TECHNISCHE UNIVERSITÄT WIEN

DIPLOMARBEIT

# On Convergence of Entropy Gradient Flow Structures for Discrete Porous Medium Equations

Ausgeführt am Institut für
*Analysis und Scientific Computing*
der Technischen Universität Wien

unter der Anleitung von
*Univ.Prof. Dr.rer.nat. Ansgar Jüngel*
und
*Ass.-Prof. dr. Jan Maas*
(Institute of Science and Technology Austria)

durch

*DI Dominik Leopold Forkert, BSc*
*Grüntal 4*
*3412 Kierling*

Datum

Unterschrift

Dominik Leopold Forkert

# ON CONVERGENCE OF ENTROPY GRADIENT FLOW STRUCTURES FOR DISCRETE POROUS MEDIUM EQUATIONS

Diploma Thesis
August 30, 2016

**Advisors:**
Ansgar Jüngel
Jan Maas

Institute of Analysis and Scientific Computing
VIENNA UNIVERSITY OF TECHNOLOGY

Institute of Science and Technology
AUSTRIA

# Preface

This thesis investigates gradient flow structures for porous medium equations and their discrete counterparts. The first structure goes back to the seminal paper [36] by Jordan, Kinderlehrer and Otto, where solutions to the Fokker-Plank equation on $\mathbb{R}^n$ were obtained as metric gradient flows for the Shannon-Boltzmann entropy functional on the 2-Wasserstein space over $\mathbb{R}^n$; that is the space of probability measures with finite second moment on $\mathbb{R}^n$, together with a certain distance which is closely related to the weak convergence of probability measures.

In [54] Otto used this deep relation between partial differential equations and gradient flows on spaces of probability measures to identify porous medium equations with Wasserstein gradient flows for the Rényi entropy.

The second gradient flow structure to be presented in this thesis was proposed independently by Maas [47] and Mielke [50], in order to provide a discrete counterpart to Wasserstein gradient flows. The key rôle plays a non local transportation distance between discrete probability measures on a finite state space, induced by a reversible continuous-time Markov chain on the aforementioned space. The corresponding Markov semigroup arises as gradient flow for the relative entropy functional with respect to the stationary distribution of the Markov chain.

Erbar and Maas [30] extended this discrete gradient flow framework to relate discrete versions of porous medium equations to the discrete relative Rényi entropies.

After having established both, Wasserstein gradient flows and their discrete counterpart for finite Markov chains, the central objective of this thesis is to investigate convergence of the latter to the former in a suitable scaling limit. To this aim, we will use a simple finite-volume discretisation for the porous medium equation with drift on the unit interval, which gives rise to corresponding discrete entropy gradient flows.

At this point, we will pursue two different strategies in order to pass to a gradient flow for the Rényi entropy on Wasserstein space over the unit interval: First, we opt to exert a $\Gamma$-convergence result for gradient flows in *EDE (energy dissipation equality)* sense, proposed by Sandier and Serfaty [60] in case of Hilbertian flows and extended to a metric setting by Serfaty [63]. This approach was already successfully used by Disser and Liero in [24] to pass from discrete entropy gradient flows to Wasserstein gradient flows related to the linear Fokker-Plank equation on the unit interval. We follow the same route and show convergence of the gradient flows related to the finite-volume discretisation to a Wasserstein gradient flow EDE sense for the Rényi entropy. In particular, this establishes convergence of the underlying finite-volume scheme via a gradient flow approach.

The final part of this thesis evolves around the passage to a limit of the gradient flow structures when the underlying entropy functionals are convex along geodesics with respect to the involved transportation metrics. The convexity of the Rényi entropy of order $m \geq 1 - 1/n$ on the 2-Wasserstein space over $\mathbb{R}^n$ is a well known result, due to McCann [49]. Curiously, the situation is vastly different for the discrete setting: We present new counterexamples which show that, in general, the discrete Rényi entropy of order $m \leq 1/4$ or $m \geq 7/4$ fails to be convex along geodesics associated to the discrete transportation metrics.

Nevertheless, it is at least possible to obtain a positive result for $m = 1$. In this case, convexity along geodesics for the discrete entropy was established by Erbar and Maas [29] and Mielke [51] for various choices of reversible finite-state Markov chains.

This enables us to investigate the limit of discrete entropy gradient flows in *EVI (evolution variational inequality)* sense, a notion which exhibits strong regularity and uniqueness properties. As backbone for this approach we provide an abstract stability result for EVI gradient flows under $\Gamma$-convergence of the underlying functionals. This is a generalisation of a stability result by Daneri and Savaré [20] to gradient flows on metric spaces which converge in the sense of *Gromov-Hausdorff*.

Gromov-Hausdorff convergence of the discrete transportation metrics on the discrete torus to the 2-Wasserstein distance on the continuous torus was obtained by Gigli and Maas [35]. We will exploit an adaptation of this result and $\Gamma$-convergence of the discrete entropy functionals to show that the corresponding gradient flows, defined on an equidistant discretisation of the unit interval, converge to a Wasserstein gradient flow in EVI sense. In particular, the resulting limit curve will be a distributional solution to the homogeneous heat equation on the unit interval.

## Structure of the Thesis

The first three chapters of this thesis will mainly provide a review of main concepts required in the later parts of this thesis. Detailed references are provided in the bibliographical notes at the end of each chapter.

Chapter 1 gives a brief overview of gradient flows in abstract metric spaces. We introduce three notions of gradient flows, related to the evolution variational inequality and the energy dissipation equality, as well as an inequality variant of the latter. In addition to basic properties, we also provide abstract existence and uniqueness results which are based on a *minimising movements* variational scheme.

In Chapter 2 we give an account of Wasserstein spaces and gradient flows therein. The corresponding distance function on these spaces of probability measures are introduced by means of an optimal transportation problem. After a short discussion of the geometry of Wasserstein spaces, we study Wasserstein gradient flows for two classical types of energy functionals which encompass entropy functionals amongst others and exhibit a tight relation to certain partial differential equations.

In Chapter 3 we present an entropic gradient flow structure for reversible continuous-time Markov chains on a finite state space. We show that porous medium equations arise as gradient flows for the discrete Rényi entropy with respect to suitable non-local transport distances. Finally, we investigate geodesic convexity of these entropy functionals and the Riemannian structure on the interior of the corresponding spaces of discrete probability measures.

Chapter 4 portrays two stability results for gradient flows under notions of $\Gamma$-convergence in the abstract metric framework of the first chapter. The first result is concerned about convergence of gradient flows in EDE sense, the second about stability of EVI gradient flows under Gromov-Hausdorff convergence of the underlying metric spaces. We briefly illustrate how the former was successfully applied to study the convergence of Cahn-Hilliard equations, and outline a simple consequence of the latter for coercive functionals.

In Chapter 5 we investigate how Wasserstein gradient flows for the porous medium equation with drift on the unit interval may be approximated by their discrete counterparts for reversible Markov chains. This approach depicts an application of the stability result for EDE gradient flows from the previous chapter.

Finally, Chapter 6 illustrates convergence to gradient flows for the homogeneous heat equation in the stronger EVI notion. To this aim, we show geodesic convexity of the involved discrete entropy functionals and outline Gromov-Hausdorff convergence of the discrete transportation metrics to the 2-Wasserstein distance.

## Acknowledgements

Foremost, I would like to express my gratitude to Jan Maas. Jan not only proposed the interesting subject of this treatise but also supported me with countless stimulating discussions which pivotally helped to shape this thesis.

I owe Ansgar Jüngel many thanks for taking care of the supervision at the Vienna University of Technology and providing valuable comments.

Finally, I thank Andreas Kreuml and Philipp Kniefacz various remarks and corrections.

# Table of Contents

# Introduction

The central equation of this thesis is the porous medium equation with drift

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho = \Delta(\rho^m) + \operatorname{div}(\rho\nabla v) \qquad \text{for } m > 0, \tag{1}$$

which admits a representation as gradient flow with respect to entropy functionals on a Wasserstein space. In a similar fashion, it is also possible to identify the discrete counterpart to (1) as gradient flow on a space of discrete probability measures, endowed with a suitable non-local transportation metric.
In both formulations the underlying functionals are defined on spaces of probability measures which do not admit any linear structure needed to study gradient flows in the classical sense by means of the equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho = -\nabla\phi(\rho). \tag{2}$$

At first glance, the derivatives in the equation above are not well defined when $\phi$ is a functional on some metric space $(X, d)$. However, metric analysis still provides certain 'moduli of the derivatives' involved in (2), namely the *metric differential*

$$|\dot\rho|(t) := \lim_{s\to t} \frac{d(\rho(s), \rho(t))}{|s-t|}$$

of a curve $\rho : \mathbb{R}^+ \to X$, and the *(local) slope*

$$|\partial\phi|(v) := \limsup_{w\to v} \frac{(\phi(v) - \phi(w))^+}{d(v,w)}$$

of a functional $\phi : X \to \mathbb{R} \cup \{+\infty\}$. Indeed, with these two definitions at hand, (2) makes sense in a metric space as well, provided that we take the modulus in both sides of the equation. However, it is not hard to see that the resulting scalar equation comes with a loss of information and need not be equivalent to (2) in Euclidean space, or more generally, when $X$ is a Hilbert space. In order to retain equivalence to the gradient flow equation in (2) in this case, it is meaningful to look at the derivative of the energy:

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi(\rho(t)) = \langle\dot\rho(t), \nabla\phi(\rho(t))\rangle \geq -|\dot\rho(t)|\,|\nabla\phi(\rho(t))| \geq -\frac{1}{2}|\dot\rho(t)|^2 - \frac{1}{2}|\nabla\phi(\rho(t))|^2. \tag{3}$$

Apparently, there is equality in the first inequality above, exactly, when $\dot\rho(t)$ and $\nabla\phi(\rho(t))$ differ by a negative factor. On the other hand, equality in the second inequality above holds iff $-|\dot\rho(t)| = |\nabla\phi(\rho(t))|$. In other words, there is equality between the left-hand and right-hand side in (3), precisely, when $\dot\rho(t)$ agrees with $-\nabla\phi(\rho(t))$. This means that we may write (2) in the equivalent form

$$\frac{1}{2}|\dot\rho|^2(t) - \frac{1}{2}|\partial\phi(\rho)|^2(t) = -\frac{\mathrm{d}}{\mathrm{d}t}\phi(\rho(t)). \tag{4}$$

This equality is known as (pointwise) *energy dissipation equality* and provides a notion of gradient flows on metric spaces which will be reviewed in Chapter 1. In addition, we will study a different generalisation of (2) to a metric setting which is based on the so called *evolution variational inequality* which is established by exploiting convexity of the underlying functional $\phi$:

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}d^2(\rho(t), y) \leq \phi(y) - \phi(\rho(t)) \qquad \forall y \in X. \tag{5}$$

Above inequality and its natural generalisation to *strongly convex* functionals turn out to provide powerful regularisation and uniqueness properties. In particular, the evolution variational inequality already implies the energy dissipation equality, whilst the converse implication need not be true in general.

Equipped with the abstract tools from the first chapter, we are ready to study gradient flows on the space of probability measures on $\mathbb{R}^n$ in Chapter 2. A suitable metric on this space is provided by the 2-*Wasserstein distance* which metrises the weak convergence of probability measures on bounded metric spaces. This distance is usually defined by means of the *Kantorovich transport problem*. For our purposes, the following characterisation of the 2-Wasserstein distance via the *Benamou-Brenier* formula plays a major rôle:

$$W_2^2(\mu_0, \mu_1) = \inf \left\{ \int_0^1 \|v_t\|^2_{L^2(\mu(t))} \, \mathrm{d}t \right\}, \tag{6}$$

where the infimum is taken over curves of probability measures $\mu(t)$ for $t \in [0,1]$, joining $\mu_0$ to $\mu_1$, and suitable functions $v : [0,1] \to \mathbb{R}^n$ solving the continuity equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu(t) + \operatorname{div}\big(v_t \mu(t)\big) = 0$$

in the sense of distributions. The distance $W_2$ provides a metric for the space of Borel probability measures with finite second moment on $\mathbb{R}^n$. This metric space is called 2-*Wasserstein space* over $\mathbb{R}^n$ and will be denoted by $\mathscr{P}_2(\mathbb{R}^n)$.

Now we are in the position to study gradient flows for certain functionals on $\mathscr{P}_2(\mathbb{R}^n)$. Of particular interest will be the *Rényi entropy functional*

$$F_m(\mu) = \begin{cases} \dfrac{1}{m-1} \displaystyle\int_{\mathbb{R}^n} \rho^m(x) \, \mathrm{d}x, & \text{if } \mu \text{ has density } \rho, \\ +\infty, & \text{otherwise.} \end{cases}$$

It turns out that the density $\rho$ of each curve $\mu : \mathbb{R}^+ \to \mathscr{P}_2(\mathbb{R}^n)$ which satisfies a weak formulation of (4) for the functional $F_m$, is a distributional solution of the homogeneous porous medium equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho = \Delta(\rho^m).$$

This relation is a crucial inspiration for Chapter 3, mainly based on the work [30] by Erbar and Maas: There we will present a gradient flow structure for the space of discrete probability measures on the finite set $\mathcal{X}^n \simeq \{1, 2, \dots n\}$, which gives rise to solutions of the *discrete porous medium equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_i = \sum_{j=1}^n \mathcal{Q}_{ij} \rho_j^m \qquad \text{for } m > 0. \tag{7}$$

Here we assume that the matrix $\mathcal{Q} \in \mathbb{R}^{n \times n}$ is the generator of a reversible time continuous Markov chain on $\mathcal{X}^n$ and $\rho$ is a curve taking values in $\mathcal{P}^n$, that is the space of probability densities with respect to the stationary distribution $\pi$ of $\mathcal{Q}$.

In order to interpolate a discrete density $\rho$ between points $i$ and $j$ in $\mathcal{X}^n$, we introduce the mean

$$\hat{\rho}_{i,j} := \frac{m-1}{m} \frac{\rho_i^m - \rho_j^m}{\rho_i^{m-1} - \rho_j^{m-1}} \tag{8}$$

for $0 < m \le 2$. Inspired by the Benamou-Brenier formula (6), we define a distance function on $\mathcal{P}^n$ by

$$\mathcal{W}^2(\rho_0, \rho_1) := \inf \left\{ \int_0^1 \sum_{i,j} (\psi_j(t) - \psi_i(t))^2 \hat{\rho}_{ij}(t) Q_{ij} \pi_i \, \mathrm{d}t \right\},$$

where the infimum is taken over all smooth curves $\rho : [0,1] \to \mathcal{P}^n$ joining $\rho_0$ to $\rho_1$, and $\psi : [0,1] \to \mathbb{R}^n$ satisfying the discrete continuity equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_i(t) + \sum_{i=1}^{n} (\psi_j(t) - \psi_i(t))\hat{\rho}_{ij}(t)Q_{ij} = 0 \qquad \forall i \in \mathcal{X}^n.$$

The definition of $\mathcal{W}$ somewhat resembles the distance on a Riemannian manifold. Indeed, it turns out that the interior of $\mathcal{P}^n$ can be endowed with a Riemannian structure which induces the distance $\mathcal{W}$. This allows us to study the gradient flow for the *discrete Rényi entropy functional*

$$\mathcal{F}_m^n(\rho) := \begin{cases} \dfrac{1}{m-1}\displaystyle\sum_{i=1}^{n} \rho_i^m \pi_i & \text{if } m \neq 1, \\ \displaystyle\sum_{i=1}^{n} \rho_i \pi_i \log \rho_i & \text{if } m = 1, \end{cases}$$

on the Riemannian manifold $\operatorname{int}\mathcal{P}^n$. Now the particular structure of the mean $\hat{\rho}_{ij}$ in (8) comes into play, which allows us to identify the gradient flow equation for the functional $\mathcal{F}_m^n$ with the corresponding discrete porous medium equation in (7).

At this point, we have established two types of gradient flow structures on distinct metric spaces: the first one for probability measures on Euclidean space in the context of the 2-Wasserstein distance, the latter one for discrete probability densities on the finite set $\mathcal{X}^n$ with respect to the distance $\mathcal{W}$. Now the question arises whether one can relate these two notions by approximation of the former by the latter. To this aim, we present an abstract framework for convergence of gradient flows in general metric spaces in Chapter 4, following results of Serfaty [63] and Daneri and Savaré [20].
The main ingredient is a certain notion of convergence of functionals $\phi^n$ on metric spaces $(X_n, d_n)$ to a limit functional $\phi$ on $(X, d)$: Let $\rho^n : \mathbb{R}^+ \to X_n$ be curves satisfying the energy dissipation equality (4). Moreover, we assume that there is a limit curve in the sense that there exists a metric space $(X, d)$ together with mappings $\iota^n : X_n \to X$ such that $\iota^n \circ \rho^n$ is pointwise convergent to a curve $\rho : \mathbb{R}^+ \to X$. In order to retain $\rho$ as solution of (4), it is enough to check the $\Gamma$-lim inf bounds

$$\liminf_{n\to\infty} \phi^n(\rho^n(t)) \geq \phi(\rho(t)), \tag{9.a}$$

$$\liminf_{n\to\infty} \int_0^t |\dot{\rho}^n|^2(r) + |\partial\phi^n(\rho^n)|^2(r)\,\mathrm{d}r \geq \int_0^t |\dot{\rho}|^2(r) + |\partial\phi(\rho)|^2(r)\,\mathrm{d}r \tag{9.b}$$

for all times $t > 0$, together with suitable initial conditions on all $\rho^n$.
In the second part of this chapter we provide a similar stability result for the evolution variational inequality (5). Here a bound of the form (9.a) is not sufficient and the stronger notion of (sequential) $\Gamma$-convergence of the functionals $\phi^n$ is required. Since the metric appears explicitly in (5), we need to relate the metrics $d_n$ to the metric $d$ by assuming that $\iota^n$ is 'almost an isometry' between $X_n$ and $X$ up to some small error $\varepsilon_n > 0$, viz.

$$\operatorname{dist}(\iota^n(X_n), X) \leq \varepsilon_n \qquad \text{and} \qquad |d(\iota^n(x), \iota^n(y)) - d_n(x, y)| \leq \varepsilon_n \quad \forall x, y \in X_n. \tag{10}$$

Provided that a sequence of such mappings exists for some $(\varepsilon_n \searrow 0)$, we say that the metric spaces $(X_n, d_n)$ are convergent to $(X, d)$ in the sense of *Gromov-Hausdorff*.

In Chapter 5 we are concerned about applying the first stability result of the previous chapter to approximate the 2-Wasserstein gradient flow for the porous medium equation (1) on the unit interval $\Omega := (0, 1)$ by its discrete counterparts as already done for linear Fokker-Plank equations by Disser and Liero [24]. The Markov chains for the corresponding gradient flow structures will be induced by a simple finite-volume scheme for (1). The resulting generator $\mathcal{Q}$ only allows for nearest-neighbour transitions; in other words, $\mathcal{Q}$ is a tridiagonal matrix, which enables us to obtain explicit expressions for the metric differential of a curve in $\mathcal{P}^n$, and the slope of the discrete Rényi entropy functional $\mathcal{F}_m^n$, both with respect to the discrete transportation distance $\mathcal{W}$.

In the next step, we consider piecewise constant interpolants $U^n$ of discrete gradient flow curves in $\mathcal{P}^n$, which (up to a subsequence) are weakly convergent to a probability density curve $U$ on $\Omega$. Now it remains to verify the $\Gamma$-$\liminf$ bounds in (9) to allow for an application of the first stability result in Chapter 4. Thus, we established that $U$ is the density curve of a gradient flow satisfying the energy dissipation equality (4) in $\mathscr{P}_2(\Omega)$. As a consequence of the results obtained in the Wasserstein framework, $U$ is a distributional solution of the porous medium equation (1) with non-flux Neumann boundary condition.

In the last chapter, we adapt our approach from Chapter 5 to the related notion of gradient flows satisfying the evolution variational inequality (5). In this case, the abstract framework of convergence is provided by the second stability result in Chapter 4. That means that we have to assure that the spaces $(\mathcal{P}^n, \mathcal{W})$ converge to $(\mathscr{P}_2(\Omega), W_2)$ in the sense of Gromov-Hausdorff as $(n \to \infty)$. To this aim, we make the simplifying assumption that the underlying discretisation of $\Omega$ is equidistant which implies that the induced generator is of following form

$$\mathcal{Q} = n^2 \begin{pmatrix} -1 & 1 & 0 & 0 & \dots & 0 \\ 1 & -2 & 1 & 0 & \dots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \dots & 0 & 1 & -2 & 1 \\ 0 & \dots & 0 & 0 & 1 & -1 \end{pmatrix}.$$

This particular structure of $\mathcal{Q}$ allows us to adapt an argument by Gigli and Maas [35], thereby obtaining suitable mappings $\iota^n : \mathcal{P}^n \to \mathscr{P}_2(\Omega)$, satisfying (10) by exploiting strong regularisation properties of the heat semigroup on the real line.

Equipped with these 'almost isometric' mappings and $\Gamma$-convergence of the discrete Rényi entropies $\mathcal{F}_m^n$ to $F_m$ as $(n \to \infty)$, it remains to obtain gradient flow curves $\rho : \mathbb{R}^+ \to \mathcal{P}^n$ which satisfy (5) for the functionals $\mathcal{F}_m^n$. This is indeed straightforward, provided that the entropies $\mathcal{F}_m^n$ are convex along geodesics in the Riemannian manifolds $\operatorname{int}\mathcal{P}^n$, i.e.

$$\mathcal{F}_m^n(\gamma(t)) \leq (1-t)\mathcal{F}_m^n(\gamma_0) + t\mathcal{F}_m^n(\gamma_1) \qquad \forall t \in [0,1] \tag{11}$$

for all points $\gamma_0, \gamma_1 \in \mathcal{P}^n$ and geodesics $\gamma : [0,1] \to \mathcal{P}^n$ connecting $\gamma_0$ to $\gamma_1$.

Whilst the verification of (11) is fairly straightforward for the case $m = 1$, it is not clear whether $\mathcal{F}_m^n$ is convex along geodesics for arbitrary $m > 0$. At least for $m \leq 1/4$ or $m \geq 7/4$ the answer is negative as counterexamples for generators of tridiagonal Toeplitz matrix form will show.

Nevertheless, we can apply the abstract convergence result for gradient flows satisfying the evolution variational inequality in the case $m = 1$. Thus, we recover a density curve $U$ in $\mathscr{P}_2(\Omega)$, which belongs to the (unique, provided we fix a starting point) gradient flow, satisfying (5) for the entropy functional $F_1$. Consequently, $U$ is also a distributional solution to the homogeneous heat equation

$$\frac{\mathrm{d}}{\mathrm{d}t} U = \Delta U \qquad \text{in } \Omega,$$

with non-flux Neumann boundary condition.

# 1 Gradient Flows in Metric Spaces

## 1.1 Absolutely Continuous Curves and Their Metric Derivative

In this section we take a glance at absolutely continuous curves, taking values in a metric space, and introduce the closely related concept of a metric derivative. Indeed, with these tools we lay the groundwork for various notions of gradient flows in metric spaces, which we are going to study in the subsequent section.

We will start with the definition of an absolutely continuous curve, taking values in a complete metric space.

**Notation**

In this chapter, by $(a,b)$ we denote a possibly unbounded interval on $\mathbb{R}$ and by $(X,d)$ an arbitrary complete metric space.

**1.1.1 Definition** Given a curve $v:(a,b)\to X$, we say that $v$ belongs to $AC^p((a,b),X)$, $1\le p\le\infty$, if there exists a function $m\in L^p((a,b),\mathbb{R})$ such that

$$d\big(v(s),v(t)\big)\le\int_s^t m(r)\,\mathrm{d}r\qquad\forall s,t\in(a,b):s\le t. \tag{1.1}$$

We say that $v$ belongs to $AC^p_{\mathrm{loc}}((a,b),X)$, $1\le p\le\infty$, if for every $t\in(a,b)$ there exists a neighbourhood $U\subseteq(a,b)$ of $t$ such that $v\big|_U\in AC^p(U,X)$.

In the case $p=1$ we say that $v$ is ***absolutely continuous*** or ***locally absolutely continuous*** and simply write $AC((a,b),X)$ or $AC_{\mathrm{loc}}((a,b),X)$ for the corresponding space, instead of $AC^1((a,b),X)$ or $AC^1_{\mathrm{loc}}((a,b),X)$, respectively.

Although metric spaces lack the linear structure of vector spaces, it is possible to define a certain generalization of a derivative of functions taking values in arbitrary metric spaces.

**1.1.2 Definition** A function $v:(a,b)\to X$ is said to be ***metrically differentiable*** at a point $t\in(a,b)$ if the limit

$$|\dot v|(t):=\lim_{s\to t}\frac{d\big(v(s),v(t)\big)}{|s-t|} \tag{1.2}$$

exists. Then $|\dot v|(t)\in\mathbb{R}$ is called the ***metric differential*** or ***metric derivative*** of $v$ at $t$.

**1.1.3 Example** (Fréchet derivative) Consider a function $v:(a,b)\to Y$ where $(Y,\|\cdot\|_Y)$ is a Banach space. Then $v$ is metrically differentiable at a point $t$ if $v$ is Fréchet differentiable at $t$, since

$$\|\mathrm{d}v(t)\|_Y=\left\|\lim_{s\to t}\frac{v(s)-v(t)}{s-t}\right\|_Y=\lim_{s\to t}\frac{\|(v(s)-v(t)\|_Y}{|s-t|}=|\dot v|\,(t).$$

◀

Concerning the following theorem, recall that the *limit inferior* of a function $\phi:X\to\mathbb{R}\cup\{+\infty\}$ at a point $x\in X$ is defined as $\liminf_{y\to x}\phi(y):=\sup_{U\in\mathfrak{U}(x)}\inf\phi(U)$, where $\mathfrak{U}(x)$ denotes the neighbourhood filter of $x$.

**1.1.4 Theorem** *For any curve $v\in AC^p((a,b),X)$ the metric differential $|\dot v|$ exists a.e. in $(a,b)$ and satisfies the following properties:*

(MD1) *The function $|\dot v|$ belongs to $L^p((a,b),\mathbb{R})$.*

(MD2) *$|\dot v|$ is an admissible integrand for the right-hand side of* (1.1.1).

(MD3) *The metric differential is minimal in the sense that $|\dot v|\le m$ holds a.e. in $(a,b)$, for each function $m\in L^p((a,b),\mathbb{R})$ satisfying* (1.1.1).

**1.1.5 Lemma** (Lipschitz and arc-length reparametrisation) *Let $v \in AC((a,b),X)$ be an absolutely continuous curve with length $L := \int_a^b |\dot{v}|(t)\,\mathrm{d}t$.*

(i) *For every $\varepsilon > 0$ and $L_\varepsilon := L + \varepsilon(b-a)$ there exists a strictly increasing, absolutely continuous map*

$$\varsigma_\varepsilon : (a,b) \to (0, L_\varepsilon) \quad \text{with} \quad \lim_{t \searrow a} \varsigma_\varepsilon(t) = 0 \quad \text{and} \quad \lim_{t \nearrow b} \varsigma_\varepsilon(t) = L_\varepsilon,$$

*and a Lipschitz curve*

$$\nu_\varepsilon : (0, L_\varepsilon) \to X, \quad \text{such that} \quad v = \nu_\varepsilon \quad \text{and} \quad |\dot{\nu}_\varepsilon| \circ \varsigma_\varepsilon = \frac{|\dot{v}|}{\varepsilon + |\dot{v}|} \in L^\infty((a,b),\mathbb{R}). \tag{1.3}$$

*Moreover, the map $\varsigma_\varepsilon$ admits a Lipschitz continuous inverse $\tau_\varepsilon : (0, L_\varepsilon) \to (a,b)$ with a Lipschitz constant $\varepsilon^{-1}$ such that $\nu_\varepsilon = v \circ \tau_\varepsilon$.*

(ii) *There exists an increasing, absolutely continuous map*

$$\varsigma : (a,b) \to [0,L] \quad \text{with} \quad \lim_{t \searrow a} \varsigma(t) = 0 \quad \text{and} \quad \lim_{t \nearrow b} \varsigma(t) = L,$$

*and a Lipschitz curve*

$$\nu : [0,L] \to X, \quad \text{such that} \quad v = \nu \circ \varsigma \quad \text{and} \quad |\dot{\nu}| = 1 \ \text{ a.e. in } [0,L]. \tag{1.4}$$

## 1.2 Different Formulations of Gradient Flows in Metric Spaces

**Notation**

Throughout this section $(M,g)$ denotes denotes a smooth, complete connected Riemannian manifold $M$ of dimension $N \geq 1$, endowed with the metric tensor $g$ and the Riemannian distance $d_g$.

By $(X,d)$ we denote an arbitrary complete metric space and by $(a,b)$ a possibly unbounded open interval in $\mathbb{R}$.

Moreover, we will assume that any extended real functional $\phi : X \to \mathbb{R} \cup \{+\infty\}$ has *proper* effective domain, i.e. the *effective domain* $\operatorname{dom}\phi := \{x \in X : \phi(x) < +\infty\}$ of $\phi$ is nonempty.

Let $\phi$ be a smooth functional on some Riemannian manifold $(M,g)$. Then a gradient flow (with respect to $F$) is just a differentiable curve $\nu : \mathbb{R}^+ \to M$, solving the *gradient flow equation*

$$\dot{\nu}(t) = -(\operatorname{grad}_g \phi)_{\nu(t)} \qquad \forall t \in \mathbb{R}^+ \tag{1.5}$$

and satisfying $\lim_{t \searrow 0} \nu(t) = v_0$ for some initial value $\nu_0 \in M$.

The aim of this section is to generalise this notion to general metric spaces. Clearly, (1.5) makes a-priori no sense for a curve, taking values in a metric space since one lacks the tools to properly define the derivative of such a curve. Nevertheless, we may characterise (1.5), only using the notions of geodesics and convexity together with the metric tools developed in the previous section.

We recall that a functional $\phi \in C^2(M)$ is *convex*, precisely, when $\operatorname{Hess}\phi$ is positive semi-definite, i.e. $\operatorname{Hess}\phi(v,v) \geq 0$ for all $v \in T_pM$ and every point $p \in M$. This naturally generalises to $\kappa$-convexity in the following sense: $\phi$ is called *$\kappa$-convex* if $\operatorname{Hess}\phi - \kappa g$ is positive semi-definite.

The following proposition gives various characterisations of $\kappa$-convexity.

**1.2.1 Proposition** ($\kappa$-convexity) *Let $(M,g)$ be a Riemannian manifold and fix $\kappa \in \mathbb{R}$. Then for every functional $\phi \in C^2(M)$ the following statements are equivalent:*

(i) Hess $\phi - \kappa g$ *is positive semi-definite;*

(ii) *for every constant speed, minimizing geodesic* $\gamma : [0,1] \to M$*, connecting two points* $\gamma_0, \gamma_1 \in M$*, we have*

$$\phi\big(\gamma(t)\big) \le (1-t)\phi(\gamma_0) + t\phi(\gamma_1) - t(1-t)\frac{\kappa}{2}d_g^2(\gamma_0,\gamma_1);$$

(iii) *for every constant speed, minimizing geodesic* $\gamma : [0,1] \to M$*, connecting two points* $\gamma_0, \gamma_1 \in M$*, we have*

$$g\big(\nabla\phi(\gamma_0), \dot\gamma(0)\big) \le \phi(\gamma_1) - \phi(\gamma_0) - \frac{\kappa}{2}d_g^2(\gamma_0,\gamma_1);$$

(iv) *for every constant speed, minimizing geodesic* $\gamma : [0,1] \to M$*, connecting two points* $\gamma_0, \gamma_1 \in M$*, we have*

$$g\big(\nabla\phi(\gamma_0), \dot\gamma(0)\big) \le g\big(\nabla\phi(\gamma_1), \dot\gamma(1)\big) - \frac{\kappa}{2}d_g^2(\gamma_0,\gamma_1).$$

Note that the formulations (iii) and (iv) of Proposition 1.2.1 only require the functional $\phi$ to belong to $C^1(M)$, whereas (ii) requires no smoothness assumption on $\phi$. Therefore, we may use (ii) to define an appropriate notion of $\kappa$-convex functionals on general metric spaces.

