THE EXPONENTIAL FORMULA FOR THE WASSERSTEIN METRIC

BY KATY CRAIG

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Mathematics

Written under the direction of

Eric Carlen

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2014

ABSTRACT OF THE DISSERTATION

The Exponential Formula for the Wasserstein Metric

by Katy Craig

Dissertation Director: Eric Carlen

Many evolutionary partial differential equations may be rewritten as the gradient flow of an energy functional, a perspective which provides useful estimates on the behavior of solutions. The notion of gradient flow requires both the specification of an energy functional and a metric with respect to which the gradient is taken. In recent years, there has been significant interest in gradient flow on the space of probability measures endowed with the Wasserstein metric. The notion of gradient in this setting in purely formal and rigorous analysis of the gradient flow typically considers a time discretization of the problem known as the *discrete gradient flow*. In this dissertation, we adapt Crandall and Liggett's Banach space method to give a new proof of the exponential formula, quantifying the rate at which solutions to the discrete gradient flow converge to solutions of the gradient flow. In the process, we use a new class of metrics—*transport metrics*—that have stronger convexity properties than the Wasserstein metric to prove an Euler-Lagrange equation characterizing the discrete gradient flow. We then apply these results to give simple proofs of properties of the gradient flow, including the contracting semigroup property and the energy dissipation inequality.

Acknowledgements

First, I thank my advisor, Eric Carlen. Throughout my Ph.D., I have come up against problems, mathematical and otherwise, that I didn't think I could solve. Each time, Eric has always seemed to have complete confidence that I would solve them and was always happy to discuss them with me and provide many helpful suggestions. I decided that if Eric thought I could solve a problem, I probably could, and this helped me always keep trying. Thank you, Eric, for your unwavering support and optimism. I am incredibly lucky to have had you as my advisor.

I also thank another important mentor, Andrea Bertozzi, who invited me to spend the spring and summer of my third year of graduate school at UCLA. Andrea introduced me to many interesting problems in applied mathematics and numerical analysis, and I look forward to going back to UCLA this fall.

I am very honored that Prof. Wilfrid Gangbo, Prof. Zheng-Chao Han, and Prof. Natasa Sesum agreed to serve on my thesis committee, and I am grateful for their suggestions and thought-provoking questions. I also thank Prof. Han the for the welcoming environment he nurtured as graduate director.

I thank Prof. Giuseppe Savaré for suggesting the problem that became this thesis and for his excellent lecture notes [25], which inspired my presentation of the Hilbert space a priori estimates in section 1.1 I also thank Prof. Luigi Ambrosio, Prof. Alessio Figalli, Prof. José Carrillo, and Prof. Ryan Hynd for many helpful conversations which greatly improved this work.

I thank my math sister, Shabnam Beheshti, and my wonderful friends at Rutgers, including Elena Caffarelli, Wes Cowan, Manuel Larenas, Glen Wilson, the members of the graduate analysis seminar, and everyone who came to the math women's tea. I thank the administrative staff in the math department, particularly Colette Claiborne,

Hope DiCapua, and Katie Guarino. I also thank Kathy Smalley and Eric Zolt for their friendship and guidance while I was living in Los Angeles.

I thank my mother, Shirley, for supporting me completely, whether I am an aspiring Broadway star or aspiring mathematician. I thank my father, Stuart, for teaching me by example to discover what goals matter to me and pursue them wholeheartedly. I thank my sisters, Ginny and Maisy, for making all of our time together like a nineties themed slumber party.

Finally, I thank my husband Nathaniel. Without him, I wouldn't have dreamed to try to get my Ph.D. in mathematics, and I certainly would have failed the contour integration portion of my written qualifying exam. Thank you for showing me what it means to be a great scientist, a generous human being, and a very fast cyclist. I look forward to where physics and math take us next.

Table of Contents

Chapter 1

Introduction: Gradient Flow on Hilbert Spaces

The theory of gradient flow provides a variational perspective for studying a wide variety of partial differential equations, contributing estimates relevant to existence, regularity, uniqueness, and stability of solutions. In general, a partial differential equation is a *gradient flow* of an energy functional $E : X \rightarrow (-\infty, +\infty]$ on a metric space (X, d) if the equation may be rewritten as

$$\frac{d}{dt} u(t) = -\nabla_d E(u(t)) , \quad u(0) = u \in X ,\tag{1.1}$$

for a generalized notion of gradient ∇_d.

A classical example of gradient flow is the heat equation, which is the gradient flow on $L^2(\Omega)$ of the Dirichlet energy

$$E(u) := \begin{cases} \frac{1}{2} \int_\Omega |\nabla u(x)|^2 dx & \text{if } u \in H_0^1(\Omega) \\ +\infty & \text{otherwise.} \end{cases}\tag{1.2}$$

Though E is not continuous with respect to the $L^2(\Omega)$ norm, even when restricted to $\mathcal{G} := H_0^1(\Omega) \cap H^2(\Omega)$, its directional derivatives are well defined for all $u \in \mathcal{G}$ by

$$\lim_{h \to 0} \frac{E(u + hv) - E(u)}{h} = \int \nabla u \cdot \nabla v = -\int \Delta u v \quad \forall v \in C_c^\infty(\Omega) .\tag{1.3}$$

Using this, we define $\nabla_{L^2} E(u)$ to be the vector field satisfying

$$(\nabla_{L^2} E(u), v)_{L^2(\Omega)} = \lim_{h \to 0} \frac{E(u + hv) - E(u)}{h} = (-\Delta u, v)_{L^2(\Omega)} \quad \forall v \in L^2(\Omega) .\tag{1.4}$$

Therefore, the gradient flow of the Dirichlet energy is the heat equation,

$$\frac{d}{dt} u(t) = -\nabla_{L^2} E(u(t)) = \Delta u(t) .$$

By linearity of the gradient and convexity of $|\cdot|^2$, the Dirichlet energy is convex on $L^2(\Omega)$. One of the benefits of the gradient flow perspective is that, simply by writing

a PDE as the gradient flow of a *convex* energy functional, one automatically obtains a variety of a priori estimates. (See section 1.1). A second benefit is that a time discretization of the gradient flow problem provides a simple variational scheme for approximating solutions.

In the case of the heat equation, this time discretization is given by replacing the time derivative with a backward finite difference of time step τ. Specifically, we define the *discrete gradient flow sequence*

$$\frac{u_n - u_{n-1}}{\tau} = \Delta u_n \ , \quad u_0 = u \in L^2(\mathbb{R}^d) \ , \quad u_n|_{\partial\Omega} = 0 \ . \tag{1.5}$$

This implicit definition of u_n can be made explicit by recognizing that, in analogy with Dirichlet's principle for the Poisson equation, u_n is characterized as the minimizer over $H_0^1(\Omega)$ of

$$\Phi(v) := \frac{1}{2\tau}\int_\Omega |v(x) - u_{n-1}(x)|^2 dx + \frac{1}{2}\int_\Omega |\nabla v(x)|^2 dx \ . \tag{1.6}$$

At first, the need for this second variational characterization of u_n may seem like a drawback of this discretization method. After all, one could replace (1.5) with the explicit scheme

$$\frac{u_n - u_{n-1}}{\tau} = \Delta u_{n-1} \ , \quad u_0 = u \in L^2(\mathbb{R}^d) \ , \quad u_n|_{\partial\Omega} = 0 \ ,$$

so that we have the explicit formula $u_n := u_{n-1} + \tau\Delta u_{n-1}$. However, the variational characterization of u_n (1.6) causes a *gain* in regularity along the discrete gradient flow sequence—given $u_0 \in L^2(\Omega)$, $u_n \in H_0^1(\Omega)$ for all $n > 0$—whereas the explicit scheme corresponds to a *decrease* in regularity. Since solutions of the gradient flow (in this case, solutions of the heat equation) likewise exhibit a gain in regularity for $t > 0$, the implicit method is the best choice of time discretization. Furthermore, one may show that, as the time step goes to zero, the discrete gradient flow sequence converges to the gradient flow. Specifically, if we define $\tau := \frac{t}{n}$,

$$\lim_{\tau \to \infty} u_n = u(t) \ . \tag{1.7}$$

Equation (1.7) is known as the *exponential formula*. If we rewrite (1.5) as

$$u_n := (\mathbf{id} - \tau\Delta)^{-1}u_{n-1} = (\mathbf{id} - \tau\Delta)^{-n}u_0 \ ,$$

(1.7) is an infinite dimensional generalization of the limit definition of the exponential

$$\lim_{n \to \infty} (\mathbf{id} - \frac{t}{n}\Delta)^{-n} u_0 = u(t) \ .$$

The importance of the variational characterization of the discrete gradient flow (1.6) becomes even clearer when one seeks to extend the notion gradient flow to other metrics (X, d) and energy functionals E. In general, the minimization problem which defines the discrete gradient flow is of the form

$$u_n := \operatorname*{argmin}_{v \in X} \left\{ \frac{1}{2\tau} d(v, u_{n-1})^2 + E(v) \right\} \ . \tag{1.8}$$

Such a minimization problem may be posed for any metric and functional, but in order for it to have a unique solution one needs the sum of the square distance and the energy functional to be sufficiently convex. This balance between convexity of the square distance and convexity of the energy functional is a recurring theme in the theory of gradient flow. Once one can show that (1.8) is well-posed, it is also useful to have a characterization of the minimizer u_n in terms of an Euler Lagrange equation. For the heat equation example above, this simply corresponds to (1.5), but in a general metric space that lacks a vector space structure, one will seek to choose the right generalized notion of gradient ∇_d so that a version of (1.5) holds.

In this thesis, we consider gradient flow in the *Wasserstein metric*, a metric on the space of probability measures that shares many properties with the $L^2(\Omega)$ norm. (One may even define a formal notion of inner product—see section 2.7.) We prove an Euler-Lagrange equation characterizing the discrete gradient flow, in analogy with (1.5) and (1.6), and use this, along with other new results, to give a new proof of the exponential formula in the Wasserstein metric.

Before turning to the Wasserstein metric, we first provide an overview of the classical theory of gradient flow on a Hilbert space. We will generalize these concepts to the Wasserstein case in chapter 2.

1.1 Hilbert Space Gradient Flow

The classical theory of gradient flow on a Hilbert space can be seen both as a nonlinear, infinite dimensional generalization of the theory of ordinary differential equations and as a particular case of the theory of monotone operators [3,4,22]. We will begin with the first perspective, in order to introduce a priori estimates associated with the gradient flow, and then switch to the second perspective, in order to adapt to the low regularity of energy functionals E often required by applications in PDE.

Table 1.1: Examples of PDEs that are gradient flows on a Hilbert space.
Suppose $m > 0$ and $F : \mathbb{R}^d \to [0, +\infty)$ is continuously differentiable, convex, and normalized so $F(0) = 0$.

	PDE	Energy Functional	Metric
Allen-Cahn	$\frac{d}{dt}u = \Delta u - F'(u)$	$E(u) = \frac{1}{2}\int \left[|\nabla u|^2 + F(u)\right]$	L^2
Cahn-Hilliard	$\frac{d}{dt}u = \Delta(\Delta u - F'(u))$	$E(u) = \frac{1}{2}\int \left[|\nabla u|^2 + F(u)\right]$	H^{-1}
Porous Media	$\frac{d}{dt}u = \Delta u^m$	$E(u) = \frac{1}{m+1}\int u^{m+1}$	H^{-1}

Given a functional $E : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ on a Hilbert space $\mathcal{H}$, we define the Hilbert space gradient in analogy with the definition of ∇_{L^2} from the previous section.

DEFINITION 1.1.1 (Hilbert space gradient). $\nabla_{\mathcal{H}} E(u) \in \mathcal{H}$ is the Hilbert space gradient of E at u if

$$(\nabla_{\mathcal{H}} E(u), v) = \lim_{h \to 0} \frac{E(u + hv) - E(u)}{h} \quad \forall v \in \mathcal{H} \ . \tag{1.9}$$

REMARK 1.1.2 (L^2 gradient and functional derivative). For a wide class of integral functionals $E(u) = \int F(x, u(x), \nabla u(x))dx$ on $L^2(\Omega)$, if u is sufficiently regular, the *functional derivative* $\frac{\delta E}{\delta u} \in L^2(\Omega)$ exists and satisfies

$$\lim_{h \to 0} \frac{E(u + hv) - E(u)}{h} = \int \frac{\delta E}{\delta u} v \quad \forall v \in C_c^\infty(\Omega) \ . \tag{1.10}$$

In this case, we may identify $\nabla_{L^2} E(u)$ with $\frac{\delta E}{\delta u}$.

With this definition of Hilbert space gradient, we may now define the Hilbert space gradient flow.

DEFINITION 1.1.3 (Hilbert space gradient flow)**.** The gradient flow of E with respect to $\mathcal{H}$ is

$$\frac{d}{dt}u(t) = -\nabla_{\mathcal{H}}E(u(t)) \ , \quad u(0) = u \ . \tag{1.11}$$

Part of the utility of gradient flow in partial differential equations is that the gradient flow structure provides several a priori estimates. We now give a formal argument for why these estimates hold for solutions of (1.11). For the purposes of these formal computations, we suppose there is a subspace $\mathcal{G} \subseteq \mathcal{H}$ such that for all $u \in \mathcal{G}$, $\nabla_{\mathcal{H}}E(u)$ exists. We also suppose that $u(t) \in C^1([0, +\infty), \mathcal{G})$, so that we may consider (1.11) in a strong sense.

In addition to these regularity assumptions, which allow us to defer questions of differentiability to the following section, we also assume that E is λ-convex:

DEFINITION 1.1.4 (λ-convex)**.** Given $\lambda \in \mathbb{R}$, functional $E : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is λ-convex in case for all $u, v \in \mathcal{H}$, $\alpha \in [0, 1]$.

$$E(\alpha v + (1 - \alpha)u) \leq \alpha E(v) + (1 - \alpha)E(u) - \alpha(1 - \alpha)\frac{\lambda}{2}|u - v|^2 \ . \tag{1.12}$$

REMARK 1.1.5. Since $|\alpha v + (1 - \alpha)u|^2 = \alpha|v|^2 + (1 - \alpha)|u|^2 + \alpha(1 - \alpha)|u - v|^2$, E is λ-convex if and only if $E - \lambda|\cdot|^2/2$ is convex. In particular, E is convex in the usual sense if E is λ-convex for $\lambda \geq 0$.

Rearranging (1.12) gives

$$E(v) - E(u) \geq \frac{1}{\alpha}\left(E(\alpha v + (1 - \alpha)u) - E(u)\right) + (1 - \alpha)\frac{\lambda}{2}|u - v|^2 \ . \tag{1.13}$$

Therefore, sending $\alpha \to 0$, we obtain the following inequality for the gradient of a λ-convex function:

$$E(v) - E(u) \geq (\nabla_{\mathcal{H}}E(u), v - u) + \frac{\lambda}{2}|u - v|^2 \ . \tag{1.14}$$

Furthermore, by interchanging the roles of u and v in (1.14) and adding the two inequalities together, we obtain the monotonicity property of the gradient,

$$(\nabla E(v) - \nabla E(u), v - u) \geq \frac{\lambda}{2}|v - u|^2 \ . \tag{1.15}$$

With these definitions and basic properties, we may now prove five key estimates for the gradient flow. The first three estimates quantify the rate at which a Lyapunov functional $F : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ decays along the gradient flow. In each case, we assume that $F(u(t))$ is continuously differentiable in time and

$$\frac{d}{dt}F(u(t)) = \left(\nabla_{\mathcal{H}} F(u(t)), \frac{d}{dt}u(t) \right) = - \left(\nabla_{\mathcal{H}} F(u(t)), \nabla_{\mathcal{H}} E(u(t)) \right) .$$

The fourth and fifth estimates quantify the regularity and stability of the flow. To simplify notation, we write ∇ for $\nabla_{\mathcal{H}}$.

1. **Energy Dissipation Identity:** $E(u(t_0)) - E(u(t_1)) = \int_{t_0}^{t_1} |\nabla E(u(s))|^2 \, ds$

    Take $F(u) = E(u)$ to obtain

    $$\frac{d}{dt}E(u(t)) = - |\nabla E(u(t))|^2 \ , \tag{1.16}$$

    and then integrate in time.

2. **Evolution Variational Inequality:**

    $$\frac{d}{dt}\frac{1}{2}|u(t) - w|^2 \le E(w) - E(u(t)) - \frac{\lambda}{2}|u(t) - w|^2$$

    Take $F(u) = \frac{1}{2}|u - w|^2$ to obtain

    $$\frac{d}{dt}\frac{1}{2}|u(t) - w|^2 = - (u(t) - w, \nabla E(u(t))) \le^{(1.14)} E(w) - E(u(t)) - \frac{\lambda}{2}|u(t) - w|^2$$

    In particular, if E attains its minimum at $\bar{u}$, the evolution variational inequality provides an upper bound for the distance between the solution of the gradient flow and the minimizer:

    $$|u(t) - \bar{u}|^2 \le e^{-\lambda t}|u(0) - \bar{u}|^2$$

3. **Exponential Decay of Gradient:** $|\nabla E(u(t)| \le e^{-\lambda t}|\nabla E(u(0))|$

    For this estimate, we impose the additional assumption that for all $u \in \mathcal{G}$, there exists $D^2 E(u) : \mathcal{G} \to \mathcal{H}$ satisfying

    $$(D^2 E(u)v, w) = \lim_{h \to 0} \frac{(\nabla E(u + hv) - \nabla E(u), w)}{h} \quad \forall v \in \mathcal{G} .$$

By inequality (1.15),

$$(D^2 E(u)v, v) = \lim_{h \to 0} \frac{(\nabla E(u + hv) - \nabla E(u), hv)}{h^2} \geq \lim_{h \to 0} \frac{\lambda}{2} \frac{|hv|^2}{h^2} = \frac{\lambda}{2}|v|^2 \ .$$

Therefore, taking $F(u) = \frac{1}{2}|\nabla E(u)|^2$,

$$\frac{d}{dt} \frac{1}{2}|\nabla E(u(t))|^2 = -(D^2 E(u(t))\nabla E(u(t)), \nabla E(u(t))) \leq -\frac{\lambda}{2}|\nabla E(u(t))|^2 \ . \quad (1.17)$$

Integrating gives the result.

4. **Instantaneous Regularization for** $\lambda \geq 0$: For all $t > 0$, $E(u(t)) \leq \frac{1}{2t}|u(0) - w|^2 + E(w)$.

For all $t > 0$ and $w \in \mathcal{G}$, $|\nabla E(u(t))|^2 \leq \frac{1}{t^2}|u(0) - w|^2 + |\nabla E(w)|^2$.

Suppose $\lambda \geq 0$. Combining the differential forms of the first three estimates, we obtain

$$\frac{d}{dt} t[E(u(t)) - E(w)] =^{(\textbf{1. EDI})} E(u(t)) - E(w) - t|\nabla E(u(t))|^2$$

$$\frac{d}{dt} \frac{1}{2}|u(t) - w|^2 \leq^{(\textbf{2. EVI})} E(w) - E(u(t))$$

$$\frac{d}{dt} \frac{t^2}{2}|\nabla E(u(t))|^2 \leq^{(\textbf{3. EDG})} t|\nabla E(u(t))|^2$$

Adding these three inequalities and integrating from 0 to t, we obtain for all $t > 0$,

$$t[E(u(t)) - E(w)] + \frac{1}{2}|u(t) - w|^2 + \frac{t^2}{2}|\nabla E(u(t))|^2 \leq \frac{1}{2}|u(0) - w|^2 \quad (1.18)$$

Furthermore, if $\nabla E(w)$ is well defined, by (1.14) and Cauchy's inequality, we have

$$E(u(t)) - E(w) \geq -\frac{1}{2}\left[t|\nabla E(w)|^2 + \frac{1}{t}|u(t) - w|^2\right] \ .$$

Therefore, (1.18) implies the above instantaneous regularization estimates.

5. **Contraction Inequality**: $|u(t) - v(t)| \leq e^{-\lambda t}|u(0) - v(0)|$

Given solutions to the gradient flow $u(t)$ and $v(t)$,

$$\frac{d}{dt} \frac{1}{2}|u(t) - v(t)|^2 = -(u(t) - v(t), \nabla E(u(t)) - \nabla E(v(t))) \leq^{(1.15)} \frac{\lambda}{2}|u(t) - v(t)|^2 \ .$$

Integrating gives the result.

1.2 Differentiability

While the Hilbert space gradient from Definition 1.1.1 is extremely useful as a heuristic tool, it does not correspond precisely to the notion of gradient from Euclidean space. In particular, the Hilbert space gradient is well-defined even if E is only Gâteaux differentiable, allowing the possibility of functionals—like the Dirichlet energy (1.2)—which are not continuous but do have a gradient.

In order to have a notion of gradient that includes functionals with low regularity, like the Dirichlet energy, and also preserves essential properties when passing from finite dimensional to infinite dimensional Hilbert spaces, it is most common to work with a generalization of the gradient known as the *subdifferential*. Recall that a functional E is *proper* if $D(E) = \{u \in \mathcal{H} : E(u) < +\infty\} \neq \emptyset$ and is *lower semicontinuous* if its sub-level sets are closed.

DEFINITION 1.2.1 (subdifferential). Given $E : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ proper and lower semicontinuous, ξ belongs to the subdifferential of E at $u \in D(E)$ in case

$$E(v) - E(u) \geq (\xi, v - u) + o(|v - u|) \text{ as } v \to u .$$

We write $\xi \in \partial E(u)$. Let $D(\partial E) := \{u \in \mathcal{H} : \partial E(u) \neq \emptyset\}$ and $|\partial E(u)| := \min_{\xi \in \partial E(u)} |\xi|$.

