Nonlocal Wasserstein distance: metric and asymptotic properties

Dejan Slepčev1 · Andrew Warren1

Received: 22 September 2022 / Accepted: 3 September 2023
© The Author(s) 2023

Abstract
The seminal result of Benamou and Brenier provides a characterization of the Wasserstein distance as the path of the minimal action in the space of probability measures, where paths are solutions of the continuity equation and the action is the kinetic energy. Here we consider a fundamental modification of the framework where the paths are solutions of nonlocal (jump) continuity equations and the action is a nonlocal kinetic energy. The resulting nonlocal Wasserstein distances are relevant to fractional diffusions and Wasserstein distances on graphs. We characterize the basic properties of the distance and obtain sharp conditions on the (jump) kernel specifying the nonlocal transport that determine whether the topology metrized is the weak or the strong topology. A key result of the paper are the quantitative comparisons between the nonlocal and local Wasserstein distance.

Mathematics Subject Classification 46E27 · 49J99 · 60J76 · 60B10 · 45G10

Contents

1 Introduction and summary of results .........................................
  1.1 Related work .............................................
    1.1.1 Nonlocal Wasserstein metric, and associated gradient flows ........
    1.1.2 Graph Wasserstein distances ................................
    1.1.3 Other related work ....................................
  1.1.3 From graphs to continuum ...................................
  2 Preliminaries ................................................
    Wasserstein metric ............................................
    2.1 Nonlocal structure: weight kernel, interpolation, vector calculus ........
    2.2 Action ................................................
    2.3 Nonlocal continuity equation ....................................
    2.4 Nonlocal Wasserstein metric ....................................

Communicated by Andrea Mondino.

Dejan Slepčev
slepcev@math.cmu.edu

Andrew Warren
awarren1@alumni.cmu.edu

1 Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Published online: 27 September 2023
2.5 Convolutions .................................................................
2.5.1 Relative Lipschitz estimate for the right convolution ...............

3 Metric structure of \( W_{2,\theta} \) distances ..............................
3.1 General lower bounds for nonlocal Wasserstein distances ..........
3.2 Expel problem for \( W_{2,\theta} \) .......................................... 
3.3 Global upper bounds for \( W_{2,\theta} \) ....................................

4 Exact nonlocalization ......................................................
4.1 Exact solution to nonlocal continuity equation ....................... 
4.2 A quantitative upper bound on the nonlocal Wasserstein distance 

5 Nonlocal Hamilton–Jacobi subsolution ................................

Appendix A: Lower semicontinuity of integral functionals on \( \mathcal{M}_{loc} \)
Appendix B: Additional lemmas ...........................................
B. 1. Lemmas for Sect. 4 .................................................
B. 2. Lemmas for Sect. 5 ...................................................
References ...........................................................................

1 Introduction and summary of results

This paper is devoted to the study of a nonlocal Wasserstein distance defined as the least nonlocal action needed to connect two measures via a nonlocal continuity equation. The standard Wasserstein distance is the minimal transportation cost to couple two measures, for the quadratic point-to-point cost:

\[
W_2(\mu_0, \mu_1) := \left( \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\pi(x, y) \right)^{1/2},
\]

where \( \mu_0 \) and \( \mu_1 \) lie in \( \mathcal{P}_2(\mathbb{R}^d) \), the space of probability measures on \( \mathbb{R}^d \) with finite second moments, and \( \Pi(\mu_0, \mu_1) \) is the space of all couplings of \( \mu_0 \) and \( \mu_1 \). The celebrated work of Benamou and Brenier [4] establishes that the Wasserstein metric has a dynamical reformulation inspired by fluid mechanics. There, one considers the space of all narrowly continuous curves \( \rho_t : [0, 1] \rightarrow \mathcal{P}_2(\mathbb{R}^d) \) and velocity vector fields \( v_t : [0, 1] \rightarrow L^2(\rho_t; \mathbb{R}^d) \) such that the continuity equation

\[
\partial_t \rho_t + \text{div}(\rho_t v_t) = 0 \quad (1.1)
\]

holds in the sense of distributions. In [4] it is shown that

\[
W_2^2(\mu_0, \mu_1) = \inf \left\{ \int_0^1 \int_{\mathbb{R}^d} |v_t(x)|^2 d\rho_t(x) dt : \rho_0 = \mu_0, \rho_1 = \mu_1, \text{ and } (\rho_t, v_t)_{t \in [0, 1]} \text{ solve (1.1)} \right\}. \quad (1.2)
\]

The integral \( \int_0^1 \int_{\mathbb{R}^d} |v_t(x)|^2 d\rho_t(x) dt \) represents the total action along \( (\rho_t, v_t)_{t \in [0, 1]} \); in other words, the 2-Wasserstein distance between \( \mu_0 \) and \( \mu_1 \) can be reformulated as the least action of a curve in \( \mathcal{P}_2(\mathbb{R}^d) \) connecting \( \mu_0 \) to \( \mu_1 \) in which the flow of mass is continuous, in the sense of satisfying Eq. 1.1.

In this article, our object of study is the class of nonlocal transportation metrics on \( \mathcal{P}(\mathbb{R}^d) \), introduced by Erbar in [19], which are defined in terms of an nonlocal action minimization problem on the space of curves in \( \mathcal{P}(\mathbb{R}^d) \). A key difference is that the curves connecting the measures are not solutions of the continuity Eq. 1.1, but solutions of nonlocal continuity equation:

\[
\partial_t \rho_t(x) + \int_{\mathbb{R}^d} v_t(x, y) \theta(\rho_t(x), \rho_t(y)) \eta(x, y) dy = 0, \quad (1.3)
\]
where \( v_t : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \) is the nonlocal velocity, \( \eta : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty) \) is the weight kernel which encodes the ability to transport mass directly from \( x \) to \( y \), and \( \theta : [0, \infty) \times [0, \infty) \to [0, \infty) \) allows one to define an “interpolated density” \( \theta(\rho_t(x), \rho_t(y)) \), which is a generalized average of \( \rho_t(x) \) and \( \rho_t(y) \).

The nonlocal total action is formally given by

\[
\frac{1}{2} \int_0^1 \int_{\mathbb{R}^d \times \mathbb{R}^d} v_t(x, y)^2 \theta(\rho_t(x), \rho_t(y)) \eta(x, y) dx dy dt. \tag{1.4}
\]

Together, 1.3 and 1.4 allow us to consider the family of nonlocal transportation distances (since the distance now depends on the choice of \( \eta \) and \( \theta \))

\[
\mathcal{W}_{\eta, \theta}^2(\mu_0, \mu_1) := \inf \left\{ \frac{1}{2} \int_0^1 \int_{\mathbb{R}^d \times \mathbb{R}^d} v_t(x, y)^2 \theta(\rho_t(x), \rho_t(y)) \eta(x, y) dx dy dt \right\} \tag{1.5}
\]

where the infimum runs over all \((\rho_t, v_t)_{t \in [0,1]}\) which solve (1.3), such that \( \rho_0 = \mu_0 \) and \( \rho_1 = \mu_1 \). This family of distances can be viewed simultaneously as nonlocal analogues of the Benamou-Brenier formulation of the \( W_2 \) metric, and also as a “continuum state space” analogue of the graph Wasserstein distance defined in [10, 37, 38], wherein the underlying space is a finite graph or irreducible Markov chain, rather than \( \mathbb{R}^d \) as in the case of \( \mathcal{W}_{\eta, \theta} \).

In this paper we investigate topological and metric properties of the family of distances \( \mathcal{W}_{\eta, \theta} \) and compare them to the Wasserstein metric \( W_2 \). Some properties have already been established in [19]: it is known that the \( \mathcal{W}_{\eta, \theta} \) distance is lower semicontinuous with respect to narrow convergence, and that if the kernel \( \eta \) has finite second moments, then the topology induced by \( \mathcal{W}_{\eta, \theta} \) is at least as fine as that of narrow convergence.

Here we show that the topology metrized by \( \mathcal{W}_{\eta, \theta} \) can be strictly stronger than that of narrow convergence. In particular we characterize for which kernels \( \mathcal{W}_{\eta, \theta} \) metrizes the narrow, the strong, or an even stronger topology on the space of measures supported within a compact domain. The key to establishing the result is the following proposition which loosely speaking characterizes the effort needed to spread mass from a point to the surrounding region.

**Proposition 1.1** Let \( \mathcal{W}_{\eta, \theta} \) be defined as in Definition 2.15. Suppose that \( \eta \) and \( \theta \) satisfy Assumption 2.1 (i-v) and Assumption 2.2 respectively. In particular, in this case \( \eta(x, y) = \eta(|x - y|) \) for some univariate function \( \eta \).

If \( v \) is any compactly supported probability measure singular to \( \delta_0 \) (the Dirac measure at the origin), then depending on the choice of \( \eta \) and \( \theta \):

(i) If \( \theta(1, 0) = 0 \) and \( \int_{B(0,1)} \eta(|y|) dy < \infty \) then \( \mathcal{W}_{\eta, \theta}(\delta_0, v) = \infty \).

(ii) If \( \theta(1, 0) > 0 \) and \( \int_{B(0,1)} \eta(|y|) dy < \infty \) then \( \infty > \mathcal{W}_{\eta, \theta}(\delta_0, v) \geq 2 \left( \int_{\mathbb{R}^d} \eta(|y|) dy \right)^{-1/2} \).

(iii) If instead \( \eta \) has algebraic blow-up at the origin, that is, there exists some \( s > 0, \delta > 0 \), and constant \( c > 0 \) such that \( \eta(|y|) \geq c |y|^{-d-s} \) when \( |y| \leq \delta \), then instead we have the estimate

\[
\mathcal{W}_{\eta, \theta}(\delta_0, m_{B(0,\delta)}) \leq C\delta^{s/2}
\]

with explicit constant \( C \), where \( m_{B(0,\delta)} \) is the uniform probability measure on \( B(0, \delta) \). In particular, \( \inf_{v \in \mathcal{P}(\mathbb{R}^d): v \perp \delta_0} \mathcal{W}_{\eta, \theta}(\delta_0, v) = 0 \).

We remark that the proposition is not a trichotomy: the case where \( \int_{B(0,1)} \eta(|y|) dy = \infty \), but \( \eta(|y|) \) does not grow strictly faster than \( |y|^{-d} \) near the origin, remains open.
We prove Proposition 1.1 in Sect. 3.2. While Proposition 1.1 shows that in case (i), the topology induced by $W_{\eta,\theta}$ on $\mathcal{P}(\mathbb{R}^d)$ is highly disconnected, we give a further structural description of the topology of $W_{\eta,\theta}$ in the other two cases. Moreover in case (ii) we establish a quantitative comparison to a combination of total variation and transportation distances. Let $W_1$ be the Monge distance, that is the optimal transportation distance with linear cost.

**Theorem 1.2** Let $W_{\eta,\theta}$ be defined as in Definition 2.15. Suppose that $\eta$ and $\theta$ satisfy Assumptions 2.1 (i–v) and 2.2 respectively, and in particular, $\eta(x, y) = \eta(|x - y|)$ for some univariate function $\eta$. If $\theta(1, 0) > 0$ and $\int_{B(0,1)} \eta(|y|) dy < \infty$, then on a compact domain, there exists an explicit constant $C$ such that for all $\mu, v \in \mathcal{P}(\mathbb{R}^d)$,

$$\frac{1}{C} TV^{1/2}(\mu, v) \leq W_{\eta,\theta}(\mu, v) \leq C \cdot TV^{1/2}(\mu, v).$$

In particular, on a compact domain, $W_{\eta,\theta}$ metrizes the strong topology on probability measures.

Conversely, in the case where $\int_{B(0,1)} \eta(|y|) dy = \infty$ but $\eta$ has finite second moment, then we merely have the lower bound

$$W_1(\mu, v) \leq C W_{\eta,\theta}(\mu, v);$$

and furthermore, if there exists some $s > 0$, $\delta > 0$, and constant $c > 0$ such that $\eta(|y|) \geq c|y|^{-d-s}$ when $|y| \leq \delta$, then on a compact domain, $W_{\eta,\theta}$ metrizes the weak topology on probability measures.

The lower bounds asserted in Theorem 1.2 are established in Sect. 3.1; the corresponding upper bound makes use of the estimates from Sect. 3.2, and is proved in Sect. 3.3.

We now turn to making the quantitative comparison between the Wasserstein distance and the nonlocal Wasserstein distances more precise. In particular we show that when the kernel of nonlocal transport $\eta$ is localized the nonlocal Wasserstein distance converges to the Wasserstein distance up to the appropriate scaling. We furthermore obtain explicit error bounds on the difference.

**Theorem 1.3** Let $W_{\eta,\theta}$ be defined as in Definition 2.15. Suppose that $\eta$ and $\theta$ satisfy Assumptions 2.1 and 2.2 respectively, and in particular, $\eta(x, y) = \eta(|x - y|)$ for some univariate function $\eta$. Let $\varepsilon \in (0, 1]$, and define $\eta_\varepsilon(|x - y|) := \varepsilon^{-d} \eta \left(\frac{|x-y|}{\varepsilon}\right)$ and similarly $\eta_\varepsilon(x, y) := \eta_\varepsilon(|x - y|)$. Suppose that $\eta$ has finite second moment $M_2(\eta) := \int |y|^2 \eta(|y|) dy$. Let $\rho_0, \rho_1 \in \mathcal{P}(\mathbb{R}^d)$. Then the following estimates hold:

(i) Suppose either that $\eta(|y|)$ is integrable and $\theta(1, 0) > 0$, or that there exists some $s > 0$, $\delta > 0$, and constant $c > 0$ such that $\eta(|y|) \geq c|y|^{-d-s}$ when $|y| \leq \delta$. Then, there exists a constant $C_{d,\theta,\eta}$ depending solely and explicitly on $d$, $\theta$, and $\eta$, such that

$$\sqrt{\frac{M_2(\eta)}{2d}} \varepsilon W_{\eta,\theta}(\rho_0, \rho_1) \leq \left(1 + \sqrt{\varepsilon}\right)^2 W_2(\rho_0, \rho_1) + C_{d,\theta,\eta} \sqrt{\varepsilon}.$$

(ii) Suppose that $\rho_0$ and $\rho_1$ are supported inside some domain of radius $R$. Then, there is a constant $C_{R,\eta}$ depending solely and explicitly on $R$ and $\eta$ such that

$$W_2^2(\rho_0, \rho_1) \leq \frac{M_2(\eta)}{2d} \varepsilon^2 W_{\eta,\theta}^2(\rho_0, \rho_1) + C_{R,\eta} \sqrt{\varepsilon}.$$
In particular, when restricting attention to probability measures supported on a compact domain, these estimates imply the Gromov-Hausdorff convergence of $\sqrt{\frac{M_2(\eta)}{2d}} \varepsilon W_{\eta, \theta}$ to $W_2$ as $\varepsilon \to 0$.

Part 1 of Theorem 1.3 is deduced, with an explicit constant for $C_{d, \theta, \eta}$, as Corollary 4.3; likewise, Part 2 of Theorem 1.3 is deduced, with an explicit constant for $C_{R, \eta}$, in Corollary 5.12.

1.1 Related work

Having stated the main results of the article, let us give some further motivating discussion regarding nonlocal Wasserstein distances, and why their topological and asymptotic properties are of interest.

1.1.1 Nonlocal Wasserstein metric, and associated gradient flows

Nonlocal Wasserstein distances were introduced in the work of Erbar [19]. A central result of [19] is that the $W_{\eta, \theta}$ gradient flow of entropy $E(\rho) = \int \ln \rho \, d\rho$ is the fractional heat equation

$$\partial_t \rho_t + (-\Delta)^{s/2} \rho = 0$$

when $\theta$ is chosen to be the logarithmic mean, and $\eta$ is chosen to be the jump kernel of an $s$-stable Levy process: $\eta(x, y) = -c|x-y|^{-s-d}$. More broadly, this result suggests that nonlocal parabolic equations may be studied in an analogous fashion to those parabolic equations (such as: the heat equation, the Fokker–Planck equation, the porous medium equation) which can be cast as $W_2$ gradient flows [3]. Erbar has made important contributions in that direction by establishing the lower-semicontinuity of the nonlocal action, showing the topology of nonlocal Wasserstein distances is at least as strong as that of the Wasserstein distance, and showing that the entropy is geodesically convex with respect to the nonlocal Wasserstein distance.

As with the regular Wasserstein distance it is of interest to consider gradient flows of the functionals that combine some or all of: entropy, potential, and interaction energy. For $\beta_i \geq 0$ for $i = 1, 2, 3$

$$G(\rho) = \beta_1 \int \ln \rho \, d\rho + \beta_2 \int U(x) \, d\rho(x) + \beta_3 \iint K(x, y) \, d\rho(x) \, d\rho(y).$$

(1.6)

where $K(x, y) = K(y, x)$. If $\beta_1 \beta_2 > 0$ and $\beta_3 = 0$ this would be a nonlocal Fokker-Planck equation, for $\beta_1 \beta_3 > 0$ this would be a nonlocal McKean-Vlasov equation. The issue that arises in nonlocal Wasserstein gradient flows is that the behavior of the solutions at low temperature, for $\beta_1 \ll 1$, crucially depends on the kernel $\eta$ and interpolation $\theta$. In particular the result of Proposition 1.1(i) indicates that when $\beta_1 = 0$, $\eta$ is integrable and $\theta$ is for example a logarithmic mean then gradient flow of the potential (or interaction) energy is unable to move a delta mass! This follows from the fact that since the potential energy is finite the gradient flow curves have finite action. This highlights the need to better understand the influence of the choice of $\eta$ and $\theta$ on the nonlocal Wasserstein metric and the resulting gradient flows.

To overcome the issues with the freezing of support for the gradient flow of potential and interaction energies, Esposito et al. [25] studied a modification of the nonlocal transportation
framework, inspired by upwind numerical schemes, which allows for the interpolation $\theta$ to depend on the velocity. For antisymmetric $v$

$$\theta(\rho(x), \rho(y), v(x, y)) = \begin{cases} \rho(x) & \text{if } v(x, y) \geq 0 \\ \rho(y) & \text{otherwise.} \end{cases}$$

They study gradient flow of the nonlocal interaction energy ($\beta_1 = \beta_2 = 0$) with respect to nonlocal Wasserstein distances both on graphs and in the continuum.

The resulting upwind nonlocal Wasserstein “distance” is not symmetric, and is shown to be a quasimetric. It provides a formal Finslerian (rather than Riemannian) differential structure on $\mathcal{P}(\mathbb{R}^d)$, which is nonetheless sufficient to develop gradient flows as curves of maximal slope.

Lastly: in [24], the authors show that the 1D aggregation equation

$$\partial_t f_t = \partial_v(f_t \partial_v W(f_t)) \quad W(v) = c|v|^3$$

can be cast as the gradient flow of the kinetic energy with respect to a nonlocal transportation metric, which they call the nonlocal collision metric. This metric falls outside the scope of this article because the action they consider is 2-homogeneous (rather than 1-homogeneous). While the aggregation equation is also known to be a 2-Wasserstein gradient flow (of the nonlocal interaction energy, rather than the kinetic energy), it is nonetheless notable that a nonlocal transportation metric is recently shown to be physically relevant in kinetic theory.

1.1.2 Graph Wasserstein distances

Maas [37], Mielke [38], and Chow et al. [10] have independently introduced a metric structure for probability measures on discrete spaces (finite graphs or Markov chains) modeled on the Benamou-Brenier formulation of the $W_2$ metric. Our setup largely follows that of Maas.

Let $\mathcal{X}$ be a finite set. Let $\pi$ be some distinguished probability measure on $\mathcal{X}$. Define

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

In other words, $\mathcal{P}_\pi(\mathcal{X})$ is the set of probability densities on $\mathcal{X}$, w.r.t. $\pi$.

Consider some irreducible Markov kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ such that $\pi$ is the unique stationary measure for $K$, that is,

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

Furthermore, assume that $K$ is reversible, that is, the detailed balance condition

$$K(x, y)\pi(x) = K(y, x)\pi(y)$$

holds. Equivalently [36, Chapter 9], one may consider a connected weighted graph on $\mathcal{X}$ with weights $w(x, y)$, where $\pi$ is the stationary distribution for the uniform random walk on $(\mathcal{X}, w)$. 
Let \( \theta(x, y) \) be an interpolation function\(^1\); define the shorthand \( \hat{\rho}(x, y) := \theta(\rho(x), \rho(y)) \). We introduce the graph continuity equation

\[
\dot{\rho}_t(x) + \sum_{y \in X} v_t(x, y) \hat{\rho}_t(x, y) K(x, y) = 0
\]

where \( v_t : X \times X \to \mathbb{R} \) is thought of as a “vector field” on \( X \), analogous to the vector field \( v_t \) appearing in the (continuum) continuity equation. The term \( \sum_{y \in X} v_t(x, y) \hat{\rho}_t(x, y) K(x, y) \) can be interpreted as a graph analogue of the term \( \text{div}(\rho v) \) from the continuity equation on \( \mathbb{R}^d \).

The action of a density-potential pair is given by

\[
A(\rho, v) := \frac{1}{2} \sum_{x, y \in X} (v_t(x, y))^2 \hat{\rho}(x, y) K(x, y) \pi(x).
\]

From here, one can define a geodesic metric on \( \mathcal{P}_\pi(X) \) in a variational fashion, by setting

\[
\mathcal{W}_{\theta, \eta, \pi}(\hat{\rho}_0, \hat{\rho}_1)^2 := \inf \left\{ \int_0^1 A(\rho_t, v_t) dt \right\}
\]

where the infimum runs over all pairs \( (\rho_t, v_t); t \in [0, 1] \) satisfying the graph continuity equation, with \( \rho_0 = \hat{\rho}_0 \) and \( \rho_1 = \hat{\rho}_1 \). This Benamou–Brenier-type formulation of a distance on a discrete base space is more technically straightforward than its continuum ancestor \( W_2 \).

For one, it is shown in [37] that, at least on the “interior” of \( \mathcal{P}_\pi(X) \) (namely, the subset \{ \( \rho(x) \in \mathcal{P}_\pi(X) : \forall x \in X, \rho(x) > 0 \} \) we can interpret \( \mathcal{W} \) as a geodesic metric arising from a bona fide Riemannian metric structure; this is in contrast to the continuum setting, where the space \( \mathcal{P}_2(\mathbb{R}^d) \) can only be understood “formally” as a Riemannian manifold. Indeed, for the metric \( \mathcal{W}_{\theta, \eta, \pi} \) and related gradient flows, numerous heuristic arguments from the Otto calculus can be translated to rigorous arguments in the discrete setting. This has been exploited to study a variety of evolution equations on discrete spaces as \( \mathcal{W}_{\theta, \eta, \pi} \) gradient flows, for instance discrete analogs of the porous medium equation [22] and the McKean-Vlasov equation [20].

The reason for the need to introduce an interpolation \( \theta \) in discrete setting is rather clear. Indeed, because mass configurations are defined on the set of nodes, and vector fields are defined on edges, the discrete analogue of the flux of the continuity equation, \( \rho v \), must combine node- and edge-defined quantities in a noncanonical fashion. Indeed, in the definition of \( \mathcal{W}_{\theta, \eta, \pi} \) the role of the flux \( \rho v \) is played by the quantity \( \theta(\rho(x), \rho(y)) v(x, y) \), where the interpolation \( \theta \) is introduced in order to define an edge-based quantity (flux) based on vertex-defined quantities (mass). While there are many possible choices for \( \theta \), which \( \theta \) one chooses can significantly alter the geometry of \( \mathcal{W}_{\theta, \eta, \pi} \).

**From graphs to continuum**

A number of works has investigated the asymptotic properties of graph Wasserstein distances as the graphs converge to a continuum limit. This is of particular interest in data science and for mesh-free numerical schemes. For instance, if one considers a sequence of finite graphs \( G_n \) equipped with the shortest-path metric, converging in some sense to a continuous domain \( \mathcal{G} \subset \mathbb{R}^d \), is it the case that \( (\mathcal{P}(G_n), \mathcal{W}_{\theta, \eta, \pi}) \) also converges to \( (\mathcal{P}(\mathcal{G}), W_2) \)? Similar Gromov-Hausdorff-type stability results are well-established for a sequence of continuous domains.

---

\(^1\) By this, we informally mean: a function which is a well-behaved, but possibly nonlinear, average of \( x \) and \( y \). Rigorously, we mean a function from \( \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R} \) satisfying all the conditions of Assumption 2.2 below.
where each respective space of probability measures is equipped with the $W_2$ metric [47]. However, the problem of discrete-to-continuum stability for the graph Wasserstein distance turns out to be considerably more delicate. In [31], it is shown that if we consider a sequence $\lambda_n$ of finer and finer $d$-dimensional regular lattices on the flat $d$-torus $T^d$, and take for our Markov chain the uniform random walk on said lattice, then under appropriate rescaling the sequence of spaces $(\mathcal{P}(X_n), W_{\theta, \eta, \pi})$ converges to $(\mathcal{P}(\mathbb{T}^d), W_2)$ in the sense of Gromov–Hausdorff. On the other hand, such convergence does not hold for an arbitrary sequence of regular meshes [33]. Despite this failure of convergence for general sequences of meshes, Garcia Trillos has shown [29] that $W_{\theta, \eta, \pi}$ corresponding to weighted random geometric graphs (e.g. where vertices are random i.i.d. samples from the Lebesgue measure on the torus) converges in the sense of Gromov–Hausdorff to $(\mathcal{P}(\mathbb{T}^d), W_2)$ as the number of vertices goes to infinity and the graph bandwidth converges to zero at appropriate rate (which is such that unweighted graph degrees go to infinity).

Our own Theorem 1.3 provides another result in this vein, whereby the 2-Wasserstein distance is recovered in the limit; but ours gives a nonlocal-to-local convergence result, rather than discrete-to-continuum.

We should also draw attention to one other question regarding the graph Wasserstein metrics: what can be said about the geodesics on $(\mathcal{P}(X), W_{\theta, \eta, \pi})$? Two works [23, 28] have independently investigated graph Wasserstein geodesics via their dual description in terms of solutions to a suitable discrete Hamilton–Jacobi equation. In the present article, an analogous Hamilton–Jacobi duality result for the nonlocal Wasserstein metric is exploited in Sect. 5 to prove Part 2 of Theorem 1.3.

### 1.1.3 Other related work

In [40], Peletier, Rossi, Savaré and Tse consider a far-reaching generalization of the setup of [19]. While [19] shows that the fractional heat equation can be viewed as the gradient flow of $\text{KL}(\cdot | \text{Leb})$ with respect to a Wasserstein-like metric, [40] investigates a large class of reversible Markov jump processes whose Kolmogorov forward equations may be cast as so-called generalized gradient systems, which are a further abstraction of the gradient flows in metric spaces studied in [3]. We note, however, that [40] imposes an assumption of uniform boundedness on the transition kernel of the Markov process, so that the results of [19] are not a special case of those established in [40].

Finally, let us remark on some related work on the subject of nonlocal conservation laws. We draw attention to two distinct uses of the term in the literature (although the works cited in the following discussion refer to some other uses of the term in the literature). On the one hand, certain authors use “nonlocal conservation law” to refer to solutions to the (local) continuity equation $\partial_t u_t + \text{div}(u_t v_t) = 0$ where $v_t$ is itself a nonlocal functional of $u_t$; see [5, 13] for a general discussion of this class of equations and an overview of related literature. In particular, we draw attention to recent work [11, 12] investigating the local limit (namely, as $\varepsilon \to 0$) of nonlocal conservation laws of the form

$$\partial_t u^\varepsilon_t + \text{div} \left( u^\varepsilon_t b \left( \eta_{\varepsilon} \ast u^\varepsilon_t \right) \right) = 0$$

where $b : \mathbb{R}_+ \to \mathbb{R}^d$, and $\eta : \mathbb{R}^d \to \mathbb{R}_+$ is some convolution kernel and $\eta_{\varepsilon}(x) := \frac{1}{\varepsilon^d} \eta \left( \frac{x}{\varepsilon} \right)$. Formally, the singular limit is given by the conservation law $\partial_t u_t + \text{div}(u_t b(u_t)) = 0$, but [12] exhibits counterexamples where $u^\varepsilon_t \rightharpoonup u_t$ (e.g. in $L^p$ for $p > 1$) even when $b$ and $\eta$ are regular. More recently, sufficient conditions for nonlocal-to-local convergence in

$\square$ Springer
one dimension have been given in [11]; but it remains the case that formally “obvious”
nonlocal-to-local convergence problems can present unexpected technical phenomena.
In a distinct line of work, the articles [16, 17] introduce new classes of “nonlocal con-
servation laws” where one replaces the divergence term in the continuity equation with a
nonlocal divergence-type operator. The nonlocal divergence and gradient in these papers, as
well as in [17, 32]), share the properties of the objects we study, but also have important dif-
ferences. More importantly the authors are interested in conservation laws where the flux
\( j \) is a nonlinear function of \( \rho \). Several ways to encode the nonlinear and nonlocal dependence on
\( \rho \) are developed. In [16] the authors present, in the same vein as in [11], sufficient conditions
which allow one to recover a local conservation law in the limit, under a suitable rescaling
of the nonlocal divergence operator.

2 Preliminaries

Wasserstein metric

Let \( \mathcal{P}_2(\mathbb{R}^d) \) denote the space of probability measures on \( \mathbb{R}^d \) with finite second moments. The
2-Wasserstein distance \( W_2 \) on \( \mathcal{P}_2(\mathbb{R}^d) \) is defined by

\[
W_2^2(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int |x - y|^2 \, d\pi(x, y)
\]

where \( \Pi(\mu, \nu) \) is the set of all transport plans (couplings) of \( \mu \) and \( \nu \). The 2-Wasserstein
distance also has a well-known dynamical formulation, due to Benamou and Brenier [4]:

\[
W_2^2(\mu, \nu) = \inf \left\{ \int_0^1 \int_{\mathbb{R}^d} |v_t(x)|^2 \, d\rho_t(x) \, dt \,: \, \partial_t \rho_t + \text{div}(\rho_t v_t) = 0, \, \rho_0 = \mu, \, \rho_1 = \nu \right\}.
\]

Here, the continuity equation \( \partial_t \rho_t + \text{div}(\rho_t v_t) = 0 \) is interpreted in a suitable distributional
sense, in particular to allow \( \rho_t \) to be a probability measure (rather than a smooth density).

The Benamou-Brenier formulation of the \( W_2 \) metric can be interpreted as showing that
the \( W_2 \) distance between \( \mu \) and \( \nu \) is given by the minimal total kinetic energy of a unit-
time flow of mass with initial and terminal distribution specified by \( \mu \) and \( \nu \) respectively.
Classically, kinetic energy is either formulated in terms of position and velocity, or position
and momentum; accordingly, as was observed in [4] (but see also further discussion and
extensions in [8, 14]), one can also rewrite the Benamou-Brenier formulation of \( W_2 \) in
“mass-flux” coordinates:

\[
W_2^2(\mu, \nu) = \inf \left\{ \int_0^1 \int_{\mathbb{R}^d} \left| \frac{d\tilde{j}_t}{d\rho_t} \right|^2 \, d\rho_t(x) \, dt \,: \, \partial_t \rho_t + \text{div}(\tilde{j}_t) = 0, \, \rho_0 = \mu, \, \rho_1 = \nu \right\}
\]

where \( \tilde{j}_t \) is a locally finite signed measure, which formally takes the place of \( \rho_t v_t \). This presentation
of \( W_2 \) has the technical advantage that the action \( \int_{\mathbb{R}^d} \left| \frac{d\tilde{j}}{d\rho} (x) \right|^2 \, d\rho(x) \) is jointly convex
and lower semicontinuous in \( \rho \) and \( \tilde{j} \) (this is a well-known consequence of Reshetnyak’s
theorem; for completeness, we provide a proof of this result which covers the case where
\( \tilde{j} \in \mathcal{M}_{loc}(\mathbb{R}^d) \) in Theorem A.1). In particular, see Remark 2.23 for a useful consequence of
these properties.
2.1 Nonlocal structure: weight kernel, interpolation, vector calculus

We equip $\mathbb{R}^d$ with an “underlying nonlocal structure”, as follows.

**Assumption 2.1 (Weight kernel)** The function $\eta : \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \setminus \{x = y\}\} \to [0, \infty)$ satisfies the following properties:

(i) $\eta$ is continuous on the set $\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \setminus \{x = y\} : \eta(x, y) > 0\}$;
(ii) $\eta$ is isotropic, that is, there exists radial profile $\eta : (0, \infty) \to [0, \infty)$ such that $\eta(x, y) = \eta(|x - y|)$.
(iii) The radial profile $\eta$ is non-increasing.
(iv) $\eta$ satisfies the tail moment bound $\int_{\mathbb{R}^d} 1 \wedge |y|^2 \eta(|y|)dy < \infty$.
Additionally, for some results we require that
(v) The support of $\eta$ contains $(0, 1)$.
Or, furthermore, that
(vi) The support of $\eta$ is equal to $(0, 1)$.

The assumption of isotropy is largely imposed to simplify the statements of our results. When combined with the tail moment bound, these suffice to guarantee that Assumption 1.1 from [19] is satisfied; the relevance for us is that we make use of several results from [19] which require this assumption. On the other hand, the arguments of Sects. 4 and 5 make use of the assumption that $\eta$ is compactly supported. Note also that under the assumption $\eta$ is isotropic, if $\eta$ is compactly supported then the support of $\eta(|y|)$ is equal to $B(0, R)$ for some $R > 0$; in assumption (vi), we fix $R = 1$ merely as a convention. Likewise, in view of assumption (iii) assumption (v) may be viewed as merely a convention unless $\eta$ is identically zero.

Since $\eta$ is taken to be isotropic, we often write $\eta(|x - y|)$ rather than $\eta(x - y)$; in other words, we abusively identify $\eta$ with its radial profile.

We write $M_p(\eta) := \int_{\mathbb{R}^d} |x - y|^p \eta(|x - y|)dy$ to denote the $p$-th central moment of $\eta(x, y)$ with $x$ fixed. Note that due to isotropy, $M_p(\eta)$ does not depend on the choice of $x$. Unless otherwise stated, we do not explicitly assume $p$th moment bounds on $\eta$, but note that Theorem 1.3 assumes that $M_2(\eta) < \infty$.

