THE EXPONENTIAL FORMULA FOR THE WASSERSTEIN METRIC *

Katy Craig[1]

Abstract. A recurring obstacle in the study of Wasserstein gradient flow is the lack of convexity of the square Wasserstein metric. In this paper, we develop a class of *transport metrics* that have better convexity properties and use these metrics to prove an Euler−Lagrange equation characterizing Wasserstein discrete gradient flow. We then apply these results to give a new proof of the exponential formula for the Wasserstein metric, mirroring Crandall and Liggett's proof of the corresponding Banach space result [M.G. Crandall and T.M. Liggett, *Amer. J. Math.* **93** (1971) 265–298]. We conclude by using our approach to give simple proofs of properties of the gradient flow, including the contracting semigroup property and energy dissipation inequality.

1. Introduction

Given a continuously differentiable, convex function $E : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, the *gradient flow* of E is the solution to the Cauchy problem

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t) = -\nabla E(u(t)), \quad u(0) \in \overline{D(E)} = \overline{\{v \in \mathbb{R}^d : E(v) < +\infty\}}. \tag{1.1}$$

Through suitable generalizations of the notion of the gradient, the theory of gradient flow has been extensively studied in Hilbert spaces [2], Banach spaces [8,9], nonpositively curved metric spaces [12], and general metric spaces [1,7], including the space of probability measures endowed with the Wasserstein metric W_2.

Gradient flow in the Wasserstein metric is of particular interest due to the sharp estimates it can provide on long-time behavior of solutions to partial differential equations [15] and the low regularity it requires, allowing one to pass seamlessly between Lagrangian and Eulerian perspectives [5,6]. For a sufficiently regular functional E and measure μ, the formal Wasserstein gradient is given by $\nabla_{W_2} E(\mu) = -\nabla \cdot (\mu \nabla \frac{\delta E}{\delta \mu})$ and the gradient flow of E corresponds to the partial differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu(t) = -\nabla_{W_2} E(\mu) = \nabla \cdot \left(\mu \nabla \frac{\delta E}{\delta \mu} \right), \tag{1.2}$$

in the duality with $C_c^\infty(\mathbb{R}^d \times (0, +\infty))$ with initial conditions $\lim_{t \to 0} \mu(t) = \mu_0$ ([1], Lem. 10.4.1, Thm. 11.1.4).

Due to the formal nature of the Wasserstein gradient, rigorous study of Wasserstein gradient flow often considers a time discretization of the problem that does not require a rigorous notion of gradient. For Euclidean gradient flow (1.1), this discretization is simply the implicit Euler method, and it is a classical result that u_n is a finite difference approximation to the gradient flow if and only if u_n solves a sequence of minimization problems,

$$(u_n - u_{n-1})/\tau = -\nabla E(u_n), u_0 = u \quad \Longleftrightarrow \quad u_n = \operatorname*{argmin}_v \{|v - u_{n-1}|^2/2\tau + E(v)\}, u_0 = u. \qquad (1.3)$$

Motivated by this, the Wasserstein *proximal map* J_τ is defined by

$$J_\tau \mu = \operatorname*{argmin}_\nu \left\{ W_2^2(\nu, \mu)/2\tau + E(\nu) \right\},$$

and the *discrete gradient flow* is defined by $J_\tau^n \mu$. Taking $\tau = \frac{t}{n}$ and sending $n \to \infty$ gives the *exponential formula* relating the discrete gradient flow to the gradient flow which solves (1.2),

$$\lim_{n \to \infty} J_{t/n}^n \mu = \mu(t).$$

Ambrosio *et al.* were the first to prove the exponential formula for the Wasserstein metric, obtaining the sharp rate of convergence of the discrete gradient flow $J_{t/n}^n \mu$ to the gradient flow $\mu(t)$ ([1], Thm. 4.0.4). In the same work, they raised the question of whether another proof might be possible, mirroring Crandall and Liggett's approach for the Banach space case [9].

A recurring difficulty in the theory of Wasserstein gradient flow is that, while $x \mapsto \frac{1}{2}|x - y|^2$ is 1-convex along geodesics in Euclidean space, the square Wasserstein metric $\mu \mapsto \frac{1}{2}W_2^2(\mu, \omega)$ is not convex along geodesics ([1], Ex. 9.1.5). In fact, it is semi-concave ([1], Thm. 7.3.2). Convexity of the square Euclidean distance ensures that $v \mapsto |v - u_{n-1}|^2/2\tau + E(v)$ is convex, so the minimization problem in (1.3) is well-posed. This argument fails in the Wasserstein case. Convexity of the square distance is also essential to concluding that the proximal map satisfies the contraction inequality $\|J_\tau u - J_\tau v\| \le \|u - v\|$, a key element in Crandall and Liggett's proof of the exponential formula [9, 12]. It is unknown if such a contraction holds in the Wasserstein case, though "almost" contraction inequalities exist, such as Ambrosio, Gigli, and Savaré's ([1], Lem. 4.2.4) and Carlen and the author's ([4], Thm. 1.3).

In order to circumvent these difficulties, Ambrosio *et al.* introduced a new class of curves − *generalized geodesics* − along which the square distance is 1-convex ([1], Lem. 9.2.1). They also introduced *pseudo-Wasserstein metrics*, which they used to study the semi-concavity and differentiability of the square Wasserstein metric ([1], Eq. 7.3.2). In this paper, we further develop these ideas, introducing a class of *transport metrics*, which are a type of pseudo-Wasserstein metric. We show that generalized geodesics, while not geodesics with respect to the Wasserstein metric, are geodesics with respect to the transport metrics. The transport metrics also satisfy the key property that the square distance is 1-convex. We use the transport metrics to prove an Euler−Lagrange equation characterizing the discrete gradient flow, analogous to equation (1.3) above. (One direction of this characterization is due to Ambrosio *et al.* [1]. We prove the converse.) We then apply the transport metrics and Euler−Lagrange equation to give a new proof of the exponential formula, in the style of Crandall and Liggett. We conclude by applying our estimates to give simple proofs of properties of the gradient flow, including the contracting semigroup property and energy dissipation inequality.

We are optimistic that our results will be useful in future study of Wasserstein gradient flow. Our Euler−Lagrange equation replaces the global minimization problem in the definition of the proximal map with a local computation of the subdifferential. Our results on the structure of transport metrics give further credence to the geometric naturalness of the assumption of convexity along generalized geodesics, a key element in the work of Ambrosio *et al.* Our new proof of the exponential formula brings together the theories of Banach space and Wasserstein gradient flow, and we believe our work may be useful in studying the behavior of the gradient flow as the functional E varies − for example, as a regularization of E is removed. The corresponding problem in the Banach space case is well-understood [3], and the analogy we establish between the Banach and Wasserstein cases may help extend these results.

2. Transport metrics and the Euler−Lagrange equation

In Sections 2.1−2.3 we recall foundational results on Wasserstein gradient flow, including Ambrosio, Gigli, and Savaré's notion of generalized geodesics and pseudo-Wasserstein metrics. In Section 2.4 we build on this theory, developing the structure of transport metrics, which have better convexity properties than the Wasserstein metric. In Section 2.5, we use the convexity of the square transport metrics to prove an Euler−Lagrange equation characterizing the discrete gradient flow.

2.1. Wasserstein metric

We begin by recalling properties of the Wasserstein metric. Let $\mathcal{P}(\mathbb{R}^d)$ denote the set of probability measures on $\mathbb{R}^d$. Given $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, a measurable function $\mathbf{t} : \mathbb{R}^d \to \mathbb{R}^d$ *transports μ onto ν* if $\nu(B) = \mu(\mathbf{t}^{-1}(B))$ for all measureable $B \subseteq \mathbb{R}^d$. We call ν the *push-forward of μ under $\mathbf{t}$* and write $\nu = \mathbf{t}\#\mu$.

For a finite product $\Pi_{i=1}^n \mathbb{R}^d$, let π^i the be projection onto the ith component and $\pi^{i,j}$ be the projection onto the ith and jth components. Given $\mu, \nu \in P(\mathbb{R}^d)$, the set of *transport plans* from μ to ν is

$$\Gamma(\mu, \nu) = \{\boldsymbol{\mu} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \pi^1\#\boldsymbol{\mu} = \mu, \pi^2\#\boldsymbol{\mu} = \nu\}.$$

The *Wasserstein distance* between μ and ν is defined to be

$$W_2(\mu, \nu) = \inf \left\{ \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x_1 - x_2|^2 \mathrm{d}\boldsymbol{\mu}(x_1, x_2) \right)^{1/2} : \boldsymbol{\mu} \in \Gamma(\mu, \nu) \right\}. \tag{2.1}$$

If $W_2(\mu, \nu) < +\infty$, the infimum is attained, and we denote the set of *optimal transport plans* by $\Gamma_0(\mu, \nu)$.

If μ does not charge sets of $d - 1$ dimensional Hausdorff measure, we say that μ *does not charge small sets*. In this case, there is a unique optimal transport plan from μ to ν of the form $(\mathbf{id} \times \mathbf{t})\#\mu$, where $\mathbf{t}$ is unique μ-a.e. and $\mathbf{id}(x) = x$ is the identity transformation [13]. In particular, there is an *optimal transport map* $\mathbf{t} = \mathbf{t}_\mu^\nu$ satisfying $\mathbf{t}\#\mu = \nu$ and $W_2(\mu, \nu) = \|\mathbf{id} - \mathbf{t}\|_{L^2(\mu)}$. (See Gigli [10] for a sharp version of this result.) Furthermore, a measurable map satisfying $\mathbf{t}\#\mu = \nu$ is optimal if and only if it is cyclically monotone μ-a.e. [13]. If ν also does not charge small sets, then $\mathbf{t}_\mu^\nu \circ \mathbf{t}_\nu^\mu = \mathbf{id}$ almost everywhere with respect to μ.

One technical difficulty when working with the Wasserstein distance on $\mathcal{P}(\mathbb{R}^d)$ is that there exist measures that are infinite distances apart. Given a fixed reference measure ω_0, define

$$\mathcal{P}_{2,\omega_0}(\mathbb{R}^d) = \{\mu \in \mathcal{P}(\mathbb{R}^d) : W_2(\mu, \omega_0) < +\infty\}.$$

By the triangle inequality, $(\mathcal{P}_{2,\omega_0}(\mathbb{R}^d), W_2)$ is a metric space. When $\omega_0 = \delta_0$, the Dirac mass at the origin, $\mathcal{P}_{2,\omega_0}(\mathbb{R}^d) = \mathcal{P}_2(\mathbb{R}^d)$, the subset of $\mathcal{P}(\mathbb{R}^d)$ with finite second moment.

We consider three classes of curves: locally absolutely continuous curves, (constant speed) geodesics, and generalized geodesics. We define the first two now and leave the third for the next section.

Definition 2.1. $\mu(t) : \mathbb{R} \to \mathcal{P}_{2,\omega_0}(\mathbb{R}^d)$ is *locally absolutely continuous* if for all $I \subseteq \mathbb{R}$ bounded, there exists $m \in L^1(I)$ so that $W_2(\mu(t), \mu(s)) \leq \int_s^t m(r)\mathrm{d}r$ for all $s \leq t \in I$.

Definition 2.2. Given a metric space (X, d), $u_\alpha : [0, 1] \to X$ is a (constant speed) *geodesic* in case $d(u_\alpha, u_\beta) = |\beta - \alpha| d(u_0, u_1)$ for all $\alpha, \beta \in [0, 1]$.

As any two measures are connected by a geodesic, $(\mathcal{P}_{2,\omega_0}(\mathbb{R}^d), W_2)$ is a geodesic metric space, and all geodesics are of the form $\mu_\alpha = \left((1 - \alpha)\pi^1 + \alpha\pi^2\right)\#\boldsymbol{\mu}$ for $\boldsymbol{\mu} \in \Gamma_0(\mu_0, \mu_1)$ ([1], Thm. 7.2.2). If μ_0 does not charge small sets, the geodesic from μ_0 to μ_1 is unique and given by $\mu_\alpha = \left((1 - \alpha)\mathbf{id} + \alpha\mathbf{t}_{\mu_0}^{\mu_1}\right)\#\mu_0$.

Given a metric space (X, d), we place the following conditions on functionals $E : X \to \mathbb{R} \cup \{+\infty\}$:

- *proper*: $D(E) = \{u \in X : E(u) < +\infty\} \neq \emptyset$

- *coercive*: There exists $\tau_0 > 0$, $u_0 \in X$ so that $\inf\left\{\frac{1}{2\tau_0}d^2(u_0, v) + E(v) : v \in X\right\} > -\infty$.
- *lower semicontinuous:* For all $u_n, u \in X$ so that $u_n \to u$, $\liminf_{n\to\infty} E(u_n) \geq E(u)$.
- λ-*convex along a curve u_α:* Given $\lambda \in \mathbb{R}$ and a curve $u_\alpha \in X$,

$$E(u_\alpha) \leq (1 - \alpha)E(u_0) + \alpha E(u_1) - \alpha(1 - \alpha)\frac{\lambda}{2}d(u_0, u_1)^2. \tag{2.2}$$

- λ-*convex along geodesics*: Given $\lambda \in \mathbb{R}$, for all $u_0, u_1 \in X$, there exists a geodesic u_α from u_0 and u_1 along which (2.2) holds. We will often simply say that E is λ-*convex*, or in the case $\lambda = 0$, *convex*.

