Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2014, Vol. 50, No. 3, 920–945
DOI: 10.1214/12-AIHP537

ANNALES
DE L'INSTITUT
HENRI
POINCARÉ
PROBABILITÉS
ET STATISTIQUES
www.imstat.org/aihp

Gradient flows of the entropy for jump processes

Matthias Erbar[1]

Institute for Applied Mathematics, University of Bonn, Endenicher Allee 60, 53115 Bonn, Germany. E-mail: erbar@iam.uni-bonn.de

Received 4 June 2012; revised 30 October 2012; accepted 6 December 2012

Abstract. 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.

Résumé. On considère une nouvelle distance entre les mesures de probabilité sur $\mathbb{R}^n$. Elle est construite à partir d'un processus de saut par une variante non-locale de la formule de Benamou–Brenier. Pour les processus de Lévy on démontre que le semigroupe engendré par l'opérateur non-local associé est le flot de gradient de l'entropie par rapport à la nouvelle distance. On démontre aussi que l'entropie est convexe le long des géodésiques dans ce cas.

MSC: Primary 60J75; secondary 35S10; 45K05; 49J45; 60G51

Keywords: Jump process; Lévy process; Gradient flow; Entropy; Optimal transport

1. Introduction

In the last two decades the theory of optimal transport has found applications to many areas of mathematics such as partial differential equations, geometry and probability. We refer the reader to the monograph [30] for an overview. In particular, optimal transport has proved very useful in the study of diffusion processes. One of the most striking examples is Otto's discovery [20,26] that many diffusion equations can be interpreted as gradient flows of a suitable free energy functional with respect to the L^2-Wasserstein distance on the space of probability measures. A prominent example is the heat equation which is the gradient flow of the Shannon entropy. By now, similar interpretations of the heat flow have been established in a variety of settings ranging from Riemannian manifolds to abstract metric measure spaces, see [2,15,17,19,25].

The aim of this article 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. We will give a gradient flow interpretation of the equation

$$\partial_t u = \mathcal{L}u, \tag{1.1}$$

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

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

[1]Supported by Bonn International Graduate School in Mathematics.

with a symmetric Lévy measure ν on $\mathbb{R}^d$. Such operators arise naturally as the generators of pure jump Lévy processes. 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 [4,8]. A prominent example of a non-local operator that our results will apply to is the fractional Laplacian $\mathcal{L} = -(-\Delta)^{\alpha/2}$ 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 (1.1) the Wasserstein distance is not appropriate. The main contribution of this article is thus the construction of a new transport distance $\mathcal{W}$ on the space of probability measures that is non-local in nature and allows to interpret equation (1.1) as the gradient flow of the relative entropy. We define this distance via a non-local variant of the dynamical characterization of the Wasserstein distance by Benamou and Brenier [7]. 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 $J(x,\mathrm{d}y) = \nu(\mathrm{d}y - x)$ and identify the evolution equation (1.1) as the gradient flow of the entropy w.r.t. the corresponding distance in the framework of gradient flows in metric spaces developed in [1]. Moreover, we show that the entropy is convex along $\mathcal{W}$-geodesics.

To motivate our interest in such a link between jump processes and optimal transport, let us highlight two observations.

The gradient flow approach has been used as a powerful tool in the study of many evolution partial differential equations. Already in Otto's original work [26] convexity properties of the entropy functional have been used to derive explicit rates of convergence to equilibrium for the porous medium equation. This approach is also well adapted to the study of functional inequalities, such as logarithmic Sobolev inequalities (see e.g. the famous result by Otto–Villani [27]). Recently, it has been shown that the gradient flow characterization provides a good framework to study stability properties of diffusion processes under changes of the driving potential or the underlying geometry [3,18].

The regularity theory for elliptic and parabolic equations involving non-local operators is under active development including both analytic and probabilistic approaches (see e.g. [6,10] 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 [5]. Equivalently, such curvature information can be encoded into convexity properties of the entropy along Wasserstein geodesics. In fact, geodesic convexity of the entropy has been used as a synthetic notion of a lower Ricci curvature bound for metric measure spaces by Lott–Villani [21] and Sturm [28,29]. In this sense the approach presented here could be used to define an alternative notion of curvature in the spirit of Lott–Villani–Sturm that might be more adapted to certain situations than the non-local Γ^2-calculus. In the discrete setting of finite Markov chains, this approach has already been used in [16] to derive new functional inequalities.

Modifications of the Wasserstein distance have been considered recently by a number of authors. In [14] Dolbeault, Nazaret and Savaré proposed a new class of transport distances 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. Very recently, Maas [22] (see also for independent related work by Mielke [24] and Chow et al. [12]) introduced a distance between probability measures on a discrete space equipped with a Markov kernel such that the law of the continuous time Markov chain evolves as the gradient flow of the entropy. Our approach is very similar in spirit to the work of Maas and generalizes it to a certain extend. On the technical side we use an adaptation of the techniques developed in [14] to our non-local setting.

Main results

Let us now discuss the content of this article in more detail. Let $(J(x,\cdot), x \in \mathbb{R}^d)$ be a jump kernel. 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. Further let m be a Radon measure on $\mathbb{R}^d$. Throughout this text J and m shall satisfy the following

Assumption 1.1. *We assume that*

(J1) *J is reversible w.r.t. m, i.e. the measure $J(x,\mathrm{d}y)m(\mathrm{d}x)$ is symmetric.*
(J2) *For every bounded continuous function $f : \mathbb{R}^d \to \mathbb{R}$ the mapping*

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

is again bounded and continuous.

(J3) *The measures $J(x,\cdot)$ are uniformly integrable*, *i.e.*

$$\sup_x \int_{A_R(x)} \big(1\wedge |x-y|^2\big) J(x,\mathrm{d}y) \to 0 \quad \textit{as } R\to\infty,$$

where $A_R(x) := \{y\colon |x-y|<1/R \textit{ or } |x-y|>R\}$.

We fix the shorthand notation Jm to denote the measure given by

$$Jm(\mathrm{d}x,\mathrm{d}y) = J(x,\mathrm{d}y)m(\mathrm{d}x).$$

Remark 1.2. *If the jump kernel is translation invariant*, *i.e.* $J(x+z,A+z)=J(x,A)$ *for all* $x,z\in\mathbb{R}^d$ *and all Borel sets* $A\subset\mathbb{R}^d\setminus\{x\}$, *we can write* $J(x,A)=\nu(A-x)$, *where* $\nu=J(0,\cdot)$. *Note that in this case Assumption* 1.1 *reduces to the requirement that* ν *is a symmetric Lévy measure*, *i.e. it satisfies* $\nu(A)=\nu(-A)$ *for all Borel sets* $A\subset\mathbb{R}^d\setminus\{0\}$ *as well as*

$$\int \big(1\wedge|z|^2\big)\nu(\mathrm{d}z)<\infty.$$

A non-local transport distance

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. The Benamou–Brenier formula [7] asserts that for two probability densities $\bar\rho_0,\bar\rho_1$ on $\mathbb{R}^d$ we have

$$W_2^2(\bar\rho_0,\bar\rho_1) = \inf_{\rho,\psi}\int_0^1\int \big|\nabla\psi_t(x)\big|^2\rho_t(x)\,\mathrm{d}x\,\mathrm{d}t, \tag{1.2}$$

where the infimum 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

$$\begin{cases}\partial_t\rho+\nabla\cdot(\rho\nabla\psi)=0,\\ \rho_0=\bar\rho_0, \qquad \rho_1=\bar\rho_1.\end{cases} \tag{1.3}$$

Here we will define a (pseudo-) metric (i.e. possibly attaining the value $+\infty$) on probability measures by giving a non-local analogue of formulas (1.2) and (1.3). 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 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

$$\theta(s,t)=\frac{s-t}{\log s-\log t} \tag{1.4}$$

but for future use we allow for more generality in the construction. For a function $\psi:\mathbb{R}^d\to\mathbb{R}$ we will denote by $\bar\nabla\psi(x,y)=\psi(y)-\psi(x)$ its discrete gradient. Following the approach of [22] 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

$$\widetilde{\mathcal{W}}(\bar\mu_0,\bar\mu_1)^2 := \inf_{\rho,\psi}\frac12\int_0^1\int\big|\bar\nabla\psi_t(x,y)\big|^2\hat\rho_t(x,y)Jm(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}t, \tag{1.5}$$

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

$$\begin{cases}\partial_t\rho+\bar\nabla\cdot(\hat\rho\bar\nabla\psi)=0,\\ \rho_0=\bar\rho_0, \qquad \rho_1=\bar\rho_1,\end{cases} \tag{1.6}$$

in the sense that for every test function $\varphi \in C_c^\infty(\mathbb{R}^d)$ we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\int \varphi\rho_t(x)m(\mathrm{d}x) - \frac{1}{2}\int \bar\nabla\varphi(x,y)\bar\nabla\psi_t(x,y)\hat\rho_t(x,y)Jm(\mathrm{d}x,\mathrm{d}y) = 0.$$

For the rigorous construction of the new transport distance we will not address the variational problem (1.5) 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.

We replace ρ by a curve $t \mapsto \mu_t = \rho_t m$ in the set of Borel probability measures $\mathscr{P}(\mathbb{R}^d)$ and ψ by a family of signed Radon measures $\boldsymbol{\nu}_t(\mathrm{d}x,\mathrm{d}y) = \bar\nabla\psi_t(x,y)\hat\rho_t(x,y)Jm(\mathrm{d}x,\mathrm{d}y)$ on the set $G = \{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d\colon\ x \neq y\}$. The couple $(\mu, \boldsymbol{\nu})$ now satisfies the linear equation

$$\begin{cases} \partial_t\mu_t + \bar\nabla\cdot\boldsymbol{\nu}_t = 0, \\ \mu_0 = \bar\mu_0, \qquad \mu_1 = \bar\mu_1 \end{cases} \tag{1.7}$$

which we understand in the sense of distributions, i.e. for all test functions $\varphi \in C_c^\infty((0,1)\times\mathbb{R}^d)$:

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

The quantity to be minimized in (1.5) can now be rewritten as

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

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) $(\mu_t, \boldsymbol{\nu}_t)_{t\in[0,1]}$ of the continuity equation (1.7).

