OPTIMAL TRANSPORT AND CONVERGENCE OF THE JKO SCHEME

Simon Foldvik

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Master of Science in Mathematical Sciences
at the University of Oxford

Hilary 2025

Abstract

We review rudimentary aspects of optimal transport and Wasserstein spaces in the locally compact category with particular emphasis on topological conditions. We give a proof that measurability is preserved under pointwise limits into second countable regular spaces by embeddings into Tychonoff cubes. This class is large enough to encompass all Polish spaces and all separable Banach spaces. Although these embeddings are well-known, we are not aware of any proof of this result in the measure theory literature utilising this technique. We construct a counterexample to show that measurability is not preserved into general topological spaces. We conclude by proving convergence of the JKO scheme for the gradient flow of the Brinkman functional on the 2-Wasserstein space, following an outline in Ambrosio, Gigli, Savaré [1]. Finding some of their estimates unclear, we suggest clarifications and adaptations of their arguments based on similar estimates derived herein.

Acknowledgements

This dissertation was made possible by the generous support of the Aker Scholarship, which fully funded my academic year at Oxford, 2024–2025.

I wish to thank Jakub Skrzeczkowski, my supervisor at the Mathematical Institute, for insightful discussions and for inviting me to research seminars at Queen's College and within the OxPDE group.

Thank you to Balliol College for providing a hospitable and vibrant academic environment, and for offering memorable experiences of the Oxford tradition.

Contents

1 Introduction

Recent decades have seen a surge in successful interpretations of partial differential equations of evolution type as gradient flows on spaces of probability measures equipped with various geometric structures. Foundational contributions include the seminal work of Jordan, Kinderlehrer, Otto on the linear Fokker–Planck equation as the gradient flow of a free energy functional on the space of probabilities equipped with the Wasserstein distance [16], and Otto's new interpretation of the porous medium equation as a gradient flow, together with their development of a Riemannian calculus on the Wasserstein spaces [25]. These developments, along with the metric theory for gradient flows in non-smooth settings developed most notably by Ambrosio, Gigli, Savaré [1], have reestablished this classical topic as an active area of current research.

In [7], the authors seek to establish existence and uniqueness of solutions to equations of the form

$$\begin{cases} \partial_t u_t - \operatorname{div}(u_t \nabla w_t) = 0 & \text{in} \quad (0,T) \times \mathbb{R}^n \\ -\Delta w_t + w_t = u_t & \text{in} \quad [0,T] \times \mathbb{R}^n \end{cases} \tag{1.0.1}$$

by the fixed-point method of Banach, but we find the arguments provided therein unclear and beyond our ability to repair. Here, $u_t \geq 0$ models a population density evolving in time and w_t is a velocity potential linked to the population density through the uniformly elliptic contraint known as Brinkman's law. We suggest instead to obtain well-posedness by recognising (1.0.1) as the gradient flow of the interaction energy

$$\mathscr{F}(\mu) := \frac{1}{2} \int_{\mathbb{R}} k_1 * \mu \, d\mu \tag{1.0.2}$$

for μ belonging to the Wasserstein space $\mathscr{P}_2(\mathbb{R})$ of probability measures on $\mathbb{R}$ with finite second moments. In (1.0.2), k_1 is a fundamental solution to the operator $-\Delta + I$ on $\mathbb{R}^n$ with $n = 1$, stemming from the elliptic constraint imposed by Brinkman. In one dimension, it takes the form $k_1(x) = (1/2)e^{-|x|}$, and consequently the *Brinkman energy* (1.0.2) is

$$\mathscr{F}(\mu) = \frac{1}{4} \iint e^{-|x-y|} \, d\mu(x) \, d\mu(y) \quad (\mu \in \mathscr{P}_2(\mathbb{R})). \tag{1.0.3}$$

In one dimension, we prove that (1.0.2) is displacement convex on $\mathscr{P}_2(\mathbb{R})$ in the sense of McCann [21], hence allowing us to build unique solutions to the corresponding Wasserstein gradient flow

$$\partial_t \mu_t = -\operatorname{grad}_{W_2} \mathscr{F}(\mu_t) \tag{1.0.4}$$

by the JKO scheme of Jordan, Kinderlehrer, Otto [16]. Originally applied to produce convergence of a discrete variational scheme to the unique smooth solution to the linear Fokker–Planck equation

$$\partial_t \rho_t = \operatorname{div}(\rho_t \nabla V) + \Delta \rho_t \tag{1.0.5}$$

by discretising in time the gradient flow for the free energy functional

$$\mathscr{E}(\rho) := \int_{\mathbb{R}^n} V\, d\rho + \int_{\mathbb{R}^n} \rho \log \rho\, dx \tag{1.0.6}$$

on $\mathscr{P}_2(\mathbb{R})$, the JKO scheme has since become a standard technical device in the subject to produce solutions to equations with a gradient flow structure in the more general metric setting [1, 30, 36].

The above considerations break down in higher dimensions and are similar in spirit to the approach taken in Carrillo et al. [3], where the authors obtain well-posedness of measure-valued solutions and long-term asymptotics for the continuity equation

$$\partial_t u_t + \partial_x (H(u_t) u_t) = 0 \quad \text{in} \quad (0, \infty) \times \mathbb{R} \tag{1.0.7}$$

driven by the Hilbert transform (see [33] for its basic theory)

$$Hu(x) := \frac{1}{\pi} \operatorname{PV} \int_{\mathbb{R}} \frac{u(z)}{x - z}\, dz, \tag{1.0.8}$$

starting at some initial Borel probability $u_0 \geq 0$. This equation arises formally from a functional of the type (1.0.2) with a singular logarithmic interaction potential and has been proposed as a simplified one-dimensional model in fluid mechanics [3, § 1.1].

That equations (1.0.1) and (1.0.7) may be treated by similar methods stem from the fact that gradient flows (1.0.4) on the Wasserstein spaces $\mathscr{P}_2(\mathbb{R}^n)$ all take the form of continuity equations

$$\partial_t \mu_t - \operatorname{div}\left(\mu_t \nabla \left(\frac{\delta \mathscr{F}}{\delta \mu} \right) \right) = 0, \tag{1.0.9}$$

where $\delta \mathscr{F} / \delta \mu$ denotes the *first variation* of the energy functional $\mathscr{F}$ on $\mathscr{P}_2(\mathbb{R}^n)$. As soon as certain convexity and continuity properties of the energy are observed, the theory of [1] applies, and this is the approach taken in Chapter 3.

Underlying these advances are tools from the by now mature field of optimal transport. Initiated by Monge's considerations of optimal mass transportation in the 18th century [22], the subject saw a resurgence with Kantorovich's relaxation of the problem in the 1940s [17] and had by the close of the 1980s gained new impetus by the works of Brenier, Cullen, Mather, and others (see Chapter 3 and the chapter-ending remarks in Villani [35] for a historical overview). The authoritative reference on the subject today is Villani [35], although many shorter introductory accounts exist. We mention in this regard Villani [36], Santambrogio [31], Ambrosio & Gigli [2], Figalli [11], and Garling [14]. The latter has the unusual virtue of highlighting the role of the axiom of dependent choice. Readers interested in computational aspects may consult the book [27] by Peyré and Cuturi and the references therein. The recent book by Düll et al. [9] provides a useful reference on Radon measures with a view towards structured population modelling.

The aforementioned standard reference [1] of Ambrosio, Gigli, Savaré on the metric and Wasserstein theories of gradient flows is technical and comprehensive, but the survey article [30] of Santambrogio may be considered an expanded synopsis of this work with the stated aim of 'giving an overview of the whole theory'. As such it serves as a useful starting point for newcomers; see also Chapters 23–25 of Villani [35] for a general reference.

Central to the success of the analysis of long-term asymptotics of the gradient flow dynamics is a supply of intricate functional inequalities and convexity estimates. Recent studies have generalised the success observed in Wasserstein spaces to other geometries. See for instance [4] for a comprehensive study of such inequalities and convexity criteria for f-divergences on spaces of probability measures in the Fisher–Rao geometry, which is an active area of ongoing research.

1.1 Outline of Chapters

In § 2.1 we review rudimentary aspects of probability theory in a locally compact setting, emphasising topological rather than metric arguments. We prove in § 2.1.1 that measurability is preserved under pointwise limits into spaces second countable regular by embeddings into Tychonoff cubes. To the best of our knowledge, a proof of this result by means of such embeddings does not occur in the literature and is particularly natural in this context since it is purely topological and does away with any explicit reference to a choice of metric. We also construct a corresponding counterexample showcasing that some condition is required in order for measurability to be preserved in the limit. In § 2.2, we recall the Monge–Kantorovich problems of optimal transport and develop enough to provide a short reference on Wasserstein spaces in § 2.3. We conclude the chapter by computing a series of examples in § 2.4.

Chapter 3 is devoted to proving convergence of the JKO scheme for the Brinkman functional (1.0.2) on the Wasserstein space $\mathscr{P}_2(\mathbb{R})$. We follow an outline given in [1, § 11.1.3] and suggest adjustments to estimates we were unable to reproduce and adaptations to the arguments based thereon.

2 Preliminaries on Optimal Transport

2.1 Rudiments of Probability Theory

The main tools of this thesis come from the theory of optimal transport, and although much can be done over Polish spaces [14, 35] (separable and completely metrisable topological spaces), we are primarily concerned with the one-dimensional situation on the real line and restrict our attention to the probabilistic framework on locally compact Hausdorff spaces (henceforth LCH) which are second countable. These assumptions are broad enough to encompass dynamics of probability measures on Euclidean spaces $\mathbb{R}^n$ and all submanifolds thereof.

This approach is also observed in Figalli [11], under the additional assumption of metrisability and camouflaged as separable, locally compact metric spaces. That these categories are the same rests on two facts. The first is that metrisable spaces are separable iff they are second countable, although the latter property is stronger and often more appropriate for general topological spaces. The second is that LCH spaces X embed openly into their Alexandroff extensions (one-point compactifications) $\widetilde{X}$. These are compact Hausdorff, hence normal, and so Urysohn's lemma applies to $\widetilde{X}$. This shows X must be Tychonoff (completely regular). An LCH space is therefore regular (but not necessarily normal [32, p. 190]). Adding the hypothesis of second countability now affords us normality [23, Theorem 32.1] and thus a topological embedding

$$X \hookrightarrow [0,1]^{\mathbb{N}} \tag{2.1.1}$$

into the Tychonoff cube [26, Proposition 1.6.14], a space which is compact and metrisable in the product topology. This yields metrisability of second countable LCH spaces and is one approach to the Urysohn metrisation theorem. We push this technique further in § 2.1.1.

Completeness is always useful, but when X is second countable LCH, the Banach algebra $C_0(X)$ of continuous real functions $X \to \mathbb{R}$ vanishing at infinity is itself separable [14, Proposition 3.4.2]. From this follows the metrisability of the closed unit ball in the continuous dual space

$$M(X) \simeq C_0(X)^* \tag{2.1.2}$$

of real regular Borel measures in the relative weak*-topology [6, Theorem V.5.1]. Combining this fact with the theorem of Banach–Alaoglu yields *sequential* precompactness of bounded sequences in $M(X)$ in the weak*-topology, a source of great joy for analysts. This is subtle, since the weak and weak*-topologies on infinite-dimensional Banach spaces are themselves never metrisable [6, § V.5]. The best one can do in general with a bounded sequence in the weak*-topology is the extraction of a convergent *subnet*, and easy examples show that a subnet need not be a subsequence. Nothing to this effect is discussed in Figalli [11].

There is no universal agreement in the literature as to exactly what constitutes a Radon measure. We adopt terminology consistent with Folland [13, p. 212] and say that a Borel measure μ on a locally compact Hausdorff space X is *Radon* iff it is finite on compacts, inner regular (or tight) on *open sets*, and outer regular on Borel sets. We say it is *regular* iff it is Radon and also inner regular on all Borel sets. A real (signed) Borel measure μ on X is regular iff its total variation measure $|\mu|$ is regular, the Banach space of all such under the total variation norm $\|\mu\|_{\mathrm{TV}}$ we denote by $M(X)$. An example of a Radon measure which is not regular is given in [13, Exercise 7.2.12].

The theory of measure and integration on spaces LCH is well-developed, and it is useful to know that for certain spaces the regularity of a Borel measure is automatic for reasons pertaining to the underlying topology. Among these, we highlight the following implications for positive Borel measures on spaces LCH, all of which are either proved or may easily be deduced from statements in Folland [13, Chapter 7]:

$$\text{Radon} \not\Longrightarrow \text{Regular}, \tag{2.1.3}$$

$$\begin{aligned} \text{Radon} + \sigma\text{-finite} &\Longrightarrow \text{Regular}, &(2.1.4)\\ \text{Radon} + \text{Finite} &\Longrightarrow \text{Regular}, &(2.1.5)\\ \text{Radon} + X\ \sigma\text{-compact} &\Longrightarrow \text{Regular}, &(2.1.6)\end{aligned}$$

$$\begin{aligned} \text{Finite on compacts} + \text{Every open set } \sigma\text{-compact} &\Longrightarrow \text{Regular}, &(2.1.7)\\ \text{Finite on compacts} + X \text{ second countable} &\Longrightarrow \text{Regular}, &(2.1.8)\\ \text{Measure finite} + X \text{ second countable} &\Longrightarrow \text{Regular}. &(2.1.9)\end{aligned}$$

Corollary 2.1.1. *On $\mathbb{R}^n$, or any topological submanifold thereof, every Borel probability measure is regular (in particular, tight).*

More generally, a collection $\mathscr{E}$ of Borel probability measures on an LCH space X is *tight* iff for every $\epsilon > 0$ there is a compact $K_\epsilon \subset\subset X$ such that

$$\mu(X \setminus K_\epsilon) \leq \epsilon \quad \text{for all} \quad \mu \in \mathscr{E}, \tag{2.1.10}$$

the idea being that all measures $\mu \in \mathscr{E}$ are up to small errors uniformly concentrated on common compact sets. Prokhorov's theorem (Theorem 2.1.10) states that tightness is both a necessary and sufficient condition for *sequential* precompactness of $\mathscr{E}$ in the narrow topology when X is Polish.

A measure $\mu \in M(X)$ for X LCH integrates every bounded Borel function $f\colon X \to \mathbb{R}$, the collection of which we denote by $B_\infty(X)$ as in Murphy [24], and there is the standard estimate

$$\left|\int f\,d\mu\right| \leq \int |f|\,d|\mu| \leq \|f\|_\infty \|\mu\|_{\mathrm{TV}}. \tag{2.1.11}$$

A linear subspace $\mathscr{F} \subseteq B_\infty(X)$ thus induces a subspace $\widehat{\mathscr{F}} := \{\widehat{f} : f \in \mathscr{F}\}$ of norm-continuous functionals in $M(X)^*$ by $\widehat{f}(\mu) := \int f\,d\mu$ for $f \in \mathscr{F}$ and

$\mu \in M(X)$. We consequently say that a net (μ_α) in $M(X)$ is $\mathscr{F}$-convergent to $\mu \in M(X)$ iff $\mu_\alpha \to \mu$ in the $\widehat{\mathscr{F}}$-weak topology induced by $\widehat{\mathscr{F}}$ on $M(X)$. Since

$$(M(X), \widehat{\mathscr{F}}\text{-weak})^* = \widehat{\mathscr{F}} \tag{2.1.12}$$

(see [26, Proposition 2.4.4]), this occurs iff

$$\int f\, d\mu_\alpha \to \int f\, d\mu \quad \text{for all} \quad f \in \mathscr{F}, \tag{2.1.13}$$

a situation we denote by $\mu_\alpha \xrightarrow{\mathscr{F}} \mu$. It is a standard result on weak topologies that the $\widehat{\mathscr{F}}$-weak topology is Hausdorff iff $\widehat{\mathscr{F}}$ is separating, meaning iff $\int f\, d\mu = 0$ for all $f \in \mathscr{F}$ implies $\mu = 0$, and this occurs iff all $\mathscr{F}$-convergent nets have unique limits.