If $(X, d) = (\mathcal{P}_{2,\omega_0}(\mathbb{R}^d), W_2)$, we consider the following notions of differentiability.

Definition 2.3. Given $E : \mathcal{P}_{2,\omega_0}(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$, the *metric slope* of E at $u \in D(E)$ is

$$|\partial E|(u) = \limsup_{v \to u} \frac{(E(u) - E(v))^+}{d(u, v)}.$$

For ease of notation, we suppose E satisfies the following assumption, which ensures that for all $\mu \in D(E)$ and $\nu \in \mathcal{P}(\mathbb{R}^d)$ there exists an optimal transport map $\mathbf{t}_\mu^\nu$ from μ to ν.

Assumption 2.4. If $\mu \in D(E)$, μ does not give mass to small sets.

Definition 2.5. Suppose $E : \mathcal{P}_{2,\omega_0}(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ satisfies domain Assumption 2.4 and is proper, coercive, lower semicontinuous, and λ-convex along geodesics. Then $\boldsymbol{\xi} \in L^2(\mu)$ belongs to the *subdifferential* of E at μ in case for all $\nu \in D(E)$,

$$E(\nu) - E(\mu) \geq \int_{\mathbb{R}^d} \langle \boldsymbol{\xi}, \mathbf{t}_\mu^\nu - \mathbf{id}\rangle \mathrm{d}\mu + \frac{\lambda}{2}W_2^2(\mu, \nu).$$

We denote this by $\boldsymbol{\xi} \in \partial E(\mu)$. In addition, $\boldsymbol{\xi}$ is a *strong subdifferential* if for all measurable $\mathbf{t}$ it satisfies

$$E(\mathbf{t}\#\mu) - E(\mu) \geq \int_{\mathbb{R}^d} \langle \boldsymbol{\xi}, \mathbf{t} - \mathbf{id}\rangle \mathrm{d}\mu + o(\|\mathbf{t} - \mathbf{id}\|_{L^2(\mu)}^2) \quad \text{as } t \xrightarrow{L^2(\mu)} \mathbf{id}.$$

2.2. Generalized geodesics

A recurring difficulty in extending results from a Hilbert space $(\mathcal{H}, \|\cdot\|)$ to the Wasserstein metric $(\mathcal{P}_{2,\omega_0}, W_2)$ is that while $y \mapsto \frac{1}{2}\|x - y\|^2$ is 1-convex along geodesics, $\mu \mapsto \frac{1}{2}W_2^2(\omega, \mu)$ is not ([1], Ex. 9.1.5). Ambrosio *et al.* circumvent this difficulty by introducing generalized geodesics ([1], Def. 9.2.2).

Definition 2.6. Given $\mu_0, \mu_1, \omega_1 \in \mathcal{P}_{2,\omega_0}(\mathbb{R}^d)$, a measure $\boldsymbol{\omega} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ is a *transport plan from μ_0 to μ_1 with base ω_1* in case $\pi^{1,2}\#\boldsymbol{\omega} \in \Gamma_0(\omega_1, \mu_0)$ and $\pi^{1,3}\#\boldsymbol{\omega} \in \Gamma_0(\omega_1, \mu_1)$.

Definition 2.7. Given $\mu_0, \mu_1, \omega_1 \in \mathcal{P}_{2,\omega_0}(\mathbb{R}^d)$ and $\boldsymbol{\omega} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ a transport plan from μ_0 to μ_1 with base ω_1, a *generalized geodesic from μ_0 to μ_1 with base ω_1* is the curve

$$\mu_\alpha = \left((1 - \alpha)\pi^2 + \alpha\pi^3\right)\#\boldsymbol{\omega}, \quad \alpha \in [0, 1].$$

For any three measures, a transport plan from μ_0 to μ_1 with base ω_1 always exists ([1], Lem. 5.3.2), hence generalized geodesics always exist. If ω_1 is absolutely continuous with respect to Lebesgue measure, the generalized geodesic is unique and defined by $\mu_\alpha = \left((1 - \alpha)\mathbf{t}_{\omega_1}^{\mu_0} + \alpha\mathbf{t}_{\omega_1}^{\mu_1}\right)\#\omega_1$. Typically, a generalized geodesic is *not* a geodesic. However, it is if the base ω_1 coincides with either μ_0 or μ_1.

In addition to the notion of generalized geodesics, Ambrosio *et al.* introduced the related notion of pseudo-Wasserstein metrics ([1], Eq. 9.2.5).

Definition 2.8. Given a transport plan $\boldsymbol{\omega}$ from μ_0 to μ_1 with base ω_1, the *pseudo-Wasserstein metric* is

$$W_{2,\boldsymbol{\omega}}(\mu_0, \mu_1) = \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x_2 - x_3|^2 \mathrm{d}\boldsymbol{\omega}(x_1, x_2, x_3) \right)^{1/2}.$$

Remark 2.9. If $\omega_1 = \mu_0$ or μ_1, this reduces to the Wasserstein metric. In general, $W_{2,\boldsymbol{\omega}}(\mu_0, \mu_1) \geq W_2(\mu_0, \mu_1)$. We also have $W_{2,\boldsymbol{\omega}}(\mu_0, \mu_1) \leq W_2(\mu_0, \omega_1) + W_2(\omega_1, \mu_1)$ by the triangle inequality for $L^2(\boldsymbol{\omega})$.

Let $\boldsymbol{\omega}$ be a transport plan from μ_0 to μ_1 with base ω_1. If μ_α is the generalized geodesic induced by $\boldsymbol{\omega}$ and $W_{2,\boldsymbol{\omega}}$ is the corresponding pseudo-Wasserstein metric, Ambrosio *et al.* showed

$$W_2^2(\omega_1, \mu_\alpha) = (1-\alpha)W_2^2(\omega_1, \mu_0) + \alpha W_2^2(\omega_1, \mu_1) - \alpha(1-\alpha)W_{2,\boldsymbol{\omega}}^2(\mu_0, \mu_1), \quad \forall \alpha \in [0,1]. \tag{2.3}$$

In particular, while $\mu \mapsto \frac{1}{2}W_2^2(\omega_1, \mu)$ is not convex along geodesics, it is 1-convex along *generalized geodesics* with base ω_1 ([1], Lem. 9.2.1). This convexity is a key element in their study of discrete gradient flow.

2.3. Gradient flow and discrete gradient flow

For $E : \mathcal{P}_{2,\omega_0}(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ and $\tau > 0$, the *quadratic perturbation* Φ and *proximal map* J_τ are defined by

$$\Phi(\tau, \mu; \nu) = \frac{1}{2\tau}W_2^2(\mu, \nu) + E(\nu), \quad J_\tau(\mu) = \underset{\nu \in \mathcal{P}_{2,\omega_0}(\mathbb{R}^d)}{\operatorname{argmin}} \Phi(\tau, \mu; \nu). \tag{2.4}$$

Let $J_0(\mu) = \mu$. The *discrete gradient flow* sequence of E is constructed *via* repeated applications of J_τ,

$$\mu_n = J_\tau(\mu_{n-1}) = J_\tau^n(\mu_0), \quad \mu_0 \in \overline{D(E)}.$$

In order to ensure the proximal map is well-defined, Ambrosio *et al.* require that the quadratic perturbation $\Phi(\tau, \mu; \cdot)$ be *convex along generalized geodesics* with base μ ([1], Def. 9.2.2).

Definition 2.10. Given $\lambda \in \mathbb{R}$, a functional $E : \mathcal{P}_{2,\omega_0}(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ is λ-*convex along a generalized geodesic* μ_α induced by $\boldsymbol{\omega}$, in case

$$E(\mu_\alpha) \leq (1-\alpha)E(\mu_0) + \alpha E(\mu_1) - \alpha(1-\alpha)\frac{\lambda}{2}W_{2,\boldsymbol{\omega}}^2(\mu_0, \mu_1), \quad \forall \alpha \in [0,1]. \tag{2.5}$$

E is *convex along generalized geodesics* if, for all $\mu_0, \mu_1, \omega \in \mathcal{P}_{2,\omega_0}(\mathbb{R}^d)$, there exists a generalized geodesic from μ_0 to μ_1 with base ω along which E is convex.

By equation (2.3), to ensure $\Phi(\tau, \mu; \cdot)$ is convex along generalized geodesics with base μ, it is enough for E to be λ-convex along generalized geodesics for $0 < \tau < \frac{1}{\lambda^-}$ (where $\lambda^- = \max\{0, -\lambda\}$). Note that if E is λ-convex along generalized geodesics, it is also λ-convex along geodesics. Going forward, we assume the following:

Assumption 2.11. E is proper, coercive, lower semicontinuous, and λ-convex along generalized geodesics.

With this, the proximal map $J_\tau : \overline{D(E)} \to D(E) : \mu \mapsto \mu_\tau$ is well-defined and continuous ([1], Thm. 4.1.2). We now define the Wasserstein gradient flow of a functional E.

Definition 2.12. Suppose $E : \mathcal{P}_{2,\omega_0}(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ is proper, coercive, lower semicontinuous, and λ-convex along generalized geodesics for $\lambda \in \mathbb{R}$. A locally absolutely continuous curve $\mu : (0, +\infty) \to \mathcal{P}_{2,\omega_0}(\mathbb{R}^d)$ is the *gradient flow* of E with initial data $\mu_0 \in \overline{D(E)}$ if $\mu(t) \xrightarrow{t \to 0} \mu_0$ and

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}W_2^2(\mu(t), \omega) + \frac{\lambda}{2}W_2^2(\mu(t), \omega) \leq E(\omega) - E(\mu(t)), \quad \forall \omega \in D(E), \text{ a.e. } t > 0. \tag{2.6}$$

We close this section by recalling two inequalities for the discrete gradient flow that are consequences of the convexity of $\Phi(\tau, \mu; \cdot)$ along generalized geodesics with base μ ([1], Thms. 3.1.6 and 4.1.2). Suppose E is proper, coercive, lower semicontinuous, and λ-convex along generalized geodesics with $0 < \tau < 1/\lambda^-$. Then for $\mu \in D(|\partial E|)$,

$$\tau^2 |\partial E|^2 (J_\tau \mu) \leq W_2^2(\mu, J_\tau \mu) \leq \frac{2\tau}{1 + \lambda\tau} \left[E(\mu) - E(J_\tau \mu) - \frac{1}{2\tau} W_2^2(\mu, J_\tau \mu) \right] \leq \frac{\tau^2}{(1 + \lambda\tau)^2} |\partial E|^2(\mu). \qquad (2.7)$$

For $\mu \in \overline{D(E)}$ and $\nu \in D(E)$,

$$\frac{1}{2\tau} [W_2^2(J_\tau \mu, \nu) - W_2^2(\mu, \nu)] + \frac{\lambda}{2} W_2^2(J_\tau \mu, \nu) \leq E(\nu) - E(J_\tau \mu) - \frac{1}{2\tau} W_2^2(\mu, J_\tau \mu). \qquad (2.8)$$

2.4. Transport metrics

We now consider further properties of the pseudo-Wasserstein metric in the particular case that the base measure does not give mass to small sets. In contrast to the previous sections, in which we reviewed existing results, the results in the current section are new and play a key role in our proof of the Euler−Lagrange equation characterizing the proximal map and the exponential formula.

First, we show the following generalization of (2.3).

