Ricci curvature and gradient flows of the entropy for jump processes

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

vorgelegt von
Matthias Erbar
aus
Koblenz

Bonn, September 2012

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

1. Gutachter: Prof. Dr. Karl-Theodor Sturm
2. Gutachter: Prof. Dr. Moritz Kaßmann

Tag der Promotion: 26.02.2013
Erscheinungsjahr: 2013

Contents

Preface

This thesis is concerned with the application of ideas from optimal transport to jump processes. The first chapter gives an introductory survey of the topics of this thesis. The later chapters are each self contained and treat results of different flavours. Chapter 2 deals with a notion of Ricci curvature that applies to finite Markov chains. The results presented here have been obtained in collaboration with Jan Maas and have been published in [EM12]. Chapter 3 deals with a gradient flow interpretation of evolution equations driven by non-local Lévy operators. Most of the results presented here have been published in the preprint [Erb12].

This seems to be the place for some brief personal words. First, I would like to thank my adviser Prof. Karl-Theodor Sturm. I benefited a lot from many discussions with him and I am grateful for his constant encouragement and support. I really enjoyed working in his group during the last years.
I would also like to thank Prof. Moritz Kaßmann, Prof. Werner Ballmann and Prof. Rolf Klein for being part of the jury.
I am very grateful to Jan Maas for the close collaboration which in particular lead to results in the first part of this thesis. It is a great pleasure to discuss and work with him – despite some differences when it comes to sports.
A big thanks is due to my colleagues in the probability groups in Bonn for creating such a warm and friendly atmosphere. Especially, I want to thank my office mate Martin Huesmann for sharing so many good and bad experiences in the life of a PhD student – and for all the tea.
Most of all, I am indebted to my family and friends, especially to my wife Michaela, for many things.
Finally, I would like to thank the Bonn International Graduate School in Mathematics and the Collaborative Research Center SFB 611 for their financial and non-financial support.

Summary

In the first part of this thesis, we present a new notion of Ricci curvature that applies to finite Markov chains. This notion relies on geodesic convexity of the entropy and is analogous to the one introduced by Lott–Villani and Sturm for geodesic metric measure spaces. In order to apply to the discrete setting the role of the L^2-Wasserstein distance is taken over by a different metric $\mathcal{W}$ on the space of probability measures having the property that the continuous time Markov chain is the gradient flow of the entropy.

Using this notion of Ricci curvature we prove discrete analogues of fundamental results by Bakry–Émery and Otto–Villani. In particular, we show that Ricci curvature bounds imply a number of functional inequalities for the invariant measure of the Markov chain. These include a modified logarithmic Sobolev inequality and a Talagrand-type transport inequality involving the distance $\mathcal{W}$.

Moreover, we prove that Ricci curvature bounds are stable under tensorisation. As a special case we obtain the sharp Ricci curvature bound for the simple random walk on the discrete hypercube $\{0,1\}^n$.

In the second part, we take a similar approach towards jump processes on $\mathbb{R}^d$. We introduce a new transport distance $\mathcal{W}$ between probability measures on $\mathbb{R}^d$ that is built from a jump kernel $J(x, \mathrm{d}y)$ of Lévy measures. It is defined via a non-local variant of the dynamical characterisation of the L^2-Wasserstein distance. We study geometric and topological properties of the distance $\mathcal{W}$. In particular, we prove that every pair of probability measures at finite distance can be connected by a geodesic.

We put particular focus on translation invariant jump kernels $J(x, A) = \nu(A - x)$ and consider the non local operator $\mathcal{L}$ given by

$$\mathcal{L}u(x) \;=\; \int \frac{1}{2}\big(u(x+z) + u(x-z) - 2u(x)\big)\nu(\mathrm{d}z)\;,$$

with a symmetric Lévy measure ν on $\mathbb{R}^d$. $\mathcal{L}$ is the generator of a pure jump Lévy process. We prove that the semigroup generated by this non-local operator is the gradient flow of the relative entropy with respect to the distance $\mathcal{W}$. This is reminiscent of the Jordan–Kinderlehrer–Otto interpretation of the heat equation as the gradient flow of the entropy w.r.t. the L^2-Wasserstein distance. Moreover, we show that the entropy is convex along $\mathcal{W}$-geodesics.

As a special case, we obtain a gradient flow characterisation of the semigroup generated by the fractional Laplacian $\mathcal{L} = -(-\Delta)^{\frac{\alpha}{2}}$.

1 Introduction

Jump processes and non-local operators arise naturally in various areas of mathematics and become increasingly important in applications ranging from discrete structures with finite Markov chains in computer science and statistical mechanics to the use of Lévy jump processes in physical modelling and mathematical finance.

In this thesis we will focus on two seemingly very different aspects of jump processes: the geometry of graphs induced by finite Markov chains and evolution equations driven by non-local operators in $\mathbb{R}^d$.

In the last two decades the theory of optimal transport emerged as a powerful tool both to study the geometry of non-smooth spaces and to study evolution PDEs associated to diffusions. On one hand, the Wasserstein geometry on the space of probability measures encodes curvature information about the underlying space via geodesic convexity of the entropy. On the other hand, diffusion equations evolve as gradient flows of entropy functionals w.r.t. the Wasserstein distance. Unfortunately, discrete spaces and non-local evolutions remain inaccessible to the existing theory.

In this thesis, we establish a link between ideas from optimal transport and jump processes by using a new non-local transport distance. We show that evolution equations driven by non-local Lévy operators are gradient flows of the entropy w.r.t. the new distance. In the discrete setting we use similar transport metrics to give a notion of Ricci curvature that applies to finite Markoc chains.

In this first chapter we give a survey-style introduction to the topics of this thesis. Our intention is twofold: on one hand, we want to motivate our interest in a link between optimal transport and jump processes and non-local operators. On the other hand, we want to place the results obtained in this thesis into the framework of the existing mathematical literature.

We will first give a short introduction to optimal transport and survey important results of the theory. In particular, we will highlight the connection of optimal transport to geometry and to partial differential equations and discuss the problems arising when dealing with discrete structures. In this context we will then present the main results of this thesis. A precise statement of the main results and a more detailed summary will be given at the beginning of Chapters 2 and 3, respectively.

1.1 Optimal transport

The theory of optimal transport originated in the work of Monge [Mon81] on a civil engineering problem. He asked for an optimal way of moving a pile of building material (e.g. sand) into a given configuration (say a sandcastle), minimising a certain cost given by the total transport distance. In modern terms his problem can be formulated as follows. Given two probability measures μ and ν on $\mathbb{R}^d$, find a minimiser of

$$\inf_{T} \int_{\mathbb{R}^d} |x - T(x)| \, \mathrm{d}\mu(x) \ ,$$

where the infimum is taken over all maps $T : \mathbb{R}^d \to \mathbb{R}^d$ pushing μ forward to ν, i.e. $T_{\#}\mu = \nu$ or $\nu(A) = \mu(T^{-1}(A))$ for all Borel sets $A \subset \mathbb{R}^d$.

This problem turned out to be extremely difficult, in particular because the constraint on T is highly non-linear. Moreover, there need not even be an admissible map, as is easily seen by considering the case where μ is a Dirac measure and ν is not.

Only more than 150 years later Kantorovich gave a first satisfactory answer (see [Kan06] for an English translation of the Russian article from 1942). He suggested the following relaxation of the transport problem with a general cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ looking for a minimiser of

$$\inf_{q} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \mathrm{d}q(x, y) \ ,$$

where now the infimum is taken over all couplings of μ and ν, i.e. probability measures q on $\mathbb{R}^d \times \mathbb{R}^d$ whose first and second marginal are μ and ν respectively. This is a linear constraint on q and if a transport map T exists, the measure $q = (Id, T)_{\#}\mu$ constitutes a coupling. Under suitable assumptions on the cost function c existence of a minimiser follows by direct methods.

These results have found numerous applications in mathematics and economics, we refer to [RR98] for a survey. Starting with the contributions of Brenier, Gangbo, McCann, Rüschendorf and others the theory of optimal transports has undergone a massive development, attracting a broad interest. Independently, Brenier [Bre91] and Rachev and Rüschendorf [RR90] solved the Monge problem with squared distance cost $c(x, y) = |x - y|^2$. They showed that if μ is absolutely continuous w.r.t. Lebesgue measure, then the optimal coupling is unique and given by a transport map $T = \nabla\varphi$ which is the gradient of a convex function. Their results are based on the powerful duality theory of Kantorovich and a fine analysis of the structure of optimal couplings. They have been extended by McCann [McC01] to optimal transport on Riemannian manifolds and to more general cost functions by several authors, we refer to the review [GM96].

Of particular importance for us later on will be a connection of the L^2-transport problem to continuum mechanics. Benamou and Brenier [BB00] gave a dynamical characterisation of the L^2-transport cost between two probabilities $\mu_0 = \rho_0(x)\mathrm{d}x$ and $\mu_1 = \rho_1(x)\mathrm{d}x$ on $\mathbb{R}^d$, showing that

$$\inf_{T_{\#}\mu_0 = \mu_1} \int_{\mathbb{R}^d} |x - T(x)|^2 \rho_0(x)\mathrm{d}x \ = \ \inf_{\rho, \psi} \int_0^1 \int_{\mathbb{R}^d} |\nabla\psi_t(x)|^2 \rho_t(x)\mathrm{d}x\mathrm{d}t \ , \qquad (1.1.1)$$

where the infimum on the right-hand side is taken over all sufficiently smooth functions $\rho : [0,1] \times \mathbb{R}^d \to \mathbb{R}_+$ and $\psi : [0,1] \times \mathbb{R}^d \to \mathbb{R}$ subject to the continuity equation

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \nabla \psi) \ = \ 0 \ , \\ \rho(0,\cdot) = \rho_0 \ , \ \rho(1,\cdot) = \rho_1 \ . \end{cases} \tag{1.1.2}$$

Intuitively, if we picture ρ_0 and ρ_1 as two configurations of a gas of particles, then a solution to the continuity equation corresponds to an evolution of the gas, where a particle at x at time t moves with velocity $\nabla \psi_t(x)$. Thus the minimal L^2-transport cost is given by the minimal total kinetic energy required to let ρ_0 evolve into ρ_1.

The optimal transport problem is also used to endow the space of probability measures over a (polish) metric space (X,d) with a distance function. We consider for $p \in [1,\infty)$ the *L^p-Wasserstein distance*

$$W_p(\mu,\nu)^p \ := \ \inf_q \int\limits_{X \times X} d(x,y)^p q(\mathrm{d}x, \mathrm{d}y) \ .$$

It is finite on the set $\mathscr{P}_p(X)$ of all probability measures μ with finite p-th moments, i.e. $\int d(x_0,x)^p \mu(\mathrm{d}x) < \infty$ for some (hence any) $x_0 \in X$ and we call the pair $(\mathscr{P}_p(X), W_p)$ the *L^p-Wasserstein space*. It inherits many geometric properties of the underlying space X. For example, let X be a geodesic space, i.e. any $x,y \in X$ can be connected by a curve $\gamma : [0,1] \to X$ with $\gamma(0) = x, \gamma(1) = y$ satisfying $d(\gamma(t),\gamma(s)) = |t-s|\, d(x,y)$ for all $s,t \in [0,1]$. Then also $\mathscr{P}_p(X)$ for $p > 1$ is a geodesic space. In fact, each geodesic in $\mathscr{P}_p(X)$ for $p > 1$ can be represented as a probability measure on the geodesics in X such that the joint law of the initial and final point constitutes a coupling of μ_0 and μ_1.

Due to Brenier's theorem we can give an explicit description of geodesics in $\mathscr{P}_2(\mathbb{R}^d)$. Given two measures $\mu_0 = \rho_0(x)\mathrm{d}x$ and μ_1 the curve

$$\mu_t = \big((1-t)Id + t\nabla\varphi\big)_{\#}\mu_0 \ ,$$

where $\nabla\varphi$ is the optimal map pushing μ_0 to μ_1, is a geodesic. This curve is called the *displacement interpolation* between ρ_0 and ρ_1. In his thesis [McC97] McCann made the important observation that certain functionals on probability measures are *displacement convex*, i.e. convex along any displacement interpolation curve. Important examples are the Boltzmann entropy and the potential energy functional defined for a measure $\mu(\mathrm{d}x) = \rho(x)\mathrm{d}x$ by

$$\mathcal{H}(\mu) \ = \ \int \rho(x)\log\rho(x)\mathrm{d}x \ , \quad \mathcal{V}(\mu) = \int V(x)\mu(\mathrm{d}x) \ ,$$

where $V : \mathbb{R}^d \to \mathbb{R}$ is a convex potential. McCann used this result to prove uniqueness of minimisers of such functionals.

1.2 The geometry of metric measure spaces

The discovery of displacement convexity revealed a deep connection between optimal transport and geometry and gave rise to a fascinating development in the study of non-smooth spaces that we shall now describe.

The key observation is a generalisation of McCann's displacement convexity to optimal transport on a Riemannian manifold $\mathcal{M}$, proved by von Renesse and Sturm [RS05] building on earlier work by Cordero-Erausquin, McCann and Schmuckenschläger [CEMS01]. Namely, they showed that the Ricci curvature of $\mathcal{M}$ is bounded below by some constant $K \in \mathbb{R}$ if and only if the entropy is K-convex along geodesics in the Wasserstein space $\mathscr{P}_2(\mathcal{M})$, i.e.

$$\mathcal{H}(\mu_t) \;\leq\; (1-t)\mathcal{H}(\mu_0) + t\mathcal{H}(\mu_1) - \frac{K}{2}t(1-t)W_2(\mu_0,\mu_1)^2 \;.$$

The reason for this is the fact that Ricci curvature controls the distortion of volume elements that are transported along geodesics. More generally, the relative entropy with respect to a weighted volume measure given for $\mu = \rho\,\mathrm{vol}$ by $\mathcal{H}(\mu|e^{-V}\,\mathrm{vol}) = \int \rho\log\rho \,\mathrm{d}\,\mathrm{vol} + \int V\rho \,\mathrm{d}\,\mathrm{vol}$ is K-convex along geodesics if and only if the so-called Bakry–Émery tensor (see [BÉ85]) is bounded below:

$$\mathrm{Ric} + \mathrm{Hess}\, V \geq K \;.$$

Note that the condition of geodesic convexity of the entropy does not use the differential structure of $\mathcal{M}$ but only requires a metric to define the L^2-Wasserstein distance and a reference measure to define the entropy. Therefore this condition can be used to define a notion of Ricci curvature lower bound on a large class of metric measure spaces. This allowed Sturm [Stu06] and Lott and Villani [LV09] in two independent contributions to solve the long-standing open problem of finding a synthetic notion of Ricci curvature for non-smooth spaces. By considering more refined convexity properties of the entropy it is possible to give a condition which combines a lower bound on the Ricci curvature with an upper bound on the dimension.

One of the remarkable features of this *curvature-dimension condition* for metric measure spaces is that it is stable under Gromov–Hausdorff convergence, which makes this theory a good framework for studying non-smooth spaces that arise as limits of Riemannian manifolds with uniform curvature and dimension bounds. Moreover, the curvature-dimension condition implies a large number of geometric and functional inequalities such as a Poincaré inequality, a logarithmic Sobolev inequality or a Brunn–Minkowski inequality. This allows one to generalise many well known results from comparison geometry to metric measure spaces, such as Bishop–Gromov volume comparison or a Bonnet–Myers diameter bound.

The theory of metric measure spaces with Ricci curvature bounds in the sense of Lott–Villani and Sturm is still under active development, see e.g. the recent preprints [AGS11a, AGS11b].

1.3 The geometry of diffusion equations

A second striking application of optimal transport that we want to highlight is its link with evolution partial differential equations. This connection will be a recurrent theme in this thesis.

In a seminal paper [JKO98], Jordan, Kinderlehrer and Otto gave a new interpretation of the Fokker–Planck equation

$$\partial_t \rho \;=\; \Delta \rho + \operatorname{div}(\rho \nabla V)\ .$$

They showed that it can be viewed as the gradient flow of the entropy with respect to the L^2-Wasserstein distance on the space of probability measures. Later, Otto [Ott01] gave a similar interpretation of the porous medium equation. Along with this discovery, he developed a powerful geometric picture viewing the Wasserstein space as a formal Riemannian manifold. This intuition, often called "Otto calculus", is based on the dynamical characterisation of the L^2-Wasserstein distance (1.1.1) and allows one to heuristically derive results about the PDE in question by performing Riemannian computations on the Wasserstein space.

In order to give a rigorous meaning to the gradient flow interpretation of a diffusion equation one has to define a notion of gradient flow in the metric space $\mathscr{P}_2(\mathbb{R}^d)$. There are several concepts that generalise gradient flows to a metric setting, for a detailed study we refer to the monograph [AGS08]. Here, we want to highlight one concept that we will use in the sequel and that gives a very strong characterisation of gradient flows of geodesically (semi-)convex functionals. To motivate this concept, consider a convex function $F : \mathbb{R}^d \to \mathbb{R}$. It is easy to see that a smooth curve $u : [0,\infty) \to \mathbb{R}^d$ satisfies the gradient flow equation $\dot{u}(t) = -\nabla F(u(t))$ if and only if it satisfies the following set of inequalities:

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left|u(t) - y\right|^2 \le F(y) - F(u(t)) \quad \forall y \in \mathbb{R}^d\ .$$

The latter condition only appeals to the function F and the metric $|\cdot|$, hence it can be taken as a definition of gradient flow in a metric space. More precisely, the result of Jordan, Kinderlehrer and Otto can be rephrased as follows. The solution (ρ_t) to the Fokker–Planck equation is the gradient flow of the entropy in the Wasserstein space $\mathscr{P}_2(\mathbb{R}^d)$, in the sense that it satisfies the so-called *Evolution Variational Inequality*, short EVI,

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} W_2(\rho_t,\sigma)^2 \le \mathcal{H}(\sigma) - \mathcal{H}(\rho_t) \quad \forall \sigma \in \mathscr{P}_2(\mathbb{R}^d)\ . \tag{1.3.1}$$

In fact, in [JKO98] the authors use a time discrete approximation scheme to characterise the gradient flow, but the Evolution Variational Inequality is already implicit in their work.

The discovery of Jordan, Kinderlehrer and Otto has been the starting point for many developments in evolution equations, probability and geometry. For an overview we refer to the monographs [AGS08, Vil09]. In particular, the gradient flow approach has proved very successful in studying diffusion processes and the PDEs associated to them. Let us highlight some results. Already in Otto's original work [Ott01], the characterisation of the porous medium equation as the gradient flow of a strictly convex functional has been used to derive exponential convergence rates in the asymptotic behaviour. This approach also yields nice geometric heuristics for famous results of Bakry–Émery [BÉ85] and Otto–Villani [OV00], which were then proved by PDE methods. In the former work it is shown that on a Riemannian manifold a lower bound on the weighted Ricci curvature $\mathrm{Ric} + \mathrm{Hess}\, V$ implies a logarithmic Sobolev inequality for the measure $e^{-V}\,\mathrm{vol}$. In the latter, the authors show that a logarithmic Sobolev inequality implies Talagrand's transport inequality. Many of these results have been generalised to the setting of metric measure spaces with curvature bounds as noted in section 1.2. In Chapter 2 we will derive discrete analogues of these results in the setting of finite Markov chains.

Besides being a source of valuable intuition the theory of gradient flows in metric spaces itself has become a powerful tool in analysing evolution partial differential equations. Constructing the metric gradient flow can yield existence of weak solutions, a strategy that has been applied e.g. by Matthes, McCann and Savaré [MMS09] to a class of nonlinear fourth order equations. Recently, it has been shown that this theory also provides a robust framework to study stability properties of diffusion processes. Ambrosio, Savaré and Zambotti [ASZ09] consider stochastic partial differential equations of Fokker–Planck type with log-concave invariant measure and show stability in law with respect to changes in the potential.

By now, the gradient flow interpretation of the heat flow has been extended to a broad variety of settings including Riemannian manifolds [Erb10, Vil09], Finsler spaces [OS09], Aleksandrov spaces [GKO10], Wiener spaces [FSS10] and metric measure spaces with Ricci curvature bounds [AGS11a, AGS11b]. In the latter case, the twofold description of the heat flow as the L^2 gradient flow of the Dirichlet energy and the Wasserstein gradient flow of the entropy is a major tool in developing a powerful calculus on metric measure spaces.

1.4 Discrete spaces

In Section 1.2 we have already seen some of the strong results in terms of functional inequalities that can be obtained on a large class of metric measure spaces using the theory of synthetic curvature bounds, based on optimal transport. In many applications the spaces of interest are discrete, such as graphs, and the relevant process is a Markov chain. A prominent example is the discrete hypercube $\{0,1\}^n$ which is the fundamental building block of many models in computer science and statistical physics.

Unfortunately, the curvature condition of Lott–Sturm–Villani typically does not apply to discrete spaces. This is due to the fact that the L^2-Wasserstein space over a graph X equipped with the graph distance does not contain geodesics. To see this, consider e.g. the graph consisting of two points a, b and a single edge $\{a, b\}$. Every probability measure is then of the form $\mu_\beta = \beta\delta_a + (1-\beta)\delta_b$ with $\beta \in [0,1]$. Now if $(\mu_{\beta(t)})_{t\in[0,1]}$ is a constant speed geodesic we have

$$|t-s|\, W_2(\mu_{\beta(0)}, \mu_{\beta(1)}) = W_2(\mu_{\beta(s)}, \mu_{\beta(t)}) = \sqrt{|\beta(t) - \beta(s)|}\ .$$

Hence β is 2-Hölder and thus constant. So every constant speed W_2-geodesic must be constant.

As a consequence, several other concepts of curvature for discrete spaces have been proposed in the literature. One of the most prominent is a notion of Ricci curvature that applies to Markov chains on metric spaces and was introduced by Ollivier [Oll07, Oll09]. This notion is also based on ideas from optimal transport and uses the L^1-Wasserstein metric W_1, which behaves better in a discrete setting than W_2. Ollivier's criterion has the advantage of being easy to check in many examples. Furthermore, in some interesting cases it yields functional inequalities with good – yet non-optimal – constants. It is not completely clear how Ollivier's notion relates to the one by Lott–Sturm–Villani (see [OV10] for a discussion).

In the setting of graphs, Ollivier's Ricci curvature has been further studied in the recent preprints [BJL11, HJL11, JL11].

Another approach has been taken by Lin and S.T. Yau [LY10], who defined Ricci curvature in terms of the heat semigroup, i.e. the semigroup associated to the continuous time random walk on a graph.

Bonciocat and Sturm [BS09] followed a different approach to modify the Lott–Sturm–Villani criterion, in which they circumvented the lack of midpoints in the L^2-Wasserstein metric by allowing for approximate midpoints. A Brunn–Minkowski inequality in this spirit has been proved on the discrete hypercube by Ollivier and Villani [OV10].

In a recent preprint, Gozlan, Roberto, Samson and Tetali [GSRT12] study convexity of the entropy along a certain class of interpolating paths in the space of probability measures over a finite graph, which generalises W_1-geodesics. The authors prove a tensorisation property and also derive different functional inequalities with optimal constants in important examples.

1.5 Finite Markov chains and non-local transport

In view of the great generality in which a gradient flow interpretation of the heat flow has been obtained, it seems natural to ask whether a similar result also holds in a discrete setting. We shall now discuss a recent result by Maas [Maa11] in which an extension of the gradient flow approach to finite Markov chains has been obtained.

Consider a finite set $\mathcal{X}$ equipped with an irreducible, reversible Markov kernel K. The law ρ_t of the continuous time Markov chain associated to K evolves according to the equation $\dot\rho = (K - I)\rho$, which can be regarded as the heat equation on $\mathcal{X}$.

In fact, $(K - I)$ is the Laplacian associated to the weighted graph structure induced by the kernel K. This gives rise to a semi group $(P_t)_{t\geq 0}$ given by $P_t = e^{t(K-I)}$ which we view as the “heat semigroup” on $\mathcal{X}$. Therefore, one can ask, whether this semigroup can also be characterised as the gradient flow of the entropy with respect to a suitable metric on the space of probability measures on $\mathcal{X}$, where the unique stationary measure of the Markov chain is the natural reference measure on $\mathcal{X}$.

Unfortunately, it turns out that also here the L^2-Wasserstein distance (build from the graph distance) is not appropriate for this purpose. Indeed, the derivative of W_2 along the semigroup P_t will typically be infinite (this can easily be seen on a similar example as above by considering a Markov chain on just two points). Thus, the Evolution Variational Inequality (1.3.1) for W_2 breaks down.

However, Maas [Maa11] showed that the heat semigroup can still be seen as the gradient flow of the entropy if the Wasserstein metric is replaced by a different metric $\mathcal{W}$ on the space of probability measures over the set $\mathcal{X}$. This new metric is constructed by Maas via a discrete analogue of the Benamou–Brenier formula (1.1.1). It is a dynamic transport distance which is non-local in the sense that the cost of moving mass between two points x, y depends on the amount of mass present at x and y. We will discuss this distance in detail in Chapter 2. A remarkable feature of the distance $\mathcal{W}$ is the fact that, unlike the Wasserstein distance, it does not require a metric on the underlying set $\mathcal{X}$ but only the Markov kernel K.

A similar gradient structure has been discovered independently both by Mielke [Mie11] in the setting of reaction-diffusion systems and by Chow et al. [CHLZ12] for Fokker–Planck equations on graphs. In two recent independent preprints Carlen and Maas as well as Mielke use a similar approach to give a gradient flow interpretation of dissipative quantum systems [CM12, Mie12a].

1.6 The results of Chapter 2

The first part of this thesis will be concerned with the application of ideas from optimal transport to the study of discrete spaces. We will present a new notion of Ricci curvature that applies to finite Markov chains.

In a continuous setting, the Wasserstein distance and the entropy have a twofold significance: they encode important geometric information about the underlying space through the powerful Lott–Sturm–Villani theory of synthetic curvature bounds and they characterise the heat flow as the Wasserstein gradient flow of the entropy. Unfortunately, as we saw in Sections 1.4 and 1.5, both the curvature condition of Lott–Sturm–Villani and the gradient flow approach based on the L^2-Wasserstein distance do not apply to discrete spaces, the former being a consequence of the non-existence of W_2-geodesics.

The aim of Chapter 2 is to develop a variant of the theory of Lott–Sturm–Villani, which does apply to discrete spaces. We consider an irreducible, reversible Markov kernel K on a finite set $\mathcal{X}$ with invariant measure π. Our approach is based on a different geometry on the space of probability measures determined by the metric $\mathcal{W}$ that was introduced in [Maa11] and depends on the kernel K. Since the heat flow associated to K is the gradient flow of the entropy with respect to $\mathcal{W}$, this new metric naturally takes over the role of the L^2-Wasserstein distance.

We will show that every pair of probability densities on $\mathcal{X}$ can be joined by a $\mathcal{W}$-geodesic. Hence, it is possible to define a notion of Ricci curvature in the spirit of Lott–Sturm–Villani by requiring geodesic (semi-)convexity of the entropy with respect to the metric $\mathcal{W}$. This possibility has already been indicated in [Maa11]. Owing to the non-local character of the distance $\mathcal{W}$, we call this notion *non-local* Ricci curvature. Convexity along $\mathcal{W}$-geodesics may be regarded as a discrete analogue of McCann's displacement convexity [McC97], which corresponds to convexity along W_2-geodesics in a continuous setting.

We shall show that this new notion of Ricci curvature shares a number of properties which make the Lott–Sturm–Villani theory so powerful. For a detailed statement and discussion of our results we refer to Section 2.1. Here, we only highlight the key points. One of our main result is that

- non-local Ricci curvature bounds are stable under tensorisation.

As a consequence, we obtain explicit curvature bounds in important examples. In particular, we obtain

- a lower bound on the non-local Ricci curvature of (the kernel of simple random walk on) the discrete hypercube $\{0,1\}^n$, which turns out to be optimal.

Our second main result is a discrete counterpart of the results by Bakry–Émery [BÉ85] and Otto–Villani [OV00]. Namely, we will prove that

- non-local Ricci curvature bounds imply a number of functional inequalities including a modified logarithmic Sobolev inequality and a Talagrand-type inequality involving the metric $\mathcal{W}$.

In the case of the discrete hypercube, we obtain sharp constants in these inequalities.

As compared to the other notions of Ricci curvature for discrete spaces that we have discussed in Section 1.4 the notion considered here appears to be the closest in spirit to the one by Lott–Sturm–Villani. Furthermore, it seems to be the first that yields natural analogues of the results by Bakry–Émery and Otto–Villani.

Independently to the results obtained in [EM12] geodesic convexity of the entropy for Markov chains has also been studied by Mielke [Mie12b]. There, good convexity estimates are obtained, in particular, for discretisations of one-dimensional Fokker–Planck equations. However, the notion of curvature bounds is not discussed and no consequences in terms of functional inequalities are considered.

1.7 The results of Chapter 3

We have seen in Section 1.3 that in the last two decades optimal transport has been applied very successfully in the study of diffusion processes and PDEs associated to them. In many applications ranging from physical modelling to financial markets we encounter discontinuous stochastic processes that propagate by jumps. Their generators are typically non-local operators.

The aim of the second part of this thesis is to build a bridge between the theory of jump processes and non-local operators on one hand and ideas from optimal transport on the other hand. Our approach is similar in spirit to the work of Maas [Maa11] for finite Markov chains and generalises it to a certain extent.

We will give a gradient flow interpretation of the equation

$$\partial_t u \;=\; \mathcal{L}u\;, \tag{1.7.1}$$

where $\mathcal{L}$ is a non-local operator given by

$$\mathcal{L}u(x) \;=\; \int \frac{1}{2}\big(u(x+z)+u(x-z)-2u(x)\big)\nu(\mathrm{d}z)\;,$$

with a symmetric Lévy measure ν on $\mathbb{R}^d$. Such operators arise naturally as the generators of pure jump Lévy processes, i.e. jump processes with independent and stationary increments. The measure $\nu(\mathrm{d}z)$ gives the intensity of jumps from x to $x+z$. For background on Lévy processes and their generators we refer to the books [Ber96, App04]. A prominent example of a non-local operator that our results will apply to is the fractional Laplacian $\mathcal{L} = -(-\Delta)^{\frac{\alpha}{2}}$ corresponding to the choice $\nu_\alpha(\mathrm{d}z) = c_\alpha\,|z|^{-\alpha-d}\,\mathrm{d}z$ with $\alpha\in(0,2)$. This is a pseudo differential operator with symbol $|\xi|^\alpha$ and the corresponding Lévy process is the α-stable process.

In order to give a gradient flow interpretation to equation (1.7.1) the Wasserstein distance is not appropriate. We take a similar approach as in Chapter 2 and construct a new transport distance $\mathcal{W}$ on the space of probability measures on $\mathbb{R}^d$ that is non-local in nature. It is defined via a non-local variant of the dynamical characterisation of the Wasserstein distance by Benamou and Brenier [BB00]. In fact, the construction of this distance is general and applies also to inhomogeneous jump processes where the intensity of jumps from x to y is given by a space dependent Lévy measure $J(x,\mathrm{d}y)$. We will show that any two probability measures at finite distance can be joined by a $\mathcal{W}$-geodesic.

We will then focus on homogeneous jump kernels corresponding to Lévy processes, i.e. $J(x,\mathrm{d}y)=\nu(\mathrm{d}y-x)$ for a symmetric Lévy measure ν satisfying suitable regularity assumptions. One of the main results of Chapter 3 is that

- the entropy is convex along geodesics of the metric $\mathcal{W}$ built from ν.

This can be see as a non-local analogue of McCann displacement convexity that we encountered in Section 1.1. Moreover, we obtain a gradient flow characterisation of equation 1.7.1. We show that

- the semigroup $P_t = e^{t\mathcal{L}}$ of the pure jump Lévy process generated by the non-local operator $\mathcal{L}$ is the gradient flow of the entropy with respect to $\mathcal{W}$ in the sense of the Evolution Varitational Inequality:

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{W}(P_t\mu,\sigma)^2 \leq \mathcal{H}(\sigma) - \mathcal{H}(P_t\mu) \quad \forall\sigma\ .$$

For a precise statement and detailed discussion of our results we refer to Section 3.1.

Our interest in such a link between jump processes and optimal transport is motivated on one hand by the many applications of the Wasserstein gradient flow approach to evolution PDE. For inhomogeneous jump kernels where the entropy is strictly $\mathcal{W}$-geodesically convex the framework developed here could be used to study the asymptotic behaviour of the associated non-local equation and derive functional inequalities for the equilibrium along the lines of the results in Chapter 2.

On the other hand our approach is also motivated by the regularity theory for elliptic and parabolic equations involving non-local operators which is under active development including both analytic and probabilistic approaches (see e.g. [CS11], [BBCK09] and references therein). In a local setting very precise regularity results can be obtained using a lower bound on the Ricci curvature of the operator in the sense of the Bakry-Émery criterion [BÉ85]. Equivalently, such curvature information can be encoded into convexity properties of the entropy along Wasserstein geodesics. In this sense the approach presented here could be used to define an alternative notion of curvature of a non-local operator in the spirit of Lott–Villani–Sturm that might be more adapted to certain situations than the non-local Γ_2-calculus.

Modifications of the Wasserstein distance in a continuous setting have been considered recently by a number of authors. For example, Buttazzo, Jimenez and Oudet [BJO09] and Brasco, Butazzo, Santambrogio [BBS11] use a Benamou–Brenier type approach to study mass transport with congestion effects in crowd movement or branched transports, respectively. In [DNS09], Dolbeault, Nazaret and Savaré proposed a new class of transport distances also based on an adaptation of the Benamou-Brenier formula to give a gradient flow interpretation to a class of transport equations with non-linear mobilities. Geodesic convexity of the entropy with respect to these distances has been studied in [CLSS10].

The distance $\mathcal{W}$ introduced in Chapter 3 seems to be the first that captures non-local effects and makes the connection to jump processes. The techniques that we use here are inspired by the ones developed in [DNS09].

2 Ricci curvature for finite Markov chains

In this chapter we study a new notion of Ricci curvature that applies to Markov chains on discrete spaces. This notion relies on geodesic convexity of the entropy and is analogous to the one introduced by Lott, Sturm, and Villani for geodesic measure spaces. In order to apply to the discrete setting, the role of the Wasserstein metric is taken over by a different metric, having the property that continuous time Markov chains are gradient flows of the entropy. Using this notion of Ricci curvature we prove discrete analogues of fundamental results by Bakry–Émery and Otto–Villani. Furthermore we show that Ricci curvature bounds are preserved under tensorisation. As a special case we obtain the sharp Ricci curvature lower bound for the discrete hypercube.

2.1 Main results

Let us start by describing our setting and results in more detail. We work with an irreducible Markov kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ on a finite set $\mathcal{X}$, i.e., we assume that

$$\sum_{y\in\mathcal{X}} K(x, y) = 1$$

for all $x \in \mathcal{X}$, and that for every $x, y \in \mathcal{X}$ there exists a sequence $\{x_i\}_{i=0}^n \in \mathcal{X}$ such that $x_0 = x$, $x_n = y$ and $K(x_{i-1}, x_i) > 0$ for all $1 \le i \le n$. Basic Markov chain theory guarantees the existence of a unique stationary probability measure (also called steady state) π on $\mathcal{X}$, i.e.,

$$\sum_{x\in\mathcal{X}} \pi(x) = 1 \qquad \text{and} \qquad \pi(y) = \sum_{x\in\mathcal{X}} \pi(x)K(x, y)$$

for all $y \in \mathcal{X}$. We assume that π is *reversible* for K, which means that the detailed balance equations

$$K(x, y)\pi(x) = K(y, x)\pi(y) \tag{2.1.1}$$

hold for $x, y \in \mathcal{X}$.

Let

$$\mathscr{P}(\mathcal{X}) := \Big\{ \rho : \mathcal{X} \to \mathbb{R}_+ \mid \sum_{x\in\mathcal{X}} \pi(x)\rho(x) = 1 \Big\}$$

be the set of *probability densities* on $\mathcal{X}$. Since by elementary Markov chain theory $\pi(x) > 0$ for all x, we can identify $\mathscr{P}(\mathcal{X})$ with the set of probability measures on

$\mathcal{X}$. The subset consisting of those probability densities that are strictly positive is denoted by $\mathscr{P}_*(\mathcal{X})$. We consider the metric $\mathcal{W}$ defined for $\rho_0, \rho_1 \in \mathscr{P}(\mathcal{X})$ by

$$\mathcal{W}(\rho_0,\rho_1)^2 := \inf_{\rho,\psi}\left\{\frac{1}{2}\int_0^1 \sum_{x,y\in\mathcal{X}} (\psi_t(x)-\psi_t(y))^2\hat{\rho}_t(x,y)K(x,y)\pi(x)\mathrm{d}t\right\},$$

where the infimum runs over all sufficiently regular curves $\rho : [0,1] \to \mathscr{P}(\mathcal{X})$ and $\psi : [0,1] \to \mathbb{R}^{\mathcal{X}}$ satisfying the 'continuity equation'

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}\rho_t(x) + \displaystyle\sum_{y\in\mathcal{X}}(\psi_t(y)-\psi_t(x))\hat{\rho}_t(x,y)K(x,y) \;=\; 0 \qquad \forall x \in \mathcal{X}\,, \\ \rho(0) = \rho_0\,, \qquad \rho(1) = \rho_1\,. \end{cases} \tag{2.1.2}$$

Here, given $\rho \in \mathscr{P}(\mathcal{X})$, we write $\hat{\rho}(x,y) := \int_0^1 \rho(x)^{1-p}\rho(y)^p\mathrm{d}p$ for the logarithmic mean of $\rho(x)$ and $\rho(y)$. The relevance of the logarithmic mean in this setting is due to the identity

$$\rho(x) - \rho(y) = \hat{\rho}(x,y)(\log\rho(x) - \log\rho(y))\,,$$

which somewhat compensates for the lack of a 'discrete chain rule'. The definition of $\mathcal{W}$ can be regarded as a discrete analogue of the Benamou–Brenier formula 1.1.1 (see [BB00]). Let us remark that if $t \mapsto \rho_t$ is differentiable at some t and ρ_t belongs to $\mathscr{P}_*(\mathcal{X})$, then the continuity equation (2.1.2) is satisfied for some $\psi_t \in \mathbb{R}^{\mathcal{X}}$, which is unique up to an additive constant (see [Maa11, Proposition 3.26]).

It has been proved in [Maa11] that the interior $\mathscr{P}_*(\mathcal{X})$ equipped with the distance $\mathcal{W}$ is a Riemannian manifold and that the heat flow associated with the continuous time Markov semigroup $P_t = e^{t(K-I)}$ is the gradient flow of the entropy

$$\mathcal{H}(\rho) = \sum_{x\in\mathcal{X}} \pi(x)\rho(x)\log\rho(x)\,, \tag{2.1.3}$$

with respect to the Riemannian structure determined by $\mathcal{W}$.

We shall show that every pair of densities $\rho_0, \rho_1 \in \mathscr{P}(\mathcal{X})$ can be joined by a constant speed geodesic. Therefore the following definition in the spirit of Lott–Sturm–Villani seems natural.