REMARK 1.2.2 (subdifferential and gradient). If $\nabla_{\mathcal{H}} E(u)$ exists, $\partial E(u) = \{\nabla_{\mathcal{H}} E(u)\}$.

LEMMA 1.2.3 (subdifferential of a convex function). *If $E : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is proper, lower semicontinuous, and λ-convex, $\xi \in \partial E(u)$ if and only if*

$$E(v) - E(u) \geq (\xi, v - u) + \frac{\lambda}{2}|u - v|^2 \quad \forall v \in \mathcal{H} . \tag{1.19}$$

Proof. If (1.19) holds then $\xi \in \partial E(u)$ by definition of the subdifferential. For the other direction, note that as $\alpha \to 0$ in the right hand side of (1.13), the definition of the subdifferential ensures that the right hand side is bounded below by

$$(\xi, v - u) + \frac{\lambda}{2}|u - v|^2 .$$

$\square$
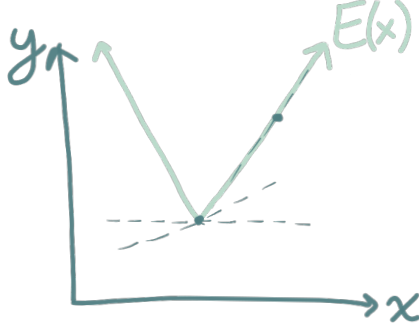
Figure 1.1: When E is convex, the elements of the subdifferential correspond to the supporting hyperplanes of E. If E is differentiable at a point, there is a unique supporting hyperplane through that point which corresponds to the gradient.

As a consequence of Lemma 1.2.3, we have the following corollary.

COROLLARY 1.2.4. *Given $E : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, and λ-convex for $\lambda \geq 0$, $0 \in \partial E(u)$ if and only if u is a minimizer of E.*

REMARK 1.2.5 (subdifferential of a convex function as monotone operator)**.** Lemma 1.2.3 is the subdifferential analogue of inequality (1.14) for the Hilbert space gradient of a λ-convex function. As before, if we interchange the roles of u and v in (1.19) and add the inequalities together, we obtain

$$(\xi - \tilde{\xi}, u - v) \geq \lambda |u - v|^2 , \quad \forall \xi \in \partial E(u), \tilde{\xi} \in \partial E(v). \tag{1.20}$$

Any (possibly multivalued) mapping $A : \mathcal{H} \to \mathcal{H}$ that satisfies

$$(f_1 - f_2, u_1 - u_2) \geq 0 , \quad \forall f_1 \in Au_1, f_2 \in Au_2$$

is called a *monotone operator*, and (1.20) shows that when E is λ-convex, $\partial E - \lambda \mathbf{id}$ is a monotone operator. Consequently, choosing the subdifferential as our notion of generalized gradient allows us to apply the rich theory of monotone operators to our study of Hilbert space gradient flow [3, 4, 22].

We now define the *gradient flow* of $E : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ in terms of the subdifferential.

DEFINITION 1.2.6 (Hilbert space gradient flow, subdifferential)**.** u is a solution of the gradient flow of E with respect to $\mathcal{H}$ if u is differentiable a.e. and

$$-\frac{d}{dt}u(t) \in \partial E(u(t)) \ \text{ a.e. } t \geq 0, \quad u(0) = u_0 . \tag{1.21}$$

With this definition of gradient flow, we have the following classical theorem.

THEOREM 1.2.7 (Hilbert Space Gradient Flow [3,22]). *Suppose* $E : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ *is proper, lower semicontinuous, and* λ-*convex. Given initial conditions* $u_0 \in \overline{D(E)}$, *there exists* $u : [0, \infty) \to \mathcal{H}$, *continuous for* $t \geq 0$ *and locally Lipschitz continuous for* $t > 0$, *which is the unique solution of*

$$-\frac{d}{dt}u(t) \in \partial E(u(t)) \quad a.e. \ t \geq 0 \ , \qquad u(0) = u_0 \ .$$

Furthermore, the following inequalities hold:

1. *Energy Dissipation Inequality: For* $t_1, t_0 \geq 0$,
   $\int_{t_0}^{t_1} |\partial E(u(s))|^2 \, ds \leq E(u(t_0)) - E(u(t_1))$.

2. *Evolution Variational Inequality: For a.e.* $t > 0$,
   $\frac{d}{dt} \frac{1}{2} |u(t) - w|^2 \leq E(w) - E(u(t)) - \frac{\lambda}{2} |u(t) - w|^2$.

3. *Exponential Decay of Subdifferential: If* $u(0) \in D(\partial E)$, $|\partial E(u(t)| \leq e^{-\lambda t} |\partial E(u(0))|$.

4. *Instantaneous Regularization for* $\lambda \geq 0$: *For* $t > 0$, $E(u(t)) \leq \frac{1}{t} |u(0) - w|^2 + E(w)$.
   For $t > 0$ *and* $w \in D(\partial E)$, $|\partial E(u(t))|^2 \leq \frac{1}{t^2} |u(0) - w|^2 + |\partial E(w)|^2$.

5. *Contraction Inequality: For two solutions* $u(t)$ *and* $v(t)$,
   $|u(t) - v(t)| \leq e^{-\lambda t} |u(0) - v(0)|$.

REMARK 1.2.8. In the above theorem, the property that $u(t)$ is locally Lipschitz continuous guarantees that $\frac{d}{dt}u(t)$ exists almost everywhere and

$$u(t) = \int_0^t u'(s)ds + u(0) \ .$$

More generally, this result holds for any reflexive Banach space $\mathcal{B}$ and any $u : [0, T] \to \mathcal{B}$ absolutely continuous [16].

REMARK 1.2.9 (equivalence of evolution variational inequality and gradient flow). In fact, $u(t)$ is a solution of the gradient flow if and only if the evolution variational inequality holds for a.e. $t > 0$. The fact that solutions of the gradient flow satisfy the evolution variational inequality follows by the same argument as in the previous section

(1.1) by simply replacing the gradient with the subdifferential. The other direction follows by expanding

$$\frac{d}{dt}\frac{1}{2}|u(t) - w|^2 = \left(w - u(t), -\frac{d}{dt}u(t)\right)$$

and noting that the evolution variational inequality reduces to the definition of $-\frac{d}{dt}u(t)$ belonging to the subdifferential of E at $u(t)$.

We close this section by applying the above theorem to the $L^2(\Omega)$ gradient flow of the Dirichlet energy described in the first section. As required by the above theorem, E is proper, lower semicontinuous, and convex. In addition to being Gâteaux differentiable on $H_0^1(\Omega) \cap H^2(\Omega)$ with $\nabla_{L^2}E(u) = -\Delta u$, one may also show that for all $u \in D(E) = H_0^1(\Omega)$,

$$\xi \in \partial E(u) \iff u \in H^2(\Omega) \text{ and } \xi = -\Delta u \ .$$

Consequently, by Theorem 1.2.7, for all $u_0 \in \overline{D(E)} = L^2(\Omega)$, there exists $u : [0, \infty) \to L^2(\Omega)$, continuous for $t \geq 0$ and locally Lipschitz continuous for $t > 0$, such that

$$\frac{d}{dt}u(t) = \Delta u(t) \quad a.e. \ t \geq 0 \ , \qquad u(0) = u_0 \ .$$

Furthermore, we have the following five estimates, with $|\cdot| = ||\cdot||_{L^2(\Omega)}$

1. For $t_1 > 0$, $\int_{t_0}^{t_1} |\Delta u(s)|^2 \, ds \leq \frac{1}{2}|\nabla u(t_0))|^2 - \frac{1}{2}|\nabla u(t_1)|^2$.

2. For a.e. $t > 0$, $\frac{d}{dt}\frac{1}{2}|u(t) - w|^2 \leq \frac{1}{2}|\nabla w|^2 - \frac{1}{2}|\nabla u(t))|^2 - \frac{\lambda}{2}|u(t) - w|^2$.

3. If $u(0) \in H^2(\Omega)$, $|\Delta u(t)| \leq |\Delta u(0)|$.

4. For $t > 0$, $u(t) \in H_0^1(\Omega)$ and $\frac{1}{2}|\nabla u(t)|^2 \leq \frac{1}{t}|u(0) - w|^2 + \frac{1}{2}|\nabla w|^2$ for all $w \in H_0^1(\Omega)$.
   For $t > 0$, $u(t) \in H^2(\Omega)$ and $|\Delta u(t)|^2 \leq \frac{1}{t^2}|u(0) - w|^2 + |\Delta w|^2$ for all $w \in H_0^1(\Omega) \cap H^2(\Omega)$.

5. For solutions $u(t)$ and $v(t)$, $|u(t) - v(t)| \leq |u(0) - v(0)|$.

1.3   Discrete Gradient Flow

As described at the beginning of the chapter, the gradient flow perspective also provides a method for approximating solutions to partial differential equations. In the

case of gradient flow on Euclidean space, this approximation corresponds the implicit Euler method. To define the *discrete gradient flow* on a general Hilbert space, we first introduce the notion of *proximal map*.

Let $\lambda^- := \max\{0, -\lambda\}$ denote the negative part of λ.

DEFINITION 1.3.1 (proximal map). Given $E : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, and λ-convex, for any $0 < \tau < \frac{1}{\lambda^-}$ we define the *proximal map* $J_\tau : \mathcal{H} \to D(E)$ by

$$J_\tau u := \operatorname*{argmin}_{v \in \mathcal{H}} \left\{ \frac{1}{2\tau} |v - u|^2 + E(v) \right\} .$$

The sum of $\frac{1}{2\tau}|v - u|^2$ and $E(v)$ is $\frac{1}{\tau} + \lambda$ convex, and we require $0 < \tau < \frac{1}{\lambda^-}$ to ensure $\frac{1}{\tau} + \lambda > 0$.

The fact that the proximal map is well-defined and single-valued for all $u \in \mathcal{H}$ is a specific case of the following classical result, which relies on several essential properties Hilbert spaces (or, more generally, reflexive Banach spaces) [5].

THEOREM 1.3.2 (Minimizers of Strictly Convex Functions). *Given $\Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, and λ-convex for $\lambda > 0$, Φ uniquely attains its minimum.*

Proof. If $\Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is proper, lower semicontinuous, and convex, then there exists $v \in \mathcal{H}$ and $c \in \mathbb{R}$ so that $\Phi(u) \geq (u, v) + c \; \forall u \in \mathcal{H}$. (This is a consequence of Hahn-Banach.) If, in addition, Φ is λ-convex for $\lambda > 0$, then $\Phi - \frac{\lambda}{2}|\cdot|^2$ is convex, so $\Phi(u) \geq (u, v) + c + \frac{\lambda}{2}|u|^2 \xrightarrow{|u| \to +\infty} +\infty$. Therefore, for any $u \in D(\Phi)$, $C := \{v \in \mathcal{H} : \Phi(v) \leq \Phi(u)\}$ is closed, convex, and bounded. By compactness of the unit ball and the fact that convex sets are weakly closed if they are strongly closed, C is weakly compact. The second fact also guarantees that lower semicontinuous, convex functions are weakly lower semicontinuous. Therefore, Φ is a weakly lower semicontinuous on a weakly compact set C, hence it achieves its minimum on C, which is its global minimum. $\square$

COROLLARY 1.3.3. *Given $E : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, and λ-convex, for any $0 < \tau < \frac{1}{\lambda^-}$, the proximal map is well-defined and single-valued for*

all $u \in \mathcal{H}$.

Proof. Apply Theorem 1.3.2 to

$$\Phi(v) := \frac{1}{2\tau}|u - v|^2 + E(v) \ . \tag{1.22}$$

$\square$

Given $E : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, and λ-convex, for any $0 < \tau < \frac{1}{\lambda^-}$ proximal map is characterized by the following Euler-Lagrange equation.

PROPOSITION 1.3.4 (Euler-Lagrange equation for the proximal map)**.** *Under the assumptions of Corollary 1.3.3, for all $u \in \mathcal{H}$*

$$v = J_\tau u \iff \frac{1}{\tau}(u - v) \in \partial E(v) \ .$$

Proof. Define Φ as in (1.22). Since Φ is convex, Lemma 1.2.4 implies that v minimizes Φ if and only if $0 \in \partial E(v) = \frac{1}{\tau}(v - u) + \partial E(v)$. $\square$

PROPOSITION 1.3.5 (contraction of the proximal map)**.** *Under the assumptions of Corollary 1.3.3, for all $u, v \in \mathcal{H}$*

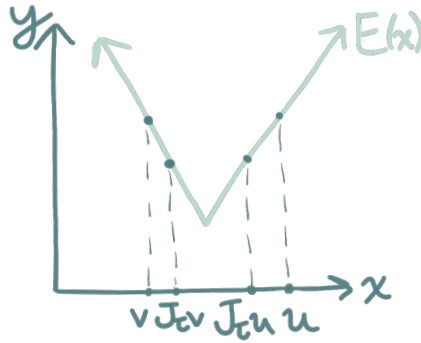$$|J_\tau u - J_\tau v| \leq (1 + \tau\lambda)^{-1}|u - v| \ .$$

Figure 1.2: By Definition 1.3.1, the proximal map J_τ sends a point $u \in \mathcal{H}$ to the point w that makes $E(w)$ as small as possible, up to a $\frac{1}{2\tau}|u - w|^2$ penalty. When E is convex, the proximal map is a contraction.

Proof. By Remark 1.2.5, for all $\xi_1 \in \partial E(u_1), \xi_\in \partial E(u_2)$,

$$(\xi_1 - \xi_2, u_1 - u_2) \geq \lambda|u_1 - u_2|^2$$

$$\implies ([u_1 + \tau\xi_1] - [u_2 + \tau\xi_2], u_1 - u_2) \geq (1 + \tau\lambda)|u_1 - u_2|^2$$

$$\implies |[u_1 + \tau\xi_1] - [u_2 + \tau\xi_2]| \geq (1 + \tau\lambda)|u_1 - u_2| .$$

If we define $v_1 = u_1 + \tau\xi_1$, $v_2 = u_2 + \tau\xi_2$,

$$\frac{1}{\tau}(u - v_1) = \xi_1 \in \partial E(u_1) , \quad \frac{1}{\tau}(v - v_2) = \xi_2 \in \partial E(u_2) .$$

By Proposition 1.3.4, this implies $u_1 = J_\tau v_1$, $u_2 = J_\tau v_2$, which gives the result. $\qquad\square$

We now use the proximal map to define the discrete gradient flow sequence.

DEFINITION 1.3.6 (Discrete Gradient Flow). Suppose $E : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is proper, lower semicontinuous, and λ-convex. Given initial conditions $u \in \mathcal{H}$, the *discrete gradient flow* with time step $0 < \tau < \frac{1}{\lambda^-}$ is defined by

$$u_n := J_\tau u_{n-1} = J_\tau^n u , \quad u_0 = u .$$

We write $J_\tau^n u$ for the nth element of the sequence to emphacize the dependence on τ.

As described at the beginning of the chapter, the convergence of the discrete gradient flow to the continuous time gradient flow is known as the *exponential formula.*

THEOREM 1.3.7 (Exponential Formula, [22]). *Suppose $E : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is proper, lower semicontinuous, and λ-convex. Given initial conditions $u \in \overline{D(E)}$,*

$$\lim_{n\to\infty} J_{t/n}^n u = u(t) \;\; \text{for all } t \geq 0 ,$$

where $u(t)$ is the solution to the gradient flow of E with initial conditions $u(0) = u$.

Theorem 1.3.7 is the infinite dimensional, nonlinear generalization of the limit definition of the matrix exponential. To see this, suppose that $\nabla_{\mathcal{H}} E$ exists, so by Remark 1.2.2 and Proposition 1.3.4,

$$\frac{u - J_\tau u}{\tau} = \nabla_{\mathcal{H}} E(J_\tau u) \implies J_\tau u = (\mathbf{id} + \tau\nabla_{\mathcal{H}} E)^{-1} u . \tag{1.23}$$

In this case, Theorem 1.3.7 may be rewritten as

$$\lim_{n\to\infty} J^n_{t/n} u = \lim_{n\to\infty} \left(\mathbf{id} + \frac{t}{n}\nabla_{\mathcal{H}} E\right)^{-1} u = u(t) \ .$$

In particular, if $\mathcal{H}$ is finite dimensional and $E(u) := (Au, u)$ for $A \in \mathcal{L}(\mathcal{H})$ positive definite, E satisfies the conditions of Theorem 1.3.7 and its gradient $\nabla_{\mathcal{H}} E(u) = Au$ is well-defined. Thus,

$$e^{-At} u := \lim_{n\to\infty} \left(\mathbf{id} + \frac{t}{n} A\right)^{-1} u$$

is the solution of the system of ordinary differential equations given by

$$\frac{d}{dt} u(t) = -Au(t) \ , \quad u(0) = u \ .$$

Chapter 2

Gradient Flow in the Wasserstein Metric: Background and New Results

Through suitable generalizations of the notion of the gradient, the theory of gradient flows has been extended to Banach spaces [11, 12], nonpositively curved metric spaces [17], and general metric spaces [1, 10], and in recent years, there has been significant interest in gradient flow with respect to the *Wasserstein metric*. The Wasserstein metric measures the distance between two probability measures according to the amount of effort it would take to rearrange one probability measure to look like the other, where effort is measured according to the square distance mass is moved. (We refer the reader to sections 2.2 and 2.3 for a precise definition of the Wasserstein metric, along with background about its geometric structure. For a comprehensive introduction to the broader field of *optimal transportation*, we refer the reader to the excellent books by Ambrosio, Gigli, and Savaré [1] and Villani [26, 27].)

The utility of Wasserstein gradient flow in the study of partial differential equations was first demonstrated by Otto in his work on the porous media equation [20, 21]. Though the porous media equation may also be studied as a gradient flow in H^{-1}, Otto showed that by considering its gradient flow in the Wasserstein metric, one obtains sharp polynomial rates of convergence of solutions to Barenblatt profiles. The heat equation provides a second example of how the Wasserstein gradient flow perspective can provide sharper estimates. As described at the beginning of chapter 1, the heat equation is the gradient flow of the Dirichlet energy

$$E(u) := \begin{cases} \frac{1}{2} \int_\Omega |\nabla u(x)|^2 dx & \text{if } u \in H_0^1(\Omega) \\ +\infty & \text{otherwise,} \end{cases} \tag{2.1}$$

with respect to $L^2(\Omega)$. The heat equation is also the gradient flow of the (negative) entropy

$$E(\mu) := \begin{cases} \int_{\mathbb{R}^d} \rho(x) \log \rho(x) dx & \text{if } \mu = \rho(x)dx \text{ and } \int |x|^2 d\mu < \infty \\ +\infty & \text{otherwise.} \end{cases} \tag{2.2}$$

with respect to the Wasserstein metric. In the Hilbertian perspective, there is no $\alpha > 0$ which satisfies

$$|\nabla_{L^2} E(u)|^2 = ||\Delta u||^2_{L^2(\Omega)} \geq \alpha ||\nabla u||^2_{L^2(\Omega)} = \alpha |E(u)| .$$

On the other hand, the analogous inequality in the Wasserstein metric does hold—it is the logarithmic Sobolev inequality,

$$|\nabla_{W_2} E(\mu)|^2 = \frac{1}{4} \int |\nabla \sqrt{\rho}|^2 dx \geq \frac{\pi}{4} \int \rho \log \rho dx = \frac{\pi}{4} E(\mu) . \tag{2.3}$$

Inequalities of the form $|\nabla E(\mu)|^2 \geq \alpha E(\mu)$ are useful for studying asymptotics of a gradient flow. In particular, combining (2.3) with the Wasserstein energy dissipation inequality (see Corollary 3.5.7),

$$E(\mu(t_0)) - E(\mu(t_1)) \geq \int_{t_0}^{t_1} |\nabla_{W_2} E(\mu(s))|^2 ds ,$$

shows by Gronwall's inequality

$$E(\mu(t_1)) \leq E(\mu(t_0)) - \frac{\pi}{4} \int_{t_0}^{t_1} E(\mu(s)) ds \implies E(\mu(t_1)) \leq E(\mu(t_0)) e^{-\pi(t_1-t_0)/4}$$

A more recent reason for interest in gradient flow in the Wasserstein metric is the low regularity of solutions it allows one to consider. When one views a partial differential equation as a gradient flow in the Wasserstein metric, solutions are given by the time evolution of probability measures. First, this allows one to study the long time behavior of solutions which leave all L^p spaces in finite time—in particular, solutions which approach a Dirac mass. Second, this allows one to consider particle approximations of a solution as the time evolution of a sum of Dirac masses, placing particle approximations and the solutions they approximate within the same Wasserstein gradient flow framework [9]. In the case of the aggregation equation, these two techniques have led to new results on blowup and confinement behavior of solutions [7,8].

In this thesis, we further develop the theory of Wasserstein gradient flow and use our results to give a new proof of the exponential formula, inspired by Crandall and Liggett's Banach space result. We then apply our proof of the exponential formula to obtain simple proofs of many of the a priori estimates for the Wasserstein gradient flow.

Table 2.1: Examples of PDEs that are gradient flows in the Wasserstein metric [13, 26] Suppose $m > 1 - \frac{1}{d}$, $V : \mathbb{R}^d \to [0, +\infty)$ is continuously differentiable and convex, and $K : \mathbb{R}^d \to [0, +\infty)$ is continuously differentiable, radial, and convex. We consider the partial differential equations in a weak sense, in the duality with $C_c^\infty(\mathbb{R}^d \times [0, T])$.