In what follows, given a choice of weight kernel $\eta$, we denote $G := \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \setminus \{x = y\} : \eta(x, y) > 0\}$. The intended interpretation is that $G$ is the set of edges we have placed on $\mathbb{R}^d$, with $\eta(x, y)$ being the edge weight between $x$ and $y$.

We also assume that the interpolation function $\theta$ satisfies the following:

**Assumption 2.2 (Interpolation function)** $\theta : [0, \infty) \times [0, \infty) \to [0, \infty)$ satisfies the following properties:

(i) Regularity: $\theta$ is continuous on $[0, \infty) \times [0, \infty)$ and $C^1$ on $(0, \infty) \times (0, \infty)$;
(ii) Symmetry: $\theta(s, t) = \theta(t, s)$ for $s, t \geq 0$;
(iii) Positivity, normalisation: $\theta(s, t) > 0$ for $s, t > 0$ and $\theta(1, 1) = 1$;
(iv) Monotonicity: $\theta(r, t) \leq \theta(s, t)$ for all $0 \leq r \leq s$ and $t \geq 0$;
(v) Positive 1-homogeneity: $\theta(\lambda s, \lambda t) = \lambda \theta(s, t)$ for $\lambda > 0$ and $s, t \geq 0$;
(vi) Concavity: the function $\theta : [0, \infty) \times [0, \infty) \to [0, \infty)$ is concave;
(vii) Connectedness: $C_\theta := \int_0^1 \frac{dr}{\theta(1-r,1+r)} \in [0, \infty)$.

**Remark 2.3** Points (i–vi) in the preceding assumption are identical to [19, Assumption 2.1], except that Erbar also assumes that $\theta$ is zero on the boundary, namely $\theta(0, t) = 0$ for all $t \geq 0$. However, a careful reading of [19] indicates that this extra assumption is never used.
(except in the sense that Erbar proves some results only for the logarithmic mean, which is indeed zero on the boundary). We do not assume \( \theta \) is zero on the boundary; moreover, we will see below that whether or not \( \theta \) is zero on the boundary has significant topological consequences for the nonlocal Wasserstein distance.

We call point (vii) “connectedness” because, as discussed in [37] if instead \( C_\theta = \infty \), the discrete \( W_\theta \) distance on the space of probability measures supported on a symmetric graph with two points becomes topologically disconnected. The assumption that \( C_\theta < \infty \) is required for several arguments in Sect. 3.

**Lemma 2.4** Any \( \theta \) satisfying Assumption 2.2 also satisfies \( \theta(r, s) \leq \frac{r + s}{2} \).

**Proof** First, note that by 1-homogeneity and the normalization \( \theta(1, 1) = 1 \), we have \( \theta(r, r) = r \) for all \( r \in \mathbb{R}_+ \). At the same time, symmetry implies that at any point along the line \( (r, r) \) in \( \mathbb{R}_+ \times \mathbb{R}_+ \) (excluding \( (0, 0) \), the directional derivative of \( \theta(r, r) \) in the direction orthogonal to the vector \( (1, 1) \) must be zero. Therefore, by concavity, \( \theta(r, s) \) is upper-bounded by the hyperplane which takes the value \( r \) at \( (r, r) \) and has directional derivative zero in the direction orthogonal to \( (1, 1) \) at every point \( (r, r) \), and the only such hyperplane is given by \( \frac{r + s}{2} \).

**Definition 2.5** (Nonlocal gradient and divergence) For any function \( \phi : \mathbb{R}^d \to \mathbb{R} \) we define its nonlocal gradient \( \nabla \phi : G \to \mathbb{R} \) by

\[
\nabla \phi(x, y) = \phi(y) - \phi(x) \text{ for all } (x, y) \in G.
\]

For any \( j \in \mathcal{M}_{loc}(G) \), its nonlocal divergence \( \nabla \cdot j : \mathcal{M}_{loc}(\mathbb{R}^d) \) is defined as the \( \eta \)-weighted adjoint of \( \nabla \), i.e.,

\[
\int_{\mathbb{R}^d} \phi d(\nabla \cdot j) = -\frac{1}{2} \int_G \nabla \phi(x, y) \eta(x, y)dj(x, y).
\]

**Remark 2.6** It is the nonlocal divergence operator \( \nabla \cdot \) which replaces the usual divergence operator in the Definition 2.12. However, we will simply write out the integral operator \( \frac{1}{2} \int_G \nabla \phi(x, y) \eta(x, y)dj(x, y) \) explicitly in the sequel; the definition here is just presented for easier comparison with articles such as [19, 25]. Moreover, it is not obvious under what circumstances the measure \( \nabla \cdot j \) exists, even assuming that the integral \( \frac{1}{2} \int_G \nabla \phi(x, y) \eta(x, y)dj(x, y) \) is finite. Additionally, we caution the reader that other conventions for the definition of nonlocal gradient and divergence operator are present in the literature, in particular our definitions are not the same as the ones presented in [15].

### 2.2 Action

We rigorously define the action in the “flux form”.

**Definition 2.7** (Action) Let \( \eta \) satisfy Assumption 2.1 and \( \theta \) satisfy Assumption 2.2. Let \( \mu \in \mathcal{P}(\mathbb{R}^d) \) and \( j \in \mathcal{M}_{loc}(G) \). Let \( m \in \mathcal{M}_{loc}(\mathbb{R}^d) \) be any reference measure. Define the action of the pair \((\mu, j)\) by

\[
\mathcal{A}_{\theta, \eta}(\mu, j; m) := \frac{1}{2} \int_G \theta \left( \frac{d(j \otimes \mu)}{d\lambda}(x, y) \right)^2 \eta(x, y)d\lambda(x, y)
\]

where \( \lambda \) is taken to be any nonnegative measure in \( \mathcal{M}_{loc}^+(G) \) such that \( |j|, \mu \otimes m, m \otimes \mu \ll \lambda \). Here, the fraction in the integrand is understood with the convention that \( 0/0 = 0 \).
If θ and η are obvious from context, and m is chosen to be Leb, the Lebesgue measure on \( \mathbb{R}^d \), we simply write \( \mathcal{A}(\mu, j) \).

**Remark 2.8** Note that \( \mathcal{A}_{\theta,\eta}(\mu, j; m) \) does not depend on the choice of \( \lambda \), satisfying this domination condition \( |j|, \mu \otimes m, m \otimes \mu \ll \lambda \), since \( \theta \) is 1-homogeneous.

**Remark 2.9** The choice of “reference measure” \( m \) in Definition 2.7 allows us to encode alternate geometries on \( \mathbb{R}^d \) besides the usual Euclidean one; phrased differently, Definition 2.7 makes sense when working with any metric measure space \( (\mathbb{R}^d, d, m) \) (however if we use a metric \( d \) other than the Euclidean one, note this would change what it means for \( \eta \) to be isotropic). In particular, if we consider a weighted graph \( G = (V, E, w, m_n) \) with vertices in \( \mathbb{R}^d \), where \( w(x_i, x_j) = \eta(x_i, x_j) \) for any \((x_i, x_j) \in E\), and \( m_n \) is a measure supported on \( V \), then selecting \( m_n \) as the reference measure in Definition 2.7 causes \( \mathcal{A}_{\eta,\theta,m_n} \) to coincide with the action associated to the graph Wasserstein distance discussed in the introduction, if we also restrict to the case where \( \mu \ll m_n \) and \( j \ll \sum_{i,j} w(x_i, x_j)m_n(x_i)m_n(x_j) \).

**Lemma 2.10** The action \( \mathcal{A}_{\theta,\eta}(\mu, j; m) \) is jointly convex in \( (\mu, j) \), and is jointly lower semicontinuous in \( (\mu, j) \) with respect to the narrow topology on \( \mathcal{P}(\mathbb{R}^d) \) and the weak* topology on \( \mathcal{M}_{loc}(G) \).

**Proof** This is proved in Corollary A.2. \( \square \)

**Remark 2.11** (Comparison with action given in Erbar) Our Definition 2.7 is superficially different from the definition of the action given in [19, Section 2]. Nonetheless, it can be seen that the two definitions are, in fact, equivalent, and so our choice of an alternate presentation of the action is largely one of taste.

Specialized to our setting, Erbar defined the action \( \mathcal{A}'(\mu, v) \) of a pair \( (\mu, v) \in \mathcal{P}(\mathbb{R}^d) \times \mathcal{M}_{loc}(G) \) by

\[
\mathcal{A}'_{\theta,\eta}(\mu, v) := \int_G \frac{\left( \frac{d\mu_1}{dz}(x, y) \right)^2}{2\theta \left( \frac{d\mu_1}{dz}(x, y), \frac{d\mu_2}{dz}(x, y) \right)} d\lambda(x, y)
\]

where \( d\mu_1 = \eta(x, y)d\mu(x)m(y) \) and \( d\mu_2 = \eta(x, y)dm(x)d\mu(y) \), and \( \lambda \) is any measure in \( \mathcal{M}_{loc}(G) \) dominating \( v, \mu^1, \) and \( \mu^2 \) (and the fraction in the integrand is understood with the convention that \( \frac{0}{0} = 0 \)). Now, given a pair \( (\mu, j) \in \mathcal{P}(\mathbb{R}^d) \times \mathcal{M}_{loc}(G) \), it is routine to check (by using the chain rule for Radon-Nikodym derivatives and the 1-homogeneity of \( \theta \)) that

\[
\mathcal{A}_{\theta,\eta}(\mu, j) = \mathcal{A}'_{\theta,\eta}(\mu, v)
\]

provided that \( v \in \mathcal{M}_{loc}(G) \) is defined so that \( \frac{\partial v}{\partial y}(x, y) = \eta(x, y) \). And conversely, if first given \( (\mu, v) \), it holds that \( \mathcal{A}_{\theta,\eta}(\mu, j) = \mathcal{A}'_{\theta,\eta}(\mu, v) \) provided that \( j \in \mathcal{M}_{loc}(G) \) is defined so that \( \frac{\partial j}{\partial y} = \frac{1}{\eta(x, y)} \); note that this latter definition (of \( j \)) is unproblematic on \( G \), since by definition \( \eta(x, y) > 0 \) everywhere on \( G \).

### 2.3 Nonlocal continuity equation

We define the weak solutions of the nonlocal continuity Eq. (1.3), in flux form, in the same way as [25, Section 2.3].
Definition 2.12 (Nonlocal continuity equation) Let $T > 0$. We say that $(\mu_t, j_t)_{t \in [0, T]}$ solves the nonlocal continuity equation provided that

(i) $\mu(\cdot) : [0, T] \to \mathcal{P}(\mathbb{R}^d)$ is continuous when $\mathcal{P}(\mathbb{R}^d)$ is equipped with the narrow topology,

(ii) $j_1 : [0, T] \to \mathcal{M}_{loc}(G)$ is Borel when $\mathcal{M}_{loc}(G)$ is equipped with the weak* topology,

(iii) $\forall \varphi \in C_c^\infty([0, T] \times \mathbb{R}^d)$

$$\int_0^T \int_{\mathbb{R}^d} \partial_t \varphi_t(x) d\mu_t(x) dt + \frac{1}{2} \int_0^T \int_G \nabla \varphi_t(x, y) \eta(x, y) d\nu_t(x, y) dt = 0. \quad (2.1)$$

We write $(\mu_t, j_1) \in \mathcal{CE}_T$ to indicate that $(\mu_t, j_1)_{t \in [0, T]}$ satisfies conditions (i), (ii), and (iii) above. Furthermore, we write $\mathcal{CE}$ to denote $\mathcal{CE}_1$, and write $(\mu_t, j_1) \in \mathcal{CE}_T(v, \sigma)$ to indicate that $(\mu_t, j_1) \in \mathcal{CE}_T$ and $\mu_0 = v$ and $\mu_1 = \sigma$.

Remark 2.13 (Comparison with nonlocal continuity equation given in [19]) In [19], a slightly different nonlocal continuity equation is considered. There, Eq. 1.3 is replaced with the following equation:

$$\forall \varphi \in C_c^\infty([0, T] \times \mathbb{R}^d)$$

$$\int_0^T \int_{\mathbb{R}^d} \partial_t \varphi_t(x) d\mu_t(x) dt + \frac{1}{2} \int_0^T \int_G \nabla \varphi_t(x, y) d\nu_t(x, y) dt = 0. \quad (2.2)$$

Here, $\nu(\cdot) : [0, T] \to \mathcal{M}_{loc}(G)$ is likewise assumed to be weak* Borel. It is evident that if $(\mu_t, j_1)_{t \in [0, T]}$ satisfies the nonlocal continuity equation in the sense of Definition 2.12, then $(\mu_t, \nu_t)$ satisfies 2.2, provided that $\nu_t \in \mathcal{M}_{loc}(G)$ is defined so that $\frac{d\nu}{dt}(x, y) = \eta(x, y)$. Conversely, if $(\mu_t, \nu_t)$ satisfies 2.2, then defining $j_1$ so that $\frac{d}{dt} \nu_t(x, y) = \frac{1}{\eta(x, y)}$, we see that $(\mu_t, j_1)_{t \in [0, T]}$ satisfies the nonlocal continuity equation in the sense of Definition 2.12. Note that this latter definition (of $j_1$) is unproblematic on $G$, since by definition $\eta(x, y) > 0$ everywhere on $G$.

It is sometimes advantageous to work with a stronger notion of solution to the nonlocal continuity equation:

Definition 2.14 Let $\rho_t(x) : [0, 1] \times \mathbb{R}^d \to \mathbb{R}_+$ be a probability density which is differentiable in $t$, and let $j_t(x, y) : [0, 1] \times G \to \mathbb{R}$. We say that $(\rho_t, j_t)_{t \in [0, 1]}$ is a classical solution to the nonlocal continuity equation provided that $j_t(x, y)$ is antisymmetric and

$$\partial_t \rho_t(x) + \int_{\mathbb{R}^d} j_t(x, y) \eta(x, y) dy = 0$$

holds pointwise in $t$ and $x$.

Note that if $(\rho_t, j_t)_{t \in [0, 1]}$ is a classical solution to the nonlocal continuity equation, then it holds that the time-dependent measures $(\rho_t dx, j_t dx dy)_{t \in [0, 1]}$ are again a solution to the nonlocal continuity equation in the sense of Definition 2.12.

2.4 Nonlocal Wasserstein metric

Definition 2.15 Let $\eta$ satisfy Assumption 2.1 (i–iv) and $\theta$ satisfy Assumption 2.2, and let $m \in \mathcal{M}_{loc}^+(\mathbb{R}^d)$. The nonlocal Wasserstein distance $\mathcal{W}_{\eta, \theta, m}$ on $\mathcal{P}(\mathbb{R}^d)$ is defined by

$$\mathcal{W}_{\eta, \theta, m}^2(v, \sigma) := \inf \left\{ \int_0^1 A_{\eta, \theta, m}(\mu_t, j_t) dt : (\mu_t, j_t) \in \mathcal{CE}(v, \sigma) \right\}.$$
We will write \( \mathcal{W}_{\eta, \theta} \) to denote the case where \( m = \text{Leb} \). Furthermore, we will usually drop the explicit reference to the choice of \( \theta \), and simply write \( \mathcal{W}_{\eta} \).

**Remark 2.16** In view of Remarks 2.11 and 2.13, our definition of the nonlocal Wasserstein distance is equivalent to a special case of the nonlocal Wasserstein distance defined in [19].

**Fact 2.17** \( \mathcal{W}_{\eta, \theta, m} \) is a pseudometric on \( \mathcal{P}(\mathbb{R}^d) \). On \( \mathcal{P}(\mathbb{R}^d) \), \( \mathcal{W}_{\eta, \theta, m}^2 \) is jointly convex, and \( \mathcal{W}_{\eta, \theta, m} \) is jointly narrowly lower semicontinuous.

The proof of this fact is exactly as in [19] and is therefore omitted.

**Lemma 2.18** (antisymmetric flux) Let \((\mu_t, \mathbf{j}_t)_{t \in [0, 1]}\) solve the nonlocal continuity equation. Let \( \mathbf{j}_t^{as} := (\mathbf{j} - \mathbf{j}_t^{T})/2 \). Then

(i) \((\mu_t, \mathbf{j}_t^{as})_{t \in [0, 1]}\) also solves the nonlocal continuity equation.

(ii) \( \mathcal{A}(\mu_t, \mathbf{j}_t^{as}) \leq \mathcal{A}(\mu_t, \mathbf{j}_t) \) for all \( t \in [0, 1] \).

In particular, given any \( \mathcal{W}_{\eta, \theta, m} \) geodesic \((\mu_t)_{t \in [0, 1]}\), there exists a tangent flux which is antisymmetric.

**Proof**

The proof of (i) is identical to the argument presented in [25, Corollary 2.8] and is therefore omitted.

(ii) This likewise follows from a minor modification of arguments given in [25, Lemma 2.6 and Corollary 2.8]. Namely, we reason from Lemma 2.10, in particular from the convexity of the action in the \( \mathbf{j} \) variable:

\[
\mathcal{A}(\mu_t, \mathbf{j}_t^{as}) \leq \frac{1}{2} \mathcal{A}(\mu_t, \mathbf{j}_t) + \frac{1}{2} \mathcal{A}(\mu_t, -\mathbf{j}_t^{T}).
\]

Now, selecting \( \lambda_t \in \mathcal{M}(G) \) so that \( \lambda_t \gg -\mathbf{j}_t^{T} \),

\[
\mathcal{A}_{\eta, \theta}(\mu_t, -\mathbf{j}_t^{T}) = \int_G 2\theta \left( \frac{d(-\mathbf{j}_t^{T})(x,y)}{d\lambda_t} \right)^2 \eta(x,y) d\lambda_t(x,y).
\]

Note that without loss of generality we can take \( \lambda_t \) to also dominate \( \mathbf{j}_t \), and furthermore we can take \( \lambda_t \) to be symmetric (by replacing \( \lambda_t \) with \( (\lambda_t + \lambda_t^{T})/2 \)). Then, computing that

\[
\frac{d(-\mathbf{j}_t^{T})}{d\lambda_t} = -\left( \frac{d\mathbf{j}_t}{d\lambda_t} \right)^T
\]

and using the fact that \( \theta \left( \frac{d(\mu \otimes m)}{d\lambda}(x,y), \frac{d(m \otimes \mu)}{d\lambda}(x,y) \right), \eta(x,y) \), and \( \lambda_t \) are all symmetric, we see that

\[
\int_G 2\theta \left( \frac{d(-\mathbf{j}_t^{T})}{d\lambda_t} \right)^2 \eta d\lambda_t = \int_G 2\theta \left( \frac{d\mathbf{j}_t}{d\lambda_t} \right)^2 \eta d\lambda_t.
\]

Consequently, \( \mathcal{A}(\mu_t, -\mathbf{j}_t^{T}) = \mathcal{A}(\mu_t, \mathbf{j}_t) \), and so \( \mathcal{A}(\mu_t, \mathbf{j}_t^{as}) \leq \mathcal{A}(\mu_t, \mathbf{j}_t) \).  

**Remark 2.19** Thus far, we have only defined \( \mathcal{W}_{\eta, \theta, m}(\mu, \nu) \) in the case where \( \mu, \nu \in \mathcal{P}(\mathbb{R}^d) \). However, it is occasionally useful to consider the nonlocal Wasserstein distance between
We say that \( k \) we denote by \( \mathcal{L} K \) and \( M \) have the following well-definedness and homogeneity result:

**Proposition 2.20** (Extension of \( W_{\eta, \theta, m} \) to general nonnegative measures) Let \( \mu, v \in \mathcal{M}^+(\mathbb{R}^d) \). Suppose that \( \|\mu\|_{TV} = \|v\|_{TV} = M > 0 \). Then,

\[
W_{\eta, \theta, m}^2(\mu, v) = MW_{\eta, \theta, m}^2 \left( \frac{\mu}{M}, \frac{v}{M} \right).
\]

In particular, \( W_{\eta, \theta, m}^2(\mu, v) < \infty \) iff \( W_{\eta, \theta, m}^2 \left( \frac{\mu}{M}, \frac{v}{M} \right) < \infty \).

**Proof** Let \((\rho_t, j_t)_{t \in [0,1]} \in CE \left( \frac{\mu}{M}, \frac{v}{M} \right)\), and suppose that \( W_{\eta, \theta, m}^2 \left( \frac{\mu}{M}, \frac{v}{M} \right) = \int_0^1 A(\rho_t, j_t) dt \). Then, \((M \rho_t, M j_t)_{t \in [0,1]}\) is also a solution to the nonlocal continuity equation, with endpoints \( \mu \) and \( v \); consequently, using the 1-homogeneity of the action,

\[
W_{\eta, \theta, m}^2(\mu, v) \leq \int_0^1 A(M \rho_t, M j_t) dt = M \int_0^1 A(\rho_t, j_t) dt = MW_{\eta, \theta, m}^2 \left( \frac{\mu}{M}, \frac{v}{M} \right).
\]

By identical reasoning, we also deduce that \( W_{\eta, \theta, m}^2 \left( \frac{\mu}{M}, \frac{v}{M} \right) \leq \frac{1}{M} W_{\eta, \theta, m}^2(\mu, v) \). \( \square \)

### 2.5 Convolutions

We make frequent use of convolution estimates in Sects. 4 and 5. A number of elementary computations relating to convolutions are deferred to Appendix B; here, we fix notation and state some basic convolution stability results concerning the \( W_\eta \) distances. Lastly, we show that a specific convolution kernel, the Laplace kernel \( K := c e^{-|x-y|} \), has the property that smoothed densities \( K \ast \mu \) have relative Lipschitz regularity of the form \( \frac{K \ast \mu(x)}{K \ast \mu(x')} \leq 1 + C|x - x'| \), a fact which we exploited repeatedly in Sects. 4 and 5.

**Definition 2.21** We say that \( (0, \infty) \to [0, \infty) \) is a convolution kernel if it is \( C^1 \) on its support and normalized so that \( \int_{\mathbb{R}^d} k(|x|) dx = 1 \).

Given a convolution kernel \( k \) and a (possibly signed, possibly vector-valued) measure \( \mu \), we denote by \( k \ast \mu \) the convolution of \( k(\cdot) \) and \( \mu \):

\[
k \ast \mu(x) := \int k(|x - y|) d\mu(y).
\]

We also use the following notation: given a convolution kernel \( k \) and a measure \( \mu \), we write \( k \ast \mu \) to denote the measure whose Lebesgue density is given by

\[
\frac{d(k \ast \mu)}{dx} = k \ast \mu.
\]

Separately, for measures \( j \in M_{loc}(G) \), we follow [19] and define the convolution \( k \ast j \in M_{loc}(G) \) of a convolution kernel \( k \) (on \( \mathbb{R}^d \)) with \( j \) as follows:

\[
k \ast j = \int_{\mathbb{R}^d} k(|z|) d\gamma(x - z, y - z) dz
\]
in other words, for all $\varphi \in C_c^\infty(G)$,

$$\int_G \varphi(x, y) d(k * j)(x, y) = \int_{\mathbb{R}^d} k(|z|) \varphi(x + z, y + z) d\jmath(x, y) dz.$$ 

Note that this definition may be understood as a special case of convolution with respect to a translation-invariant group action on a space: in this case, $\mathbb{R}^d$ acts on $G$ by the translation $(x, y) \mapsto (x + z, y + z)$, and this action is indeed translation-invariant since $\eta(|x - y|)$ is preserved under this translation.

**Proposition 2.22** (Stability of action and metric under convolution) Let $k$ be any convolution kernel.

(i) For any $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\jmath \in \mathcal{M}_{loc}(G)$, $A_{\theta, \eta}(k * \mu, k * \jmath) \leq A_{\theta, \eta}(\mu, \jmath)$.

(ii) Let $(\mu_t, \jmath_t)_{t \in [0, 1]} \in CE$. Then, $(k * \mu_t, k * \jmath_t)_{t \in [0, 1]} \in CE$ also.

(iii) ([19, Proposition 4.8]) For any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$, $W_{\theta, \eta}(k * \mu_0, k * \mu_1) \leq W_{\theta, \eta}(\mu_0, \mu_1)$.

**Proof** (i) In Lemma A.3, we show that

$$A_{\theta, \eta}(k * \mu, k * \jmath) \leq \int_{\mathbb{R}^d} A_{\theta, \eta}(\mu_z, \jmath_z) k(z) dz.$$ 

The result now follows according to the reasoning given in [19, proof of Proposition 2.8]. (ii) follows by identical reasoning to [19, proof of Proposition 4.8]. Finally, (iii) follows by combining (i) and (ii): indeed, letting $(\mu_t, \jmath_t)_{t \in [0, 1]}$ be an action-minimizing solution to the nonlocal continuity equation with endpoints $\mu_0$ and $\mu_1$, we see that

$$W_{\theta, \eta}^2(\mu_0, \mu_1) = \int_0^1 \mathcal{A}(\mu_t, \jmath_t) dt \geq \int_0^1 \mathcal{A}(k * \mu_t, k * \jmath_t) dt \geq W_{\theta, \eta}(k * \mu_0, k * \mu_1).$$

\[\square\]

**Remark 2.23** Similarly, it is known [3, Lemmas 8.1.9 and 8.1.0] that if $(\rho_t, \jmath_t)_{t \in [0, 1]}$ solves the (local) continuity equation $\partial_t \rho_t + \text{div} \jmath_t = 0$ in the sense of distributions, and $k$ is a convolution kernel, then $(k * \rho_t, k * \jmath_t)_{t \in [0, 1]}$ is again a solution of $\partial_t \rho_t + \text{div} \jmath_t = 0$ in the sense of distributions; and similarly,

$$\int_{\mathbb{R}^d} \left| \frac{d(k * \jmath_t)}{d(k * \mu_t)} \right|^2 d(k * \mu_t) \leq \int_{\mathbb{R}^d} \left| \frac{d\jmath_t}{d\mu_t} \right|^2 d\mu_t.$$ 

Therefore, applying the mass-flux presentation of the $W_2$ metric described at the beginning of Sect. 2 above, we can reason exactly as in the proof of part (iii) of the previous proposition to deduce that for any convolution kernel $k$ and probability measures $\mu_0$ and $\mu_1$,

$$W_2(k * \mu_0, k * \mu_1) \leq W_2(\mu_0, \mu_1).$$

### 2.5.1 Relative Lipschitz estimate for the right convolution

Let $K(x) = c_K e^{-|x|}$ where $\frac{1}{c_K} = \int_{\mathbb{R}^d} e^{-|x|} dx$. Let $K_\delta(x) = \frac{1}{\delta} K \left( \frac{x}{\delta} \right)$, where $\delta > 0$.

**Lemma 2.24** Consider $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. Let $\mu_\delta = K_\delta \ast \mu$ Then

$$\ln \frac{\mu_\delta(y)}{\mu_\delta(x)} \leq \frac{1}{\delta} |y - x|.$$
Furthermore if \(|x - y| \leq \delta\) then

\[
\mu_\delta(y) \leq \mu_\delta(x) \left(1 + \frac{3}{\delta}|y - x|\right)
\]

**Proof** Let \(h = y - x\). We can assume \(h \neq 0\). Compute that

\[
|\ln \mu_\delta(x + h) - \ln \mu_\delta(x)| \leq \ln \left(\int e^{-|x - z|/\delta} e^{h|z|/\delta} d\mu(z)\right) = \frac{|h|}{\delta}.
\]

Therefore

\[
\frac{\ln \mu_\delta(y)}{\ln \mu_\delta(x)} \leq \frac{1}{\delta} |y - x|.
\]

Hence, for \(|y - x| < \delta\),

\[
\mu_\delta(y) \leq \mu_\delta(x) e^{|y - x|/\delta} \leq \mu_\delta(x) \left(1 + \frac{3}{\delta}|y - x|\right).
\]

\(\square\)

3 Metric structure of \(\mathcal{W}_{\eta, \theta}\) distances

3.1 General lower bounds for nonlocal Wasserstein distances

In this subsection, we consider nonlocal Wasserstein distances \(\mathcal{W}_{\eta, \theta, m}\) with general reference measure \(m\), since this complicates our analysis only minimally. As in Definition 2.15, we assume that \(\eta\) satisfies Assumption 2.1 (i–iv), and that \(\theta\) satisfies Assumption 2.2.

We first recall the following result of Erbar (which we specialize somewhat), which gives a partial characterization of the topology induced by the nonlocal Wasserstein distance.

**Proposition 3.1** ([19, Proposition 4.5]) Suppose that \(\rho_0, \rho_1 \in \mathcal{P}(\mathbb{R}^d)\). The nonlocal Wasserstein distance \(\mathcal{W}_{\eta, \theta, m}\) with arbitrary reference measure \(m\) of Definition 2.15 satisfies

\[
\sqrt{\frac{2}{C}} W_1(\rho_0, \rho_1) \leq \mathcal{W}_{\eta, \theta, m}(\rho_0, \rho_1).
\]

Here \(\tilde{C} = \sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} |x - y|^2 \eta(|x - y|) dm(y)\).

In particular, observe that this \(W_1\) lower bound is vacuous in the case where \(m = \text{Leb}\) and the second moment of \(\eta\) is infinite.

When the lower bound in the previous proposition is non-vacuous, this shows, in particular, that the topology induced by \(\mathcal{W}_{\eta, \theta}\) as strong or stronger than the narrow topology on \(\mathcal{P}(\mathbb{R}^d)\).

What we show below is that, when \(\eta\) is integrable, the topology is strictly stronger. More precisely we show that the topology induced by \(\mathcal{W}_{\eta, \theta, m}\) is at least as strong as the strong topology on \(\mathcal{P}(\mathbb{R}^d)\). This indicates that the nonlocal Wasserstein distances are fundamentally different from the standard Wasserstein distances.

**Proposition 3.2** Suppose that \(\rho_0, \rho_1 \in \mathcal{P}(\mathbb{R}^d)\). The nonlocal Wasserstein distance \(\mathcal{W}_{\eta, \theta, m}\) with arbitrary reference measure \(m\) of Definition 2.15 satisfies

\[
\sqrt{\frac{2}{C}} TV(\rho_0, \rho_1) \leq \mathcal{W}_{\eta, \theta, m}(\rho_0, \rho_1).
\]
Here $C = \sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} \eta(|x - y|)dm(y)$.  

This lower bound is vacuous when $\eta$ is non-integrable. This suggests that the nonlocal transportation distance induces a weaker topology when $\eta$ is non-integrable; we address this point later on in Lemma 3.11.

**Proof** Recall that one of the several equivalent definitions of the TV norm is as follows:

$$TV(\rho_0, \rho_1) = \sup_{A \in \mathcal{B}(\mathbb{R}^d)} |\rho_0(A) - \rho_1(A)|.$$  

Let $A$ be some measurable set such that $|\rho_0(A) - \rho_1(A)| > TV(\rho_0, \rho_1) - \frac{\varepsilon}{2}$. Any mass-flux pair $(\rho_t, j_t)$ connecting $\rho_0$ to $\rho_1$ must therefore move at least $TV(\rho_0, \rho_1) - \varepsilon$ of mass from $A$ to $A^c$ (or vice versa). Without loss of generality, we can take $A$ to be compact.

Let $(\rho_t, j_t) \in CE(\rho_0, \rho_1)$ be an action-minimizing mass-flux pair, so that $\mathcal{W}^2_{\eta, \theta, m}(\rho_0, \rho_1) = \int_0^1 A(\rho_t, j_t)dt$; without loss of generality we can assume, by [19, Proposition 4.3], that $(\rho_t, j_t)$ is unit speed in the sense that $\mathcal{W}^2_{\eta, \theta, m}(\rho_0, \rho_1) = A(\rho_t, j_t)$ for almost all $t \in [0, 1]$. We may also assume, without loss of generality, that $j_t$ is antisymmetric for almost all $t$, by Lemma 2.18.