There are at least three relevant modes of weak convergence for measures in $M(X)$, and for a net (μ_α) and a measure μ in $M(X)$ we shall say that:

- $\mu_\alpha \to \mu$ *vaguely* iff $\mu_\alpha \xrightarrow{C_c} \mu$,
- $\mu_\alpha \to \mu$ weakly* iff $\mu_\alpha \xrightarrow{C_0} \mu$,
- $\mu_\alpha \to \mu$ *narrowly* iff $\mu_\alpha \xrightarrow{C_b} \mu$.

Lemma 2.1.2. *Since $\int f\, d\mu = 0$ for all $f \in C_c(X)$ implies $\mu = 0$ for $\mu \in M(X)$ with X LCH, each of the above three topologies are Hausdorff.*

For X LCH, there are inclusions

$$C_c(X) \hookrightarrow C_0(X) \hookrightarrow C_b(X), \tag{2.1.14}$$

the first of which is dense and the second proper (in general). There is an induced adjoint sequence

$$C_c(X)^* \xleftarrow{\sim} C_0(X)^* \leftarrow C_b(X)^*, \tag{2.1.15}$$

where the first arrow turns out to be a Banach space isomorphism. Care must nevertheless be exercised to observe that weak*-convergence in $C_c(X)^*$ and $C_0(X)^*$ might differ (Example 2.1.3), since the adjoint map $C_c(X)^* \xleftarrow{\sim} C_0(X)^*$ may fail to be an isomorphism in the weak*-topologies (it is bijective and weak*-continuous, but its inverse may fail to be weak*-continuous even though it is norm-continuous). Recalling the Riesz isomorphism

$$M(X) \simeq C_0(X)^* \tag{2.1.16}$$

(see Appendix C in Conway [6] or Chapter 6 in Rudin [29] for efficient treatments), it follows that similar precautions must be taken for the weak convergence of measures.

Example 2.1.3 ($C_c \not\Rightarrow C_0$)**.** Consider the sequence $\mu_k := k\delta_k$ of scaled Diracs on $\mathbb{R}$. Then

$$\int f\, d\mu_k = kf(k) \to 0 \quad \text{for all} \quad f \in C_c(\mathbb{R}), \tag{2.1.17}$$

but $\int \operatorname{sinc}\, d\mu_k = \sin k$ (recall $\operatorname{sinc}(x) := \sin(x)/x$) has no limit as $k \to \infty$. So $(\int f\, d\mu_k)_{k\in\mathbb{N}}$ need not even converge for all $f \in C_0(\mathbb{R})$.

Example 2.1.4 ($C_0 \not\Rightarrow C_b$)**.** For the Dirac measures $(\delta_k)_{k\in\mathbb{N}}$ on $\mathbb{R}$,

$$\int f\, d\delta_k = f(k) \to 0 \quad \text{for all} \quad f \in C_0(\mathbb{R}), \tag{2.1.18}$$

but $1 \in C_b(\mathbb{R})$ and $\int 1\, d\delta_k = 1 \not\to 0$.

Not only does the last example show that C_0-convergence may fail to imply C_b-convergence, but also that the C_0-limit of probability measures may fail to be a probability. The interpretation is that C_0-convergence allows mass to escape to infinity, a defect which is not observed with C_b-convergence:

$$\mu(X) = \int 1\, d\mu = \lim \int 1\, d\mu_\alpha = 1 \tag{2.1.19}$$

if $\mu_\alpha \to \mu$ narrowly and all $\mu_\alpha(X) = 1$, and the family $\{\mu_\alpha\}$ is tight if the convergence is sequential. That μ must be a probability is also pointed out in Figalli [11, Remark 2.1.6], but they offer no explanation as to why μ must be positive. We prove this.

Proof that μ is positive. The convergence $\int f\, d\mu = \lim \int f\, d\mu_k$ holds for all $f \in C_b(X)$ and μ therefore defines a positive functional on $C_c(X)$. By the Riesz theorem, there is a unique positive Radon measure ν on $\mathscr{B}(X)$ such that $\mu(f) = \nu(f)$ for all $f \in C_c(X)$. For $0 \le f \le 1$ in $C_c(X)$, one estimates $\int f\, d\nu = \int f\, d\mu \le |\mu|(X)$, and therefore $\nu(X) = \sup \int f\, d\nu \le |\mu|(X)$ over such f. The measure ν is therefore finite, hence regular. It is now clear that $\mu = \nu$ since they agree as functionals on $C_0(X)$. So μ must be positive. ■

Lemma 2.1.5. *If a sequence $(\mu_k)_{k\in\mathbb{N}}$ of probability measures converges narrowly to a measure $\mu \in M(X)$, then μ is a probability and $\{\mu_k\}$ is tight.*

Remark 2.1.6. Tightness would follow from Prokhorov's theorem (below), but we prefer to construct a direct proof.

Proof. It remains to prove the tightness assertion. The target measure μ is tight by regularity, so given $\epsilon > 0$ we pick $K_\epsilon \subset\subset X$ with $\mu(X \setminus K_\epsilon) \le \epsilon$. Since X is LCH, we find a precompact open set $U \supseteq K$ (so $\overline{U}$ is compact) and an

$f \in C_c(X, [0,1])$ with $f = 1$ on K_ϵ and $\operatorname{supp} f \subset\subset U$. Then $1 - f \geq 0$ belongs to $C_b(X)$, vanishes on K_ϵ, takes the constant value 1 on $X \setminus U$, and

$$\int (1-f)\, d\mu_k \to \int (1-f)\, d\mu. \tag{2.1.20}$$

Estimating $0 \leq \int(1-f)\, d\mu \leq \mu(X \setminus K_\epsilon) \leq \epsilon$ and choosing N so that $|\int(1-f)\, d\mu_k - \int(1-f)\, d\mu| \leq \epsilon$ for $k \geq N$, we find furthermore

$$0 \leq \int (1-f)\, d\mu_k \leq 2\epsilon \quad (k \geq N). \tag{2.1.21}$$

Recalling the anatomy of $1 - f$, we write

$$\begin{aligned} \int (1-f)\, d\mu_k &= \int_{X\setminus U} (1-f)\, d\mu_k + \int_{U \setminus K_\epsilon} (1-f)\, d\mu_k & (2.1.22) \\ &= \mu_k(X \setminus U) + \int_{U\setminus K_\epsilon} (1-f)\, d\mu_k, & (2.1.23) \end{aligned}$$

from which

$$\mu_k(X \setminus U) \leq 2\epsilon - \int_{U\setminus K_\epsilon} (1-f)\, d\mu_k \leq 2\epsilon \quad (k \geq N). \tag{2.1.24}$$

Since $X \setminus \overline{U} \subseteq X \setminus U$, the initial segment $\mu_0, \ldots, \mu_N$ is tight one by one, a finite union of compact sets is compact, and estimate (2.1.24) is stable upon enlarging U, we find upon so doing the estimate

$$\mu_k(X \setminus \overline{U}) \leq 2\epsilon \quad (k \in \mathbb{N}). \tag{2.1.25}$$

■

What goes wrong in Example 2.1.3 is the lack of boundedness, a technicality which is suppressed when focus is restricted to the set of probability measures. We identify and prove:

Lemma 2.1.7 (Uniformly bounded vague implies weak*)**.** *If* (μ_α) *is a* bounded *net in* $M(X)$ *for* X *LCH and* $\mu \in M(X)$ *is such that*

$$\int f\, d\mu_\alpha \to \int f\, d\mu \tag{2.1.26}$$

holds for all $f \in C_c(X)$*, then it holds for all* $f \in C_0(X)$*.*

Proof. Suppose $\mu_\alpha \xrightarrow{C_c} \mu$ and consider $f \in C_0(X)$. For $g \in C_c(X)$, an $\epsilon/3$-argument using (2.1.11) gives

$$\begin{aligned} \left| \int f\, d\mu_\alpha - \int f\, d\mu \right| \leq \|f - g\|_\infty \|\mu_\alpha\|_{\mathrm{TV}} & \quad (2.1.27) \\ + |\mu_\alpha(g) - \mu(g)| & \quad (2.1.28) \\ + \|g - f\|_\infty \|\mu\|_{\mathrm{TV}}. & \quad (2.1.29) \end{aligned}$$

Letting $\epsilon > 0$ and taking g such that $\|g - f\|_\infty \le \epsilon$, we find α_0 (depending on g, f) with $|\mu_\alpha(g) - \mu(g)| \le \epsilon$ for $\alpha \ge \alpha_0$, and then

$$\left| \int f \, d\mu_\alpha - \int f \, d\mu \right| \le \epsilon(1 + \|\mu\|_{\mathrm{TV}} + \|\mu_\alpha\|_{\mathrm{TV}}) \quad (\alpha \ge \alpha_0). \tag{2.1.30}$$

The claim follows since (μ_α) is assumed bounded. ■

Existence of minimisers for the Kantorovich problem (§ 2.2) follows from standard continuity and compactness arguments. We recall in this regard that a mapping f of a topological space X into the extended real line $\overline{\mathbb{R}} := [-\infty, +\infty]$ is *lower semicontinuous* (henceforth LSC) iff

$$f(x) \le \liminf f(x_\alpha) \tag{2.1.31}$$

for every net (x_α) convergent to x in X. It is *sequentially LSC* iff (2.1.31) is satisfied whenever the net $(x_\alpha)_{\alpha\in\mathbb{N}}$ is a sequence. For metrisable topological spaces, these notions coincide. Equivalent characterisations and structural properties of the space of functions LSC may be found in [26, § 1.5], [13, § 7.2].

Lemma 2.1.8 (Measure evaluation is LSC)**.** *If $f\colon X \to \overline{\mathbb{R}}_+$ is LSC on the LCH space X, then the evaluation $\mu \mapsto \int f \, d\mu$ is LSC with respect to vague convergence on $\mathscr{P}(X)$: If $\mu_\alpha \to \mu$ in $\mathscr{P}(X)$ vaguely for some net (μ_α), then*

$$\int f \, d\mu \le \liminf \int f \, d\mu_\alpha. \tag{2.1.32}$$

Remark 2.1.9. Figalli [11, pp. 22–23] offers a proof that $\mu \mapsto \int f \, d\mu$ is *sequentially* LSC under the additional assumption that X be metrisable. Adapting this argument, we give a simpler proof based on the theory of Radon integrals, allowing us to lift the metrisability requirement on X and also to prove the statement for general nets.

Proof. By [13, Corollary 7.13], we have

$$\int f \, d\mu = \sup \int g \, d\mu \tag{2.1.33}$$

over all $g \in C_c(X)$ with $0 \le g \le f$. This gives a sequence (g_k) in $C_c(X)$ with $0 \le g_k \le f$ and $\int g_k \, d\mu \to \int f \, d\mu$. Using the vague convergence $\mu_\alpha \to \mu$, we estimate

$$\int g_k \, d\mu = \lim_\alpha \int g_k \, d\mu_\alpha \le \liminf_\alpha \int f \, d\mu_\alpha, \tag{2.1.34}$$

the last inequality owing to $g_k \le f$. Sending $k \to \infty$, we find

$$\int f \, d\mu \le \liminf \int f \, d\mu_\alpha. \tag{2.1.35}$$

■

We now recall the compactness result of Prokhorov. The reader may look up a proof in any standard reference, for instance [14, 35, 11]. Further equivalent statements are given in [14, Theorem 18.6.2].

Theorem 2.1.10 (Prokhorov)**.** *A collection $\mathscr{E} \subseteq \mathscr{P}(X)$ of probability measures on a Polish space X is sequentially precompact in $\mathscr{P}(X)$ for the induced C_b-topology iff it is tight.*

Remark 2.1.11. The result applies when X is second countable LCH. Indeed, the Alexandroff extension $\widetilde{X} := X \cup \{\infty\}$ is compact Hausdorff and countable at infinity. To see this, let $(K_m)_{m\in\mathbb{N}}$ be a compact sequential exhaustion of X [20, Proposition 4.76]: A sequence of compact sets $K_m \subset \operatorname{int} K_{m+1}$ with $X = \bigcup K_m$. Then the $\widetilde{X} \setminus K_m$ by construction form a countable neighbourhood basis at ∞. The space $\widetilde{X}$ is thus second countable compact Hausdorff, hence metrisable by the discussion at the beginning of § 2.1. Any metric on $\widetilde{X}$ is automatically complete by compactness, hence X is completely metrisable (not by the pullback metric) since it is a G_δ (in fact, open) subset of $\widetilde{X}$ [14, Theorem 2.6.1]. Further characterisations of which locally compact metrisable spaces are Polish are given in Garling [14, Theorem 3.4.1].

Corollary 2.1.12. *Spaces second countable LCH are Polish.*

2.1.1 Measurability Through the Tychonoff Cube

It is natural in the applications to consider integration of maps from time intervals $[0, T]$ into various spaces of functions. Such theories were initiated in the first half of the 20th century by Bochner, Pettis, and others; the classic reference [8] provides extensive historical comments on these developments in the style of Dunford & Schwartz and is current up to the end of the 1970s. A more modern and up to date treatment may be found in [15].

A standard result in real analysis says that measurability is preserved under pointwise limits for functions taking values in any of the spaces of classical measure theory. We give an example to the effect that this is no longer true for maps valued in general topological spaces.

Let $\mathbb{R}_f$ denote the real numbers in the finite complement topology. A set $W \subseteq \mathbb{R}_f$ is open iff $W = \mathbb{R} \setminus A$ for A finite or $A = \mathbb{R}$. In particular, every finite set is closed.

Consider an injective sequence (t_k) in $\mathbb{R}_f$ and *any* point $t \in \mathbb{R}$. Then $t_k \to t$ in $\mathbb{R}_f$. Hence an injective sequence in $\mathbb{R}_f$ converges to every point.

Proof. Take an open set $W = \mathbb{R} \setminus A$ containing t. Then $A \neq \mathbb{R}$, so A is finite. Pick N such that $t_k \notin A$ for $k \geq N$. This is possible since the sequence is injective. Hence $t_k \in W$ whenever $k \geq N$. ∎

Consider $\mathbb{R}$ with its standard topology. Define constant functions $u_k : \mathbb{R} \to \mathbb{R}_f$ in such a way that no constant value appears twice, say $u_k(x) := 1/k$ for

$x \in \mathbb{R}$ and $k \geq 1$. The u_k are measurable $\mathbb{R} \to \mathbb{R}_f$ with respect to the corresponding Borel σ-algebras. The insight is that (u_k) converges pointwise to *any* function $u\colon \mathbb{R} \to \mathbb{R}_f$, hence we are free to choose a pathological u which is not measurable.