	PDE	Energy Functional, $\mu = \rho(x)dx$
Porous Media	$\frac{d}{dt}\mu = \Delta\mu^m$	$E(\mu) = \frac{1}{m-1}\int \rho(x)^m dx$
Fokker Planck	$\frac{d}{dt}\mu = \Delta\mu + \nabla \cdot (\mu\nabla V)$	$E(\mu) = \int \rho(x)\log\rho(x) + V(x)\rho(x)dx$
Aggregation	$\frac{d}{dt}u = \nabla \cdot (\mu\nabla K * \mu)$	$E(\mu) = \frac{1}{2}\int\int \rho(x)K(x-y)\rho(y)dxdy$
DLSS	$\frac{d}{dt}u = -\nabla \cdot (\mu\nabla\frac{\Delta\sqrt{\mu}}{\sqrt{\mu}})$	$E(\mu) = \frac{1}{4}\int \frac{|\nabla\rho(x)|^2}{\rho(x)}dx$

2.1   Summary of Results

The exponential formula in the Wasserstein metric was first proved by Ambrosio, Gigli, and Savaré in their book on gradient flow [1]. Through a careful analysis of affine interpolations of functions of the discrete gradient flow, they obtained the sharp rate of convergence of the discrete gradient flow to the gradient flow and used this to develop many properties of the gradient flow. They also raised the question of whether it might be possible to obtain the same results using a method similar to Crandall and Liggett, bringing together the Banach space and Wasserstein theories.

At first glance, an adaptation of Crandall and Liggett's method to the Wasserstein metric seems unlikely. For E convex, the generalization of ∇E in the Banach space case is an *accretive operator*, which, by definition, is an operator for which the proximal map satisfies

$$||J_\tau u - J_\tau v|| \leq ||u - v|| \,. \tag{2.4}$$

While such an inequality does hold in metric spaces of nonpositive curvature [17], the Wasserstein metric is nonnegatively curved [1, Theorem 7.3.2], and it is unknown if such a contraction holds in this case. Still, there exist "almost" contraction inequalities for the Wasserstein metric, such as [1, Lemma 4.2.4], developed by Ambrosio, Gigli, and Savaré, or [6, Theorem 1.3], developed by Carlen and the author. Using this second contraction inequality, Carlen and the author showed that many of the remarkable features of solutions to the porous media and fast diffusion equations, such as convergence to Barenblatt profiles, are also present in the discrete gradient flow [6].

In this thesis, we use a new almost contraction inequality to adapt Crandall and Liggett's proof of the exponential formula to the Wasserstein metric. A fundamental difference between our method and Crandall and Liggett's is that our almost contraction inequality involves the *square* distance, rather than the distance itself. This prevents us from applying the triangle inequality, as they did, to control the distance between different elements of the the discrete gradient flow. Furthermore, unlike in the Hilbertian case where

$$x \mapsto \frac{1}{2}||x - y||^2$$

is 1-convex along geodesics, the square Wasserstein metric

$$\mu \mapsto \frac{1}{2}W_2^2(\mu, \omega)$$

is not [1, Example 9.1.5]. In fact, it satisfies the opposite inequality [1, Theorem 7.3.2].

The lack of convexity of the square Wasserstein distance is a recurring difficulty when extending results from Hilbert and Banach spaces to the Wasserstein metric. Ambrosio, Gigli, and Savaré circumvented this by introducing a different class of curves— *generalized geodesics*—along which the square distance is 1-convex [1]. We further develop this idea, introducing a class of *transport metrics* $W_{2,\omega}$, with respect to which the generalized geodesics are truly geodesics. The transport metrics satisfy the key property that

$$\mu \mapsto \frac{1}{2}W_{2,\omega}^2(\mu, \omega)$$

is convex along the geodesics induced by $W_{2,\omega}$. This turns out to be the essential fact needed to control the discrete gradient flow and adapt Crandall and Liggett's method to the Wasserstein case.

In sections 2.2 through 2.5, we recall general facts about the Wasserstein metric and functionals defined on this metric space. We will often impose the following assumptions on our functionals.

ASSUMPTION 2.1.1 (optional domain assumption). *$E(\mu) < +\infty$ only if μ is absolutely continuous with respect to Lebesgue measure.*

This assumption 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 (see section 2.2). This is purely for notational convenience. In section A.2 we describe how to remove this assumption.

ASSUMPTION 2.1.2 (convexity assumption). *E is proper, coercive, lower semicontinuous, and λ-convex along generalized geodesics for $\lambda \in \mathbb{R}$.*

This assumption is essential. In particular, the fact that E is λ-convex along generalized geodesics ensures that E is λ-convex in the transport metric $W_{2,\omega}$.

In section 2.6, we define the transport metric $W_{2,\omega}$ and the corresponding subdifferential $\partial_{2,\omega}$ and study their properties. In section 2.7, we define gradient flow in the Wasserstein metric and describe the formal inner product structure of the Wasserstein metric, from which perspective Wasserstein gradient flow is analogous to the Hilbert space gradient flow discussed in chapter 1. In section 2.8, we recall basic facts about the Wasserstein discrete gradient flow and the proximal map J_τ, and the associated minimization problem. In section 2.9, we reframe the minimization problem in terms of the transport metrics, allowing us to prove an Euler-Lagrange equation for minimizer J_τ. In section 2.10, we recall Ambrosio, Gigli, and Savaré's discrete variational inequality [1, Theorem 4.1.2] and prove a stronger version using transport metrics.

In section 3.1, we begin our proof of the exponential formula by proving a new asymmetric almost contraction inequality. In section 3.2, we apply our Euler-Lagrange equation to obtain an expression relating proximal maps with different time steps. In sections 3.3 and 3.4, we combine these results to bound the distance between gradient

flow sequences with different time steps via an asymmetric induction in the style of Rasmussen [23]. Finally, in section 3.5, we prove the exponential formula and quantify the convergence of the discrete gradient flow to the gradient flow.

We close section 3.5 by applying our estimates to give simple proofs of properties of the continuous gradient flow, including the contracting semigroup property and the energy dissipation inequality. Finally, in section 3.6, we extend our results, which only applied to gradient flows with initial conditions $\mu \in D(|\partial E|)$, to include initial conditions $\mu \in \overline{D(E)}$. (See Definition 2.5.1 of the *metric slope* $|\partial E|$.)

In the appendix, we describe two extensions. In section, A.1, we adapt our proof of the exponential formula to include discrete gradient flows with varying time steps. In section A.2, we describe how to remove the optional domain assumption 2.1.1, which we imposed for notational convenience.

2.2   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 sets $B \subseteq \mathbb{R}^d$. We call ν the *push-forward of μ under* $\mathbf{t}$ and write $\nu = \mathbf{t}\#\mu$.
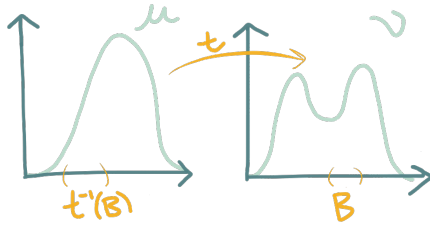
Figure 2.1: $\mathbf{t} : \mathbb{R}^d \to \mathbb{R}^d$ *transports μ onto ν* if $\nu(B) = \mu(\mathbf{t}^{-1}(B))$ for all measurable B.

Consider a measure $\boldsymbol{\mu} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. (We distinguish probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ or $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$, from probability measures on $\mathbb{R}^d$ by writing them in bold font.) Let π^1 be the projection onto the first component of $\mathbb{R}^d \times \mathbb{R}^d$, and let π^2 be the projection onto the second component. The first and second *marginals* of $\boldsymbol{\mu}$ are $\pi^1\#\boldsymbol{\mu} \in \mathcal{P}(\mathbb{R}^d)$ and $\pi^2\#\boldsymbol{\mu} \in \mathcal{P}(\mathbb{R}^d)$.

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

$$W_2(\mu, \nu) := \left(\inf\left\{\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\boldsymbol{\mu}(x, y) : \boldsymbol{\mu} \in \Gamma(\mu, \nu)\right\}\right)^{1/2} . \tag{2.5}$$

When $W_2(\mu, \nu) < +\infty$, there exist plans which attain the infimum. We denote this set of *optimal transport plans* by $\Gamma_0(\mu, \nu)$.

When μ is absolutely continuous with respect to Lebesgue measure, there is a unique optimal transport plan from μ to ν of the form $(\mathbf{id} \times \mathbf{t}) \# \mu$, where $\mathbf{id}(x) = x$ is the identity transformation and $\mathbf{t}$ is unique μ-a.e. [18]. In particular, there is a map $\mathbf{t}$ satisfying $\mathbf{t} \# \mu = \nu$ and

$$W_2(\mu, \nu) = \left(\int_{\mathbb{R}^d} |\mathbf{id} - \mathbf{t}|^2 d\mu\right)^{1/2} .$$

We denote this unique *optimal transport map* by $\mathbf{t}_\mu^\nu$. Furthermore, a Borel measurable map $\mathbf{t}$ that transports μ to ν is optimal if and only if it is *cyclically monotone* μ-a.e. [18], i.e. if there exists $N \subseteq \mathbb{R}^d$ with $\mu(N) = 0$ such that for every finite sequence of distinct points $\{x_1, \ldots, x_m\} \subseteq \mathbb{R}^d \setminus N$,

$$\mathbf{t}(x_1) \cdot (x_2 - x_1) + \mathbf{t}(x_2) \cdot (x_3 - x_2) + \cdots + \mathbf{t}(x_m) \cdot (x_1 - x_m) \leq 0 \ .$$

If, in addition, ν is absolutely continuous with respect to Lebesgue measure, then $t_\mu^\nu \circ t_\nu^\mu = \mathbf{id}$ 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. Throughout this paper, we denote by ω_0 some fixed reference measure and 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.

2.3 Geodesics and Generalized Geodesics

DEFINITION 2.3.1 (constant speed geodesic)**.** Given a metric space (X, d), a *constant speed geodesic* $u : [0, 1] \to X$ is a curve satisfying

$$d(u_\alpha, u_\beta) = |\beta - \alpha| d(u_0, u_1) , \quad \text{for all } \alpha, \beta \in [0, 1] .$$

We will often refer to constant speed geodesics simply as *geodesics.*

By [1][Theorem 7.2.2], all geodesics in $\mathcal{P}_{2,\omega_0}(\mathbb{R}^d)$ are curves of the form

$$\mu_\alpha = \left((1 - \alpha)\pi^1 + \alpha\pi^2 \right) \#\boldsymbol{\mu} , \quad \boldsymbol{\mu} \in \Gamma_0(\mu_0, \mu_1) .$$

If μ_0 is absolutely continuous with respect to Lebesgue measure, the geodesic from μ_0 to μ_1 is unique and of the form

$$\mu_\alpha = \left((1 - \alpha)\mathbf{id} + \alpha\mathbf{t}_{\mu_0}^{\mu_1} \right) \#\mu_0 .$$

We now recall Amrbosio, Gigli, and Savaré's notion of generalized geodesics [1, Definition 9.2.2]. Given a finite product $\mathbb{R}^d \times \mathbb{R}^d \times \cdots \times \mathbb{R}^d$, let π^i the be projection onto the ith component and $\pi^{i,j}$ be the projection onto the ith and jth components.

DEFINITION 2.3.2 (generalized geodesic)**.** Given $\mu_0, \mu_1, \omega \in \mathcal{P}_{2,\omega_0}(\mathbb{R}^d)$, a *generalized geodesic from μ_0 to μ_1 with base ω* is a curve $\mu_\alpha : [0, 1] \to \mathcal{P}(\mathbb{R}^d)$ of the form

$$\mu_\alpha := \left((1 - \alpha)\pi^2 + \alpha\pi^3 \right) \#\boldsymbol{\mu},$$

where $\boldsymbol{\mu} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ satisfies

$$\pi^{1,2}\#\boldsymbol{\mu} \in \Gamma_0(\omega, \mu_0) \text{ and } \pi^{1,3}\#\boldsymbol{\mu} \in \Gamma_0(\omega, \mu_1) . \tag{2.6}$$

We refer to any $\boldsymbol{\mu} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ that satisfies (2.6) as *a plan that induces a generalized geodesic from μ_0 to μ_1 with base ω.*

REMARK 2.3.3. Such a $\boldsymbol{\mu}$ always exists [1, Lemma 5.3.2]. If the base ω equals either μ_0 or μ_1, then μ_α is a geodesic joining μ_0 and μ_1.

REMARK 2.3.4. If ω is absolutely continuous with respect to Lebesgue measure, the generalized geodesic from μ_0 to μ_1 with base ω is unique and of the form

$$\mu_\alpha = ((1-\alpha)\mathbf{t}_\omega^{\mu_0} + \alpha\mathbf{t}_\omega^{\mu_1})\,\#\omega.$$

Since $((1-\alpha)\mathbf{t}_\omega^{\mu_0} + \alpha\mathbf{t}_\omega^{\mu_1})$ is a convex combination of optimal transport maps, it is cyclically monotone, hence it is the optimal transport map from ω to μ_α.

2.4   Convexity

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

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

- *coercive:* There exists $\tau_0 > 0$, $u_0 \in X$ such that

$$\inf\left\{\frac{1}{2\tau_0}d^2(u_0, v) + E(v) : v \in X\right\} > -\infty \ .$$

- *lower semicontinuous:* For all $u_n, u \in X$ such 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 $u_\alpha \in X$,

$$E(u_\alpha) \leq (1-\alpha)E(u_0) + \alpha E(u_1) - \alpha(1-\alpha)\frac{\lambda}{2}d(u_0, u_1)^2 \ , \quad \forall \alpha \in [0,1] \ . \quad (2.7)$$

- *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 (2.7) holds. We will often simply say that E is *λ-convex*, or in the case $\lambda = 0$, *convex*.

Fix $\omega_0 \in \mathcal{P}(\mathbb{R}^d)$ and suppose $(X, d) = (\mathcal{P}_{2,\omega_0}(\mathbb{R}^d), W_2)$. In this setting, convexity is often referred to as *displacement convexity* [19]. This setting also allows us to define the stronger notion of convexity along generalized geodesics [1, Definition 9.2.2].

DEFINITION 2.4.1 (λ-convex along generalized geodesics)**.** 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 if

$$E(\mu_\alpha) \leq (1-\alpha)E(\mu_0) + \alpha E(\mu_1) - \alpha(1-\alpha)\frac{\lambda}{2}\int |x_2 - x_3|^2 d\boldsymbol{\mu} \ , \qquad (2.8)$$

where $\boldsymbol{\mu}$ is the plan that induces the generalized geodesic. 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 μ_α from μ_0 to μ_1 with base ω along which E is convex.

REMARK 2.4.2. This definition is slightly different from E being λ-convex along all of the curves μ_α according to equation (2.7), since

$$W_2^2(\mu_0, \mu_1) \leq \int |x_2 - x_3|^2 d\boldsymbol{\mu}(x) \ . \qquad (2.9)$$

When $\lambda > 0$, equation (2.8) is stronger, and when $\lambda < 0$, it is weaker.

REMARK 2.4.3. When $\omega = \mu_0$ or μ_1, μ_α is simply the geodesic from μ_0 to μ_1 and equality holds in (2.9). Therefore, λ-convexity along generalized geodesics implies λ-convexity along geodesics.

2.5 Differentiability

DEFINITION 2.5.1 (metric slope)**.** Given a metric space (X, d) and a functional $E : X \to \mathbb{R} \cup \{+\infty\}$, the *metric slope* of E at $u \in D(E)$ is given by

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

The above notion of differentiability merely relies on the metric space structure. If $(X, d) = (\mathcal{P}_{2,\omega_0}(\mathbb{R}^d), W_2)$, we may also consider a stronger notion of differentiability known as the subdifferential [1, Definition 10.1.1]. For ease of notation, we assume E satisfies domain assumption 2.1.1, so that, for any $\mu \in D(E)$, $\nu \in \mathcal{P}_{2,\omega_0}(\mathbb{R}^d)$, there exists a unique optimal transport map $\mathbf{t}_\mu^\nu$ from μ to ν. We explain how to extend these results to the general case in section A.2.

DEFINITION 2.5.2 (Wasserstein subdifferential)**.** Consider $E : \mathcal{P}_{2,\omega_0}(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, and satisfying domain assumption 2.1.1. Given

$\mu \in D(|\partial E|)$, $\boldsymbol{\xi} \in L^2(\mu)$ belongs to the *Wasserstein subdifferential* of E at μ, written $\boldsymbol{\xi} \in \partial E(\mu)$, in case

$$E(\nu) - E(\mu) \geq \int_{\mathbb{R}^d} \langle \boldsymbol{\xi}, \mathbf{t}_\mu^\nu - \mathbf{id} \rangle d\mu + o(W_2(\mu, \nu)) \quad \text{as } \nu \xrightarrow{W_2} \mu .$$

REMARK 2.5.3 (Wasserstein subdifferential and metric slope)**.** Given E satisfying domain assumption 2.1.1 and convexity assumption 2.1.2, $\mu \in D(|\partial E|)$ if and only if $\partial E(\mu)$ is nonempty [1, Lemma 10.1.5]. In this case,

$$|\partial E|(\mu) = \min\{||\xi||_{L^2(\mu)} : \xi \in \partial E(\mu)\} .$$

Finally, we recall the definition of the strong subdifferential from [1, 10.1.1]. This quantifies the rate of change of E when approaching μ via any transport map, not necessarily an optimal one.

DEFINITION 2.5.4 (strong subdifferential)**.** Consider $E : \mathcal{P}_{2,\omega_0}(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, and satisfying domain assumption 2.1.1. $\boldsymbol{\xi} \in \partial E(\mu)$ is a *strong subdifferential* in case for all measurable maps $\mathbf{t} : \mathbb{R}^d \to \mathbb{R}^d$ such that $||\mathbf{t} - \mathbf{id}||_{L^2(\mu)} < +\infty$,

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

2.6 Transport Metrics

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

$$x \mapsto \frac{1}{2}||x - y||^2$$

is 1-convex along geodesics,

$$\mu \mapsto \frac{1}{2}W_2^2(\mu, \omega)$$

is not [1, Example 9.1.5].

Ambrosio, Gigli, and Savaré circumvent this difficulty by introducing the notion of generalized geodesics and showing that

$$\mu \mapsto \frac{1}{2}W_2^2(\mu, \omega)$$

is 1-convex along generalized geodesics with base ω [1, Lemma 9.2.1]. In this section, we introduce a class of metrics whose geodesics correspond exactly to the generalized geodesics with a given base. Furthermore, these metrics satisfy the property that the square distance is convex with respect to its own constant speed geodesics. This convexity turns out to be extremely useful for some of the key estimates in our adaptation of Crandall and Liggett's proof of the exponential formula.

For simplicity of notation, we make the following assumption on the measure ω:

ASSUMPTION 2.6.1 (ω doesn't charge small sets)**.**
$\omega \in \mathcal{P}(\mathbb{R}^d)$ *is absolutely continuous with respect to Lebesgue measure.*

This ensures the existence of an optimal transport maps $\mathbf{t}_\omega^\mu$ from ω to any $\mu \in \mathcal{P}_{2,\omega}(\mathbb{R}^d)$ [18]. We use these optimal transport maps to define the $(2, \omega)$-transport distance. See section A.2 for how to extend this definition for ω are not absolutely continuous with respect to Lebesgue measure.

DEFINITION 2.6.2 ($(2, \omega)$-transport metric)**.** The $(2, \omega)$-*transport metric* is

$$W_{2,\omega} : \mathcal{P}_{2,\omega}(\mathbb{R}^d) \times \mathcal{P}_{2,\omega}(\mathbb{R}^d) \to \mathbb{R} ,$$

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

REMARK 2.6.3. If $\mu = \omega$ or $\nu = \omega$, this reduces to the Wasserstein metric. In general, $W_{2,\omega}(\mu, \nu) \geq W_2(\mu, \nu)$.

In the following proposition, we prove a few key properties of transport metrics. In particular, we show that the geodesics of the $W_{2,\omega}$ metric are exactly the generalized geodesics with base ω, and that the function $\mu \mapsto W_{2,\omega}(\nu, \mu)^2$ is convex for any $\nu \in \mathcal{P}_{2,\omega}(\mathbb{R}^d)$.

PROPOSITION 2.6.4 (properties of the $(2, \omega)$-transport metric)**.**

(i) $W_{2,\omega}$ *is a metric on* $\mathcal{P}_{2,\omega}(\mathbb{R}^d)$*.*

(ii) *The constant speed geodesics with respect to the $W_{2,\omega}$ metric are exactly the generalized geodesics with base ω. Furthermore, these generalized geodesics μ_α satisfy*

$$W_{2,\omega}^2(\nu, \mu_\alpha) = (1-\alpha)W_{2,\omega}^2(\nu, \mu_0) + \alpha W_{2,\omega}^2(\nu, \mu_1) - \alpha(1-\alpha)W_{2,\omega}^2(\mu_0, \mu_1)$$

(2.10)

for all $\nu \in P_\omega(\mathbb{R}^d)$.

(iii) *Generalized geodesics with base ω are the unique constant speed geodesics in the $W_{2,\omega}$ metric. Consequently, a functional E is λ-convex along generalized geodesics with base ω if and only if it is λ-convex in the $W_{2,\omega}$ metric. In particular, the function $\mu \mapsto W_{2,\omega}^2(\nu, \mu)$ is 2-convex in the $W_{2,\omega}$ metric for any $\nu \in \mathcal{P}_{2,\omega}(\mathbb{R}^d)$.*

Proof.