Definition 2.1.1. *We say that K has* non-local Ricci curvature bounded from below *by $\kappa \in \mathbb{R}$ if for any constant speed geodesic $\{\rho_t\}_{t\in[0,1]}$ in $(\mathscr{P}(\mathcal{X}), \mathcal{W})$ we have*

$$\mathcal{H}(\rho_t) \le (1-t)\mathcal{H}(\rho_0) + t\mathcal{H}(\rho_1) - \frac{\kappa}{2}t(1-t)\mathcal{W}(\rho_0,\rho_1)^2\,.$$

In this case, we shall use the notation

$$\mathrm{Ric}(K) \ge \kappa\,.$$

Remark 2.1.2. Instead of requiring convexity along all geodesics it will be shown to be equivalent to require that every pair of densities $\rho_0, \rho_1 \in \mathscr{P}(\mathcal{X})$ can be joined by a constant speed geodesic along which the entropy is κ-convex. Another equivalent condition would be to impose a lower bound on the Hessian of $\mathcal{H}$ in the interior $\mathscr{P}_*(\mathcal{X})$ (see Theorem 2.4.5 below for the details).

One of the main results of this chapter is a tensorisation result for non-local Ricci curvature, which we will now describe. For $1 \leq i \leq n$, let K_i be an irreducible and reversible Markov kernel on a finite set $\mathcal{X}_i$, and let π_i denote the corresponding invariant probability measure. Let $K_{(i)}$ denote the lift of K_i to the product space $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$, defined for $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ by

$$K_{(i)}(\mathbf{x}, \mathbf{y}) = \begin{cases} K_i(x_i, y_i), & \text{if } x_j = y_j \text{ for all } j \neq i, \\ 0, & \text{otherwise.} \end{cases}$$

For a sequence $\{\alpha_i\}_{1 \leq i \leq n}$ of non-negative numbers with $\sum_{i=1}^n \alpha_i = 1$, we consider the weighted product chain, determined by the kernel

$$K_\alpha := \sum_{i=1}^n \alpha_i K_{(i)} \ .$$

Its reversible probability measure is the product measure $\pi = \pi_1 \otimes \cdots \otimes \pi_n$.

Theorem 2.1.3 (Tensorisation of Ricci bounds)**.** *Assume that* $\mathrm{Ric}(K_i) \geq \kappa_i$ *for* $i = 1, \ldots, n$. *Then we have*

$$\mathrm{Ric}(K_\alpha) \ \geq \ \min_i \alpha_i \kappa_i \ .$$

Tensorisation results have also been obtained for other notions of Ricci curvature, including the ones by Lott–Sturm–Villani [Stu06, Proposition 4.16] and Ollivier [Oll09, Proposition 27]. In both cases the proof does not extend to our setting, and completely different ideas are needed here.

As a consequence, we obtain a lower bound on the non-local Ricci curvature for (the kernel K_n of the simple random walk on) the discrete hypercube $\{0,1\}^n$, which turns out to be optimal.

Corollary 2.1.4. *For* $n \geq 1$ *we have* $\mathrm{Ric}(K_n) \geq \frac{2}{n}$.

The hypercube is a fundamental building block for applications in mathematical physics and theoretical computer science, and the problem of proving "displacement convexity" on this space has been an open problem that motivated the recent paper by Ollivier and Villani [OV10], in which a Brunn–Minkowski inequality was obtained.

The second main result of this chapter is the fact that non-local Ricci bounds imply a number of functional inequalities, which are natural discrete counterparts to powerful inequalities in a continuous setting. In particular, we obtain discrete counterparts to the results by Bakry–Émery [BÉ85] and Otto–Villani [OV00].

To state the results we consider the Dirichlet form

$$\mathcal{E}(\varphi,\psi) \;=\; \frac{1}{2}\sum_{x,y\in\mathcal{X}}\big(\varphi(x)-\varphi(y)\big)\big(\psi(x)-\psi(y)\big)K(x,y)\pi(x)$$

defined for functions $\varphi,\psi:\mathcal{X}\to\mathbb{R}$. Furthermore, we consider the functional

$$\mathcal{I}(\rho) \;=\; \mathcal{E}(\rho,\log\rho)$$

defined for $\rho\in\mathscr{P}(\mathcal{X})$, with the convention that $\mathcal{I}(\rho)=+\infty$ if ρ does not belong to $\mathscr{P}_*(\mathcal{X})$. Its significance here is due to the fact that it is the time-derivative of the entropy along the heat flow: $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(P_t\rho)=-\mathcal{I}(P_t\rho)$. In this sense, $\mathcal{I}$ can be regarded as a discrete version of the Fisher information.

Theorem 2.1.5 (Functional inequalities)**.** *Let K be an irreducible and reversible Markov kernel on a finite set $\mathcal{X}$.*

(i) If $\mathrm{Ric}(K)\geq\kappa$ for some $\kappa\in\mathbb{R}$, then the HWI-inequality

$$\mathcal{H}(\rho) \;\leq\; \mathcal{W}(\rho,\mathbf{1})\sqrt{\mathcal{I}(\rho)}-\frac{\kappa}{2}\mathcal{W}(\rho,\mathbf{1})^2 \qquad (\mathrm{H}\mathcal{W}\mathrm{I}(\kappa))$$

holds for all $\rho\in\mathscr{P}(\mathcal{X})$.

(ii) If $\mathrm{Ric}(K)\geq\lambda$ for some $\lambda>0$, then the modified logarithmic Sobolev inequality

$$\mathcal{H}(\rho) \;\leq\; \frac{1}{2\lambda}\mathcal{I}(\rho) \qquad (\mathrm{MLSI}(\lambda))$$

holds for all $\rho\in\mathscr{P}(\mathcal{X})$.

(iii) If K satisfies (MLSI(λ)) *for some $\lambda>0$, then the modified Talagrand inequality*

$$\mathcal{W}(\rho,\mathbf{1}) \;\leq\; \sqrt{\frac{2}{\lambda}\mathcal{H}(\rho)} \qquad (\mathrm{T}_{\mathcal{W}}(\lambda))$$

holds for all $\rho\in\mathscr{P}(\mathcal{X})$.

(iv) If K satisfies ($\mathrm{T}_{\mathcal{W}}(\lambda)$) *for some $\lambda>0$, then the Poincaré inequality*

$$\|\varphi\|^2_{L^2(\mathcal{X},\pi)} \;\leq\; \frac{1}{\lambda}\mathcal{E}(\varphi,\varphi) \qquad (\mathrm{P}(\lambda))$$

holds for all functions $\varphi:\mathcal{X}\to\mathbb{R}$.

Here, $\mathbf{1}$ denotes the density of the stationary measure π.

The first inequality in Theorem 2.1.5 is a discrete counterpart to the HWI-inequality from Otto and Villani [OV00], with the difference that the L^2-Wasserstein metric has been replaced by $\mathcal{W}$.

The second result is as a discrete version of the celebrated criterion by Bakry-Émery [BÉ85], who proved the corresponding result on Riemannian manifolds. Classically, the Bakry-Émery criterion applies to weighted Riemannian manifolds $(\mathcal{M}, e^{-V} \mathrm{vol}_{\mathcal{M}})$, and asks for a lower bound on the generalised Ricci curvature given by $\mathrm{Ric}_{\mathcal{M}} + \mathrm{Hess}\, V$. As in our setting we allow for general K and π, the potential V is already incorporated in K and π, and our notion of Ricci curvature can be thought of as the analogue of this generalised Ricci curvature.

The modified logarithmic Sobolev inequality (MLSI) is motivated by the fact that it yields an explicit rate of exponential decay of the entropy along the heat flow. It has been extensively studied (see, for example, [BT06, CDPP09]), along with different discrete logarithmic Sobolev inequalities in the literature (for example, [AL00, BL98]).

The third part is a discrete counterpart to a famous result by Otto and Villani [OV00], who showed that the logarithmic Sobolev inequality implies the so-called T_2-inequality; recall that the T_p-inequality is the analogue of $\mathrm{T}_{\mathcal{W}}$, in which $\mathcal{W}$ is replaced by the L^p-Wasserstein metric W_p (see Section 1.1 for the definition), for $1 \leq p < \infty$. These inequalities have been extensively studied in recent years. We refer to [GL10] for a survey and to [ST09] for a study of the T_1-inequality in a discrete setting.

The modified Talagrand inequality $\mathrm{T}_{\mathcal{W}}$ that we consider is new. This inequality combines some of the good properties of T_1 and T_2, as we shall now discuss.

Like T_1, it is weak enough to be applicable in a discrete setting. In fact, we shall prove that $\mathrm{T}_{\mathcal{W}}(\lambda)$ holds on the discrete hypercube $\{0,1\}^n$ with the optimal constant $\lambda = \frac{2}{n}$. By contrast, the T_2-inequality does not even hold on the two-point space, and it has been an open problem to find an adequate substitute.

Like T_2, and unlike T_1, $\mathrm{T}_{\mathcal{W}}$ is strong enough to capture spectral information. In fact, the fourth part in Theorem 2.1.5 asserts that it implies a Poincaré inequality with constant λ.

Furthermore, we shall show that $\mathrm{T}_{\mathcal{W}}$ yields good bounds on the sub-Gaussian constant, in the sense that

$$\mathbb{E}_\pi\left[e^{t(\varphi - \mathbb{E}_\pi[\varphi])}\right] \leq \exp\left(\frac{t^2}{4\lambda}\right) \tag{2.1.4}$$

for all $t > 0$ and all functions $\varphi : \mathcal{X} \to \mathbb{R}$ that are Lipschitz constant 1 with respect to the graph norm. Here, we use the notation $\mathbb{E}_\pi[\varphi] = \sum_{x \in \mathcal{X}} \varphi(x)\pi(x)$. As is well known, this estimate yields the concentration inequality

$$\pi\left(\varphi - \mathbb{E}_\pi[\varphi] \geq h\right) \leq e^{-\lambda h^2}$$

for all $h > 0$. The proof of (2.1.4) relies on the fact, proved in Section 3.4, that the metric $\mathcal{W}$ can be bounded from below by W_1 (with respect to the graph metric), so that $\mathrm{T}_{\mathcal{W}}(\lambda)$ implies a $T_1(2\lambda)$-inequality, which is known to be equivalent to the sub-Gaussian inequality [BG99].

The proof of Theorem 2.1.5 follows the approach by Otto and Villani. On a technical level, the proofs are simpler in the discrete case, since heuristic arguments from

Otto and Villani are essentially rigorous proofs in our setting, and no additional PDE arguments are required as in [OV00].

To summarise, we have the following sequence of implications, for any $\lambda > 0$:

$$\mathrm{Ric}(K) \geq \lambda \ \Rightarrow \ \mathrm{MLSI}(\lambda) \ \Rightarrow \ \mathrm{T}_{\mathcal{W}}(\lambda) \ \Rightarrow \ \begin{cases} \mathrm{P}(\lambda) \\ \mathrm{T}_1(2\lambda) \ . \end{cases}$$

Organisation of this chapter

In Section 2.2 we collect basic properties of the metric $\mathcal{W}$ and formulate an equivalent definition that is more convenient to work with in some situations. Geodesics in the $\mathcal{W}$-metric are studied in Section 2.3. In particular, it is shown that every pair of densities can be joined by a constant speed geodesic. In Section 2.4 we present the definition of non-local Ricci curvature and give a characterisation in terms of the Hessian of the entropy. Section 2.5 contains a criterion that allows us to give lower bounds on the Ricci curvature in some basic examples, including the discrete circle and the discrete hypercube. A tensorisation result is contained in Section 2.6. Finally, we introduce new versions of well-known functional inequalities in Section 2.7 and prove implications between these and known inequalities.

2.2 The metric $\mathcal{W}$

In this section we shall study some basic properties of the metric $\mathcal{W}$. Throughout we shall work with an irreducible and reversible Markov kernel K on a finite set $\mathcal{X}$. The unique steady state will be denoted by π, and we shall write $P_t := e^{t(K-I)}$, $t \geq 0$, to denote the corresponding Markov semigroup.

We start by introducing some notation.

2.2.1 Notation

For $\varphi \in \mathbb{R}^{\mathcal{X}}$ we consider the *discrete gradient* $\nabla\varphi \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ defined by

$$\nabla\varphi(x,y) := \varphi(y) - \varphi(x) \ .$$

For $\Psi \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ we consider the *discrete divergence* $\nabla \cdot \Psi \in \mathbb{R}^{\mathcal{X}}$ defined by

$$(\nabla \cdot \Psi)(x) := \frac{1}{2}\sum_{y\in\mathcal{X}}(\Psi(x,y) - \Psi(y,x))K(x,y) \in \mathbb{R} \ .$$

With this notation we have

$$\Delta := \nabla \cdot \nabla = K - I \ ,$$

and the integration by parts formula

$$\langle \nabla\psi, \Psi\rangle_\pi = -\langle \psi, \nabla \cdot \Psi\rangle_\pi$$

holds. Here we write, for $\varphi, \psi \in \mathbb{R}^{\mathcal{X}}$ and $\Phi, \Psi \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$,

$$\langle \varphi, \psi \rangle_\pi = \sum_{x\in\mathcal{X}} \varphi(x)\psi(x)\pi(x) \ ,$$

$$\langle \Phi, \Psi \rangle_\pi = \frac{1}{2} \sum_{x,y\in\mathcal{X}} \Phi(x,y)\Psi(x,y)K(x,y)\pi(x) \ .$$

From now on we shall fix a function $\theta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ satisfying the following assumptions:

Assumption 2.2.1. *The function θ has the following properties:*

(A1) (Regularity): θ is continuous on $\mathbb{R}_+ \times \mathbb{R}_+$ and C^∞ on $(0,\infty)\times(0,\infty)$;

(A2) (Symmetry): $\theta(s,t) = \theta(t,s)$ for $s,t \geq 0$;

(A3) (Positivity, normalisation): $\theta(s,t) > 0$ for $s,t > 0$ and $\theta(1,1) = 1$;

(A4) (Zero at the boundary): $\theta(0,t) = 0$ for all $t \geq 0$;

(A5) (Monotonicity): $\theta(r,t) \leq \theta(s,t)$ for all $0 \leq r \leq s$ and $t \geq 0$;

(A6) (Positive homogeneity): $\theta(\lambda s, \lambda t) = \lambda\theta(s,t)$ for $\lambda > 0$ and $s,t \geq 0$;

(A7) (Concavity): the function $\theta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is concave.

It is easily checked that these assumptions imply that θ is bounded from above by the arithmetic mean:

$$\theta(s,t) \ \leq \ \frac{s+t}{2} \qquad \forall s,t \geq 0 \ . \tag{2.2.1}$$

In the next result we collect some properties of the function θ, which turn out be very useful in obtaining non-local Ricci curvature bounds.

Lemma 2.2.2. *For all $s,t,u,v > 0$ we have*

$$s\cdot\partial_1\theta(s,t) + t\cdot\partial_2\theta(s,t) \ = \ \theta(s,t) \ , \tag{2.2.2}$$

$$s\cdot\partial_1\theta(u,v) + t\cdot\partial_2\theta(u,v) - \theta(s,t) \ \geq \ 0 \ . \tag{2.2.3}$$

Proof. The equality (2.2.2) follows immediately from the homogeneity (A6) by noting that the left-hand side equals $\frac{\mathrm{d}}{\mathrm{d}r}\big|_{r=1}\theta(rs,rt)$. Let us prove (2.2.3). Note that by the concavity (A7) of θ the gradient $\nabla\theta$ is a monotone operator from $\mathbb{R}^2_+$ to $\mathbb{R}^2$. Hence, for all $s,t,x,y > 0$ we have

$$(s-x)\Big(\partial_1\theta(s,t) - \partial_1\theta(x,y)\Big) + (t-y)\Big(\partial_2\theta(s,t) - \partial_2\theta(x,y)\Big) \ \leq \ 0 \ .$$

By the homogeneity (A6) both $\partial_1\theta$ and $\partial_2\theta$ are 0-homogeneous. Taking now, in particular $x = \varepsilon u, y = \varepsilon v$ and letting $\varepsilon \to 0$ we obtain

$$s\Big(\partial_1\theta(s,t) - \partial_1\theta(u,v)\Big) + t\Big(\partial_2\theta(s,t) - \partial_2\theta(u,v)\Big) \ \leq \ 0 \ .$$

From this we deduce (2.2.3) by an application of (2.2.2). □

The most important example for our purposes is the logarithmic mean defined by

$$\theta(s,t) := \int_0^1 s^{1-p} t^p \mathrm{d}p = \frac{s-t}{\log s - \log t} \ ,$$

the latter expression being valid if $s, t > 0$ and $s \neq t$. For $\rho \in \mathscr{P}(\mathcal{X})$ and $x, y \in \mathcal{X}$ we define

$$\hat{\rho}(x,y) = \theta(\rho(x), \rho(y)) \ .$$

For a fixed $\rho \in \mathscr{P}(\mathcal{X})$ it will be useful to consider the Hilbert space $\mathcal{G}_\rho$ consisting of all (equivalence classes of) functions $\Psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, endowed with the inner product

$$\langle \Phi, \Psi \rangle_\rho := \frac{1}{2} \sum_{x,y \in \mathcal{X}} \Phi(x,y) \Psi(x,y) \hat{\rho}(x,y) K(x,y) \pi(x) \ . \tag{2.2.4}$$

Here we identify functions coinciding on the set $\{(x,y) \in \mathcal{X} \times \mathcal{X} : \ \hat{\rho}(x,y)K(x,y) > 0\}$. The operator ∇ can then be considered as a linear operator $\nabla : L^2(\mathcal{X}) \to \mathcal{G}_\rho$, whose negative adjoint is the *ρ-divergence operator* $(\nabla_\rho \cdot) : \mathcal{G}_\rho \to L^2(\mathcal{X})$ given by

$$(\nabla_\rho \cdot \Psi)(x) := \frac{1}{2} \sum_{y \in \mathcal{X}} (\Psi(x,y) - \Psi(y,x)) \hat{\rho}(x,y) K(x,y) \ .$$

2.2.2 Equivalent Definitions of the metric $\mathcal{W}$

We shall now state the definition of the metric $\mathcal{W}$ introduced in [Maa11]. Here and in the rest of this chapter we will use the shorthand notation

$$\mathcal{A}(\rho, \psi) := \|\nabla \psi\|_\rho^2 = \frac{1}{2} \sum_{x,y \in \mathcal{X}} (\psi(y) - \psi(x))^2 \hat{\rho}(x,y) K(x,y) \pi(x) \ ,$$

for $\rho \in \mathscr{P}(\mathcal{X})$ and $\psi \in \mathbb{R}^{\mathcal{X}}$.

Definition 2.2.3. *For $\bar{\rho}_0, \bar{\rho}_1 \in \mathscr{P}(\mathcal{X})$ we define*

$$\mathcal{W}(\bar{\rho}_0, \bar{\rho}_1)^2 := \inf \left\{ \int_0^1 \mathcal{A}(\rho_t, \psi_t) \mathrm{d}t \ : \ (\rho, \psi) \in \mathcal{CE}_1(\bar{\rho}_0, \bar{\rho}_1) \right\} ,$$

where for $T > 0$, $\mathcal{CE}_T(\bar{\rho}_0, \bar{\rho}_1)$ denotes the collection of pairs (ρ, ψ) satisfying the following conditions:

$$\begin{cases} (i) & \rho : [0,T] \to \mathbb{R}^{\mathcal{X}} \text{ is } C^\infty \ ; \\ (ii) & \rho_0 = \bar{\rho}_0 \ , \qquad \rho_T = \bar{\rho}_1 \ ; \\ (iii) & \rho_t \in \mathscr{P}(\mathcal{X}) \text{ for all } t \in [0,T] \ ; \\ (iv) & \psi : [0,T] \to \mathbb{R}^{\mathcal{X}} \text{ is measurable} \ ; \\ (v) & \text{For all } x \in \mathcal{X} \text{ and all } t \in (0,T) \text{ we have} \\ & \dot{\rho}_t(x) + \sum_{y \in \mathcal{X}} \big(\psi_t(y) - \psi_t(x)\big) \hat{\rho}_t(x,y) K(x,y) = 0 \ . \end{cases} \tag{2.2.5}$$

Using the notation introduced above, the continuity equation in (v) can be written as

$$\dot{\rho}_t + \nabla \cdot (\hat{\rho} \nabla \psi) = 0 \,. \tag{2.2.6}$$

Definition 2.2.3 is the same as the one in [Maa11], except that slightly different regularity conditions have been imposed on ρ. We shall shortly see that both definitions are equivalent.

The following results on the metric $\mathcal{W}$ have been proved in [Maa11].

Theorem 2.2.4. *The following assertions hold:*

(i) The space $(\mathscr{P}(\mathcal{X}), \mathcal{W})$ is a complete metric space, compatible with the Euclidean topology.

(ii) The restriction of $\mathcal{W}$ to $\mathscr{P}_*(\mathcal{X})$ is the Riemannian distance induced by the following Riemannian structure:

- *the tangent space of $\rho \in \mathscr{P}_*(\mathcal{X})$ can be identified with the set*

$$T_\rho := \{\nabla \psi \ : \ \psi \in \mathbb{R}^{\mathcal{X}}\}$$

by means of the following identification: given a smooth curve $(-\varepsilon, \varepsilon) \ni t \mapsto \rho_t \in \mathscr{P}_*(\mathcal{X})$ with $\rho_0 = \rho$, there exists a unique element $\nabla \psi_0 \in T_\rho$, such that the continuity equation (2.2.5)*(v) holds at $t = 0$.*

- *The Riemannian metric on T_ρ is given by the inner product*

$$\langle \nabla \varphi, \nabla \psi \rangle_\rho = \frac{1}{2} \sum_{x,y \in \mathcal{X}} (\varphi(x) - \varphi(y))(\psi(x) - \psi(y)) \hat{\rho}(x,y) K(x,y) \pi(x) \,.$$

(iii) If θ is the logarithmic mean, i.e., $\theta(s,t) = \int_0^1 s^{1-p} t^p \,\mathrm{d}p$, then the heat flow is the gradient flow of the entropy, in the sense that for any $\rho \in \mathscr{P}(\mathcal{X})$ and $t > 0$, we have $\rho_t := P_t \rho \in \mathscr{P}_*(\mathcal{X})$ and

$$D_t \rho_t = -\operatorname{grad} \mathcal{H}(\rho_t) \,. \tag{2.2.7}$$

Remark 2.2.5. If ρ belongs to $\mathscr{P}_*(\mathcal{X})$, then the gradient flow equation (2.2.7) also holds for $t = 0$.

Remark 2.2.6. The relevance of the logarithmic mean can be seen as follows. The heat equation $\dot{\rho}_t = \Delta \rho_t = \nabla \cdot (\nabla \rho_t)$ can be rewritten as a continuity equation (2.2.6) provided that

$$\nabla \psi = -\frac{\nabla \rho}{\hat{\rho}} \,.$$

On the other hand, an easy computation (see [Maa11, Proposition 4.2 and Corollary 4.3]) shows that under the identification above, the gradient of the entropy is given by

$$\operatorname{grad}_{\mathcal{W}} \mathcal{H}(\rho) = \nabla \log \rho \ .$$

Combining these observations, we infer that the heat flow is the gradient flow of the entropy with respect to $\mathcal{W}$, precisely when

$$\frac{\nabla \rho}{\hat{\rho}} = \nabla \log \rho \ ,$$

that is, when θ is the logarithmic mean.

This argument shows that the same heat flow can also be identified as the gradient flow of the functional $\mathcal{F}(\rho) = \sum_{x \in \mathcal{X}} f(\rho(x))\pi(x)$ for any smooth function $f : \mathbb{R} \to \mathbb{R}$ with $f'' > 0$, if one replaces the logarithmic mean by $\theta(r, s) = \frac{r-s}{f'(r)-f'(s)}$. We refer to [Maa11] for the details.

Our next aim is to provide an equivalent formulation of the definition of $\mathcal{W}$, which may seem less intuitive at first sight, but offers several technical advantages. First, the continuity equation becomes linear in V and ρ, which allows us to exploit the concavity of θ. Second, this formulation is more stable so that we can prove existence of minimisers in the class $\mathcal{CE}'_0(\bar{\rho}_0, \bar{\rho}_1)$. Similar ideas have already been developed in a continuous setting in [DNS09], where a general class of transportation metrics was constructed based on the usual continuity equation in $\mathbb{R}^n$.

An important role will be played by the function $\alpha : \mathbb{R} \times \mathbb{R}^2_+ \to \mathbb{R} \cup \{+\infty\}$ defined by

$$\alpha(x, s, t) = \begin{cases} 0 \ , & \theta(s,t) = 0 \text{ and } x = 0 \ , \\ \frac{x^2}{2\theta(s,t)} \ , & \theta(s,t) \neq 0 \ , \\ +\infty \ , & \theta(s,t) = 0 \text{ and } x \neq 0 \ . \end{cases}$$

The following observation will be useful.

Lemma 2.2.7. *The function α is lower semicontinuous and convex.*

Proof. This is easily checked using (A7) and the convexity of the function $(x, y) \mapsto \frac{x^2}{y}$ on $\mathbb{R} \times (0, \infty)$. □

Given $\rho \in \mathscr{P}(\mathcal{X})$ and $V \in \mathbb{R}^{\mathcal{X} \times \mathcal{X}}$ we define

$$\mathcal{A}'(\rho, V) := \sum_{x,y \in \mathcal{X}} \alpha(V(x,y), \rho(x), \rho(y)) K(x,y)\pi(x) \ ,$$

and we set

$$\mathcal{CE}'_T(\bar{\rho}_0, \bar{\rho}_1) := \{(\rho, \psi) \ : \ (i'), (ii), (iii), (iv'), (v') \text{ hold} \} \ ,$$

where

$$\begin{cases} (i') & \rho : [0,T] \to \mathbb{R}^{\mathcal{X}} \text{ is continuous ;} \\ (iv') & V : [0,T] \to \mathbb{R}^{\mathcal{X}\times\mathcal{X}} \text{ is locally integrable ;} \\ (v') & \text{For all } x \in \mathcal{X} \text{ we have in the sense of distributions} \\ & \dot{\rho}_t(x) + \frac{1}{2}\sum_{y\in\mathcal{X}} \big(V_t(x,y) - V_t(y,x)\big)K(x,y) = 0 \ . \end{cases} \tag{2.2.8}$$

The continuity equation in (v') can equivalently be written as

$$\dot{\rho}_t + \nabla \cdot V = 0 \ .$$

As an immediate consequence of Lemma 2.2.7 we obtain the following convexity of $\mathcal{A}'$.

Corollary 2.2.8. *Let $\rho^i \in \mathscr{P}(\mathcal{X})$ and $V^i \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ for $i = 0,1$. For $\tau \in [0,1]$ set $\rho^\tau := (1-\tau)\rho^0 + \tau\rho^1$ and $V^\tau := (1-\tau)V^0 + \tau V^1$. Then we have*

$$\mathcal{A}'(\rho^\tau, V^\tau) \ \leq \ (1-\tau)\mathcal{A}'(\rho^0, V^0) + \tau\mathcal{A}'(\rho^1, V^1) \ .$$

Now we have the following reformulation of Definition 2.2.3.

Lemma 2.2.9. *For $\bar{\rho}_0, \bar{\rho}_1 \in \mathscr{P}(\mathcal{X})$ we have*

$$\mathcal{W}(\bar{\rho}_0,\bar{\rho}_1)^2 = \inf\left\{ \int_0^1 \mathcal{A}'(\rho_t, V_t)\mathrm{d}t \ : \ (\rho, V) \in \mathcal{CE}_1'(\bar{\rho}_0,\bar{\rho}_1) \right\} \ .$$

Furthermore, if $\bar{\rho}_0, \bar{\rho}_1 \in \mathscr{P}_*(\mathcal{X})$, condition (iv) in (2.2.5) can be reinforced into: "$\psi : [0,T] \to \mathbb{R}^{\mathcal{X}}$ is C^∞".

Proof. The inequality "$\geq$" follows easily by noting that the infimum is taken over a larger set. Indeed, given a pair $(\rho,\psi) \in \mathcal{CE}_1(\bar{\rho}_0,\bar{\rho}_1)$ we obtain a pair $(\rho, V) \in \mathcal{CE}_1'(\bar{\rho}_0,\bar{\rho}_1)$ by setting $V_t(x,y) = \nabla\psi_t(x,y)\hat{\rho}_t(x,y)$ and we have $\mathcal{A}'(\rho_t, V_t) = \mathcal{A}(\rho_t,\psi_t)$.

To show the opposite inequality "$\leq$", we fix an arbitrary pair $(\rho, V) \in \mathcal{CE}_1'(\bar{\rho}_0,\bar{\rho}_1)$. It is sufficient to show that for every $\varepsilon > 0$ there exists a pair $(\rho^\varepsilon, \psi^\varepsilon) \in \mathcal{CE}_1(\bar{\rho}_0,\bar{\rho}_1)$ such that

$$\int_0^1 \mathcal{A}(\rho_t^\varepsilon, \psi_t^\varepsilon)\mathrm{d}t \ \leq \ \int_0^1 \mathcal{A}'(\rho_t, V_t)\mathrm{d}t + \varepsilon \ .$$

For this purpose we first regularise (ρ, V) by a mollification argument. We thus define $(\widetilde{\rho}, \widetilde{V}) : [-\varepsilon, 1+\varepsilon] \to \mathscr{P}(\mathcal{X}) \times \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ by

$$(\widetilde{\rho}_t, \widetilde{V}_t) \ = \ \begin{cases} (\rho(0), 0) \ , & t \in [-\varepsilon, \varepsilon) \ , \\ (\rho(\frac{t-\varepsilon}{1-2\varepsilon}), \frac{1}{1-2\varepsilon}V(\frac{t-\varepsilon}{1-2\varepsilon})) \ , & t \in [\varepsilon, 1-\varepsilon) \ , \\ (\rho(1), 0) \ , & t \in [1-\varepsilon, 1+\varepsilon] \ , \end{cases}$$

and take a non-negative smooth function $\eta : \mathbb{R} \to \mathbb{R}_+$ which vanishes outside of $[-\varepsilon, \varepsilon]$, is strictly positive on $(-\varepsilon, \varepsilon)$ and satisfies $\int \eta(s)\mathrm{d}s = 1$. For $t \in [0,1]$ we define

$$\rho_t^\varepsilon \;=\; \int \eta(s)\widetilde{\rho}_{t+s}\mathrm{d}s \;, \qquad V_t^\varepsilon \;=\; \int \eta(s)\widetilde{V}_{t+s}\mathrm{d}s \;.$$

Now $t \mapsto \rho_t^\varepsilon$ is C^∞ and using the continuity of ρ it is easy to check that $(\rho^\varepsilon, V^\varepsilon) \in \mathcal{CE}_1'(\bar{\rho}_0, \bar{\rho}_1)$. Moreover, using the convexity from Corollary 2.2.8 we can estimate

$$\begin{aligned}\int_0^1 \mathcal{A}'(\rho_t^\varepsilon, V_t^\varepsilon)\mathrm{d}t \;&\le\; \int_0^1 \int \eta(s)\mathcal{A}'(\widetilde{\rho}_{t+s}, \widetilde{V}_{t+s})\mathrm{d}s\mathrm{d}t\\ &\le\; \int_{-\varepsilon}^{1+\varepsilon} \mathcal{A}'(\widetilde{\rho}_t, \widetilde{V}_t)\mathrm{d}t \;=\; \frac{1}{1-2\varepsilon}\int_0^1 \mathcal{A}'(\rho_t, V_t)\mathrm{d}t \;.\end{aligned}$$

To proceed further, we may assume without loss of generality that $V(x,y) = 0$ whenever $K(x,y) = 0$. The fact that $\int_0^1 \mathcal{A}'(\rho_t, V_t)\mathrm{d}t$ is finite implies that the set $\{t : \hat{\rho}_t(x,y) = 0 \text{ and } V_t(x,y) \neq 0\}$ is negligible for all $x, y \in \mathcal{X}$. Taking properties (A3) and (A4) of the function θ into account, this implies that for the convolved quantities the corresponding set $\{t : \hat{\rho}_t^\varepsilon(x,y) = 0 \text{ and } V_t^\varepsilon(x,y) \neq 0\}$ is empty for all $x, y \in \mathcal{X}$. Hence there exists a measurable function $\Psi^\varepsilon : [0,1] \to \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ satisfying

$$V_t^\varepsilon(x,y) \;=\; \Psi_t^\varepsilon(x,y)\hat{\rho}_t^\varepsilon(x,y) \qquad \text{for all } x,y \in \mathcal{X} \text{ and all } t \in [0,1] \;. \tag{2.2.9}$$

It remains to find a function $\psi^\varepsilon : [0,1] \to \mathbb{R}^{\mathcal{X}}$ such that $\nabla_{\rho_t^\varepsilon} \cdot \Psi_t^\varepsilon = \nabla_{\rho_t^\varepsilon} \cdot \nabla\psi_t^\varepsilon$. Let $\mathcal{P}_\rho$ denote the orthogonal projection in $\mathcal{G}_\rho$ onto the range of ∇. Then there exists a measurable function $\psi^\varepsilon : [0,1] \to \mathbb{R}^{\mathcal{X}}$ such that $\mathcal{P}_{\rho_t^\varepsilon}\Psi_t^\varepsilon = \nabla\psi_t^\varepsilon$. The orthogonal decomposition

$$\mathcal{G}_{\rho_t^\varepsilon} \;=\; \mathrm{Ran}(\nabla) \oplus^\perp \mathrm{Ker}(\nabla_{\rho_t^\varepsilon}^*) \tag{2.2.10}$$

implies that $\nabla_{\rho_t^\varepsilon} \cdot \Psi_t^\varepsilon = \nabla_{\rho_t^\varepsilon} \cdot \nabla\psi_t^\varepsilon$, hence $(\rho^\varepsilon, \psi^\varepsilon) \in \mathcal{CE}_1(\bar{\rho}_o, \bar{\rho}_1)$. Using the decomposition (2.2.10) once more, we infer that $\langle \nabla\psi_t^\varepsilon, \nabla\psi_t^\varepsilon\rangle_{\rho_t^\varepsilon} \le \langle \Psi_t^\varepsilon, \Psi_t^\varepsilon\rangle_{\rho_t^\varepsilon}$. This implies $\mathcal{A}(\rho_t^\varepsilon, \psi_t^\varepsilon) \le \mathcal{A}'(\rho_t^\varepsilon, V_t^\varepsilon)$ and finishes the proof of the first assertion.

If $\bar{\rho}_0$ and $\bar{\rho}_1$ belong to $\mathscr{P}_*(\mathcal{X})$, one can follow the argument in [Maa11, Lemma 3.30] and construct a curve $(\breve{\rho}, \breve{V}) \in \mathcal{CE}_1'(\bar{\rho}_0, \bar{\rho}_1)$ such that $\breve{\rho}_t \in \mathscr{P}_*(\mathcal{X})$ for $t \in [0,1]$ and

$$\int_0^1 \mathcal{A}'(\breve{\rho}_t, \breve{V}_t)\mathrm{d}t \;\le\; \int_0^1 \mathcal{A}'(\rho_t, V_t)\mathrm{d}t + \varepsilon \;.$$

Then one can apply the argument above. In this case, $\rho_t^\varepsilon(x) > 0$ for all $x \in \mathcal{X}$ and $t \in [0,1]$, and therefore the function $\Psi^\varepsilon : [0,1] \to \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ is C^∞. Furthermore, since the orthogonal projection P_ρ depends smoothly on $\rho \in \mathscr{P}_*(\mathcal{X})$, the function $\psi^\varepsilon : [0,1] \to \mathbb{R}^{\mathcal{X}}$ is smooth as well. □

Remark 2.2.10. In [Maa11] the metric $\mathcal{W}$ has been defined as in Definition 2.2.3, with the difference that (i) in (2.2.8) was replaced by "$\rho : [0,T] \to \mathscr{P}(\mathcal{X})$ is piecewise C^1". Therefore Lemma 2.2.9 shows, in particular, that Definition 2.2.3 coincides with the original definition of $\mathcal{W}$ from [Maa11].

2.2.3 Basic Properties of $\mathcal{W}$

As an application of Lemma 2.2.9 we shall prove the following convexity result, which is a discrete counterpart of the well-known fact that the squared L^2-Wasserstein distance over Euclidean space is convex with respect to linear interpolation (see, for example, [DNS09, Theorem 5.11]).

Proposition 2.2.11 (Convexity of the squared distance)**.** *For $i, j = 0, 1$, let $\rho_i^j \in \mathscr{P}(\mathcal{X})$, and for $\tau \in [0,1]$ set $\rho_i^\tau := (1-\tau)\rho_i^0 + \tau\rho_i^1$. Then*

$$\mathcal{W}(\rho_0^\tau, \rho_1^\tau)^2 \le (1-\tau)\mathcal{W}(\rho_0^0, \rho_1^0)^2 + \tau\mathcal{W}(\rho_0^1, \rho_1^1)^2 \ .$$

Proof. Let $\varepsilon > 0$. For $j = 0, 1$ we may take a pair $(\rho^j, V^j) \in \mathcal{CE}'(\rho_0^j, \rho_1^j)$ with

$$\int_0^1 \mathcal{A}'(\rho_t^j, V_t^j)\mathrm{d}t \le \mathcal{W}^2(\rho_0^j, \rho_1^j) + \varepsilon$$

in view of Lemma 2.2.9. For $\tau \in [0,1]$ we set

$$\rho_t^\tau := (1-\tau)\rho_t^0 + \tau\rho_t^1 \ , \qquad V_t^\tau := (1-\tau)V_t^0 + \tau V_t^1 \ .$$

It then follows that $(\rho^\tau, V^\tau) \in \mathcal{CE}'_1(\rho_0^\tau, \rho_1^\tau)$, hence by Corollary 2.2.8,

$$\begin{aligned}\mathcal{W}(\rho_0^\tau, \rho_1^\tau)^2 &\le \int_0^1 \mathcal{A}'(\rho_t^\tau, V_t^\tau)\mathrm{d}t \\ &\le (1-\tau)\int_0^1 \mathcal{A}'(\rho_t^0, V_t^0)\mathrm{d}t + \tau\int_0^1 \mathcal{A}'(\rho_t^1, V_t^1)\mathrm{d}t \\ &= (1-\tau)\mathcal{W}(\rho_0^0, \rho_1^0)^2 + \tau\mathcal{W}(\rho_0^1, \rho_1^1)^2 + \varepsilon \ .\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, this completes the proof. □

In this section we compare $\mathcal{W}$ to some commonly used metrics. A first result of this type (see [Maa11, Lemma 3.10]) gives a lower bound on $\mathcal{W}$ in terms of the total variation metric

$$d_{TV}(\rho_0, \rho_1) = \sum_{x\in\mathcal{X}} \pi(x)|\rho_0(x) - \rho_1(x)| \ .$$

Here, more generally, we shall compare $\mathcal{W}$ to various Wasserstein distances. Given a metric d on $\mathcal{X}$ and $1 \le p < \infty$, recall that the L^p-Wasserstein metric $W_{p,d}$ on $\mathscr{P}(\mathcal{X})$ is defined by

$$W_{p,d}(\rho_0, \rho_1) := \inf\left\{ \left(\sum_{x,y\in\mathcal{X}} d(x,y)^p q(x,y) \right)^{\frac{1}{p}} \,\middle|\, q \in \Gamma(\rho_0, \rho_1) \right\} , \tag{2.2.11}$$

where $\Gamma(\rho_0, \rho_1)$ denotes the set of all couplings between ρ_0 and ρ_1, i.e.,

$$\Gamma(\rho_0, \rho_1) := \Big\{ q : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+ \ \Big| \ \sum_{y \in \mathcal{X}} q(x, y) = \rho_0(x)\pi(x) \ , \\ \sum_{x \in \mathcal{X}} q(x, y) = \rho_1(y)\pi(y) \Big\} \ .$$

It is well known (see, for example, [Vil09, Theorem 4.1]) that the infimum in (2.2.11) is attained; as usual we shall denote the collection of minimisers by $\Gamma_o(\rho_0, \rho_1)$.