Let $\lambda_t$ be some measure which dominates all of $j_t$, $m \otimes \rho_t$, and $\rho_t \otimes m$. The action for $(\rho_t, j_t)$ is

$$A_{\theta}(\rho_t, j_t) = \frac{1}{2} \iint_{G} \left( \theta \left( \frac{d(j_t \otimes \rho_t)}{d\lambda_t}(x, y), \frac{d(m \otimes \rho_t)}{d\lambda_t}(x, y) \right)^2 \right) \eta(x, y)d\lambda_t(x, y).$$

Applying the Cauchy-Schwarz inequality, and using the fact that $\frac{d(j_t \otimes \rho_t)}{d\lambda_t} = \frac{d(j_t \otimes 1)}{d\lambda_t}$, we get

$$A_{\theta}(\rho_t, j_t) \geq \frac{1}{2} \left( \iint_{G} \frac{d(j_t \otimes 1)}{d\lambda_t}(x, y) \eta(x, y)d\lambda_t(x, y) \right)^2 \times \left( \iint_{G} \theta \left( \frac{d(\rho_t \otimes m)}{d\lambda_t}(x, y), \frac{d(m \otimes \rho_t)}{d\lambda_t}(x, y) \right) \eta(x, y)d\lambda_t(x, y) \right)^{-1}.$$  

By Lemma 2.4, $\theta$ automatically satisfies $\theta(r, s) \leq (r + s)/2$, so we find that

$$\left( \iint_{G} \theta \left( \frac{d(\rho_t \otimes m)}{d\lambda_t}(x, y), \frac{d(m \otimes \rho_t)}{d\lambda_t}(x, y) \right) \eta(x, y)d\lambda_t(x, y) \right)^{-1} \leq \frac{1}{2} \left( \iint_{G} \left( \frac{d(\rho_t \otimes m)}{d\lambda_t}(x, y) + \frac{d(m \otimes \rho_t)}{d\lambda_t}(x, y) \right) \eta(x, y)d\lambda_t(x, y) \right).$$

By the Radon-Nikodym theorem, we have the estimate

$$\iint_{G} \frac{d(\rho_t \otimes m)}{d\lambda_t}(x, y) \eta(x, y)d\lambda_t(x, y) = \iint_{G} \eta(x, y)d(\rho_t \otimes m)(x, y) \leq \int_{\mathbb{R}^d} \left[ \sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} \eta(|x - y|)dm(y) \right] d\rho_t(x)$$

$$= \sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} \eta(|x - y|)dm(y) =: C.$$
The same estimate works if we replace \( \rho_t(y) \) with \( \rho_t(x) \); hence, we conclude that
\[
\int_G \theta \left( \frac{d(\rho_t \otimes m)}{d\lambda_t}(x, y), \frac{d(m \otimes \rho_t)}{d\lambda_t}(x, y) \right) \eta(x, y) d\lambda_t(x, y) \leq C.
\]
Therefore, (applying the Radon–Nikodym theorem once more)
\[
A_\theta(\rho_t, j_t) \geq \frac{1}{2C} \left( \int_G \frac{d|j_t|}{d\lambda_t}(x, y) \eta(x, y) d\lambda_t \right)^2 = \frac{1}{2C} \left( \int_G \eta(x, y) d|j_t|(x, y) \right)^2.
\]

Let \( \xi_\delta \) be a smooth cutoff function for the set \( A \), more precisely \( \xi_\delta = 1 \) on \( A \), \( 0 \) on \( A^c \), and smooth on \( \mathbb{R}^d \) (the existence of such a \( \xi_\delta \) is guaranteed by the smooth version of Urysohn’s lemma, see [35, Proposition 2.2.5]). We use \( \xi_\delta(x) \) as a test function in the nonlocal continuity equation: by [25, Lemma 2.15] we find that
\[
\int_{\mathbb{R}^d} \xi_\delta(x) d\rho_1(x) - \int_{\mathbb{R}^d} \xi_\delta(x) d\rho_0 = -\frac{1}{2} \int_0^1 \int_G (\xi_\delta(y) - \xi_\delta(x)) \eta(x, y) d|j_t|(x, y) dt.
\]
Note that \( |\xi_\delta(y) - \xi_\delta(x)| \leq 1 \), so
\[
\left| \int_{\mathbb{R}^d} \xi_\delta(x) d\rho_1(x) - \int_{\mathbb{R}^d} \xi_\delta(x) d\rho_0 \right| \leq \frac{1}{2} \int_0^1 \int_G (\xi_\delta(y) - \xi_\delta(x)) \eta(x, y) d|j_t|(x, y) dt \leq \frac{1}{2} \int_0^1 \int_G \eta(x, y) d|j_t|(x, y) dt.
\]

Now, selecting \( \delta > 0 \) so that
\[
\left| \left( \int_{\mathbb{R}^d} \xi_\delta(x) d\rho_1(x) - \int_{\mathbb{R}^d} \xi_\delta(x) d\rho_0 \right) - (\rho_1(A) - \rho_0(A)) \right| < \frac{\varepsilon}{2}
\]
we can compute that
\[
TV(\rho_0, \rho_1) \leq |\rho_1(A) - \rho_0(A)| + \frac{\varepsilon}{2}
\]
\[
\leq \left| \int_{\mathbb{R}^d} \xi_\delta(x) d\rho_1(x) - \int_{\mathbb{R}^d} \xi_\delta(x) d\rho_0 \right| + \frac{\varepsilon}{2}
\]
\[
\leq \frac{1}{2} \int_0^1 \int_G \eta(x, y) d|j_t|(x, y) dt + \frac{\varepsilon}{2}
\]
\[
\leq \frac{1}{2} \sqrt{2C} \int_0^1 \sqrt{\mathcal{A}(\rho_t, j_t)} dt + \varepsilon.
\]

Since \( \mathcal{W}_{\eta, \theta, m}(\rho_0, \rho_1) := \int_0^1 \sqrt{\mathcal{A}(\rho_t, j_t)} dt \), and \( \varepsilon > 0 \) was arbitrary, we conclude that
\[
\mathcal{W}_{\eta, \theta, m}(\rho_0, \rho_1) \geq \sqrt{\frac{2}{C}} TV(\rho_0, \rho_1) \quad C = \sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} \eta(|x - y|) dm(y)
\]
as desired. \( \square \)

We also have the following technical corollary, which will be used in Proposition 3.6.

**Corollary 3.3** Suppose that \( C := \sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} \eta(|x - y|) dm(y) < \infty \). Let \( \rho_0, \rho_1 \in \mathcal{P}(\mathbb{R}^d) \), and suppose that \( \mathcal{W}_{\eta, \theta, m}(\rho_0, \rho_1) < \infty \). Let \( (\rho_t, j_t) \in \mathcal{CE}(\rho_0, \rho_1) \) be a constant-speed action-minimizing mass-flux pair, so that \( \mathcal{W}_{\eta, \theta, m}^2(\rho_0, \rho_1) = \int_{t_0}^{t_1} \mathcal{A}(\rho_t, j_t) dt \) for all \( 0 \leq t_0 < t_1 \leq 1 \). Then, for any Borel \( A \subset \mathbb{R}^d \), the function \( t \mapsto \rho_t(A) \) is \( \frac{1}{2} \)-Hölder continuous.
Finally, since \( \nu \) and hence the expel cost is zero: \( \inf_{\rho} (\rho A - \nu) \leq 0 \). Let \( (\tilde{\rho}, \tilde{\mathbf{j}}) \) denote the uniform reparametrization of \( (\rho_t, \mathbf{j}_t)_{t \in [0, t_1]} \) into a unit-time solution to the nonlocal continuity equation; compute that

\[
\int_{t_0}^{t_1} A(\rho_t, \mathbf{j}_t) dt = (t_1 - t_0) \int_0^1 A(\tilde{\rho}_t, \tilde{\mathbf{j}}_t) dt.
\]

But, \( (\tilde{\rho}_t, \tilde{\mathbf{j}}_t)_{t \in [0, 1]} \) is also an action-minimizing solution to the nonlocal continuity equation—otherwise, we could locally replace \( (\rho_t, \mathbf{j}_t)_{t \in [0, t_1]} \) and get a lower-action mass-flux pair connecting \( \rho_0 \) and \( \rho_1 \), which is ruled out by assumption. Therefore, by Proposition 3.2,

\[
(t_1 - t_0)^{1/2} W_{\theta, \eta, m}(\rho_0, \rho_1) \geq \sqrt{\frac{2}{C}} TV(\rho_0, \rho_1).
\]

Now, let \( A \) be any Borel set. Since \( TV(\rho_0, \rho_1) = \sup_{A \in B(\mathbb{R}^d)} |\rho_0(A) - \rho_1(A)| \), we find that

\[
(t_1 - t_0)^{1/2} W_{\theta, \eta, m}(\rho_0, \rho_1) \geq \sqrt{\frac{2}{C}} |\rho_0(A) - \rho_1(A)|.
\]

Finally, since \( W_{\theta, \eta, m}(\rho_0, \rho_1) \leq W_{\theta, \eta, m}(\rho_0, \rho_1) \), we find that

\[
|\rho_0(A) - \rho_1(A)| \leq \sqrt{\frac{2}{C}} W_{\theta, \eta, m}(\rho_0, \rho_1)(t_1 - t_0)^{1/2}
\]

which shows that \( t \mapsto \rho_t(A) \) is \( \frac{1}{2} \)-Hölder continuous, as desired. \( \square \)

### 3.2 Expel problem for \( W_{\eta, \theta} \)

In this subsection we consider the expel problem for nonlocal Wasserstein distances. That is, given a Dirac mass, say at the origin \( \delta_0 \) for concreteness, we wish to estimate \( \inf_{v, x_0, \delta} W_{\eta, \theta}(\delta_0, v) \). Throughout, we only consider the case where the reference measure is the Lebesgue measure. In this subsection, we also assume that \( \eta \) satisfies Assumption 2.1 (i–v), and that \( \theta \) satisfies Assumption 2.2.

We shall make repeated use of an adaptation of a specific computation from [37], which we present separately as Lemma B.1.

**Lemma 3.4** (expel cost upper bound) Let \( 0 < \delta < \varepsilon \) where \( \varepsilon \leq 1 \), and let \( x_0 \in \mathbb{R}^d \). Suppose that there is some \( s > 0 \) and some constant \( c_s > 0 \) such that \( \eta(x, y) \geq c_s |x - y|^{-d-s} \) when \( |x - y| \leq \frac{\delta}{\varepsilon} \). Let \( m_B(x_0, \delta) \) denote the uniform probability measure on the ball \( B(x_0, \delta) \). Then,

\[
W_{\eta, \varepsilon}(\delta x_0, m_B(x_0, \delta)) \leq \frac{C_0}{C_{d,s}} \left( \frac{\delta}{\varepsilon} \right)^{s/2}
\]

where \( C_0 := \int_0^1 \frac{1}{\sqrt{\theta(1-r, 1+r)}} dr \) and \( C_{d,s} \) is given explicitly in the proof.

An important consequence of the result above is that \( W_{\eta, \varepsilon}(\delta x_0, m_B(x_0, \delta)) \to 0 \) as \( \delta \to 0 \) and hence the expel cost is zero: \( \inf_{v, x_0, \delta} W_{\eta, \theta}(\delta_0, v) = 0 \).

**Proof** Given \( x_0 \in \mathbb{R}^d \) and \( r > s > 0 \) we write \( \mathcal{A}(x_0, r) \) to denote \( B(x_0, r) \setminus B(x_0, \frac{r}{2}) \), that is, the annulus of outer radius \( r \) and inner radius \( \frac{r}{2} \). We let \( m_{\mathcal{A}(x_0, r)} \) denote the uniform probability measure on \( \mathcal{A}(x_0, r) \).

\( \square \) Springer
Fix \( \delta > 0 \). Applying Lemma B.1 with \( A = \mathfrak{A}(x_0, \delta 2^{-n}) \) and \( B = \mathfrak{A}(x_0, 2^{-n-1}) \), we find that

\[
\mathcal{W}_{\eta, \epsilon}(\mathfrak{m}_{\mathfrak{A}(x_0, \delta 2^{-n})}, \mathfrak{m}_{\mathfrak{A}(x_0, 2^{-n-1})}) \leq \frac{C_\theta}{4 \sqrt{|\mathfrak{A}(x_0, \delta 2^{-n})| \eta_\epsilon(\delta \frac{3}{2} 2^{-n})}}.
\]

Let us write this upper bound in a more explicit fashion. We know that

\[
|\mathfrak{A}(x_0, \delta 2^{-n})| = \alpha_d (\delta 2^{-n})^d - \alpha_d (\delta 2^{-n-1})^d = \alpha_d \delta^d \left( \frac{1}{2^{nd}} - \frac{1}{2^{nd+d}} \right)
\]

where \( \alpha_d \) is the volume of the \( d \)-dimensional unit ball. On the other hand, \( \eta_\epsilon(\delta \frac{3}{2} 2^{-n}) = \frac{1}{\epsilon^d} \eta(\delta \frac{3}{2} 2^{-n}) \). Suppose now that on \( B(x_0, \delta \epsilon) \), \( \eta(x, y) \geq \epsilon \cdot |x - y|^{-d-s} \) where \( s > 0 \). Then

\[
\eta_\epsilon(\delta \frac{3}{2} 2^{-n}) \geq \frac{1}{\epsilon^d} \alpha_d \left( \frac{\delta}{\epsilon} \right)^{-d-s} \left( \frac{3}{2} 2^{-n} \right)^{-d-s}
\]

so that

\[
|\mathfrak{A}(x_0, \delta 2^{-n})| \eta_\epsilon(\delta \frac{3}{2} 2^{-n}) \geq \alpha_d \delta^d \left( \frac{1}{2^{nd}} - \frac{1}{2^{nd+d}} \right) \cdot \frac{1}{\epsilon^d} c_s \left( \frac{\delta}{\epsilon} \right)^{-d-s} \left( \frac{3}{2} 2^{-n} \right)^{d+s} = \alpha_d \left( \frac{\delta}{\epsilon} \right)^{-s} 2^{ns} \left( 1 - \frac{1}{2^d} \right) c_s \left( \frac{3}{2} \right)^{-d-s}.
\]

Putting \( \tilde{C}_{d,s} = 2 \left( \alpha_d c_s \left( \frac{3}{2} \right)^{-d-s} \right)^{1/2} \), this shows that

\[
\mathcal{W}_{\eta, \epsilon}(\mathfrak{m}_{\mathfrak{A}(x_0, \delta 2^{-n})}, \mathfrak{m}_{\mathfrak{A}(x_0, 2^{-n-1})}) \leq \frac{C_\theta}{\tilde{C}_{d,s}} \left( \frac{\delta}{\epsilon} \right)^{s/2} 2^{-ns/2}.
\]

Summing the geometric series, we find

\[
\sum_{n=0}^{\infty} \mathcal{W}_{\eta, \epsilon}(\mathfrak{m}_{\mathfrak{A}(x_0, \delta 2^{-n})}, \mathfrak{m}_{\mathfrak{A}(x_0, 2^{-n-1})}) \leq \frac{C_\theta}{\tilde{C}_{d,s}} \frac{1}{1 - 2^{s/2}} \left( \frac{\delta}{\epsilon} \right)^{s/2}.
\]

It follows that \( \mathcal{W}_{\eta, \epsilon}(\delta x_0, \mathfrak{m}_{\mathfrak{A}(x_0, \delta)}) \leq \frac{C_\theta}{\tilde{C}_{d,s}} \frac{1}{1 - 2^{s/2}} \left( \frac{\delta}{\epsilon} \right)^{s/2} \). To see why, observe that \( (\mathfrak{m}_{\mathfrak{A}(x_0, \delta 2^{-n})})_{n \in \mathbb{N}} \) converges to \( \delta x_0 \) in \( W_1 \) and thus in the narrow topology. Since \( \mathcal{W}_{\eta, \epsilon} \) is jointly l.s.c. with respect to the narrow topology, we find that

\[
\mathcal{W}_{\eta, \epsilon}(\delta x_0, \mathfrak{m}_{\mathfrak{A}(x_0, \delta)}) \leq \liminf_{k \to \infty} \mathcal{W}_{\eta, \epsilon}(\mathfrak{m}_{\mathfrak{A}(x_0, \delta 2^{-k})}, \mathfrak{m}_{\mathfrak{A}(x_0, \delta)})
\]

\[
\leq \liminf_{k \to \infty} \sum_{n=0}^{k} \mathcal{W}_{\eta, \epsilon}(\mathfrak{m}_{\mathfrak{A}(x_0, \delta 2^{-n})}, \mathfrak{m}_{\mathfrak{A}(x_0, 2^{-n-1})})
\]

\[
\leq \frac{C_\theta}{\tilde{C}_{d,s}} \frac{1}{1 - 2^{s/2}} \left( \frac{\delta}{\epsilon} \right)^{s/2}.
\]

Finally, we can easily upper bound \( \mathcal{W}_{\eta, \epsilon}(\mathfrak{m}_{\mathfrak{A}(x_0, \delta)}, \mathfrak{m}_{B(x_0, \delta)}) \). We use yet another construction based on the \( \mathcal{W} \) geodesic in the two-point space: we use exactly the same computation.
as in the proof of Lemma B.1. Indeed, consider the curve \( \rho_t : [0, 1] \to \mathcal{P}(\mathbb{R}^d) \) defined by

\[
d\rho_t \over d\text{Leb}(x) = \begin{cases} 
1 & x \in \mathcal{A}(x_0, \delta) \\
\frac{1 - \gamma_t}{2|\mathcal{A}(x_0, \delta)|} & x \in B(x_0, \delta^{-1}) \\
0 & \text{else.}
\end{cases}
\]

Additionally, let \( j \) be chosen exactly as in the proof of Lemma B.1. This \( (\rho_t, j) \) is constructed so that \( \rho_0 = m_{\mathcal{A}(x_0, \delta)} \) and \( \rho_1 = m_{B(x_0, \frac{1}{2})} \), and so that the mass on \( \mathcal{A}(x_0, \delta) \) is decreasing uniformly on the set, and continuously in time; therefore, there is a \( t_0 \in (0, 1) \) such that \( \rho_{t_0} \) has uniform distribution on \( B(x_0, \delta) \).

In particular, it follows that

\[
\mathcal{W}_{\eta, \epsilon}(m_{\mathcal{A}(x_0, \delta)}, m_{B(x_0, \delta)}) \leq \sqrt{\int_0^1 A(\rho_t, j)dt}.
\]

Note however that \( \sqrt{\int_0^1 A(\rho_t, j)dt} \) is none other than the upper bound for \( \mathcal{W}_{\eta, \epsilon}(m_{\mathcal{A}(x_0, \delta)}, m_{B(x_0, \frac{1}{2})}) \), so we have that

\[
\mathcal{W}_{\eta, \epsilon}(m_{\mathcal{A}(x_0, \delta)}, m_{B(x_0, \delta)}) \leq \frac{C_0}{4\sqrt{|B(x_0, \frac{\delta}{2})| |\eta\epsilon(\frac{3}{2}\delta)|}} \leq \frac{C_0}{4\sqrt{\alpha_d(\frac{\delta}{2})^d \frac{1}{\epsilon} \sigma_2(\frac{3}{2}\delta)^{-d-\epsilon}}} = \frac{C_0}{21^{-d/2}\tilde{C}_{d,s}(\frac{\delta}{\epsilon})^{s/2}}.
\]

Therefore, by the triangle inequality, we have that

\[
\mathcal{W}_{\eta, \epsilon}(\delta_{x_0}, m_{B(x_0, \delta)}) \leq \frac{C_0}{C_{d,s}} \left(\frac{\delta}{\epsilon}\right)^{s/2}
\]

where \( C_{d,s} = \tilde{C}_{d,s} \left(\frac{1}{1-2^{s/2}} + 2^{d/2-1}\right)^{-1} \).

The previous lemma computed an upper bound on the expel cost for general interpolation \( \theta \), in the case where \( \eta(|\cdot|) \) is non-integrable in \( B(0, \delta) \). It is also possible to provide an expel upper bound in the case where \( \theta \) is non-zero on the boundary—this condition is satisfied, for instance, by the arithmetic mean, but not by the logarithmic mean.

**Lemma 3.5** Let \( 0 < \delta < \epsilon \) where \( \epsilon \leq 1 \), and let \( x_0 \in \mathbb{R}^d \). Let \( m_{B(x_0, \delta)} \) denote the uniform probability measure on the ball \( B(x_0, \delta) \). Suppose that \( \theta(1, 0) = \kappa_\theta > 0 \). (Note that if \( \theta(r, s) = \frac{r+s}{2} \), then \( \kappa_\theta = \frac{1}{2} \).) In this case,

\[
\mathcal{W}_{\eta, \epsilon}(\delta_{x_0}, m_{B(x_0, \delta)}) \leq \frac{1}{\sqrt{\kappa_\theta \alpha_d(\frac{\delta}{\epsilon})^d \eta(\frac{\delta}{\epsilon})}}
\]

where \( \alpha_d \) is the volume of the \( d \)-dimensional unit ball.

**Proof** Let \( 0 < \delta < \epsilon \). Let \( g : [0, 1] \to [0, 1] \) be a function to be determined later, such that \( g(0) = 0 \) and \( g(1) = 1 \). Let \( \gamma := \text{Leb} + \delta_{x_0} \). Consider the curve \( \rho_t : [0, 1] \to \mathcal{P}(\mathbb{R}^d) \) defined by

\[
d\rho_t \over d\gamma(x) = \begin{cases} 
g(t) & x = x_0 \\
g(t) & x \in B(x_0, \delta) \setminus \{x_0\} \\
0 & \text{else.}
\end{cases}
\]
Nonlocal Wasserstein distance: metric and asymptotic properties

Page 23 of 66

238

Note that with our given boundary conditions on \( g(t), \rho_0 \) is the uniform measure on \( B(x_0, \delta_0) \), and \( \rho_1 = \delta_{x_0} \). Note also that by construction,

\[
\frac{d}{dt} \frac{d\rho_t}{d\gamma}(x) = \begin{cases} 
    \frac{g'(t)}{|B(x_0, \delta)|} x = x_0, \\
    \frac{-g'(t)}{|B(x_0, \delta)|} x \in B(x_0, \delta) \setminus \{x_0\}, \\
    0 \quad \text{else.}
\end{cases}
\]

Let \( j_t(x, y) \) be a flux so that \((\rho_t, j_t)\) solves the nonlocal continuity equation; in particular we set

\[
\frac{d}{d\gamma \otimes \gamma} j_t(x, y) = \begin{cases} 
    \frac{-2\eta_t(x, y)|B(x_0, \delta)|}{g'(t)} (x, y) \in \{x_0\} \times B(x_0, \delta) \setminus \{x_0\}, \\
    \frac{2\eta_t(x, y)|B|}{g'(t)} (x, y) \in B(x_0, \delta) \setminus \{x_0\} \times \{x_0\}, \\
    0 \quad \text{else.}
\end{cases}
\]

Together, since \( \gamma \otimes \gamma \) dominates all of \( j_t, \rho_t \otimes \text{Leb}, \) and \( \text{Leb} \otimes \rho_t \), the action of \((\rho_t, j_t)\) is then

\[
\mathcal{A}(\rho_t, j_t) := \iint_{\mathcal{G}} 2\theta \left( \frac{d}{d\gamma \otimes \gamma} j_t(x, y) \right)^2 \eta_\varepsilon(x, y) d\gamma(x) d\gamma(y).
\]

Observe that

\[
\frac{d}{d\gamma \otimes \gamma} (\rho_t \otimes \text{Leb}) = \frac{d}{d((\text{Leb} + \delta_{x_0}) \otimes (\text{Leb} + \delta_{x_0}))}(x, y) = \begin{cases} 
    \frac{d\rho_t}{d(\text{Leb} + \delta_{x_0})}(x) & y \neq x_0, \\
    0 & y = x_0
\end{cases}
\]

and similarly for \( \frac{d}{d\gamma \otimes \gamma} (\rho_t \otimes \text{Leb}) \). Moreover, note that \( \frac{d}{d\gamma \otimes \gamma} j_t(x, y) = 0 \) off of \( \{x_0\} \times B(x_0, \delta) \setminus \{x_0\} \cup B(x_0, \delta) \setminus \{x_0\} \times \{x_0\} \). Consequently,

\[
\mathcal{A}(\rho_t, j_t) = \int_{\{x_0\} \times B(x_0, \delta) \setminus \{x_0\}} 2\theta \left( \frac{d\rho_t}{d(\text{Leb} + \delta_{x_0})}(x) \right)^2 \eta_\varepsilon(x, y) d\gamma(x) d\gamma(y)
\]

\[
+ \iint_{B(x_0, \delta) \setminus \{x_0\}} 2\theta \left( \frac{d\rho_t}{d(\text{Leb} + \delta_{x_0})}(x) \right)^2 \eta_\varepsilon(x, y) d\gamma(x) d\gamma(y)
\]

\[
= \int_{B(x_0, \delta) \setminus \{x_0\}} \frac{\left( \frac{g'(t)}{2\eta_t(x_0, y)|B(x_0, \delta)|} \right)^2}{\theta(g(t), 0)} \eta_\varepsilon(x_0, y) dy.
\]

Clearly, if it were the case that \( \theta(g(t), 0) = 0 \), then the action of \((\rho_t, j_t)\) would be infinite. However, since we have instead assumed that \( \theta(1, 0) = \kappa_0 > 0 \), and so by 1-homogeneity, \( \theta(g(t), 0) = \kappa_0 g(t) \) for all \( t \), and hence

\[
\mathcal{A}(\rho_t, j_t) = \int_{B(x_0, \delta) \setminus \{x_0\}} \frac{\left( \frac{g'(t)}{2\eta_t(x_0, y)|B(x_0, \delta)|} \right)^2}{\kappa_0 g(t)} \eta_\varepsilon(x_0, y) dy.
\]
\[
\begin{align*}
= & \frac{1}{4\kappa_\theta |B(x_0, \delta)|^2} \int_{B(x_0, \delta) \setminus \{x_0\}} \frac{(g'(t))^2}{g(t)\eta_\varepsilon(x_0, y)} \, dy \\
\leq & \frac{1}{4\kappa_\theta |B(x_0, \delta)|\eta_\varepsilon(\delta)} \left(\frac{g'(t)}{g(t)}\right)^2.
\end{align*}
\]

Consequently,
\[
\int_0^1 A(\rho_t, j_t) \leq \frac{1}{4\kappa_\theta |B(x_0, \delta)|e^{-d} \eta(\frac{\delta}{\varepsilon})} \int_0^1 \left(\frac{g'(t)}{g(t)}\right)^2 \, dt.
\]

Finally, we select \( g(t) = t^2 \). With this choice, \( \int_0^1 \left(\frac{g'(t)}{g(t)}\right)^2 \, dt = 4 \). And since \( |B(x_0, \delta)| = \alpha_d \delta^d \), we conclude that
\[
\mathcal{W}_{\eta, \varepsilon}(\delta_{x_0}, m_{B(x_0, \delta)}) \leq \frac{1}{\sqrt{\kappa_\theta \alpha_d \left(\frac{\delta}{\varepsilon}\right)^d \eta\left(\frac{\delta}{\varepsilon}\right)}}
\]
as desired. \( \square \)

**Proposition 3.6** If \( \int_{\mathbb{R}^d} \eta(|y|) \, dy < \infty \), and \( \theta(1, 0) = 0 \), then for all probability measures \( \nu \in \mathcal{P}(\mathbb{R}^d) \) which are singular to \( \delta_{x_0} \), \( \mathcal{W}_{\eta, \theta}(\delta_{x_0}, \nu) = \infty \).

**Proof** Let \((\rho_t, j_t)_{t \in [0, 1]} \) solve the nonlocal continuity equation, and let \( \rho_0 = \delta_0 \) and \( \rho_1 = \nu \).

We assume for the sake of contradiction that \( \int_0^1 A(\rho_t, j_t) \, dt < \infty \).

We define the set
\[
\mathcal{T} = \{ t \in [0, 1] : \rho_t(\{0\}) > 0 \land [j_t(\{0\}) \times \mathbb{R}^d \setminus \{0\}] > 0 \}.
\]

Note that this set is measurable since \( \rho_t : [0, 1] \to \mathcal{P}(\mathbb{R}^d) \) is narrowly continuous and \( j_t : [0, 1] \to \mathcal{M}_{loc}(G) \) is a Borel function. We claim that for any \( t \in \mathcal{T} \), it holds that
\[
A(\rho_t, j_t) = \infty.
\]

To see this, let \( \lambda_t \in \mathcal{M}^+(G) \) be any measure such that \( \rho_t \otimes \text{Leb} + \text{Leb} \otimes \rho_t + |j_t| \ll \lambda_t \). In particular, if \( \rho_t(\{0\}) \neq 0 \) (and, so, for any \( t \in \mathcal{T} \), the fact that \( \rho_t \otimes \text{Leb} \ll \lambda_t \) implies that \( \lambda_t \res \{0\} \times \mathbb{R}^d \) is not identically zero.

At the same time, compute that
\[
\frac{d(\text{Leb} \otimes \rho_t)}{d\lambda_t}(x, y) = \frac{d(\text{Leb} \otimes \rho_t)}{d(\text{Leb} \otimes \rho_t + \rho_t \otimes \text{Leb})}(x, y) \frac{d(\text{Leb} \otimes \rho_t + \rho_t \otimes \text{Leb})}{d\lambda_t}(x, y)
\]

Note that for all \( t \in \mathcal{T} \) we may select a representative of \( \frac{d(\text{Leb} \otimes \rho_t)}{d(\text{Leb} \otimes \rho_t + \rho_t \otimes \text{Leb})} \) so that
\[
\frac{d(\text{Leb} \otimes \rho_t)}{d(\text{Leb} \otimes \rho_t + \rho_t \otimes \text{Leb})}(0, y) = 0 \ \forall \ y \in \mathbb{R}^d
\]
since for all \( t \in \mathcal{T} \) and open, bounded \( U \subset \mathbb{R}^d \),
\[
\rho_t \otimes \text{Leb}(\{0\} \times U) = \rho_t(\{0\})\text{Leb}(U) > 0
\]
and so \( \text{Leb} \otimes \rho_t + \rho_t \otimes \text{Leb})(\{0\} \times U) > 0, \) but \( \text{Leb} \otimes \rho_t(\{0\} \times U) = 0 \). This implies that (up to a choice of a.e.-equivalent representative) \( \frac{d\text{Leb} \otimes \rho_t}{d\lambda_t}(0, y) = 0 \) for all \( y \in \mathbb{R}^d \).

Therefore, compute as follows: for all \( t \in \mathcal{T} \),
\[
A_{\eta, \varepsilon}(\rho_t, j_t) = \int_G \int \frac{1}{2\theta} \left( \frac{d(\rho_t \otimes \text{Leb})}{d\lambda_t}(x, y) \right)^2 \eta_\varepsilon(x, y)d\lambda_t(x, y).
\]

Springer
\[ \forall \int_{[0] \times \mathbb{R}^d} \eta(x, y) d\lambda_t(x, y) \geq 2 \theta \left( \frac{d \rho_t \otimes \text{Leb}}{d \lambda_t}, \frac{d \text{Leb} \otimes \rho_t}{d \lambda_t} \right) \eta(x, y) d\lambda_t(x, y) \]
\[ = \int_{[0] \times \mathbb{R}^d} \left( \frac{dj_t}{d\lambda_t} \right)^2 \eta(x, y) d\lambda_t(x, y) \]
\[ = \infty. \]

Therefore, if \( \int_0^1 A(\rho_t, j_t) < \infty \), it must be the case that \( \text{Leb}(\Xi) = 0 \).