**1.2.2 Definition** Let $(X,d)$ be a metric space and fix $\kappa \in \mathbb{R}$. Then a functional $\phi : X \to \mathbb{R} \cup \{+\infty\}$ is called ***κ-convex along a curve*** $\gamma : [0,1] \to X$, connecting two point $\gamma_0, \gamma_1 \in X$ if

$$\phi\big(\gamma(t)\big) \le (1-t)\phi(\gamma_0) + t\phi(\gamma_1) - \frac{\kappa}{2}t(1-t)d^2(\gamma_0,\gamma_1) \qquad \forall t \in [0,1]. \tag{1.6}$$

In particular, $\phi$ is called ***geodesically κ-convex*** if for every pair of points $\gamma_0, \gamma_1 \in \operatorname{dom}\phi$ there exists a geodesic $\gamma : [0,1] \to X$, connecting $\gamma_0$ and $\gamma_1$ such that $\phi$ is $\kappa$-convex along $\gamma$.

Now we introduce the notion of slopes which generalise the modulus of a gradient to a general metric setting.

**1.2.3 Definition** We call

$$|\partial\phi|(x) := \limsup_{y \to x} \frac{(\phi(x) - \phi(y))^+}{d(x,y)} \qquad \forall x \in \operatorname{dom}\phi$$

the ***slope*** of $\phi$.

**1.2.4 Proposition** (Slopes are upper gradients)

(i) *The slope* $|\partial\phi|$ *is a* weak upper gradient *for* $\phi$*, i.e. for every curve* $v \in AC((a,b),X)$ *with the properties*
- ⇝ $t \mapsto |\partial\phi|(v(t))\,|\dot v|(t)$ *belongs to* $L^1((a,b),\mathbb{R})$ *and*
- ⇝ $\phi \circ v$ *is of essential bounded variation on* $(a,b)$*, .i.e. there exists a function* $\varphi : (a,b) \to \mathbb{R} \cup \{+\infty\}$ *of bounded variation such that* $\phi(v(t)) = \varphi(t)$ *a.e. in* $(a,b)$*,*

*one has* $|\varphi'(t)| \le g(v(t))\,|\dot v|(t)$ *a.e. in* $(a,b)$*.*

(ii) *If* $\phi$ *is lower semicontinuous and geodesically* $\kappa$*-convex for some* $\kappa \in \mathbb{R}$*, then the slope* $|\partial\phi|$ *is also a* strong upper gradient *for* $\phi$*, i.e. for every absolutely continuous curve* $v : (a,b) \to X$ *the function* $|\partial\phi| \circ v$ *is Borel measurable and*

$$|\phi(v(t)) - \phi(v(s))| \le \int_s^t |\partial\phi|(v(r))\,|\dot v|(r)\,\mathrm{d}r \qquad \forall s,t \in (a,b) : s \le t. \tag{1.7}$$

*Moreover, the slope admits the representation*

$$|\partial\phi|(x) = \sup_{y \ne x}\left(\frac{\phi(x) - \phi(y)}{d(x,y)} + \frac{\kappa^-}{2}d(x,y)\right)^+ \qquad \forall x \in \operatorname{dom}\phi,$$

*where* $\kappa^- := \min\{\kappa, 0\}$*.*

Now we are ready to some reformulations of (1.5). To this aim, recall that the *upper right-hand Dini derivative* of a function $f : (a, b) \to \mathbb{R}$ is given by

$$\frac{\mathrm{d}^+}{\mathrm{d}t} f(t) := \limsup_{h \searrow 0} \frac{f(t+h) - f(t)}{h} \qquad \forall t \in (a, b), \tag{1.8.a}$$

whereas the *upper left-hand Dini derivative* of $f$ is given by

$$\frac{\mathrm{d}^-}{\mathrm{d}t} f(t) := \limsup_{h \nearrow 0} \frac{f(t+h) - f(t)}{h} \qquad \forall t \in (a, b). \tag{1.8.b}$$

**1.2.5 Proposition** (Gradient flows on Riemannian manifolds) *Let $\phi \in C^1(M)$ be a smooth functional on a Riemannian manifold $(M, g)$ and let $\nu : (a, b) \to M$ is a continuous curve. For every time $t \in (a, b)$ where $\nu$ is differentiable at $t$, the following two statements are equivalent:*

(i) *$\nu$ satisfies the* gradient flow equation *at $t$, i.e*

$$\frac{\mathrm{d}}{\mathrm{d}t} \nu(t) = -(\mathrm{grad}_g \phi)_{\nu(t)};$$

(ii) *$\nu$ satisfies the* energy dissipation equality *at $t$, i.e.*

$$\frac{\mathrm{d}}{\mathrm{d}t} \phi\big(\nu(t)\big) = -\frac{1}{2} |\partial \phi|^2 \big(\nu(t)\big) - \frac{1}{2} |\dot{\nu}(t)|_g^2;$$

*If $\phi$ is a geodesically $\kappa$-convex functional for some $\kappa \in \mathbb{R}$ according to* Definition 1.2.2*, then any of above statements is equivalent to any of:*

(iii) *for every point $y \in M$ and every geodesic $\gamma_t : [0, 1] \to M$, connecting $\nu(t)$ to $y$, we have*

$$\frac{1}{2} \frac{\mathrm{d}^+}{\mathrm{d}t} d_g^2\big(\nu(t), y\big) \leq \frac{\mathrm{d}^+}{\mathrm{d}s} \phi\big(\gamma_t(s)\big)\Big|_{s=0}.$$

(iv) *for every point $y \in M$, $\nu$ satisfies the evolution variational inequality*

$$\frac{1}{2} \frac{\mathrm{d}^+}{\mathrm{d}t} d_g^2\big(\nu(t), y\big) \leq \phi(y) - \phi\big(\nu(t)\big) - \frac{\kappa}{2} d_g^2\big(\nu(t), y\big).$$

Proof We start by proofing the equivalence of (i) and (ii): The chain rule, together with the inequalities of Cauchy-Schwarz and Young imply the estimate

$$-\frac{\mathrm{d}}{\mathrm{d}t} \phi\big(\nu(t)\big) = \Big\langle -(\mathrm{grad}_g \phi)_{\nu(t)}, \dot{\nu}(t) \Big\rangle_g \leq \big|(\mathrm{grad}_g \phi)_{\nu(t)}\big|_g \, |\dot{\nu}(t)|_g \leq \frac{1}{2} \Big( \big|(\mathrm{grad}_g \phi)_{\nu(t)}\big|_g^2 + |\dot{\nu}(t)|_g^2 \Big). \tag{1.9}$$

Note that there is equality in the first inequality of (1.9), precisely, when $\mathrm{grad}_g \phi$ at $\nu(t)$ and $\dot{\nu}(t)$ are linearly dependent by a negative scalar. On the other hand, equality in the second inequality holds iff the norms of both vectors agree. Combining both conditions implies that there is equality in (1.9), exactly, when the gradient flow equation of (i) holds.

Let us check that (i) implies (iii): Since $\gamma_t$ as defined in (iii) is a geodesic with end point $y$, we have $\nabla_{\dot{\gamma}_t} \dot{\gamma}_t = 0$ and $\gamma_t'(1) = 0$ in the first variation formula (B.2) from Appendix B. Therefore, we obtain

$$\frac{1}{2} \frac{\mathrm{d}^+}{\mathrm{d}t} d_g^2\big(\nu(t), y\big) \leq \int_0^1 |\dot{\gamma}_t(s)|_g^2 \, \mathrm{d}s = -\Big\langle \dot{\gamma}_t(0), \gamma_t'(0) \Big\rangle_g = -\big\langle \dot{\gamma}_t(0), \dot{v}(t) \big\rangle_g. \tag{1.10}$$

Moreover, the gradient flow equality in (i) implies

$$\lim_{s\searrow 0}\frac{\phi\big(\gamma_t(s)\big)-\phi\big(\gamma_t(0)\big)}{s}=\nabla_{\dot\gamma_t(0)}\phi=\Big\langle(\mathrm{grad}_g\phi)_{\nu(t)},\dot\gamma_t(0)\Big\rangle_g=\langle-\dot\nu(t),\dot\gamma_t(0)\rangle_g. \tag{1.11}$$

Together, both (1.10) and (1.11) yield

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}d_g^2\big(\nu(t),y\big)\le\frac{\mathrm{d}^+}{\mathrm{d}s}\phi\big(\gamma_t(s)\big)\Big|_{s=0},$$

which is precisely the inequality in (iii).

Next, we show that (iii) implies (iv): Note that $\phi$ is $\kappa$-convex along the geodesic $\gamma_t$ connecting $\nu(t)$ to $y$. Thus, (1.6) corresponds to

$$\frac{\phi\big(\gamma_t(s)\big)-\phi\big(\gamma_t(0)\big)}{s}\le\phi\big(\gamma_t(1)\big)-\phi\big(\gamma_t(0)\big)-\frac{\kappa}{2}(1-t)d_g^2\big(\gamma_t(0),\gamma_t(1)\big)\qquad\forall s\in(0,1].$$

Now passing to the limit above as $(s\searrow 0)$ and using the inequality in (iii) results in the evolution variational inequality

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}d_g^2\big(\nu(t),y\big)\le\frac{\mathrm{d}^+}{\mathrm{d}s}\phi\big(\gamma_t(s)\big)\Big|_{s=0}\le\phi\big(\gamma_t(1)\big)-\phi\big(\gamma_t(0)\big)-\frac{\kappa}{2}d_g^2\big(\gamma_t(0),\gamma_t(1)\big).$$

Finally, it remains to go from (iv) to (i): Fix $w\in T_{\nu(t)}M$ and $y=\exp_{\nu(t)}(\varepsilon w)$. Provided that we choose $\varepsilon>0$ small enough, $\gamma_t(s)=\exp_{\nu(t)}(s\varepsilon w)$ is the unique length minimising geodesic joining $\nu(t)$ to $y$ with constant speed $|\dot\gamma|_g=\varepsilon\,|w|_g$. In particular, there is equality in (1.10) for this choice of $\gamma_t$, to wit

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}d_g^2\big(\nu(t),y\big)=-\langle\dot\gamma_t(0),\dot v(t)\rangle_g=-\langle\varepsilon w,\dot v(t)\rangle_g.$$

Together with the the evolution variational inequality in (iv) we obtain

$$\begin{aligned}-\langle\varepsilon w,\dot v(t)\rangle_g&\le\phi\big(\gamma_t(1)\big)-\phi\big(\gamma_t(0)\big)-\frac{\kappa}{2}d_g^2\big(\gamma_t(0),\gamma_t(1)\big)=\\&=\phi\big(\exp_{\nu(t)}(\varepsilon w)\big)-\phi\big(\nu(t)\big)-\frac{\kappa}{2}\varepsilon^2\,|w|_g^2,\end{aligned}$$

where we used that $\gamma_t$ has constant speed $|\dot\gamma|_g=\varepsilon\,|w|_g$in the equality from the first to the second line. Reordering the terms of this inequality and dividing both sides by $\varepsilon$ yields

$$\frac{\phi\big(\exp_{\nu(t)}(\varepsilon w)\big)-\phi\big(\nu(t)\big)}{\varepsilon}\ge-\langle w,\dot v(t)\rangle_g+\frac{\kappa}{2}\varepsilon\,|w|_g^2,$$

where we may pass the limit as $(\varepsilon\searrow 0)$ to arrive at

$$\Big\langle(\mathrm{grad}_g\phi)_{\nu(t)},w\Big\rangle_g=\nabla_w\phi\ge\langle w,-\dot v(t)\rangle_g. \tag{1.12}$$

Since this inequality holds for all $w\in T_{\nu(t)}M$, there is actually equality in (1.12) and we conclude that the gradient flow equation in (i) is satisfied. ■

Note that we may carry over the gradient flow characterisations (ii) to (iv) in Proposition 1.2.5 to general metric spaces by interpreting the modulus $|\dot\nu(t)|$ of the velocity field $\dot\nu$ as metric derivative $|\dot\nu|(t)$ in the corresponding (in-)equalities. We only need to clarify the meaning of a geodesics in a metric space: Recall that the Hopf-Rinow theorem (cf. Theorem B.3.4 in Appendix B) implies that for any two points in the connected Riemannian manifold $(M,g)$, there exists a length minimizing geodesic $\gamma:[0,1]\to M$, connecting these two points. Furthermore, we can assume without restriction that $\gamma$ is a constant speed curve. This gives rise to the following definition.

**1.2.6 Definition** Let $(X, d)$ be a metric space. A curve $\gamma : [0,1] \to X$ is called ***constant-speed geodesic*** if

$$d\big(\gamma(s), \gamma(t)\big) = |t-s|\, d\big(\gamma(0), \gamma(1)\big) \qquad \forall s, t \in [0,1].$$

We call $(X, d)$ a ***geodesic space*** if for every pair of points $\gamma_0, \gamma_1 \in X$, there exists a constant-speed geodesic $\gamma : [0,1] \to X$, joining $\gamma_0$ to $\gamma_1$.

Clearly, every connected Riemannian manifold, endowed with the Riemannian distance, is a geodesic space by the aforementioned Hopf–Rinow theorem.

Regarding the following definitions, it is sufficient to require a curve $v$ to belong to $AC_{\mathrm{loc}}(\mathbb{R}^+, X)$. Then Theorem 1.1.4 assures that the metric derivative $|\dot{v}|$ exists a.e. in $\mathbb{R}^+$ and is Borel measurable. Moreover, the reverse triangle inequality implies for fixed $y \in X$ that

$$\big|d\big(v(s), y\big) - d\big(v(t), y\big)\big| \le d\big(v(s), v(t)\big) \qquad \forall s, t \in \mathbb{R}^+.$$

As a result, we obtain that the real-valued function $t \mapsto d(v(t), y)$ is locally absolutely continuous and therefore differentiable a.e. in $\mathbb{R}^+$. Note that this holds also true for the mapping $t \mapsto d^2(v(t), y)$. Thus we can establish the following definitions, inspired by the characterisations of a gradient flow on a Riemannian manifold as in Proposition 1.2.5.

**1.2.7 Definition** Assume that $\phi : X \to \mathbb{R} \cup \{+\infty\}$ is a functional with proper effective domain $\operatorname{dom} \phi$.

(EDI) A curve $v \in AC_{\mathrm{loc}}(\mathbb{R}^+, X)$, starting from $\lim\limits_{t \searrow 0} v(t) = v_0 \in \operatorname{dom} \phi$, satisfies the ***energy dissipation inequality (EDI)*** if

$$\frac{1}{2} \int_0^t |\dot{v}|^2(r) \,\mathrm{d}r + \frac{1}{2} \int_0^t |\partial \phi|^2\big(v(r)\big) \,\mathrm{d}r \le \phi(x_0) - \phi\big(v(t)\big) \qquad \forall t \in \mathbb{R}^+, \tag{1.13.a}$$

and

$$\frac{1}{2} \int_s^t |\dot{v}|^2(r) \,\mathrm{d}r + \frac{1}{2} \int_s^t |\partial \phi|^2\big(v(r)\big) \,\mathrm{d}r \le \phi\big(v(s)\big) - \phi\big(v(t)\big) \qquad \text{a.e. } s, t \in \mathbb{R}^+ : s \le t. \tag{1.13.b}$$

We call such a curve ***gradient flow in the EDI sense***.

(EDE) A curve $v \in AC_{\mathrm{loc}}(\mathbb{R}^+, X)$, starting from $\lim\limits_{t \searrow 0} v(t) = v_0 \in \operatorname{dom} \phi$, satisfies the ***energy dissipation equality (EDE)*** if

$$\frac{1}{2} \int_s^t |\dot{v}|^2(r) \,\mathrm{d}r + \frac{1}{2} \int_s^t |\partial \phi|^2\big(v(r)\big) \,\mathrm{d}r = \phi\big(v(s)\big) - \phi\big(v(t)\big) \qquad \forall s, t \in \mathbb{R}^+ : s \le t. \tag{1.14}$$

Such a curve is called ***gradient flow in the EDE sense***.

($\mathrm{EVI}_\kappa$) A curve $v \in AC_{\mathrm{loc}}(\mathbb{R}^+, X)$, starting from $\lim_{t \searrow 0} v(t) = v_0 \in \overline{\operatorname{dom} \phi}$, satisfies the ***evolution variational inequality (EVI)*** with respect to a given $\kappa \in \mathbb{R}$ if

$$\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} d^2\big(v(t), y\big) \le \phi(y) - \phi\big(v(t)\big) - \frac{\kappa}{2} d^2\big(v(t), y\big) \qquad \forall y \in X, \text{ a.e. } t \in \mathbb{R}^+. \tag{1.15}$$

We say such a curve is a ***gradient flow in the EVI sense*** with respect to $\kappa$.

## 1.3 Properties of Gradient Flows in Metric Spaces

**Notation**

As before, our minimal assumption in this section is that the functional $\phi : X \to \mathbb{R} \cup \{+\infty\}$ has proper effective domain, i.e. $\operatorname{dom}\phi$ is nonempty, in a complete metric space $(X, d)$. Moreover, we assume that $\phi$ is lower semicontinuous.

---

Before we will compare the different notions of gradient flows introduced in the previous section, we will show some properties unique to EVI gradient flows. We start with two useful characterisations of EVI, which avoid differentiation and do not assume any absolute continuity of the curve.

**1.3.1 Lemma** (characterisation of EVI) *For $\kappa \in \mathbb{R}$ a continuous curve $v : \mathbb{R}^+ \to \overline{\operatorname{dom}\phi}$ satisfies the EVI, precisely, when*

$$\frac{1}{2}e^{\kappa t}d^2\big(v(t), y\big) - \frac{1}{2}e^{\kappa s}d^2\big(v(s), y\big) \le \big(\phi(y) - \phi(v(t))\big)\int_s^t e^{\kappa r}\,\mathrm{d}r \qquad \forall y \in \operatorname{dom}\phi,\ \forall s, t \in \mathbb{R}^+ : s \le t; \tag{1.16}$$

Above characterisation allows us to obtain a slightly refined version of **(1.15)** which holds for *all* $t > 0$. Indeed, **(1.16)** we can divide both sides of **(1.16)** by $(t - s)$ and pass to the limit superior as $(t \searrow s)$ to obtain the following pointwise variant of the EVI:

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}d^2\big(v(t), y\big) \le \phi(y) - \phi\big(v(t)\big) - \frac{\kappa}{2}d^2\big(v(t), y\big) \qquad \forall y \in \operatorname{dom}\phi,\ \forall t \in \mathbb{R}^+. \tag{1.17}$$

Here $\frac{\mathrm{d}^+}{\mathrm{d}t}$ denotes again the upper right-hand Dini derivative as defined in **(1.8.a)**

Gradient flows in the EVI sense have extensive contraction and regularising properties. The following theorem summarises some of the crucial results.

**1.3.2 Theorem** *Let $v, w :\in AC_{\mathrm{loc}}(\mathbb{R}^+, X)$ two be curves , both of which solve the evolution variational inequality* **(1.15)** *with respect to some $\kappa \in \mathbb{R}$ and assume that $\phi$ is a lower semicontinuous functional with proper effective domain. Then $v$ and $w$ satisfy the following properties:*

(i) $\kappa$-contraction and uniqueness:

$$d\big(v(t), w(t)\big) \le \mathrm{e}^{-\kappa(t-s)}\, d\big(v(s), w(s)\big) \qquad s, t \in \mathbb{R}^+ : s \le t.$$

*In particular, for every initial datum $v_0 \in \overline{\operatorname{dom}\phi}$ there exists at most one gradient flow in the EVI sense.*

(ii) Regularising effects: *The curve $v$ is locally Lipschitz and $v(t)$ belongs to $\operatorname{dom}|\partial\phi| \subseteq \operatorname{dom}\phi$ for all times $t > 0$. For every initial datum $\lim_{t\searrow 0} v(t) = v_0 \in \overline{\operatorname{dom}\phi}$ and all times $t > 0$ the following a priori estimate holds:*

$$\frac{1}{2}e^{\kappa t}d^2\big(v(t), y\big) + \big(\phi\big(v(t) - \phi(y)\big)\big)\int_0^t e^{\kappa r}\,\mathrm{d}r + \frac{1}{2}\,|\partial\phi|^2\big(v(t)\big)\Big(\int_0^t e^{\kappa r}\,\mathrm{d}r\Big)^2 \le \frac{1}{2}d^2(v_0, y) \qquad \forall y \in \operatorname{dom}\phi. \tag{1.18}$$

(iii) Energy identity: *The map $\phi \circ v$ is locally Lipschitz; the right limits*

$$|\dot v|(t+) := \lim_{s\searrow t}\frac{d\big(v(s), v(t)\big)}{s - t} \qquad \text{and} \qquad \frac{\mathrm{d}}{\mathrm{d}t}\phi\big(v(t+)\big) := \lim_{s\searrow t}\frac{\phi\big(v(s)\big) - \phi\big(v(t)\big)}{s - t}$$

*exist for all times $t > 0$ and satisfies the **energy identity***

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi\big(v(t+)\big) = -|\dot{v}|^2(t+) = -|\partial\phi|^2\big(v(t)\big) \qquad \forall t \in \mathbb{R}^+. \tag{1.19}$$

The following result shows that the existence of a gradient flow in the EVI sense implies geodesic $\kappa$-convexity of the underlying functional.

**1.3.3 Proposition** *Let $\phi : X \to \mathbb{R} \cup \{+\infty\}$ be a functional with proper effective domain and fix $\kappa \in \mathbb{R}$. Assume for every initial value $x \in \overline{\mathrm{dom}\,\phi}$ there exists a gradient flow $v_x \in AC_{\mathrm{loc}}(\mathbb{R}^+, X)$ in the EVI sense, starting from $\lim_{t\searrow 0} v_x(t) = x$. Then $\phi$ is $\kappa$-convex along* every *geodesic in $\overline{\mathrm{dom}\,\phi}$.*

Now we turn to some comparison results between the different notions of gradient flows in metric spaces. When comparing (EDI) and (EDE), the following implications are obvious:

**1.3.4 Facts** Let $\phi : X \to \mathbb{R} \cup \{+\infty\}$ be a functional with proper effective domain and $v : \mathbb{R}^+ \to X$ be a locally absolutely continuous curve, starting from $v_0 \in \mathrm{dom}\,\phi$.

(i) If $v$ is a gradient flow in the EDE sense, then the definitions immediately imply that $v$ satisfies (EDI).

(ii) If $\phi$ is lower semicontinuous and $\kappa$-convex for some $\kappa \in \mathbb{R}$, then the inverse implication does also hold: If $v$ is gradient flow in the EDI sense, we may use Proposition 1.2.4.ii and the AM-GM inequality to infer

$$\phi\big(v(t)\big) - \phi\big(v(s)\big) \le \int_s^t |\dot{v}|(r)\,|\partial\phi|\big(v(r)\big)\,\mathrm{d}r \le \frac{1}{2}\int_s^t |\dot{v}|^2(r)\,\mathrm{d}r + \frac{1}{2}\int_s^t |\partial\phi|^2\big(v(r)\big)\,\mathrm{d}r \qquad \text{a.e. } s \le t.$$

Hence, $v$ is also a solution to (EDE).

On the other hand, the following result, namely, that (EVI$_\kappa$) is the strongest of the three notions of gradient flows in a metric setting, is a non-trivial consequence of Proposition ii.

**1.3.5 Proposition** *Let $\phi : X \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous functional with proper effective domain and $v \in AC_{\mathrm{loc}}(\mathbb{R}^+, X)$ be a curve, starting from $x_0 \in X$. If $v$ is a gradient flow in the EVI sense with respect to some $\kappa \in \mathbb{R}$, then $v$ satisfies* (EDI) *and* (EDE)*.*

The following elementary example shows that the implication in Proposition 1.3.5 cannot be reversed.

**1.3.6 Example** Consider the space $\big(\mathbb{R}^2, \|\cdot\|_\infty\big)$ and define a smooth functional on $\mathbb{R}^2$ by $\phi(x_1, x_2) := x_1$. Clearly, $\phi$ is convex and $|\partial\phi| = \|\nabla\phi\|_\infty \equiv 1$.

Next define a family $(v_i)_{i\in\mathbb{R}^+}$ of smooth curves with joint initial datum $(0,0)$ by

$$v_i : [0, +\infty) \to \mathbb{R}^2, \qquad v_i(t) := \left(-t, \frac{t}{1+i}\right) \qquad \forall i \in [0, +\infty),$$

and note that $\phi\big(v_i(t)\big) = -t$ and $|\dot{v}|(t) = \|v'(t)\|_\infty = 1$ for all $i \in \mathbb{R}^+$. Now it is immediate to check that every $v_i$ satisfies (1.14) as well as (1.13). On the other hand, the lack of uniqueness of the flow curve $v_i$ and Proposition 1.3.2 imply that the family $(v_i)_{i\in\mathbb{R}^+}$ does not belong to (EVI$_\kappa$) as $v(t) := (t, 0)$ depicts the unique gradient flow for $\phi$ in the EVI sense.

In particular, Proposition 1.2.5 cannot be applied since the norm $\|\cdot\|_\infty$ does not induce any inner product on $\mathbb{R}^n$. ◀

The difference between (EDI) and (EDE) in a general metric setting is more subtle. We just refer to Example 3.15 in [3]. By making use of the *minimizing movement scheme* which will be introduced in the next section, this example shows that there exists a gradient flow in the EDI sense, which does not satisfy (EDE).

## 1.4 Existence of Gradient Flows in Metric Spaces

In this short section we investigate a discrete approximating scheme which plays a major role in the existence theory of gradient flows in metric spaces.

At first we introduce a uniform partition of $\mathbb{R}^+$:

**Notation**

Denote by $P_\tau := (n\tau)_{n\in\mathbb{N}_0}$ the uniform partition of $\mathbb{R}^+$ into left-open, right-closed intervals $I_\tau^n := \big((n-1)\tau, n\tau\big]$, $n \in \mathbb{N}$ of size $\tau > 0$.

**1.4.1 Definition** Let a lower semicontinuous functional $\phi : X \to \mathbb{R} \cup \{+\infty\}$ with proper effective domain $\operatorname{dom}\phi$ be given, where $(X,d)$ is a Polish metric space. Define the functional

$$\Phi : \mathbb{R}^+ \times \operatorname{dom}\phi \times X \longrightarrow \mathbb{R} \cup \{+\infty\} \tag{1.20.a}$$

$$(\tau, M, x) \longmapsto \frac{1}{2\tau} d^2(x,M) + \phi(x). \tag{1.20.b}$$

For any given *time step* $\tau > 0$ and *discrete initial datum* $M_\tau^0 \in \operatorname{dom}\phi$, a ***τ-discrete minimizing movement*** starting from $M_\tau^0$ is a sequence $(M_\tau^n)_{n\in\mathbb{N}}$ in $\operatorname{dom}\phi$ which satisfies

$$\Phi(\tau, M_\tau^{n-1}, M_\tau^n) \leq \Phi(\tau, M_\tau^{n-1}, x) \qquad \forall x \in X,\ \forall n \in \mathbb{N}.$$

A ***discrete solution*** is the piecewise constant interpolant

$$\overline{M}_\tau(t) := \sum_{n=1}^{\infty} M_\tau^n \mathbb{1}_{I_\tau^n}(t) \qquad \forall t \in \mathbb{R}^+.$$

In general, the existence of a discrete minimizing movement $(M_\tau^n)_{n\in\mathbb{N}}$ cannot be assured without further assumption on the functional $\phi$. However, in case of existence of such a sequence for every $\tau > 0$, one hopes to find a limit curve as $(\tau \searrow 0)$ which satisfies the definition of a gradient flow in some sense.

For instance, a first convergence result could be obtained if one requires all sublevel sets of $\phi$ to be boundedly compact and some regularity to hold:

**1.4.2 Theorem** (Existence of EDI gradient flows) *Let $(X,d)$ be a Polish metric space and $\phi : X \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous functional, bounded from below with proper effective domain* $\operatorname{dom}\phi$*. Assume that $\phi$ and $\Phi$, defined in* (1.20)*, satisfy the following properties:*

(COMP) *The sublevel sets of $\phi$ are* boundedly compact, *i.e. every closed bounded subset of $\{x \in X : \phi(x) \leq c\}$ is compact for any $c \in \mathbb{R}$;*

(REG1) *there exists $\overline{\tau} > 0$ such that $\Phi(\tau, M, \cdot)$ admits a minimum for every choice of $\tau \in (0, \overline{\tau})$ and $M \in \overline{\operatorname{dom}\phi}$;*

(REG2) *the slope $|\partial\phi| : \operatorname{dom}\phi \to \mathbb{R}_0^+ \cup \{+\infty\}$ is lower semicontinuous;*

(REG3) *for every sequence $(x_n)_{n\in\mathbb{N}}$, converging to some $x \in X$, such that $|\partial\phi|(x_n)$ and $\phi(x_n)$ are bounded from above for all $n \in \mathbb{N}$, we have $\lim_{n\to\infty} \phi(x_n) = \phi(x)$.*

*Then the following statements hold:*

(i) *For every discrete initial datum $M_\tau^0 = u_0 \in \overline{\operatorname{dom}\phi}$ and every time step $\tau \in (0, \overline{\tau})$, there exists a discrete minimizing movement $(M_\tau^n)_{n\in\mathbb{N}}$ in* $\operatorname{dom}\phi$.

(ii) *The corresponding discrete solutions $\overline{M}_\tau$ are locally uniformly convergent to a curve $u \in AC_{\mathrm{loc}}(\mathbb{R}^+, X)$ as $(\tau \searrow 0)$.*

(iii) *The limit curve $u$ is a gradient flow in EDI sense, starting from $u_0$.*

We already saw in Fact 1.3.4.ii that geodesic $\kappa$-convexity of the functional $\phi$ implies that any EVI gradient flow additionally satisfies (EDE). Moreover, assumptions (REG1) to (REG3) in Theorem 1.4.2 can be removed in favour of said $\kappa$-convexity.

**1.4.3 Theorem** (Existence of EDI gradient flows) *Let $(X,d)$ be a Polish metric space and $\phi : X \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous functional with proper effective domain* $\mathrm{dom}\,\phi$. *Assume that $\phi$ is $\kappa$-convex for some $\kappa \in \mathbb{R}$ and* (COMP) *is satisfied.*

*Then* (REG1) *to* (REG3)*, as well as the following statements hold:*

(i) *For every discrete initial datum $M^0_\tau = u_0 \in \overline{\mathrm{dom}\,\phi}$ and every time step $\tau \in (0,\overline{\tau})$, there exists a discrete minimizing movement $(M^n_\tau)_{n\in\mathbb{N}}$ in* $\mathrm{dom}\,\phi$.

(ii) *The corresponding discrete solutions $\overline{M}_\tau$ are locally uniformly convergent to a curve $u \in AC_{\mathrm{loc}}(\mathbb{R}^+,X)$ as $(\tau \searrow 0)$.*

(iii) *The limit curve $u$ is a gradient flow in EDE sense, starting from $u_0$.*

For the remainder of this section we are interested in the stronger notion of EVI gradient flows. We have already observed in Section 1.3 (see Proposition 1.3.3) that gradient flows in the sense of ($\mathrm{EVI}_\kappa$) are closely related to the $\kappa$-convexity of the underlying functional $\phi$. However, in a general metric setting the existence a EVI gradient flow does not only depend on $\phi$ but also on the geometrical structure of the metric space $(X,d)$. We cite the following result due to Ambrosio, Gigli, Savaré, which assumes that the functional $\phi$ is $\kappa$-convex along a suitable class of curves in $X$.