In our setting there are various metrics on $\mathcal{X}$ that are natural to consider. In particular,

- the *graph distance* d_g with respect to the graph structure on $\mathcal{X}$ induced by K (i.e., $\{x, y\}$ is an edge iff $K(x, y) > 0$).

- the metric $d_{\mathcal{W}}$, that is, the restriction of $\mathcal{W}$ from $\mathscr{P}(\mathcal{X})$ to $\mathcal{X}$ under the identification of points in $\mathcal{X}$ with the corresponding Dirac masses:

$$d_{\mathcal{W}}(x, y) := \mathcal{W}\bigg(\frac{\mathbf{1}_{\{x\}}}{\pi(x)}, \frac{\mathbf{1}_{\{y\}}}{\pi(y)} \bigg) \ .$$

The induced L^p-Wasserstein distances will be denoted by $W_{p,g}$ and $W_{p,\mathcal{W}}$ respectively.

We shall now prove lower and upper bounds for the metric $\mathcal{W}$ in terms of suitable Wasserstein metrics. We start with the lower bounds. Let us remark that, unlike most other results in this chapter, the second inequality in the following result relies on the normalisation $\sum_{y \in \mathcal{X}} K(x, y) = 1$.

Proposition 2.2.12 (Lower bounds for $\mathcal{W}$)**.** *For all probability densities $\rho_0, \rho_1 \in \mathscr{P}(\mathcal{X})$ we have*

$$\frac{1}{\sqrt{2}} d_{TV}(\rho_0, \rho_1) \le \sqrt{2} W_{1,g}(\rho_0, \rho_1) \le \mathcal{W}(\rho_0, \rho_1) \ . \tag{2.2.12}$$

Proof. Note that $d_{tr} \le d_g$, where $d_{tr}(x, y) = \mathbf{1}_{x \ne y}$ denotes the trivial distance. Therefore, the first bound follows from the fact that d_{TV} is the L^1-Wasserstein distance induced by d_{tr} (see [Vil03, Theorem 1.14]).

In order to prove the second bound, we fix $\varepsilon > 0$, take $\bar{\rho}_0, \bar{\rho}_1 \in \mathscr{P}(\mathcal{X})$ and $(\rho, \psi) \in \mathcal{CE}_1(\bar{\rho}_0, \bar{\rho}_1)$ with

$$\bigg(\int_0^1 \mathcal{A}(\rho_t, \psi_t) \mathrm{d}t \bigg)^{\frac{1}{2}} \le \mathcal{W}(\bar{\rho}_0, \bar{\rho}_1) + \varepsilon \ .$$

Using the continuity equation from (2.2.5) we obtain for any $\varphi : \mathcal{X} \to \mathbb{R}$,

$$\begin{aligned}
&\Big| \sum_{x\in\mathcal{X}} \varphi(x)(\rho_0(x) - \rho_1(x))\pi(x) \Big| \\
&= \Big| \int_0^1 \sum_{x\in\mathcal{X}} \varphi(x)\dot\rho_t(x)\pi(x)\mathrm{d}t \Big| \\
&= \Big| \int_0^1 \sum_{x,y\in\mathcal{X}} \varphi(x)\big(\psi_t(x) - \psi_t(y)\big)\hat\rho_t(x,y)K(x,y)\pi(x)\mathrm{d}t \Big| \\
&= \Big| \int_0^1 \langle \nabla\varphi, \nabla\psi_t\rangle_{\rho_t}\mathrm{d}t \Big| \\
&\le \Big(\int_0^1 \|\nabla\varphi\|_{\rho_t}^2 \mathrm{d}t \Big)^{1/2} \Big(\int_0^1 \|\nabla\psi_t\|_{\rho_t}^2 \mathrm{d}t \Big)^{1/2} \\
&= \Big(\int_0^1 \|\nabla\varphi\|_{\rho_t}^2 \mathrm{d}t \Big)^{1/2} (\mathcal{W}(\bar\rho_0, \bar\rho_1) + \varepsilon)\ .
\end{aligned}$$

Let $[\varphi]_{\mathrm{Lip}}$ denote the Lipschitz constant of φ with respect to the graph distance d_g, i.e.,

$$[\varphi]_{\mathrm{Lip}} := \sup_{x\neq y} \frac{|\varphi(x) - \varphi(y)|}{d_g(x,y)}\ .$$

Applying the inequality (2.2.1) and using the fact that $d_g(x,y) = 1$ if $x \neq y$ and $K(x,y) > 0$, we infer that

$$\begin{aligned}
\|\nabla\varphi\|_{\rho_t}^2 &= \frac{1}{2} \sum_{x,y\in\mathcal{X}} \big(\varphi(x) - \varphi(y)\big)^2 K(x,y)\hat\rho_t(x,y)\pi(x) \\
&\le \frac{1}{4}[\varphi]_{\mathrm{Lip}}^2 \sum_{x,y\in\mathcal{X}} K(x,y)\big(\rho_t(x) + \rho_t(y)\big)\pi(x) \\
&= \frac{1}{2}[\varphi]_{\mathrm{Lip}}^2 \sum_{x\in\mathcal{X}} \rho_t(x)\pi(x) \sum_{y\in\mathcal{X}} K(x,y) \\
&= \frac{1}{2}[\varphi]_{\mathrm{Lip}}^2\ .
\end{aligned}$$

The Kantorovich–Rubinstein Theorem (see, for example, [Vil03, Theorem 1.14]) yields

$$W_{1,g}(\bar\rho_0, \bar\rho_1) = \sup_{\varphi:[\varphi]_{\mathrm{Lip}}\le 1} \Big| \sum_{x\in\mathcal{X}} \varphi(x)(\bar\rho_0(x) - \bar\rho_1(x))\pi(x) \Big| \le \frac{\mathcal{W}(\bar\rho_0, \bar\rho_1) + \varepsilon}{\sqrt{2}}\ ,$$

which completes the proof, since $\varepsilon > 0$ is arbitrary. □

Before stating the upper bounds, we provide a simple relation between d_g and $d_{\mathcal{W}}$.

Lemma 2.2.13. *For $x, y \in \mathcal{X}$ we have*

$$d_{\mathcal{W}}(x,y) \leq \frac{c}{\sqrt{k}} d_g(x,y) \ ,$$

where

$$c = \int_{-1}^{1} \frac{\mathrm{d}r}{\sqrt{2\theta(1-r,1+r)}} < \infty \qquad \text{and} \qquad k = \min_{(x,y)\ :\ K(x,y)>0} K(x,y) \ .$$

If θ is the logarithmic mean, then $c \approx 1.56$.

Proof. Let $\{x_i\}_{i=0}^n$ be a sequence in $\mathcal{X}$ with $x_0 = x$, $x_n = y$ and $K(x_i, x_{i+1}) > 0$ for all i. We shall use the fact, proved in [Maa11, Theorem 2.4], that the $\mathcal{W}$-distance between two Dirac measures on a two-point space $\{a, b\}$ with transition probabilities $K(a,b) = K(b,a) = p$ is equal to $\frac{c}{\sqrt{p}}$. The concavity of θ readily implies that c is finite. Furthermore, it follows from [Maa11, Lemma 3.14] and its proof, that for any pair $x, y \in \mathcal{X}$ with $K(x,y) > 0$, one has

$$\mathcal{W}\left(\frac{\mathbf{1}_{\{x\}}}{\pi(x)}, \frac{\mathbf{1}_{\{y\}}}{\pi(y)}\right) \leq c\sqrt{\frac{\max\{\pi(x), \pi(y)\}}{K(x,y)\pi(x)}} \leq \frac{c}{\sqrt{k}} \ .$$

Using the triangle inequality for $\mathcal{W}$ we obtain

$$d_{\mathcal{W}}(x,y) = \mathcal{W}\left(\frac{\mathbf{1}_{\{x\}}}{\pi(x)}, \frac{\mathbf{1}_{\{y\}}}{\pi(y)}\right) \leq \sum_{i=0}^{n-1} \mathcal{W}\left(\frac{\mathbf{1}_{\{x_i\}}}{\pi(x_i)}, \frac{\mathbf{1}_{\{x_{i+1}\}}}{\pi(x_{i+1})}\right) \leq \frac{nc}{\sqrt{k}} \ ,$$

hence the result follows by taking the infimum over all such sequences $\{x_i\}_{i=0}^n$. □

Now we turn to upper bounds for $\mathcal{W}$ in terms of L^2-Wasserstein distances.

Proposition 2.2.14 (Upper bounds for $\mathcal{W}$)**.** *For all probability densities $\rho_0, \rho_1 \in \mathscr{P}(\mathcal{X})$ we have*

$$\mathcal{W}(\rho_0, \rho_1) \leq W_{2,\mathcal{W}}(\rho_0, \rho_1) \leq \frac{c}{\sqrt{k}} W_{2,g}(\rho_0, \rho_1) \ , \tag{2.2.13}$$

where c and k are as in Lemma 2.2.13.

Proof. We shall prove the first bound, the second one being an immediate consequence of Lemma 2.2.13. For this purpose, we fix $\bar{\rho}_0, \bar{\rho}_1 \in \mathscr{P}(\mathcal{X})$ and take $q \in \Gamma_o(\bar{\rho}_0, \bar{\rho}_1)$. For all $u, v \in \mathcal{X}$, take a curve $(\rho^{u,v}, V^{u,v}) \in \mathcal{CE}'\big(\frac{\mathbf{1}_{\{u\}}}{\pi(u)}, \frac{\mathbf{1}_{\{v\}}}{\pi(v)}\big)$ with

$$\int_0^1 \mathcal{A}'(\rho_t^{u,v}, V_t^{u,v}) \mathrm{d}t \leq d_{\mathcal{W}}(u,v)^2 + \varepsilon \ ,$$

and consider the convex combination of these curves, weighted according to the optimal plan q, i.e.,

$$\rho_t := \sum_{u,v\in\mathcal{X}} q(u,v)\rho_t^{u,v} , \qquad V_t := \sum_{u,v\in\mathcal{X}} q(u,v)V_t^{u,v} .$$

It then follows that the resulting curve (ρ, V) belongs to $\mathcal{CE}'_1(\bar\rho_0, \bar\rho_1)$. Using the convexity result from Lemma 2.2.7 we infer that

$$\begin{aligned}\mathcal{W}(\bar\rho_0,\bar\rho_1)^2 \le \int_0^1 \mathcal{A}'(\rho_t,V_t)\mathrm{d}t &\le \sum_{u,v\in\mathcal{X}} q(u,v)\int_0^1 \mathcal{A}'(\rho_t^{u,v},V_t^{u,v})\mathrm{d}t\\ &\le \sum_{u,v\in\mathcal{X}} q(u,v)(d_{\mathcal{W}}(u,v)^2+\varepsilon)\\ &= W_{2,\mathcal{W}}(\bar\rho_0,\bar\rho_1)^2+\varepsilon .\end{aligned}$$

which implies the result. □

2.3 Geodesics

In this section we show that the metric space $(\mathscr{P}(\mathcal{X}), \mathcal{W})$ is a geodesic space, in the sense that any two densities $\rho_0, \rho_1 \in \mathscr{P}(\mathcal{X})$ can be connected by a *(constant speed) geodesic*, that is, a curve $\gamma : [0,1] \to \mathscr{P}(\mathcal{X})$ satisfying

$$\mathcal{W}(\gamma_s,\gamma_t) = |s-t|\mathcal{W}(\gamma_0,\gamma_1)$$

for all $0 \le s,t \le 1$.

Let us first give an equivalent characterisation of the infimum in Lemma 2.2.9, which is invariant under reparametrisation.

Lemma 2.3.1. *For any $T > 0$ and $\bar\rho_0, \bar\rho_1 \in \mathscr{P}(\mathcal{X})$ we have*

$$\mathcal{W}(\bar\rho_0,\bar\rho_1) \;=\; \inf\left\{\int_0^T \sqrt{\mathcal{A}'(\rho_t,V_t)}dt \;:\; (\rho,V)\in\mathcal{CE}'_T(\bar\rho_0,\bar\rho_1)\right\} . \qquad (2.3.1)$$

Proof. Taking Lemma 2.2.9 into account, this follows from a standard reparametrisation argument. See [AGS08, Lemma 1.1.4] or [DNS09, Theorem 5.4] for details in similar situations. □

Theorem 2.3.2. *For all $\bar\rho_0, \bar\rho_1 \in \mathscr{P}(\mathcal{X})$ the infimum in Lemma 2.2.9 is attained by a pair $(\rho, V) \in \mathcal{CE}'_1(\bar\rho_0, \bar\rho_1)$ satisfying $\mathcal{A}'(\rho_t, V_t) = \mathcal{W}(\bar\rho_0, \bar\rho_1)^2$ for a.e. $t \in [0, 1]$. In particular, the curve $(\rho_t)_{t\in[0,1]}$ is a constant speed geodesic.*

Proof. We will show existence of a minimising curve by a direct argument. Let $(\rho^n, V^n) \in \mathcal{CE}'_1(\bar\rho_0, \bar\rho_1)$ be a minimising sequence. Thus we can assume that

$$\sup_n \int_0^1 \mathcal{A}'(\rho_t^n, V_t^n)\mathrm{d}t \ < \ C$$

for some finite constant C. Without loss of generality we assume that $V_t^n(x,y) = 0$ when $K(x,y) = 0$. For $x, y \in \mathcal{X}$, define the sequence of signed Borel measures $\nu_{x,y}^n$ on $[0,1]$ by $\nu_{x,y}^n(\mathrm{d}t) := V_t^n(x,y)\mathrm{d}t$. For every Borel set $B \subset [0,1]$ we can give the following bound on the total variation of these measures:

$$\|\nu_{x,y}^n\|(B) \ \le \ \int_B |V_t^n(x,y)| \,\mathrm{d}t \le \ \sqrt{2C'} \int_B \sqrt{\alpha(V_t^n(x,y), \rho_t^n(x), \rho_t^n(y))}\mathrm{d}t \ ,$$

where we used the fact that $\rho(x) \le \max\{\pi(z)^{-1} : z \in \mathcal{X}\} =: C' < \infty$ for $\rho \in \mathscr{P}(\mathcal{X})$. Using Hölder's inequality we obtain

$$\begin{aligned}\sum_{x,y\in\mathcal{X}} \|\nu_{x,y}^n\|(B) K(x,y)\pi(x) \ &\le \ \sqrt{2C' \,\mathrm{Leb}(B)} \left(\int_0^1 \mathcal{A}'(\rho_t^n, V_t^n)\mathrm{d}t \right)^{\frac{1}{2}} \\ &\le \ \sqrt{2CC' \,\mathrm{Leb}(B)}\ . \end{aligned} \tag{2.3.2}$$

In particular, the total variation of the measures $\nu_{x,y}^n$ is bounded uniformly in n. Hence we can extract a subsequence (still indexed by n) such that for all $x, y \in \mathcal{X}$ the measures $\nu_{x,y}^n$ converge weakly* to some finite signed Borel measure $\nu_{x,y}$. The estimate (2.3.2) also shows that $\nu_{x,y}$ is absolutely continuous with respect to the Lebesgue measure. Thus there exists $V : [0,1] \to \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ such that $\nu_{x,y}(\mathrm{d}t) := V_t(x,y)\mathrm{d}t$. We claim that, along the same subsequence, ρ^n converges pointwise to a function $\rho : [0,1] \to \mathscr{P}(\mathcal{X})$. Indeed, using the continuity of $t \mapsto \rho_t^n$ one derives from the continuity equation (v') in (2.2.8) that for $s \in [0,1]$ and every $x \in \mathcal{X}$,

$$\rho_s^n - \rho_0^n \ = \ \frac{1}{2} \int_0^s \sum_{y\in\mathcal{X}} (V_t^n(y,x) - V_t^n(x,y))K(x,y)\mathrm{d}t\ . \tag{2.3.3}$$

The weak* convergence of $\nu_{x,y}^n$ implies (see [AGS08, Prop. 5.1.10]) the convergence of the right-hand side of (2.3.3). Since $\rho_0^n = \bar\rho_0$ for all n, this yields the desired convergence of ρ_s^n for all s, and one easily checks that $(\rho, V) \in \mathcal{CE}'_1(\rho_0, \rho_1)$. The weak* convergence of $\nu_{x,y}^n$ further implies that the measures $\rho_t^n(x)\mathrm{d}t$ converge weakly* to $\rho_t(x)\mathrm{d}t$. Applying a general result on the lower-semicontinuity of integral functionals (see [But89, Thm. 3.4.3]) and taking into account Lemma 2.2.7, we obtain

$$\int_0^1 \mathcal{A}'(\rho_t, V_t)\mathrm{d}t \ \le \ \liminf_n \int_0^1 \mathcal{A}'(\rho_t^n, V_t^n)\mathrm{d}t \ = \ \mathcal{W}(\bar\rho_0, \bar\rho_1)^2\ .$$

Hence the pair (ρ, V) is a minimiser of the variational problem in the definition of $\mathcal{W}$. Finally, Lemma 2.3.1 yields

$$\int_0^1 \sqrt{\mathcal{A}'(\rho_t, V_t)}\mathrm{d}t \;\geq\; \mathcal{W}(\bar\rho_0, \bar\rho_1) \;=\; \left(\int_0^1 \mathcal{A}'(\rho_t, V_t)\mathrm{d}t\right)^{\frac{1}{2}} \;,$$

which implies that $\mathcal{A}'(\rho_t, V_t) = W(\bar\rho_0, \bar\rho_1)^2$ for a.e. $t \in [0,1]$.

The fact that $(\rho_t)_t$ is a constant speed geodesic follows now by another application of Lemma 2.3.1. □

We shall now give a characterisation of absolutely continuous curves in the metric space $(\mathscr{P}(\mathcal{X}), \mathcal{W})$ and relate their length to their minimal action. First we recall some notions from the theory of analysis in metric spaces. A curve $(\rho_t)_{t\in[0,T]}$ in $\mathscr{P}(\mathcal{X})$ is called *absolutely continuous w.r.t.* $\mathcal{W}$ if there exists $m \in L^1(0,T)$ such that

$$\mathcal{W}(\rho_s, \rho_t) \;\leq\; \int_s^t m(r)\mathrm{d}r \quad \text{for all } 0 \leq s \leq t \leq T\;.$$

If (ρ_t) is absolutely continuous, then its *metric derivative*

$$|\rho_t'| \;:=\; \lim_{h\to 0} \frac{\mathcal{W}(\rho_{t+h}, \rho_t)}{|h|}$$

exists for a.e. $t \in [0,T]$ and satisfies $|\rho_t'| \leq m(t)$ a.e. (see [AGS08, Theorem 1.1.2]).

Proposition 2.3.3 (Metric velocity)**.** *A curve $(\rho_t)_{t\in[0,T]}$ is absolutely continuous with respect to $\mathcal{W}$ if and only if there exists a measurable function $V : [0,T] \to \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ such that $(\rho, V) \in \mathcal{CE}'_T(\rho_0, \rho_T)$ and*

$$\int_0^T \sqrt{\mathcal{A}'(\rho_t, V_t)}dt \;<\; \infty\;.$$

In this case we have $|\rho_t'|^2 \leq \mathcal{A}'(\rho_t, V_t)$ for a.e. $t \in [0,T]$ and there exists an almost everywhere uniquely defined function $\widetilde V : [0,1] \to \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ such that $(\rho, \widetilde V) \in \mathcal{CE}'_T(\rho_0, \rho_T)$ and $|\rho_t'|^2 = \mathcal{A}'(\rho_t, \widetilde V_t)$ for a.e. $t \in [0,T]$.

Proof. The proof follows from the very same arguments as in [DNS09, Thm. 5.17]. To construct the velocity field $\widetilde V$, the curve ρ is approximated by curves (ρ^n, V^n) which are piecewise minimising. The velocity field $\widetilde V$ is then defined as a subsequential limit of the velocity fields V^n. In our case, existence of this limit is guaranteed by a compactness argument similar to the one in the proof of Theorem 2.3.2. □

For later use we state an explicit formula for the geodesic equations in $\mathscr{P}_*(\mathcal{X})$ from [Maa11, Proposition 3.4]. Since the interior $\mathscr{P}_*(\mathcal{X})$ of $\mathscr{P}(\mathcal{X})$ is Riemannian by Theorem 2.2.4, local existence and uniqueness of geodesics is guaranteed by standard Riemannian geometry.

Proposition 2.3.4. *Let $\bar\rho \in \mathscr{P}_*(\mathcal{X})$ and $\bar\psi \in \mathbb{R}^{\mathcal{X}}$. On a sufficiently small time interval around 0, the unique constant speed geodesic with $\rho_0 = \bar\rho$ and initial tangent vector $\nabla\psi_0 = \nabla\bar\psi$ satisfies the following equations:*

$$\begin{cases} \partial_t \rho_t(x) + \sum_{y\in\mathcal{X}} (\psi_t(y) - \psi_t(x))\hat\rho_t(x,y) K(x,y) = 0 \ , \\ \partial_t \psi_t(x) + \frac{1}{2} \sum_{y\in\mathcal{X}} \big(\psi_t(x) - \psi_t(y)\big)^2 \partial_1\theta(\rho_t(x), \rho_t(y)) K(x,y) = 0 \ . \end{cases} \tag{2.3.4}$$

2.4 Ricci curvature

In this section we initiate the study of a notion of Ricci curvature lower boundedness in the spirit of Lott, Sturm, and Villani [LV09, Stu06]. Furthermore, we present a characterisation, which we shall use to prove Ricci bounds in concrete examples.

As before, we fix an irreducible and reversible Markov kernel K on a finite set $\mathcal{X}$ with steady state π. The associated Markov semigroup shall be denoted by $(P_t)_{t\geq 0}$.

Assumption 2.4.1. *Throughout the remainder of this chapter we assume that θ is the logarithmic mean.*

We are now ready to state the definition, which has already been given in [Maa11, Definition 1.3].

Definition 2.4.2. *We say that K has* non-local Ricci curvature bounded from below *by $\kappa \in \mathbb{R}$ and write $\mathrm{Ric}(K) \geq \kappa$, if the following holds: for every constant speed geodesic $(\rho_t)_{t\in[0,1]}$ in $(\mathscr{P}(\mathcal{X}), \mathcal{W})$ we have*

$$\mathcal{H}(\rho_t) \leq (1-t)\mathcal{H}(\rho_0) + t\mathcal{H}(\rho_1) - \frac{\kappa}{2} t(1-t)\mathcal{W}(\rho_0,\rho_1)^2 \ . \tag{2.4.1}$$

An important role in our analysis is played by the quantity $\mathcal{B}(\rho,\psi)$, which is defined for $\rho \in \mathscr{P}_*(\mathcal{X})$ and $\psi \in \mathbb{R}^{\mathcal{X}}$ by

$$\begin{aligned} \mathcal{B}(\rho,\psi) &:= \frac{1}{2}\big\langle \hat\Delta\rho \cdot \nabla\psi, \nabla\psi \big\rangle_\pi - \big\langle \hat\rho \cdot \nabla\psi \ , \ \nabla\Delta\psi \big\rangle_\pi \\ &= \frac{1}{4} \sum_{x,y,z\in\mathcal{X}} \big(\psi(x) - \psi(y)\big)^2 \Big(\partial_1\theta\big(\rho(x),\rho(y)\big)\big(\rho(z) - \rho(x)\big)K(x,z) \\ &\qquad\qquad + \partial_2\theta\big(\rho(x),\rho(y)\big)\big(\rho(z) - \rho(y)\big)K(y,z)\Big) K(x,y)\pi(x) \\ &\quad - \frac{1}{2} \sum_{x,y,z\in\mathcal{X}} \Big(K(x,z)\big(\psi(z) - \psi(x)\big) - K(y,z)\big(\psi(z) - \psi(y)\big)\Big) \\ &\qquad\qquad \times \big(\psi(x) - \psi(y)\big)\hat\rho(x,y)K(x,y)\pi(x) \ , \end{aligned} \tag{2.4.2}$$

where

$$\hat\Delta\rho(x,y) := \partial_1\theta(\rho(x),\rho(y))\Delta\rho(x) + \partial_2\theta(\rho(x),\rho(y))\Delta\rho(y) \ .$$

The significance of $\mathcal{B}(\rho,\psi)$ is mainly due to the following result:

Proposition 2.4.3. *For $\rho \in \mathscr{P}_*(\mathcal{X})$ and $\psi \in \mathbb{R}^{\mathcal{X}}$ we have*

$$\big\langle \mathrm{Hess}\,\mathcal{H}(\rho)\nabla\psi\ ,\ \nabla\psi\big\rangle_\rho = \mathcal{B}(\rho,\psi)\ .$$

Proof. Take (ρ,ψ) satisfying the geodesic equations (2.3.4), so that

$$\big\langle \mathrm{Hess}\,\mathcal{H}(\rho_t)\nabla\psi_t\ ,\ \nabla\psi_t\big\rangle_{\rho_t} = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{H}(\rho_t)\ .$$

Using the continuity equation we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(\rho_t) &= -\big\langle 1+\log\rho_t, \nabla\cdot(\hat\rho_t\nabla\psi_t)\big\rangle_\pi\\ &= \big\langle \nabla\log\rho_t, \hat\rho_t\cdot\nabla\psi_t\big\rangle_\pi\\ &= \big\langle \nabla\rho_t, \nabla\psi_t\big\rangle_\pi\ .\end{aligned}$$

Furthermore,

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{H}(\rho_t) &= \big\langle \nabla\partial_t\rho_t, \nabla\psi_t\big\rangle_\pi + \big\langle \nabla\rho_t, \nabla\partial_t\psi_t\big\rangle_\pi\\ &= -\big\langle \partial_t\rho_t, \Delta\psi_t\big\rangle_\pi - \big\langle \Delta\rho_t, \partial_t\psi_t\big\rangle_\pi\ .\end{aligned}$$

Using the continuity equation we obtain

$$\begin{aligned}\big\langle \partial_t\rho_t, \Delta\psi_t\big\rangle_\pi &= -\big\langle \nabla\cdot(\hat\rho\nabla\psi_t), \Delta\psi_t\big\rangle_\pi\\ &= \big\langle \hat\rho_t\nabla\psi_t, \nabla\Delta\psi_t\big\rangle_\pi = \big\langle \nabla\psi_t, \nabla\Delta\psi_t\big\rangle_{\rho_t}\ .\end{aligned}$$

Furthermore, applying the geodesic equations (2.3.4) and the detailed balance equations (2.1.1), we infer that

$$\begin{aligned}&\big\langle \Delta\rho_t, \partial_t\psi_t\big\rangle_\pi\\ &= -\frac{1}{2}\sum_{x,y,z\in\mathcal{X}} \big(\psi_t(x)-\psi_t(y)\big)^2\partial_1\theta\big(\rho_t(x),\rho_t(y)\big)\\ &\qquad\qquad \times\big(\rho_t(z)-\rho_t(x)\big)K(x,y)K(x,z)\pi(x)\\ &= -\frac{1}{4}\sum_{x,y,z\in\mathcal{X}} \big(\psi_t(x)-\psi_t(y)\big)^2\Big(\partial_1\theta\big(\rho_t(x),\rho_t(y)\big)\big(\rho_t(z)-\rho_t(x)\big)K(x,z)\\ &\qquad\qquad + \partial_2\theta\big(\rho_t(x),\rho_t(y)\big)\big(\rho_t(z)-\rho_t(y)\big)K(y,z)\Big)K(x,y)\pi(x)\\ &= -\frac{1}{2}\big\langle \hat\Delta\rho_t\ \cdot\ \nabla\psi_t, \nabla\psi_t\big\rangle_\pi\ .\end{aligned}$$

Combining the latter three identities, we arrive at

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{H}(\rho_t) = -\big\langle \nabla\psi_t, \nabla\Delta\psi_t\big\rangle_{\rho_t} + \frac{1}{2}\big\langle \hat\Delta\rho_t\ \cdot\ \nabla\psi_t, \nabla\psi_t\big\rangle_\pi\ ,$$

which is the desired identity. □

Our next aim is to show that κ-convexity of $\mathcal{H}$ along geodesics is equivalent to a lower bound of the Hessian of $\mathcal{H}$ in $\mathscr{P}_*(\mathcal{X})$. Since the Riemannian metric on $(\mathscr{P}(\mathcal{X}), \mathcal{W})$ degenerates at the boundary, this is not an obvious result. In particular, in order to prove the implication "(4) $\Rightarrow$ (3)" below we cannot directly apply the equivalence between the so-called EVI (2.4.4) and the usual gradient flow equation, which holds on complete Riemannian manifolds (see, for example, [Vil09, Proposition 23.1]). Therefore, we take a different approach, based on an argument by Daneri and Savaré [DS08], which avoids delicate regularity issues for geodesics. An additional benefit of this approach is that we expect it to apply in a more general setting where the underlying space $\mathcal{X}$ is infinite, and finite-dimensional Riemannian techniques do not apply at all.

Remark 2.4.4. The quantity $\mathcal{B}(\rho, \psi)$ arises naturally in the Eulerian approach to the Wasserstein metric, as developed in [DS08, OW05]. In fact, in a crucial argument from [DS08], the authors consider a certain two-parameter family of measures (ρ_t^s) and functions (ψ_t^s) on a Riemannian manifold $\mathcal{M}$, and show that

$$\partial_s \mathcal{H}(\rho_t^s) + \frac{1}{2}\partial_t \int_{\mathcal{M}} |\nabla \psi_t^s|^2 \mathrm{d}\rho_t^s = -sB(\rho_t^s, \psi_t^s) \;, \tag{2.4.3}$$

where

$$B(\rho, \psi) := \int_{\mathcal{M}} \Big(\frac{1}{2}\Delta(|\nabla\psi|^2) - \langle \nabla\psi, \nabla\Delta\psi\rangle \Big) \mathrm{d}\rho \;.$$

Since Bochner's formula asserts that

$$B(\rho, \psi) = \int_{\mathcal{M}} |D^2\psi|^2 + \mathrm{Ric}(\nabla\psi, \nabla\psi) \mathrm{d}\rho \;,$$

one obtains a lower bound on B if the Ricci curvature is bounded from below. The lower bound on B can be used to prove an evolution variational inequality, which in turn yields convexity of the entropy along W_2-geodesics.

In our setting, the quantity $\mathcal{B}(\rho, \psi)$ can be regarded as a discrete analogue of $B(\rho, \psi)$. Therefore the inequality $\mathcal{B}(\rho, \psi) \geq \kappa \mathcal{A}(\rho, \psi)$ could be interpreted as a one-sided Bochner inequality, which allows us to adapt the strategy from [DS08] to the discrete setting.

In the following result and the rest of this chapter we shall use the notation

$$\frac{\mathrm{d}^+}{\mathrm{d}t} f(t) = \limsup_{h \downarrow 0} \frac{f(t+h) - f(t)}{h} \;.$$

Theorem 2.4.5. *Let $\kappa \in \mathbb{R}$. For an irreducible and reversible Markov kernel $(\mathcal{X}, K)$ the following assertions are equivalent:*

(i) $\mathrm{Ric}(K) \geq \kappa$ *;*

(ii) *For all* $\rho, \nu \in \mathscr{P}(\mathcal{X})$*, the following 'evolution variational inequality' holds for all* $t \geq 0$*:*

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}^2(P_t\rho,\nu) + \frac{\kappa}{2}\mathcal{W}^2(P_t\rho,\nu) \leq \mathcal{H}(\nu) - \mathcal{H}(P_t\rho)\ ; \tag{2.4.4}$$

(iii) *For all* $\rho, \nu \in \mathscr{P}_*(\mathcal{X})$*,* (2.4.4) *holds for all* $t \geq 0$*;*

(iv) *For all* $\rho \in \mathscr{P}_*(\mathcal{X})$ *and* $\psi \in \mathbb{R}^{\mathcal{X}}$ *we have*

$$\mathcal{B}(\rho,\psi) \geq \kappa\mathcal{A}(\rho,\psi)\ .$$

(v) *For all* $\rho \in \mathscr{P}_*(\mathcal{X})$ *we have*

$$\operatorname{Hess}\mathcal{H}(\rho) \geq \kappa\ ;$$

(vi) *For all* $\bar\rho_0, \bar\rho_1 \in \mathscr{P}_*(\mathcal{X})$ *there exists a constant speed geodesic* $(\rho_t)_{t\in[0,1]}$ *satisfying* $\rho_0 = \bar\rho_0$*,* $\rho_1 = \bar\rho_1$*, and* (2.4.1)*.*

Proof. "(3) ⇒ (2)": This is a special case of [DS08, Theorem 3.3].

"(2) ⇒ (1)": This follows by applying [DS08, Theorem 3.2] to the metric space $(\mathscr{P}(\mathcal{X}), \mathcal{W})$ and the functional $\mathcal{H}$.

"(1) ⇒ (6)": This is clear in view of Theorem 2.3.2.

"(6) ⇒ (5)": Take $\rho \in \mathscr{P}_*(\mathcal{X})$ and $\psi \in \mathbb{R}^{\mathcal{X}}$ and consider the unique solution $(\rho_t, \psi_t)_{t\in(-\varepsilon,\varepsilon)}$ to the geodesic equations with $\rho_0 = \rho$ and $\psi_0 = \psi$ on a sufficiently small time interval around 0. Using the local uniqueness of geodesics and (6), we infer that

$$\operatorname{Hess}\mathcal{H}(\rho)(\nabla\psi) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\Big|_{t=0}\mathcal{H}(\rho_t) \geq \kappa\|\nabla\psi\|_\rho^2$$

(see, for example, the implication "$(ii) \Leftrightarrow (i)$" in [Vil09, Proposition 16.2]).

"(5) ⇒ (4)": This follows from Proposition 2.4.3.

"(4) ⇒ (3)": We follow [DS08]. In view of Lemma 2.2.9 we can find a smooth curve $(\rho^\cdot, \psi^\cdot) \in \mathcal{CE}_1(\nu,\rho)$ satisfying

$$\int_0^1 \mathcal{A}(\rho^s, \psi^s)\mathrm{d}s < \mathcal{W}(\rho,\nu)^2 + \varepsilon\ . \tag{2.4.5}$$

Note in particular that $s \mapsto \rho^s$ and $s \mapsto \psi^s$ are sufficiently regular to apply Lemma 2.4.6 below. Using the notation from this lemma, we infer that

$$\frac{1}{2}\partial_t\mathcal{A}(\rho_t^s, \psi_t^s) + \partial_s\mathcal{H}(\rho_t^s) = -s\mathcal{B}(\rho_t^s, \psi_t^s)\ .$$

Using the assumption that $\mathcal{B} \geq \kappa\mathcal{A}$ we infer that

$$\frac{1}{2}\partial_t\Big(e^{2\kappa st}\mathcal{A}(\rho_t^s, \psi_t^s)\Big) + \partial_s\Big(e^{2\kappa st}\mathcal{H}(\rho_t^s)\Big) \leq 2\kappa t e^{2\kappa st}\mathcal{H}(\rho_t^s)\ .$$

Integration with respect to $t \in [0, h]$ and $s \in [0, 1]$ yields

$$\frac{1}{2}\int_0^1 \Big(e^{2\kappa sh}\mathcal{A}(\rho_h^s, \psi_h^s) - \mathcal{A}(\rho_0^s, \psi_0^s)\Big)\mathrm{d}s + \int_0^h \Big(e^{2\kappa t}\mathcal{H}(\rho_t^1) - \mathcal{H}(\rho_t^0)\Big)\mathrm{d}t \leq 2\kappa \int_0^1 \int_0^h te^{2\kappa st}\mathcal{H}(\rho_t^s)\mathrm{d}t\mathrm{d}s \ .$$

Arguing as in [DS08, Lemma 5.1] we infer that

$$\int_0^1 e^{2\kappa sh}\mathcal{A}(\rho_h^s, \psi_h^s)\mathrm{d}s \geq m(\kappa h)\mathcal{W}^2(P_h\rho, \nu) \ ,$$

where $m(\kappa) = \frac{\kappa e^\kappa}{\sinh(\kappa)}$. Using (2.4.5) together with the fact that the entropy decreases along the heat flow, we infer that

$$\frac{m(\kappa h)}{2}\mathcal{W}^2(P_h\rho, \nu) - \frac{1}{2}\mathcal{W}^2(\rho, \nu) - \varepsilon + E_\kappa(h)\mathcal{H}(P_h\rho) - h\mathcal{H}(\nu) \leq 2\kappa \int_0^1 \int_0^h te^{2\kappa st}\mathcal{H}(\rho_t^s)\mathrm{d}t\mathrm{d}s \ , \tag{2.4.6}$$

where $E_\kappa(h) := \int_0^h e^{2\kappa t}\, dt$. Since $\mathcal{H}$ is bounded, it follows that

$$\lim_{h\downarrow 0}\frac{1}{h}\int_0^1 \int_0^h te^{2\kappa st}\mathcal{H}(\rho_t^s)\mathrm{d}t\mathrm{d}s = 0 \ .$$

Furthermore,

$$\lim_{h\downarrow 0}\frac{1}{h}\Big(E_\kappa(h)\mathcal{H}(P_h\rho) - h\mathcal{H}(\nu)\Big) = \mathcal{H}(\rho) - \mathcal{H}(\nu) \ .$$

Since $\varepsilon > 0$ is arbitrary, (2.4.6) implies that

$$\frac{\mathrm{d}^+}{\mathrm{d}h}\bigg|_{h=0}\left(\frac{m(\kappa h)}{2}\mathcal{W}^2(P_h\rho, \nu)\right) + \mathcal{H}(\rho) - \mathcal{H}(\nu) \leq 0 \ .$$

Taking into account that

$$\frac{\mathrm{d}^+}{\mathrm{d}h}\bigg|_{h=0}\left(\frac{m(\kappa h)}{2}\mathcal{W}^2(P_h\rho, \nu)\right) = \frac{\kappa}{2}\mathcal{W}^2(\rho, \nu) + \frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}h}\bigg|_{h=0}\mathcal{W}^2(P_h\rho, \nu) \ ,$$

we obtain (2.4.4) for $t = 0$, which clearly implies (2.4.4) for all $t \geq 0$. □

The following result, which is used in the proof of Theorem 2.4.5, is a discrete analogue of (2.4.3) and the proof proceeds along the lines of [DS08, Lemma 4.3]. Since the details are slightly different in the discrete setting, we present a proof for the convenience of the reader.