Consider a non-Lebesgue measurable Vitali set $V \subset \mathbb{R}$ ([34, § 1.3] or [13, § 1.1]) and let $u := 1_V$. Then $u_k \to u$ pointwise by the above, and each u_k is measurable. But u is by construction not measurable $\mathbb{R} \to \mathbb{R}_f$, since this would require V be Lebesgue.

An example of non-measurability where the range space is Hausdorff may be found in Cohn [5, Exercise 8.1.14].

The foregoing examples show that measurability need not be preserved under pointwise limits into general topological spaces. We identify the second countable regular spaces as a class where this pathology is ruled out, which is broad enough to encompass all Polish spaces and all separable Banach spaces. As mentioned at the beginning of § 2.1, such spaces are metrisable, and it is well-known that metrisability is enough to preserve measurability in the limit ([19, p. 117] or [5, Proposition 8.1.10]). Although the metrisable spaces form a more general class than ours, we are not aware of any proofs by embeddings into the Tychonoff cube.

Lemma 2.1.13. *If $u_k \to u$ pointwise on a measurable space $(X, \mathscr{A})$ with values in a second countable regular topological space Y, then u is measurable.*

Proof. Let $T := \prod_{\mathbb{N}}[0,1]_n$ be the Tychonoff cube equipped with the product topology $\mathscr{T}_{\mathrm{prod}}$, and choose as at the beginning of § 2.1 a topological embedding

$$Y \overset{\Phi}{\hookrightarrow} T. \tag{2.1.36}$$

Let $\pi_n\colon T \to [0,1]_n$ be the canonical projections. These are continuous, and so each $\pi_n \circ \Phi \circ u_k$ is measurable $X \to [0,1]_n$. The convergence $u_k \to u$ is preserved under continuity, thus

$$\pi_n \circ \Phi \circ u_k \to \pi_n \circ \Phi \circ u \tag{2.1.37}$$

pointwise for all n. Each $\pi_n \circ \Phi \circ u$ is consequently measurable $X \to [0,1]_n$. Let $\mathscr{B}_n$ denote the Borel σ-algebra on $[0,1]_n$. The product σ-algebra $\bigotimes_n \mathscr{B}_n$ on $T = \prod_{\mathbb{N}}[0,1]_n$ is (by definition) the smallest σ-algebra with respect to which each π_n is measurable. It is basic that measurability of $\pi_n \circ \Phi \circ u$ for each n implies measurability into the product:

$$(X, \mathscr{A}) \xrightarrow{\Phi \circ u} (T, \otimes_n \mathscr{B}_n). \tag{2.1.38}$$

Let $\mathscr{B}(T)$ be the Borel σ-algebra of T in the product topology $\mathscr{T}_{\mathrm{prod}}$. The factor spaces $[0,1]_n$ are second countable, hence there is an inclusion

$$\mathscr{B}(T) \subseteq \bigotimes_{n \in \mathbb{N}} \mathscr{B}_n. \tag{2.1.39}$$

This shows $\Phi \circ u$ is measurable $(X, \mathscr{A}) \to (T, \mathscr{B}(T))$. Measuability of u follows since Φ is an embedding:

$$\begin{tikzcd} & (T, \mathscr{T}_{\mathrm{prod}}) \\ (X, \mathscr{A}) \arrow[ur, dashed, bend left] \arrow[r, "u"'] & (Y, \mathscr{T}(Y)) \arrow[u, hook, "\Phi"'] \end{tikzcd} . \tag{2.1.40}$$

■

2.2 The Monge and Kantorovich Problems

2.2.1 Historical Introduction

The *Monge problem* for measurable transformations $T\colon X \to Y$ and Borel measures μ, ν on X and Y, respectively, is to minimise the *Monge functional*

$$M(T) := \int_X c(x, T(x))\, d\mu(x) \tag{2.2.1}$$

over all such T transforming μ into ν.

Definition 2.2.1. A measurable $T\colon X \to Y$ transforming μ to ν in the sense that $T_\sharp \mu = \nu$, meaning $\nu(B) = \mu(T^{-1}(B))$ for all $B \in \mathscr{B}(Y)$, is known as a *transport map* from μ to ν.

The problem originates with Monge's 18th century considerations of optimal mass transportation [22]. One imagines (for instance in the case $X = Y = \mathbb{R}^n$) that some initial mass is distributed according to μ and needs to be rearranged into the configuration prescribed by ν. Associated with moving all mass from $x \in X$ to $T(x) \in Y$ is a cost $c(x, T(x))$, and the Monge functional (2.2.1) may be interpreted as the total cost endured in this process. Minimisers need not exist for the Monge functional for the simple reason that there need not exist any T transforming μ into ν (take μ Dirac and ν non-atomic), but when they do, they are referred to as *optimal transport maps* associated with the cost function $c\colon X \times Y \to \mathbb{R} \cup \{+\infty\}$.

In the 1940s, Kantorovich [17] considered a relaxation of Monge's problem by instead of moving *all* mass from x to $T(x)$ as required by a transport map, one instead construes the possibility of moving mass from a single source to several destinations. An appropriate way to encode this process is to construe probability measures $\pi \in \mathscr{P}(X \times Y)$ whose first and second marginals are μ and ν, respectively, such that the mass moved from region $A \in \mathscr{B}(X)$ to the region $B \in \mathscr{B}(Y)$ is given by $\pi(A \times B)$. The cost endured in this process is given by the *Kantorovich functional*

$$K(\pi) := \int_{X \times Y} c(x, y)\, d\pi(x, y), \tag{2.2.2}$$

and the *Kantorovich problem* is to minimise K over all such transport plans π.

Definition 2.2.2 (Couplings)**.** The set of all probability measures $\pi \in \mathscr{P}(X \times Y)$ with first marginal $\mu \in \mathscr{P}(X)$ and second marginal $\nu \in \mathscr{P}(Y)$, meaning

$$\pi^1_\sharp \pi = \mu \quad \text{and} \quad \pi^2_\sharp \pi = \nu, \tag{2.2.3}$$

is denoted $\Pi(\mu, \nu)$ and referred to as the set of *couplings* from μ to ν.

Remark 2.2.3*.* The relations (2.2.3) mean (by definition) that

$$\pi(A \times Y) = \mu(A) \quad \text{and} \quad \pi(X \times B) = \nu(B) \tag{2.2.4}$$

for all Borel sets $A \in \mathscr{B}(X)$ and $B \in \mathscr{B}(Y)$ for the canonical projections

$$\begin{tikzcd} & X \times Y \arrow[dl, bend right, "\pi^1"'] \arrow[dr, bend left, "\pi^2"] & \\ X & & Y \end{tikzcd} \quad \rightsquigarrow \quad \begin{tikzcd} & \mathscr{P}(X \times Y) \arrow[dl, bend right, "\pi^1_\sharp"'] \arrow[dr, bend left, "\pi^2_\sharp"] & \\ \mathscr{P}(X) & & \mathscr{P}(Y) \end{tikzcd} \quad . \tag{2.2.5}$$

For general topological spaces, this holds iff

$$\int_X f(x)\, d\mu(x) = \int_{X \times Y} f(x)\, d\pi(x, y) \tag{2.2.6}$$

for all $f \in B_\infty(X)$ bounded Borel, and similarly with ν. When X, Y are second countable LCH, it is by (2.1.9) enough to verify (2.2.6) for $f \in C_c(X)$.

Remark 2.2.4 (Monge couplings)*.* Every measurable transformation $T : X \to Y$ taking μ to ν induces a coupling $\pi_T \coloneqq (\mathrm{id} \times T)_\sharp \mu \in \Pi(\mu, \nu)$ when X, Y are second countable. The associated Kantorovich cost of a Monge coupling π_T is

$$K(\pi_T) = \int_X c(x, T(x))\, d\mu(x) = M(T), \tag{2.2.7}$$

as one sees upon applying the pushforward formula for measures. In particular,

$$\inf_\pi K(\pi) \leq \inf_{T_\sharp \mu = \nu} M(T). \tag{2.2.8}$$

Consequently, if a Kantorovich minimiser arises from a transport map, this also yields a solution to the Monge problem.

We note that $\mathscr{B}(X \times Y) = \mathscr{B}(X) \otimes \mathscr{B}(Y)$ when both spaces X, Y are second countable [13, Theorem 7.20].

2.2.2 Existence of Minimisers

The Kantorovich functional (2.2.2) is nothing but the evaluation $\pi \to \int c\, d\pi$ from Lemma 2.1.8 on couplings, and we see immediately that it is LSC as soon as this is true for the cost. We shall therefore not be surprised to find that minimisers always exist under these conditions. We need a lemma.

Lemma 2.2.5. *The set $\Pi(\mu,\nu)$ of couplings from $\mu \in \mathscr{P}(X)$ to $\nu \in \mathscr{P}(Y)$ is tight and closed under sequential C_b-convergence in the space $\mathscr{P}(X \times Y)$ of Borel probability measures on $X \times Y$ for spaces X, Y second countable LCH.*

Proof. The measures μ, ν are automatically regular by (2.1.9). Let $\epsilon > 0$ and pick compact sets $K_\mu \subset\subset X$ and $K_\nu \subset\subset Y$ such that $\mu(X \setminus K_\mu) \leq \epsilon/2$ and $\nu(Y \setminus K_\nu) \leq \epsilon/2$. Then $K_\mu \times K_\nu$ is compact in $X \times Y$, and for $\pi \in \Pi(\mu,\nu)$, one estimates

$$\pi\big((X \times Y) \setminus (K_\mu \times K_\nu)\big) \leq \pi\big((X \setminus K_\mu) \times Y\big) + \pi\big(X \times (Y \setminus K_\nu)\big) \tag{2.2.9}$$
$$= \mu(X \setminus K_\mu) + \nu(Y \setminus K_\nu) \tag{2.2.10}$$
$$\leq \epsilon. \tag{2.2.11}$$

This proves $\Pi(\mu,\nu)$ is tight.

If $\pi_k \xrightarrow{C_b} \pi$ for some sequence $(\pi_k)_{k\in\mathbb{N}}$ in $\Pi(\mu,\nu)$ and $\pi \in \mathscr{P}(X \times Y)$, then $\pi \in \Pi(\mu,\nu)$. Indeed, for $g \in C_b(X)$,

$$\int_X g\, d\mu = \int_{X\times Y} g \circ \pi^1\, d\pi_k \to \int_{X \times Y} g \circ \pi^1\, d\pi = \int_X g\, d(\pi^1_\sharp \pi). \tag{2.2.12}$$

This gives $\pi^1_\sharp \pi = \mu$ by Remark 2.2.3, and $\pi^2_\sharp \pi = \nu$ is proved similarly. ∎

Take a minimising sequence (π_k) in $\Pi(\mu,\nu)$ for the Kantorovich problem:

$$K(\pi_k) \to \inf_{\pi \in \Pi(\mu,\nu)} K(\pi). \tag{2.2.13}$$

Extract by Lemma 2.2.5 and Prokhorov's theorem a subsequence, still denoted (π_k), C_b-convergent to some $\pi^* \in \mathscr{P}(X \times Y)$. Then $\pi^* \in \Pi(\mu,\nu)$ still, and

$$K(\pi^*) \leq \liminf K(\pi_k) = \inf_\pi K(\pi) \tag{2.2.14}$$

by Lemma 2.1.8. This shows π^* is a minimiser and concludes the standard existence argument [35, Chapter 4].

Theorem 2.2.6. *If the cost $c\colon X \times Y \to \overline{\mathbb{R}}_+$ is LSC, then there always exists at least one minimiser for the Kantorovich problem: For all $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$, there is $\pi^* \in \Pi(\mu,\nu)$ such that*

$$K(\pi^*) = \min_\pi K(\pi). \tag{2.2.15}$$

Here we assume X, Y are second countable LCH.

See [11, Example 2.3.5] for an example of measures μ, ν where there are infinitely many couplings in $\Pi(\mu,\nu)$, all of which are optimal for the quadratic cost on $\mathbb{R}^2$. This is bad news for minimisers of the W_2-distance (Definition 2.3.2): They need not be unique. The problem is the singularity of μ. Indeed, Brenier's theorem (see [11, § 2.5.3] and [35, p. 209]) states that if $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^n)$ and $\mu \ll dx$, then there exists a unique optimal coupling $\pi^* \in \Pi(\mu, \nu)$ for the quadratic cost $c(x, y) = |x - y|^2$. This must also be optimal for the W_2-distance. Brenier's theorem also describes the structure of this coupling: It is Monge (Remark 2.2.4) and given by a transport map which is the gradient of a convex function, thus $\pi^* = (\mathrm{id} \times T)_\sharp \mu$, where $T = \nabla \phi$ for some convex $\phi \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.

2.3 Wasserstein Spaces

In this section X is a second countable LCH space, the model example being $\mathbb{R}^n$.

2.3.1 Generalities

A probability measure $\mu \in \mathscr{P}(X)$ has *finite pth moment* for $1 \leq p < \infty$ iff

$$\int_X d(x, x_0)^p \, d\mu(x) < \infty \tag{2.3.1}$$

for some (and then all) $x_0 \in X$. In Euclidean spaces this reduces to the requirement

$$\int_{\mathbb{R}^n} |x|^p \, d\mu(x) < \infty, \tag{2.3.2}$$

where the finiteness does not depend on the choice of norm. Letting

$$\mathscr{P}_p(X) := \{\mu \in \mathscr{P}(X) : \mu \text{ has finite } p\text{th moment}\}, \tag{2.3.3}$$

we see from the Hölder inequality (with exponent p/q) that for $1 \leq q \leq p < \infty$ there is an estimate

$$\left(\int d(x, x_0)^q \, d\mu(x) \right)^{1/q} \leq \left(\int d(x, x_0)^p \, d\mu(x) \right)^{1/p}, \tag{2.3.4}$$

hence an inclusion

$$\mathscr{P}_q(X) \hookrightarrow \mathscr{P}_p(X) \quad (1 \leq q \leq p < \infty). \tag{2.3.5}$$

Every $\mu \in \mathscr{P}(X)$ with finite pth moment thus has finite qth moment for all $1 \leq q \leq p < \infty$.

Remark 2.3.1. It is useful to know that a family (μ_α) of probability measures with uniformly bounded p-moments on $\mathbb{R}^n$ is tight. Indeed, assuming $\sup_\alpha \int |x|^p \, d\mu_\alpha(x) \leq C$, then

$$\mu_\alpha(\{|x| \geq R\}) \leq \frac{1}{R^p} \int |x|^p \, d\mu_\alpha(x) \leq \frac{C}{R^p} \quad (R > 0) \tag{2.3.6}$$

by the Chebychev inequality, and this gives tightness. The converse fails.

