z_{j_*-1}) + d^*(z_{j_*-1}, w_{j_*+1}) + d(w_{j_*+1}, z_k) \\
&\geq d^*(z_1, z_k).
\end{aligned}$$

In this case we collect the starting points z_1 of chains of this form in the set $X_0 \subset \{x_i\}$ (which is equivalent to collecting the endpoints z_k). Denoting

$$\mu_n^0 := \frac{1}{n} \sum_{x \in X_0} \delta_x,$$

the sum of the costs of these chains is bounded below by $\frac{1}{2} W_1^*(\mu_n^0, \mu_n^0)$.

Case 2b: Chains such that $w_1 \in \{y_i\}$ and $w_k \in \{y_i\}$.
This case is completely analogous to Case 2a; we collect the starting points w_1 of such chains in $Y_0 \subset \{y_i\}$ and define

$$\nu_n^0 := \frac{1}{n} \sum_{y \in Y_0} \delta_y.$$

Then the sum of the costs of these chains is bounded below by $\frac{1}{2} W_1^*(\nu_n^0, \nu_n^0)$.

Case 3: Chains such that $z_1 \in \{u_i\} \cup \{u_i^*\}$ and $z_k = z_1^*$.
The cost of these cyclic chains is redundant. They can be avoided by an appropriate choice of the measures ρ_n, η_n, namely by choosing the points from ρ_n, ρ_n^* and η_n, η_n^* that occur in these chains to coincide so that $z_j = w_j$.

Observe that each chain in Case 2a contains at least two points in $\{v_i\} \cup \{v_i^*\}$. This means that the number of points in X_0 is at most $2n_2$, and hence

$$(\nu_n + \mu_n^0)(X) \leq \frac{n - 2n_2 + 2n_2}{n} = 1.$$

Analogously for the chains in Case 2b, so that $(\mu_n + \nu_n^0)(X) \leq 1$.
Thus we have a lower bound

$$\hat{W}_1\big(\mu_n + \rho_n + \rho_n^*, \nu_n + \eta_n + \eta_n^*\big) \geq W_1\big(\mu_n^1, \nu_n^1\big) + \frac{1}{2} W_1^*\big(\mu_n^0, \mu_n^0\big) + \frac{1}{2} W_1^*\big(\nu_n^0, \nu_n^0\big).$$

Via the optimal coupling of μ_n and μ, the decomposition $\mu_n = \mu_n^1 + \mu_n^0$ induces a decomposition $\mu = \mu^1 + \mu^0$ such that

$$W_1(\mu^1, \mu_n^1) \leq \varepsilon, \quad W_1(\mu^0, \mu_n^0) \leq \varepsilon,$$

and similarly for $\nu_n = \nu_n^1 + \nu_n^0$ and $\nu = \nu^1 + \nu^0$. This finally yields

$$\begin{aligned} W_1^0(\mu, \nu) &= \hat{W}_1\big(\mu + \rho + \rho^*, \nu + \eta + \eta^*\big) \\ &\geq \hat{W}_1\big(\mu_n + \rho_n + \rho_n^*, \nu_n + \eta_n + \eta_n^*\big) - 6\varepsilon \\ &\geq W_1\big(\mu_n^1, \nu_n^1\big) + \frac{1}{2} W_1^*\big(\mu_n^0, \mu_n^0\big) + \frac{1}{2} W_1^*\big(\nu_n^0, \nu_n^0\big) - 6\varepsilon \\ &\geq W_1\big(\mu^1, \nu^1\big) + \frac{1}{2} W_1^*\big(\mu^0, \mu^0\big) + \frac{1}{2} W_1^*\big(\nu^0, \nu^0\big) - 10\varepsilon. \end{aligned} \tag{4.2.2}$$

Since $\varepsilon > 0$ was arbitrary, this proves the claim.

For the general case of real masses, one can approximate Borel measures by sums of Dirac measures (with rational masses) in the weak topology. By continuity of $\tilde{W}_1, W_1$ and W_1^* with respect to weak convergence, one can apply the rational case and go to the limit in (4.2.2). □

4.3 Induced Length Metric: Definitions

Here, we continue to assume that X is a length space.

In the last section we introduced the function W_p^0 as an attempt to define a metric on the space $\mathcal{P}^{sub}(Y)$. However, it does not satisfy the triangle inequality. To overcome this problem, we will pass to the induced metric given by the procedure in Lemma 2.1.2 and further to the induced length metric.

Corollary 4.3.1. *The function*

$$W_p^\flat(\mu,\nu) := \inf\left\{\sum_{i=1}^n W_p^0(\eta_{i-1},\eta_i)\,\middle|\, n\in\mathbb{N}, \eta_i\in\mathcal{P}_p^{sub}(Y), \eta_0=\mu, \eta_n=\nu\right\} \tag{4.3.1}$$

is a pseudo-metric on $\mathcal{P}_p^{sub}(Y)$, and it is the biggest pseudo-metric below W_p^0.

This however is only a means to an end, namely to define a *length* pseudo-metric.

Definition 4.3.2. i) Given a curve $(\eta_s)_{s\in[0,1]}\subset\mathcal{P}_p^{sub}(Y)$, we define its $W_p^\flat$-length by

$$L_p^\flat(\eta) := \sup\left\{\sum_{i=1}^n W_p^\flat(\eta_{s_{i-1}},\eta_{s_i})\,\middle|\, n\in\mathbb{N}, 0=s_0<\ldots<s_n=1\right\}.$$

ii) For two measures $\mu,\nu\in\mathcal{P}_p^{sub}(Y)$, the induced length pseudo-metric is now obtained by

$$W_p^\sharp(\mu,\nu) := \inf\left\{L_p^\flat(\eta)\,\middle|\,\eta\colon[0,1]\to\mathcal{P}_p^{sub}(Y)\ W_p^\flat\text{-continuous}, \eta_0=\mu, \eta_1=\nu\right\}. \tag{4.3.2}$$

It will be called *transportation-annihilation distance.*

Remark 4.3.3. Both, $W_p^\flat$ and $W_p^\sharp$ are a priori only pseudo-metrics; the former the biggest one below W_p^0, the latter the smallest *intrinsic* one above $W_p^\flat$. In what follows, it will turn out however that both indeed are metrics and for $p=1$ they coincide.

Lemma 4.3.4. *i) For $1\le p\le q<\infty$ and every $\mu,\nu\in\mathcal{P}^{sub}(Y)$:*

$$W_p^0(\mu,\nu)\le W_q^0(\mu,\nu).$$

The same is true for the distances $W_p^\flat, W_p^\sharp$.

ii) If X is bounded, we additionally have that for $1\le p\le q<\infty$ and every $\mu,\nu\in\mathcal{P}^{sub}(Y)$:

$$W_q^0(\mu,\nu)\le W_p^0(\mu,\nu)^{\frac{p}{q}}\operatorname{diam}(X)^{\frac{q-p}{p}}.$$

The same is true for the distances $W_p^\flat, W_p^\sharp$.

Proof. i) It is a consequence of the same inequality for $\tilde{W}_p$ which in turn is true because it is (by our identification with $\hat{W}_p$) a Kantorovich-Wasserstein metric, for which this inequality is but an application of Hölder's inequality. Indeed, for $1 \leq p \leq q$, Hölder's inequality gives both, $\tilde{\mathcal{P}}_q(Y|X) \subset \tilde{\mathcal{P}}_p(Y|X)$ and

$$\tilde{W}_p(\sigma,\tau) \leq \tilde{W}_q(\sigma,\tau)$$

for every $\sigma,\tau \in \tilde{\mathcal{P}}_q(Y|X)$. Hence, it follows that the same is true when taking infima, so that

$$W_p^0(\mu,\nu) = \inf_{\substack{\sigma,\tau\in\tilde{\mathcal{P}}_p(Y|X)\\ \sigma^0=\mu,\tau^0=\nu}} \tilde{W}_p(\sigma,\tau) \leq \inf_{\substack{\sigma,\tau\in\tilde{\mathcal{P}}_q(Y|X)\\ \sigma^0=\mu,\tau^0=\nu}} \tilde{W}_q(\sigma,\tau) = W_q^0(\mu,\nu).$$

Now by the same reasoning, we see that also $W_p^\flat(\mu,\nu) \leq W_q^\flat(\mu,\nu)$, because

$$\sum_{i=1}^n W_p^0(\eta_{i-1},\eta_i) \leq \sum_{i=1}^n W_q^0(\eta_{i-1},\eta_i)$$

and $\mathcal{P}_q^{sub}(Y) \subset \mathcal{P}_p^{sub}(Y)$. Finally, the case $W_p^\sharp$ follows again in the same way.

ii) This is again true due to the same result for Kantorovich-Wasserstein metrics. In the case of bounded spaces, all the spaces $\tilde{\mathcal{P}}_p(Y|X)$ for every possible p coincide, and so do the $\mathcal{P}_p^{sub}(Y)$'s, so that we can argue in the same way as in part i), given the starting point

$$\tilde{W}_q \leq \tilde{W}_p^{\frac{p}{q}} \operatorname{diam}(X)^{\frac{q-p}{q}}$$

for $1 \leq p \leq q$. □

4.4 Induced Length Metric: Comparison

Also here we continue to assume that X is a length space.

Just like the identification of $\tilde{W}_p$ with the Kantorovich-Wasserstein metric $\hat{W}_p$ on the doubled space gave us immediately some properties about the space of charged probabilities, it will now be convenient to compare the previously defined metrics $W_p^\flat$ and $W_p^\sharp$ to a Kantorovich-Wasserstein metric on the *one-point completion* we introduced in Subsection 2.2. We will do so by providing similar representation formulas as the one in Lemma 4.2.6 for W_p^0. To state them, let us introduce some more auxiliary transport cost functions.

Definition 4.4.1. i) W_p' will denote the L^p-Kantorovich-Wasserstein distance on $\mathcal{P}_p(Y')$ induced by the distance d'.

ii) Extending each subprobability measure $\mu \in \mathcal{P}_p^{sub}(Y)$ to a probability measure $\mu' \in \mathcal{P}_p(Y')$ by $\mu' := \mu + (1-\mu(Y))\delta_\partial$ induces a bijection between $\mathcal{P}_p^{sub}(Y)$ and $\mathcal{P}_p(Y')$ (see the lemma below). The induced distance on $\mathcal{P}_p^{sub}(Y)$ will again be denoted by W_p', i.e.

$$W_p'(\mu,\nu) := W_p'(\mu',\nu').$$

iii) For subprobability measures μ, ν with equal mass $\mu(Y) = \nu(Y)$ we will also make use of the transportation cost

$$W_p^\dagger(\mu,\nu)^p := \inf_{q \in \mathrm{Cpl}(\mu,\nu)} \int_{Y\times Y} d^\dagger(x,y)^p \, \mathrm{d}q(x,y) \tag{4.4.1}$$

induced by $d^\dagger$ (which was defined in (2.2.1)).

iv) For a subprobability $\mu \in \mathcal{P}^{sub}(Y)$ define

$$W_p'(\mu,0)^p := W_p'(\mu,\delta_\partial)^p = \int_Y d'(x,\partial)^p \, \mathrm{d}\mu(x), \tag{4.4.2}$$

with 0 denoting the subprobability measure with vanishing total mass.

Lemma 4.4.2. *i) The map*

$$\mathcal{P}_p^{sub}(Y) \to \mathcal{P}_p(Y'), \quad \mu \mapsto \mu' := \mu + (1-\mu(Y))\delta_\partial$$

is a bijection, and it is an isometry when we equip both spaces with W_p'.

ii) If Y is additionally totally bounded, W_p' metrizes the vague convergence in $\mathcal{P}_p^{sub}(Y)$.

Proof. i) The map is clearly a bijection with inverse $\mathcal{P}_p(Y') \ni \mu' \mapsto \mu := \mu'|_Y \in \mathcal{P}_p^{sub}(Y)$. By definition of W_p' on $\mathcal{P}_p^{sub}(Y)$ it is an isometry.

ii) Let us show that W_p' metrizes the vague convergence in $\mathcal{P}_p^{sub}(Y)$. Given a vaguely converging sequence $\mu_n \to \mu_*$ in $\mathcal{P}_p^{sub}(Y)$, define $\mu_n' := \mu_n + (1-\mu_n(Y))\delta_\partial \in \mathcal{P}_p(Y')$. This is a sequence of probability measures on a compact space; hence, for every subsequence, Prokhorov's theorem provides a converging further subsequence. Since the restriction of all these limits to Y has to coincide with μ_*, the whole sequence μ_n' converges weakly to $\mu_*' := \mu_* + (1-\mu_*(Y))\delta_\partial$, so that $W_p'(\mu_n', \mu_*') \to 0$. Then also $W_p'(\mu_n, \mu_*) \to 0$.

Assume conversely that $W_p'(\mu_n, \mu_*) \to 0$. By definition this means that we have convergence $W_p'(\mu_n', \mu_*') \to 0$, which in turn assures that $\mu_n' \to \mu_*'$ weakly in Y'. Then the restrictions to Y converge vaguely. □

Remark 4.4.3. Note that without the assumption of total boundedness the vague convergence in Y would *not* imply the weak convergence of the corresponding probability measures on Y' since they could lose mass at infinity instead of at the boundary.

Remark 4.4.4. One could equally well define

$$W_p''(\mu,\nu) := \inf\{W_p'(\check{\mu},\check{\nu}) \,|\, \check{\mu},\check{\nu} \in \mathcal{M}(Y'), \check{\mu}|_Y = \mu, \check{\nu}|_Y = \nu\}.$$

For $p = 1$ the metrics W_1' and W_1'' coincide, but for $p > 1$ this is no longer true. This is due to the fact that $(d(x,\partial) + d(y,\partial))^p = d(x,\partial)^p + d(y,\partial)^p$ only for $p = 1$. Intuitively speaking, for $p > 1$, it makes a difference if we transport mass *through* the boundary point, or *to* it – however, for the latter we need to allow for masses bigger than 1. Take for instance $X = \mathbb{R}$, $Y = (-3,3)$ and $\mu = \delta_{-2}, \nu = \delta_2$. Then $\mu' = \mu$ and $\nu' = \nu$, so that $W_p'(\mu,\nu)^p = d'(-2,2)^p = 2^p$, whereas $W_p''(\mu,\nu)^p \le W_p'(\mu+\delta_\partial, \nu+\delta_\partial)^p = d'(-2,\partial)^p + d'(2,\partial)^p = 2$.
The metric W_2'' coincides with Figalli & Gigli's metric Wb_2 [FG10].

We start by characterizing the metric W'_p in terms of L^p-transportation and -annihilation costs.