Lemma 2.4.6. *Let $\{\rho^s\}_{s\in[0,1]}$ be a smooth curve in $\mathscr{P}(\mathcal{X})$. For each $t \geq 0$, set $\rho_t^s := e^{st\Delta}\rho^s$, and let $\{\psi_t^s\}_{s\in[0,1]}$ be a smooth curve in $\mathbb{R}^{\mathcal{X}}$ satisfying the continuity equation*

$$\partial_s \rho_t^s + \nabla \cdot (\hat{\rho}_t^s \cdot \nabla \psi_t^s) = 0 \ , \qquad s \in [0,1] \ .$$

Then the identity

$$\frac{1}{2}\partial_t \mathcal{A}(\rho_t^s, \psi_t^s) + \partial_s \mathcal{H}(\rho_t^s) = -s\mathcal{B}(\rho_t^s, \psi_t^s)$$

holds for every $s \in [0,1]$ and $t \geq 0$.

Proof. First of all, we have

$$\begin{aligned} \partial_s \mathcal{H}(\rho_t^s) &= \big\langle 1 + \log \rho_t^s \ , \ \partial_s \rho_t^s \big\rangle_\pi \\ &= -\big\langle 1 + \log \rho_t^s \ , \ \nabla \cdot (\hat{\rho}_t^s \cdot \nabla \psi) \big\rangle_\pi \\ &= \big\langle \nabla \log \rho_t^s \ , \ \hat{\rho}_t^s \cdot \nabla \psi_t^s \big\rangle_\pi \\ &= \big\langle \nabla \rho_t^s \ , \ \nabla \psi_t^s \big\rangle_\pi \\ &= -\big\langle \psi_t^s \ , \ \Delta \rho_t^s \big\rangle_\pi \ . \end{aligned} \tag{2.4.7}$$

Furthermore,

$$\begin{aligned} \frac{1}{2}\partial_t \mathcal{A}(\rho_t^s, \psi_t^s) &= \big\langle \hat{\rho}_t^s \cdot \partial_t \nabla \psi_t^s \ , \ \nabla \psi_t^s \big\rangle_\pi + \frac{1}{2}\big\langle \partial_t \hat{\rho}_t^s \cdot \nabla \psi_t^s \ , \ \nabla \psi_t^s \big\rangle_\pi \\ &=: I_1 + I_2 \ . \end{aligned}$$

In order to simplify I_1 we claim that

$$-\nabla \cdot \big((\partial_t \hat{\rho}_t^s) \cdot \nabla \psi_t^s\big) - \nabla \cdot \big(\hat{\rho}_t^s \cdot \partial_t \nabla \psi_t^s\big) = \Delta \rho_t^s - s\Delta\big(\nabla \cdot (\hat{\rho}_t^s \cdot \nabla \psi_t^s)\big) \ , \tag{2.4.8}$$

$$\partial_t \hat{\rho}_t^s = s\hat{\Delta}\rho_t^s \ . \tag{2.4.9}$$

To show (2.4.8), note that the left-hand side equals $\partial_t \partial_s \rho_t^s$, while the right-hand side equals $\partial_s \partial_t \rho_t^s$. The identity (2.4.9) follows from a straightforward calculation.

Integrating by parts repeatedly and using (2.4.7), (2.4.8) and (2.4.9), we obtain

$$\begin{aligned} I_1 &= -\big\langle \psi_t^s \ , \ \nabla \cdot (\hat{\rho}_t^s \cdot \partial_t \nabla \psi_t^s) \big\rangle_\pi \\ &= \big\langle \psi_t^s \ , \ \Delta \rho_t^s \big\rangle_\pi - s\big\langle \psi_t^s \ , \ \Delta\big(\nabla \cdot (\hat{\rho}_t^s \nabla \psi_t^s)\big) \big\rangle_\pi + \big\langle \psi_t^s \ , \ \nabla \cdot \big((\partial_t \hat{\rho}_t^s) \nabla \psi_t^s\big) \big\rangle_\pi \\ &= -\partial_s \mathcal{H}(\rho_t^s) + s\big\langle \hat{\rho}_t^s \cdot \nabla \psi_t^s \ , \ \nabla \Delta \psi_t^s \big\rangle_\pi - s\big\langle \hat{\Delta}\rho_t^s \cdot \nabla \psi_t^s \ , \ \nabla \psi_t^s \big\rangle_\pi \ . \end{aligned}$$

Taking into account that

$$I_2 = \frac{s}{2}\big\langle \hat{\Delta}\rho_t^s \cdot \nabla \psi_t^s \ , \ \nabla \psi_t^s \big\rangle_\pi \ ,$$

the result follows by summing the expressions for I_1 and I_2. □

The evolution variational inequality (2.4.4) has been extensively studied in the theory of gradient flows in metric spaces [AGS08]. It readily implies a number of interesting properties for the associated gradient flow (see, for example, [DS08, Section 3]). Among them we single out the following κ-contractivity property.

Proposition 2.4.7 (κ-Contractivity of the heat flow)**.** *Let $(\mathcal{X}, K)$ be an irreducible and reversible Markov kernel satisfying* $\mathrm{Ric}(K) \geq \kappa$ *for some $\kappa \in \mathbb{R}$. Then the associated continuous time Markov semigroup $(P_t)_{t\geq 0}$ satisfies*

$$\mathcal{W}(P_t\rho, P_t\sigma) \leq e^{-\kappa t}\mathcal{W}(\rho, \sigma)$$

for all $\rho, \sigma \in \mathscr{P}(\mathcal{X})$ and $t \geq 0$.

Proof. This follows by applying [DS08, Proposition 3.1] to the functional $\mathcal{H}$ on the metric space $(\mathscr{P}(\mathcal{X}), \mathcal{W})$. □

2.5 Examples

In this section we give explicit lower bounds on the non-local Ricci curvature in several examples. Moreover, we present a simple criterion (see Proposition 2.5.4) for proving non-local Ricci curvature bounds. Although the assumptions seem restrictive, the criterion allows us to obtain the sharp Ricci bound for the discrete hypercube. Moreover, it can be combined with the tensorisation result from Section 2.6 in order to prove Ricci bounds in other nontrivial situations. To get started let us consider a particularly simple example.

Example 2.5.1 (The complete graph)*.* Let $\mathcal{K}^n$ denote the complete graph on n vertices and let K_n be the simple random walk on $\mathcal{K}^n$ with transition kernel $K(x,y) = \frac{1}{n}$ for all $x, y \in \mathcal{K}^n$. Note that in this case π is the uniform measure. We will show that $\mathrm{Ric}(K_n) \geq \frac{1}{2} + \frac{1}{2n}$. In view of Theorem 2.4.5 we have to show $\mathcal{B}(\rho, \psi) \geq (\frac{1}{2} + \frac{1}{2n})\mathcal{A}(\rho, \psi)$ for all $\rho \in \mathscr{P}_*(\mathcal{X})$ and $\psi \in \mathbb{R}^{\mathcal{X}}$. Recall the definition (2.4.2) of the quantity $\mathcal{B}$. We calculate explicitly:

$$\begin{aligned}\big\langle \hat{\rho} \cdot \nabla\psi \ , \ \nabla\Delta\psi \big\rangle_\pi \ &= \ \frac{1}{2}\frac{1}{n^3} \sum_{x,y,z\in\mathcal{X}} \hat{\rho}(x,y)\nabla\psi(y,x)\Big[\nabla\psi(x,z) - \nabla\psi(y,z)\Big] \\ &= \ -\frac{1}{2}\frac{1}{n^2}\sum_{x,y} \hat{\rho}(x,y)\big(\nabla\psi(x,y)\big)^2 \ = \ -\mathcal{A}(\rho,\psi) \ .\end{aligned}$$

With the notation $\hat{\rho}_i(x,y) = \partial_i\theta(\rho(x), \rho(y))$ and using equation (2.2.2) we obtain further

$$\begin{aligned}\big\langle \hat{\Delta}\rho \cdot \nabla\psi, \nabla\psi \big\rangle_\pi &= \frac{1}{2}\frac{1}{n^3}\sum_{x,y,z} \big(\nabla\psi(x,y)\big)^2 \Big[\hat{\rho}_1(x,y)(\rho(z)-\rho(x)) \\ &\qquad + \hat{\rho}_2(x,y)(\rho(z)-\rho(y))\Big] \\ &= -\mathcal{A}(\rho,\psi) + \frac{1}{2}\frac{1}{n^3}\sum_{x,y,z} \big(\nabla\psi(x,y)\big)^2 \Big[\hat{\rho}_1(x,y)\rho(z) \\ &\qquad + \hat{\rho}_2(x,y)\rho(z)\Big] \ .\end{aligned}$$

Keeping only the terms with $z = x$ (resp. $z = y$) in the last sum and using (2.2.2) again, we see

$$\big\langle \hat{\Delta}\rho \cdot \nabla\psi, \nabla\psi \big\rangle_\pi \ \geq \ \left(\frac{1}{n} - 1\right)\mathcal{A}(\rho,\psi) \ .$$

Summing up, we obtain $\mathcal{B} \geq (\frac{1}{2}(\frac{1}{n} - 1) + 1)\mathcal{A}$, which yields the claim.

For the rest of this section we let K be an irreducible and reversible Markov kernel on a finite set $\mathcal{X}$. In order to state the criterion and to perform calculations, it will be convenient to write a Markov chain in terms of allowed moves rather than jumps from point to point.

Let G be a set of maps from $\mathcal{X}$ to itself (the allowed moves) and consider a function $c : \mathcal{X} \times G \to \mathbb{R}_+$ (representing the jump rates).

Definition 2.5.2. *We call the pair (G, c) a* mapping representation *of K if the following properties hold:*

(i) The generator $\Delta = K - I$ can be written in the form

$$\Delta\psi(x) \ = \ \sum_{\delta\in G} \nabla_\delta\psi(x)c(x,\delta) \ , \tag{2.5.1}$$

where

$$\nabla_\delta\psi(x) = \psi(\delta x) - \psi(x) \ .$$

(ii) For every $\delta \in G$ there exists a unique $\delta^{-1} \in G$ satisfying $\delta^{-1}(\delta(x)) = x$ for all x with $c(x,\delta) > 0$.

(iii) For every $F : \mathcal{X} \times G \to \mathbb{R}$ we have

$$\sum_{x\in\mathcal{X},\delta\in G} F(x,\delta)c(x,\delta)\pi(x) \ = \ \sum_{x\in\mathcal{X},\delta\in G} F(\delta x, \delta^{-1})c(x,\delta)\pi(x) \ . \tag{2.5.2}$$

Remark 2.5.3. This definition is close in spirit to the recent work [CDPP09], where Γ_2-type calculations have been performed in order to prove strict convexity of the entropy along the heat flow in a discrete setting. Here, we essentially compute the second derivatives of the entropy along $\mathcal{W}$-geodesics. Since the geodesic equations are more complicated than the heat equation, the expressions that we need to work with are somewhat more involved.

Every irreducible, reversible Markov chain has a mapping representation. In fact, an explicit mapping representation can be obtained as follows. For $x, y \in \mathcal{X}$ consider the bijection $t_{\{x,y\}} : \mathcal{X} \to \mathcal{X}$ that interchanges x and y and keeps all other points fixed. Then let G be the set of all these "transpositions" and set $c(x, t_{\{x,y\}}) = K(x, y)$ and $c(x, t_{\{y,z\}}) = 0$ for $x \notin \{y, z\}$. Then (G, c) defines a mapping representation. However, in examples it is often more natural to work with a different mapping representation involving a smaller set G, as we shall see below.

It will be useful to formulate the expressions for $\mathcal{A}$ and $\mathcal{B}$ in this formalism. For this purpose, we note that (2.5.1) implies that

$$\sum_{y\in\mathcal{X}} F(x,y)K(x,y) \;=\; \sum_{\delta\in G} F(x,\delta x)c(x,\delta)$$

for any $F : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ vanishing on the diagonal. As a consequence we obtain

$$\mathcal{A}(\rho,\psi) \;=\; \frac{1}{2} \sum_{x\in\mathcal{X},\delta\in G} \big(\nabla_\delta \psi(x)\big)^2 \hat\rho(x,\delta x)c(x,\delta)\pi(x) \tag{2.5.3}$$

and

$$\begin{aligned}\langle \hat\rho \nabla\psi, \nabla\Delta\psi\rangle_\pi \;=\; \frac{1}{2} \sum_{x\in\mathcal{X}} \sum_{\delta,\eta\in G} \nabla_\delta\psi(x) \Big[&\nabla_\eta \psi(\delta x) c(\delta x,\eta) \\ &- \nabla_\eta\psi(x)c(x,\eta)\Big] \hat\rho(x,\delta x)c(x,\delta)\pi(x)\ .\end{aligned} \tag{2.5.4}$$

Setting for convenience $\partial_i\theta(\rho(x),\rho(y)) =: \hat\rho_i(x,y)$ for $i = 1, 2$ we further get

$$\begin{aligned}\frac{1}{2}\langle \hat\Delta\rho\nabla\psi, \nabla\psi\rangle_\pi \;=\; \frac{1}{4} \sum_{x,\delta,\eta} \big(\nabla_\delta\psi(x)\big)^2 \Big[&\hat\rho_1(x,\delta x) \nabla_\eta\rho(x) c(x,\eta) \\ &+ \hat\rho_2(x,\delta x)\nabla_\eta\rho(\delta x)c(\delta x,\eta)\Big] c(x,\delta)\pi(x)\ .\end{aligned} \tag{2.5.5}$$

Now the expression for $\mathcal{B}(\rho, \psi)$ is obtained as the difference between the two preceding two expressions.

We are now ready to state the announced criterion, which shall be used in Examples 2.5.6 and 2.5.8 below. Intuitively, condition (ii) expresses a certain 'spatial homogeneity', saying that the jump rate in a given direction is the same before and after another jump.

Proposition 2.5.4. *Let K be an irreducible and reversible Markov kernel on a finite set $\mathcal{X}$ and let (G, c) be a mapping representation. Consider the following conditions:*

(i) $\delta \circ \eta \;=\; \eta \circ \delta$ *, for all* $\delta, \eta \in G$,

(ii) $c(\delta x, \eta) \;=\; c(x, \eta)$ *, for all* $x \in \mathcal{X},\ \delta, \eta \in G$,

(iii) $\delta \circ \delta = id$ *, for all* $\delta \in G$.

If (i) and (ii) are satisfied, then $\mathrm{Ric}(K) \geq 0$*. If moreover (iii) is satisfied, then* $\mathrm{Ric}(K) \geq 2C$*, where*

$$C := \min\{c(x,\delta) : x \in \mathcal{X}, \delta \in G \text{ such that } c(x,\delta) > 0\} \ .$$

Remark 2.5.5. Note that requiring *(i)* and *(iii)* simultaneously imposes a very strong restriction on the graph associated with K. We prefer to state the result in this form in order to give a unified proof which applies both to the discrete circle and the discrete hypercube, with optimal constant in the latter case.

Proof of Proposition 2.5.4. In view of Theorem 2.4.5 it is sufficient to show that $\mathcal{B}(\rho,\psi) \geq 0$ resp. $\mathcal{B}(\rho,\psi) \geq 2C\mathcal{A}(\rho,\psi)$ for all $\rho \in \mathscr{P}_*(\mathcal{X})$ and $\psi \in \mathbb{R}^{\mathcal{X}}$. First recall that

$$\mathcal{B}(\rho,\psi) \;=\; -\langle \hat{\rho}\nabla\psi, \nabla\Delta\psi\rangle_\pi + \frac{1}{2}\langle \hat{\Delta}\rho\nabla\psi, \nabla\psi\rangle_\pi \;=:\; T_1 + T_2 \ .$$

Using (2.5.4) and conditions (i) and (ii) we can write the first summand as

$$\begin{aligned} T_1 &= -\frac{1}{2}\sum_{x,\delta,\eta} \nabla_\delta\psi(x)\Big[\nabla_\eta\psi(\delta x) - \nabla_\eta\psi(x)\Big]\hat{\rho}(x,\delta x)c(x,\delta)c(x,\eta)\pi(x) \\ &= -\frac{1}{2}\sum_{x,\delta,\eta} \nabla_\delta\psi(x)\Big[\nabla_\delta\psi(\eta x) - \nabla_\delta\psi(x)\Big]\hat{\rho}(x,\delta x)c(x,\delta)c(x,\eta)\pi(x) \ . \end{aligned}$$

In a similar way we shall write the second summand. Starting from (2.5.5) and invoking (ii) and equation (2.2.2) from Lemma 2.2.2, we obtain

$$\begin{aligned} T_2 &= \frac{1}{4}\sum_{x,\delta,\eta} \big(\nabla_\delta\psi(x)\big)^2\Big[\hat{\rho}_1(x,\delta x)\nabla_\eta\rho(x) \\ &\qquad\qquad + \hat{\rho}_2(x,\delta x)\nabla_\eta\rho(\delta x)\Big]c(x,\delta)c(x,\eta)\pi(x) \\ &= \frac{1}{4}\sum_{x,\delta,\eta} \big(\nabla_\delta\psi(x)\big)^2\Big[\hat{\rho}_1(x,\delta x)\rho(\eta x) \\ &\qquad\qquad + \hat{\rho}_2(x,\delta x)\rho(\eta\delta x) - \hat{\rho}(x,\delta x)\Big]c(x,\delta)c(x,\eta)\pi(x) \ . \end{aligned}$$

Using the reversibility of K in the form of (2.5.2), and again condition (ii) we can write

$$\begin{aligned} T_2 \;=\; & \frac{1}{4}\sum_{x,\delta,\eta}\Bigg(\big(\nabla_\delta\psi(\eta x)\big)^2\Big[\hat\rho_1(\eta x,\delta\eta x)\rho(x)+\hat\rho_2(\eta x,\delta\eta x)\rho(\delta x)\Big] \\ & \qquad - \big(\nabla_\delta\psi(x)\big)^2\hat\rho(x,\delta x)\Bigg)c(x,\delta)c(x,\eta)\pi(x)\ . \end{aligned}$$

Adding a zero, we obtain

$$\begin{aligned} T_2 \;=\; & \frac{1}{4}\sum_{x,\delta,\eta}\Big(\big(\nabla_\delta\psi(\eta x)\big)^2-\big(\nabla_\delta\psi(x)\big)^2\Big)\,\hat\rho(x,\delta x)c(x,\delta)c(x,\eta)\pi(x) \\ & +\frac{1}{4}\sum_{x,\delta,\eta}\big(\nabla_\delta\psi(\eta x)\big)^2\Big[\hat\rho_1(\eta x,\delta\eta x)\rho(x)+\hat\rho_2(\eta x,\delta\eta x)\rho(\delta x) \\ & \qquad\qquad -\hat\rho(x,\delta x)\Big]c(x,\delta)c(x,\eta)\pi(x) \\ =: \;& T_3+T_4\ . \end{aligned}$$

Invoking the inequality (2.2.3) from Lemma 2.2.2, we immediately see that $T_4 \geq 0$. Hence we get

$$\begin{aligned} \mathcal{B}(\rho,\psi) \;\geq\; & T_1+T_3 \\ =\; & \frac{1}{4}\sum_{x,\delta,\eta}\big(\nabla_\delta\psi(\eta x)-\nabla_\delta\psi(x)\big)^2\hat\rho(x,\delta x)c(x,\delta)c(x,\eta)\pi(x) \\ \geq\; & 0\ . \end{aligned}$$

If moreover, condition (iii) is satisfied, the latter estimate can be improved by keeping only the terms with $\eta = \delta$ in the last sum. We thus obtain

$$\begin{aligned} \mathcal{B}(\rho,\psi) \;\geq\; & \frac{C}{4}\sum_{x,\delta}\big(2\nabla_\delta\psi(x)\big)^2\hat\rho(x,\delta x)c(x,\delta)\pi(x) \\ =\; & 2C\mathcal{A}(\rho,\psi)\ . \end{aligned}$$

□

Let us now consider some examples to which Proposition 2.5.4 can be applied.

Example 2.5.6 (The discrete circle). Consider the simple random walk on the discrete circle $C_n = \mathbb{Z}/n\mathbb{Z}$ of n sites with transition kernel $K(m, m-1) = K(m, m+1) = \frac{1}{2}$ for $m \in C_n$. We have the following mapping representation for K. Set $G = \{+, -\}$ where $+(m) = m + 1$ and $-(m) = m - 1$ and let $c(m, +) = c(m, -) = \frac{1}{2}$ for all m. Proposition 2.5.4 immediately yields that $\mathrm{Ric}(K) \geq 0$.

A similar argument can be used to show that the simple random walk on a discrete torus $\mathcal{X} = C_n \times \cdots \times C_n$ has non-negative non-local Ricci curvature. Alternatively, in Proposition 2.6.4 we deduce such a result from the previous example with the help of the tensorisation property proven in Theorem 2.6.2.

Example 2.5.7 (Triangular lattices on the torus). Consider the simple random walk on the two-dimensional discrete torus $T_n = C_n \times C_n$ with transition kernel

$$K\big((m,l),(m+e,l)\big) = K\big((m,l),(m,l+e)\big) = K\big((m,l),(m+e,l+e)\big) = \frac{1}{6} ,$$

where $e = +1, -1$. The corresponding graph structure on the torus is a periodic triangular lattice. A mapping representation for K is obtained by setting $G = \{h_e, v_e, d_e \mid e = +1, -1\}$, where $h_e(m,l) = (m+e,l)$, $v_e(m,l) = (m,l+e)$, $d_e(m,l) = (m+e,l+e)$, and $c(x,\delta) = \frac{1}{6}$ for all $x \in T_n$ and $\delta \in G$. Proposition 2.5.4 then yields that $\mathrm{Ric}(K) \geq 0$.

Example 2.5.8 (The discrete hypercube). Let $\mathcal{Q}^n = \{0,1\}^n$ be the hypercube endowed with the usual graph structure and let K_n be the kernel of the simple random walk on $\mathcal{Q}^n$. The natural mapping representation is given by $G = \{\delta_1, \ldots, \delta_n\}$, where $\delta_i : \mathcal{Q}^n \to \mathcal{Q}^n$ is the map that flips the i-th coordinate, and $c(x, \delta_i) = \frac{1}{n}$ for all $x \in \mathcal{Q}^n$. Here the criterion from Proposition 2.5.4 yields $\mathrm{Ric}(K_n) \geq \frac{2}{n}$. We shall see in Section 2.7 that this bound is optimal.

Alternatively, we can use the fact that $\mathcal{Q}^n$ is a product space and use the tensorisation property Theorem 2.6.2 below. This will allow us to consider asymmetric random walks on the hypercube, as well.

2.6 Basic Constructions

In this section we show how non-local Ricci curvature bounds transform under some basic operations on a Markov kernel. The main result is Theorem 2.6.2, which yields Ricci bounds for product chains. We start with a simple result that shows how Ricci bounds behave when adding laziness.

Let K be an irreducible and reversible Markov kernel on a finite set $\mathcal{X}$. For $\lambda \in (0,1)$ we consider the *lazy* Markov kernel defined by $K_\lambda := (1-\lambda)I + \lambda K$. Clearly, K_λ is irreducible and reversible with the same invariant measure π. With this notation, we have the following result:

Proposition 2.6.1 (Laziness). *Let $\lambda \in (0,1)$. If $\mathrm{Ric}(K) \geq \kappa$ for some $\kappa \in \mathbb{R}$, then the lazy kernel K_λ satisfies*

$$\mathrm{Ric}(K_\lambda) \geq \lambda\kappa .$$

Proof. Writing $\mathcal{A}_\lambda$ and $\mathcal{B}_\lambda$ to denote the lazy versions of $\mathcal{A}$ and $\mathcal{B}$, a direct calculation shows that

$$\mathcal{A}_\lambda(\rho,\psi) = \lambda\mathcal{A}(\rho,\psi) , \qquad \mathcal{B}_\lambda(\rho,\psi) = \lambda^2\mathcal{B}(\rho,\psi)$$

for all $\rho \in \mathscr{P}_*(\mathcal{X})$ and $\psi \in \mathbb{R}^{\mathcal{X}}$. As a consequence,

$$\mathcal{B}_\lambda(\rho,\psi) - \lambda\kappa\mathcal{A}_\lambda(\rho,\psi) = \lambda^2(\mathcal{B}(\rho,\psi) - \kappa\mathcal{A}(\rho,\psi)) .$$

The result thus follows from Theorem 2.4.5. □

We now give a tensorisation property of lower Ricci bounds with respect to products of Markov chains. For $i = 1,\dots,n$, let $(\mathcal{X}_i, K_i)$ be an irreducible, reversible finite Markov chain with steady state π_i, and let α_i be a non-negative number satisfying $\sum_{i=1}^n \alpha_i = 1$. The product chain K_α on the product space $\mathcal{X} = \prod_i \mathcal{X}_i$ is defined for $\mathbf{x} = (x_1,\dots,x_n)$ and $\mathbf{y} = (y_1,\dots,y_n)$ by

$$K_\alpha(\mathbf{x},\mathbf{y}) \;=\; \begin{cases} \sum\limits_{i=1}^n \alpha_i K_i(x_i,x_i) \ , & \text{if } x_i = y_i \ \forall i \ , \\ \alpha_i K_i(x_i,y_i) \ , & \text{if } x_i \neq y_i \text{ and } x_j = y_j \ \forall j \neq i \ , \\ 0 \ , & \text{otherwise} \ . \end{cases}$$

Note that the steady state of K_α is the product $\pi = \pi_1 \otimes \cdots \otimes \pi_n$ of the steady states of K_i.

Theorem 2.6.2 (Tensorisation)**.** *Assume that* $\mathrm{Ric}(K_i) \geq \kappa_i$ *for* $i = 1,\dots,n$. *Then we have*

$$\mathrm{Ric}(K_\alpha) \;\geq\; \min_i \alpha_i\kappa_i \ .$$

Proof. In view of Theorem 2.4.5 we have to show that for any $\rho \in \mathscr{P}_*(\mathcal{X})$ and $\psi : \mathcal{X} \to \mathbb{R}$:

$$\mathcal{B}(\rho,\psi) \;\geq\; (\min_i \alpha_i\kappa_i)\mathcal{A}(\rho,\psi) \ .$$

We will use a mapping representation for the Markov kernel K_α as introduced in Section 2.5. Let (G_i, c_i) be mapping representations of K_i for $i = 1,\dots,n$. To each $\delta \in G_i$ we associate a map $\bar\delta : \mathcal{X} \to \mathcal{X}$ by letting δ act on the i-th coordinate. Let us set $G = \bigcup_i \{\bar\delta : \delta \in G_i\}$ and define $c : \mathcal{X} \times G \to \mathbb{R}_+$ by

$$c(x,\bar\delta) \;:=\; \alpha_i c_i(x_i,\delta) \ , \quad \text{for } \delta \in G_i \ .$$

One easily checks that (G,c) is a mapping representation of K_α. Recalling the expressions (2.5.4),(2.5.5) which constitute $\mathcal{B}$ in mapping representation, we write

$$\mathcal{B}(\rho,\psi) \;=:\; \sum_{x\in\mathcal{X},\delta,\eta\in G} F(x,\delta,\eta) \ .$$

Taking into account the product structure of the chain we can write

$$\mathcal{B}(\rho,\psi) \;=\; \sum_{i,j=1}^n \mathcal{B}_{i,j} \quad \text{with} \quad \mathcal{B}_{i,j} \;=\; \sum_{x\in\mathcal{X}} \sum_{\delta\in G_i,\eta\in G_j} F(x,\bar\delta,\bar\eta) \ .$$

The proof will be finished if we prove the following two assertions:

(i) $\mathcal{B}_{i,j} \geq 0$ for all $i \neq j$,

(ii) $\sum_{i=1}^n \mathcal{B}_{i,i} \geq (\min_i \alpha_i \kappa_i)\mathcal{A}(\rho, \psi)$.

To show (i), first note that for $\delta \in G_i$ and $\eta \in G_j$ the maps $\bar{\delta}$ and $\bar{\eta}$ act on different coordinates if $i \neq j$. Thus we have $\bar{\delta} \circ \bar{\eta} = \bar{\eta} \circ \bar{\delta}$ and furthermore $c(\bar{\delta}x, \bar{\eta}) = c(x, \bar{\eta})$. Note that these are precisely the properties used in the proof Proposition 2.5.4, hence the assertion here follows from the same arguments.

Let us now show (ii). We set $\check{\mathcal{X}}_i = \prod_{j\neq i} \mathcal{X}_j$. For $\check{x}_i \in \check{\mathcal{X}}_i$ we let $\rho^{\check{x}_i}, \psi^{\check{x}_i} : \mathcal{X}_i \to \mathbb{R}$ denote the functions ρ and ψ where all variables except x_i are fixed to $\check{x}_i$. Note that $\rho^{\check{x}_i}$ does not necessarily belong to $\mathscr{P}(\mathcal{X}_i)$, but this will be irrelevant in the calculation below, and we shall use expressions such as $\mathcal{A}(\rho^{\check{x}_i}, \psi^{\check{x}_i})$ by abuse of notation. We also set $\check{\pi}_i = \bigotimes_{j\neq i} \pi_j$. Once more using the product structure of the chain c, we see:

$$\begin{aligned}
&\mathcal{A}(\rho, \psi) \\
&= \frac{1}{2}\sum_{i=1}^n \sum_{x\in\mathcal{X},\delta\in G_i} \big(\nabla_{\bar{\delta}}\psi(x)\big)^2 \hat{\rho}(x, \bar{\delta}x) c(x, \bar{\delta})\pi(x) \\
&= \frac{1}{2}\sum_{i=1}^n \sum_{\check{x}_i\in\check{\mathcal{X}}_i} \sum_{x_i\in\mathcal{X}_i,\delta\in G_i} \big(\nabla_{\delta}\psi^{\check{x}_i}(x_i)\big)^2 \widehat{\rho^{\check{x}_i}}(x_i, \delta x_i)\alpha_i c_i(x_i, \delta)\pi_i(x_i)\check{\pi}_i(\check{x}_i) \\
&= \sum_{i=1}^n \alpha_i \sum_{\check{x}_i\in\check{\mathcal{X}}_i} \mathcal{A}_i(\rho^{\check{x}_i}, \psi^{\check{x}_i})\check{\pi}_i(\check{x}_i) \ ,
\end{aligned}$$

where $\mathcal{A}_i$ (resp. $\mathcal{B}_i$) denotes the function $\mathcal{A}$ (resp. $\mathcal{B}$) associated with the ith chain. Similarly, we obtain

$$\begin{aligned}
\mathcal{B}_{i,i} &= \alpha_i^2 \sum_{\check{x}_i\in\check{\mathcal{X}}_i} \mathcal{B}_i(\rho^{\check{x}_i}, \psi^{\check{x}_i})\check{\pi}_i(\check{x}_i) \\
&\geq \alpha_i^2\kappa_i \sum_{\check{x}_i\in\check{\mathcal{X}}_i} \mathcal{A}_i(\rho^{\check{x}_i}, \psi^{\check{x}_i})\check{\pi}_i(\check{x}_i) \ ,
\end{aligned}$$

where the last inequality holds by assumption on the curvature bound for K_i. Summing over $i = 1, \ldots, n$ we obtain (ii). □

We shall now apply Theorem 2.6.2 to asymmetric random walks on the discrete hypercube. For $p, q \in (0, 1)$ let $K_{p,q}$ be the Markov kernel on the two point space $\{0, 1\}$ defined by $K(0, 1) = p, K(1, 0) = q$. The asymmetric random walk is the n-fold product chain on $\mathcal{Q}^n$ denoted by $K_{p,q,n}$ where $\alpha_i = \frac{1}{n}$. Note that the steady state of $K_{p,q,n}$ is the Bernoulli measure

$$\Big((1-\lambda)\delta_{\{0\}} + \lambda\delta_{\{1\}}\Big)^{\otimes n}$$

with parameter $\lambda = \frac{p}{p+q}$. We then have the following bound on the non-local Ricci curvature:

Proposition 2.6.3. *For $n \geq 1$ we have*

$$\mathrm{Ric}(K_{p,q,n}) \;\geq\; \frac{1}{n}\left(\frac{p+q}{2} + \sqrt{pq}\right) .$$

Proof. The two-point space $\mathcal{Q}^1 = \{0,1\}$ has been analysed in detail in [Maa11]. In particular, [Maa11, Proposition 2.12] asserts that $\mathrm{Ric}(K_{p,q,1}) \geq \kappa_{p,q,1}$, where

$$\kappa_{p,q,1} \;=\; \frac{p+q}{2} + \inf_{-1<\beta<1}\left\{\frac{1}{1-\beta^2}\frac{q(1+\beta)-p(1-\beta)}{\log q(1+\beta) - \log p(1-\beta)}\right\} .$$

In order to estimate the right-hand side, we use the logarithmic-geometric mean inequality to obtain for $\beta \in (-1,1)$,

$$\frac{1}{1-\beta^2}\frac{q(1+\beta)-p(1-\beta)}{\log q(1+\beta) - \log p(1-\beta)} \;\geq\; \sqrt{\frac{pq}{1-\beta^2}} \;\geq\; \sqrt{pq}$$

We thus infer that $\mathrm{Ric}(K_{p,q,1}) \geq \frac{p+q}{2} + \sqrt{pq}$. The general bound then follows immediately from Theorem 2.6.2. □

We shall see in Section 2.7 that this bound is sharp if $p = q$. If $p \neq q$, it should be possible to improve this bound by obtaining a sharper bound in the minimisation problem in the proof above.

As another application of the tensorisation result, we prove non-negativity of the non-local Ricci curvature for the simple random walk on a discrete torus of arbitrary size in any dimension $d \geq 1$.

Let $\mathbf{c} := \{c_n\}_{n=1}^d$ be a sequence of natural numbers and consider the discrete torus

$$T_{\mathbf{c}} := C_{c_1} \times \ldots \times C_{c_d} .$$

The simple random walk $K_{\mathbf{c}}$ on $T_{\mathbf{c}}$ is the d-fold product of simple random walks on the circles of length $c_1, \ldots, c_d$.

Proposition 2.6.4 (d-dimensional torus)**.** *For any $d \geq 1$ and $\mathbf{c} := \{c_n\}_{n=1}^d \in \mathbb{N}^d$ we have*

$$\mathrm{Ric}(K_{\mathbf{c}}) \geq 0 .$$

Proof. This follows from Example 2.5.6 and Theorem 2.6.2. □

2.7 Functional Inequalities

The aim of this section is to prove discrete counterparts to the celebrated theorems by Bakry–Émery and Otto–Villani. Along the way we prove a discrete version of the HWI-inequality, which relates the L^2-Wasserstein distance to the entropy and the

Fisher information. As announced in the introduction, we shall follow the approach from Otto–Villani, which relies on the fact that the heat flow is the gradient flow of the entropy. Therefore, the role of the L^2-Wasserstein distance will be taken over by the distance $\mathcal{W}$.

We fix a finite set $\mathcal{X}$ and an irreducible and reversible Markov kernel K with steady state π. Recall that the relative entropy of a density $\rho \in \mathscr{P}(\mathcal{X})$ is defined by

$$\mathcal{H}(\rho) \;=\; \sum_{x\in\mathcal{X}} \rho(x)\log\rho(x)\pi(x)\ .$$

As before, we consider a discrete analogue of the Fisher information, given for $\rho \in \mathscr{P}_*(\mathcal{X})$ by

$$\mathcal{I}(\rho) \;=\; \frac{1}{2}\sum_{x,y\in\mathcal{X}} \big(\rho(x)-\rho(y)\big)\big(\log\rho(x)-\log\rho(y)\big)K(x,y)\pi(x)\ .$$

If $\rho(x) = 0$ for some $x \in \mathcal{X}$, we set $\mathcal{I}(\rho) = +\infty$. Note that this quantity can be rewritten in the form $\mathcal{I}(\rho) = \|\nabla\log\rho\|_\rho^2$ using the definition of the logarithmic mean. The relevance of $\mathcal{I}$ in this setting is due to the fact that it describes the entropy dissipation along the heat flow:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(P_t\rho) \;=\; -\,\mathcal{I}(P_t\rho)\ . \tag{2.7.1}$$

The following proposition gives an upper bound for the speed of the heat flow measured in the metric $\mathcal{W}$.

Proposition 2.7.1. *Let $\rho,\sigma \in \mathscr{P}(\mathcal{X})$. For all $t > 0$ we have*

$$\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}(P_t\rho,\sigma) \;\le\; \sqrt{\mathcal{I}(P_t\rho)}\ . \tag{2.7.2}$$

In particular, the metric derivative of the heat flow with respect to $\mathcal{W}$ satisfies the bound $|(P_t\rho)'| \le \sqrt{\mathcal{I}(P_t\rho)}$. If ρ belongs to $\mathscr{P}_*(\mathcal{X})$, then (2.7.2) holds at $t = 0$ as well.

Proof. Let us set $\rho_t := P_t\rho$. Elementary Markov chain theory guarantees that $\rho_t \in \mathscr{P}_*(\mathcal{X})$ for all $t > 0$ and that the map $t \mapsto \rho_t$ is smooth. To prove (2.7.2) we use the triangle inequality and obtain

$$\begin{aligned}\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}(\rho_t,\sigma) \;&=\; \limsup_{s\searrow 0}\frac{1}{s}\big(\mathcal{W}(\rho_{t+s},\sigma)-\mathcal{W}(\rho_t,\sigma)\big)\\ &\le\; \limsup_{s\searrow 0}\frac{1}{s}\mathcal{W}(\rho_t,\rho_{t+s})\ .\end{aligned}$$

Note that the couple $(\rho_r, -\log \rho_r)_{r\in[0,1]}$ solves the continuity equation (2.1.2). From the definition of $\mathcal{W}$ we thus obtain the estimate

$$\begin{aligned} \limsup_{s\searrow 0} \frac{1}{s}\mathcal{W}(\rho_t, \rho_{t+s}) &\leq \limsup_{s\searrow 0} \frac{1}{s} \int_t^{t+s} \|\nabla \log \rho_r\|_{\rho_r} \mathrm{d}r \\ &= \limsup_{s\searrow 0} \frac{1}{s} \int_t^{t+s} \sqrt{\mathcal{I}(\rho_r)} \mathrm{d}r \\ &= \sqrt{\mathcal{I}(\rho_t)} \ . \end{aligned}$$

The last equality holds since $r \mapsto \sqrt{\mathcal{I}(\rho_r)}$ is a continuous function. □

Let us now recall from Section 2.1 the functional inequalities that will be studied. Recall that $\mathbf{1} \in \mathscr{P}(\mathcal{X})$ denotes the density of the stationary distribution, which is everywhere equal to 1.