Proposition 2.13. *Fix* $\mu_0, \mu_1, \omega_1 \in \mathcal{P}_{2,\omega_0}(\mathbb{R}^d)$. *Let* $\boldsymbol{\omega}$ *be a transport plan from* μ_0 *to* μ_1 *with base* ω_1 *and let* μ_α *be the generalized geodesic induced by* $\boldsymbol{\omega}$. *Then, for all* $\nu \in P_{2,\omega_0}(\mathbb{R}^d)$ *there exists a transport plan* $\tilde{\boldsymbol{\omega}}$ *from* ν *to* μ_α *with base* ω_1 *so that*

$$W_{2,\tilde{\boldsymbol{\omega}}}^2(\nu, \mu_\alpha) = (1 - \alpha) W_{2,\boldsymbol{\omega}}^2(\nu, \mu_0) + \alpha W_{2,\boldsymbol{\omega}}^2(\nu, \mu_1) - \alpha(1 - \alpha) W_{2,\boldsymbol{\omega}}^2(\mu_0, \mu_1), \quad \forall \alpha \in [0, 1].$$

Proof. Fix $\boldsymbol{\mu} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ so that $\pi^{1,2} \# \boldsymbol{\mu} = \pi^{1,2} \# \boldsymbol{\omega} \in \Gamma_0(\omega_1, \mu_0)$, $\pi^{1,3} \# \boldsymbol{\mu} = \pi^{1,3} \# \boldsymbol{\omega} \in \Gamma_0(\omega_1, \mu_1)$, and $\pi^{1,4} \# \boldsymbol{\mu} \in \Gamma_0(\omega_1, \nu)$ ([1], Lem. 5.3.4). Define $\tilde{\boldsymbol{\omega}} = (\pi^1 \times \pi^4 \times [(1 - \alpha)\pi^2 + \alpha\pi^3]) \# \boldsymbol{\mu}$. Then $\tilde{\boldsymbol{\omega}}$ is a transport plan from ν to μ_α with base ω_1 and, by the corresponding identity for $L^2(\boldsymbol{\mu})$,

$$W_{2,\tilde{\boldsymbol{\omega}}}^2(\nu, \mu_\alpha) = \|x_4 - (1 - \alpha)x_2 + \alpha x_3\|_{L^2(\boldsymbol{\mu})} = (1 - \alpha) W_{2,\boldsymbol{\omega}}^2(\nu, \mu_0) + \alpha W_{2,\boldsymbol{\omega}}^2(\nu, \mu_1) - \alpha(1 - \alpha) W_{2,\boldsymbol{\omega}}^2(\mu_0, \mu_1). \quad \square$$

Now we consider the pseudo-Wasserstein metrics in the particular case that the base ω does not charge small sets. (We drop the subscript on ω to ease notation.) In this case, it the pseudo-Wasserstein metric becomes a true metric, and to emphasize this point, we call it the $(2, \omega)$-transport metric.

Definition 2.14. *The* $(2, \omega)$-*transport metric* $W_{2,\omega} : \mathcal{P}_{2,\omega}(\mathbb{R}^d) \times \mathcal{P}_{2,\omega}(\mathbb{R}^d) \to \mathbb{R}$ *is given by*

$$W_{2,\omega}(\mu_0, \mu_1) = \left(\int |\mathbf{t}_\omega^{\mu_0} - \mathbf{t}_\omega^{\mu_1}|^2 \mathrm{d}\omega \right)^{1/2}.$$

We show that the generalized geodesics with base ω are exactly the constant speed geodesics for this metric. This allows us to consider functionals which are convex with respect to the transport metric and define a notion of subdifferential with respect to this metric. The map $\mu_0 \mapsto \mathbf{t}_\omega^{\mu_0}$ is a geodesic preserving isometry from $\mathcal{P}_{2,\omega}(\mathbb{R}^d)$ to $L^2(\omega)$, so the square transport metric is 1-convex with respect to its own geodesic structure. (The 1-convexity of the square transport metric can also be seen as a special case of Prop. 2.13).

Proposition 2.15.

(i) $W_{2,\omega}$ *is a metric on* $\mathcal{P}_{2,\omega}(\mathbb{R}^d)$.
(ii) *The constant speed geodesics with respect to* $W_{2,\omega}$ *are the generalized geodesics with base* ω.
(iii) $\mu \mapsto \mathbf{t}_\omega^\mu$ *is a geodesic preserving isometry from* $\mathcal{P}_{2,\omega}(\mathbb{R}^d)$ *to* $L^2(\omega)$, *hence for all* $\nu \in P_{2,\omega}(\mathbb{R}^d)$,

$$W_{2,\omega}^2(\nu, \mu_\alpha) = (1 - \alpha) W_{2,\omega}^2(\nu, \mu_0) + \alpha W_{2,\omega}^2(\nu, \mu_1) - \alpha(1 - \alpha) W_{2,\omega}^2(\mu_0, \mu_1) \quad \forall \alpha \in [0, 1]. \qquad (2.9)$$

Proof. First, we show (i). $W_{2,\omega}$ is symmetric and nonnegative by definition. It is non-degenerate since $W_{2,\omega}(\mu, \nu) \geq W_2(\mu, \nu)$. It satisfies the triangle inequality since $L^2(\omega)$ satisfies the triangle inequality.

Next, we show that generalized geodesics with base ω are constant speed geodesics in $W_{2,\omega}$. Let $\mu_\alpha = ((1-\alpha)\mathbf{t}_\omega^{\mu_0} + \alpha\mathbf{t}_\omega^{\mu_1})\#\omega$ be the generalized geodesic with base ω from μ_0 to μ_1. Since $(1-\alpha)\mathbf{t}_\omega^{\mu_0} + \alpha\mathbf{t}_\omega^{\mu_1}$ is a convex combination of cyclically monotone maps, $\mathbf{t}_\omega^{\mu_\alpha} = (1-\alpha)\mathbf{t}_\omega^{\mu_0} + \alpha\mathbf{t}_\omega^{\mu_1}$, hence

$$W_{2,\omega}(\mu_\alpha^{\mu\to\nu}, \mu_\beta^{\mu\to\nu}) = \|[(1-\alpha)\mathbf{t}_\omega^\mu + \alpha\mathbf{t}_\omega^\nu] - [(1-\beta)\mathbf{t}_\omega^\mu + \beta\mathbf{t}_\omega^\nu]\|_{L^2(\omega)} = |\beta - \alpha|W_{2,\omega}(\mu, \nu).$$

Since the isometry $\mu \mapsto \mathbf{t}_\omega^\mu$ sends $\mu_\alpha \mapsto t_\omega^{\mu_\alpha} = (1-\alpha)\mathbf{t}_\omega^{\mu_0} + \alpha\mathbf{t}_\omega^{\mu_1}$, equation (2.9) holds for μ_α by the corresponding identity for $L^2(\omega)$. It remains to show that μ_α is the unique geodesic from μ_0 to μ_1. Suppose $\tilde{\mu}_\alpha$ is another. Setting $\nu = \tilde{\mu}_\alpha$ in equation (2.9) gives

$$\begin{aligned}
W_{2,\omega}^2(\tilde{\mu}_\alpha, \mu_\alpha) &= (1-\alpha)W_{2,\omega}^2(\tilde{\mu}_\alpha, \mu_0) + \alpha W_{2,\omega}^2(\tilde{\mu}_\alpha, \mu_1) - \alpha(1-\alpha)W_{2,\omega}^2(\mu_0, \mu_1) \\
&= (1-\alpha)\alpha^2 W_{2,\omega}^2(\mu_0, \mu_1) + \alpha(1-\alpha)^2 W_{2,\omega}^2(\mu_0, \mu_1) - \alpha(1-\alpha)W_{2,\omega}^2(\mu_0, \mu_1) = 0. \qquad \square
\end{aligned}$$

Remark 2.16 (Convexity)**.** By (ii), if E is λ-convex along generalized geodesics with base ω, it is λ-convex in the $W_{2,\omega}$ metric. By (iii), the function $\mu \mapsto W_{2,\omega}^2(\nu, \mu)$ is 2-convex in the $W_{2,\omega}$ metric for any $\nu \in \mathcal{P}_{2,\omega}(\mathbb{R}^d)$. Note the difference between (2.3), which ensures $\mu \mapsto W_2^2(\omega, \mu)$ is convex along generalized geodesics with base ω, and (2.9), which ensures $\mu \mapsto W_{2,\omega}^2(\nu, \mu)$ is convex along generalized geodesics with base ω for all $\nu \in P_{2,\omega}(\mathbb{R}^d)$.

Remark 2.17 (Lower semicontinuity)**.** By Remark 2.9, if μ_n converges to μ in $W_{2,\omega}$, then the sequence converges in W_2. Therefore, if E is lower semicontinuous in W_2, E is lower semicontinuous in $W_{2,\omega}$.

Using the isometry from $\mathcal{P}_{2,\omega}(\mathbb{R}^d)$ to $L^2(\omega)$, we define the $W_{2,\omega}$ subdifferential.

Definition 2.18. Given $E : \mathcal{P}_{2,\omega}(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, and λ-convex with respect to $W_{2,\omega}$, $\boldsymbol{\xi} \in L^2(\omega)$ belongs to the $W_{2,\omega}$ *subdifferential* of E at μ in case for all $\nu \in \mathcal{P}_{2,\omega}(\mathbb{R}^d)$,

$$E(\nu) - E(\mu) \geq \int \langle \boldsymbol{\xi}, \mathbf{t}_\omega^\nu - \mathbf{t}_\omega^\mu \rangle \mathrm{d}\omega + \frac{\lambda}{2}W_{2,\omega}^2(\mu, \nu). \tag{2.10}$$

We denote this by $\boldsymbol{\xi} \in \partial_{2,\omega}E(\mu)$.

Remark 2.19 (Characterization of minimizers)**.** Given E as in Definition 2.18 with $\lambda \geq 0$, μ is a minimizer of E if and only if $0 \in \partial_{2,\omega}E(\mu)$.

By equation (2.3) and Remark 2.16, the square Wasserstein distance from ω, $\mu \mapsto W_2^2(\omega, \mu)$, is 2-convex in $W_{2,\omega}$. Thus, we may compute its subdifferential with respect to this metric.

Proposition 2.20. *Let $F(\mu) = W_2^2(\omega, \mu)$. Then $2(\mathbf{t}_\omega^\mu - \mathbf{id}) \in \partial_{2,\omega}F(\mu)$.*

Proof. $W_2^2(\omega, \nu) - W_2^2(\omega, \mu) = \int |\mathbf{t}_\omega^\nu - \mathbf{id}|^2 d\omega - \int |\mathbf{t}_\omega^\mu - \mathbf{id}|^2 \mathrm{d}\omega = \int 2\langle \mathbf{t}_\omega^\mu - \mathbf{id}, \mathbf{t}_\omega^\nu - \mathbf{t}_\omega^\mu \rangle \mathrm{d}\omega + W_{2,\omega}^2(\mu, \nu).$ $\square$

Finally, we relate the transport metric subdifferential to the strong subdifferential from Definition 2.5.

Lemma 2.21. *Suppose E satisfies domain Assumption 2.4 and is proper, lower semicontinuous, and λ-convex along generalized geodesics. Then if $\boldsymbol{\xi} \in \partial E(\mu)$ is a strong subdifferential, $\boldsymbol{\xi} \circ \mathbf{t}_\omega^\mu \in \partial_{2,\omega}E(\mu)$.*

Proof. If $\boldsymbol{\xi} \in \partial E(\mu)$, then $\boldsymbol{\xi} \in L^2(\mu)$, hence $\boldsymbol{\xi} \circ \mathbf{t}_\omega^\mu \in L^2(\omega)$. Furthermore, for all $\nu \in \mathcal{P}_{2,\omega}(\mathbb{R}^d)$,

$$E(\nu) - E(\mu) \geq \int_{\mathbb{R}^d} \langle \boldsymbol{\xi}, \mathbf{t}_\omega^\nu \circ \mathbf{t}_\mu^\omega - \mathbf{id} \rangle \mathrm{d}\mu + o(\|\mathbf{t}_\omega^\nu \circ \mathbf{t}_\mu^\omega - \mathbf{id}\|_{L^2(\mu)}) = \int_{\mathbb{R}^d} \langle \boldsymbol{\xi} \circ \mathbf{t}_\omega^\mu, \mathbf{t}_\omega^\nu - \mathbf{t}_\omega^\mu \rangle \mathrm{d}\omega + o(W_{2,\omega}(\mu, \nu)).$$

As in ([1], Sect. 10.1.1, B), this implies $E(\nu) - E(\mu) \geq \int \langle \boldsymbol{\xi} \circ \mathbf{t}_\omega^\mu, \mathbf{t}_\omega^\nu - \mathbf{t}_\omega^\mu \rangle \mathrm{d}\omega + \frac{\lambda}{2}W_{2,\omega}(\mu, \nu).$ $\square$

We close this subsection with an analogue of inequality (2.8) for transport metrics. Note that since $W_2^2(J_\tau\mu, \nu) \leq W_{2,\mu}^2(J_\tau\mu, \nu)$, this is stronger than (2.8). We require this strength in our proof of Theorem 3.4.