Definition 2.3.2 (Wasserstein metrics)**.** For $1 \le p < \infty$, the p-*Wasserstein distance* between $\mu, \nu \in \mathscr{P}_p(X)$ is by definition

$$W_p(\mu,\nu) := \left(\inf_\pi \int_{X\times X} d(\cdot,\cdot)^p \, d\pi\right)^{1/p}, \tag{2.3.7}$$

the infimum reaching over all couplings $\pi \in \Pi(\mu,\nu)$.

Remark 2.3.3. By (2.3.4), if $1 \le q \le p < \infty$, then

$$W_q(\mu,\nu) \le W_p(\mu,\nu) \quad (\mu,\nu \in \mathscr{P}_p(X)). \tag{2.3.8}$$

That these quantities are finite follows from the triangle inequality, producing for each $\pi \in \Pi(\mu,\nu)$ and $(a+b)^p \le 2^{p-1}(a^p + b^p)$ the estimate

$$\int d(x,y)^p \, d\pi(x,y) \le 2^{p-1}\Big(\int d(x,x_0)^p \, d\mu(x) + \int d(x_0,y)^p \, d\nu(y)\Big) < \infty. \tag{2.3.9}$$

That they are metrics is proved in for instance [35, p. 94] or [11, pp. 58–60]. Since each d^p is continuous $X \times X \to \overline{\mathbb{R}}_+$, it follows from Theorem 2.2.6 the existence of at least one optimal coupling $\pi^* \in \Pi(\mu,\nu)$ realising the infimum in the Wasserstein distance:

$$W_p(\mu,\nu)^p = \int_{X\times X} d(x,y)^p \, d\pi^*(x,y). \tag{2.3.10}$$

Such couplings need not be unique (§ 2.2.2). Convergence in W_p is a weak one:

Proposition 2.3.4 ([35, Theorem 6.9])**.** *With $1 \le p < \infty$ and (X,d) Polish, one has $W_p(\mu_k,\mu) \to 0$ for a sequence $(\mu_k)_{k\in\mathbb{N}}$ and measure μ in $\mathscr{P}_p(X)$ iff $\mu_k(f) \to \mu(f)$ for all $f \in C(X)$ with the growth estimate*

$$|f(x)| \le C(1 + d(x,x_0)^p) \quad (x \in X) \tag{2.3.11}$$

for some (and then all) $x_0 \in X$.

Remark 2.3.5 (Convergence of moments). Since the $d(\cdot,x_0)^p$ realise this growth, one can show [35, Definition 6.8] the above convergence holds iff

$$\mu_k \xrightarrow{C_b} \mu \quad \text{and} \quad \int d(\cdot,x_0)^p \, d\mu_k \to \int d(\cdot,x_0)^p \, d\mu. \tag{2.3.12}$$

2.3.2 Geodesics

A curve $(\mu_t)_{t\in[a,b]}$ in $\mathscr{P}_p(\mathbb{R}^n)$ is a *constant-speed geodesic* iff it satisfies [30, p. 98]

$$W_p(\mu_s,\mu_t) = \frac{|s-t|}{b-a} W_p(\mu_a,\mu_b) \quad (s,t \in [a,b]). \tag{2.3.13}$$

Clearly every such curve is Lipschitz. Further discussion about the regularity of geodesic curves may be found in [30, § 2.2] (without proofs) and in [1] (with proofs). See also [35, pp. 148–162]. Of particular interest are those curves in $\mathscr{P}_p(\mathbb{R}^n)$ which are absolutely continuous, since these exhibit precise relationships with distributional solutions of continuity equations $\partial_t\mu_t + \operatorname{div}(\mu_t v_t) = 0$. An overview of these results with pointers to connections with the formula of Benamou–Brenier is found surrounding [30, Theorem 4.6], but the definite reference is [1, Chapter 8].

Our interest in geodesic curves stem from the fact that they provide appropriate interpolations in the space $\mathscr{P}_p(\mathbb{R}^n)$, and functionals which are convex along such trajectories (originally introduced by McCann [21]) satisfy certain uniqueness properties in connection with gradient flows (Chapter 3).

Proposition 2.3.6. *Each optimal coupling $\pi^* \in \Pi(\mu,\nu)$ between $\mu,\nu \in \mathscr{P}_2(\mathbb{R})$ yields a constant-speed geodesic $(\mu_t)_{t\in[0,1]}$ from μ to ν by the formula*

$$\mu_t := \pi^t_\sharp \pi^* \quad (t \in [0,1]), \tag{2.3.14}$$

where $\pi^t := (1-t)\pi_1 + t\pi_2$ is the interpolation between the canonical projections $\pi_j \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ as in (2.2.5)*. Hence*

$$\pi^t(x,y) := (1-t)x + ty \quad (x, y \in \mathbb{R}^n, t \in [0,1]). \tag{2.3.15}$$

Remark 2.3.7. Compare (2.3.14) and (2.3.18) with McCann [21, Definition 1.1].

Proof. See [11, § 3.1.1] or [1, Theorem 7.2.2]. ■

For conceptual clarity, we suggest the name *optimal coupling geodesics* for constant-speed geodesics (2.3.14) arising from optimal couplings. In $\mathbb{R}^n$, there is a useful converse in the sense that *all* constant-speed geodesics in $\mathscr{P}_2(\mathbb{R}^n)$ are of this form. A proof of this fact may be found in [35, Corollary 7.22], [1, Theorem 7.2.2].

Corollary 2.3.8. *If $\mu \in \mathscr{P}_2(\mathbb{R}^n)$ has a Lebesgue density, then the W_2-geodesics emanating from μ are unique and given by transport maps.*

Proof. If $\nu \in \mathscr{P}_2(\mathbb{R}^n)$, the uniqueness of the constant-speed geodesic $(\mu_t)_{t\in[0,1]}$ from $\mu \ll dx$ to ν follows since it is given by (2.3.14) for an optimal coupling $\pi^* \in \Pi(\mu,\nu)$, and by Brenier's theorem this optimal coupling is unique (see the discussion at the end of § 2.2.2 and also [35, Corollary 7.23]). Moreover, Brenier's theorem shows that $\pi^* = (\mathrm{id} \times T)_\sharp\mu$ for some optimal transport map $T_\sharp\mu = \nu$ on $\mathbb{R}^n$.

Consider the action of μ_t on a test function $f \in C_c(\mathbb{R}^n)$:

$$\int f\, d\mu_t = \int f \circ \pi^t\, d\pi^* = \int f \circ \pi^t \circ (\mathrm{id} \times T)\, d\mu. \tag{2.3.16}$$

Letting $T_t \coloneqq (1-t)\,\mathrm{id}_{\mathbb{R}^n} + tT$, this takes the form

$$\int f\,d\mu_t = \int f\big((1-t)x + tT(x)\big)\,d\mu(x) = \int f \circ T_t\,d\mu, \tag{2.3.17}$$

from which it is clear that

$$\mu_t = (T_t)_\sharp \mu \quad (t \in [0,1]). \tag{2.3.18}$$

■

A functional $\mathscr{E}\colon \mathscr{P}_2(\mathbb{R}^n) \to \mathbb{R} \cup \{+\infty\}$ is *weakly geodesically convex* [11, § 4.3] iff for each $\mu_0, \mu_1 \in \mathscr{P}_2(\mathbb{R}^n)$, the estimate $\mathscr{E}(\mu_t) \le (1-t)\mathscr{E}(\mu_0) + t\mathscr{E}(\mu_1)$ holds for all $t \in [0,1]$ along some constant-speed geodesic $(\mu_t)_{t\in[0,1]}$ connecting μ_0 to μ_1. It is *geodesically convex* iff this estimate holds along all such geodesics.

2.3.3 Convolutions

The spaces $\mathscr{P}(\mathbb{R}^n)$ and $\mathscr{P}_2(\mathbb{R}^n)$ are stable under smoothing by convolution. By a mollifier we mean a smooth function $\phi \ge 0$ on $\mathbb{R}^n$ with compact support such that $\int \phi\,dx = 1$. To each such we associate the corresponding mollifier family (ϕ_ϵ) by $\phi_\epsilon(x) \coloneqq \epsilon^{-n}\phi(x/\epsilon)$ for $\epsilon > 0$. It is well-known that convolution against (ϕ_ϵ) produces smooth approximations, standard properties of which we summarise in [12, Appendix A]. We study it here for the spaces of probabilities following no particular reference.

Every finite Borel measure μ on $\mathbb{R}^n$ defines a tempered distribution by $\mu(f) \coloneqq \int f\,d\mu$ for f in the Schwartz space $\mathscr{S}(\mathbb{R}^n)$ of rapidly decreasing smooth functions, hence the convolution $\mu * f$, defined by

$$(\mu * f)(x) \coloneqq \mu(\tau_x \check{f}) = \int_{\mathbb{R}^n} f(x-y)\,d\mu(y) \quad (x \in \mathbb{R}^n), \tag{2.3.19}$$

is in this case a smooth function of polynomial growth [28, Theorem 7.19]. Here and elsewhere, $(\tau_x f)(y) \coloneqq f(y-x)$ and $\check{f}(x) \coloneqq f(-x)$ are the translation and reflection operators acting on functions.

The same formula (2.3.19) makes sense whenever $f \in B_\infty(\mathbb{R}^n)$ is bounded Borel, but there is no longer any reason to expect $\mu * f$ to be smooth. See [18, pp. 176, 275] and [28, § 9.14] for further properties of the convolution algebra $M(\mathbb{R}^n)$.

Lemma 2.3.9. *$\mathscr{P}(\mathbb{R}^n)$ is stable under mollification, and one has for each $\mu \in \mathscr{P}(\mathbb{R}^n)$ that $\mu * \phi_\epsilon \ge 0$ is a C^∞ Lebesgue density,*

$$\int_{\mathbb{R}^n} \mu * \phi_\epsilon\,dx = 1 \quad (\epsilon > 0), \tag{2.3.20}$$

such that $\mu * \phi_\epsilon \to \mu$ in the narrow topology of measures as $\epsilon \downarrow 0$.

Proof. By Tonelli,

$$\int \mu * \phi_\epsilon \, dx = \iint \phi_\epsilon(x-y) \, dx \, d\mu(y) = 1, \tag{2.3.21}$$

where we used translation-invariance and $\int \phi_\epsilon \, dx = 1$. This gives stability.

We now show $\int f(\mu * \phi_\epsilon) \, dx \to \int f \, d\mu$ for each $f \in C_b(\mathbb{R}^n)$. By Fubini,

$$\int f(\mu * \phi_\epsilon) \, dx = \int f * \check{\phi}_\epsilon \, d\mu \tag{2.3.22}$$

for the reflection $\check{\phi}_\epsilon(x) = \phi_\epsilon(-x)$. By regularity, pick $K \subset\subset \mathbb{R}^n$ such that $\mu(\mathbb{R}^n \backslash K)$ is small. Since $f * \check{\phi}_\epsilon \to f$ uniformly on compact sets [12, Appendix A],

$$\left| \int f * \check{\phi}_\epsilon \, d\mu - \int f \, d\mu \right| \le \int |f * \check{\phi}_\epsilon - f| \, d\mu \to 0 \quad \text{as} \quad \epsilon \downarrow 0 \tag{2.3.23}$$

by splitting the integral as one over K and one over $\mathbb{R}^n \setminus K$. The first becomes small by uniform convergence, and the second becomes small since $\mu(\mathbb{R}^n \setminus K)$ is small, f is bounded, and there is stability of mollification:

$$\|f * \check{\phi}_\epsilon\|_{L^\infty(\mathbb{R}^n)} \le \|f\|_{L^\infty(\mathbb{R}^n)} \quad (\epsilon > 0). \tag{2.3.24}$$

■

We now show stability of $\mathscr{P}_2(\mathbb{R}^n)$ as well.

Lemma 2.3.10. *If $\mu \in \mathscr{P}_2(\mathbb{R}^n)$ and $\phi \ge 0$ is $C_c(\mathbb{R}^n)$ with $\int \phi \, dx = 1$, then $\mu * \phi \in \mathscr{P}_2(\mathbb{R}^n)$, and there is an estimate*

$$\int_{\mathbb{R}^n} |x|^2 \, d(\mu * \phi)(x) \le \sum_{k=0}^{2} \binom{2}{k} \int_{\mathbb{R}^n} |x|^k \, d\phi(x) \int_{\mathbb{R}^n} |y|^{2-k} \, d\mu(y) \tag{2.3.25}$$

for the measure $d\phi \coloneqq \phi \, dx$.

Proof. One verifies as above that $\mu * \phi \ge 0$ is a C^0-density. With $\mu_\phi \coloneqq \mu * \phi \, dx$,

$$\int |x|^2 \, d\mu_\phi(x) = \int |x|^2 (\mu * \phi)(x) \, dx = \iint |x|^2 \phi(x-y) \, d\mu(y) \, dx. \tag{2.3.26}$$

Using Tonelli and changing variables, this becomes

$$\int |x|^2 \, d\mu_\phi(x) = \iint |x+y|^2 \phi(x) \, dx \, d\mu(y). \tag{2.3.27}$$

Using $|x+y|^2 \le |x|^2 + 2|x||y| + |y|^2$, this is bounded by

$$\begin{aligned} \iint |x|^2 \phi(x) \, dx \, d\mu(y) + 2 \int |y| \int |x| \phi(x) \, dx \, d\mu(y) \\ + \int \phi(x) \, dx \int |y|^2 \, d\mu(y). \end{aligned} \tag{2.3.28}$$

This is finite and yields (2.3.25) since ϕ has compact support and every $\mu \in \mathscr{P}_2(\mathbb{R}^n)$ has finite first moment by (2.3.4). ■

2.4 A Few Examples

In simple cases, it is possible to compute explicitly the set of transport maps and couplings between measures. We give several such examples and study the *Brinkman energy*

$$\mathscr{F}(\mu) = \frac{1}{2}\int_{\mathbb{R}} (k_1 * \mu)\, d\mu \quad (\mu \in \mathscr{P}_2(\mathbb{R})) \tag{2.4.1}$$

along Wasserstein geodesics (§ 2.3.2), where $k_1(x) = (1/2)e^{-|x|}$ is a fundamental solution to the operator $-\partial_x^2 + I$ (§ 3.2) that we hold fixed. The convolution

$$(k_1 * \mu)(x) = \int_{\mathbb{R}} k_1(x-y)\, d\mu(y) \geq 0 \quad (x \in \mathbb{R}) \tag{2.4.2}$$

exists for every $x \in \mathbb{R}$ and defines a bounded continuous function. It is thus μ-integrable and so $\mathscr{F} \geq 0$ is well-defined for every $\mu \in \mathscr{P}_2(\mathbb{R})$. If μ has density $\rho \geq 0$ with respect to Lebesgue, (2.4.1) takes the form of an interaction energy

$$\mathscr{F}(\rho) = \frac{1}{4}\iint e^{-|x-y|}\rho(x)\rho(y)\, dx\, dy. \tag{2.4.3}$$

One easily verifies the bounds $0 \leq \mathscr{F} \leq 1/4$, the maximum being achieved at every Dirac: $\mathscr{F}(\delta_x) = 1/4$ for all $x \in \mathbb{R}$. By (2.4.38), we also learn that

$$\inf_{\mu \in \mathscr{P}_2(\mathbb{R})} \mathscr{F}(\mu) = 0 \tag{2.4.4}$$

upon considering the uniform densities $\mu \sim (1/t)1_{[0,t]}$ as $t \to +\infty$.