(i) $W_{2,\omega}$ is symmetric and nonnegative by definition. It is non-degenerate since

$$0 = W_{2,\omega}(\mu, \nu) \geq W_2(\mu, \nu) \implies \mu = \nu \ .$$

$W_{2,\omega}$ satisfies the triangle inequality since $L^2(\omega)$ satisfies the triangle inequality:

$$W_{2,\omega}(\mu, \nu) = ||\mathbf{t}_\omega^\mu - \mathbf{t}_\omega^\nu||_{L^2(\omega)} \leq ||\mathbf{t}_\omega^\mu - \mathbf{t}_\omega^\rho||_{L^2(\omega)} + ||\mathbf{t}_\omega^\rho - \mathbf{t}_\omega^\nu||_{L^2(\omega)}$$

$$= W_{2,\omega}(\mu, \rho) + W_{2,\omega}(\rho, \nu)$$

(ii) 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 at time $\alpha \in [0,1]$. By Remark 2.3.4, $\mathbf{t}_\omega^{\mu_\alpha} = (1-\alpha)\mathbf{t}_\omega^{\mu_0} + \alpha\mathbf{t}_\omega^{\mu_1}$. Consequently,

$$W_{2,\omega}(\mu_\alpha^{\mu \to \nu}, \mu_\beta^{\mu \to \nu}) = \left(\int |((1-\alpha)\mathbf{t}_\omega^\mu + \alpha\mathbf{t}_\omega^\nu) - ((1-\beta)\mathbf{t}_\omega^\mu + \beta\mathbf{t}_\omega^\nu)|^2 d\omega\right)^{1/2}$$

$$= \left(\int |((\beta-\alpha)\mathbf{t}_\omega^\mu + (\alpha-\beta)\mathbf{t}_\omega^\nu|^2 d\omega\right)^{1/2}$$

$$= |\beta - \alpha|W_{2,\omega}(\mu, \nu)$$

This shows that μ_α is a constant speed geodesic. The second result follows from

the corresponding identity of the $L^2(\omega)$ norm.

$$W_{2,\omega}^2(\nu, \mu_\alpha) = ||(1-\alpha)\mathbf{t}_\omega^{\mu_0} + \alpha\mathbf{t}_\omega^{\mu_1} - \mathbf{t}_\omega^\nu||_{L^2(\omega)}^2$$

$$= (1-\alpha)||\mathbf{t}_\omega^{\mu_0} - \mathbf{t}_\omega^\nu||_{L^2(\omega)}^2 + \alpha||\mathbf{t}_\omega^{\mu_1} - \mathbf{t}_\omega^\nu||_{L^2(\omega)}^2$$

$$- \alpha(1-\alpha)||\mathbf{t}_\omega^{\mu_0} - \mathbf{t}_\omega^{\mu_1}||_{L^2(\omega)}^2$$

$$= (1-\alpha)W_{2,\omega}^2(\mu_0, \nu) + \alpha W_{2,\omega}^2(\mu_1, \nu) - \alpha(1-\alpha)W_{2,\omega}^2(\mu_0, \mu_1)$$

(iii) Suppose $\tilde{\mu}_\alpha$ is a constant speed geodesic in the $W_{2,\omega}$ metric from μ_0 to μ_1. 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. Setting $\nu = \tilde{\mu}_\alpha$ in equation (2.10) gives

$$W_{2,\omega}^2(\tilde{\mu}_\alpha, \mu_\alpha) = (1-\alpha)W_{2,\omega}^2(\tilde{\mu}_\alpha, \mu_0) + \alpha W_{2,\omega}^2(\tilde{\mu}_\alpha, \mu_1) - \alpha(1-\alpha)W_{2,\omega}^2(\mu_0, \mu_1) \ .$$

Using the fact that $\tilde{\mu}_\alpha$ is a constant speed geodesic shows

$$W_{2,\omega}^2(\tilde{\mu}_\alpha, \mu_\alpha) = (1-\alpha)\alpha^2 W_{2,\omega}^2(\mu_1, \mu_0) + \alpha(1-\alpha)^2 W_{2,\omega}^2(\mu_0, \mu_1)$$

$$- \alpha(1-\alpha)W_{2,\omega}^2(\mu_0, \mu_1)$$

$$= (\alpha + (1-\alpha) - 1)(1-\alpha)\alpha W_{2,\omega}^2(\mu_0, \mu_1)$$

$$= 0 \ .$$

Therefore $\tilde{\mu}_\alpha = \mu_\alpha$ and generalized geodesics are the unique constant speed geodesics in the $W_{2,\omega}$ metric.

$\square$

We may define the subdifferential with respect to $W_{2,\omega}$ in analogy with the Wasserstein subdifferential, Definition 2.5.2.

DEFINITION 2.6.5 ($W_{2,\omega}$ subdifferential)**.** Given $E : \mathcal{P}_{2,\omega}(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ proper and lower semicontinuous in $W_{2,\omega}$, $\boldsymbol{\xi} \in L^2(\omega)$ belongs to the $W_{2,\omega}$ subdifferential $\partial_{2,\omega}E(\mu)$ in case

$$E(\nu) - E(\mu) \geq \int \langle \boldsymbol{\xi}, \mathbf{t}_\omega^\nu - \mathbf{t}_\omega^\mu \rangle d\omega + o(W_{2,\omega}(\mu, \nu)) \qquad \text{as } \nu \to \mu \ .$$

REMARK 2.6.6 (lower semicontinuity in W_2 vs. $W_{2,\omega}$). By Remark 2.6.3, 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, i.e. $\liminf_{n\to\infty} E(\mu_n) \geq E(\mu)$ for all sequences which converge in W_2, the inequality continues to hold for the possibly smaller set of sequences which converge in $W_{2,\omega}$, hence E is lower semicontinuous in $W_{2,\omega}$.

REMARK 2.6.7 (additivity of $W_{2,\omega}$ subdifferential). If $\boldsymbol{\xi}_1 \in \partial_{2,\omega} E_1(\mu)$ and $\boldsymbol{\xi}_2 \in \partial_{2,\omega} E_2(\mu)$,

$$E_1(\nu) + E_2(\nu) - E_1(\mu) - E_2(\mu) \geq \int \langle \boldsymbol{\xi}_1 + \boldsymbol{\xi}_2, \mathbf{t}_\omega^\nu - \mathbf{t}_\omega^\mu \rangle d\omega + o(W_{2,\omega}(\mu,\nu)) \ ,$$

so $\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2 \in \partial_{2,\omega}(E_1 + E_2)(\mu)$.

The next proposition provides a characterization of the $W_{2,\omega}$ subdifferential for functionals that are convex in $W_{2,\omega}$, in analogy with [1, Equation (10.1.7)].

PROPOSITION 2.6.8 ($W_{2,\omega}$ subdifferential for convex function). *Given E λ-convex with respect to $W_{2,\omega}$ and satisfying the conditions of Definition 2.6.5, $\boldsymbol{\xi} \in \partial_{2,\omega} E(\mu)$ if and only if*

$$E(\nu) - E(\mu) \geq \int \langle \boldsymbol{\xi}, \mathbf{t}_\omega^\nu - \mathbf{t}_\omega^\mu \rangle d\omega + \frac{\lambda}{2} W_{2,\omega}^2(\mu,\nu) \quad \forall \nu \ . \tag{2.11}$$

Proof. If (2.11) holds, then $\boldsymbol{\xi} \in \partial_{2,\omega} E(\mu)$ by Definition 2.6.5. For the converse, assume $\boldsymbol{\xi} \in \partial_{2,\omega} E(\mu)$. Define $\mu_\alpha = ((1-\alpha)\mathbf{t}_\omega^\mu + \alpha \mathbf{t}_\omega^\nu) \# \omega$ to be the generalized geodesic from μ to ν with base ω. Since E is λ convex in the $W_{2,\omega}$ metric,

$$\frac{E(\mu_\alpha) - E(\mu)}{\alpha} \leq E(\nu) - E(\mu) - \frac{\lambda}{2}(1-\alpha) W_\omega^2(\mu,\nu) \ . \tag{2.12}$$

By Proposition 2.6.4, $W_{2,\omega}(\mu,\mu_\alpha) = \alpha W_{2,\omega}(\mu,\nu)$, and by Remark 2.3.4, $\mathbf{t}_\omega^{\mu_\alpha} = (1-\alpha)\mathbf{t}_\omega^\mu + \alpha \mathbf{t}_\omega^\nu$. Combining these with the definition of $\boldsymbol{\xi} \in \partial_{2,\omega} E(\mu)$ gives

$$
\begin{aligned}
\liminf_{\alpha \to 0} \frac{E(\mu_\alpha) - E(\mu)}{\alpha} &\geq \liminf_{\alpha \to 0} \frac{1}{\alpha} \int \langle \boldsymbol{\xi}, \mathbf{t}_\omega^{\mu_\alpha} - \mathbf{t}_\omega^\mu \rangle d\omega \\
&= \liminf_{\alpha \to 0} \frac{1}{\alpha} \int \langle \boldsymbol{\xi}, (1-\alpha)\mathbf{t}_\omega^\mu + \alpha \mathbf{t}_\omega^\nu - \mathbf{t}_\omega^\mu \rangle d\omega \\
&= \int \langle \boldsymbol{\xi}, \mathbf{t}_\omega^\nu - \mathbf{t}_\omega^\mu \rangle d\omega
\end{aligned}
$$

Sending $\alpha \to 0$ in equation (2.12) shows

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

$\square$

COROLLARY 2.6.9. *Given E satisfying the conditions of Definition 2.6.5 with $\lambda \geq 0$, μ is a minimizer for E if and only if $0 \in \partial_{2,\omega} E(\mu)$.*

PROPOSITION 2.6.10 ($W_{2,\omega}$ subdifferential of $W_2^2(\omega, \cdot)$). *The $W_{2,\omega}$ subdifferential of $W_2^2(\omega, \cdot)$ evaluated at μ contains the element $2(\mathbf{t}_\omega^\mu - \mathbf{id})$.*

Proof.

$$\begin{aligned}
W_2^2(\omega, \nu) - W_2^2(\omega, \mu) &= \int |\mathbf{t}_\omega^\nu - \mathbf{id}|^2 d\omega - \int |\mathbf{t}_\omega^\mu - \mathbf{id}|^2 d\omega \\
&= \int |\mathbf{t}_\omega^\nu - \mathbf{t}_\omega^\mu|^2 + 2\langle \mathbf{t}_\omega^\nu, \mathbf{t}_\omega^\mu \rangle - 2\langle \mathbf{t}_\omega^\nu, \mathbf{id} \rangle + 2\langle \mathbf{t}_\omega^\mu, \mathbf{id} \rangle - 2|\mathbf{t}_\omega^\mu|^2 d\omega \\
&= W_{2,\omega}^2(\mu, \nu) + \int 2\langle \mathbf{t}_\omega^\nu, \mathbf{t}_\omega^\mu - \mathbf{id} \rangle + 2\langle \mathbf{t}_\omega^\mu, \mathbf{id} - \mathbf{t}_\omega^\mu \rangle d\omega \\
&= W_{2,\omega}^2(\mu, \nu) + \int 2\langle \mathbf{t}_\omega^\nu - \mathbf{t}_\omega^\mu, \mathbf{t}_\omega^\mu - \mathbf{id} \rangle d\omega
\end{aligned}$$

By Proposition 2.6.8, this implies that $2(\mathbf{t}_\omega^\mu - \mathbf{id}) \in \partial_{2,\omega} W_\omega^2(\omega, \mu)$. $\square$

Finally, if E has a strong subdifferential (Definition 2.5.4), E is subdifferentiable with respect to $W_{2,\omega}$.

LEMMA 2.6.11 (strong subdifferential vs. $W_{2,\omega}$ subdifferential). *Given E satisfying the conditions of Definition 2.5.4, if $\boldsymbol{\xi} \in \partial E(\mu)$ is a strong subdifferential, then $\boldsymbol{\xi} \circ \mathbf{t}_\omega^\mu \in \partial_{2,\omega} E(\mu)$.*

Proof. If E has a strong subdifferential $\boldsymbol{\xi}$ at μ, $\boldsymbol{\xi} \in L^2(\mu)$, hence $\boldsymbol{\xi} \circ \mathbf{t}_\omega^\mu \in L^2(\omega)$. Furthermore,

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

Therefore, $\boldsymbol{\xi} \circ \mathbf{t}_\omega^\mu \in \partial_{2,\omega} E(\mu)$. $\square$

2.7   Gradient Flow

With the notion of differentiability defined in section 2.5, we may now define *gradient flow* in the Wasserstein metric. We begin with some heuristic motivation, in analogy with our treatment of Hilbert space gradient flow in section 1.1.

Following Ambrosio, Gigli, and Savaré [1, Chapters 1 and 8], we consider the following class of curves in $\mathcal{P}_{2,\omega_0}(\mathbb{R}^d)$.

DEFINITION 2.7.1 (absolutely continuous curve)**.** Given an open interval $I \subseteq \mathbb{R}$, $\mu(t) : I \to \mathcal{P}_{2,\omega_0}(\mathbb{R}^d)$ is *absolutely continuous* if there exists $m \in L^1(I)$ so that

$$W_2(\mu(t), \mu(s)) \leq \int_s^t m(r)dr \quad \forall s, t \in I \quad s \leq t \ .$$

Likewise, $\mu(t) : \mathbb{R}^d \to \mathcal{P}_{2,\mu_0}(\mathbb{R}^d)$ is *locally absolutely continuous* if it is absolutely continuous on all bounded intervals.

By [1][Theorem 8.3.1], if $\mu(t)$ is absolutely continuous, there exists a Borel vector field $v(x, t)$ such that the continuity equation

$$\frac{\partial \mu}{\partial t} + \nabla \cdot (\mu v) = 0 \ ,$$

holds in the duality with $C_c^\infty(\mathbb{R}^d \times I)$ and if we define $|\mu'|(t) = \lim_{s \to t} W_2(\mu(t), \mu(s))/|t - s|$,

$$v(x, t) \in \overline{\{\nabla \psi : \psi \in C_c^\infty(\mathbb{R}^d)\}}^{L_2(\mu(t))} \ , \quad \int |v(x, t)|^2 d\mu(t) = |\mu'|(t) \ , \quad \text{for a.e. } t \in I \ .$$

$$(2.13)$$

Motivated by this result, we now formally define the Wasserstein inner product and gradient. We identify the tangent space at a measure μ with the space of absolutely continuous curves $\mu : [0, 1] \to \mathcal{P}_{2,\omega_0}(t)$ such that $\mu(0) = \mu$, and we suppose that the corresponding velocity fields $v(x, t)$ are given by $\nabla \psi$ for $\psi \in C_c^\infty(\mathbb{R}^d)$, rather than just the limit of such functions with respect to $L^2(\mu(t))$. With this, we define the Wasserstein inner product at μ by

$$\left(\frac{\partial \mu}{\partial t}, \frac{\partial \tilde{\mu}}{\partial t} \right)_\mu := \int \nabla \psi(x) \cdot \nabla \tilde{\psi}(x) d\mu \ . \tag{2.14}$$

This formal inner product induces the Wasserstein distance between two measures as the minimum energy of all connecting curves [2],

$$W_2^2(\mu_0, \mu_1) = \inf_{\mu(t)} \left\{ \int_0^1 \left\| \frac{\partial \mu}{\partial t} \right\|_{\mu(t)}^2 dt : \mu(0) = \mu_0, \mu(1) = \mu_1, \ \frac{\partial \mu}{\partial t} + \nabla \cdot (\mu v) = 0 \right\} .$$

Suppose $\mu(t)$ is a geodesic from μ_0 to μ_1 (Definition 2.3.1). Then $\mu(t)$ is absolutely continuous, so by (2.13) its velocity field satisfies $\int |v(x,t)|^2 d\mu(t) = W_2(\mu_1, \mu_0)$. Thus, $\int_0^1 \|\partial\mu/\partial t\|_{\mu(t)}^2 dt = \int_0^1 \int |v(x,t)|^2 d\mu(t)dt = W_2(\mu_0, \mu_1)$. Therefore, constant speed geodesics are length minimizing geodesics.

In analogy with the Hilbert space gradient from Definition 1.1.1, we formally define the gradient $\nabla_{W_2} E(\mu)$ with respect to the Wasserstein metric by

$$\left(\nabla_{W_2} E(\mu), \frac{\partial \mu}{\partial t} \right)_\mu = \lim_{t \to 0} \frac{E(\mu(t)) - E(\mu)}{t} . \tag{2.15}$$

In particular, when E is an integral functional of the form

$$E(\mu) = \begin{cases} \int_{\mathbb{R}^d} F(x, \rho(x), \nabla\rho(x))dx & \text{for } \mu = \rho \ dx, \ \rho(x) \in C^1(\mathbb{R}^d) \\ +\infty & \text{otherwise,} \end{cases} \tag{2.16}$$

for F sufficiently regular, the Wasserstein gradient of E may be expressed in terms of the functional derivative $\frac{\delta E}{\delta \mu}$ as

$$\nabla_{W_2} E(\mu) = -\nabla \cdot \left(\mu \nabla \frac{\delta E}{\delta \mu} \right) .$$

This pseudo-Riemannian structure of the Wasserstein metric was discovered by Otto. Along with Jordan and Kinderlehrer, he demonstrated its great utility as a heuristic tool for studying partial differential equations [14, 20, 21]. Note that for an integral functional of the form (2.16), the Wasserstein subdifferential is given by

$$\partial E(\mu) = \left\{ \nabla \frac{\delta E}{\delta \mu} \right\} .$$

Thus, while we identify the Wasserstein gradient of E with members of the tangent space $\frac{\partial \mu}{\partial t}$, we identify the Wasserstein subdifferential with the corresponding velocity field $v(x,t)$ [1][Lemma 10.4.1].

With intuition from this formal perspective, we now turn to the rigorous definition of gradient flow. Given the equivalence of the evolution variational inequality and the

Hilbertian gradient flow (see Remark 1.2.9), we follow Ambrosio, Gilgi, and Savaré and define the Wasserstein gradient flow via the analogous the evolution variational inequality [1, Equation (4.0.3)].

DEFINITION 2.7.2 (gradient flow). Suppose E satisfies convexity assumption 2.1.2, so E is λ-convex 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 a functional E with initial data $\mu \in \overline{D(E)}$ if $\mu(t) \xrightarrow{t \to 0} \mu$ and

$$\frac{1}{2}\frac{d}{dt}W_2^2(\mu(t), \omega) + \frac{\lambda}{2}W_2^2(\mu(t), \omega) \leq E(\omega) - E(\mu(t)), \quad \forall \omega \in D(E), \text{ a.e. } t > 0 . \quad (2.17)$$

We will sometimes refer to $\mu(t)$ as the *continuous* gradient flow, to distinguish it from the discrete gradient flow we define in the following section.

As in the Hilbertian case, the above definition of Wasserstein gradient flow in terms of an evolution variational inequality is equivalent to a differential definition [1, Lemma 10.4.1, Theorem 11.1.4]. In particular, $\mu(t)$ is a gradient flow of E according to Definition 2.7.2 if and only if $\mu(t) \xrightarrow{t \to 0} \mu$ and its velocity field satisfies

$$v(x, t) \in -\partial E(\mu(t)) \text{ a.e. } t > 0 .$$

If in addition E is an integral functional of the form (2.16), this is also equivalent to

$$\frac{d}{dt}\mu(t) = -\nabla_W E(\mu(t)) ,$$

in the duality with $C_c^\infty(\mathbb{R}^d \times (0, +\infty))$.

2.8 Discrete Gradient Flow

Given a functional E, a time step $\tau > 0$, and $\mu, \nu \in \mathcal{P}_{2,\omega_0}(\mathbb{R}^d)$ the *quadratic perturbation* of E is

$$\Phi(\tau, \mu; \nu) := \frac{1}{2\tau}W_2^2(\mu, \nu) + E(\nu) . \quad (2.18)$$

The *proximal set* $J_\tau : \mathcal{P}_{2,\omega_0}(\mathbb{R}^d) \to 2^{\mathcal{P}_{2,\omega_0}(\mathbb{R}^d)}$ corresponding to E is

$$J_\tau(\mu) := \operatorname*{argmin}_{\nu \in \mathcal{P}_{2,\omega_0}(\mathbb{R}^d)} \left\{ \frac{1}{2\tau}W_2^2(\mu, \nu) + E(\nu) \right\} . \quad (2.19)$$

We define $J_0(\mu) := \mu$.

For the remainder of this section, we consider functionals that satisfy the convexity assumption 2.1.2 for some $\lambda \in \mathbb{R}$. In order to jointly consider the cases $\lambda \geq 0$ and $\lambda < 0$, we define the negative part of λ,

$$\lambda^- = \begin{cases} -\lambda & \text{if } \lambda < 0 \\ \\ 0 & \text{if } \lambda \geq 0 . \end{cases}$$

In the case $\lambda \geq 0$, we interpret $\frac{1}{\lambda^-} = +\infty$.

Suppose $\mu \in \overline{D(E)}$ and $0 < \tau < \frac{1}{\lambda^-}$. (When $\lambda < 0$, the size restriction $0 < \tau < \frac{1}{\lambda^-}$ ensures that $0 < 1 + \lambda\tau < 1$.) Then there exists a unique element in $J_\tau(\mu)$ and the *proximal map* $J_\tau : \overline{D(E)} \to D(E) : \mu \mapsto \mu_\tau$ is continuous [1, Theorem 4.1.2].