Definition. *For $\bar\mu_0, \bar\mu_1 \in \mathscr{P}(\mathbb{R}^d)$ we define*

$$\mathcal{W}(\bar\mu_0,\bar\mu_1)^2 := \inf\left\{\int_0^1 \mathcal{A}(\mu_t,\boldsymbol{\nu}_t)\,\mathrm{d}t\colon\ (\mu,\boldsymbol{\nu}) \in \mathcal{CE}_{0,1}(\bar\mu_0,\bar\mu_1)\right\}.$$

It is unclear whether $\mathcal{W}$ coincides with $\widetilde{\mathcal{W}}$ defined in (1.5) in full generality. However, we will give a positive answer for the more restricted case of a sufficiently regular translation invariant jump kernel (see Proposition 5.14). We can now state the first main result of this article.

Theorem 1.3. *$\mathcal{W}$ 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 $\mathscr{P}_\tau := \{\mu \in \mathscr{P}(\mathbb{R}^d)\colon\ \mathcal{W}(\mu,\tau) < \infty\}$ equipped with the distance $\mathcal{W}$ is a complete geodesic space.*

Gradient flow of the entropy
We now concentrate on a translation invariant jump kernel J. We assume that

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

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 (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. As usual we define the relative entropy of a measure $\mu\in\mathscr{P}(\mathbb{R}^d)$ by

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

if $\mu=\rho m$ is absolutely continuous and $\mathcal{H}(\mu)=+\infty$ otherwise. In the classical local 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 (1.5), (1.6) the gradient flow of the functional $\mathcal{F}$ should be given by the equation

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

If we now consider the relative entropy $\mathcal{H}$ we have $f'(r)=1+\log r$. Taking into account (1.4) we see that the corresponding gradient flow is given by

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

which is a weak formulation of (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 (1.1) as the gradient flow of the entropy in terms of the Evolution Variational Inequality (EVI).

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 (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 5 for a precise statement) we prove the following

Theorem 1.4. *The semigroup P generated by $\mathcal{L}$ is the gradient flow of the relative entropy in the sense that it satisfies the Evolution Variational Inequality* (*EVI*): *For any $\mu\in\mathscr{P}^*=\{\tau\in\mathscr{P}(\mathbb{R}^d):\ \mathcal{H}(\tau)>-\infty\}$ and $\sigma\in\mathscr{P}_\mu\cap\mathscr{P}^*$ we have*

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}^2\big(P_t[\mu],\sigma\big)+\mathcal{H}\big(P_t[\mu]\big)\le\mathcal{H}(\sigma)\quad\forall t>0.\tag{1.8}$$

Moreover 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

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

Among several ways to characterize gradient flows in metric spaces, the EVI is one of the strongest. For example it implies geodesic convexity of the entropy (see [13]). Convexity of the entropy along $\mathcal{W}$-geodesics can be seen as a non-local analogue of McCann's displacement convexity [23], which corresponds to convexity along geodesics of the L^2-Wasserstein distance. For the choice $\nu_\alpha(\mathrm{d}y)=c_\alpha|y|^{-\alpha-d}\,\mathrm{d}y$ with $\alpha\in(0,2)$ and a suitable constant c_α we obtain the following

Corollary 1.5. *The fractional heat equation*

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

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

We expect that a similar result should also hold for semigroups associated to suitable non-homogeneous jump kernels J. It would be desirable to find examples of kernels where the entropy is strictly geodesically convex. This could be exploited to derive new functional inequalities and rates of convergence to equilibrium for the corresponding evolution equation, as has been done in the discrete setting of finite Markov chains in [16]. However, establishing a stronger EVI(κ) in concrete examples does not seem to be an easy task and we will address this question in a forthcoming publication. Moreover, we expect that the approach presented here can be generalized in order to give a gradient flow interpretation to evolution equations associated to Lévy-type operators with both non-local and diffusion part.

Organization of the paper

In Section 2 we study the action functional $\mathcal{A}$ and establish various properties needed in the sequel. Section 3 is devoted to an analysis of the non-local continuity equation (1.7). In Section 4 we define the metric $\mathcal{W}$ and prove Theorem 1.3. Finally, we focus on translation invariant jump kernels and present the proof of Theorem 1.4 in Section 5.

2. The action functional

In this section we introduce and study an action functional on pairs of measures. 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 $\mathcal{M}_{\mathrm{loc}}(G)$ 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 fulfill the following assumptions:

Assumption 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.*

It is easy to check that these assumptions imply

$$\theta(s, t) \leq \frac{s+t}{2} \quad \forall s, t \geq 0. \tag{2.1}$$

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

$$\theta(s, t) = \int_0^1 s^\alpha t^{1-\alpha}\,\mathrm{d}\alpha = \frac{s-t}{\log s - \log t}, \tag{2.2}$$

the latter expression being valid for positive $s \neq t$. 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

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

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

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

The following observation will be useful.

Lemma 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 $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\boldsymbol{\nu} \in \mathcal{M}_{\mathrm{loc}}(G)$. To μ we associate two Radon measures in $\mathcal{M}_{\mathrm{loc}}(G)$ by setting:

$$\mu^1(\mathrm{d}x, \mathrm{d}y) := J(x, \mathrm{d}y)\mu(\mathrm{d}x), \qquad \mu^2(\mathrm{d}x, \mathrm{d}y) := J(y, \mathrm{d}x)\mu(\mathrm{d}y). \tag{2.3}$$

We can always choose a measure $\sigma \in \mathcal{M}_{\mathrm{loc}}(G)$ 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 $\sigma := |\mu^1| + |\mu^2| + |\boldsymbol{\nu}|$. 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

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

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 $Jm \in \mathcal{M}_{\mathrm{loc}}(G)$ the measure given by $Jm(\mathrm{d}x, \mathrm{d}y) = J(x, \mathrm{d}y)m(\mathrm{d}x)$.

Lemma 2.3. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ be absolutely continuous w.r.t. m with density ρ. Further let $\boldsymbol{\nu} \in \mathcal{M}_{\mathrm{loc}}(G)$ 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*

$$\mathcal{A}(\mu, \boldsymbol{\nu}) = \frac{1}{2}\int \big|w(x,y)\big|^2 \hat{\rho}(x,y) Jm(\mathrm{d}x, \mathrm{d}y). \tag{2.4}$$

Proof. Choose $\lambda \in \mathcal{M}_{\mathrm{loc}}(G)$ 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,

$$\mathcal{A}(\mu, \boldsymbol{\nu}) = \int \alpha\big(\widetilde{w}, \widetilde{\rho}^1, \widetilde{\rho}^2\big)\,\mathrm{d}\lambda < \infty. \tag{2.5}$$

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

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

i.e. $\theta(\widetilde{\rho}^1, \widetilde{\rho}^2) = 0$ λ-a.e. on A. Now the finiteness of the integral in (2.5) 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 (2.4) now follows immediately from the homogeneity of α. □

Lemma 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 $\mathcal{P}(\mathbb{R}^d)$ and $\boldsymbol{\nu}_n \rightharpoonup^* \boldsymbol{\nu}$ weakly* in $\mathcal{M}_{\mathrm{loc}}(G)$. Then*

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

Proof. Note that by Assumption 1.1 the weak convergence of μ_n to μ implies the weak* convergence of μ_n^i to μ^i in $\mathcal{M}_{\mathrm{loc}}(G)$ for $i = 1, 2$. Now the claim follows immediately from a general result on integral functionals, Proposition 2.5. □

Proposition 2.5 ([9], Theorem 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*

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

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

The next estimate will be crucial for establishing compactness of families of curves with bounded action in Section 3.

Lemma 2.6. *There exists a constant $C > 0$ such that for all $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\boldsymbol{\nu} \in \mathcal{M}_{\mathrm{loc}}(G)$ we have*:

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

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

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

Proof. To prove the first statement we 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 $A = \{(x, y) | \alpha(w, \rho^1, \rho^2) = \infty\}$ has zero measure with respect to λ. We can now estimate:

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

The last inequality follows, since by the estimate (2.1) and Assumption 1.1 we have:

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

The second statement follows immediately from the first one by noting that

$$a:=\min\bigl\{|x-y|:\ (x,y)\in K\bigr\}>0$$

and estimating

$$|\boldsymbol{\nu}|(K)\le\frac{1}{a}\int\bigl(1\wedge|x-y|\bigr)|\boldsymbol{\nu}|(\mathrm{d}x,\mathrm{d}y).$$

□

Lemma 2.7 (Convexity of the action). *Let $\mu^j\in\mathcal{P}(\mathbb{R}^d)$ and $\boldsymbol{\nu}^j\in\mathcal{M}_{\mathrm{loc}}(G)$ 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*:

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

Proof. Let us fix a reference measure $\lambda\in\mathcal{M}_{\mathrm{loc}}(G)$ 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 $\boldsymbol{\nu}^\tau=w^\tau\lambda$ with $w^\tau=\tau w^1+(1-\tau)w^0$. From the convexity of the action density function α we obtain:

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

□

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

$$J(x-z,A-z)=J(x,A)\quad\forall x,z\in\mathbb{R}^d,A\in\mathcal{B}\bigl(\mathbb{R}^d\bigr).\tag{2.6}$$

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 measurable function $k:\mathbb{R}^d\to\mathbb{R}_+$ satisfying $\int k(z)\,\mathrm{d}z=1$. Given a measure $\mu\in\mathcal{P}(\mathbb{R}^d)$, its convolution is defined as usual by

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

On the other hand, let $\boldsymbol{\nu}\in\mathcal{M}_{\mathrm{loc}}(G)$ be a measure satisfying

$$\int\bigl(1\wedge|x-y|\bigr)|\boldsymbol{\nu}|(\mathrm{d}x,\mathrm{d}y)<\infty.\tag{2.7}$$

Then we define a measure $\boldsymbol{\nu}*k\in\mathcal{M}_{\mathrm{loc}}(G)$ still satisfying (2.7) by setting for all compact sets $K\subset G$

$$(\boldsymbol{\nu}*k)(K):=\int k(z)\boldsymbol{\nu}\left(K-\begin{pmatrix}z\\z\end{pmatrix}\right)\mathrm{d}z.\tag{2.8}$$

In particular, for every continuous 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 2.8. *Assume that J satisfies (2.6) and let k be a convolution kernel. Then for every $\mu \in \mathscr{P}(\mathbb{R}^d)$, $\boldsymbol{\nu} \in \mathcal{M}_{\mathrm{loc}}(G)$ with $\mathcal{A}(\mu,\boldsymbol{\nu}) < \infty$ we have*

$$\mathcal{A}(\mu * k, \boldsymbol{\nu} * k) \le \mathcal{A}(\mu,\boldsymbol{\nu}).$$

Proof. Note that since $\mathcal{A}(\mu,\boldsymbol{\nu})$ is finite, $\boldsymbol{\nu}$ satisfies (2.7) by Lemma 2.6 and $\boldsymbol{\nu} * k$ is defined. Let us introduce the maps $\tau_z : x \mapsto x + z$ for $z \in \mathbb{R}^d$ and let us denote by μ_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 2.7, together with its lower semicontinuity, Lemma 2.4, we see that