Example 2.4.1 (Transport of Diracs)**.** Given arbitrary topological spaces X and Y with points $x_0 \in X$ and $y_0 \in Y$, the set of *all* transport maps from δ_{x_0} to δ_{y_0} is precisely

$$\{T\colon X \to Y \text{ Borel and } T(x_0) = y_0\}. \tag{2.4.5}$$

It is immaterial what such T do to points besides x_0.

Suppose X, Y are second countable Hausdorff. Each T in (2.4.5) yields a Monge coupling (Remark 2.2.4) $\pi_T \coloneqq (\mathrm{id} \times T)_\sharp \delta_{x_0} \in \Pi(\delta_{x_0}, \delta_{y_0})$ with cost

$$K(\pi_T) = \int_{X\times Y} c\, d\pi_T = \int_X c \circ (\mathrm{id} \times T)\, d\delta_{x_0} = c(x_0, y_0) \tag{2.4.6}$$

for any $\mathscr{B}(X \times Y)$-measurable $c\colon X \times Y \to \overline{\mathbb{R}}_+$. All these π_T are in fact the same measure $\delta_{(x_0,y_0)}$, and

$$\Pi(\delta_{x_0}, \delta_{y_0}) = \{\delta_{(x_0,y_0)}\}. \tag{2.4.7}$$

Proof. If $\pi \in \Pi(\delta_{x_0}, \delta_{y_0})$, then $\pi(\{x_0\} \times Y) = \delta_{x_0}(\{x_0\}) = 1$ and similarly $\pi(X \times \{y_0\}) = 1$. The intersection $\{(x_0, y_0)\} = (\{x_0\} \times Y) \cap (X \times \{y_0\})$ thus has full measure and so π is concentrated on $\{(x_0, y_0)\}$. ■

Consequently, for any transport map $T_\sharp \delta_{x_0} = \delta_{y_0}$ and any coupling $\pi \in \Pi(\delta_{x_0}, \delta_{y_0})$, we find

$$K(\pi) = c(x_0, y_0) \quad \text{and} \quad M(T) = c(x_0, y_0). \tag{2.4.8}$$

This gives a complete solution to the Monge and Kantorovich problems for Dirac measures in spaces second countable Hausdorff. In particular,

$$W_p(\delta_{x_0}, \delta_{y_0}) = \left(\inf_\pi \int |x - y|^p \, d\pi(x, y) \right)^{1/p} = |x_0 - y_0| \tag{2.4.9}$$

for any $1 \leq p < \infty$ and any $x_0, y_0 \in \mathbb{R}^n$ for the p-Wasserstein distances in $\mathbb{R}^n$, and similarly for separable metric spaces.

Corollary 2.4.2. *For separable metric spaces* (X, d)*, there are isometric embeddings*

$$\begin{cases} X \hookrightarrow \mathscr{P}_p(X) \\ x \mapsto \delta_x \end{cases} \tag{2.4.10}$$

into the p*-Wasserstein spaces for* $1 \leq p < \infty$.

Example 2.4.3 (Transport from Dirac to arbitrary measures)**.** Drawing a support diagram in $X \times Y$, it is easy to guess that

$$\Pi(\delta_{x_0}, \mu) = \{\delta_{x_0} \otimes \mu\} \tag{2.4.11}$$

for all $\mu \in \mathscr{P}(Y)$. We prove this. Given $\pi \in \Pi(\delta_{x_0}, \mu)$, it is sufficient by uniqueness of the product measure to prove

$$\pi(A \times B) = \delta_{x_0}(A)\mu(B) \quad \text{for all} \quad A \in \mathscr{B}(X), \quad B \in \mathscr{B}(Y). \tag{2.4.12}$$

Arguing as in Example 2.4.1, one finds that π is concentrated on $\{x_0\} \times Y$. This takes care of (2.4.12) when $x_0 \notin A$. For $x_0 \in A$, the set $A \times B$ decomposes into the disjoint union of $\{x_0\} \times B$ and $(A \setminus \{x_0\}) \times B$. This gives $\pi(A \times B) = \pi(\{x_0\} \times B)$. To prove this agrees with $\mu(B)$, use the marginal formula (2.2.6) and the concentration of π to compute

$$\mu(B) = \int_Y 1_B \, d\mu = \int_{\{x_0\} \times Y} 1_B(y) \, d\pi(x, y) \tag{2.4.13}$$

$$= \int_{X \times Y} 1_{\{x_0\} \times B} \, d\pi = \pi(\{x_0\} \times B), \tag{2.4.14}$$

completing the proof.

This gives a complete solution to the Kantorovich problem for transporting a Dirac to an arbitrary measure $\mu \in \mathscr{P}(Y)$, which by (2.4.11) and Tonelli comes at the minimal (and only) cost

$$\inf_{\pi} K(\pi) = K(\delta_{x_0} \otimes \mu) = \int_Y c(x_0, y)\, d\mu(y). \tag{2.4.15}$$

Incidentally, this gives a formula for Wasserstein distances in separable metric spaces (X, d) in terms of moments:

$$W_p(\delta_{x_0}, \mu) = \Big(\int_Y d(x_0, y)^p \, d\mu(y) \Big)^{1/p} \quad (x_0 \in X, \mu \in \mathscr{P}_p(X)), \tag{2.4.16}$$

and in $\mathbb{R}^n$,

$$W_p(\delta_0, \mu) = \Big(\int_{\mathbb{R}^n} |y|^p \, d\mu(y) \Big)^{1/p}. \tag{2.4.17}$$

Combining Example 2.4.3 with Proposition 2.3.6 gives a constant-speed geodesic $(\mu_t)_{t\in[0,1]}$ from a Dirac to any measure $\mu \in \mathscr{P}_2(\mathbb{R})$ by $\mu_t = \pi^t_\sharp(\delta_0 \otimes \mu)$. Its action on $g \in B_\infty(\mathbb{R})$ bounded Borel is

$$\int g\, d\mu_t = \int g \circ \pi^t(0, y)\, d\mu(y) = \int g(ty)\, d\mu(y). \tag{2.4.18}$$

The unique positive Borel measure with this action is the rescaled probability

$$\mu_t(B) = \mu(t^{-1}B) \quad (t > 0, B \in \mathscr{B}(\mathbb{R})). \tag{2.4.19}$$

If μ has Lebesgue density $\rho \geq 0$, then μ_t has density $\rho_t \geq 0$, where

$$\rho_t = \frac{1}{t}\rho(\cdot/t) \quad (t > 0). \tag{2.4.20}$$

This is seen by using $\mu \sim \rho$ in (2.4.18):

$$\int g\, d\mu_t = \int g(tx)\rho(x)\, dx = \frac{1}{t}\int g(x)\rho(x/t)\, dx \quad (g \in B_\infty(\mathbb{R})), \tag{2.4.21}$$

and (2.4.20) follows. Clearly $\operatorname{supp}\rho_t = t \operatorname{supp}\rho$, and so we expect (ρ_t) to spike at zero in the limit as $t \downarrow 0$.

Lemma 2.4.4. *For the geodesic densities $(\rho_t)_{t\in[0,1]}$ of (2.4.20), one has $\rho_t \to \delta_0$ in $\mathscr{P}_2(\mathbb{R})$, hence*

$$\int_{\mathbb{R}} g(x)\rho_t(x)\, dx \to g(0) \quad \textit{as} \quad t \downarrow 0 \tag{2.4.22}$$

for all $g \in C(\mathbb{R})$ with $|g(x)| \leq C(1 + x^2)$.

Proof. Of course, $\mu_t \to \delta_0$ in $\mathscr{P}_2(\mathbb{R})$ also when μ has no density. An easy proof uses that (μ_t) is a constant-speed geodesic:

$$W_2(\mu_t, \delta_0) = tW_2(\mu, \delta_0) \to 0 \quad \text{as} \quad t \downarrow 0. \tag{2.4.23}$$

Another proof is afforded by (2.4.17) along with (2.4.18), yielding

$$W_2^2(\mu_t, \delta_0) = t^2 \int_{\mathbb{R}} x^2 \, d\mu(x) \to 0 \quad \text{as} \quad t \downarrow 0. \tag{2.4.24}$$

This gives (2.4.22) by Proposition 2.3.4. ∎

We now compute the Brinkman energy $\mathscr{F}$ (2.4.1) along the geodesic $(\mu_t)_{t\in[0,1]}$ from δ_0 to any $\mu \in \mathscr{P}_2(\mathbb{R})$. First,

$$(k_1 * \mu_t)(tx) = \int k_1(tx - y) \, d\mu_t(y) = \int k_1\big(t(x-y)\big) \, d\mu(y) \tag{2.4.25}$$

by (2.4.18), hence $(k_1 * \mu_t)(t\cdot) = k_1(t\cdot) * \mu$ and

$$\mathscr{F}(\mu_t) = \frac{1}{2} \int (k_1 * \mu_t)(x) \, d\mu_t(x) \tag{2.4.26}$$

$$= \frac{1}{2} \int (k_1 * \mu_t)(tx) \, d\mu(x) \tag{2.4.27}$$

$$= \frac{1}{2} \int_{\mathbb{R}} (k_1(t\cdot) * \mu) \, d\mu. \tag{2.4.28}$$

Example 2.4.5 (Dirac to Dirac)**.** As seen above, the optimal coupling between two Diracs δ_p and δ_q in $\mathbb{R}$ is $\delta_p \otimes \delta_q = \delta_{(p,q)}$, hence (2.3.14) gives the constant-speed geodesic

$$\mu_t = \pi^t_\sharp \delta_{(p,q)} = \delta_{\pi^t(p,q)} = \delta_{(1-t)p+tq} \quad (t \in [0,1]), \tag{2.4.29}$$

a family of Diracs travelling along the straight line segment from p to q. Our functional is constant along this geodesic:

$$\mathscr{F}\big(\delta_{(1-t)p+tq}\big) = \frac{1}{4} \quad (t \in [0,1]). \tag{2.4.30}$$

Example 2.4.6 (Dirac to Gaussian)**.** The constant-speed geodesic from δ_0 to the standard Gaussian (whose density we denote by g) given by (2.3.14) is the curve $(g_t)_{t\in[0,1]}$ of densities

$$g_t(x) = \frac{1}{t} g(x/t) = \frac{1}{t\sqrt{2\pi}} e^{-(x/t)^2/2}. \tag{2.4.31}$$

This is a classical approximation of identity, and we see from Lemma 2.4.4 that $g_t \to \delta_0$ narrowly. By Lemma 2.4.8 below, the Brinkman energy

$$\mathscr{F}(g_t) = \frac{1}{4\pi t^2} \int_{\mathbb{R}^2} e^{-|x-y|} e^{-\frac{1}{2}(x^2+y^2)/t^2} \, d(x,y) \tag{2.4.32}$$

is convex for $t \in [0,1]$.

In numerical simulations, one may truncate the integral representation (2.4.28) at $R > 0$ with uniform error control in $t \in [0,1]$. Indeed, writing

$$\mathscr{F}(g_t) = \frac{1}{2}\Big(\int_{-R}^{R} (k_1 * g_t) g_t \, dx + \int_{|x|\geq R} (k_1 * g_t) g_t \, dx\Big), \tag{2.4.33}$$

letting $E_t(R)$ be the parenthesised error term, and observing $\|k_1 * g_t\|_{L^\infty} \leq 1/2$, one estimates

$$|E_t(R)| \leq \frac{1}{\sqrt{2\pi}} \int_R^\infty e^{-\frac{1}{2}x^2} \, dx = \sqrt{\frac{\pi}{2}}\Big(1 - \operatorname{erf}\Big(\frac{R}{\sqrt{2}}\Big)\Big) \quad (t \in [0,1]). \tag{2.4.34}$$

Here $\operatorname{erf}(x) := (2/\sqrt{\pi}) \int_0^x e^{-s^2} \, ds$ is the error function, for which efficient implementations are available on most modern computing systems.

Example 2.4.7 (Dirac to uniform)**.** According to (2.4.20), the geodesic density from δ_0 to the uniform density $\rho := 1_{[0,1]}$ is given by

$$\rho_t := \frac{1}{t} 1_{[0,1]}(\cdot/t) = \frac{1}{t} 1_{[0,t]}, \tag{2.4.35}$$

the uniform density on $[0,t]$. For this we compute as in (2.4.28) first

$$\big(k_1(t\cdot) * 1_{[0,1]}\big)(x) = \int_{[0,1]} k_1\big(t(x-y)\big) \, dy, \tag{2.4.36}$$

and then the Brinkman energy along the geodesic:

$$\mathscr{F}(\rho_t) = \frac{1}{2} \int (k_1(t\cdot) * \rho)\rho \, dx = \frac{1}{4} \int_{[0,1]^2} e^{-t|x-y|} \, d(x,y). \tag{2.4.37}$$

This works out to

$$\mathscr{F}(\rho_t) = \frac{1}{2}\Big(\frac{1}{t} - \frac{1-e^{-t}}{t^2}\Big) \quad (t > 0), \tag{2.4.38}$$

which is continuous, convex, and decreasing on $[0,1]$ (in fact, on $[0,\infty)$), taking the value $\mathscr{F}(\delta_0) = 1/4$ at the left endpoint and $\mathscr{F}(1_{[0,1]}) = 1/(2e)$ at the right endpoint. We also observe $\mathscr{F}(\rho_t) \downarrow 0$ as $t \to +\infty$.

In all of the above examples, the geodesic energy $\mathscr{F}(\mu_t)$ is convex. This is no coincidence. We prove it now for all geodesics emanating from δ_0, but shall in § 3.3 see that it in fact holds along geodesics connecting any two measures in $\mathscr{P}_2(\mathbb{R})$.

Lemma 2.4.8 (Geodesic convexity from δ_0)**.** *The energy $\mathscr{F}$ is convex along the W_2-geodesic $(\mu_t)_{t\in[0,1]}$ from the Dirac measure δ_0 to any $\mu \in \mathscr{P}_2(\mathbb{R})$.*

Proof. By (2.4.28), one has

$$\mathscr{F}(\mu_t) = \frac{1}{4} \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-t|x-y|} \, d\mu(x) \, d\mu(y) \quad (t \in [0,1]). \tag{2.4.39}$$

Conjuring up the inequality $e^{-tz} - te^{-z} \leq 1 - t$ for $t \in [0,1]$ and all $z \geq 0$ (the map $t \mapsto t + e^{-tz} - te^{-z}$ is for each $z > 0$ strictly convex on $[0,1]$ with value 1 at both endspoints), replacing z with $|x-y|$, and integrating yields

$$\mathscr{F}(\mu_t) \leq \frac{1}{4}(1-t) + \frac{t}{4} \iint e^{-|x-y|} \, d\mu(x) \, d\mu(y) \quad (t \in [0,1]). \tag{2.4.40}$$

Now $\mathscr{F}(\mu_t) \leq (1-t)\mathscr{F}(\delta_0) + t\mathscr{F}(\mu)$ follows upon recalling $\mathscr{F}(\delta_0) = 1/4$. ■

3 Convergence of the JKO Scheme

We prove here convergence of the JKO scheme for the Brinkman functional

$$\mathscr{F}(\mu) = \frac{1}{2}\int_{\mathbb{R}} (k_1 * \mu)\, d\mu \tag{3.0.1}$$

on the Wasserstein space $\mathscr{P}_2(\mathbb{R})$ in one dimension, following the outline in Ambrosio, Gigli, Savaré (henceforth AGS) [1, § 11.1.3]. We suggest suitable adaptations where we were unable to reproduce their estimates. Here, $k_1(x) = (1/2)e^{-|x|}$.