Lemma 4.4.5. *For all $\mu, \nu \in \mathcal{P}_p^{sub}(Y)$*

$$W'_p(\mu,\nu)^p = \inf\Big\{ W_p(\mu_1,\nu_1)^p + W_p^\dagger(\mu_2,\nu_2)^p + W'_p(\mu_0,0)^p + W'_p(\nu_0,0)^p \;\Big|\; \mu = \mu_1+\mu_2+\mu_0,\ \nu = \nu_1+\nu_2+\nu_0,\ (\mu+\nu_0)(Y) \le 1,\ (\nu+\mu_0)(Y) \le 1 \Big\}. \quad (4.4.3)$$

In the case $p = 1$, contributions from the term $W_p^\dagger(\mu_2,\nu_2)^p$ can be avoided, in other words, one can always choose $\mu_2 = \nu_2 = 0$.

Proof. The derivation of this formula is straightforward. The transport decomposes into trivial transports within ∂ (which do not appear in the formula), transports between Y and ∂ (given by $W'_p(\mu_0,0)^p + W'_p(\nu_0,0)^p$), and transports within Y, and the latter ones into transports using d and $d^\dagger$ (given by $W_p(\mu_1,\nu_1)^p + W_p^\dagger(\mu_2,\nu_2)^p$). One can construct these decompositions more explicitly like in the proof of Lemma 4.2.5. The resulting couplings are still optimal between their marginals. The inequalities in the constraints are due to the fact that we compose the *probability* measures μ', ν' instead of the subprobabilities μ, ν and the trivial transport within ∂ can be omitted.

For the vanishing of the $W_p^\dagger$-term note that in the case $p = 1$ one has $[d'(x,\partial) + d'(x,\partial)]^p = d'(x,\partial)^p + d'(x,\partial)^p$, meaning that the term can be absorbed in the annihilation terms $W'_p(\mu_0,0)^p + W'_p(\nu_0,0)^p$. ☐

The following lemma discusses the connection between our two *annihilation costs* W'_p and W^*_p.

Lemma 4.4.6. *i) For all $\mu, \nu \in \mathcal{P}_1(Y)$*

$$W_1^*(\mu,\nu) = \inf\Big\{ W_1(\mu,\xi) + W_1(\xi,\nu) \;\Big|\; \xi \in \mathcal{P}(\partial Y) \Big\}.$$

ii) For all $p \ge 1$ and all $\mu \in \mathcal{P}_p(Y)$

$$W'_p(\mu,0) = \inf\big\{ W_p(\mu,\xi) \;\big|\; \xi \in \mathcal{P}(\partial Y) \big\}.$$

iii) For all $p \ge 1$ and all $\mu \in \mathcal{P}_p(Y)$

$$2^{-1+1/p}\, W'_p(\mu,0) \le W^*_p(\mu) \le W'_p(\mu,0).$$

In particular, $W_1^*(\mu) = W'_1(\mu,0)$.

Proof. i) By triangle inequality, we have that for every $\xi \in \mathcal{P}(\partial Y)$

$$W_1^*(\mu,\nu) \le W_1(\mu,\xi) + W_1(\xi,\nu).$$

Making use of Lemma 4.1.5, we consider the measures as given on the different copies, $\mu \in \mathcal{P}(Y^+)$ and $\nu \in \mathcal{P}(Y^-)$. Take now a $\hat{W}_1$-optimal coupling $q \in \mathrm{Cpl}(\mu, \nu) \subset \mathcal{P}(Y^+ \times Y^-)$. Let $\varepsilon > 0$ and

$$G_\varepsilon(x, y) := \{\gamma \in C^0([0,1], \hat{X}) \,\big|\, \gamma_0 = x, \gamma_1 = y, |L(\gamma) - d(x, y)| \leq \varepsilon\}$$

be the set of ε-geodesics in $\hat{X}$ connecting x and y. Given a curve γ in $\hat{X}$ with $\gamma_0 \in Y^+$ and $\gamma_1 \in Y^-$, define $\alpha(\gamma) := \inf\{s > 0 \,|\, \gamma_s \notin Y^+\}$ and $z(\gamma) := \gamma_{\alpha(\gamma)}$. Then $z(\gamma) \in \partial Y$, and given a measurable selection $\Gamma_\varepsilon \colon \hat{X} \times \hat{X} \to C^0([0,1], \hat{X})$ with $\Gamma_\varepsilon(x, y) \in G_\varepsilon(x, y)$ (which exists by our measurable selection Lemma 2.5.6), we define the "boundary crossing points" $\mathcal{Z} := z \circ \Gamma_\varepsilon \colon Y^+ \times Y^- \to \partial Y$. Using the projection $\mathrm{pr}_1 \colon \hat{X} \times \hat{X} \to \hat{X}, (x, y) \mapsto x$, we get a map

$$(\mathrm{pr}_1, \mathcal{Z}) \colon Y^+ \times Y^- \to Y^+ \times \partial Y$$

and define the push-forward measure $Q_1 := (\mathrm{pr}_1, \mathcal{Z})_\# q \in \mathcal{P}(Y^+ \times \partial Y)$.

Let us check that this is a coupling between μ and $\xi := \mathcal{Z}_\# q \in \mathcal{P}(\partial Y)$: Given a measurable set $A \subset Y^+$,

$$\begin{aligned}(\mathrm{pr}_1, \mathcal{Z})^{-1}(A \times \partial Y) =& \{(x, y) \in Y^+ \times Y^- \,\big|\, (\mathrm{pr}_1(x, y), \mathcal{Z}(x, y)) \in A \times \partial Y\} \\ =& \{(x, y) \in Y^+ \times Y^- \,\big|\, (x, \mathcal{Z}(x, y)) \in A \times \partial Y\} \\ =& A \times Y^-,\end{aligned}$$

which yields that $Q_1(A \times \partial Y) = q(A \times Y^-) = \mu(A)$. On the other hand, given $B \subset \partial Y$ measurable,

$$\begin{aligned}(\mathrm{pr}_1, \mathcal{Z})^{-1}(Y^+ \times B) =& \{(x, y) \in Y^+ \times Y^- \,\big|\, (\mathrm{pr}_1(x, y), \mathcal{Z}(x, y)) \in Y^+ \times B\} \\ =& \{(x, y) \in Y^+ \times Y^- \,\big|\, (x, \mathcal{Z}(x, y)) \in Y^+ \times B\} \\ =& (Y^+ \times Y^-) \cap \mathcal{Z}^{-1}(B) \\ =& \mathcal{Z}^{-1}(B),\end{aligned}$$

hence in this case we have $Q_1(Y^+ \times B) = q(\mathcal{Z}^{-1}(B)) = \xi(B)$. Analogously one sees that $Q_2 := (\mathcal{Z}, \mathrm{pr}_2)_\# q$ as a coupling between ξ and ν.

Now what is left to prove is that ξ is an "almost-midpoint". Since for $y \in \partial Y$ we have $d(x, y) = d^*(x, y)$, together with Lemma 2.1.7 we get that

$$\begin{aligned}W_1(\mu, \xi) &\leq \int_{Y^+ \times \partial Y} d(x, y) \,\mathrm{d}Q_1(x, y) \\ &= \int_{Y^+ \times Y^-} d^*(\mathrm{pr}_1(x, y), \mathcal{Z}(x, y)) \,\mathrm{d}q(x, y) \\ &= \int_{Y^+ \times Y^-} d^*(x, \Gamma_\varepsilon(x, y)_t|_{t = \alpha(\Gamma_\varepsilon(x, y))}) \,\mathrm{d}q(x, y) \\ &\leq \int_{Y^+ \times Y^-} \alpha(\Gamma_\varepsilon(x, y)) d^*(x, y) + \varepsilon \,\mathrm{d}q(x, y).\end{aligned}$$

Using in the same way Q_2 as a coupling between ξ and ν, we finally get that

$$\begin{aligned}W_1(\mu,\xi)+W_1(\xi,\nu)\leq&\int_{Y^+\times Y^-}\alpha(\Gamma(x,y))d^*(x,y)+\varepsilon\,\mathrm{d}q(x,y)\\&+\int_{Y^+\times Y^-}(1-\alpha(\Gamma(x,y)))d^*(x,y)+\varepsilon\,\mathrm{d}q(x,y)\\=&\int_{Y^+\times Y^-}d^*(x,y)+2\varepsilon\,\mathrm{d}q(x,y)\\=&W_1^*(\mu,\nu)+2\varepsilon.\end{aligned}$$

ii) Given an arbitrary $\xi\in\mathcal{P}(\partial Y)$ and a W_p-optimal coupling $q\in\mathrm{Cpl}(\mu,\xi)$, we get

$$W_p(\mu,\xi)^p=\int_{X\times X}d(x,y)^p\,\mathrm{d}q(x,y)\geq\int_{X\times X}d'(x,\partial)^p\,\mathrm{d}q(x,y)=W_p'(\mu,0)^p.$$

For the other inequality, similarly as in part i), we will define a map and use its push-forward measure. Given $x\in Y$, let

$$G_\varepsilon(x):=\Big\{z\in\partial Y\;\Big|\;|d(x,z)-d'(x,\partial)|\leq\varepsilon\Big\}$$

be the boundary points closest to x. Let us show that the graph of the multivalued map G_ε is closed: Take a sequence (x_n,z_n) with $z_n\in G_\varepsilon(x)$ that converges to (x,z) in $X\times X$. Then $z\in\partial Y$ by the closedness of the boundary, and

$$|d(x,z)-d'(x,\partial)|=\lim_{n\to\infty}|d(x_n,z_n)-d'(x_n,\partial)|\leq\varepsilon,$$

hence $z\in G_\varepsilon(x)$. Thus we can apply the measurable selection Theorem 2.5.4 and get a measurable function $\Phi_\varepsilon\colon Y\to\partial Y$ such that $\Phi_\varepsilon(x)\in G_\varepsilon(x)$. Then for the measure $\xi_\varepsilon:=(\Phi_\varepsilon)_\#\mu$ we see that

$$W_p(\mu,\xi_\varepsilon)^p\leq\int_X d(x,\Phi_\varepsilon(x))^p\,\mathrm{d}\mu(x)\leq\int_X\big(d'(x,\partial)+\varepsilon\big)^p\,\mathrm{d}\mu(x).$$

First of all observe that the moment bound of μ implies that also the d'-moment $\int d'(\cdot,\partial)^p\,\mathrm{d}\mu$ is finite. Since μ is a probability measure, constant functions are integrable, thus also the sum $d'(\cdot,\partial)+\varepsilon$ is in $L^p(\mu)$. This sum converges pointwise to $d'(\cdot,\partial)$ as $\varepsilon\to0$, and it is dominated by $d'(\cdot,\partial)+1\in L^p(\mu)$. By the dominated convergence theorem we get convergence in $L^p(\mu)$ as $\varepsilon\to0$, i.e.

$$\lim_{\varepsilon\to0}W_p(\mu,\xi_\varepsilon)^p\leq\int_X d'(x,\partial)^p\,\mathrm{d}\mu(x)=W_p'(\mu,0).$$

iii) The triangle inequality for d^* implies that $W_p(\mu,\xi)+W_p(\xi,\mu)\geq W_p^*(\mu,\mu)$ for all $\xi\in\mathcal{P}(\partial Y)$. Thus $W_p'(\mu,0)\geq\frac{1}{2}W_p^*(\mu,\mu)=W_p^*(\mu)$. An estimate in the other direction is obtained as follows

$$W_p^*(\mu)^p=2^{-p}W_p^*(\mu,\mu)^p=2^{-p}\int_{X\times X}\left(\inf_{z\in X\setminus Y}\big(d(x,z)+d(z,y)\big)\right)^p\mathrm{d}q(x,y)$$

$$\begin{aligned}
&\geq 2^{-p}\int_{X\times X}\left(\inf_{z\in X\setminus Y} d(x,z)+\inf_{w\in X\setminus Y} d(w,y)\right)^p \mathrm{d}q(x,y)\\
&\geq 2^{1-p}\int_{X\times X}\left(\inf_{z\in X\setminus Y} d(x,z)\right)^p \mathrm{d}q(x,y)\\
&= 2^{1-p}W_p'(\mu,0)^p,
\end{aligned}$$

where q denotes any W_p^*-optimal coupling of μ and μ. □

Remark 4.4.7. In general, $W_p^*(\mu)$ and $W_p'(\mu,0)$ will not coincide. Our lower bound for $W_p^*(\mu)/W_p'(\mu,0)$ is sharp. For instance, let $Y=(0,2)\subset X=\mathbb{R}$ and $\mu=\frac{1}{2}(\delta_1+\delta_\varepsilon)$ for some $\varepsilon\in(0,1)$. Then $W_p'(\mu,0)^p=\frac{1}{2}(d'(1,\partial)^p+d'(\varepsilon,\partial)^p)=\frac{1}{2}(1+\varepsilon^p)$ whereas $W_p^*(\mu)^p=\left(\frac{1+\varepsilon}{2}\right)^p$. Thus

$$\frac{W_p^*(\mu)}{W_p'(\mu,0)}=2^{-1+\frac{1}{p}}\frac{1+\varepsilon}{(1+\varepsilon^p)^{\frac{1}{p}}}\longrightarrow 2^{-1+\frac{1}{p}}\quad\text{as } \varepsilon\to 0.$$

Now we are in position to compare $W_p^\flat, W_p^\sharp, W_p'$. It turns out that for $p=1$ they all coincide.

Theorem 4.4.8. *i) For all $\mu,\nu\in\mathcal{P}_1^{sub}(Y)$*

$$W_1^\flat(\mu,\nu)=W_1^\sharp(\mu,\nu)=W_1'(\mu,\nu).$$

ii) More generally, for all $p\geq 1$ and all $\mu,\nu\in\mathcal{P}_p^{sub}(Y)$

$$W_1'(\mu,\nu)\leq W_p^\flat(\mu,\nu)\leq W_p^\sharp(\mu,\nu)\leq W_p'(\mu,\nu).$$

In particular, $W_p^\flat, W_p^\sharp, W_p^0$ do not vanish outside the diagonal.

Proof. i) According to Lemma 4.4.5 and Lemma 4.4.6,

$$W_1'(\mu,\nu)=\inf\left\{W_1(\mu_1,\nu_1)+W_1^*(\mu_0)+W_1^*(\nu_0)\;\middle|\;\mu=\mu_1+\mu_0,\nu=\nu_1+\nu_0\right\}\tag{4.4.4}$$

for all subprobability measures $\mu,\nu\in\mathcal{P}_1^{sub}(Y)$. Together with Lemma 4.2.6i) this implies $W_1'(\mu,\nu)\leq W_1^0(\mu,\nu)$. As $W_1^\flat$ is the biggest metric below W_1^0, we have $W_1'\leq W_1^\flat$. Using the fact that W_1' is a *length* metric, this yields

$$\begin{aligned}
W_1^\sharp(\mu,\nu)&=\inf_{\substack{\eta:\,\mu\rightsquigarrow\nu\\ W_1^\flat\text{-cont.}}}\ \sup_{0=s_0<\ldots<s_n=1}\sum_{i=1}^n W_1^\flat(\eta_{s_{i-1}},\eta_{s_i})\\
&\geq\inf_{\substack{\eta:\,\mu\rightsquigarrow\nu\\ W_1^\flat\text{-cont.}}}\ \sup_{0=s_0<\ldots<s_n=1}\sum_{i=1}^n W_1'(\eta_{s_{i-1}},\eta_{s_i})\\
&\geq\inf_{\substack{\eta:\,\mu\rightsquigarrow\nu\\ W_1'\text{-cont.}}}\ \sup_{0=s_0<\ldots<s_n=1}\sum_{i=1}^n W_1'(\eta_{s_{i-1}},\eta_{s_i})=W_1'(\mu,\nu).
\end{aligned}$$

Since $W_1^\sharp$ is the length metric induced by $W_1^\flat$, one gets $W_1^\flat \le W_1^\sharp$.