In [1, Theorem 3.1.6], Ambrosio, Gigli, and Savaré unite the notions of subdifferential and proximal map through the following chain of inequalities. Recall that $|\partial E| : \mathcal{P}_{2,\omega_0} \to \mathbb{R} \cup \{+\infty\}$ is the metric slope—see Definition 2.5.1.

Theorem AGS1. *Given E satisfying convexity assumption 2.1.2 and $\mu \in D(|\partial E|)$ and $0 < \tau < \frac{1}{\lambda^-}$,*

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

$$\leq \frac{\tau^2}{(1 + \lambda\tau)^2}|\partial E|^2(\mu) . \tag{2.20}$$

The *discrete gradient flow* sequence with time step τ is constructed via repeated applications of the proximal map,

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

We write J_τ^n to indicate n repeated applications of the proximal map, so that $\mu_n = J_\tau^n \mu_0$.

2.9   Euler-Lagrange Equation

THEOREM 2.9.1 (Euler-Lagrange equation)**.** *Assume that E satisfies assumptions 2.1.1 and 2.1.2 and $\omega \in D(E)$. Then for $0 < \tau < \frac{1}{\lambda^-}$, ν is the unique minimizer of the quadratic perturbation $\Phi(\tau, \omega; \cdot)$, if and only if*

$$\frac{1}{\tau}(\mathbf{t}_\nu^\omega - \mathbf{id}) \in \partial E(\nu) \text{ is a strong subdifferential.} \tag{2.21}$$

Hence, ω_τ is characterized by the fact that $\frac{1}{\tau}(\mathbf{t}_{\omega_\tau}^\omega - \mathbf{id}) \in \partial E(\omega_\tau)$.

We assume $\omega \in D(E)$ and E satisfies domain assumption 2.1.1 to ease notation. See section A.2 for how the assumption on ω can be relaxed to $\omega \in \overline{D(E)}$ and the domain assumption can be removed.

Proof of Theorem 2.9.1. The fact that

$$\nu \text{ minimizes } \Phi(\tau, \omega; \nu) \implies \frac{1}{\tau}(\mathbf{t}_\nu^\omega - \mathbf{id}) \in \partial E(\nu) \text{ is a strong subdifferential}$$

is proved in [1, Lemma 10.1.2] using a type of argument introduced by Otto [20, 21]. To see the other direction, note that if

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

then by Lemma 2.6.11,

$$\frac{1}{\tau}(\mathbf{id} - \mathbf{t}_\omega^\nu) \in \partial_{2,\omega} E(\nu) \ .$$

Combining Remark 2.6.7 and Proposition 2.6.10 shows

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

Since $W_2^2(\omega, \cdot) = W_{2,\omega}^2(\omega, \cdot)$ is 2-convex in the $W_{2,\omega}$ metric and E is λ-convex in the $W_{2,\omega}$ metric, $\Phi(\tau, \omega; \cdot)$ is $\left(\frac{1}{\tau} + \lambda\right)$-convex in the $W_{2,\omega}$ metric, with $\left(\frac{1}{\tau} + \lambda\right) > 0$. Therefore, by Corollary 2.6.9, when $0 < \tau < \frac{1}{\lambda^-}$, $0 \in \partial_{2,\omega} \Phi(\tau, \omega; \nu)$, and ν minimizes $\Phi(\tau, \omega; \cdot)$ □

2.10 Discrete Variational Inequality

The notion of a discrete variational inequality was introduced in [1] to provide quantitative control over the discrete gradient flow for functionals that are convex along generalized geodesics. This inequality follows from the fact that if E is λ-convex along generalized geodesics, then for all $\mu \in D(E)$, $0 < \tau < \frac{1}{\lambda^-}$, $\nu \mapsto \Phi(\tau, \mu; \nu)$ is $(1/\tau + \lambda)$-convex along generalized geodesics with base μ. In particular, this inequality is an "above the tangent line" or Talagrand inequality for the convex function Φ.

THEOREM AGS2. *Given E satisfying convexity assumption 2.1.2, for all $0 < \tau < \frac{1}{\lambda^{-}}$, $\mu \in \overline{D(E)}$, and $\nu \in D(E)$,*

$$\frac{1}{2h}[W_2^2(\mu_h, \nu) - W_2^2(\mu, \nu)] + \frac{\lambda}{2}W_2^2(\mu_h, \nu) \leq E(\nu) - E(\mu_h) - \frac{1}{2h}W_2^2(\mu, \mu_h) \quad (2.22)$$

For our purposes, we require not only control of the Wasserstein metric along the discrete gradient flow, but also control over transport metrics along discrete gradient flow. Luckily, the convexity of E along generalized geodesics implies something slightly stronger than Theorem AGS2. In particular, we may obtain an "above the tangent line" inequality for Φ with respect to the $W_{2,\mu}$ transport metric.

In the next theorem, we assume the base point $\mu << \mathcal{L}^d$ so that the transport metric $W_{2,\mu}$ is well defined by Definition 2.6.2. As before, this assumption is only for ease of notation, and we describe how to remove it in section A.2.

THEOREM 2.10.1 (discrete variational inequality). *Suppose E satisfies convexity assumption 2.1.2. Then for all $\mu \in \overline{D(E)}$ and $\nu \in D(E)$,*

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

or, equivalently,

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

Proof. The following proof is nearly identical to [1, Theorem 4.1.2 (ii)], except for the use of the transport metric $W_{2,\mu}$. By the convexity of E and $\frac{1}{2\tau}W_2^2(\cdot, \mu)$ along generalized geodesics with base μ, the functional $\nu \mapsto \Phi(\tau, \mu; \nu)$ is convex in the $W_{2,\mu}$ transport metric. Thus, for any generalized geodesic μ_α from μ_τ to ν with base μ, since μ_τ is the minimizer of $\Phi(\tau, \mu; \cdot)$,

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

Rearranging and dividing by α,

$$0 \leq \Phi(\tau, \mu; \nu) - \Phi(\tau, \mu; \mu_\tau) - \frac{1 + \lambda\tau}{2\tau}(1 - \alpha)W_{2,\mu}^2(\mu_\tau, \nu) \,.$$

Sending $\alpha \to 0$ and expanding Φ according to its definition gives the result. $\qquad\square$

Chapter 3

Exponential Formula for the Wasserstein Metric

Given E satisfying convexity assumption 2.1.2, we aim to show that, as the time step goes to zero, the discrete gradient flow converges to the continuous gradient flow

$$\lim_{n \to \infty} J_{t/n}^n \mu = \mu(t) \ . \tag{3.1}$$

The key difficulty in showing (3.1) lies in proving that the limit exists. We accomplish this via a Crandall and Liggett type method, using recursive inequalities to prove the sequence is Cauchy and then invoking the completeness of W_2 [1, Prop 7.1.5].

First we consider initial data $\mu \in D(|\partial E|)$. In section 3.6, we extend our results to $\mu \in \overline{D(E)}$.

3.1 Almost Contraction Inequality

In this subsection, we use the discrete variational inequality Theorem AGS2 to prove an almost contraction inequality for the discrete gradient flow. (Theorem AGS2 is sufficient for this purpose—we use the stronger discrete variational inequality of Theorem 2.10.1 in a later section.)

Our approach is similar to previous work of Carlen and the author [6], though instead of symmetrizing the contraction inequality, we leave the inequality in an asymmetric form that is more compatible with the asymmetric induction in sections 3.3 and 3.4. The asymmetry is useful a second time when we consider gradient flow with initial conditions $\nu \in \overline{D(E)}$—see section 3.6.

For the $\lambda \leq 0$ case, we follow the proof of [1, Lemma 4.2.4]. For the $\lambda > 0$ case, we use a new approach. In this case, we rely on the fact that $\lambda > 0$ implies E is bounded below [1, Lemma 2.4.8].

THEOREM 3.1.1 (almost contraction inequality). *Suppose E satisfies convexity assumption 2.1.2, $\mu \in D(|\partial E|)$, and $\nu \in \overline{D(E)}$. If $\lambda > 0$, then for all $\tau > 0$,*

$$(1 + \lambda\tau)^2 W_2^2(\mu_\tau, \nu_\tau) \leq W_2^2(\mu, \nu) + \tau^2 |\partial E|^2(\mu) + 2\lambda\tau^2 \left[E(\nu) - \inf E\right] \qquad (3.2)$$

If $\lambda \leq 0$, then for all $0 < \tau < -\frac{1}{\lambda}$,

$$(1 + \lambda\tau)^2 W_2^2(\mu_\tau, \nu_\tau) \leq W_2^2(\mu, \nu) + \tau^2 |\partial E|^2(\mu) . \qquad (3.3)$$

When $\lambda > 0$, $(1 + \lambda\tau)^2$ may be large, and we must compensate with extra terms on the right hand side of (3.2) that are not needed when $\lambda \leq 0$.

Proof. By Theorem AGS2, recalled for the reader's convenience in section 2.10,

$$(1 + \lambda\tau)W_2^2(\mu_\tau, \nu_\tau) - W_2^2(\mu, \nu_\tau) \leq 2\tau \left(E(\nu_\tau) - E(\mu_\tau) - \frac{1}{2\tau} W_2^2(\mu, \mu_\tau)\right) , \qquad (3.4)$$

$$(1 + \lambda\tau)W_2^2(\nu_\tau, \mu) - W_2^2(\nu, \mu) \leq 2\tau \left(E(\mu) - E(\nu_\tau) - \frac{1}{2\tau} W_2^2(\nu, \nu_\tau)\right) . \qquad (3.5)$$

Consider the case $\lambda > 0$. Dropping the $-\frac{1}{2\tau} W_2^2(\nu, \nu_\tau)$ term from (3.5), dividing by $(1 + \lambda\tau)$, and adding to (3.4) gives

$$(1 + \lambda\tau)W_2^2(\mu_\tau, \nu_\tau) - \frac{1}{1 + \lambda\tau} W_2^2(\mu, \nu)$$

$$\leq 2\tau \left(E(\nu_\tau) - \frac{1}{1 + \lambda\tau} E(\nu_\tau) + \frac{1}{1 + \lambda\tau} E(\mu) - E(\mu_\tau) - \frac{1}{2\tau} W_2^2(\mu, \mu_\tau)\right)$$

$$(1 + \lambda\tau)^2 W_2^2(\mu_\tau, \nu_\tau) - W_2^2(\mu, \nu)$$

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

Since $\lambda > 0$, E is bounded below [1, Lemma 2.4.8]. Applying Theorem AGS1 and the fact that $E(\mu_\tau) \leq E(\mu)$, we have

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

$$\leq \tau^2 |\partial E|^2(\mu) + 2\lambda\tau^2 \left[E(\nu) - \inf E\right] ,$$

which gives the result.

Now consider the case $\lambda \leq 0$. Adding (3.4) and (3.5) and then applying Theorem AGS1 gives

$$(1 + \lambda\tau)W_2^2(\mu_\tau, \nu_\tau) - W_2^2(\nu, \mu) + \lambda\tau W_2^2(\nu_\tau, \mu)$$

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

$$\leq \frac{\tau^2}{1 + \lambda\tau}|\partial E|^2(\mu) - W_2^2(\nu, \nu_\tau) . \tag{3.6}$$

Since for $a, b > 0$ and $0 < \epsilon < 1$, the convex function

$$\phi(\epsilon) := \frac{a^2}{\epsilon} + \frac{b^2}{1 - \epsilon}$$

has the minimum value $(a + b)^2$, attained at $\epsilon = a/(a + b)$, we have

$$(a + b)^2 \leq \frac{a^2}{\epsilon} + \frac{b^2}{1 - \epsilon} .$$

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

$$W_2^2(\nu_\tau, \mu) \leq (W_2(\nu_\tau, \nu) + W_2(\nu, \mu))^2 \leq -\frac{1}{\lambda\tau}W_2^2(\nu_\tau, \nu) + \frac{1}{1 + \lambda\tau}W_2^2(\nu, \mu) . \tag{3.7}$$

Multiplying by $-\lambda\tau$, summing with (3.6), multiplying the total by $(1 + \lambda\tau)$, and using the fact that $-\lambda\tau < 1$, we obtain

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

which gives the result. $\qquad\square$

3.2 Relation Between Proximal Maps with Different Time Steps

We now apply the Euler-Lagrange equation, Theorem 2.9.1, to prove a theorem relating the proximal map with a large time step τ to the proximal map with a small time step h. Assumption 2.1.1 is purely for notational convenience. See Theorem A.2.7 for the general case.

THEOREM 3.2.1. *Suppose E satisfies assumptions 2.1.1 and 2.1.2. Then if $\mu \in D(E)$ and $0 < h \leq \tau < \frac{1}{\lambda^-}$,*

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

Figure 3.1: The Hilbertian analogue of Theorem 3.2.1 can be stated as follows: if v is the point on the geodesic from u to $J_\tau u$ at time $\frac{\tau-h}{\tau}$ (i.e. $v = \frac{\tau-h}{\tau}J_\tau u + \frac{h}{\tau}u$) then $J_\tau u = J_h v$.

We may restate the above theorem in terms of the nth step of the discrete gradient flow as follows.

COROLLARY 3.2.2. *Under the assumptions of the previous theorem, if $\mu \in D(E)$, $n \geq 1$,*

$$J_\tau^n \mu = J_\tau(J_\tau^{n-1}\mu) = J_h\left[\left(\frac{\tau-h}{\tau}\mathbf{t}^{J_\tau^n \mu}_{J_\tau^{n-1}\mu} + \frac{h}{\tau}\mathbf{id}\right)\#J_\tau^{n-1}\mu\right] .$$

Proof of Theorem 3.2.1. By Theorem 2.9.1,

$$\boldsymbol{\xi} := \frac{1}{\tau}(\mathbf{t}^{\mu}_{\mu_\tau} - \mathbf{id}) \in \partial E(\mu_\tau)$$

is a strong subdifferential. Next, since $h/\tau < 1$,

$$(\mathbf{id} + h\boldsymbol{\xi}) = \left(\mathbf{id} + \frac{h}{\tau}(\mathbf{t}^{\mu}_{\mu_\tau} - \mathbf{id})\right) = \left(\frac{\tau-h}{\tau}\mathbf{id} + \frac{h}{\tau}\mathbf{t}^{\mu}_{\mu_\tau}\right) . \tag{3.8}$$

is cyclically monotone. Consequently, if we define $\nu := (\mathbf{id} + h\boldsymbol{\xi})\#\mu_\tau$, the transport map is the optimal transport map, $\mathbf{t}^{\nu}_{\mu_\tau} = \mathbf{id} + h\boldsymbol{\xi}$. Rearranging shows

$$\frac{1}{h}(\mathbf{t}^{\nu}_{\mu_\tau} - \mathbf{id}) = \boldsymbol{\xi} \in \partial E(\mu_\tau) ,$$

so by a second application of Theorem 2.9.1, $\mu_\tau = \nu_h$.

We now rewrite ν as it appears in the theorem. By equation (3.8), $(\mathbf{id} + h\boldsymbol{\xi}) = \left(\frac{\tau-h}{\tau}\mathbf{id} + \frac{h}{\tau}\mathbf{t}^{\mu}_{\mu_\tau}\right) = \left(\frac{\tau-h}{\tau}\mathbf{t}^{\mu_\tau}_{\mu} + \frac{h}{\tau}\mathbf{id}\right)\circ\mathbf{t}^{\mu}_{\mu_\tau}$. Thus, $\nu = (\mathbf{id}+h\boldsymbol{\xi})\#\mu_\tau = \left(\frac{\tau-h}{\tau}\mathbf{t}^{\mu_\tau}_{\mu} + \frac{h}{\tau}\mathbf{id}\right)\#\mu.$

$\square$

After proving Theorem 3.2.1, we discovered another proof of the same result in [15, 17]. 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.
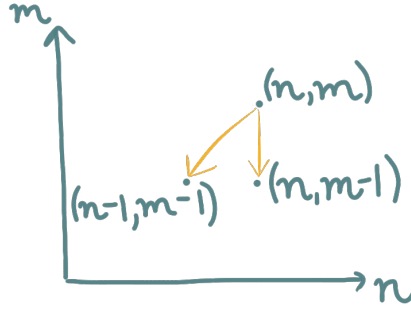

Figure 3.2: The recursive inequality in Theorem 3.3.1 is asymmetric: the (n, m)th term is controlled in terms of the $(n-1, m-1)$th term and the $(n, m-1)$th term.

A fundamental difference between Crandall and Liggett's recursive inequality and Theorem 3.3.1 is that the former involves the distance while the latter involves the square distance. (This is a consequence of the fact that our contraction inequality Theorem 3.1.1 involves the square distance plus error terms.) Therefore, where Crandall and Liggett are able to use the triangle inequality, we have to use the convexity of the square transport metrics. The bulk of the proof is devoted to passing from the transport metrics back to the Wasserstein metric.

THEOREM 3.3.1 (asymmetric recursive inequality). *Suppose E satisfies convexity assumption 2.1.2 and $\mu \in D(|\partial E|)$. If $0 < h \leq \tau < \frac{1}{\lambda^-}$,*

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

To consider $\lambda \geq 0$ and $\lambda < 0$ jointly in the following theorem, 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 and $J_h^m \mu$ by J^m. First, note that

$$(1 - \lambda^- h)^2 W_2^2(J^n, J^m)$$

$$= (1 - \lambda^- h)^2 W_2^2(J_h(\mu_{\frac{\tau-h}{\tau}}^{J^{n-1} \to J^n}), J^m) \qquad \text{by Theorem 3.2.1}$$

$$\leq W_2^2(\mu_{\frac{\tau-h}{\tau}}^{J^{n-1} \to J^n}, J^{m-1}) + h^2 |\partial E|^2(J^{m-1}) \qquad \text{by Theorem 3.1.1}$$

$$\leq W_{2,J^{n-1}}^2(\mu_{\frac{\tau-h}{\tau}}^{J^{n-1} \to J^n}, J^{m-1}) + h^2 |\partial E|^2(J^{m-1})$$

By Proposition 2.6.4, the $W_{2,J^{n-1}}$ metric 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

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

$$+ h^2 |\partial E|^2(J^{m-1}) . \qquad (3.9)$$

The first term on the right hand side coincides with the standard Wasserstein metric. To control the second term, we use the stronger version of the discrete variational inequality Theorem 2.10.1. Specifically, replacing (μ, ν) in Theorem 2.10.1 with (J^{m-1}, J^n) and (J^{n-1}, J^{m-1}) gives

$$(1 - \lambda^- h) W_{2,J^{m-1}}^2(J^m, J^n) - W_2^2(J^{m-1}, J^n)$$

$$\leq 2h \left[E(J^n) - E(J^m) - \frac{1}{2h} W_2^2(J^{m-1}, J^m) \right]$$

$$(1 - \lambda^- \tau) W_{2,J^{n-1}}^2(J^n, J^{m-1}) - W_2^2(J^{n-1}, J^{m-1})$$

$$\leq 2\tau \left[E(J^{m-1}) - E(J^n) - \frac{1}{2\tau} W_2^2(J^{n-1}, J^n) \right]$$

Multiplying the first inequality by τ, the second inequality by h, adding them together,

and then applying Theorem AGS1 gives

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

$$\leq \tau W^2_2(J^{m-1}, J^n) + hW^2_2(J^{n-1}, J^{m-1})$$

$$+ 2\tau h\left[E(J^{m-1}) - E(J^m) - \frac{1}{2h}W^2_2(J^{m-1}, J^m)\right] - hW^2_2(J^{n-1}, J^n)$$

$$\leq \tau W^2_2(J^{m-1}, J^n) + hW^2_2(J^{n-1}, J^{m-1}) + \frac{\tau h^2}{1 - \lambda^- h}|\partial E|^2(J^{m-1}) - hW^2_2(J^{n-1}, J^n) \, .$$

$$(3.10)$$

As in equation (3.7) we have,

$$\lambda^- \tau W^2_{2,J^{n-1}}(J^{m-1}, J^n) \leq W^2_2(J^n, J^{n-1}) + \frac{\lambda^- \tau}{1 - \lambda^- \tau}W^2_2(J^{n-1}, J^{m-1}) \, .$$

Multiplying this by h and adding it to (3.10) gives

$$\tau(1 - \lambda^- h)W^2_{2,J^{m-1}}(J^m, J^n) + hW^2_{2,J^{n-1}}(J^n, J^{m-1})$$

$$\leq \tau W^2_2(J^{m-1}, J^n) + \frac{h}{1 - \lambda^- \tau}W^2_2(J^{n-1}, J^{m-1}) + \frac{\tau h^2}{1 - \lambda^- h}|\partial E|^2(J^{m-1}) \, .$$

Rearranging and dividing by h gives the upper bound

$$W^2_{2,J^{n-1}}(J^{m-1}, J^n) \leq \frac{\tau}{h}\left(W^2_2(J^{m-1}, J^n) - (1 - \lambda^- h)W^2_{2,J^{m-1}}(J^m, J^n)\right)$$

$$+ \frac{1}{1 - \lambda^- \tau}W^2_2(J^{n-1}, J^{m-1}) + \frac{\tau h}{1 - \lambda^- h}|\partial E|^2(J^{m-1}) \, . \quad (3.11)$$

We now combine this with equation (3.9) to prove the theorem. Substituting (3.11) into (3.9) and using $-(1 - \lambda^- h) \leq -(1 - \lambda^- h)^2$ gives

$$(1 - \lambda^- h)^2 W^2_2(J^n, J^m)$$

$$\leq \frac{h}{\tau}W^2_2(J^{n-1}, J^{m-1}) + h^2|\partial E|^2(J^{m-1})$$

$$+ \frac{\tau - h}{\tau}\left[\frac{\tau}{h}\left(W^2_2(J^{m-1}, J^n) - (1 - \lambda^- h)^2 W^2_2(J^m, J^n)\right)\right.$$

$$+ \frac{1}{1 - \lambda^- \tau}W^2_2(J^{m-1}, J^{n-1}) + \left.\frac{\tau h}{1 - \lambda^- h}|\partial E|^2(J^{m-1})\right] \, .$$

Simplifying and rearranging,

$$\frac{\tau}{h}(1 - \lambda^- h)^2 W^2_2(J^n, J^m)$$

$$\leq \left(\frac{h}{\tau} + \frac{\tau - h}{\tau(1 - \lambda^- \tau)}\right)W^2_2(J^{n-1}, J^{m-1}) + \frac{\tau - h}{h}W^2_2(J^n, J^{m-1}) + h^2|\partial E|^2(J^{m-1})$$

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

Therefore,

$$(1 - \lambda^- h)^2 W_2^2(J^n, J^m)$$

$$\leq \frac{h}{\tau} \frac{1 - \lambda^- h}{1 - \lambda^- \tau} W_2^2(J^{n-1}, J^{m-1}) + \frac{\tau - h}{\tau} W_2^2(J^n, J^{m-1}) + \left[\frac{h^3}{\tau} + \frac{h^2}{1 - \lambda^- h} \right] |\partial E|^2(J^{m-1})$$

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

since $0 < h \leq \tau \leq \frac{1}{\lambda^-}$.