Theorem 2.22. *Suppose E is proper, coercive, lower semicontinuous, and λ-convex along generalized geodesics for $\lambda \in \mathbb{R}$. Then for all $\mu \in \overline{D(E)}$ and $\nu \in D(E)$,*

$$\frac{1}{2\tau}[W_{2,\mu}^2(J_\tau\mu, \nu) - W_2^2(\mu, \nu)] + \frac{\lambda}{2}W_{2,\mu}^2(J_\tau\mu, \nu) \leq E(\nu) - E(J_\tau\mu) - \frac{1}{2\tau}W_2^2(\mu, J_\tau\mu).$$

Proof. If μ does not charge small sets, so $W_{2,\mu}$ is a well-defined metric, this is simply the Talagrand inequality for $\Phi(\tau, \mu; \cdot)$ in the $W_{2,\mu}$ metric. Otherwise, since both E and $\frac{1}{2\tau}W_2^2(\mu, \cdot)$ are convex along generalized geodesics with base μ, so is $\Phi(\tau, \mu; \cdot)$. Thus, for any generalized geodesic μ_α from $J_\tau\mu$ to ν with base μ, since $J_\tau\mu$ is the minimizer of $\Phi(\tau, \mu; \cdot)$,

$$\Phi(\tau, \mu; J_\tau\mu) \leq \Phi(\tau, \mu; \mu_\alpha) \leq (1-\alpha)\Phi(\tau, \mu; J_\tau\mu) + \alpha\Phi(\tau, \mu; \nu) - \frac{1 + \lambda\tau}{2\tau}\alpha(1-\alpha)W_{2,\mu}^2(J_\tau\mu, \nu).$$

Rearranging and dividing by α gives $0 \leq \Phi(\tau, \mu; \nu) - \Phi(\tau, \mu; J_\tau\mu) - \frac{1+\lambda\tau}{2\tau}(1-\alpha)W_{2,\mu}^2(J_\tau\mu, \nu)$. Sending $\alpha \to 0$ and expanding Φ according to its definition gives the result. $\square$

2.5. Euler−Lagrange equation

In this section, we use our results on transport metrics from the previous section to prove an Euler−Lagrange equation characterizing $J_\tau\mu$, defined in equation (2.4) as the minimizer of the quadratic perturbation $\Phi(\tau, \mu; \cdot)$. The fact that $J_\tau\mu$ satisfies $\frac{1}{\tau}(\mathbf{t}_{J_\tau\mu}^\mu - \mathbf{id}) \in \partial E(J_\tau\mu)$ was proved by Ambrosio, Gigli, and Savaré using a type of argument introduced by Otto ([1], Lem. 10.1.2) [14,15]. The converse is new.

Theorem 2.23 (Euler−Lagrange equation)**.** *Assume that E satisfies domain Assumption 2.4 and convexity Assumption 2.11 for $\lambda \in \mathbb{R}$. Then for $\mu \in D(E)$ and $0 < \tau < \frac{1}{\lambda^-}$, ν is the unique minimizer of the quadratic perturbation $\Phi(\tau, \mu; \cdot)$ if and only if*

$$\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu) \text{ is a strong subdifferential.} \tag{2.11}$$

Hence, $J_\tau\mu$ is characterized by the fact that $\frac{1}{\tau}(\mathbf{t}_{J_\tau\mu}^\mu - \mathbf{id}) \in \partial E(J_\tau\mu)$ is a strong subdifferential.

We assume $\mu \in D(E)$ and E satisfies domain Assumption 2.4 to ease notation. See Theorem A.5 for how the assumption $\mu \in D(E)$ can be relaxed to $\mu \in \overline{D(E)}$ and the domain assumption can be removed.

Proof of Theorem 2.23. Suppose $\frac{1}{\tau}(\mathbf{t}_\nu^\mu - \mathbf{id}) \in \partial E(\nu)$ is a strong subdifferential. By Lemma 2.21, we have $\frac{1}{\tau}(\mathbf{id} - \mathbf{t}_\mu^\nu) \in \partial_{2,\mu}E(\nu)$. By additivity of the subdifferential and Proposition 2.20,

$$\frac{1}{2\tau}2(\mathbf{t}_\mu^\nu - \mathbf{id}) + \frac{1}{\tau}(\mathbf{id} - \mathbf{t}_\mu^\nu) = 0 \in \partial_{2,\mu}\Phi(\tau, \mu; \nu).$$

Since $W_2^2(\mu, \cdot)$ is 2-convex in the $W_{2,\mu}$ metric and E is λ-convex in the $W_{2,\mu}$ metric, $\Phi(\tau, \mu; \cdot)$ is $\left(\frac{1}{\tau} + \lambda\right)$-convex in the $W_{2,\mu}$ metric, with $\left(\frac{1}{\tau} + \lambda\right) > 0$. Therefore, by Remark 2.19, $0 \in \partial_{2,\mu}\Phi(\tau, \mu; \nu)$ implies ν minimizes $\Phi(\tau, \mu; \cdot)$. See ([1], Lem. 10.1.2) for the converse. $\square$

3. Exponential formula for the wasserstein metric

We now apply our results on transport metrics and the Euler−Lagrange equation to give a new proof of the exponential formula in the style of Crandall and Liggett [9]. In Section 3.1, we prove a new "almost" contraction inequality, analogous to the key inequality $\|J_\tau u - J_\tau v\| \leq \|u - v\|$ from Crandall and Liggett's proof, with higher order terms. In Section 3.2, we apply our Euler−Lagrange equation to control the behavior of the discrete gradient flow $J_\tau^n \mu$ as the time step τ varies. In particular, proximal maps with different time steps can be related by considering intermediate points along a geodesic between μ and $J_\tau \mu$. The convexity of the square transport metric $W_{2,\mu}^2$ along this geodesic allows us to control the behavior of the geodesic in terms of its endpoints. In Sections 3.3 and 3.4, we use these ideas to bound the distance between discrete gradient flow sequences with different time steps *via* an asymmetric induction in the style of Rasmussen [16]. Finally, in Section 3.5, we conclude that the discrete gradient flow converges to the gradient flow. We close Section 3.5 by applying our estimates to give simple proofs of properties of the gradient flow, including the contracting semigroup property and the energy dissipation inequality. (Note that we do not consider gradient flow with respect to the transport metrics, but instead use the transport metrics for intermediate estimates of the Wasserstein discrete gradient flow.)

3.1. Almost contraction inequality

In this subsection, we use the convexity of $\Phi(\tau, \mu; \cdot)$ along generalized geodesics with base μ, in the form of inequality (2.8), to prove an almost contraction inequality for the discrete gradient flow. (Inequality (2.8) is sufficient for this purpose − we use the stronger version in Theorem 2.22 later). Our approach is similar to previous work of Carlen and the author [4], though instead of symmetrizing the contraction inequality, we leave the inequality in an asymmetric form that is compatible with the asymmetric induction in Theorems 3.4−3.6.

For the $\lambda \leq 0$ case, our argument follows the first three steps in the proof of ([1], Lem. 4.2.4). For the $\lambda > 0$ case, we use a new approach.

Theorem 3.1 (Almost contraction inequality). *Suppose E satisfies convexity Assumption 2.11, $\mu \in D(|\partial E|)$, and $\nu \in \overline{D(E)}$. Then we have the following inequalities for all $0 < \tau < \frac{1}{\lambda^-}$,*

$$\text{if } \lambda > 0: \quad (1 + \lambda\tau)^2 W_2^2(J_\tau\mu, J_\tau\nu) \leq W_2^2(\mu, \nu) + \tau^2|\partial E|^2(\mu) + 2\lambda\tau^2\left[E(\nu) - \inf E\right],$$
$$\text{if } \lambda \leq 0: \quad (1 + \lambda\tau)^2 W_2^2(J_\tau\mu, J_\tau\nu) \leq W_2^2(\mu, \nu) + \tau^2|\partial E|^2(\mu).$$

Proof. Throughout the proof we abbreviate $J_\tau\mu$ by μ_τ and $J_\tau\nu$ by ν_τ. By inequality (2.8),

$$(1 + \lambda\tau)W_2^2(\mu_\tau, \nu_\tau) - W_2^2(\mu, \nu) \leq 2\tau[E(\nu_\tau) - E(\mu_\tau) - (1/2\tau)W_2^2(\mu, \mu_\tau) \tag{3.1}$$
$$(1 + \lambda\tau)W_2^2(\nu_\tau, \mu) - W_2^2(\nu, \mu) \leq 2\tau[E(\mu) - E(\nu_\tau) - (1/2\tau)W_2^2(\nu, \nu_\tau)]. \tag{3.2}$$

Suppose $\lambda > 0$. Dropping $-\frac{1}{2\tau}W_2^2(\nu, \nu_\tau)$ from (3.2), multiplying (3.1) by $(1 + \lambda\tau)$, and adding the two together,

$$(1 + \lambda\tau)^2 W_2^2(\mu_\tau, \nu_\tau) - W_2^2(\mu, \nu) \leq 2\tau\left[(1 + \lambda\tau)E(\nu_\tau) - E(\nu_\tau) + E(\mu) - (1 + \lambda\tau)\left[E(\mu_\tau) + (1/2\tau)W_2^2(\mu, \mu_\tau)\right]\right]$$

Since $\lambda > 0$, E is bounded below ([1], Lem. 2.4.8). Applying inequality (2.7) and the fact that $E(\nu_\tau) \leq E(\nu)$,

$$(1 + \lambda\tau)^2 W_2^2(\mu_\tau, \nu_\tau) - W_2^2(\mu, \nu) \leq \tau^2\left[\frac{1}{1 + \lambda\tau}|\partial E|^2(\mu) + 2\lambda[E(\nu_\tau) - \inf E]\right] \leq \tau^2[|\partial E|^2(\mu) + 2\lambda\left[E(\nu) - \inf E\right]],$$

which gives the result.

Now suppose $\lambda \leq 0$. Adding (3.1) and (3.2) and applying inequality (2.7),

$$(1 + \lambda\tau)W_2^2(\mu_\tau, \nu_\tau) - W_2^2(\nu, \mu) + \lambda\tau W_2^2(\nu_\tau, \mu) \leq 2\tau\left[E(\mu) - E(\mu_\tau) - (1/2\tau)W_2^2(\mu, \mu_\tau)\right] - W_2^2(\nu, \nu_\tau)$$
$$\leq \frac{\tau^2}{1 + \lambda\tau}|\partial E|^2(\mu) - W_2^2(\nu, \nu_\tau). \tag{3.3}$$

For $a, b > 0$ and $0 < \epsilon < 1$, the convex function $\phi(\epsilon) = a^2/\epsilon + b^2/(1 - \epsilon)$ attains its minimum $(a + b)^2$ at $\epsilon = a/(a + b)$, hence $a^2/\epsilon + b^2/(1 - \epsilon) \geq (a + b)^2$.

Consequently, with $\epsilon = -\lambda\tau$, we obtain

$$W_2^2(\nu_\tau, \mu) \leq (W_2(\nu_\tau, \nu) + W_2(\nu, \mu))^2 \leq -W_2^2(\nu_\tau, \nu)/\lambda\tau + W_2^2(\nu, \mu)/(1 + \lambda\tau). \tag{3.4}$$

Multiplying by $-\lambda\tau$, adding to (3.3), and multiplying by $(1 + \lambda\tau)$, we obtain the result

$$(1 + \lambda\tau)^2 W_2^2(\mu_\tau, \nu_\tau) \leq W_2^2(\mu, \nu) + \tau^2|\partial E|^2(\mu). \qquad \square$$

Iterating the inequalities in the above theorem and applying inequality (2.7) gives the following corollary.

Corollary 3.2. *Suppose E satisfies convexity Assumption 2.11, $\mu \in D(|\partial E|)$, and $\nu \in \overline{D(E)}$. Then we have the following inequalities for all $0 < \tau < -\frac{1}{\lambda}$,*

$$\text{if } \lambda > 0: \quad W_2^2(J_\tau^n \mu, J_\tau^n \nu) \leq (1 + \lambda\tau)^{-2n} W_2^2(\mu, \nu) + n\tau^2 \left(|\partial E|^2(\mu) + 2\lambda \left[E(\nu) - \inf E\right]\right), \tag{3.5}$$

$$\text{if } \lambda \leq 0: \quad W_2^2(J_\tau^n \mu, J_\tau^n \nu) \leq (1 + \lambda\tau)^{-2n} W_2^2(\mu, \nu) + n\tau^2(1 + \lambda\tau)^{-2n}|\partial E|^2(\mu). \tag{3.6}$$

3.2. Proximal Map with large vs. small time steps

We now apply the Euler−Lagrange equation, Theorem 2.23, to relate the proximal map with a large time step τ to the proximal map with a small time step h.

Theorem 3.3. *Suppose E satisfies convexity Assumption 2.11. Then if $\mu \in D(E)$ and $0 < h \leq \tau < \frac{1}{\lambda^-}$,*

$$J_\tau \mu = J_h \left[\left(\frac{\tau - h}{\tau}\mathbf{t}_\mu^{J_\tau \mu} + \frac{h}{\tau}\mathbf{id}\right) \# \mu\right].$$

Proof of Theorem 3.3. For simplicity of notation, we prove the result in the case E satisfies domain Assumption 2.4. See Theorem A.6 for the general result. As in the previous proof, we abbreviate $J_\tau\mu$ by μ_τ.

By Theorem 2.23, $\boldsymbol{\xi} = \frac{1}{\tau}(\mathbf{t}_{\mu_\tau}^\mu - \mathbf{id}) \in \partial E(\mu_\tau)$ is a strong subdifferential. Since $h/\tau \leq 1$,

$$(\mathbf{id} + h\boldsymbol{\xi}) = \left(\mathbf{id} + \frac{h}{\tau}(\mathbf{t}_{\mu_\tau}^\mu - \mathbf{id})\right) = \left(\frac{\tau - h}{\tau}\mathbf{id} + \frac{h}{\tau}\mathbf{t}_{\mu_\tau}^\mu\right) \tag{3.7}$$

is cyclically monotone. Consequently, if we define $\nu = (\mathbf{id} + h\boldsymbol{\xi})\#\mu_\tau$, then $\mathbf{t}_{\mu_\tau}^\nu = \mathbf{id} + h\boldsymbol{\xi}$. Rearranging shows $\frac{1}{h}(\mathbf{t}_{\mu_\tau}^\nu - \mathbf{id}) = \boldsymbol{\xi} \in \partial E(\mu_\tau)$, so by a second application of Theorem 2.23, $\mu_\tau = \nu_h$.