**1.4.4 Theorem** (Existence of EVI gradient flows) *Let $(X,d)$ be a Polish metric space and $\phi : X \to \mathbb{R}\cup\{+\infty\}$ be a lower semicontinuous functional with proper effective domain* $\mathrm{dom}\,\phi$. *Assume that $\Phi$, defined in* (1.20)*, satisfies the following property:*

(GCON) *For every triple of points $M, x_0, x_1 \in \mathrm{dom}\,\phi$ there exists a curve $\gamma : [0,1] \to X$ with end-points $\gamma(0) = x_0$, $\gamma(1) = x_1$ such that $\Phi(\tau, M, \cdot)$ is $(\tau^{-1} + \kappa)$-convex along $\gamma$ for every $\tau > 0$ with $\frac{1}{\tau} > -\min\{0,\kappa\}$.*

*Then the following statements hold:*

(i) *For every discrete initial datum $M^0_\tau = u_0 \in \overline{\mathrm{dom}\,\phi}$ and every time step $\tau > 0$ with $1 + \tau\kappa > 0$ there exists a discrete minimizing movement $(M^n_\tau)_{n\in\mathbb{N}}$ in* $\mathrm{dom}\,\phi$.

(ii) *The corresponding discrete solutions $\overline{M}_\tau$ converge locally uniformly to a limit curve $u \in AC_{\mathrm{loc}}(\mathbb{R}^+,X)$ as $(\tau \searrow 0)$.*

(iii) *The limit curve $u$ is the unique gradient flow in EVI sense, starting from $u_0$.*

(iv) *For every $T \geq 0$ there exists a constant $C_{\kappa,T} > 0$ such that the following error estimate holds:*

$$d\big(u(t), \overline{M}_\tau(t)\big) \leq C_{\kappa,T}\, |\partial\phi|(u_0)\tau \qquad \forall t \in [0,T].$$

Finally, we mention a geometrical class of metric spaces in which the property (GCON) seems to be very natural.

**1.4.5 Remark** (Non-positively curved geodesic spaces) We call a metric space $(X,d)$ ***geodesic space*** if for every pair of points $\gamma_0, \gamma_1 \in X$ there exists a constant-speed geodesic $\gamma$ joining $\gamma_0$ to $\gamma_1$.

Then a geodesic space is said to be ***non-positively curved (NPC)*** in the sense of Alexandrov if for every constant speed geodesic $\gamma$ and every point $x \in X$ the following inequality holds:

$$d^2\big(\gamma(t),x\big) \leq (1-t)d^2\big(\gamma(0),x\big) + td^2\big(\gamma(1),x\big) - t(1-t)d^2\big(\gamma(0),\gamma(1)\big) \qquad \forall t \in [0,1]. \tag{1.21}$$

Clearly, above inequality holds precisely when the functional $\frac{1}{2}d^2(\cdot,x)$ is 1-convex along $\gamma$.

In Riemannian geometry, (1.21) may be also characterised by means of the sectional curvature: A connected Riemmanian manifold $(M,g)$ is an NPC space, precisely, when the sectional curvature tensor $K$ is bounded from above by 0 on $M$.

Now assume that a geodesically $\kappa$-convex functional $\phi : X \to \mathbb{R}\cup\{+\infty\}$ on an NPC space $(X,d)$ is given and fix $\tau > 0$. Then for every pair of points $\gamma_0, \gamma_1 \in X$ there exists a geodesic $\gamma : [0,1] \to X$, connecting $\gamma_0$ to $\gamma_1$ such that (1.6) holds. Therefore, we may add the inequalities (1.21) multiplied by $\frac{1}{2\tau}$ and (1.6) up to obtain that the functional $\Phi(\tau, M, \cdot)$ is $(\tau^{-1} + \kappa)$-convex along $\gamma$. Thus, in NPC spaces geodesics seem to be the natural choice of curves required in (GCON).

However, it turns out that Wasserstein spaces over $\mathbb{R}^n$ do not satisfy (1.21). Hence, a different class of curves has to be considered in such spaces.

## 1.5 Bibliographical Notes

The main reference for the content presented in this chapter is the first part in the monograph [4] by Ambrosio, Gigli and Savaré, which gives a thorough account on abstract gradient flows in metric spaces – with the exception of Proposition 1.3.5 which depicts a more recent result by Savaré [62]. For a proof of the characterisation of the evolution variational inequality (EVI) in Lemma 1.3.1 we refer to Clément and Desch [18]. Proposition 1.2.5, describing gradient flows on manifolds, is taken from chapter 23 in Villani's text book [71].

Proofs for most results presented in the first three sections of this chapter may also be found in the author's work [33].

For a pedagogical introduction to this topic in the context of optimal transport we mention Ambrosio and Gigli [3], Daneri and Savaré [20], Ambrosio [6]. In addition, chapters devoted to gradient flows may be found in Villani's and Santambrogio's respective monographs [71] and [61] about optimal transport theory.

The study of the geometry of abstract metric spaces goes back to the work [2] of Alexandrov and is extensively covered by the monograph Burago, Burago and Ivanov [15]. An introduction which focuses on non-positively curved spaces is provided by Papadopoulos [56].

Gradient flows in EDI/EDE sense dates back to a series of papers initiated by [22], [23] by De Giorgi, Degiovanni, Marino and Tosques. This notion is commonly known in literature by the term *curves of maximal slope*. The approach by evolution variational inequalities (EVI) was introduced in the form of integral solutions to evolution equations in Banach spaces by Bénilan [9].

The minimising movements variational scheme was introduced by Di Giorgi [21] to provide a general method of approximating gradient flows. Adapting the fundamental generation result [19] by Crandall and Liggett, Mayer obtained a first existence result for non-positively curved spaces in [48]. In [4] a relaxation of geodesics to a more general class of curves (including generalised geodesics) proved to be crucial to obtain an existence result which is applicable to Wasserstein spaces.

There are several extensions of above results; we mention Sturm [67] for a generation result in metric measure spaces.

# 2 Gradient Flows in Wasserstein Spaces

## 2.1 The Kantorovich Transportation Problem

Our starting point in this chapter is the *Kantorovich transportation problem*.

Recall that for a finite family of measurable spaces $(X_i, \mathcal{A}_i)_{i\in I}$, the *tensor-product $\sigma$-algebra* $\bigotimes_{i\in I}\mathcal{A}_i$ is generated by $\bigcup_{i\in I}(\pi^i)^{-1}(\mathcal{A}_i)$. Then the set $\big(\bigtimes_{i\in I} X_i, \bigotimes_{i\in I}\mathcal{A}_i\big)$ is called the *product measurable space* of the family $(X_i, \mathcal{A}_i)_{i\in I}$.

Since the projection $\pi^i : \bigtimes_{i\in I} X \to X_i$ is a measurable function by definition, we may consider the *pushforward* $\pi^i_\#\sigma := \sigma\circ(\pi^i)^{-1}$ of an arbitrary measure $\sigma$ on $\big(\bigtimes_{i\in I} X_i, \bigotimes_{i\in I}\mathcal{A}_i\big)$, which then induces a measure on $(X_i, \mathcal{A}_i)$.

**Notation**

For a finite family of probability measure spaces $(X_i, \mathcal{A}_i, \mu_i)_{i\in I}$, we denote by $\Pi(\mu_i)_{i\in I}$ the set of all probability measures $\sigma$ on $\big(\bigtimes_{i\in I} X_i, \bigotimes_{i\in I}\mathcal{A}_i\big)$ such that $\pi^i_\#\sigma = \mu_i$ for all $i\in I$.

**2.1.1 Definition** (Kantorovich transport problem) Let $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ be measure spaces and assume that there is given a measurable map $h : X_1\times X_2 \to \mathbb{R}_0^+\cup\{+\infty\}$. The elements of $\Pi(\mu_1,\mu_2)$ are called ***admissible (transport) plans***. We say that $\sigma_{\text{opt}}\in\Pi(\mu_1,\mu_2)$ is an ***optimal (transport) plan*** if $\sigma_{\text{opt}}$ minimises the functional

$$K(\mu_1,\mu_2,\sigma) := \int_{X_1\times X_2} h\,\mathrm{d}\sigma \in \mathbb{R}_0\cup\{+\infty\}\,, \tag{2.1}$$

i.e. $K(\mu_1,\mu_2,\sigma_{\text{opt}}) = \inf_{\sigma\in\Pi(\mu_1,\mu_2)} K(\mu_1,\mu_2,\sigma)$. $\Pi_{\text{opt}}(\mu_1,\mu_2)$ denotes the subset of all optimal transport plans in $\Pi(\mu_1,\mu_2)$. The map $h$ is called ***cost function*** of the Kantorovich problem (2.1).

It is clear that there always exists an admissible plan for the Kantorovich problem since the product measure $\mu_1\times\mu_2$ belongs to $\Pi(\mu_1,\mu_2)$. However, note that $\inf_{\sigma\in\Pi(\mu_1,\mu_2)} K(\mu_1,\mu_2,\sigma)$ need not be finite. Nevertheless, the Kantorovich problem has a solution under rather general assumptions:

**2.1.2 Theorem** *For any two Borel probability measures $\mu_1$ and $\mu_2$ on Polish spaces $(X_1,d_1)$ and $(X_2,d_2)$, and any lower semicontinuous cost function $h : X_1\times X_2\to\mathbb{R}_0\cup\{+\infty\}$ the Kantorovich problem admits an optimal plan.*

Now we bring up an important relationship of the Kantorovich problem and its dual problem.

**2.1.3 Definition** (Kantorovich duality) In the setting of Definition 2.1.1, we formulate the related *dual problem* as follows: Consider the functional

$$J(\mu,\nu,\varphi,\psi) := \int_{X_1}\varphi\,\mathrm{d}\mu + \int_{X_2}\psi\,\mathrm{d}\nu \qquad \forall(\varphi,\psi)\in L^1(X_1,\mu,\mathbb{R})\times L^1(X_2,\nu,\mathbb{R})\,.$$

Then a pair $(\tilde\varphi,\tilde\psi)$ in the set

$$\Phi_h := \big\{(\varphi,\psi)\in L^1(X_1,\mu,\mathbb{R})\times L^1(X_2,\nu,\mathbb{R}) : \varphi(x_1)+\psi(x_2)\le h(x_1,x_2)\big\}$$

is called ***optimal*** if $J(\mu,\nu,\tilde\varphi,\tilde\psi) = \sup_{(\varphi,\psi)\in\Phi_h} J(\mu,\nu,\varphi,\psi)$.

Now the *Kantorovich duality* asserts that under certain conditions the optimal value of the functional $K$ equals the optimal value of $J$. Here we cite the following version.

**2.1.4 Proposition** (Kantorovich duality) *Let $h : X_1\times X_2\to\mathbb{R}_0^+\cup\{+\infty\}$ be a lower semicontinuous cost function. Then*

$$\min_{\sigma\in\Pi(\mu_1,\mu_2)} K(\mu_1,\mu_2,\sigma) = \sup_{(\varphi,\psi)\in\Phi_h} J(\mu,\nu,\varphi,\psi). \tag{2.2}$$

Note that the value on the right-hand side of (2.2) need not be attained in $\Phi_h$ and the value $+\infty$ is not excluded.

The statement also holds for other function classes than $\Phi_h$. For instance, one may consider

$$\Psi_h := \{(\varphi, \psi) \in \mathcal{B}_b(X_1) \times \mathcal{B}_b(X_2) : \varphi(x_1) + \psi(x_2) \le h(x_1, x_2)\},$$

where $\mathcal{B}_b(X_i)$ denotes all bounded Borel measurable real-valued functions on $X_i$, for $i \in \{1, 2\}$. Then one may invoke *monotone convergence* to obtain the following more general result.

**2.1.5 Proposition** *Let $h : X_1 \times X_2 \to \mathbb{R}_0^+ \cup \{+\infty\}$ be a lower semicontinuous cost function. Then*

$$\min_{\sigma \in \Pi(\mu_1,\mu_2)} K(\mu_1, \mu_2, \sigma) = \sup_{(\varphi,\psi)\in\Psi_h} J(\mu, \nu, \varphi, \psi). \tag{2.3}$$

At the end of this section, we mention the *Monge problem*, a transportation problem closely related to the one of Kantorovich.

**2.1.6 Definition** (Monge Problem) In the setting of the Kontorovich problem, consider two measure spaces $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ and let $h : X_1 \times X_2 \to \mathbb{R}_0^+ \cup \{+\infty\}$ be a measurable cost function. We denote by $\mathrm{T}(\mu_1, \mu_2)$ the class of all measurable maps $T : X_1 \to X_2$ with pushforward $T_{\#}\mu_1 = \mu_2$. The elements in $\mathrm{T}(\mu_1, \mu_2)$ are called ***admissible (transport) maps***.

A transport map $T_{\mathrm{opt}} \in \mathrm{T}(\mu_1, \mu_2)$ is called ***optimal*** if it minimises the *Monge problem*

$$M(\mu_1, \mu_2, T) := \int_{X_1} h\big(x, T(x)\big)\, \mathrm{d}\mu_1(x) \in \mathbb{R}_0 \cup \{+\infty\}, \tag{2.4}$$

i.e. $M(\mu_1, \mu_2, T_{\mathrm{opt}}) = \inf_{T\in\mathrm{T}(\mu_1,\mu_2)} M(\mu_1, \mu_2, T)$. The subset of all optimal transport maps in $\mathrm{T}(\mu_1, \mu_2)$ is denoted by $\mathrm{T}_{\mathrm{opt}}(\mu_1, \mu_2)$.

It is clear, that every given transport map $T_{\min} \in \mathrm{T}(\mu_1, \mu_2)$ induces an admissible plan in the set $\Pi(\mu_1, \mu_2)$ of the corresponding Kantorovich problem by means of the pushforward $(\mathrm{Id}, T)_{\#}\mu_1$. However, unlike the Kantorovich problem, the Monge problem may not admit an optimal solution in even very simple settings.

For example, one may consider $X_1 = X_2 = [-1, 1]$ with the *quadratic cost function* $h(x_1, x_2) := |x_1 - x_2|^2$ and measures $\mu_1 := \delta_0$, $\mu_2 := 2^{-1}(\delta_{-1} + \delta_1)$. Then $\Pi(\mu_1, \mu_2)$ consists only of one admissible plan $\sigma = 2^{-1}(\delta_{(0,-1)} + \delta_{(0,1)})$. Hence, the Kantorovich problem admits a unique solution.

On the other hand, there exists no admissible transport map $T \in \mathrm{T}(\mu_1, \mu_2)$ in the corresponding Monge problem since $T$ would be required to take values at $\pm 1$ at the same time (see Figure 2.1).

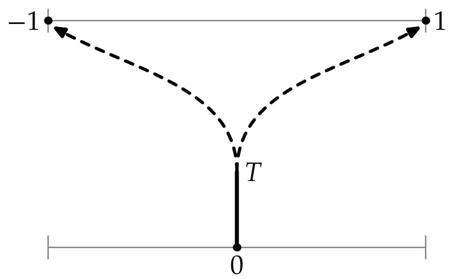

**Figure 2.1** When $\mu_1$ charges single points, a single-valued transport map $T$ may not exist. This means that the Monge problem does not admit any split of mass, whereas mass splitting transport plans are generally admissible in the corresponding Kantorovich problem.

The key argument in the example given above is the fact that the measure $\mu_1$ gives mass to a single point. If one avoids such situations, one can expect results on the existence of an optimal transport map.

To this aim, recall that every convex function $\varphi : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz continuous. In particular, the coordinate functions $\varphi^i$, $i \in \{1, \dots n\}$ are absolutely continuous on every compact interval $[a, b] \subset \mathbb{R}$. Therefore, the gradient $\nabla\varphi$ exists a.e. in $\mathbb{R}^n$. Moreover, we denote by $\varphi^\star(x) := \sup_{y\in\mathbb{R}^n} \{\langle x, y\rangle - \varphi(y)\}$ the *convex conjugate* of $\varphi$.

**2.1.7 Theorem** (Brenier) *Let $\mu_1$ and $\mu_2$ be Borel probability measures with finite second moment on $\mathbb{R}^n$, i.e. $|\cdot|^2$ belongs to $L^1(\mathbb{R}^n, \mu_i, \mathbb{R})$ for $i \in \{1, 2\}$ and $\mu_1(B) = 0$ for every Lebesgue null set $B \in \mathcal{B}(\mathbb{R}^n)$. Consider* *the quadratic cost function $h : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_0^+$, $h(x_1, x_2) := |x_1 - x_2|^2$. Then the following statements hold:*

(i) *There exists a Borel measurable transport map $T \in \mathrm{T}(\mu_1, \mu_2)$, such that $T = \nabla\varphi$ for some convex function $\varphi : \mathbb{R}^n \to \mathbb{R}$. The transport map $T$ is uniquely determined up a $\mu_1$-null set in $\mathbb{R}^n$.*

(ii) *$T = \nabla\varphi$ is the unique optimal transport map of the corresponding Monge problem.*

(iii) *Under the additional assumption that $\mu_2(B) = 0$ for every Lebesgue null set $B \in \mathcal{B}(\mathbb{R}^n)$, there exists a map $\varsigma \in \mathrm{T}(\mu_2, \mu_1)$ with $\varsigma = \nabla\varphi^\star$ such that $\nabla\varphi^\star \circ \nabla\varphi = \mathrm{Id}$ $\mu_1$-a.e. in $\mathbb{R}^n$ and $\nabla\varphi \circ \nabla\varphi^\star = \mathrm{Id}$ $\mu_2$-a.e. in $\mathbb{R}^n$.*

## 2.2 The Structure of Wasserstein Spaces

Consider a Borel probability measure $\mu$ on a Polish space $(X, d)$ and set $d\!\!\!^{-}_y(x) := d(x, y)$ for $x, y \in X$. It is clear that in the case that $d\!\!\!^{-}_{y_0}$ belongs to $L^p(X, \mu, \mathbb{R})$ for some $y_0 \in X$, the triangle inequality

$$d\!\!\!^{-}_y(x) = d(x, y) \leq d(x, y_0) + d(y, y_0) \qquad \forall x, y \in X$$

implies $d\!\!\!^{-}_y \in L^p(\mu, X, \mathbb{R})$ for *all* $y \in X$. This justifies the following notation.

**Notation**

Given a metric space $(X, d)$, $\mathcal{P}_p(X)$ denotes the set of all Borel probability measures $\mu$ such that $x \mapsto d(x, y_0)$ belongs to $L^p(X, \mu, \mathbb{R})$ for some $y_0 \in X$. The set $\mathcal{P}_p(X)$ does not depend on the choice of $y_0 \in X$.

The set $\mathcal{P}_p(X)$ can be equipped with a certain family of metrics. The idea is to consider the Kantorovich problem (2.1) with the metric $d$ or, more generally, $d^p$, $p \geq 1$ as cost function.

**2.2.1 Definition** Let $(X, d)$ be a Polish space. Then for every $p \geq 1$, the function

$$W_p(\mu, \nu) := \inf_{\sigma \in \Pi(\mu,\nu)} \Big( \int_{X\times X} d^p(x, y)\, \mathrm{d}\sigma(x, y) \Big)^{1/p} \qquad \forall \mu, \nu \in \mathcal{P}_p(X) \tag{2.5}$$

is called ***Wasserstein distance*** or sometimes also ***Kantorovich distance*** of order $p$ on $\mathcal{P}_p(X)$. The space $(\mathcal{P}_p(X), W_p)$ is called ***Wasserstein space*** of order $p$ over $X$.

Let us verify that $W_p$ defines a metric on $\mathcal{P}_p(X)$ for all $p \geq 1$.

**2.2.2 Proposition** *Let $(X, d)$ be a Polish space. Then for every $p \geq 1$ the Wasserstein distance $W_p$ defines a metric on $\mathcal{P}_p(X)$.*

The proof of this result is not completely straightforward. Indeed, the verification that $W_p$ satisfies the triangle inequality, is based on the following crucial lemma which assures that certain compatible measures can be "glued together".

**2.2.3 Lemma** (Coupling) *Let $X_1, X_2, X_3$ be Polish spaces and assume, there are two Borel probability measures $\sigma_{1,2}$ and $\sigma_{2,3}$ on the product spaces $X_1 \times X_2$ and $X_2 \times X_3$ with projections ${}^{1,2}\pi^2 : X_1 \times X_2 \to X_2$ and ${}^{2,3}\pi^2 : X_2 \times X_3 \to X_2$ such that*

$${}^{1,2}\pi^2_\#\sigma_{1,2} = {}^{2,3}\pi^2_\#\sigma_{2,3}.$$

*Then on the product space $X_1 \times X_2 \times X_3$ there exists a Borel probability measure $\eta$ with projections $\pi^{1,2} : X_1 \times X_2 \times X_3 \to X_1 \times X_2$ and $\pi^{2,3} : X_1 \times X_2 \times X_3 \to X_2 \times X_3$ such that*

$$\pi^{1,2}_\#\eta = \sigma_{1,2} \qquad \textit{and} \qquad \pi^{2,3}_\#\eta = \sigma_{2,3}.$$

Next we show two important estimates for the Wasserstein distances.

**2.2.4 Proposition** *Let $(X,d)$ be a Polish space. Then for every $p \geq 1$ and all $\mu,\nu \in \mathcal{P}_p(X)$ we have the following inequalities:*

(i) *For every choice of $p \leq q$ the estimate $W_p(\mu,\nu) \leq W_q(\mu,\nu)$ holds.*

(ii) *For every $y_0 \in X$ we have*

$$W_p(\mu,\nu) \leq 2^{1/q}\Big(\int_X d^p(x,y_0)\,\mathrm{d}|\mu-\nu|(x)\Big)^{1/p} \quad \forall p,q \in \mathbb{R}^+ : \frac{1}{p}+\frac{1}{q}=1.$$

Now we investigate a useful characterisation of the Wasserstein distance $W_1$. The following result is closely related to the Kantorovich duality, introduced in Theorem 2.1.3.

**2.2.5 Theorem** (Kantorovich-Rubinstein) *Let $(X,d)$ be a Polish space. Then we have*

$$W_1(\mu,\nu) = \sup\Big\{\int_X f\,\mathrm{d}(\mu-\nu) : f \in Lip_1(X)\Big\} \qquad \forall \mu,\nu \in \mathcal{P}_1(X).$$

Wasserstein spaces inherit to a great extend the topological structure of the underlying metric space:

**2.2.6 Proposition** *Let $(X,d)$ be a Polish space. Then for every $p \geq 1$ the Wasserstein space $(\mathcal{P}_p(X),W_p)$ inherits the following properties from $X$:*

(i) *The space $(\mathcal{P}_p(X),W_p)$ is Polish.*

(ii) *The space $(\mathcal{P}_p(X),W_p)$ is compact if $(X,d)$ is compact.*

Finally, we illustrate the relation between the topology generated by the $p$-Wasserstein distance and the $w^*$-topology on $\mathcal{P}_2(X)$. In fact, the Wasserstein distances almost metrisise the $w^*$-topology in the following way.

**2.2.7 Proposition** *Let $(X,d)$ be a Polish space and fix $p \geq 1$. Then a sequence $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{P}_p(X)$ converges weakly to a Borel probability measure $\mu$ on $X$, precisely, when*

$$\lim_{n\to\infty} W_p(\mu_n,\mu) = 0 \qquad \text{and} \qquad \limsup_{n\to\infty}\int_X d^p(x,x_0)\,\mathrm{d}\mu_n(x) \leq \int_X d^p(x,x_0)\,\mathrm{d}\mu(x) \tag{2.6}$$

*for any $x_0 \in X$.*

In case, the metric space $(X,d)$ is bounded, i.e. $x \mapsto d(x,x_0)$ is a bounded function on $X$, the second condition in (2.6) is always satisfied. In other words, weak convergence of Borel probability measures on a bounded Polish space is metrised by the Wasserstein distances.

## 2.3 Geodesics in the 2-Wasserstein Space

**Notation**

Throughout this section $(M,g)$ denotes denotes a smooth, complete connected Riemannian manifold $M$ of dimension $N \geq 1$, endowed with the metric tensor $g$ and the corresponding Riemannian distance $d_g$.

The following result gives a useful characterisation of geodesics in $(\mathcal{P}_2(M),W_2)$ by means of the exponential mapping on a Riemannian manifold (cf.Section B.3 in Appendix B). Recall that an element of the tangent bundle $TM$ may be thought as pair $(x,v)$, consisting of a point $x \in M$ and a tangent vector $v \in T_xM$.

**2.3.1 Proposition** (Geodesics in the 2-Wasserstein space over a Riemannian manifold) *For every $t \in [0,1]$ define a mapping $\mathrm{Exp}(t) : TM \to M$ via $\mathrm{Exp}(t)(x,v) := \exp_x(tv)$. A curve $\mu : [0,1] \to \mathcal{P}_2(M)$ is a geodesic connecting $\mu_0$ to $\mu_1$ in $(\mathcal{P}_2(M),W_2)$, precisely, when there exists a probability measure $\sigma \in \mathcal{P}(TM)$ such that*

$$\int_{TM} |v|_g^2 \, \mathrm{d}\sigma(x,v) = W_2^2(\mu_0,\mu_1) \qquad \textit{and} \qquad \big(\mathrm{Exp}(t)\big)_\# \sigma = \mu(t) \qquad \forall t \in [0,1].$$

*In particular, $(\mathscr{P}_2(M), W_2)$ is a geodesic space.*

On the contrary, we have the following analogous result for 2-Wasserstein spaces over Hilbert spaces.

**2.3.2 Proposition** (Geodesics in the 2-Wasserstein space over a Hilbert space) *Let $X$ be a (possibly infinite dimensional) Hilbert space. A curve $\mu : [0,1] \to \mathscr{P}_2(X)$ is a geodesic connecting $\mu_0$ to $\mu_1$ in $(\mathscr{P}_2(X), W_2)$, precisely, when there exists an optimal plan $\sigma \in \Pi(\mu_0,\mu_1)$ such that*

$$\mu(t) = \big((1-t)\pi^1 + t\pi^2\big)_\# \sigma \qquad \forall t \in [0,1]. \tag{2.7}$$

*In case, $\sigma$ is induced by a transport map $T$, formula* (2.7) *reduces to*

$$\mu(t) = \big((1-t)\,\mathrm{Id} + tT\big)_\# \mu_0 \qquad \forall t \in [0,1]. \tag{2.8}$$

**2.3.3 Theorem** (Continuity equation) *For every absolutely continuous curve $\mu : [0,1] \to \mathscr{P}_2(M)$, there exists a Borel measurable family $(v_t)_{t\in[0,1]}$ of vector fields $v_t \in \mathcal{T}(M)$ such that $\|v_t\|_{L^2(\mu(t))} \le |\dot\mu|(t)$ for a.e. $t \in [0,1]$ and the* ***continuity equation***

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu(t) + \mathrm{div}\big(v_t\mu(t)\big) = 0 \tag{2.9}$$

*holds in the sense of distributions.*
*Conversely, if $\big(\mu(t), v_t\big)_{t\in[0,1]}$ is satisfying the continuity equation* (2.9) *in the sense of distributions and $\int_0^1 \|v_t\|_{L^2(\mu(t))}\,\mathrm{d}t < +\infty$, then there exists an absolutely continuous curve $\varpi : [0,1] \to \mathscr{P}_2(M)$ such that $\mu$ agrees with $\varpi$ $\mathscr{L}^1$-a.e. on $[0,1]$ and $|\dot\varpi|(t) \le \|v_t\|_{L^2(\mu(t))}$ for a.e. $t \in [0,1]$.*

**2.3.4 Corollary** (Benamou-Brenier formula) *For every pair of points $\mu_0, \mu_1 \in \mathscr{P}_2(M)$, the 2-Wasserstein distance between these two points is given by*

$$W_2(\mu_0,\mu_1) = \min\Big\{\int_0^1 \|v_t\|_{L^2(\mu(t))}\,\mathrm{d}t\Big\} = \min\Big\{\int_0^1 \|v_t\|^2_{L^2(\mu(t))}\,\mathrm{d}t\Big\}^{1/2}, \tag{2.10}$$

*where the minimum is taken over all weakly continuous distributional solutions $(\mu(t), v_t)_{t\in[0,1]}$ of the continuity equation* (2.9) *such that the curve $\mu$ is joining $\mu_0$ to $\mu_1$.*

Recall from Remark 1.4.5 that the characterising inequality in (1.21), together with geodesic $\kappa$-convexity of $\phi$, immediately implies that geodesics satisfy (GCON). This is a key assumption for the existence of gradient flows in EVI sense.

To clarify, to what extend the Wasserstein spaces inherit metric curvature properties like (1.21) from the underlying metric space, we recall the definition of an NPC space from Remark 1.4.5 and introduce its counterpart:

**2.3.5 Definition** A geodesic space $(X,d)$ is called ***positively curved (PC) in the sense of Alexandrov*** if for every constant speed geodesic $\gamma : [0,1] \to X$, connecting $\gamma_0$ and $\gamma_1$, the following inequality holds:

$$d^2\big(\gamma(t),x\big) \ge (1-t)d^2\big(\gamma_0,x\big) + td^2\big(\gamma_1,x\big) - t(1-t)d^2\big(\gamma_0,\gamma_1\big) \qquad \forall t \in [0,1],\ \forall x \in X. \tag{2.11}$$

A geodesic space $(X,d)$ is called ***non positively curved (NPC) in the sense of Alexandrov*** if for every constant speed geodesic $\gamma : [0,1] \to X$, connecting $\gamma_0$ and $\gamma_1$, the reverse inequality of (2.11) holds.

Euclidean space $\mathbb{R}^n$ is flat in the sense that $\mathbb{R}^n$ is both positively and non positively curved.

Let us investigate above definitions in the context of 2-Wasserstein spaces: For instance, it is straightforward to obtain convexity of the mapping $(\mu, \nu) \mapsto W_2^2(\mu, \nu)$, due to the linearity of the Kantorovich transport problem. Indeed, for measures $\mu_0, \mu_1, \nu_0, \nu_1 \in \mathcal{P}_2(X)$, the convex combinations

$$\mu_t = t\mu_0 + (1-t)\mu_1 \qquad \text{and} \qquad \nu_t = t\nu_0 + (1-t)\nu_1 \tag{2.12}$$

also belong to $\mathcal{P}_2(X)$ and the corresponding transport plan

$$\sigma_t = t\sigma_0 + (1-t)\sigma_1$$

with $\sigma_i \in \Pi_{\mathrm{opt}}(\mu_i, \nu_i)$ is admissible with respect to $\mu_t$ and $\nu_t$ for all $t \in [0,1]$. Hence,

$$W_2^2(\mu_t, \nu_t) \leq \int_{X\times X} d^2(x,y)\,\mathrm{d}\sigma_t(x,y) = tW_2^2(\mu_0,\nu_0) + (1-t)W_2^2(\mu_1,\nu_1) \qquad \forall t \in [0,1]. \tag{2.13}$$

However, we already noticed in this section that geodesics in the 2-Wasserstein space over a Hilbert space are given by **(2.7)**, rather than convex combinations like in **(2.12)**. In fact, it turns out that more or less the converse of **(2.13)** is true if $X$ is a PC space and we use geodesics to connect the measures instead.

**2.3.6 Proposition** *Let $(X,d)$ be a geodesic space. If $(X,d)$ is positively curved, then $(\mathcal{P}_2(X), W_2)$ is positively curved as well.*

Somewhat surprisingly, an analogous statement for NPC spaces does not hold true in general as the following example shows.