Finally, applying Theorem AGS1 and the fact that $(1 - \lambda^- h)^{-1} \leq (1 - \lambda^- h)^{-2}$ gives the result:

$$(1 - \lambda^- h)^2 W_2^2(J^n, J^m)$$

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

$\square$

3.4   Inductive Bound

The following inductive bound follows the simplification of Crandall and Liggett's method introduced by Rasmussen [23, 28]. A key difference is that, in the Banach space case, one works with the distance, rather than the square distance. While this complicated matters in the previous theorem, in simplifies the induction in the following theorem.

We begin by bounding the distance between the 0th and nth terms of the discrete gradient flow.

LEMMA 3.4.1. *Given E as in Assumption 2.1.2 and $\mu \in D(|\partial E|)$, for all $0 < \tau < \frac{1}{\lambda^-}$*

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

Proof. This is follows from the triangle inequality, Theorem AGS1, and the inequalities $\frac{1}{1+\tau\lambda} \leq \frac{1}{1-\tau\lambda^-}$ and $1 \leq \frac{1}{1-\tau\lambda^-}$.

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

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

$\square$

THEOREM 3.4.2 (a Rasmussen type inductive bound). *Suppose E satisfies convexity assumption 2.1.2. Then if $\mu \in D(|\partial E|)$ and $0 < h \leq \tau < \frac{1}{\lambda^-}$,*

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

$$(3.12)$$

Proof. We proceed by induction. The base case, when either $n = 0$ or $m = 0$, follows from the linear growth estimate Lemma 3.4.1. We assume the inequality holds for $(n-1, m)$ and (n, m) and show that this implies it holds for $(n, m+1)$.

First, we apply the Asymmetric Recursive Inequality, Theorem 3.3.1,

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

Next, we divide by $(1 - \lambda^- h)^2$ and apply the inductive hypothesis.

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

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 since $0 < h \leq \tau \leq \frac{1}{\lambda^-}$,

$$(1 - \lambda^- h)^{-2} \leq (1 - \lambda^- \tau)^{-2} \leq (1 - \lambda^- \tau)^{-2n} .$$

Using these estimates, we may group together the three terms and obtain the following

bound.

$$W_2^2(J_\tau^n \mu, J_h^{m+1} \mu)$$

$$\leq \left\{ \frac{h}{\tau} \left[(n\tau - mh)^2 - 2(n\tau - mh)\tau + \tau^2 + \tau hm + 2\tau^2(n-1) \right] \right.$$

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

We now consider the convex combination (plus an additional $2h^2$ term) within the brackets.

$$\frac{h}{\tau} \left[(n\tau - mh)^2 - 2(n\tau - mh)\tau + \tau^2 + \tau hm + 2\tau^2(n-1) \right]$$

$$+ \frac{\tau - h}{\tau} \left[(n\tau - mh)^2 + \tau hm + 2\tau^2 n \right] + 2h^2$$

$$= \frac{h}{\tau} \left[(n\tau - mh)^2 + \tau hm + 2\tau^2 n \right]$$

$$+ \frac{\tau - h}{\tau} \left[(n\tau - mh)^2 + \tau hm + 2\tau^2 n \right] + \frac{h}{\tau} \left[-2(n\tau - mh)\tau - \tau^2 \right] + 2h^2$$

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

$$= (n\tau - mh)^2 - 2(n\tau - mh)h + \tau hm - \tau h + 2\tau^2 n + 2h^2$$

$$= (n\tau - (m+1)h)^2 + h^2 + \tau h(m+1) - \tau h + 2\tau^2 n$$

$$\leq (n\tau - (m+1)h)^2 + \tau h(m+1) + 2\tau^2 n \; .$$

Therefore,

$$W_2^2(J_\tau^n \mu, J_h^{m+1} \mu)$$

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

$$\square$$

3.5 Exponential Formula for the Wasserstein Metric

We now combine our previous results to prove the exponential formula for the Wasserstein metric.

THEOREM 3.5.1 (exponential formula). *Suppose E satisfies convexity assumption 2.1.2. For $\mu \in D(|\partial E|)$, $t \geq 0$, the discrete gradient flow sequence $J_{t/n}^n \mu$ converges as*

$n \to \infty$. *Denote the limit by $\mu(t)$. The convergence is uniform in t on compact subsets of $[0, +\infty)$, and when $n \geq 2\lambda^- t$, the distance between $J_{t/n}^n$ and $\mu(t)$ is bounded by*

$$W_2(J_{t/n}^n \mu, \mu(t)) \leq \sqrt{3} \frac{t}{\sqrt{n}} e^{3\lambda^- t} |\partial E|(\mu) . \tag{3.13}$$

REMARK 3.5.2 (range of $S(t)$). Given $\mu \in D(|\partial E|)$, we may use the fact that $|\partial E|$ is lower semicontinuous [1, Corollary 2.4.10] and Theorem AGS1 to conclude

$$|\partial E|(\mu(t)) \leq \liminf_{n \to \infty} |\partial E|(J_{t/n}^n \mu) \leq \liminf_{n \to \infty} (1 - \lambda^- t/n)^{-n} |\partial E|(\mu) = e^{\lambda^- t} |\partial E|(\mu) .$$

Therefore, $\mu(t) \in D(|\partial E|)$.

We have shown $W_2(J_{t/n}^n \mu, \mu(t)) \leq O(n^{-1/2})$, which agrees with the rate Crandall and Liggett obtained in a Banach space [11]. By a different method, Ambrosio, Gigli, and Savaré showed $W_2(J_{t/n}^n \mu, \mu(t)) \leq O(n^{-1})$ [1, Theorem 4.0.4], which agrees with the optimal rate in a Hilbert space [24]. Our rate improves upon the rate obtained by Clément and Desch [10], $d(J_{t/n}^n \mu, \mu(t)) \leq O(n^{-1/4})$, though they considered the more general case of gradient flow on a metric space (X, d). Still, they also required that Φ be $(\frac{1}{\tau} + \lambda)$ convex.

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. This brings together the Banach space theory with the Wasserstein theory, and it is hoped that this method will help extend the abstract theory of Wasserstein gradient flow to a broader class of functionals.

REMARK 3.5.3 (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 step size goes to zero. See section A.1.

Our estimates lead to a simple proof of the fact that $\mu(t)$ is a λ-contracting semigroup, as originally shown in [1, Proposition 4.3.1].

THEOREM 3.5.4 ($S(t)$ is a λ-contracting semigroup). *Given E satisfying convexity assumption 2.1.2, the function $S(t)$ on $[0, +\infty)$,*

$$S(t) : D(|\partial E|) \to D(|\partial E|) : \mu \mapsto \mu(t)$$

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

Next, we apply the semigroup property (ii) to conclude that $E(\mu(t))$ is nonincreasing.

COROLLARY 3.5.5. *For all $\mu \in D(|\partial E|)$, $E(\mu(t))$ is non-increasing for $t \in [0, +\infty)$.*

Combining the previous results, we prove that $S(t)$ is the continuous gradient flow, in the sense of Definition 2.7.2.

THEOREM 3.5.6 ($\mu(t)$ is the continuous gradient flow)**.** *Given E satisfying convexity assumption 2.1.2 and $\mu \in D(|\partial E|)$, $\mu(t)$ is the continuous gradient flow for E with initial conditions μ. Furthermore,*

$$W_2(\mu(t), \mu(s)) \leq |t - s| e^{\lambda^- t} e^{\lambda^- s} |\partial E|(\mu) , \tag{3.14}$$

so $\mu(t)$ is locally Lipschitz on $[0, +\infty)$.

Finally, we use our method to give a simple proof of the energy dissipation inequality, which shows the regularizing effect of the gradient flow.

COROLLARY 3.5.7 (Energy Dissipation Inequality)**.** *Given E satisfying convexity assumption 2.1.2 and $\mu \in D(|\partial E|)$, for all $t_0, t_1 \geq 0$,*

$$\int_{t_0}^{t_1} |\partial E|^2(\mu(s)) ds \leq E(\mu(t_0)) - E(\mu(t_1)) .$$

We now turn to the proofs of these results.

Proof of Theorem 3.5.1. By Theorem 3.4.2, for fixed $t \geq 0$, if we define $\tau := \frac{t}{n}$, $h := \frac{t}{m}$, with $m \geq n > 2t\lambda^-$, so $0 \leq h \leq \tau < \frac{1}{2\lambda^-}$,

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

$$\tag{3.15}$$

In the second inequality, we use that $(1 - \alpha)^{-1} \leq e^{2\alpha}$ 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. The estimate (3.15) shows that 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 (3.15) gives the error estimate

$$W_2^2(J_{t/n}^n \mu, S(t)\mu) \leq 3 \frac{t^2}{n} e^{6\lambda^- t} |\partial E|^2(\mu) . \tag{3.16}$$

$\square$

Proof of Theorem 3.5.4.

(i) follows from Lemma 3.4.1, since

$$W_2(S(t)\mu, \mu) = \lim_{n \to \infty} W_2(J_{t/n}^n \mu, \mu) \leq \lim_{n \to \infty} \frac{t}{(1 - \lambda^- t/n)^n} |\partial E(\mu)| = t e^{\lambda^- t} |\partial E(\mu)| \xrightarrow{t \to 0} 0 .$$

We now turn to the contraction property (iii). Our proof of the $\lambda > 0$ case is new, using the almost contraction inequality, Theorem 3.1.1. For completeness, we recall the proof of [1, Proposition 4.3.1], which shows the $\lambda \leq 0$ case.

Iterating the contraction inequality from Theorem 3.1.1 for $\lambda > 0$ and applying Theorem AGS1,

$$W_2^2(J_{t/n}^n \mu, J_{t/n}^n \nu) \leq (1 + \lambda(t/n))^{-2n} W_2^2(\mu, \nu)$$
$$+ \sum_{i=1}^{n} \frac{(t/n)^2}{(1 + \lambda(t/n))^{2i}} \left(|\partial E|^2(J_{t/n}^{n-i} \mu) + 2\lambda \left[E(J_{t/n}^{n-i} \nu) - \inf E \right] \right)$$
$$\leq (1 + \lambda(t/n))^{-2n} W_2^2(\mu, \nu)$$
$$+ n(t/n)^2 \left(|\partial E|^2(\mu) + 2\lambda \left[E(\nu) - \inf E \right] \right) . \tag{3.17}$$

Likewise, for $\lambda \leq 0$, $n > -t\lambda$, we have

$$W_2^2(J_{t/n}^n \mu, J_{t/n}^n \nu) \leq (1 + \lambda(t/n))^{-2n} W_2^2(\mu, \nu) + \sum_{i=1}^{n} \frac{(t/n)^2}{(1 + \lambda(t/n))^{2i}} |\partial E|^2(J_{t/n}^{n-i} \mu)$$
$$\leq (1 + \lambda(t/n))^{-2n} W_2^2(\mu, \nu) + \frac{n(t/n)^2}{(1 + \lambda(t/n))^{2n}} |\partial E|^2(\mu) . \tag{3.18}$$

Sending $n \to \infty$ in both cases shows

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

We now prove the semigroup property (ii). First, we show that $S(t)^m\mu = S(mt)\mu$ for fixed $m \in \mathbb{N}$. To consider $\lambda \geq 0$ and $\lambda < 0$ jointly, we replace λ by $-\lambda^- \leq 0$, since any function that is λ convex is also $-\lambda^-$ convex.

First, note that

$$W_2(S(t)^m\mu, (J_{t/n}^n)^m\mu) = W_2(S(t)^m\mu, J_{t/n}^n(J_{t/n}^n)^{m-1}\mu)$$
$$\leq W_2(S(t)^m\mu, J_{t/n}^n S(t)^{m-1}\mu) + W_2(J_{t/n}^n S(t)^{m-1}\mu, J_{t/n}^n(J_{t/n}^n)^{m-1}\mu)$$
$$(3.19)$$

Remark 3.5.2 ensures $S(t)^{m-1}\mu \in D(|\partial E|)$, so by Theorem 3.5.1, $J_{t/n}^n S(t)^{m-1}\mu \xrightarrow{n\to\infty} S(t)^m\mu$. Consequently, we may choose n large enough so that the first term is arbitrarily small for fixed $m \in \mathbb{N}$.

We bound the second term in (3.19) using (3.18). By Remark 3.5.2, $|\partial E|^2(S(t)^{m-1}\mu) \leq e^{2(m-1)\lambda^- t}|\partial E|^2(\mu)$. Therefore,

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

Thus, taking square roots of both sides and combining with (3.19) shows that for all $\epsilon > 0$, these exists n large enough so that

$$W_2(S(t)^m\mu, (J_{t/n}^n)^m\mu) \leq \epsilon + e^{4\lambda^- t}W_2(S(t)^{m-1}\mu, (J_{t/n}^n)^{m-1}\mu) \,.$$

Iterating this shows that for n large enough

$$W_2(S(t)^m\mu, (J_{t/n}^n)^m\mu) \leq \epsilon \sum_{i=0}^{m-1} e^{4i\lambda^- t} + e^{4m\lambda^- t}W_2(\mu, \mu) \leq \epsilon\left(me^{2(m-1)\lambda^- t}\right) \,. \quad (3.20)$$

We now apply this to show $S(t)^m\mu = S(mt)\mu$. By the triangle inequality,

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

The first term can be made arbitrarily small by (3.20). Since $W_2((J_{t/n}^n)^m\mu, S(mt)\mu) = W_2((J_{tm/nm}^{nm})\mu, S(mt)\mu)$, by Theorem 3.5.1 we may choose n large so the second term is arbitrarily small. Therefore, $S(t)^m\mu = S(mt)\mu$.

This shows shows that 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 $t, s \geq 0$. □

Proof of Corollary 3.5.5. For $t \geq 0$, the lower semicontinuity of E and definition of the proximal map (2.19) imply

$$E(S(t)\mu) \leq \liminf_{n\to\infty} E(J^n_{t/n}\mu) \leq \liminf_{n\to\infty} E(\mu) = E(\mu) \ .$$

The result then follows from the semigroup property, Theorem 3.5.4 (ii). □

Proof of Theorem 3.5.6. First, we show that $S(t)\mu$ is locally Lipschitz continuous in t. Given $t, s \geq 0$, define $\tau := \frac{t}{n}, h := \frac{s}{m}$ for m and n large enough so that $0 \leq h \leq \tau < \frac{1}{\lambda^-}$. By Theorem 3.4.2,

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

(3.21)

Sending $n, m \to \infty$ and taking the square root of both sides gives

$$W_2(S(t)\mu, S(s)\mu) \leq |t - s|e^{\lambda^- t}e^{\lambda^- s}|\partial E|(\mu) \ . \tag{3.22}$$

We now turn to the proof that $S(t)\mu$ is the continuous gradient flow for E with initial conditions μ in the sense of Definition 2.7.2. We already showed $S(t)\mu \xrightarrow{t \to 0} \mu$ in part (i) of Theorem 3.5.4, so it remains to show that $S(t)\mu$ satisfies (2.17).

Iterating Theorem AGS2 with $\tau = t/n < \frac{1}{\lambda^-}$ shows that for all $\omega \in D(E)$,

$$(1 + \lambda t/n)^n W_2^2(J^n_{t/n}\mu, \omega) \leq W_2^2(\mu, \omega) + 2(t/n)\sum_{i=1}^{n}[E(\omega) - E(J^i_{t/n}\mu)](1 + \lambda t/n)^{i-1}$$

(3.23)

Consider the piecewise constant function

$$g_n(s) := [E(J^i_{t/n}\mu) - E(\omega)](1 + \lambda t/n)^{i-1} \quad \text{for } s \in ((i-1)t/n, it/n] \ , 1 \leq i \leq n \ .$$

We may rewrite the second term on the right hand side of (3.23) as $-2\int_0^t g_n(s)ds$. 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. Applying Fatou's lemma,

$$\liminf_{n\to\infty} \int_0^t g_n(s)ds \geq \int_0^t \liminf_{n\to\infty} g_n(s)ds . \tag{3.24}$$

By Theorem 3.4.2, for m, n large enough so that $\frac{s}{m} \leq \frac{t}{n} < \frac{1}{2\lambda^-}$,

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

$$\leq \left(\left(\frac{t}{n}\right)^2 + \frac{ts + 2t^2}{n}\right) e^{4\lambda^- t}e^{4\lambda^- s}|\partial E|(\mu) ,$$

where the second inequality follows from the fact that $s \in ((i-1)t/n, it/n]$. Since $J_{s/m}^m\mu \xrightarrow{m\to\infty} S(s)\mu$, this shows that $J_{t/n}^i\mu \xrightarrow{n\to\infty} S(s)\mu$. Combining with the lower semicontinuity of E gives

$$\liminf_{n\to\infty} g_n(s)ds = \liminf_{n\to\infty}[E(J_{t/n}^i\mu) - E(\omega)](1 + \lambda t/n)^{i-1} \geq [E(S(s)\mu) - E(\omega)]e^{\lambda s} . \tag{3.25}$$

Likewise, combining (3.24) and (3.25) shows that taking $\liminf_{n\to\infty}$ of (3.23) gives

$$e^{\lambda t}W_2^2(S(t)\mu, \omega) \leq W_2^2(\mu, \omega) + 2\int_0^t [E(\omega) - E(S(s)\mu)]e^{\lambda s}ds . \tag{3.26}$$

By part (ii) of Theorem 3.5.4, we have for all $t, t_0 \geq 0$,

$$e^{\lambda t}W_2^2(S(t + t_0)\mu, \omega) \leq W_2^2(S(t_0)\mu, \omega) + 2\int_0^t [E(\omega) - E(S(s + t_0)\mu)]e^{\lambda s}ds \tag{3.27}$$

$$= W_2^2(S(t_0)\mu, \omega) + 2\int_{t_0}^{t+t_0} [E(\omega) - E(S(s)\mu)]e^{\lambda(s-t_0)}ds . \tag{3.28}$$

Hence,

$$e^{\lambda(t+t_0)}W_2^2(S(t + t_0)\mu, \omega) - e^{\lambda t_0}W_2^2(S(t_0)\mu, \omega) \leq 2\int_{t_0}^{t+t_0} [E(\omega) - E(S(s)\mu)]e^{\lambda s}ds . \tag{3.29}$$

It remains to divide (3.29) by t and send $t \to 0$ to get (2.17). The left hand side will converge for a.e. t_0 since the function $f(t) = \frac{1}{2}W_2^2(S(t)\mu, \omega)$ is locally Lipschitz.

In particular, for $t, s \in [0, T]$,

$$
\begin{aligned}
|f(t) - f(s)| &\leq \left| \frac{1}{2} W_2^2(S(t)\mu, \omega) - \frac{1}{2} W_2^2(S(s)\mu, \omega) \right| \\
&\leq \frac{1}{2} \left(W_2(S(t)\mu, \omega) + W_2(S(s)\mu, \omega) \right) |W_2(S(t)\mu, \omega) - W_2(S(s)\mu, \omega)| \\
&\leq \left[\max_{t \in [0,T]} W_2(S(t)\mu, \omega) \right] W_2(S(t)\mu, S(s)\mu) \\
&\leq \left[\max_{t \in [0,T]} W_2(S(t)\mu, \mu) + W_2(\mu, \omega) \right] W_2(S(t)\mu, S(s)\mu) \\
&\leq C_{T,\omega} |t - s| \, ,
\end{aligned}
$$

where the last inequality follows by the fact that $S(t)\mu$ is locally Lipschitz (3.22).