Definition 2.7.2. *The Markov kernel K satisfies*

(i) a modified logarithmic Sobolev *inequality with constant $\lambda > 0$ if for all $\rho \in \mathscr{P}(\mathcal{X})$*

$$\mathcal{H}(\rho) \leq \frac{1}{2\lambda}\mathcal{I}(\rho) \ . \tag{MLSI(λ)}$$

(ii) an H$\mathcal{W}$I *inequality with constant $\kappa \in \mathbb{R}$ if for all $\rho \in \mathscr{P}(\mathcal{X})$*

$$\mathcal{H}(\rho) \leq \mathcal{W}(\rho, \mathbf{1})\sqrt{\mathcal{I}(\rho)} - \frac{\kappa}{2}\mathcal{W}(\rho, \mathbf{1})^2 \ . \tag{H$\mathcal{W}$I(κ)}$$

(iii) a modified Talagrand *inequality with constant $\lambda > 0$ if for all $\rho \in \mathscr{P}(\mathcal{X})$*

$$\mathcal{W}(\rho, \mathbf{1}) \leq \sqrt{\frac{2}{\lambda}\mathcal{H}(\rho)} \ . \tag{T$_\mathcal{W}$(λ)}$$

(iv) a Poincaré *inequality with constant $\lambda > 0$ if for all $\varphi \in \mathbb{R}^X$ with $\sum_x \varphi(x)\pi(x) = 0$*

$$\|\varphi\|_\pi^2 \leq \frac{1}{\lambda}\|\nabla\varphi\|_\pi^2 \ . \tag{P(λ)}$$

The following result is a discrete analogue of a result by Otto and Villani [OV00].

Theorem 2.7.3. *Assume that* $\mathrm{Ric}(K) \geq \kappa$ *for some* $\kappa \in \mathbb{R}$. *Then K satisfies* H$\mathcal{W}$I(κ).

Proof. Fix $\rho \in \mathscr{P}(\mathcal{X})$. Without restriction we can assume that $\rho > 0$ since otherwise $\mathcal{I}(\rho) = +\infty$ and there is nothing to prove. Let $\rho_t = P_t\rho$ where $P_t = e^{t(K-I)}$ is the heat semigroup. From Theorem 2.4.5 and the lower bound on the Ricci curvature we know that the curve (ρ_t) satisfies EVI(κ), i.e., equation (2.4.4). Choosing, in particular, $\nu = \mathbf{1}$ and $t = 0$ in the EVI we obtain the inequality

$$\mathcal{H}(\rho) \ \leq \ -\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\bigg|_{t=0}\mathcal{W}(\rho_t, \mathbf{1})^2 - \frac{\kappa}{2}\mathcal{W}(\rho, \mathbf{1})^2 \ .$$

To finish the proof we show that

$$-\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\bigg|_{t=0}\mathcal{W}(\rho_t, \mathbf{1})^2 \ \leq \ \mathcal{W}(\rho, \mathbf{1})\sqrt{\mathcal{I}(\rho)} \ .$$

Indeed, using the triangle inequality we estimate

$$\begin{aligned}
-\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\bigg|_{t=0}\mathcal{W}(\rho_t, \mathbf{1})^2 \ &= \ \liminf_{s\searrow 0}\frac{1}{2s}\left(\mathcal{W}(\rho, \mathbf{1})^2 - \mathcal{W}(\rho_s, \mathbf{1})^2\right) \\
&\leq \ \limsup_{s\searrow 0}\frac{1}{2s}\left(\mathcal{W}(\rho, \rho_s)^2 + 2\mathcal{W}(\rho, \rho_s)\cdot\mathcal{W}(\rho, \mathbf{1})\right) \ ,
\end{aligned}$$

Using the estimate (2.7.2) from Proposition 2.7.1 with $\sigma = \rho$ and $t = 0$ we see that the second term on the right-hand side is bounded by $\mathcal{W}(\rho, \mathbf{1})\sqrt{\mathcal{I}(\rho)}$ while the first term vanishes. □

The following result is now a simple consequence.

Theorem 2.7.4 (Discrete Bakry–Émery Theorem)**.** *Assume that* Ric(K) $\geq \lambda$ *for some* $\lambda > 0$. *Then* K *satisfies* MLSI(λ).

Proof. By Theorem 2.7.3 K satisfies H$\mathcal{W}$I(λ). From this we derive MLSI(λ) by an application of Young's inequality :

$$xy \ \leq \ cx^2 + \frac{1}{4c}y^2 \qquad \forall x, y \in \mathbb{R}\ ,\ c > 0\ ,$$

in which we set $x = \mathcal{W}(\rho, 1),\ y = \sqrt{\mathcal{I}(\rho)}$ and $c = \frac{\lambda}{2}$. □

Theorem 2.7.5 (Discrete Otto–Villani Theorem)**.** *Assume that* K *satisfies* MLSI(λ) *for some* $\lambda > 0$. *Then* K *also satisfies* $\mathrm{T}_{\mathcal{W}}(\lambda)$.

Proof. It is sufficient to prove that $\mathrm{T}_{\mathcal{W}}(\lambda)$ holds for any $\rho \in \mathscr{P}_*(\mathcal{X})$. The inequality for general ρ can then be obtained by an easy approximation argument taking into account the continuity of $\mathcal{W}$ with respect to the Euclidean metric.

So, fix $\rho \in \mathscr{P}_*(\mathcal{X})$ and set $\rho_t = P_t\rho$. First note that as $t \to \infty$, we have

$$\mathcal{H}(\rho_t) \to 0 \quad \text{and} \quad \mathcal{W}(\rho, \rho_t) \to \mathcal{W}(\rho, \mathbf{1})\ . \tag{2.7.3}$$

Indeed, by elementary Markov chain theory, we know that as $t \to \infty$, one has $\rho_t \to \mathbf{1}$ in, say, the Euclidean distance. The claim follows immediately from the continuity of $\mathcal{H}$ and $\mathcal{W}$ with respect to the Euclidean distance, the latter being a consequence of, for example, Proposition 2.2.14.

We now define the function $F : \mathbb{R}_+ \to \mathbb{R}_+$ by

$$F(t) \ := \ \mathcal{W}(\rho, \rho_t) + \sqrt{\frac{2}{\lambda}\mathcal{H}(\rho_t)}\ .$$

Obviously we have $F(0) = \sqrt{\frac{2}{\lambda}\mathcal{H}(\rho)}$ and by (2.7.3) we have that $F(t) \to \mathcal{W}(\rho, \mathbf{1})$ as $t \to \infty$. Hence it is sufficient to show that F is non-increasing. To this end we show that its upper right derivative is non-positive. If $\rho_t \neq \mathbf{1}$ we deduce from Proposition 2.7.1 that

$$\frac{\mathrm{d}^+}{\mathrm{d}t}F(t) \ \le \ \sqrt{\mathcal{I}(\rho_t)} - \frac{\mathcal{I}(\rho_t)}{\sqrt{2\lambda\mathcal{H}(\rho_t)}} \ \le \ 0\ ,$$

where we used MLSI(λ) in the last inequality. If $\rho_t = \mathbf{1}$, then the relation also holds true, since this implies that $\rho_r = \mathbf{1}$ for all $r \ge t$. □

In a classical continuous setting it is well known that a logarithmic Sobolev inequality implies a Poincaré inequality by linearisation. Let us make this explicit in the present discrete context. Fix $\varphi \in \mathbb{R}^{\mathcal{X}}$ satisfying $\sum_x \varphi(x)\pi(x) = 0$ and for sufficiently small $\varepsilon > 0$ set $\rho^\varepsilon = \mathbf{1} + \varepsilon\varphi \in \mathscr{P}_*(\mathcal{X})$. One easily checks that as $\varepsilon \to 0$ we have:

$$\frac{1}{\varepsilon^2}\mathcal{H}(\rho^\varepsilon) \longrightarrow \frac{1}{2}\|\varphi\|_\pi^2\ , \qquad \frac{1}{\varepsilon^2}\mathcal{I}(\rho^\varepsilon) \longrightarrow \|\nabla\varphi\|_\pi^2\ .$$

Thus assuming MLSI(λ) holds and applying it to ρ^ε we get the Poincaré inequality P(λ). In [OV00] it has been shown that the Poincaré inequality can also be obtained from Talagrand's inequality by linearisation. The same is true for the modified Talagrand inequality involving the distance $\mathcal{W}$.

Proposition 2.7.6. *Assume that K satisfies $T_{\mathcal{W}}(\lambda)$ for some $\lambda > 0$. Then K also satisfies P(λ). In particular,* Ric$(K) \ge \lambda$ *implies* P(λ).

Proof. Assume that $\mathrm{T}_{\mathcal{W}}(\lambda)$ holds and let us show P(λ). The second assertion of the proposition then follows from Theorem 2.7.4 and Theorem 2.7.5. So fix $\varphi \in \mathbb{R}^{\mathcal{X}}$ satisfying $\sum_x \varphi(x)\pi(x) = 0$ and for sufficiently small $\varepsilon > 0$ set $\rho^\varepsilon = \mathbf{1} + \varepsilon\varphi \in \mathscr{P}_*(\mathcal{X})$. Let $(\rho^\varepsilon_\cdot, V^\varepsilon_\cdot) \in \mathcal{CE}'_1(\rho^\varepsilon, \mathbf{1})$ be an action minimising curve. Now we write, using the continuity equation,

$$\begin{aligned}\sum_x \varphi(x)^2\pi(x) \ &= \ \frac{1}{\varepsilon}\left[\sum_x \varphi(x)\big(\rho^\varepsilon(x) - 1\big)\pi(x)\right]\\ &= \ \frac{1}{2\varepsilon}\int_0^1 \sum_{x,y} \nabla\varphi(x,y)V_t^\varepsilon(x,y)K(x,y)\pi(x)\mathrm{d}t\ .\end{aligned}$$

Using Hölder's inequality we can estimate

$$\begin{aligned}\sum_x \varphi(x)^2\pi(x) &\leq \frac{1}{\varepsilon}\left(\int_0^1 \|\nabla\varphi\|^2_{\rho_t^\varepsilon}\mathrm{d}t\right)^{\frac{1}{2}}\left(\int_0^1 \mathcal{A}'(\rho_t^\varepsilon, V_t^\varepsilon)\mathrm{d}t\right)^{\frac{1}{2}}\\ &= \left(\frac{1}{2}\sum_{x,y}\big(\nabla\varphi(x,y)\big)^2 g^\varepsilon(x,y)K(x,y)\pi(x)\right)^{\frac{1}{2}}\frac{1}{\varepsilon}\mathcal{W}(\rho^\varepsilon,\mathbf{1})\ ,\end{aligned}$$

where $g^\varepsilon \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ is defined by $g^\varepsilon(x,y) = \int_0^1 \hat{\rho}_t^\varepsilon(x,y)\mathrm{d}t$. Using $\mathrm{T}_{\mathcal{W}}(\lambda)$ we arrive at

$$\|\varphi\|_\pi^2 \ \leq\ \|(\nabla\varphi)\sqrt{g^\varepsilon}\|_\pi \frac{1}{\varepsilon}\sqrt{\frac{2}{\lambda}\mathcal{H}(\rho^\varepsilon)}\ .$$

The proof will be finished if we show that as ε goes to 0

$$\frac{1}{\varepsilon}\sqrt{\frac{2}{\lambda}\mathcal{H}(\rho^\varepsilon)} \ \longrightarrow\ \sqrt{\frac{1}{\lambda}}\|\varphi\|_\pi\ , \qquad \|(\nabla\varphi)\sqrt{g^\varepsilon}\|_\pi \ \longrightarrow\ \|\nabla\varphi\|_\pi\ .$$

As before, the first statement is easily checked. For the second statement it is sufficient to show that $\rho_t^\varepsilon \to \mathbf{1}$ uniformly in t as $\varepsilon \to 0$, as this implies that $g^\varepsilon \to 1$. Since $\mathcal{W}(\rho^\varepsilon, \mathbf{1}) \to 0$ as $\varepsilon \to 0$, this follows immediately from the estimate

$$\mathcal{W}(\rho^\varepsilon,\mathbf{1}) \ \geq\ \sup_t \mathcal{W}(\rho_t^\varepsilon,\mathbf{1}) \ \geq\ \sup_t \sum_x \pi(x)\,|\rho_t^\varepsilon(x) - 1|\ ,$$

where we used that $(\rho_t^\varepsilon)_{t\in[0,1]}$ is a geodesic and the fact that $\mathcal{W}$ is an upper bound for the total variation distance (see Proposition 2.2.12). □

In the following result we use the probabilistic notation

$$\mathbb{E}_\pi[\varphi] = \sum_{x\in\mathcal{X}} \varphi(x)\pi(x)$$

for functions $\varphi : \mathcal{X} \to \mathbb{R}$.

Proposition 2.7.7. *Assume that K satisfies $\mathrm{T}_{\mathcal{W}}(\lambda)$ for some $\lambda > 0$. Then the $T_1(2\lambda)$ inequality holds with respect to the graph distance:*

$$W_{1,g}(\rho,\mathbf{1}) \ \leq\ \sqrt{\frac{1}{\lambda}\mathcal{H}(\rho)}\ . \tag{2.7.4}$$

Furthermore, the sub-Gaussian inequality

$$\mathbb{E}_\pi\big[e^{t(\varphi - \mathbb{E}_\pi[\varphi])}\big] \leq \exp\left(\frac{t^2}{4\lambda}\right) \tag{2.7.5}$$

holds for all $t > 0$ and every function $\varphi : \mathcal{X} \to \mathbb{R}$ that is 1-Lipschitz with respect to the graph distance on $\mathcal{X}$.

Proof. The T_1-inequality (2.7.4) follows immediately from Proposition 2.2.12. The inequalities (2.7.4) and (2.7.5) are equivalent, as has been shown in [BG99]. □

Arguing again exactly as in [OV00], we infer that a modified Talagrand inequality implies a modified log-Sobolev inequality (with some loss in the constant), provided that the non-local Ricci curvature is not too bad.

Proposition 2.7.8. *Suppose that* K *satisfies* $\mathrm{T}_{\mathcal{W}}(\lambda)$ *for some* $\lambda > 0$ *and that* $\mathrm{Ric}(K) \geq \kappa$ *for some* $\kappa > -\lambda$. *Then* K *satisfies* $\mathrm{MLSI}(\widetilde{\lambda})$, *where*

$$\widetilde{\lambda} = \max\left\{ \frac{\lambda}{4}\Big(1 + \frac{\kappa}{\lambda}\Big)^2 \ , \ \kappa \right\} .$$

Proof. This is an immediate consequence of the $\mathrm{H}\mathcal{W}\mathrm{I}(\kappa)$-inequality and an elementary computation (see [OV00, Corollary 3.1]). □

As an application of the results proved in this section, we will show how non-local Ricci curvature bounds can be used to recover functional inequalities with sharp constants in an important example.

Example 2.7.9 (Discrete hypercube). In Example 2.5.8 and Proposition 2.6.3 we proved that the Markov kernel K_n associated with the simple random walk on the discrete hypercube $\mathcal{Q}^n = \{0,1\}^n$ has non-local Ricci curvature bounded from below by $\frac{2}{n}$. Applying Theorem 2.7.4 and Proposition 2.7.7 in this setting, we obtain the following result. We shall write $y \sim x$ if $K(x,y) > 0$.

Corollary 2.7.10. *The simple random walk on* $\mathcal{Q}^n$ *has the following properties:*

(i) the modified log-Sobolev inequality $\mathrm{MLSI}(\frac{2}{n})$ *holds, that is, for all* $\rho \in \mathscr{P}_*(\mathcal{Q}^n)$ *we have*

$$\sum_{x \in \mathcal{Q}^n} \rho(x) \log \rho(x) \ \leq \ \frac{1}{8} \sum_{x \in \mathcal{Q}^n, y \sim x} (\rho(x) - \rho(y)) \left(\log \rho(x) - \log \rho(y)\right) \ .$$

(ii) the Poincaré inequality $\mathrm{P}(\frac{2}{n})$ *holds, that is, for all* $\varphi : \mathcal{Q}^n \to \mathbb{R}$ *we have*

$$\sum_{x \in \mathcal{Q}^n} \varphi(x)^2 \ \leq \ \frac{1}{4} \sum_{x \in \mathcal{Q}^n, y \sim x} (\varphi(x) - \varphi(y))^2 \ .$$

(iii) The sub-Gaussian inequality (2.7.5) *holds with* $\lambda = \frac{2}{n}$.

In all cases the constants are optimal (see [BT06, Example 3.7] and [BHT06, Proposition 2.3] respectively). Moreover, the optimality in (3) implies that the constant $\lambda = \frac{2}{n}$ in the modified Talagrand inequality for the discrete cube is sharp, as well.

We finish this chapter by remarking that modified logarithmic Sobolev inequalities for appropriately rescaled product chains on the discrete hypercube $\{-1,1\}^n$ can be used to prove a similar inequality for Poisson measures by passing to the limit $n \to \infty$ (see [Led01, Section 5.4] for an argument along these lines involving a slightly different modified log Sobolev inequality). All of the functional inequalities in Theorem 2.1.5 are compatible with this limit. However, the sub-Gaussian estimate will (of course) not hold for the limiting Poisson law. This does not contradict the results in this section, since the sub-Gaussian estimates here are obtained using the lower bound for $\mathcal{W}$ in terms of W_1, which relies on the normalisation assumption $\sum_y K(x,y) = 1$, which does not hold in the Poissonian limit.

3 Gradient flows of the entropy for jump processes

In this chapter we introduce a new transport distance between probability measures on $\mathbb{R}^d$ that is built from a Lévy jump kernel. It is defined via a non-local variant of the Benamou-Brenier formula. We study geometric and topological properties of this distance, in particular we prove existence of geodesics. For translation invariant jump kernels we identify the semigroup generated by the associated non-local operator as the gradient flow of the relative entropy w.r.t. the new distance and show that the entropy is convex along geodesics.

3.1 Main results

Let us now discuss the setting and the main results of this chapter in more detail. Let ν be a symmetric Lévy measure on $\mathbb{R}^d$, i.e. a Borel measure satisfying $\nu(\{0\}) = 0$ and $\nu(A) = \nu(-A)$ for all $A \subset \mathbb{R}^d$ as well as

$$\int_{\mathbb{R}^d} (1 \wedge |z|^2)\nu(\mathrm{d}z) < \infty \ .$$

Associated to ν is a non-local operator $\mathcal{L}$ given by

$$\mathcal{L}u(x) \ = \ \int \frac{1}{2}\big(u(x+z) + u(x-z) - 2u(x)\big)\nu(\mathrm{d}z) \ .$$

The operator $\mathcal{L}$ is the general form of the generator of a pure jump Lévy process. Here the jump measure $\nu(\mathrm{d}z)$ gives the intensity of jumps from x to $x+z$. The Lévy process generated by $\mathcal{L}$ is characterised in the sense of the Lévy-Khintchine formula (see e.g. [App04]) by the parameters $(0, 0, \nu)$, i.e. drift and diffusion part vanish. We will be particularly interested in the choice $\nu(\mathrm{d}z) = c_\alpha |z|^{-\alpha-d}\,\mathrm{d}z$ with $\alpha \in (0, 2)$ and a suitable constant c_α. In this case $\mathcal{L} = -(-\Delta)^{\frac{\alpha}{2}}$ is the fractional Laplacian, the pseudo differential operator with symbol $|\xi|^\alpha$. The associated Lévy process is the symmetric α-stable process.

The main result of this chapter is a characterisation of the evolution equation

$$\partial_t u \ = \ \mathcal{L}u \tag{3.1.1}$$

as the gradient flow of the entropy with respect to a suitable metric on the space of probability measures. The usual Wasserstein distance is not appropriate for this purpose and one of the main contributions of this chapter is the construction of a new transport distance between probability measures on $\mathbb{R}^d$ which is non-local in nature and allows for the desired gradient flow interpretation.

A non-local transport distance

The construction of the new transport distance is more general and does not only apply to Lévy processes but also to inhomogeneous jump processes where the intensity of jumps from x to y is given by a space dependent measure $J(x, \mathrm{d}y)$.

We fix a jump kernel $(J(x,\cdot), x \in \mathbb{R}^d)$. By this we mean that for all $x \in \mathbb{R}^d$ $J(x,\cdot)$ is a Radon measure on $\mathbb{R}^d \setminus \{x\}$ depending measurably on x. Throughout this text J shall satisfy the following

Assumption 3.1.1. *For every bounded continuous function $f : \mathbb{R}^d \to \mathbb{R}$ the mapping*

$$x \ \mapsto \ \int f(y)(1 \wedge |x-y|^2)J(x, \mathrm{d}y)$$

is again bounded and continuous.

In particular $(J(x,\cdot), x \in \mathbb{R}^d)$ is a so called Lévy kernel (see e.g. [App04, Ch. 3.5]). Further let m be a Radon measure on $\mathbb{R}^d$. We assume that J is reversible w.r.t. m, i.e. the measure $J(x, \mathrm{d}y)m(\mathrm{d}x)$ is symmetric.

Let us first give a heuristic description of the new distance before we sketch the rigorous construction. The construction is motivated by the dynamical characterisation of the L^2-Wasserstein distance via the Benamou–Brenier formula discussed in Section 1.1 and is based on non-local analogues of the formulas (1.1.1) and (1.1.2).

For a function $\psi : \mathbb{R}^d \to \mathbb{R}$ we will denote by $\overline{\nabla}\psi(x,y) = \psi(y) - \psi(x)$ its discrete gradient. In order to obtain a metric with the desired properties it is necessary to introduce a function $\theta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ satisfying Assumption 3.2.1 below and to consider the mean $\hat{\rho}(x,y) := \theta(\rho(x), \rho(y))$ of a given density $\rho : \mathbb{R}^d \to \mathbb{R}$ at different points. We will be mostly interested in the logarithmic mean

$$\theta(s,t) = \frac{s-t}{\log s - \log t} \tag{3.1.2}$$

but for future use we allow for more generality in the construction. Following the approach of [Maa11] one is led to consider the following 'distance'. Given probability measures $\bar{\mu}_0 = \bar{\rho}_0 m$ and $\bar{\mu}_1 = \bar{\rho}_1 m$ set

$$\widetilde{\mathcal{W}}(\bar{\mu}_0, \bar{\mu}_1)^2 \ := \ \inf_{\rho,\psi} \frac{1}{2}\int_0^1 \int \left|\overline{\nabla}\psi_t(x,y)\right|^2 \hat{\rho}_t(x,y) J(x,\mathrm{d}y)m(\mathrm{d}x)\mathrm{d}t \ , \tag{3.1.3}$$

where the infimum is taken over all functions ρ and ψ satisfying the 'continuity equation'

$$\begin{cases} \partial_t \rho_t + \overline{\nabla}\cdot(\hat{\rho}_t \overline{\nabla}\psi_t) \ = \ 0 \ , \\ \rho_0 = \bar{\rho}_0 \ , \ \ \rho_1 = \bar{\rho}_1 \ , \end{cases} \tag{3.1.4}$$

in distribution sense, i.e. for every test function $\varphi \in C_c^\infty((0,1)\times\mathbb{R}^d)$ we have

$$\int_0^1 \int \partial_t\varphi\rho_t \mathrm{d}m\mathrm{d}t - \frac{1}{2}\int_0^1 \int \overline{\nabla}\varphi_t(x,y)\overline{\nabla}\psi_t(x,y)\hat{\rho}_t(x,y)J(x,\mathrm{d}y)m(\mathrm{d}x)\mathrm{d}t \ = \ 0 \ .$$

For the rigorous construction of the new transport distance we will not address the variational problem (3.1.3) directly. Instead, we will adopt a measure theoretic point of view and recast it in the more natural relaxed setting of time-dependent families of measures. Let us briefly sketch this approach.

Denote by $\mathscr{P}(\mathbb{R}^d)$ the space of Borel probability measures on $\mathbb{R}^d$. We define the set $G := \{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d \ : \ x \neq y\}$ and fix the measure $\gamma(\mathrm{d}x, \mathrm{d}y) = J(x, \mathrm{d}y)m(\mathrm{d}x)$. We replace ρ by a continuous curve $t \mapsto \mu_t = \rho_t m$ in $\mathscr{P}(\mathbb{R}^d)$ and ψ_t induces a family of signed Radon measures $\boldsymbol{\nu}_t(\mathrm{d}x, \mathrm{d}y) = \overline{\nabla}\psi_t(x,y)\hat{\rho}_t(x,y)\gamma(\mathrm{d}x, \mathrm{d}y)$ on G. The couple $(\mu, \boldsymbol{\nu})$ now satisfies the linear equation

$$\begin{cases} \partial_t \mu_t + \overline{\nabla} \cdot \boldsymbol{\nu}_t \ = \ 0 \ , \\ \mu_0 = \bar{\mu}_0, \ \ \mu_1 = \bar{\mu}_1 \end{cases} \tag{3.1.5}$$

in distribution sense, i.e. for all test functions $\varphi \in C_c^\infty((0,1) \times \mathbb{R}^d)$:

$$\int_0^1 \int \partial_t \varphi \mathrm{d}\mu_t \mathrm{d}t + \frac{1}{2} \int_0^1 \int \overline{\nabla}\varphi(x,y)\boldsymbol{\nu}_t(\mathrm{d}x, \mathrm{d}y)\mathrm{d}t \ = \ 0 \ .$$

The quantity to be minimised in (3.1.3) can now be rewritten as

$$\frac{1}{2} \int_0^1 \int \left| \frac{\mathrm{d}\boldsymbol{\nu}_t}{\mathrm{d}\gamma}(x,y) \right|^2 \theta \left(\frac{\mathrm{d}\mu_t}{\mathrm{d}m}(x), \frac{\mathrm{d}\mu_t}{\mathrm{d}m}(y) \right)^{-1} \gamma(\mathrm{d}x, \mathrm{d}y)\mathrm{d}t \ .$$

We will define a distance $\mathcal{W}$ by proceeding as follows. To any $\mu \in \mathscr{P}(\mathbb{R}^d)$ we associate two Radon measures on G by setting $\mu^1(\mathrm{d}x, \mathrm{d}y) = J(x, \mathrm{d}y)\mu(\mathrm{d}x)$ and $\mu^2(\mathrm{d}x, \mathrm{d}y) = J(y, \mathrm{d}x)\mu(\mathrm{d}y)$. Given a Radon measure $\boldsymbol{\nu}$ on G we choose a reference measure σ on G such that $\boldsymbol{\nu} = w\sigma$ and $\mu^i = \rho^i \sigma, \ i = 1,2$ are all absolutely continuous w.r.t. σ. Then we define the action functional by

$$\mathcal{A}(\mu, \boldsymbol{\nu}) \ := \ \frac{1}{2} \int \left| \frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\sigma} \right|^2 \theta \left(\frac{\mathrm{d}\mu^1}{\mathrm{d}\sigma}, \frac{\mathrm{d}\mu^2}{\mathrm{d}\sigma} \right)^{-1} \mathrm{d}\sigma \ .$$

Assumptions on θ will guarantee that the map $(w,s,t) \mapsto w^2\theta(s,t)^{-1}$ is homogeneous, hence the definition of $\mathcal{A}$ is independent of the choice of σ. Given two measures $\bar{\mu}_0, \bar{\mu}_1 \in \mathscr{P}(\mathbb{R}^d)$ we denote by $\mathcal{CE}_{0,1}(\bar{\mu}_0, \bar{\mu}_1)$ the set of all sufficiently regular solutions (to be made precise in section 3.3) $(\mu_t, \boldsymbol{\nu}_t)_{t\in[0,1]}$ of the continuity equation (3.1.5).

Definition. *For $\bar{\mu}_0, \bar{\mu}_1 \in \mathscr{P}(\mathbb{R}^d)$ we define*

$$\mathcal{W}(\bar{\mu}_0, \bar{\mu}_1)^2 \ := \ \inf \left\{ \int_0^1 \mathcal{A}(\mu_t, \boldsymbol{\nu}_t)\mathrm{d}t \ : \quad (\mu, \boldsymbol{\nu}) \in \mathcal{CE}_{0,1}(\bar{\mu}_0, \bar{\mu}_1) \right\} \ .$$

It is unclear whether $\mathcal{W}$ coincides with $\widetilde{\mathcal{W}}$ defined in (3.1.3) in full generality. However, we will give a positive answer in the case of a sufficiently regular translation invariant jump kernel. (see Proposition 3.5.11). We can now state the first main result.

Theorem 3.1.2. *$\mathcal{W}$ defines a (pseudo-) metric on $\mathscr{P}(\mathbb{R}^d)$. The topology it induces is stronger than the topology of weak convergence. For each $\tau \in \mathscr{P}(\mathbb{R}^d)$ the set $\mathscr{P}_\tau := \{\mu \in \mathscr{P}(\mathbb{R}^d) \ : \ \mathcal{W}(\mu, \tau) < \infty\}$ equipped with the distance $\mathcal{W}$ is a complete geodesic space.*

As the Wasserstein distance, the distance $\mathcal{W}$ can take the value $+\infty$ on $\mathscr{P}(\mathbb{R}^d)$. We will study in more detail the "α-stable distance" $\mathcal{W}_\alpha$ associated to the jump kernel $J_\alpha(x, \mathrm{d}y) = c_\alpha \left|x - y\right|^{-\alpha-d} \mathrm{d}y$. In particular, we will show that $\mathcal{W}_\alpha(\mu_0, \mu_1) < \infty$ if μ_0 and μ_1 have finite moments of order α, see Proposition 3.4.14.

Gradient flow of the entropy

We now concentrate on a translation invariant jump kernel J. We assume that

$$J(x + z, A + z) = J(x, A) \quad \forall x, z \in \mathbb{R}^d,\ A \subset \mathbb{R}^d$$

and that m is Lebesgue measure. In this case we have $J(x, A) = \nu(A - x)$ for a symmetric Lévy measure ν on $\mathbb{R}^d$ and the underlying jump process is a Lévy process.

Let us give a short formal argument why the evolution equation (3.1.1) can be seen as the gradient flow of the relative entropy w.r.t. the distance $\mathcal{W}$ if we choose θ to be the logarithmic mean. In the classical setting many partial differential equations of the form

$$\partial_t \rho - \nabla \cdot \big(\rho \nabla f'(\rho)\big) \ = \ 0$$

can, at least formally, be seen as the gradient flow of the integral functional $\mathcal{F}$ given by $\mathcal{F}(\rho) = \int f(\rho)\mathrm{d}m$ w.r.t. the L^2-Wasserstein distance. By the same formal argument, in the new geometry determined by the distance $\widetilde{\mathcal{W}}$ via (3.1.3), (3.1.4) the gradient flow of the functional $\mathcal{F}$ should be given by the equation

$$\partial_t \rho - \overline{\nabla} \cdot \big(\hat{\rho} \overline{\nabla} f'(\rho)\big) \ = \ 0 \ .$$

If we now consider the relative entropy $\mathcal{H}$ we have $f'(r) = 1 + \log r$. Taking into account (3.1.2) we see that the corresponding gradient flow is given by

$$\partial_t \rho - \overline{\nabla} \cdot \big(\overline{\nabla} \rho\big) \ = \ 0 \ ,$$

which is a weak formulation of (3.1.1). In particular we see that the role of the logarithmic mean is to compensate the lack of a chain rule for the discrete gradient.

Our second main result is a rigorous characterisation of the evolution equation (3.1.1) as the gradient flow of the entropy in terms of the Evolution Variational Inequality (EVI). For $\mu \in \mathscr{P}(\mathbb{R}^d)$ we define the relative entropy by

$$\mathcal{H}(\mu) = \int \rho(x) \log \rho(x) \mathrm{d}x$$

if μ is absolutely continuous with density ρ and $(\rho \log \rho)_+$ is integrable. Otherwise, we set $\mathcal{H}(\mu) = +\infty$.

We formulate our result in terms of the semigroup $P_t = \exp(t\mathcal{L})$ generated by the operator $\mathcal{L}$. We assume that the equation (3.1.1) has a fundamental solution $\psi : (0, \infty) \times \mathbb{R}^d \to \mathbb{R}_+$. The semigroup P_t then acts on $\mathscr{P}(\mathbb{R}^d)$ via convolution:

$$P_t[\mu] \ := \ \mu * \psi_t \ .$$

Under certain further regularity assumptions on the kernel ψ (see Section 3.5 for a precise statement) we prove the following

Theorem 3.1.3. *The semigroup P generated by $\mathcal{L}$ is the gradient flow of the relative entropy in the sense that it satisfies the EVI: For any $\mu \in \mathscr{P}^* = \{\mu \in \mathscr{P}(\mathbb{R}^d) \ : \ \mathcal{H}(\mu) > -\infty\}$ and $\sigma \in \mathscr{P}_\mu \cap \mathscr{P}^*$ we have*

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}(P_t[\mu], \sigma)^2 + \mathcal{H}(P_t[\mu]) \ \leq \ \mathcal{H}(\sigma) \quad \forall t > 0 \ . \tag{3.1.6}$$

Moreover the entropy is convex along $\mathcal{W}$-geodesics. More precisely, let $\mu_0, \mu_1 \in \mathscr{P}(\mathbb{R}^d)$ such that $\mathcal{W}(\mu_0, \mu_1) < \infty$ and let $(\mu_t)_{t\in[0,1]}$ be a geodesic connecting μ_0 and μ_1. Then we have

$$\mathcal{H}(\mu_t) \ \leq \ (1 - t)\mathcal{H}(\mu_0) + t\mathcal{H}(\mu_1) \ .$$

The statement about geodesic convexity of the entropy is a direct consequence of the EVI in a general setting of metric spaces (see [DS08]). Convexity of the entropy along $\mathcal{W}$-geodesics can be seen as a non-local analogue of McCann's displacement convexity [McC97], which corresponds to convexity along geodesics of the L^2-Wasserstein distance. For the choice $\nu(\mathrm{d}y) = c_\alpha \left|y\right|^{-\alpha-d} \mathrm{d}y$ with $\alpha \in (0, 2)$ and a suitable constant c_α we obtain the following

Corollary 3.1.4. *The fractional heat equation*

$$\partial_t u + (-\Delta)^{\frac{\alpha}{2}} u \ = \ 0$$

is the gradient flow of the relative entropy w.r.t. the metric $\mathcal{W}_\alpha$ built from the jump kernel $J_\alpha(x, dy) = c_\alpha \left|y - x\right|^{-\alpha-d} \mathrm{d}y$.

Organisation of this chapter

In Section 3.2 we study the action functional $\mathcal{A}$ and establish various properties needed in the sequel. Section 3.3 is devoted to an analysis of the non-local continuity equation (3.1.5). In Section 3.4 we define the metric $\mathcal{W}$ and prove Theorem 3.1.2. Then we study in more detail the α-stable distance $\mathcal{W}_\alpha$. Finally, we focus on translation invariant jump kernels and present the proof of Theorem 3.1.3 in Section 3.5.

3.2 The action functional

In this section we introduce and study an action functional on pairs of measures. This functional will depend on a jump kernel. Throughout this section we fix a jump kernel J satisfying Assumption 3.1.1.

Let us first introduce some notation. We denote by $\mathscr{P}(\mathbb{R}^d)$ the space of Borel probability measures on $\mathbb{R}^d$ equipped with the topology of weak convergence. We let $G = \{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d | x \neq y\}$ and denote by $\mathcal{M}_{loc}(G)$ the space of signed Radon measures on the open set G equipped with the weak* topology in duality with continuous functions with compact support in G.

The definition of the action functional and later the metric will depend on the choice of a function $\theta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$. We will always require it to fulfil the following assumptions:

Assumption 3.2.1. *The function θ has the following properties:*

(A1) (Regularity): θ is continuous on $\mathbb{R}_+ \times \mathbb{R}_+$ and C^1 on $(0,\infty) \times (0,\infty)$;

(A2) (Symmetry): $\theta(s,t) = \theta(t,s)$ for $s,t \geq 0$;

(A3) (Positivity, normalisation): $\theta(s,t) > 0$ for $s,t > 0$ and $\theta(1,1) = 1$;

(A4) (Zero at the boundary): $\theta(0,t) = 0$ for all $t \geq 0$;

(A5) (Monotonicity): $\theta(r,t) \leq \theta(s,t)$ for all $0 \leq r \leq s$ and $t \geq 0$;

(A6) (Positive homogeneity): $\theta(\lambda s, \lambda t) = \lambda \theta(s,t)$ for $\lambda > 0$ and $s,t \geq 0$;

(A7) (Concavity): the function $\theta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is concave.

These assumptions are identical to Assumption 2.2.1 from Chapter 2. It is easy to check that they imply

$$\theta(s,t) \ \leq \ \frac{s+t}{2} \quad \forall s,t \geq 0 \ . \tag{3.2.1}$$

In view of applications to gradient flows of the entropy we will be mostly interested in a particular choice of θ, namely the logarithmic mean given by

$$\theta(s,t) \ = \ \int_0^1 s^\alpha t^{1-\alpha} \mathrm{d}\alpha \ = \ \frac{s-t}{\log s - \log t} \ , \tag{3.2.2}$$

the latter expression being valid for $s,t > 0$ with $s \neq t$. However, for future use we will allow for more generality in the choice of θ. Given a function $\rho : \mathbb{R}^d \to \mathbb{R}_+$ we will often write

$$\hat{\rho}(x,y) \ := \ \theta(\rho(x), \rho(y)) \ .$$

We introduce a function $\alpha : \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+ \cup \{\infty\}$, called the action density function, by setting

$$\alpha(w,s,t) := \begin{cases} \frac{w^2}{2\theta(s,t)} , & \theta(s,t) \neq 0 , \\ 0 , & \theta(s,t) = 0 \text{ and } w = 0 , \\ +\infty , & \theta(s,t) = 0 \text{ and } w \neq 0 . \end{cases}$$

We recall the following observation from Chapter 2 which will be useful in the sequel.

Lemma 3.2.2. *The function α is lower semicontinuous, convex and positively homogeneous, i.e.*

$$\alpha(\lambda w, \lambda s, \lambda t) \; = \; \lambda \alpha(w,s,t) \quad \forall w \in \mathbb{R} , \; s,t \geq 0 , \; \lambda \geq 0 .$$

Proof. This is easily checked using (A6),(A7) and the convexity of the function $(x,y) \mapsto \frac{x^2}{y}$ on $\mathbb{R} \times (0,\infty)$. □

We will now define an action functional on pairs of measures $(\mu, \boldsymbol{\nu})$ where $\mu \in \mathscr{P}(\mathbb{R}^d)$ and $\boldsymbol{\nu} \in \mathcal{M}_{loc}(G)$. To μ we associate two Radon measures in $\mathcal{M}_{loc}(G)$ by setting:

$$\mu^1(\mathrm{d}x, \mathrm{d}y) \; := \; J(x, \mathrm{d}y)\mu(\mathrm{d}x) \; , \; \mu^2(\mathrm{d}x, \mathrm{d}y) \; := \; J(y, \mathrm{d}x)\mu(\mathrm{d}y) \; . \tag{3.2.3}$$

We can always choose a measure $\sigma \in \mathcal{M}_{loc}(G)$ such that $\mu^i = \rho^i \sigma, \; i = 1,2$ and $\boldsymbol{\nu} = w\sigma$ are all absolutely continuous with respect to σ. For example take the sum of the total variations $\sigma := \left|\mu^1\right| + \left|\mu^2\right| + |\boldsymbol{\nu}|$. We can then define the *action functional* by

$$\mathcal{A}(\mu, \boldsymbol{\nu}) \; := \; \int \alpha\big(w, \rho^1, \rho^2\big) \mathrm{d}\sigma \; .$$

Note that this definition is independent of the choice of σ since α is positively homogeneous. Hence we can also write the action functional as

$$\mathcal{A}(\mu, \boldsymbol{\nu}) \; = \; \int \alpha \left(\frac{\mathrm{d}\lambda_1}{\mathrm{d}\,|\lambda|}, \frac{\mathrm{d}\lambda_2}{\mathrm{d}\,|\lambda|}, \frac{\mathrm{d}\lambda_3}{\mathrm{d}\,|\lambda|} \right) \mathrm{d}\,|\lambda| \; , \tag{3.2.4}$$

where λ is the vector valued measure given by $\lambda = (\boldsymbol{\nu}, \mu^1, \mu^2)$.