**2.3.7 Example** The Wasserstein space $(\mathcal{P}^2(\mathbb{R}^2), W_2)$ is not an NPC space.

Proof Define the probability measures

$$\mu_0 := \frac{1}{2}\big(\delta_{(1,1)} + \delta_{(5,3)}\big),$$

$$\mu_1 := \frac{1}{2}\big(\delta_{(-1,1)} + \delta_{(-5,3)}\big),$$

$$\nu := \frac{1}{2}\big(\delta_{(0,0)} + \delta_{(0,-4)}\big).$$

Clearly, $\mu_i$ belongs to $\mathcal{P}^2(\mathbb{R}^2)$ for every $i \in \{1,2,3\}$. Since all admissible plans in each of the sets $\Pi(\mu_0,\mu_1), \Pi(\mu_0,\nu), \Pi(\mu_0,\nu)$ are concentrated on at most four points in $\mathbb{R}^4$, explicit computations of the distances are elementary and one obtains

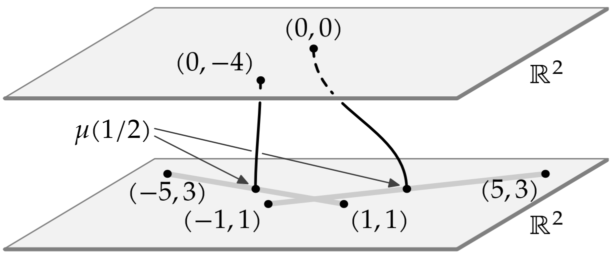


**Figure 2.2** For $t = 1/2$, the links between the atoms of the measures $\mu(t)$ and $\nu$ represent an optimal plan of transportation with respect to quadratic costs $|\cdot|^2$. Note that the depicted optimal plan is not uniquely determined, due to particular symmetry of the transport problem at $t = 1/2$.

$$W_2^2(\mu_0,\mu_1) = 40, \quad W_2^2(\mu_0,\nu) = 30, \quad W_2^2(\mu_1,\nu) = 30.$$

Moreover, one easily shows that

$$\mu(t) := \frac{1}{2}\big(\delta_{(1-6t,1+2t)} + \delta_{(5-6t,3-2t)}\big) \qquad \forall t \in [0,1]$$

depicts a constant-speed geodesic with end-points $\mu(0) = \mu_0$ and $\mu(1) = \mu_1$ (see Figure 2.2). Now

$$W_2^2(\mu(1/2),\nu) = 40 > 20 = \frac{30}{2} + \frac{30}{2} - \frac{40}{4} = \frac{W_2^2(\mu_0,\nu)}{2} + \frac{W_2^2(\mu_1,\nu)}{2} - \frac{W_2^2(\mu_0,\mu_1)}{4}$$

shows that inequality **(1.21)** does not hold. ◀

## 2.4 Potential Energy and Internal Energy Functionals

In this section we will study two classical classes of functionals, defined on the 2-Wasserstein space over $\mathbb{R}^n$.

**2.4.1 Definition** Let $\upsilon : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper, lower semicontinuous, and bounded from below. Then the ***potential energy functional*** $V : \mathscr{P}_2(\mathbb{R}^n) \to \mathbb{R} \cup \{+\infty\}$ associated to $\upsilon$ is defined as

$$V(\mu) := \int_{\mathbb{R}^n} \upsilon \, \mathrm{d}\mu.$$

Let $f : \mathbb{R}_0^+ \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex lower semicontinuous function. Assume that $f$ satisfies $f(0) = 0$ and $\liminf_{z \searrow 0} \frac{f(z)}{z^\alpha}$ belongs to $\mathbb{R} \cup \{+\infty\}$ for some $\alpha > \frac{n}{n+2}$. Set $f'(\infty) := \lim_{z \to \infty} \frac{f(z)}{z}$. Then the ***(lower semicontinuous envelope of the) internal energy functional*** $F : \mathscr{P}_2(\mathbb{R}^n) \to \mathbb{R} \cup \{+\infty\}$ associated to $f$ is defined as

$$F(\mu) := \int_{\mathbb{R}^n} f(\rho(x)) \, \mathrm{d}x + f'(\infty)\mu_{\mathrm{sing}}(\mathbb{R}^n), \tag{2.14}$$

where $\mathrm{d}\mu(x) = \rho(x)\,\mathrm{d}x + \mathrm{d}\mu_{\mathrm{sing}}(x)$ is the Lebesgue decomposition of $\mu$ in an absolutely continuous part with density $\rho$ and a singular part $\mu_{\mathrm{sing}}$ with respect to the Lebesgue measure on $\mathbb{R}^n$.

Let us check that the internal energy functional $F$ is well defined.

**2.4.2 Facts**

(i) The condition $\liminf_{z \searrow 0} \frac{f(z)}{z^\alpha} \in \mathbb{R} \cup \{+\infty\}$ assures that the integral in (2.14) does not attain the value $-\infty$: Indeed, we may assume without loss of generality that $\alpha < 1$ and then this condition implies the lower bound $f(z) \geq az + bz^\alpha$ for some $a, b \in \mathbb{R}$. Hence, to show that $(az + bz^\alpha)\rho$ belongs to $L^1(\mathbb{R}_0^+, \mathbb{R})$, it is enough to invoke Hölder's inequality to find the estimate

$$\int_{\mathbb{R}^n} \rho^\alpha(x)\,\mathrm{d}x = \int_{\mathbb{R}^n} \rho^\alpha(x) \left(\frac{1+|x|}{1+|x|}\right)^{2\alpha} \mathrm{d}x \leq$$
$$\leq \Big(\int_{\mathbb{R}^n} \rho(x)\big(1+|x|\big)^2 \,\mathrm{d}x\Big)^\alpha \Big(\int_{\mathbb{R}^n} \big(1+|x|\big)^{-\frac{2\alpha}{1-\alpha}} \,\mathrm{d}x\Big)^{1-\alpha} < +\infty.$$

(ii) Regarding $f'(\infty)$, we need to verify that the limit $\lim_{z\to\infty} \frac{f(z)}{z}$ exists and actually stays away from $-\infty$. The latter follows directly from the convexity of $f$, while for the former we may use the monotonicity of $f$ on some interval $(c, +\infty)$, in case $\lim_{x\to\infty} f(x) = +\infty$.

The following result shows that the potential energy and the internal energy (or a linear combination of both) are lower semicontinuous functionals.

**2.4.3 Proposition** (Lower semicontinuity of the energy functionals) *Both the potential energy functional $V$ and the internal energy functional $F$ are lower semicontinuous on the 2-Wasserstein space $(\mathscr{P}_2(\mathbb{R}^n), W_2)$.*

Under additional assumptions one can show that both the potential energy and the internal energy functional are geodesically convex on $(\mathscr{P}_2(\mathbb{R}^n), W_2)$. However, this does not facilitate applying the existence theory for EVI gradient flows from Section 1.4: Namely, to apply Theorem 1.4.4, we need to consider an appropriate class of curves in $(\mathscr{P}_2(\mathbb{R}^n), W_2)$ which satisfies (GCON). In Remark 1.4.5 we already noted that geodesics suit our needs in NPC spaces. Unfortunately, Example 2.3.7 shows that $(\mathscr{P}_2(\mathbb{R}^n), W_2)$ does not inherit the non positive curvature from $\mathbb{R}^n$. The remedy for this dilemma consists in considering a broader class of interpolating curves which satisfy (GCON).

**2.4.4 Definition** Let probability measures $\mu, \nu_0, \nu_1 \in \mathcal{P}_2(\mathbb{R}^n)$ be given. Then we call a curve $\nu : [0,1] \to \mathcal{P}_2(\mathbb{R}^n)$, defined by

$$\nu(t) := \big((1-t)\pi^2 + t\pi^3\big)_{\#}\eta \qquad \forall t \in [0,1], \tag{2.15}$$

where $\eta \in \Pi(\mu, \nu_0, \nu_1)$ such that $\pi^{1,2}_{\#}\eta \in \Pi_{\mathrm{opt}}(\mu, \nu_0)$ and $\pi^{1,3}_{\#}\eta \in \Pi_{\mathrm{opt}}(\mu, \nu_1)$, a ***generalised geodesic***, connecting $\nu_0$ to $\nu_1$ with ***base point*** $\mu$. The existence of the joint measure $\eta$ is guaranteed by coupling (Lemma 2.2.3).

In case $\mu = \nu_0$, (2.15) reduces to (2.7) which is the definition of a geodesic in the 2-Wasserstein space over a Hilbert space.

In case, there exist optimal transport maps $T_0 \in \mathrm{T}_{\mathrm{opt}}(\mu, \nu_0)$ and $T_1 \in \mathrm{T}_{\mathrm{opt}}(\mu, \nu_1)$, (2.15) may be also written in a more convenient form which avoids the joint measure $\eta$:

$$\nu(t) = \big((1-t)T_0 + tT_1\big)_{\#}\mu \qquad \forall t \in [0,1].$$

Generalised geodesics possess the following crucial convexity property as an advantage over ordinary geodesics.

**2.4.5 Proposition** *Let a probability measure $\mu \in \mathcal{P}_2(\mathbb{R}^n)$ be given. Then the function $\frac{1}{2}W_2^2(\mu, \cdot)$ is 1-convex along generalised geodesics with base point $\mu$. This means that every generalised geodesic $\nu : [0,1] \to \mathcal{P}_2(\mathbb{R}^n)$, connecting $\nu_0$ to $\nu_1$ with base point $\mu$, satisfies the inequality*

$$W_2^2\big(\nu(t), \mu\big) \le (1-t)W_2^2(\nu_0, \mu) + tW_2^2(\nu_1, \mu) - t(1-t)W_2^2(\nu_0, \nu_1) \qquad \forall t \in [0,1]. \tag{2.16}$$

On the other hand, it turns out that the potential energy and the internal energy functional are convex not only along geodesics in $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ but also along generalised geodesics.

**2.4.6 Proposition** (Convexity of the energy functionals)

(i) *For $\kappa \ge 0$, the potential energy functional $V$ associated to $\upsilon$ is $\kappa$-convex along generalised geodesics in $(\mathcal{P}_2(\mathbb{R}^n), W_2)$, precisely, when $\upsilon$ is $\kappa$-convex.*

(ii) *Assume that the mapping $z \mapsto z^n f(z^{-n})$ is convex and non increasing on $\mathbb{R}^+$. Then the internal energy functional $F$ associated to $f$ is convex along generalised geodesics in $(\mathcal{P}_2(\mathbb{R}^n), W_2)$.*

Note that part (ii) of the result above depends on the dimension of the underlying Euclidean space $\mathbb{R}^n$.

Finally, we are in the position of putting the ideas of Remark 1.4.5 to use, despite $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ not having the NPC property. To this aim, we recognise (2.16), as an analogue of (1.21) and combine it with Proposition 2.4.5; thus, we arrive at the following existence result for EVI gradient flows for the potential energy and the internal energy functional on the 2-Wasserstein space over $\mathbb{R}^n$.

**2.4.7 Theorem** *Consider the mixed energy functional $\phi = F + V$ where $F$ denotes the internal energy functional and $V$ denotes the potential energy functional. Fix $\kappa \ge 0$ such that $\phi$ is $\kappa$-convex along generalised geodesics. Then for every $\mu_0 \in \mathcal{P}_2(\mathbb{R}^n)$ there exists a curve $\mu \in AC_{\mathrm{loc}}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^n))$, starting from $\lim_{t \searrow 0} \mu(t) = \mu_0$, which is the unique gradient flow in the EVI sense for $\phi$ .*

Apparently, this result is still quite abstract. In order to reformulate the theorem above in more tangible terms, we collect some estimates which guarantee, inter alia, that $F$ and $V$ are $\kappa$-convex along generalised geodesics by Proposition 2.4.6.

**2.4.8 Assumptions** We make the following assumptions on the energy functionals defined in Definition 2.4.1 for some $\kappa \ge 0$:

(i) Re potential energy functional $V$ : $\upsilon$ is lower semicontinuous $\kappa$-convex for some $\kappa \ge 0$ such that $\operatorname{dom} f$ has nonempty interior $\operatorname{int}\operatorname{dom} \phi$.

(ii) Re internal energy functional $F$ : $f$ is differentiable such that the function $z \mapsto z^n f(z^{-n})$ is convex as well and non increasing on $\mathbb{R}^+$. In addition, there exists a constant $C > 0$ such that $f$ satisfies the *doubling condition*

$$f(x+y) \leq C\big(f(x) + f(y) + 1\big) \qquad \forall x, y \in \mathbb{R}^n.$$

Apparently, the notion of the metric slope introduced in Definition 1.2.3 does not appear in (EVI$_\kappa$). Nevertheless, the next result is not only useful for studying gradient flows in the EDI sense but plays also a rôle in the identification of the gradient flow in Theorem 2.4.7 as solution of certain partial differential equations.

**2.4.9 Lemma** (Slope of the mixed energy functional) *Consider the mixed energy functional $\phi := F + V$. Let the functions $f$ and $\upsilon$ associated to their respective energy functionals $F$ and $V$ satisfy Assumptions 2.4.8. Let $\mu \in \mathscr{P}_2(\mathbb{R}^n)$ be absolutely continuous with density given by $\mathrm{d}\mu = \rho\, \mathrm{d}x$ such that $\phi(\mu) < +\infty$.*

*Then $|\partial\phi|(\mu) < +\infty$, precisely, when $L_f(\rho) := \rho f'(\rho) - f(\rho)$ belongs to $W^{1,1}_{\mathrm{loc}}(\Omega)$ where $\Omega := \operatorname{int}\operatorname{dom}\phi$, and there exists a function $v \in L^2(\mu)$ such that*

$$v\rho = \nabla\big(L_f(\rho)\big) + \rho\nabla\upsilon \qquad \text{and} \qquad \int_{\mathbb{R}^n} |v|^2 \,\mathrm{d}\mu = |\partial\phi|(\mu). \tag{2.17}$$

Let $\mu \in AC_{\mathrm{loc}}(\mathbb{R}^+, \mathscr{P}_2(\mathbb{R}^n))$ be the unique EVI gradient flow from Theorem 2.4.7. Then the regularising effects for EVI gradient flows obtained in Theorem 1.3.2.ii assure that $\mu(t)$ belongs to $\operatorname{dom}|\partial\phi|$ and $\operatorname{dom}\phi$ for all $t > 0$. In particular, $F(\mu) < +\infty$ and $V(\mu) < +\infty$, which implies that $\mu(t)$ is absolutely continuous with density given by $\mathrm{d}\mu(t) = \rho(t)\,\mathrm{d}x$ for every $t > 0$, due to the definition of the energy functionals $F$ and $V$.

Hence, Lemma 2.4.9 implies the existence of a mapping $v : \mathbb{R}^+ \to L^2(\mu)$ such that $\rho(t)$ and $v(t)$ satisfy the identities in (2.17) at each time $t > 0$. Using a notion of subdifferential calculus in the 2-Wasserstein space over $\mathbb{R}$, it turns out that the pair $(\rho, v)$ satisfies the continuity equation (2.9) in the sense of distributions. In other words, the curve $\rho$ is a distributional solution to the evolution equation $\dot{\rho} = \operatorname{div}(v\rho)$ where $v\rho$ is given by the first equation in (2.17).

We highlight this crucial observation in the following theorem.

**2.4.10 Theorem** *Let the functions $f$ and $\upsilon$ associated to their respective energy functionals $F$ and $V$ satisfy Assumptions 2.4.8. Then for every $\mu_0 \in \mathscr{P}_2(\mathbb{R}^n)$ there exists a curve $\mu \in AC_{\mathrm{loc}}(\mathbb{R}^+, \mathscr{P}_2(\mathbb{R}^n))$, starting from $\lim_{t\searrow 0}\mu(t) = \mu_0$, which is the unique gradient flow in the EVI sense for the mixed energy functional $\phi := F + V$ satisfying the following properties:*

(i) *At each time $t > 0$, $\mu(t)$ is absolutely continuous with density given by $\mathrm{d}\mu(t) = \rho(t)\,\mathrm{d}x$;*

(ii) *the curve $\mu$ is locally Lipschitz and $L_f(\rho(t))$, defined in Lemma 2.4.9, belongs to $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^n)$ for a.e. $t > 0$;*

(iii) *the curve $\rho$ is a distributional solution of the evolution equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho = \operatorname{div}\big(\nabla(L_f(\rho)) + \rho\nabla\upsilon\big) \qquad \text{in } \mathbb{R}^+ \times \mathbb{R}^n, \tag{2.18.a}$$

$$\lim_{t\searrow 0}\rho(t)\,\mathrm{d}x = \mathrm{d}\mu_0 \qquad \text{in } \mathscr{P}_2(\mathbb{R}^n). \tag{2.18.b}$$

At the end of this section we introduce an specific internal energy functional which plays a crucial rôle in the following chapters of this thesis.

**2.4.11 Definition** For $m > 0$ define

$$f_m(x) := \begin{cases} \dfrac{1}{m-1}x^m, & \text{if } m \neq 1, \qquad (2.19.a)\\ x\log x, & \text{if } m = 1. \qquad (2.19.b)\end{cases}$$

Then the internal energy functional associated to $f_m$ is called the ***(the lower semicontinuous envelope of the continuous) Rényi entropy functional*** and is denoted by $F_m$. For $m > 1$ the Rényi entropy takes the form

$$F_m(\mu) = \begin{cases} \dfrac{1}{m-1} \displaystyle\int_{\mathbb{R}^n} \rho^m(x)\,\mathrm{d}x, & \text{if } \mu \ll \mathcal{L}^n,\ \mathrm{d}\mu(x) = \rho(x)\,\mathrm{d}x, \qquad (2.20.a) \\ +\infty, & \text{otherwise.} \qquad (2.20.b) \end{cases}$$

For $m = 1$, the functional is also called ***(continuous) Shannon entropy functional*** and is usually written in the form

$$F_1(\mu) = \begin{cases} \displaystyle\int_{\mathbb{R}^n} \rho \log \rho(x)\,\mathrm{d}x, & \text{if } \mu \ll \mathcal{L}^n,\ \mathrm{d}\mu = \rho(x)\,\mathrm{d}x, \qquad (2.21.a) \\ +\infty, & \text{otherwise.} \qquad (2.21.b) \end{cases}$$

For $m \geq 1$ the Rényi entropy functional $F_m$ has superlinear growth at infinity, i.e. $f'_m(\infty) = +\infty$. Therefore, $F_m$ is possibly finite only at probability measures which are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$. This establishes equivalence between $F_m$ and the internal energy functional associated to $f_m$ as given in (2.14).

2.4.12 **Remark** Apparently, the Shannon entropy functional $F_1$ cannot be obtained by simply passing to the limit $\lim_{m\to 1} F_m$. However, it is still possible to recover $F_1$ from $F_m$ in the following way: Let us assume for simplicity that $\rho$ is a probability density with respect to the Lebesgue measure on $\mathbb{R}^n$, which is essentially bounded and has compact support. Then $F(\mu)$ takes finite values for all $m > 0$ and L'Hôpital's rule implies

$$\lim_{m\to 1} \left( \frac{1}{1-m} \log \int_{\mathbb{R}^n} \rho^m(x)\,\mathrm{d}x \right) = \lim_{m\to 1} \left( -\int_{\mathbb{R}^n} \rho^m(x)\,\mathrm{d}x \right)^{-1} \int_{\mathbb{R}^n} \rho^m(x) \log \rho(x)\,\mathrm{d}x =$$
$$= -\int_{\mathbb{R}^n} \rho(x) \log \rho(x)\,\mathrm{d}x = -F_1(\rho).$$

The expression in the parenthesis on the left-hand side of this equation corresponds to a definition of the Rényi entropy usually encountered in *information theory*. In a similar fashion, the continuous Shannon entropy is usually defined with an negative sign in front of the functional $F_1$.

Another way of obtaining the Shannon entropy $F_1$ is to consider the limit

$$\lim_{m\to 1} \left( F_m - \frac{1}{m-1} \right) = \lim_{m\to 1} \int_{\mathbb{R}^n} \frac{\rho^m(x) - \rho(x)}{m-1}\,\mathrm{d}x = \int_{\mathbb{R}^n} \rho(x) \log \rho(x)\,\mathrm{d}x = F_1(\rho). \qquad (2.22)$$

Since the the correction term on the left-hand side of this equation is just a constant for fixed $m \neq 1$, the term does not contribute to any of the considered gradient flow notions.

In the following proposition we apply the results, developed so far in this section, to the Rényi entropy as defined in (2.20). It is straightforward to check whether the functional $F_m$ on $\mathcal{P}_2(\mathbb{R}^n)$ satisfies the Assumptions 2.4.8.ii for internal energy functionals. Indeed, let us assume for the moment that $m \neq 1$ . Then the mapping

$$z \mapsto z^n f_m(z^{-n}) = \frac{1}{m-1} z^{n(1-m)}$$

is convex and non increasing on $\mathbb{R}^+$, precisely, when $m \geq 1 - \frac{1}{n}$. Moreover, the doubling condition follows easily from convexity and $m$-homogeneity of $f_m$. In case $m = 1$, these properties hold, due to the relation between $F_1$ and $F_m$ in (2.22).

With these considerations in mind, we are ready to identify EVI gradient flows for $F_m$ as solutions to the homogeneous porous medium equation in $\mathbb{R}^n$. However, we still may take a potential energy functional into account. Of particular interest is an *infinite potential well* in $\mathbb{R}^n$, that is the potential $\upsilon$ takes the form of the characteristic function $\chi_{\overline{\Omega}}$ of a convex domain $\Omega \subseteq \mathbb{R}^n$ in the sense of convex analysis, i.e.

$$\chi_{\overline{\Omega}}(x) := \begin{cases} 0 & \text{if } x \in \overline{\Omega}, \\ +\infty & \text{otherwise.} \end{cases}$$

Supposed that $\mu : \mathbb{R}^+ \to \mathcal{P}_2(\mathbb{R}^n)$ is the gradient flow in EVI sense for such an infinite potential well functional $V$, then $\mu$ is necessarily concentrated on $\overline{\Omega}$ since $\mu(t) \in \operatorname{dom} V$ for all times $t > 0$. As a result, we obtain that EVI gradient flows for $F_m + V$ are solutions to the homogeneous porous medium equation in $\Omega$. In this setting, homogeneous Neumann boundary conditions appear naturally. Indeed, the fact that the solution at each time is a probability measure on $\Omega$ results in no flux at the boundary.

Above considerations lead directly to the following corollary of Theorem 2.4.10.

**2.4.13 Corollary** (Gradient flow associated to the continuous porous medium equation)

*Let $\Omega$ be convex domain in $\subseteq \mathbb{R}^n$. Let $\upsilon = \upsilon$ associated to the potential energy functional $F$ satisfy* Assumptions 2.4.8.i *and assume that $\upsilon(x) = +\infty$ for all $x \in \mathbb{R}^n \setminus \overline{\Omega}$. Then for every $\mu_0 \in \mathcal{P}_2(\overline{\Omega})$ there exists a curve $\mu \in AC_{\text{loc}}(\mathbb{R}^+, \mathcal{P}_2(\mathbb{R}^n))$, starting from $\lim_{t\searrow 0} \mu(t) = \mu_0$, which is the unique gradient flow in the EVI sense for the mixed energy functional $\phi := F_m + V$ satisfying the following properties:*

(i) *At each time $t > 0$, $\mu(t)$ is absolutely continuous with density given by $\mathrm{d}\mu(t) = \rho(t)\,\mathrm{d}x$ such that* $\operatorname{supp} \rho(t) \subseteq \overline{\Omega}$;

(ii) *the curve $\mu$ is locally Lipschitz and $\rho^m(t)$ belongs to $W^{1,1}_{\text{loc}}(\Omega)$ for a.e. $t > 0$;*

(iii) *the curve $\rho$ is a distributional solution of the following porous medium equation with drift and non-flux Neumann boundary condition:*

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\rho &= \Delta(\rho^m) + \operatorname{div}(\rho \nabla \upsilon) && \text{in } \mathbb{R}^+ \times \Omega, \\ \lim_{t\searrow 0} \rho(t)\,\mathrm{d}x &= \mathrm{d}\mu_0 && \text{in } \mathcal{P}_2(\mathbb{R}^n). \end{aligned}$$

## 2.5 Bibliographical Notes

This chapter is mainly based on the second part of the monographs [4] by Ambrosio, Gigli and Savaré and [71] by Villani. The former considers Wasserstein spaces over $\mathbb{R}^n$, whereas the latter is concerned about Wasserstein spaces over smooth manifolds as well. Proposition 2.3.1 is a slightly more recent result which appeared first in [3] by Ambrosio and Gigli. Elementary proofs for the Kantorovich duality (Proposition 2.1.4) and the Kantorovich-Rubinstein theorem (Theorem 2.2.5) were obtained by Edwards in [26] and [27], respectively.

Most of those proofs are also presented in the author's work [33].

Concise treatments of optimal transport are provided by the lecture notes [3] as well as the survey article [13] by Bogachev and Kolesnikov. Santambrogio's recent text book [61] outlines several applications in population dynamics, economics, image processing processing, amongst others. In addition we mention Villani's first book [70] on this topic and [59] by Rachev and Rüschendorf.

The definition of Wasserstein spaces in based on the problem of optimal transportation introduced by Kantorovich in the seminal papers [40], [39]. The closely related transport problem of Monge [52] is considerably older.

The term *Wasserstein distance* goes back to an influential paper [25] by Dobrushin who had wrongly accredited the discovery of this metric to Vasershtein [68]. Henceforth, the term *Kantorovich distance* which is used mainly in Russian literature, seems to be historically more accurate.

The dynamic characterisation of the 2-Wasserstein distance via Lemma 2.3.4 is due to Benamou and Brenier [8].

The investigation of geodesics in Wasserstein spaces over $\mathbb{R}^n$ and gedodesic convexity of the functionals in Section 2.4 thereon goes back to McCann [49], who also coined the respective terms *displacement interpolation* and *displacement convexity*.

The study of gradient flows in Wasserstein spaces started with the seminal paper [36] by Jordan, Kinderlehrer and Otto, where gradient flows for the entropy on the 2-Wasserstein space were identified as solutons of Fokker-Plank equations. Later Otto extended this result in [54] to Wasserstein gradient flows for the Rényi entropy which turn out to be solutions of the porous medium equation.

In the celebrated works [45], [46] of Lott and Villani on the one side, and [65], [66] of Sturm on the other side, geodesic convexity of entropy functionals on Wasserstein spaces plays a crucial rôle in the study of a synthetic notion of Ricci curvature for metric measure spaces.

# 3 Entropy Gradient Flows for Continuous-time Markov Chains

## 3.1 A Metric Structure Induced by Markov Chains

In this section we introduce a discrete counterpart to Wasserstein spaces. Instead of a metric space we consider a finite discrete set $\mathcal{X}^n$ with $n$ distinct elements, together with a Markov chain which gives rise to a Wasserstein-like distance function on the class of probability measures on $\mathcal{X}^n$.

We start with some basic notation.

**Notation**

By $\mathcal{X}^n$ we denote a finite set of cardinality $n \in \mathbb{N}$. For simplicity, we will identify $\mathcal{X}^n$ with the well-ordered set $\{1, 2, ...n\}$.

The matrix $\mathcal{Q} \in \mathbb{R}^{n\times n}$ denotes the infinitesimal generator of an *continuous-time Markov chain* on the state space $\mathcal{X}^n$. $\pi$ is the associated *stationary distribution* on $\mathcal{X}^n$, determined by the equation $\pi\mathcal{Q} = 0$. We assume that the Markov chain is *irreducible* and *reversible*, i.e. the stationary distribution satisfies the *detailed balance condition* $\pi_i\mathcal{Q}_{ij} = \pi_j\mathcal{Q}_{ji}$ for all states $i, j \in \mathcal{X}^n$ and it is possible to get from any state to any state, respectively. In this thesis an *irreducible and reversible continuous-time Markov chain* is always denoted by the triple $(\mathcal{X}^n, \mathcal{Q}, \pi)$.

In this chapter we use three different notions of gradients: On a Riemannian manifold $(M, g)$ we denote the *gradient of a smooth function* $f \in C^\infty(M)$ by $\operatorname{grad}_g f$. On the discrete space $\mathcal{X}^n$ the *discrete gradient* of a function $\psi : \mathcal{X}^n \to \mathbb{R}$ is denoted by $\nabla\psi_{ij} := \psi_j - \psi_i$. Finally, $\nabla f$ denotes the usual *Euclidean gradient* of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, which should not be confused with the Levi-Civita connection $\nabla_X Y$ of vector fields $X, Y$ on a Riemannian manifold.

The Euclidean space $\mathbb{R}^n$ will be always endowed with the standard smooth structure and the standard Euclidean scalar product.

---

**3.1.1 Definition** A function $\vartheta : \mathbb{R}_0^+ \times \mathbb{R}_0^+ \to \mathbb{R}_0^+$ is called ***weight function*** if $\vartheta$ has the following properties:

(W1) $\vartheta$ is continuous and $\vartheta$, restricted to $\mathbb{R}^+ \times \mathbb{R}^+$, is infinitely differentiable;

(W2) $\vartheta$ is symmetric, i.e. $\vartheta(s, t) = \vartheta(t, s)$ for all $s, t \in \mathbb{R}_0^+$;

(W3) $\vartheta$ is strictly positive on $\mathbb{R}^+ \times \mathbb{R}^+$;

(W4) $\vartheta$ is monotone in the following sense:

$$\vartheta(r, s) \leq \vartheta(r, t) \qquad \forall r, s, t \in \mathbb{R}_0^+ : s \leq t;$$

(W5) $\vartheta$ is concave.

Of later interest will be the following particular weight function

$$\theta_m(s, t) := \begin{cases} \dfrac{m-1}{m}\dfrac{s^m - t^m}{s^{m-1} - t^{m-1}}, & \text{if } m \in \mathbb{R}^+ \setminus \{1\}; \qquad \text{(3.1.a)} \\[2ex] \dfrac{s - t}{\log s - \log t}, & \text{if } m = 1. \qquad \text{(3.1.b)} \end{cases}$$

We collect some important properties of $\theta_m$.

**3.1.2 Facts**

(i) For $m \neq 1$, the mapping $\theta_m$ admits the following integral representation:

$$\theta_m(s, t) = \int_0^1 \left((1 - \alpha)s^{m-1} + \alpha t^{m-1}\right)^{1/(m-1)} \mathrm{d}\alpha \qquad \forall s, t \in \mathbb{R}^+. \tag{3.2}$$

This equation follows easily from the fact that an antiderivative of the integrand in (3.2) is given by

$$\frac{m-1}{m}\frac{\left(s^{m-1}-\alpha(s^{m-1}-t^{m-1})\right)^{m/(m-1)}}{t^{m-1}-s^{m-1}}.$$

(ii) In the case $m=1$, the equation in (3.2) becomes

$$\theta_1(s,t)=\int_0^1 s^{1-\alpha}t^{\alpha}\,\mathrm{d}\alpha \qquad \forall s,t\in\mathbb{R}_0^+. \tag{3.3}$$

In particular, we have

$$\lim_{m\to 1}(s,t)=\theta_1(s,t) \qquad \forall s,t\in\mathbb{R}_0^+.$$

(iii) For $0<m\leq 2$ the mapping $\theta_m$ is a weight function. Indeed, $\theta_m$ satisfies (W1) to (W3) in Definition 3.1.1 for every choice of $m\in\mathbb{R}^+$. To show (W3), we appeal to the integral representation in (3.2).Denote by $f_m:\mathbb{R}_0^+\times\mathbb{R}_0^+\to\mathbb{R}_0^+$ the integrand of the representation given in (3.2), namely

$$f_m(s,t):=\Big((1-\alpha)s^{m-1}+\alpha t^{m-1}\Big)^{1/(m-1)}. \tag{3.4}$$

Then the Hessian of $f_m$ is given by

$$\mathrm{Hess} f_m(s,t)=(1-\alpha)\alpha(2-m)\Big((1-\alpha)s^{m-1}+\alpha t^{m-1}\Big)^{1/(m-1)-2}s^{m-3}t^{m-3}\begin{pmatrix}-t^2 & st\\ st & -s^2\end{pmatrix},$$

Assuming $m\leq 2$, all principal minors of $\mathrm{Hess} f_m$ are nonpositive for all $s,t\in\mathbb{R}^+$, as well as all $\alpha\in[0,1]$. Therefore, the matrix $\mathrm{Hess} f_m$ is negative semi-definite by *Sylvester's criterion*, and the integrand $f_m$ is concave for all $\alpha\in[0,1]$; the concavity of $\theta_m$ follows.
For $m=1$ the argument follows along similar lines.

Note that for $m>2$, $\mathrm{Hess} f_m$ is positive semi-definite for all $s,t\in\mathbb{R}^+$, $\alpha\in[0,1]$. In this case, $\theta_m$ is only convex and cannot depict a weight function.