However, we claim that \( \text{Leb}(\Xi) > 0 \). Indeed, consider the following. Let \( \xi \in C_c^\infty(\mathbb{R}^d, [0, 1]) \) such that \( \xi(0) = 1 \) and let \( \xi_\delta(x) := \xi \left( \frac{x}{\delta} \right) \). Note that \( 0 \leq \xi_\delta(x) \leq 1 \), and that as \( \delta \to 0 \), \( \xi_\delta(x) \) converges pointwise to the indicator \( 1_{\{x = 0\}} \). Plugging \( \xi_\delta(x) \) into the continuity equation, we find that for any \( \tau \in (0, 1] \),
\[ \int_{\mathbb{R}^d} \xi_\delta(x) d\rho_\tau(x) - \int_{\mathbb{R}^d} \xi_\delta(x) d\rho_0(x) = -\int_0^\tau \int_G (\xi_\delta(y) - \xi_\delta(x)) \eta(x, y) d[j_t](x, y) dt \]
and so
\[ \left| \int_{\mathbb{R}^d} \xi_\delta(x) d\rho_\tau(x) - \int_{\mathbb{R}^d} \xi_\delta(x) d\rho_0(x) \right| \leq \int_0^\tau \int_G |\xi_\delta(y) - \xi_\delta(x)| \eta(x, y) d|j_t|(x, y) dt. \]
\[ = \int_0^\tau \int_G |\xi_\delta(y) - \xi_\delta(x)| \frac{d|j_t|}{d\lambda_t}(x, y) \eta(x, y) d\lambda_t(x, y) dt. \]

Using the fact that \( \frac{d|j_t|}{d\lambda_t} = \left| \frac{dj_t}{d\lambda_t} \right| \), and observing that
\[ \left| \frac{dj_t}{d\lambda_t} \right| = \sqrt{2 \theta \left( \frac{d \rho_t \otimes \text{Leb}}{d \lambda_t}, \frac{d \text{Leb} \otimes \rho_t}{d \lambda_t} \right) \frac{\left( \frac{dj_t}{d\lambda_t} \right)^2}{\theta \left( \frac{d \rho_t \otimes \text{Leb}}{d \lambda_t}, \frac{d \text{Leb} \otimes \rho_t}{d \lambda_t} \right)}} \]
whenever the right hand side is well defined, we deduce (now using the convention that \( 0 \cdot \infty = \infty \)) that for any \( \tau \in (0, 1] \),
\[ \int_0^\tau \int_G |\xi_\delta(y) - \xi_\delta(x)| \frac{d|j_t|}{d\lambda_t} \eta d\lambda_t(x, y) dt \leq \int_0^\tau \int_G \left|\xi_\delta(y) - \xi_\delta(x)\right| \sqrt{2 \theta \left( \frac{d \rho_t \otimes \text{Leb}}{d \lambda_t}, \frac{d \text{Leb} \otimes \rho_t}{d \lambda_t} \right)} \frac{\left( \frac{dj_t}{d\lambda_t} \right)^2}{\theta \left( \frac{d \rho_t \otimes \text{Leb}}{d \lambda_t}, \frac{d \text{Leb} \otimes \rho_t}{d \lambda_t} \right)} \eta d\lambda_t(x, y) dt. \]

Applying Hölder’s inequality, and using the fact that \( \theta(r, s) \leq \frac{r + s}{2} \), we find that this last expression is bounded above by
\[ \left( \int_0^\tau \int_G \left( \frac{d \rho_t \otimes \text{Leb}}{d \lambda_t}(x, y) + \frac{d \text{Leb} \otimes \rho_t}{d \lambda_t}(x, y) \right) \eta(x, y) d\lambda_t(x, y) dt \right)^{1/2} \]
\begin{align*}
&\frac{1}{\sqrt{2\pi C}} \left( \int_0^\tau \int_G \frac{(\xi_\delta(y) - \xi_\delta(x))^2 \left( \frac{d\mu}{d\lambda_t} \right)^2 (x, y)}{2\theta \left( \frac{d(\rho_t \otimes \text{Leb})}{d\lambda_t} (x, y), \frac{d(\text{Leb} \otimes \rho_t)}{d\lambda_t} (x, y) \right)} \eta_\epsilon(x, y) d\lambda_t(x, y) dt \right)^{1/2}.
\end{align*}

The term in the left parentheses is bounded above, in turn, by $2\tau \int_{\mathbb{R}^d} \eta(|y|) dy$. To see this, compute that
\begin{align*}
\int_0^\tau \int_G \frac{d\rho_t \otimes \text{Leb}}{d\lambda_t}(x, y) \eta(x, y) d\lambda_t(x, y) dt &= \int_0^\tau \int_G \eta(x, y) d(\rho_t \otimes \text{Leb})(x, y) dt \\
&\leq \int_0^\tau \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} \eta(x, y) dy \right) d\rho_t(x) dt \\
&= \tau C.
\end{align*}

The computation for $\int_0^\tau \int_G \frac{\text{Leb} \otimes \rho_t}{d\lambda_t}(x, y) \eta(x, y) d\lambda_t(x, y) dt$ is identical. Therefore, we find that
\begin{align*}
\left\| \int_{\mathbb{R}^d} \xi_\delta(x) d\rho_\tau(x) - \int_{\mathbb{R}^d} \xi_\delta(x) d\rho_0(x) \right\| &\leq \sqrt{2\tau C} \left( \int_0^\tau \int_G \frac{(\xi_\delta(y) - \xi_\delta(x))^2 \left( \frac{d\mu}{d\lambda_t}(x, y) \right)^2 \eta_\epsilon(x, y) d\lambda_t(x, y)}{2\theta \left( \frac{d(\rho_t \otimes \text{Leb})}{d\lambda_t} (x, y), \frac{d(\text{Leb} \otimes \rho_t)}{d\lambda_t} (x, y) \right)} dt \right)^{1/2}.
\end{align*}

Now, since $(\xi_\delta(y) - \xi_\delta(x))^2 \leq 1$, and $\left( \frac{d\mu}{d\lambda_t}(x, y) \right)^2 \eta_\epsilon(x, y)$ is integrable with respect to $d\lambda_t(x, y) dt$ on $G \times [0, \tau]$ (assuming that $\int_0^\tau A(\rho_t, j) dt < \infty$), and $(\xi_\delta(y) - \xi_\delta(x))^2$ converges pointwise to $1_{\{0\} \times \mathbb{R}^d \setminus \{0\}}$, we can apply the dominated convergence theorem to deduce that as $\delta \to 0$,
\begin{align*}
\int_0^\tau \int_G \frac{(\xi_\delta(y) - \xi_\delta(x))^2 \left( \frac{d\mu}{d\lambda_t}(x, y) \right)^2 \eta_\epsilon(x, y) d\lambda_t(x, y)}{2\theta \left( \frac{d(\rho_t \otimes \text{Leb})}{d\lambda_t} (x, y), \frac{d(\text{Leb} \otimes \rho_t)}{d\lambda_t} (x, y) \right)} dt &\to \int_0^\tau \int_{\mathbb{R}^d \setminus \{0\}} \frac{(\xi_\delta(y) - \xi_\delta(x))^2 \left( \frac{d\mu}{d\lambda_t}(x, y) \right)^2 \eta_\epsilon(x, y) d\lambda_t(x, y)}{2\theta \left( \frac{d(\rho_t \otimes \text{Leb})}{d\lambda_t} (x, y), \frac{d(\text{Leb} \otimes \rho_t)}{d\lambda_t} (x, y) \right)} dt.
\end{align*}

At the same time, as $\delta \to 0$, $\left| \int_{\mathbb{R}^d} \xi_\delta(x) d\rho_\tau(x) - \int_{\mathbb{R}^d} \xi_\delta(x) d\rho_0(x) \right| \to |\rho_\tau(\{0\}) - \rho_0(\{0\})|$. So in the limit we find that
\begin{align*}
&\left| \rho_\tau(\{0\}) - \rho_0(\{0\}) \right| \\
&\leq \sqrt{2\tau C} \left( \int_0^\tau \int_{\mathbb{R}^d \setminus \{0\}} \frac{(\xi_\delta(y) - \xi_\delta(x))^2 \left( \frac{d\mu}{d\lambda_t}(x, y) \right)^2 \eta_\epsilon(x, y) d\lambda_t(x, y)}{2\theta \left( \frac{d(\rho_t \otimes \text{Leb})}{d\lambda_t} (x, y), \frac{d(\text{Leb} \otimes \rho_t)}{d\lambda_t} (x, y) \right)} dt \right)^{1/2}.
\end{align*}

In particular, if $\tau$ is taken to be the first time $\tau$ such that $\rho_\tau(\{0\}) = 0$ (note that such a first $t$ exists, thanks to Corollary 3.3), then we find that
\begin{align*}
\frac{1}{\sqrt{2\tau C}} &< \int_0^\tau \int_{\mathbb{R}^d \setminus \{0\}} \frac{(\xi_\delta(y) - \xi_\delta(x))^2 \left( \frac{d\mu}{d\lambda_t}(x, y) \right)^2 \eta_\epsilon(x, y) d\lambda_t(x, y)}{2\theta \left( \frac{d(\rho_t \otimes \text{Leb})}{d\lambda_t} (x, y), \frac{d(\text{Leb} \otimes \rho_t)}{d\lambda_t} (x, y) \right)} dt
\end{align*}
which implies, in particular, that on a subset of \([0, \tau]\) of positive measure,

$$\frac{1}{\sqrt{2\pi C}} \leq \int_{[0] \times \mathbb{R}^d \setminus [0]} 2\theta \left( \frac{d\rho_t \otimes \text{Leb}}{d\lambda_t}(x, y), \frac{d(L\text{e} \otimes \rho_t)}{d\lambda_t}(x, y) \right) \eta_\varepsilon(x, y) d\lambda_t(x, y)$$

which means that on a subset of \([0, \tau]\) of positive measure, \(|\nabla \rho_t([0] \times \mathbb{R}^d \setminus [0])| > 0.\) However, for all of \([0, \tau]\), we know that \(\rho_t([0]) > 0.\) Therefore, \(\text{Leb}(\Sigma) > 0.\)

In the remainder of the article, we will often need to rule out, by assumption, the phenomenon presented by Proposition 3.6. We do this by moreover demanding that the basic assumptions of at least one of Lemmas 3.4 or 3.5 are satisfied, instead. For clarity, we now state this as an additional Assumption which may be imposed on \(\eta\) and \(\theta\), alongside Assumptions 2.1 and 2.2.

**Assumption 3.7** Regarding the weight kernel \(\eta\) and interpolation function \(\theta\), at least one of the following sets of assumptions holds:

(i) There exists an \(s > 0\) and constant \(c_s > 0\) such that: for all \(x, y \in \mathbb{R}^d\) satisfying \(|x - y| \leq \frac{1}{6}\), it holds that \(\eta(x, y) \geq c_s |x - y|^{-d-s}\); or,

(ii) The constant \(\theta(1, 0)\), which we denote by \(\kappa_\theta\), is strictly positive.

We remark that the constant \(\frac{1}{6}\) appearing in the Assumption above is chosen for simple convenience.

### 3.3 Global upper bounds for \(\mathcal{W}_{\eta, \theta}\)

Throughout this subsection, we assume that \(\eta\) satisfies Assumption 2.1(i–v), and that \(\theta\) satisfies Assumption 2.2.

We record the following estimate, which will ultimately be used to show that on a compact domain, the topology induced by \(\mathcal{W}_{\eta, \theta}\) is no stronger than the narrow topology.

**Lemma 3.8** Suppose that there is some \(s > 0\) and some constant \(c_s > 0\) such that for all \(x, y \in \mathbb{R}^d\) satisfying \(|x - y| \leq \frac{1}{2}\), it holds that \(\eta(x, y) \geq c_s |x - y|^{-d-s}\). Then, for all \(x, y \in \mathbb{R}^d\),

\[
\mathcal{W}_{\eta, \varepsilon}(\delta_x, \delta_y) \leq \Phi \left( \frac{|x - y|}{\varepsilon} \right)
\]

where

\[
\Phi(t) = \begin{cases} 
C_{d, \theta, s} t^{s/2} & 0 \leq t < \frac{3}{8} \\
C_{d, \theta, s} \left( \frac{3}{8} \right)^{s/2} t^{3/2} & \frac{3}{8} \leq t
\end{cases}
\]

and \(C_{d, \theta, s}\) depends only on \(d, \theta, \) and \(s\), and is explicitly given in the proof.

**Proof** Let \(m_{B(x_0, \delta)}\) denote the uniform probability measure on the ball \(B(x_0, \delta)\). Then,

\[
\mathcal{W}_{\eta, \varepsilon}(\delta_x, \delta_y) \leq \mathcal{W}_{\eta, \varepsilon}(\delta_x, m_{B(x_0, \delta)}) + \mathcal{W}_{\eta, \varepsilon}(m_{B(x_0, \delta)}, m_{B(y, \delta)}) + \mathcal{W}_{\eta, \varepsilon}(m_{B(y, \delta)}, \delta_y).
\]

In order to estimate \(\mathcal{W}_{\eta, \varepsilon}(m_{B(x, \delta)}, m_{B(y, \delta)})\), we use Lemma B.1. Indeed, applying Lemma B.1 in the case where \(A = B(x_0, \delta)\) and \(B = B(x_1, \delta)\), and \(|x_0 - x_1| + 2\delta \leq \frac{1}{2}\), it follows that

\[
\mathcal{W}_{\varepsilon, \eta}(m_{B(x_0, \delta)}, m_{B(x_1, \delta)}) \leq \frac{C_{\theta}}{4\sqrt{\alpha_d \delta^d \eta_\varepsilon(|x_0 - x_1| + 2\delta)}},
\]
In particular, since \( \eta(|x - y|) \geq c_s|x - y|^{-d - s} \), we find that

\[
\mathcal{W}_{\varepsilon, \eta}(m_{B(x_0, \delta)}, m_{B(x_1, \delta)}) \leq \frac{C_\theta}{4\sqrt{\alpha_d \delta^d c_s \left( \frac{|x_0 - x_1| + 2\delta}{\varepsilon} \right)^{-d - s}}. 
\]

Note that \( 2\delta < |x_0 - x_1| \leq \frac{7}{2} - 2\delta \) but otherwise \( \delta \) is arbitrary. In particular, if we pick \( \delta = \frac{1}{6}|x_0 - x_1| \), this leads to the constraint on \( x_0 \) and \( x_1 \) that \( |x_0 - x_1| < \frac{3}{8}\varepsilon \), and the estimate

\[
\mathcal{W}_{\varepsilon, \eta}(m_{B(x_0, \delta)}, m_{B(x_1, \delta)}) \leq \frac{C_\theta}{4\sqrt{\alpha_d \delta c_s 8^{-d-s}}} \left( \frac{4}{3} \frac{|x_0 - x_1|}{\varepsilon} \right)^{s/2} \text{ when } |x_0 - x_1| < \frac{3}{8}\varepsilon. 
\]

At the same time, we know from Lemma 3.4 that

\[
\mathcal{W}_{\varepsilon, \eta}(\delta_{x_0}, m_{B(x_0, \delta)}) \leq \frac{C_\theta}{C_{\delta, s}} \left( \frac{4}{3} \frac{|x_0 - x_1|}{\varepsilon} \right)^{s/2} 
\]

and similarly for \( x_1 \), so altogether,

\[
\mathcal{W}_{\varepsilon, \eta}(\delta_{x_0}, \delta_{x_1}) \leq \frac{1}{2} C_{\delta, s} \left( \frac{|x_0 - x_1|}{\varepsilon} \right)^{s/2} \text{ when } |x_0 - x_1| < \frac{3}{8}\varepsilon 
\]

where

\[
C_{\delta, s} = \frac{C_\theta}{2\sqrt{\alpha_d \delta c_s 8^{-d-s}}} \left( \frac{4}{3} \right)^{s/2} + \frac{4}{C_{\delta, s}} \frac{1}{6^{s/2}}. 
\]

For arbitrary \( x, y \in \mathbb{R}^d \), it suffices to repeatedly apply the triangle inequality. Namely, construct a sequence \( x = x_0, x_1, \ldots, x_k = y \) so that \( |x_i - x_{i+1}| < \frac{3}{8}\varepsilon \) for each \( i \in \{0, \ldots, k - 1\} \); note in particular we can take

\[
k = \left\lceil \frac{|x - y|}{\frac{3}{8}\varepsilon} \right\rceil. 
\]

Thus, for \( |x - y| \geq \frac{3}{8}\varepsilon \), we have that

\[
\mathcal{W}_{\varepsilon, \eta}(\delta_x, \delta_y) \leq \sum_{i=1}^{k-1} \mathcal{W}_{\varepsilon, \eta}(\delta_{x_i}, \delta_{x_{i+1}}) \leq k \frac{C_{\delta, s}}{2} \left( \frac{3}{8} \right)^{s/2} \leq C_{\delta, s} \left( \frac{3}{8} \right)^{s/2} \left( \frac{|x - y|}{\frac{3}{8}\varepsilon} + 1 \right)
\]

\[
\leq C_{\delta, s} \left( \frac{3}{8} \right)^{s/2} \frac{8}{3} \frac{|x - y|}{\varepsilon}. 
\]

Note that when \( |x - y| = \frac{3}{8}\varepsilon \), we have \( C_{\delta, s} \left( \frac{3}{8} \right)^{s/2} \frac{8}{3} \frac{|x - y|}{\varepsilon} = C_{\delta, s} \left( \frac{3}{8} \right)^{s/2} \), while

\[
\frac{1}{2} C_{\delta, s} \left( \frac{|x_0 - x_1|}{\varepsilon} \right)^{s/2} = \frac{1}{2} C_{\delta, s} \left( \frac{3}{8} \right)^{s/2} ; \text{ so by defining the continuous, nondecreasing function}
\]

\[
\Phi : [0, \infty) \rightarrow [0, \infty) \quad \Phi(t) = \begin{cases} 
C_{\delta, s} t^{s/2} & 0 \leq t < \frac{3}{8} \\
C_{\delta, s} \left( \frac{3}{8} \right)^{s/2} \frac{8}{3} t & \frac{3}{8} \leq t
\end{cases}
\]

we see that

\[
\mathcal{W}_{\varepsilon, \eta}(\delta_x, \delta_y) \leq \Phi \left( \frac{|x - y|}{\varepsilon} \right).
\]
More generally, one has the following upper bound:

**Lemma 3.9** Let $\varepsilon > 0$. Suppose that $\eta$ and $\theta$ satisfy Assumption 3.7. Then for all $x, y \in \mathbb{R}^d$,

$$W_{\eta, \varepsilon}(\delta x, \delta y) \leq \frac{C_{d, \theta}}{\sqrt{\eta \left(\frac{\varepsilon}{2}\right)}} |x - y| + C_{d, \theta, \eta}$$

where $C_{d, \theta, \eta}$ is an explicit constant given in the proof.

**Proof** To estimate the distance between delta masses at $x$ and $y$ when $|x - y| \gg \varepsilon$, we first spread the mass to a ball of width $\delta$, comparable to $\varepsilon$, and then jump between identical balls placed at distance comparable to $\varepsilon$ along the line segment between $x$ and $y$. The number of the balls is comparable to $|x - y|/\varepsilon$, which explains the scaling of the right hand side in our estimate.

Let $m_{B(x_0, \delta)}$ denote the uniform probability measure on the ball $B(x_0, \delta)$. Then,

$$W_{\eta, \varepsilon}(\delta x, \delta y) \leq W_{\eta, \varepsilon}(\delta x, m_{B(x_0, \delta)}) + W_{\eta, \varepsilon}(m_{B(x_0, \delta)}, m_{B(y, \delta)}) + W_{\eta, \varepsilon}(m_{B(y, \delta)}, \delta y).$$

In other to estimate $W_{\eta, \varepsilon}(m_{B(x_0, \delta)}, m_{B(y, \delta)})$, we use Lemma B.1. Indeed, applying Lemma B.1 in the case where $A = B(x_0, \delta)$ and $B = B(x_1, \delta)$, and $|x_0 - x_1| + 2\delta \leq \frac{\varepsilon}{2}$, it follows that

$$W_{\eta, \varepsilon}(m_{B(x_0, \delta)}, m_{B(x_1, \delta)}) \leq \frac{C_{\theta}}{4\sqrt{\alpha_d \delta^d \eta \left(\frac{\varepsilon}{2}\right)}} \frac{C_{\theta}}{4\sqrt{\alpha_d \delta^d \eta \left(\frac{\varepsilon}{2}\right)}}.$$ 

Then, in order to estimate $W_{\eta, \varepsilon}(m_{B(x_0, \delta)}, m_{B(y, \delta)})$, it suffices to select a sequence $x = x_0, x_1, \ldots, x_k = y$ so that $\delta < \frac{|x_i - x_{i+1}|}{\varepsilon} \leq \frac{\varepsilon}{2} - 2\delta$ for all $i \in \{0, \ldots, k - 1\}$; in particular, we can take

$$k = \left\lceil \frac{|x - y|}{\frac{\varepsilon}{2} - 2\delta} \right\rceil.$$ 

Hence,

$$W_{\eta, \varepsilon}(m_{B(x, \delta)}, m_{B(y, \delta)}) \leq k \frac{C_{\theta}}{4\sqrt{\alpha_d \delta^d \eta \left(\frac{\varepsilon}{2}\right)}} = \left(\left\lceil \frac{|x - y|}{\frac{\varepsilon}{2} - 2\delta} \right\rceil\right) \frac{C_{\theta}}{4\sqrt{\alpha_d \delta^d \eta \left(\frac{\varepsilon}{2}\right)}}.$$

Now, note that

$$\eta \left(\frac{\varepsilon}{2}\right) = \frac{1}{\varepsilon^d} \eta \left(\frac{\varepsilon}{2}\right) = \frac{1}{\varepsilon^d} \eta \left(\frac{1}{2}\right).$$

For convenience, we also select $\delta = \frac{\varepsilon}{6}$. Plugging this in, we get

$$W_{\eta, \varepsilon}(m_{B(x, \delta)}, m_{B(y, \delta)}) \leq \left(\left\lceil \frac{|x - y|}{\frac{\varepsilon}{6}} \right\rceil\right) \frac{1}{4\sqrt{\alpha_d \left(\frac{1}{6}\right)^d \eta \left(\frac{1}{2}\right)}} \frac{C_{\theta}}{\sqrt{\eta \left(\frac{1}{2}\right)}}.$$

In other words, putting

$$C_{d, \theta} = \frac{\sqrt{\alpha_d \left(\frac{1}{6}\right)^d \eta \left(\frac{1}{2}\right)}}{2 \left\lceil \frac{|x - y|}{\frac{\varepsilon}{6}} \right\rceil^{d/2}},$$

we can take

$$|x - y| \leq \frac{\varepsilon}{6}.$$

The number of the balls is comparable to $|x - y|/\varepsilon$, which explains the scaling of the right hand side in our estimate.
we see that
\[
\mathcal{W}_{\eta, \varepsilon}(m_B(x, \delta), m_B(y, \delta)) \leq \frac{C_{d, \theta}}{\sqrt{\eta(\frac{1}{2})}} \varepsilon |x - y| + \frac{1}{6} \frac{C_{d, \theta}}{\sqrt{\eta(\frac{1}{2})}}.
\]

Finally, we can compute as follows:
\[
\begin{align*}
\mathcal{W}_{\eta, \varepsilon}(\delta_x, \delta_y) & \leq \mathcal{W}_{\eta, \varepsilon}(\delta_x, m_B(x, \delta)) + \mathcal{W}_{\eta, \varepsilon}(m_B(x, \delta), m_B(y, \delta)) + \mathcal{W}_{\eta, \varepsilon}(m_B(y, \delta), \delta_y) \\
& \leq \frac{C_{d, \theta}}{\sqrt{\eta(\frac{1}{2})}} \varepsilon |x - y| + \frac{1}{6} \frac{C_{d, \theta}}{\sqrt{\eta(\frac{1}{2})}} + 2 \mathcal{W}_{\eta, \varepsilon}(\delta_x, m_B(x, \xi))
\end{align*}
\]

where we have used the fact that \( \delta = \frac{\varepsilon}{6} \). In particular, applying Lemma 3.4, we find that if \( \eta(x, y) \geq c_s |x - y|^{-d-s} \) when \( |x - y| \leq \frac{1}{6} \), it holds that
\[
\mathcal{W}_{\eta, \varepsilon}(\delta_x, \delta_y) \leq \frac{C_{d, \theta}}{\sqrt{\eta(\frac{1}{2})}} \varepsilon |x - y| + \frac{1}{6} \frac{C_{d, \theta}}{\sqrt{\eta(\frac{1}{2})}} + 2 \frac{C_{d, \theta}}{\sqrt{\eta(\frac{1}{2})}} \left( \frac{1}{6} \right)^{s/2};
\]

applying Lemma 3.5, we find that if \( \theta(1, 0) = \kappa \theta > 0 \), then
\[
\mathcal{W}_{\eta, \varepsilon}(\delta_x, \delta_y) \leq \frac{C_{d, \theta}}{\sqrt{\eta(\frac{1}{2})}} \varepsilon |x - y| + \frac{1}{6} \frac{C_{d, \theta}}{\sqrt{\eta(\frac{1}{2})}} + \frac{2}{\sqrt{\kappa \alpha_d (\frac{1}{6})^d \eta(\frac{1}{6})}}.
\]

So altogether, we have that
\[
\mathcal{W}_{\eta, \varepsilon}(\delta_x, \delta_y) \leq \frac{C_{d, \theta}}{\sqrt{\eta(\frac{1}{2})}} \varepsilon |x - y| + C_{d, \theta, \eta}
\]

where we have the case-wise definition of \( C_{d, \theta, \eta} \) (if both conditions obtain, either case can be chosen for the value of \( C_{d, \theta, \eta} \)):
\[
C_{d, \theta, \eta} := \begin{cases} \\
\frac{1}{6} \frac{C_{d, \theta}}{\sqrt{\eta(\frac{1}{2})}} + \frac{2}{\sqrt{\kappa \alpha_d (\frac{1}{6})^d \eta(\frac{1}{6})}} \theta(1, 0) = \kappa \theta > 0. \quad \text{when } |x - y| \leq \frac{1}{6}; \\
\frac{1}{6} \frac{C_{d, \theta}}{\sqrt{\eta(\frac{1}{2})}} + \frac{2}{\sqrt{\kappa \alpha_d (\frac{1}{6})^d \eta(\frac{1}{6})}} \eta(x, y) \geq c_s |x - y|^{-d-s} \end{cases}
\]

In order to proceed, we prove the following disintegration inequality for the \( \mathcal{W}_{\eta} \) metric, which is of independent interest, in addition to being needed in the proof of Lemma 3.11 below. An analogous result was established in [21, Proposition 2.14] in the discrete case; however, their proof does not readily adapt to our continuum setting.

**Theorem 3.10** (Disintegration inequality) Let \( \mu, \nu \in \mathcal{P}(\mathbb{R}^d) \). Then,
\[
\mathcal{W}^2_{\eta}(\mu, \nu) \leq \min_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathcal{W}^2_{\eta}(\delta_x, \delta_y) d\pi(x, y).
\]

Morally speaking, this theorem is just an instance of Jensen’s inequality, since \( \mathcal{W}^2_{\eta} \) is a convex l.s.c. function, and \( \mathcal{W}^2_{\eta}(\mu, \nu) = \mathcal{W}^2_{\eta}(\int \delta_x d\pi(x, y), \int \delta_y d\pi(x, y)) \). And indeed, in the discrete case, the proof of [21, Proposition 2.14] proceeds rather directly from Jensen’s
inequality, albeit applied to the action $A(\rho, j)$ rather than to $W$. However, standard proofs of Jensen’s inequality (see for instance [41]) require the underlying space (in this case $\mathcal{P}(\mathbb{R}^d)^2$) to carry the structure of a topological vector space, which we do not have here. We are aware of one more abstract version of Jensen’s inequality [45] which does not require a t.v.s. structure, but in our situation a direct proof turns out to be readily available.

**Proof** Let $\pi \in \Pi(\mu, \nu)$. Let $(X(\omega), Y(\omega))$ and $(X_i(\omega), Y_i(\omega)), i = 1, 2, \ldots$ be i.i.d. random variables distributed according to $\pi$. In particular, $(X, Y)_\#P = \pi$. By the joint convexity of $W^2_\eta$, we have that

$$W^2_\eta\left(\frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)}, \frac{1}{n} \sum_{i=1}^n \delta_{Y_i(\omega)}\right) \leq \frac{1}{n} \sum_{i=1}^n W^2_\eta(\delta_{X_i(\omega)}, \delta_{Y_i(\omega)}).$$

At the same time, by change of variables, we observe that

$$\int W^2_\eta(\delta_X(\omega), \delta_Y(\omega))dP(\omega) = \int W^2_\eta(\delta_x, \delta_y)d\pi(x, y).$$

Now, suppose that $W^2_\eta(\delta_X(\omega), \delta_Y(\omega))$ is an integrable random variable; since $W^2_\eta$ is nonnegative, otherwise $\int W^2_\eta(\delta_x, \delta_y)d\pi(x, y) = \infty$, in which case the theorem holds trivially. Applying the strong law of large numbers to the i.i.d. random variables $W^2_\eta(\delta_X(\omega), \delta_Y(\omega))$, we see that with probability 1,

$$\frac{1}{n} \sum_{i=1}^n W^2_\eta(\delta_{X_i(\omega)}, \delta_{Y_i(\omega)}) \rightarrow \int W^2_\eta(\delta_X(\omega), \delta_Y(\omega))dP(\omega).$$

Therefore,

$$\liminf_{n \to \infty} W^2_\eta\left(\frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)}, \frac{1}{n} \sum_{i=1}^n \delta_{Y_i(\omega)}\right) \leq \int W^2_\eta(\delta_x, \delta_y)d\pi(x, y).$$

The Glivenko–Cantelli theorem tells us that with probability 1, $\frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)} \rightarrow^* \mu$ and $\frac{1}{n} \sum_{i=1}^n \delta_{Y_i(\omega)} \rightarrow^* \nu$, so since $W^2_\eta$ is jointly l.s.c. with respect to narrow convergence,

$$W^2_\eta(\mu, \nu) \leq \liminf_{n \to \infty} W^2_\eta\left(\frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)}, \frac{1}{n} \sum_{i=1}^n \delta_{Y_i(\omega)}\right).$$

This shows that $W^2_\eta(\mu, \nu) \leq \int W^2_\eta(\delta_x, \delta_y)d\pi(x, y)$. But since $\pi \in \Pi(\mu, \nu)$ was arbitrary, we find that $W^2_\eta(\nu_0, \nu_1) \leq \inf_{\pi \in \Pi(\nu_0, \nu_1)} \int W^2_\eta(\delta_x, \delta_y)d\pi(x, y)$. Finally, the fact that the infimum is actually attained follows from the fact that $c(x, y) := W^2_\eta(\delta_x, \delta_y)$ is a nonnegative l.s.c. cost function, so standard Monge-Kantorovich theory applies, for instance [43, Theorem 1.7].

We now use the disintegration inequality and the estimates on the nonlocal transport between delta masses to obtain the initial, crude, upper bound on nonlocal transport between general probability measures in terms of the Wasserstein distance.

**Lemma 3.11** Let $\nu_0, \nu_1$ be any measures in $\mathcal{P}_2(\mathbb{R}^d)$. Suppose that $\eta$ and $\theta$ satisfy Assumption 3.7. Then,

$$W^2_{\eta, \varepsilon}(\nu_0, \nu_1) \leq 2 \frac{C^2_{d, \theta}}{\eta(1/2)} \frac{1}{\varepsilon^2} W^2_2(\nu_0, \nu_1) + 2C^2_{d, \theta, \eta}$$
where \( C_{d,\theta} = \frac{C_d}{\sqrt{\pi \sqrt{d}}} \), \( C_\theta = \int_{0}^{1} \frac{1}{\sqrt{(1-r,1+r)}} dr \), and \( C_{d,\theta,\eta} \) is the constant from Lemma 3.9.

Furthermore, in the case where Assumption 3.7 (i) is satisfied, one has the alternative upper bound

\[
\mathcal{W}^2_{\eta,\varepsilon}(v_0,v_1) \leq \min_{\pi \in \Pi(v_0,v_1)} \int_{(\mathbb{R}^d)^2} \Phi\left( \frac{|x-y|}{\varepsilon} \right)^2 d\pi(x,y)
\]

where \( \Phi \) is the function from the statement of Lemma 3.8.

**Remark 3.12** This lemma actually helps to address a question posed by Erbar. In [19], it is mentioned that it is unclear for which probability measures \( v_0 \) and \( v_1 \) on \( \mathbb{R}^d \) we have that \( \mathcal{W}(v_0, v_1) < \infty \) (in the case of the Wasserstein distances \( W_p \), these are precisely the measures which have finite \( p \)th moments). The proposition we are about to prove gives a sufficient condition on \( \eta \) and \( \theta \) ensuring that \( \mathcal{W}(v_0, v_1) < \infty \) for all \( v_0, v_1 \in \mathcal{P}_2(\mathbb{R}^d) \).

**Proof** In Proposition 3.10, we proved the disintegration inequality

\[
\mathcal{W}^2_{\eta,\varepsilon}(v_0,v_1) \leq \min_{\pi \in \Pi(v_0,v_1)} \int_{(\mathbb{R}^d)^2} \mathcal{W}^2_{\eta,\varepsilon}(\delta_x, \delta_y)d\pi(x,y).
\]

On the other hand,

\[
W^2_2(v_0,v_1) = \min_{\pi \in \Pi(v_0,v_1)} \int_{(\mathbb{R}^d)^2} |x-y|^2 d\pi(x,y).
\]

Therefore, let \( \overline{\pi} \) be \( W_2 \)-optimal plan for \((v_0,v_1)\). Using the assumption that \( \eta(x,y) \geq c_s |x-y|^{-d-s} \) when \( |x-y| \leq \frac{1}{6} \), or \( \theta(1,0) = \kappa_\theta > 0 \) (or both), it follows that

\[
\mathcal{W}^2_{\eta,\varepsilon}(v_0,v_1) \leq \int_{(\mathbb{R}^d)^2} \mathcal{W}^2_{\eta,\varepsilon}(\delta_x, \delta_y)d\pi(x,y)
\]

(Lemma 3.9) \leq \int_{(\mathbb{R}^d)^2} \left[ \frac{C_{d,\theta}}{\sqrt{\eta\left(\frac{1}{2}\right)}} \frac{1}{\varepsilon} |x-y| + C_{d,\theta,\eta} \right]^2 d\overline{\pi}(x,y)

\leq 2 \left( \frac{C_{d,\theta}}{\sqrt{\eta\left(\frac{1}{2}\right)}} \frac{1}{\varepsilon} \right)^2 \int_{(\mathbb{R}^d)^2} |x-y|^2 d\overline{\pi}(x,y) + 2C^2_{d,\theta,\eta}

= 2 \left( \frac{C_{d,\theta}}{\sqrt{\eta\left(\frac{1}{2}\right)}} \frac{1}{\varepsilon} \right)^2 W^2_2(v_0,v_1) + 2C^2_{d,\theta,\eta}.

Alternatively, in the case where \( \eta(|x-y|) \geq c_s |x-y|^{-d-s} \) when \( |x-y| \leq \frac{1}{6} \), we can apply Lemma 3.8, and deduce that

\[
\mathcal{W}^2_{\eta,\varepsilon}(v_0,v_1) \leq \min_{\pi \in \Pi(v_0,v_1)} \int_{(\mathbb{R}^d)^2} \mathcal{W}^2_{\eta,\varepsilon}(\delta_x, \delta_y)d\pi(x,y)
\]

\leq \min_{\pi \in \Pi(v_0,v_1)} \int_{(\mathbb{R}^d)^2} \Phi\left( \frac{|x-y|}{\varepsilon} \right)^2 d\pi(x,y).

\[\square\]
Let us mention the following consequence of Lemma 3.11. Together with Proposition 3.2, this shows that when $\eta$ is integrable, on a bounded domain the topology induced by $\mathcal{W}_\eta$ is equivalent to the strong topology on probability measures.

**Proposition 3.13** (TV upper bound on a bounded set) Suppose that $\eta$ and $\theta$ satisfy Assumption 3.7, and that $\nu_0, \nu_1 \in \mathcal{P}(\mathbb{R}^d)$ are both supported inside some bounded set $K \subset \mathbb{R}^d$. Then, there exists some constant $C$ (independent of $\nu_0$ and $\nu_1$, but allowed to depend on $d, \theta, \eta$, and the diameter of $K$) such that

$$\mathcal{W}_\eta^2(\nu_0, \nu_1) \leq C \cdot TV(\nu_0, \nu_1).$$

**Proof** The idea is that we simply rerun the argument for Lemma 3.11, but allow the mass in the “overlap” between $\nu_0$ and $\nu_1$ to stay put.

To wit, define the measure $\Theta := \min\{\nu_0, \nu_1\}$. We suppose that $\Theta$ is not identically zero, since otherwise the desired inequality holds trivially. Observe that

$$\|\nu_0 - \Theta\|_{TV} = \|\nu_1 - \Theta\|_{TV} = 2TV(\nu_0, \nu_1),$$

so in particular

$$\mathcal{W}_\eta\left(\frac{\nu_0 - \Theta}{2TV(\nu_0, \nu_1)}, \frac{\nu_1 - \Theta}{2TV(\nu_0, \nu_1)}\right)$$

is well-defined. Moreover, by Lemma 3.11 (with $\varepsilon = 1$) it holds that

$$\mathcal{W}_\eta^2\left(\frac{\nu_0 - \Theta}{2TV(\nu_0, \nu_1)}, \frac{\nu_1 - \Theta}{2TV(\nu_0, \nu_1)}\right) \leq 2\frac{C^2_{d,\theta}}{\eta\left(\frac{1}{2}\right)} W^2_2\left(\frac{\nu_0 - \Theta}{2TV(\nu_0, \nu_1)}, \frac{\nu_1 - \Theta}{2TV(\nu_0, \nu_1)}\right) + 2C^2_{d,\theta,\eta}.$$ 

By the 1-homogeneity of the action $A_{\eta, \theta}$, and also the 1-homogeneity of $W^2_2$, this implies that

$$\mathcal{W}_\eta^2(\nu_0 - \Theta, \nu_1 - \Theta) \leq 2\frac{C^2_{d,\theta}}{\eta\left(\frac{1}{2}\right)} W^2_2(\nu_0 - \Theta, \nu_1 - \Theta) + 4C^2_{d,\theta,\eta}TV(\nu_0, \nu_1).$$

At the same time, any solution to the nonlocal continuity equation with endpoints $\nu_0 - \Theta$ and $\nu_1 - \Theta$ extends trivially to a solution to the nonlocal continuity equation with endpoints $\nu_0$ and $\nu_1$: this is because the nonlocal continuity equation is additive, and so we can just add on the constant solution $(\Theta, 0)_{t \in [0,1]}$ to the NCE. By the convexity of the action, this implies that

$$\mathcal{W}_\eta^2(\nu_0, \nu_1) \leq 2TV(\nu_0, \nu_1)\mathcal{W}_\eta^2\left(\frac{\nu_0 - \Theta}{2TV(\nu_0, \nu_1)}, \frac{\nu_1 - \Theta}{2TV(\nu_0, \nu_1)}\right) = \mathcal{W}_\eta^2(\nu_0 - \Theta, \nu_1 - \Theta).$$

Therefore,

$$\mathcal{W}_\eta^2(\nu_0, \nu_1) \leq 2\frac{C^2_{d,\theta}}{\eta\left(\frac{1}{2}\right)} W^2_2(\nu_0, \nu_1) + 4C^2_{d,\theta,\eta}TV(\nu_0, \nu_1).$$

as desired.