Now we are going to show that $W_1^\sharp \le W_1'$. To do so, starting from an almost-geodesic in the representation of W_1' given by (4.4.4) we define a new curve connecting μ and ν and estimate its $W_1^\flat$-length by using a clever decomposition in the representation formula for W_1^0 given by Lemma 4.2.6ii).
Let $\varepsilon > 0$ and take a decomposition $\mu = \mu_1 + \mu_0$, $\nu = \nu_1 + \nu_0$ in (4.4.4) such that

$$W_1'(\mu,\nu) + \varepsilon \ge W_1(\mu_1,\nu_1) + W_1^*(\mu_0) + W_1^*(\nu_0).$$

Then we take an ε-W_1-geodesic $(\eta_{s,1})_{s\in[0,1]}$ connecting μ_1 and ν_1 that is supported on ε-geodesics in Y. Define

$$\tilde\eta'_{s,0} := \begin{cases} (1-2s)\mu_0 + 2s\mu_0(Y')\,\delta_\partial, & s\in[0,\frac12] \\ (2s-1)\nu_0 + 2(1-s)\mu_0(Y')\,\delta_\partial, & s\in(\frac12,1]. \end{cases}$$

This is a curve connecting μ_0 and ν_0. Take the restriction $\tilde\eta_{s,0} := \tilde\eta'_{s,0}|_Y$ and define

$$\tilde\eta_s := \eta_{s,1} + \tilde\eta_{s,0},$$

which is a curve connecting μ and ν. To estimate the $W_1^\flat$-length of the restricted curve $\tilde\eta_s$ it is useful to get a bound on $W_1^0(\tilde\eta_s,\tilde\eta_t)$. Consider the case $0\le s\le t\le \frac12$: We can rewrite

$$\tilde\eta_s = \underbrace{(\eta_{s,1}+\tilde\eta'_{t,0})}_{\text{“}\mu_1\text{”in Lemma 4.2.6}} + \underbrace{2(t-s)\mu_0}_{\text{“}\mu_0\text{”}} \quad\text{and}\quad \tilde\eta_t = \underbrace{(\eta_{t,1}+\tilde\eta'_{t,0})}_{\text{“}\nu_1\text{”}} + \underbrace{0}_{\text{“}\nu_0\text{”}}.$$

This is an admissible decomposition in Lemma 4.2.6ii) as for instance

$$\begin{aligned} \text{“}(\mu+\nu_0)(X)\text{”} = (\eta_{s,1}+\tilde\eta_{s,0}+0)\,(X) &= (\eta_{s,1}+(1-2s)\mu_0)(X) \\ &\le (\eta_{0,1}+\mu_0)(X) \le \mu(X) \le 1, \end{aligned}$$

and similarly for “$(\nu+\mu_0)(X)$”. Thus, in this case the representation given by Lemma 4.2.6ii) yields

$$\begin{aligned} W_1^0(\tilde\eta_s,\tilde\eta_t) &\le W_1\left(\eta_{s,1}+\tilde\eta_{t,0},\eta_{t,1}+\tilde\eta_{t,0}\right) + W_1^*(2(t-s)\mu_0) \\ &= W_1(\eta_{s,1},\eta_{t,1}) + W_1^*(2(t-s)\mu_0) \\ &= |t-s|W_1(\eta_{0,1},\eta_{1,1}) + |t-s|\varepsilon + 2|t-s|W_1^*(\mu_0), \end{aligned}$$

where we made use of the translation invariance of the Kantorovich-Wasserstein metric for $p=1$, see (2.5.3).

In the case $\frac12 \le s\le t\le 1$ we analogously rewrite

$$\tilde\eta_s = \underbrace{(\eta_{s,1}+\tilde\eta_{s,0})}_{\text{“}\mu_1\text{”}} + \underbrace{0}_{\text{“}\mu_0\text{”}} \quad\text{and}\quad \tilde\eta_t = \underbrace{(\eta_{t,1}+\tilde\eta_{s,0})}_{\text{“}\nu_1\text{”}} + \underbrace{2(t-s)\nu_0}_{\text{“}\nu_0\text{”}},$$

and end up with

$$W_1^0(\tilde\eta_s,\tilde\eta_t) \le |t-s|W_1(\eta_{0,1},\eta_{1,1}) + |t-s|\varepsilon + 2|t-s|W_1^*(\nu_0).$$

To compute the length of the curve $\tilde{\eta}_s$, we enforce the partitions to visit the time step $\frac{1}{2}$, and then use the above estimates for W_1^0:

$$
\begin{aligned}
L_1^\flat(\tilde{\eta}) &= \sup\left\{ \sum_{k=0}^{n} W_1^\flat(\tilde{\eta}_{s_{i-1}}, \tilde{\eta}_{s_i}) \,\middle|\, n \in \mathbb{N}, 0 = s_0 < \cdots <_n = 1 \right\} \\
&= \sup\left\{ \sum_{k=0}^{n} W_1^\flat(\tilde{\eta}_{s_{i-1}}, \tilde{\eta}_{s_i}) \,\middle|\, n \in \mathbb{N}, 0 = s_0 < s_{i^*} = \frac{1}{2} < \cdots <_n = 1 \right\} \\
&\le \sup\left\{ \sum_{k=0}^{n} W_1^0(\tilde{\eta}_{s_{i-1}}, \tilde{\eta}_{s_i}) \,\middle|\, n \in \mathbb{N}, 0 = s_0 < s_{i^*} = \frac{1}{2} < \cdots <_n = 1 \right\} \\
&\le \sup\Big\{ \sum_{s_i \le \frac{1}{2}} |s_i - s_{i-1}| W_1(\eta_{0,1}, \eta_{1,1}) + |s_i - s_{i-1}|\varepsilon + 2|s_i - s_{i-1}| W_1^*(\mu_0) \\
&\qquad + \sum_{s_i \ge \frac{1}{2}} |s_i - s_{i-1}| W_1(\eta_{0,1}, \eta_{1,1}) + |s_i - s_{i-s}|\varepsilon + 2|s_i - s_{i-1}| W_1^*(\nu_0) \Big\} \\
&= \frac{1}{2} W_1(\eta_{0,1}, \eta_{1,1}) + \frac{1}{2}\varepsilon + W_1^*(\mu_0) + \frac{1}{2} W_1(\eta_{0,1}, \eta_{1,1}) + \frac{1}{2}\varepsilon + W_1^*(\nu_0) \\
&= W_1(\mu_1, \nu_1) + \varepsilon + W_1^*(\mu_0) + W_1^*(\nu_0) \\
&\le W_1'(\mu, \nu) + 2\varepsilon.
\end{aligned}
$$

This finally yields $W_1^\sharp(\mu,\nu) \le L_1^\flat(\tilde{\eta}) \le W_1'(\mu,\nu) + 2\varepsilon$. Since ε was arbitrary, this proves $W_1^\sharp \le W_1'$.
By the fact that $W_1^\flat$ is the biggest metric below W_1^0 and we now know that $W_1^\sharp = W_1' \le W_1^0$, we also get $W_1^\flat \ge W_1^\sharp$.

ii) Thanks to i) and Lemma 4.3.4 we know that $W_1' = W_1^\flat \le W_p^\flat$. Further, since $W_p^\sharp$ is the length metric induced by $W_p^\flat$ we also have $W_p^\flat \le W_p^\sharp$. Hence the only thing left to show is that $W_p^\sharp \le W_p'$.

The idea to do so is that locally (along a geodesic) the contribution of $W_p^\dagger$ is negligible, so that we can compare W_p' and $W_p^\flat$ on a small scale and then carry it over to the induced length metrics.
Let subprobabilities μ, ν be given as well as an ε-W_p'-geodesic $(\eta_t')_{t\in[0,1]}$ connecting the measures $\mu' := \mu + (1-\mu(Y))\delta_\partial$ and $\nu' := \nu + (1-\nu(Y))\delta_\partial$. By the continuity of W_p' and W_p^* with respect to weak convergence we can assume without loss of generality that μ and ν have compact supports and for $\alpha > 0$ small

$$\eta_t(Y) \le 1 - \alpha$$

for all $t \in (0,1)$. Again we use the notation that measures without primes are the restrictions to Y. We thus have $\eta_t(\partial) = 0$, whereas $\eta_t'(\partial) \ge \alpha$. Choose $\delta > 0$ such that $\eta_t(B_\delta'(\partial)) \le \frac{\alpha}{2}$. Let Π be the probability measure on $C^0([0,1], Y')$ supported on ε-geodesics such that $\eta_t' = (\mathrm{e}_t)_\#\Pi$ and denote by L the essential supremum of $d'(\gamma_0, \gamma_1)$ under Π (which is finite thanks to the compact supports of μ and ν). Let $\delta' := \frac{\delta}{L}$.

We consider η_s and η_t for $|s-t| \leq \delta'$. Using that $d^\dagger(x,y)^p \geq d'(x,\partial)^p + d'(y,\partial)^p$, we see that in the decomposition (4.4.3) it is actually cheaper to annihilate mass at the boundary:

$$\begin{aligned} W_p'(\eta_s,\eta_t)^p = &\inf\Big\{W_p(\eta_{s,1},\eta_{t,1})^p + W_p^\dagger(\eta_{s,2},\eta_{t,2})^p + W_p'(\eta_{s,0},0)^p + W_p'(\eta_{t,0},0)^p \,\Big|\\ &\qquad \eta_s = \eta_{s,1}+\eta_{s,2}+\eta_{s,0}, \eta_t = \eta_{t,1}+\eta_{t,2}+\eta_{t,0},\\ &\qquad (\eta_s+\eta_{t,0})(Y)\leq 1, (\eta_t+\eta_{s,0})(Y)\leq 1\Big\}\\ \geq &\inf\Big\{W_p(\eta_{s,1},\eta_{t,1})^p + W_p'(\eta_{s,0}+\eta_{s,2},0)^p + W_p'(\eta_{t,0}+\eta_{t,2},0)^p \,\Big|\\ &\qquad \eta_s = \eta_{s,1}+\eta_{s,2}+\eta_{s,0}, \eta_t = \eta_{t,1}+\eta_{t,2}+\eta_{t,0},\\ &\qquad (\eta_s+\eta_{t,0})(Y)\leq 1, (\eta_t+\eta_{s,0})(Y)\leq 1\Big\}. \end{aligned}$$

Since $W^\dagger$ only occurs where $d^\dagger$ is smaller than d, its contribution comes from ε-geodesics in $B_\delta'(\partial)$, so that by our choice of δ we know that $\eta_{s,2}(Y) = \eta_{s,2}(B_\delta'(\partial)) \leq \frac{\alpha}{2}$ and the same for $\eta_{t,2}$. Hence for α small enough we have $(\eta_s + (\eta_{t,2}+\eta_{t,0}))(Y) \leq 1$, so that $\eta_s = \eta_{s,1} + \tilde\eta_{s,0}$ with $\tilde\eta_{s,0} := \eta_{s,0} + \eta_{s,2}$ is an admissible decomposition in (4.4.3). In particular, the above inequality is an *equality.* Note that we cannot use this trick for $s=0, t=1$ because then the constraint might not be satisfied. Thanks to Lemma 4.4.6 we thus have

$$\begin{aligned} W_p'(\eta_s,\eta_t)^p \geq &\inf\Big\{W_p(\eta_{s,1},\eta_{t,1})^p + W_p^*(\tilde\eta_{s,0})^p + W_p^*(\tilde\eta_{t,0})^p \,\Big|\\ &\qquad \eta_s = \eta_{s,1}+\tilde\eta_{s,0}, \eta_t = \eta_{t,1}+\tilde\eta_{t,0}, (\eta_s+\tilde\eta_{t,0})(Y)\leq 1,\\ &\qquad (\eta_t+\tilde\eta_{s,0})(Y)\leq 1\Big\}\\ \geq\ & W_p^0(\eta_s,\eta_t)^p \geq W_p^\flat(\eta_s,\eta_t)^p. \end{aligned}$$

Hence, the W_p'-length of the curve $(\eta_t)_{t\in[0,1]}$ dominates its $W_p^\flat$-length. As this curve is an almost-geodesic for W_p', going to the induced length metrics this finally proves

$$W_p'(\mu,\nu) + \varepsilon \geq W_p^\sharp(\mu,\nu).$$

Since ε was arbitrary, the proof is finished. □

Remark 4.4.9. As we have seen in the proof of part i), $W_1' \leq W_1^0$, and in particular W_1^0 does not vanish outside the diagonal.

Let us give some simple examples illustrating Theorem 4.4.8.

Example 4.4.10. Let $X = \mathbb{R}, Y = (-1,1), \mu = \delta_x, \nu = \delta_y$ for $x, y \in Y$. Then

$$W_p^0(\mu,\nu) = W_p(\mu,\nu) = |x-y|,$$

and for every $p \geq 1$

$$W_p'(\mu,\nu) = d'(x,y) = \min\{|x-y|, 2-|x-y|\}. \tag{4.4.5}$$

Hence, by the independence on p on the right-hand side of (4.4.5), Theorem 4.4.8 yields

$$W_p^\flat(\mu,\nu) = W_p^\sharp(\mu,\nu) = \min\{|x-y|, 2-|x-y|\}.$$

Example 4.4.11. Let $X = \mathbb{R}, Y = (-2,2), \mu = \frac{1}{2n+1}\delta_{-1/2}, \nu = \frac{1}{2n+1}\delta_{+1/2}$ for $n \in \mathbb{N}$. Then

$$W_p'(\mu,\nu)^p = W_p(\mu,\nu)^p = \frac{1}{2n+1}.$$

Taking

$$\sigma := \left(\frac{1}{2n+1}\sum_{k=0}^{n}\delta_{\frac{2k}{2n+1}-\frac{1}{2}}, \frac{1}{2n+1}\sum_{k=1}^{n}\delta_{\frac{2k}{2n+1}-\frac{1}{2}}\right)$$

and

$$\tau := \left(\frac{1}{2n+1}\sum_{k=0}^{n}\delta_{\frac{2k+1}{2n+1}-\frac{1}{2}}, \frac{1}{2n+1}\sum_{k=0}^{n-1}\delta_{\frac{2k+1}{2n+1}-\frac{1}{2}}\right),$$

we see that

$$W_p^0(\mu,\nu)^p \leq \tilde{W}_p(\sigma,\tau)^p = \left(\frac{1}{2n+1}\right)^p,$$

so that

$$W_p^\flat(\mu,\nu) \leq W_p^0(\mu,\nu) \leq \left(\frac{1}{2n+1}\right) < \left(\frac{1}{2n+1}\right)^{\frac{1}{p}} = W_p'(\mu,\nu),$$

for $p > 1$, $n \geq 1$. In particular, the lower estimate for $W_p^\flat$ in assertion ii) of the previous Theorem is sharp.