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

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 (2.3). Using the invariance property (2.6) it is immediate to check that $\mu^i_z = (\tau_z \times \tau_z)_*\mu^i$ for $i = 1, 2$. Now choose $\lambda \in \mathcal{M}_{\mathrm{loc}}(G)$ 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

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

□

3. A non-local continuity equation

In this section we will consider the continuity equation

$$\partial_t \mu_t + \bar{\nabla}\cdot \boldsymbol{\nu}_t = 0 \quad \text{on } (0,T)\times\mathbb{R}^d. \tag{3.1}$$

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 $\mathcal{M}_{\mathrm{loc}}(G)$ respectively such that

$$\int_0^T \int \big(1 \wedge |x-y|\big)|\boldsymbol{\nu}_t|(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}t < \infty. \tag{3.2}$$

We suppose that (3.1) holds in the sense of distributions. More precisely, we require that for all $\varphi \in C_c^\infty((0,T)\times\mathbb{R}^d)$:

$$\int_0^T \int \partial_t\varphi_t(x)\mu_t(\mathrm{d}x)\,\mathrm{d}t + \frac{1}{2}\int_0^T \int \bar{\nabla}\varphi_t(x,y)\boldsymbol{\nu}_t(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}t = 0. \tag{3.3}$$

Recall that for a function $\varphi : \mathbb{R}^d \to \mathbb{R}$ we denote by $\bar{\nabla}\varphi(x,y) = \varphi(y) - \varphi(x)$ the discrete gradient. Note that $|\bar{\nabla}\varphi(x,y)| \le \|\varphi\|_{C^1}(1 \wedge |x-y|)$. Hence the integrability assumption (3.2) ensures that the second term in (3.3) is well-defined. The following is an adaptation of [1], Lemma 8.1.2.

Lemma 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 $\mathcal{M}_{\mathrm{loc}}(G)$ satisfying* (3.1) *and* (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 $\varphi \in C_c^\infty([0,T]\times\mathbb{R}^d)$ and all $0 \le t_0 \le t_1 \le T$ we have*:

$$\int \varphi_{t_1}\,\mathrm{d}\widetilde{\mu}_{t_1} - \int \varphi_{t_0}\,\mathrm{d}\widetilde{\mu}_{t_0} = \int_{t_0}^{t_1}\int \partial_t\varphi\,\mathrm{d}\mu_t\,\mathrm{d}t + \frac{1}{2}\int_{t_0}^{t_1}\int \bar{\nabla}\varphi\,\mathrm{d}\boldsymbol{\nu}_t\,\mathrm{d}t. \tag{3.4}$$

Proof. Let us set

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

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) shows that the distributional derivative of $\mu_t(\xi)$ is given by

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

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

$$\bigl|\dot{\mu}_t(\xi)\bigr| \le \frac{1}{2}\int |\bar{\nabla}\xi|\,\mathrm{d}|\boldsymbol{\nu}_t| \le \frac{1}{2}\|\xi\|_{C^1}V(t). \tag{3.5}$$

Based on (3.5) we can argue as in [1], Lemma 8.1.2, to obtain existence of a weakly continuous representative $t \mapsto \widetilde{\mu}_t$.

To prove (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

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

Now equation (3.3) implies

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

Thanks to the continuity of $t \mapsto \widetilde{\mu}_t$ we can pass to limit as $\varepsilon \to 0$ and obtain (3.4). □

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

Definition 3.2. *We denote by $\mathcal{CE}_T$ the set of all pairs $(\mu,\boldsymbol{\nu})$ satisfying the following conditions*:

$$\begin{cases} \text{(i)} & \mu : [0,T] \to \mathscr{P}(\mathbb{R}^d) \textit{ is weakly continuous;} \\ \text{(ii)} & (\boldsymbol{\nu}_t)_{t\in[0,T]} \textit{ is a Borel family of measures in } \mathcal{M}_{\mathrm{loc}}(G); \\ \text{(iii)} & \int_0^T \int (1\wedge|x-y|)|\boldsymbol{\nu}_t|(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}t < \infty; \\ \text{(iv)} & \textit{We have in the sense of distributions:} \\ & \partial_t\mu_t + \bar{\nabla}\cdot\boldsymbol{\nu}_t = 0. \end{cases} \tag{3.6}$$

Moreover, we will denote by $\mathcal{CE}_T(\bar{\mu}_0,\bar{\mu}_1)$ the set of pairs $(\mu,\boldsymbol{\nu}) \in \mathcal{CE}_T$ satisfying in addition: $\mu_0 = \bar{\mu}_0$, $\mu_1 = \bar{\mu}_1$.

Remark 3.3. *The continuity equation can also be tested against more general test functions. E.g. let $(\mu,\boldsymbol{\nu}) \in \mathcal{CE}_1$ and let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be bounded and Lipschitz. Approximating φ with functions in $C_c^\infty(\mathbb{R}^d)$ with uniformly bounded C^1-norm and using the integrability assumption* (iii) *in* (3.6) *to pass to the limit in* (3.4) *we obtain*

$$\int \varphi\,\mathrm{d}\mu_1 - \int \varphi\,\mathrm{d}\mu_0 = \int_0^1 \int \bar{\nabla}\varphi\,\mathrm{d}\boldsymbol{\nu}_t\,\mathrm{d}t.$$

A similar reasoning will often be used later on.

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

Proposition 3.4 (Compactness of solutions to the continuity equation). *Let $(\mu^n, \boldsymbol{\nu}^n)$ be a sequence in $\mathcal{CE}_T$ such that $(\mu^n_0)_n$ is tight and*

$$\sup_n \int_0^T \mathcal{A}\big(\mu^n_t, \boldsymbol{\nu}^n_t\big)\,\mathrm{d}t < \infty. \tag{3.7}$$

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

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

Moreover, along this subsequence we have:

$$\int_0^T \mathcal{A}(\mu_t, \boldsymbol{\nu}_t)\,\mathrm{d}t \le \liminf_n \int_0^T \mathcal{A}\big(\mu^n_t, \boldsymbol{\nu}^n_t\big)\,\mathrm{d}t.$$

Proof. For each n define the measure $\boldsymbol{\nu}^n \in \mathcal{M}_{\mathrm{loc}}(G \times (0,T))$ given by $\boldsymbol{\nu}^n(\mathrm{d}x, \mathrm{d}y, \mathrm{d}t) := \boldsymbol{\nu}^n_t(\mathrm{d}x, \mathrm{d}y)\,\mathrm{d}t$. From Lemma 2.6 and (3.7) we infer immediately that

$$\sup_n \int_0^T \int \big(1 \wedge |x-y|\big)\big|\boldsymbol{\nu}^n_t\big|(\mathrm{d}x, \mathrm{d}y)\,\mathrm{d}t < \infty. \tag{3.8}$$

Moreover, arguing exactly as in Lemma 2.6, we obtain that for every compact set $K \subset G$ and every Borel set $B \subset [0,T]$ we have

$$\begin{aligned}
\sup_n \big|\boldsymbol{\nu}^n\big|(K \times B) &\le \sup_n \int_B \big|\boldsymbol{\nu}^n_t\big|(K)\,\mathrm{d}t \qquad (3.9)\\
&\le \sqrt{A}C(K)\sqrt{\mathrm{Leb}(B)}, \qquad (3.10)
\end{aligned}$$

where A denotes the supremum in (3.7). 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 $\mathcal{M}_{\mathrm{loc}}(G \times [0,T])$. The estimate (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 (3.2).

Let $0 \le t_0 \le t_1 \le T$ and $\xi \in C_c^\infty(\mathbb{R}^d)$. We claim that

$$\int_{t_0}^{t_1} \int \bar{\nabla}\xi\,\mathrm{d}\boldsymbol{\nu}^n_t\,\mathrm{d}t \overset{n\to\infty}{\longrightarrow} \int_{t_0}^{t_1} \int \bar{\nabla}\xi\,\mathrm{d}\boldsymbol{\nu}_t\,\mathrm{d}t. \tag{3.11}$$

Note, that $\mathbf{1}_{(t_0,t_1)}\bar{\nabla}\xi$ is not continuous and not compactly supported in $G \times [0,T]$. In order to prove (3.11), we argue by approximation. Let $F \subset \mathbb{R}^d$ be a compact set supporting ξ, then $\bar{\nabla}\xi$ is supported in $N := (F \times \mathbb{R}^d) \cup (\mathbb{R}^d \times F)$. For $R > 0$ we define the sets $A_R := [t_0, t_0 + \frac{1}{R}] \cup [t_1 - \frac{1}{R}, t_1]$ and $D_R := \{(x,y) \in G\colon\ |x-y| < R^{-1}\}$. Moreover, we define the set

$$M_R := \big(D_R \cap (F \times F)\big) \cup \big(B_R^c \times F\big) \cup \big(F \times B_R^c\big),$$

where $B_R = \{x \in \mathbb{R}^d\colon\ |x| < R\}$. For each R sufficiently large we can find a continuous compactly supported function $\varphi_R : G \times [0,T] \to [0,1]$ such that

$$\{\varphi_R < 1\} \cap \big(N \times [t_0, t_1]\big) \subset \big(M_R \times [t_0, t_1]\big) \cup (N \times A_R) =: S_R.$$

The convergence (3.11) holds if we replace $\mathbf{1}_{(t_0,t_1)}\bar{\nabla}\xi$ by the continuous and compactly supported function $\varphi_R \cdot \mathbf{1}_{(t_0,t_1)}\bar{\nabla}\xi$. Thus, it remains to show that

$$\sup_n \left| \int_{t_0}^{t_1} \int (1-\varphi_R)\bar{\nabla}\xi \, \mathrm{d}\boldsymbol{v}_t^n \, \mathrm{d}t \right| \longrightarrow 0,$$

as $R \to \infty$. Arguing as in Lemma 2.6, we estimate

$$\begin{aligned} &\left| \int_{t_0}^{t_1} \int (1-\varphi_R)\bar{\nabla}\xi \, \mathrm{d}\boldsymbol{v}_t^n \, \mathrm{d}t \right| \\ &\quad \le \|\xi\|_{C^1} \int_{S_R} \big(1 \wedge |x-y|\big) \, \mathrm{d}\big|\boldsymbol{v}_t^n\big| \, \mathrm{d}t \\ &\quad \le \|\xi\|_{C^1} \sqrt{A} \left(\int_{S_R} \big(1 \wedge |x-y|^2\big) J(x, \mathrm{d}y) \mu_t^n(\mathrm{d}x) \, \mathrm{d}t \right)^{1/2}. \end{aligned}$$

From Assumption 1.1 we deduce that the integral in the last line goes to zero uniformly in n as $R \to \infty$.