In the case where the measure μ is absolutely continuous w.r.t. m the next lemma shows that the action takes a more intuitive form. For this we denote by $Jm \in \mathcal{M}_{loc}(G)$ the measure given by $Jm(\mathrm{d}x, \mathrm{d}y) = J(x, \mathrm{d}y)m(\mathrm{d}x)$.

Lemma 3.2.3. *Let $\mu \in \mathscr{P}(\mathbb{R}^d)$ be absolutely continuous w.r.t. m with density ρ. Further, let $\boldsymbol{\nu} \in \mathcal{M}_{loc}(G)$ such that $\mathcal{A}(\mu, \boldsymbol{\nu}) < \infty$. Then there exist a function $w : G \to \mathbb{R}$ such that $\boldsymbol{\nu} = w\hat{\rho}Jm$ and we have*

$$\mathcal{A}(\mu, \boldsymbol{\nu}) \; = \; \frac{1}{2} \int |w(x,y)|^2 \, \hat{\rho}(x,y) J(x, \mathrm{d}y) m(\mathrm{d}x) \; . \tag{3.2.5}$$

Proof. Choose $\lambda \in \mathcal{M}_{loc}(G)$ such that $Jm = h\lambda$ and $\boldsymbol{\nu} = \widetilde{w}\lambda$ are both absolutely continuous w.r.t. λ. Note that $\mu^i = \rho^i Jm$, $i = 1,2$ with $\rho^1(x,y) = \rho(x)$ and $\rho^2(x,y) = \rho(y)$. Further, we denote by $\widetilde{\rho}^i$ the density of μ^i w.r.t λ. Now by definition,

$$\mathcal{A}(\mu,\boldsymbol{\nu}) \;=\; \int \alpha\big(\widetilde{w},\widetilde{\rho}^1,\widetilde{\rho}^2\big)\mathrm{d}\lambda \;<\; \infty\;. \tag{3.2.6}$$

Let $A \subset G$ such that $\int_A \theta(\rho^1,\rho^2)\mathrm{d}Jm = 0$. From the homogeneity of θ we conclude

$$0 \;=\; \int_A \theta(\rho^1,\rho^2)\mathrm{d}Jm \;=\; \int_A \theta(\widetilde{\rho}^1,\widetilde{\rho}^2)\mathrm{d}\lambda\;,$$

i.e. $\theta(\widetilde{\rho}^1,\widetilde{\rho}^2) = 0$ λ-a.e. on A. Now the finiteness of the integral in (3.2.6) implies that $\widetilde{w} = 0$ λ-a.e. on A. In other words $\boldsymbol{\nu}(A) = 0$ and hence $\boldsymbol{\nu}$ is absolutely continuous w.r.t. the measure $\hat{\rho}Jm$. Formula (3.2.5) now follows immediately from the homogeneity of α. □

We will now establish several important properties of the action functional.

Lemma 3.2.4 (Lower semicontinuity of the action)**.** *$\mathcal{A}$ is lower semicontinuous w.r.t. weak convergence of measures. More precisely, assume that $\mu_n \rightharpoonup \mu$ weakly in $\mathscr{P}(\mathbb{R}^d)$ and $\boldsymbol{\nu}_n \rightharpoonup^* \boldsymbol{\nu}$ weakly* in $\mathcal{M}_{loc}(G)$. Then*

$$\mathcal{A}(\mu,\boldsymbol{\nu}) \;\le\; \liminf_n \mathcal{A}(\mu_n,\boldsymbol{\nu}_n)\;.$$

Proof. Note that by Assumption 3.1.1 the weak convergence of μ_n to μ implies the weak* convergence of μ_n^i to μ^i in $\mathcal{M}_+(G)$ for $i = 1,2$. Now the claim follows immediately from the representation (3.2.4) and a general result on integral functionals, Proposition 3.2.5. □

Proposition 3.2.5 ([But89, Thm. 3.4.3])**.** *Let Ω be a locally compact Polish space and let $f : \Omega \times \mathbb{R}^n \to [0,+\infty]$ be a lower semicontinuous function such that $f(\omega,\cdot)$ is convex and positively 1-homogeneous for every $\omega \in \Omega$. Then the functional*

$$F(\lambda) \;=\; \int_\Omega f\left(\omega,\frac{\mathrm{d}\lambda}{\mathrm{d}\left|\lambda\right|}(\omega)\right)\left|\lambda\right|(\mathrm{d}\omega)$$

is sequentially weak lower semicontinuous on the space of vector valued signed Radon measures $\mathcal{M}_{loc}(\Omega,\mathbb{R}^n)$.*

The next estimate will be very useful in the sequel.

Lemma 3.2.6. *i) There exists a constant $C > 0$ such that for all $\mu \in \mathscr{P}(\mathbb{R}^d)$ and $\boldsymbol{\nu} \in \mathcal{M}_{loc}(G)$ we have:*

$$\int_G \big(1 \wedge |x-y|\,\big)\left|\boldsymbol{\nu}\right|(\mathrm{d}x,\mathrm{d}y) \;\le\; C\sqrt{\mathcal{A}(\mu,\boldsymbol{\nu})}\;.$$

ii) For each compact set $K \subset G$ there exists a constant $C(K) > 0$ such that for all $\mu \in \mathscr{P}(\mathbb{R}^d)$ and $\boldsymbol{\nu} \in \mathcal{M}_{loc}(G)$ we have:

$$|\boldsymbol{\nu}|(K) \ \le\ C(K)\sqrt{\mathcal{A}(\mu,\boldsymbol{\nu})}\ .$$

Proof. To prove i) let us define the measure $\lambda = |\mu^1| + |\mu^2| + |\boldsymbol{\nu}|$ and write $\mu^i = \rho^i\lambda$, $\boldsymbol{\nu} = w\lambda$. We can assume that $\mathcal{A}(\mu,\boldsymbol{\nu}) < \infty$ as otherwise there is nothing to prove. This implies that the set $A = \{(x,y) \mid \alpha(w,\rho^1,\rho^2) = \infty\}$ has zero measure with respect to λ. We can now estimate:

$$\begin{aligned}
&\int_G \big(1\wedge|x-y|\big)\,|\boldsymbol{\nu}|(\mathrm{d}x,\mathrm{d}y)\\
\le\ &\int_G \big(1\wedge|x-y|\big)\,|w|\,\mathrm{d}\lambda\\
=\ &\int_{A^c} \big(1\wedge|x-y|\big)\sqrt{2\theta(\rho^1,\rho^2)}\sqrt{\alpha(w,\rho^1,\rho^2)}\mathrm{d}\lambda\\
\le\ &\left(\int_G \big(1\wedge|x-y|^2\big)2\theta(\rho^1,\rho^2)\mathrm{d}\lambda\right)^{\frac12}\left(\int_G \alpha(w,\rho^1,\rho^2)\mathrm{d}\lambda\right)^{\frac12}\\
\le\ &C\sqrt{\mathcal{A}(\mu,\boldsymbol{\nu})}\ .
\end{aligned}$$

The last inequality follows, since by the estimate (3.2.1) and Assumption 3.1.1 we have:

$$\begin{aligned}
\int_G \big(1\wedge|x-y|^2\big)\theta(\rho^1,\rho^2)\mathrm{d}\lambda\ &\le\ \int_G \big(1\wedge|x-y|^2\big)\frac{1}{2}(\rho^1+\rho^2)\mathrm{d}\lambda\\
&=\ \int_G \big(1\wedge|x-y|^2\big)J(x,\mathrm{d}y)\mu(\mathrm{d}x)\\
&\le\ \sup_x \int (1\wedge|x-y|^2)J(x,\mathrm{d}y)\ <\ \infty\ .
\end{aligned}$$

To prove ii) we note that by a similar argument

$$\begin{aligned}
|\boldsymbol{\nu}|(K)\ &\le\ \left(\int_K 2\theta(\rho^1,\rho^2)\mathrm{d}\lambda\right)^{\frac12}\sqrt{\mathcal{A}(\mu,\boldsymbol{\nu})}\\
&\le\ \left(\int_{K\cup K^t} J(x,\mathrm{d}y)\mu(\mathrm{d}x)\right)^{\frac12}\sqrt{\mathcal{A}(\mu,\boldsymbol{\nu})}\ ,
\end{aligned}$$

where $K^t := \{(y,x)|(x,y)\in K\}$. The first factor in second line of the previous estimate is bounded independently of μ by Assumption 3.1.1. □

Lemma 3.2.7 (Convexity of the action)**.** *Let* $\mu^j \in \mathscr{P}(\mathbb{R}^d)$ *and* $\boldsymbol{\nu}^j \in \mathcal{M}_{loc}(G)$ *for* $j = 0, 1$. *For* $\tau \in [0, 1]$ *set* $\mu^\tau = \tau\mu^1 + (1 - \tau)\mu^0$ *and* $\boldsymbol{\nu}^\tau = \tau\boldsymbol{\nu}^1 + (1 - \tau)\boldsymbol{\nu}^0$. *Then we have :*

$$\mathcal{A}(\mu^\tau, \boldsymbol{\nu}^\tau) \;\le\; \tau\mathcal{A}(\mu^1, \boldsymbol{\nu}^1) + (1 - \tau)\mathcal{A}(\mu^0, \boldsymbol{\nu}^0) \;.$$

Proof. Let us fix a reference measure $\lambda \in \mathcal{M}_{loc}(G)$ such that $\mu^{j,i}, \boldsymbol{\nu}^j$ for $j = 0, 1$ and $i = 1, 2$ are all absolutely continuous w.r.t. λ and write $\mu^{j,i} = \rho^{j,i}\lambda$ and $\boldsymbol{\nu}^j = w^j\lambda$. Note that $\mu^{\tau,i} = \rho^{\tau,i}\lambda$ with $\rho^{\tau,i} = \tau\rho^{1,i} + (1 - \tau)\rho^{0,i}$ and that $\boldsymbol{\nu}^\tau = w^\tau\lambda$ with $w^\tau = \tau w^1 + (1 - \tau)w^0$. From the convexity of the action density function α we obtain:

$$\begin{aligned}
\mathcal{A}(\mu^\tau, \boldsymbol{\nu}^\tau) &= \int \alpha(w^\tau, \rho^{\tau,1}, \rho^{\tau,2})\mathrm{d}\lambda \\
&\le \tau \int \alpha(w^1, \rho^{1,1}, \rho^{1,2})\mathrm{d}\lambda + (1 - \tau) \int \alpha(w^0, \rho^{0,1}, \rho^{0,2})\mathrm{d}\lambda \\
&= \tau\mathcal{A}(\mu^1, \boldsymbol{\nu}^1) + (1 - \tau)\mathcal{A}(\mu^0, \boldsymbol{\nu}^0) \;.
\end{aligned}$$

□

We will now show that the action functional enjoys a monotonicity property under convolution if we assume that the jump kernel is translation invariant in the sense that

$$J(x - z, A - z) \;=\; J(x, A) \qquad \forall x, z \in \mathbb{R}^d, A \in \mathcal{B}(\mathbb{R}^d) \;. \tag{3.2.7}$$

For the rest of this section we also assume that m is Lebesgue measure. We first need to fix a way of convoluting measures on $\mathbb{R}^d$ and on G in a consistent manner. Let k be a convolution kernel, i.e. a function $k : \mathbb{R}^d \to \mathbb{R}_+$ satisfying $\int k(z)\mathrm{d}z = 1$. Given a measure $\mu \in \mathscr{P}(\mathbb{R}^d)$, its convolution is defined as usual by

$$(\mu * k)(A) \;:=\; \int k(z)\mu(A - z)\mathrm{d}z \qquad \forall A \in \mathcal{B}(\mathbb{R}^d) \;.$$

On the other hand given a measure $\boldsymbol{\nu} \in \mathcal{M}_{loc}(G)$ we define $\boldsymbol{\nu} * k \in \mathcal{M}_{loc}(G)$ by setting for all Borel measurable sets $B \subset G$

$$(\boldsymbol{\nu} * k)(B) \;:=\; \int k(z)\boldsymbol{\nu}(B - \begin{pmatrix} z \\ z \end{pmatrix})\mathrm{d}z \;. \tag{3.2.8}$$

Note that this implies in particular that for every bounded function $f : G \to \mathbb{R}$ with compact support in G we have:

$$\int f(x, y)(\boldsymbol{\nu} * k)(\mathrm{d}x, \mathrm{d}y) \;=\; \int\int k(z)f(x + z, y + z)\boldsymbol{\nu}(\mathrm{d}x, \mathrm{d}y)\mathrm{d}z \;.$$

We now have the following monotonicity property under convolution.

Proposition 3.2.8. *Assume that J satisfies* (3.2.7) *and let k be a convolution kernel. Then for every $\mu \in \mathscr{P}(\mathbb{R}^d), \boldsymbol{\nu} \in \mathcal{M}_{loc}(G)$ we have*

$$\mathcal{A}(\mu * k, \boldsymbol{\nu} * k) \ \leq \ \mathcal{A}(\mu, \boldsymbol{\nu}) \ . \tag{3.2.9}$$

Proof. We can assume without restriction that $\mathcal{A}(\mu, \boldsymbol{\nu})$ is finite as otherwise there is nothing to prove. Let us introduce the maps $\tau_z : x \mapsto x + z$ for $z \in \mathbb{R}^d$ and let us denote by $\mu_z, \boldsymbol{\nu}_z$ the push forward $(\tau_z)_*\mu = \mu(\cdot - z)$, resp. $(\tau_z \times \tau_z)_*\boldsymbol{\nu} = \boldsymbol{\nu}(\cdot - \binom{z}{z})$. Using the convexity of the action functional, Lemma 3.2.7, together with its lower semicontinuity, Lemma 3.2.4, we see that

$$\mathcal{A}(\mu * k, \boldsymbol{\nu} * k) \ \leq \ \int \mathcal{A}(\mu_z, \boldsymbol{\nu}_z) k(z) \mathrm{d}z \ .$$

Thus the proof is complete if we show that $\mathcal{A}(\mu_z, \boldsymbol{\nu}_z) = \mathcal{A}(\mu, \boldsymbol{\nu})$ for all $z \in \mathbb{R}^d$. To this end, recall the definition (3.2.3). Using the the invariance property (3.2.7) it is immediate to check that $\mu^i_z = (\tau_z \times \tau_z)_*\mu^i$ for $i = 1, 2$. Now choose $\lambda \in \mathcal{M}_{loc}(G)$ with $\mu^i = \rho^i\lambda$ and $\boldsymbol{\nu} = w\lambda$. Then for all $z \in \mathbb{R}^d$ we have $(\mu_z)^i = (\mu^i)_z = \rho^i(\cdot - \binom{z}{z})\lambda_z$ and $\boldsymbol{\nu}_z = w(\cdot - \binom{z}{z})\lambda_z$. Hence we finally obtain

$$\begin{aligned}\mathcal{A}(\mu_z, \boldsymbol{\nu}_z) \ &= \ \int \alpha\left(w(\cdot - \begin{pmatrix} z \\ z \end{pmatrix}), \rho^1(\cdot - \begin{pmatrix} z \\ z \end{pmatrix}), \rho^2(\cdot - \begin{pmatrix} z \\ z \end{pmatrix})\right) \mathrm{d}\lambda_z \\ &= \ \int \alpha(w, \rho^1, \rho^2) \mathrm{d}\lambda \ = \ \mathcal{A}(\mu, \boldsymbol{\nu}) \ .\end{aligned}$$

□

3.3 A non-local continuity equation

In this section we will consider the continuity equation

$$\partial_t \mu_t + \overline{\nabla} \cdot \boldsymbol{\nu}_t \ = \ 0 \qquad \text{on } (0, T) \times \mathbb{R}^d \ . \tag{3.3.1}$$

Here $(\mu_t)_{t \in [0,T]}$ and $(\boldsymbol{\nu}_t)_{t \in [0,T]}$ are Borel families of measures in $\mathscr{P}(\mathbb{R}^d)$ and $\mathcal{M}_{loc}(G)$ respectively such that

$$\int_0^T \int \big(1 \wedge |x - y|\big) \, |\boldsymbol{\nu}_t| \, (\mathrm{d}x, \mathrm{d}y) \mathrm{d}t \ < \ \infty \ . \tag{3.3.2}$$

We suppose that (3.3.1) holds in the sense of distributions. More precisely, we require that for all $\varphi \in C_c^\infty((0, T) \times \mathbb{R}^d)$:

$$\int_0^T \int \partial_t \varphi_t(x) \mu_t(\mathrm{d}x) \mathrm{d}t + \frac{1}{2} \int_0^T \int \overline{\nabla} \varphi_t(x, y) \boldsymbol{\nu}_t(\mathrm{d}x, \mathrm{d}y) \mathrm{d}t \ = \ 0 \ . \tag{3.3.3}$$

Recall that for a function $\varphi : \mathbb{R}^d \to \mathbb{R}$ we denote by $\overline{\nabla}\varphi(x, y) = \varphi(y) - \varphi(x)$ the discrete gradient. Note that (3.3.2) is a natural integrability assumption one should make to ensure that the second term in (3.3.3) is well-defined. The following is an adaptation of [AGS08, Lemma 8.1.2].

Lemma 3.3.1. *Let $(\mu_t)_{t\in[0,T]}$ and $(\boldsymbol{\nu}_t)_{t\in[0,T]}$ be Borel families of measures in $\mathscr{P}(\mathbb{R}^d)$ and $\mathcal{M}_{loc}(G)$ satisfying* (3.3.1) *and* (3.3.2)*. Then there exists a weakly continuous curve $(\widetilde{\mu}_t)_{t\in[0,T]}$ such that $\widetilde{\mu}_t = \mu_t$ for a.e. $t \in [0,T]$. Moreover, for every test function $\varphi \in C_c^\infty([0,T]\times\mathbb{R}^d)$ and all $0 \le t_0 \le t_1 \le T$ we have :*

$$\int \varphi_{t_1} \mathrm{d}\widetilde{\mu}_{t_1} - \int \varphi_{t_0} \mathrm{d}\widetilde{\mu}_{t_0} \;=\; \int_{t_0}^{t_1} \int \partial_t \varphi \mathrm{d}\mu_t \mathrm{d}t + \frac{1}{2}\int_{t_0}^{t_1} \int \overline{\nabla}\varphi \mathrm{d}\boldsymbol{\nu}_t \mathrm{d}t \;. \tag{3.3.4}$$

Proof. Let us set

$$V(t) \;:=\; \int \big(1 \wedge |x-y|\big)\, |\boldsymbol{\nu}_t|\,(\mathrm{d}x,\mathrm{d}y) \;.$$

By assumption $t \mapsto V(t)$ belongs to $L^1(0,T)$. Fix $\xi \in C_c^\infty(\mathbb{R}^d)$. We claim that the map $t \mapsto \mu_t(\xi) = \int \xi \mathrm{d}\mu_t$ belongs to $W^{1,1}(0,T)$. Indeed, using test functions of the form $\varphi(t,x) = \eta(t)\xi(x)$ with $\eta \in C_c^\infty(0,T)$, equation (3.3.3) shows that the distributional derivative of $\mu_t(\xi)$ is given by

$$\dot{\mu}_t(\xi) \;=\; \frac{1}{2}\int \overline{\nabla}\xi \mathrm{d}\boldsymbol{\nu}_t$$

for a.e. $t \in (0,T)$ and we can estimate

$$|\dot{\mu}_t(\xi)| \;\le\; \frac{1}{2}\int \big|\overline{\nabla}\xi\big|\, \mathrm{d}\,|\boldsymbol{\nu}_t| \;\le\; \frac{1}{2}\|\xi\|_{C^1} V(t) \;. \tag{3.3.5}$$

Based on (3.3.5) we can argue as in [AGS08, Lemma 8.1.2] to obtain existence of a weakly continuous representative $t \mapsto \widetilde{\mu}_t$.

To prove (3.3.4) fix $\varphi \in C_c^\infty([0,T]\times\mathbb{R}^d)$ and choose $\eta_\varepsilon \in C_c^\infty(t_0,t_1)$ such that

$$0 \le \eta_\varepsilon \le 1 \;, \quad \lim_{\varepsilon\to 0} \eta_\varepsilon(t) = 1_{(t_0,t_1)}(t) \;\forall t \in [0,T] \;, \quad \lim_{\varepsilon\to 0} \eta'_\varepsilon = \delta_{t_0} - \delta_{t_1} \;.$$

Now equation (3.3.3) implies

$$-\int_0^T \eta'_\varepsilon \int \varphi \mathrm{d}\widetilde{\mu}_t \mathrm{d}t \;=\; \int_0^T \eta_\varepsilon \int \partial_t\varphi \mathrm{d}\mu_t \mathrm{d}t + \frac{1}{2}\int_0^T \eta_\varepsilon \int \overline{\nabla}\varphi \mathrm{d}\boldsymbol{\nu}_t \mathrm{d}t \;.$$

Thanks to the continuity of $t \mapsto \widetilde{\mu}_t$ we can pass to limit as $\varepsilon \to 0$ and obtain (3.3.4). □

In view of the previous lemma it makes sense to define solutions to the continuity equation in the following way.

Definition 3.3.2. *We denote by $\mathcal{CE}_T(\bar{\mu}_0,\bar{\mu}_1)$ the set of all pairs $(\mu,\boldsymbol{\nu})$ satisfying the following conditions:*

$$\begin{cases} (i) & \mu : [0,T] \to \mathscr{P}(\mathbb{R}^d) \text{ is weakly continuous} \;;\\ (ii) & \mu_0 = \bar{\mu}_0 \;, \quad \mu_T = \bar{\mu}_1 \;;\\ (iii) & (\boldsymbol{\nu}_t)_{t\in[0,T]} \text{ is a Borel family of measures in } \mathcal{M}_{loc}(G) \;;\\ (iv) & \int_0^T \int \big(1\wedge|x-y|\big)\,|\boldsymbol{\nu}_t|\,(\mathrm{d}x,\mathrm{d}y)\mathrm{d}t \;<\; \infty \;;\\ (v) & \text{We have in the sense of distributions:}\\ & \partial_t\mu_t + \overline{\nabla}\cdot\boldsymbol{\nu}_t \;=\; 0 \;. \end{cases} \tag{3.3.6}$$

The following result will allow us to extract subsequential limits from sequences of solutions to the continuity equation which have bounded action.

Proposition 3.3.3 (Compactness of solutions to the CE)**.** *Let $(\mu^n, \boldsymbol{\nu}^n)$ be a sequence in $\mathcal{CE}_T(\bar\mu_0, \bar\mu_1)$ such that*

$$A \ := \ \sup_n \int_0^T \mathcal{A}(\mu_t^n, \boldsymbol{\nu}_t^n)\mathrm{d}t \ < \ \infty \ . \tag{3.3.7}$$

Then there exists a couple $(\mu, \boldsymbol{\nu}) \in \mathcal{CE}_T(\bar\mu_0, \bar\mu_1)$ such that up to extraction of a subsequence

$$\begin{aligned} \mu_t^n &\rightharpoonup \mu_t \quad \text{weakly in } \mathscr{P}(\mathbb{R}^d) \text{ for all } t \in [0,T] \ , \\ \boldsymbol{\nu}^n &\rightharpoonup^* \boldsymbol{\nu} \quad \text{weakly* in } \mathcal{M}_{log}(G \times (0,T)) \ . \end{aligned}$$

Moreover along this subsequence we have :

$$\int_0^T \mathcal{A}(\mu_t, \boldsymbol{\nu}_t)\mathrm{d}t \ \le \ \liminf_n \int_0^T \mathcal{A}(\mu_t^n, \boldsymbol{\nu}_t^n)\mathrm{d}t \ .$$

Proof. For each n define the measure $\boldsymbol{\nu}^n := \int_0^T \boldsymbol{\nu}_t^n \mathrm{d}t \in \mathcal{M}_{loc}(G \times (0,T))$. From Lemma 3.2.6 and (3.3.7) we infer immediately that

$$\sup_n \int_0^T \int \big(1 \wedge |x-y|\big) \left|\boldsymbol{\nu}^n\right| (\mathrm{d}x, \mathrm{d}y)\mathrm{d}t \ < \ \infty \ . \tag{3.3.8}$$

Moreover, arguing exactly as in Lemma 3.2.6, we obtain that for every compact set $K \subset G$ and $B \subset [0,T]$ we have

$$\begin{aligned} \sup_n \left|\boldsymbol{\nu}^n\right| (K \times B) \ &\le \ \sup_n \int_B \left|\boldsymbol{\nu}_t^n\right| (K)\mathrm{d}t & (3.3.9) \\ &\le \ \sqrt{A}C(K)\sqrt{\mathrm{Leb}(B)} \ . & (3.3.10) \end{aligned}$$

In particular, $\boldsymbol{\nu}^n$ has total variation uniformly bounded on every compact subset of $G \times [0,T]$. Hence, we can extract a subsequence (still indexed by n) such that $\boldsymbol{\nu}^n \rightharpoonup^* \boldsymbol{\nu}$ in $\mathcal{M}_{loc}(G \times [0,T])$. The estimate (3.3.9) also shows that $\boldsymbol{\nu}$ can be desintegrated w.r.t. Lebesgue measure on $[0,T]$ and we can write $\boldsymbol{\nu} = \int_0^T \boldsymbol{\nu}_t \mathrm{d}t$ for a Borel family $(\boldsymbol{\nu}_t)$ still satisfying (3.3.2).

Let $0 \le t_0 \le t_1 \le T$ and $\xi \in C_c^\infty(\mathbb{R}^d)$. We claim that

$$\int_{t_0}^{t_1} \int \overline{\nabla}\xi \mathrm{d}\boldsymbol{\nu}_t^n \mathrm{d}t \ \overset{n\to\infty}{\longrightarrow} \ \int_{t_0}^{t_1} \int \overline{\nabla}\xi \mathrm{d}\boldsymbol{\nu}_t \mathrm{d}t \ . \tag{3.3.11}$$

Note, that $\mathbf{1}_{(t_0,t_1)}\overline{\nabla}\xi$ is not continuous and not compactly supported in $G \times [0,T]$. In order to prove (3.3.11), we argue by approximation. Let $F \subset \mathbb{R}^d$ be a compact set supporting ξ. For $R > 0$ large we define the sets $A_R := [t_0, t_0 + \frac{1}{R}] \cup [t_1 - \frac{1}{R}, t_1]$ and $D_R := \{(x,y) \in G : |x-y| > R^{-1}\}$. Moreover, we define the sets

$$M_R \ := \ \Big(D_R \cap (F \times F)\Big) \cup (B_R^c \times F) \cup (F \times B_R^c) \ ,$$

where B_R denotes the ball of radius R in $\mathbb{R}^d$. Let $\varphi_R : G \times [0,T] \to [0,1]$ be a continuous function supported in $K \times [t_0, t_1]$ for a compact set $K \subset G$ such that the set, where we have $\varphi < 1$ is contained in the set

$$S_R \ := \ \Big(\big((B_R \times B_R)^c \cup D_R\big) \times [0,T]\Big) \cup (G \times A_R) \ .$$

The convergence (3.3.11) holds if we replace $\overline{\nabla}\xi$ by the continuous and compactly supported function $\varphi_R \cdot \overline{\nabla}\xi$. Thus, it remains to show that

$$\sup_n \left| \int_{t_0}^{t_1} \int (1 - \varphi_R) \overline{\nabla}\xi \mathrm{d}\boldsymbol{\nu}_t^n \mathrm{d}t \right| \ \longrightarrow \ 0 \ ,$$

as $R \to \infty$. Argueing as in Lemma 3.2.6, we estimate

$$\begin{aligned} \left| \int_{t_0}^{t_1} \int (1 - \varphi_R) \overline{\nabla}\xi \mathrm{d}\boldsymbol{\nu}_t^n \mathrm{d}t \right| \ &\leq \ \|\xi\|_{C^1} \int_{S_R} (1 \wedge |x-y|) \mathrm{d} \left|\boldsymbol{\nu}_t^n\right| \mathrm{d}t \\ &\leq \ \|\xi\|_{C^1} \sqrt{A} \left(\int_{S_R} (1 \wedge |x-y|^2) J(x, \mathrm{d}y) \mu_t^n(\mathrm{d}x) \mathrm{d}t \right)^{\frac{1}{2}} \ . \end{aligned}$$

From Assumption 3.1.1 we deduce that the second factor in the last line goes to zero uniformly in n as $R \to \infty$.

Combining now the convergence (3.3.11) with (3.3.4) for the choice $\varphi(t,x) = \xi(x)$ and $t_0 = 0, t_1 = t$ we infer that μ_t^n converges weakly to some $\mu_t \in \mathscr{P}(\mathbb{R}^d)$ for every $t \in [0,T]$. It is easily checked that the couple $(\mu, \boldsymbol{\nu})$ belongs to $\mathcal{CE}_T(\bar{\mu}_0, \bar{\mu}_1)$. As in Lemma 3.2.4 the lower semicontinuity now follows from Proposition 3.2.5 by considering $\int_0^T \mathcal{A}(\mu_t, \boldsymbol{\nu}_t)\mathrm{d}t$ as an integral functional on the space $\mathcal{M}_{loc}(G \times [0,T])$. □

3.4 A non-local transport distance

We are now ready to give the definition of the distance $\mathcal{W}$. We will then establish various properties, in particular existence of geodesics. Moreover, we will characterise absolutely continuous curves in the metric space $(\mathscr{P}, \mathcal{W})$. Finally, we will investigate the distance $\mathcal{W}_\alpha$ associated to the α-stable jump kernel J_α in more detail.

3.4.1 Construction and properties of the distance $\mathcal{W}$

We fix a jump kernel J satisfying Assumption 3.1.1 and make the following

Definition 3.4.1. *For $\bar\mu_0, \bar\mu_1 \in \mathscr{P}(\mathbb{R}^d)$ we define*

$$\mathcal{W}(\bar\mu_0,\bar\mu_1)^2 \ := \ \inf\left\{ \int_0^1 \mathcal{A}(\mu_t,\boldsymbol{\nu}_t)\mathrm{d}t \ : \quad (\mu,\boldsymbol{\nu}) \in \mathcal{CE}_1(\bar\mu_0,\bar\mu_1)\right\} \ . \tag{3.4.1}$$

Let us first give an equivalent characterisation of the infimum in (3.4.1).

Lemma 3.4.2. *For any $T > 0$ and $\bar\mu_0, \bar\mu_1 \in \mathscr{P}(\mathbb{R}^d)$ we have :*

$$\mathcal{W}(\bar\mu_0,\bar\mu_1) \ = \ \inf\left\{ \int_0^T \sqrt{\mathcal{A}(\mu_t,\boldsymbol{\nu}_t)}\mathrm{d}t \ : \quad (\mu,\boldsymbol{\nu}) \in \mathcal{CE}_T(\bar\mu_0,\bar\mu_1)\right\} \ . \tag{3.4.2}$$

Proof. This follows from a standard reparametrisation argument. See [AGS08, Lem. 1.1.4] or [DNS09, Thm. 5.4] for details in similar situations. □

The next result shows that the infimum in the definition above is in fact a minimum.

Proposition 3.4.3. *Let $\bar\mu_0, \bar\mu_1 \in \mathscr{P}(\mathbb{R}^d)$ be such that $W := \mathcal{W}(\bar\mu_0,\bar\mu_1)$ is finite. Then the infimum in (3.4.1) is attained by a curve $(\mu,\boldsymbol{\nu}) \in \mathcal{CE}_1(\bar\mu_0,\bar\mu_1)$ satisfying $\mathcal{A}(\mu_t,\boldsymbol{\nu}_t) = W^2$ for a.e. $t \in [0,1]$.*

Proof. Existence of a minimising curve $(\mu,\boldsymbol{\nu}) \in \mathcal{CE}_1(\bar\mu_0,\bar\mu_1)$ follows immediately by the direct method taking into account Proposition 3.3.3. Invoking Lemma 3.4.2 and Jensen's inequality, we see that this curve satisfies

$$\int_0^1 \sqrt{\mathcal{A}(\mu_t,\boldsymbol{\nu}_t)}\mathrm{d}t \ \geq \ W \ = \ \left(\int_0^1 \mathcal{A}(\mu_t,\boldsymbol{\nu}_t)\mathrm{d}t\right)^{\frac12} \ \geq \ \int_0^1 \sqrt{\mathcal{A}(\mu_t,\boldsymbol{\nu}_t)}\mathrm{d}t \ .$$

Hence we must have $\mathcal{A}(\mu_t,\boldsymbol{\nu}_t) = W^2$ for a.e. $t \in [0,T]$. □

We now prove Theorem 3.1.2, the first main result announced in the introduction, which we recall here for convenience.

Theorem 3.4.4. *$\mathcal{W}$ defines a (pseudo-) metric on $\mathscr{P}(\mathbb{R}^d)$. The topology it induces is stronger than the weak topology and bounded sets w.r.t. $\mathcal{W}$ are weakly compact. Moreover, the map $(\mu_0,\mu_1) \mapsto \mathcal{W}(\mu_0,\mu_1)$ is lower semicontinuous w.r.t. weak convergence. For each $\tau \in \mathscr{P}(\mathbb{R}^d)$ the set $\mathscr{P}_\tau := \{\mu \in \mathscr{P}(\mathbb{R}^d) \ : \ \mathcal{W}(\mu,\tau) < \infty\}$ equipped with the distance $\mathcal{W}$ is a complete geodesic space.*

Proof. Symmetry of $\mathcal{W}$ is obvious from the fact that $\alpha(w,\cdot,\cdot) = \alpha(-w,\cdot,\cdot)$. Equation (3.3.4) from Lemma 3.3.1 shows that two curves in $\mathcal{CE}_1$ can be concatenated to obtain a curve in $\mathcal{CE}_2$. Hence the triangle inequality follows easily using Lemma 3.4.2. To see that $\mathcal{W}(\bar\mu_0,\bar\mu_1) > 0$ whenever $\bar\mu_0 \neq \bar\mu_1$, assume that $\mathcal{W}(\bar\mu_0,\bar\mu_1) = 0$ and choose a minimising curve $(\mu, \boldsymbol{\nu}) \in \mathcal{CE}_1(\bar\mu_0, \bar\mu_1)$. Then we must have $\mathcal{A}(\mu_t, \boldsymbol{\nu}_t) = 0$ and hence $\boldsymbol{\nu}_t = 0$ for a.e. $t \in (0,1)$. From the continuity equation in the form (3.3.4) we infer $\bar\mu_0 = \bar\mu_1$.

Let us now show that the topology induced by $\mathcal{W}$ is stronger than the weak one. Let $\mu_n, \mu \in \mathscr{P}(\mathbb{R}^d)$ with $\mathcal{W}(\mu_n, \mu) \to 0$ and choose minimising curves $(\mu^n, \boldsymbol{\nu}^n) \in \mathcal{CE}_1(\mu_n, \mu)$. Fix a function $\varphi : \mathbb{R}^d \to \mathbb{R}$ bounded in C^1. Using the continuity equation in the form (3.3.4) and Lemma 3.2.6 we estimate:

$$
\begin{aligned}
\left|\int \varphi \mathrm{d}\mu_n - \int \varphi \mathrm{d}\mu\right| &= \frac{1}{2}\left|\int_0^1 \int \overline{\nabla}\varphi \mathrm{d}\boldsymbol{\nu}_t^n \mathrm{d}t\right| \\
&\leq \|\varphi\|_{C^1} \int_0^1 \int \big(1 \wedge |x-y|\big)\, |\boldsymbol{\nu}_t^n|\,(\mathrm{d}x,\mathrm{d}y)\mathrm{d}t \\
&\leq \|\varphi\|_{C^1} C \int_0^1 \sqrt{\mathcal{A}(\mu_t^n, \boldsymbol{\nu}_t^n)}\mathrm{d}t \;=\; \|\varphi\|_{C^1} C \cdot \mathcal{W}(\mu_n, \mu)\;.
\end{aligned}
$$

This implies $\mu_n \rightharpoonup \mu$ weakly.

The compactness assertion and lower semicontinuity of $\mathcal{W}$ follow immediately from Proposition 3.3.3. Let us now fix $\tau \in \mathscr{P}(\mathbb{R}^d)$ and let $\bar\mu_0, \bar\mu_1 \in \mathscr{P}_\tau$. By the triangle inequality we have $\mathcal{W}(\bar\mu_0, \bar\mu_1) < \infty$ and hence Proposition 3.4.3 yields existence of a minimising curve $(\mu, \boldsymbol{\nu}) \in \mathcal{CE}_1(\bar\mu_0, \bar\mu_1)$. The curve $t \mapsto \mu_t$ is then a constant speed geodesic in $\mathscr{P}_\tau$ since it satisfies

$$
\mathcal{W}(\mu_s, \mu_t) \;=\; \int_s^t \sqrt{\mathcal{A}(\mu_r, \boldsymbol{\nu}_r)}\mathrm{d}r \;=\; (t-s)\mathcal{W}(\mu_0, \mu_1) \quad \forall 0 \leq s \leq t \leq 1\;.
$$

To show completeness let $(\mu^n)_n$ be a Cauchy sequence in $\mathscr{P}_\tau$. In particular the sequence is bounded w.r.t. $\mathcal{W}$ and we can find a subsequence (still indexed by n) and μ^∞ such that $\mu^n \rightharpoonup \mu^\infty$ weakly. Invoking lower semicontinuity of $\mathcal{W}$ and the Cauchy condition we infer $\mathcal{W}(\mu^n, \mu^\infty) \to 0$ as $n \to \infty$ and $\mu^\infty \in \mathscr{P}_\tau$. □

It is yet unclear when precisely the distance $\mathcal{W}$ is finite. This seems to depend heavily on the structure of the jump kernel J. In section 3.4.3 we will study in more detail the distance $\mathcal{W}_\alpha$ built from the α-stable jump kernel J_α and give sufficient conditions for finiteness in terms of moments.

The following result shows that under certain assumptions the distance $\mathcal{W}$ can be bounded from below by the L^1-Wasserstein distance. Recall that the L^1-Wasserstein distance is defined for $\mu_0, \mu_1 \in \mathscr{P}(\mathbb{R}^d)$ by

$$
W_1(\mu_0, \mu_1) \;:=\; \inf_\pi \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|\, \pi(\mathrm{d}x, \mathrm{d}y)\;,
$$

where the infimum is taken over all probability measures $\pi \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d)$ whose first and second marginal are μ_0 and μ_1 respectively (see e.g. [Vil09, Chap. 6]).