Lemma 4.4.12. *For all $\mu,\nu \in \mathcal{P}_1^{sub}(Y)$*

$$W_1^\sharp(\mu,\nu) = \inf\left\{W_1(\mu_1,\nu_1) + W_1^*(\mu_0) + W_1^*(\nu_0) \,\middle|\, \mu = \mu_1 + \mu_0, \nu = \nu_1 + \nu_0\right\}.$$

Proof. This is a result of the identification of $W_1^\sharp$ with W_1' done in Theorem 4.4.8 together with the characterization of W_1' shown in Lemma 4.4.5 and the identification of the annihilation costs in Lemma 4.4.6. □

4.5 Induced Length Metric: Topology

Here we assume that X is a *compact* length space. Then it is in particular a geodesic space.

A useful feature of $W_p^\sharp$ is that it metrizes vague convergence of subprobability measures.

Proposition 4.5.1. *For every $p \in [1,\infty)$, $W_p^\sharp$ is a complete, separable, geodesic metric on $\mathcal{P}_p^{sub}(Y)$.*

Furthermore, for $\mu_n, \mu \in \mathcal{P}_p^{sub}(Y)$, $n \in \mathbb{N}$, the following are equivalent:

(i) $\mu_n \to \mu$ vaguely on Y.

(ii) $W_p^\sharp(\mu_n,\mu) \to 0$ as $n \to \infty$

Proof. By Lemma 4.3.4, on a bounded space all the $W_p^\sharp$-metrics are equivalent, so it suffices to prove the equivalence for $p = 1$.

(ii) $\Rightarrow$ *(i)*: Given a sequence with $W_1^\sharp(\mu_n, \mu) \to 0$, we know by Theorem 4.4.8 that then also $W_1'(\mu_n', \nu') \to 0$. Since W_1' is a Kantorovich-Wasserstein metric and thus metrizes weak convergence, we get $\mu_n' \to \mu'$ weakly on Y'. Hence, turning to the restrictions to Y, we end up with vague convergence $\mu_n \to \mu$.

(i) $\Rightarrow$ *(ii)*: For the sake of a contradiction assume that $W_1^\sharp(\mu_n, \mu)$ does not converge to zero. Then we can take a subsequence such that $W_1^\sharp(\mu_{n_k}, \mu) \to c \in (0, \infty]$. Extending the measures to probability measures on Y', we can use Prokhorov's theorem on $\mathcal{P}(Y')$ to extract a weakly converging subsequence $\mu'_{n_{k_\ell}} \to \nu' \in \mathcal{P}(Y')$. Then $\nu := \nu'|_Y$ is a vague limit of the sequence $(\mu_{n_{k_\ell}})_{\ell \in \mathbb{N}}$ and we know that

$$W_1^\sharp(\mu_{n_{k_\ell}}, \nu) = W_1'(\mu'_{n_{k_\ell}}, \nu') \longrightarrow 0 \quad \text{as } \ell \to \infty.$$

In particular, $W_1^\sharp(\mu, \nu) = \lim_{\ell \to \infty} W_1^\sharp(\mu, \mu_{n_{k_\ell}}) \neq 0$, so $\mu \neq \nu$. Hence there exists a function $f \in C_c^0(Y)$ such that

$$\int_Y f \, \mathrm{d}\mu \neq \int_Y f \, \mathrm{d}\nu = \lim_{\ell \to \infty} \int_Y f \, \mathrm{d}\mu_{n_{k_\ell}},$$

so μ_n cannot converge vaguely to μ on Y.

Now with this characterization of vague convergence we can finish the proof. Since the vague topology is complete, so is $(\mathcal{P}_p^{sub}(Y), W_p^\sharp)$. By definition, $W_p^\sharp$ is a length metric, so let us prove the existence of midpoints to show that it is actually geodesic. Let $\mu, \nu \in \mathcal{P}_p^{sub}(Y)$ and $\varepsilon_n > 0$ such that $\varepsilon_n \to 0$ as $n \to \infty$. Then take ε_n-midpoints $\eta_n \in \mathcal{P}_p^{sub}(Y)$ between μ and ν. Again switching to the compact space Y', by Prokhorov's theorem we get a weakly converging subsequence $\eta'_{n_k} \rightharpoonup \eta'_*$ in $\mathcal{P}_p(Y')$. Then the restrictions on Y converge vaguely, which by the above means that

$$W_p^\sharp(\mu, \eta_{n_k}) \longrightarrow W_p^\sharp(\mu, \eta_*).$$

But then η_* is indeed a midpoint between μ and ν. □

Remark 4.5.2.
a) In particular, this implies that $\mu_n \to \mu$ weakly on Y if and only if $W_p^\sharp(\mu_n, \mu) \to 0$ and $\mu_n(Y) \to \mu(Y)$.
b) The implication "(ii)$\Rightarrow$(i)" holds true for all length spaces X without requiring their compactness.

The following simple estimate will make it possible to prove the continuity of W_p^0 with respect to weak convergence plus convergence of moments of subprobability measures.

Lemma 4.5.3. *Let $\mu, \nu \in \mathcal{P}_1^{sub}(Y)$ with $\mu(Y) \geq \nu(Y)$. Then, for any $z \in X \setminus Y$,*

$$W_1^0(\mu, \nu) \leq \inf \left\{ W_1(\mu_1, \nu) + \int_X d(x, z) \, \mathrm{d}\mu_0(x) \;\middle|\; \mu = \mu_1 + \mu_0, \mu_1(Y) = \nu(Y) \right\}.$$

Proof. Taking a decomposition such that $\nu_1 = \nu$, $\nu_0 = 0$, Lemma 4.2.6 yields $W_1^0(\mu,\nu) \leq W_1(\mu_1,\nu) + W_1^*(\mu_0)$. Using now

$$\begin{aligned} W_1^*(\mu_0,\mu_0) &= \inf_q \int_{X\times X} d^*(x,y)\,\mathrm{d}q(x,y) \\ &\leq \inf_q \int_{X\times X} \big[d(x,z)+d(z,y)\big]\,\mathrm{d}q(x,y) \\ &= 2\int_X d(x,z)\,\mathrm{d}\mu_0(x), \end{aligned}$$

the proof is complete. □

Lemma 4.5.4. *For $\mu^{(n)}, \mu^* \in \mathcal{P}^{sub}(Y)$ the following are equivalent:*

(i) $\mu^{(n)} \to \mu^*$ weakly on Y

(ii) $W_p^0(\mu^{(n)},\mu^*) \to 0$ and $\mu^{(n)}(Y) \to \mu^*(Y)$

Proof. (i) ⇒ (ii): Assume $\mu^{(n)} \to \mu^*$ weakly on Y. It again suffices to prove the result for $p=1$. We want to use Lemma 4.5.3 to show continuity. In order to apply this lemma, we have to decompose the *larger* measure. We will proceed in three steps. First we will consider only sequences $(\mu^{(n)})$ with $\mu^{(n)}(Y) \geq \mu^*(Y)$ for all $n \in \mathbb{N}$. Define $\lambda_n := \frac{\mu^*(Y)}{\mu^{(n)}(Y)}$ and $\mu_1^{(n)} := \lambda_n \mu^{(n)}$. Then $\mu_1^{(n)}(Y) = \mu^*(Y)$, $\lambda_n \to 1$, and for $f \in C_b^0$

$$\begin{aligned} \left|\int_X f\,\mathrm{d}\mu_1^{(n)} - \int_X f\,\mathrm{d}\mu^*\right| \leq& \left|\int_X \lambda_n f\,\mathrm{d}\mu^{(n)} - \int_X f\,\mathrm{d}\mu^{(n)}\right| + \left|\int_X f\,\mathrm{d}\mu^{(n)} - \int_X f\,\mathrm{d}\mu^*\right| \\ =& |\lambda_n - 1| \left|\int_X f\,\mathrm{d}\mu^{(n)}\right| + \left|\int_X f\,\mathrm{d}\mu^{(n)} - \int_X f\,\mathrm{d}\mu^*\right| \longrightarrow 0. \end{aligned}$$

Hence, we have convergence in the Kantorovich-Wasserstein metric: $W_1(\mu_1^{(n)},\mu^*) \to 0$. Writing $\mu_0^{(n)} := (1-\lambda_n)\mu^{(n)}$, by Lemma 4.5.3 we finally have

$$W_1^0(\mu^{(n)},\mu^*) \leq W_1(\mu_1^{(n)},\mu^*) + \int_X d(x,z)\,\mathrm{d}\mu_0^{(n)}(x) \longrightarrow 0.$$

Now, for the case that $\mu^{(n)}(Y) \leq \mu^*(Y)$, let $\lambda'_n := \frac{\mu^{(n)}(Y)}{\mu^*(Y)}$ and $\mu_{1,n}^* := \lambda'_n \mu^*$. Then $\mu_{1,n}^*(Y) = \mu^{(n)}(Y)$ and $\lambda'_n \to 1$. Given $f \in C_b^0$, by

$$\left|\int_X f\,\mathrm{d}\mu_{1,n}^* - \int_X f\,\mathrm{d}\mu^*\right| \leq |\lambda'_n - 1| \left|\int_X f\,\mathrm{d}\mu^*\right| \longrightarrow 0,$$

we see that $\mu_{1,n}^* \rightharpoonup \mu^*$. In a next step this yields

$$\left|\int_X f\,\mathrm{d}\mu_{1,n}^* - \int_X f\,\mathrm{d}\mu^{(n)}\right|$$

$$\leq \left| \int_X f \, \mathrm{d}\mu_{1,n}^* - \int_X f \, \mathrm{d}\mu^* \right| + \left| \int_X f \, \mathrm{d}\mu^* - \int_X f \, \mathrm{d}\mu^{(n)} \right| \longrightarrow 0,$$

i.e. $\mu_{1,n}^* - \mu^{(n)} \rightharpoonup 0$. Hence, using again Lemma 4.5.3, we see that

$$W_1^0(\mu^{(n)}, \mu^*) \leq W_1(\mu^{(n)}, \mu_{1,n}^*) + \int_X d(x,z) \, \mathrm{d}\mu_{0,n}^*(x) \longrightarrow 0.$$

Since a sequence converges if and only if every subsequence has a convergent subsequence, we now can conclude that $a_n := W_1^0(\mu^{(n)}, \mu^*)$ converges to 0. Indeed, take a subsequence a_{n_k}. Then we can take a further subsequence $a_{n_{k_\ell}}$ such that either $\mu^{(n_{k_\ell})}(Y) \geq \mu^*(Y)$ for every $\ell \in \mathbb{N}$, or $\mu^{(n_{k_\ell})}(Y) \leq \mu^*(Y)$ for every $\ell \in \mathbb{N}$. But then the above ensures convergence of these subsequences to 0.

(ii) $\Rightarrow$ *(i)*: Conversely, now assume that $\mu^{(n)}(Y) \to \mu^*(Y)$ and $W_p^0(\mu^{(n)}, \mu^*) \to 0$. Let $\rho^{(n)}, \eta^{(n)} \in \mathcal{P}^{sub}(X)$ such that $(2\rho^{(n)} + \mu^{(n)})(X) = 1 = (2\eta^{(n)} + \mu^*)(X)$, and $W_p^0(\mu^{(n)}, \mu^*) = \tilde{W}_p((\mu^{(n)} + \rho^{(n)}, \rho^{(n)}), (\mu^* + \eta^{(n)}, \eta^{(n)}))$. Let $\mu^{(n_k)}$ be any subsequence and consider the corresponding subsequences $\rho^{(n_k)}, \eta^{(n_k)}$. Compactness of $\hat{X}$ implies that there exists a sub-subsequence $(n_{k_\ell})_\ell$ such that

$$\eta^{(n_{k_\ell})} \rightharpoonup \eta^* \text{ and } \mu^{(n_{k_\ell})} \rightharpoonup \tilde{\mu}^* \text{ and } \rho^{(n_{k_\ell})} \rightharpoonup \rho^*$$

with suitable limits points $\eta^*, \tilde{\mu}^*, \rho^*$. Then we have

$$\begin{aligned} \tilde{W}_p\left((\tilde{\mu}^* + \rho^*, \rho^*), (\mu^* + \eta^*, \eta^*)\right) \leq\, & \tilde{W}_p\Big((\tilde{\mu}^* + \rho^*, \rho^*), (\mu^{(n_{k_\ell})} + \rho^{(n_{k_\ell})}, \rho^{(n_{k_\ell})})\Big) \\ & + \tilde{W}_p\Big((\mu^{(n_{k_\ell})} + \rho^{(n_{k_\ell})}, \rho^{(n_{k_\ell})}), (\mu^* + \eta^{(n_{k_\ell})}, \eta^{(n_{k_\ell})})\Big) \\ & + \tilde{W}_p\Big((\mu^* + \eta^{(n_{k_\ell})}, \eta^{(n_{k_\ell})}), (\mu^* + \eta^*, \eta^*)\Big) \longrightarrow 0. \end{aligned}$$

Hence $\rho^* = \eta^*$ and in particular $\tilde{\mu}^* = \mu^*$. This way we see that every subsequence of $\mu^{(n)}$ has a further subsequence which converges to μ^*, so that also the whole sequence converges to μ^*. □

Remark 4.5.5. Without assuming compactness in Lemma 4.5.4, we are still able to get that $W_p^0(\mu^{(n)}, \mu^*) \to 0$ for $\mu^{(n)}, \mu^* \in \mathcal{P}_p^{sub}(Y)$ if $\mu^{(n)} \to \mu^*$ weakly in Y and

$$\int_Y d(x, x_0)^p \, \mathrm{d}\mu^{(n)}(x) \to \int_Y d(x, x_0)^p \, \mathrm{d}\mu^*(x)$$

for some $x_0 \in Y$.

Chapter 5

Heat Flow with Dirichlet Boundary Conditions

Thanks to the characterization of the heat flow on the glued space as (3.1.2), we can use the glued space to infer some properties on the heat flow with Dirichlet boundary conditions. However, since the gluing does not preserve Ricci curvature bounds, we have to impose the $\mathrm{RCD}(K,\infty)$ condition on the glued space to get interesting consequences.