We now rewrite ν as it appears in the theorem. By equation (3.7), $(\mathbf{id} + h\boldsymbol{\xi}) = \left(\frac{\tau-h}{\tau}\mathbf{id} + \frac{h}{\tau}\mathbf{t}_{\mu_\tau}^\mu\right) = \left(\frac{\tau-h}{\tau}\mathbf{t}_\mu^{\mu_\tau} + \frac{h}{\tau}\mathbf{id}\right) \circ \mathbf{t}_{\mu_\tau}^\mu$. Therefore, $\nu = (\mathbf{id} + h\boldsymbol{\xi})\#\mu_\tau = \left(\frac{\tau-h}{\tau}\mathbf{t}_\mu^{\mu_\tau} + \frac{h}{\tau}\mathbf{id}\right)\#\mu$. $\qquad \square$

After proving Theorem 3.3, we discovered another proof of the same result, independently obtained by Jost and Mayer [11, 12]. It is non-variational and quite different from the proof given above, and we hope our proof is of independent interest.

3.3. Asymmetric recursive inequality

The following inequality bounds the Wasserstein distance between discrete gradient flow sequences with different time steps in terms of a convex combination of earlier elements of the sequences, plus a small error term. A fundamental difference between Crandall and Liggett's recursive inequality and our Theorem 3.4 is that theirs involves the distance while ours involves the square distance. (This is a consequence of the fact that our contraction inequality Thm. 3.1 involves the square distance plus error terms.) Therefore, where Crandall and Liggett were able to use the triangle inequality to control the distance to intermediate points on a geodesic in terms of the distance to its endpoints, we have to use the convexity of the square transport metric. The bulk of our proof is devoted to passing from the transport metric back to the Wasserstein metric.

Theorem 3.4. *Suppose E satisfies convexity Assumption* 2.11. *Then if $\mu \in D(|\partial E|)$ and $0 < h \le \tau < \frac{1}{\lambda^-}$,*

$$(1 - \lambda^- h)^2 W_2^2(J_\tau^n \mu, J_h^m \mu)$$
$$\le \frac{h}{\tau}(1 - \lambda^- \tau)^{-1} W_2^2(J_\tau^{n-1}\mu, J_h^{m-1}\mu) + \frac{\tau - h}{\tau} W_2^2(J_h^{m-1}\mu, J_\tau^n \mu) + 2h^2(1 - \lambda^- h)^{-2m}|\partial E|^2(\mu).$$

To consider $\lambda \ge 0$ and $\lambda < 0$ jointly, we replace λ by $-\lambda^-$: any function that is λ convex is also $-\lambda^-$ convex.

Proof. To simplify notation, we abbreviate $J_\tau^n \mu$ by J^n, $J_h^m \mu$ by J^m, and $[([\tau - h]/\tau)\mathbf{t}_{J^{n-1}}^{J^n} + (h/\tau)\mathbf{id}]\#J^{n-1}$ by $\mu_{\frac{\tau-h}{\tau}}^{J^{n-1} \to J^n}$, since the latter is the geodesic from J^{n-1} to J^n at time $(\tau - h)/\tau$. With this, we have

$$(1 - \lambda^- h)^2 W_2^2(J^n, J^m) = (1 - \lambda^- h)^2 W_2^2(J_h(\mu_{\frac{\tau-h}{\tau}}^{J^{n-1} \to J^n}), J^m) \qquad \text{by Theorem 3.3,}$$
$$\le W_2^2(\mu_{\frac{\tau-h}{\tau}}^{J^{n-1} \to J^n}, J^{m-1}) + h^2|\partial E|^2(J^{m-1}) \qquad \text{by Theorem 3.1,}$$
$$\le W_{2, J^{n-1}}^2(\mu_{\frac{\tau-h}{\tau}}^{J^{n-1} \to J^n}, J^{m-1}) + h^2|\partial E|^2(J^{m-1}) \qquad \text{by Remark 2.9.}$$

By Proposition 2.13, $W_{2, J^{n-1}}^2$ is convex along generalized geodesics with base J^{n-1}. In particular, it is convex along the geodesic $\mu_{\frac{\tau-h}{\tau}}^{J^{n-1} \to J^n}$, which gives

$$(1 - \lambda^- h)^2 W_2^2(J^n, J^m) \le \frac{h}{\tau} W_{2, J^{n-1}}^2(J^{n-1}, J^{m-1}) + \frac{\tau - h}{\tau} W_{2, J^{n-1}}^2(J^{m-1}, J^n) + h^2|\partial E|^2(J^{m-1}). \qquad (3.8)$$

By Remark 2.9, $W_{2, J^{n-1}}^2(J^{n-1}, J^{m-1}) = W_2^2(J^{n-1}, J^{m-1})$. To control the second term, we claim that

$$W_{2, J^{n-1}}^2(J^{m-1}, J^n) \le \frac{\tau}{h} \left(W_2^2(J^{m-1}, J^n) - (1 - \lambda^- h) W_{2, J^{m-1}}^2(J^m, J^n) \right)$$
$$+ \frac{1}{1 - \lambda^- \tau} W_2^2(J^{n-1}, J^{m-1}) + \frac{\tau h}{1 - \lambda^- h}|\partial E|^2(J^{m-1}). \qquad (3.9)$$

Substituting (3.9) into (3.8), simplifying and rearranging, and using $(1 - \lambda^- h)^2 \le (1 - \lambda^- h)$ gives

$$(1 - \lambda^- h)^2 (\tau/h) W_2^2(J^n, J^m) \le \frac{1 - \lambda^- h}{1 - \lambda^- \tau} W_2^2(J^{n-1}, J^{m-1}) + \frac{\tau - h}{h} W_2^2(J^{m-1}, J^n)$$
$$+ \left[\frac{h(\tau - h)}{1 - \lambda^- h} + h^2 \right] |\partial E|^2(J^{m-1}).$$

Multiplying by h/τ and using both $1 - \lambda^- h \le 1$ and $|\partial E|^2(J^{m-1}) \le (1 - \lambda^- h)^{-2(m-1)}|\partial E|^2(\mu)$ gives the result

$$(1 - \lambda^- h)^2 W_2^2(J^n, J^m) \le \frac{h}{\tau} \frac{1}{1 - \lambda^- \tau} W_2^2(J^{n-1}, J^{m-1}) + \frac{\tau - h}{\tau} W_2^2(J^{m-1}, J^n) + 2h^2(1 - \lambda^- h)^{-2m}|\partial E|^2(\mu).$$

It remains to show (3.9). Replacing (μ, ν) in Theorem 2.22 with (J^{m-1}, J^n) and (J^{n-1}, J^{m-1}) gives

$$(1 - \lambda^- h) W_{2, J^{m-1}}^2(J^m, J^n) - W_2^2(J^{m-1}, J^n) \le 2h \left[E(J^n) - E(J^m) - W_2^2(J^{m-1}, J^m)/(2h) \right], \qquad (3.10)$$
$$(1 - \lambda^- \tau) W_{2, J^{n-1}}^2(J^n, J^{m-1}) - W_2^2(J^{n-1}, J^{m-1}) \le 2\tau \left[E(J^{m-1}) - E(J^n) - W_2^2(J^{n-1}, J^n)/(2\tau) \right]. \qquad (3.11)$$

Multiplying (3.10) by τ, (3.11) by h, adding them together, and applying inequality (2.7) gives

$$\tau(1 - \lambda^- h) W_{2, J^{m-1}}^2(J^m, J^n) + h(1 - \lambda^- \tau) W_{2, J^{n-1}}^2(J^n, J^{m-1})$$
$$\le \tau W_2^2(J^{m-1}, J^n) + h W_2^2(J^{n-1}, J^{m-1})$$
$$+ 2\tau h \left[E(J^{m-1}) - E(J^m) - W_2^2(J^{m-1}, J^m)/(2h) \right] - h W_2^2(J^{n-1}, J^n)$$
$$\le \tau W_2^2(J^{m-1}, J^n) + h W_2^2(J^{n-1}, J^{m-1}) + \frac{\tau h^2}{1 - \lambda^- h}|\partial E|^2(J^{m-1}) - h W_2^2(J^{n-1}, J^n). \qquad (3.12)$$

As in equation (3.4) we have, $\lambda^- \tau W_{2,J^{n-1}}^2(J^{m-1}, J^n) \leq W_2^2(J^n, J^{n-1}) + \frac{\lambda^- \tau}{1 - \lambda^- \tau} W_2^2(J^{n-1}, J^{m-1})$. Multiplying by h and adding to (3.12) gives

$$\tau(1 - \lambda^- h)W_{2,J^{m-1}}^2(J^m, J^n) + hW_{2,J^{n-1}}^2(J^n, J^{m-1})$$

$$\leq \tau W_2^2(J^{m-1}, J^n) + \frac{h}{1 - \lambda^- \tau} W_2^2(J^{n-1}, J^{m-1}) + \frac{\tau h^2}{1 - \lambda^- h} |\partial E|^2(J^{m-1}).$$

Rearranging and dividing by h gives the desired bound (3.9).   □

3.4. Distance between discrete gradient flows with different time steps

The following bound controls the distance between discrete gradient flow sequences with time steps τ and h. It is inspired by Rasmussen's simplification of Crandall and Liggett's method [16, 18]. Unlike in the Banach space case, we work with the square distance instead of the distance itself. While this complicated matters in the previous theorem, it simplifies the induction in the following theorem.

We begin with the base case of the induction, bounding the distance between the 0th and nth terms.

Lemma 3.5. *Suppose E satisfies convexity Assumption* 2.11. *Then if $\mu \in D(|\partial E|)$ and $0 < \tau < \frac{1}{\lambda^-}$,*

$$W_2(J_\tau^n \mu, \mu) \leq \frac{n\tau}{(1 - \tau\lambda^-)^n} |\partial E(\mu)|$$

Proof. This follows from the triangle inequality, (2.7), and the inequalities $\frac{1}{1 + \tau\lambda} \leq \frac{1}{1 - \tau\lambda^-}$ and $1 \leq \frac{1}{1 - \tau\lambda^-}$.

$$W_2(J_\tau^n \mu, \mu) \leq \sum_{i=1}^n W_2(J_\tau^i \mu, J_\tau^{i-1} \mu) \leq \sum_{i=1}^n \frac{\tau}{1 + \tau\lambda} |\partial E(J_\tau^{i-1} \mu)| \leq \sum_{i=1}^n \frac{\tau}{(1 + \tau\lambda)^i} |\partial E(\mu)| \leq \frac{n\tau}{(1 - \tau\lambda^-)^n} |\partial E(\mu)|.$$

□

Theorem 3.6. *Suppose E satisfies convexity Assumption* 2.11. *Then if $\mu \in D(|\partial E|)$ and $0 < h \leq \tau < \frac{1}{\lambda^-}$,*

$$W_2^2(J_\tau^n \mu, J_h^m \mu) \leq \left[(n\tau - mh)^2 + \tau hm + 2\tau^2 n \right] (1 - \lambda^- \tau)^{-2n}(1 - \lambda^- h)^{-2m} |\partial E|^2(\mu). \tag{3.13}$$

Proof. We proceed by induction. The base case, when either $n = 0$ or $m = 0$, follows from Lemma 3.5. We assume the inequality holds for $(n - 1, m)$ and (n, m) and show that this implies it holds for $(n, m + 1)$.