If we divide the right hand side of (3.29) by t and send $t \to 0$, it will also converge, since

$$
E(\omega) - E(S(t_0)) \leq E(\omega) - E(S(s)) \leq E(\omega) - E(S(t_0 + t))
$$

for $s \in [t_0, t + t_0]$ and $\liminf_{t \to 0} -E(S(t_0 + t)\mu) \leq -E(S(t_0))$. Therefore,

$$
\frac{d}{dt} e^{\lambda t} W_2^2(S(t)\mu, \omega) \leq 2e^{\lambda t} [E(\omega) - E(S(t)\mu)] \quad \text{for Lebesgue a.e. } t > 0, \forall \omega \in D(E) \, .
$$

Rearranging shows this is equivalent to (2.17):

$$
\frac{1}{2} \frac{d}{dt} W_2^2(S(t)\mu, \omega) + \frac{\lambda}{2} W_2^2(S(t)\mu, \omega) \leq [E(\omega) - E(S(t)\mu)] \quad \text{for a.e. } t > 0, \forall \omega \in D(E) \, .
$$

$\square$

Proof of Corollary 3.5.7. By the semigroup property, Theorem 3.5.4 *(ii)*, it is enough to prove the result for $t_0 = 0$, $t_1 = t$. Theorem AGS1 provides the following bounds on the discrete gradient flow:

$$
\frac{\tau}{2} (1 + \lambda \tau) |\partial E|^2(\mu_\tau) + \frac{1}{2\tau} W_2^2(\mu, \mu_\tau) \leq E(\mu) - E(\mu_\tau) \, ,
$$

$$
\frac{\tau}{2} |\partial E|^2(\mu_\tau) \leq \frac{1}{2\tau} W_2^2(\mu, \mu_\tau) \, .
$$

Combining these shows

$$
\tau |\partial E|^2(\mu_\tau) + \frac{\lambda \tau^2}{2} |\partial E|^2(\mu_\tau) \leq E(\mu) - E(\mu_\tau) \, .
$$

Summing this inequality along the first n elements of the discrete gradient flow,

$$\sum_{i=1}^{n} \tau |\partial E|^2(J_\tau^i \mu) + \frac{\lambda \tau^2}{2} |\partial E|^2(J_\tau^i \mu) \leq E(\mu) - E(J_\tau^n \mu) . \tag{3.30}$$

As in the proof of the previous theorem, if we define the piecewise function

$$g_n(s) := |\partial E|^2(J_\tau^i \mu) + \frac{\lambda \tau}{2} |\partial E|^2(J_\tau^i \mu) \quad \text{for } s \in ((i-1)\tau, i\tau] , 1 \leq i \leq n ,$$

the left hand side of (3.30) becomes $\int_0^{\tau n} g_n(s)ds$. Suppose $\tau = t/n$. Then taking $\liminf_{n \to \infty}$ of both side of (3.30) and applying Fatou's lemma and the lower semicontinuity of E shows

$$\int_0^t \liminf_{n \to \infty} g_n(s)ds \leq E(\mu) - E(S(t)\mu) .$$

Since $\tau |\partial E|^2(J_\tau^i \mu) \leq \frac{2}{1+\lambda \tau}(E(\mu) - E(J_\tau \mu) - \frac{1}{2\tau} W_2^2(\mu, J_\tau \mu)) \xrightarrow{\tau \to 0} 0$ [1, Lemma 3.1.2], the second term in $g_n(s)$ goes to zero pointwise. As in the proof of the previous theorem, $J_{t/n}^i \mu \xrightarrow{n \to \infty} S(s)\mu$. Thus, by the lower semicontinuity of $|\partial E|$,

$$\int_0^t |\partial E|^2(S(s)\mu)ds \leq E(\mu) - E(S(t)\mu) .$$

$\square$

3.6 Gradient Flow with Initial Conditions $\mu \in \overline{D(E)}$

In this section, we describe how to extend our results to accommodate initial data $\mu \in \overline{D(E)}$, rather than just $\mu \in D(|\partial E|)$. In doing this, we repeatedly use two facts. First, we may approximate any $\nu \in \overline{D(E)}$ by elements in $D(|\partial E|)$, since $\nu_\tau \in D(|\partial E|)$ by Theorem AGS1, and $\nu_\tau \xrightarrow{\tau \to 0} \nu$ by [1, Lemma 3.1.2]. Second, we may iterate the contraction inequality from Theorem 3.1.1 as in (3.17) and (3.18) to show that, for all $\nu \in \overline{D(E)}, \mu \in D(|\partial E|)$,

$$W_2^2(J_{t/n}^n \mu, J_{t/n}^n \nu) \leq (1 + \lambda(t/n))^{-2n} W_2^2(\mu, \nu) + n(t/n)^2 \left(|\partial E|^2(\mu) + 2\lambda \left[E(\nu) - \inf E\right]\right) \tag{3.31}$$

when $\lambda > 0$ and

$$W_2^2(J_{t/n}^n \mu, J_{t/n}^n \nu) \leq (1 - \lambda^-(t/n))^{-2n} W_2^2(\mu, \nu) + \frac{n(t/n)^2}{(1 - \lambda^-(t/n))^{2n}} |\partial E|^2(\mu) \tag{3.32}$$

when $\lambda \leq 0$. (As usual, we will at times consider $\lambda \geq 0$ and $\lambda < 0$ jointly, by replacing λ by $-\lambda^-$: any function that is λ convex is also $-\lambda^-$ convex.)

Combining these two facts with our previous results for continuous gradient flow with initial data in $D(|\partial E|)$ gives the following corollaries.

COROLLARY 3.6.1. *For all $\nu \in \overline{D(E)}$, the limit of the discrete gradient flow sequence $J_{t/n}^n \nu$ exists. Denote this limit by $\nu(t)$. The convergence is uniform in t on compact subsets of $[0, +\infty)$.*

COROLLARY 3.6.2. *For $\nu, \tilde{\nu} \in \overline{D(E)}$, continuous gradient flow is a λ-contracting semigroup:*

(i) $\lim_{t\to 0} S(t)\nu = S(0)\nu = \nu$

(ii) $S(t+s)\nu = S(t)S(s)\nu$ for $t, s \geq 0$

(iii) $W_2(S(t)\nu, S(t)\tilde{\nu}) \leq e^{-\lambda t} W_2(\nu, \tilde{\nu})$

COROLLARY 3.6.3. *For all $\nu \in \overline{D(E)}$, $E(\nu(t))$ is non-increasing for $t \in [0, +\infty)$.*

COROLLARY 3.6.4. *$\nu(t)$ is the continuous gradient flow for E with initial conditions ν, in the sense of Definition 2.7.2. For $t \in (0, +\infty)$, $\nu(t)$ belongs to $D(E)$ and is locally Lipschitz continuous.*

COROLLARY 3.6.5. *Given E satisfying convexity assumption 2.1.2 and $\mu \in \overline{D(E)}$*

$$\int_{t_0}^{t_1} |\partial E|^2(\mu(s))ds \leq E(\mu(t_0)) - E(\mu(t_1)) \quad \text{for all } t_0, t_1 \geq 0 \ .$$

Furthermore, $|\partial E|(S(t)\mu) < +\infty$ for all $t > 0$.

Proof of Corollary 3.6.1. By the triangle inequality, for all $\mu \in D(|\partial E|)$,

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

Fix $\epsilon > 0$ and $\mu \in D(|\partial E|)$ so that $e^{-2\lambda t} W_2(\mu, \nu) < \epsilon$ for all $t \in [0, T]$. By (3.32), we may choose n, m large enough, uniformly in $t \in [0, T]$, so that both the first and third terms are less than 2ϵ. By Theorem 3.5.1, we may choose n, m large enough, uniformly in $t \in [0, T]$, so that the second term is less than ϵ.

Thus, the sequence $J_{t/n}^n \nu$ is Cauchy uniformly in $t \in [0,T]$, so the limit exists and convergence is uniform for $t \in [0,T]$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Corollary 3.6.2. First, we prove the continuous time contraction property (iii). Sending $n \to \infty$ in (3.31) and (3.32) shows that for any $\nu \in \overline{D(E)}, \mu \in D(|\partial E|)$,

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

By the triangle inequality, for any $\mu \in D(|\partial E|)$,

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

Sending $\mu \to \tilde{\nu}$ gives the result.

Next, we prove (i). By the triangle inequality and (iii),

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

Choosing μ arbitrarily close to ν and sending $t \to 0$, the result follows from the corresponding result for $\mu \in D(|\partial E|)$, Theorem 3.5.4 (i).

Finally, we show (ii). By (iii) and the corresponding result for $\mu \in D(|\partial E|)$, Theorem 3.5.4 (ii),

$$W_2(S(t+s)\nu, S(t)S(s)\nu) \leq W_2(S(t+s)\nu, S(t+s)\mu) + W_2(S(t+s)\mu, S(t)S(s)\mu)$$
$$+ W_2(S(t)S(s)\mu, S(t)S(s)\nu)$$
$$\leq 2e^{-\lambda(t+s)} W_2(\mu, \nu) \ .$$

Sending $\mu \to \nu$ shows $S(t+s)\nu = S(t)S(s)\nu$. $\qquad\qquad\qquad\qquad\qquad\square$

Proof of Corollary 3.6.3. The same argument for Corollary 3.5.5 applies. $\qquad\square$

Proof of Corollary 3.6.4. To show that for $t \in (0, +\infty)$, $S(t)\nu$ belongs to $D(E)$ and is locally Lipschitz continuous, we follow the proof of [17, Lemma 2.8, Theorem 2.9], which we recall for the reader's convenience. Though developed for metric spaces of

nonpositive curvature, it applies to our setting without modification. Without loss of generality, we suppose $\lambda^- > 0$.

Iterating the discrete variational inequality, Theorem AGS2, and using that $E(J_\tau^i \nu) \le E(J_\tau^{i-1} \nu)$,

$$W_2^2(J_\tau^n \nu, \omega) \le \frac{1}{(1 - \lambda^- \tau)^n} W_2^2(\nu, \omega) + 2\tau \left[E(\omega) - E(J_\tau^n \nu)\right] \sum_{i=1}^{n} \frac{1}{(1 - \lambda^- \tau)^i} , \quad (3.33)$$

for all $\omega \in D(E)$. Since $0 < \tau < \frac{1}{\lambda^-}$,

$$\sum_{i=1}^{n} \frac{1}{(1 - \lambda^- \tau)^i} \le \frac{(1 - \lambda^- \tau)^{-n} - 1}{\lambda^- \tau} ,$$

Since E is lower semicontinuous, setting $\tau = s/n$ and sending $n \to \infty$ in (3.33) gives

$$W_2^2(S(s)\nu, \omega) \le e^{\lambda^- s} W_2^2(\nu, \omega) + 2\frac{e^{\lambda^- s} - 1}{\lambda^-} \left[E(\omega) - E(S(s)\nu)\right] .$$

This shows $S(s)\nu \in D(E)$ for $s > 0$. By the semigroup property, Corollary 3.6.2 (ii),

$$W_2^2(S(t + s)\nu, \omega) \le e^{\lambda^- s} W_2^2(S(t)\nu, \omega) + 2\frac{e^{\lambda^- s} - 1}{\lambda^-} \left[E(\omega) - E(S(t + s)\nu)\right] .$$

Taking $\omega = S(t)\nu \in D(E)$ and rearranging gives

$$\frac{W_2^2(S(t + s)\nu, S(t)\nu)}{s^2} \le 2\frac{e^{\lambda^- s} - 1}{\lambda^- s} \frac{E(S(t)\nu) - E(S(t + s)\nu)}{s} . \quad (3.34)$$

By Corollary 3.6.3, $E(S(t)\nu)$ is non-increasing. Since $E(S(t)\mu)$ is finite for $t > 0$, we conclude that the limit as $s \to 0$ of

$$\frac{E(S(t)\nu) - E(S(t + s)\nu)}{s}$$

exists for a.e. $t > 0$. Since $\lim_{s \to 0} \frac{e^{\lambda^- s} - 1}{\lambda^- s} = 1$, for any such t, there exists an ϵ such that for $|s| < \epsilon$, the right hand side of (3.34) is bounded. Therefore, for any $t_0 > 0$, if the limit of the right hand side exists at $t \in (0, t_0]$, we may use properties (ii) and (iii) of Corollary 3.6.2 to conclude,

$$W_2(S(t_0 + s)\nu, S(t)\nu) = W_2(S(t_0 - t)S(t + s)\nu, S(t_0 - t)S(t)\nu)$$

$$\le e^{\lambda^-(t_0 - t)} W_2(S(t + s)\nu, S(t)\nu)$$

$$\le e^{\lambda^-(t_0 - t)} C_0 s \text{ for } |s| < \epsilon .$$

Therefore, $S(t)\nu$ is locally Lipschitz on $(0, +\infty)$.

Finally, we show that $S(t)\nu$ is the continuous gradient flow for E with initial conditions ν, in the sense of Definition 2.7.2. We already showed $\lim_{t\to 0} S(t)\nu = S(0)\nu = \nu$ in Corollary 3.6.2 (i), so it remains to show that for all $\omega \in D(E)$

$$\frac{1}{2}\frac{d}{dt}W_2^2(S(t)\nu, \omega) + \frac{\lambda}{2}W_2^2(S(t)\nu, \omega) \leq E(\omega) - E(S(t)\nu) \quad \text{for Lebesgue a.e. } t > 0 \ .$$

$$(3.35)$$

Since $S(t)\nu$ is locally Lipschitz on $(0, +\infty)$ and $J_{t/n}^n\nu \to S(t)\nu$ uniformly in $t \in [0, T]$, this follows by the same argument as in the proof of Theorem 3.5.6, with a small modification in the proof that $J_{t/n}^i\nu \xrightarrow{n\to\infty} S(s)\nu$. This follows by approximating ν by $\mu \in D(|\partial E|)$, for which we have $J_{t/n}^i\mu \xrightarrow{n\to\infty} S(s)\mu$. $\hfill\square$

Proof of Corollary 3.6.5. The same argument as for Corollary 3.5.7 applies to prove the inequality. The inequality implies that $|\partial E|^2(S(s)\mu) < +\infty$ for almost every $s \geq 0$. By the semigroup property, Theorem 3.6.2 (ii), and Remark 3.5.2, $|\partial E|(S(t_1)\mu) \leq e^{\lambda^-(t_1-t_0)}|\partial E|(S(t_0)\mu)$ for $t_0 < t_1$. Therefore, $|\partial E|(\mu_t) < +\infty$ for all $t > 0$. $\hfill\square$

Appendix

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 h of varying time steps. For simplicity of notation, we write $J^m := \prod_{k=1}^{m} J_{h_k}\mu$ and $J^n := J_\tau^n \mu$.

First, we prove a generalization of the asymmetric recursive inequality, Theorem 3.3.1.

THEOREM A.1.1 (asymmetric recursive inequality)**.** *Suppose E satisfies assumptions 2.1.1 and 2.1.2 and $\mu \in D(|\partial E|)$. If $0 < h_i \leq \tau < \frac{1}{\lambda^-}$,*

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

Proof. To simplify notation, we abbreviate $J_\tau^n \mu$ by J^n and $\prod_{i=1}^{m} J_{h_i}\mu$ by J^m. First, note that

$$(1 - \lambda^- h_m)^2 W_2^2(J^n, J^m)$$

$$= (1 - \lambda^- h_m)^2 W_2^2(J_{h_m}(\mu_{\frac{\tau - h_m}{\tau}}^{J^{n-1} \to J^n}), J^m) \qquad \text{by Theorem 3.2.1}$$

$$\leq W_2^2(\mu_{\frac{\tau - h_m}{\tau}}^{J^{n-1} \to J^n}, J^{m-1}) + h_m^2|\partial E|^2(J^{m-1}) \qquad \text{by Theorem 3.1.1}$$

$$\leq W_{2,J^{n-1}}^2(\mu_{\frac{\tau - h_m}{\tau}}^{J^{n-1} \to J^n}, J^{m-1}) + h_m^2|\partial E|^2(J^{m-1})$$

By Proposition 2.6.4, the $W_{2,J^{n-1}}$ metric is convex along generalized geodesics with

base J^{n-1}. In particular, it is convex along the geodesic $\mu_{\frac{\tau-h_m}{\tau}}^{J^{n-1}\to J^n}$, which gives

$$(1-\lambda^- h_m)^2 W_2^2(J^n, J^m) \leq \frac{h_m}{\tau} W_{2,J^{n-1}}^2(J^{m-1}, J^{n-1}) + \frac{\tau-h_m}{\tau} W_{2,J^{n-1}}^2(J^{m-1}, J^n)$$

$$+ h_m^2 |\partial E|^2(J^{m-1}) \ . \tag{A.1}$$

The first term on the right hand side coincides with the standard Wasserstein metric. To control the second term, we use the stronger version of the discrete variational inequality Theorem 2.10.1. Specifically, replacing (μ, ν) in Theorem 2.10.1 with (J^{m-1}, J^n) and (J^{n-1}, J^{m-1}) gives

$$(1-\lambda^- h_m) W_{2,J^{m-1}}^2(J^m, J^n) - W_2^2(J^{m-1}, J^n)$$

$$\leq 2h_m \left[E(J^n) - E(J^m) - \frac{1}{2h_m} W_2^2(J^{m-1}, J^m) \right]$$

$$(1-\lambda^- \tau) W_{2,J^{n-1}}^2(J^n, J^{m-1}) - W_2^2(J^{n-1}, J^{m-1})$$

$$\leq 2\tau \left[E(J^{m-1}) - E(J^n) - \frac{1}{2\tau} W_2^2(J^{n-1}, J^n) \right]$$

Multiplying the first inequality by τ, the second inequality by h_m, adding them together, and then applying Theorem AGS1 gives

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

$$\leq \tau W_2^2(J^{m-1}, J^n) + h_m W_2^2(J^{n-1}, J^{m-1})$$

$$+ 2\tau h_m \left[E(J^{m-1}) - E(J^m) - \frac{1}{2h_m} W_2^2(J^{m-1}, J^m) \right] - h_m W_2^2(J^{n-1}, J^n)$$

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

$$\tag{A.2}$$

As in equation (3.7) 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 this by h_m and adding it to (A.2) gives

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

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

Rearranging and dividing by h_m gives the upper bound

$$W_{2,J^{n-1}}^2(J^{m-1}, J^n) \le \frac{\tau}{h_m} \left(W_2^2(J^{m-1}, J^n) - (1 - \lambda^- h_m)W_{2,J^{m-1}}^2(J^m, J^n) \right)$$
$$+ \frac{1}{1 - \lambda^- \tau}W_2^2(J^{n-1}, J^{m-1}) + \frac{\tau h_m}{1 - \lambda^- h_m}|\partial E|^2(J^{m-1}) . \quad \text{(A.3)}$$

We now combine this with equation (A.1) to prove the theorem. Substituting (A.3) into (A.1) and using $-(1 - \lambda^- h_m) \le -(1 - \lambda^- h_m)^2$ gives

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

Simplifying and rearranging,

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

Therefore,

$$(1 - \lambda^- h_m)^2 W_2^2(J^n, J^m)$$
$$\le \frac{h_m}{\tau} \frac{1 - \lambda^- h_m}{1 - \lambda^- \tau} W_2^2(J^{m-1}, J^{n-1}) + \frac{\tau - h_m}{\tau} W_2^2(J^{m-1}, J^n)$$
$$+ \left[\frac{h_m^3}{\tau} + \frac{h_m^2}{1 - \lambda^- h_m} \right] |\partial E|^2(J^{m-1})$$
$$\le \frac{h_m}{\tau} \frac{1}{1 - \lambda^- \tau} W_2^2(J^{m-1}, J^{n-1}) + \frac{\tau - h_m}{\tau} W_2^2(J^{m-1}, J^n) + \frac{2h_m^2}{1 - \lambda^- h_m}|\partial E|^2(J^{m-1}) ,$$

since $0 \le h_k \le \tau \le \frac{1}{\lambda^-}$ for all $k = 1, \dots, m$. Finally, applying Theorem AGS1 and then

the fact that $(1 - \lambda^- h_m)^{-1} \leq (1 - \lambda^- h_m)^{-2}$ gives the result:

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

$\square$

Next, we prove a generalization of Lemma 3.4.1, bounding the distance between the 0th and mth terms of the discrete gradient flow sequence.

THEOREM A.1.2 (generalization of Lemma 3.4.1). *Suppose E satisfies convexity assumption 2.1.2 and $\mu \in D(|\partial E|)$, for all $0 < h_i < \frac{1}{\lambda^-}$*

$$W_2(J^m, \mu) \leq |\partial E(\mu)| \sum_{i=1}^{m} h_i \prod_{k=1}^{m}(1 - \lambda^- h_k)^{-1} .$$

Proof.

$$W_2(J^m, \mu) \leq \sum_{i=1}^{m} W_2(J^i, J^{i-1}) \leq \sum_{i=1}^{m} \frac{h_i}{1 + h_i \lambda} |\partial E(J^{i-1})|$$
$$\leq \sum_{i=1}^{m} h_i |\partial E(\mu)| \prod_{k=1}^{i}(1 - \lambda^- h_k)^{-1} \leq |\partial E(\mu)| \sum_{i=1}^{m} h_i \prod_{k=1}^{m}(1 - \lambda^- h_k)^{-1} .$$

$\square$

Finally, we prove the generalization of the inductive bound, Theorem 3.4.2.

THEOREM A.1.3 (a Rasmussen type inductive bound). *Suppose E satisfies assumptions 2.1.1 and 2.1.2 and $\mu \in D(|\partial E|)$. If $0 < h_k \leq \tau < \frac{1}{\lambda^-}$ and $S_m := \sum_{k=1}^{m} h_k$,*

$$W_2^2(J^n, J^m) \leq \left[(n\tau - S_m)^2 + \tau S_m + 2\tau^2 n\right](1 - \lambda^- \tau)^{-2n} \left\{ \prod_{k=1}^{m}(1 - \lambda^- h_k)^{-2} \right\} |\partial E|^2(\mu) .$$

(A.4)

Proof. We proceed by induction. The base cases $(n, 0)$ and $(0, m)$ follow from the linear growth estimate Lemma A.1.2. We assume the inequality holds for $(n-1, m)$ and (n, m) and show that this implies it holds for $(n, m+1)$.