3.1 Motivation

A gradient flow of a smooth function $f\colon \mathbb{R}^n \to \mathbb{R}$ starting at $u_0 \in \mathbb{R}^n$ is a smooth curve (u_t) satisfying

$$\begin{cases} \partial_t u_t = -\nabla f(u_t) \quad (t \in \mathbb{R}) \\ u(0) = u_0 \end{cases}. \tag{3.1.1}$$

Along the gradient flow, there holds

$$\partial_t f(u_t) = \nabla f(u_t) \cdot \partial_t u_t = -|\nabla f(u_t)|^2 \leq 0, \tag{3.1.2}$$

hence the values of f decrease as the flow progresses. A good summary of gradient flows of strongly convex functionals in a Euclidean setting along with their exponential convergence towards minimisers is found in [4, § 3.1]. See [30, § 2.1] for extensions to subdifferential flows of convex functionals.

It is well-known that the heat equation

$$\partial_t u_t = \Delta u_t \tag{3.1.3}$$

may be viewed as the gradient flow of the Dirichlet energy

$$\frac{1}{2}\int_{\mathbb{R}^n} |\nabla u|^2\, dx \tag{3.1.4}$$

on the Hilbert space $L^2(\mathbb{R}^n)$. This is proved for instance in Figalli [11, § 3.2] by observing that the Dirichlet subdifferential is non-empty at $u \in L^2(\mathbb{R}^n)$ precisely when $\Delta u \in L^2$, in which case the only subdifferential point is $-\Delta u$.

It was discovered by Jordan, Kinderlehrer, and Otto (henceforth JKO) at the end of the 1990s [16] that the heat equation may similarly be viewed as the gradient flow of the negative Boltzmann entropy

$$\int_{\mathbb{R}^n} \rho \log \rho\, dx \tag{3.1.5}$$

on the Wasserstein space $\mathscr{P}_2(\mathbb{R}^n)$ with respect to the W_2-metric.

The method of proof in JKO [16] is based upon discretising in time the gradient flow equation and is derived thus. Given an energy functional $f\colon H \to \mathbb{R}$ smooth enough (here H is a Hilbert space), consider its gradient flow

$$u'(t) = -\nabla f(u(t)) \quad (t > 0) \tag{3.1.6}$$

starting at $u(0) = u_0$. The idea is to introduce a time-step $\tau > 0$, employ the implicit Euler scheme, and use the recurrence

$$\frac{u_{k+1}^{(\tau)} - u_k^{(\tau)}}{\tau} = -\nabla f(u_{k+1}^{(\tau)}) \tag{3.1.7}$$

as a discrete approximation of the gradient flow for the energy f. Formally, the next point $u_{k+1}^{(\tau)}$ is found by solving $\nabla g_k^{(\tau)}(u_{k+1}^{(\tau)}) = 0$ for the augmented functional

$$g_k^{(\tau)}(v) \coloneqq \frac{1}{2\tau}\|v - u_k^{(\tau)}\|_H^2 + f(v) \quad (v \in H). \tag{3.1.8}$$

It therefore makes sense to look for minimisers of the variational problem

$$\inf_{v\in H}\left(\frac{1}{2\tau}\|v - u_k^{(\tau)}\|_H^2 + f(v)\right) \tag{3.1.9}$$

at each step. Constructing sequences $(u_k^{(\tau)})_{k\in\mathbb{N}}$ of minimisers in the above variational scheme and studying their convergence to limit curves as $\tau \to 0$ is in the literature known as the *JKO scheme*. The authors of [16] replace the Hilbert distance with the Wasserstein metric to obtain instead the recurrence

$$u_{k+1}^{(\tau)} \in \operatorname*{arg\,min}_{v\in\mathscr{P}_2(\mathbb{R}^n)}\left(\frac{1}{2\tau}W_2^2(v, u_k^{(\tau)}) + f(v)\right) \tag{3.1.10}$$

when f is defined on the Wasserstein space.

There is a formal Riemannian calculus on $\mathscr{P}_2(\mathbb{R}^n)$ first discovered by Otto [25], in which the Wasserstein gradient of the Brinkman functional (3.0.1) takes the form

$$\operatorname{grad}_{W_2}\mathscr{F}(\rho) = -\operatorname{div}\left(\rho\nabla\left(\frac{\delta\mathscr{F}}{\delta\rho}\right)\right), \tag{3.1.11}$$

where $\delta\mathscr{F}/\delta\rho$ is the variational derivative. We do not elaborate on this here (see for instance [11, Chapter 4] and the references therein) and remark only that formally

$$\frac{\delta\mathscr{F}}{\delta\rho} = k_1 * \rho. \tag{3.1.12}$$

On a formal level, this shows that the Wasserstein gradient flow of $\mathscr{F}$ on $\mathscr{P}_2(\mathbb{R})$ is

$$\partial_t\rho_t - \operatorname{div}(\rho_t\nabla(k_1 * \rho)) = 0, \tag{3.1.13}$$

and the connection with (1.0.1) is apparent as $w \coloneqq k_1 * \rho$ will under suitable regularity assumptions solve $-w'' + w = \rho$ since k_1 is fundamental (§ 3.2).

3.2 Bessel Potentials

By the Malgrange–Ehrenpreis theorem [28, Theorem 8.5], the constant-coefficient operator $\mathscr{L} := -\Delta + I$ admits a fundamental solution u in the space $\mathscr{D}'(\mathbb{R}^n)$ of Schwartz distributions:

$$-\Delta u + u = \delta_0 \quad \text{in} \quad \mathscr{D}'(\mathbb{R}^n). \tag{3.2.1}$$

Such u are not unique in $\mathscr{D}'(\mathbb{R}^n)$ since $\mathscr{L}$ has non-trivial kernel. Indeed, $\mathscr{L} u_\alpha = 0$ for $u_\alpha(x) := e^{\alpha \cdot x}$ for any $\alpha \in \mathbb{R}^n$ with $|\alpha| = 1$. By elliptic regularity [28, Theorem 8.12], every $v \in \mathscr{D}'(\mathbb{R}^n)$ solving $\mathscr{L} v = 0$ is $C^\infty(\mathbb{R}^n)$, so any two fundamental solutions (3.2.1) differ by a smooth function in $\ker \mathscr{L}$.

In one dimension, one checks directly that

$$k_1(x) := \frac{1}{2} e^{-|x|} \quad (x \in \mathbb{R}) \tag{3.2.2}$$

is fundamental for $\mathscr{L}$. It satisfies $k_1 \geq 0$ with $\int k_1 \, dx = 1$ and belongs to $C_0(\mathbb{R}) \cap W^{1,p}(\mathbb{R})$ for all $1 \leq p \leq \infty$. However, $k_1'(x) = -(1/2)(\operatorname{sgn} x) e^{-|x|}$ is not continuous (at the origin), and $k_1'' = k_1 - \delta_0$ is not in $L^1_{\text{loc}}(\mathbb{R})$.

See Evans [10, p. 191] for a similar analysis and an integral formula for the Bessel potentials in $\mathbb{R}^n$.

3.3 The Brinkman Energy

Basic properties of the Brinkman energy were listed at the beginning of § 2.4. We now establish its convexity and lower semicontinuity properties.

Lemma 3.3.1. *$\mathscr{F}$ is LSC with respect to vague convergence in $\mathscr{P}_2(\mathbb{R})$.*

Proof. Suppose $\mu_k \to \mu$ in $\mathscr{P}_2(\mathbb{R})$ vaguely. Since

$$\mathscr{F}(\mu_k) = \frac{1}{2} \int_{\mathbb{R}} (k_1 * \mu_k) \, d\mu_k = \frac{1}{2} \int_{\mathbb{R}^2} k_1(x-y) \, d(\mu_k \otimes \mu_k)(x,y) \tag{3.3.1}$$

for the product measures, the conclusion follows from Lemma 2.1.8 as soon as it is seen that $\mu_k \otimes \mu_k \to \mu \otimes \mu$ vaguley in $\mathscr{P}(\mathbb{R}^2)$. We prove this.

Let $f \in C_c(\mathbb{R}^2)$ be supported in a large square $[a,b]^2$. Define

$$F_k(y) := \int_{\mathbb{R}} f(x,y) \, d\mu_k(x), \quad F(y) := \int_{\mathbb{R}} f(x,y) \, d\mu(x). \tag{3.3.2}$$

Then $F_k, F \in C_c(\mathbb{R})$, all with support in $[a,b]$. We estimate

$$|\mu_k(F_k) - \mu(F)| \leq |\mu_k(F_k) - \mu_k(F)| + |\mu_k(F) - \mu(F)|. \tag{3.3.3}$$

The second terms vanish since $\mu_k(F) \to \mu(F)$ by assumption. The first terms vanish as soon as it is seen that $F_k \to F$ uniformly. This rests on the following observations. One checks $F_k \to F$ pointwise since $\mu_k \to \mu$ vaguely, and $(F_k)_{k\in\mathbb{N}}$ is uniformly equicontinuous since f is uniformly continuous. Consider any compact set $K \subset\subset \mathbb{R}$. Any subsequence of (F_k) has a further subsequence converging uniformly to F on K by Arzelà–Ascoli. Arguing by contradiction, this implies $F_k \to F$ uniformly on K. The uniform convergence escalates to all of $\mathbb{R}$ since the F_k, F are supported in $[a,b]$. Estimate (3.3.3) therefore gives

$$\int_{\mathbb{R}^2} f\, d(\mu_k \otimes \mu_k) \to \int_{\mathbb{R}^2} f\, d(\mu \otimes \mu). \tag{3.3.4}$$

■

In (2.4.4) we saw that $\inf \mathscr{F}(\mu) = 0$ over $\mathscr{P}_2(\mathbb{R})$. One might hope to find minimisers by considering a minimising sequence (μ_k) and obtain some form of compactness and appeal to semicontinuity. This does not obtain. Indeed, taking $\mu_k \sim \rho_k$ uniform on $[0,k]$, we know from (2.4.38) that $(\mu_k)_{k\in\mathbb{N}}$ is minimising for $\mathscr{F}$.

Lemma 3.3.2. *No subsequence of (ρ_k) converges in $\mathscr{P}_2(\mathbb{R})$, and no subsequence converges vaguely to any probability measure on $\mathbb{R}$.*

Proof. If $\rho_{k_n} \to \mu$ vaguely for some $\mu \in \mathscr{P}(\mathbb{R})$, then

$$\int_{\mathbb{R}} f\, d\mu = \lim_n \int f\, d\rho_{k_n} = \lim_n ⨍_{[0,k_n]} f\, dx = 0 \tag{3.3.5}$$

for the sequence of averages and all $f \in C_c(\mathbb{R})$. Hence μ would be a vanishing Radon integral on $C_c(\mathbb{R})$, and therefore $\mu = 0$. Contradiction. ■

As promised at the end of § 2.4, we prove that the Brinkman functional is convex along constant-speed geodesics.

Lemma 3.3.3. *One has $\mathscr{F}(\mu_t) \le (1-t)\mathscr{F}(\mu) + t\mathscr{F}(\nu)$ for all $t \in [0,1]$ whenever $(\mu_t)_{t\in[0,1]}$ is a constant-speed geodesic connecting $\mu \ll dx$ to ν in $\mathscr{P}_2(\mathbb{R})$.*

Proof. Recall from Corollary 2.3.8 that the constant-speed geodesic connecting μ to ν is unique and given by the unique optimal transport map $T_\sharp\mu = \nu$ since μ has a density. Explicitly, $\mu_t = (T_t)_\sharp\mu$ for $T_t := (1-t)\,\mathrm{id}_{\mathbb{R}} + tT$. We compute

$$(k_1 * \mu_t)(x) = \int_{\mathbb{R}} k_1(x-y)\, d\mu_t(y) = \int_{\mathbb{R}} k_1(x - T_t(y))\, d\mu(y), \tag{3.3.6}$$

and then

$$\begin{aligned}
\mathscr{F}(\mu_t) &= \frac{1}{2}\int (k_1 * \mu_t)(T_t(x))\, d\mu(x) && (3.3.7)\\
&= \frac{1}{2}\iint k_1(T_t(x) - T_t(y))\, d\mu(y)\, d\mu(x) && (3.3.8)\\
&= \frac{1}{2}\iint k_1\Big((1-t)(x-y) + t(T(x) - T(y))\Big)\, d\mu(y)\, d\mu(x). && (3.3.9)
\end{aligned}$$

The standard remark is that this is convex if k_1 is convex, but our k_1 is only piecewise convex on the half-lines $(-\infty, 0]$ and $[0, +\infty)$. This is where one-dimensionality enters the picture and the proof breaks down in higher dimensions: Optimal transport in one dimension is μ-essentially increasing (apply for instance the convex gradient $T = \nabla\phi$ afforded by Brenier in § 2.2.2), and for this reason $x - y$ and $T(x) - T(y)$ always have the same sign except perhaps on a $(\mu \otimes \mu)$-null set. Splitting the integral over $\{x \leq y\}$ and $\{x > y\}$ and using convexity of k_1 on each half-line yields

$$\mathscr{F}(\mu_t) \leq \frac{1}{2} \int_{\mathbb{R}^2} (1-t)k_1(x-y) + tk_1(T(x) - T(y))\, d(\mu \otimes \mu)(x, y). \quad (3.3.10)$$

This gives the desired estimate upon applying $T_\sharp\mu = \nu$. ■

3.4 Well-Posedness of the Variational Scheme

Given a time-step $\tau > 0$, define the augmented energy functional

$$J_\tau(\mu|\nu) := \frac{1}{2\tau} W_2^2(\mu, \nu) + \mathscr{F}(\mu) \geq 0 \quad (\mu, \nu \in \mathscr{P}_2(\mathbb{R})) \quad (3.4.1)$$

for the JKO scheme. Our goal in this section is to prove the well-posedness of the one-step variational movement for J_τ.

Lemma 3.4.1. *To each $\nu \in \mathscr{P}_2(\mathbb{R})$ corresponds a minimiser $\mu^* \in \mathscr{P}_2(\mathbb{R})$ of the variational problem*

$$J_\tau(\mu^*|\nu) = \inf_{\mu \in \mathscr{P}_2(\mathbb{R})} J_\tau(\mu|\nu) \geq 0. \quad (3.4.2)$$

We need a lemma.