After extraction of another subsequence we can assume $\mu_0^n \rightharpoonup \mu_0$ weakly for some $\mu_0 \in \mathscr{P}(\mathbb{R}^d)$. Using this, the convergence (3.11) and the continuity equation in the form (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 finite non-negative measure $\mu_t \in \mathcal{M}(\mathbb{R}^d)$ for every $t \in [0,T]$. We now argue that μ_t is a probability measure for all t. From the above reasoning we obtain that for any $\xi \in C_c^\infty(\mathbb{R}^d)$ and any $t \in [0,1]$:

$$\int \xi \, \mathrm{d}\mu_t = \int \xi \, \mathrm{d}\mu_0 + \frac{1}{2} \int_0^t \int \bar{\nabla}\xi \, \mathrm{d}\boldsymbol{v}_s \, \mathrm{d}s. \tag{3.12}$$

For $R > 0$ we choose functions $\xi_R \in C_c^\infty(\mathbb{R}^d)$ with $0 \le \xi \le 1$, $\xi = 1$ on B_R and $\|\xi\|_{C^1} \le 1$. Since $\boldsymbol{v}$ satisfies the integrability assumption (3.2) we observe that as $R \to \infty$ we have

$$\left| \int_0^t \int \bar{\nabla}\xi_R \, \mathrm{d}\boldsymbol{v}_s \, \mathrm{d}s \right| \le \int_0^t \int \big(1 \wedge |x-y|\big) \mathbf{1}_{(B_R \times B_R)^c} |\boldsymbol{v}_s|(\mathrm{d}x, \mathrm{d}y) \, \mathrm{d}s \to 0.$$

Hence we deduce from (3.12) that $\mu_t(\mathbb{R}^d) = \mu_0(\mathbb{R}^d) = 1$. It is now easily checked that the couple $(\mu, \boldsymbol{v})$ belongs to $\mathcal{CE}_T$. As in Lemma 2.4 the lower semicontinuity follows from Proposition 2.5 by considering $\int_0^T \mathcal{A}(\mu_t, \boldsymbol{v}_t) \, \mathrm{d}t$ as an integral functional on the space $\mathcal{M}_{\mathrm{loc}}(G \times [0,T])$. $\square$

4. A non-local transport distance

In this section we define the distance $\mathcal{W}$. We will establish various properties, in particular existence of geodesics. Moreover, we will characterize absolutely continuous curves in the metric space $(\mathscr{P}(\mathbb{R}^d), \mathcal{W})$.

Definition 4.1. *For $\bar{\mu}_0, \bar{\mu}_1 \in \mathscr{P}(\mathbb{R}^d)$ we define*

$$\mathcal{W}(\bar{\mu}_0, \bar{\mu}_1)^2 := \inf \left\{ \int_0^1 \mathcal{A}(\mu_t, \boldsymbol{v}_t) \, \mathrm{d}t \colon (\mu, \boldsymbol{v}) \in \mathcal{CE}_1(\bar{\mu}_0, \bar{\mu}_1) \right\}. \tag{4.1}$$

Note that the definition of the distance $\mathcal{W}$ depends on the choice of a jump kernel J satisfying Assumption 1.1 and a function θ satisfying Assumption 2.1. However, we will suppress this dependence in the notation.

Let us give an equivalent characterization of the infimum in (4.1).

Lemma 4.2. *For any $T > 0$ and $\bar{\mu}_0, \bar{\mu}_1 \in \mathscr{P}(\mathbb{R}^d)$ we have*:

$$\mathcal{W}(\bar{\mu}_0, \bar{\mu}_1) = \inf \left\{ \int_0^T \sqrt{\mathcal{A}(\mu_t, \boldsymbol{v}_t)} \, \mathrm{d}t \colon (\mu, \boldsymbol{v}) \in \mathcal{CE}_T(\bar{\mu}_0, \bar{\mu}_1) \right\}. \tag{4.2}$$

Proof. This follows from a standard reparametrization argument. See [1], Lemma 1.1.4, or [14], Theorem 5.4, for details in similar situations. □

The next result shows that the infimum in the definition above is in fact a minimum.

Proposition 4.3. *Let $\bar{\mu}_0, \bar{\mu}_1 \in \mathcal{P}(\mathbb{R}^d)$ be such that $W := \mathcal{W}(\bar{\mu}_0, \bar{\mu}_1)$ is finite. Then the infimum in (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 minimizing curve $(\mu, \boldsymbol{\nu}) \in \mathcal{CE}_1(\bar{\mu}_0, \bar{\mu}_1)$ follows immediately by the direct method taking into account Proposition 3.4. Invoking Lemma 4.2 and Jensen's inequality we see that this curve satisfies

$$\int_0^1 \sqrt{\mathcal{A}(\mu_t, \boldsymbol{\nu}_t)}\, \mathrm{d}t \geq W = \left(\int_0^1 \mathcal{A}(\mu_t, \boldsymbol{\nu}_t)\, \mathrm{d}t\right)^{1/2} \geq \int_0^1 \sqrt{\mathcal{A}(\mu_t, \boldsymbol{\nu}_t)}\, \mathrm{d}t.$$

Hence we must have $\mathcal{A}(\mu_t, \boldsymbol{\nu}_t) = W^2$ for a.e. $t \in [0, T]$. □

We now prove the first main result Theorem 1.3 announced in the Introduction which we recall here for convenience.

Theorem 4.4. *$\mathcal{W}$ defines a (pseudo-) metric on $\mathcal{P}(\mathbb{R}^d)$. 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 $\tau \in \mathcal{P}(\mathbb{R}^d)$ the set $\mathcal{P}_\tau := \{\mu \in \mathcal{P}(\mathbb{R}^d) \colon \mathcal{W}(\mu, \tau) < \infty\}$ 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 (3.4) from Lemma 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 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 minimizing 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 (3.4) we infer $\bar{\mu}_0 = \bar{\mu}_1$.

The compactness assertion and lower semicontinuity of $\mathcal{W}$ follow immediately from Proposition 3.4. These in turn imply that the topology induced by $\mathcal{W}$ is stronger than the weak one.

Let us now fix $\tau \in \mathcal{P}(\mathbb{R}^d)$ and let $\bar{\mu}_0, \bar{\mu}_1 \in \mathcal{P}_\tau$. By the triangle inequality we have $\mathcal{W}(\bar{\mu}_0, \bar{\mu}_1) < \infty$ and hence Proposition 4.3 yields existence of a minimizing 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 $\mathcal{P}_\tau$ 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 $\mathcal{P}_\tau$. In particular the sequence is bounded w.r.t. $\mathcal{W}$ and we can find a subsequence (still indexed by n) and $\mu^\infty \in \mathcal{P}(\mathbb{R}^d)$ such that $\mu^n \rightharpoonup \mu^\infty$ weakly. Invoking lower semicontinuity of $\mathcal{W}$ and the Cauchy condition we infer that $\mathcal{W}(\mu^n, \mu^\infty) \to 0$ as $n \to \infty$ and that $\mu^\infty \in \mathcal{P}_\tau$. □

It is yet unclear when precisely the distance $\mathcal{W}$ is finite. However, we will see in the next section that the distance is finite e.g. along trajectories of the semigroup associated to a translation invariant and sufficiently regular jump kernel.

The following result shows that under certain assumptions the distance $\mathcal{W}$ can be bounded from below by the L^1-Wasserstein distance. Let us fix a Lipschitz function $f : [0, \infty) \to [0, \infty)$ which is non-decreasing and concave with $f(0) = 0$. Then $d(x, y) := f(|x - y|)$ defines a metric on $\mathbb{R}^d$. Recall that the L^1-Wasserstein distance induced by d is defined for $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^d)$ by

$$W_{1,d}(\mu_0, \mu_1) := \inf_\pi \int d(x, y)\pi(\mathrm{d}x, \mathrm{d}y),$$

where the infimum is taken over all probability measures $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ whose first and second marginal are μ_0 and μ_1 respectively (see e.g. [30], Chapter 6).

Proposition 4.5. *Assume that the jump kernel J satisfies*

$$M_d^2 := \sup_x \int f\big(|x-y|\big)^2 J(x, \mathrm{d}y) < \infty. \tag{4.3}$$

Then for any $\mu_0, \mu_1 \in \mathscr{P}(\mathbb{R}^d)$ we have the bound

$$W_{1,d}(\mu_0, \mu_1) \le \frac{M_d}{\sqrt{2}} \mathcal{W}(\mu_0, \mu_1).$$

Proof. We can assume that $\mathcal{W}(\mu_0, \mu_1) < \infty$. Take a minimizing curve $(\mu, \boldsymbol{\nu}) \in \mathcal{CE}_1(\mu_0, \mu_1)$ and let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a bounded function that is 1-Lipschitz w.r.t. d. Since f is Lipschitz, φ is also Lipschitz w.r.t. Euclidean distance. Taking into account Remark 3.3 and arguing as in Lemma 2.6, we estimate

$$\begin{aligned}
&\left| \int \varphi \, \mathrm{d}\mu_1 - \int \varphi \, \mathrm{d}\mu_0 \right| \\
&\quad = \frac{1}{2} \left| \int_0^1 \int \bar{\nabla}\varphi \, \mathrm{d}\boldsymbol{\nu}_t \, \mathrm{d}t \right| \\
&\quad \le \frac{1}{2} \int_0^1 \int d(x, y) |\boldsymbol{\nu}_t|(\mathrm{d}x, \mathrm{d}y) \, \mathrm{d}t \\
&\quad \le \frac{1}{\sqrt{2}} \left(\int_0^1 \mathcal{A}(\mu_t, \boldsymbol{\nu}_t) \, \mathrm{d}t \right)^{1/2} \left(\int_0^1 \int d(x, y)^2 J(x, \mathrm{d}y) \mu_t(\mathrm{d}x) \, \mathrm{d}t \right)^{1/2} \\
&\quad \le \frac{M_d}{\sqrt{2}} \mathcal{W}(\mu_0, \mu_1).
\end{aligned}$$

Taking the supremum over all bounded 1-Lipschitz functions φ yields the claim by Kantorovich–Rubinstein duality (see [30], Theorems 5.10, 5.16). □

Remark 4.6. *We highlight two special cases of the previous result.*

(i) *Let d_* denote the bounded metric on $\mathbb{R}^d$ given by $d_*(x, y) = 1 \wedge |x-y|$. Then Assumption 1.1 guarantees that $M_{d_*} < \infty$ and we can always bound $\mathcal{W}$ from below by W_{1,d_*}.*
(ii) *Let $\alpha \in (0, 2)$ and consider the α-stable jump kernel $J_\alpha(x, \mathrm{d}y) = |x-y|^{-\alpha-d}\,\mathrm{d}y$. Then for any $\beta < \frac{\alpha}{2}$ the corresponding distance $\mathcal{W}_\alpha$ can be bounded below by W_{1,d_β} where d_β is the metric given by $d_\beta(x, y) = |x-y| \wedge |x-y|^\beta$.*

The convexity and monotonicity properties of the action functional established in Section 2 extend naturally to the distance function.