Proposition 3.4.5. *Assume that the jump kernel J satisfies*

$$M^2 \ := \ \sup_x \int |x-y|^2 \, J(x, \mathrm{d}y) \ < \ \infty \ . \tag{3.4.3}$$

Then for any $\mu_0, \mu_1 \in \mathscr{P}(\mathbb{R}^d)$ we have the bound

$$W_1(\mu_0, \mu_1) \ \leq \ \frac{M}{\sqrt{2}} \mathcal{W}(\mu_0, \mu_1) \ .$$

Proof. We can assume that $\mathcal{W}(\mu_0, \mu_1) < \infty$. Take a minimising curve $(\mu, \boldsymbol{\nu}) \in \mathcal{CE}_1(\mu_0, \mu_1)$ and let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a 1-Lipschitz function. Using the continuity equation in the form (3.3.4) and arguing similar as in Lemma 3.2.6, we estimate

$$\begin{aligned}
&\left| \int \varphi \mathrm{d}\mu_n - \int \varphi \mathrm{d}\mu \right| \\
&= \frac{1}{2} \left| \int_0^1 \int \overline{\nabla} \varphi \mathrm{d}\boldsymbol{\nu}_t^n \mathrm{d}t \right| \\
&\leq \frac{1}{2} \int_0^1 \int |x-y| \, |\boldsymbol{\nu}_t^n| \, (\mathrm{d}x, \mathrm{d}y) \mathrm{d}t \\
&\leq \frac{1}{\sqrt{2}} \left(\int_0^1 \mathcal{A}(\mu_t^n, \boldsymbol{\nu}_t^n) \mathrm{d}t \right)^{\frac{1}{2}} \left(\int_0^1 \int |x-y|^2 \, J(x, \mathrm{d}y) \mu_t(\mathrm{d}x) \mathrm{d}t \right)^{\frac{1}{2}} \\
&\leq \frac{M}{\sqrt{2}} \mathcal{W}(\mu_n, \mu) \ .
\end{aligned}$$

Taking the supremum over all 1-Lipschitz functions φ yields the claim by Kantorovich–Rubinstein duality (see e.g. [Vil09, 5.16]). □

The convexity and monotonicity properties of the action functional established in Section 3.2 extend naturally to the distance function.

Proposition 3.4.6 (Convexity of the squared distance)**.** *Let $\mu_0^j, \mu_1^j \in \mathscr{P}(\mathbb{R}^d)$ for $j = 0, 1$. For $\tau \in [0,1]$ and $k = 0, 1$ set $\mu_k^\tau = \tau \mu_k^1 + (1-\tau)\mu_k^0$. Then we have :*

$$\mathcal{W}(\mu_0^\tau, \mu_1^\tau)^2 \ \leq \ \tau \mathcal{W}(\mu_0^1, \mu_1^1)^2 + (1-\tau) \mathcal{W}(\mu_0^0, \mu_1^0)^2 \ .$$

Proof. We can assume without restriction that $\mathcal{W}(\mu_0^j, \mu_1^j)$ is finite and choose minimising curves $(\mu^j, \boldsymbol{\nu}^j) \in \mathcal{CE}_1(\mu_0^j, \mu_1^j)$. Then for $t \in [0,1]$ set $\mu_t^\tau = \tau \mu_t^1 + (1-\tau)\mu_t^0$ and $\boldsymbol{\nu}_t^\tau = \tau \boldsymbol{\nu}_t^1 + (1-\tau)\boldsymbol{\nu}_t^0$. Observe that $(\mu^\tau, \boldsymbol{\nu}^\tau) \in \mathcal{CE}_1(\mu_0^\tau, \mu_1^\tau)$. From the definition of $\mathcal{W}$ and the convexity of $\mathcal{A}$ as stated in Lemma 3.2.7 we infer

$$\begin{aligned}
\mathcal{W}(\mu_0^\tau, \mu_1^\tau)^2 \ &\leq \ \int_0^1 \mathcal{A}(\mu_t^\tau, \boldsymbol{\nu}_t^\tau) \mathrm{d}t \ \leq \ \int_0^1 \tau \mathcal{A}(\mu_t^1, \boldsymbol{\nu}_t^1) + (1-\tau) \mathcal{A}(\mu_t^0, \boldsymbol{\nu}_t^0) \mathrm{d}t \\
&= \ \tau \mathcal{W}(\mu_0^1, \mu_1^1)^2 + (1-\tau) \mathcal{W}(\mu_0^0, \mu_1^0)^2 \ .
\end{aligned}$$

□

Proposition 3.4.7 (Monotonicity under convolution)**.** *Let* $\mu_0, \mu_1 \in \mathscr{P}(\mathbb{R}^d)$*. Assume that* J *satisfies* (3.2.7) *and let* m *be Lebesgue measure. Let* k *be a convolution kernel. Then we have*

$$\mathcal{W}(\mu_0 * k, \mu_1 * k) \ \leq \ \mathcal{W}(\mu_0, \mu_1) \ .$$

If we set $k_\varepsilon(x) = \varepsilon^{-d} k(x/\varepsilon)$*, then as* $\varepsilon \searrow 0$ *we have*

$$\mathcal{W}(\mu_0 * k_\varepsilon, \mu_1 * k_\varepsilon) \ \longrightarrow \ \mathcal{W}(\mu_0, \mu_1) \ .$$

Proof. Assume that $\mathcal{W}(\mu_0, \mu_1)$ is finite, as otherwise there is nothing to prove. Let $(\mu, \boldsymbol{\nu}) \in \mathcal{CE}_1(\mu_0, \mu_1)$ be a minimising curve according to Proposition 3.4.3. Define $\widetilde{\mu}_t = \mu_t * k$, $\widetilde{\boldsymbol{\nu}}_t = \boldsymbol{\nu}_t * k$. We claim that $(\widetilde{\mu}, \widetilde{\boldsymbol{\nu}}) \in \mathcal{CE}_1(\mu_0 * k, \mu_1 * k)$. Indeed, let us show that the continuity equation (v) in (3.3.6) holds for $(\widetilde{\mu}, \widetilde{\boldsymbol{\nu}})$. The other properties are equally easy to verify. So let $\varphi \in C_c^\infty((0,1)\times\mathbb{R}^d)$ and set $\widetilde{\varphi}(t,x) = \int \varphi(t, x+z) k(z) \mathrm{d}z$. Using the continuity equation for $(\mu, \boldsymbol{\nu})$ and (3.2.8) we obtain

$$\begin{aligned}
\int \partial_t \varphi \mathrm{d}\widetilde{\mu}_t \mathrm{d}t &= \int \partial_t \varphi(t, x+z) k(z) \mathrm{d}z \mu_t(\mathrm{d}x) \mathrm{d}t \\
&= \int \partial_t \widetilde{\varphi} \mathrm{d}\mu_t \mathrm{d}t \ = \ -\frac{1}{2} \int \overline{\nabla} \widetilde{\varphi} \mathrm{d}\boldsymbol{\nu}_t \mathrm{d}t \\
&= -\frac{1}{2} \int \overline{\nabla} \varphi(t, x+z, y+z) k(z) \boldsymbol{\nu}_t(\mathrm{d}x, \mathrm{d}y) \mathrm{d}z \mathrm{d}t \\
&= -\frac{1}{2} \int \overline{\nabla} \varphi \mathrm{d}\widetilde{\boldsymbol{\nu}}_t \mathrm{d}t \ .
\end{aligned}$$

Now the first assertion follows immediately from Proposition 3.2.8. This in turn together with weak lower semicontinuity of $\mathcal{W}$ (see Theorem 3.4.4) yields the second assertion. □

3.4.2 Absolutely continuous curves and tangent structure

We now give a characterisation of absolutely continuous curves with respect to $\mathcal{W}$ and consider a notion of tangent bundle.

A curve $(\mu_t)_{t\in[0,T]}$ in $\mathscr{P}(\mathbb{R}^d)$ is called absolutely continuous w.r.t. $\mathcal{W}$ if there exists $m \in L^1(0,T)$ such that

$$\mathcal{W}(\mu_s, \mu_t) \ \leq \ \int_s^t m(r) \mathrm{d}r \quad \forall\ 0 \leq s \leq t \leq T \ . \tag{3.4.4}$$

For an absolutely continuous curve the metric derivative defined by

$$|\mu_t'| \ := \ \lim_{h\to 0} \frac{\mathcal{W}(\mu_{t+h}, \mu_t)}{|h|}$$

exists for a.e. $t \in [0,T]$ and is the minimal m in (3.4.4) (see [AGS08, Thm. 1.1.2]).

Proposition 3.4.8 (Metric velocity). *A curve $(\mu_t)_{t\in[0,T]}$ is absolutely continuous with respect to $\mathcal{W}$ if and only if there exists a Borel family $(\boldsymbol{\nu}_t)_{t\in[0,T]}$ such that $(\mu,\boldsymbol{\nu}) \in \mathcal{CE}_T$ and*

$$\int_0^T \sqrt{\mathcal{A}(\mu_t,\boldsymbol{\nu}_t)}dt \;<\; \infty\;.$$

In this case we have $|\mu_t'|^2 \leq \mathcal{A}(\mu_t,\boldsymbol{\nu}_t)$ for a.e. $t \in [0,T]$. Moreover, there exists a unique Borel family $\widetilde{\boldsymbol{\nu}}_t$ with $(\mu,\widetilde{\boldsymbol{\nu}}) \in \mathcal{CE}_T$ such that

$$\left|\mu_t'\right|^2 = \mathcal{A}(\mu_t,\widetilde{\boldsymbol{\nu}}_t) \qquad \text{for a.e. } t \in [0,T]\;. \tag{3.4.5}$$

Proof. The proof follows from the very same arguments as in [DNS09, Thm. 5.17]. □

We can describe the optimal velocity measures $\widetilde{\boldsymbol{\nu}}_t$ appearing in the preceding proposition in more detail. We define

$$\begin{aligned} T_\mu\mathscr{P}(\mathbb{R}^d) \;:=\; \Big\{&\boldsymbol{\nu} \in \mathcal{M}_{loc}(G) \;:\; \mathcal{A}(\mu,\boldsymbol{\nu}) < \infty\;, \\ &\mathcal{A}(\mu,\boldsymbol{\nu}) \;\leq\; \mathcal{A}(\mu,\boldsymbol{\nu}+\boldsymbol{\eta})\;\forall\boldsymbol{\eta}\;:\;\overline{\nabla}\cdot\boldsymbol{\eta} = 0\Big\}\;. \end{aligned} \tag{3.4.6}$$

Here $\overline{\nabla}\cdot\boldsymbol{\eta} = 0$ is understood in a weak sense, i.e.

$$\frac{1}{2}\int \overline{\nabla}\xi(x,y)\boldsymbol{\eta}(\mathrm{d}x,\mathrm{d}y) \;=\; 0 \qquad \forall \xi \in C_c^\infty(\mathbb{R}^d)\;.$$

Corollary 3.4.9. *Let $(\mu,\boldsymbol{\nu}) \in \mathcal{CE}_T$ such that the curve $t \mapsto \mu_t$ is absolutely continuous w.r.t. $\mathcal{W}$. Then $\boldsymbol{\nu}$ satisfies (3.4.5) if and only if $\boldsymbol{\nu}_t \in T_{\mu_t}\mathscr{P}(\mathbb{R}^d)$ for a.e. $t \in [0,T]$.*

If μ is absolutely continuous with respect to m we can give an explicit description of $T_\mu\mathscr{P}(\mathbb{R}^d)$ as a subspace of an L^2 space. For this recall that we denote by $Jm \in \mathcal{M}_{loc}(G)$ the measure given by $Jm(\mathrm{d}x,\mathrm{d}y) = J(x,\mathrm{d}y)m(\mathrm{d}x)$.

Proposition 3.4.10. *Let $\mu = \rho m \in \mathscr{P}(\mathbb{R}^d)$. Then we have $\boldsymbol{\nu} \in T_\mu\mathscr{P}(\mathbb{R}^d)$ if and only if $\boldsymbol{\nu} = w\hat{\rho}Jm$ is absolutely continuous w.r.t. the measure $\hat{\rho}Jm$ and*

$$w \;\in\; \overline{\{\overline{\nabla}\varphi \mid \varphi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\hat{\rho}Jm)} \;=:\; T_\rho\;.$$

Proof. If $\mathcal{A}(\mu,\boldsymbol{\nu})$ is finite we infer from Lemma 3.2.3 that $\boldsymbol{\nu} = w\hat{\rho}Jm$ for some density $w : G \to \mathbb{R}$ and that $\mathcal{A}(\mu,\boldsymbol{\nu}) = \|w\|^2_{L^2(\hat{\rho}Jm)}$. Now the optimality condition in (3.4.6) is equivalent to

$$\|w\|_{L^2(\hat{\rho}Jm)} \;\leq\; \|w+v\|_{L^2(\hat{\rho}Jm)} \qquad \forall v \in N_\rho\;,$$

where $N_\rho := \{v \in L^2(\hat{\rho}Jm) \;:\; \int \overline{\nabla}\xi v\hat{\rho}\;\mathrm{d}Jm = 0\;\forall \xi \in C_c^\infty(\mathbb{R}^d)\}$. This implies the assertion of the proposition after noting that N_ρ is the orthogonal complement in L^2 of T_ρ. □

In the light of the formal Riemannian interpretation of the distance $\mathcal{W}$ it seems natural to view $T_\mu\mathscr{P}(\mathbb{R}^d)$ as the tangent space to $\mathscr{P}(\mathbb{R}^d)$ at the measure μ. This is reminiscent of Otto's Riemannian interpretation of the L^2-Wasserstein space [Ott01]. The results obtained here are in close analogy to the notion of tangent bundle to the Wasserstein space studied in [AGS08, Sec. 8.4].

3.4.3 The α-stable distance $\mathcal{W}_\alpha$

In this section we focus on the jump kernel associated to the α-stable process and establish some special properties of the corresponding pseudo distance.

For $\alpha \in (0,2)$ consider the jump kernel

$$J_\alpha(x,\mathrm{d}y) \;=\; \frac{c_{\alpha,d}}{|x-y|^{\alpha+d}}\mathrm{d}y\;,$$

where $c_{\alpha,d}$ is a constant depending on α and the dimension d.

Definition 3.4.11. *We denote by $\mathcal{W}_\alpha$ the distance built from the jump kernel J_α according to Definition 2.2.3 where we choose θ to be the logarithmic mean.*

Before we investigate this distance, let us collect some facts about the kernel J_α and related objects.

The jump kernel J_α gives rise to a non-local operator which, for a suitable choice of the constant $c_{\alpha,d}$ coincides with the fractional Laplacian, i.e.

$$-(-\Delta)^{\frac{\alpha}{2}}u(x) \;=\; \int u(y)-u(x)-(y-x)\cdot\nabla u(x)\mathbf{1}_{\{|x-y|\leq 1\}}J_\alpha(x,\mathrm{d}y)\;.$$

This is a pseudo differential operator with symbol $|\xi|^\alpha$. This means that

$$\mathcal{F}(-(-\Delta)^{\frac{\alpha}{2}}u)(\xi) \;=\; |\xi|^\alpha\,\mathcal{F}u(\xi)\;,$$

where $\mathcal{F}$ denotes the Fourier transform. It is well known that the fractional heat equation possesses a fundamental solution, i.e a function $\psi : (0,\infty)\times\mathbb{R}^d \to \mathbb{R}_+$ satisfying

$$\begin{aligned}\partial_t\psi + (-\Delta)^{\frac{\alpha}{2}}\psi \;&=\; 0 \quad \text{in } (0,\infty)\times\mathbb{R}^d\;,\\ \psi(t,\cdot)\;&\longrightarrow\;\delta_0 \quad \text{as } t\to 0\;.\end{aligned} \tag{3.4.7}$$

Although for general α the function ψ has no closed expression, its Fourier transform can be given explicitly:

$$\mathcal{F}(\psi_t)(\xi) \;=\; \int e^{i\xi\cdot x}\psi(t,x)\mathrm{d}x \;=\; e^{-t|\xi|^\alpha}\;. \tag{3.4.8}$$

This implies the following scaling behaviour

$$\psi(t,x) \;=\; t^{-\frac{d}{\alpha}}\cdot\psi(1,x\cdot t^{-\frac{1}{\alpha}})\;. \tag{3.4.9}$$

The function ψ is known to be smooth and strictly positive on $(0,\infty)\times\mathbb{R}^d$. Moreover, one has the following heat kernel bounds (see e.g. [CK03, Thm. 1.1]): There exist a constant $C>0$ such that for all x,t we have

$$\frac{1}{C}\cdot\left(t^{-\frac{d}{\alpha}}\wedge\frac{t}{|x|^{\alpha+d}}\right)\;\le\;\psi(t,x)\;\le\;C\cdot\left(t^{-\frac{d}{\alpha}}\wedge\frac{t}{|x|^{\alpha+d}}\right)\;. \tag{3.4.10}$$

Note that the fractional Laplacian is also the generator of the symmetric α-stable Lévy process. The measure $q_t(\mathrm{d}x)=\psi_t(x)\mathrm{d}x$ is the law at time t of this process started in 0.

Now we are ready to study the distance $\mathcal{W}_\alpha$. We will first show that it also enjoys a scaling property.

Lemma 3.4.12. *For $h>0$ consider the map $s^h(x)=h\cdot x$. For any $\mu_0,\mu_1\in\mathscr{P}(\mathbb{R}^d)$ we have*

$$\mathcal{W}_\alpha(s^h_*\mu_0,s^h_*\mu_1)\;=\;h^{\frac{\alpha}{2}}\cdot\mathcal{W}_\alpha(\mu_0,\mu_1)\;.$$

Proof. Let $(\mu,\boldsymbol{\nu})\in\mathcal{CE}_1(\mu_0,\mu_1)$ be a minimising curve for $\mathcal{W}_\alpha$. We set $\mu^h_t:=s^h_*\mu_t$ and $\boldsymbol{\nu}^h_t=(s^h\times s^h)_*\boldsymbol{\nu}_t$ to obtain a pair $(\mu^h,\boldsymbol{\nu}^h)\in\mathcal{CE}_1(s^h_*\mu_0,s^h_*\mu^h_1)$. To calculate the action of this curve, recall that for a given $\mu\in\mathscr{P}(\mathbb{R}^d)$ we defined the measures $\mu^1(\mathrm{d}x,\mathrm{d}y):=J_\alpha(x,\mathrm{d}y)\mu(\mathrm{d}x)$ and $\mu^2(\mathrm{d}x,\mathrm{d}y):=J_\alpha(y,\mathrm{d}x)\mu(\mathrm{d}y)$. It is easily checked that

$$(s^h_*\mu)^i\;=\;h^{-\alpha}\cdot(s^h\times s^h)_*\mu^i\;,\quad i=1,2\;. \tag{3.4.11}$$

We choose a reference measure σ such that $\mu^i_t=\rho^i_t\sigma$ and $\boldsymbol{\nu}_t=w_t\sigma$ are all absolutely continuous w.r.t. σ and set $\sigma^h:=(s^h\times s^h)_*\sigma$. By (3.4.11) we have that $(\mu^h_t)^i=\rho^{h,i}_t\sigma^h$ and $\boldsymbol{\nu}^h_t=w^h_t\sigma^h$ with $\rho^{h,i}_t(x,y)=h^{-\alpha}\cdot\rho^i_t(\frac{x}{h},\frac{y}{h})$ and $w^h_t(x,y)=w_t(\frac{x}{h},\frac{y}{h})$. Thus we obtain

$$\mathcal{A}(\mu^h_t,\boldsymbol{\nu}^h_t)\;=\;\int\frac{|w^h_t|^2}{2\theta(\rho^{h,1}_t,\rho^{h,2}_t)}\mathrm{d}\sigma^h\;=\;h^\alpha\cdot\int\frac{|w_t|^2}{2\theta(\rho^1_t,\rho^2_t)}\mathrm{d}\sigma=\;h^\alpha\cdot\mathcal{A}(\mu_t,\boldsymbol{\nu}_t)\;.$$

Integrating over t, we obtain the estimate

$$\mathcal{W}_\alpha(s^h_*\mu_0,s^h_*\mu_1)\;\le\;h^{\frac{\alpha}{2}}\cdot\mathcal{W}_\alpha(\mu_0,\mu_1)\;.$$

We can apply a similar argument scaling a geodesic between $s^h_*\mu_0$ and $s^h_*\mu_1$ by h^{-1} to obtain equality. □

Our next result shows that any two Dirac masses have finite distance.

Lemma 3.4.13. *For any $x,y\in\mathbb{R}^d$ we have that*

$$\mathcal{W}_\alpha(\delta_x,\delta_y)\;=\;c\cdot|x-y|^{\frac{\alpha}{2}}\;,$$

where $c:=\mathcal{W}_\alpha(\delta_0,\delta_v)<\infty$ for any $v\in\mathbb{R}^d$ with $|v|=1$.

Proof. By rotational symmetry of the jump kernel J_α it is obvious that the definition of c does not depend on the choice of v. We will show that c is finite, the rest will then follow from Lemma 3.4.12.

So fix $v \in \mathbb{R}^d$ with $|v| = 1$. For $t > 0$ consider the measure $q_t(\mathrm{d}x) = \psi_t(x)\mathrm{d}x$, where ψ is the fractional heat kernel given by (3.4.8). We set $q_0 = \delta_0$. We first note that

$$\mathcal{W}_\alpha(\delta_0, q_1) < \infty \ . \tag{3.4.12}$$

Indeed, if we set set $\boldsymbol{\nu}_t(\mathrm{d}x, \mathrm{d}y) = \overline{\nabla}\psi_t(x,y)J_\alpha(x,\mathrm{d}y)\mathrm{d}x$, we have by the fractional heat equation (3.4.7) that $(q, \boldsymbol{\nu}) \in \mathcal{CE}_1(\delta_0, q_1)$. We compute

$$\mathcal{A}(q_t, \boldsymbol{\nu}_t) = \int \overline{\nabla}\psi_t(x,y)\overline{\nabla}\log\psi_t(x,y)\frac{c_\alpha}{|x-y|^{d+\alpha}}\mathrm{d}x\mathrm{d}y \ =: \ \mathcal{I}(q_t) \ .$$

From the explicit Fourier representation (3.4.8) and the heat kernel bounds (3.4.10) one can check that $\mathcal{I}(q_t) < \infty$ for $t > 0$. Moreover, using the scaling property (3.4.9) of the heat kernel ψ, it is easily checked that $\mathcal{I}(q_t) = t^{-1}\mathcal{I}(q_1)$. Hence, using Lemma 3.4.2, we estimate

$$\begin{aligned}
\mathcal{W}_\alpha(\delta_0, q_1) \ &\leq \ \int_0^1 \sqrt{\mathcal{A}(q_t, \boldsymbol{\nu}_t)}\mathrm{d}t \ = \ \int_0^1 \sqrt{\mathcal{I}(q_t)}\mathrm{d}t \\
&= \ \sqrt{\mathcal{I}(q_1)}\int_0^1 t^{-\frac{1}{2}}\mathrm{d}t \ = \ 2\sqrt{\mathcal{I}(q_1)} \ < \ \infty \ .
\end{aligned}$$

By translation invariance we immediately deduce that

$$\mathcal{W}_\alpha(\delta_v, q_1^v) < \infty \ , \tag{3.4.13}$$

where $q_1^v = q_1(\cdot - v)$. We will now show that

$$\mathcal{W}_\alpha(q_1, q_1^v) \ < \ \infty \ . \tag{3.4.14}$$

For $z \in \mathbb{R}^d$ we denote by $Q_z := z + [0,1]^d$ the unit hypercube at z. Furthermore, we set $\sigma_z := \mathbf{1}_{Q_z} \cdot q_1$ and $\sigma_z^v := \mathbf{1}_{Q_{z+v}} \cdot q_1^v = \sigma_z(\cdot - v)$ and $m_z := q_1(Q_z) > 0$. We consider the following interpolating path, for $t \in [0,1]$ we define:

$$\mu_t := (1-t)q_1 + tq_1^v \ , \quad \boldsymbol{\nu}_t := \sum_{z\in\mathbb{Z}^d}\frac{2}{m_z}\sigma_z \otimes \sigma_z^v \ .$$

We claim that $(\mu, \boldsymbol{\nu}) \in \mathcal{CE}_1(q_1, q_1^v)$. Indeed, for any $\varphi \in C_c^\infty(\mathbb{R}^d)$ we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\int \varphi \mathrm{d}\mu_t \ &= \ \int \varphi \mathrm{d}(q_1^v - q_1) \\
&= \ \sum_{z\in\mathbb{Z}^d}\int \varphi\mathrm{d}(\sigma_z^v - \sigma_z) \\
&= \ \sum_{z\in\mathbb{Z}^d}\int \big(\varphi(y) - \varphi(x)\big)\frac{1}{m_z}\sigma_z^v(\mathrm{d}y)\sigma_z(\mathrm{d}x) \\
&= \ \frac{1}{2}\int \overline{\nabla}\varphi \mathrm{d}\boldsymbol{\nu}_t \ .
\end{aligned}$$

Note that $\mu_t(\mathrm{d}x) = \rho_t(x)\mathrm{d}x$ with $\rho_t(x) = (1-t)\psi_1(x) + t\psi_1(x-v)$. Thus we can write the action as

$$\mathcal{A}(\mu_t, \boldsymbol{\nu}_t) = \sum_{z\in\mathbb{Z}^d} \int \frac{|\psi_1(x)\psi_1(y-v)|^2}{\theta(\rho_t(x),\rho_t(y))} \frac{2}{c_\alpha m_z^2} |x-y|^{\alpha+d} \mathbf{1}_{Q_z}(x)\mathbf{1}_{Q_{z+v}}(y)\mathrm{d}x\mathrm{d}y \ .$$

To estimate this quantity, first note that for any $x \in Q_z, y \in Q_{z+v}$ we obviously have $|x-y| \leq |v| + 2\sqrt{d} =: C_1$. Furthermore, by monotonicity (A5) and homogeneity (A6), we have the following inequality for the function θ. For all $a,b,c,d > 0$ and $t \in (0,1)$ we have

$$\frac{(ab)^2}{\theta((1-t)a+tc,(1-t)d+tb)} \leq \frac{ab}{\theta(1-t,t)} \max(a,b) \ .$$

Using the fact that $\sup_{x\in Q_z} |\psi_1(x)| \leq C_2 |z|^{-\alpha-d}$ for a suitable constant C_2 by the upper heat kernel bound in (3.4.10), we obtain

$$\mathcal{A}(\mu_t, \boldsymbol{\nu}_t) \leq C_1 C_2 \frac{2}{c_\alpha} \frac{1}{\theta(1-t,t)} \sum_{z\in\mathbb{Z}^d} |z|^{-\alpha-d} = C_3 \frac{1}{\theta(1-t,t)}$$

with a finite constant C_3. Hence, the distance between q_1 and q_1^v can be bounded as

$$\begin{aligned}\mathcal{W}_\alpha(q_1, q_1^v) &\leq \int_0^1 \sqrt{\mathcal{A}(\mu_t,\boldsymbol{\nu}_t)}\mathrm{d}t \leq C_3 \int_0^1 \frac{\mathrm{d}t}{\sqrt{\theta(1-t,t)}} \\ &\leq C_3 \int_0^1 \frac{\mathrm{d}t}{\sqrt{t\wedge(1-t)}} < \infty \ ,\end{aligned}$$

where we have used the monotonicity of θ (A5) in the last step. Using the triangle inequality and combining (3.4.12), (3.4.13) and (3.4.14), we complete the proof. □

We end this section by giving a sufficient condition for two measure to have finite $\mathcal{W}_\alpha$ distance in terms of their moments. For $\mu \in \mathscr{P}(\mathbb{R}^d)$ we define the *α-moment* by

$$m_\alpha(\mu) := \int |x|^\alpha \mu(\mathrm{d}x) \ .$$

Moreover we denote the set of probability measures with finite α-moment by

$$\mathscr{P}_\alpha(\mathbb{R}^d) := \left\{\mu \in \mathscr{P}(\mathbb{R}^d) \mid m_\alpha(\mu) < \infty\right\} \ .$$

The following result is a non-local analogue of the fact that finiteness of p-th moments implies finiteness of the L^p-Wasserstein distance.

Proposition 3.4.14. *For any $\mu_0, \mu_1 \in \mathscr{P}_\alpha(\mathbb{R}^d)$ we have that*

$$\mathcal{W}_\alpha(\mu_0,\mu_1) \leq c\cdot\left(\sqrt{m_\alpha(\mu_0)} + \sqrt{m_\alpha(\mu_1)}\right) < \infty \ ,$$

where c is the constant from Lemma 3.4.13.

Proof. Using the convexity of the squared distance as stated in Proposition 3.4.6 and Lemma 3.4.13, we estimate

$$\begin{aligned}\mathcal{W}_\alpha(\delta_0,\mu_0)^2 \;&\le\; \int \mathcal{W}_\alpha(\delta_0,\delta_x)^2\mu_0(\mathrm{d}x)\\ &=\; c^2\int |x|^\alpha\,\mu_0(\mathrm{d}x) \;=\; c^2\cdot m_\alpha(\mu_0)\;.\end{aligned}$$

The same estimate holds for μ_1 and we conclude by the triangle inequality. □

3.5 Geodesic convexity and gradient flow of the entropy

In this section we focus on a translation invariant jump kernel J and identify the evolution equation generated by the associated non-local operator as the gradient flow of the relative entropy with respect to the distance $\mathcal{W}$.

Assumption 3.5.1. *Throughout the remainder of this chapter we assume that θ is the logarithmic mean.*

First, we have to make precise what we mean by gradient flow. Among several possibilities to define the notion of gradient flow in a metric space the so called "Evolution Variational Inequality"(EVI) is one of the most powerful and restrictive concepts. We refer to [AGS08] for a comprehensive study of gradient flows in metric spaces. We adopt the following

Definition 3.5.2. *Let (X,d) be a metric space and $F:X\to(-\infty,\infty]$ a lower semicontinuous function such that its proper domain $D(F):=\{x\in X\mid F(x)<\infty\}$ is dense in X. Further let $(S_t)_{t\ge0}$ be a C^0-semigroup on X and $\lambda\in\mathbb{R}$. S is called the (λ-)gradient flow of F if $S_t(X)\subset D(F)$ for all $t>0$, the map $t\mapsto F(S_t(u))$ is non-increasing in $(0,\infty)$ for all $u\in X$ and if for all $u\in X, v\in D(F), t>0$:*

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}d^2(S_t(u),v)+\frac{\lambda}{2}d^2(S_t(u),v)+F(S_t(u)) \;\le\; F(v)\;. \tag{3.5.1}$$

Here and in the following we will use the notation

$$\frac{\mathrm{d}^+}{\mathrm{d}t}f(t):=\limsup_{h\searrow0}\frac{f(t+h)-f(t)}{h}\;.$$

We will only consider translation invariant jump kernels. More precisely, from now on we make the following

Assumption 3.5.3. *Assume that m is Lebesgue measure on $\mathbb{R}^d$ and that J satisfies*

$$J(x+z,A+z) \;=\; J(x,A)\;,\quad \forall x,z\in\mathbb{R}^d, A\in\mathcal{B}(\mathbb{R}^d)\;.$$

Let us denote by ν the Borel measure on $\mathbb{R}^d$ given by $J(0,\cdot)$. Then Assumption 3.1.1 together with the reversibility assumption simply reduces to the requirement that ν is a symmetric Lévy measure, i.e. it satisfies $\nu(A) = \nu(-A)$ for all $A \subset \mathbb{R}^d$ as well as

$$C_\nu \ := \ \int (1 \wedge |z|^2)\, \nu(\mathrm{d}z) \ < \ \infty \ . \tag{3.5.2}$$

The jump kernel J gives rise to a non-local operator $\mathcal{L}$ given by

$$\mathcal{L}u(x) \ = \ \frac{1}{2}\int \big(u(x+z) + u(x-z) - 2u(x)\big)\nu(\mathrm{d}z) \ .$$

We will use the shorthand notation $\delta u(x,z) := \frac{1}{2}\big(u(x+z) + u(x-z) - 2u(x)\big)$.

Note that $\mathcal{L}$ is also the generator of the Lévy process $(X_t)_{t\geq 0}$ with vanishing drift and diffusion and with Lévy measure ν (see e.g. [App04] or [Ber96] for background on Lévy processes). Let us denote by q_t the law of the Lévy process X at time t started in 0. This law q_t can be given explicitly in terms of its Fourier transformation. Namely, we have

$$\int e^{ix\cdot\xi} q_t(\mathrm{d}x) \ = \ \mathbb{E}\big[\exp(i\langle\xi, X_t\rangle)\big] \ = \ \exp(-t\eta(\xi)) \ ,$$

where η is given by the Lévy-Khintchine formula:

$$\eta(\xi) \ = \ \int e^{i\langle y,\xi\rangle} - 1 - i\langle y,\xi\rangle \mathbf{1}_{\{|y|\leq 1\}}\nu(\mathrm{d}y) \ .$$

The generator $\mathcal{L}$ is a pseudo differential operator with symbol η. This means that $\mathcal{F}(\mathcal{L}u)(\xi) = \eta(\xi)\mathcal{F}(u)(\xi)$, where $\mathcal{F}$ denotes the Fourier transform.

Given $\mu \in \mathscr{P}(\mathbb{R}^d)$ we define its relative entropy w.r.t. a measure γ by

$$\mathcal{H}(\mu|\gamma) \ := \ \begin{cases} \int \rho \log\rho \,\mathrm{d}\gamma \ , & \text{if } \mu = \rho\gamma \text{ and } \int(\rho\log\rho)_+\mathrm{d}\gamma < \infty \\ +\infty \ , & \text{else.} \end{cases}$$

We will use the shorthand notation $\mathcal{H}(\mu) := \mathcal{H}(\mu|m)$ for the relative entropy w.r.t. Lebesgue measure. As before, we will denote by $Jm \in \mathcal{M}_{loc}(G)$ the measure given by $Jm(\mathrm{d}x,\mathrm{d}y) = J(x,\mathrm{d}y)m(\mathrm{d}x)$. For a probability measure $\mu \in \mathscr{P}(\mathbb{R}^d)$ we define a non-local analogue of the Fisher information by

$$\mathcal{I}(\mu) \ := \ \begin{cases} \frac{1}{2}\int \overline{\nabla}\rho\overline{\nabla}\log\rho \,\mathrm{d}(Jm), & \text{if } \mu = \rho m \text{ and } \rho > 0 \ , \\ +\infty \ , & \text{else .} \end{cases} \tag{3.5.3}$$

Throughout this section we will make the following assumption on ν in terms of the law of the associated Lévy process.

Assumption 3.5.4. *For any $t > 0$ the measure q_t is absolutely continuous w.r.t. m with density ψ_t, where $\psi : (0,\infty) \times \mathbb{R}^d \to \mathbb{R}_+$ is smooth, bounded and strictly positive. We assume that ψ is a fundamental solution to the equation $\partial_t u = \mathcal{L}u$, i.e.*

$$\begin{aligned} \partial_t \psi &= \mathcal{L}\psi \quad \text{in } (0,\infty) \times \mathbb{R}^d , \\ \psi(t,\cdot) &\longrightarrow \delta_0 \quad \text{as } t \to 0 . \end{aligned}$$

Moreover, we assume that

$$\mathcal{H}(q_t) \in (-\infty, \infty) \quad \forall t > 0 , \tag{3.5.4}$$

$$\int_0^t \sqrt{\mathcal{I}(q_s)}\mathrm{d}s \ < \ \infty \quad \forall t > 0 , \tag{3.5.5}$$

$$\int_s^r |\delta\psi_t(x,z)|\, \nu(\mathrm{d}z) m(\mathrm{d}x)\mathrm{d}t \ < \ \infty \quad \forall 0 < s < r . \tag{3.5.6}$$

We will also assume a control on the moment of the Lévy measure.

Assumption 3.5.5. *There exists a constant $\beta > 0$ such that*

$$M_\beta \ := \ \int \mathbf{1}_{\{|x|>1\}}\, |x|^\beta\, \nu(\mathrm{d}x) \ < \ \infty .$$

Remark 3.5.6. The assumptions on the regularity of ψ are made in order to make the presentation of the proofs in this section as simple as possible and could be weakened. In [AGS11a] e.g. similar calculations as here are performed under very mild assumptions in a local setting. Still, Assumption 3.5.4 is fulfilled e.g. for $\nu_\alpha(\mathrm{d}y) = c_\alpha\, |y|^{-\alpha-d}$ for $\alpha \in (0,2)$ corresponding to the fractional Laplacian $-(-\Delta)^{\frac{\alpha}{2}}$ with symbol $\eta(\xi) = |\xi|^\alpha$. This can be checked using the explicit Fourier representation (3.4.8) and the heat kernel bounds (3.4.10).

Assumption 3.5.5 is only used in Proposition 3.5.7 to ensure lower semicontinuity of the entropy w.r.t. $\mathcal{W}$-convergence. For ν_α it is satisfied for any $\beta < \alpha$.

The Lévy process generated by the operator $\mathcal{L}$ gives rise to a convolution semigroup $(P_t)_{t\geq 0}$ acting on $\mathscr{P}(\mathbb{R}^d)$ via

$$P_t[\mu] \ := \ \mu * q_t \ = \ \mu * \psi_t \ = \ \int \mu(\cdot - z)\psi_t(z)\mathrm{d}z .$$

For $\boldsymbol{\nu} \in \mathcal{M}(G)$ we set

$$P_t[\boldsymbol{\nu}] \ := \ \boldsymbol{\nu} * \psi_t ,$$

with the convolution being understood in the sense of (3.2.8). Proposition 3.4.7 shows that P is a C^0-semigroup in the sense that $\mathcal{W}(P_t[\mu],\mu) \to 0$ as $t \to 0$.

In order to characterise the semigroup P_t as the gradient flow of the entropy, we want to apply Definition 3.5.2 in the case where the space X is (a subspace of) the space of probability measures $\mathscr{P}(\mathbb{R}^d)$ equipped with the distance $\mathcal{W}$ and the functional F is the relative entropy $\mathcal{H}$. Let us denote

$$\mathscr{P}^* \;:=\; \{\mu \in \mathscr{P}(\mathbb{R}^d) \;:\; \mathcal{H}(\mu) > -\infty\}\;.$$

We set $X := \mathscr{P}_\tau = \{\mu \in \mathscr{P}(\mathbb{R}^d) \;:\; \mathcal{W}(\mu,\tau) < \infty\}$ for some $\tau \in \mathscr{P}^*$. The next result ensures that this choice fits well into the setting of Definition 3.5.2.

Proposition 3.5.7. *Let $\tau \in \mathscr{P}^*$. For any $\mu \in \mathscr{P}(\mathbb{R}^d)$ with $\mathcal{W}(\mu,\tau) < \infty$ we have $\mathcal{H}(\mu) > -\infty$, i.e. $\mathscr{P}_\tau \subset \mathscr{P}^*$. Moreover, the entropy functional $\mathcal{H} : \mathscr{P}_\tau \to (-\infty,\infty]$ is lower semicontinuous w.r.t. convergence in the metric $\mathcal{W}$.*

Proof. To prepare for the proof let us fix a measure $\gamma(\mathrm{d}x) := \exp(-V(x))\mathrm{d}x$ with $V(x) := \max(1,|x|^{\frac{\beta}{2}}) + c$. Here β is the constant from Assumption 3.5.5 and the constant c is chosen such that γ is a probability measure. We can assume that $\beta < 1$. Using the inequality $\left||y|^{\frac{\beta}{2}} - |x|^{\frac{\beta}{2}}\right| \leq |y-x|^{\frac{\beta}{2}}$, it is easy to check that

$$\left|\overline{\nabla} V(x,y)\right| \;=\; |V(y) - V(x)| \;\leq\; \min(|y-x|\,,|y-x|^{\frac{\beta}{2}})\;. \tag{3.5.7}$$

Now note that for any $\mu \in \mathscr{P}(\mathbb{R}^d)$ we have

$$\mathcal{H}(\mu) \;=\; \mathcal{H}(\mu|\gamma) - \int V(x)\mu(\mathrm{d}x) \tag{3.5.8}$$

Moreover, $\mathcal{H}(\mu|\gamma) \geq 0$ since γ is a probability measure.