Lemma 3.4.2. *$J_\tau(\cdot|\nu)$ is LSC with respect to narrow convergence on $\mathscr{P}_2(\mathbb{R})$ for each $\nu \in \mathscr{P}_2(\mathbb{R})$.*

Proof. Assume $\mu_k \to \mu$ narrowly in $\mathscr{P}_2(\mathbb{R})$. Then $W_2^2(\mu, \nu) \leq \liminf W_2^2(\mu_k, \nu)$ by [1, Lemma 7.1.4]. Also, $\mathscr{F}(\mu) \leq \liminf \mathscr{F}(\mu_k)$ by Lemma 3.3.1. This gives $J_\tau(\mu|\nu) \leq \liminf J_\tau(\mu_k|\nu)$. ■

Proof of Lemma 3.4.1. Let (μ_k) be a minimising sequence, so $J_\tau(\mu_k|\nu) \to \inf_\mu J_\tau(\mu|\nu)$. Then

$$\frac{1}{2\tau} W_2^2(\mu_k, \nu) = J_\tau(\mu_k|\nu) - \mathscr{F}(\mu_k) \quad (3.4.3)$$

is the difference between two bounded sequences, hence is bounded:

$$W_2(\mu_k, \nu) \leq C \quad \text{for all} \quad k \in \mathbb{N}. \quad (3.4.4)$$

By (2.4.17), we have a uniform control on the second moments:

$$\int_{\mathbb{R}} x^2 \, d\mu_k(x) = W_2^2(\delta_0, \mu_k) \le \big(W_2(\delta_0, \nu) + C\big)^2 \quad (k \in \mathbb{N}). \tag{3.4.5}$$

Now (μ_k) is tight by Remark 2.3.1. We get a limit measure $\mu^* \in \mathscr{P}(\mathbb{R})$ such that $\mu_k \to \mu^*$ narrowly along a subsequence by Prokhorov. We observe

$$\int x^2 \, d\mu^*(x) \le \liminf \int x^2 \, d\mu_k(x) < \infty \tag{3.4.6}$$

by (3.4.5) and Lemma 2.1.8. This shows $\mu^* \in \mathscr{P}_2(\mathbb{R})$. By Lemma 3.4.2,

$$J_\tau(\mu^* | \nu) \le \liminf J_\tau(\mu_k | \nu) = \inf_{\mu} J_\tau(\mu | \nu). \tag{3.4.7}$$

Hence μ^* is a minimiser. ■

3.5 Convergence of the JKO Scheme

With notation as in § 3.4 and given any initial measure $\mu_0 \in \mathscr{P}_2(\mathbb{R})$, we form a *JKO sequence* $(\mu_k^{(\tau)})_{k\in\mathbb{N}}$ for $\mathscr{F}$ starting at μ_0 with time-step $\tau > 0$ by iteratively solving the variational problem (3.4.2):

$$\begin{cases} \mu_0^{(\tau)} \coloneqq \mu_0 \\ \mu_{k+1}^{(\tau)} \in \operatorname*{arg\,min}_{\mu \in \mathscr{P}_2(\mathbb{R})} J_\tau\Big(\mu \,\Big|\, \mu_k^{(\tau)}\Big) \quad (k \in \mathbb{N}) \end{cases}. \tag{3.5.1}$$

This is well-defined by Lemma 3.4.1 and the axiom of dependent choice. We fix such a sequence for each $\tau > 0$ for the rest of this chapter.

To each $\tau > 0$ corresponds a uniform partition $0 < \tau < 2\tau < \cdots$ of the positive real numbers $\mathbb{R}_+ \coloneqq [0, +\infty)$. Following AGS [1, § 11.1.3], we define two compatible interpolations of each JKO sequence $(\mu_k^{(\tau)})_{k\in\mathbb{N}}$:

- Let $\overline{\mu}_\tau \colon \mathbb{R}_+ \to \mathscr{P}_2(\mathbb{R})$ be the piecewise-constant interpolation such that $\overline{\mu}_\tau(0) \coloneqq \mu_0$ and $\overline{\mu}_\tau(t) \coloneqq \mu_{k+1}^{(\tau)}$ if $k \in \mathbb{N}$ is such that $t \in (k\tau, (k+1)\tau]$.

- Let $\widehat{\mu}_\tau \colon \mathbb{R}_+ \to \mathscr{P}_2(\mathbb{R})$ be obtained from the JKO sequence $(\mu_k^{(\tau)})_{k\in\mathbb{N}}$ by connecting each $\mu_k^{(\tau)}$ to $\mu_{k+1}^{(\tau)}$ by means of a constant-speed geodesic on $[k\tau, (k+1)\tau]$.

We spend the rest of this section establishing Corollary 3.5.8, following and adapting the outline in AGS [1, pp. 291–292].

3.5.1 A Priori Estimates

We include the following standard estimates on the JKO scheme for the sake of completeness, following [1, p. 291].

Since $\mu_{k+1}^{(\tau)}$ minimises $J_\tau(\cdot|\mu_k^{(\tau)})$ in (3.5.1), we find

$$\frac{1}{2\tau}W_2^2\left(\mu_{k+1}^{(\tau)},\mu_k^{(\tau)}\right)+\mathscr{F}\left(\mu_{k+1}^{(\tau)}\right)\leq\mathscr{F}\left(\mu_k^{(\tau)}\right)\quad(k\in\mathbb{N}).\tag{3.5.2}$$

From this follows

$$\mathscr{F}\left(\mu_0^{(\tau)}\right)\geq\mathscr{F}\left(\mu_1^{(\tau)}\right)\geq\cdots,\tag{3.5.3}$$

hence $\mathscr{F}$ is decreasing along the JKO sequence $(\mu_k^{(\tau)})_{k\in\mathbb{N}}$. Each $\overline{\mu}_\tau(t)$ is some $\mu_k^{(\tau)}$, hence there is a uniform control

$$\sup_{t\geq 0}\mathscr{F}(\overline{\mu}_\tau(t))\leq\mathscr{F}(\mu_0)\quad(\tau>0).\tag{3.5.4}$$

Rearranging (3.5.2) yields

$$\frac{1}{2\tau}\sum_{k=0}^{N}W_2^2\left(\mu_{k+1}^{(\tau)},\mu_k^{(\tau)}\right)\leq\mathscr{F}\left(\mu_0^{(\tau)}\right)-\mathscr{F}\left(\mu_{N+1}^{(\tau)}\right)\leq\mathscr{F}(\mu_0)\tag{3.5.5}$$

and then

$$\frac{\tau}{2}\sum_{k\geq 0}\left(\frac{W_2(\mu_{k+1}^{(\tau)},\mu_k^{(\tau)})}{\tau}\right)^2\leq\mathscr{F}(\mu_0).\tag{3.5.6}$$

It follows from this and Cauchy–Schwarz that for $0\leq k\leq l-1$,

$$W_2\left(\mu_k^{(\tau)},\mu_l^{(\tau)}\right)\leq\sum_{j=k}^{l-1}W_2\left(\mu_j^{(\tau)},\mu_{j+1}^{(\tau)}\right)\tag{3.5.7}$$

$$\leq\left(\tau\sum_{j=k}^{l-1}\left(\frac{W_2(\mu_j^{(\tau)},\mu_{j+1}^{(\tau)})}{\tau}\right)^2\right)^{1/2}((l-k)\tau)^{1/2}\tag{3.5.8}$$

$$\leq(2\mathscr{F}(\mu_0))^{1/2}(|k-l|\tau)^{1/2}.\tag{3.5.9}$$

By symmetry, we have

$$W_2\left(\mu_k^{(\tau)},\mu_l^{(\tau)}\right)\leq(2\mathscr{F}(\mu_0))^{1/2}(|k-l|\tau)^{1/2}\quad(k,l\in\mathbb{N}).\tag{3.5.10}$$

3.5.2 Compactness

We now make explicit the compactness assertions outlined in AGS [1, p. 292].

Let $\mathcal{M}_T$ denote the collection of all $\overline{\mu}_\tau(t)$ for $\tau>0$ and $t\in[0,T]$. Each such $\overline{\mu}_\tau(t)$ is μ_0 or $\mu_k^{(\tau)}$ for some $k\tau\leq T$ with $t\in(k\tau,(k+1)\tau]$, hence

$$\operatorname{diam}_{W_2}\mathcal{M}_T=\sup_{\substack{0\leq s,t\leq T\\ \tau>0}}W_2(\overline{\mu}_\tau(s),\overline{\mu}_\tau(t))\leq C_T\tag{3.5.11}$$

for $C_T:=(4T\mathscr{F}(\mu_0))^{1/2}$ by (3.5.10).

Corollary 3.5.1. *$\mathcal{M}_T$ is C_b-precompact in $\mathscr{P}(\mathbb{R})$.*

Proof. $\mathcal{M}_T$ is W_2-bounded by (3.5.11). There is then a uniform control on its second moments as in (3.4.5). It is therefore tight by Remark 2.3.1, hence precompact by Prokhorov. ■

We fix a sequence (τ_k) of time-steps $\tau_k \downarrow 0$ for the remainder of this section. For each $t \in [0, T]$, it follows that the sequence $(\overline{\mu}_{\tau_k}(t))_{k\in\mathbb{N}}$ lies in a fixed compact subset of $\mathscr{P}(\mathbb{R})$ for the narrow topology, and we obtain a limit measure $\mu_t \in \mathscr{P}(\mathbb{R})$ such that

$$\overline{\mu}_{\tau_k}(t) \xrightarrow{C_b} \mu_t \tag{3.5.12}$$

along a subsequence, still denoted (τ_k).

Lemma 3.5.2. *There is an estimate*

$$W_2(\widehat{\mu}_{\tau_k}(t), \overline{\mu}_{\tau_k}(t)) \leq C\tau_k^{1/2} \quad (k \in \mathbb{N}). \tag{3.5.13}$$

Remark 3.5.3. This is stated without proof in AGS [1, p. 292].

Proof. Assume without loss of generality that $t > 0$ and pick $l \in \mathbb{N}$ such that $t \in (l\tau, (l+1)\tau]$. Write $\tau := \tau_k$ for simplicity. Then $\widehat{\mu}_\tau$ is a constant-speed geodesic connecting $\mu_l^{(\tau)}$ to $\mu_{l+1}^{(\tau)}$ on $[l\tau, (l+1)\tau]$, and $\overline{\mu}_\tau(t) = \mu_{l+1}^{(\tau)}$, hence

$$\begin{aligned} W_2(\widehat{\mu}_\tau(t), \overline{\mu}_\tau(t)) &= W_2\Big(\widehat{\mu}_\tau(t), \widehat{\mu}_\tau((l+1)\tau)\Big) && (3.5.14)\\ &= \frac{(l+1)\tau - t}{\tau} W_2\Big(\mu_l^{(\tau)}, \mu_{l+1}^{(\tau)}\Big) && (3.5.15)\\ &\leq W_2\Big(\mu_l^{(\tau)}, \mu_{l+1}^{(\tau)}\Big) && (3.5.16) \end{aligned}$$

since $0 \leq ((l+1)\tau - t)/\tau \leq 1$. By (3.5.10), this is bounded by $(2\mathscr{F}(\mu_0))^{1/2}\tau^{1/2}$. ■

We would now like to conclude that also

$$\widehat{\mu}_{\tau_k}(t) \xrightarrow{C_b} \mu_t \quad \text{in} \quad \mathscr{P}(\mathbb{R}) \tag{3.5.17}$$

as in (3.5.12) for the geodesic interpolations along the same subsequence. This is not so clear from the proof in AGS [1, p. 292], and we suggest the following approach based on the β-metric, which is not mentioned there.

By Garling [14, Theorem 18.5.5], the C_b-topology on $\mathscr{P}(\mathbb{R})$ is metrised by the *bounded Lipschitz metric*

$$\beta(\mu, \nu) := \sup_{\|f\|_{\mathrm{BL}} \leq 1} \left| \int f \, d\mu - \int f \, d\nu \right| \quad (\mu, \nu \in \mathscr{P}(\mathbb{R})), \tag{3.5.18}$$

the supremum ranging over all f in the class $\mathrm{BL}(\mathbb{R})$ of bounded Lipschitz functions $\mathbb{R} \to \mathbb{R}$ with $\|f\|_{\mathrm{BL}} := \|f\|_\infty + \mathrm{Lip}\, f \leq 1$. It is proved in the same reference that the C_b-topology, the β-metric topology, and $\mathrm{BL}(\mathbb{R})$-weak topology on $\mathscr{P}(\mathbb{R})$ all agree. In particular, β metrises narrow convergence.

We prove:

Lemma 3.5.4. *$\beta(\mu, \nu) \leq W_2(\mu, \nu)$ for all $\mu, \nu \in \mathscr{P}_2(\mathbb{R})$.*

Proof. Let $\pi \in \Pi(\mu, \nu)$ be optimal for the W_2-distance and consider $\|f\|_{\mathrm{BL}} \leq 1$. By the marginal formulas,

$$\left| \int f(x)\, d\mu(x) - \int f(y)\, d\nu(y) \right| \leq \int_{\mathbb{R}^2} |f(x) - f(y)|\, d\pi(x, y). \tag{3.5.19}$$

Applying the Schwarz inequality and using $\mathrm{Lip}\, f \leq 1$, this is bounded by

$$\mathrm{Lip}(f) \int |x - y|\, d\pi(x, y) \leq W_2(\mu, \nu). \tag{3.5.20}$$

The estimate follows. ∎

With Lemma 3.5.4 and (3.5.13) at hand, we find

$$\beta\big(\widehat{\mu}_{\tau_k}(t), \overline{\mu}_{\tau_k}(t)\big) \leq C\tau_k^{1/2} \quad (k \in \mathbb{N}) \tag{3.5.21}$$

also for the β-metric. Combined with the convergence (3.5.12), this gives the desired narrow convergence $\widehat{\mu}_{\tau_k}(t) \to \mu_t$ in $\mathscr{P}(\mathbb{R})$ along the same subsequence since β metrises the C_b-topology.

We summarise our findings thus far:

Corollary 3.5.5. *For any sequence $\tau_k \downarrow 0$, the sequence $(\widehat{\mu}_{\tau_k})_{k \in \mathbb{N}}$ is pointwise precompact in $C([0, T], \mathscr{P}(\mathbb{R}))$ for $\mathscr{P}(\mathbb{R})$ in the narrow topology.*

The next step is to seek equicontinuity and appeal to Arzelà–Ascoli to obtain uniform precompactness of the geodesic interpolations on compact intervals. To this end, AGS [1, p. 292] claims the Hölder estimate

$$W_2\big(\widehat{\mu}_\tau(s), \widehat{\mu}_\tau(t)\big) \leq C|s - t|^{1/2} \quad (s, t \in [0, T]) \tag{3.5.22}$$

for $C \geq 0$ independent of $\tau > 0$. This would yield equicontinuity of $(\widehat{\mu}_{\tau_k})_{k \in \mathbb{N}}$ with respect to the W_2-metric directly, hence for the β-metric by Lemma 3.5.4. We were not able to reproduce this estimate and obtain instead:

Lemma 3.5.6. *There is an estimate*

$$W_2\big(\widehat{\mu}_\tau(s), \widehat{\mu}_\tau(t)\big) \leq C\Big(|s - t|^{1/2} + \tau^{1/2}\Big) \quad (s, t \in [0, T]) \tag{3.5.23}$$

for $C \geq 0$ depending only on $\mathscr{F}(\mu_0) \geq 0$.