Proposition 4.7 (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*:

$$\mathcal{W}\big(\mu_0^\tau, \mu_1^\tau\big)^2 \le \tau \mathcal{W}\big(\mu_0^1, \mu_1^1\big)^2 + (1-\tau)\mathcal{W}\big(\mu_0^0, \mu_1^0\big)^2.$$

Proof. We can assume that $\mathcal{W}(\mu_0^j, \mu_1^j)$ is finite and choose minimizing 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 $(\mu^\tau, \boldsymbol{\nu}^\tau)_t \in \mathcal{CE}_1(\mu_0^\tau, \mu_1^\tau)$. From the definition of $\mathcal{W}$ and the convexity of $\mathcal{A}$ as stated in Lemma 2.7 we infer

$$\begin{aligned}
\mathcal{W}\big(\mu_0^\tau, \mu_1^\tau\big)^2 &\le \int_0^1 \mathcal{A}\big(\mu_t^\tau, \boldsymbol{\nu}_t^\tau\big) \, \mathrm{d}t \le \int_0^1 \big[\tau\mathcal{A}\big(\mu_t^1, \boldsymbol{\nu}_t^1\big) + (1-\tau)\mathcal{A}\big(\mu_t^0, \boldsymbol{\nu}_t^0\big)\big] \, \mathrm{d}t \\
&= \tau\mathcal{W}\big(\mu_0^1, \mu_1^1\big)^2 + (1-\tau)\mathcal{W}\big(\mu_0^0, \mu_1^0\big)^2.
\end{aligned}$$

□

Proposition 4.8 (Monotonicity under convolution). *Let $\mu_0, \mu_1 \in \mathscr{P}(\mathbb{R}^d)$. Assume that J satisfies (2.6) and let m be Lebesgue measure. Let k be a convolution kernel. Then we have*

$$\mathcal{W}(\mu_0 * k, \mu_1 * k) \le \mathcal{W}(\mu_0, \mu_1).$$

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 proof. Let $(\mu, \boldsymbol{\nu}) \in \mathcal{CE}_1(\mu_0, \mu_1)$ be a minimizing curve according to Proposition 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 (3.6) 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 (2.8) we obtain

$$\begin{aligned}
\int \partial_t \varphi \,\mathrm{d}\widetilde{\mu}_t \,\mathrm{d}t &= \int \partial_t \varphi(t, x+z) k(z)\,\mathrm{d}z \mu_t(\mathrm{d}x)\,\mathrm{d}t \\
&= \int \partial_t \widetilde{\varphi} \,\mathrm{d}\mu_t \,\mathrm{d}t = -\frac{1}{2} \int \bar{\nabla} \widetilde{\varphi} \,\mathrm{d}\boldsymbol{\nu}_t \,\mathrm{d}t \\
&= -\frac{1}{2} \int \bar{\nabla} \varphi(t, x+z, y+z) k(z) \boldsymbol{\nu}_t(\mathrm{d}x, \mathrm{d}y)\,\mathrm{d}z\,\mathrm{d}t \\
&= -\frac{1}{2} \int \bar{\nabla} \varphi \,\mathrm{d}\widetilde{\boldsymbol{\nu}}_t \,\mathrm{d}t.
\end{aligned}$$

Now the first assertion follows immediately from Proposition 2.8. This in turn together with weak lower semicontinuity of $\mathcal{W}$ (see Theorem 4.4) yields the second assertion. □

We now give a characterization 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 $g \in L^1(0,T)$ such that

$$\mathcal{W}(\mu_s, \mu_t) \le \int_s^t g(r)\,\mathrm{d}r \quad \forall 0 \le s \le t \le T. \tag{4.4}$$

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 g in (4.4), see [1], Theorem 1.1.2.

Proposition 4.9 (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*

$$\int_0^T \sqrt{\mathcal{A}(\mu_t, \boldsymbol{\nu}_t)}\,\mathrm{d}t < \infty.$$

In this case we have $|\mu_t'|^2 \le \mathcal{A}(\mu_t, \boldsymbol{\nu}_t)$ 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

$$|\mu_t'|^2 = \mathcal{A}(\mu_t, \widetilde{\boldsymbol{\nu}}_t) \quad \textit{for a.e. } t \in [0,T]. \tag{4.5}$$

Proof. The proof follows from the very same arguments as in [14], Theorem 5.17. □

We can describe the optimal velocity measures $\widetilde{\boldsymbol{\nu}}_t$ appearing in the preceding proposition in more detail. We define

$$T_\mu\mathcal{P}\big(\mathbb{R}^d\big) := \big\{\boldsymbol{\nu}\in\mathcal{M}_{\mathrm{loc}}(G)\colon\ \mathcal{A}(\mu,\boldsymbol{\nu})<\infty, \mathcal{A}(\mu,\boldsymbol{\nu})\le\mathcal{A}(\mu,\boldsymbol{\nu}+\boldsymbol{\eta})\ \forall\boldsymbol{\eta}\colon\ \bar{\nabla}\cdot\boldsymbol{\eta}=0\big\}. \tag{4.6}$$

Here $\bar{\nabla}\cdot\boldsymbol{\eta}=0$ is understood in a weak sense, i.e.

$$\frac{1}{2}\int\bar{\nabla}\xi(x,y)\boldsymbol{\eta}(\mathrm{d}x,\mathrm{d}y)=0 \quad \forall\xi\in C_c^\infty\big(\mathbb{R}^d\big).$$

Corollary 4.10. *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 (4.5) if and only if $\boldsymbol{\nu}_t\in T_{\mu_t}\mathcal{P}(\mathbb{R}^d)$ for a.e. $t\in[0,T]$.*

If μ is absolutely continuous with respect to m we can give an explicit description of $T_\mu\mathcal{P}(\mathbb{R}^d)$ as a subspace of an L^2 space.

Proposition 4.11. *Let $\mu=\rho m\in\mathcal{P}(\mathbb{R}^d)$. Then we have $\boldsymbol{\nu}\in T_\mu\mathcal{P}(\mathbb{R}^d)$ if and only if $\boldsymbol{\nu}=w\hat{\rho}Jm$ is absolutely continuous w.r.t. the measure $\hat{\rho}Jm$ and*

$$w\in\overline{\big\{\bar{\nabla}\varphi|\varphi\in C_c^\infty\big(\mathbb{R}^d\big)\big\}}^{L^2(\hat{\rho}Jm)} =: T_\rho.$$

Proof. If $\mathcal{A}(\mu,\boldsymbol{\nu})$ is finite we infer from Lemma 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 (4.6) is equivalent to

$$\|w\|_{L^2(\hat{\rho}Jm)}\le\|w+v\|_{L^2(\hat{\rho}Jm)} \quad \forall v\in N_\rho,$$

where $N_\rho := \{v\in L^2(\hat{\rho}Jm)\colon \int\bar{\nabla}\xi v\hat{\rho}\,\mathrm{d}Jm=0\ \forall\xi\in C_c^\infty(\mathbb{R}^d)\}$. 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 $T_\mu\mathcal{P}(\mathbb{R}^d)$ as the tangent space to $\mathcal{P}(\mathbb{R}^d)$ at the measure μ. This is reminiscent of Otto's Riemannian interpretation of the L^2-Wasserstein space [26]. The results obtained here are in close analogy to the notion of tangent bundle to the Wasserstein space studied in [1], Section 8.4.

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 5.1. *Throughout this section 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 [1] for a comprehensive study of gradient flows in metric spaces. We adopt the following

Definition 5.2. *Let (X,d) be a metric space and $F:X\to(-\infty,\infty]$ a lower semicontinuous function such that its proper domain $D(F):=\{x\in X|F(x)<\infty\}$ is dense in X. Further let $(S_t)_{t\ge0}$ be a strongly continuous semigroup on (X,d) 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 $u\in X$, $v\in D(F)$, $t>0$:*

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}d^2\big(S_t(u),v\big)+\frac{\lambda}{2}d^2\big(S_t(u),v\big)+F\big(S_t(u)\big)\le F(v). \tag{5.1}$$

Here and in the following we will use the notation

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

We will only consider translation invariant jump kernels. More precisely, from now on we make the following

Assumption 5.3. *Assume that m is Lebesgue measure on $\mathbb{R}^d$ and that J satisfies*

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

Recall from Remark 1.2 that Assumption 1.1 implies that the measure $\nu := J(0,\cdot)$ is a symmetric Lévy measure, i.e. it satisfies $\nu(A) = \nu(-A)$ for all $A \subset \mathbb{R}^d \setminus \{0\}$ as well as

$$C_\nu := \int \big(1 \wedge |z|^2\big)\nu(\mathrm{d}z) < \infty. \tag{5.2}$$

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 $\delta u(x,z) := \frac{1}{2}(u(x+z)+u(x-z)-2u(x))$.

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. [4] or [8] 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

$$\int \mathrm{e}^{\mathrm{i}x\cdot\xi} q_t(\mathrm{d}x) = \mathbb{E}\big[\exp\big(\mathrm{i}\langle \xi, X_t\rangle\big)\big] = \exp\big(-t\eta(\xi)\big),$$

where η is given by the Lévy–Khintchine formula:

$$\eta(\xi) = \int \mathrm{e}^{\mathrm{i}\langle y,\xi\rangle} - 1 - \mathrm{i}\langle y, \xi\rangle \mathbf{1}_{\{|y|\leq 1\}}\nu(\mathrm{d}y).$$

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. Recall that we denote by $Jm \in \mathcal{M}_{\mathrm{loc}}(G)$ 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

$$\mathcal{I}(\mu) := \begin{cases} \frac{1}{2}\int \bar\nabla\rho \bar\nabla \log\rho \,\mathrm{d}Jm, & \text{if } \mu = \rho m \text{ and } \rho > 0, \\ +\infty, & \text{else.} \end{cases} \tag{5.3}$$

The following observation will be useful

Lemma 5.4. *Let $\mu, \tau \in \mathscr{P}(\mathbb{R}^d)$ such that $\mathcal{I}(\mu) < \infty$. Then we have*

$$\mathcal{I}(\mu * \tau) \leq \mathcal{I}(\mu).$$

Proof. This follows from convexity of the map $(r,s) \mapsto (r-s)(\log r - \log s)$ by an application of Jensen's inequality. □

Throughout this section we will make the following assumption on ν in terms of the law of the associated Lévy process.