Lastly, by combining [43, Equation (5.1)] with [30, Theorem 4], we see that

$$W^2_2(\nu_0, \nu_1) \leq \text{diam}(K) \cdot W^1(\nu_0, \nu_1) \leq \text{diam}(K)^2 \cdot TV(\nu_0, \nu_1).$$
which allows us to deduce that
\[ W^2_\eta(v_0, v_1) \leq \left( \frac{C^2_{d,\theta} \cdot \text{diam}(K)^2}{\eta(\frac{1}{2})} + 4C^2_{d,\theta,\eta} \right) TV(v_0, v_1). \]

\[ \Box \]

4 Exact nonlocalization

4.1 Exact solution to nonlocal continuity equation

We start by introducing a way to use solutions of the continuity equation to create solutions of the nonlocal continuity equation with a kernel \( \eta \), (1.3). Namely we discovered that given a solution of the continuity equation in the flux form one can convolve it by a specific, \( \eta \)-dependent kernel \( \zeta \) so that the convolved flow is an exact solution of the nonlocal continuity equation. We first present the solution in a formal way and then justify it for weak solutions below.

Let \( \eta(s) \) denote the radial profile of a kernel \( \eta(x, y) \) satisfying Assumption 2.1. Define
\[ \zeta(r) = \int_r^\infty s \eta(s) ds. \]

One can check that under this assumption, \( \zeta \) is integrable in \( \mathbb{R}^d \) even when \( \eta \) is not.

Consider a solution of the continuity equation
\[ \partial_t \rho + \text{div}(J) = 0. \]

Let \( \rho_\zeta = \rho \ast \zeta \) and \( J_\zeta = J \ast \zeta \). Then
\[ \partial_t \rho_\zeta + \text{div}(J_\zeta) = 0. \]

Let \( j(x, y) = (y - x) \cdot (J(y) + J(x)) \). We claim that
\[ \partial_t \rho_\zeta + \int j(x, y) \eta(x - y) dy = 0. \]

Namely note that
\[ \nabla \zeta(|x - y|) = \eta(|x - y|)(x - y). \]

Thus, using symmetry of \( \eta \),
\[ \int j(x, y) \eta(|x - y|) dy = -\int \eta(|x - y|)(x - y) \cdot (J(y) + J(x)) dy \]
\[ = \int \nabla \zeta(|x - y|) \cdot J(y) dy - J(x) \int \eta(|x - y|)(x - y) dy \]
\[ = \text{div} \left( \int \zeta(|x - y|) J(y) dy \right) + 0 = \text{div}(J_\zeta). \]

While the preceding argument is formal, and written for strong solutions, making the argument rigorous and extending to weak solutions is straightforward, and is done in the lemma below.

In what follows, we use slightly more burdensome notation: given a specific kernel \( \eta \), we write \( \zeta_\eta(r) := \int_r^\infty s \eta(s) ds \); so in particular \( \zeta_{(\eta_\epsilon)}(r) := \int_r^\infty s \eta_\epsilon(s) ds \). Note however that
it need not hold that $\zeta_\eta(|x|)$ is a convolution kernel, i.e. in may not be normalized when integrating on $\mathbb{R}^d$. We therefore introduce the (normalized) convolution kernels (that these are the correct normalization constants is shown in Lemma B.3):

$$\bar{\zeta}_\eta := \frac{d}{M_2(\eta)} \zeta_\eta; \quad \bar{\zeta}(\eta_\varepsilon) := \frac{d}{\varepsilon^2 M_2(\eta)} \zeta(\eta_\varepsilon).$$

Note that $\bar{\zeta}_\eta$ is a convolution kernel supported on the unit ball, while $\bar{\zeta}(\eta_\varepsilon)$ is a convolution kernel supported on the ball of radius $\varepsilon$.

**Lemma 4.1** (exact nonlocalization) Assume that $\eta$ satisfies Assumption 2.1. Suppose that for all test functions $\varphi_t \in C^\infty_c((0, 1) \times \mathbb{R}^d)$, $(\rho_t, \bar{\varphi}_t)_{t \in [0, 1]} \in [0, 1] \to \mathcal{P}(\mathbb{R}^d) \times \mathcal{M}_{loc}(\mathbb{R}^d; \mathbb{R}^d)$ satisfies

$$\int_0^1 \int_{\mathbb{R}^d} \partial_t \varphi_t(x) d\rho_t(x) dt + \int_0^1 \int_{\mathbb{R}^d} \nabla \varphi_t(x) \cdot d\bar{\varphi}_t(x) dt = 0.$$

Then, it holds that $(\bar{\varphi}_t \ast \rho_t, \bar{\varphi}_t \ast \bar{\varphi}_t)_{t \in [0, 1]}$ solves the nonlocal continuity equation, where the measure $\bar{\varphi}_t : [0, 1] \to \mathcal{M}_{loc}(G)$ is defined in the following way: for all test functions $\Phi(x, y) \in C^\infty_c(G)$, we define

$$\int_G \Phi(x, y) d\bar{\varphi}_t(x, y) := \frac{d}{\varepsilon^2 M_2(\eta)} \int_G \Phi(x, y)(y - x) \cdot \left( (d\bar{\varphi}_t(x) dy + d\bar{\varphi}_t(y) dx) \right).$$

**Proof** Suppose that for all test functions $\varphi_t \in C^\infty_c((0, 1) \times \mathbb{R}^d)$, it holds that $(\rho_t, \bar{\varphi}_t)_{t \in [0, 1]} \in [0, 1] \to \mathcal{P}(\mathbb{R}^d) \times \mathcal{M}_{loc}(\mathbb{R}^d; \mathbb{R}^d)$ satisfies

$$\int_0^1 \int_{\mathbb{R}^d} \partial_t \varphi_t(x) d\rho_t(x) dt + \int_0^1 \int_{\mathbb{R}^d} \nabla \varphi_t(x) \cdot d\bar{\varphi}_t(x) dt = 0.$$

Then, it also holds that $(\bar{\varphi}_t \ast \rho_t, \bar{\varphi}_t \ast \bar{\varphi}_t)_{t \in [0, 1]}$ is also a solution to this form of the continuity equation:

$$\int_0^1 \int_{\mathbb{R}^d} \partial_t \varphi_t(x) d(\bar{\varphi}_t \ast \rho_t)(x) dt + \int_0^1 \int_{\mathbb{R}^d} \nabla \varphi_t(x) \cdot d(\bar{\varphi}_t \ast \bar{\varphi}_t)(x) dt = 0.$$

Our goal is to show that

$$\int_0^1 \int_{\mathbb{R}^d} \partial_t \varphi_t(x) d\bar{\varphi}_t(x) dy + \frac{1}{2} \int_0^1 \int_G \nabla \varphi_t(x, y) \eta_\varepsilon(x, y) d\bar{\varphi}_t(x, y) dt = 0,$$

so we claim that

$$\frac{1}{2} \int_0^1 \int_G \nabla \varphi_t(x, y) \eta_\varepsilon(x, y) d\bar{\varphi}_t(x, y) dt = \int_0^1 \int_{\mathbb{R}^d} \nabla \varphi_t(x) \cdot d\bar{\varphi}_t \ast \bar{\varphi}_t(x) dt,$$

which establishes the theorem. Recalling that $\tilde{\nabla} \varphi_t(x, y) := \varphi_t(y) - \varphi_t(x)$, we see that for each $t \in [0, 1]$,

$$\int_G \tilde{\nabla} \varphi_t(x, y) \eta_\varepsilon(x, y) d\bar{\varphi}_t(x, y) = \frac{1}{\varepsilon^2 \alpha_d \sigma_\eta} \int_G \varphi_t(y) \eta_\varepsilon(x, y)(y - x) \cdot (d\bar{\varphi}_t(x) dy + d\bar{\varphi}_t(y) dx)$$

$$- \frac{1}{\varepsilon^2 \alpha_d \sigma_\eta} \int_G \varphi_t(x) \eta_\varepsilon(x, y)(y - x) \cdot (d\bar{\varphi}_t(x) dy + d\bar{\varphi}_t(y) dx).$$
(Note that the two integrals on the right hand side are well-defined, since \( \varphi_t(x) \) is smooth and compactly supported in \( \mathbb{R}^d \), \( \eta_e(x, y)(y - x) \) is integrable and compactly supported in \( x \) for each \( y \), and \( \vec{j}_t \in M_{loc}(\mathbb{R}^d; \mathbb{R}^d) \).) So first, compute (using the fact that \( \nabla_y \zeta(\eta_e)(|x - y|) = (x - y)\eta_e(|x - y|) \)) that
\[
\frac{d}{\varepsilon^2 M_2(\eta)} \int_G \varphi_t(y)\eta_e(x, y)(y - x) \cdot \tilde{d}\vec{j}_t(x)dy
= -\frac{d}{\varepsilon^2 M_2(\eta)} \int_G \varphi_t(y)\nabla_y \zeta(\eta_e)(|x - y|) \cdot \tilde{d}\vec{j}_t(x)dy
= \frac{d}{\varepsilon^2 M_2(\eta)} \int_G \zeta(\eta_e)(|x - y|)\nabla \varphi_t(y) \cdot \tilde{d}\vec{j}_t(x)dy
= \int_{\mathbb{R}^d} \nabla \varphi_t(y) \cdot \left( \int_{\mathbb{R}^d} \zeta(\eta_e)(|x - y|) \tilde{d}\vec{j}_t(x) \right) dy
= \int_{\mathbb{R}^d} \nabla \varphi_t(y) \cdot d \left( \tilde{\zeta}(\eta_e) \ast \tilde{\vec{j}_t} \right)(y).
\]

By identical reasoning,
\[
\frac{d}{\varepsilon^2 M_2(\eta)} \int_G \varphi_t(x)\eta_e(x, y)(y - x) \cdot d\tilde{\vec{j}_t}(y)dx = -\int_{\mathbb{R}^d} \nabla \varphi_t(x) \cdot d \left( \tilde{\zeta}(\eta_e) \ast \tilde{\vec{j}_t} \right)(x).
\]

Next, compute that
\[
\int_G \varphi_t(y)\eta_e(x, y)(y - x) \cdot d\tilde{\vec{j}_t}(y)dy = \int_{\mathbb{R}^d} \varphi_t(y) \left( \int_{\mathbb{R}^d} \eta_e(x, y)(y - x)dx \right) \cdot d\tilde{\vec{j}_t}(y)
= 0
\]

since the function \( \eta_e(x, y)(y - x) \) is radially anti-symmetric around \( y \), implying that \( \int_{\mathbb{R}^d} \eta_e(x, y)(y - x)dx = 0 \). By identical reasoning, it also holds that
\[
\int_G \varphi_t(x)\eta_e(x, y) d \left( \tilde{\vec{j}_t} \cdot (y - x) \right)(x)dy = 0.
\]

Therefore, for all \( t \in [0, 1] \),
\[
\int_G \nabla \varphi_t(x, y)\eta_e(x, y)d\tilde{\vec{j}_t}(x, y) = 2\int_{\mathbb{R}^d} \nabla \varphi_t(x) \cdot \left( \tilde{\zeta}(\eta_e) \ast \tilde{\vec{j}_t}(y) \right)dx,
\]

which establishes the claim. \( \square \)

### 4.2 A quantitative upper bound on the nonlocal Wasserstein distance

In this section we establish a bound on the nonlocal Wasserstein distance of the form \( W_{\eta, \varepsilon} \leq C W_2 + O(\varepsilon) \), where \( C \) is the exact proportionality constant that also appears in the matching lower bound on \( W_{\eta, \varepsilon} \) presented in Corollary 5.12.

**Proposition 4.2** (bounding \( W \) by \( W_2 \)) Assume that \( \eta \) and \( \theta \) satisfy Assumptions 2.1 and 2.2 respectively. Let \( K \) denote the convolution kernel \( K(x) = c_K e^{-|x|^2} \), where \( c_K \) is a normalizing constant, and let \( \tilde{\zeta}(\eta_e) \) denote the convolution kernel \( \tilde{\zeta}(\eta_e)(x) = \frac{d}{\varepsilon^2 M_2(\eta)} \int_{||x||}^{\infty} \eta_e(t)dt \). Let
We then smooth $\rho_s$ and all $t\in\mathbb{R}$, $\theta(\bar{\xi}\in\mathbb{R})$, respectively. By \cite[Lemma 8.1.9]{3}, $\rho_t \in \mathcal{P}_2(\mathbb{R}^d)$, $(\rho_t^n, J_t^n)_{t\in[0,1]}$ also solves the continuity equation in flux form. We then smooth $\rho_t^n$ and $J_t^n$ again, using the kernel $\tilde{\xi}(\eta\epsilon)$, so that $(\tilde{\xi}(\eta\epsilon) \ast \rho_t^n, \tilde{\xi}(\eta\epsilon) \ast J_t^n)$ again solves the continuity equation in flux form. By Lemma \ref{B.4}, we know (since $\eta\epsilon(|x-y|)$ is supported on $B(0,\epsilon)$) that if $\epsilon < s$, then the corresponding Lebesgue density $\bar{\xi}(\eta\epsilon) \ast \rho_t^n$ has local relative Lipschitz regularity of the form

$$\frac{\tilde{\xi}(\eta\epsilon) \ast \rho_t^n(y)}{\tilde{\xi}(\eta\epsilon) \ast \rho_t^n(x)} \leq \left(1 + \frac{3}{s} \right) \left(1 + \frac{3}{s} |x-y| \right)$$

when $|x-y| < s$.

Define $\bar{J}_t : [0, 1] \to \mathcal{M}_{loc}(G)$ as in Lemma \ref{4.1} with respect to $\tilde{J}_t^n$, namely

$$d\bar{J}_t(x, y) = \frac{d}{\varepsilon^2 M_2(\eta)}(y-x) \cdot \left( d\tilde{J}_t^n(x)dy + d\tilde{J}_t^n(y)dx \right).$$

In this case, since $\tilde{J}_t^n$ has a density with respect to the Lebesgue measure given by $\tilde{J}_t^n$, it follows that $\bar{J}_t$ has density with respect to the product Lebesgue measure restricted to $G$, given by

$$\bar{J}_t(x, y) := \frac{d\bar{J}_t}{dxdy} = \frac{d}{\varepsilon^2 M_2(\eta)}(y-x) \cdot \left[ \tilde{J}_t^n(x) + \tilde{J}_t^n(y) \right].$$

Furthermore, by Lemma \ref{4.1}, we know that $(\tilde{\xi}(\eta\epsilon) \ast \rho_t^n, \bar{J}_t)$ solves the nonlocal continuity equation.

Now, let us compare the nonlocal action $\mathcal{A}_{\eta,\epsilon}(\tilde{\xi}(\eta\epsilon) \ast \rho_t^n, \bar{J}_t)$ with the local action $\mathcal{A}_{\rho_t^n, \bar{J}_t}$. Relying on the homogeneity of the interpolation $\theta$, we observe (using Lemma \ref{B.4}) that

$$\theta(\tilde{\xi}(\eta\epsilon) \ast \rho_t^n(x), \tilde{\xi}(\eta\epsilon) \ast \rho_t^n(y)) = \tilde{\xi}(\eta\epsilon) \ast \rho_t^n(y) \theta \left( \frac{\tilde{\xi}(\eta\epsilon) \ast \rho_t^n(x)}{\tilde{\xi}(\eta\epsilon) \ast \rho_t^n(y)}, 1 \right)$$

$$\geq \tilde{\xi}(\eta\epsilon) \ast \rho_t^n(y) \theta \left( \frac{1}{1 + \frac{3}{s} \epsilon} \frac{1}{1 + \frac{3}{s} |x-y|}, 1 \right)$$

$$\geq \tilde{\xi}(\eta\epsilon) \ast \rho_t^n(y) \cdot \frac{1}{1 + \frac{3}{s} \epsilon} \frac{1}{1 + \frac{3}{s} |x-y|}.$$
Therefore,
\[
\mathcal{A}_{\eta,\varepsilon}(\tilde{\zeta}(\eta) \ast \rho_i^s \ast \bar{j}_t) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{j_t(x, y)^2}{2\theta(\xi_{\eta \varepsilon})} \frac{\eta_{e}(x, y)dx dy}{\xi_{\eta \varepsilon}(x) \ast \rho^s_i(y)} = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{\eta_{e}(x, y)dx dy}{\xi_{\eta \varepsilon}(x) \ast \rho^s_i(y)} \mathcal{M}(\tilde{j}_t^s(x) + \tilde{j}_t^s(y)) \cdot (y - x)^2.
\]

By [29, Corollary 2.16], together with Lemma B.2, we have that
\[
\int_{\mathbb{R}^d} \left(\tilde{j}_t^s(y) \cdot (y - x)\right)^2 \eta_{e}(x, y)dx = \epsilon^2 \frac{M_2(\eta)}{d} |\tilde{j}_t^s(y)|^2.
\]

Consequently, also using the fact that the support of $\eta_{e}(x, y)$ has diameter $\varepsilon$,
\[
\left(1 + \frac{3}{s} \varepsilon \right)^2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left(1 + \frac{3}{s} |x - y|\right) \left(\frac{\mathcal{M}(\tilde{j}_t^s(y))}{\xi_{\eta \varepsilon}} \ast \rho_i^s(y)\right)^2 \eta_{e}(x, y)dx dy
\]
\[
\leq \left(1 + \frac{3}{s} \varepsilon \right)^3 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left(\frac{\mathcal{M}(\tilde{j}_t^s(y))}{\xi_{\eta \varepsilon}} \ast \rho_i^s(y)\right)^2 \eta_{e}(x, y)dx dy
\]
\[
\leq \left(1 + \frac{3}{s} \varepsilon \right)^3 \int_{\mathbb{R}^d} \frac{(\mathcal{M}(\tilde{j}_t^s(y))}{\xi_{\eta \varepsilon}} \ast \rho_i^s(y)\right)^2 \eta_{e}(x, y)dx dy
\]

Likewise,
\[
\left(1 + \frac{3}{s} \varepsilon \right)^2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left(1 + \frac{3}{s} |x - y|\right) \left(\frac{\mathcal{M}(\tilde{j}_t^s(x))}{\xi_{\eta \varepsilon}} \ast \rho_i^s(y)\right)^2 \eta_{e}(x, y)dx dy
\]
\[
\leq \left(1 + \frac{3}{s} \varepsilon \right)^3 \int_{\mathbb{R}^d} \frac{(\mathcal{M}(\tilde{j}_t^s(x))}{\xi_{\eta \varepsilon}} \ast \rho_i^s(x)\right)^2 dx
\]

and so we deduce that
\[
\mathcal{A}_{\eta,\varepsilon}(\tilde{\zeta}(\eta) \ast \rho_i^s \ast \bar{j}_t) \leq \frac{2d}{\epsilon^2 \mathcal{M}(\eta)} \left(1 + \frac{3}{s} \varepsilon \right)^3 \int_{\mathbb{R}^d} \frac{(\mathcal{M}(\tilde{j}_t^s(y))}{\xi_{\eta \varepsilon}} \ast \rho_i^s(y)\right)^2 dy.
\]

From Lemma B.4, we know that $\rho_i^s(x) \left(\frac{1}{\mathcal{M}_{\eta \varepsilon}(x)}\right) \leq (\tilde{\zeta}(\eta) \ast \rho_i^s)(x)$. Therefore,
\[
\int_{\mathbb{R}^d} \frac{(\mathcal{M}(\tilde{j}_t^s(y))}{\xi_{\eta \varepsilon}} \ast \rho_i^s(y)\right)^2 dy \leq \left(1 + \frac{3}{s} \varepsilon \right)^3 \int_{\mathbb{R}^d} \frac{(\mathcal{M}(\tilde{j}_t^s(y))}{\rho_i^s(y)\right)^2 dy
\]

and hence
\[
\mathcal{A}_{\eta,\varepsilon}(\tilde{\zeta}(\eta) \ast \rho_i^s \ast \bar{j}_t) \leq \frac{2d}{\epsilon^2 \mathcal{M}(\eta)} \left(1 + \frac{3}{s} \varepsilon \right)^4 \mathcal{A}(\rho_i^s \ast \bar{j}_t).
\]
We now take $(\rho_t, \tilde{j}_t)_{t\in[0,1]}$ to be a curve of least action for $W_2$, connecting $\rho_0$ and $\rho_1$. Then, by using the fact that $\mathcal{A}(\rho^*_t, \tilde{j}_t^*) \leq \mathcal{A}(\rho_t, \tilde{j}_t)$, and integrating in $t$, we have that

$$W(\check{\xi}_{(q_e)} * \rho^*_0, \check{\xi}_{(q_e)} * \rho^*_1) \leq \frac{1}{\varepsilon} \left( \frac{2d}{M_2(\eta)} \right)^{1/2} \left( 1 + \frac{3}{s} \varepsilon \right)^2 W_2(\rho_0, \rho_1),$$

as desired. \hfill \Box

**Corollary 4.3** Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$. Assume that $\eta$ and $\theta$ satisfy Assumptions 2.1, 2.2, and 3.7. Then,

$$\varepsilon W_{\eta,\varepsilon}(\mu_0, \mu_1) \leq \left( \frac{2d}{M_2(\eta)} \right)^{1/2} W_2(\mu_0, \mu_1) + O(\sqrt{\varepsilon}).$$

Explicitly,

$$W_{\varepsilon,\eta}(\mu_0, \mu_1) \leq \frac{1}{\varepsilon} \left( \frac{2d}{M_2(\eta)} \right)^{1/2} \left( 1 + \sqrt{\varepsilon} \right)^2 W_2(\mu_0, \mu_1) + 2\sqrt{2c_{d,\theta} W_2(\mu_0, \mu_1)} + 2\sqrt{2c_{d,\theta,\eta} \varepsilon \sqrt{\varepsilon} W_2(\mu_0, \mu_1)},$$

where $c_{d,\theta} = \frac{c_0}{\sqrt{\varepsilon} \sqrt{\eta(\frac{1}{2})}}$, $c_\theta = \frac{1}{\sqrt{\eta}} \int_0^1 \frac{1}{\sqrt{\theta(1-r,1+r)}} dr$, and $c_{d,\theta,\eta}$ is the constant from Lemma 3.9.

**Proof** We saw in Lemma 3.11 that for arbitrary $\nu_0, \nu_1 \in \mathcal{P}_2(\mathbb{R}^d),

$$W_{\eta,\varepsilon}^2(\nu_0, \nu_1) \leq \frac{2}{\varepsilon} \frac{c_{d,\theta}}{\eta(\frac{1}{2})} \varepsilon^2 W_2^2(\nu_0, \nu_1) + 2c_{d,\theta,\eta}^2,$$

where $c_{d,\theta} = \frac{c_0}{\sqrt{\varepsilon} \sqrt{\eta(\frac{1}{2})}}$, $c_\theta = \frac{1}{\sqrt{\eta}} \int_0^1 \frac{1}{\sqrt{\theta(1-r,1+r)}} dr$, and $c_{d,\theta,\eta}$ is the constant from Lemma 3.9. In particular, take $\nu_0 = \mu_0$ and $\nu_1 = \check{\xi}_{(q_e)} * K_s * \mu_0$. It follows that

$$W_{\eta,\varepsilon}^2(\mu_0, \check{\xi}_{(q_e)} * K_s * \mu_0) \leq \frac{2}{\eta(\frac{1}{2})} \frac{c_{d,\theta}}{\varepsilon^2} W_2^2(\mu_0, \check{\xi}_{(q_e)} * K_s * \mu_0) + 2c_{d,\theta,\eta}^2.$$

By identical reasoning,

$$W_{\eta,\varepsilon}^2(\mu_1, \check{\xi}_{(q_e)} * K_s * \mu_1) \leq \frac{2}{\eta(\frac{1}{2})} \frac{c_{d,\theta}}{\varepsilon^2} W_2^2(\mu_1, \check{\xi}_{(q_e)} * K_s * \mu_1) + 2c_{d,\theta,\eta}^2.$$

Now since, by the triangle inequality,

$$W_{\eta,\varepsilon}(\mu_0, \mu_1) \leq W_{\eta,\varepsilon}(\mu_0, \check{\xi}_{(q_e)} * K_s * \mu_0) + W_{\eta,\varepsilon}(\check{\xi}_{(q_e)} * K_s * \mu_0, \check{\xi}_{(q_e)} * K_s * \mu_1) + W_{\eta,\varepsilon}(\check{\xi}_{(q_e)} * K_s * \mu_1, \mu_1) + W_{\eta,\varepsilon}(\mu_0, \check{\xi}_{(q_e)} * K_s * \mu_1) + W_{\eta,\varepsilon}(\check{\xi}_{(q_e)} * K_s * \mu_1, \mu_1),$$

we can use the previous proposition to see that

$$W_{\eta,\varepsilon}(\mu_0, \mu_1) \leq \frac{1}{\varepsilon} \left( \frac{2d}{M_2(\eta)} \right)^{1/2} \left( 1 + \frac{3}{s} \varepsilon \right)^2 W_2(\mu_0, \mu_1).$$
\[ + \sqrt{2 C_{d, \theta} \left( \frac{1}{\eta} \right) \varepsilon^2 W_2^2(\mu_0, \tilde{\xi}_{(\eta)} \ast K_s \ast \mu_0) + 2 C_{d, \theta, \eta}^2} \]

\[ + \sqrt{2 C_{d, \theta} \left( \frac{1}{\eta} \right) \frac{1}{\varepsilon^2} W_2^2(\mu_1, \tilde{\xi}_{(\eta)} \ast K_s \ast \mu_1) + 2 C_{d, \theta, \eta}^2}. \]

It remains to estimate \( W_2(\mu_0, \tilde{\xi}_{(\eta)} \ast K_s \ast \mu_0) \) and \( W_2(\mu_1, \tilde{\xi}_{(\eta)} \ast K_s \ast \mu_1) \). To do so, we can use two successive convolution estimates, since

\[ W_2(\mu_0, \tilde{\xi}_{(\eta)} \ast K_s \ast \mu_0) \leq W_2(\mu_0, K_s \ast \mu_0) + W_2(K_s \ast \mu_0, \tilde{\xi}_{(\eta)} \ast K_s \ast \mu_0) \]

and similarly for \( \mu_1 \). Thanks to the estimates from Lemmas B.5 and B.6, we know that

\[ W_2(\mu_0, K_s \ast \mu_0) \leq \left( \int |y|^2 cK e^{-|y|} dy \right)^{1/2} = (d^2 + d)^{1/2} s \]

and

\[ W_2(K_s \ast \mu_0, \tilde{\xi}_{(\eta)} \ast K_s \ast \mu_0) \leq \left( \frac{d}{d + 2 M_2(\eta)} \right)^{1/2} \varepsilon \]

and of course the same holds for \( \mu_1 \). Therefore, putting \( s = \sqrt{\varepsilon} \), we deduce the estimate

\[ \mathcal{W}_{\varepsilon, \eta}(\mu_0, \mu_1) \leq \frac{1}{\varepsilon} \left( \frac{2d}{M_2(\eta)} \right)^{1/2} \left( 1 + \sqrt{\varepsilon} \right)^2 W_2(\mu_0, \mu_1) \]

\[ + 2 \sqrt{2} \left( \frac{C_{d, \theta}}{\sqrt{\eta}} \right) \left( \frac{1}{\varepsilon} \right) \left( (d^2 + d)^{1/2} \sqrt{\varepsilon} + \left( \frac{d}{d + 2 M_2(\eta)} \right)^{1/2} \varepsilon \right) \]

\[ + 2 \sqrt{2} C_{d, \theta, \eta} \]

as desired. \( \square \)

### 5 Nonlocal Hamilton–Jacobi subsolution

We first first sketch the argument of this section. Our aim is to prove a bound of the following form:

\[ W_2(\mu_0, \mu_1) \leq \varepsilon \sqrt{\frac{M_2(\eta)}{2d}} \mathcal{W}_{\eta, \varepsilon}(\mu_0, \mu_1) + \text{error terms}. \]

Ultimately, we will show that the error terms are of order \( \sqrt{\varepsilon} \). Our strategy is to use Hamilton–Jacobi duality for \( W_2 \) (and also \( \mathcal{W}_{\eta, \varepsilon} \)). Simplifying somewhat:

(i) Let \((\phi_t)_{t \in [0,1]}\) be a (viscosity) solution to the Hamilton–Jacobi equation \( \partial_t \phi_t + |\nabla \phi_t|^2 = 0 \). Duality theory for \( W_2 \) (specifically \([46, \text{Proposition 5.48}]\)) tells us that

\[ \frac{1}{2} W_2^2(\mu_0, \mu_1) = \max_{\phi \in C_b} \left\{ \int \phi_1 d\mu_1 - \int \phi_0 d\mu_0 : \partial_t \phi_t + \frac{1}{2} |\nabla \phi_t|^2 = 0 \right\}. \]
(ii) We expect that a similar duality theorem holds for $\mathcal{W}_{\eta,\varepsilon}$ (but there is a different notion of “nonlocal Hamilton–Jacobi equation”):
\[
\frac{1}{2} \mathcal{W}_{\eta,\varepsilon}^2(\mu_0, \mu_1) = \max \left\{ \int \phi_1 d\mu_1 - \int \phi_0 d\mu_0 : (\phi_\varepsilon)_{\varepsilon \in [0, 1]} \text{ is a n.l. HJ subsolution} \right\}.
\]

(iii) We will use solutions of the Hamilton–Jacobi equation to construct subsolutions to the nonlocal Hamilton–Jacobi equations thus obtaining a lower bound on $\mathcal{W}_{\eta,\varepsilon}^2$.

Unfortunately, things are not so simple an we will need to introduce a layer of approximations. Because of the (conjectured, at this point, based on the outcome of Sect. 4) asymptotic proportionality constant between $W_2$ and $\mathcal{W}_{\eta,\varepsilon}$, the constant prefactor in any mapping which takes a (local) HJ solution and gives us a nonlocal HJ subsolution must be $\frac{2 M_{2(\eta)}}{\varepsilon^2}$.

Before proceeding, we present several preparatory lemmas.

**Definition 5.1** (space of space-time bounded Lipschitz functions) We define
\[
BL([0, 1] \times \mathbb{R}^d) := \{ \phi(x) : [0, 1] \times \mathbb{R}^d \to \mathbb{R} \mid \text{Lip}_{[0, 1] \times \mathbb{R}^d}(\phi) + \| \phi \|_{L^\infty([0, 1] \times \mathbb{R}^d)} < \infty \}.
\]

**Lemma 5.2** Suppose that $\mu_0$ and $\mu_1$ are probability measures which are both supported within $B(0, R)$ inside $\mathbb{R}^d$. Then, in the Kantorovich duality
\[
\frac{1}{2} W_2^2(\mu_0, \mu_1) = \sup_{\phi \in \text{BL}(\mathbb{R}^d)} \left\{ \int \phi^c(y) d\mu_1(y) - \int \phi(x) d\mu_0(x) \right\}
\]
(\text{where } \phi^c(y) := \inf_x \{ \phi(x) + \frac{1}{2} |x - y|^2 \}, \text{ the optimal potential, that is},
\[
\arg\max_{\phi \in \text{BL}(\mathbb{R}^d)} \left\{ \int \phi^c(y) d\mu_1(y) - \int \phi(x) d\mu_0(x) \right\}
\]
has $\text{Lip}(\phi) \leq R$.

Note that by Rademacher’s theorem, if $\phi$ is Lipschitz then $\nabla \phi$ exists Lebesgue-almost everywhere; the lemma therefore also shows that for the optimal Kantorovich potential, $|\nabla \phi| \leq R$.

**Proof** This is an easy refinement of standard results concerning Kantorovich duality, such as [47, Theorem 5.10].

Indeed, as discussed on [43, p. 11], if $c(x, y)$ is any continuous cost function and $\psi^c$ is any $c$-convex (resp. $c$-concave) function, it holds automatically that any modulus of continuity for $c(x, y)$ is also a modulus of continuity for $\psi^c$. In the case of $c(x, y) = \frac{1}{2} |x - y|^2$ on a domain of diameter $R$, we can take $|c(x, y) - c(x', y')| \leq R|x - x'|$ as a crude global modulus of continuity in the $x$ variable (and of course the same reasoning applies to the $y$ variable). Since the optimal Kantorovich potential $\phi$ can always be taken to be $c$-convex, the claim follows.

The Kantorovich duality formula for $W_2$ also has a “dynamic” counterpart in terms of solutions to a Hamilton–Jacobi equation. This fact was initially observed in [6, 39]; here we just give a proof for convenience.

**Corollary 5.3** Suppose that $\mu_0$ and $\mu_1$ are probability measures which are both supported within some domain of radius $R$ inside $\mathbb{R}^d$. Then,
\[
\frac{1}{2} W_2^2(\mu_0, \mu_1)
\]
\[ t \in \mathbb{R} \] 

where on the right hand side (for which we do prove a partial duality result), as follows.