(iv) Directly from (3.1) follows that $\theta_m$ is homogeneous, i.e.

$$\theta_m(\alpha s,\alpha t)=\alpha\theta_m(s,t) \qquad \forall\alpha>0,\ \forall s,t\in\mathbb{R}_0^+.$$

(v) $\theta_m$ is monotonous in $m$; more precisely, we have

$$\theta_m(s,t)\leq\theta_k(s,t) \qquad \forall s,t,m,k\in\mathbb{R}^+: m\leq k.$$

To see this, let $f_n$ be defined as in (3.4). It is enough to show that $\mapsto f_m$ is monotonous for all $\alpha\in[0,1]$. However, this follows readily from Jensen's inequality for two points if we additionally assume that $k>1$ and $m\neq 0$:

$$f_m^{k-1}(s,t)=\Big((1-\alpha)s^{m-1}+\alpha t^{m-1}\Big)^{\frac{k-1}{m-1}}\leq(1-\alpha)s^{k-1}+\alpha t^{k-1}=f_k^{k-1}(s,t).$$

For $k<1$ we get a similar estimate applying the concave version of Jensen's inequality. Finally, the cases $m=1$ or $k=1$ follow by continuity.

(vi) The weight function $\theta_m$ vanishes at the boundary $\{0\}\times\mathbb{R}_0^+\cup\mathbb{R}_0^+\times\{0\}$ precisely when $0<m\leq 1$.

(vii) The homogeneity of $\theta_m$ implies the useful identity

$$(s,t)\cdot\nabla\theta_m(s,t)=\theta_m(s,t) \qquad \forall s,t,\in\mathbb{R}^+,$$

since the left-hand side equals $\frac{\mathrm{d}}{\mathrm{d}r}\Big|_{r=1} \theta_m(rs, rt) = \frac{\mathrm{d}}{\mathrm{d}r}\Big|_{r=1} r\theta_m(s,t) = \theta_m(s,t)$.

(viii) For $0 < m \leq 2$ the weight function $\theta_m$ satisfies the estimate

$$(s,t) \cdot \nabla\theta_m(u,v) \geq \theta_m(s,t) \qquad \forall s,t,u,v \in \mathbb{R}^+. \tag{3.5}$$

Indeed, due to our assumption on $m$, the gradient $\nabla\theta_m$ is a monotonous function, i.e.

$$(s-x, t-y) \cdot \big(\nabla\theta_m(s,t) - \nabla\theta_m(x,y)\big) \leq 0.$$

Now, taking $(x,y) = \varepsilon(u,v)$ with $\varepsilon > 0$ and passing to the limit ($\varepsilon \searrow 0$) in the inequality above, yields

$$(s,t) \cdot \big(\nabla\theta_m(s,t) - \nabla\theta_m(u,v)\big) \leq 0,$$

where we used that $\nabla\theta_m(\varepsilon u, \varepsilon v) = \nabla\theta_m(u,v)$. Finally, applying Fact 3.1.2.vii to this inequality results in (3.5).

**Notation**

Let $(\mathcal{X}^n, \mathcal{Q}, \pi)$ be an irreducible and reversible continuous-time Markov chain. Note that irreducibility implies that the corresponding Markov semigroup $e^{t\mathcal{Q}}$ has strictly positive entries for all times $t > 0$. Since the stationary distribution $\pi$ is invariant under $e^{t\mathcal{Q}}$, this means that the stationary distribution $\pi$ is strictly positive on $\mathcal{X}^n$ in this case. We will introduce a metric on the class of all discrete probability densities with respect to a stationary distribution $\pi$ on $\mathcal{X}^n$, denoted by

$$\mathcal{P}^n := \left\{\rho : \mathcal{X}^n \to \mathbb{R}_0^+ : \sum_{i=1}^n \rho_i \pi_i = 1\right\}.$$

Recalling the definition of the weight function $\theta_m$ in (3.1), we will also make use of the shorthand notation $\hat{\rho}_{ij} := \theta_m(\rho_i, \rho_j)$ for a discrete probability density $\rho \in \mathcal{P}^n$.

Since the stationary distribution $\pi$ has full support, the set $\mathcal{P}^n$ represents an $(n-1)$-simplex with vertices $\big((1/\pi_i)e_i\big)_{1\leq i\leq n}$ in $\mathbb{R}^n$. In this context, the probability measure corresponding to $\rho$ is also known as *barycentric coordinates* of the point $\rho$ in the $(n-1)$-simplex $\mathcal{P}^n$.

Note that $\mathcal{P}^n$ is a subset of an affine subspace in $\mathbb{R}^n$ which is orthogonal to $\pi^\top$. Since $\pi$ is the stationary distribution of the infinitesimal generator $\mathcal{Q}$, this means that this particular affine subspace is just $\operatorname{ran}\mathcal{Q} + \rho$ for any $\rho \in \mathcal{P}^n$.

Simplices like $\mathcal{P}^n$ are simple examples of topological manifolds with boundary. However, $\mathcal{P}^n$ depicts no smooth manifold with boundary, due to neighbourhoods of the vertices in the simplex being not diffeomorphic to the half-space $\mathbb{R}_+^{n-1}$. Here the closely related concept of a manifold with corners provides a remedy.

Moreover, with the standard Euclidean scalar product at hand, it is not hard to endow $\operatorname{int}\mathcal{P}^n$ with the standard Riemannian structure. However, we will pursue a slightly different direction and endow the interior of $\mathcal{P}^n$ with a metric tensor which is induced by the underlying Markov chain $(\mathcal{X}^n, \mathcal{Q}, \pi)$. To this aim, it will be useful to identify tangent vectors with certain discrete gradients by means of the following preliminary result.

**3.1.3 Lemma** *Let $(\mathcal{X}^n, \mathcal{Q}, \pi)$ be an irreducible and reversible continuous-time Markov chain and fix a weight function $\vartheta$. Let $\varrho : (-\varepsilon, \varepsilon) \to \operatorname{int}\mathcal{P}^n$ be a differentiable curve. Then there exists a unique family of discrete gradients $\nabla\psi(t)$ for some $\psi : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ such that the discrete continuity equation*

$$\dot{\varrho}_i(t) + \sum_j \nabla\psi_{ij}(t)\hat{\rho}_{ij}(t)\mathcal{Q}_{ij} = 0 \tag{3.6}$$

*is satisfied for all times $t \in (-\varepsilon, \varepsilon)$.*

Proof We start by rewriting the discrete continuity equation in (3.6) in terms of the matrix-vector formulation

$$\dot{\varrho} = B(\varrho)\psi,$$

where we dropped the dependency of the terms on the time variable $t$ and the matrix $B(\varrho) \in \mathbb{R}^{n\times n}$ is given by

$$B_{ij}(\varrho) := \begin{cases} -\mathcal{Q}_{ij}\hat{\varrho}_{ij} & \text{if } i \neq j, \quad (3.7.a)\\ \sum_{k\neq i} \mathcal{Q}_{ik}\hat{\varrho}_{ik} & \text{if } i = j. \quad (3.7.b)\end{cases}$$

Using that $\pi\mathcal{Q} = 0$, it is straightforward to verify that $\pi B(\varrho) = 0$. This means that $\operatorname{ran} B(\varrho) \subseteq \pi^\perp$ where the subspace $\pi^\perp$ is spanned exactly by the shifted simplex $\mathcal{P}^n - \pi^\top$. Since every vector tangent to a curve in $\mathcal{P}^n$ belongs to $\pi^\perp$, it remains to show that the range of $B(\varrho)$ agrees with $\pi^\perp$. To this end, it is useful to work with the matrix $A(\varrho) := \operatorname{diag} \pi B(\varrho)$. Note that $A(\rho)$ is symmetric whilst $B(\varrho)$ need not be.

Now we use the trivial identity $2\psi_i\psi_j = \psi_i^2+\psi_j^2-(\psi_i-\psi_j)^2$ and the fact that $\sum\limits_i A_{ij}(\varrho) = \sum\limits_j A_{ij}(\varrho) = 0$ to compute

$$2\psi^\top A(\varrho)\psi = \sum_{i,j} 2\psi_i\psi_j A_{ij}(\varrho) =$$

$$= \sum_{i,j} \psi_i^2 A_{ij}(\varrho) + \sum_{i,j} \psi_j^2 A_{ij}(\varrho) - \sum_{i,j} (\psi_i - \psi_j)^2 A_{ij}(\varrho) = -\sum_{i,j} (\psi_i - \psi_j)^2 A_{ij}(\varrho).$$

Since $A_{ij}(\varrho) \leq 0$ for all $i \neq j$, this equation shows that every $\psi \in \ker A(\varrho)$ belongs to the 1-dimensional subspace $\operatorname{span}\{(1,\dots 1)^\top\}$ of $\mathbb{R}^n$. It remains to note that $\ker B(\varrho)$ agrees with $\ker A(\varrho)$ to prove the claim. ■

**3.1.4 Proposition** *Let $(\mathcal{X}^n, \mathcal{Q}, \pi)$ be an irreducible and reversible continuous-time Markov chain and fix a weight function $\vartheta$.*

(i) *The set $\mathcal{P}^n$ is a compact and connected embedded hypersurface with corners in $\mathbb{R}^n$.*

(ii) *The manifold $\mathcal{P}^n$ admits a global trivialization of the tangent bundle, i.e. $T\mathcal{P}^n \simeq \mathcal{P}^n \times \mathbb{R}^{n-1}$. For every point $\rho \in \operatorname{int}\mathcal{P}^n$ we have*

$$T_\rho\mathcal{P}^n \simeq \{\nabla\psi : \psi \in \mathbb{R}^n\}, \tag{3.8}$$

*by means of the following identification: Every tangent vector $\dot{\varrho}(0)$ of a smooth curve $\varrho : (-\varepsilon,\varepsilon) \to \operatorname{int}\mathcal{P}^n$ with $\rho = \varrho(0)$ is uniquely identified with a discrete gradient $\nabla\psi(0)$ via the discrete continuity equation in* (3.6).

*By slight abuse of notation, we will simply write $\nabla\psi \in T_\rho\mathcal{P}^n$.*

(iii) *Define a metric tensor $g : T_p\mathcal{P}^n \times T_p\mathcal{P}^n \to \mathbb{R}$ – using the identification* (3.8) *– by*

$$\langle\nabla\psi,\nabla\phi\rangle_g := \frac{1}{2}\sum_{i,j} \nabla\psi_{ij}\nabla\phi_{ij}\hat{\rho}_{ij}\mathcal{Q}_{ij}\pi_i \qquad \forall\, \nabla\psi,\nabla\phi \in T_\rho\mathcal{P}^n. \tag{3.9}$$

*Then $(\operatorname{int}\mathcal{P}^n, g)$ is a Riemannian manifold.*

Proof The statement in (i) follows from the fact that $\pi$ as stationary distribution of an irreducible Markov chain is nowhere vanishing, whereas (ii) is a direct consequence of Lemma 3.1.3.

It remains to prove that the metric tensor $g$ in (iii) defined by (3.9) is positive definite: This claim can be shown easily by writing $g$ with respect to the standard scalar product on $\mathbb{R}^n$:

$$\langle \nabla\psi, \nabla\phi \rangle_g = \psi^\top A(\rho)\psi,$$

where the symmetric matrix $A(\rho) \coloneqq \operatorname{diag}\pi B(\rho)$ with $B(\rho)$ as in (3.7) was already encountered in the proof of Lemma 3.1.3. Alternatively, one can express $A(\rho)$ more tangible as

$$A_{ij}(\rho) = \begin{cases} -\mathcal{Q}_{ij}\hat{\rho}_{ij}\pi_i, & \text{if } i \neq j; \qquad \text{(3.10.a)} \\ \sum_{k\neq i} \mathcal{Q}_{ik}\hat{\rho}_{ik}\pi_i, & \text{if } i = j \qquad \text{(3.10.b)} \end{cases}$$

Since

$$\sum_{j\neq i} \left|A_{ij}(\rho)\right| = A_{ii}(\rho) \geq 0 \qquad \forall i \in \mathcal{X}^n,$$

it becomes clear that $A(\rho)$ is diagonally dominant and consequently also positive definite. ∎

The following remark shows that the metric tensor $g$ defined by (3.9) degenerates at the boundary of $\mathcal{P}^n$.

**3.1.5 Remark** (Degeneracy of the metric tensor at the boundary) Let $(\mathcal{X}^n, \mathcal{Q}, \pi)$ be a Markov chain and let the the weight function be given by $\theta_m$ as in (3.1). Then the metric tensor $g$ is degenerated at a point $\rho \in \mathcal{P}^n$ precisely when the matrix $A(\rho) \in \mathbb{R}^{n\times n}$ defined in (3.10) has rank $A(\rho) < n-1$. Clearly, Proposition 3.1.4.iii implies that this can only happen at the boundary $\partial\mathcal{P}^n$.
More precisely, $g$ is always degenerated at the vertices $\big((1/\pi_i)e_i\big)_{1\leq i\leq n}$. If for instance $m \leq 1$, then $\hat{\rho}_{ij}$ vanishes at the boundary $\partial\mathcal{P}^n$ for some $i \neq j$, due to Fact 3.1.2.vi. Therefore, $g$ is degenerated at the whole boundary $\partial\mathcal{P}^n$ in this case.

The Riemannian manifold $(\operatorname{int}\mathcal{P}^n, g)$ is naturally equipped with a metric such that the metric topology agrees with the topology corresponding to the smooth manifold $\operatorname{int}\mathcal{P}^n$. This metric between two points $\rho_0, \rho_1 \in \operatorname{int}\mathcal{P}^n$ is given by minimizing the length of all smooth curves $\rho : [0,1] \to \operatorname{int}\mathcal{P}^n$ connecting $\rho_0$ to $\rho_1$, where the length of such a curve $\rho(t)$ is given by

$$L(\rho) \coloneqq \int_0^1 \sqrt{g\big(\dot{\rho}(t), \dot{\rho}(t)\big)}\,\mathrm{d}t. \tag{3.11}$$

Equivalently, one may consider the *energy functional*

$$E(\rho) \coloneqq \int_0^1 g\big(\dot{\rho}(t), \dot{\rho}(t)\big)\,\mathrm{d}t \tag{3.12}$$

instead of the length functional $L$ since Jensen's inequality implies $L(\rho) \leq \sqrt{E(\rho)}$ with equality of $L(\rho)$ and $\sqrt{E(\rho)}$ precisely when the curve $\rho$ has constant speed.
This notion of the length or energy of a curve clearly depends on the metric tensor $g$ involved. Therefore it comes to some surprise that it allows to defines a distance on the whole manifold with corners.

**3.1.6 Proposition** *For any two points $\rho_0, \rho_1 \in \mathcal{P}^n$ define*

$$\mathcal{W}(\rho_0, \rho_1) \coloneqq \inf\Big(\int_0^1 |\nabla\psi(t)|_g^2\,\mathrm{d}t\Big)^{1/2}, \tag{3.13}$$

*where the infimum is taken over all vector fields $\nabla\psi$ along (piecewise) smooth curves $\rho : [0,1] \to \mathcal{P}^n$, connecting $\rho_0$ to $\rho_1$ such that*

(CE1) $\nabla\psi : [0,1] \to \mathbb{R}^n \times \mathbb{R}^n$ *is Borel measurable;*

(CE2) *the pair $(\rho, \nabla\psi)$ satisfies the discrete continuity equation*

$$\dot{\rho}_i(t) + \sum_j \nabla\psi_{ij}(t)\hat{\rho}_{ij}(t)\mathcal{Q}_{ij} = 0 \qquad \forall t \in (0,1). \tag{3.14}$$

*Then $(\mathcal{P}^n, \mathcal{W})$ is a Polish metric space. Moreover, $(\mathcal{P}^n, \mathcal{W})$ is a geodesic space in the metric sense, i.e. for every pair of points $\rho_0, \rho_1 \in \mathcal{P}^n$ there exists a curve $\rho : [0,1] \to \mathcal{P}^n$, connecting $\rho_0$ to $\rho_1$ in such a way that*

$$\mathcal{W}(\rho(s), \rho(t)) = |s-t|\,\mathcal{W}(\rho_0, \rho_1) \qquad \forall s,t \in [0,1].$$

Sketch of proof We will only show that $\mathcal{W}$ defines an extended metric on $\mathcal{P}^n$, i.e $\mathcal{W}$ takes values in $[0, +\infty]$. The prove of the other claims is rather involved and requires a thorough analysis of the two-point space $\mathcal{P}^2$.

The symmetry of $\mathcal{W}$ is clear from the definition. In order to show that that $\mathcal{W}$ is positive definite, it is enough to establish an estimate of the form

$$|\rho^0 - \rho^1| \le C\,\mathcal{W}(\rho^0, \rho^1) \qquad \forall \rho^0, \rho^1 \in \mathcal{P}^n \tag{3.15}$$

for some constant $C > 0$. Accordingly, let $\rho^0, \rho^1 \in \mathcal{P}^n$ where we may assume that $\mathcal{W}(\rho^0, \rho^1) < +\infty$. By definition of $\mathcal{W}$, for every $\varepsilon > 0$ there exists a pair $\rho, \nabla\psi$ satisfying the discrete continuity equation (3.14) such that

$$\int_0^1 |\nabla\psi(t)|_g^2 \,\mathrm{d}t \le \mathcal{W}^2(\rho^0, \rho^1) + \varepsilon. \tag{3.16}$$

Now we will again use the matrix notation from the proofs of Lemma 3.1.3 and Proposition 3.1.4: Namely, dropping any dependency on $t$, the discrete continuity equation takes the form $\dot{\rho} = B(\rho)\psi$, whereas the Riemannian metric may be written as $\langle \nabla\psi, \nabla\phi \rangle_g = \psi^\top A(\rho)\psi$ for matrices $A(\rho)$ and $B(\rho)$ defined in (3.10) and (3.7), respectively. Thus, we can appeal to the relation $A(\rho) = \operatorname{diag}\pi B(\rho)$ to infer for any $\eta \in \mathbb{R}^n$ the estimate

$$\min_j \pi_j \Big|\sum_i \eta_i(\rho_i^0 - \rho_i^1)\Big| \le \Big|\sum_i \eta_i(\rho_i^0 - \rho_i^1)\pi_i\Big| = \Big|\int_0^1 \dot{\rho}^\top(\operatorname{diag}\pi)\eta \,\mathrm{d}t\Big| =$$
$$= \Big|\int_0^1 \big(B(\rho)\psi\big)^\top(\operatorname{diag}\pi)\eta\,\mathrm{d}t\Big| = \Big|\int_0^1 \psi^\top A(\rho)\eta\,\mathrm{d}t\Big| \le \Big(\int_0^1 \psi^\top A(\rho)\psi\,\mathrm{d}t\Big)^{1/2}\Big(\int_0^1 \eta^\top A(\rho)\eta\,\mathrm{d}t\Big)^{1/2},$$

where we used the Cauchy-Schwarz inequality for the positive bilinear form $(\psi, \eta) \mapsto \psi^\top A(\rho)\eta$ in the last inequality above. Since we have the rough estimate

$$\eta^\top A(\rho)\eta = \frac{1}{2}\sum_{i,j}(\eta_i - \eta_j)^2 \mathcal{Q}_{ij}\hat{\rho}_{ij}\pi_i \le 2\,|\vartheta|_\infty\,|\eta|_\infty^2 \sum_{\substack{i,j\\ i\neq j}} \mathcal{Q}_{ij}\pi_i$$

and $\pi$ is strictly positive, we obtain

$$\Big|\sum_i \eta_i(\rho_i^0 - \rho_i^1)\Big| \le C\,|\eta|_\infty \Big(\int_0^1 \psi^\top A(\rho)\psi\,\mathrm{d}t\Big)^{1/2}$$

for a suitable constant $C > 0$. Now taking (3.16) into account and letting $(\varepsilon \searrow 0)$, this inequality becomes

$$\Big|\sum_i \eta_i(\rho_i^0 - \rho_i^1)\Big| \le C\,|\eta|_\infty \mathcal{W}(\rho^0,\rho^1). \tag{3.17}$$

Choosing $\eta_i = \mathrm{sgn}(\rho_i^0 - \rho_i^1)$ in (3.17), finally establishes (3.15).

It remains to show that $\mathcal{W}$ satisfies the triangle inequality: Note that a reparametrisation à la Lemma 1.1.5 allows us to take the infimum in (3.13) only over pairs $(\hat\rho, \nabla\hat\psi)$ where the curve $\hat\rho$ has constant speed. As a result, the energy functional $E(\hat\rho)$ in (3.11) agrees with the length functional $L(\hat\rho)$ in (3.11). Thus, we may invoke the triangle inequality in $L^1(0,1)$ to obtain

$$\int_0^1 \left|\nabla\hat\psi^{1,2}(t)\right|_g^2 \mathrm{d}t = \int_0^1 \left|\nabla\hat\psi^{1,2}(t)\right|_g \mathrm{d}t \le \int_0^1 \left|\nabla\hat\psi^{1}(t)\right|_g \mathrm{d}t + \int_0^1 \left|\nabla\hat\psi^{2}(t)\right|_g \mathrm{d}t, \tag{3.18}$$

where $(\hat\rho^{1,2}, \nabla\hat\psi^{1,2})$ denotes the reparametrised composition of constant speed pairs $(\hat\rho^1, \nabla\hat\psi^1)$ and $(\hat\rho^2, \nabla\hat\psi^2)$. Now taking the infimum in (3.18) over such curves $\hat\rho^1$ connecting points $\rho_0$ to $\rho_1$ and $\hat\rho^2$ connecting points $\rho_1$ to $\rho_2$, results in the sought triangle inequality

$$\mathcal{W}(\rho_0,\rho_2) \le \mathcal{W}(\rho_0,\rho_1) + \mathcal{W}(\rho_1,\rho_2).$$ ∎

The Riemannian nature of $\mathrm{int}\,\mathcal{P}^n$ assures that $\mathrm{int}\,\mathcal{P}^n\mathcal{W}$ is a geodesic space by the Hopf-Rinow theorem (see Theorem B.3.4 in Appendix B). Moreover, the constant-speed geodesics may be characterised by a system of first-order equations as the following result shows.

**3.1.7 Proposition** (Geodesic equations) *For every point $\rho_0 \in \mathrm{int}\,\mathcal{P}^n$ and $\psi_0 \in \mathbb{R}^n$, there exists a sufficiently small $\varepsilon > 0$ such that the unique constant-speed geodesic $\rho : (-\varepsilon,\varepsilon) \to \mathrm{int}\,\mathcal{P}^n$ f, starting from $\rho(0) = \rho_0$ with initial tangent vector $\nabla\psi(0) = \nabla\psi_0$ satisfies the following system of equations:*

$$\dot\rho_i(t) + \sum_j \nabla\psi_{ij}(t)\hat\rho_{ij}\mathcal{Q}_{ij} = 0, \tag{3.19.a}$$

$$\dot\psi_i(t) + \frac{1}{2}\sum_j \big(\nabla\psi_{ij}(t)\big)^2 \partial_1\vartheta(\rho_i(t),\rho_j(t))\mathcal{Q}_{ij} = 0. \tag{3.19.b}$$

Proof All statements are standard of results in Riemannian geometry summarised in Appendix B: Existence and uniqueness of the constant-speed geodesic in $\mathrm{int}\,\mathcal{P}^n$ follows by Proposition B.3.2 and the fact that every geodesic may be reparametrised to constant speed such that $\varepsilon = 1$.

The characterisation of such a geodesic by (3.19) just corresponds to the geodesic equations in local coordinates as given in Definition B.3.1. ∎

In addition to the metric $\mathcal{W}$, one can also consider Wasserstein distances on $\mathcal{P}^n$: To this end, it is natural to endow $\mathcal{X}^n$ with the graph distance induced by the infinitesimal generator $\mathcal{Q}$, i.e. the length of the shortest path in the graph with vertex set $\mathcal{X}^n$ and edge set

$$\left\{(i,j) \in \mathcal{X}^n \times \mathcal{X}^n : \mathcal{Q}_{ij} > 0\right\}.$$

We will write $W_p^{\mathrm{gra}}$ for the $p$-Wasserstein distance with respect to the graph distance on $(\mathcal{X}^n, \mathcal{Q}, \pi)$.

In the next result we provide lower and upper bounds for the discrete transportation metric $\mathcal{W}$ in terms of $W_1^{\mathrm{gra}}$ and $W_2^{\mathrm{gra}}$.

**3.1.8 Proposition** *Let $(\mathcal{X}^n, \mathcal{Q}, \pi)$ be a Markov chain and assume that the weight function $\vartheta$ vanishes on the boundary $\{0\} \times \mathbb{R}_0^+$. Then there exists a constant $C > 0$ only dependent on the choice of $\vartheta$ such that we have the bounds*

$$\sqrt{2}W_1^{\mathrm{gra}}(\rho_0,\rho_1) \le \mathcal{W}(\rho_0,\rho_1) \le C\big(\min_{\mathcal{Q}_{ij}>0}\mathcal{Q}_{ij}\big)^{-1/2} W_2^{\mathrm{gra}}(\rho_0,\rho_1) \qquad \forall \rho_0,\rho_1 \in \mathcal{P}^n.$$

## 3.2 The Discrete Porous Medium Equation as Entropy Gradient Flow

**Notation**

From this section onwards, the Riemannian and metric structure of $\operatorname{int}\mathcal{P}^n$ and $\mathcal{P}^n$, respectively, which were introduced in the last section, will be always induced by the weight function $\theta_m$.

In this section we will develop a gradient flow structure which may be regarded as a discrete analogue of the Wasserstein gradient flows. We will focus on a gradient flow structure for a mixed energy functional which consists of the discrete Rényi entropy together with an discrete potential energy functional.

**3.2.1 Definition** Let $(\mathcal{X}^n, \mathcal{Q}, \pi)$ be an irreducible continuous-time Markov chain. For $m > 0$ the ***(discrete) Rényi entropy functional*** $\mathcal{F}_m^n : \mathcal{P}^n \to \mathbb{R}$ associated to the stationary distribution $\pi$ is defined by

$$\mathcal{F}_m^n(\rho) := \sum_{i=1}^n f_m(\rho_i)\pi_i,$$

where $f_m$ is given as in (2.19). For $m = 1$ the Rényi entropy takes the particular form

$$\mathcal{F}_1^n(\rho) := \sum_{i=1}^n \rho_i \pi_i \log \rho_i,$$

which is also known as ***(discrete) Shannon entropy functional*** associated to $\pi$.
For $\upsilon \in \mathbb{R}^n$, the ***(discrete) potential energy functional*** $\mathcal{V}^n : \mathcal{P}^n \to \mathbb{R}$ associated with $\upsilon$ is defined by

$$\mathcal{V}^n(\rho) := \sum_{i=1}^n \upsilon_i \rho_i \pi_i.$$

There are several possibilities to proceed with the smooth manifold $\mathcal{P}^n$ or its interior: We may consider $(\mathcal{P}^n, \mathcal{W})$ as a metric space and follow along the lines of the purely metric theory developed in Chapter 1, namely gradient flows in the sense of EDI, EDE, or EVI.
However, for now we will pursue a different approach which makes use of the underlying Riemannian structure of $\operatorname{int}\mathcal{P}^n$. To this end, we invoke the *fundamental theorem on flows* stated in Appendix B, which – together with the compactness of $\mathcal{P}^n$ – implies the existence of a global gradient flow for the mixed energy functional $\Phi^n = \mathcal{F}_m^n + \mathcal{V}^n$. To characterises the flow curves in a fashion similar to Theorem 2.4.10, we introduce the notions of discrete Laplacian and discrete divergence.

**3.2.2 Definition** Let $(\mathcal{X}^n, \mathcal{Q}, \pi)$ be a Markov chain. Given a function $\rho : \mathcal{X}^n \to \mathbb{R}$,

$$\Delta \psi_i := \sum_{j=1}^n \mathcal{Q}_{ij}(\psi_j - \psi_i) = \sum_{j=1}^n \mathcal{Q}_{ij}\psi_j$$

denotes the ***discrete Laplacian*** of $\rho$ associated to $\mathcal{Q}$. Given a function $\psi : \mathcal{X}^n \times \mathcal{X}^n \to \mathbb{R}$,

$$\nabla \cdot \psi_i := \frac{1}{2}\sum_{j=1}^n \mathcal{Q}_{ij}(\psi_{ij} - \psi_{ji})$$

denotes the ***discrete divergence*** of $\psi$ associated to $\mathcal{Q}$.

**3.2.3 Proposition** (Gradient flow associated to the discrete porous medium equation) *Let an irreducible continuous-time Markov chain $(\mathcal{X}^n, \mathcal{Q}, \pi)$ and an discrete potential $\upsilon \in \mathbb{R}^n$ be given; fix $0 < m \leq 2$.*

*Then for every $\rho_0 \in \operatorname{int}\mathcal{P}^n$ there exists a unique differentiable curve $\rho : (-\varepsilon, \varepsilon) \to \operatorname{int}\mathcal{P}^n$, starting from $\rho(0) = \rho_0$, such that the following two equivalent statements hold:*

(i) *$\rho$ satisfies the gradient flow equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = -(\operatorname{grad}_g \Phi^n)_{\rho(t)} \tag{3.20}$$

*for the functional $\Phi^n = \mathcal{F}_m^n + \mathcal{V}^n$;*

(ii) *$\rho$ satisfies the* ***discrete porous medium equation with drift***

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = \nabla\cdot(\hat{\rho}(t)\,\nabla f'(\rho(t)) + \hat{\rho}(t)\,\nabla\upsilon) = \Delta\left(\rho^m(t)\right) + \nabla\cdot(\hat{\rho}(t)\,\nabla\upsilon)\,. \tag{3.21}$$

*If at least one of the following conditions is satisfied additionally:*

- *$m \le 1$ and the Markov chain allows only nearest-neighbour interactions, i.e. $\mathcal{Q}_{ij} = 0$ for all $|i-j| > 1$;*
- *the porous medium equation is homogeneous, i.e. $\nabla\upsilon \equiv 0$;*

*then for every $\rho_0 \in \mathcal{P}^n$, there exists a unique differentiable curve $\rho : \mathbb{R}^+ \to \operatorname{int}\mathcal{P}^n$, starting from $\lim_{t\searrow 0}\rho(t) = \rho_0$ such that (i) and (ii) hold.*

Proof Let us start by showing that the equations (3.20) and (3.21) give rise to the same solution. To this aim, pick a pair $(\varrho, \nabla\psi)$ satisfying the continuity equation (3.14). Then we may use summation by parts and the detailed balance condition $\mathcal{Q}_{ij}\pi_i = \mathcal{Q}_{ji}\pi_j$ to compute

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\Phi^n(\varrho(t)) &= \frac{\mathrm{d}}{\mathrm{d}t}\sum_i \big(f_m(\varrho_i(t)) + \upsilon_i\varrho_i(t)\big)\pi_i = -\sum_{i,j}\big(f_m'(\varrho_i(t)) + \upsilon_i\big)\nabla\psi_{ij}(t)\hat{\varrho}_{ij}(t)\mathcal{Q}_{ij}\pi_i = \\ &= -\frac{1}{2}\sum_{i,j}\big(f_m'(\varrho_i(t)) + \upsilon_i\big)\hat{\varrho}_{ij}(t)\mathcal{Q}_{ij}\pi_i\big(\nabla\psi_{ij}(t) - \nabla\psi_{ji}(t)\big) = \big\langle\nabla f_m'(\varrho(t)) + \nabla\upsilon, \nabla\psi(t)\big\rangle_g.\end{aligned}$$

This means we have identified $\nabla f_m'(\rho) + \nabla\upsilon$ as gradient of the functional $\Phi^n$ at the point $\rho$ in the Riemannian manifold $(\operatorname{int}\mathcal{P}^n, g)$. Using the continuity equation (3.14) once again to identify

$$(\operatorname{grad}_g\Phi^n)_{\rho(t)} = \nabla f_m'(\varrho(t)) + \nabla\upsilon$$

with the velocity of a curve $\rho$ in $\operatorname{int}\mathcal{P}^n$, to wit

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_i(t) = -\sum_j\big(\nabla f_m'(\rho(t)) + \nabla\upsilon\big)\hat{\rho}_{ij}(t)\mathcal{Q}_{ij} = -\nabla\cdot\big(\hat{\rho}_{ij}(t)\,\nabla f_m'(\rho(t)) + \hat{\rho}_{ij}(t)\,\nabla\upsilon\big). \tag{3.22}$$

To verify the last equality in (3.21), it is enough to check that $\hat{\rho}_{ij}\,\nabla f_m'(\rho) = \nabla(\rho^m)$ by definition of the weight function $\theta_m$.