Assumption 5.5. *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

$$\mathcal{H}(q_t) \in (-\infty,\infty), \qquad \mathcal{I}(q_t) < \infty \quad \forall t > 0, \tag{5.4}$$

$$\int_0^t \sqrt{\mathcal{I}(q_s)}\,\mathrm{d}s < \infty \quad \forall t > 0, \tag{5.5}$$

$$\int_s^r \int\int \bigl|\delta\psi_t(x,z)\bigr|\nu(\mathrm{d}z)m(\mathrm{d}x)\,\mathrm{d}t < \infty \quad \forall 0 < s < r. \tag{5.6}$$

We will also assume a control on the moment of the Lévy measure.

Assumption 5.6. *There exists a constant $\beta > 0$ such that*

$$M_\beta := \int \mathbf{1}_{\{|x|>1\}}|x|^\beta \nu(\mathrm{d}x) < \infty.$$

Remark 5.7. *The assumptions on the regularity of ψ could possibly be weakened, however, we prefer to keep the presentation simple here. In* [2] *similar calculations as here are performed under very mild assumptions in a local setting. Still, Assumption 5.5 is fulfilled e.g. for the choice $\nu_\alpha(\mathrm{d}y) = c_\alpha |y|^{-\alpha-d}\,\mathrm{d}y$ for $\alpha \in (0,2)$ corresponding to the fractional Laplacian $\mathcal{L} = -(-\Delta)^{\alpha/2}$. This can be checked using the explicit Fourier representation $\mathcal{F}(\psi_t)(\xi) = \exp(-t|\xi|^\alpha)$, which implies e.g. the following heat kernel bounds* (*see e.g.* [11], *Theorem* 1.1):

$$\frac{1}{C} \cdot \left(t^{-d/\alpha} \wedge \frac{t}{|x|^{\alpha+d}}\right) \le \psi(t,x) \le C \cdot \left(t^{-d/\alpha} \wedge \frac{t}{|x|^{\alpha+d}}\right). \tag{5.7}$$

Assumption 5.6 is only used in Proposition 5.8 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 $(P_t)_{t\ge 0}$ 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 $\boldsymbol{\nu} \in \mathcal{M}_{\mathrm{loc}}(G)$ we set

$$P_t[\boldsymbol{\nu}] := \boldsymbol{\nu} * \psi_t,$$

with the convolution being understood in the sense of (2.8).

In order to characterise the semigroup P_t as the gradient flow of the entropy, we want to apply Definition 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

$$\mathscr{P}^* := \bigl\{\mu \in \mathscr{P}\bigl(\mathbb{R}^d\bigr) : \mathcal{H}(\mu) > -\infty\bigr\}.$$

We set $X := \mathscr{P}_\tau = \{\mu \in \mathscr{P}(\mathbb{R}^d)\colon \mathcal{W}(\mu,\tau) < \infty\}$ for some $\tau \in \mathscr{P}^*$. The next two results ensure that this choice fits well into the setting of Definition 5.2.

Proposition 5.8. *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 $\mathcal{H}\colon \mathscr{P}_\tau \to (-\infty,\infty]$ 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 $V(x) := \max(1, |x|^{\beta/2}) + c$. Here β is the constant from Assumption 5.6 and the constant c is chosen such that γ is a probability measure. We can assume that $\beta < 1$. Using the inequality $||y|^{\beta/2} - |x|^{\beta/2}| \le |y-x|^{\beta/2}$, it is easy to check that

$$\big|\bar{\nabla} V(x,y)\big| = \big|V(y) - V(x)\big| \le \min\big(|y-x|, |y-x|^{\beta/2}\big). \tag{5.8}$$

Now note that for any $\mu \in \mathscr{P}(\mathbb{R}^d)$ we have

$$\mathcal{H}(\mu) = \mathcal{H}(\mu|\gamma) - \int V(x)\mu(\mathrm{d}x). \tag{5.9}$$

Moreover, $\mathcal{H}(\mu|\gamma) \ge 0$ since γ is a probability measure.

Let us now show the first statement of the proposition. By (5.9) 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 2.6 or Proposition 4.5 and using (5.8) we obtain

$$\begin{aligned}
\left|\int V_n\,\mathrm{d}\mu - \int V_n\,\mathrm{d}\tau\right| &\le \frac{\mathcal{W}(\mu,\tau)}{\sqrt{2}} \cdot \left(\int_0^1 \int \big|\bar{\nabla} V_n(x,y)\big|^2 J(x,\mathrm{d}y)\mu_s(\mathrm{d}x)\,\mathrm{d}s\right)^{1/2} \\
&\le \frac{\mathcal{W}(\mu,\tau)}{\sqrt{2}} \cdot \left(\int_0^1 \int \min\big(|z|^2, |z|^\beta\big)\nu(\mathrm{d}z)\mu_s(\mathrm{d}x)\,\mathrm{d}s\right)^{1/2} \\
&\le \mathcal{W}(\mu,\tau) \cdot \sqrt{\frac{M_\beta + C_\nu}{2}}.
\end{aligned}$$

Here M_β is the constant from Assumption 5.6 and C_ν is given by (5.2). Letting $n \to \infty$, monotone convergence yields

$$\left|\int V\,\mathrm{d}\mu - \int V\,\mathrm{d}\tau\right| \le \mathcal{W}(\mu,\tau) \cdot \sqrt{\frac{M_\beta + C_\nu}{2}}$$

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 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. [1], Lemma 9.4.3). Furthermore, arguing as before, we obtain the estimate

$$\left|\int V\,\mathrm{d}\mu_n - \int V\,\mathrm{d}\mu\right| \le \mathcal{W}(\mu_n,\mu) \cdot \sqrt{\frac{M_\beta + C_\nu}{2}} \longrightarrow 0.$$

In view of (5.9) this finishes the proof. □

Lemma 5.9. *$(P_t)_t$ is a strongly continuous semigroup on $\mathscr{P}_\tau$, i.e. for any $\mu \in \mathscr{P}_\tau$ we have that $P_t[\mu] \in \mathscr{P}_\tau$ for all $t > 0$ and $\mathcal{W}(P_t[\mu],\mu) \to 0$ as $t \to 0$.*

Proof. For $s \in [0,t]$ we put $\mu_s := P_s[\mu] = \rho_s m$, where

$$\rho_s(x) = \int \psi_s(x-z)\mu(\mathrm{d}z).$$

Further set $\boldsymbol{v}_s = \bar{\nabla}\rho_s Jm$. Since by Assumption 5.5 ψ is a fundamental solution to $\partial_t u = \mathcal{L}u$ we easily check that $(\mu, \boldsymbol{v}) \in \mathcal{CE}_t(\mu, P_t[\mu])$. The action is given as

$$\mathcal{A}(\mu_s, \boldsymbol{v}_s) = \int \frac{|\bar{\nabla}\rho_s|^2}{\hat{\rho}_s}\,\mathrm{d}Jm = \mathcal{I}(\mu_s).$$

By Lemma 5.4 we have $\mathcal{I}(\mu_s) \le \mathcal{I}(q_s)$ and using Lemma 4.2, we estimate

$$\mathcal{W}\big(P_t[\mu], \mu\big) \le \int_0^t \sqrt{\mathcal{A}(\mu_s, \boldsymbol{v}_s)}\,\mathrm{d}s \le \int_0^t \sqrt{\mathcal{I}(q_s)}\,\mathrm{d}s.$$

By (5.5) the last expression is finite and tends to 0 as $t \to 0$. □

Let us now state a result giving the entropy production along the semigroup P.

Proposition 5.10. *Let $\mu \in \mathcal{P}^*$. 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*

$$\mathcal{H}\big(P_t[\mu]\big) - \mathcal{H}\big(P_s[\mu]\big) = -\int_s^t \mathcal{I}\big(P_r[\mu]\big)\,\mathrm{d}r \quad \forall t \ge s > 0. \tag{5.10}$$

In particular the map $t \mapsto \mathcal{H}(P_t[\mu])$ is non-increasing.

Proof. Since $\mathcal{W}(P_t[\mu], \mu) < \infty$ by Lemma 5.9, Proposition 5.8 gives that $P_t[\mu] \in \mathcal{P}_*$. Note that $P_t[\mu] = \rho_t m$ is absolutely continuous with density

$$\rho_t(x) = \int \psi_t(x - z)\mu(\mathrm{d}z).$$

Finiteness of $\mathcal{H}(P_t[\mu])$ and $\mathcal{I}(P_t[\mu])$ thus follows immediately from (5.4) and convexity of the map $r \mapsto r\log r$ respectively Lemma 5.4.

We prove (5.10) by approximating $\mathcal{H}$ with functionals $\mathcal{H}_n$. Let us set

$$f_n(u) := \int_0^u \max\big(1 + \log(r), -n\big)\,\mathrm{d}r. \tag{5.11}$$

Then we have $f_n(u) \searrow u\log(u)$ and $f_n'(u) \searrow 1 + \log(u)$ as $n \to \infty$. For $\mu = \rho m \in \mathcal{P}(\mathbb{R}^d)$ we set $\mathcal{H}_n(\mu) := \int f_n(\rho)\,\mathrm{d}m$. From Assumption 5.5 we deduce that ρ satisfies $\partial_t \rho = \mathcal{L}\rho$. Now we calculate

$$\begin{aligned}
\mathcal{H}_n\big(P_t[\mu]\big) - \mathcal{H}_n\big(P_s[\mu]\big) &= \int f_n(\rho_t) - f_n(\rho_s)\,\mathrm{d}m = \int\int_s^t f_n'(\rho_r)\partial_r\rho_r\,\mathrm{d}r\,\mathrm{d}m \\
&= \int\int_s^t f_n'(\rho_r)\mathcal{L}\rho_r\,\mathrm{d}r\,\mathrm{d}m \\
&= -\frac{1}{2}\int_s^t\int \bar{\nabla} f_n'(\rho_r)\bar{\nabla}\rho_r\,\mathrm{d}Jm\,\mathrm{d}r.
\end{aligned}$$

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 (5.6) holds with ψ replaced by ρ. The integration by parts in the last line can be justified by using again (5.6) and (5.4). Letting finally $n \to \infty$, we obtain (5.10) 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 5.2. Our strategy of proof is inspired by an argument developed in [13] and used in a similar form in [14], Theorem 5.29. The following two results are a restatement of Theorem 1.4 from the introduction.