Let us now show the first statement of the proposition. By (3.5.8) we have that the integral $\int V\mathrm{d}\tau$ is finite and we have to show that $\int V\mathrm{d}\mu$ is finite as well. Let $(\mu_s,\boldsymbol{\nu}_s)_s \in [0,1]$ be a minimising curve in $\mathcal{CE}_1(\tau,\mu)$. For $n \in \mathbb{N}$ we define the function $V_n(x) := \max(V(x),n)$. Arguing similar as in the proof of Lemma 3.2.6 or Proposition 3.4.5 and using (3.5.7) we obtain

$$\begin{aligned}\left|\int V_n\mathrm{d}\mu - \int V_n\mathrm{d}\tau\right| \;&\leq\; \mathcal{W}(\mu,\tau)\cdot\left(\frac{1}{2}\int_0^1\int \left|\overline{\nabla} V_n(x,y)\right|^2 J(x,\mathrm{d}y)\mu_s(\mathrm{d}x)\mathrm{d}s\right)^{\frac{1}{2}}\\ &\leq\; \mathcal{W}(\mu,\tau)\cdot\left(\frac{1}{2}\int_0^1\int \min(|z|^2\,,|z|^\beta)\nu(\mathrm{d}z)\mu_s(\mathrm{d}x)\mathrm{d}s\right)^{\frac{1}{2}}\\ &\leq\; \mathcal{W}(\mu,\tau)\cdot\sqrt{\frac{M_\beta + C_\nu}{2}}\;.\end{aligned}$$

Here M_β is the constant from Assumption 3.5.5 and C_ν is given by (3.5.2). Letting $n \to \infty$, monotone convergence yields

$$\left|\int V\mathrm{d}\mu - \int V\mathrm{d}\tau\right| \;\leq\; \mathcal{W}(\mu,\tau)\cdot\sqrt{M_\beta + C_\nu}$$

and in particular finiteness of the integral $\int V \mathrm{d}\mu$.

To prove the lower semicontinuity statement, fix $\mu \in \mathscr{P}_\tau$ and a sequence (μ_n) such that $\mathcal{W}(\mu_n, \mu) \to 0$. By Theorem 3.4.4 we have $\mu_n \rightharpoonup \mu$ weakly and it is well known that $\mathcal{H}(\cdot|\gamma)$ is lower semicontinuous w.r.t. weak convergence of probability measures (see e.g. [AGS08, Lemma 9.4.3]). Furthermore, arguing as before, we obtain the estimate

$$\left|\int V \mathrm{d}\mu_n - \int V \mathrm{d}\mu\right| \;\leq\; \mathcal{W}(\mu_n, \mu) \cdot \sqrt{\frac{M_\beta + C_\nu}{2}} \;\longrightarrow\; 0\ .$$

In view of (3.5.8) this finishes the proof. □

Let us now state a result giving the entropy production along the semigroup P.

Proposition 3.5.8. *Let $\mu \in \mathscr{P}^*$. For every $t > 0$ we have $\mathcal{H}(P_t[\mu]) \in (-\infty, \infty)$ and $\mathcal{I}(P_t[\mu]) < \infty$. Moreover, we have the energy identity*

$$\mathcal{H}(P_t[\mu]) - \mathcal{H}(P_s[\mu]) \;=\; -\int_s^t \mathcal{I}(P_r[\mu])\,\mathrm{d}r \quad \forall t \geq s > 0\ . \tag{3.5.9}$$

In particular the map $t \mapsto \mathcal{H}(P_t[\mu])$ is non-increasing.

Proof. Note that $P_t[\mu] = \rho_t m$ is absolutely continuous w.r.t. Lebesgue measure for every $t > 0$ where

$$\rho_t(x) = \int \psi_t(x - z)\mu(\mathrm{d}z)\ .$$

Finiteness of $\mathcal{H}(P_t[\mu])$ and $\mathcal{I}(P_t[\mu])$ follows immediately from (3.5.4), (3.5.5) and convexity of the maps $r \mapsto r\log r$ and $(r,s) \mapsto (r-s)(\log r - \log s)$.

We prove (3.5.9) by approximating $\mathcal{H}$ with functionals $\mathcal{H}_n$. Let us set

$$f_n(u) := \int_0^u \max(1 + \log(r), -n)\,\mathrm{d}r\ . \tag{3.5.10}$$

Then we have $f_n(u) \searrow u\log(u)$ and $f_n'(u) \searrow 1 + \log(u)$ as $n \to \infty$. For $\mu = \rho m \in \mathscr{P}(\mathbb{R}^d)$ we set $\mathcal{H}_n(\mu) := \int f_n(\rho)\mathrm{d}m$. From Assumption 3.5.4 we deduce that ρ satisfies $\partial_t\rho = \mathcal{L}\rho$. Now we calculate

$$\begin{aligned}
\mathcal{H}_n(P_t[\mu]) - \mathcal{H}_n(P_s[\mu]) \;&=\; \int f_n(\rho_t) - f_n(\rho_s)\,\mathrm{d}m \\
&=\; \int\int_s^t f_n'(\rho_r)\partial_r\rho_r\,\mathrm{d}r\mathrm{d}m \;=\; \int\int_s^t f_n'(\rho_r)\mathcal{L}\rho_r\,\mathrm{d}r\mathrm{d}m \\
&=\; -\frac{1}{2}\int_s^t\int \overline{\nabla} f_n'(\rho_r)\overline{\nabla}\rho_r\,\mathrm{d}(Jm)\mathrm{d}r\ .
\end{aligned}$$

The interchange of integrals in the second line is justified since $f_n'(\rho_r)$ is bounded and $\mathcal{L}\rho_r(x)$ is integrable in $(s,t)\times\mathbb{R}^d$. The latter follows from the fact that (3.5.6) holds with ψ replaced by ρ. The integration by parts in the last line can be justified by using again (3.5.6) and (3.5.5).

Letting finally $n\to\infty$, we obtain (3.5.9) by monotone convergence of both the left and right hand sides. □

We will now show that the semigroup (P_t) is the gradient flow of the relative entropy with respect to the distance $\mathcal{W}$ in the sense of Definition 3.5.2. Our strategy of proof is inspired by an argument developed in [DS08] and used in a similar form in [DNS09, Thm. 5.29]. The following two results are a restatement of Theorem 3.1.3 from the introduction.

Theorem 3.5.9. *Let $\mu\in\mathscr{P}^*$. Then $P_t[\mu]\in\mathscr{P}_\mu$ and $\mathcal{H}(P_t[\mu])<\infty$ for all $t>0$ and the map $t\mapsto\mathcal{H}(\mu_t)$ is non-increasing. Moreover, for any $\sigma\in\mathscr{P}_\mu$ the Evolution Variational Inequality holds:*

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}(P_t[\mu],\sigma)^2+\mathcal{H}(P_t[\mu])\ \le\ \mathcal{H}(\sigma)\quad\forall t>0\ . \tag{3.5.11}$$

Proof. Finiteness and monotonicity of $\mathcal{H}(P_t[\mu])$ were already proved in Proposition 3.5.8. In order to estimate $\mathcal{W}(P_t[\mu],\mu)$, denote by ρ_t the density of $P_t[\mu]$ w.r.t. m and set $\mu_s=P_s[\mu]$ and $\boldsymbol{\nu}_s(\mathrm{d}x,\mathrm{d}y)=\overline{\nabla}\rho_sJ(x,\mathrm{d}y)m(\mathrm{d}x)$ for $s\in[0,t]$. Then $(\mu,\boldsymbol{\nu})\in\mathcal{CE}_t(\mu,P_t[\mu])$ and thus by (3.5.5)

$$\mathcal{W}(P_t[\mu],\mu)^2\ \le\ \int_0^t\sqrt{\mathcal{A}(\mu_s,\boldsymbol{\nu}_s)}\mathrm{d}s\ =\ \int_0^t\sqrt{\mathcal{I}(\mu_s)}\mathrm{d}s\ <\ \infty\ .$$

To prove the second statement, it is sufficient by the semigroup property of P_t to assume $\mathcal{H}(\mu)<\infty$ and prove the inequality at $t=0$. So let $\sigma\in\mathscr{P}_\mu$ with $\mathcal{H}(\sigma)<\infty$ and let $(\sigma_s,\boldsymbol{\nu}_s)_{s\in[0,1]}$ be a minimising curve connecting $\sigma_0=\sigma$ to $\sigma_1=\mu$. We set

$$\begin{aligned}\mu^\varepsilon_{s,t}\ &=\ \rho^\varepsilon_{s,t}m\ :=\ P_{st+\varepsilon}[\sigma_s]\quad\text{and}\\ \widetilde{\boldsymbol{\nu}}^\varepsilon_{s,t}\ &=\ \widetilde{v}^\varepsilon_{s,t}Jm\ :=\ P_{st+\varepsilon}[\boldsymbol{\nu}_s]\ .\end{aligned}$$

The couple $(\mu^\varepsilon_{s,t},\widetilde{\boldsymbol{\nu}}^\varepsilon_{s,t})$ does not satisfy the continuity equation. Hence we make the correction

$$\boldsymbol{\nu}^\varepsilon_{s,t}=v^\varepsilon_{s,t}Jm:=(\widetilde{v}^\varepsilon_{s,t}-t\overline{\nabla}\rho^\varepsilon_{s,t})Jm\ .$$

We will need the following result whose proof we postpone for the moment.

Claim 1. We have $(\mu^\varepsilon_{\cdot,t},\boldsymbol{\nu}^\varepsilon_{\cdot,t})\in\mathcal{CE}_1(P_\varepsilon[\sigma],P_{t+\varepsilon}[\mu])$ and moreover,

$$\mathcal{H}(P_{\varepsilon+t}[\mu])-\mathcal{H}(P_\varepsilon[\sigma])\ =\ -\frac{1}{2}\int_0^1\int\overline{\nabla}\log\rho^\varepsilon_{s,t}\mathrm{d}\boldsymbol{\nu}^\varepsilon_{s,t}\mathrm{d}s\ . \tag{3.5.12}$$

From the definition of the distance $\mathcal{W}$ we now obtain the estimate

$$\mathcal{W}(P_{t+\varepsilon}[\mu], P_\varepsilon[\sigma])^2 \;\le\; \int_0^1 \mathcal{A}(\mu^\varepsilon_{s,t}, \boldsymbol{\nu}^\varepsilon_{s,t})\,\mathrm{d}s\;. \tag{3.5.13}$$

Recall the notation $\hat\rho(x,y) = \theta(\rho(x),\rho(y))$ with θ being the logarithmic mean here. We can further estimate

$$\begin{aligned}
\mathcal{A}(\mu^\varepsilon_{s,t}, \boldsymbol{\nu}^\varepsilon_{s,t}) \;&=\; \int \frac{\left|v^\varepsilon_{s,t}\right|^2}{2\hat\rho^\varepsilon_{s,t}}\,\mathrm{d}(Jm)\\
&=\; \int \Big(\left|\widetilde v^\varepsilon_{s,t}\right|^2 - 2t\overline\nabla\rho^\varepsilon_{s,t}v^\varepsilon_{s,t} - t^2\left|\overline\nabla\rho^\varepsilon_{s,t}\right|^2\Big)\frac{1}{2\hat\rho^\varepsilon_{s,t}}\,\mathrm{d}(Jm)\\
&\le\; \mathcal{A}(\mu^\varepsilon_{s,t}, \widetilde{\boldsymbol{\nu}}^\varepsilon_{s,t}) - t\int \overline\nabla\log\rho^\varepsilon_{s,t}v^\varepsilon_{s,t}\,\mathrm{d}(Jm)\\
&\le\; \mathcal{A}(\sigma_s, \boldsymbol{\nu}_s) - t\int \overline\nabla\log\rho^\varepsilon_{s,t}\mathrm{d}\boldsymbol{\nu}^\varepsilon_{s,t}\;,
\end{aligned}$$

where we have dropped the quadratic term in t and used the monotonicity under convolution (Proposition 3.2.8) in the last inequality. Integration over s from 0 to 1 and using (3.5.12) gives

$$\frac{1}{2}\mathcal{W}(P_{t+\varepsilon}[\mu], P_\varepsilon[\sigma])^2 \;\le\; \frac{1}{2}\mathcal{W}(\mu,\sigma)^2 - t\cdot\big(\mathcal{H}(P_{t+\varepsilon}[\mu]) - \mathcal{H}(P_\varepsilon[\sigma])\big)\;.$$

By lower semicontinuity of $\mathcal{W}$ (see Theorem 3.4.4) and continuity of $\mathcal{H}$ along the semigroup we can take the limit $\varepsilon \to 0$ and obtain

$$\frac{1}{2}\mathcal{W}(P_t[\mu], \sigma)^2 \;\le\; \frac{1}{2}\mathcal{W}(\mu,\sigma)^2 - t\cdot\big(\mathcal{H}(P_t[\mu]) - \mathcal{H}(\sigma)\big)\;.$$

Finally, rearranging terms and letting $t \searrow 0$ yields (3.5.11).

Proof of Claim 1. For the proof we first need two estimates. First, note that

$$\int_0^1 \mathcal{I}(\mu^\varepsilon_{s,t})\,\mathrm{d}s < \infty\;. \tag{3.5.14}$$

Indeed, by convexity of the map $(u,v)\mapsto(u-v)(\log u - \log v)$ we have the estimate $\mathcal{I}(\mu * \psi_t) \le \mathcal{I}(\psi_t m)$ for every $\mu \in \mathscr{P}(\mathbb{R}^d)$. Hence, we conclude from Proposition 3.5.8 that

$$\int_0^1 \mathcal{I}(\mu^\varepsilon_{s,t})\,\mathrm{d}s \;\le\; \int_0^1 \mathcal{I}(\psi_{\varepsilon+st}m)\,\mathrm{d}s \;=\; \mathcal{H}(\psi_\varepsilon m) - \mathcal{H}(\psi_{\varepsilon+t}m) \;<\; \infty\;.$$

From this we conclude that the curve $(\mu^\varepsilon_{\cdot,t}, \boldsymbol{\nu}^\varepsilon_{\cdot,t})$ has finite action. Indeed,

$$\begin{aligned} A \;&:=\; \int_0^1 \int \frac{\left|v^\varepsilon_{s,t}\right|^2}{2\hat\rho^\varepsilon_{s,t}} \,\mathrm{d}(Jm)\,\mathrm{d}s \\ &\le\; \int_0^1 \int 2\frac{\left|\widetilde{v}^\varepsilon_{s,t}\right|^2}{2\hat\rho^\varepsilon_{s,t}} + 2t^2\frac{\left|\overline{\nabla}\rho^\varepsilon_{s,t}\right|^2}{2\hat\rho^\varepsilon_{s,t}} \,\mathrm{d}(Jm)\,\mathrm{d}s \\ &\le\; 2\int_0^1 \mathcal{A}(\sigma_s, \boldsymbol{\nu}_s)\mathrm{d}s + 2t^2 \int_0^1 \mathcal{I}(\mu^\varepsilon_{s,t})\,\mathrm{d}s \;<\; \infty\;, \end{aligned}$$

where we use Proposition 3.2.8 in the last inequality. Using Lemma 3.2.6 and the previous estimate we see that $\boldsymbol{\nu}^\varepsilon_{\cdot,t}$ satisfies the integrability condition (iv) in Definition 3.3.2. The other conditions are also easily checked. Hence, we see that $(\mu^\varepsilon_{\cdot,t}, \boldsymbol{\nu}^\varepsilon_{\cdot,t}) \in \mathcal{CE}_1(P_\varepsilon[\sigma], P_{\varepsilon+t}[\mu])$.

Now let us prove (3.5.12). By a simple convolution argument we can assume that $\rho^\varepsilon_{s,t}$ is differentiable in s. Let f_n be the function defined by (3.5.10) and set $f(u) = u\log(u)$ for $u \ge 0$. Now we calculate

$$\mathcal{H}_n(P_{\varepsilon+t}[\mu]) - \mathcal{H}_n(P_\varepsilon[\sigma]) \;=\; \int\int_0^1 f_n'(\rho^\varepsilon_{s,t})\partial_s\rho^\varepsilon_{s,t}\,\mathrm{d}s\,\mathrm{d}m\;.$$

Note that the map $x \mapsto f_n'(\rho^\varepsilon_{s,t}(x))$ is bounded and Lipschitz uniformly in $s \in [0,1]$. Using the integrability condition (iv) from Definition 3.3.2 we can approximate it by functions in $C_c^\infty((0,1)\times\mathbb{R}^d)$ and obtain by the continuity equation

$$\mathcal{H}_n(P_{\varepsilon+t}[\mu]) - \mathcal{H}_n(P_\varepsilon[\sigma]) \;=\; -\frac{1}{2}\int_0^1\int \overline{\nabla} f_n'(\rho^\varepsilon_{s,t})\mathrm{d}\boldsymbol{\nu}^\varepsilon_{s,t}\mathrm{d}s\;. \tag{3.5.15}$$

By monotone convergence the left hand side of (3.5.15) converges to the left hand side of (3.5.12). It remains to prove convergence of the right hand side. Using Hölder inequality, we estimate

$$\begin{aligned} &\left|\int_0^1\int \overline{\nabla}(f'(\rho^\varepsilon_{s,t}) - f_n'(\rho^\varepsilon_{s,t}))\mathrm{d}\boldsymbol{\nu}^\varepsilon_{s,t}\mathrm{d}s\right| \\ &\le\; \int_0^1\int \left|\overline{\nabla}(f'(\rho^\varepsilon_{s,t}) - f_n'(\rho^\varepsilon_{s,t}))\right|\left|w^\varepsilon_{s,t}\right|\mathrm{d}(Jm)\mathrm{d}s \\ &\le\; A^{\frac{1}{2}}\left(\int_0^1\int \left|\overline{\nabla}(f'(\rho^\varepsilon_{s,t}) - f_n'(\rho^\varepsilon_{s,t}))\right|^2 2\hat\rho^\varepsilon_{s,t}\mathrm{d}(Jm)\mathrm{d}s\right)^{\frac{1}{2}}\;. \end{aligned}$$

The integrand in the last term is bounded as

$$\left|\overline{\nabla}(f'(\rho^\varepsilon_{s,t}) - f_n'(\rho^\varepsilon_{s,t}))\right|^2\hat\rho^\varepsilon_{s,t} \;\le\; \left|\overline{\nabla} f'(\rho^\varepsilon_{s,t})\right|^2\hat\rho^\varepsilon_{s,t} \;=\; \overline{\nabla}\log\rho^\varepsilon_{s,t}\overline{\nabla}\rho^\varepsilon_{s,t}\;.$$

With the help of (3.5.14) and dominated convergence we conclude convergence of the right hand side of (3.5.15) to the right hand side of (3.5.12). □

□

Corollary 3.5.10. *The entropy is convex along $\mathcal{W}$-geodesics. More precisely, let $\mu_0, \mu_1 \in \mathscr{P}^*$ such that $\mathcal{W}(\mu_0, \mu_1) < \infty$ and let $(\mu_t)_{t\in[0,1]}$ be a geodesic connecting μ_0 and μ_1. Then we have*

$$\mathcal{H}(\mu_t) \ \leq \ (1-t)\mathcal{H}(\mu_0) + t\mathcal{H}(\mu_1) \ .$$

Proof. This is a direct consequence of Theorem 3.5.9 and the fact, proved in [DS08, Thm. 3.2], that in a general setting the Evolution Variational Inequality implies geodesic convexity. □

We finish by giving an equivalent and more intuitive definition of the distance $\mathcal{W}$ in the present setting of a translation invariant jump kernel J. We show that it coincides with $\widetilde{\mathcal{W}}$ defined in (3.1.3). We introduce the following shorthand notation. Given functions $\rho : \mathbb{R}^d \to \mathbb{R}_+$ and $\psi : \mathbb{R}^d \to \mathbb{R}$ we write

$$\mathcal{A}'(\rho, \psi) := \frac{1}{2}\int \big(\psi(y) - \psi(x)\big)^2 \hat{\rho}(x,y) J(x, \mathrm{d}y) m(\mathrm{d}x) \ .$$

For two probability densities $\bar{\rho}_0, \bar{\rho}_1$ w.r.t. m and $T > 0$ let us denote by $\mathcal{CE}'_T(\bar{\rho}_0, \bar{\rho}_1)$ the collection of pairs (ρ, ψ) satisfying the following conditions:

$$\begin{cases} (i) & \rho : [0,T] \times \mathbb{R}^d \to \mathbb{R}_+ \text{ is measurable ;} \\ (ii) & \rho_t \text{ is a probability density for all } t \in [0,T] \text{ ;} \\ (iii) & \text{The curve } t \mapsto \mu_t := \rho_t m \text{ is weakly continuous ;} \\ (iv) & \psi : [0,T] \times \mathbb{R}^d \to \mathbb{R} \text{ is measurable ;} \\ (v) & \partial_t \rho_t + \overline{\nabla} \cdot (\hat{\rho}_t \overline{\nabla} \psi_t) \ = \ 0 \ , \ \rho_0 = \bar{\rho}_0 \ , \ \rho_T = \bar{\rho}_1 \ . \end{cases} \tag{3.5.16}$$

Here the continuity equation (v) is understood in the sense that for every test function $\varphi \in C_c^\infty((0,T) \times \mathbb{R}^d)$ we have

$$\int_0^1 \int \partial_t \varphi \rho_t \mathrm{d}m \mathrm{d}t + \frac{1}{2}\int_0^1 \int \overline{\nabla}\varphi(x,y) \overline{\nabla}\psi_t(x,y) \hat{\rho}_t(x,y) J(x, \mathrm{d}y) m(\mathrm{d}y) \mathrm{d}t \ = \ 0 \ .$$

Proposition 3.5.11. *In addition to Assumptions 3.5.3 and 3.5.4 assume that the jump kernel is given as $J(x, \mathrm{d}y) = j(y-x)\mathrm{d}y$ for a function $j : \mathbb{R}^d \setminus \{0\} \to \mathbb{R}^+$ that is strictly positive. Let $\bar{\mu}_i = \bar{\rho}_i m \in \mathscr{P}(\mathbb{R}^d)$ for $i = 0, 1$ such that $\mathcal{I}(\bar{\mu}_i)$ is finite. Then we have*

$$\mathcal{W}(\bar{\mu}_0, \bar{\mu}_1)^2 = \inf \left\{ \int_0^1 \mathcal{A}'(\rho_t, \psi_t) \mathrm{d}t \ : \ (\rho, \psi) \in \mathcal{CE}'_1(\bar{\rho}_0, \bar{\rho}_1) \right\} .$$

Note that the assumptions above on the jump kernel J are satisfied by the kernel J_α associated to the fractional Laplacian.

Proof. The inequality '≤' follows easily by noting that the infimum in the definition of $\mathcal{W}$ is taken over a larger set. Indeed, given a pair $(\rho,\psi)\in\mathcal{CE}'_1(\bar\rho_0,\bar\rho_1)$ such that $\int_0^1\mathcal{A}'(\rho_t,\psi_t)\mathrm{d}t$ is finite we set $\mu_t=\rho_t m$ and define $\boldsymbol{\nu}_t\in\mathcal{M}_{loc}(G)$ by setting $\boldsymbol{\nu}_t(\mathrm{d}x,\mathrm{d}y)=\overline{\nabla}\psi_t(x,y)\hat\rho_t(x,y)J(x,\mathrm{d}y)m(\mathrm{d}x)$. Then we have $\mathcal{A}'(\rho_t,\psi_t)=\mathcal{A}(\mu_t,\boldsymbol{\nu}_t)$ and it is easily checked using Lemma 3.2.6 that $(\mu,\boldsymbol{\nu})\in\mathcal{CE}_1(\bar\mu_0,\bar\mu_1)$.

Let us now prove the opposite inequality '≥'. To this end, note that by a reparametrisation argument similar to Lemma 3.4.2 the square root of the infimum on the right hand side coincides with

$$\inf\left\{\int_0^T\sqrt{\mathcal{A}'(\rho_t,\psi_t)}\mathrm{d}t\ :\ (\rho,\psi)\in\mathcal{CE}'_T(\bar\rho_0,\bar\rho_1)\right\}\ .$$

We set $\mu_t^{i,\varepsilon}:=P_t[\bar\mu_i]=\rho_t^{i,\varepsilon}m$ and $\psi_t^{i,\varepsilon}=\log\rho_t^{i,\varepsilon}$ for $i=0,1$ and $t\in(0,\varepsilon]$. It is easily checked, that the pair $(\rho^{i,\varepsilon},\psi^{i,\varepsilon})$ belongs to $\mathcal{CE}'_\varepsilon(\bar\rho_i,\rho_1^{i,\varepsilon})$. Using the monotonicity of $\mathcal{I}$ under convolution as in the proof of Claim 1 we infer that

$$L^{i,\varepsilon}\ :=\ \int_0^\varepsilon\sqrt{\mathcal{A}'(\rho_t^{i,\varepsilon},\psi_t^{i,\varepsilon})}\mathrm{d}t\ =\ \int_0^\varepsilon\sqrt{\mathcal{I}(\mu_t^{i,\varepsilon})}\mathrm{d}t\ \le\ \varepsilon\sqrt{\mathcal{I}(\bar\mu_i)}\ .$$

Now let $(\mu,\boldsymbol{\nu})\in\mathcal{CE}_1(\bar\mu_0,\bar\mu_1)$ be a minimising curve and set $\mu_t^\varepsilon:=P_\varepsilon[\mu_t]=\rho_t^\varepsilon m$. Proposition 3.4.8 and the proof of Proposition 3.4.7 show that the curve $t\mapsto\mu_t^\varepsilon$ is absolutely continuous w.r.t. $\mathcal{W}$ and thus there is a family of optimal velocity measures $\widetilde{\boldsymbol{\nu}}^\varepsilon$. By Proposition 3.4.10 we have that $\widetilde{\boldsymbol{\nu}}_t^\varepsilon=w_t^\varepsilon\hat\rho_t^\varepsilon Jm$ where w_t^ε belongs to $T_{\rho_t^\varepsilon}$. Note that $\rho_t^\varepsilon>0$ by Assumption 3.5.4 and thus $\hat\rho_t^\varepsilon>0$ for all $t\in(0,1)$ and moreover $j>0$. Hence, it is easily checked that any limit of discrete gradients in L^2 w.r.t. the measure $\hat\rho_t^\varepsilon Jm(\mathrm{d}x,\mathrm{d}y)=\hat\rho_t^\varepsilon(x,y)j(y-x)\mathrm{d}x\mathrm{d}y$ coincides again a.e. with a discrete gradient. Thus we have $w_t^\varepsilon=\overline{\nabla}\psi_t^\varepsilon$ a.e. for a suitable function $\psi^\varepsilon:(0,1)\times\mathbb{R}^d\to\mathbb{R}$. Now observe that $(\rho^\varepsilon,\psi^\varepsilon)\in\mathcal{CE}'_1(\rho_0^\varepsilon,\rho_1^\varepsilon)$ and

$$\begin{aligned}L^\varepsilon\ &:=\ \int_0^1\sqrt{\mathcal{A}'(\rho_t^\varepsilon,\psi_t^\varepsilon)}\mathrm{d}t\ =\ \int_0^1\sqrt{\mathcal{A}(\mu_t^\varepsilon,\widetilde{\boldsymbol{\nu}}_t^\varepsilon)}\mathrm{d}t\\ &\le\ \int_0^1\sqrt{\mathcal{A}(\mu_t,\boldsymbol{\nu}_t)}\mathrm{d}t\ =\ \mathcal{W}(\bar\mu_0,\bar\mu_1)\ ,\end{aligned}$$

where we have used Proposition 3.2.8 in the second line. Finally we concatenate the three curves $(\rho^{0,\varepsilon},\psi^{0,\varepsilon}),(\rho^\varepsilon,\psi^\varepsilon)$ and $(\rho^{1,\varepsilon},\psi^{1,\varepsilon})$ to obtain a curve $(\widetilde\rho^\varepsilon,\widetilde\psi^\varepsilon)\in\mathcal{CE}'_{1+2\varepsilon}(\bar\rho_0,\bar\rho_1)$ which satisfies

$$\begin{aligned}\int_0^{1+2\varepsilon}\sqrt{\mathcal{A}'(\widetilde\rho_t^\varepsilon,\widetilde\psi_t^\varepsilon)}\mathrm{d}t\ &=\ L^{0,\varepsilon}+L^\varepsilon+L^{1,\varepsilon}\\ &\le\ \mathcal{W}(\bar\mu_0,\bar\mu_1)+\varepsilon(\mathcal{I}(\bar\mu_0)+\mathcal{I}(\bar\mu_1))\ .\end{aligned}$$

Letting ε go to zero now yields the claim. □

Bibliography

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008.

[AGS11a] ———, *Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below*, Preprint at arXiv:1106.2090 (2011).

[AGS11b] ———, *Metric measure spaces with Riemannian Ricci curvature bounded from below*, Preprint at arXiv:1109.0222 (2011).

[AL00] C. Ané and M. Ledoux, *On logarithmic Sobolev inequalities for continuous time random walks on graphs*, Probab. Theory Related Fields **116** (2000), no. 4, 573–602.

[App04] D. Applebaum, *Lévy processes and stochastic calculus*, Cambridge studies in advanced mathematics, vol. 93, Cambridge University Press, 2004.

[ASZ09] L. Ambrosio, G. Savaré, and L. Zambotti, *Existence and stability for Fokker-Planck equations with log-concave reference measure*, Probab. Theory Related Fields **145** (2009), no. 3-4, 517–564.

[BB00] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math. **84** (2000), no. 3, 375–393.

[BBCK09] M. Barlow, R. Bass, Z.-G. Chen, and M. Kassmann, *Non-local Dirichlet forms and symmetric jump processes*, Trans. Amer. Math. Soc. **361** (2009), no. 4, 1963–1999.

[BBS11] L. Brasco, G. Buttazzo, and F. Santambrogio, *A Benamou-Brenier approach to branched transport*, SIAM J. Math. Anal. **43** (2011), no. 2, 1023–1040.

[BÉ85] D. Bakry and M. Émery, *Diffusions hypercontractives*, Séminaire de probabilités, XIX, 1983/84, Lecture Notes in Math., vol. 1123, Springer, Berlin, 1985, pp. 177–206.

[Ber96] J. Bertoin, *Lévy processes*, Cambridge Tracts in Mathematics, vol. 121, Cambridge University Press, Cambridge, 1996.

[BG99] S. G. Bobkov and F. Götze, *Discrete isoperimetric and Poincaré-type inequalities*, Probab. Theory Related Fields **114** (1999), no. 2, 245–277.

[BHT06] S. G. Bobkov, C. Houdré, and P. Tetali, *The subgaussian constant and concentration inequalities*, Israel J. Math. **156** (2006), 255–283.

[BJL11] F. Bauer, J. Jost, and S. Liu, *Ollivier-Ricci curvature and the spectrum of the normalized graph Laplace operator*, Preprint at arXiv:1105.3803 (2011).

[BJO09] G. Buttazzo, C. Jimenez, and E. Oudet, *An optimization problem for mass transportation with congested dynamics*, SIAM J. Control Optim. **48** (2009), no. 3, 1961–1976.

[BL98] S. G. Bobkov and M. Ledoux, *On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures*, J. Funct. Anal. **156** (1998), no. 2, 347–365.

[Bre91] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math. **44** (1991), no. 4, 375–417.

[BS09] A.-I. Bonciocat and K.-Th. Sturm, *Mass transportation and rough curvature bounds for discrete spaces*, J. Funct. Anal. **256** (2009), no. 9, 2944–2966.

[BT06] S. G. Bobkov and P. Tetali, *Modified logarithmic Sobolev inequalities in discrete settings*, J. Theoret. Probab. **19** (2006), no. 2, 289–336.

[But89] G. Buttazzo, *Semicontinuity, relaxation and integral representation in the calculus of variations*, Pitman Research Notes in Mathematics Series, Longman Scientific and Technical, Harlow, 1989.

[CDPP09] P. Caputo, P. Dai Pra, and G. Posta, *Convex entropy decay via the Bochner-Bakry-Emery approach*, Ann. Inst. Henri Poincaré Probab. Stat. **45** (2009), no. 3, 734–753.

[CEMS01] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger, *A Riemannian interpolation inequality à la Borell, Brascamp and Lieb*, Invent. Math. **146** (2001), no. 2, 219–257.

[CHLZ12] S.-N. Chow, W. Huang, Y. Li, and H. Zhou, *Fokker-Planck equations for a free energy functional or Markov process on a graph*, Arch. Ration. Mech. Anal. **203** (2012), 969–1008.

[CK03] Z.-Q. Chen and T. Kumagai, *Heat kernel estimates for stable-like processes on d-sets*, Stochastic Process. Appl. **108** (2003), no. 1, 27–62.

[CLSS10] J. A. Carrillo, S. Lisini, G. Savaré, and D. Slepčev, *Nonlinear mobility continuity equations and generalized displacement convexity*, J. Funct. Anal. **258** (2010), no. 4, 1273–1309.

[CM12] E. Carlen and J. Maas, *An analog of the 2-Wasserstein metric in non-commutative probability under which the fermionic Fokker–Planck equation is gradient flow for the entropy*, Preprint at arXiv:1203.5377 (2012).

[CS11] L. Caffarelli and L. Silvestre, *The Evans–Krylov theorem for non local fully non linear equations*, Ann. of Math. **174** (2011), no. 2, 1163–1187.

[DNS09] J. Dolbeault, B. Nazaret, and G. Savaré, *A new class of transport distances between measures*, Calc. Var. Partial Differential Equations **34** (2009), no. 2, 193–231.

[DS08] S. Daneri and G. Savaré, *Eulerian calculus for the displacement convexity in the Wasserstein distance*, SIAM J. Math. Anal. **40** (2008), no. 3, 1104–1122.

[EM12] M. Erbar and J. Maas, *Ricci curvature of finite Markov chains via convexity of the entropy*, Arch. Ration. Mech. Anal., Online first (2012).

[Erb10] M. Erbar, *The heat equation on manifolds as a gradient flow in the Wasserstein space*, Ann. Inst. Henri Poincaré Probab. Stat. **46** (2010), no. 1, 1–23.

[Erb12] ———, *Gradient flows of the entropy for jump processes*, Preprint at arXiv:1204.2190 (2012).

[FSS10] S. Fang, J. Shao, and K.-Th. Sturm, *Wasserstein space over the Wiener space*, Probab. Theory Related Fields **146** (2010), no. 3-4, 535–565.

[GKO10] N. Gigli, K. Kuwada, and S.-I. Ohta, *Heat flow on Alexandrov spaces*, Preprint at arXiv:1008.1319 (2010).

[GL10] N. Gozlan and C. Léonard, *Transport inequalities. A survey.*, Markov Process. Related Fields **16** (2010), no. 4, 635–736.

[GM96] W. Gangbo and R. J. McCann, *The geometry of optimal transportation*, Acta Math. **177** (1996), no. 2, 113–161.

[GSRT12] N. Gozlan, P. Samson, C. Roberto, and P. Tetali, *Displacement of entropy and related inequalities on graphs*, Preprint at arxiv:1207.5116 (2012).

[HJL11] B. Hua, J. Jost, and S. Liu, *Geometric analysis aspects of infinite semiplanar graphs with nonnegative curvature*, Preprint at arXiv:1107.2826 (2011).

[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.

[JL11] J. Jost and S. Liu, *Ollivier's Ricci curvature, local clustering and curvature dimension inequalities on graphs*, Preprint at arXiv:1103.4037 (2011).

[Kan06] L. V. Kantorovich, *On the translocation of masses*, Journal of Mathematical Sciences **133** (2006), no. 4, 1381–1382.

[Led01] M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001.

[LV09] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. **169** (2009), no. 3, 903–991.

[LY10] Y. Lin and S.-T. Yau, *Ricci curvature and eigenvalue estimate on locally finite graphs*, Math. Res. Lett. **17** (2010), no. 2, 343–356.

[Maa11] J. Maas, *Gradient flows of the entropy for finite Markov chains*, J. Funct. Anal. **261** (2011), no. 8, 2250–2292.

[McC97] R. J. McCann, *A convexity principle for interacting gases*, Adv. Math. **128** (1997), no. 1, 153–179.

[McC01] ______, *Polar factorization of maps on Riemannian manifolds*, Geom. Funct. Anal. **11** (2001), no. 3, 589–608.

[Mie11] A. Mielke, *A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems*, Nonlinearity **24** (2011), no. 4, 1329–1346.

[Mie12a] ______, *Dissipative quantum mechanics using GENERIC*, Preprint (2012).

[Mie12b] ______, *Geodesic convexity of the relative entropy in reversible Markov chains*, Calc. Var. Partial Differential Equations, Online first (2012).

[MMS09] D. Matthes, R. J. McCann, and G. Savaré, *A family of nonlinear fourth order equations of gradient flow type*, Comm. Partial Differential Equations **34** (2009), no. 10-12, 1352–1397.

[Mon81] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, De l'Imprimerie Royale, 1781.

[Oll07] Y. Ollivier, *Ricci curvature of metric spaces*, C. R. Math. Acad. Sci. Paris **345** (2007), no. 11, 643–646.

[Oll09] ______, *Ricci curvature of Markov chains on metric spaces*, J. Funct. Anal. **256** (2009), no. 3, 810–864.

[OS09] S.-I. Ohta and K.-Th. Sturm, *Heat flow on Finsler manifolds*, Comm. Pure Appl. Math. **62** (2009), no. 10, 1386–1433.

[Ott01] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174.

[OV00] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), no. 2, 361–400.

[OV10] Y. Ollivier and C. Villani, *A curved Brunn-Minkowski inequality on the discrete hypercube*, Preprint at arXiv:1011.4779 (2010).

[OW05] F. Otto and M. Westdickenberg, *Eulerian calculus for the contraction in the Wasserstein distance*, SIAM J. Math. Anal. **37** (2005), no. 4, 1227–1255.

[RR90] S. T. Rachev and L. Rüschendorf, *A characterization of random variables with minimum L^2-distance*, J. Multivariate Anal. **32** (1990), no. 1, 48–54.

[RR98] ———, *Mass transportation problems. Vol. I*, Probability and its Applications, Springer-Verlag, New York, 1998.

[RS05] M.-K. von Renesse and K.-Th. Sturm, *Transport inequalities, gradient estimates, entropy, and Ricci curvature*, Comm. Pure Appl. Math. **58** (2005), no. 7, 923–940.

[ST09] M. Sammer and P. Tetali, *Concentration on the discrete torus using transportation*, Combin. Probab. Comput. **18** (2009), no. 5, 835–860.

[Stu06] K.-Th. Sturm, *On the geometry of metric measure spaces. I and II*, Acta Math. **196** (2006), no. 1, 65–177.

[Vil03] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.

[Vil09] ———, *Optimal transport, old and new*, Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer-Verlag, Berlin, 2009.