Throughout this chapter we assume that $(X,d,\mathfrak{m})$ is an infinitesimally Hilbertian metric measure space, and $Y \subset X$ is a dense, open subset with $\mathfrak{m}(\partial Y)=0$. Recall that then also the glued space $\hat{X}$ is infinitesimally Hilbertian.

5.1 Gradient Flow Description

Let us define an entropy for charged probabilities. It will turn out that it equals the relative entropy on the glued space up to an additive constant, so that convexity of this entropy is equivalent to the $\mathrm{CD}(K,\infty)$ condition on the glued space.

Definition 5.1.1. The *charged entropy* is

$$\begin{aligned}&\widetilde{\mathrm{Ent}}_{\mathfrak{m}} \colon \tilde{\mathcal{P}}_2(Y|X) \to (-\infty,\infty],\\ &\widetilde{\mathrm{Ent}}_{\mathfrak{m}}(\sigma) := \mathrm{Ent}_{\mathfrak{m}}(\sigma^+) + \mathrm{Ent}_{\mathfrak{m}}(\sigma^-).\end{aligned}$$

We will say that $(X,Y,d,\mathfrak{m})$ has *charged Ricci curvature bounded below by* $K \in \mathbb{R}$ if the charged entropy is K-convex in $(\tilde{\mathcal{P}}_2(Y|X),\tilde{W}_2)$, i.e. if for every $\sigma,\tau \in \tilde{\mathcal{P}}_2(Y|X)$ there is a $\tilde{W}_2$-geodesic $(\eta_t)_{t\in[0,1]} \subset \tilde{\mathcal{P}}_2(Y|X)$ connecting σ and τ such that

$$\widetilde{\mathrm{Ent}}_{\mathfrak{m}}(\eta_t) \le (1-t)\widetilde{\mathrm{Ent}}_{\mathfrak{m}}(\sigma) + t\widetilde{\mathrm{Ent}}_{\mathfrak{m}}(\tau) - \frac{K}{2}t(1-t)\tilde{W}_2(\sigma,\tau)^2.$$

The identification between the space of charged measures and the probability measures on the doubled space now yields the comparability of the charged entropy with the relative entropy on the doubled space, so that the "charged Ricci curvature bound" is nothing than the Ricci curvature bound on the doubled space.

Lemma 5.1.2. *The charged entropy* $\widetilde{\mathrm{Ent}}_{\mathfrak{m}}$ *is* K*-convex in* $\tilde{\mathcal{P}}_2(Y|X)$ *if and only if the entropy* $\widehat{\mathrm{Ent}}_{\hat{\mathfrak{m}}}$ *is* K*-convex in* $\mathcal{P}_2(\hat{X})$ *(i.e.* $\hat{X}$ *is an* $\mathrm{RCD}(K,\infty)$ *space).*

Proof. Recall the identifications maps from Lemma 4.1.4. Let $\hat{\sigma} \in \mathcal{P}_2(\hat{X})$ with $\hat{\sigma} = \hat{\xi}\hat{\mathfrak{m}}$. We will show that the entropy of $\hat{\sigma}$ in $\mathcal{P}_2(\hat{X})$ equals that of $\Psi(\hat{\sigma})$ in $\tilde{\mathcal{P}}_2(Y|X)$ up to an additive constant, and then the result follows by Lemma 4.1.5 and the fact that K-convexity is preserved if you add a constant to the functional. We have

$$\begin{aligned}\widehat{\mathrm{Ent}}_{\hat{\mathfrak{m}}}(\hat{\sigma}) &= \int_{\hat{X}} \hat{\xi}\log\hat{\xi}\,\mathrm{d}\hat{\mathfrak{m}}\\ &=\frac{1}{2}\int_{Y^+} \hat{\xi}|_{Y^+}\log\hat{\xi}|_{Y^+}\,\mathrm{d}\mathfrak{m} + \frac{1}{2}\int_{Y^-} \hat{\xi}|_{Y^-}\log\hat{\xi}|_{Y^-}\,\mathrm{d}\mathfrak{m} + \int_Z \hat{\xi}|_Z\log\hat{\xi}|_Z\,\mathrm{d}\mathfrak{m}\\ &=\frac{1}{2}\int_{X^+} \hat{\xi}|_{X^+}\log\hat{\xi}|_{X^+}\,\mathrm{d}\mathfrak{m} + \frac{1}{2}\int_{X^-} \hat{\xi}|_{X^-}\log\hat{\xi}|_{X^+}\,\mathrm{d}\mathfrak{m}\end{aligned}$$

On the other hand, to compute $\widetilde{\mathrm{Ent}}_{\mathfrak{m}}(\Psi(\hat{\sigma}))$, let us first identify the density of $\Psi(\hat{\sigma})^i$ with respect to $\mathfrak{m}$: For a Borel-measurable set $A \subset X$

$$\begin{aligned}\Psi(\hat{\sigma})^i(A) &= \hat{\sigma}(\iota_i(A)\cap Y^i) + \frac{1}{2}\hat{\sigma}(\iota_i(A)\cap Z)\\ &= \int_{\iota_i(A)\cap Y^i}\mathrm{d}\hat{\sigma} + \frac{1}{2}\int_{\iota_i(A)\cap Z}\mathrm{d}\hat{\sigma}\\ &= \int_{\iota_i(A)\cap Y^i}\frac{1}{2}\hat{\xi}\,\mathrm{d}\mathfrak{m} + \frac{1}{2}\int_{\iota_i(A)\cap Z}\hat{\xi}\,\mathrm{d}\mathfrak{m}\\ &= \frac{1}{2}\int_{\iota_i(A)\cap X^i}\hat{\xi}|_{X^i}\,\mathrm{d}\mathfrak{m},\end{aligned}$$

so that

$$\Psi(\hat{\sigma})^i = \frac{1}{2}\left(\hat{\xi}|_{X^i}\circ\iota_i\right)\mathfrak{m}.$$

Thus

$$\begin{aligned}\widetilde{\mathrm{Ent}}_{\mathfrak{m}}(\Psi(\hat{\sigma})) =&\ \mathrm{Ent}_{\mathfrak{m}}(\Psi(\hat{\sigma})^+) + \mathrm{Ent}_{\mathfrak{m}}(\Psi(\hat{\sigma})^-)\\ =&\int_X \frac{1}{2}\left(\hat{\xi}|_{X^+}\circ\iota_+\right)\log\left(\frac{1}{2}\left(\hat{\xi}|_{X^+}\circ\iota_+\right)\right)\mathrm{d}\mathfrak{m}\\ &+\int_X \frac{1}{2}\left(\hat{\xi}|_{X^-}\circ\iota_-\right)\log\left(\frac{1}{2}\left(\hat{\xi}|_{X^-}\circ\iota_-\right)\right)\mathrm{d}\mathfrak{m}\\ =&\int_X \frac{1}{2}\left(\hat{\xi}|_{X^+}\circ\iota_+\right)\log\left(\left(\hat{\xi}|_{X^+}\circ\iota_+\right)\right)\mathrm{d}\mathfrak{m} + \int_X \frac{1}{2}\left(\hat{\xi}|_{X^+}\circ\iota_+\right)\log\left(\frac{1}{2}\right)\mathrm{d}\mathfrak{m}\\ &+\int_X \frac{1}{2}\left(\hat{\xi}|_{X^-}\circ\iota_-\right)\log\left(\left(\hat{\xi}|_{X^-}\circ\iota_-\right)\right)\mathrm{d}\mathfrak{m} + \int_X \frac{1}{2}\left(\hat{\xi}|_{X^-}\circ\iota_-\right)\log\left(\frac{1}{2}\right)\mathrm{d}\mathfrak{m}\\ =&\int_X \frac{1}{2}\left(\hat{\xi}|_{X^+}\circ\iota_+\right)\log\left(\left(\hat{\xi}|_{X^+}\circ\iota_+\right)\right)\mathrm{d}\mathfrak{m}\\ &+\int_X \frac{1}{2}\left(\hat{\xi}|_{X^-}\circ\iota_-\right)\log\left(\left(\hat{\xi}|_{X^-}\circ\iota_-\right)\right)\mathrm{d}\mathfrak{m}\end{aligned}$$

$$+\log\frac{1}{2}\underbrace{\int_X \frac{1}{2}\left(\hat{\xi}|_{X^+}\circ\iota_+\right)+\frac{1}{2}\left(\hat{\xi}|_{X^-}\circ\iota_-\right)\mathrm{d}\mathfrak{m}}_{=1}$$
$$=\widehat{\mathrm{Ent}}_{\hat{\mathfrak{m}}}(\hat{\sigma})+\log\frac{1}{2}.$$

□

Lemma 5.1.3. *Assume that* $(X, Y, d, \mathfrak{m})$ *has charged Ricci curvature bounded below by* $K \in \mathbb{R}$. *Then* $(X, d, \mathfrak{m})$ *is an* $\mathrm{RCD}(K, \infty)$ *space.*

Proof. Due to the isometric embedding of $\mathcal{P}_2(X)$ into $\tilde{\mathcal{P}}_2(Y|X)$, a geodesic $(\mu_t)_{[0,1]}$ in $\mathcal{P}_2(X)$ yields a geodesic $\tilde{\mu}_t := (\frac{1}{2}\mu_t, \frac{1}{2}\mu_t)$ in $\tilde{\mathcal{P}}_2(Y|X)$. Thanks to the charged Ricci curvature, we know that

$$\widetilde{\mathrm{Ent}}_{\mathfrak{m}}(\tilde{\mu}_t) \leq (1-t)\widetilde{\mathrm{Ent}}_{\mathfrak{m}}(\tilde{\mu}_0) + t\widetilde{\mathrm{Ent}}_{\mathfrak{m}}(\tilde{\mu}_1) - \frac{K}{2}t(1-t)\tilde{W}_2(\mu_0, \mu_1)^2.$$

Thanks to

$$\widetilde{\mathrm{Ent}}_{\mathfrak{m}}(\tilde{\mu}_t) = 2\,\mathrm{Ent}_{\mathfrak{m}}\left(\frac{1}{2}\mu_t\right) = \mathrm{Ent}_{\mathfrak{m}}(\mu_t) + \log\frac{1}{2},$$

this means

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_t) \leq (1-t)\,\mathrm{Ent}_{\mathfrak{m}}(\mu_0) + t\,\mathrm{Ent}_{\mathfrak{m}}(\mu_1) - \frac{K}{2}t(1-t)W_2^2(\mu_0, \mu_1).$$

□

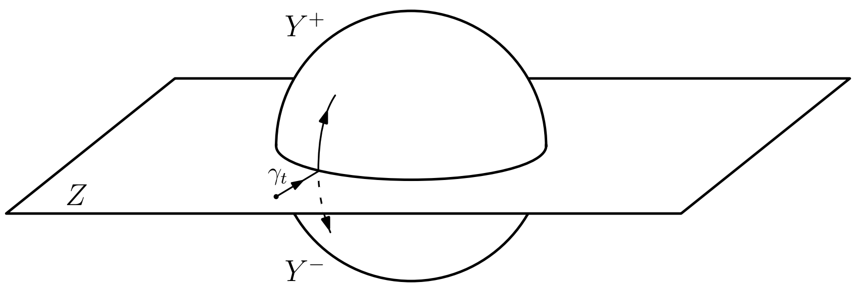

Figure 5.1: Branching geodesic in the case that $Y \subset X$ is not dense.

Remark 5.1.4. If $(X, d, \mathfrak{m})$ is infinitesimally Hilbertian and if $\mathfrak{m}$ has full topological support then the K-convexity of $\widetilde{\mathrm{Ent}}_{\mathfrak{m}}$ actually implies that $\overline{Y} = X$. Indeed, it was shown in [RS14] that the space is then essentially non-branching. If Y would not be dense, then we could start a geodesic in $Z = X \setminus Y$ that could split at the gluing edge into both copies, yielding a branching geodesic, see Figure 5.1.

As an example of a space, whose charged entropy is convex we give convex subsets of Riemannian manifolds with a Ricci curvature bound.

Example 5.1.5. Let (M,g) be a complete Riemannian manifold with Ricci curvature bounded below by $K \in \mathbb{R}$. Take an open, bounded, convex subset $Y \subset M$ with smooth, compact boundary. Consider the closure $X := \overline{Y}$ with the Riemannian distance d and the Riemannian volume measure $\mathfrak{m}$ obtained by restriction to X. Then the metric measure space $(X,d,\mathfrak{m})$ satisfies the $\mathrm{RCD}(K,\infty)$-condition and $(X,Y,d,\mathfrak{m})$ has charged Ricci curvature bounded below by K. Indeed, as a Riemannian manifold with lower Ricci curvature bound K, M is an $\mathrm{RCD}(K,\infty)$ space. As a *convex* subset, also $\overline{Y}$ with the restricted distance and measure is an $\mathrm{RCD}(K,\infty)$ space. By Theorem 3.2.1, the doubling of the manifold is an $\mathrm{RCD}(K,\infty)$ space, so that by the identification of the entropies in the previous lemma we get the convexity of the charged entropy.

Proposition 5.1.6. *Assume that $(X,Y,d,\mathfrak{m})$ has charged Ricci curvature bounded below by $K \in \mathbb{R}$.*

i) For each $\sigma_0 \in \tilde{\mathcal{P}}_2(Y|X)$, there exists a unique EVI_K-gradient flow $(\sigma_t)_{t>0}$ for the Boltzmann entropy $\widetilde{\mathrm{Ent}}_{\mathfrak{m}}$ in $\big(\tilde{\mathcal{P}}_2(Y|X), \tilde{W}_2\big)$. We will also denote it by $\tilde{\mathscr{P}}_t\sigma$.

ii) For each $\mu_0 \in \mathcal{P}_2^{sub}(Y)$, the heat flow $(\mu_t)_{t>0}$ on Y with Dirichlet boundary conditions is obtained as the effective flow

$$\mu_t = \sigma_t^+ - \sigma_t^-$$

where $(\sigma_t)_{t>0}$ is the EVI_K-flow as above starting in any $\sigma_0 \in \tilde{\mathcal{P}}_2(Y|X)$ with $\mu_0 = \sigma_0^+ - \sigma_0^-$.

iii) For each $\nu_0 \in \mathcal{P}_2(X)$, the heat flow $(\nu_t)_{t>0}$ on X is obtained as the total flow

$$\nu_t = \sigma_t^+ + \sigma_t^-$$

where $(\sigma_t)_{t>0}$ is the EVI_K-flow as above starting in any $\sigma_0 \in \tilde{\mathcal{P}}_2(Y|X)$ with $\nu_0 = \sigma_0^+ + \sigma_0^-$.

iv) For each $\sigma_0 \in \tilde{\mathcal{P}}_2(Y|X)$, the EVI_K-flow $(\sigma_t)_{t>0}$ from i) can be characterized as

$$\sigma_t = \Big(\frac{\nu_t + \mu_t}{2}, \frac{\nu_t - \mu_t}{2}\Big)$$

where $(\nu_t)_{t>0}$ will denote the heat flow on X starting in $\nu_0 = \sigma_0^+ + \sigma_0^-$ and $(\mu_t)_{t>0}$ will denote the heat flow on Y with Dirichlet boundary conditions starting in $\mu_0 = \sigma_0^+ - \sigma_0^-$.