Theorem 5.11. *Let $\mu \in \mathcal{P}^*$. Then $\mathcal{H}(P_t[\mu]) < \infty$ for all $t > 0$ and the map $t \mapsto \mathcal{H}(\mu_t)$ is non-increasing. Moreover, for any $\sigma \in \mathcal{P}_\mu$ the Evolution Variational Inequality holds*:

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}\big(P_t[\mu], \sigma\big)^2 + \mathcal{H}\big(P_t[\mu]\big) \le \mathcal{H}(\sigma) \quad \forall t > 0. \tag{5.12}$$

Proof. Finiteness and monotonicity of $\mathcal{H}(P_t[\mu])$ were already proved in Proposition 5.10. To prove the Evolution Variational Inequality, 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 $\sigma \in \mathcal{P}_\mu$ with $\mathcal{H}(\sigma) < \infty$ and let $(\sigma_s, \boldsymbol{v}_s)_{s\in[0,1]}$ be a minimising curve connecting $\sigma_0 = \sigma$ to $\sigma_1 = \mu$. We set

$$\mu^\varepsilon_{s,t} = \rho^\varepsilon_{s,t} m := P_{st+\varepsilon}[\sigma_s] \quad \text{and}$$
$$\widetilde{\boldsymbol{v}}^\varepsilon_{s,t} = \widetilde{v}^\varepsilon_{s,t} Jm := P_{st+\varepsilon}[\boldsymbol{v}_s].$$

The couple $(\mu^\varepsilon_{s,t}, \widetilde{\boldsymbol{v}}^\varepsilon_{s,t})$ does not satisfy the continuity equation. Hence we make the correction

$$\boldsymbol{v}^\varepsilon_{s,t} = v^\varepsilon_{s,t} Jm := \big(\widetilde{v}^\varepsilon_{s,t} - t\bar{\nabla}\rho^\varepsilon_{s,t}\big) Jm.$$

We will need the following result whose proof we postpone for the moment.

Claim 5.12. *We have $(\mu^\varepsilon_{\cdot,t}, \boldsymbol{v}^\varepsilon_{\cdot,t}) \in \mathcal{CE}_1(P_\varepsilon[\sigma], P_{t+\varepsilon}[\mu])$ and moreover,*

$$\mathcal{H}\big(P_{\varepsilon+t}[\mu]\big) - \mathcal{H}\big(P_\varepsilon[\sigma]\big) = -\frac{1}{2}\int_0^1 \int \bar{\nabla} \log \rho^\varepsilon_{s,t} \, \mathrm{d}\boldsymbol{v}^\varepsilon_{s,t} \, \mathrm{d}s. \tag{5.13}$$

From the definition of the distance $\mathcal{W}$ we now obtain the estimate

$$\mathcal{W}\big(P_{t+\varepsilon}[\mu], P_\varepsilon[\sigma]\big)^2 \le \int_0^1 \mathcal{A}\big(\mu^\varepsilon_{s,t}, \boldsymbol{v}^\varepsilon_{s,t}\big)\, \mathrm{d}s. \tag{5.14}$$

Recall the notation $\hat{\rho}(x, y) = \theta(\rho(x), \rho(y))$ with θ being the logarithmic mean here. We can further estimate

$$\begin{aligned}
\mathcal{A}\big(\mu^\varepsilon_{s,t}, \boldsymbol{v}^\varepsilon_{s,t}\big) &= \int \frac{|v^\varepsilon_{s,t}|^2}{2\hat{\rho}^\varepsilon_{s,t}}\, \mathrm{d}Jm \\
&= \int \big(\big|\widetilde{v}^\varepsilon_{s,t}\big|^2 - 2t\bar{\nabla}\rho^\varepsilon_{s,t} v^\varepsilon_{s,t} - t^2\big|\bar{\nabla}\rho^\varepsilon_{s,t}\big|^2\big)\frac{1}{2\hat{\rho}^\varepsilon_{s,t}}\, \mathrm{d}Jm \\
&\le \mathcal{A}\big(\mu^\varepsilon_{s,t}, \widetilde{\boldsymbol{v}}^\varepsilon_{s,t}\big) - t\int \bar{\nabla}\log\rho^\varepsilon_{s,t} v^\varepsilon_{s,t}\, \mathrm{d}Jm \\
&\le \mathcal{A}(\sigma_s, \boldsymbol{v}_s) - t\int \bar{\nabla}\log\rho^\varepsilon_{s,t}\, \mathrm{d}\boldsymbol{v}^\varepsilon_{s,t},
\end{aligned}$$

where we have dropped the quadratic term in t and used the monotonicity under convolution (Proposition 2.8) in the last inequality. Integration over s from 0 to 1 and using (5.13) gives

$$\frac{1}{2}\mathcal{W}\big(P_{t+\varepsilon}[\mu], P_\varepsilon[\sigma]\big)^2 \le \frac{1}{2}\mathcal{W}(\mu, \sigma)^2 - t\cdot\big(\mathcal{H}\big(P_{t+\varepsilon}[\mu]\big) - \mathcal{H}\big(P_\varepsilon[\sigma]\big)\big).$$

By lower semicontinuity of $\mathcal{W}$ (see Theorem 4.4) and continuity of $\mathcal{H}$ along the semigroup we can take the limit $\varepsilon \to 0$ and obtain

$$\frac{1}{2}\mathcal{W}\big(P_t[\mu], \sigma\big)^2 \le \frac{1}{2}\mathcal{W}(\mu, \sigma)^2 - t\cdot\big(\mathcal{H}\big(P_t[\mu]\big) - \mathcal{H}(\sigma)\big).$$

Finally, rearranging terms and letting $t \searrow 0$ yields (5.12).

Proof of Claim 5.12. For the proof we first need two estimates. First, note that

$$\int_0^1 \mathcal{I}\big(\mu_{s,t}^\varepsilon\big)\,\mathrm{d}s < \infty. \tag{5.15}$$

Indeed, using Lemma 5.4 and Proposition 5.10, we estimate

$$\int_0^1 \mathcal{I}\big(\mu_{s,t}^\varepsilon\big)\,\mathrm{d}s \le \int_0^1 \mathcal{I}(q_{\varepsilon+st})\,\mathrm{d}s = \mathcal{H}(q_\varepsilon) - \mathcal{H}(q_{\varepsilon+t}) < \infty.$$

From this we conclude that the curve $(\mu_{\cdot,t}^\varepsilon, \boldsymbol{v}_{\cdot,t}^\varepsilon)$ has finite action. Indeed,

$$\begin{aligned} A &:= \int_0^1 \int \frac{|v_{s,t}^\varepsilon|^2}{2\hat{\rho}_{s,t}^\varepsilon}\,\mathrm{d}Jm\,\mathrm{d}s \\ &\le \int_0^1 \int 2\frac{|\widetilde{v}_{s,t}^\varepsilon|^2}{2\hat{\rho}_{s,t}^\varepsilon} + 2t^2\frac{|\bar{\nabla}\rho_{s,t}^\varepsilon|^2}{2\hat{\rho}_{s,t}^\varepsilon}\,\mathrm{d}Jm\,\mathrm{d}s \\ &\le 2\int_0^1 \mathcal{A}(\sigma_s, \boldsymbol{v}_s)\,\mathrm{d}s + 2t^2\int_0^1 \mathcal{I}\big(\mu_{s,t}^\varepsilon\big)\,\mathrm{d}s < \infty, \end{aligned}$$

where we use Proposition 2.8 in the last inequality. Using Lemma 2.6 and the previous estimate we see that $\boldsymbol{v}_{\cdot,t}^\varepsilon$ satisfies the integrability condition (iv) in Definition 3.2. The other conditions are also easily checked. Hence, we see that $(\mu_{\cdot,t}^\varepsilon, \boldsymbol{v}_{\cdot,t}^\varepsilon) \in \mathcal{CE}_1(P_\varepsilon[\sigma], P_{\varepsilon+t}[\mu])$.

Now let us prove (5.13). By a simple convolution argument we can assume that $\rho_{s,t}^\varepsilon$ is differentiable in s. Let f_n be the function defined by (5.11) and set $f(u) = u\log(u)$ for $u \ge 0$. Now we calculate

$$\mathcal{H}_n\big(P_{\varepsilon+t}[\mu]\big) - \mathcal{H}_n\big(P_\varepsilon[\sigma]\big) = \int\int_0^1 f_n'\big(\rho_{s,t}^\varepsilon\big)\partial_s\rho_{s,t}^\varepsilon\,\mathrm{d}s\,\mathrm{d}m.$$

Note that the map $x \mapsto f_n'(\rho_{s,t}^\varepsilon(x))$ is bounded and Lipschitz uniformly in $s \in [0,1]$. Using the integrability condition (iii) from Definition 3.2 we can approximate it by functions in $C_c^\infty((0,1)\times\mathbb{R}^d)$ and obtain by the continuity equation

$$\mathcal{H}_n\big(P_{\varepsilon+t}[\mu]\big) - \mathcal{H}_n\big(P_\varepsilon[\sigma]\big) = -\frac{1}{2}\int_0^1\int \bar{\nabla} f_n'\big(\rho_{s,t}^\varepsilon\big)\,\mathrm{d}\boldsymbol{v}_{s,t}^\varepsilon\,\mathrm{d}s. \tag{5.16}$$

By monotone convergence the left-hand side of (5.16) converges to the left-hand side of (5.13). It remains to prove convergence of the right-hand side. Using Hölder inequality, we estimate

$$\begin{aligned} &\left|\int_0^1\int \bar{\nabla}\big(f'\big(\rho_{s,t}^\varepsilon\big) - f_n'\big(\rho_{s,t}^\varepsilon\big)\big)\,\mathrm{d}\boldsymbol{v}_{s,t}^\varepsilon\,\mathrm{d}s\right| \\ &\quad\le \int_0^1\int \big|\bar{\nabla}\big(f'\big(\rho_{s,t}^\varepsilon\big) - f_n'\big(\rho_{s,t}^\varepsilon\big)\big)\big|\big|w_{s,t}^\varepsilon\big|\,\mathrm{d}Jm\,\mathrm{d}s \\ &\quad\le A^{1/2}\bigg(\int_0^1\int \big|\bar{\nabla}\big(f'\big(\rho_{s,t}^\varepsilon\big) - f_n'\big(\rho_{s,t}^\varepsilon\big)\big)\big|^2 2\hat{\rho}_{s,t}^\varepsilon\,\mathrm{d}Jm\,\mathrm{d}s\bigg)^{1/2}. \end{aligned}$$