By Theorem 3.4,

$$(1 - \lambda^- h)^2 W_2^2(J_\tau^n \mu, J_h^{m+1} \mu)$$
$$\leq \frac{h}{\tau}(1 - \lambda^- \tau)^{-1} W_2^2(J_\tau^{n-1} \mu, J_h^m \mu) + \frac{\tau - h}{\tau} W_2^2(J_h^m \mu, J_\tau^n \mu) + 2h^2(1 - \lambda^- h)^{-2(m+1)} |\partial E|^2(\mu).$$

Dividing by $(1 - \lambda^- h)^2$ and applying the inductive hypothesis,

$$W_2^2(J_\tau^n \mu, J_h^{m+1} \mu) \leq \frac{h}{\tau} \left[((n-1)\tau - mh)^2 + \tau hm + 2\tau^2(n-1) \right] (1 - \lambda^- \tau)^{-2(n-1)-1}(1 - \lambda^- h)^{-2(m+1)} |\partial E|^2(\mu)$$

$$+ \frac{\tau - h}{\tau} \left[(n\tau - mh)^2 + \tau hm + 2\tau^2 n \right] (1 - \lambda^- \tau)^{-2n}(1 - \lambda^- h)^{-2(m+1)} |\partial E(\mu)|^2$$

$$+ 2h^2(1 - \lambda^- h)^{-2(m+1)-2} |\partial E|^2(\mu).$$

To control the first term, note that $(1 - \lambda^- \tau)^{-2(n-1)-1} = (1 - \lambda^- \tau)^{-2n+1} < (1 - \lambda^- \tau)^{-2n}$ and

$$\left[((n-1)\tau - mh)^2 + \tau hm + 2\tau^2(n-1) \right] = \left[(n\tau - mh)^2 - 2(n\tau - mh)\tau + \tau^2 + \tau hm + 2\tau^2(n-1) \right].$$

To control the third term, note that $(1 - \lambda^- h)^{-2} \le (1 - \lambda^- \tau)^{-2} \le (1 - \lambda^- \tau)^{-2n}$. Using these estimates, we may group together the three terms and obtain the following bound.

$$W_2^2(J_\tau^n \mu, J_h^{m+1}\mu)$$
$$\le \left\{ \frac{h}{\tau} \left[(n\tau - mh)^2 - 2(n\tau - mh)\tau + \tau^2 + \tau hm + 2\tau^2(n-1) \right] + \frac{\tau - h}{\tau} \left[(n\tau - mh)^2 + \tau hm + 2\tau^2 n \right] + 2h^2 \right\}$$
$$\cdot (1 - \lambda^- \tau)^{-2n}(1 - \lambda^- h)^{-2(m+1)}|\partial E|^2(\mu).$$

Simplifying and bounding the quantity within the brackets gives the result,

$$\frac{h}{\tau} \left[(n\tau - mh)^2 - 2(n\tau - mh)\tau + \tau^2 + \tau hm + 2\tau^2(n-1) \right] + \frac{\tau - h}{\tau} \left[(n\tau - mh)^2 + \tau hm + 2\tau^2 n \right] + 2h^2$$
$$= \frac{h}{\tau} \left[(n\tau - mh)^2 + \tau hm + 2\tau^2 n \right] + \frac{\tau - h}{\tau} \left[(n\tau - mh)^2 + \tau hm + 2\tau^2 n \right] + \frac{h}{\tau} \left[-2(n\tau - mh)\tau - \tau^2 \right] + 2h^2$$
$$= \left[(n\tau - mh)^2 + \tau hm + 2\tau^2 n \right] - 2(n\tau - mh)h - \tau h + 2h^2$$
$$= (n\tau - mh)^2 - 2(n\tau - mh)h + \tau hm - \tau h + 2\tau^2 n + 2h^2$$
$$= (n\tau - (m+1)h)^2 + \tau hm - \tau h + 2\tau^2 n + h^2$$
$$\le (n\tau - (m+1)h)^2 + \tau h(m+1) + 2\tau^2 n. \qquad \square$$

3.5. Exponential formula for the wasserstein metric

We now combine our previous results to conclude the exponential formula for the Wasserstein metric. We prove the quantitative bound $W_2(J_{t/n}^n \mu, \mu(t)) \le O(n^{-1/2})$, which agrees with the rate Crandall and Liggett obtained in the Banach space case[9]. By a different method, Ambrosio, Gigli, and Savaré showed $W_2(J_{t/n}^n \mu, \mu(t)) \le O(n^{-1})$ ([1], Thm. 4.0.4), which agrees with the optimal rate in a Hilbert space [17]. Our rate improves upon the rate obtained by Clément and Desch [7], $d(J_{t/n}^n \mu, \mu(t)) \le O(n^{-1/4})$, though they considered the more general case of a metric space (X, d) that satisfies inequality (2.8), with W_2 replaced by d.

Though we do not obtain the optimal rate of convergence, we demonstrate that Crandall and Liggett's approach extends to the Wasserstein metric, providing a simple and robust route to the exponential formula and properties of continuous gradient flow. In particular, it is hoped that this method may be used to study the behavior of the gradient flow as the functional E varies − for example, as a regularization of E is removed. The corresponding problem in the Banach space case is well-understood [3], and the analogy we establish between the Banach and Wasserstein cases may help extend these results.

Theorem 3.7 (Exponential formula). *Suppose E satisfies convexity Assumption 2.11. For $\mu \in \overline{D(E)}$, $t \ge 0$, the discrete gradient flow sequence $J_{t/n}^n \mu$ converges as $n \to \infty$. The convergence is uniform in t on compact subsets of $[0, +\infty)$, and the limit $\mu(t)$ is the gradient flow of E with initial conditions μ_0 in the sense of Definition 2.12. When $\mu_0 \in D(|\partial E|)$ and $n \ge 2\lambda^- t$, the distance between $J_{t/n}^n$ and $\mu(t)$ is bounded by*

$$W_2(J_{t/n}^n \mu, \mu(t)) \le \sqrt{3}\frac{t}{\sqrt{n}}\mathrm{e}^{3\lambda^- t}|\partial E|(\mu_0). \tag{3.14}$$

Remark 3.8 (Varying time steps). For any partition of the interval $[0, t]$ into n time steps $\tau_1, \ldots, \tau_n$, the corresponding discrete gradient flow with varying time steps $\Pi_{i=1}^n J_{\tau_i}\mu$ converges to $\mu(t)$ as the maximum time step goes to zero. See Theorem A.4.

We may apply our estimates to give shorts proofs of several known properties of the gradient flow ([1], Thm. 2.4.15, Prop. 4.3.1, Cor. 4.3.3). Our proofs merely use the fact that $\mu(t) = \lim_{n\to\infty} J_{t/n}^n \mu_0$, not that $\mu(t)$ is a solution to the gradient flow in the sense of Definition 2.12.

Theorem 3.9. *Suppose E satisfies convexity Assumption* 2.11. *Then the function $S(t)$ on $(0, +\infty)$, $S(t)$:* $\overline{D(E)} \to D(|\partial E|)$: *$\mu \mapsto \mu(t)$ is a λ-contracting semigroup,* i.e.

(i) $\lim_{t \to 0} S(t)\mu = S(0)\mu = \mu$,
(ii) $S(t+s) = S(t)S(s)\mu$ *for $t, s \geq 0$,*
(iii) $W_2(S(t)\mu, S(t)\nu) \leq e^{-\lambda t} W_2(\mu, \nu)$.

Furthermore, $\mu(t)$ is locally Lipschitz on $(0, +\infty)$ and if $\mu \in D(|\partial E|)$,

$$W_2(\mu(t), \mu(s)) \leq |t - s| e^{\lambda^- t} e^{\lambda^- s} |\partial E|(\mu). \tag{3.15}$$

Finally, an energy dissipation inequality holds,

$$\int_{t_0}^{t_1} |\partial E|^2(\mu(s)) \mathrm{d}s \leq E(\mu(t_0)) - E(\mu(t_1)). \tag{3.16}$$

We now turn to the proofs of these results.

Proof of Theorem 3.7. To ease notation, we drop the subscript on the initial data μ_0 and we abbreviate $J_\tau \mu$ by μ_τ. We also use that $\mu \in \overline{D(E)}$ implies $\mu_\tau \in D(|\partial E|)$ (2.7) and $\lim_{\tau \to 0} \mu_\tau = \mu$ ([1], Lem. 3.1.2).

First, we prove the error estimate for $\mu \in D(|\partial E|)$. By Theorem 3.6, if we define $\tau = \frac{t}{n}$, $h = \frac{t}{m}$, with $m \geq n > 2t\lambda^-$, so $0 \leq h \leq \tau < \frac{1}{2\lambda^-}$,

$$W_2^2(J_{t/n}^n \mu, J_{t/m}^m \mu) \leq 3\frac{t^2}{n}(1 - \lambda^- t/n)^{-2n}(1 - \lambda^- t/m)^{-2m} |\partial E|^2(\mu) \leq 3\frac{t^2}{n} e^{8\lambda^- t} |\partial E|^2(\mu). \tag{3.17}$$

In the second inequality, we use $(1 - \alpha)^{-1} \leq e^{2\alpha}$ for $\alpha \in [0, 1/2]$. Thus, the sequence $J_{t/n}^n \mu$ is Cauchy, and $\lim_{n \to \infty} J_{t/n}^n \mu$ exists ([1], Prop. 7.1.5). The convergence is uniform in t on compact subsets of $[0, +\infty)$. If $\mu(t)$ denotes the limit, then sending $m \to \infty$ in the first inequality of (3.17) gives (3.14).

Now suppose $\mu \in \overline{D(E)}$. By the triangle inequality,

$$W_2(J_{t/n}^n \mu, J_{t/m}^m \mu) \leq W_2(J_{t/n}^n \mu, J_{t/n}^n \mu_\tau) + W_2(J_{t/n}^n \mu_\tau, J_{t/m}^m \mu_\tau) + W_2(J_{t/m}^m \mu_\tau, J_{t/m}^m \mu).$$

We may choose n, m large enough, uniformly in $t \in [0, T]$, so that the second term arbitrarily small. By Corollary 3.2, we may choose n, m large enough and τ small enough, uniformly in $t \in [0, T]$, so that the first and third terms are arbitrarily small. Thus, the sequence $J_{t/n}^n \mu$ is Cauchy uniformly in $t \in [0, T]$, so the limit exists and convergence is uniform for $t \in [0, T]$.

We now show that $\mu(t)$ is the gradient flow of E with initial conditions μ in the sense of Definition 2.12. As our proof of Theorem 3.9 merely uses the fact that $\mu(t) = \lim_{n \to \infty} J_{t/n}^n \mu$, we may leverage these results. By (i), $\lim_{t \to 0} S(t)\mu = \mu$, so it remains to show that $S(t)\mu$ satisfies inequality (2.6) from Definition 2.12.

Define the piecewise constant function $g_n(s) = [E(J_\tau^i \mu) - E(\omega)](1 + \lambda\tau)^{i-1}$ for $s \in ((i-1)\tau, i\tau]$, $1 \leq i \leq n$. Iterating inequality (2.8) for $0 < \tau < 1/\lambda^-$ shows that for all $\omega \in D(E)$,

$$(1 + \lambda\tau)^n W_2^2(J_\tau^n \mu, \omega) \leq W_2^2(\mu, \omega) + 2\tau \sum_{i=1}^n [E(\omega) - E(J_\tau^i \mu)](1 + \lambda\tau)^{i-1} = W_2^2(\mu, \omega) - 2\int_0^{\tau n} g_n(s) \mathrm{d}s$$

Set $\tau = t/n$ and suppose the following claim holds:

Claim. $\liminf_{n \to \infty} \int_0^t g_n(s) \mathrm{d}s \geq \int_0^t [E(S(s)\mu) - E(\omega)] e^{\lambda s} \mathrm{d}s$.
Then taking the liminf of the above inequality gives

$$e^{\lambda t} W_2^2(S(t)\mu, \omega) \leq W_2^2(\mu, \omega) + 2\int_0^t [E(\omega) - E(S(s)\mu)] e^{\lambda s} \mathrm{d}s. \tag{3.18}$$

Applying the semigroup property (ii) of Theorem 3.9 and multiplying through by $e^{\lambda t_0}$,

$$e^{\lambda(t+t_0)}W_2^2(S(t+t_0)\mu, \omega) - e^{\lambda t_0}W_2^2(S(t_0)\mu, \omega) \leq 2\int_{t_0}^{t+t_0}[E(\omega) - E(S(s)\mu)]e^{\lambda s}\mathrm{d}s.$$

Dividing by t, sending $t \to 0$, and dividing by $2e^{\lambda t_0}$ gives inequality (2.6). The left hand side converges for a.e. t_0 since the function $f(t) = \frac{1}{2}W_2^2(S(t)\mu, \omega)$ is locally Lipschitz by Theorem 3.9.

We conclude by proving the claim. Since $E(J_{t/n}^i\mu) \geq E(J_{t/n}^n\mu)$ and $\liminf_{n\to\infty} E(J_{t/n}^n\mu) \geq E(S(t)\mu)$, $g_n(s)$ is bounded below. By Fatou's lemma, it is enough to show $\liminf_{n\to\infty} g_n(s)\mathrm{d}s \geq [E(S(s)\mu) - E(\omega)]e^{\lambda s}$, and by the lower semicontinuity of E, this holds if $\lim_{n\to\infty} J_{t/n}^i\mu = S(s)\mu$. By the triangle inequality,

$$W_2(J_{t/n}^i\mu, S(s)\mu) \leq W_2(J_{t/n}^i\mu, J_{t/n}^i\mu_\tau) + W_2(J_{t/n}^i\mu_\tau, S(s)\mu_\tau) + W_2(S(s)\mu_\tau, S(s)\mu).$$

By Corollary 3.2, we may choose n large enough and τ small enough, uniformly in $t \in [0, T]$, so that the first term is arbitrarily small. Likewise, the third term may be made arbitrarily small by the contraction inequality (iii) of Theorem 3.9. Thus it remains to show that $\lim_{n\to\infty} J_{t/n}^i\mu_\tau = S(s)\mu_\tau$. This follows by Theorem 3.6, since for $s \in ((i-1)t/n, it/n]$ and m, n large enough,

$$W_2^2(J_{t/n}^i\mu_\tau, J_{s/m}^m\mu_\tau) \leq \left(\left(i\frac{t}{n} - s\right)^2 + \frac{t}{n}s + 2\frac{t^2}{n^2}i\right)e^{4\lambda^-(t+s)}|\partial E|(\mu_\tau)$$

$$\leq \left(\frac{3t^2}{n^2} + \frac{ts + 2t^2}{n}\right)e^{4\lambda^-(t+s)}|\partial E|(\mu_\tau). \qquad \square$$

Proof of Theorem 3.9. Throughout, we abbreviate $J_\tau\mu$ by μ_τ and use that $\mu \in \overline{D(E)}$ implies $\mu_\tau \in D(|\partial E|)$ by (2.7) and $\lim_{\tau\to 0}\mu_\tau = \mu$ by ([1], Lem. 3.1.2).