First, we apply the Asymmetric Recursive Inequality, Theorem A.1.1,

$$(1 - \lambda^- h_{m+1})^2 W_2^2(J^n, J^{m+1})$$

$$\leq \frac{h_{m+1}}{\tau}(1 - \lambda^- \tau)^{-1} W_2^2(J^{n-1}, M^m) + \frac{\tau - h_{m+1}}{\tau} W_2^2(J^n, J^m)$$

$$+ 2h_{m+1}^2 \prod_{k=1}^{m+1}(1 - \lambda^- h_k)^{-2} |\partial E|^2(\mu) \ .$$

Next, we divide by $(1 - \lambda^- h_{m+1})^2$ and apply the inductive hypothesis.

$$W_2^2(J^n \mu, J^{m+1}\mu)$$

$$\leq \left\{ \frac{h_{m+1}}{\tau}\left[((n-1)\tau - S_m)^2 + \tau S_m + 2\tau^2(n-1)\right](1 - \lambda^- \tau)^{-2(n-1)-1} \right.$$

$$\left. + \frac{\tau - h_{m+1}}{\tau}\left[(n\tau - S_m)^2 + \tau S_m + 2\tau^2 n\right](1 - \lambda^- \tau)^{-2n} + 2h_{m+1}^2 \right\}$$

$$\cdot (1 - \lambda^- h_{m+1})^{-2}\left[\prod_{k=1}^{m+1}(1 - \lambda^- h_k)^{-2}\right]|\partial E|^2(\mu) \ .$$

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

$$\left[((n-1)\tau - S_m)^2 + \tau S_m + 2\tau^2(n-1)\right]$$

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

To control the third term, note that since $0 < h_{m+1} \leq \tau \leq \frac{1}{\lambda^-}$,

$$(1 - \lambda^- h_{m+1})^{-2} \leq (1 - \lambda^- \tau)^{-2} \leq (1 - \lambda^- \tau)^{-2n} \ .$$

Using these estimates, we may group together the three terms and obtain the following bound.

$$W_2^2(J^n, J^{m+1}) \leq \left\{ \frac{h_{m+1}}{\tau}[(n\tau - S_m)^2 - 2(n\tau - S_m)\tau + \tau^2 + \tau S_m + 2\tau^2(n-1)] \right.$$

$$\left. + \frac{\tau - h_{m+1}}{\tau}[(n\tau - S_m)^2 + \tau S_m + 2\tau^2 n] + 2h_{m+1}^2 \right\}$$

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

We now consider the convex combination (plus the additional $2h_{m+1}^2$ term) within the

first brackets.

$$\frac{h_{m+1}}{\tau} \left[(n\tau - S_m)^2 - 2(n\tau - S_m)\tau + \tau^2 + \tau S_m + 2\tau^2(n-1) \right]$$

$$+ \frac{\tau - h_{m+1}}{\tau} \left[(n\tau - S_m)^2 + \tau S_m + 2\tau^2 n \right] + 2h_{m+1}^2$$

$$= \frac{h_{m+1}}{\tau} \left[(n\tau - S_m)^2 + \tau S_m + 2\tau^2 n \right] + \frac{\tau - h_{m+1}}{\tau} \left[(n\tau - S_m)^2 + \tau S_m + 2\tau^2 n \right]$$

$$+ \frac{h_{m+1}}{\tau} \left[-2(n\tau - S_m)\tau + \tau^2 - 2\tau^2 \right] + 2h_{m+1}^2$$

$$= \left[(n\tau - S_m)^2 + \tau S_m + 2\tau^2 n \right] - 2(n\tau - S_m)h_{m+1} - \tau h_{m+1} + 2h_{m+1}^2$$

$$= (n\tau - S_m)^2 - 2(n\tau - S_m)h_{m+1} + \tau S_m + 2\tau^2 n - \tau h_{m+1} + 2h_{m+1}^2$$

$$= (n\tau - S_{m+1})^2 - h_{m+1}^2 + \tau S_m + 2\tau^2 n - \tau h_{m+1} + 2h_{m+1}^2$$

$$= (n\tau - S_{m+1})^2 + \tau S_m + 2\tau^2 n + h_{m+1}^2 - \tau h_{m+1}$$

$$\leq (n\tau - S_{m+1})^2 + \tau S_{m+1} + 2\tau^2 n \ .$$

In the last line, we again use that $h_{m+1} \leq \tau$. This shows

$$W_2^2(J^n \mu, J^{m+1}\mu)$$

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

$\square$

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

THEOREM A.1.4 (exponential formula, varying time steps)**.** *Suppose E satisfies assumptions 2.1.1 and 2.1.2. 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 continuous gradient flow $S(t)\mu$ 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 error 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. The result holds trivially for $t = 0$.

Suppose $0 < h_k \leq \tau < \frac{1}{2\lambda^-}$. By the triangle inequality, Theorem 3.4.2, Theorem A.1.3, and the fact that $(1 - \alpha)^{-1} \leq e^{2\alpha}$ for $\alpha \in [0, 1/2]$,

$$W_2\left(S(t)\mu, \prod_{i=1}^{m} J_{h_i}\mu\right)$$

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

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

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

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

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

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

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

Define $\tau := |\boldsymbol{h}|$ and let n be the greatest integer less than or equal to $\frac{t}{|\boldsymbol{h}|} = \frac{t}{\tau}$. Then,

$$\frac{t}{\tau} - 1 < n \leq \frac{t}{\tau} \implies t - \tau < n\tau \leq t \implies -\tau < n\tau - t \leq 0 .$$

Consequently,

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

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

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

$\square$

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

In this section, we discuss how to extend our results for functionals that do not satisfy the domain assumption 2.1.1. We used this assumption to ensure that for all $\omega \in D(E)$, there exists an optimal transport map $\mathbf{t}_\omega^\mu$ from ω to any $\mu \in \mathcal{P}_{2,\omega}(\mathbb{R}^d)$ [18]. This allowed us to define the $(2,\omega)$-transport metric for any $\omega \in D(E)$. The convexity of $W_{2,\omega}$ along generalized geodesics with base ω and the triangle inequality with reference point ω

$$W_{2,\omega}(\mu, \nu) \leq W_2(\mu, \omega) + W_2(\omega, \nu) \tag{A.5}$$

allowed us to prove the stronger discrete variational inequality (Theorem 2.10.1) and the asymmetric recursive inequality (Theorem 3.3.1). In addition, the subdifferential with respect to the transport metric allowed us to prove the Euler-Lagrange equation for the proximal map (Theorem 2.9.1), with which we were able to relate proximal maps with different time steps (Theorem 3.2.1).

If ω is not absolutely continuous with respect to Lebesgue measure, the optimal transport maps used in Definition 2.6.2 of $W_{2,\omega}$ may no longer exist. However, inspired by [1, Equation (7.3.2)], we are able to define a pseudo-metric and use it in the same manner to prove the intermediate results that lead to the exponential formula.

Definition and convexity properties of pseudo-metrics $W_{2,\boldsymbol{\omega}}$

DEFINITION A.2.1 (pseudo-metric $W_{2,\boldsymbol{\omega}}$). Fix $\omega \in \mathcal{P}(\mathbb{R}^d)$. Given $\mu, \nu \in \mathcal{P}_{2,\omega}(\mathbb{R}^d)$, choose $\boldsymbol{\omega} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ satisfying

$$\pi^{1,2}\#\boldsymbol{\omega} \in \Gamma_0(\omega, \mu) \text{ and } \pi^{1,3}\#\boldsymbol{\omega} \in \Gamma_0(\omega, \nu) \ .$$

(Recall from Definition 2.3.2 that we call $\boldsymbol{\omega}$ a *plan that induces a generalized geodesic from μ to ν with base ω*.) For this plan, we define the pseudo-metric $W_{2,\boldsymbol{\omega}}$ to be

$$W_{2,\boldsymbol{\omega}}(\mu, \nu) := \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x_2 - x_3|^2 d\boldsymbol{\omega} \right)^{1/2}$$

REMARK A.2.2. If $\mu = \omega$ or $\nu = \omega$, this reduces to the Wasserstein metric. In general, $W_{2,\boldsymbol{\omega}}(\mu, \nu) \geq W_2(\mu, \nu)$. If ω is absolutely continuous with respect to Lebesgue measure, $\boldsymbol{\omega}$ is unique and can be written as $\boldsymbol{\omega} = (\mathbf{id}, \mathbf{t}_\omega^\mu, \mathbf{t}_\omega^\nu)\#\omega$. Consequently, Definition A.2.1 extends Definition 2.6.2.

REMARK A.2.3. By the triangle inequality for the $L^2(\boldsymbol{\omega})$ norm,

$$W_{2,\boldsymbol{\omega}}(\mu,\nu) = ||x_2 - x_3||_{L^2(\boldsymbol{\omega})} \leq ||x_2 - \mathbf{id}||_{L^2(\boldsymbol{\omega})} + ||\mathbf{id} - x_3||_{L^2(\boldsymbol{\omega})} = W_2(\mu,\omega) + W_2(\omega,\nu) ,$$

so the analogue of the inequality (A.5) holds.

REMARK A.2.4. The pseudo-metric $W_{2,\boldsymbol{\omega}}$ occurs naturally in Definition 2.4.1 of convexity along generalized geodesics. 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 in case

$$E(\mu_\alpha) \leq (1-\alpha)E(\mu_0) + \alpha E(\mu_1) - \alpha(1-\alpha)\frac{\lambda}{2}W_{2,\boldsymbol{\omega}}(\mu_0,\mu_1) , \tag{A.6}$$

where $\boldsymbol{\omega}$ is the plan that induces the generalized geodesic μ_α.

We recall the following property of the pseudo-metric from [1, Lemma 9.2.1].

PROPOSITION A.2.5 (convexity of the pseudo-metric $W_{2,\boldsymbol{\omega}}$). *If μ_α is a generalized geodesic from μ to ν with base ω, induced by the plan $\boldsymbol{\omega}$, then*

$$W_{2,\boldsymbol{\omega}}(\omega,\mu_\alpha) = (1-\alpha)W_2(\omega,\mu) + \alpha W_2(\omega,\nu) - \alpha(1-\alpha)W_{2,\boldsymbol{\omega}}(\mu,\nu) .$$

Finally, we recall notation from [1, Section 7.2]:

$$\pi_\alpha^{i \to j} := (1-\alpha)\pi^i + \alpha\pi^j$$

$$\pi_\alpha^{k,i \to j} := (1-\alpha)\pi^{k,i} + \alpha\pi^{k,j}$$

Euler-Lagrange equation

In the following theorem, we generalize the result of Theorem 2.9.1 by removing the requirement that E satisfy domain assumption 2.1.1.

Our proof uses the following consequence of [1, Lemmas 5.3.2]: given $\mu^1, \mu^2, \mu^3, \mu^4 \in \mathcal{P}_{2,\omega_0}(\mathbb{R}^d)$ and $\gamma^{1,2} \in \Gamma(\mu^1,\mu^2)$, $\gamma^{2,3} \in \Gamma(\mu^2,\mu^3)$, and $\gamma^{3,4} \in \Gamma(\mu^3,\mu^4)$, there exists $\boldsymbol{\mu} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ such that

$$\pi^{1,2}\#\boldsymbol{\mu} = \gamma^{1,2}, \pi^{2,3}\#\boldsymbol{\mu} = \gamma^{2,3}, \text{ and } \pi^{3,4}\#\boldsymbol{\mu} = \gamma^{3,4} .$$

We will often permute the roles of the indices $1, 2, 3,$ and 4.

THEOREM A.2.6 (Euler-Lagrange equation). *Suppose E satisfies convexity assumption 2.1.2. 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 $\gamma \in \Gamma_0(\mu, \omega)$, if we define $\gamma_\tau := \rho_\tau \# \gamma$, $\rho_\tau(x_1, x_2) := (x_1, (x_2 - x_1)/\tau)$, then the following holds for all $\nu \in D(E), \boldsymbol{\gamma} \in \Gamma(\gamma_\tau, \nu)$.*

$$E(\nu) - E(\mu) \geq \int \langle x_2, x_3 - x_1 \rangle d\boldsymbol{\gamma} + o\left(||x_1 - x_3||_{L^2(\boldsymbol{\gamma})}\right) . \tag{A.7}$$

Proof. [1, Lemma 10.3.4] shows that if μ is the unique minimizer of $\Phi(\tau, \omega; \cdot)$, then for all $\gamma \in \Gamma_0(\mu, \omega)$, (A.7) holds.

We now prove the converse. Suppose that for all $\gamma \in \Gamma_0(\mu, \omega)$, (A.7) holds. 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 $\pi^{1,3} \# \boldsymbol{\omega} \in \Gamma_0(\mu, \omega)$ and $\pi^{2,3} \# \boldsymbol{\omega} \in \Gamma_0(\nu, \omega)$, so $\mu_\alpha = \pi_\alpha^{1 \to 2} \# \boldsymbol{\omega}$.

Applying (A.7) with $\nu = \mu_\alpha$ shows

$$E(\mu_\alpha) - E(\mu) \geq \int \langle x_2, x_3 - x_1 \rangle d\boldsymbol{\gamma}_\alpha + o\left(||x_1 - x_3||_{L^2(\boldsymbol{\gamma}_\alpha)}\right) \quad \forall \boldsymbol{\gamma}_\alpha \in \Gamma(\gamma_\tau, \mu_\alpha) . \tag{A.8}$$

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

$$
\begin{aligned}
& E(\mu_\alpha) - E(\mu) \\
& \geq \int \left\langle \frac{x_3 - x_1}{\tau}, ((1 - \alpha)x_1 + \alpha x_2) - x_1 \right\rangle d\boldsymbol{\omega} + o\left(||x_1 - (1 - \alpha)x_1 - \alpha x_2||_{L^2(\boldsymbol{\omega})}\right) \\
& = \alpha \int \left\langle \frac{x_3 - x_1}{\tau}, x_2 - x_1 \right\rangle d\boldsymbol{\omega} + o\left(\alpha ||x_1 - x_2||_{L^2(\boldsymbol{\omega})}\right) \\
& = \alpha \int \left\langle \frac{x_3 - x_1}{\tau}, x_2 - x_1 \right\rangle d\boldsymbol{\omega} + o(\alpha)
\end{aligned}
\tag{A.9}
$$

By definition of convexity along μ_α, $E(\nu) - E(\mu) \geq \frac{1}{\alpha} \left[E(\mu_\alpha) - E(\mu)\right] + (1 - \alpha)\frac{\lambda}{2}||x_1 - x_2||_{L^2(\boldsymbol{\omega})}$. Using (A.9), we may bound this from below:

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

Sending $\alpha \to 0$,

$$E(\nu) - E(\mu) \geq \int \left\langle \frac{x_3 - x_1}{\tau}, x_2 - x_1 \right\rangle d\boldsymbol{\omega} + \frac{\lambda}{2}||x_1 - x_2||_{L^2(\boldsymbol{\omega})} . \tag{A.10}$$

A similar inequality holds for $W_2^2(\cdot, \omega)$,

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

Combining with (A.10) and using that $\lambda + \frac{1}{\tau} > 0$,

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

Since $\nu \in D(E)$ was arbitrary, μ is the minimizer of $\Phi(\tau, \omega; \cdot)$.

$\square$

Relation between proximal maps with different time steps

We now provide the generalization of Theorem 3.2.1.

THEOREM A.2.7. *Suppose E satisfies convexity assumption 2.1.2. 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 $\omega := \mu_{\frac{h}{\tau}}^{\mu_\tau \to \mu}$. To prove the desired result, we must show $\mu_\tau = \omega_h$.

By [1, Lemma 7.2.1], there exists a unique plan $\gamma^{\mu_\tau \to \omega} \in \Gamma_0(\mu_\tau, \omega)$ and there exists $\gamma \in \Gamma_0(\mu_\tau, \mu)$ such that

$$
\gamma^{\mu_\tau \to \omega} = \pi_{\frac{h}{\tau}}^{1,1\to 2} \# \gamma . \tag{A.11}
$$

Since μ_τ is the unique minimizer of the quadratic perturbation $\Phi(\tau, \mu; \cdot)$, Theorem A.2.6 implies that for all $\gamma \in \Gamma_0(\mu_\tau, \mu)$, if we define $\gamma_\tau := \rho_\tau \# \gamma$, $\rho_\tau(x_1, x_2) := (x_1, (x_2 - x_1)/\tau)$, then the following holds for all $\nu \in D(E), \boldsymbol{\gamma} \in \Gamma(\gamma_\tau, \nu)$,

$$
E(\nu) - E(\mu_\tau) \geq \int \langle x_2, x_3 - x_1 \rangle d\boldsymbol{\gamma} + o\left(||x_1 - x_3||_{L^2(\boldsymbol{\gamma})} \right) . \tag{A.12}
$$

By a second application of Theorem A.2.6, to prove $\mu_\tau = \omega_h$, it's enough to show that for all $\tilde{\gamma} \in \Gamma_0(\mu_\tau, \omega)$, if we define $\tilde{\gamma}_h := \rho_h \# \gamma$, $\rho_h(x_1, x_2) := (x_1, (x_2 - x_1)/h)$, then the following holds for all $\tilde{\nu} \in D(E), \tilde{\boldsymbol{\gamma}} \in \Gamma(\tilde{\gamma}_h, \tilde{\nu})$,

$$E(\tilde{\nu}) - E(\mu_\tau) \geq \int \langle x_2, x_3 - x_1 \rangle d\tilde{\boldsymbol{\gamma}} + o\left(\|x_1 - x_3\|_{L^2(\tilde{\gamma})}\right) . \tag{A.13}$$

Since $\gamma^{\mu_\tau \to \omega}$ is the unique plan in $\Gamma_0(\mu_\tau, \omega)$, it is enough to show that (A.13) holds for $\tilde{\gamma} = \gamma^{\mu_\tau \to \omega}$.

Since (A.12) holds for all $\gamma \in \Gamma_0(\mu_\tau, \mu)$, in particular, it holds for the γ satisfying (A.11). A brief computation shows

$$\rho_h \circ \pi_{\frac{h}{\tau}}^{1,1 \to 2}(x_1, x_2) = \rho_h\left(x_1, \frac{\tau - h}{\tau} x_1 + \frac{h}{\tau} x_2\right) = \left(x_1, \frac{1}{h}\left[\frac{\tau - h}{\tau} x_1 + \frac{h}{\tau} x_2 - x_1\right]\right)$$
$$= \left(x_1, \frac{x_2 - x_1}{\tau}\right) = \rho_\tau(x_1, x_2) .$$

Consequently, for these choices of $\tilde{\gamma}$ and γ,

$$\tilde{\gamma}_h = \rho_h \# \gamma^{\mu_\tau \to \nu} = \rho_h \circ \pi_{\frac{h}{\tau}}^{1,1 \to 2} \# \gamma = \rho_\tau \# \gamma = \gamma_\tau .$$

Therefore, the fact that (A.12) holds for γ implies that (A.13) holds for $\tilde{\gamma}$, which proves the result. $\square$

References

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[2] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[3] Haïm Brézis. Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations. In *Contributions to nonlinear functional analysis (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1971)*, pages 101–156. Academic Press, New York, 1971.

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

[5] Haim Brezis. *Functional analysis, Sobolev spaces and partial differential equations.* Universitext. Springer, New York, 2011.

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

[7] 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(2):229–271, 2011.

[8] J. A. Carrillo, M. Di Francesco, A. Figalli, T. Laurent, and D. Slepčev. Confinement in nonlocal interaction equations. *Nonlinear Anal.*, 75(2):550–558, 2012.

[9] José Antonio Carrillo, Young-Pil Choi, and Maxime Hauray. The derivation of swarming models: Mean-field limit and Wasserstein distances. *preprint*.

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

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

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

[13] Ugo Gianazza, Giuseppe Savaré, and Giuseppe Toscani. The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation. *Arch. Ration. Mech. Anal.*, 194(1):133–220, 2009.

[14] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[15] Jürgen Jost. Convex functionals and generalized harmonic maps into spaces of nonpositive curvature. *Comment. Math. Helv.*, 70(4):659–673, 1995.

[16] Yukio Kōmura. Nonlinear semi-groups in Hilbert space. *J. Math. Soc. Japan*, 19:493–507, 1967.

[17] Uwe F. Mayer. Gradient flows on nonpositively curved metric spaces and harmonic maps. *Comm. Anal. Geom.*, 6(2):199–253, 1998.

[18] Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80(2):309–323, 1995.

[19] Robert J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.

[20] Felix Otto. Doubly degenerate diffusion equations as steepest descent, manuscript. 1996.

[21] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[22] A. Pazy. Semi-groups of nonlinear contractions in Hilbert space. In *Problems in nonlinear analysis*. CIME Varenna, Cremonese, 1971.

[23] Soren Rasmussen. Non-linear semi-groups, evolution equations and product integral representations. *Aarhus Universitet,*, (2), 1971.

[24] Jim Rulla. Error analysis for implicit approximations to solutions to Cauchy problems. *SIAM J. Numer. Anal.*, 33(1):68–87, 1996.

[25] Giuseppe Savaré. Application of optimal transport to evolutionary PDEs. Carnegie Mellon Center for Nonlinear Analysis, June 7-12, 2010. Lecture Notes.

[26] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[27] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[28] Kōsaku Yosida. *Functional analysis*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the sixth (1980) edition.