The integrand in the last term is bounded as

$$\big|\bar{\nabla}\big(f'\big(\rho_{s,t}^\varepsilon\big) - f_n'\big(\rho_{s,t}^\varepsilon\big)\big)\big|^2\hat{\rho}_{s,t}^\varepsilon \le \big|\bar{\nabla} f'\big(\rho_{s,t}^\varepsilon\big)\big|^2\hat{\rho}_{s,t}^\varepsilon = \bar{\nabla}\log\rho_{s,t}^\varepsilon\bar{\nabla}\rho_{s,t}^\varepsilon.$$

With the help of (5.15) and dominated convergence we conclude convergence of the right-hand side of (5.16) to the right-hand side of (5.13). □

□

Corollary 5.13. *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*

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

Proof. This is a direct consequence of Theorem 5.11 and the fact, proved in [13], Theorem 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 (1.5). 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

$$\mathcal{A}'(\rho, \psi) := \frac{1}{2} \int \left|\bar{\nabla}\psi(x,y)\right|^2 \hat{\rho}(x,y) J m(\mathrm{d}x, \mathrm{d}y).$$

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:

$$\begin{cases} \text{(i)} & \rho : [0,T] \times \mathbb{R}^d \to \mathbb{R}_+ \text{ is measurable;} \\ \text{(ii)} & \rho_t \text{ is a probability density for all } t \in [0,T]; \\ \text{(iii)} & \text{The curve } t \mapsto \mu_t := \rho_t m \text{ is weakly continuous;} \\ \text{(iv)} & \psi : [0,T] \times \mathbb{R}^d \to \mathbb{R} \text{ is measurable;} \\ (v) & \partial_t \rho_t + \bar{\nabla} \cdot (\hat{\rho}_t \bar{\nabla}\psi_t) = 0,\ \rho_0 = \bar{\rho}_0,\ \rho_T = \bar{\rho}_1. \end{cases} \tag{5.17}$$

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

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

Proposition 5.14. *In addition to Assumptions 5.3 and 5.5 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 $i = 0, 1$. 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 of Proposition 5.14. The inequality '$\le$' 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 $\boldsymbol{\nu}_t \in \mathcal{M}_{\mathrm{loc}}(G)$ by setting $\boldsymbol{\nu}_t(\mathrm{d}x, \mathrm{d}y) = \bar{\nabla}\psi_t(x,y)\hat{\rho}_t(x,y) J(x, \mathrm{d}y) m(\mathrm{d}x)$. Then we have $\mathcal{A}'(\rho_t, \psi_t) = \mathcal{A}(\mu_t, \boldsymbol{\nu}_t)$ and it is easily checked using Lemma 2.6 that $(\mu, \boldsymbol{\nu}) \in \mathcal{CE}_1(\bar{\mu}_0, \bar{\mu}_1)$.

Let us now prove the opposite inequality '$\ge$'. To this end, note that by a reparametrisation argument similar to Lemma 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 Lemma 5.4, we infer that

$$L^{i,\varepsilon} := \int_0^\varepsilon \sqrt{\mathcal{A}'\big(\rho_t^{i,\varepsilon}, \psi_t^{i,\varepsilon}\big)}\,\mathrm{d}t = \int_0^\varepsilon \sqrt{\mathcal{I}\big(\mu_t^{i,\varepsilon}\big)}\,\mathrm{d}t \le \int_0^\varepsilon \sqrt{\mathcal{I}(q_t)}\,\mathrm{d}t.$$

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 4.9 and the proof of Proposition 4.8 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 4.11 we have that $\widetilde{\boldsymbol{\nu}}_t^\varepsilon = w_t^\varepsilon \hat{\rho}_t^\varepsilon Jm$ where w_t^ε belongs to $T_{\rho_t}^\varepsilon$. In particular, there exists a sequence of functions ψ_n such that $\bar{\nabla}\psi_n \to w_t^\varepsilon$ almost surely w.r.t. the measure $\hat{\rho}_t^\varepsilon Jm$. Note that $\rho_t^\varepsilon > 0$ by Assumption 5.5 and thus $\hat{\rho}_t^\varepsilon > 0$ for all $t \in (0,1)$ and moreover $j > 0$. Hence, we have $\bar{\nabla}\psi_n \to w_t^\varepsilon$ also m^2-almost surely and it is easily checked that any a.s. limit of discrete gradients coincides again a.e. with a discrete gradient. Thus there exist a function $\psi^\varepsilon : (0,1) \times \mathbb{R}^d \to \mathbb{R}$ such that $w_t^\varepsilon = \bar{\nabla}\psi_t^\varepsilon$ a.e. Now observe that $(\rho^\varepsilon, \psi^\varepsilon) \in \mathcal{CE}_1'(\rho_0^\varepsilon, \rho_1^\varepsilon)$ and

$$L^\varepsilon := \int_0^1 \sqrt{\mathcal{A}'\big(\rho_t^\varepsilon, \psi_t^\varepsilon\big)}\,\mathrm{d}t = \int_0^1 \sqrt{\mathcal{A}\big(\mu_t^\varepsilon, \widetilde{\boldsymbol{\nu}}_t^\varepsilon\big)}\,\mathrm{d}t \le \int_0^1 \sqrt{\mathcal{A}(\mu_t, \boldsymbol{\nu}_t)}\,\mathrm{d}t = \mathcal{W}(\bar{\mu}_0, \bar{\mu}_1),$$

where we have used Proposition 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

$$\int_0^{1+2\varepsilon} \sqrt{\mathcal{A}'\big(\widetilde{\rho}_t^\varepsilon, \widetilde{\psi}_t^\varepsilon\big)}\,\mathrm{d}t = L^{0,\varepsilon} + L^\varepsilon + L^{1,\varepsilon} \le \mathcal{W}(\bar{\mu}_0, \bar{\mu}_1) + 2\int_0^\varepsilon \sqrt{\mathcal{I}(q_t)}\,\mathrm{d}t.$$

By Assumption 5.5 the second term in the last line goes to zero as $\varepsilon \to 0$ which yields the claim. □

Acknowledgments

The author is grateful to Jan Maas and Karl-Theodor Sturm for stimulating discussions on this paper and related questions.

References

[1] L. Ambrosio, N. Gigli and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd edition. *Lectures in Mathematics*. Birkhäuser, Basel, 2008. MR2401600

[2] L. Ambrosio, N. Gigli and G. Savaré. Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. Preprint, 2011. Available at arXiv:1106.2090.

[3] 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) 517–564. MR2529438

[4] D. Applebaum. *Lévy Processes and Stochastic Calculus*. *Cambridge Studies in Advanced Mathematics* **93**. Cambridge Univ. Press, Cambridge, 2004. MR2072890

[5] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de Probabilités XIX* 177–206. *Lecture Notes in Math.* **1123**. Springer, Berlin, 1985. MR0889476

[6] M. Barlow, R. Bass, Z.-G. Chen and M. Kassmann. Non-local Dirichlet forms and symmetric jump processes. *Trans. Amer. Math. Soc.* **361** (2009) 1963–1999. MR2465826

[7] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math.* **84** (2000) 375–393. MR1738163

[8] J. Bertoin. *Lévy Processes*. *Cambridge Tracts in Mathematics* **121**. Cambridge Univ. Press, Cambridge, 1996. MR1406564

[9] G. Buttazzo. *Semicontinuity, Relaxation and Integral Representation in the Calculus of Variations*. *Pitman Research Notes in Mathematics Series*. Longman Scientific and Technical, Harlow, 1989. MR1020296

[10] L. Caffarelli and L. Silvestre. The Evans–Krylov theorem for nonlocal fully nonlinear equations. *Ann. of Math. (2)* **174** (2011) 1163–1187. MR2831115

[11] Z.-Q. Chen and T. Kumagai. Heat kernel estimates for stable-like processes on d-sets. *Stochastic Process. Appl.* **108** (2003) 27–62. MR2008600

[12] 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. MR2928139

[13] S. Daneri and G. Savaré. Eulerian calculus for the displacement convexity in the Wasserstein distance. *SIAM J. Math. Anal.* **40** (2008) 1104–1122. MR2452882

[14] J. Dolbeault, B. Nazaret and G. Savaré. A new class of transport distances between measures. *Calc. Var. Partial Differential Equations* **34** (2009) 193–231. MR2448650

[15] M. Erbar. The heat equation on manifolds as a gradient flow in the Wasserstein space. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** (2010) 1–23. MR2641767

[16] M. Erbar and J. Maas. Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.* **206** (2012) 997–1038. MR2989449
[17] S. Fang, J. Shao and K.-Th. Sturm. Wasserstein space over the Wiener space. *Probab. Theory Related Fields* **146** (2010) 535–565. MR2574738
[18] N. Gigli. On the heat flow on metric measure spaces: Existence, uniqueness and stability. *Calc. Var. Partial Differential Equations* **39** (2010) 101–120. MR2659681
[19] N. Gigli, K. Kuwada and S.-I. Ohta. Heat flow on Alexandrov spaces. *Comm. Pure Appl. Math.* **66** (2013) 307–331. MR3008226
[20] R. Jordan, D. Kinderlehrer and F. Otto. The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.* **29** (1998) 1–17. MR1617171
[21] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)* **169** (2009) 903–991. MR2480619
[22] J. Maas. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.* **261** (2011) 2250–2292. MR2824578
[23] R. McCann. A convexity principle for interacting gases. *Adv. Math.* **128** (1997) 153–179. MR1451422
[24] A. Mielke. Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Partial Differential Equations* **48** (2013) 1–31. MR3090532
[25] S.-I. Ohta and K.-Th. Sturm. Heat flow on Finsler manifolds. *Comm. Pure Appl. Math.* **62** (2009) 1386–1433. MR2547978
[26] F. Otto. The geometry of dissipative evolution equations: The porous medium equation. *Comm. Partial Differential Equations* **26** (2001) 101–174. MR1842429
[27] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** (2000) 361–400. MR1760620
[28] K.-Th. Sturm. On the geometry of metric measure spaces. I. *Acta Math.* **196** (2006) 65–131. MR2237206
[29] K.-Th. Sturm. On the geometry of metric measure spaces. II. *Acta Math.* **196** (2006) 133–177. MR2237207
[30] C. Villani. *Optimal Transport: Old and New*. *Grundlehren der Mathematischen Wissenschaften* **338**. Springer, Berlin, 2009. MR2459454