(iii) follows by sending $n \to \infty$ in Corollary 3.2. In particular, for $\mu, \nu \in \overline{D(E)}$,

$$W_2^2(S(t)\mu_\tau, S(t)\nu) \leq e^{-\lambda t}W_2^2(\mu_\tau, \nu).$$

The result then follows by the triangle inequality and the continuity of the proximal map, as $\tau \to 0$,

$$W_2(S(t)\mu, S(t)\nu) \leq W_2(S(t)\mu, S(t)\mu_\tau) + W_2(S(t)\mu_\tau, S(t)\nu) \leq e^{-\lambda t}W_2(\mu, \mu_\tau) + e^{-\lambda t}W_2(\mu_\tau, \nu).$$

Next, we show estimate (3.15) on the modulus of continuity for $S(t)\mu$ when $\mu \in D(|\partial E|)$. Given $t \geq s \geq 0$, define $\tau = \frac{t}{n}, h = \frac{s}{m}$ for m, n large enough so $0 \leq h \leq \tau < \frac{1}{\lambda^-}$. By Theorem 3.6,

$$W_2^2(J_{t/n}^n\mu, J_{s/m}^m\mu) \leq \left[(t-s)^2 + \frac{ts}{n} + 2\frac{t^2}{n}\right](1 - \lambda^- t/n)^{-2n}(1 - \lambda^- s/m)^{-2m}|\partial E|^2(\mu). \qquad (3.19)$$

Sending $m, n \to \infty$ and taking the square root of both sides gives (3.15). If $\mu \in \overline{D(E)}$, then by (iii), $\lim_{\tau\to 0} S(t)\mu_\tau = S(t)\mu$ uniformly for $t \in [0, T]$, so $S(t)\mu$ is also continuous for $t \in [0, T]$. Thus (i) holds.

Next, we turn to (ii). Note that it is enough to show $S(t)^m\mu = S(mt)\mu$ for fixed $m \in \mathbb{N}$. If we have this, then for any $l, k, r, s \in \mathbb{N}$,

$$S\left(\frac{l}{k} + \frac{r}{s}\right)\mu = S\left(\frac{ls + rk}{ks}\right)\mu = \left[S\left(\frac{1}{ks}\right)\right]^{ls+rk}\mu = \left[S\left(\frac{1}{ks}\right)\right]^{ls}\left[S\left(\frac{1}{ks}\right)\right]^{rk}\mu = S\left(\frac{l}{k}\right)S\left(\frac{r}{s}\right)\mu.$$

Since $S(t)\mu$ is continuous in $t \in [0, +\infty)$, $S(t+s)\mu = S(t)S(s)\mu$ for all $t, s \geq 0$. Likewise, it is enough to show the result for $\mu \in D(|\partial E|)$. If $\mu \in \overline{D(E)}$, then by (iii), $\lim_{\tau\to 0} S(t)\mu_\tau = S(t)\mu$ uniformly for $t \in [0, T]$, so $S(t+s)\mu = \lim_{\tau\to 0} S(t+s)\mu_\tau = \lim_{\tau\to 0} S(t)S(s)\mu_\tau = S(t)S(s)\mu$.

We proceed by induction. $S(t)^m\mu = S(mt)\mu$ for $m = 1$. Suppose that $S(t)^{m-1}\mu = S((m-1)t)\mu$. We show that we may choose n large enough to make the right hand side of the following inequality arbitrarily small:

$$W_2(S(mt)\mu, S(t)^m\mu) \le W_2(S(mt)\mu, (J^n_{t/n})^m\mu) + W_2((J^n_{t/n})^m\mu, J^n_{t/n}S(t)^{m-1}\mu) + W_2(J^n_{t/n}S(t)^{m-1}\mu, S(t)^m\mu)$$

Since $W_2(S(mt)\mu, (J^n_{t/n})^m\mu) = W_2(S(mt)\mu, J^{nm}_{tm/nm}\mu)$ and $\lim_{n\to\infty} J^{nm}_{tm/nm}\mu = S(mt)\mu$, the first term may be made arbitrarily small. Since $\lim_{n\to\infty} J^n_{t/n}S(t)^{m-1}\mu = S(t)^m\mu$, so may the third term.

To bound the second term, note that by the lower semicontinuity of $|\partial E|$ ([1], Cor. 2.4.10) and (2.7),

$$|\partial E|(\mu(t)) \le \liminf_{n\to\infty} |\partial E|(J^n_{t/n}\mu) \le \liminf_{n\to\infty}(1 - \lambda^- t/n)^{-n}|\partial E|(\mu) = e^{\lambda^- t}|\partial E|(\mu). \tag{3.20}$$

Hence, $|\partial E|^2(S(t)^{m-1}\mu) \le e^{2(m-1)\lambda^- t}|\partial E|^2(\mu)$. Combining this with Corollary 3.2,

$$W_2^2(J^n_{t/n}S(t)^{m-1}\mu, J^n_{t/n}(J^n_{t/n})^{m-1}\mu)$$

$$\le (1 - \lambda^-(t/n))^{-2n}W_2^2(S(t)^{m-1}\mu, (J^n_{t/n})^{m-1}\mu) + \frac{n(t/n)^2 e^{2(m-1)\lambda^- t}}{(1 - \lambda^-(t/n))^{2n}}|\partial E|^2(\mu).$$

By the inductive hypothesis, $\lim_{n\to\infty}(J^n_{t/n})^{m-1}\mu = \lim_{n\to\infty} J^{n(m-1)}_{t(m-1)/n(m-1)}\mu = S((m-1)t)\mu = S(t)^{m-1}\mu$. Thus, we may choose n large enough to make the second term arbitrarily small.

Finally, we show the energy dissipation inequality (3.16). By (ii), it is enough to prove the result for $t_0 = 0$, $t_1 = t$. By inequality (2.7),

$$\tau(1 + \lambda\tau/2)|\partial E|^2(\mu_\tau) \le E(\mu) - E(\mu_\tau). \tag{3.21}$$

Define $g_n(s) = (1 + \lambda\tau/2)|\partial E|^2(J^i_\tau\mu)$ for $s \in ((i-1)\tau, i\tau]$, $1 \le i \le n$. Summing (3.21),

$$\int_0^{\tau n} g_n(s)\mathrm{d}s = \sum_{i=1}^n \tau(1 + \lambda\tau/2)|\partial E|^2(J^i_\tau\mu) \le E(\mu) - E(J^n_\tau\mu). \tag{3.22}$$

Let $\tau = t/n$. As in the proof of the previous theorem, $\lim_{n\to\infty} J^i_{t/n}\mu = S(s)\mu$. Taking the liminf of (3.22) and applying Fatou's lemma along with the lower semicontinuity of E and $|\partial E|$ ([1], Cor. 2.4.10) gives

$$\int_0^t |\partial E|^2(S(s)\mu) = \int_0^t \liminf_{n\to\infty} g_n(s)\mathrm{d}s \le E(\mu) - E(S(t)\mu).$$

The lower semicontinuity of E and the fact that $J^n_{t/n}\mu \in D(E)$ ensure $E(S(t)\mu) < +\infty$ for all $t \ge 0$. Then, the energy dissipation inequality implies that $|\partial E|^2(S(s)\mu) < +\infty$ for almost every $s \ge 0$. By (iii) and (3.20), $|\partial E|(S(t)\mu) \le e^{\lambda^-(t-s)}|\partial E|(S(s)\mu)$ for $0 < s < t$. Therefore, $S(t): \overline{D(E)} \to D(|\partial E|)$. $\qquad\square$

APPENDIX A. GENERALIZATIONS

A.1. Varying time steps

This section contains generalizations of the previous theorems to the case where we replace m time steps of size h with a sequence of varying time steps. For simplicity of notation, we write $J^m = \prod_{k=1}^m J_{h_k}\mu$, $J^n = J^n_\tau\mu$, $S_m = \sum_{k=1}^m h_k$, and $P_m = \prod_{k=1}^m (1 - \lambda^- h_k)^{-1}$.

In Theorems A.1−A.3 below, we suppose E satisfies convexity Assumption 2.11, $\mu \in D(|\partial E|)$, and $0 < h_i \le \tau < \frac{1}{\lambda^-}$. The first result is a generalization of Theorem 3.4, the second is a generalization of Lemma 3.5, and the third is a generalization of Theorem 3.6.

Theorem A.1.

$$(1 - \lambda^- h_m)^2 W_2^2(J^n, J^m) \leq \frac{h_m}{\tau}(1 - \lambda^- \tau)^{-1} W_2^2(J^{n-1}, J^{m-1}) + \frac{\tau - h_m}{\tau} W_2^2(J^{m-1}, J^n) + 2h_m^2 P_m^2 |\partial E|^2(\mu).$$

Lemma A.2. $W_2(J^m, \mu) \leq |\partial E(\mu)| S_m P_m.$

Theorem A.3. $W_2^2(J^n, J^m) \leq \left[(n\tau - S_m)^2 + \tau S_m + 2\tau^2 n\right](1 - \lambda^- \tau)^{-2n} P_m^2 |\partial E|^2(\mu).$

The proof of each of these results is a straightforward generalization of the previous proof. (Simply replace h with h_m, use $|\partial E|^2(J^m) \leq P_m^2 |\partial E|^2(\mu)$, and, in the third result, replace mh by S_m.)

We combine these results in the following theorem to prove the convergence of the discrete gradient flow with varying time steps.

Theorem A.4. *Suppose E satisfies convexity Assumption 2.11. For $\mu \in D(|\partial E|)$ and any partition of the interval $\{0 = t_0 < t_1 < \cdots < t_i < t_{i+1} < \ldots t_m = t\}$ corresponding to time steps $h_i = t_i - t_{i+1}$ the discrete gradient flow sequence sequence $\prod_{i=1}^m J_{h_i}\mu$ converges to the gradient flow $\mu(t)$ as $|\boldsymbol{h}| = \max_{1 \leq i \leq m} h_i \to 0$. The convergence is uniform in t on compact subsets of $[0, +\infty)$. When $|\boldsymbol{h}| \leq \frac{1}{2\lambda^-}$, the distance between the approximating sequence $\prod_{i=1}^m J_{h_i}\mu$ and the continuous gradient flow $S(t)\mu$ is bounded by*

$$W_2\left(S(t)\mu, \prod_{i=1}^m J_{h_i}\mu\right) \leq 2\left[|\boldsymbol{h}|^2 + 3|\boldsymbol{h}|t\right]^{1/2} e^{4\lambda^- t} |\partial E|(\mu)$$

Proof. Let $\tau = |\boldsymbol{h}|$, so $0 < h_k \leq \tau < \frac{1}{2\lambda^-}$, and let n be the greatest integer less than or equal to t/τ, so $t/\tau - 1 < n \leq t/\tau$, hence $t - \tau < n\tau \leq t$. By the triangle inequality, Theorem A.3, and the fact that $(1 - \alpha)^{-1} \leq e^{2\alpha}$ for $\alpha \in [0, 1/2]$,

$$W_2\left(S(t)\mu, \prod_{i=1}^m J_{h_i}\mu\right) \leq W_2(S(t)\mu, J_\tau^n \mu) + W_2\left(J_\tau^n \mu, \prod_{i=1}^m J_{h_i}\mu\right) = \lim_{l \to \infty} W_2(J_{t/l}^l \mu, J_\tau^n \mu) + W_2\left(J_\tau^n \mu, \prod_{i=1}^m J_{h_i}\mu\right)$$

$$\leq \lim_{l \to \infty}\left[(n\tau - t)^2 + \tau t + 2\tau^2 n\right]^{1/2}(1 - \lambda^- \tau)^{-n}(1 - \lambda^- t/l)^{-l}|\partial E|(\mu)$$

$$+ \left[(n\tau - t)^2 + \tau t + 2\tau^2 n\right]^{1/2}(1 - \lambda^- \tau)^{-n}\left\{\prod_{k=1}^m (1 - \lambda^- h_k)^{-1}\right\}|\partial E|(\mu)$$

$$\leq 2\left[|\boldsymbol{h}|^2 + 3|\boldsymbol{h}|t\right]^{1/2} e^{4\lambda^- t}|\partial E|(\mu) \qquad \square$$

A.2. Allowing $E(\mu) < +\infty$ when μ charges small sets

In this section, we give proofs of the Euler−Lagrange equation (Thm. 2.23) and the relation between proximal maps with small and large time steps (Thm. 3.3) in the general case, in which measures may give mass to small sets. In this context, optimal transport maps may no longer exist, so the transport metrics (Def. 2.14) may no longer be well-defined.