However, in this work, we do not attempt to prove this duality formula directly. Rather, for technical reasons, we introduce a "smoothed version" of the nonlocal Wasserstein distance (for which we do prove a partial duality result), as follows.

\[ \mathcal{W}_{\eta, \varepsilon}^2(\mu_0, \mu_1) = \sup \left\{ \int \phi_1(x) d\mu_1(x) - \int \phi_0(x) d\mu_0(x) : \phi_t \in \text{HJ}^{1}_{\text{NL}} \right\}. \]

Proof Let \( \phi_0 \in BL(\mathbb{R}^d) \). By [26, Theorem 10.3.3], the unique viscosity solution of \( \partial_t \phi_t + \frac{1}{2} |\nabla \phi_t|^2 = 0 \) with initial condition \( \phi_0 \) is given by the Hopf-Lax formula \( \phi_t(x) = \inf_y \{ \phi_0(y) + \frac{1}{2} |x-y|^2 \} \) (which at time 1 just returns \( \phi_0^c \) for \( c = \frac{1}{2} |x-y|^2 \), and so by comparing with [47, Theorem 5.10], we see that

\[
\sup_{\phi_0 \in BL(\mathbb{R}^d)} \int \phi^c(y) d\mu_1(y) - \int \phi(x) d\mu_0(x) = \sup_{\phi_0 \in BL(\mathbb{R}^d)} \left\{ \int \phi_1 d\mu_1 - \int \phi_0 d\mu_0 : \partial_t \phi_t + \frac{1}{2} |\nabla \phi_t|^2 = 0 \right\}
\]

where on the right hand side \( \phi_t \) solves \( \partial_t \phi_t + \frac{1}{2} |\nabla \phi_t|^2 = 0 \) in the viscosity sense. By [26, Lemma 3.3.2], we know that \( \phi_t \in BL([0,1] \times \mathbb{R}^d) \).

Furthermore, the proof of [26, Lemma 3.3.2] indicates that \( \text{Lip}(\phi_t) \leq \text{Lip}(\phi_0) \) for all \( t \in [0,1] \). Therefore, by Lemma 5.2 it holds that \( \text{Lip}(\phi_t) \leq R \), for all \( t \in [0,1] \).

Definition 5.4 Following [23, Definition 3.1], we say that \( \phi_t(x) : [0,1] \times \mathbb{R}^d \to \mathbb{R} \) is a nonlocal Hamilton–Jacobi (HJ) sub-solution, and write \( \phi_t(x) \in \text{HJ}^1_{\text{NL}} \), if, for a.e. \( t \in (0,1) \), if, for a.e. \( x \in [0,1] \times \mathbb{R}^d \), the partial derivative \( \partial_t \phi_t(x) \) exists for every \( x \in \mathbb{R}^d \), and \( \sup_{x \in \mathbb{R}^d} |\partial_t \phi_t(x)| < \infty \); and we have, for all probability measures \( \mu \in \mathcal{P}(\mathbb{R}^d) \), and for any (hence all) \( \lambda \) such that \( \mu \ll \lambda \),

\[
\int \partial_t \phi_t(x) d\mu(x) + \frac{1}{4} \int (\phi_t(y) - \phi_t(x))^2 \theta \left( \frac{d\mu}{d\lambda}(x), \frac{d\mu}{d\lambda}(y) \right) \eta_\varepsilon(x,y) d\lambda(x) d\lambda(y) \leq 0.
\]

Remark 5.5 (Nonlocal Hamilton–Jacobi solutions) In the present work, it is the notion of nonlocal HJ sub-solution which is relevant. Nonetheless, we mention here an associated notion of solution. Following [28], we say that \( \phi_t(x) \) is a nonlocal Hamilton–Jacobi solution if it is a nonlocal Hamilton–Jacobi sub-solution, and moreover and we have that

\[
\sup_{\mu \in \mathcal{P}(\mathbb{R}^d) ; \lambda \gg \mu} \left\{ \int \partial_t \phi_t(x) d\mu(x)
\right. \\
+ \frac{1}{4} \int (\phi_t(y) - \phi_t(x))^2 \theta \left( \frac{d\mu}{d\lambda}(x), \frac{d\mu}{d\lambda}(y) \right) \eta_\varepsilon(x,y) d\lambda(x) d\lambda(y) \right\} = 0.
\]

We leave investigation of this rather atypical PDE to future work.

The duality formula we expect to hold for \( \mathcal{W}_{\eta, \varepsilon} \) is

\[
\frac{1}{2} \mathcal{W}_{\eta, \varepsilon}^2(\mu_0, \mu_1) = \sup \left\{ \int \phi_1(x) d\mu_1(x) - \int \phi_0(x) d\mu_0(x) : \phi_t \in \text{HJ}^1_{\text{NL}} \right\}.
\]
Definition 5.6 Let $K$ denote the convolution kernel $c_Ke^{-|x|}$. The “$s$-smoothed nonlocal Wasserstein distance”, denoted $\mathcal{W}_{\eta,e,s}$, is defined as follows: given $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^d)$, and denoting $\mu_s^\ast := K_s \ast \mu_t$ and $j_t^\ast := K_s \ast j_t$,

$$\mathcal{W}_{\eta,e,s}^2(\mu_0^\ast, \mu_1^\ast) := \inf_{(\mu_t,j_t)_{t\in[0,1]}} \left\{ \int_0^1 A(\mu_t^\ast, j_t^\ast)dt : (\mu_t^\ast, j_t^\ast)_{t\in[0,1]} \in \mathcal{CE}(\mu_0^\ast, \mu_1^\ast) \right\}.$$

In other words, $\mathcal{W}_{\eta,e,s}$ is defined in the same variational fashion as $\mathcal{W}_{\eta,e}$, except we restrict to the class of a.c. curves which have been smoothed using the mollification kernel $K_s$. Similarly, we also have a notion of “smoothed nonlocal HJ subsolution.”

Definition 5.7 We say that $\phi_t(x) : [0, 1] \times \mathbb{R}^d \to \mathbb{R}$ is an $s$-smoothed nonlocal Hamilton–Jacobi subsolution, and write $\phi_t(x) \in HJ_{NL}^{1,s}$ if the weak partial derivative $\partial_t \phi_t(x) : \mathbb{R}^d \to \mathbb{R}$ exists, and there exists some $p \in [1, \infty)$ such that $\|\partial_t \phi_t\|_{L_p(\mathbb{R}^d)} < \infty$, for Lebesgue almost all $t \in (0, 1)$; and we have, for all probability measures $\mu \in \mathcal{P}(\mathbb{R}^d)$, for almost all $t \in (0, 1)$,

$$\int \partial_t \phi_t(x)d(K_s \ast \mu)(x) + \frac{1}{4} \iint (\phi_t(y) - \phi_t(x))^2 \theta \left( \frac{d(K_s \ast \mu)}{d\text{Leb}}(x), \frac{d(K_s \ast \mu)}{d\text{Leb}}(y) \right) \eta_e(x, y)dxdy \leq 0.$$

Remark 5.8 The regularity assumption we have imposed on $\partial_t \phi_t(x)$ is chosen so that $\|\partial_t \phi_t dK_s \ast \mu_t\| < \infty$ for almost all $t \in (0, 1)$. Indeed, $K_s \ast \mu$ has density $K_s \ast \mu$ with respect to the Lebesgue measure, and since $K_s \ast \mu$ is a probability measure, $K_s \ast \mu \in L^1(\mathbb{R}^d)$. At the same time, Young’s convolution inequality implies that $K_s \ast \mu \in L^{\infty}(\mathbb{R}^d)$ since $e^{-|x|} \in L^{\infty}$. Hence by interpolation, $K_s \ast \mu \in L^q$ for all $q \in [1, \infty)$, so it suffices that $\partial_t \phi_t(x) \in L^p(\mathbb{R}^d)$ for some $p \in [1, \infty)$, for almost all $t \in (0, 1)$. In our present situation, it will always be the case that $\phi_t(x) \in BL([0, 1] \times \mathbb{R}^d)$, so in particular we can take $p = \infty$.

Lemma 5.9 For all $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$, it holds that

$$\mathcal{W}_{\eta,e}(K_s \ast \mu_0, K_s \ast \mu_1) \leq \mathcal{W}_{\eta,e,s}(K_s \ast \mu_0, K_s \ast \mu_1) \leq \mathcal{W}_{\eta,e}(\mu_0, \mu_1).$$

Proof That $\mathcal{W}_{\eta,e}(K_s \ast \mu_0, K_s \ast \mu_1) \leq \mathcal{W}_{\eta,e,s}(K_s \ast \mu_0, K_s \ast \mu_1)$ holds is immediate from the fact that the infimum in the definition of $\mathcal{W}_{\eta,e,s}$ runs only over a.c. curves connecting $K_s \ast \mu_0$ and $K_s \ast \mu_1$ which happen to be $K_s$-smoothings of (not necessarily a.c.) curves between $\mu_0$ and $\mu_1$; whereas, for $\mathcal{W}_{\eta,e}(K_s \ast \mu_0, K_s \ast \mu_1)$, the infimum runs over all a.c. curves connecting $K_s \ast \mu_0$ and $K_s \ast \mu_1$.

That $\mathcal{W}_{\eta,e,s}(K_s \ast \mu_0, K_s \ast \mu_1) \leq \mathcal{W}_{\eta,e}(\mu_0, \mu_1)$ follows from the fact that mollifying a.c. curves reduces their total action, see Proposition 2.22; and that the class of time-dependent mass-flux pairs $(\mu_t, j_t)_{t\in[0,1]}$ considered in the infimum defining $\mathcal{W}_{\eta,e,s}$ includes all a.c. curves connecting $\mu_0$ and $\mu_1$.

Proposition 5.10 Assume that $\eta$ and $\theta$ satisfy Assumptions 2.1 (i–iv) and 2.2 respectively. Let $s > 0$. Denote $\mu_s^\ast := K_s \ast \mu$ for any $\mu \in \mathcal{P}(\mathbb{R}^d)$. The following duality inequality for $\mathcal{W}_{\eta,e,s}$ holds:

$$\frac{1}{2} \mathcal{W}_{\eta,e,s}^2(\mu_0^\ast, \mu_1^\ast) \geq \sup \left\{ \int \phi_1(x)d\mu_1^\ast(x) - \int \phi_0(x)d\mu_0^\ast(x) : \phi_t \in HJ_{NL}^{1,s} \cap BL([0, 1] \times \mathbb{R}^d) \right\}.$$
Proof We follow rather closely the argument given in [23, Section 3]. Following the exposition there, we introduce the shorthand notation
\[ \langle \phi, \mu \rangle := \int_{\mathbb{R}^d} \phi(x) d\mu(x); \quad \langle \Phi, j \rangle := \int_G \Phi(x, y) \eta(x, y) d\eta(x, y). \]
We also use the notation \( \mu^s := K_s * \mu \) and \( j^s := K_s * j \). In this argument, we allow \( \mu_t \) to take values in \( \mathcal{M}^+(\mathbb{R}^d) \) rather than just \( \mathcal{P}(\mathbb{R}^d) \). Note that in this situation, the action \( A(\mu, j) \) and the nonlocal continuity equation are still well-defined. Additionally, there will be no loss of generality in assuming that, for paths we consider, \( A(\mu^s, j^s) < \infty \) for almost all \( t \in [0, 1] \); since otherwise, we will find that \( \mathcal{V}^2_{t,\varepsilon,s}(K_s * \mu_0, K_s * \mu_1) = \infty \), and so the proposition holds trivially. In particular, quantification over “all” \((\mu_t, j_t)_{t \in [0,1]}\) shall be understood to mean that:

- both \( \mu_t : [0, 1] \to \mathcal{M}^+(\mathbb{R}^d) \) is continuous, and \( j_t : [0, 1] \to \mathcal{M}_{loc}(G) \) is Borel measurable, w.r.t. the respective weak* topologies on \( \mathcal{M}^+(\mathbb{R}^d) \) and \( \mathcal{M}_{loc}(G) \); and

- \( A(\mu_t^s, j_t^s) < \infty \) for almost all \( t \in [0, 1] \).

By [19, Lemma 3.1], we know that, if \((\mu_t^s, j_t^s)_{t \in [0,1]}\) satisfies the nonlocal continuity equation in the sense of distributions, then for all \( \phi_t(x) \in C_c^\infty([0, 1] \times \mathbb{R}^d) 
\]
\[
\langle \phi_1, \mu_1^s \rangle - \langle \phi_0, \mu_0^s \rangle - \int_0^1 \left( \langle \partial_t \phi_t, \mu_t^s \rangle + \frac{1}{2} \langle \langle \bar{\nabla} \phi_t, j_t^s \rangle \rangle \right) dt = 0.
\]

Now consider, more generally, the case where \( \tilde{\phi}_t(x) \in BL([0, 1] \times \mathbb{R}^d) \). Reasoning as in [19, Remark 3.3], we approximate \( \tilde{\phi}_t(x) \) by functions \( \phi_t \in C_c^\infty([0, 1] \times \mathbb{R}^d) \) which are uniformly bounded in \( C^1([0, 1] \times \mathbb{R}^d) \) norm, and use the fact that
\[
|\langle \langle \bar{\nabla} \phi_t, j_t^s \rangle \rangle| \leq \left( \sup_{t \in [0,1]} \|\phi_t\|_{C^1(\mathbb{R}^d)} \right) \int \int (1 + |x - y|) \eta_t(x, y) d\eta_t^s(x, y) (x, y) \in L^1([0, 1])
\]
to pass to the limit in the nonlocal continuity equation, and deduce that more generally,
\[
\langle \phi_1, \mu_1^s \rangle - \langle \phi_0, \mu_0^s \rangle - \int_0^1 \left( \langle \partial_t \phi_t, \mu_t^s \rangle + \frac{1}{2} \langle \langle \bar{\nabla} \phi_t, j_t^s \rangle \rangle \right) dt = 0. \tag{5.1}
\]

Therefore, for any two \( s \)-smoothed probability measures \( \tilde{\mu}_0^s \) and \( \tilde{\mu}_1^s \) with \( \mathcal{V}^2_{t,\varepsilon,s}(\tilde{\mu}_0^s, \tilde{\mu}_1^s) < \infty \), we have that
\[
\frac{1}{2} \mathcal{V}^2_{t,\varepsilon,s}(\tilde{\mu}_0^s, \tilde{\mu}_1^s) = \inf_{(\mu_t, j_t)_{t \in [0,1]}, \mu_0 = \tilde{\mu}_0, \mu_1 = \tilde{\mu}_1} \left\{ \int_0^1 \frac{1}{2} A(\mu_t^s, j_t^s) dt : (\mu_t^s, j_t^s)_{t \in [0,1]} \text{satisfies (5.1)} \right\}.
\]

Introducing a Lagrange multiplier for the constraint “\((\mu_t^s, j_t^s)_{t \in [0,1]} \) satisfies (5.1)” we see that
\[
\frac{1}{2} \mathcal{V}^2_{t,\varepsilon,s}(\tilde{\mu}_0^s, \tilde{\mu}_1^s) = \inf_{(\mu_t, j_t)_{t \in [0,1]}, \mu_0 = \tilde{\mu}_0, \mu_1 = \tilde{\mu}_1} \sup_{\phi_t \in BL([0,1] \times \mathbb{R}^d)} \left\{ \int_0^1 \frac{1}{2} A(\mu_t^s, j_t^s) dt + \langle \phi_1, \mu_1^s \rangle - \langle \phi_0, \mu_0^s \rangle - \int_0^1 \left( \langle \partial_t \phi_t, \mu_t^s \rangle + \frac{1}{2} \langle \langle \bar{\nabla} \phi_t, j_t^s \rangle \rangle \right) dt \right\}.
\]
(Note that the the inner supremum takes the value \( +\infty \) unless (5.1) holds.)

Springer
Using the general fact that sup inf ≤ inf sup, we see that
\[
\frac{1}{2} \mathcal{W}^2_{\eta,s}(\tilde{\mu}_0^s, \tilde{\mu}_1^s) \geq \sup_{\phi_t \in BL([0,1] \times \mathbb{R}^d)} \inf_{(\mu_t, j_t) \in [0,1]} \left\{ \langle \phi_1, \mu_1^s \rangle - \langle \phi_0, \tilde{\mu}_0^s \rangle + \int_0^1 \left( \frac{1}{2} \mathcal{A}(\mu_t^s, j_t^s) - \langle \partial_t \phi_t, \mu_t^s \rangle - \frac{1}{2} \langle \nabla \phi_t, j_t^s \rangle \right) dt \right\}
\]
which in turn implies that (now letting the infimum quantify over a larger set, without fixed endpoints \(\tilde{\mu}_0^s\) and \(\tilde{\mu}_1^s\))
\[
\frac{1}{2} \mathcal{W}^2_{\eta,s}(\tilde{\mu}_0^s, \tilde{\mu}_1^s) \geq \sup_{\phi_t \in BL([0,1] \times \mathbb{R}^d)} \inf_{(\mu_t, j_t) \in [0,1]} \left\{ \langle \phi_1, \mu_1^s \rangle - \langle \phi_0, \tilde{\mu}_0^s \rangle \right\}
\]
Observe that due to the 1-homogeneity in \((\mu_t^s, j_t^s)\) of both the total action and nonlocal continuity equation, the inner infimum evaluates to \(-\infty\) unless \(\phi_t\) is chosen so that, for all \((\mu_t, j_t) \in [0,1],\)
\[
\int_0^1 \left( \frac{1}{2} \mathcal{A}(\mu_t^s, j_t^s) - \langle \partial_t \phi_t, \mu_t^s \rangle - \frac{1}{2} \langle \nabla \phi_t, j_t^s \rangle \right) dt \geq 0
\]
since otherwise we can just replace \((\mu_t^s, j_t^s)\) with \((\lambda \mu_t^s, \lambda j_t^s)\) and then send \(\lambda \to \infty\).

At the same time, since \(\mu_t^s\) has full support, and \(\mathcal{A}(\mu_t^s, j_t^s) < \infty\) by assumption, we have by [19, Lemma 2.3] that \(j_t^s \ll dx \otimes dy\). So, we can compute that
\[
\frac{1}{2} \mathcal{A}(\mu_t^s, j_t^s) - \frac{1}{2} \langle \nabla \phi_t, j_t^s \rangle = \frac{1}{4} \iint \left( \frac{d\mu_t^s}{dx} (x,y) - \nabla \phi_t (x,y) \theta \left( \frac{d\mu_t^s}{dx} (x), \frac{d\mu_t^s}{dy} (y) \right) \right)^2 \theta \left( \frac{d\mu_t^s}{dx} (x), \frac{d\mu_t^s}{dy} (y) \right) dy dx
\]
\[
- \frac{1}{4} \iint \left( \nabla \phi_t (x,y) \right)^2 \left( \frac{d\mu_t^s}{dx} (x), \frac{d\mu_t^s}{dy} (y) \right) \eta_e (x,y) dy dx
\]
Therefore, we conclude that
\[
\int_0^1 \left( \int \partial_t \phi_t d\mu_t^s + \frac{1}{4} \iint \left( \nabla \phi_t (x,y) \right)^2 \left( \frac{d\mu_t^s}{dx} (x), \frac{d\mu_t^s}{dy} (y) \right) \eta_e (x, y) dy dx \right) dt
\]
\[
\leq \int_0^1 \frac{1}{4} \iint \left( \frac{d\mu_t^s}{dx} (x,y) - \nabla \phi_t (x,y) \theta \left( \frac{d\mu_t^s}{dx} (x), \frac{d\mu_t^s}{dy} (y) \right) \right)^2 \theta \left( \frac{d\mu_t^s}{dx} (x), \frac{d\mu_t^s}{dy} (y) \right) \eta_e (x, y) dy dx dt.
\]
(5.2)
Again, this condition holds provided that the inner infimum is \(\geq 0\) (as opposed to \(-\infty\)). Of course, if it is the case that the inner infimum is nonnegative (for a given \(\phi_t (x) \in BL([0, 1] \times \mathbb{R}^d)\)), we certainly have that
\[
\langle \phi_1, \tilde{\mu_1}^s \rangle - \langle \phi_0, \tilde{\mu_0}^s \rangle + \inf_{(\mu_t, j_t) \in [0,1]} \int_0^1 \left( \frac{1}{2} \mathcal{A}(\mu_t^s, j_t^s) - \langle \partial_t \phi_t, \mu_t^s \rangle - \frac{1}{2} \langle \nabla \phi_t, j_t^s \rangle \right) dt \geq \langle \phi_1, \tilde{\mu_1}^s \rangle - \langle \phi_0, \tilde{\mu_0}^s \rangle.
\]
Therefore, we deduce the duality relation
\[
\frac{1}{2} \mathcal{W}_{\eta, s, \varepsilon}^2 (\tilde{\mu}_0, \tilde{\mu}_1) \geq \sup_{\mu \in B L([0,1] \times \mathbb{R}^d)} \{(\phi_1, \tilde{\mu}_1) - (\phi_0, \tilde{\mu}_0) : \forall (\mu_t, \tilde{\mu}_t) \in [0,1] \text{ (5.2) holds}\}.
\]

In turn, since in general, \(\exists \text{ an } s\text{-smooth nonlocal Hamilton–Jacobi subsolution, belonging to } B L([0,1] \times \mathbb{R}^d)\),

\[
\int_0^1 \int_0^1 \int \left( \frac{\partial \phi_t}{\partial x} (x, y) - \overline{\phi}_t (x, y) \theta \left( \frac{\partial \mu_t}{\partial x} (x), \frac{\partial \mu_t}{\partial y} (y) \right) \right) ^2 \eta_\varepsilon (x, y) dxdy dt \geq 0,
\]

this implies that
\[
\frac{1}{2} \mathcal{W}_{\eta, s, \varepsilon}^2 (\tilde{\mu}_0, \tilde{\mu}_1) \geq \sup_{\mu \in B L([0,1] \times \mathbb{R}^d)} \{|(\phi_1, \tilde{\mu}_1) - (\phi_0, \tilde{\mu}_0) : \forall (\mu_t, \tilde{\mu}_t) \in [0,1] \text{ (5.2) holds} | (\mu_t)_{t \in [0,1]} \}
\]

However, the quantity (†) is independent of \(j_r\). Therefore the statement reduces to quantification over \((\mu_t)_{t \in [0,1]}\) is for all weak*ly continuous “curves” in \(M^+ (\mathbb{R}^d)\), such that there exists a \((\tilde{\mu}_t)_{t \in [0,1]}\) for which \(A (\mu_t, j_r) < \infty\) for almost all \(t \in [0,1]\); but we do not even require satisfaction of the nonlocal continuity equation, so we can always just take \(j_r = 0\).

From this, by restricting the quantification only to “curves” \((\mu_t)_{t \in [0,1]}\) which are constant and belong to \(P (\mathbb{R}^d)\), we deduce that
\[
\frac{1}{2} \mathcal{W}_{\eta, s, \varepsilon}^2 (\tilde{\mu}_0, \tilde{\mu}_1) \geq \sup_{\mu \in P (\mathbb{R}^d)} \{|(\phi_1, \tilde{\mu}_1) - (\phi_0, \tilde{\mu}_0) : \forall (\mu_t, \tilde{\mu}_t) \in [0,1] \text{ (5.2) holds} | t \in [0,1] \}
\]

as desired. \(\square\)

**Proposition 5.11** Let \(\varepsilon \in (0,1)\) and \(s \geq \varepsilon\). Assume that \(\eta\) and \(\theta\) satisfy Assumptions 2.1 (i–iv) and 2.2 respectively. Assume also that \(M_5 (\eta) < \infty\). Suppose that \((\phi_t)_{t \in [0,1]} \in B L([0,1] \times \mathbb{R}^d)\) is a Hamilton–Jacobi subsolution in the sense that
\[
\partial_t \phi_t (x) + \frac{1}{2} |\nabla \phi_t (x)|^2 \leq 0 \text{ a.e.}
\]

Then,
\[
C = \frac{d^2}{M_2 (\eta)^2} \left[ \frac{3}{8} M_3 (\eta) + \sqrt{\left( \frac{M_2 (\eta)}{d} + \frac{3}{2} M_3 (\eta) \right) \left( M_4 (\eta) + \frac{3}{2} M_5 (\eta) \right) + \frac{1}{4} \left( M_4 (\eta) + \frac{3}{2} M_5 (\eta) \right)} \right]
\]

\(A = \sup_{t \in [0,1]} \| \nabla \phi_t \|_{L^\infty(\mathbb{R}^d)}\)

is an \(s\)-smooth nonlocal Hamilton–Jacobi subsolution, belonging to \(B L([0,1] \times \mathbb{R}^d)\).

**Proof** Let \(\phi_t \in B L([0,1] \times \mathbb{R}^d)\) satisfy \(\partial_t \phi_t (x) + \frac{1}{2} |\nabla \phi_t (x)|^2 \leq 0\) almost everywhere. Then \(\phi_t^s (x) := (K_s * \phi_t) (x)\) also satisfies
\[
\partial_t \phi_t^s (x) + \frac{1}{2} |\nabla \phi_t^s (x)|^2 \leq 0 \text{ (t, } x\text{-a.e.}.)
\]
by Lemma B.7. Additionally, we observe that $\phi_t^s \in BL([0, 1] \times \mathbb{R}^d)$, thanks to Lemma B.10.

We claim that some slight modification of

$$\tilde{\phi}_t^{s,e} := \frac{2d}{\varepsilon^2 M_2(\eta)} \phi_t^s,$$

more precisely

$$\tilde{\phi}_t^{s,e} := \tilde{\phi}_t^{s,e} - \frac{CA^2}{\varepsilon s} t$$

where $C > 0$ and $A > 0$ are some constants to be determined later on, is a subsolution to the $s$-smoothed nonlocal/HJ equation.

First, observe that since $\partial_t \phi_t^s(x) + \frac{1}{2} |\nabla \phi_t^s(x)|^2 \leq 0$, it follows that

$$\frac{2d}{\varepsilon^2 M_2(\eta)} \partial_t \phi_t^s(x) + \frac{2d}{\varepsilon^2 M_2(\eta)} \frac{1}{2} |\nabla \phi_t^s(x)|^2 \leq 0$$

and so

$$\partial_t \tilde{\phi}_t^{s,e} + \frac{1}{4} \varepsilon^2 \frac{M_2(\eta)}{d} |\nabla \tilde{\phi}_t^{s,e}(x)|^2 \leq 0.$$

Second, note that replacing $\tilde{\phi}_t^{s,e}$ with $\tilde{\phi}_t^{s,e} := \tilde{\phi}_t^{s,e} - \frac{CA^2}{\varepsilon s} t$ leaves the gradient unchanged (that is, $\nabla \tilde{\phi}_t^{s,e}(x) = \nabla \tilde{\phi}_t^{s,e}(x)$), whereas $\partial_t (\tilde{\phi}_t^{s,e} - \frac{CA^2}{\varepsilon s} t) = \partial_t \tilde{\phi}_t^{s,e} - \frac{CA^2}{\varepsilon s}$. In particular, given that

$$\partial_t \tilde{\phi}_t^{s,e} + \frac{1}{4} \varepsilon^2 \frac{M_2(\eta)}{d} |\nabla \tilde{\phi}_t^{s,e}(x)|^2 \leq 0$$

we have that

$$\partial_t \left( \tilde{\phi}_t^{s,e} - \frac{CA^2}{\varepsilon s} t \right) + \frac{1}{4} \varepsilon^2 \frac{M_2(\eta)}{d} |\nabla \left( \tilde{\phi}_t^{s,e} - \frac{CA^2}{\varepsilon s} t \right)(x)|^2 \leq - \frac{CA^2}{\varepsilon s}$$

so we know that $\tilde{\phi}_t^{s,e}$ is also a (local) HJ subsolution. At the same time, it is clear that $\tilde{\phi}_t^{s,e} \in BL([0, 1] \times \mathbb{R}^d)$, from the fact that $\phi_t^s \in BL([0, 1] \times \mathbb{R}^d)$.

Given some arbitrary $\mu \in \mathcal{P}(\mathbb{R}^d)$, we denote $\rho := K_s * \mu$. Note that $\rho$ is a Lebesgue density. Then, we see that

$$\int \partial_t \tilde{\phi}_t^{s,e}(x) d\rho(x) + \frac{1}{4} \int \int (\tilde{\phi}_t^{s,e}(y) - \tilde{\phi}_t^{s,e}(x))^2 \theta(\rho(x), \rho(y)) \eta_e(x, y) dx dy$$

$$\leq - \frac{1}{4} \varepsilon^2 \frac{M_2(\eta)}{d} \int |\nabla \tilde{\phi}_t^{s,e}(x)|^2 d\rho(x) - \frac{CA^2}{\varepsilon s}$$

$$+ \frac{1}{4} \int \int (\tilde{\phi}_t^{s,e}(y) - \tilde{\phi}_t^{s,e}(x))^2 \theta(\rho(x), \rho(y)) \eta_e(x, y) dx dy$$

$$\leq - \frac{1}{4} \varepsilon^2 \frac{M_2(\eta)}{d} \int |\nabla \tilde{\phi}_t^{s,e}(x)|^2 d\rho(x) - \frac{CA^2}{\varepsilon s}$$

$$+ \frac{1}{4} \int \int (\tilde{\phi}_t^{s,e}(y) - \tilde{\phi}_t^{s,e}(x))^2 \left(1 + \frac{3}{2s} |y - x| \right) \rho(x) \eta_e(x, y) dx dy$$

where we have used the fact that $K = c_K e^{-|x|}$, which implies (thanks to Lemma 2.24) that

$$\theta(\rho(x), \rho(y)) = \rho(x) \theta \left(1, \frac{\rho(y)}{\rho(x)} \right) \leq \rho(x) \theta \left(1, 1 + \frac{3}{s} |y - x| \right) \leq \rho(x) \left(1 + \frac{3}{2s} |y - x| \right).$$
Next, we replace \((\tilde{\phi}_t^{s,\varepsilon}(y) - \tilde{\phi}_t^{s,\varepsilon}(x))^2\) with an expression involving \(|\nabla \tilde{\phi}_t^{s,\varepsilon}|^2\), using a Taylor approximation. Note that such a Taylor series approximation requires sufficient regularity of \(\tilde{\phi}_t^{s,\varepsilon}\) (which is why we are using the smooth potential \(\tilde{\phi}_t^{s,\varepsilon}\) as our candidate \(s\)-smoothed nonlocal HJ subsolution).

To wit,
\[
|\tilde{\phi}_t^{s,\varepsilon}(y) - \tilde{\phi}_t^{s,\varepsilon}(x)| - |\nabla \tilde{\phi}_t^{s,\varepsilon}(x) \cdot (x - y)| \leq \frac{1}{2} \| D^2 \tilde{\phi}_t^{s,\varepsilon} \|_\infty |x - y|^2
\]
and so
\[
(\tilde{\phi}_t^{s,\varepsilon}(y) - \tilde{\phi}_t^{s,\varepsilon}(x))^2 \leq \frac{1}{4} \| D^2 \tilde{\phi}_t^{s,\varepsilon} \|_\infty^2 |x - y|^4
+
\| D^2 \tilde{\phi}_t^{s,\varepsilon} \|_\infty |x - y|^2 |\nabla \tilde{\phi}_t^{s,\varepsilon}(x) \cdot (x - y)| + |\nabla \tilde{\phi}_t^{s,\varepsilon}(x) \cdot (x - y)|^2.
\]

And, since \(|\nabla \tilde{\phi}_t^{s,\varepsilon}(x) \cdot (x - y)| \leq |\nabla \tilde{\phi}_t^{s,\varepsilon}(x)||x - y|\), it follows that
\[
\frac{1}{4} \int \int (\tilde{\phi}_t^{s,\varepsilon}(y) - \tilde{\phi}_t^{s,\varepsilon}(x))^2 \left( 1 + \frac{3}{2s} |x - y| \right) \rho(x) \eta_\varepsilon(x, y) dxdy
\leq \frac{1}{4} \int \int |\nabla \tilde{\phi}_t^{s,\varepsilon}(x) \cdot (x - y)|^2 \left( 1 + \frac{3}{2s} |x - y| \right) \rho(x) \eta_\varepsilon(x, y) dxdy
+
\frac{1}{4} \int \int \| D^2 \tilde{\phi}_t^{s,\varepsilon} \|_\infty |x - y|^2 |\nabla \tilde{\phi}_t^{s,\varepsilon}(x) \cdot (x - y)| \left( 1 + \frac{3}{2s} |x - y| \right) \rho(x) \eta_\varepsilon(x, y) dxdy
+
\frac{1}{16} \int \int \| D^2 \tilde{\phi}_t^{s,\varepsilon} \|_\infty^2 |x - y|^4 \left( 1 + \frac{3}{2s} |x - y| \right) \rho(x) \eta_\varepsilon(x, y) dxdy
\]
\[= I_\varepsilon + II_\varepsilon + III_\varepsilon.
\]

Since (as we will explicitly show below) the latter two terms are higher order in \(\varepsilon\), we initially focus our attention on the first term \(I_\varepsilon\).

In what follows, recall our notation \(M_\rho(\eta) \coloneqq \int |x - y|^\rho \eta(|x - y|) dxdy\); note that this does not depend on the choice of \(x \in \mathbb{R}^d\). Additionally, observe that \(M_\rho(\eta_\varepsilon) = \varepsilon^\rho M_\rho(\eta)\).