In order to prove this proposition, we will provide a simple lemma characterizing the heat flow of charged measures in terms of the heat flows of their effective and total measures.

Lemma 5.1.7. *Let $\sigma \in \tilde{\mathcal{P}}(Y|X)$. Then*

$$\tilde{\mathscr{P}}_t\sigma = \left(\mathscr{P}_t\frac{\sigma^+ + \sigma^-}{2} + \mathscr{P}_t^0\frac{\sigma^+ - \sigma^-}{2}, \mathscr{P}_t\frac{\sigma^+ + \sigma^-}{2} - \mathscr{P}_t^0\frac{\sigma^+ - \sigma^-}{2}\right).$$

Proof. We do the calculation in the equivalent setting of the doubled space $\hat{X}$. Let $\hat{\sigma} \in \mathcal{P}(\hat{X})$. Then

$$
\begin{aligned}
&\int_{\hat{X}} u \,\mathrm{d}\hat{\mathscr{P}}_t\hat{\sigma} = \int_{\hat{X}} \hat{P}_t u \,\mathrm{d}\hat{\sigma} \\
&= \int_{X^+} P_t \frac{u^+ + u^-}{2} + P_t^0 \frac{u^+ - u^-}{2} \,\mathrm{d}\sigma^+ + \int_{X^-} P_t \frac{u^+ + u^-}{2} - P_t^0 \frac{u^+ - u^-}{2} \,\mathrm{d}\sigma^- \\
&= \int_{X^+} \frac{u^+ + u^-}{2} \,\mathrm{d}\mathscr{P}_t\sigma^+ + \int_{X^+} \frac{u^+ - u^-}{2} \,\mathrm{d}\mathscr{P}_t^0\sigma^+ + \int_{X^-} \frac{u^+ + u^-}{2} \,\mathrm{d}\mathscr{P}_t\sigma^- \\
&\quad - \int_{X^-} \frac{u^+ - u^-}{2} \,\mathrm{d}\mathscr{P}_t^0\sigma^- \\
&= \int_{X^+} \frac{1}{2}u^+ \,\mathrm{d}\mathscr{P}_t\sigma^+ + \int_{X^+} \frac{1}{2}u^- \,\mathrm{d}\mathscr{P}_t\sigma^+ + \int_{X^+} \frac{1}{2}u^+ \,\mathrm{d}\mathscr{P}_t^0\sigma^+ - \int_{X^+} \frac{1}{2}u^- \,\mathrm{d}\mathscr{P}_t^0\sigma^+ \\
&\quad + \int_{X^-} \frac{1}{2}u^+ \,\mathrm{d}\mathscr{P}_t\sigma^- + \int_{X^-} \frac{1}{2}u^- \,\mathrm{d}\mathscr{P}_t\sigma^- - \int_{X^-} \frac{1}{2}u^+ \,\mathrm{d}\mathscr{P}_t^0\sigma^- + \int_{X^-} \frac{1}{2}u^- \,\mathrm{d}\mathscr{P}_t^0\sigma^- \\
&= \int_{X^+} \frac{1}{2}u^+ \,\mathrm{d}\mathscr{P}_t\sigma^+ + \int_{X^-} \frac{1}{2}u^- \,\mathrm{d}\mathscr{P}_t\sigma^+ + \int_{X^+} \frac{1}{2}u^+ \,\mathrm{d}\mathscr{P}_t^0\sigma^+ - \int_{X^-} \frac{1}{2}u^- \,\mathrm{d}\mathscr{P}_t^0\sigma^+ \\
&\quad + \int_{X^+} \frac{1}{2}u^+ \,\mathrm{d}\mathscr{P}_t\sigma^- + \int_{X^-} \frac{1}{2}u^- \,\mathrm{d}\mathscr{P}_t\sigma^- - \int_{X^+} \frac{1}{2}u^+ \,\mathrm{d}\mathscr{P}_t^0\sigma^- + \int_{X^-} \frac{1}{2}u^- \,\mathrm{d}\mathscr{P}_t^0\sigma^- \\
&= \int_{X^+} u^+ \,\mathrm{d}\mathscr{P}_t \frac{\sigma^+ + \sigma^-}{2} + \int_{X^+} u^+ \,\mathrm{d}\mathscr{P}_t^0 \frac{\sigma^+ - \sigma^-}{2} + \int_{X^-} u^- \,\mathrm{d}\mathscr{P}_t \frac{\sigma^+ + \sigma^-}{2} \\
&\quad - \int_{X^-} u^- \,\mathrm{d}\mathscr{P}_t^0 \frac{\sigma^+ - \sigma^-}{2} \\
&= \int_{X^+} u^+ \,\mathrm{d}\left(\mathscr{P}_t \frac{\sigma^+ + \sigma^-}{2} + \mathscr{P}_t^0 \frac{\sigma^+ - \sigma^-}{2} \right) \\
&\quad + \int_{X^-} u^- \,\mathrm{d}\left(\mathscr{P}_t \frac{\sigma^+ + \sigma^-}{2} - \mathscr{P}_t^0 \frac{\sigma^+ - \sigma^-}{2} \right).
\end{aligned}
$$

We again relied heavily on the fact that we glue together copies of the *same* space, making it possible to "switch" indices when necessary. To do it rigorously, one should use the identification maps $\iota_\pm : X \to X^\pm$. $\square$

Proof of Proposition 5.1.6. This will follow from the identification with the glued space and the properties shown in Section 3.1, in particular Theorem 3.1.16. Let us provide the details.

i) Given $\sigma_0 \in \tilde{\mathcal{P}}(Y|X)$, consider $\hat{\sigma} := \Phi(\sigma_0) \in \mathcal{P}(\hat{X})$, with the isometry Φ given in Lemma 4.1.4. Since $\hat{X}$ is an $\mathrm{RCD}(K, \infty)$ space by the convexity of $\widetilde{\mathrm{Ent}}_{\mathfrak{m}}$ and Lemma 5.1.2, the EVI_K-gradient flow $\hat{\sigma}_t \in \mathcal{P}(\hat{X})$ (of the relative entropy $\widehat{\mathrm{Ent}}_{\hat{\mathfrak{m}}}$ in $(\mathcal{P}_2(\hat{X}), \hat{W}_2)$) starting in $\hat{\sigma}$ exists. Again by the identification of the entropies in Lemma 5.1.2, the flow $\sigma_t := \Psi(\hat{\sigma}_t)$ is the EVI_K-gradient flow of $\widetilde{\mathrm{Ent}}_{\mathfrak{m}}$ in $\tilde{\mathcal{P}}(Y|X)$.

ii) Let $\mu_0 \in \mathcal{P}_2^{sub}(X)$, and let $\sigma_0 \in \tilde{\mathcal{P}}(Y|X)$ such that $\mu_0 = \sigma_0^+ - \sigma_0^-$. Consider $\sigma_t := \tilde{\mathscr{P}}_t\sigma_0$. By Lemma 5.1.7 we have

$$\sigma_t^+ - \sigma_t^- = \mathscr{P}_t^0(\sigma_0^+ - \sigma_0^-) = \mathscr{P}_t^0\mu_0.$$

This also shows the independence of the chosen σ_0, as the right-hand side is independent of it.

iii) As in ii).

iv) Let $\sigma_0 \in \tilde{\mathcal{P}}_2(Y|X)$ and define $\mu_0 := \sigma_0^+ - \sigma_0^-$ and $\nu_0 := \sigma_0^+ + \sigma_0^-$. Then, again by Lemma 5.1.7,

$$\begin{aligned}\sigma_t = \tilde{\mathscr{P}}_t \sigma_0 &= \left(\mathscr{P}_t \frac{\sigma_0^+ + \sigma_0^-}{2} + \mathscr{P}_t^0 \frac{\sigma_0^+ + \sigma_0^-}{2}, \mathscr{P}_t \frac{\sigma_0^+ + \sigma_0^-}{2} + \mathscr{P}_t^0 \frac{\sigma_0^+ - \sigma_0^-}{2} \right) \\ &= \left(\mathscr{P}_t \frac{\mu_0}{2} + \mathscr{P}_t^0 \frac{\nu_0}{2}, \mathscr{P}_t \frac{\mu_0}{2} + \mathscr{P}_t^0 \frac{\nu_0}{2} \right) \\ &= \left(\frac{\mu_t + \nu_t}{2}, \frac{\mu_t - \nu_t}{2} \right).\end{aligned}$$

□

Remark 5.1.8. a) As in [Sav14, after Cor. 4.3, Thm. 4.4] (based on [AGS15, Prop. 3.2, Thm. 3.5]) one can extend the flow to measures without finite second moment.

b) In the situation of Example 5.1.5, the "heat flow on X" will be the heat flow on $\overline{Y} \subset M$ with Neumann boundary conditions at ∂Y.

From the charged Ricci curvature condition we can deduce a number of contraction results in the various metrics that occurred in Chapter 4.

Proposition 5.1.9. *Assume that $(X, Y, d, \mathfrak{m})$ has charged Ricci curvature bounded below by $K \in \mathbb{R}$. Then the* EVI_K*-flows $(\sigma_t)_{t>0}$ and $(\tau_t)_{t>0}$ of the charged entropy in $\tilde{\mathcal{P}}_2(Y|X)$ are K-contractive in all L^p-transportation distances:*

$$\tilde{W}_p(\sigma_t, \tau_t) \le e^{-Kt} \cdot \tilde{W}_p(\sigma_0, \tau_0)$$

for all $t > 0$ and all $p \in [1, \infty)$.

Proof. This is again a direct consequence of the identification, since the glued space is an $\mathrm{RCD}(K, \infty)$ space and thus satisfies the desired Wasserstein contraction. □

Theorem 5.1.10. *Assume that $(X, Y, d, \mathfrak{m})$ has charged Ricci curvature bounded below by $K \in \mathbb{R}$. For all $\mu_0, \nu_0 \in \mathcal{P}_p^{sub}(Y)$, all $t > 0$ and all $p \in [1, \infty)$*

$$W_p^0(\mu_t, \nu_t) \le e^{-Kt} \cdot W_p^0(\mu_0, \nu_0)$$

where $\mu_t := \mathscr{P}_t^0 \mu_0$ and $\nu_t := \mathscr{P}_t^0 \nu_0$ denote the heat flows on Y with Dirichlet boundary conditions starting in μ_0 and ν_0, resp.

Proof. Given $\mu_0, \nu_0 \in \mathcal{P}_p^{sub}(Y)$ and $\varepsilon > 0$, we may choose $\sigma_0, \tau_0 \in \tilde{\mathcal{P}}_p(Y|X)$ with $\mu_0 = \sigma_0^+ - \sigma_0^-$ and $\nu_0 = \tau_0^+ - \tau_0^-$ such that

$$\tilde{W}_p(\sigma_0, \tau_0) \le W_p^0(\mu_0, \nu_0) + \varepsilon.$$

Thus, by the very definition of W_p^0 and by the previous proposition,

$$W_p^0\big(\mu_t,\nu_t\big) \le \tilde{W}_p\big(\sigma_t,\tau_t\big) \le e^{-Kt}\cdot \tilde{W}_p\big(\sigma_0,\tau_0\big) \le e^{-Kt}\cdot \Big(W_p^0\big(\mu_0,\nu_0\big)+\varepsilon\Big).$$

Since $\varepsilon > 0$ was arbitrary, this proves the claim. □

Corollary 5.1.11. *Assume that $(X, Y, d, \mathfrak{m})$ has charged Ricci curvature bounded below by $K \in \mathbb{R}$. Let $\mu_0, \nu_0 \in \mathcal{P}_p^{sub}(Y)$, and let $\mu_t := \mathscr{P}_t^0\mu_0$ and $\nu_t := \mathscr{P}_t^0\nu_0$ denote the heat flows on Y with Dirichlet boundary conditions starting in μ_0 and ν_0, resp. Then for all $t > 0$ and all $p \in [1,\infty)$ we have both*

$$W_p^\flat\big(\mu_t,\nu_t\big) \le e^{-Kt}\cdot W_p^\flat\big(\mu_0,\nu_0\big),$$

and

$$W_p^\sharp\big(\mu_t,\nu_t\big) \le e^{-Kt}\cdot W_p^\sharp\big(\mu_0,\nu_0\big).$$

In particular, $W_1'\big(\mu_t,\nu_t\big) \le e^{-Kt}\cdot W_1'\big(\mu_0,\nu_0\big)$.

Proof. Observe that

$$\begin{aligned}
W_p^\flat(\mu_t,\nu_t) &= \inf\left\{\sum_{i=1}^n W_p^0(\eta_{i-1},\eta_i)\,\Big|\, n\in\mathbb{N}, \eta_i\in\mathcal{P}_p^{sub}(Y), \eta_0=\mu_t, \eta_n=\nu_t\right\}\\
&\le \inf\left\{\sum_{i=1}^n W_p^0(\mathscr{P}_t^0\xi_{i-1},\mathscr{P}_t^0\xi_i)\,\Big|\, n\in\mathbb{N}, \xi_i\in\mathcal{P}_p^{sub}(Y), \xi_0=\mu_0, \xi_n=\nu_0\right\}\\
&\le e^{-Kt}\inf\left\{\sum_{i=1}^n W_p^0(\xi_{i-1},\xi_i)\,\Big|\, n\in\mathbb{N}, \xi_i\in\mathcal{P}_p^{sub}(Y), \xi_0=\mu_0, \xi_n=\nu_0\right\}\\
&= e^{-Kt}W_p^\flat(\mu_0,\nu_0).
\end{aligned}$$

This also implies that for a curve $(\eta_s)_{s\in[0,1]} \subset \mathcal{P}_p^{sub}(Y)$ its length satisfies $L_p^\flat(\mathscr{P}_t\eta) \le e^{-Kt}L_p^\flat(\eta)$, so that eventually

$$\begin{aligned}
W_p^\sharp(\mu_t,\nu_t) = \inf_{\eta:\mu_t\rightsquigarrow\nu_t} L_p^\flat(\eta) &\le \inf_{\xi:\mu_0\rightsquigarrow\nu_0} L_p^\flat(\mathscr{P}_t\xi)\\
&\le e^{-Kt}\inf_{\xi:\mu_0\rightsquigarrow\nu_0} L_p^\flat(\xi) = e^{-Kt}W_p^\sharp(\mu_0,\nu_0).
\end{aligned}$$

□

5.2 Gradient Estimates and Bochner's Inequality

The charged Ricci curvature bound will not only imply the Wasserstein contraction results for the heat flow with Dirichlet boundary conditions, but also a gradient estimate which involves *both* semigroups, P_t (with Neumann boundary condition) and P_t^0 (with Dirichlet boundary condition), and a Bochner inequality involving both Laplacians. Before proving them, we will show that they are equivalent to each other. We continue to work with an infinitesimally Hilbertian metric measure space $(X, d, \mathfrak{m})$ with dense, open subset $Y \subset X$ with $\mathfrak{m}(\partial Y) = 0$.