To conclude the first part of the proof, we still need to infer existence and uniqueness of the curve $\rho : (-\varepsilon, \varepsilon) \to \operatorname{int}\mathcal{P}^n$ solving (3.22). However, this follows immediately from the fundamental theorem on flows (cf.Theorem B.2.5), noting that $(\operatorname{grad}_g\Phi^n)_\rho$ is a smooth vector field on $\operatorname{int}\mathcal{P}^n$.

To prove the second part of the proposition, we appeal to Proposition B.2.8 to infer existence of a global solution $\rho : [0,\infty) \to \mathcal{P}^n$. To this end, we have to verify that the vector field

$$V_\rho := \Delta\left(\rho^m\right) + \nabla\cdot(\hat{\rho}\,\nabla\upsilon)\,,$$

depicting the right-hand side of (3.21), is nowhere outward pointing for all boundary points $\rho \in \partial\mathcal{P}^n$. Let us assume for the moment that the potential $\upsilon$, i.e. the vector field takes the form $V_\rho = \Delta\left(\rho^m\right)$. If $\rho \in \partial\mathcal{P}^n$, then we have $\rho_i = 0$ for at least one $i \in \mathcal{X}^n$. Therefore, we obtain for such components that

$$\Delta(\rho^m)_i = \sum_j \mathcal{Q}_{ij}\rho_j^m = \sum_{j\neq i} \mathcal{Q}_{ij}\rho_j^m \geq 0. \tag{3.23}$$

This proves the claim for all points in the boundary except the corner points with vanishing components except $\rho_i = \pi_i$ for one $i \in \mathcal{X}^n$. Here the vector field takes the form

$$\Delta(\rho^m)_i = \sum_j \mathcal{Q}_{ij}\rho_j^m = \mathcal{Q}_{ii}\rho_i^m < 0.$$

Thus, $V_\rho$ is nowhere outward pointing for all $\rho \in \partial\mathcal{P}^n$.
Finally, we investigate the vector field $V_\rho$ when there is given a nonvanishing potential $\upsilon$ for $m \leq 1$. In this case, Fact 3.1.2.vi immediately implies that the term $\nabla\cdot(\hat{\rho}\nabla\upsilon)$ in (3.23) vanishes for all $\rho \in \partial\mathcal{P}^n$. Hence, this instance reduces the case already discussed. ∎

## 3.3 Geodesic Convexity of the Entropy Functional

We already encountered the notion of convexity of a functional along geodesics in section Section 2.3. In the light of Proposition 3.1.6 it seems appropriate to investigate geodesic convexity of the Rényi entropy functional $\mathcal{F}_m^n$ in the metric space $(\mathcal{P}^n, \mathcal{W})$.
Moreover, the Riemannian structure on $\operatorname{int}\mathcal{P}^n$ provides us with additional tools which allow different characterisations of the synthetic definition (1.6) by means of second-order calculus like the Hessian of $\mathcal{F}_m^n$.
We recall that the Hessian of the smooth functional $\mathcal{F}_m^n$ on the Riemannian manifold $(\mathcal{P}^n, g)$ is a tensor field $\operatorname{Hess}\mathcal{F}_m^n \in T_2^0(\operatorname{int}\mathcal{P}^n)$, defined by $\operatorname{Hess}\mathcal{F}_m^n(X,Y) = \nabla_X\nabla_Y\mathcal{F}_m^n$. However, the computation of the Levi-Civita connection $\nabla$ in coordinates requires the inverse of the coordinate matrix $g_{ij}$ of the metric tensor. The computation of this inverse may be avoided if we just consider the quadratic form of the Hessian of $\mathcal{F}_m^n$ along a geodesic $\rho(t)$ since in this case we can exploit the identity

$$\operatorname{Hess}\mathcal{F}_m^n\big(\rho(t)\big)\big(\dot{\rho}(t),\dot{\rho}(t)\big) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{F}_m^n\big(\rho(t)\big). \tag{3.24}$$

As every geodesic $\rho(t)$ is uniquely characterised by its initial value $\rho(0)$ along with its initial speed $\dot{\rho}(0)$, we may use (3.24) to calculate the quadratic form of the Hessian of $\mathcal{F}_m^n$ as follows:

**3.3.1 Lemma** *At every point $\rho \in \operatorname{int}\mathcal{P}^n$ the quadratic form $\nabla\psi \mapsto \operatorname{Hess}\mathcal{F}_m^n(\rho)(\nabla\psi,\nabla\psi)$ is given by*

$$\mathcal{B}_m(\rho,\nabla\psi) := \frac{1}{4}\sum_{i,j,k}(\nabla\psi_{ji})^2\mathcal{Q}_{ij}\pi_i\Big(\partial_1\theta_m(\rho_i,\rho_j)(\rho_k^m-\rho_i^m)\mathcal{Q}_{ik} + \partial_2\theta_m(\rho_i,\rho_j)(\rho_k^m-\rho_j^m)\mathcal{Q}_{jk}\Big) - \tag{3.25.a}$$

$$-\frac{m}{2}\sum_{i,j,k}\nabla\psi_{ji}\hat{\rho}_{ij}\mathcal{Q}_{ij}\pi_i\Big(\rho_i^{m-1}\mathcal{Q}_{ik}\nabla\psi_{ik} - \rho_j^{m-1}\mathcal{Q}_{jk}d\nabla\psi_{jk}\Big). \tag{3.25.b}$$

Proof Let the pair $(\rho,\psi)$ be a solution to to the geodesic equations (3.19). Following along the lines of the first part of the proof for Proposition 3.2.3, we may use (3.19.a) and the definition of the weight function $\theta_m$ to infer

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}_m^n(\rho(t)) = \big\langle\nabla f_m'(\rho(t)),\nabla\psi(t)\big\rangle_g = \frac{1}{2}\sum_{i,j}\nabla\rho_{ij}^m(t)\nabla\psi_{ij}(t)\mathcal{Q}_{ij}\pi_i.$$

Hence, the second derivative of the discrete Rényi entropy is given by

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{F}_m^n(\rho(t)) = \frac{1}{2}\sum_{i,j}\big(\nabla\rho_{ij}^m(t)\nabla\dot{\psi}_{ij}(t) + m\nabla\dot{\rho}_{ij}^{m-1}(t)\nabla\psi_{ij}(t)\big)\mathcal{Q}_{ij}\pi_i.$$

Using the geodesic equations to identify both $\dot{\rho}$ and $\dot{\psi}$, we obtain in a somewhat unwieldy computation

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{F}_m^n(\rho(t)) = \frac{1}{4}\sum_{i,j,k}(\nabla\psi_{ji})^2\mathcal{Q}_{ij}\pi_i\Big(\partial_1\theta_m(\rho_i,\rho_j)\nabla\rho_{ik}^m\mathcal{Q}_{ik} - \partial_1\theta_m(\rho_j,\rho_i)\nabla\rho_{kj}^m\mathcal{Q}_{jk}\Big) -$$
$$-\frac{m}{2}\sum_{i,j,k}\nabla\psi_{ji}\hat{\rho}_{ij}\mathcal{Q}_{ij}\pi_i\Big(\rho_i^{m-1}\mathcal{Q}_{ik}\nabla\psi_{ik} - \rho_j^{m-1}\mathcal{Q}_{jk}d\nabla\psi_{jk}\Big).$$

Finally, applying the trivial identities $\partial_1\theta_m(\rho_i,\rho_j) = \partial_1\theta_m(\rho_j,\rho_i)$ and $\nabla\rho_{kj}^m = -\nabla\rho_{jk}^m$ to the first line of this equation, establishes that

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{F}_m^n\big(\rho(t)\big) = \operatorname{Hess}\mathcal{F}_m^n\big(\rho(t)\big)\big(\nabla\psi(t),\nabla\psi(t)\big),$$

is given by **(3.25)**. ■

The following result shows that $\operatorname{Hess}\mathcal{F}_m^n$ may be used to characterise geodesic convexity of the Rényi entropy functional $\mathcal{F}_m^n$ on the whole manifold $\mathcal{P}^n$, despite the metric tensor $g$ being possibly degenerated at the boundary $\partial\mathcal{P}^n$.

**3.3.2 Proposition** *For every $\kappa \in \mathbb{R}$ then following statements are equivalent:*

(i) *$\mathcal{F}_m^n$ is geodesically $\kappa$-convex in $(\mathcal{P}^n,\mathcal{W})$;*

(ii) *for every given constant speed geodesic $\rho : [0,1] \to \mathcal{P}^n$, the functional $\mathcal{F}_m^n$ is geodesically $\kappa$-convex in $(\mathcal{P}^n,\mathcal{W})$;*

(iii) *for every initial value $\rho_0 \in \mathcal{P}^n$, the solution $\rho(t)$ to the porous medium equation with $\rho(0) = \rho_0$ from Proposition 3.2.3 satisfies (EVI$_\kappa$);*

(iv) *At every point $\rho \in \operatorname{int}\mathcal{P}^n$, the linear operator $\operatorname{Hess}\mathcal{F}_m^n(\rho) - \kappa\,\mathrm{Id}$ is positive semidefinite, i.e.*

$$\mathcal{B}_m(\rho,\nabla\psi) \geq \kappa\,|\nabla\psi|_g^2 \qquad \forall\rho \in \operatorname{int}\mathcal{P}^n,\ \forall\,\nabla\psi \in T_\rho\mathcal{P}^n.$$

**3.3.3 Example** (Geodesic convexity of the entropy on the two-point space) Consider the two-point space $(\mathcal{X}^r r2,\mathcal{Q},\pi)$. Then the infinitesimal generator $\mathcal{Q}$ has the form

$$\mathcal{Q} = \begin{pmatrix} -p & p \\ q & -q \end{pmatrix},$$

for some $p,q \in \mathbb{R}^+$, whereas the stationary distribution may be written as $\pi = \frac{1}{p+q}(q,p)$. Due to the particular simple structure of this Markov chain, one can give a tangible characterisation of the largest possible $\kappa \in \mathbb{R}$ such that the Rényi entropy functional $\mathcal{F}_m^n$ is geodesically $\kappa$-convex. For $0 < m \leq 2$ and $m \neq 1$ this optimal value is given by

$$\kappa_{\max} = \inf_{-1<\alpha<1}\left\{\frac{pm}{2}\theta_m(1-\alpha,1+\alpha)\big((1-\alpha)^{m-2} + (1+\alpha)^{m-2}\big) + (1-\alpha)^{m-1} + (1+\alpha)^{m-1}\right\}. \tag{3.26}$$

For the particular case $m = 1$, above formula simplifies to $\kappa_{\max} = 2p$. ◀

Above example suggests that for every irreducible and reversible continuous-time Markov chain there may exist a constant $\kappa \in \mathbb{R}$ such that the Rényi entropy functional $\mathcal{F}_m^n$ is $\kappa$-convex. However, the following examples show that this is certainly not the case for every choice of $m \in \mathbb{R}^+$.

**3.3.4 Example** *Consider the discrete circle $\mathbb{Z}/n\mathbb{Z} \simeq \mathcal{X}^n$ of length $n \geq 7$, endowed with the normalised discrete Laplacian*

$$\mathcal{Q}_{ij} := \begin{cases} q & \text{if } |i-j| = 1, \\ -2q & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases}$$

*where the stationary distribution takes the form $\pi = \frac{1}{n}$.*
*Fix $m \in \mathbb{R}^+ \setminus (1/4, 7/4)$. Then the Rényi entropy functional $\mathcal{F}_m^n$ is not geodesically convex on $(\mathcal{X}^n, \mathcal{Q}, \pi)$. Additionally, there exists no constant $\kappa \in \mathbb{R}$ such that $\mathcal{F}_m^n$ is geodesically $\kappa$-convex for all $n \geq 7$.*

Proof We start with evaluating the quadratic form $\mathcal{B}_m(\rho, \nabla\psi)$ of the Hessian as given in Lemma 3.3.1. We compute the first expression (3.25.a) as

$$\frac{q^2}{n} \sum_i |\psi_i - \psi_{i+1}|^2 \left( \partial_1 \theta_m(\rho_{i+1}, \rho_i) \left( \frac{\rho_i^m + \rho_{i+2}^m}{2} - \rho_{i+1}^m \right) + \right.$$
$$\left. + \partial_2 \theta_m(\rho_{i+1}, \rho_i) \left( \frac{\rho_{i-1}^m + \rho_{i+1}^m}{2} - \rho_i^m \right) \right),$$

whereas the second expression (3.25.b) becomes

$$\frac{q^2}{n} \sum_i \theta_m(\rho_{i+1}, \rho_i) \Big( |\psi_i - \psi_{i+1}|^2 \left( m\rho_{i+1}^{m-1} + m\rho_i^{m-1} \right) +$$
$$+ (\psi_i - \psi_{i+1}) \big( (\psi_{i+2} - \psi_{i+1}) m\rho_{i+1}^{m-1} + (\psi_i - \psi_{i-1}) m\rho_i^{m-1} \big) \Big).$$

On the other hand, the norm of a tangent vector $\nabla\psi \in T_\rho(\mathcal{P}^n)$ is given by

$$|\nabla\psi|_g^2 = \frac{q}{n} \sum_i |\psi_i - \psi_{i+1}|^2 \hat{\rho}_{i+1,i}.$$

Now we are ready to obtain estimates for $\mathcal{B}_m(\rho, \nabla\psi)$ and $|\nabla\psi|_g^2$. To this aim, we discern between different cases:

(i) The case $m \in [7/4, 2)$: Choosing $\psi = (0, 1, 2, 3, \ldots 3, 0, 0)$ and $\rho = (A, \alpha A, \alpha A, A, \varepsilon, \ldots \varepsilon)$, $A, B, \varepsilon > 0$ results in (3.25.a) taking the form

$$\partial_1 \theta_m(\alpha A, A)\big(A^m - (\alpha A)^m\big) + \partial_2 \theta_m(\alpha A, A)\big((\alpha A)^m - 2A^m\big) + \frac{1}{2}\big(A^m - (\alpha A)^m\big) + \mathcal{O}(\varepsilon) \quad \text{(3.27.a)}$$

and (3.25.b) becoming

$$2\theta_m(\alpha A, A) m A^{m-1} + \mathcal{O}(\varepsilon). \quad \text{(3.27.b)}$$

Now using the identities

$$\theta_m(\alpha A, A) = \frac{(m-1)}{m} \frac{\alpha^m - 1}{\alpha^{m-1} - 1} A,$$
$$\partial_1 \theta_m(\alpha A, A) = \frac{m-1}{m} \frac{(m-1)\alpha^{m-2} - m\alpha^{m-1} + \alpha^{2(m-1)}}{(\alpha^{m-1} - 1)^2},$$
$$\partial_2 \theta_m(\alpha A, A) = \frac{m-1}{m} \frac{(m-1)\alpha^m - m\alpha^{m-1} + 1}{(\alpha^{m-1} - 1)^2},$$

both (3.27.a) and (3.27.b) simplify to

$$\mathcal{B}_m(\rho,\nabla\psi) = \Big(m\alpha^{m-1} - (m-1)\alpha^{m-2} - \alpha^{2(m-1)}\frac{\alpha^m-2}{\alpha^m-1}\big((m-1)\alpha^m - m\alpha^{m-1}+1\big) - \tag{3.28.a}$$

$$-\frac{m}{2(m-1)}(\alpha^{m-1}-1)^2 + 2m(\alpha^{m-1}-1)\Big)\frac{m-1}{m}\frac{\alpha^m-1}{(\alpha^{m-1}-1)^2}\frac{q^2}{n}A^m + \mathcal{O}(\varepsilon). \tag{3.28.b}$$

In particular, the sign in (3.28) does not depend on the choice of $A$. Indeed, a numerical analysis of (3.28) shows that there exist $m_0 < \frac{7}{4}$ and $\alpha_m > 1$ such that (3.28) attains negative values for every $m > m_0$.
Since

$$|\nabla\psi|_g^2 = \left(\frac{2(m-1)}{m}\frac{\alpha^m-1}{\alpha^{m-1}-1} + \alpha\right)\frac{q}{n}A + \mathcal{O}(\varepsilon)$$

and $n = 2(1+\alpha)A + \mathcal{O}(\varepsilon)$, this means that there exists no $\kappa \in \mathbb{R}$ such that $\mathcal{F}_m^n$ is $\kappa$-convex along geodesics for all $n \geq 7$.

(ii) The case $m \in [2,+\infty)$ : For $m \geq 2$ above argument can be simplified by choosing a vector $\psi = (0,1,2,...2,0)$ and $\rho = (\varepsilon, A, \varepsilon, ...\varepsilon)$. Then

$$\partial_1\theta_m(\varepsilon, A) = \partial_2\theta_m(A,\varepsilon) = \begin{cases} \frac{1}{2} & \text{if } m = 2, \\ \mathcal{O}(\varepsilon) & \text{if m > 2,} \end{cases}$$

implies

$$\mathcal{B}_m(\rho,\nabla\psi) = C_m\frac{m-1}{m}q^2A^{m-1} + \mathcal{O}(\varepsilon) \qquad \text{and} \qquad |\nabla\psi|_g^2 = q + \mathcal{O}(\varepsilon),$$

where $C_m = 1$ for $m = 2$ and $C_m = 2$ for $m > 2$. This implies

$$\kappa \leq -C_m\frac{m-1}{m}qn^{m-1}$$

as a negative upper bound for semi-convexity along geodesics.

(iii) The case $m \in (0,1/4)$: Consider $\psi = (0,1,...1,0,0)$ and $\rho = (A,A,\varepsilon,...\varepsilon)$. Then, noting that $\partial_1\theta_m(A,A) = \partial_2\theta_m(A,A) = \frac{1}{2}$, one immediately obtains

$$\mathcal{B}_m(\rho,\nabla\psi) = \left(2m - \frac{1}{2}\right)q^2A^{m-1} + \mathcal{O}(\varepsilon) \qquad \text{and} \qquad |\nabla\psi|_g^2 = q + \mathcal{O}(\varepsilon).$$

This establishes the following negative upper bound for the semi-convexity along geodesics:

$$\kappa \leq \left(2m - \frac{1}{2}\right)n^{m-1}.$$

◀

## 3.4 Bibliographical Notes

The gradient flow structure for reversible continuous-time Markov chain presented in this chapter was proposed by Maas [47]. In the followup [29] by Erbar and Maas the metric geometry of these discrete gradient flow structures – in particular a synthetic notion of Ricci curvature in the spirit of the works by Lott, Sturm and Villani – was investigated. In [30] the same two authors adapted the notion to discrete entropic gradient flows for discrete porous medium equations. Essentially all the content in this chapter is based on this article – with the exception of Example 3.3.4 which appears here for the first time for cases apart from $m = 2$.

Independently, the essentially same gradient flow structure from an Onsager point of view was discovered in [50] by Mielke. There a gradient system for the discrete Shannon entropy and reversible Markov chain with generator $\mathcal{Q}$ arises from the equation

$$\dot{u} = \mathcal{Q}^\top u = -K(u)\nabla\mathcal{F}_1^n(u). \tag{3.29}$$

Here $\mathcal{F}_1^n$ denotes the standard Euclidean gradient of the discrete entropy functional and $K$ is an corresponding *Onsager matrix*. Accordingly, the metric tensor given by the inverse of $K$ induces a Riemannian structure on the space of discrete probability measures $u \in \mathbb{R}^n$ with nowhere vanishing support. Criteria for geodesic convexity of the entropy functional for this gradient flow structure was were obtained by the same author in [51] as well as by Liero and Mielke [44] and by Erbar and Maas [29].

In the related work [17] a similar gradient flow structure for Fokker-Plank equations on graphs was proposed by Chow et al.

As seen in the recent work [32] by Ferreira, Santos and Valencia-Guevara, the weight function $\theta_m$ associated to the discrete Rényi entropy in this chapter also appears naturally in a weak formulation of the fractional porous medium equation in a periodic setting. In this paper the authors propose a minimising movements scheme related the the Rényi entropy to obtain a gradient flow structure for the fractional porous medium equation on the higher dimensional torus.

Recently, non-local transportation distances related to the discrete entropy gradient flow structures above were considered by Solomon et al. [64] for applications to computational problems of graph theoretical nature.

# 4 Stability Results for Gradient Flows under $\Gamma$-Convergence

In this chapter we return to the general metric setting of Chapter 1 and investigate the stability of gradient flows with respect to perturbations of the underlying functional. To be more precise, assume that $(\phi^n)_{n\in\mathbb{N}}$ is a sequence of functionals $\phi^n : X \to \mathbb{R} \cup \{+\infty\}$, each admitting a gradient flow either in the EDE or EVI metric sense. Then, under suitable conditions, we expect a limit functional $\phi : X \to \mathbb{R} \cup \{+\infty\}$ to inherit the gradient flow property from the $\phi^n$.

## 4.1 $\Gamma$-Convergence of Gradient Flows in the EDE sense

Before we state a stability result for EDE gradient flows, we first need to clarify the meaning of convergence of a sequence $(\phi^n)_{n\in\mathbb{N}}$ of functionals.

**4.1.1 Definition** Let $(X_n, \iota^n)_{n\in\mathbb{N}}$ be a sequence of topological spaces $X_n$ together with (not necessarily continuous) mappings $\iota^n : X_n \to X$ where the codomain of all $\iota^n$ is another topological space $X$. Then a sequence $(\phi^n)_{n\in\mathbb{N}}$ of functionals $\phi^n : X \to \mathbb{R} \cup \{+\infty\}$ is said to be ***sequentially* $\boldsymbol{\Gamma}$-lim inf *convergent*** to a functional $\phi : X \to \mathbb{R} \cup \{+\infty\}$ if for every sequence $(x_n)_{n\in\mathbb{N}}$ of points $x_n \in X_n$ such that $\lim_{n\to\infty} \iota^n(x^n) = x$ for some $x \in \operatorname{dom}\phi$, one has

$$\liminf_{n\to\infty} \phi^n(x_n) \geq \phi(x). \tag{4.1}$$

$(\phi^n)_{n\in\mathbb{N}}$ is said to be ***sequentially* $\boldsymbol{\Gamma}$-lim sup *convergent*** to $\phi : X \to \mathbb{R} \cup \{+\infty\}$ if for every $x \in \operatorname{dom}\phi$ there exists a sequence $(x_n)_{n\in\mathbb{N}}$ of points $x_n \in X_n$ such that $\lim_{n\to\infty} \iota^n(x^n) = x$ and

$$\limsup_{n\to\infty} \phi^n(x_n) \leq \phi(x). \tag{4.2}$$

If $(\phi^n)_{n\in\mathbb{N}}$ is sequentially $\Gamma$-lim inf convergent and sequentially $\Gamma$-lim sup convergent to $\phi$, then $(\phi^n)_{n\in\mathbb{N}}$ is just called ***sequentially* $\boldsymbol{\Gamma}$*-convergent*** to $\phi$.

Typically, one encounters the notation of $\Gamma$-convergence in the situation when $X_n = X$ and $\iota^n$ is the identity mapping for all $n \in \mathbb{N}$. In the context of Gromov-Hausdorff convergence, however, the $\iota^n$ are usually chosen to be $\varepsilon_n$-isometries for some null sequence $(\varepsilon_n)_{n\in\mathbb{N}}$.

Concerning the next result, recall that a slope $|\partial\phi|$ is a *strong upper gradient* for a functional $\phi$ on a metric space $(X, d)$ if for every absolutely continuous curve $v : (a, b) \to X$ the function $|\partial\phi| \circ v$ is Borel measurable and satisfies (1.7).

**4.1.2 Proposition** (Stability of EDE gradient flows) *For every $n \in \mathbb{N}$ let $\iota^n : X_n \to X$ be a mapping between complete metric spaces $(X_n, d_n)$ and $(X, d)$. Let $\tau$ be a topology (not necessarily generated by $d$) on $X$. Let $\phi^n : X_n \to \mathbb{R} \cup \{+\infty\}$ and $\phi : X \to \mathbb{R} \cup \{+\infty\}$ be functionals such that the slopes $|\partial\phi^n|$ and $|\partial\phi|$ are strong upper gradients for all $n \in \mathbb{N}$. Let $(v^n)_{n\in\mathbb{N}}$ be a sequence of locally absolutely continuous curves $v^n : \mathbb{R}^+ \to X_n$, each satisfying* (EDE), *such that $(\iota^n \circ v^n)_{n\in\mathbb{N}}$ is pointwise convergent to a limit curve $v : \mathbb{R}^+ \to X$ with respect to $\tau$.*

*Assume that the following prerequisites are satisfied:*

⇝ *The functionals $(\phi^n)_{n\in\mathbb{N}}$ and $\phi$ satisfy the* $\Gamma$-lim inf *bound*

$$\liminf_{n\to\infty} \phi^n\big(v^n(t)\big) \geq \phi\big(v(t)\big) \qquad \forall t \in \mathbb{R}^+. \tag{4.3}$$

⇝ *The initial value $\lim_{t\searrow 0} v(t) = v_0 \in X$ of the limit curve $v$ exists and satisfies*

$$\limsup_{n\to\infty} \phi^n(v_0^n) \leq \phi(v_0), \tag{4.4}$$

*where $v_0^n = \lim_{t\searrow 0} v^n(t)$ for $n \in \mathbb{N}$;*

⇝ *the metric derivative and the slope satisfy the following lower bound:*

$$\liminf_{n\to\infty} \int_0^t |\dot{v}^n|^2(r) + |\partial\phi^n|^2(v^n(r))\,\mathrm{d}r \geq \int_0^t |\dot{v}^n|^2(r) + |\partial\phi|^2(v(r))\,\mathrm{d}r \qquad \forall t \in \mathbb{R}^+. \tag{4.5}$$

*Then the limit curve $v$ is a gradient flow in the EDE sense and satisfies the following properties:*

- $(v^n)_{n\in\mathbb{N}}$ *is a* recovery sequence *for $v$, i.e.*

$$\lim_{n\to\infty} \phi^n\big(v^n(t)\big) = \phi\big(v(t)\big) \qquad \forall t \in \mathbb{R}^+;$$

- *If the metric derivative and the slope satisfy the reinforced* $\Gamma$-$\liminf$ *bounds*

$$\liminf_{n\to\infty} \int_0^t |\dot{v}^n|^2(r)\,\mathrm{d}r \geq \int_0^t |\dot{v}|^2(r)\,\mathrm{d}r \quad \textit{and} \quad \liminf_{n\to\infty} \int_0^t |\partial\phi^n|^2(v^n(r))\,\mathrm{d}r \geq \int_0^t |\partial\phi|^2(v(r))\,\mathrm{d}r \tag{4.6}$$

*for all $t > 0$, then the slope $|\partial\phi^n| \circ v^n$ and the metric derivative $|\dot{v}^n|$ converge in $L^2_{\mathrm{loc}}(\mathbb{R}^+)$ to their respective limits $|\partial\phi| \circ v$ and $|\dot{v}|$ as $(n \to \infty)$.*

Note that **(4.3)** is actually a weaker assumption than sequential $\Gamma$-$\liminf$ convergence of $(\phi^n)_{n\in\mathbb{N}}$ to $\phi$.

Proof First note that **(4.5)** implies that $|\dot{v}|^2$ is locally integrable and therefore the limit curve $v$ belongs to $AC^2_{\mathrm{loc}}(\mathbb{R}^+, X)$. Moreover, we may invoke **(1.7)** together with the AM-GM inequality to obtain

$$\phi(v(s)) - \phi(v(t)) \leq \int_s^t |\dot{v}|(r)\,|\partial\phi|(v(r))\,\mathrm{d}r \leq \frac{1}{2}\int_s^t |\dot{v}|^2(r) + |\partial\phi|^2(v(r))\,\mathrm{d}r \qquad \forall s,t \in \mathbb{R}^+ : s \leq t. \tag{4.7}$$

We need to show the converse inequality of **(4.7)**, which means precisely that $v$ satisfies (EDI). To this aim, it suffices to show that

$$\phi(v_0) - \phi(v(t)) = \frac{1}{2}\int_0^t |\dot{v}|^2(r) + |\partial\phi|^2(v(r))\,\mathrm{d}r \qquad \forall t \in \mathbb{R}^+. \tag{4.8}$$

Then the fact that $v$ is a solution to (EDE) for arbitrary $s \leq t$ follows by additivity of the integral. To show **(4.8)**, we recall that each $v^n$ satisfies the energy dissipation equality

$$\phi^n(v_0) - \phi^n(v^n(t)) = \frac{1}{2}\int_0^t |\dot{v}^n|^2(r)\,\mathrm{d}r + \frac{1}{2}\int_0^t |\partial\phi^n|^2(v^n(r))\,\mathrm{d}r \qquad \forall t \in \mathbb{R}^+, \tag{4.9}$$

where we already let $(s \searrow 0)$. Now taking the limit inferior as $(n \to \infty)$ in above equality results in the following estimates:
For the left-hand side, we may use **(4.4)** to obtain

$$\phi(v_0) - \limsup_{n\to\infty} \phi^n(v^n(t)) = \liminf_{n\to\infty} \big(\phi^n(v^n_0) - \phi^n(v^n(t))\big). \tag{4.10}$$

Note that **(4.7)** with $(s \searrow 0)$ gives an upper bound for the right-hand side of **(4.5)**, which results in the inequality

$$\liminf_{n\to\infty} \frac{1}{2}\int_0^t |\dot{v}^n|^2(r) + |\partial\phi^n|^2(v^n(r))\,\mathrm{d}r \geq \phi(v_0) - \phi(v(t)).$$

Together with **(4.10)**, this inequality gives us the following estimate as we take the limit inferior $(n \to \infty)$ in **(4.9)**:

$$\phi(v_0) - \limsup_{n\to\infty} \phi^n(v^n(t)) \geq \frac{1}{2}\int_0^t |\dot{v}^n|^2(r) + |\partial\phi^n|^2(v^n(r))\,\mathrm{d}r \geq \phi(v_0) - \phi(v(t)), \tag{4.11}$$

which corresponds to the inequality $\limsup_{n\to\infty} \phi^n(v^n(t)) \leq \phi(v(t))$. Now sequential $\Gamma$-lim inf convergence of $(\phi^n)_{n\in\mathbb{N}}$ comes into play, i.e. we have $\liminf_{n\to\infty} \phi^n(v^n(t)) \geq \phi(v(t))$. Together, both inequalities imply $\lim_{n\to\infty} \phi^n(v^n(t)) = \phi(v(t))$ for all $t \in \mathbb{R}^+$. In particular, there is actually equality in **(4.11)**, which shows that $v$ satisfies (EDE).

It remains to show that $|\dot{v}^n|$ and $|\partial\phi^n| \circ v^n$ converge in $L^2_{\mathrm{loc}}$ to their respective limits as $(n \to \infty)$: Using the energy dissipation equality **(4.9)** and the corresponding equality for the limit curve $v$, we infer that

$$\lim_{n\to\infty} \int_0^t |\dot{v}^n|^2(r) + |\partial\phi^n|^2(v^n(r))\,\mathrm{d}r = \int_0^t |\dot{v}|^2(r) + |\partial\phi|^2(v(r))\,\mathrm{d}r \qquad \forall t \in \mathbb{R}^+. \tag{4.12}$$

Consequently, **(4.6)** implies

$$\begin{aligned}\int_0^t |\dot{v}|^2(r) + |\partial\phi|^2(v(r))\,\mathrm{d}r &\leq \liminf_{n\to\infty} \int_0^t |\dot{v}^n|^2(r)\,\mathrm{d}r + \liminf_{n\to\infty} \int_0^t |\partial\phi^n|^2(v^n(r))\,\mathrm{d}r = \\ &= \lim_{n\to\infty} \int_0^t |\dot{v}^n|^2(r)\,\mathrm{d}r + |\partial\phi^n|^2(v^n(r))\,\mathrm{d}r = \int_0^t |\dot{v}|^2(r) + |\partial\phi|^2(v(r))\,\mathrm{d}r,\end{aligned}$$

which means that

$$\lim_{n\to\infty} \int_0^t |\dot{v}^n|^2(r)\,\mathrm{d}r + |\partial\phi^n|^2(v^n(r))\,\mathrm{d}r = \liminf_{n\to\infty} \int_0^t |\dot{v}^n|^2(r)\,\mathrm{d}r + \liminf_{n\to\infty} \int_0^t |\partial\phi^n|^2(v^n(r))\,\mathrm{d}r \quad \forall t \in \mathbb{R}^+.$$

Therefore, we obtain that

$$\lim_{n\to\infty} \int_0^t |\dot{v}^n|^2(r)\,\mathrm{d}r = \int_0^t |\dot{v}|^2(r)\,\mathrm{d}r \quad \text{and} \quad \lim_{n\to\infty} \int_0^t |\partial\phi^n|^2(v^n(r))\,\mathrm{d}r = \int_0^t |\partial\phi|^2(v(r))\,\mathrm{d}r \qquad \forall t \in \mathbb{R}^+,$$

which in turn implies that $|\dot{v}^n|$ and $|\partial\phi^n| \circ v^n$ converge in $L^2_{\mathrm{loc}}$ to $|\dot{v}|$ and $|\partial\phi| \circ v$, respectively. ■

In the the following example we sketch briefly how above stability result for gradient flows may be applied to a family of Cahn-Hillard equations with increasingly sharp phase transition.