Theorem A.5 (Euler−Lagrange equation). *Suppose E satisfies convexity Assumption 2.11. Then for $\omega \in \overline{D(E)}$, $0 < \tau < \frac{1}{\lambda^-}$, μ is the unique minimizer of the quadratic perturbation $\Phi(\tau, \omega; \cdot)$ if and only if for all $\gamma \in \Gamma_0(\mu, \omega)$, defining $\gamma_\tau = \rho_\tau \# \gamma$, $\rho_\tau(x_1, x_2) = (x_1, (x_2 - x_1)/\tau)$, we have*

$$E(\nu) - E(\mu) \geq \int \langle x_2, x_3 - x_1 \rangle d\gamma + o\left(\|x_1 - x_3\|_{L^2(\gamma)}\right) \quad \forall \gamma \in \Gamma(\gamma_\tau, \nu), \quad \nu \in D(E), \quad \nu \to \mu. \qquad (A.1)$$

Proof. ([1], Lem. 10.3.4) shows that if μ is the unique minimizer of $\Phi(\tau, \omega; \cdot)$, then (A.1) holds for all $\gamma \in \Gamma_0(\mu, \omega)$.

We now prove the converse. Suppose that for all $\gamma \in \Gamma_0(\mu, \omega)$, (A.1) holds, and fix $\nu \in D(E)$. There exists some generalized geodesic μ_α from μ to ν with base ω along which E is λ-convex. Let $\boldsymbol{\omega}$ be the plan that induces

this generalized geodesic, with $\pi^{1,2}\#\boldsymbol{\omega} \in \Gamma_0(\omega,\mu)$ and $\pi^{1,3}\#\boldsymbol{\omega} \in \Gamma_0(\omega,\nu)$, so $\mu_\alpha = [(1-\alpha)\pi^2 + \alpha\pi^3]\#\boldsymbol{\omega}$. Applying (A.1) with $\gamma = \pi^{2,1}\#\boldsymbol{\omega}$ and $\nu = \mu_\alpha$ shows

$$E(\mu_\alpha) - E(\mu) \geq \int \langle x_2, x_3 - x_1\rangle \mathrm{d}\boldsymbol{\gamma}_\alpha + o\left(\|x_1 - x_3\|_{L^2(\boldsymbol{\gamma}_\alpha)}\right) \quad \forall\boldsymbol{\gamma}_\alpha \in \Gamma(\gamma_\tau, \mu_\alpha), \mu_\alpha \to \mu. \tag{A.2}$$

Since $(x_2, \frac{x_1 - x_2}{\tau}, (1-\alpha)x_2 + \alpha x_3)\#\boldsymbol{\omega} \in \Gamma(\gamma_\tau, \mu_\alpha)$,

$$E(\mu_\alpha) - E(\mu) \geq \int \langle (x_1 - x_2)/\tau, ((1-\alpha)x_2 + \alpha x_3) - x_2\rangle \, \mathrm{d}\boldsymbol{\omega} + o\left(\|x_2 - (1-\alpha)x_2 - \alpha x_3\|_{L^2(\boldsymbol{\omega})}\right)$$

$$= \alpha \int \langle (x_1 - x_2)/\tau, x_3 - x_2\rangle \, \mathrm{d}\boldsymbol{\omega} + \alpha o\left(\|x_2 - x_3\|_{L^2(\boldsymbol{\omega})}\right).$$

By definition of convexity along a generalized geodesic, $E(\nu) - E(\mu) \geq \frac{1}{\alpha}\left[E(\mu_\alpha) - E(\mu)\right] + (1-\alpha)\frac{\lambda}{2}\|x_2 - x_3\|_{L^2(\boldsymbol{\omega})}$. Using the above inequality, we may bound this from below,

$$E(\nu) - E(\mu) \geq \int \langle (x_1 - x_2)/\tau, x_3 - x_2\rangle \, \mathrm{d}\boldsymbol{\omega} + o(1) + (1-\alpha)(\lambda/2)\|x_2 - x_3\|_{L^2(\boldsymbol{\omega})}.$$

Sending $\alpha \to 0$,

$$E(\nu) - E(\mu) \geq \int \langle (x_1 - x_2)/\tau, x_3 - x_2\rangle \, \mathrm{d}\boldsymbol{\omega} + (\lambda/2)\|x_2 - x_3\|_{L^2(\boldsymbol{\omega})}. \tag{A.3}$$

Likewise, we have

$$W_2^2(\nu, \omega) - W_2^2(\mu, \omega) = \int |x_1 - x_3|^2 \mathrm{d}\boldsymbol{\omega} - \int |x_1 - x_2|^2 \mathrm{d}\boldsymbol{\omega} = \|x_2 - x_3\|_{L^2(\boldsymbol{\omega})}^2 + 2\int \langle x_3 - x_2, x_2 - x_1\rangle \mathrm{d}\boldsymbol{\omega}.$$

Combining this with the fact that $\lambda + 1/\tau > 0$,

$$\Phi(\tau, \omega; \nu) - \Phi(\tau, \omega; \mu) \geq \int \langle (x_3 - x_1)/\tau + (x_1 - x_3)/\tau, x_2 - x_1\rangle \mathrm{d}\boldsymbol{\omega} = 0$$

Since $\nu \in D(E)$ was arbitrary, μ minimizes $\Phi(\tau, \omega; \cdot)$.   □

We now turn to the proof of Theorem 3.3. For simplicity of notation, we abbreviate $J_\tau\mu$ by μ_τ.

Theorem A.6. *Suppose E satisfies convexity Assumption 2.11. Then if $\mu \in \overline{D(E)}$ and $0 < h \leq \tau < \frac{1}{\lambda^-}$,*

$$J_\tau\mu = J_h\left[\mu_{\frac{h}{\tau}}^{\mu_\tau \to \mu}\right],$$

where $\mu_{\frac{h}{\tau}}^{\mu_\tau \to \mu}$ is any geodesic from μ_τ to μ at time $\frac{h}{\tau}$.

Proof. Choose any geodesic $\mu_\alpha^{\mu_\tau \to \mu}$ from μ_τ to μ, and define $\omega = \mu_{\frac{h}{\tau}}^{\mu_\tau \to \mu}$. We must show $\mu_\tau = \omega_h$.

By ([1], Lem. 7.2.1), there exists a unique plan $\gamma^{\mu_\tau \to \omega} \in \Gamma_0(\mu_\tau, \omega)$ and there exists $\gamma \in \Gamma_0(\mu_\tau, \mu)$ so that $\gamma^{\mu_\tau \to \omega} = (\frac{\tau - h}{\tau}\pi^{1,1} + \frac{h}{\tau}\pi^{1,2})\#\gamma$. By Theorem A.5, for all $\gamma \in \Gamma_0(\mu_\tau, \mu)$,

$$E(\nu) - E(\mu_\tau) \geq \int \langle x_2, x_3 - x_1\rangle \mathrm{d}\boldsymbol{\gamma} + o\left(\|x_1 - x_3\|_{L^2(\boldsymbol{\gamma})}\right) \quad \forall\boldsymbol{\gamma} \in \Gamma(\gamma_\tau, \nu), \nu \to \mu_\tau. \tag{A.4}$$

To prove $\mu_\tau = \omega_h$, by a second application of Theorem A.5, it's enough to show that for all $\tilde{\gamma} \in \Gamma_0(\mu_\tau, \omega)$,

$$E(\tilde{\nu}) - E(\mu_\tau) \geq \int \langle x_2, x_3 - x_1\rangle d\tilde{\boldsymbol{\gamma}} + o\left(\|x_1 - x_3\|_{L^2(\tilde{\boldsymbol{\gamma}})}\right) \quad \forall\tilde{\boldsymbol{\gamma}} \in \Gamma(\tilde{\gamma}_h, \tilde{\nu}), \tilde{\nu} \to \mu_\tau. \tag{A.5}$$

Since $\gamma^{\mu_\tau \to \omega}$ is the unique plan in $\Gamma_0(\mu_\tau, \omega)$, it is enough to show that (A.5) holds for $\tilde{\gamma} = \gamma^{\mu_\tau \to \omega}$.

As (A.4) holds for all $\gamma \in \Gamma_0(\mu_\tau, \mu)$, in particular, it holds for γ so that $\gamma^{\mu_\tau \to \omega} = (\frac{\tau-h}{\tau}\pi^{1,1} + \frac{h}{\tau}\pi^{1,2})\#\gamma$. Furthermore,

$$\rho_h\left(x_1, \frac{\tau-h}{\tau}x_1 + \frac{h}{\tau}x_2\right) = \left(x_1, \frac{1}{h}\left[\frac{\tau-h}{\tau}x_1 + \frac{h}{\tau}x_2 - x_1\right]\right) = \left(x_1, \frac{x_2-x_1}{\tau}\right) = \rho_\tau(x_1, x_2).$$

Consequently, $\tilde\gamma_h = \rho_h\#\gamma^{\mu_\tau \to \nu} = \rho_h \circ (\frac{\tau-h}{\tau}\pi^{1,1} + \frac{h}{\tau}\pi^{1,2})\#\gamma = \rho_\tau\#\gamma = \gamma_\tau$. Therefore, the fact that (A.4) holds for γ implies that (A.5) holds for $\tilde\gamma$. This proves the result. $\qquad\square$

References

[1] L. Ambrosio, N. Gigli, and G. Savaré, Gradient Flows in Metric Spaces and in the Space of Probability Measures. *Lect. Math. ETH Zürich 2nd edition*. Birkhäuser Verlag, Basel (2008).

[2] H. Brézis, Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert. *North-Holland Mathematics Studies*, No. 5. Notas de Matemática (50). North-Holland Publishing Co., Amsterdam (1973).

[3] H. Brézis and A. Pazy, Convergence and approximation of semigroups of nonlinear operators in Banach spaces. *J. Functional Analysis* **9** (1972) 63–74.

[4] E.A. Carlen and K. Craig, Contraction of the proximal map and generalized convexity of the Moreau−Yosida regularization in the 2-Wasserstein metric. *Math. Mech. Complex Systems* **1** (2013) 33–65.

[5] J.A. Carrillo, M. Di Francesco, A. Figalli, T. Laurent, and D. Slepčev, Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations. *Duke Math. J.* **156** (2011) 229–271.

[6] J. A. Carrillo, M. Di Francesco, A. Figalli, T. Laurent and D. Slepčev, Confinement in nonlocal interaction equations. *Nonlin. Anal.* **75** (2012) 550–558.

[7] Ph. Clément and W. Desch, A Crandall-Liggett approach to gradient flows in metric spaces. *J. Abstr. Differ. Equ. Appl.* **1** (2010) 46–60.

[8] M.G. Crandall, Semigroups of nonlinear transformations in Banach spaces. In *Contributions to nonlinear functional analysis Proc. of Sympos., Math. Res. Center, Univ. Wisconsin, Madison.* Publ. Math. Res. Center Univ. Wisconsin, No. 27. Academic Press, New York (1971) 157–179.

[9] M.G. Crandall and T.M. Liggett, Generation of semi-groups of nonlinear transformations on general Banach spaces. *Amer. J. Math.* **93** (1971) 265–298.

[10] N. Gigli, On the inverse implication of Brenier-McCann theorems and the structure of $(P_2(M), W_2)$. *Methods Appl. Anal.* **18** (2011) 127–158.

[11] J. Jost, Convex functionals and generalized harmonic maps into spaces of nonpositive curvature. *Comment. Math. Helv.* **70** (1995) 659–673.

[12] U.F. Mayer, Gradient flows on nonpositively curved metric spaces and harmonic maps. *Comm. Anal. Geom.* **6** (1998) 199–253.

[13] R.J. McCann, Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **80** (1995) 309–323.

[14] F. Otto, Doubly degenerate diffusion equations as steepest descent, manuscript (1996).

[15] F. Otto, The geometry of dissipative evolution equations: the porous medium equation. *Commun. Partial Differ. Equ.* **26** (2001) 101–174.

[16] S. Rasmussen, Non-linear Semi-Groups, Evolution Equations and Product Integral Representations. Aarhus Universitet (1971).

[17] J. Rulla, Error analysis for implicit approximations to solutions to Cauchy problems. *SIAM J. Numer. Anal.* **33** (1996) 68–87.

[18] K. Yosida, Functional Analysis. *Classics in Mathematics.* Springer-Verlag, Berlin (1995). Reprint of the sixth edition (1980).