Recall the definitions of $D_p(\mathcal{E}^0)$ and $D_p(\Delta^0)$ given in (2.4.3) and (2.4.4). In this section we will extend every function (e.g. in $D(\Delta^0), D(\mathcal{E}^0) \subset L^2(Y, \mathfrak{m}|_Y)$) to X by zero.

Proposition 5.2.1. *Assume that $\mathfrak{m}(X) < \infty$. For each $p \in [1, 2]$, the following properties are equivalent to each other:*

(i) For all $t > 0$, and all $f \in D_p(\mathcal{E}^0)$

$$\left|\nabla P_t^0 f\right|^p \leq e^{-Kpt} \cdot P_t\left(|\nabla f|^p\right) \quad \mathfrak{m}\text{-a.e. in } X \quad (\text{“}p\text{-gradient estimate”}). \tag{5.2.1}$$

Note that different semigroups appear on the left and right hand side.

(ii) For all $f \in D_p(\Delta^0)$ with $\Delta^0 f \in D_p(\mathcal{E}^0)$ and every $\varphi \in D_\infty(\Delta)$ with $\varphi \geq 0$

$$\frac{1}{p}\int_X \Delta\varphi |\nabla f|^p \,\mathrm{d}\mathfrak{m} - \int_{\{|\nabla f| \neq 0\}} \varphi |\nabla f|^{p-2} \nabla f \cdot \nabla \Delta^0 f \,\mathrm{d}\mathfrak{m} \geq K \int_X \varphi |\nabla f|^p \,\mathrm{d}\mathfrak{m} \tag{5.2.2}$$

(“p-Bochner inequality”).

Proof. Given $t > 0$ and functions $f \in L^p(Y, \mathfrak{m}|_Y)$ and $\varphi \in L^\infty(X, \mathfrak{m})$ with $\varphi \geq 0$, we define $F \colon [0, t] \to \mathbb{R}$ by

$$F(s) := \int_X e^{-Kps} P_s\varphi |\nabla P_{t-s}^0 f|^p \,\mathrm{d}\mathfrak{m}.$$

The main task is to show that the derivative of F is

$$\begin{aligned} F'(s) = &\int_X e^{-Kps} \Delta P_s\varphi |\nabla P_{t-s}^0 f|^p \,\mathrm{d}\mathfrak{m} \\ &- p\int_X e^{-Kps} P_s\varphi |\nabla P_{t-s}^0 f|^{p-2} \nabla P_{t-s}^0 f \cdot \nabla \Delta^0 P_{t-s}^0 f \,\mathrm{d}\mathfrak{m} \\ &- pK\int_X e^{-Kps} P_s\varphi |\nabla P_{t-s}^0 f|^p \,\mathrm{d}\mathfrak{m}. \end{aligned} \tag{5.2.3}$$

Inequalities (5.2.1) and (5.2.2) then both express the monotonicity of F – one time “by definition”, the other time by positivity of the derivative. This strategy of proof is given in [Han18, Thm. 3.5].

Consider the case $p > 1$ first. Let us carefully compute the derivative. Adding zeros, we arrive at

$$\begin{aligned} \frac{F(s+\varepsilon) - F(s)}{\varepsilon} = &\frac{1}{\varepsilon}\int_X e^{-Kp(s+\varepsilon)} P_{(s+\varepsilon)}\varphi |\nabla P_{t-(s+\varepsilon)}^0 f|^p - e^{-Kps} P_s\varphi |\nabla P_{t-s}^0 f|^p \,\mathrm{d}\mathfrak{m} \\ = &\frac{1}{\varepsilon}\int_X \left(e^{-Kp(s+\varepsilon)} - e^{-Kps}\right) P_{(s+\varepsilon)}\varphi |\nabla P_{t-(s+\varepsilon)}^0 f|^p \,\mathrm{d}\mathfrak{m} \\ &+ \frac{1}{\varepsilon}\int_X \left(P_{(s+\varepsilon)}\varphi - P_s\varphi\right) e^{-Kps} |\nabla P_{t-(s+\varepsilon)}^0 f|^p \,\mathrm{d}\mathfrak{m} \end{aligned}$$

$$+\frac{1}{\varepsilon}\int_X \Big(|\nabla P^0_{t-(s+\varepsilon)}f|^p - |\nabla P^0_{t-s}f|^p\Big)\, e^{-Kps} P_s\varphi \,\mathrm{d}\mathfrak{m}.$$

The first term yields the derivative of the exponential function. By the weak-$*$-convergence in L^∞ provided in Lemma 2.4.1iii), we know that the second term converges to

$$\int_X e^{-Kps} P_s\Delta\varphi |\nabla P^0_{t-s}f|^p.$$

For the last term, we use that by Taylor expansion around the point $b\in(0,\infty)$ of the function $g\colon (0,\infty)\to(0,\infty), g(a):=a^q, q\in(0,\infty)$, we get

$$a^q - b^q = (a-b)\left(qb^{q-1}+o(1)\right) \text{ for } a\to b.$$

Applying this to the last term, we can write

$$\begin{aligned}
&\frac{1}{\varepsilon}\Big(|\nabla P^0_{t-(s+\varepsilon)}f|^p - |\nabla P^0_{t-s}f|^p\Big)e^{-Kps}P_s\varphi\\
&= \frac{1}{\varepsilon}\left(|\nabla P^0_{t-(s+\varepsilon)}f|^2 - |\nabla P^0_{t-s}f|^2\right)\left(\frac{p}{2}(|\nabla P^0_{t-s}f|^2)^{\frac{p}{2}-1}+o(1)\right)e^{-Kps}P_s\varphi\\
&= \nabla\frac{P^0_{t-(s+\varepsilon)}f - P^0_{t-s}f}{\varepsilon}\cdot\nabla(P^0_{t-(s+\varepsilon)}f + P^0_{t-s}f)\left(\frac{p}{2}(|\nabla P^0_{t-s}f|^2)^{\frac{p}{2}-1}+o(1)\right)e^{-Kps}P_s\varphi.
\end{aligned}$$

By Lemma 2.4.2 we have convergence $\mathfrak{m}$-*almost everywhere* for a subsequence:

$$\begin{aligned}
&\nabla\frac{P^0_{t-(s+\varepsilon)}f - P^0_{t-s}f}{\varepsilon}\cdot\nabla(P^0_{t-(s+\varepsilon)}f + P^0_{t-s}f)\left(\frac{p}{2}(|\nabla P^0_{t-s}f|^2)^{\frac{p}{2}-1}+o(1)\right)e^{-Kps}P_s\varphi\\
&\qquad\longrightarrow -\nabla\Delta^0 P^0_{t-s}f\cdot\nabla(2P^0_{t-s}f)\,\frac{p}{2}|\nabla P^0_{t-s}f|^{p-2}e^{-Kps}P_s\varphi \;\; \mathfrak{m}\text{-a.e.}
\end{aligned}$$

The limit is integrable since

$$\begin{aligned}
&\int_X \left|pe^{-Kps}P_s\varphi\nabla\Delta^0 P^0_{t-s}f\cdot\nabla(P^0_{t-s}f)\,|\nabla P^0_{t-s}f|^{p-2}\right|\mathrm{d}\mathfrak{m}\\
&\qquad\le pe^{-Kps}\|P_s\varphi\|_{L^\infty}\|\nabla P^0_{t-s}f\|^{p-1}_{L^p}\|\nabla\Delta^0 P^0_{t-s}f\|_{L^p} < \infty.
\end{aligned}$$

Hence we can interchange differentiation and integration, getting the desired derivative (5.2.3) for $s\in(0,t)$. So far we only needed $f\in L^p$ and $\varphi\in L^\infty$. To get the differentiability also in the end points $s=0$ and $s=t$, we need $f\in D_p(\Delta^0), \varphi\in D(\Delta)$.

Now, for the case $p=1$ we approximate by a sequence $p_k\searrow 1$ as $k\to\infty$. Given $f\in L^1(X,\mathfrak{m})$ and $M>0$, we define the truncated function

$$f_M := \min\{M, \max\{f,-M\}\}\in L^\infty(X,\mathfrak{m}).$$

Then we have $f_M\in L^{p_k}(X,\mathfrak{m})$ for every $k\in\mathbb{N}$ and by the continuity of

$$p\mapsto\int_X P_s\varphi|\nabla P^0_{t-s}f_M|^p,$$

we can take the limit $p_k\to 1$, and similarly for the other terms, and get the formula for the derivative of F for $p=1$ for bounded functions. Now we can take the limit $M\to\infty$ by the dominated convergence theorem.

(i)⇒(ii): Using the gradient estimate on $|\nabla P^0_{s-r}(P^0_{t-s}f)|^p$, and the symmetry of P^0_t with respect to $\mathfrak{m}$, we see that for $0 \le r < s \le t$

$$\begin{aligned} F(r) &= \int_X e^{-Kpr} P_r\varphi |\nabla P^0_{s-r}(P^0_{t-s}f)|^p \,\mathrm{d}\mathfrak{m} \\ &\le \int_X e^{-Kpr} P_r\varphi \, e^{-Kp(s-r)} P_{s-r}|\nabla P^0_{t-s}f|^p \,\mathrm{d}\mathfrak{m} = F(s). \end{aligned}$$

As a monotone function, $F' \ge 0$. Evaluating it at $s = 0$ and taking the limit $t \to 0$ we arrive at the p-Bochner inequality.

(ii)⇒(i): First we will make the extra assumption that $f \in D_p(\Delta^0)$ and $\Delta^0 f \in D_p(\mathcal{E}^0)$.Take $\varphi \in D_\infty(\Delta)$, $\varphi \ge 0$. The p-Bochner inequality tells us that F' is non-negative, so that F is monotone. Hence we have the "weak form" of the gradient estimate, $F(0) \le F(t)$. Since φ is arbitrary, we get the pointwise a.e. version.
Now for general $f \in D_p(\mathcal{E}^0)$, we consider the mollified function $\mathfrak{h}^0_\delta f$ as defined in Lemma 2.4.4, for $\delta > 0$. Thanks to Lemma 2.4.4 and Corollary 2.4.3, this function satisfies all the additional assumptions, so we get the gradient estimate for $\mathfrak{h}^0_\delta f$. The convergence result in Corollary 2.4.6 yields convergence almost everywhere for a subsequence, so we are finished. □

The main result of this section is that these inequalities indeed hold if we assume that $(X, Y, d, \mathfrak{m})$ has a charged Ricci curvature bound.

Theorem 5.2.2. *Assume that $(X, Y, d, \mathfrak{m})$ has charged Ricci curvature bounded below by K. Then:*

i) Both properties (i) and (ii) of Proposition 5.2.1 are satisfied, actually for all $p \in [1, \infty)$ and without the assumption that $\mathfrak{m}(X) < \infty$.

ii) The flows from Proposition 5.1.6 and the heat semigroups for functions are related to each other by

$$\mathscr{P}_t \nu_0 = (P_t v)\mathfrak{m}, \qquad \mathscr{P}^0_t \mu_0 = (P^0_t w)\mathfrak{m},$$

for $\nu_0 = v\mathfrak{m} \in \mathcal{P}_2(X)$ and $\mu_0 = w\mathfrak{m} \in \mathcal{P}_2^{sub}(X)$.

Proof. i) Once more switching to the doubled space $\hat{X}$, we can use that it is an $\mathrm{RCD}(K, \infty)$ space and hence satisfies a gradient estimate with $p = 2$. By [Sav14, Cor. 4.3] we have the improved gradient estimate for $p \in [1, 2]$ and by Jensen's inequality one easily obtains the gradient estimate for $p > 2$ from that. Now we take a function $f \in D(\mathcal{E}^0)$ and define

$$u := \begin{cases} f, & \text{on } X^+ \\ -f, & \text{on } X^-. \end{cases}$$

Then $u \in D(\hat{\mathcal{E}})$ and $|\nabla u| = |\nabla f|$ on each X^i. Thus, inserting u in the gradient estimate on $\hat{X}$ yields on the upper half X^+:

$$|\nabla P^0_t f|^p = |\nabla \hat{P}_t u|^p \le e^{-pKt} \hat{P}_t |\nabla u|^p = e^{-pKt} P_t |\nabla f|^p.$$

ii) This follows directly from the duality of the heat semigroups (2.3.5). □

5.3 Halfspaces

Let us add an equivalent characterization of the charged Ricci curvature bound which is more geometric. Given a metric measure space $(V, d_V, \mathfrak{m}_V)$ we say that an open subset $U \subset V$ is a *halfspace* if there exists a measure-preserving isometry $\psi : V \to V$ with invariant set $\partial U = \{x \in V : \psi(x) = x\}$ such that $\psi(U) = V \setminus \overline{U}$. We call two metric measure spaces $(V, d_V, \mathfrak{m}_V)$ and $(W, d_W, \mathfrak{m}_W)$ *mms-isomorphic* if there exists a measure-preserving isometry $\xi\colon (V, d_V, \mathfrak{m}_V) \to (W, d_W, \mathfrak{m}_W)$.

An easy consequence of this definition is that halfspaces are weakly convex; this observation is due to Martin Kell.