From [29, Corollary 2.16] (together with Lemma B.2), we know that
\[
\frac{1}{4} \int \int |\nabla \tilde{\phi}_t^{s,\varepsilon}(x) \cdot (x - y)|^2 \rho(x) \eta_\varepsilon(x, y) dxdy = \frac{1}{4} \varepsilon^2 \frac{M_2(\eta)}{d} \int |\nabla \tilde{\phi}_t^{s,\varepsilon}(x)|^2 \rho(x) dx.
\]
Similarly, compute that
\[
\int \int |\nabla \tilde{\phi}_t^{s,\varepsilon}(x) \cdot (x - y)|^2 \left( \frac{3}{2s} |x - y| \right) \rho(x) \eta_\varepsilon(x, y) dxdy
\leq \frac{3}{2s} \int \int |\nabla \tilde{\phi}_t^{s,\varepsilon}(x)|^2 |x - y|^3 \rho(x) \eta_\varepsilon(x, y) dxdy
= \frac{3}{2s} \varepsilon^3 M_3(\eta) \int |\nabla \tilde{\phi}_t^{s,\varepsilon}(x)|^2 \rho(x) dx.
\]
Consequently,
\[
I_\varepsilon \leq \left( \frac{1}{4} \varepsilon^2 \frac{M_2(\eta)}{d} + \frac{3}{8} \varepsilon^3 M_3(\eta) \right) \int |\nabla \tilde{\phi}_t^{s,\varepsilon}(x)|^2 \rho(x) dx.
\]
and so

\[
\frac{1}{4} \int \int (\hat{\phi}^{s,e}_t(y) - \check{\phi}^{s,e}_t(x))^2 \left( 1 + \frac{3}{2s} |y - x| \right) \rho(x) \eta(x, y) dxdy \\
\leq \left( \frac{1}{4} \frac{M_2(\eta)}{d} \epsilon^2 + \frac{3}{8} \frac{M_3(\eta)}{s} \epsilon^3 \right) \int \int |\nabla \hat{\phi}^{s,e}_t(x)|^2 \rho(x) dxdy + II_\epsilon + III_\epsilon.
\]

Therefore, we find that

\[
\begin{aligned}
\int \partial_t \hat{\phi}^{s,e}_t(x) d\rho(x) + \frac{1}{4} \int \int (\hat{\phi}^{s,e}_t(y) - \check{\phi}^{s,e}_t(x))^2 \theta(\rho(x), \rho(y)) \eta(x, y) dxdy \\
\leq \frac{1}{8} \frac{M_2(\eta)}{d} \int \int |\nabla \hat{\phi}^{s,e}_t(x)|^2 d\rho(x) \\
- \frac{CA^2}{\epsilon s} + \left( \frac{1}{4} \frac{M_2(\eta)}{d} \epsilon^2 + \frac{3}{8} \frac{M_3(\eta)}{s} \epsilon^3 \right) \int \int |\nabla \check{\phi}^{s,e}_t(x)|^2 \rho(x) dxdx \\
+ II_\epsilon + III_\epsilon \\
= - \frac{CA^2}{\epsilon s} + \frac{3}{8} \frac{M_3(\eta)}{s} \epsilon^3 \int \int |\nabla \hat{\phi}^{s,e}_t(x)|^2 \rho(x) dxdx + II_\epsilon + III_\epsilon \\
\leq 0.
\end{aligned}
\]

Hence, in order for \( \hat{\phi}^{s,e}_t \) to be an \( s \)-smoothed nonlocal HJ subsolution, it suffices to show that for \( s > \epsilon \) and an appropriate choice of \( A \) and \( C \),

\[
- \frac{CA^2}{\epsilon s} + \frac{3}{8} \frac{M_3(\eta)}{s} \epsilon^3 \int \int |\nabla \hat{\phi}^{s,e}_t(x)|^2 \rho(x) dxdx + II_\epsilon + III_\epsilon \leq 0.
\]

To see when this occurs, we first use the estimate

\[
|\nabla \hat{\phi}^{s,e}_t(x)|^2 = |\nabla \hat{\phi}^{s,e}_t(x)|^2 = |\nabla \left( \frac{2d}{\epsilon^2 M_2(\eta)} \phi^s_t \right)(x)|^2 \\
= \frac{4d^2}{\epsilon^4 M_2(\eta)^2} |\nabla \phi^s_t(x)|^2
\]

(Lemma B.9) \( \leq \frac{4d^2}{\epsilon^4 M_2(\eta)^2} |\nabla \phi_t(x)|^2 \). (x-a.s)

Consequently, for any probability density \( \rho(x) dxdx \), it holds that

\[
\frac{3}{8} \frac{M_3(\eta)}{s} \epsilon^3 \int \int |\nabla \hat{\phi}^{s,e}_t(x)|^2 \rho(x) dxdx \leq \frac{3}{2} \frac{d^2 M_3(\eta)}{2 \epsilon s M_2(\eta)^2} |\nabla \phi_t|^2.
\]

We also note that this implies that

\[
II_\epsilon \leq \left( \frac{d}{\epsilon^2 M_2(\eta)} + \frac{3}{2} \frac{d^2 M_3(\eta)}{2 \epsilon s M_2(\eta)^2} \right) |\nabla \phi_t|^2.
\]

Having analyzed the term \( I_\epsilon \) from the Taylor series expansion, we consider the first terms \( II_\epsilon \) and \( III_\epsilon \), both of which involve the quantity \( \| D^2 \hat{\phi}^{s,e}_t \|_\infty \). For the term \( III_\epsilon \) using the fact that \( \| D^2 \hat{\phi}^{s,e}_t \|_\infty = \| D^2 \check{\phi}^{s,e}_t \|_\infty = \frac{2d}{\epsilon^2 M_2(\eta)} \| D^2 \phi^s_t \|_\infty \),
and also, by Lemma B.8 that \(\|D^2 \phi_t^\varepsilon\|_{\infty}^2 \leq s^{-2} \|\nabla \phi_t\|_{\infty}^2\), we then see that

\[
III_\varepsilon = \frac{1}{16} \iint |x-y|^4 \left(1 + \frac{3}{2s}|y-x|\right) \rho(x) \eta_\varepsilon(x,y) \, dx \, dy
\]

\[
\leq \frac{1}{4} \frac{d^2}{(\varepsilon^2 M_2(\eta))^2} \frac{1}{s^2} \|\nabla \phi_t\|_{\infty}^2 \left(M_4(\eta) \varepsilon^4 + \frac{3}{2} M_5(\eta) \varepsilon^5 \right) \int \rho(x) \, dx
\]

\[
= \frac{1}{4} \frac{d^2}{M_2(\eta)^2} \frac{1}{s^2} \|\nabla \phi_t\|_{\infty}^2 \left(M_4(\eta) + \frac{3}{2} M_5(\eta) \varepsilon \right) .
\]

It remains to consider the second “mixed” term in the Taylor series expansion. Using Holder’s inequality, we observe that

\[
II_\varepsilon := \frac{1}{4} \iint |x-y|^2 |\nabla \phi_t^\varepsilon(x-y) \cdot (x-y)| \left(1 + \frac{3}{2s}|y-x|\right) \rho(x) \eta_\varepsilon(x,y) \, dx \, dy
\]

\[
\leq \left( \frac{1}{4} \iint |\nabla \phi_t^\varepsilon(x-y)|^2 \left(1 + \frac{3}{2s}|y-x|\right) \rho(x) \eta_\varepsilon(x,y) \, dx \, dy \right)^{1/2}
\]

\[
\times \left( \frac{1}{4} \iint |D^2 \phi_t^\varepsilon|_{\infty}^2 |x-y|^2 \left(1 + \frac{3}{2s}|y-x|\right) \rho(x) \eta_\varepsilon(x,y) \, dx \, dy \right)^{1/2}
\]

\[
= (I_\varepsilon \times 4 \cdot III_\varepsilon)^{1/2}.
\]

So, plugging in Eqs. 5.3 and 5.4, we see that

\[
II_\varepsilon \leq \left( \frac{d}{\varepsilon M_2(\eta)} + \frac{3}{2} \frac{d^2 M_3(\eta)}{s M_2(\eta)^2} \|\nabla \phi_t\|_{\infty}^2 \left(M_4(\eta) + \frac{3}{2} M_5(\eta) \varepsilon \right) \right)^{1/2}
\]

\[
= \frac{d^2}{\varepsilon s M_2(\eta)^2} \sqrt{M_2(\eta)} \varepsilon + \frac{3}{8} \frac{d^2 M_3(\eta)}{s M_2(\eta)^2} \|\nabla \phi_t\|_{\infty}^2 \left(M_4(\eta) + \frac{3}{2} M_5(\eta) \varepsilon \right) \|\nabla \phi_t\|_{\infty}^2.
\]

In sum: we find that it suffices to pick \(C\) in such a way that

\[
- \frac{CA^2}{\varepsilon s} \leq \frac{3}{8} \frac{d^2 M_3(\eta)}{s M_2(\eta)^2} \|\nabla \phi_t\|_{\infty}^2
\]

\[
+ \frac{d^2}{\varepsilon s M_2(\eta)^2} \sqrt{M_2(\eta)} \varepsilon + \frac{3}{8} \frac{d^2 M_3(\eta)}{s M_2(\eta)^2} \left(M_4(\eta) + \frac{3}{2} M_5(\eta) \varepsilon \right) \|\nabla \phi_t\|_{\infty}^2 \leq 0.
\]

Using the fact that \(s \geq \varepsilon\) and \(\varepsilon \in (0,1]\), we see that it suffices to pick

\[
C = \frac{d^2}{M_2(\eta)^2} \left[ \frac{3}{8} M_3(\eta) + \sqrt{M_2(\eta)} \varepsilon \left(M_4(\eta) + \frac{3}{2} M_5(\eta) \right) \right] + \frac{1}{4} \left(M_4(\eta) + \frac{3}{2} M_5(\eta) \right)
\]

and \(A^2 \geq \sup \|\nabla \phi_t\|_{\infty}^2\). \(\square\)

**Corollary 5.12** Suppose that \(\eta\) and \(\theta\) satisfy Assumptions 2.1 and 2.2 respectively. Let \(\varepsilon \in (0,1]\). Let \(\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^d)\) and assume both \(\mu_0\) and \(\mu_1\) are supported inside some set with diameter at most \(R\). Then, we have that

\[
W_2^2(\mu_0, \mu_1) \leq \varepsilon^2 \frac{M_2(\eta)}{2d} \mathcal{W}_{\eta, \varepsilon}^2(\mu_0, \mu_1) + \left(\frac{7}{4}dR^2 + 8dR\right) \sqrt{\varepsilon}.
\]
Proof Suppose by Corollary 5.3 that \((\phi_t)_{t \in [0, 1]}\) is the optimal HJ subsolution for \((\mu_0, \mu_1)\), that is, \((\phi_t)_{t \in [0, 1]}\) satisfies

\[
\frac{1}{2} W^2_2(\mu_0, \mu_1) = \mathop{\text{argmax}}_{\phi \in BL((0, 1] \times \mathbb{R}^d)} \left\{ \int \phi_t d\mu_1 - \int \phi_0 d\mu_0 : \partial_t \phi_t + \frac{1}{2} |\nabla \phi_t|^2 = 0 \text{ in viscosity sense} \right\}.
\]

Note that \((\phi_t)_{t \in [0, 1]}\) is also a (not necessarily optimal) Hamilton–Jacobi solution for \((K_{2s} \ast \mu_0, K_{2s} \ast \mu_1)\), and so

\[
\frac{1}{2} W^2_2(\mu_0, \mu_1) \geq \frac{1}{2} W^2_2(K_{2s} \ast \mu_0, K_{2s} \ast \mu_1) \geq \int \phi_t d(K_{2s} \ast \mu_1) - \int \phi_0 d(K_{2s} \ast \mu_0).
\]

Furthermore,

\[
\int \phi_t d(K_{2s} \ast \mu_1) - \int \phi_0 d(K_{2s} \ast \mu_0) = \int \phi_t^i d(K_s \ast \mu_1) - \int \phi_0^i d(K_s \ast \mu_0) = \int \phi_t^{2s} d\mu_1 - \int \phi_0^{2s} d\mu_0.
\]

At the same time, \(\phi_t^i \to \phi_1\) and \(\phi_0^i \to \phi_0\) uniformly on compact sets as \(s \to 0\). Quantitatively, by Corollary 5.3 one has the estimate

\[
|\phi_i - K_s \ast \phi_1| \leq \epsilon \cdot \text{Lip} \phi \leq \epsilon \cdot RM_1(K)
\]

where \(M_1(K) := \int_{\mathbb{R}^d} c_K |x| e^{-|x|^2} dx\). Hence

\[
\int |\phi_i - \phi_t^{2s}| d\mu_i \leq 2 \epsilon \cdot RM_1(K) \quad i = 0, 1
\]

and therefore

\[
\int \phi_t^{2s} d\mu_1 - \int \phi_0^{2s} d\mu_0 \geq \int \phi_1 d\mu_1 - \int \phi_0 d\mu_0 - 4 \epsilon \cdot RM_1(K).
\]

Together, this implies that

\[
\frac{1}{2} W^2_2(\mu_0, \mu_1) \geq \int \phi_t^i d(K_s \ast \mu_1) - \int \phi_0^i d(K_s \ast \mu_0) \geq \frac{1}{2} W^2_2(\mu_0, \mu_1) - 4 \epsilon \cdot RM_1(K).
\]

Since \(\phi_t\) is a viscosity solution of the Hamilton–Jacobi equation and is Lipschitz, it is also a Lebesgue \((x, t)\)-almost everywhere solution, by Rademacher’s theorem and [26, Theorem 10.1.1]. Therefore, we can apply Proposition 5.11, and deduce (defining \(\tilde{\phi}_t^{s, \epsilon} := \frac{2d}{\epsilon^2 M_2(\eta)} \phi_t^{s, \epsilon}\), as in Proposition 5.11) that

\[
\frac{d}{\epsilon^2 M_2(\eta)} W^2_2(\mu_0, \mu_1) \geq \int \tilde{\phi}_t^{s, \epsilon} d(K_s \ast \mu_1) - \int \tilde{\phi}_0^{s, \epsilon} d(K_s \ast \mu_0) \geq \frac{d}{\epsilon^2 M_2(\eta)} (W^2_2(\mu_0, \mu_1) - 8 \epsilon \cdot RM_1(K))
\]

and likewise (defining \(\tilde{\phi}_t^{s, \epsilon} := \tilde{\phi}_t^{s, \epsilon} - \frac{CA^2}{\epsilon^2} t\), as in Proposition 5.11)

\[
\int \tilde{\phi}_t^{s, \epsilon} d(K_s \ast \mu_1) - \int \tilde{\phi}_0^{s, \epsilon} d(K_s \ast \mu_0) = \int (\tilde{\phi}_t^{s, \epsilon} - \frac{CA^2}{\epsilon^2}) d(K_s \ast \mu_1) - \int \tilde{\phi}_0^{s, \epsilon} d(K_s \ast \mu_0)
\]
Observe that $\phi_{i}^{s, \epsilon} \in BL([0, 1] \times \mathbb{R}^{d})$. At the same time, we know, by Lemma 5.3, that $\sup_{t} \| \nabla \phi_{i} \|_{\infty}^{2} \leq R^{2}$; so if we put $A = R$ and $C$ as in Proposition 5.11, then by Proposition 5.11, $\phi_{i}^{s}$ is an $s$-smooth nonlocal HJ subsolution. Therefore, we have that

$$\int \phi_{i}^{s, \epsilon} d(K_{s} * \mu_{0}) - \int \phi_{0}^{s, \epsilon} d(K_{s} * \mu_{0}) \leq \sup_{\phi_{i}(x) \in H^{0}_{r,s}(BL([0,1] \times \mathbb{R}^{d}) \setminus BL([0,1] \times \mathbb{R}^{d}) \setminus BL([0,1] \times \mathbb{R}^{d}), \phi_{0}(x) \in \bar{H}^{0}_{r,s}(BL([0,1] \times \mathbb{R}^{d}) \setminus BL([0,1] \times \mathbb{R}^{d}) \setminus BL([0,1] \times \mathbb{R}^{d}))} \int \phi_{1} d(K_{s} * \mu_{0})$$

(Proposition 5.10) \leq \frac{1}{2} W_{\eta, s}^{2}(K_{s} * \mu_{0}, K_{s} * \mu_{1}).

This implies that

$$\frac{d}{\epsilon^{2} M_{2}(\eta)} (W_{2}^{2}(\mu_{0}, \mu_{1}) - 8s R M_{1}(K)) - \frac{C R^{2}}{\epsilon s} \leq \frac{1}{2} W_{\eta, s}^{2}(K_{s} * \mu_{0}, K_{s} * \mu_{1})$$

and thus

$$W_{2}^{2}(\mu_{0}, \mu_{1}) - 8s R M_{1}(K) \leq \frac{\epsilon^{2} M_{2}(\eta)}{2d} \left( W_{\eta, s}^{2}(K_{s} * \mu_{0}, K_{s} * \mu_{1}) + \frac{C R^{2}}{\epsilon s} \right).$$

Now, since $\eta$ is supported on $B(0, 1)$, all of the higher moments $M_{3}(\eta), M_{4}(\eta),$ and $M_{5}(\eta)$ are bounded above by $M_{2}(\eta)$, we find that

$$C \leq \frac{d^{2}}{M_{2}(\eta)} \left[ \frac{3}{8} + \sqrt{\frac{1}{d} + \frac{5}{2}} \right]$$

so that

$$W_{2}^{2}(\mu_{0}, \mu_{1}) \leq \frac{\epsilon^{2} M_{2}(\eta)}{2d} \left[ W_{\eta, s}^{2}(K_{s} * \mu_{0}, K_{s} * \mu_{1}) + \frac{7}{4} d R^{2} \frac{\epsilon}{s} + 8s R M_{1}(K) \right].$$

Since $W_{\eta, s}^{2}(K_{s} * \mu_{0}, K_{s} * \mu_{1}) \leq W_{\eta, s}^{2}(\mu_{0}, \mu_{1})$ by Lemma 5.9, we deduce that

$$W_{2}^{2}(\mu_{0}, \mu_{1}) \leq \frac{\epsilon^{2} M_{2}(\eta)}{2d} \left[ W_{\eta, s}^{2}(\mu_{0}, \mu_{1}) + \frac{7}{4} d R^{2} \frac{\epsilon}{s} + 8s R M_{1}(K) \right].$$

Finally, we set $s = \sqrt{\epsilon}$, and use the fact that $M_{1}(K) = d$, by Lemma B.6.

Acknowledgements The authors thank Giuseppe Buttazzo, Matthias Erbar, Wilfrid Gangbo, Nicolás García Trillos, Giovanni Leoni, and Jan Maas for helpful discussions. The authors are grateful to NSF for support via Grants DMS 1814991 and DMS 2206069. Part of this work was done while the authors were visiting the Simons Institute for the Theory of Computing. The authors are thank the institute for hospitality. Furthermore they are thankful to CNA of CMU for support.

Funding Open Access funding provided by Carnegie Mellon University.

Data Availability There are no data generated or used by this research.
Appendix A: Lower semicontinuity of integral functionals on $\mathcal{M}_{loc}$

In this appendix, we establish the following variant of Reshetnyak’s theorem as well as some direct consequences of this theorem which are used in the main body of the article (for other variants on this theorem, we refer the reader to [1, 2, 9, 42, 44]).

**Theorem A.1 (locally finite, topological and sequential “Reshetnyak’s theorem”)** Let $\Omega$ be a locally compact Polish space, and let $f : \Omega \times \mathbb{R}^n \to [0, \infty]$ be a (topologically) lower semicontinuous function such that for every $\omega \in \Omega$, the function $f(\omega, \cdot)$ is convex and positively 1-homogeneous. Then the functional

$$F : \mathcal{M}_{loc}(\Omega, \mathbb{R}^n) \to [0, \infty]$$

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

is convex, and both topologically and sequentially weak* lower semicontinuous.

In the statement of this theorem, the “weak* topology on $\mathcal{M}_{loc}(\Omega, \mathbb{R}^d)$” has the following sense: we consider $\mathcal{M}_{loc}(\Omega, \mathbb{R}^n)$ as the dual space of $C_c(\Omega, \mathbb{R}^n)$, where $C_c(\Omega, \mathbb{R}^n)$ is understood as a locally convex t.v.s. which is a direct limit of the Banach spaces $C_c(K, \mathbb{R}^n)$ for every compact $K \subset \Omega$. (See also discussion on this point in discussion in [2, Chapter 1].) The reason why the statement of the theorem carefully specifies that $F$ is both topologically and sequentially weak* l.s.c. is that in this setting, it is not known to the authors whether the two notions coincide. (In particular, $\mathcal{M}_{loc}(\Omega, \mathbb{R}^d)$ is not the weak* dual of a separable Banach space, so we cannot apply [9, Proposition 1.1.6 (iii)].) For our immediate purposes, the sequential weak* l.s.c. property is used in variational arguments, while $F$ being topologically weak* l.s.c. implies directly that $F$ is weak* Borel measurable.

**Proof** Convexity of $F$ follows directly, regardless of the underlying domain, from the convexity and 1-homogeneity of $f$.

Let $K \subset \Omega$ be compact. Consider the functional

$$\lambda \mapsto \int_K f \left( \omega, \frac{d\pi_K \lambda}{d|\pi_K \lambda|}(\omega) \right) d|\pi_K \lambda|(\omega)$$

namely precomposition of the continuous projection $\pi_K$ with the functional $F(\cdot \upharpoonright K)$ (which has domain $\mathcal{M}(K, \mathbb{R}^n)$). The functional $F(\cdot \upharpoonright K)$ is known, by Reshetnyak’s theorem (more precisely, [9, Theorem 3.4.3]), to be sequentially weak* lower semicontinuous on $\mathcal{M}(K, \mathbb{R}^n)$. Since $\mathcal{M}(K, \mathbb{R}^n)$ is the dual of a separable Banach space, and $F(\cdot \upharpoonright K)$ is convex, $F(\cdot \upharpoonright K)$ is also topologically weak* lower semicontinuous on $\mathcal{M}(K, \mathbb{R}^n)$. Since the precomposition of a continuous map with a (topologically) l.s.c. map is again l.s.c., it follows that $F(\cdot \upharpoonright K) \circ \pi_K$ is a topologically lower semicontinuous map from $\mathcal{M}_{loc}(\Omega, \mathbb{R}^n)$ to $[0, \infty]$. 

---

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. 

---

Nonlocal Wasserstein distance: metric and asymptotic properties

Page 53 of 66   238
Now, consider a sequence \((K_n)\) of compact sets which exhaust \(\Omega\). Observe that
\[
\int_{\Omega} f \left( \omega, \frac{d\lambda}{d|\lambda|}(\omega) \right) d|\lambda|(\omega) = \sup_n \int_{K_n} f \left( \omega, \frac{d\lambda}{d|\lambda|}(\omega) \right) d|\lambda|(\omega)
\]
\[
= \sup_n \int_{K_n} f \left( \omega, \frac{d\pi_{K_n}\lambda}{d|\pi_{K_n}\lambda|}(\omega) \right) d|\pi_{K_n}\lambda|(\omega).
\]
Consequently, \(\int_{\Omega} f \left( \omega, \frac{d\lambda}{d|\lambda|}(\omega) \right) d|\lambda|(\omega)\) is a supremum of topologically l.s.c. functionals, and therefore is topologically l.s.c. on \(\mathcal{M}_{\text{loc}}(\Omega, \mathbb{R}^n)\) as well [9, Proposition 1.1.2 (ii)].

In fact, from this one can also deduce that \(\int_{\Omega} f \left( \omega, \frac{d\lambda}{d|\lambda|}(\omega) \right) d|\lambda|(\omega)\) is also sequentially weak* l.s.c.: this follows from the fact that a function is sequentially l.s.c. with respect to some topology \(\tau\) iff it is l.s.c. with respect to the sequential topology induced by \(\tau\), and this topology is at least as fine as \(\tau\) itself (see discussion in [9], especially [9, Proposition 1.1.5]).

\[\Box\]

**Corollary A.2** (compare [19, Lemma 2.4] and [25, Lemma 2.9]) The action \(A(\mu, j; m)\) is jointly convex, and jointly topologically and sequentially weak* l.s.c., on \(\mathcal{P}(\mathbb{R}^d) \times \mathcal{M}_{\text{loc}}(G) \times \mathcal{M}_{\text{loc}}^+(\mathbb{R}^d)\), also more generally on the larger space \(\mathcal{M}_{\text{loc}}^+(\mathbb{R}^d) \times \mathcal{M}_{\text{loc}}(G) \times \mathcal{M}_{\text{loc}}^+(\mathbb{R}^d)\).

**Proof** Joint convexity follows exactly as in [19, Lemma 2.7] and [25, Lemma 2.12].

We apply Theorem A.1 in the case where the space of vector measures is \(\mathcal{M}_{\text{loc}}(\mathbb{R}^{2d}; \mathbb{R}^3)\) and
\[
f \left( (x, y), \frac{d\lambda}{d|\lambda|}(x, y) \right) = \frac{\left( \frac{d\lambda}{d|\lambda|} \right)^2}{2\theta \left( \frac{d\lambda}{d|\lambda|}, \frac{d\lambda}{d|\lambda|} \right)} \eta(x, y) 1_G(x, y)
\]
with the convention that \(\frac{0}{0} = 0\), and that \(\theta(r, s) = 0\) if either \(r\) or \(s\) is negative. Precomposing this \(\int_{\mathbb{R}^{2d}} f \left( (x, y), \frac{d\lambda}{d|\lambda|}(x, y) \right) d|\lambda|(x, y)\) with the continuous embedding
\[
\mathcal{M}_{\text{loc}}^+(\mathbb{R}^d) \times \mathcal{M}_{\text{loc}}(G) \times \mathcal{M}_{\text{loc}}^+(\mathbb{R}^d) \hookrightarrow \mathcal{M}_{\text{loc}}(\mathbb{R}^{2d}; \mathbb{R}^3)
\]
\[(\mu, j, m) \mapsto (\mu \times m, m \times \mu, j)\]
demonstrates that \(A(\mu, j; m)\) is topologically l.s.c. on \(\mathcal{M}_{\text{loc}}(\mathbb{R}^d) \times \mathcal{M}_{\text{loc}}(G) \times \mathcal{M}_{\text{loc}}^+(\mathbb{R}^d)\) as desired, and sequential lower semicontinuity follows exactly as before. \(\Box\)

As an application, we give a proof of the following convolution inequality.

**Lemma A.3** Let \(\mu \in \mathcal{P}(\mathbb{R}^d)\) and \(j \in \mathcal{M}_{\text{loc}}(G)\), and suppose that \(A_{\eta, \theta}(\mu, j) < \infty\). Let \(k\) be a convolution kernel, and for each \(z \in \mathbb{R}^d\), let \(\mu_z\) and \(j_z\) denote the \(z\)-translates of \(\mu\) and \(j\) respectively: namely for all \(\psi \in C^\infty_c(\mathbb{R}^d)\) and \(\varphi \in C^\infty_c(G)\),
\[
\int_{\mathbb{R}^d} \psi(x) d\mu_z(x) := \int_{\mathbb{R}^d} \psi(x + z) d\mu(x);
\]
\[
\int_{G} \varphi(x, y) d\mu_z(x, y) = \int_{G} \varphi(x + z, y + z) d\mu(z, y).
\]

Then,
\[
A_{\eta, \theta}(k * \mu, k * j) \leq \int_{\mathbb{R}^d} A_{\eta, \theta}(\mu_z, j_z) k(z) dz.
\]

\[\Box\] Springer
We note that morally, this is just an application of Jensen’s inequality in a topological vector space, but we are unaware of a version of Jensen’s inequality in a t.v.s. that applies to this setting. Briefly, the relevant version of Jensen’s inequality in a locally convex t.v.s. that we are aware of is [41, Theorem 3.10], which applies to a convex l.s.c. function with support which is both convex, and either is closed or has non-empty interior. The issue here is that the action $A(\mu, j)$ is defined only when the first argument $\mu$ is a nonnegative measure, and the cone of nonnegative measures has empty interior inside the t.v.s. of signed measures. Nor can we expect that the support of $A$ is closed, at least for a general choice of $\theta$. Therefore we give an ad hoc proof, essentially the same (albeit with additional complications) to the proof of Proposition 3.10.

**Proof** Let $(\Omega, A, \mathbb{P})$ be a probability space, and let $Z : \Omega \to \mathbb{R}^d$ be a random variable with distribution $k(|z|)dz$. Furthermore, let $Z_1, Z_2, \ldots$ be i.i.d. copies of $Z$. Then, $(\mu_Z, j_Z)$ is a $\mathcal{P}(\mathbb{R}^d) \times \mathcal{M}_{loc}(G)$-valued random variable, and since $A_{\eta, \theta}$ is topologically l.s.c. and hence Borel, it follows that $A_{\eta, \theta}(\mu_Z, j_Z)$ is a $[0, \infty]$-valued random variable with distribution $\mathcal{A}_{\eta, \theta}(\mu_Z, j_Z)k(|z|)dz$.

Since $A_{\eta, \theta}$ is jointly convex, we observe that

$$A_{\eta, \theta} \left( \frac{1}{n} \sum_{i=1}^{n} \mu_{Z_i}, \frac{1}{n} \sum_{i=1}^{n} j_{Z_i} \right) \leq \frac{1}{n} \sum_{i=1}^{n} A_{\eta, \theta}(\mu_{Z_i}, j_{Z_i}).$$

Now, suppose that $\int_{\mathbb{R}^d} A_{\eta, \theta}(\mu_z, j_z)k(|z|)dz < \infty$, since otherwise the lemma holds trivially. Since the action is nonnegative, the strong law of large numbers shows that

$$\frac{1}{n} \sum_{i=1}^{n} A_{\eta, \theta}(\mu_{Z_i}, j_{Z_i}) \to \int_{\mathbb{R}^d} A_{\eta, \theta}(\mu_z, j_z)k(|z|)dz$$

almost surely. Therefore, if we can verify that $\frac{1}{n} \sum_{i=1}^{n} \mu_{Z_i} \xrightarrow{\text{a.s.}} k * \mu$ and $\frac{1}{n} \sum_{i=1}^{n} j_{Z_i} \xrightarrow{\text{a.s.}} k * j$ almost surely, it then follows from the joint sequential lower semicontinuity of $A_{\eta, \theta}$ that

$$A_{\eta, \theta}(k * \mu, k * j) \leq \liminf_{n \to \infty} A_{\eta, \theta} \left( \frac{1}{n} \sum_{i=1}^{n} \mu_{Z_i}, \frac{1}{n} \sum_{i=1}^{n} j_{Z_i} \right)$$

and the lemma is proved.

We therefore verify that $\frac{1}{n} \sum_{i=1}^{n} \mu_{Z_i} \xrightarrow{\text{a.s.}} k * \mu$ and $\frac{1}{n} \sum_{i=1}^{n} j_{Z_i} \xrightarrow{\text{a.s.}} k * j$ almost surely; note that this is a similar result to the Glivenko-Cantelli theorem.

Let $\psi \in C_0(\mathbb{R}^d)$. Then $\int_{\mathbb{R}^d} \psi(x)d\mu_Z(x)$ is an $L^1$ random variable, in fact $L^\infty$ since for all $\omega \in \Omega$

$$\left| \int_{\mathbb{R}^d} \psi(x)d\mu_{Z(\omega)}(x) \right| \leq \|\psi\|_\infty.$$ 

It follows from the strong law of large numbers that

$$\int_{\mathbb{R}^d} \psi(x)d \left( \frac{1}{n} \sum_{i=1}^{n} \mu_{Z_i}(x) \right) = \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}^d} \psi(x)d\mu_{Z_i}(x)$$

(a.s.) $\to \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} \psi(x)d\mu_z(x) \right) k(|z|)dz$

$$= \int_{\mathbb{R}^d} \psi(x)d \left( \int_{\mathbb{R}^d} \mu_zk(|z|)dz \right)(x)$$
= \int_{\mathbb{R}^d} \psi(x) d(k \ast \mu)(x).

By [34, Lemma 4.8 (i)] this implies that $\frac{1}{n} \sum_{i=1}^{n} \mu_{Z_i} \overset{\ast}{\to} k \ast \mu$ almost surely (more specifically, in duality with $C_0(\mathbb{R}^d)$). We will use similar, albeit more involved, reasoning to show that $\frac{1}{n} \sum_{i=1}^{n} j_{Z_i} \overset{\ast}{\to} k \ast j$ almost surely as well.

Consider the Jordan decomposition $j = j^+ - j^-$. By globalization of [19, Theorem 2.8] this is a measurable operation, and so both $j_Z^+$ and $j_Z^-$ are $\mathcal{M}_{loc}^+(G)$-valued random variables. Note also that

$$k \ast j = \int_{\mathbb{R}^d} j_k(|z|) dz = \int_{\mathbb{R}^d} (j_Z^+ - j_Z^-) k(|z|) dz = k \ast j^+ - k \ast j^-.$$ 

Let $K \subset G$ be any compact set, and let $\varphi \in C_c(K)$. Then, we claim that both of

$$\int_K \int_{\mathbb{R}^d} \varphi(x, y) \eta(|x - y|) d(j^+_Z(x, y)) \quad \text{and} \quad \int_K \int_{\mathbb{R}^d} \varphi(x, y) \eta(|x - y|) d(j^-_Z(x, y))$$

are $L^1$ random variables, in fact $L^\infty$. It suffices to show that $|\eta(|x - y|) d(j^+_Z(x, y))|(K)$ is uniformly bounded.

In [19, Lemma 2.6] it is shown that for some uniform constant $C$, the measure $\nu = \eta(|x - y|) d(j^+ \ast \mu)(x, y)$ satisfies the property

$$|\nu|(K) \leq \frac{1}{d} \int_G (1 \wedge |x - y|) d\nu(x, y) \leq C A(\mu, j)$$

where $a = \min\{|x - y| : (x, y) \in K\}$. Denoting

$$K - \left( \frac{z}{a} \right) := \{(x - z, y - z) : (x, y) \in K\}$$

we see that $a$ is invariant under translation of $K$; and $\eta(|x - y|)$ is invariant under translation of $(x, y)$ by $z$, so in fact for all $z \in \mathbb{R}^d$ uniformly,

$$|\eta(|x - y|) d(j^+ Z(x, y))|(K) \leq C A(\mu_z, j_z).$$

Furthermore, as observed in [19], $A(\mu, j)$ is invariant under translation of $(\mu, j)$ since we have taken the reference measure on $\mathbb{R}^d$ to be the Lebesgue measure. Therefore, it holds (uniformly in $z$) that

$$\left| \int_K \varphi(x, y) \eta(|x - y|) d(j^+_Z(x, y)) \right| \leq \|\varphi\|_{L^\infty(K)} |\eta(|x - y|) d(j)_Z(x, y)|(K)$$

and the same holds for $j^-_Z$ also.