**4.1.3 Example** (Convergence of Cahn-Hillard equations) We consider the the one-dimensional *Cahn-Hillard equation*

$$\dot{u}^\varepsilon = \Delta\big(W'(u^\varepsilon) - \varepsilon^2\Delta u^\varepsilon\big) \qquad \text{in } \mathbb{R}^+ \times \mathbb{T}, \tag{4.13.a}$$

$$u^\varepsilon = u_0 \qquad \text{on } \{0\} \times \mathbb{T}, \tag{4.13.b}$$

where $W(z) := \frac{1}{4}(1-z^2)^2$ is the double-well potential and $\mathbb{T} := \mathbb{R}/\mathbb{Z}$ is the 1-torus. Equation **(4.13)** is used to model phase-separation of two components of a fluid with $\varepsilon \in (0,1]$ corresponding to length of the transition region.

It is possible to identify the solution $u^\varepsilon$ to **(4.13)** as a gradient flow in the *homogeneous negative Sobolev space*

$$\dot{H}^{-1}(\mathbb{T}) := \left\{v \in H^{-1}(\mathbb{T}) : \langle v, 1\rangle = 0\right\},$$

where $\langle\cdot,\cdot\rangle$ denotes the dual pairing between $H^1(\mathbb{T})$ and $H^{-1}(\mathbb{T})$. When endowed with the norm

$$\|v\|_{-1}^2 := \sup_{\varphi \in H^1(\mathbb{T})} \left\{ 2\langle v, \varphi \rangle - \|\nabla \varphi\|_{L^2(\mathbb{T})}^2 \right\} \qquad \forall v \in \dot{H}^{-1}(\mathbb{T}),$$

the space $\dot{H}^{-1}(\mathbb{T})$ becomes a Banach space.

Now (4.13) corresponds to the $\dot{H}^{-1}$-gradient flow for the *Allen-Cahn energy functional*

$$E^\varepsilon(v) := \begin{cases} \int_{\mathbb{T}} \frac{\varepsilon^2}{2} |\nabla v|^2(r) + W(v(r))\, \mathrm{d}r & \text{if } \nabla v \in L^2(\mathbb{T}) \text{ and } W(v) \in L^1(\mathbb{T}), \\ +\infty & \text{otherwise.} \end{cases}$$

Supposed that we are given suitable initial data $u_0 \in \dot{H}^{-1}(\mathbb{T})$ such that $E^\varepsilon(u_0) < +\infty$, the energy dissipation equality, which characterises the $\dot{H}^{-1}$-gradient flow for $E^\varepsilon$, takes the form

$$\frac{1}{2} \int_0^t |\dot{u}^\varepsilon|^2(r)\, \mathrm{d}r + \frac{1}{2} \int_0^t |\partial E^\varepsilon|^2(u^\varepsilon(r))\, \mathrm{d}r = E^\varepsilon(u_0) - E^\varepsilon(u(t)) \qquad \forall t \in \mathbb{R}_0^+, \tag{4.14}$$

where we assume that the curve $u^\varepsilon$ belongs to $AC^2\big(\mathbb{R}_0^+, \dot{H}^{-1}(\mathbb{T})\big)$. Note that (4.14) corresponds to (EDE) by additivity of the integral.

As $(\varepsilon \searrow 0)$, the phase-transition in (4.13) becomes a sharp interface and one would naïvely expect the limit equation (4.13) with $\varepsilon = 0$ to correspond to the internal energy functional

$$E^0(v) = \int_{\mathbb{T}} W(v(r))\, \mathrm{d}r \qquad \text{if } W(v) \in L^1(\mathbb{T}).$$

However, since the double-well potential $W$ is not convex, the functional $E^0$ is not convex nor lower semicontinuous in the $\dot{H}^{-1}$-topology and the gradient flow dynamics is not well-posed (cf. also with Section 2.4 where internal energy functionals are discussed in the context of Wasserstein spaces).

A more sensible alternative is given by the lower semicontinuous envelope of $E^0$ defined as

$$E^{**}(v) := \begin{cases} \int_{\mathbb{T}} W^{**}(v(r))\, \mathrm{d}r & \text{if } W^{**}(v) \in L^1(\mathbb{T}), \\ +\infty & \text{otherwise,} \end{cases}$$

where $W^{**}$ denotes the convex envelope of $W$. It is not difficult to prove that $E^\varepsilon$ is sequentially Γ-convergent to $E^{**}$. For instance, the Γ-lim inf bound follows from the fact that the lower semicontinuity of $E^{**}$, together with $E^\varepsilon \geq E^{**}$, implies

$$\liminf_{\varepsilon \searrow 0} E^\varepsilon(v^\varepsilon) \geq \liminf_{\varepsilon \searrow 0} E^{**}(v^\varepsilon) \geq E^{**}(v)$$

for all $v^\varepsilon$ convergent to some $v$ in $\dot{H}^{-1}(\mathbb{T})$ as $(\varepsilon \searrow 0)$.

Provided that the initial data is well-prepared in the sense that $\lim_{\varepsilon \searrow 0} E^\varepsilon(u_0) = E^{**}(u_0)$, one can invoke the stability result Proposition 4.1.2 to show that the limit curve given by

$$\lim_{\varepsilon \searrow 0} u^\varepsilon = u \qquad \text{in } C\big([0,T], \dot{H}^{-1}(\mathbb{T})\big) \tag{4.15}$$

for all $T > 0$, belongs to $AC^2\big(\mathbb{R}_0^+, \dot{H}^{-1}(\mathbb{T})\big)$ and satisfies the energy dissipation equality

$$\frac{1}{2} \int_0^t |\dot{u}|^2(r)\, \mathrm{d}r + \frac{1}{2} \int_0^t |\partial E^{**}|^2(u(r))\, \mathrm{d}r = E^{**}(u_0) - E^{**}(u(t)) \qquad \forall t \in \mathbb{R}_0^+.$$

Moreover, $u$ is a solution in the sense sense of distributions to

$$\begin{aligned} \dot{u} &= \Delta(\nabla W^{**}(u)) && \text{in } \mathbb{R}^+ \times \mathbb{T}, \\ u &= u_0 && \text{on } \{0\} \times \mathbb{T}. \end{aligned}$$

In order to apply the ideas of Proposition 4.1.2, one has first to establish convergence in the form of (4.15). Then it easy to show that the metric derivatives $|\dot{u}^\varepsilon|$ satisfy (4.5). Indeed, we may use the Banach space structure of $\dot{H}^{-1}(\mathbb{T})$ to write the metric derivative $|\dot{u}^\varepsilon|$ as moduli of the Fréchet derivative $\dot{u}^\varepsilon$ (cf. also Example 1.1.3). Therefore, Fatou's lemma implies the estimate

$$\liminf_{\varepsilon\searrow 0}\int_0^t |\dot{u}^\varepsilon|^2(r)\,\mathrm{d}r = \liminf_{\varepsilon\searrow 0}\int_0^t \|\dot{u}^\varepsilon(r)\|_{-1}^2\,\mathrm{d}r \geq \int_0^t \|\dot{u}(r)\|_{-1}^2\,\mathrm{d}r = \int_0^t |\dot{u}|^2(r)\,\mathrm{d}r \qquad \forall t\in\mathbb{R}^+.$$

However, considerable work is involved to prove that the slope $|\partial E^\varepsilon|\circ u^\varepsilon$ satisfies the lim inf-bound of (4.6). We omit the details. ◀

## 4.2 $\Gamma$-Convergence of Gradient Flows in the EVI sense

Usually, the existence of a strong upper gradient required to apply Proposition 4.1.2 is established via geodesic $\kappa$-convexity of the functional (see Proposition 1.2.4). In this $\kappa$-convex setting, however, the stronger notion of a gradient flow in EVI sense may be applicable.
In the following stability result for EVI gradient flows the specific boundedness assumptions (4.3) and (4.4) are replaced by the stronger concepts of Gromov-Hausdorff convergence of metric spaces and $\Gamma$-convergence of the functionals.

For convenience, we introduce Gromov-Hausdorff convergence via the notion of $\varepsilon$-isometries.

**4.2.1 Definition** Let $(X,d_X)$ and $(Y,d_Y)$ be metric spaces. A mapping $\iota: Y\to X$ is called ***$\varepsilon$-isometry*** for $\varepsilon>0$ if

$$\big|d_X\big(\iota(y_1),\iota(y_2)\big)-d_Y(y_1,y_2)\big|\leq\varepsilon \qquad \forall y_1,y_2\in Y,$$

and for every $x\in X$ there exists a point $y\in Y$ such that $d_X(\iota(y),x)\leq\varepsilon$.
A sequence$(X^n)_{n\in\mathbb{N}}$ of compact metric spaces $X_n$ is called ***convergent in the Gromov-Hausdorff sense*** to a compact metric space $X$ if there exist $\varepsilon_n$-isometries $\iota^n: X_n\to X$ for some null sequence $(\varepsilon_n)_{n\in\mathbb{N}}$.

**4.2.2 Proposition** (Stability of EVI gradient flows) *Let $(X_n,d_n)_{n\in\mathbb{N}}$ be s a sequence of complete metric spaces, converging in the Gromov-Hausdorff sense to a complete metric space $(X,d)$, i.e. there exists a sequence $(\iota^n)_{n\in\mathbb{N}}$ of $\varepsilon_n$-isometries $\iota_n: X_n\to X$ for some null sequence $(\varepsilon_n)_{n\in\mathbb{N}}$. Let $(\phi^n)_{n\in\mathbb{N}}$ be a sequence of functionals $\phi^n: X_n\to\mathbb{R}\cup\{+\infty\}$, sequentially $\Gamma$-convergent to some $\phi: X\to\mathbb{R}\cup\{+\infty\}$. For a fixed constant $\kappa\in\mathbb{R}$, let $(v^n)_{n\in\mathbb{N}}$ be a sequence of locally absolutely continuous curves $v^n:\mathbb{R}^+\to X_n$, each satisfying* (EVI$_\kappa$)*, such that $(\iota^n\circ v^n)_{n\in\mathbb{N}}$ is pointwise convergent to a limit curve $v:\mathbb{R}^+\to X$ with respect to $d$. Assume that the initial values $\lim_{t\searrow 0}v^n(t)=v_0^n\in\overline{\operatorname{dom}\phi^n}$ satisfy $\lim_{n\to\infty}\iota^n(v_0^n)=v_0\in\overline{\operatorname{dom}\phi}$.*
*Then the limit curve $v$ is a gradient flow in the EVI sense. In addition, $(v^n)_{n\in\mathbb{N}}$ is a recovery sequence for $v$, i.e.*

$$\lim_{n\to\infty}\phi^n(v^n(t))=\phi(v(t)) \qquad \forall t\in\mathbb{R}^+. \tag{4.16}$$

Proof The proof is arranged into two steps: First, we will check that $v$ satisfies the EVI. In the second step, we will prove that $(v^n)_{n\in\mathbb{N}}$ is actually a recovery sequence for $v$.
Throughout the proof we will use the shorthand notations $v_\iota^n := \iota^n\circ v^n$ and $x_\iota^n := \iota^n\circ x^n$.

(i) Identification of the limit curve: Fix a point $x\in\operatorname{dom}\phi$. Then the sequential $\Gamma$-convergence of $(\phi^n)_{n\in\mathbb{N}}$ implies that there exists a sequence $(x_\iota^n)_{n\in\mathbb{N}}$ which converges to $x$ such that $\lim_{n\to\infty}\phi^n(x^n)=\phi(x)$. We invoke the alternative characterisation of the EVI as given in Lemma 1.3.1: In particular, choosing $y=x^n$ results in

$$\frac{1}{2}e^{\kappa t}d_n^2\big(v^n(t),x^n\big) - \frac{1}{2}e^{\kappa s}d_n^2\big(v^n(s),x^n\big) \le \big((\phi^n(x^n) - \phi^n\big(v^n(t)\big)\big)\int_s^t e^{\kappa r}\,\mathrm{d}r \qquad \forall s,t \in \mathbb{R}^+ : s \le t.$$

Using the fact that each $\iota^n$ is an $\varepsilon$-isometry, above inequality yields

$$\frac{1}{2}e^{\kappa t}d^2\big(v_\iota^n(t),x_\iota^n\big) - \frac{1}{2}e^{\kappa s}d^2\big(v_\iota^n(s),x_\iota^n\big) \le \big(\phi^n(x^n) - \phi^n(v^n(t))\big)\int_s^t e^{\kappa r}\,\mathrm{d}r + \eta_t^n \quad \forall s,t \in \mathbb{R}^+ : s \le t. \tag{4.17}$$

for some $\eta_t^n > 0$ such that $\lim\limits_{n\to\infty}\eta_t^n = 0$ for all $t > 0$. Recall that the Γ-convergence of $(\phi^n)_{n\in\mathbb{N}}$ implies

$$\liminf_{n\to\infty}\phi^n(v^n(t)) \ge \phi(v(t)) \quad \forall t \in \mathbb{R}^+.$$

Therefore, we can pass to the limit inferior in **(4.17)** as $(n \to \infty)$ to obtain

$$\frac{1}{2}e^{\kappa t}d^2(v(t),x) - \frac{1}{2}e^{\kappa s}d^2(v(s),x) \le \big(\phi(x) - \phi(v(t))\big)\int_s^t e^{\kappa r}\,\mathrm{d}r \qquad \forall x \in \operatorname{dom}\phi,\ \forall s,t \in \mathbb{R}^+ : s \le t. \tag{4.18}$$

Thus the limit curve $v$ belongs to $AC_{\mathrm{loc}}(\mathbb{R}^+,X)$ and is a solution to the EVI with respect to $\kappa$. For $v$ to satisfy ($\mathrm{EVI}_\kappa$), it only remains to show that $\lim_{t\searrow 0} v(t) = v_0$. However, this follows readily if we let **(4.17)** first attain $(s \searrow 0)$, and afterwards pass to the limit inferior as $(n \to \infty)$; thus, arriving at

$$\frac{1}{2}e^{\kappa t}d^2(v(t),x) - \frac{1}{2}d^2(v_0,x) \le \big(\phi(x) - \phi(v(t))\big)\int_0^t e^{\kappa r}\,\mathrm{d}r \qquad \forall x \in \operatorname{dom}\phi,\ \forall t \in \mathbb{R}^+. \tag{4.19}$$

Now we can pass to the limit superior in **(4.19)** as $(t \searrow 0)$, where we use the lower semicontinuity of $\phi$ to arrive at

$$\limsup_{t\searrow 0} d^2\big(v(t),x\big) \le d^2(v_0,x) \qquad \forall x \in \operatorname{dom}\phi.$$

Letting $(x \to v_0)$ in this inequality, shows that $v_0$ is the initial datum for the limit curve $v$.

(ii) Establishing the recovery sequence: The sequential Γ-convergence of $(\phi^n)_{n\in\mathbb{N}}$ yields for every fixed time $t > 0$ a sequence $(\overline{v}_t^n)_{n\in\mathbb{N}}$ such that $(\iota^n(\overline{v}_t^n)$ is convergent to $v(t)$ and $\lim_{n\to\infty}\phi^n(\overline{v}_t^n) = \phi\big(v(t)\big)$. Therefore, setting $y = \overline{v}_t^n$, the pointwise variant of the EVI, which was derived in **(1.17)**, takes the form

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}d_n^2\big(v^n(t),\overline{v}_t^n\big) \le \phi^n(\overline{v}_t^n) - \phi^n\big(v^n(t)\big) - \frac{\kappa}{2}d_n^2\big(v^n(t),\overline{v}_t^n\big) \qquad \forall n \in \mathbb{N},\ \forall t \in \mathbb{R}^+. \tag{4.20}$$

To take a meaningful limit for the inequality above, we first need to ensure that the left-hand side is bounded from below by 0 as $(t \searrow 0)$. To this aim, we use the reverse triangle inequality to obtain

$$\frac{d_n\big(v^n(t+h),z\big) - d_n\big(v^n(t),z\big)}{h} \ge -\frac{d_n\big(v^n(t+h),v^n(t)\big)}{h} \qquad \forall z \in X,\ \forall t \in \mathbb{R}^+.$$

for all $h > 0$. Hence, by passing to the limit superior in the inequality above as $(h \searrow 0)$, we infer that

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}d_n^2\big(v^n(t),z\big) = d_n\big(v^n(t),z\big)\frac{\mathrm{d}^+}{\mathrm{d}t}d_n\big(v^n(t),z\big) \ge -d_n\big(v^n(t),z\big)\,|\dot{v}^n|(t+) \qquad \forall z \in X,\ \forall t \in \mathbb{R}^+, \tag{4.21}$$

where we used the product rule in the first equality above.

The energy identity of Theorem 1.3.2.iii, together with (4.28), implies that $|\dot{v}^n|(t+) \le C_2(S,T)$ for all $n \in \mathbb{N}$ $t \in [S,T]$. Therefore, we can set again $z = \overline{v}_t^n$ in (4.21) which, together with (4.20), implies the inequality

$$\phi^n(v^n(t)) \le \phi^n(\overline{v}_t^n) - \frac{\kappa}{2} d_n^2\big(v^n(t), \overline{v}_t^n\big) + C_2(S,T) d_n\big(v^n(t), \overline{v}_t^n\big) \qquad \forall n \in \mathbb{N},\ \forall t \in [S,T].$$

Utilizing the $\varepsilon_n$-isometries $\iota^n$, we can replace the metric $d^n$ in above inequality by $d$ and a small error:

$$\phi^n((v^n(t)) \le \phi^n(\overline{v}_t^n) + \frac{|\kappa|}{2}\big(d(v_\iota^n(t), \iota(\overline{v}_t^n)) + \varepsilon_n\big) + C_2(S,T)\big(d(v_\iota^n(t), \iota(\overline{v}_t^n)) + \varepsilon_n\big)^2.$$

Now, passing to the limit superior as $(n \to \infty)$, yields the estimate

$$\limsup_{n\to\infty} \phi(v^n(t)) \le \phi(v(t)) \qquad \forall t \in \mathbb{R}^+.$$

Since the sequential $\Gamma$-convergence of $(\phi^n)_{n\in\mathbb{N}}$ gives us the converse inequality

$$\phi(v(t)) \le \liminf_{n\to\infty} \phi(v^n(t)) \qquad \forall t \in \mathbb{R}^+,$$

we finally arrive at (4.16). ■

In the result above we required $(v^n)_{n\in\mathbb{N}}$ to be pointwise convergent to a limit curve. This requirement is unneeded, however, when all functionals $\phi^n$ are coercive. We will see in the following corollary that in such a case ($\mathrm{EVI}_\kappa$) actually provides uniform bounds for $(v^n)_{n\in\mathbb{N}}$ which are strong enough to deduce the existence of a limit curve by means of an Arzelà-Ascoli argument.

**4.2.3 Definition** Let $(X,d)$ be a metric space. We say that a functional $\phi : X \to \mathbb{R} \cup \{+\infty\}$ is ***coercive*** if all sublevel sets $\{x \in X : \phi(x) \le c\}$, $c \in \mathbb{R}$ are relative compact.

To keep the formulation of the following result simple, we restrict ourselves to the situation when $X^n = X$ and $\iota^n$ is the identity mapping for all $n \in \mathbb{N}$.

**4.2.4 Corollary** *Let $(X,d)$ be a complete and separable metric space and fix $\kappa \in \mathbb{R}$. Let $(\phi^n)_{n\in\mathbb{N}}$ be a sequence of coercive functionals $\phi^n : X \to \mathbb{R} \cup \{+\infty\}$, sequentially $\Gamma$-convergent to $\phi : X \to \mathbb{R} \cup \{+\infty\}$. Assume that the $\phi^n$ are uniformly bounded from below, i.e. there exists a constant $C \in \mathbb{R}$ such that*

$$\phi^n(x) \ge C \qquad \forall x \in X,\ \forall n \in \mathbb{N}. \tag{4.22}$$

*Let $(v^n)_{n\in\mathbb{N}}$ be a sequence of locally absolutely continuous curves $v^n : \mathbb{R}^+ \to X$, each satisfying* ($\mathrm{EVI}_\kappa$)*. Assume that the initial values $\lim_{t\searrow 0} v^n(t) = v_0^n \in \overline{\operatorname{dom}\phi^n}$ converge to $v_0 \in \overline{\operatorname{dom}\phi}$. Then there exists a limit curve $v : \mathbb{R}^+ \to X$, starting from $\lim_{t\searrow 0} v(t) = v_0$, which is a gradient flow in the EVI sense. Additionally, $(v^n)_{n\in\mathbb{N}}$ is a recovery sequence for $v$, i.e.* (4.16) *holds.*

$$\lim_{n\to\infty} \phi^n(v^n(t)) = \phi(v(t)) \qquad \forall t \in \mathbb{R}^+. \tag{4.23}$$

For the proof we will need the following version of the Arzelà-Ascoli theorem.

**4.2.5 Lemma** (Metric variant of the Arzelà-Ascoli theorem) *Let $(X,d)$ be a metric space and assume that $\omega : [S,T] \times [S,T] \to \mathbb{R}_0^+$ is a symmetric function which satisfies*

$$\lim_{(r,s)\to(t,t)} \omega(r,s) = 0 \qquad \forall t \in [S,T].$$

*Let $(\nu^n)_{n\in\mathbb{N}}$ be a sequence of curves $\nu^n : [S,T] \to X$ such that for every $t \in [S,T]$ the set $\{\nu^n(t) : n \in \mathbb{N}\}$ is relatively compact in $X$. Then following statements hold:*

(i) *If $(\nu^n)_{n\in\mathbb{N}}$ satisfies*

$$\limsup_{n\to\infty} d(\nu^n(s),\nu^n(t)) \le \omega(s,t) \qquad \forall s,t\in[S,T], \tag{4.24}$$

*then there exists a subsequence $(\nu_{n^k})_{k\in\mathbb{N}}$, converging pointwise to a continuous limit curve $\nu:[S,T]\to X$ such that $d(\nu(s),\nu(t))\le\omega(s,t)$ for all $s,t\in[S,T]$.*

(ii) *If $(\nu^n)_{n\in\mathbb{N}}$ satisfies*

$$\sup_{n\in\mathbb{N}} d(\nu^n(s),\nu^n(t)) \le \omega(s,t) \qquad \forall s,t\in[S,T], \tag{4.25}$$

*then there exists a subsequence $(\nu^{n_k})_{k\in\mathbb{N}}$, converging uniformly to a continuous limit curve $\nu:[S,T]\to X$ such that $d\big(\nu(s),\nu(t)\big)\le\omega(s,t)$ for all $s,t\in[S,T]$.*

Proof of Corollary 4.2.4 We will only prove the corollary for $\kappa\le 0$ and $C\ge 0$.
In the first part of the proof we show the existence of the limit curve $v$. In order to apply Lemma 4.2.5.ii, we need to establish uniform bounds for $(v^n)_{n\in\mathbb{N}}$: Fix a point $x\in\operatorname{dom}\phi$. Then the sequential $\Gamma$-convergence of $(\phi^n)_{n\in\mathbb{N}}$ that there exists a sequence $(x_n)_{n\in\mathbb{N}}$ which converges to $x$ such that $\lim_{n\to\infty}\phi^n(x_n)=\phi(x)$. We invoke the alternative characterisation of the EVI as given in Lemma 1.3.1. Choosing $y=x_n$ results in

$$\frac{1}{2}e^{\kappa t}d^2\big(v^n(t),x_n\big)-\frac{1}{2}e^{\kappa s}d^2\big(v^n(s),x_n\big)\le\big(\phi(x_n)-\phi(v^n(t))\big)\int_s^t e^{\kappa r}\,\mathrm{d}r \qquad \forall s,t\in\mathbb{R}^+: s\le t. \tag{4.26}$$

In particular, this implies the estimate

$$d^2\big(v^n(t),x_n\big)\le\Big(d^2(v_0^n,x_n)+2\phi^n(x_n)\int_0^t e^{\kappa r}\,\mathrm{d}r\Big)e^{-\kappa t} \qquad \forall n\in\mathbb{N},\ \forall t\in\mathbb{R}^+.$$

Since $(v_0^n)_{n\in\mathbb{N}}$ and $\big(\phi^n(x_n)\big)_{n\in\mathbb{N}}$ are convergent, for every positive time $T>0$ there exists a constant $C_1(T)>0$, independent of $n$, such that

$$d\big(v^n(t),x_n\big)\le C_1(T) \qquad \forall n\in\mathbb{N},\ \forall t\in[0,T]. \tag{4.27}$$

Now the regularising estimate of Theorem 1.3.2.ii comes into play: Choosing again $y=x_n$ in (1.18) yields

$$\begin{aligned}\frac{1}{2}e^{\kappa t}d^2\big(v^n(t),x_n\big)+\phi\big(v^n(t)\big)\int_0^t e^{\kappa r}\,\mathrm{d}r+\frac{1}{2}\,|\partial\phi|^2(v^n(t))\Big(\int_0^t e^{\kappa r}\,\mathrm{d}r\Big)^2\le\\ \le\frac{1}{2}d^2\big(v_0,x_n\big)+\phi(x_n)\int_0^t e^{\kappa r}\,\mathrm{d}r\le\frac{1}{2}C_1(T)+\phi(x_n)\int_0^t e^{\kappa r}\,\mathrm{d}r\end{aligned}$$

for all $n\in\mathbb{N}$ and all $t\in[0,T]$. As before, the convergence of $\big(\phi^n(x_n)\big)_{n\in\mathbb{N}}$ implies that for every choice of $0<S<T$ there exists a constant $C_2(S,T)>0$ such that

$$\phi^n(v^n(t))\le C_2(S,T) \quad\text{and}\quad |\partial\phi^n|(v^n(t))\le C_2(S,T) \qquad \forall t\in[0,T]. \tag{4.28}$$

In particular, the coercivity of each $\phi^n$, together with the first bound in (4.28), shows that the set $v^n(t):n\in\mathbb{N}$ is relative compact in $X$ for every time $t>0$.
Furthermore, the energy identity (1.19), together with the second bound in (4.28) results in

$$d\big(v^n(s),v^n(t)\big)\le\int_s^t|\dot{v}^n|(r)\,\mathrm{d}r=\int_s^t|\partial\phi^n|\,(v^n(r))\,\mathrm{d}r\le C_2(S,T)\,|t-s| \qquad \forall s,t\in[S,T]: s\le t$$

for all $n \in \mathbb{N}$. Hence, $(v^n)_{n\in\mathbb{N}}$ is uniformly Lipschitz on each compact interval $[S,T] \subset \mathbb{R}^+$. Hence, we may apply Lemma 4.2.5.ii with $\omega(s,t) = C_3(T)\,|t-s|$ to obtain a subsequence $(v^{n_k})_{k\in\mathbb{N}}$, converging locally uniformly to a limit curve $v : \mathbb{R}^+ \to X$ which is locally Lipschitz.

Now we are in the position to invoke Proposition 4.2.2 to infer that $v$ is a gradient flow in the EVI sense. Moreover, the $\kappa$-contraction property given in Theorem 1.3.2.i implies that the the whole sequence $(v^n)_{n\in\mathbb{N}}$ converges locally uniformly to $u$ as $(n \to \infty)$.

Finally, we apply Proposition 4.2.2 once again to establish a recovery sequence as in (4.16). ∎

## 4.3 Bibliographical Notes

The EDE stability result Proposition 4.1.2 is due to Serfaty [63]. In [60] the same author together with Sandier had already applied this framework in a Hilbert space setting to the Ginzburg-Landau heat flow. A stability result, closely related to the one in [63], was already obtained by Ortner [53]. The application to Cahn-Hillard equations, outlined in Example 4.1.3 also taken from [63], was done in a formal manner by Le [41]. Independently, Gigli used an abstract Γ-convergence result in the same spirit to to study the heat flow on compact metric measure spaces in [34].

The EVI stability result Proposition 4.2.2 is an adaptation of an Γ-convergence result by Daneri and Savaré [20] (stated there in the form of Corollary 4.2.4) to an abstract Gromov-Hausdorff framework. For a comprehensive view on Γ-convergence and its application, we refer to Braides' survey [14].

For a proof of the metric varant of the Arzelà-Ascoli theorem (Lemma 4.2.5) see section 3.3 in [4] by Ambrosio, Gigli and Savaré and chapter 27 of Villani's monograph [71]. In the same chapter of the latter reference a thorough discussion of the notion of Gromov-Hausdorff convergence may be found as well.

# 5 Limit Passage of EDE Gradient Flows for Discrete Porous Medium Equations

## 5.1 Finite-Volume Discretisation

**Notation**

In this chapter we will slightly abuse the notation in the following way: Any probability measure $\mu \in \mathcal{P}_2(\Omega)$ which is absolutely continuous with respect to the Lebesgue measure $\mathcal{L}^1$, will be identified with its density $U$. For instance, we will also write $F_m(U)$ instead of $F_m(\mu)$ for the continuous Rényi entropy functional $F_m$.

Likewise, we will use a similar notation for functionals which are defined on $\mathcal{P}^n$: Every discrete probability measure $u_i^n = \pi_i \rho_i \in \mathcal{P}^n$ will be identified with its density $\rho_i$ and instead of e.g. $\mathcal{F}_m^n(\rho)$, we will also write $\mathcal{F}_m^n(u^n)$.

In addition, we will drop the dependence of functions on the spatial variable $x$ or the time variable $t$ to keep equations neat.

In this section we will introduce a discretisation for the continuous one-dimensional porous medium equation with drift

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t,x) = \Delta u^m(t,x) + \mathrm{div}\big(u(t,x)\nabla\upsilon(x)\big) \qquad \forall (t,x) \in (0,T) \times [0,1],\ m > 0. \tag{5.1}$$

subject to the non-flux Neumann boundary condition

$$\nabla u^m(t,0) + u(t,0)\nabla\upsilon(0) = \nabla u^m(t,1) + u(t,1)\nabla\upsilon(1) = 0 \qquad \forall t \in (0,T).$$

Note that we encountered this type of equation already in the more general form of (2.18). Here, however, the internal energy functional is given by the Rényi entropy introduced in Definition 2.4.11.

In the first part of this chapter we introduce a finite volume scheme for (5.1), which then is recognised as a gradient flow for the discrete Rényi entropy.

5.1.1 **Definition** (Finite volume scheme) Given a partition $0 = x_1^n < x_2^n < \ldots < x_n^n = 1$ of $\Omega := (0,1)$, define $n+1$ midpoints between the $x_i^n$ via

$$\sigma_0^n := 0 \qquad \sigma_i^n := \frac{x_{i+1}^n + x_i^n}{2} \text{ for } i \in \{1, \ldots n-1\}, \qquad \sigma_n^n := 1.$$

Moreover, we introduce ***control volumes*** $\omega_i^n := [\sigma_{i-1}^n, \sigma_i^n)$ and denote their length by $h_i^n$.

Integrating the porous medium equation (5.1) over a control volume $\omega_i^n$ results in

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\omega_i^n} u(t,x)\,\mathrm{d}x = \nabla(u^m)(t,\sigma_i^n) - \nabla(u^m)(t,\sigma_{i-1}^n) \ + \ u(t,\sigma_i^n)\nabla\upsilon(\sigma_i^n) - u(t,\sigma_{i-1}^n)\nabla\upsilon(\sigma_{i-1}^n)$$

for all $i \in \{1, \ldots n-1\}$. In the finite-volume scheme the integral $\int_{\omega_i^n} u(t,x)\,\mathrm{d}x$ is approximated by $u_i^n(t)$ and the ***flux*** $\nabla(u^m)(t,\sigma_i^n)$ is to be approximated by by finite differences.