Proof. Assume first $0 < s \le t \le T$. Pick $k, l \in \mathbb{N}$ such that $s \in (k\tau, (k+1)\tau]$ and $t \in (l\tau, (l+1)\tau]$.

If $k = l$, use that $\widehat{\mu}_\tau$ is a constant-speed geodesic connecting $\mu_k^{(\tau)}$ to $\mu_{k+1}^{(\tau)}$ on $[k\tau, (k+1)\tau]$ to estimate

$$W_2(\widehat{\mu}_\tau(s), \widehat{\mu}_\tau(t)) = \frac{t-s}{\tau} W_2\Big(\mu_k^{(\tau)}, \mu_{k+1}^{(\tau)}\Big) \le W_2\Big(\mu_k^{(\tau)}, \mu_{k+1}^{(\tau)}\Big). \tag{3.5.24}$$

By (3.5.10), this is bounded by $(2\mathscr{F}(\mu_0))^{1/2}\tau^{1/2}$.

If $k < l$, apply the triangle inequality to find

$$W_2(\widehat{\mu}_\tau(s), \widehat{\mu}_\tau(t)) \le W_2\Big(\widehat{\mu}_\tau(s), \mu_{k+1}^{(\tau)}\Big) \tag{3.5.25}$$

$$+ W_2\Big(\mu_{k+1}^{(\tau)}, \mu_l^{(\tau)}\Big) + W_2\Big(\mu_l^{(\tau)}, \widehat{\mu}_\tau(t)\Big). \tag{3.5.26}$$

Using again that $\widehat{\mu}_\tau$ is a constant-speed geodesic on $(k\tau, (k+1)\tau]$ bounds the first term by $(2\mathscr{F}(\mu_0))^{1/2}\tau^{1/2}$ as in the previous case. The third term is bounded similarly. Applying (3.5.10) bounds the middle term:

$$W_2\Big(\mu_{k+1}^{(\tau)}, \mu_l^{(\tau)}\Big) \le \big(2\mathscr{F}(\mu_0)\big)^{1/2}\big(l\tau - (k+1)\tau\big)^{1/2}. \tag{3.5.27}$$

This gives (3.5.23) upon observing $l\tau - (k+1)\tau \le t - s$.

The case $s = 0$, $t \in (0, T]$ is handled by the same techniques. ■

Applying (3.5.23) to the sequence $(\widehat{\mu}_{\tau_k})$ along with Lemma 3.5.4 produces the asymptotic Hölder estimates

$$\beta(\widehat{\mu}_{\tau_k}(s), \widehat{\mu}_{\tau_k}(t)) \le C\Big(|s-t|^{1/2} + \tau_k^{1/2}\Big) \quad (s, t \in [0, T], k \in \mathbb{N}) \tag{3.5.28}$$

in the β-metric. The desired equicontinuity is no longer direct as claimed in AGS, but still observed. We prove:

Lemma 3.5.7. *The sequence $(\widehat{\mu}_{\tau_k})_{k\in\mathbb{N}}$ is equicontinuous in $C([0,T], \mathscr{P}(\mathbb{R}))$ for the uniform β-metric and any sequence of time-steps $\tau_k \downarrow 0$.*

Proof. Let $\epsilon > 0$ and pick $N \in \mathbb{N}$ so large that $0 < \tau_k \le \delta$ for $k \ge N$, where $\delta > 0$ is chosen so that $2C\delta^{1/2} \le \epsilon$ with $C \ge 0$ independent of (τ_k) as in (3.5.28). With $|s - t| \le \delta$, we find

$$\beta(\widehat{\mu}_{\tau_k}(s), \widehat{\mu}_{\tau_k}(t)) \le \epsilon \quad (k \ge N) \tag{3.5.29}$$

by the same estimate, and this takes care of the tail $(\widehat{\mu}_{\tau_k})_{k\ge N}$. The initial segment is taken care of by observing each $\widehat{\mu}_{\tau_k}$ with $k < N$ is uniformly continuous $[0,T] \to \mathscr{P}(\mathbb{R})$ in the β-metric since $[0,T]$ is compact, and this gives equicontinuity also of $\{\widehat{\mu}_{\tau_0}, \dots, \widehat{\mu}_{\tau_{N-1}}\}$. ■

The Arzelà–Ascoli theorem now gives us the main result.

Corollary 3.5.8. *If $\tau_k \downarrow 0$ is any vanishing sequence of time steps and $(\widehat{\mu}_{\tau_k})_{k\in\mathbb{N}}$ is the corresponding sequence where $\widehat{\mu}_{\tau_k} \in C([0,+\infty), \mathscr{P}_2(\mathbb{R}))$ interpolates any JKO sequence $(\mu_n^{(\tau_k)})_{n\in\mathbb{N}}$ produced by (3.5.1) starting at arbitrary $\mu_0 \in \mathscr{P}_2(\mathbb{R})$ by constant-speed geodesics, then there exists for every $T > 0$ a limit curve $\mu_T\colon [0,T] \to \mathscr{P}(\mathbb{R})$, continuous in the narrow topology, such that*

$$\sup_{0\le t\le T} \beta(\widehat{\mu}_{\tau_k}(t), \mu_T(t)) \to 0 \quad \text{as} \quad k \to \infty \tag{3.5.30}$$

along some subsequence.

3.6 Conclusive Remarks

Our Corollary 3.5.8 is a more precise version of a substatement in AGS [1, Theorem 11.1.6], in which the β-metric does not appear, and where it is further claimed that the limit curve $t \mapsto \mu_t$ is absolutely continuous and belongs to the space $\mathrm{AC}^2_{\mathrm{loc}}(\mathbb{R}_+, \mathscr{P}_2(\mathbb{R}))$. Furthermore, it is claimed that it satisfies a continuity equation

$$\partial_t\mu_t + \operatorname{div}(\mu_t v_t) = 0 \tag{3.6.1}$$

with some subdifferential control on the velocity field v_t.

It would be natural to elucidate the relationship between gradient flows of the Brinkman functional $\mathscr{F}$ and distributional solutions of the corresponding continuity equation. This would entail describing the minimal selection of its subdifferential, as is done for the model functionals of internal, potential, and interaction types in [1, § 10.4] under smoothness assumptions which are not applicable here. Similar difficulties are overcome for a singular logarithmic potential in [3, § 3.2].

As remarked in the introduction to Chapter 11 in AGS [1], there are at least four possible approaches to gradient flows which are natural for the Wasserstein spaces. Their connections are not trivial. The limit curves obtained in Corollary 3.5.8 are in the spirit of De Giorgi's minimising movements. It would be desirable to relate these to curves of maximal slope, subdifferential formulations, and the EVI formulation (see [1, pp. 277–280]).

Since the Brinkman functional $\mathscr{F}$ is bounded, LSC, and geodesically convex, [1, Theorem 11.1.4] applies: For any $\mu_0 \in \mathscr{P}_2(\mathbb{R})$, there is at most one EVI gradient flow $\mu\colon \mathbb{R}_+ \to \mathscr{P}_2(\mathbb{R})$ of $\mathscr{F}$ satisfying $\mu_t \to \mu_0$ in $\mathscr{P}_2(\mathbb{R})$ as $t \downarrow 0$, which is characterised by the *evolution variational inequalities* (EVI)

$$\frac{1}{2}\partial_t W_2^2(\mu_t, \nu) \le \mathscr{F}(\nu) - \mathscr{F}(\mu_t) \quad \text{for ae. } t > 0 \tag{3.6.2}$$

for all $\nu \in \mathscr{P}_2(\mathbb{R})$. The uniqueness stems from the fact that if μ, ν are two EVI gradient flows of $\mathscr{F}$ such that $\mu_t \to \mu_0$ and $\nu_t \to \nu_0$ as $t \downarrow 0$, there is

$$W_2(\mu_t, \nu_t) \le W_2(\mu_0, \nu_0) \quad (t \ge 0). \tag{3.6.3}$$

Our convexity estimates for the Brinkman functional relied crucially on the one-dimensionality of the problem, and the proof breaks down in higher dimensions. However, the JKO convergence did not rely on this convexity and goes through also in $\mathbb{R}^n$. We thus obtain existence in any dimension, but uniqueness is more exclusive and is a natural topic of further investigation.

Bibliography

[1] Luigi Ambrosio, Nicola Gigli and Giuseppe Savaré. *Gradient Flows. In Metric Spaces and in the Space of Probability Measures.* Second Edition. Lectures in Mathematics. ETH Zürich. Basel: Birkhäuser, 2008.

[2] Luigi Ambrosio et al. 'A User's Guide to Optimal Transport'. In: *Modelling and Optimisation of Flows on Networks: Cetraro, Italy 2009, Editors: Benedetto Piccoli, Michel Rascle* (2013), pp. 1–155.

[3] José A. Carrillo, Lucas C. F. Ferreira and Juliana C. Precioso. 'A mass-transportation approach to a one dimensional fluid mechanics model with nonlocal velocity'. In: *Advances in Mathematics* 231 (1 2012), pp. 306–327.

[4] José A. Carrillo et al. *Fisher–Rao Gradient Flow: Geodesic Convexity and Functional Inequalities.* July 2024. arXiv: `2407.15693`.

[5] Donald L. Cohn. *Measure Theory.* Reprinted 1993 Edition. Boston, Massachusetts: Birkhäuser, 1980.

[6] John B. Conway. *A Course in Functional Analysis.* Second Edition. Graduate Texts in Mathematics 96. Corrected fourth printing, 1997. New York: Springer, 1990.

[7] Tomasz Dębiec and Markus Schmidtchen. 'Incompressible Limit for a Two-Species Tumour Model with Coupling Through Brinkman's Law in One Dimension'. In: *Acta Applicandae Mathematicae* 169 (Oct. 2020), pp. 593–611.

[8] J. Diestel and J. J. Uhl. *Vector Measures.* Mathematical Surveys 15. Providence, Rhode Island: American Mathematical Society, 1977.

[9] Christian Düll et al. *Spaces of Measures and their Applications to Structured Population Models.* Vol. 36. Cambridge Monographs on Applied and Computational Mathematics. Cambridge: Cambridge University Press, 2021.

[10] Lawrence C. Evans. *Partial Differential Equations.* Second Edition, reprinted with corrections. Vol. 19. Graduate Studies in Mathematics. Providence, Rhode Island: American Mathematical Society, 2015.

[11] Alessio Figalli and Federico Glaudo. *An Invitation to Optimal Transport, Wasserstein Distances, and Gradient Flows.* Second Edition. EMS Textbooks in Mathematics. Berlin, Germany: EMS Press, 2023.

[12] Simon Foldvik. 'Weak Solutions of the Linear Transport Equation for Rank Two Tensor Fields Under Sobolev Regularity'. Master's thesis. University of Oslo, 2024.

[13] Gerald B. Folland. *Real Analysis: Modern Techniques and Their Applications.* Second Edition. Pure and Applied Mathematics. New York: John Wiley & Sons, Inc., 1999.

[14] D. J. H. Garling. *Analysis on Polish Spaces and an Introduction to Optimal Transportation*. Vol. 89. London Mathematical Society Student Texts. Cambridge: Cambridge University Press, 2017.

[15] Tuomas Hytönen et al. *Analysis in Banach Spaces. Volume I: Martingales and Littlewood–Paley Theory*. First Edition. A Series of Modern Surveys in Mathematics 63. Cham, Switzerland: Springer, 2016.

[16] Richard Jordan, David Kinderlehrer and Felix Otto. 'The Variational Formulation of the Fokker–Planck Equation'. In: *SIAM journal on mathematical analysis* 29 (1 1998), pp. 1–17.

[17] L. V. Kantorovich. 'On the Translocation of Masses'. In: *Journal of Mathematical Sciences* 133.4 (2006). English translation of original published in Dokl. Akad. Nauk. USSR 37(7–8), 227–229, 1942, pp. 1381–1382.

[18] Serge Lang. *Fundamentals of Differential Geometry*. First Edition. Graduate Texts in Mathematics 191. Corrected second printing, 2001. New York: Springer, 1999.

[19] Serge Lang. *Real and Functional Analysis*. Third Edition. Graduate Texts in Mathematics 142. New York: Springer, 1993.

[20] John M. Lee. *Introduction to Topological Manifolds*. Second Edition. Graduate Texts in Mathematics 202. New York: Springer, 2011.

[21] Robert J. McCann. 'A Convexity Principle for Interacting Gases'. In: *Advances in mathematics* 128 (1 1997), pp. 153–179.

[22] Gaspard Monge. 'Mémoire sur la théorie des déblais et des remblais'. In: *Histoire de l'Académie Royale des Sciences*. Paris: Imprimerie Royale, 1781, pp. 666–704.

[23] James R. Munkres. *Topology*. Second Edition. Upper Saddle River: Prentice Hall, 2000.

[24] Gerald J. Murphy. *C^*-Algebras and Operator Theory*. Boston: Academic Press, 1990.

[25] Felix Otto. 'The Geometry of Dissipative Evolution Equations: The Porous Medium Equation'. In: *Communications in Partial Differential Equations* 26 (1–2 2001), pp. 101–174.

[26] Gert K. Pedersen. *Analysis Now*. Revised Printing. Graduate Texts in Mathematics 118. New York: Springer, 1995.

[27] Gabriel Peyré and Marco Cuturi. 'Computational Optimal Transport with Applications to Data Sciences'. In: Foundations and Trends in Machine Learning 11 (5–6 2019), pp. 355–607.

[28] Walter Rudin. *Functional Analysis*. Second Edition. International series in pure and applied mathematics. New York: McGraw-Hill, 1991.

[29] Walter Rudin. *Real and Complex Analysis*. Third Edition. New York: McGraw-Hill, 1987.

[30] Filippo Santambrogio. '{Euclidean, metric, and Wasserstein} gradient flows: an overview'. In: *Bulletin of Mathematical Sciences* 7 (Mar. 2017), pp. 87–154.

[31] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians. Calculus of Variations, PDEs, and Modeling*. First Edition. Progress in Nonlinear Differential Equations and Their Applications 87. Cham: Birkhäuser, 2015.

[32] L. A. Steen and J. A. Jr. Seebach. *Counterexamples in Topology*. Reprint of the Second Edition. Dover Books on Mathematics. New York: Dover Publications, 1995.

[33] Elias M. Stein and Rami Shakarchi. *Functional Analysis: Introduction to Further Topics in Analysis*. Vol. 4. Princeton Lectures in Analysis. Princeton, N.J.: Princeton University Press, 2011.

[34] Elias M. Stein and Rami Shakarchi. *Real Analysis: Measure Theory, Integration, & Hilbert Spaces*. Vol. 3. Princeton Lectures in Analysis. Princeton, N.J.: Princeton University Press, 2005.

[35] Cédric Villani. *Optimal Transport. Old and New*. Vol. 338. Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg: Springer, 2009.

[36] Cédric Villani. *Topics in Optimal Transportation*. Vol. 58. Graduate Studies in Mathematics. Providence, Rhode Island: American Mathematical Society, 2003.