Lemma 5.3.1. *Let $(V, d_V, \mathfrak{m}_V)$ be a geodesic metric measure space and let $U \subset V$ be a halfspace. Then $\overline{U}$ is weakly convex, i.e. for any two points $x, y \in \overline{U}$ there exists a geodesic staying in $\overline{U}$.*

Proof. Define $\varphi\colon V \to V$ by

$$\varphi(x) := \begin{cases} x, & \text{if } x \in U \\ \psi(x), & \text{otherwise,} \end{cases}$$

where ψ is the measure-preserving isometry in the definition of a halfspace. This function is 1-Lipschitz. Indeed, given $x, y \in U$ or $x, y \in V \setminus U$, one trivially has $d_V(\varphi(x), \varphi(y)) = d_V(x, y)$. For $x \in U, y \in V \setminus U$, let $(\gamma_t)_{t\in[0,1]}$ be a geodesic connecting x and y. Choose $t^* \in [0, 1]$ such that $\gamma_{t^*} \in \partial U$. Then $\psi(\gamma_{t^*}) = \gamma_{t^*}$ and hence

$$\begin{aligned} d_V(\varphi(x), \varphi(y)) &\le d_V(\varphi(x), \varphi(\gamma_{t^*})) + d_V(\varphi(\gamma_{t^*}), \varphi(y)) \\ &= d_V(x, \gamma_{t^*}) + d_V(\gamma_{t^*}, y) = d_V(x, y). \end{aligned}$$

If we now have two points $x, y \in \overline{U}$ and a geodesic $(\gamma_t)_{t\in[0,1]}$ connecting them, we can consider the curve $\tilde{\gamma}_t := \varphi(\gamma_t)$. This is a curve lying completely in $\overline{U}$ and connecting x and y. It remains to show that it is a geodesic. The Lipschitz-continuity of φ implies that

$$|\dot{\tilde{\gamma}}_r| \le |\dot{\gamma}_r|,$$

so the length of the curve $\tilde{\gamma}$ is *less* than the one of γ. But since γ is geodesic connecting x and y, also $\tilde{\gamma}$ is a geodesic. □

Theorem 5.3.2. *Let $(X, d, \mathfrak{m})$ be a metric measure space, and $Y \subset X$ an open local* $\mathrm{RCD}(K, \infty)$ *space. The following properties are equivalent*

(i) $(X, Y, d, \mathfrak{m})$ *has charged Ricci curvature bounded below by K.*

(ii) Y *is a halfspace in some* $\mathrm{RCD}(K, \infty)$*-space $(V, d_V, \mathfrak{m}_V)$ in the sense that there is a halfspace $\tilde{Y} \subset V$ and a measure-preserving isometry $\xi\colon (Y, d, \mathfrak{m}|_Y) \to (\tilde{Y}, d_V, \mathfrak{m}_V|_{\tilde{Y}})$.*

(iii) ∂Y *is covered by open sets X_i such that $Y \cap X_i$ for each i is mms-isomorphic to a halfspace W_i in some* $\mathrm{RCD}(K, \infty)$*-space $(V_i, d_i, \mathfrak{m}_i)$.*

Proof of Theorem 5.3.2. (i)⇒(ii): Consider the doubling of X, $V := \hat{X}$. Then we can view Y as an open subset of $\hat{X}$ by identifying it with Y^+. Now define $\psi : V \to V$ as the "mirror mapping"

$$\psi(x) := \begin{cases} \iota_- \circ \iota_+^{-1}(x), & \text{if } x \in X^+ \\ \iota_+ \circ \iota_-^{-1}(x), & \text{if } x \in X^-. \end{cases}$$

It is easy to see that ψ is a measure-preserving isometry. Further, let $x \in X^+$ such that $\psi(x) = x$, i.e. $\iota_- \circ \iota_+^{-1}(x) = x$. This in particular means $x \in Z$ since for $x \in Y^+$ we would have $\iota_- \circ \iota_+^{-1}(x) \in Y^-$, which would contradict $\psi(x) = x \in Y^+$. Finally observe that $\psi(Y) = \psi(Y^+) = \iota_-(Y) = Y^- = V \setminus Y^+$.

(ii)⇒(iii): Take $i = 1, V_1 := V$.

(ii)⇒(i): Thanks to ξ, we can define a measure-preserving isometry

$$\varphi\colon (V, d_V, \mathfrak{m}_V) \to (\hat{X}, \hat{d}, \hat{\mathfrak{m}})$$

by mapping $Y \cong \tilde{Y}$ to Y^+, $\psi(Y)$ to Y^- and ∂Y to $Z = X \setminus Y \subset \hat{X}$, where ψ is the map given in the definition of a halfspace. Since curvature-dimension conditions are preserved under measure-preserving isometries, $\hat{X}$ is an $\mathrm{RCD}(K,\infty)$ space. Lemma 5.1.2 then tells us that (i) is satisfied.

(iii)⇒(i): We want to show that $\hat{X}$ is an $\mathrm{RCD}(K,\infty)$ space by using the local-to-global property. Given $x \in \partial Y$, choose i such that $x \in X_i$. Then we can identify $(Y \cap X_i)^+ \cup (Y \cap X_i)^- \subset \hat{X}$ with $\hat{W}_i \subset V_i$ via ξ_i. Given measures $\mu_0, \mu_1 \in \mathcal{P}(\hat{X})$ supported in $(Y \cap X_i)^+ \cup (Y \cap X_i)^-$, then $\nu_\ell := (\xi_i)_\# \mu_\ell \in \mathcal{P}(V_i), \ell = 0, 1$, are supported in $\hat{W}_i$. Since V_i is an $\mathrm{RCD}(K,\infty)$ space, there is a geodesic $\nu_t \in \mathcal{P}(V_i)$ connecting ν_0 and ν_1 such that the entropy $\mathrm{Ent}_{\mathfrak{m}_{V_i}}$ is convex. Pulling back this curve via $\mu_t := (\xi_i^{-1})_\# \nu_t$ provides us with a geodesic in $\mathcal{P}(\hat{X})$ such that $\widehat{\mathrm{Ent}}_{\hat{\mathfrak{m}}}$ is convex. Combining this convex optimal transport near the boundary (i.e. the gluing edge) together with the local RCD property of X (and hence X^+ and X^-), we have that $\hat{X}$ is a local $\mathrm{RCD}(K,\infty)$ space and by the local-to-global property also an $\mathrm{RCD}(K,\infty)$ space. □

Bibliography

[AB64] M. F. Atiyah and R. Bott, *The index problem for manifolds with boundary*, Differential Analysis, Bombay Colloq., 1964, 1964, pp. 175–186. MR0185606

[AG13] L. Ambrosio and N. Gigli, *A user's guide to optimal transport*, Modelling and optimisation of flows on networks, 2013, pp. 1–155.

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Second, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008. MR2401600

[AGS14a] L. Ambrosio, N. Gigli, and G. Savaré, *Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below*, Invent. Math. **195** (2014), no. 2, 289–391.

[AGS14b] ______, *Metric measure spaces with Riemannian Ricci curvature bounded from below*, Duke Math. J. **163** (2014), no. 7, 1405–1490.

[AGS15] L. Ambrosio, N. Gigli, and G. Savaré, *Bakry-Émery curvature-dimension condition and Riemannian Ricci curvature bounds*, Ann. Probab. **43** (2015), no. 1, 339–404. MR3298475

[Ale55] A. D. Alexandrow, *Die innere Geometrie der konvexen Flächen*, Akademie-Verlag, Berlin, 1955. Translated from thr Russian version, 1948. MR0071041

[BBI01] D. Burago, Yu. Burago, and S. Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics, vol. 33, American Mathematical Society, Providence, RI, 2001. MR1835418

[BÉ85] D. Bakry and M. Émery, *Diffusions hypercontractives*, Séminaire de probabilités, XIX, 1983/84, 1985, pp. 177–206.

[BH91] N. Bouleau and F. Hirsch, *Dirichlet forms and analysis on Wiener space*, de Gruyter Studies in Mathematics, vol. 14, Walter de Gruyter & Co., Berlin, 1991.

[BH99] M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 319, Springer-Verlag, Berlin, 1999. MR1744486

[Bog07] V. I. Bogachev, *Measure theory. Vol. I, II*, Springer-Verlag, Berlin, 2007. MR2267655

[Che99] J. Cheeger, *Differentiability of Lipschitz functions on metric measure spaces*, Geom. Funct. Anal. **9** (1999), no. 3, 428–517. MR1708448

[EH15] M. Erbar and M. Huesmann, *Curvature bounds for configuration spaces*, Calc. Var. Partial Differential Equations **54** (2015), no. 1, 397–430. MR3385165

[EN00] K.-J. Engel and R. Nagel, *One-parameter semigroups for linear evolution equations*, Graduate Texts in Mathematics, vol. 194, Springer-Verlag, New York, 2000. With contributions by S. Brendle, M. Campiti, T. Hahn, G. Metafune, G. Nickel, D. Pallara, C. Perazzoli, A. Rhandi, S. Romanelli and R. Schnaubelt. MR1721989

[FG10] A. Figalli and N. Gigli, *A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions*, J. Math. Pures Appl. (9) **94** (2010), no. 2, 107–130. MR2665414

[FOT94] M. Fukushima, Y. Oshima, and M. Takeda, *Dirichlet forms and symmetric Markov processes*, De Gruyter Studies in Mathematics, vol. 19, Walter de Gruyter & Co., Berlin, 1994. MR1303354

[GGKMS18] F. Galaz-García, M. Kell, A. Mondino, and G. Sosa, *On quotients of spaces with Ricci curvature bounded below*, J. Funct. Anal. **275** (2018), no. 6, 1368–1446. MR3820328

[Gig15] N. Gigli, *On the differential structure of metric measure spaces and applications*, Mem. Amer. Math. Soc. **236** (2015), no. 1113, vi+91. MR3381131

[Gig18] ———, *Nonsmooth differential geometry—an approach tailored for spaces with Ricci curvature bounded from below*, Mem. Amer. Math. Soc. **251** (2018), no. 1196, v+161. MR3756920

[Han18] B.-X. Han, *New characterizations of Ricci curvature on RCD metric measure spaces*, Discrete Contin. Dyn. Syst. **38** (2018), no. 10, 4915–4927. MR3834700

[Han19] ———, *Conformal Transformation on Metric Measure Spaces*, Potential Anal. **51** (2019), no. 1, 127–146. MR3981445

[JKO98] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal. **29** (1998), no. 1, 1–17. MR1617171

[Kan42] L. Kantorovitch, *On the translocation of masses*, C. R. (Doklady) Acad. Sci. URSS (N.S.) **37** (1942), 199–201. MR0009619

[Kan58] ———, *On the translocation of masses*, Management Sci. **5** (1958), 1–4. MR0096552

[Ket13] C. Ketterer, *Ricci curvature bounds for warped products*, J. Funct. Anal. **265** (2013), no. 2, 266–299. MR3056704

[Ket15] ———, *Cones over metric measure spaces and the maximal diameter theorem*, J. Math. Pures Appl. (9) **103** (2015), no. 5, 1228–1275. MR3333056

[KMV16] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov, *A new optimal transport distance on the space of finite Radon measures*, Adv. Differential Equations **21** (2016), no. 11-12, 1117–1164. MR3556762

[Kos02] N. N. Kosovskiĭ, *Gluing of Riemannian manifolds of curvature $\geq \kappa$*, Algebra i Analiz **14** (2002), no. 3, 140–157 (Russian). Translation in St. Petersburg Math. J. 14 (2003), no. 3, 467–478. MR1921991

[LMS18] M. Liero, A. Mielke, and G. Savaré, *Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures*, Invent. Math. **211** (2018), no. 3, 969–1117. MR3763404

[LV09] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2) **169** (2009), no. 3, 903–991.

[Mai11] E. Mainini, *A description of transport cost for signed measures*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **390** (2011), no. Teoriya Predstavleniĭ, Dinamicheskie Sistemy, Kombinatornye Metody. XX, 147–181, 308–309. MR2870233

[Man89] M. Mandelkern, *Metrization of the one-point compactification*, Proc. Amer. Math. Soc. **107** (1989), no. 4, 1111–1115. MR991703

[Mon81] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, Histoire de l'Académie Royale des Sciences de Paris (1781).

[MR92] Z. M. Ma and M. Röckner, *Introduction to the theory of (non-symmetric) Dirichlet forms*, Universitext (1979), Springer-Verlag, 1992.

[Ott01] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174. MR1842429

[Pau05] G. Paulik, *Gluing spaces and analysis*, Bonner Mathematische Schriften [Bonn Mathematical Publications], vol. 372, Universität Bonn, Mathematisches Institut, Bonn, 2005. Dissertation, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, 2005. MR2204257

[Per91] G. Perelman, *Alexandrov's spaces with curvatures bounded from below II*, unpublished preprint (1991), available at `https://anton-petrunin.github.io/papers/alexandrov/perelmanASWCBFB2+.pdf`.

[Pet97] A. Petrunin, *Applications of quasigeodesics and gradient curves*, Comparison geometry (Berkeley, CA, 1993–94), 1997, pp. 203–219. MR1452875

[Pog73] A. V. Pogorelov, *Extrinsic geometry of convex surfaces*, American Mathematical Society, Providence, R.I., 1973. Translated from the Russian by Israel Program for Scientific Translations, Translations of Mathematical Monographs, Vol. 35. MR0346714

[PR14] B. Piccoli and F. Rossi, *Generalized Wasserstein distance and its application to transport equations with source*, Arch. Ration. Mech. Anal. **211** (2014), no. 1, 335–358. MR3182483

[PS18] A. Profeta and K.-T. Sturm, *Heat Flow with Dirichlet Boundary Conditions via Optimal Transport and Gluing of Metric Measure Spaces*, arXiv e-prints (2018Sep), arXiv:1809.00936, available at `1809.00936`.

[PV16] S. Pigola and G. Veronelli, *The smooth Riemannian extension problem*, arXiv e-prints (2016Jun), arXiv:1606.08320, available at `1606.08320`.

[Riz18] L. Rizzi, *A counterexample to gluing theorems for MCP metric measure spaces*, Bull. Lond. Math. Soc. **50** (2018), no. 5, 781–790. MR3873493

[RS14] T. Rajala and K.-T. Sturm, *Non-branching geodesics and optimal maps in strong* $CD(K,\infty)$*-spaces*, Calc. Var. Partial Differential Equations **50** (2014), no. 3-4, 831–846. MR3216835

[Sav14] G. Savaré, *Self-improvement of the Bakry-Émery condition and Wasserstein contraction of the heat flow in* $\mathrm{RCD}(K,\infty)$ *metric measure spaces*, Discrete Contin. Dyn. Syst. **34** (2014), no. 4, 1641–1661.

[Sch12] A. Schlichting, *Gluing Riemannian manifolds with curvature operators at least k*, ArXiv e-prints (October 2012), available at `1210.2957`.

[Sch14] A. Schlichting, *Smoothing singularities of riemannian metrics while preserving lower curvature bounds*, Ph.D. Thesis, 2014.

[Stu06a] K.-T. Sturm, *On the geometry of metric measure spaces. I*, Acta Math. **196** (2006), no. 1, 65–131.

[Stu06b] ———, *On the geometry of metric measure spaces. II*, Acta Math. **196** (2006), no. 1, 133–177.

[Stu18] ———, *Ricci tensor for diffusion operators and curvature-dimension inequalities under conformal transformations and time changes*, J. Funct. Anal. **275** (2018), no. 4, 793–829. MR3807777

[Vil03] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. MR1964483

[Vil09] C. Villani, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer-Verlag, Berlin, 2009.

[vRS05] M.-K. von Renesse and K.-T. Sturm, *Transport inequalities, gradient estimates, entropy, and Ricci curvature*, Comm. Pure Appl. Math. **58** (2005), no. 7, 923–940.