Now, it follows from the strong law of large numbers that

$$\int_K \varphi(x, y) \eta(|x - y|) d \left( \frac{1}{n} \sum_{i=1}^{n} j^+_Z(x, y) \right) = \frac{1}{n} \sum_{i=1}^{n} \int_K \varphi(x, y) \eta(|x - y|) d(j^+_Z(x, y))$$

(a.s.) $\to \int_{\mathbb{R}^d} \left( \int_K \varphi(x, y) \eta(|x - y|) d(j^+_Z(x, y)) \right) k(|z|) dz$

$$= \int_K \varphi(x, y) \eta(|x - y|) d \left( \int_{\mathbb{R}^d} j^+_Z k(|z|) dz \right) (x, y)$$
\[
\int_K \varphi(x, y) \eta(|x - y|) d(k * j^+)(x, y).
\]

By [34, Lemma 4.8 (i)] this implies that almost surely,
\[
\eta(|x - y|) d \left( \frac{1}{n} \sum_{i=1}^{n} j^+_{Z_t}(x, y) \right) \rightharpoonup K \overset{*}{\rightarrow} \eta(|x - y|) d(k * j^+)(x, y) \rightharpoonup K
\]
(where the convergence is in duality with \( C_C(K) \); since \( j \) is supported on \( G := \{(x, y) : \eta(|x - y|) > 0\} \) (and therefore, so is \( j \) for every \( z \), since \( G \) is translation-invariant) and \( \eta \) is a continuous function of compact support when restricted to any \( K \subset G \), this implies that almost surely,
\[
\left( \frac{1}{n} \sum_{i=1}^{n} j^+_{Z_t}(x, y) \right) \rightharpoonup_k k * j^+ \rightharpoonup K
\]
again in duality with \( C_C(K) \). Since \( K \) was arbitrary, by the characterization of weak* convergence in \( M_{loc}(G) \) we have that almost surely,
\[
\left( \frac{1}{n} \sum_{i=1}^{n} j^+_{Z_t}(x, y) \right) \rightharpoonup_k k * j^+ \rightharpoonup K
\]

Finally, this implies that \( \left( \frac{1}{n} \sum_{i=1}^{n} j^+_{Z_t}(x, y) \right) \rightharpoonup_k k \ast j \) almost surely. \( \Box \)

Appendix B: Additional lemmas

Lemma B.1 (estimate adapted from the two point \( \mathcal{W} \) space) Let \( \varepsilon > 0 \). Let \( A \) and \( B \) be disjoint, bounded subsets of \( \mathbb{R}^d \) of positive Lebesgue measure, such that \( \sup_{x \in A, y \in B} |x - y| < \varepsilon \). Suppose without loss of generality that \( |A| \leq |B| \). Let \( m_A \) and \( m_B \) denote the uniform probability measures supported on \( A \) and \( B \), respectively. Then,
\[
\mathcal{W}_{\eta, \varepsilon}(m_A, m_B) \leq \frac{C_\theta}{4 \sqrt{|A| \eta_\varepsilon(\sup_{x \in A, y \in B} |x - y|)}}
\]
where \( C_\theta := \int_0^1 \frac{1}{\sqrt{\theta(1-r,1+r)}} dr \).

Proof We use a construction (as suggested by [29]) from Lemma 2.3 (and Theorem 2.4) from [37]. If the 2-point \( \mathcal{W}_\theta \) space is connected, then the proof Lemma 2.3 therein indicates that for any \( \varepsilon_0 > 0 \), there exist differentiable functions \( \gamma : [0, 1] \rightarrow [-1, 1] \) and \( \chi : [0, 1] \rightarrow \mathbb{R} \) such that \( \gamma(0) = -1 \), \( \gamma(1) = 1 \), and
\[
\forall t \in (0, 1) \quad \gamma_t' = 4 \theta(1 + \gamma_t, 1 - \gamma_t) \chi_t,
\]
and,
\[
\int_0^1 \theta(1 + \gamma_t, 1 - \gamma_t) \chi_t^2 dt < \mathcal{W}_\theta^2(\delta_0, \delta_1) + \varepsilon_0.
\]
(Here \( \mathcal{W}_\theta(\delta_0, \delta_1) \) denotes the \( \mathcal{W}_\theta \) distance between two Dirac masses on the two-point graph, with edge weight 1/2.)
Consider the curve $\rho_t : [0, 1] \to \mathcal{P}(\mathbb{R}^d)$ defined by

$$
\frac{d\rho_t}{d\text{Leb}}(x) = \begin{cases} 
\frac{1-\gamma}{2|A|} & x \in A \\
\frac{1+\gamma}{2|B|} & x \in B \\
0 & \text{else}.
\end{cases}
$$

Note also that by construction,

$$
\frac{d}{dt} \frac{d\rho_t}{d\text{Leb}}(x) = \begin{cases} 
-\frac{\theta(1+\gamma, 1-\gamma)\chi_t}{8|A|} & x \in A \\
\frac{\theta(1+\gamma, 1-\gamma)\chi_t}{8|B|} & x \in B \\
0 & \text{else}.
\end{cases}
$$

Let $j_t(x, y)$ be a flux so that $(\rho_t, j_t)$ solves the nonlocal continuity equation; in particular we set

$$
\frac{d j_t}{d(\text{Leb} \otimes \text{Leb})}(x, y) = \begin{cases} 
\frac{\theta(1+\gamma, 1-\gamma)\chi_t}{16\eta_t(x, y)|A||B|} (x, y) \in A \times B \\
\frac{\theta(1+\gamma, 1-\gamma)\chi_t}{16\eta_t(x, y)|A||B|} (x, y) \in B \times A \\
0 & \text{else}.
\end{cases}
$$

Now, observe that $\text{Leb} \otimes \text{Leb}$ dominates all of $j_t, \rho_t \otimes \text{Leb}$, and $\text{Leb} \otimes \rho_t$. Furthermore, note that

$$
\frac{d(\rho_t \otimes \text{Leb})}{d(\text{Leb} \otimes \text{Leb})}(x, y) = \frac{d\rho_t}{d\text{Leb}}(x),
$$

and similarly

$$
\frac{d(\text{Leb} \otimes \rho_t)}{d(\text{Leb} \otimes \text{Leb})}(x, y) = \frac{d\rho_t}{d\text{Leb}}(y).
$$

Together with the fact that $j_t$ is antisymmetric, it follows that

$$
\mathcal{A}(\rho_t, j_t) = \int_A \int_B \frac{(\theta(1+\gamma, 1-\gamma)\chi_t)^2}{16\eta_t(x, y)|A||B|} \eta_e(x, y) dx dy.
$$

Moreover, from the homogeneity and monotonicity of $\theta$, and the fact that $|A| \leq |B|$, we have

$$
\mathcal{A}(\rho_t, j_t) \leq \int_A \int_B \frac{(\theta(1+\gamma, 1-\gamma)\chi_t)^2}{16\eta_t(x, y)|A||B|} \eta_e(x, y) dx dy
$$

and simplifying, we see that

$$
\mathcal{A}(\rho_t, j_t) \leq \int_A \int_B \frac{(\theta(1+\gamma, 1-\gamma)\chi_t)^2}{2^7\eta_t(x, y)|A|^2|B|} \eta_e(x, y) dx dy.
$$

On the other hand, the action computation from the 2-point space in [37] tells us that

$$
\frac{1}{4} \int_0^1 \theta(1 + \gamma_t, 1 - \gamma_t) \chi_t^2 dt < \mathcal{W}_0^2(\delta_0, \delta_1) + \varepsilon_0.
$$

Thus,

$$
\int_0^1 \mathcal{A}(\rho_t, j_t) dt \leq \frac{1}{2^7|A|^2|B|\eta_e(\sup_{x \in A, y \in B} |x - y|)} \int_0^1 \int_A \int_B \theta(1 + \gamma_t, 1 - \gamma_t) \chi_t^2 dxdy dt
$$

@ Springer
\[= \frac{1}{27} |A| \bar{\nu}_{\epsilon}(\sup_{x \in A, y \in B} |x - y|) \int_0^1 \theta(1 + \gamma t, 1 - \gamma t) \chi_t^2 dt \]

\[< \frac{25}{27} |A| \bar{\nu}_{\epsilon}(\sup_{x \in A, y \in B} |x - y|). \]

But \(\epsilon_0\) is arbitrary, so

\[\mathcal{W}_{\eta, \epsilon}(m_A, m_B) = \sqrt{\int_0^1 A(\rho_t, j_t) dt} \leq \frac{\mathcal{W}(\delta_0, \delta_1)}{2^{45}}.\]

We can rephrase this in terms of the constant \(C_\theta\), which is defined by

\[C_\theta := \int_0^1 \sqrt{\theta(1 - r) \theta(1 + r)} dr \]

or, equivalently (see [37]), \(C_\theta\) is \(\sqrt{2}\) times the \(W_\theta\) distance between a Dirac mass and the uniform distribution on the two-point space, in other words, \(C_\theta = \sqrt{2} \mathcal{W}(\delta_0, \delta_1)\); so that

\[\mathcal{W}_{\eta, \epsilon}(m_A, m_B) \leq \frac{C_\theta}{4^{\sqrt{|A| \bar{\nu}_{\epsilon}(\sup_{x \in A, y \in B} |x - y|)}}}.\]

\[\Box\]

B. 1. Lemmas for Sect. 4

Lemma B.2 Let \(\alpha_d := V o l(B(0, 1)) \in \mathbb{R}^d\). Then, \(M_p(\eta) := \int_{\mathbb{R}^d} |y|^p \eta(|y|) dy = d\alpha_d \int_0^\infty r^{d+p-1} \eta(r) dr.\)

Proof Recall that the Hausdorff measure of \(\partial B(0, r)\) is \(\frac{d}{r} V o l(B(0, r))\); in turn, \(V o l(B(0, r)) = \alpha_d r^d\). Now, simply compute that

\[\int_{\mathbb{R}^d} |y|^p \eta(|y|) dy = \int_0^\infty \int_{\partial B(0, r)} r^p \eta(r) d\mathcal{H}^{d-1} dr = d\alpha_d \int_0^\infty r^{p+d-1} \eta(r) dr.\]

\[\Box\]

Lemma B.3 Given any kernel \(\eta(|x - y|)\) satisfying Assumption 2.1 (i–iv), let \(\zeta_{\eta}(x, y) := \int_{|x - y|} s \eta(s) ds.\) Then, for all \(x \in \mathbb{R}^d,\)

\[\int_{\mathbb{R}^d} \zeta_{\eta}(x, y) dy = \frac{M_2(\eta)}{d}\]

and, concerning the rescaled kernel \(\eta_{\epsilon}(|x - y|),\) we have moreover that

\[\int_{\mathbb{R}^d} \zeta_{\eta_{\epsilon}}(x, y) dy = e^2 \frac{M_2(\eta)}{d}.\]

Proof (i) Compute that

\[\int_{\mathbb{R}^d} \zeta_{\eta}(x, y) dy = \int_{\mathbb{R}^d} \left[ \int_{|x - y|} s \eta(s) ds \right] dy = \int_{\mathbb{R}^d} \left[ \int_{|y|} s \eta(s) ds \right] dy = \int_0^\infty \int_{\partial B(0, r)} \left[ \int_r |s| \eta(s) ds \right] d\mathcal{H}^{d-1} dr.\]
Notice that $\int_r^\infty s \eta(s)ds$ is rotationally invariant, and recall that the Hausdorff measure of $\partial B(0, r) = \frac{d}{r} \text{Vol}(B(0, r))$; in turn, $\text{Vol}(B(0, r)) = \alpha_d r^d$. Hence,

$$\int_0^\infty \int_{\partial B(0, r)} \left[ \int_r^\infty s \eta(s)ds \right] r^{d-1} dr = d\alpha_d \int_0^\infty \left[ \int_r^\infty s \eta(s)ds \right] r^{d-1} dr.$$  

In turn, using integration by parts, we observe that

$$\int_0^\infty \left[ \int_r^\infty s \eta(s)ds \right] r^{d-1} dr = \left[ \left[ \int_r^\infty s \eta(s)ds \right] r^d \right]_0^\infty - \int_0^\infty (-r\eta(r)) \frac{r^d}{d} dr$$

$$= \frac{1}{d} \int_0^\infty r^{d+1} \eta(r) dr.$$  

The claim now follows by way of Lemma B.2.  

(ii) By replacing $\eta$ with $\eta_\varepsilon$, the previous part shows that

$$\int_{\mathbb{R}^d} \zeta(\eta_\varepsilon)(x, y) dy = \frac{M_2(\eta_\varepsilon)}{d}.$$  

So we conclude by computing that

$$M_2(\eta_\varepsilon) = \int_{\mathbb{R}^d} |y|^2 \frac{1}{\varepsilon} \eta \left( \frac{|y|}{\varepsilon} \right) dy = \frac{1}{\varepsilon^2} \int_{\mathbb{R}^d} |y|^2 \eta(|y|) dy = \frac{\varepsilon^2 M_2(\eta)}{d}.$$  

\[\square\]

The preceding lemma indicates that $\zeta_\eta$ and $\zeta(\eta_\varepsilon)$ are not convolution kernels (since they are not appropriately normalized). We therefore introduce their normalizations

$$\tilde{\zeta}_\eta := \frac{d}{M_2(\eta)} \zeta_\eta; \quad \tilde{\zeta}(\eta_\varepsilon) := \frac{d}{\varepsilon^2 M_2(\eta)} \zeta(\eta_\varepsilon).$$  

Note that under Assumption 2.1, $\tilde{\zeta}_\eta$ is a convolution kernel supported on the unit ball, while $\tilde{\zeta}(\eta_\varepsilon)$ is a convolution kernel supported on the ball of radius $\varepsilon$.

**Lemma B.4** Assume that $\eta$ satisfies Assumption 2.1. Let $K(x) = c_K e^{-|x|}$, where $c_K$ is a normalizing constant; and let $\tilde{\zeta}(\eta_\varepsilon)(x) = \frac{1}{\varepsilon^d \alpha_d \sigma_\eta} \int_0^\infty t \eta_\varepsilon(t) dt$. Let $\mu \in \mathcal{P}(\mathbb{R}^d)$. Then, for $0 < \varepsilon < \delta$, and $|x - y| < \varepsilon$,

$$K_\delta \ast \mu(x) \left( \frac{1}{1 + \frac{3}{\delta} \varepsilon} \right) \leq \tilde{\zeta}(\eta_\varepsilon) \ast (K_\delta \ast \mu)(x) \leq K_\delta \ast \mu(x) \left( 1 + \frac{3}{\delta} \varepsilon \right)$$  

and

$$\frac{1}{(1 + \frac{3}{\delta} \varepsilon)^2} \frac{1}{1 + \frac{3}{\delta} |x - y|} \leq \frac{\tilde{\zeta}(\eta_\varepsilon) \ast (K_\delta \ast \mu)(y)}{\tilde{\zeta}(\eta_\varepsilon) \ast (K_\delta \ast \mu)(x)} \leq \left( 1 + \frac{3}{\delta} \varepsilon \right)^2 \left( 1 + \frac{3}{\delta} |x - y| \right).$$  

**Proof** We know that for $|x - y| < \delta$,

$$\frac{K_\delta \ast \mu(y)}{K_\delta \ast \mu(x)} \leq 1 + \frac{3}{\delta} |x - y|.$$  

Since (for all $|x - y| < \delta$)

$$K_\delta \ast \mu(x) \left( \frac{1}{1 + \frac{3}{\delta} |x - y|} \right) \leq K_\delta \ast \mu(y) \leq K_\delta \ast \mu(x) \left( 1 + \frac{3}{\delta} |x - y| \right)$$
and $\zeta(\eta \epsilon)$ is a convolution kernel supported on the ball of radius $\epsilon$, and $\epsilon < \delta$, it follows that for all $x \in X$,

$$K_\delta \ast \mu(x) \left( \frac{1}{1 + \frac{3}{\delta} \epsilon} \right) \leq \zeta(\eta \epsilon) \ast (K_\delta \ast \mu)(x) \leq K_\delta \ast \mu(x) \left( 1 + \frac{3}{\delta} \epsilon \right).$$

Therefore, if $|x - y| < \delta$,

$$\frac{K_\delta \ast \mu(y)}{K_\delta \ast \mu(x)} \left( \frac{1}{1 + \frac{3}{\delta} \epsilon} \right) \leq \frac{\zeta(\eta \epsilon) \ast (K_\delta \ast \mu)(y)}{\zeta(\eta \epsilon) \ast (K_\delta \ast \mu)(x)} \leq \frac{K_\delta \ast \mu(y)}{K_\delta \ast \mu(x)} \left( 1 + \frac{3}{\delta} \epsilon \right)$$

and so

$$\frac{1}{(1 + \frac{3}{\delta} \epsilon)^2} \frac{1}{1 + \frac{3}{\delta} |x - y|} \leq \frac{\zeta(\eta \epsilon) \ast (K_\delta \ast \mu)(y)}{\zeta(\eta \epsilon) \ast (K_\delta \ast \mu)(x)} \leq \left( 1 + \frac{3}{\delta} \epsilon \right)^2 \left( 1 + \frac{3}{\delta} |x - y| \right).$$

\[\square\]

**Lemma B.5** (W2 convolution estimates) (i) Let $k$ be any radially symmetric convolution kernel. Then, for any $\mu \in P_2(\mathbb{R}^d)$,

$$W_2(\mu, k_s \ast \mu) \leq (M_2(k))^{1/2} s.$$  

(ii) In the special case of $\zeta(\eta \epsilon)$, under Assumption 2.1 (i–iv) we have that

$$W_2(\mu, \tilde{\zeta}(\eta \epsilon) \ast \mu) \leq \left( \frac{d}{d + 2} \frac{M_4(\eta)}{M_2(\eta)} \right)^{1/2} \epsilon.$$  

**Proof** (i) Consider the coupling defined by $k_s(|x - y|)d\mu(x)d\mu(x)$. Let us first check that this is indeed a coupling between $\mu$ and $k_s \ast \mu$.

To see that the first marginal is $\mu$, simply compute that for any $A \subset \mathbb{R}^d$,

$$\int \int_A k_s(|x - y|)d\mu(x)d\mu(x) = \int_A \int \mathbb{R}^d k_s(|x - y|)dyd\mu(x) = \int_A d\mu(x)$$

since $k_s(|x - y|)$ is normalized. On the other hand, $\int \mathbb{R}^d k_s(|x - y|)d\mu(x)$ is definitionally equal to the density of $k_s \ast \mu$. So the second marginal of $k_s(|x - y|)d\mu(x)d\mu(x)$ is indeed $k_s \ast \mu$. For any $A \subset \mathbb{R}^d$,

$$\int_A \int \mathbb{R}^d k_s(|x - y|)d\mu(x)d\mu(x) = \int_A (k_s \ast \mu)(y)dy = (k_s \ast \mu)(A).$$

Therefore,

$$W_2^2(\mu, k_s \ast \mu) \leq \int \int |x - y|^2 k_s(|x - y|)d\mu(x)d\mu(x).$$

Since for all $x \in \mathbb{R}^d$,

$$\int |x - y|^2 k_s(|x - y|)dy = \int |y|^2 k_s(|y|)dy = s^2 \int |y|^2 k(|y|)dy$$
we find that
\[ W_2^2(\mu, k_x * \mu) \leq \int \left( s^2 \int |y|^2 k(|y|) dy \right) d\mu(x) = s^2 \int |y|^2 k(|y|) dy. \]

(ii) By identical reasoning to part (i), we find that
\[ \bar{\zeta}(\eta \epsilon) (|x - y|) d\mu(x) dy \]
is indeed a coupling between \( \mu \) and \( \bar{\zeta}(\eta \epsilon) * \mu \); and furthermore,
\[ W_2^2(\mu, \bar{\zeta}(\eta \epsilon) * \mu) \leq \int |y|^2 \bar{\zeta}(\eta \epsilon)(|y|) dy. \]

So, observe that
\[ \int |y|^2 \bar{\zeta}(\eta \epsilon)(|y|) dy = \int |y|^2 \left( \frac{d}{\epsilon^2 M_2(\eta)} \int_{|y|}^\infty s \eta(s) ds \right) dy; \]
by performing a similar computation to Lemma B.3, we see that
\[ \int |y|^2 \left( \int_{|y|}^\infty s \eta(s) ds \right) dy = \int |y|^2 \left( \frac{d}{\epsilon^2 M_2(\eta)} \int_{|y|}^\infty s \eta(s) ds \right) dy; \]
(Lemma B.2) = \frac{\epsilon^4}{d + 2} M_4(\eta).

Hence, \( W_2^2(\mu, \bar{\zeta}(\eta \epsilon) * \mu) \leq \frac{d \epsilon^2 M_4(\eta)}{d + 2 M_2(\eta)} \) as desired.

\[ \square \]

**Lemma B.6** (Moments and normalizing constant for the Laplace kernel) Let \( N \in \mathbb{N} \). Concerning the kernel \( K = c_K e^{-|x|} \), we have that \( c_K = (\alpha_d d!)^{-1} \) and
\[ M_N(K) = (N + d - 1)!/(d - 1)! . \]

**Proof** Simply compute that
\[ c_K^{-1} = \int_{\mathbb{R}^d} e^{-|x|} d\nu = \int_0^\infty \int_{\partial B(0,r)} e^{-r} d\mathcal{H}^{d-1} dr = d\alpha_d \int_0^\infty e^{-r} r^{d-1} dr = d\alpha_d (d-1)! \]
and similarly
\[ M_N(K) = c_K \int_{\mathbb{R}^d} |x|^N e^{-|x|} d\nu = c_K d\alpha_d \int_0^\infty e^{-r} r^{N+d-1} dr \]
\[ = c_K d\alpha_d (N + d - 1)!/(d - 1)! . \]
\[ \square \]
B.2. Lemmas for Sect. 5

Lemma B.7 (Hamilton–Jacobi subsolutions closed under convolution in $x$) Let $\phi_t(x) \in BL([0, 1] \times \mathbb{R}^d)$. Suppose that for almost all $(t, x) \in [0, 1] \times \mathbb{R}^d$,
\[
\partial_t \phi_t(x) + H(\nabla \phi_t(x)) \leq 0
\]
where $H : \mathbb{R}^d \to \mathbb{R}_+$ is convex. Then, if $k$ is any convolution kernel, it also holds $(t, x)$-a.e. that
\[
\forall (t, x) \in [0, 1] \times \mathbb{R}^d \quad \partial_t (k \ast \phi_t(x)) + H(\nabla (k \ast \phi_t(x))) \leq 0
\]
where $k \ast \phi_t(x) := \int_{\mathbb{R}^d} k(x - y) \phi(y) dy$.

Proof First, since $\phi_t(x) \in BL([0, 1] \times \mathbb{R}^d)$, we have by Rademacher’s theorem that $\phi_t(x)$ is (jointly) differentiable almost everywhere, and in particular
\[
|\partial_t \phi_t(x)| \leq |\partial_t \phi_t(x) + \nabla \phi_t(x)| \leq \text{Lip}_{[0, 1] \times \mathbb{R}^d} \phi \quad (t, x) - \text{a.e.},
\]
so by [7, Ex. 5.8.135] we are justified in differentiating under the integral sign with respect to the measure $k(x - y)dy$:
\[
\partial_t (k \ast \phi_t(x)) = \partial_t \int_{\mathbb{R}^d} \phi_t(y) k(x - y) dy = \int_{\mathbb{R}^d} \partial_t \phi_t(y) k(x - y) dy = k \ast (\partial_t \phi_t)(x)
\]
$(t, x) - \text{a.e.}$.

Second, using the convexity of $H$, we compute that
\[
H(\nabla (k \ast \phi_t(x))) = H(k \ast \nabla \phi_t(x))
\]
\[
= H\left(\int_{\mathbb{R}^d} \nabla \phi_t(y) k(x - y) dy\right)
\]
(Jensen) \leq \int_{\mathbb{R}^d} H(\nabla \phi_t(y)) k(x - y) dy.
\]
Using the fact that $\partial_t \phi_t(x) + H(\nabla \phi_t(x)) \leq 0$, we see that
\[
\int_{\mathbb{R}^d} H(\nabla \phi_t(y)) k(x - y) dy \leq -\int_{\mathbb{R}^d} (\partial_t \phi_t(y)) k(x - y) dy
\]
(a.e.) = $-\partial_t (k \ast \phi_t(x))$.

Therefore, $H(\nabla (k \ast \phi_t(x))) \leq -\partial_t (k \ast \phi_t(x))$ a.e. as desired. \hfill \Box

Lemma B.8 (Cf. [27, (0.10) “generalized Young’s inequality”]) Let $K = c_K e^{-|x|}$, and let $\phi \in L^\infty(\mathbb{R}^d)$. Then
\[
\| \nabla (K_s \ast \phi) \|_\infty \leq s^{-1} \| \phi \|_\infty.
\]
Moreover, for any $\phi \in BL(\mathbb{R}^d)$,
\[
\| D^2 (K_s \ast \phi) \|_\infty \leq s^{-1} \| \nabla \phi \|_\infty.
\]

Proof Simply compute as follows:
\[
|\nabla (K_s \ast \phi)(x)| \leq \int_{\mathbb{R}^d} |\nabla K_s(x - y) \phi(y)| dy
\]
For each $x$ in $\mathbb{R}^d$, let $\phi(x) := \nabla K\left(\frac{|x|}{s}\right)$. Hence, $\nabla v(x) \leq s^{-1}\|\phi\|_\infty$ for uniformly for fixed $s$. Therefore it remains only to check that $\nabla^2 v(x)$ is Lipschitz in $x$ for uniformly for fixed $t$, and vice versa. By Lemma B.9 (together with Rademacher’s theorem), we know that $\text{Lip}_{[0,1]}(\mathbb{R}^d) \leq \|\phi_t(x)\|_{L^\infty([0,1] \times \mathbb{R}^d)}$. Therefore, $\|k \ast \phi_t(x)\|_{L^\infty([0,1] \times \mathbb{R}^d)} \leq \|\phi_t(x)\|_{L^\infty([0,1] \times \mathbb{R}^d)}$. In order to check that $\text{Lip}_{[0,1]}(\mathbb{R}^d) \leq \|\phi_t(x)\|_{L^\infty([0,1] \times \mathbb{R}^d)}$, it suffices to check that $\phi_t(x)$ is Lipschitz in $x$ for uniformly for fixed $t$, and vice versa. By Lemma B.9 (together with Rademacher’s theorem), we know that $\text{Lip}_{\mathbb{R}^d} k \ast \phi_t(x) \leq \text{Lip}_{\mathbb{R}^d} \phi_t(x)$. Therefore it remains only to check that $k \ast \phi_t(x)$ is Lipschitz in $t$ for each fixed $x$.

So compute that

$$\|k \ast \phi_t(x) - k \ast \phi_t(y)\|_{L^\infty([0,1] \times \mathbb{R}^d)} \leq \sup_{y \in \mathbb{R}^d} \text{Lip}_{[0,1]}(\phi_t(y) \cdot |t_1 - t_2| \leq \text{Lip}_{[0,1]}(\phi_t(y) \cdot |t_1 - t_2|.$$
4. Benamou, J.-D., Brenier, Y.: A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. Numer. Math. **84**, 375–393 (2000)

5. Blandin, S., Goatin, P.: Well-posedness of a conservation law with non-local flux arising in traffic flow modeling. Numer. Math. **132**, 217–241 (2016)

6. Bobkov, S.G., Gentil, I., Ledoux, M.: Hypercontractivity of Hamilton–Jacobi equations. J. de Math. Pures et Appl. **80**, 669–696 (2001)

7. Bogachev, V.I.: Measure Theory, vol. 1. Springer, Berlin (2007)

8. Brenier, Y.: Extended Monge–Kantorovich theory. In: Caffarelli, L.A., Salsa, S. (eds.) Optimal Transportation and Applications: Lectures given at the CIME Summer School held in Martina Franca, Italy, September 2–8, 2003. Springer, pp. 91–121 (2003)

9. Buttazzo, G.: Semicontinuity, relaxation and integral representation in the calculus of variations, Pitman Res. Notes Math. Ser., 207 Longman Scientific & Technical (1989)

10. Chow, S.-N., Huang, W., Li, Y., Zhou, H.: Fokker–Planck equations for a free energy functional or Markov process on a graph. Arch. Ration. Mech. Anal. **203**, 969–1008 (2012)

11. Colombo, M., Crippa, G., Marconi, E., Spinolo, L.V.: Local limit of nonlocal traffic models: convergence results and total variation blow-up. Ann. de l’Institut Henri Poincaré C Anal. Non Linéaire **38**, 1653–1666 (2021)

12. Colombo, M., Crippa, G., Spinolo, L.V.: On the singular local limit for conservation laws with nonlocal fluxes. Arch. Ration. Mech. Anal. **233**, 1131–1167 (2019)

13. Crippa, G., Lécureux-Mercier, M.: Existence and uniqueness of measure solutions for a system of continuity equations with non-local flow. Nonlinear Differ. Equ. Appl. **20**, 523–537 (2013)

14. Dolbeault, J., Nazaret, B., Savaré, G.: A new class of transport distances between measures. Calc. Var. Partial. Differ. Equ. **34**, 193–231 (2009)

15. Du, Q., Gunzburger, M., Lehoucq, R.B., Zhou, K.: Analysis and approximation of nonlocal diffusion problems with volume constraints. SIAM Rev. **54**, 667–696 (2012)

16. Du, Q., Huang, Z., LeFloch, P.G.: Nonlocal conservation laws. A new class of monotonicity-preserving models. SIAM J. Numer. Anal. **55**, 2465–2489 (2017)

17. Du, Q., Kamm, J.R., Lehoucq, R.B., Parks, M.L.: A new approach for a nonlocal, nonlinear conservation law. SIAM J. Appl. Math. **72**, 464–487 (2012)

18. Dubins, L., Freedman, D.: Measurable sets of measures. Pac. J. Math. **14**, 1211–1222 (1964)

19. Erbar, M.: Gradient flows of the entropy for jump processes. Ann. de l’IHPC Probab. et Stat. **50**, 920–945 (2014)

20. Erbar, M., Fathi, M., Laschos, V., Schlichting, A.: Gradient flow structure for McKean–Vlasov equations on discrete spaces. Discrete Contin. Dyn. Syst.-A **36**, 67–99 (2016)

21. Erbar, M., Maas, J.: Ricci curvature of finite Markov chains via convexity of the entropy. Arch. Ration. Mech. Anal. **206**, 997–1038 (2012)

22. Erbar, M., Maas, J.: Gradient flow structures for discrete porous medium equations. Discrete Contin. Dyn. Syst.-A **34**, 13–55 (2014)

23. Erbar, M., Maas, J., Wirth, M.: On the geometry of geodesics in discrete optimal transport. Calc. Var. Partial. Differ. Equ. **58**, 19 (2019)

24. Esposito, A., Gvalani, R.S., Schlichting, A., Schmidthchen, M.: On a novel gradient flow structure for the aggregation equation. arXiv:2112.08317 (2021)

25. Esposito, A., Patacchini, F.S., Schlichting, A., Slepčev, D.: Nonlocal-interaction equation on graphs: gradient flow structure and continuum limit. Arch. Ration. Mech. Anal. **240**, 699–760 (2021)

26. Evans, L.C.: Partial Differential Equations, vol. 19, 2nd edn. American Mathematical Society, Providence (2010)

27. Folland, G.B.: Introduction to Partial Differential Equations. Princeton University Press, Princeton (1995)

28. Gangbo, W., Li, W., Mou, C.: Geodesics of minimal length in the set of probability measures on graphs. ESAIM Control Optim. Calc. Var. **25**, 78 (2019)

29. Garcia Trillos, N.: Gromov–Hausdorff limit of Wasserstein spaces on point clouds. Calc. Var. Partial. Differ. Equ. **59**, 1–43 (2020)

30. Gibbs, A.L., Su, F.E.: On choosing and bounding probability metrics. Int. Stat. Rev. **70**, 419–435 (2002)

31. Gigli, N., Maas, J.: Gromov–Hausdorff convergence of discrete transportation metrics. SIAM J. Math. Anal. **45**, 879–899 (2013)

32. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. Multiscale Model. Simul. **7**, 1005–1028 (2009)

33. Glimm, P., Kopfer, E., Maas, J.: Scaling limits of discrete optimal transport. SIAM J. Math. Anal. **52**, 2759–2802 (2020)

34. Kallenberg, O.: Random Measures, Theory and Applications, vol. 77. Springer, Berlin (2017)

35. Lee, J.M.: Smooth Manifolds. Springer, Berlin (2012)
36. Levin, D.A., Peres, Y.: Markov Chains and Mixing Times, vol. 107. American Mathematical Society, Providence (2017)
37. Maas, J.: Gradient flows of the entropy for finite Markov chains. J. Funct. Anal. 261, 2250–2292 (2011)
38. Mielke, A.: A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. Nonlinearity 24, 1329 (2011)
39. Otto, F., Villani, C.: Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. J. Funct. Anal. 173, 361–400 (2000)
40. Peletier, M.A., Rossi, R., Savaré, G., Tse, O.: Jump processes as generalized gradient flows. Calc. Var. Partial. Differ. Equ. 61, 1–85 (2022)
41. Perlman, M.D.: Jensen’s inequality for a convex vector-valued function on an infinite-dimensional space. J. Multivar. Anal. 4, 52–65 (1974)
42. Reshetnyak, Y.: General theorems on semicontinuity and on convergence with a functional. Sib. Math. J. 8, 801–816 (1967)
43. Santambrogio, F.: Optimal Transport for Applied Mathematicians. Springer, Berlin (2015)
44. Spector, D.: Simple proofs of some results of Reshetnyak. In: Proceedings of the American Mathematical Society, pp. 1681–1690 (2011)
45. Terán, P.: Jensen’s inequality for random elements in metric spaces and some applications. J. Math. Anal. Appl. 414, 756–766 (2014)
46. Villani, C.: Topics in Optimal Transportation, vol. 58. American Mathematical Society, Providence (2003)
47. Villani, C.: Optimal Transport: Old and New, vol. 338. Springer, Berlin (2008)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.