Introduce the ***rate coefficients***

$$\alpha_i := \frac{1}{(x_{i+1}^n - x_i^n)h_i^n} \qquad \text{and} \qquad \beta_i := \frac{1}{(x_{i+1}^n - x_i^n)h_{i+1}^n} \qquad \forall i \in \{1, \ldots n-1\}, \tag{5.2}$$

in the interior and $\alpha_0 = \beta_0 = \alpha_n = \beta_n := 0$ at the boundary. Then, provided that we change over from $u_i^n : (0,T) \to \mathbb{R}$ to the density curves $\rho_i := u_i^n/h_i^n : (0,T) \to \mathbb{R}$, we may write the ***finite volume discretisation*** of the porous medium equation with drift in the form

$$\dot{\rho}_i(t) = \alpha_{i-1}\rho_{i-1}^m(t) - (\alpha_i + \beta_{i-1})\rho_i^m(t) + \beta_i\rho_{i+1}^m(t) + \tag{5.3.a}$$

$$+ \alpha_{i-1}\hat{\rho}_{i-1,i}(t)(\upsilon_{i-1} - \upsilon_i) + \beta_i\hat{\rho}_{i,i+1}(t)(_{i+1} - \upsilon_i) \tag{5.3.b}$$

for all $i \in \{1, \dots n\}$ and $t \in (0, T)$, where $\hat{\rho}_{i,j}$ denotes the weight $\theta_m$ between $\rho_i$ and $\rho_j$ as already defined in (3.1).

The rate coefficients give rise to a birth-death process, defined by the infinitesimal generator

$$\mathcal{Q} := \begin{pmatrix} -\alpha_1 & \alpha_1 & 0 & 0 & \dots & 0 \\ \beta_1 & -\alpha_2 - \beta_1 & \alpha_2 & 0 & \dots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \dots & 0 & \beta_{n-2} & -\alpha_{n-1} - \beta_{n-2} & \alpha_{n-1} \\ 0 & \dots & 0 & 0 & \beta_{n-1} & -\beta_{n-1} \end{pmatrix}. \tag{5.4}$$

Using the notation from Section 3.2, we may use the particular structure of $\mathcal{Q}$ to write (5.3) in the compact form

$$\dot{\rho}_i(t) = \sum_{j=1}^{n} \mathcal{Q}_{ji}(\rho_j^m(t) + \hat{\rho}_{ij}(t)\,\nabla\upsilon_{ij}) \qquad \forall t \in (0, T).$$

The stationary distribution of this continuous-time Markov chain is characterised by the equation $\pi\mathcal{Q} = 0$, which can be also written more concisely as $\alpha_i\pi_i = \beta_i\pi_{i+1}$; thus is given by $\pi_i = h_i^n$.

The next step on the towards a gradient flow which is a solution to (5.1) is as follows: We need to relate the spatial finite volume discretisation described above to the gradient flow for a linear combination of two functionals introduced in Section 3.2: the discrete Rényi entropy functional $\mathcal{F}_m^n$ and the discrete potential energy functional $\mathcal{V}^n$. Basically, the existence of such a gradient flow was already shown in Proposition 3.2.3. However, in order to obtain a global flow, we either need to assume that $m \leq 1$ or that the potential energy functional $\mathcal{V}^n$ vanishes.

**5.1.2 Assumptions** We make the following assumptions on the porous medium equation in (5.1):

(i) Re Rényi entropy and potential energy functionals: The bound $0 < m \leq 3/2$ is satisfied. In case $m > 1$, we additionally require the porous medium equation to be homogeneous, i.e. $\nabla\upsilon \equiv 0$. Otherwise, $\upsilon$ is required to be a bounded convex function in $C^1(\overline{\Omega})$. Then the discretisation of $\upsilon$ is given by

$$\upsilon_i^n = \frac{1}{h_i^n} \int_{\omega_i^n} \upsilon(x)\,\mathrm{d}x. \tag{5.5}$$

(ii) Re initial condition: Let $U_0$ be a probability density with respect to the Lebesgue measure on $\Omega$ such that $U_0 \in L^m(\Omega)$ for $m \neq 1$ or $U_0 \log U_0 \in L^1(\Omega)$ for $m = 1$. Then the initial value of the finite-volume scheme is given by

$$u_i^n(0) = \int_{\omega_i^n} U_0(x)\,\mathrm{d}x. \tag{5.6}$$

Note that the Riemannian structure on $\operatorname{int}\mathcal{X}^n$ introduced in Proposition 3.1.4 in general does not extend to boundary of $\mathcal{X}^n$. However, as already noticed in the proof of Proposition 3.2.3, Assumptions 5.1.2 assures that $\rho(t)$ stays inside $\operatorname{int}\mathcal{P}^n$ for all times $t > 0$. Hence, we have the following result.

**5.1.3 Proposition** *Let $(u_i^n)_{1 \leq i \leq n}$ be the spatial finite volume discretisation given by* (5.3) *such that* Assumptions 5.1.2 *are satisfied. Let $(\mathcal{X}^n, \mathcal{Q}, h_i^n)$ be the irreducible continuous-time Markov chain corresponding to* (5.4)*. Then $i \mapsto u_i^n(t)$ is a discrete probability measure on $\mathcal{X}^n$ for any time $t > 0$, which induces a gradient flow*

*$\rho : (0,T) \to \operatorname{int}\mathcal{P}^n$ in the Riemannian sense for the mixed energy functional $\Phi^n = \mathcal{F}_m^n + \mathcal{V}^n$ via the relation $u_i = h_i^n \rho_i$, starting from $\lim_{t\searrow 0} \rho_i(t) = u_i^n(0)/h_i^n$.*

*In particular, $\rho$ is continuously differentiable and satisfies the energy dissipation equality (EDE)*

$$\frac{1}{2}\int_s^t |\rho'(r)|_g^2 \,\mathrm{d}r + \frac{1}{2}\int_s^t |\partial\Phi^n|^2\big(\rho(r)\big)\,\mathrm{d}r = \Phi^n\big(\rho(s)\big) - \Phi^n\big(\rho(t)\big) \qquad \forall s,t \in (0,T) : s \le t. \tag{5.7}$$

Proof The first part of this proposition is an immediate consequence of Proposition 3.2.3. The energy dissipation equality (5.7) follows due to Proposition 1.2.5. ■

The particularly simple structure of $\mathcal{Q}$ allows to give an explicit characterisation of the velocity of a smooth curve.

**5.1.4 Lemma** *Under Assumptions 5.1.2, let $\rho : (0,T) \to \operatorname{int}\mathcal{P}^n$ be the smooth gradient flow induced by the finite volume discretisation $u^n$ via the relation $u_i^n = h_i^n \rho_i$. Then the velocity $\dot\rho$ may be identified via the discrete continuity equation* (3.7) *with the discrete vector field $\nabla\psi : (0,T) \to \mathcal{X}^n \times \mathcal{X}^n$ given by*

$$\nabla\psi_{i,j} = \psi_j - \psi_i = \sum_{k=j}^{i-1} \frac{q_k}{\alpha_k h_k^n \hat\rho_{k+1,k}} \qquad \text{with} \qquad q_k := \sum_{l=1}^{k} \dot u_l^n, \qquad \text{for} \qquad j < i. \tag{5.8}$$

Proof For our choice of the infinitesimal generator $\mathcal{Q}$ the continuity equation in the interior of $X$ reads

$$\dot\rho_i = \big(\psi_i - \psi_{i-1}\big)\hat\rho_{i,i-1}\beta_{i-1} + \big(\psi_i - \psi_{i+1}\big)\hat\rho_{i,i+1}\alpha_i \qquad \forall i \in \{2, \ldots n-1\}, \tag{5.9}$$

while at the boundary we have

$$\dot\rho_1 = \big(\psi_1 - \psi_2\big)\hat\rho_{1,2}\alpha_1 \qquad \text{and} \qquad \dot\rho_n = \big(\psi_n - \psi_{n-1}\big)\hat\rho_{n,n-1}\beta_{n-1}.$$

Solving for $\psi_i - \psi_{i+1}$ in (5.9), we arrive at

$$\psi_i - \psi_{i+1} = \frac{1}{\hat\rho_{i,i+1}} \sum_{k=1}^{i} \frac{\dot\rho_k}{\alpha_k} \prod_{l=k+1}^{i} \frac{\beta_{l-1}}{\alpha_l} \qquad \forall i \in \{1, \ldots n-1\}. \tag{5.10}$$

Invoking the detailed balance equation $\alpha_i h_i^n = \beta_i h_{i+1}^n$, (5.10) may be also written in the form

$$\psi_i - \psi_{i+1} = \frac{1}{\hat\rho_{i,i+1}} \sum_{k=1}^{i} \frac{\dot u_k}{\alpha_k h_k^n} \prod_{l=k+1}^{i} \frac{\beta_{l-1}}{\alpha_l} = \frac{1}{\hat\rho_{i,i+1}} \sum_{k=1}^{i} \frac{\dot u_k}{\alpha_i h_{k+1}^n} \prod_{l=k+1}^{i-1} \frac{\beta_l}{\alpha_l} = \frac{1}{\hat\rho_{i,i+1}} \sum_{k=1}^{i} \frac{\dot u_k}{\alpha_i h_i^n}. \tag{5.11}$$

Now, for $i < j$ the general expression $\psi_i - \psi_j$ arises just as telescopic sum over (5.11). ■

Now we are ready to introduce the following functionals:

**5.1.5 Definition** Let $\rho : (0,T) \to \operatorname{int}\mathcal{P}^n$ be the smooth gradient flow induced by the finite volume discretisation $u^n$ via the relation $u_i^n = h_i^n \rho_i$. Then the ***discrete dissipation potential*** is defined as

$$\mathcal{R}^n(u^n, \dot u^n) := \frac{1}{2}\langle \dot u^n, \dot u^n \rangle_g = \frac{1}{2} \sum_{i=1}^{n-1} \frac{q_i^2}{\alpha_i h_i^n \hat\rho_{i,i+1}} \qquad \text{with} \qquad q_i := \sum_{l=1}^{i} \dot u_l^n,$$

whereas the ***(generalised) discrete Fisher information*** of the functional $\Phi^n = \mathcal{F}_m^n + \mathcal{V}^n$ is given by

$$\mathcal{I}^n(u^n) := \frac{1}{2}\langle \operatorname{grad}_g \Phi^n(u^n), \operatorname{grad}_g \Phi^n(u^n) \rangle_g = \frac{1}{2} \sum_{i=1}^{n-1} \frac{\hat\rho_{i+1,i}}{x_{i+1}^n - x_i^n} \left| \frac{\rho_{i+1}^m - \rho_i^m}{\hat\rho_{i+1,i}} + v_{i+1}^n - v_i^n \right|^2.$$

## 5.2 Interpolation of the Discrete State Space

In this section we will investigate the embedding of the discrete state space into a continuous Wasserstein space. To this aim, one has to construct suitable interpolants on $\Omega$, which converge weakly to limits in such a way that estimates for the dissipation functional and the Fisher information are obtainable.

**5.2.1 Definition** We introduce the piecewise constant interpolant $U^n : \Omega \to \mathbb{R}$ of the control volume $u_i^n$ by

$$U^n(x) := \rho_i = \frac{u_i^n}{h_i^n} \qquad \text{for } x \in \omega_i^n. \tag{5.12}$$

Note that $\max\{a,b\}$ and $\min\{a,b\}$ are respective upper and lower bounds for the weight function $\theta_m(a,b)$. This gives rise to the following upper and lower interpolants of $u_i^n$, also involving the adjacent control volumes:

$$\breve{U}^n(x) := \begin{cases} \max\{\rho_1,\rho_2\} & \text{for } x \in \omega_1^n, \\ \max\{\rho_{i-1},\rho_i,\rho_{i+1}\} & \text{for } x \in \omega_i^n,\ i \in \{2,\dots n-1\}, \\ \max\{\rho_{n-1},\rho_n\} & \text{for } x \in \omega_n^n. \end{cases}$$

$$\widehat{U}^n(x) := \begin{cases} \min\{\rho_1,\rho_2\} & \text{for } x \in \omega_1^n, \\ \min\{\rho_{i-1},\rho_i,\rho_{i+1}\} & \text{for } x \in \omega_i^n,\ i \in \{2,\dots n-1\}, \\ \min\{\rho_{n-1},\rho_n\} & \text{for } x \in \omega_n^n. \end{cases}$$

Note that $\breve{U}^n$ and $\widehat{U}^n$ are respective upper and lower bounds for $\hat{\rho}$, i.e.

$$\widehat{U}^n(x) \le \hat{\rho}_{i\pm1,i} = \theta_m\left(\frac{u_{i\pm1}}{h_{i\pm1}^n}, \frac{u_i^n}{h_i^n}\right) \le \breve{U}^n(x) \qquad \forall x \in \omega_i^n,\ i \in \{2,\dots n-1\}. \tag{5.13}$$

We introduce yet another interpolant of $u_i^n$ which is constant on intervals $(x_i^n, x_{i+1}^n)$:

$$\widetilde{U}^n(x) := \hat{\rho}_{i+1,i} \qquad \text{for } x \in (x_i^n, x_{i+1}^n),\ i \in \{1,\dots m-1\}.$$

Note that (5.13) implies that the interpolants are related in the way $\widehat{U}^n \le \widetilde{U}^n \le \breve{U}^n$.
It remains to define piecewise constant interpolants for the discrete gradient of the density $\rho_i$ and for the discrete fluxes $q_i$ via

$$G^n(x) := \frac{\rho_{i+1}^m - \rho_i^m}{x_{i+1}^n - x_i^n}, \quad H^n(x) := \frac{\upsilon_{i+1}^n - \upsilon_i^n}{x_{i+1}^n - x_i^n}, \quad Q^n(x) := q_i \quad \text{for } x \in (x_i^n, x_{i+1}^n),\ i \in \{1,\dots m-1\},$$

whereas at the boundary one sets $Q^n(\sigma_0) := Q^n(\sigma_n) := 0$. Here the $x_i^n$ denote the partition of our finite volume scheme and the $\sigma_i^n$ denote the corresponding midpoints.

**5.2.2 Facts** Let the interpolants $U^n, G^n$ and $Q^n$ be given as above.

(i) The interpolant $U^n$ is defined in a way such that an evaluation of the Rényi entropy functional $F_m$ corresponds to

$$F_m(U^n) = \int_\Omega f_m(U^n)\,\mathrm{d}x = \sum_{i=1}^n f_m(\rho_i)h_i^n = \mathcal{F}_m^n(\rho).$$

In the same vain, the evaluation of the potential energy functional $V$ results in

$$V(U^n) = \int_\Omega \upsilon \cdot U^n\,\mathrm{d}x = \sum_{i=1}^n \upsilon_i^n \rho_i h_i^n = \mathcal{V}^n(\rho).$$

(ii) Recalling the definition of $q_k$ in **(5.8)**, we immediately obtain the following relation between the interpolants $U^n$ and $Q^n$:

$$\dot{U}^n(t,x_i^n) = \frac{Q^n(t,\sigma_i) - Q^n(t,\sigma_{i-1})}{h_i^n} \qquad \forall i \in \{1,\dots n-1\}. \tag{5.14}$$

The following lemma provides first estimates for the dissipation functional $\mathcal{R}^n$ and the Fisher information $\mathcal{I}^n$ with respect to the interpolants defined above.

**5.2.3 Lemma** *For every smooth gradient flow $\rho : (0,T) \to \operatorname{int}\mathcal{P}^n$, induced by the finite volume discretisation $u^n$ via $u_i = h_i^n \rho_i$, the following lower estimates hold for all times $t \in [0,T]$:*

$$\mathcal{R}^n\big(u^n(t), \dot{u}^n(t)\big) \geq \frac{1}{2}\int_\Omega \frac{|Q^n(t)|^2}{\breve{U}^n(t)}\mathrm{d}x \qquad \text{and} \qquad \mathcal{I}^n\big(u^n(t)\big) \geq \frac{1}{2}\int_\Omega \widehat{U}^n(t)\left|\frac{G^n(t)}{\widetilde{U}^n(t)} + H^n(t)\right|^2 \mathrm{d}x.$$

Proof We start with the estimate for the dissipation functional $\mathcal{R}^n$: By recalling the definition of the rate coefficient $\alpha_i$ in **(5.2)**, we obtain

$$\lambda_i := \alpha_i h_i^n \hat{\rho}_{i+1,i} = \frac{\theta_m(\rho_{i+1},\rho_i)}{x_{i+1}^n - x_i^n} \leq \frac{\max\{\rho_{i+1},\rho_i\}}{x_{i+1}^n - x_i^n}. \tag{5.15}$$

Now we have to compare the interpolant $Q^n$ which is constant of every interval $(x_i^n, x_{i+1}^n)$ with $\breve{U}^n$ which is constant on every $\omega_i^n$. To this end, we count each control volume $\omega_i^n$ twice in the definition of $\mathcal{R}^n$ and apply the estimate from **(5.15)**, which results in

$$\mathcal{R}^n(u^n,\dot{u}^n) = \frac{1}{4}\frac{q_1^2}{\lambda_1} + \frac{1}{4}\sum_{i=2}^{n-1}\Big(\frac{q_{i-1}^2}{\lambda_{i-1}} + \frac{q_i^2}{\lambda_i}\Big) + \frac{1}{4}\frac{q_{n-1}^2}{\lambda_{n-1}} \geq \frac{x_2^n - x_1^n}{4}\frac{q_1^2}{\max\{\rho_2,\rho_1\}} +$$
$$+\sum_{i=2}^{n-1}\left(\frac{x_i^n - x_{i-1}^n}{4}\frac{q_{i-1}^2}{\max\{\rho_i,\rho_{i-1}\}} + \frac{x_{i+1}^n - x_i^n}{4}\frac{q_i^2}{\max\{\rho_{i+1},\rho_i\}}\right) + \frac{x_n^n - x_{n-1}^n}{4}\frac{q_{n-1}^2}{\max\{\rho_{n-1},\rho_n\}}.$$

Now using that $x_i^n - \sigma_{i-1}^n = \frac{1}{2}(x_i^n - x_{i-1}^n)$ and $\sigma_i^n - x_i^n = \frac{1}{2}(x_{i+1}^n - x_i^n)$, we eventually arrive at

$$\mathcal{R}^n(u,\dot{u}) \geq \frac{1}{2}\sum_{i=1}^{n}\left(\frac{1}{\breve{U}^n}\int_{\omega_i^n}|Q^n|^2\,\mathrm{d}x\right) = \frac{1}{2}\int_\Omega \frac{|Q^n|^2}{\breve{U}^n}\mathrm{d}x,$$

which shows the first inequality.

The proof of the inequality for the discrete Fisher information $\mathcal{I}^n$ follows along the lines of the first part. Indeed, we have the elementary estimate **(5.13)** can be used to obtain

$$\mathcal{I}^n(u^n) = \frac{1}{2}\sum_{i=1}^{n-1}\frac{\hat{\rho}_{i+1,i}}{x_{i+1}^n - x_i^n}\left|\frac{\rho_{i+1}^m - \rho_i^m}{\hat{\rho}_{i+1,i}} + v_{i+1}^n - v_i^n\right|^2 =$$
$$= \frac{1}{2}\sum_{i=1}^{n-1}\big((\sigma_i^n - x_i^n)\hat{\rho}_{i+1,i} + (x_{i+1}^n - \sigma_i^n)\hat{\rho}_{i+1,i}\big)\left|\frac{1}{\widetilde{U}^n(\sigma_i^n)}\frac{\rho_{i+1}^m - \rho_i^m}{x_{i+1}^n - x_i^n} + \frac{v_{i+1}^n - v_i^n}{x_{i+1}^n - x_i^n}\right|^2 \geq$$
$$\geq \frac{1}{2}\sum_{i=1}^{n}\left(\widehat{U}^n\int_{\omega_i^n}\left|\frac{G^n}{\widetilde{U}^n} + H^n\right|^2\mathrm{d}x\right) = \frac{1}{2}\int_\Omega \widehat{U}^n\left|\frac{G^n}{\widetilde{U}^n} + H^n\right|^2\mathrm{d}x.$$

This establishes the second inequality. ■

With above lemma at hand, the following main result of this section is obvious.

**5.2.4 Proposition** *Let $\rho : (0,T) \to \operatorname{int} \mathcal{P}^n$ be the smooth gradient flow induced by the finite volume discretisation $u^n$ via $u_i = h_i^n \rho_i$ such that* Assumptions 5.1.2 *are fulfilled. Then the continuous functional $\phi = F_m + V$ satisfies the following inequality for the interpolants introduced in* Definition 5.2.1*:*

$$\phi\big(U^n(T)\big) + \frac{1}{2} \int_0^T \int_\Omega \frac{|Q^n|^2}{\widetilde{U}^n} + \widehat{U}^n \left| \frac{G^n}{\widetilde{U}^n} + H^n \right|^2 \mathrm{d}x \, \mathrm{d}t \le \phi\big(U^n(0)\big). \tag{5.16}$$

## 5.3 Passage to the Limit

In this section we will show that the sequence $(U^n)_{n\in\mathbb{N}}$ of interpolants admits a subsequence which weakly converges to an absolutely continuous limit curve $U$ taking values in the 2-Wasserstein space $\mathcal{P}_2(\overline{\Omega})$. Assuming $0 < m \le 3/2$, we will then identify $U$ as EDE gradient flow for the mixed energy functional $\phi = F_m + V$.

To archive these goals, we will follow the ideas of Proposition 4.1.2. However, we will first need to establish the prerequisites of aforementioned stability result. We start with the extraction of a converging subsequence.

**5.3.1 Proposition** *Let $(U^n)_{n\in\mathbb{N}}$ be a sequence of interpolants $U^n$ satisfying* (5.16)*. Then there exists a subsequence $(U^{n_k})_{k\in\mathbb{N}}$ and a continuous curve $\mu : (0,T) \to \mathcal{P}_2(\Omega)$ such that*

$$U^{n_k}(t) \xrightarrow{w^*} \mu(t) \qquad \forall t \in (0,T).$$

*Moreover, the discrete functional $\Phi^{n_k} = \mathcal{F}_m^{n_k} + \mathcal{V}^{n_k}$ and the continuous functional $\phi = F_m + V$ are related by the inequality*

$$\liminf_{k\to\infty} \Phi^{n_k}\big(u^{n_k}(t)\big) \ge \phi\big(\mu(t)\big) \qquad \forall t \in (0,T), \tag{5.17}$$

*and the limit measure $\mu(t)$ has again a density $U(t)$ with respect to the Lebesgue measure $\mathcal{L}$ for all times $t \in (0,T)$.*

Note that (5.17) is actually weaker than sequential $\Gamma$-$\liminf$ convergence as defined in Definition 4.1.1.

Proof The aim of this proof is to apply the Arzelà-Ascoli theorem in the version of Lemma 4.2.5.i to extract a weakly converging subsequence out of $(U^n)_{n\in\mathbb{N}}$. To this end, we need to show that the familiy of interpolants satisfies (4.24) with respect to the 1-Wasserstein distance. Here comes the Kantorovich-Rubinshtein theorem (Theorem 2.2.5) into play, which gives a useful characterisation of the $W_1$-distance:

$$W_1(\mu_1, \mu_2) = \sup\left\{ \int_\Omega \psi \, \mathrm{d}\mu_1 - \int_\Omega \psi \, \mathrm{d}\mu_2 : \psi \in Lip_1(\Omega) \right\} \qquad \forall \mu_1, \mu_2 \in \mathcal{P}_2(X). \tag{5.18}$$

We may invoke (5.18) to estimate the $W_1$-distance between $\mu_1 = U^n(t_1)$ and $\mu_2 = U^n(t_2)$ for two different times $0 < t_1 < t_2 < T$. Therefore, let $(\psi_k)_{k\in\mathbb{N}}$ be a sequence in $Lip_1(\Omega)$ such that the supremum in (5.18) is attained over $(\psi_k)_{k\in\mathbb{N}}$. As usual, we define the piecewise constant interpolant for every $\psi_k$ to be $\Psi_k^n(x) = \psi_i(x_k^n)$ for each $x \in \omega_k^n$. Thus, we obtain

$$W_1\big(U^n(t_1), U^n(t_2)\big) \le \tag{5.19.a}$$

$$\le \int_\Omega \Psi_k^n(x) U^n(t_1, x) \, \mathrm{d}x - \int_\Omega \Psi_k^n(x) U^n(t_2, x) \, \mathrm{d}x + 2 \left\| \psi_k - \Psi_k^n \right\|_\infty + \varepsilon_n = \tag{5.19.b}$$

$$= \int_{t_1}^{t_2} \int_\Omega \Psi_k^n(x) \dot{U}^n(t, x) \, \mathrm{d}x \, \mathrm{d}t + 2 \left\| \psi_k - \Psi_k^n \right\|_\infty + \varepsilon_n \tag{5.19.c}$$

where the error term $\varepsilon_n$ vanishes as $(n \to \infty)$. To estimate the integral in (5.19.c) we may use summation by parts to infer

$$\int_{t_1}^{t_2}\int_\Omega \Psi_k^n(x)\dot{U}^n(t,x)\,\mathrm{d}x\,\mathrm{d}t = \int_{t_1}^{t_2}\sum_{i=1}^n\int_{\omega_i^n}\psi_k(x_i^n)\frac{Q^n(t,\sigma_i)-Q^n(t,\sigma_{i-1})}{h_i^n}\,\mathrm{d}x\,\mathrm{d}t = \tag{5.20.a}$$

$$= -\int_{t_1}^{t_2}\sum_{i=1}^n\big(\psi_k(x_i^n)-\psi_k(x_{i-1}^n)\big)Q^n(t,\sigma_i)\,\mathrm{d}t = -\int_{t_1}^{t_2}\int_\Omega \psi_k'(x)Q^n(t,x)\,\mathrm{d}x\,\mathrm{d}t \le \tag{5.20.b}$$

$$\le \int_0^T \mathbb{1}_{(t_1,t_2)}(x)\sum_{i=1}^{n-1}(x_{i+1}^n-x_i^n)\,|q_i(t)|\,\mathrm{d}t \le (t_2-t_1)^{1/2}\bigg(\int_0^T\sum_{i=1}^{n-1}(x_{i+1}^n-x_i^n)^2\,|q_i(t)|^2\,\mathrm{d}t\bigg)^{1/2}, \tag{5.20.c}$$

where we used Hölder's inequality in the last line above. It remains to find an estimate for the expression in the rightmost parentheses in (5.20.c) which is uniform with respect to all $n \in \mathbb{N}$. To this aim, we may exploit the trivial bounds $h_{i+1}^n \ge (x_{i+1}^n - x_i^n)/2$ and $h_i^n \ge (x_{i+1}^n - x_i^n)/2$, as well as the 1-homogeneity of $\theta_m$ to obtain

$$(x_{i+1}^n - x_i^n)\theta_m(\rho_{i+1},\rho_1) = \theta_m\bigg(\frac{x_{i+1}^n-x_i^n}{h_i^n}u_i^n, \frac{x_{i+1}^n-x_i^n}{h_{i+1}^n}u_{i+1}^n\bigg) \le 2\max\left\{u_i^n, u_{i+1}^n\right\} \le 2. \tag{5.21}$$

Now plugging this estimate into (5.20.c) gives

$$\int_{t_1}^{t_2}\int_\Omega \Psi_k^n(x)\dot{U}^n(t,x)\,\mathrm{d}x\,\mathrm{d}t \le 2(t_2-t_1)^{1/2}\bigg(\int_0^T\sum_{i=1}^{n-1}\frac{(x_{i+1}^n-x_i^n)\,|q_i(t)|^2}{\hat{\rho}_{i+1,i}}\,\mathrm{d}t\bigg)^{1/2} =$$
$$= 4(t_2-t_1)^{1/2}\bigg(\int_0^T \mathcal{R}^n(u^n,\dot{u}^n)\,\mathrm{d}t\bigg)^{1/2}.$$

Note that $\mathcal{R}^n(u^n,\dot{u}^n)$ is uniformly bounded in $L^1(0,T)$ for all $n \in \mathbb{N}$, due to the fact that each $\rho_i = u_i^n/h_i^n$ is a solution to the corresponding EDE (5.7) with initial condition (5.6).
Finally, combining the estimate above with (5.19) shows that the family $(U^n)_{n\in\mathbb{N}}$ satisfies (4.24) with respect to the function $\omega(s,t) = 4\,|s-t|\,\big(\int_0^T \mathcal{R}^n(u^n,\dot{u}^n)\,\mathrm{d}t\big)^{1/2}$.

Next, we come to the proof of (5.17): Indeed, this inequality follows from the lower semi-continuity of the Rényi entropy functional $F_m$ and the potential energy functional $V$ (cf. Proposition 2.4.3) and the fact that $\Phi^n(u^n) = \phi(U^n)$.

For the last statement in Proposition 5.3.1, namely the existence of a density $U(t)$ for $\mu(t)$, it is enough to invoke the definition of the Rényi entropy $F_m$ and show that $F_m(\mu(t)) < +\infty$ for all times $t \in (0,T)$. Note that this follows already from the bound $\phi(\mu(t)) < +\infty$. Since the discrete functional $\Phi^n(u^n(t))$ is uniformly bounded in $t$ by $\Phi^n(u_0)$ as a consequence of $u^n$ satisfying the EDE (5.7), we may invoke (5.17) to arrive at the result. ■

In the next result we show that along with $U^n$ all the other interpolants introduced in Definition 5.2.1 admit a converging subsequence as well.

**5.3.2 Lemma** *Let $(\widetilde{U}^{n_k})_{k\in\mathbb{N}}$, $(\widehat{U}^{n_k})_{k\in\mathbb{N}}$, and $(\widetilde{U}^{n_k})_{k\in\mathbb{N}}$ be the sequences of interpolants associated with the converging subsequence $(U^{n_k})_{k\in\mathbb{N}}$ of Proposition 5.3.1. Then each of the three sequences converges weakly to $U$ on $\Omega\times(0,T)$, i.e.*

$$\lim_{k\to\infty}\int_{\Omega\times(0,T)}\psi(x,t)\Upsilon^{n_k}(x,t)\,\mathrm{d}(x,t) = \int_{\Omega\times(0,T)}\psi(x,t)U(x,t)\,\mathrm{d}(x,t) \qquad \forall\psi\in C_b\big(\Omega\times(0,T)\big), \tag{5.22}$$

*where $\Upsilon^{n_k}$ corresponds to one of the interpolants $\widetilde{U}^{n_k}$, $\widehat{U}^{n_k}$, or $\widetilde{U}^{n_k}$.*

Proof Supposed, we show first that $\lim_{k\to\infty}(U^{n_k} - \mathrm{Y}^{n_k}) = 0$ in $L^1(\Omega\times[0,T])$, then the estimate

$$\int\limits_{\Omega\times[0,T]} \psi\,|U - \mathrm{Y}^{n_k}|\,\mathrm{d}(x,t) \le \int\limits_{\Omega\times[0,T]} \psi\,|U - U^{n_k}|\,\mathrm{d}(x,t) + \|\psi\|_\infty \int\limits_{\Omega\times[0,T]} |U^{n_k} - \mathrm{Y}n_k|\,\mathrm{d}(x,t),$$

implies (5.22) as $(k\to\infty)$.

In the following, we argue for $\mathrm{Y}^{n_k} = \breve{U}^{n_k}$; the proof for $\widehat{U}^{n_k}$ follows in a similar fashion, whereas weak convergence of $\widetilde{U}^{n_k}$ follows by the elementary bound $\breve{U}^{n_k} \le \widetilde{U}^{n_k} \le \widehat{U}^{n_k}$.

It remains to compute

$$\|U^{n_k} - \breve{U}^{n_k}\|_{L^1(\Omega\times[0,T])} = \sum_{i=1}^{n} h_i^{n_k}\,|\rho_i - \breve{U}^{n_k}(x_i)| = \sum_{i\in I_+} h_i^{n_k}\,|\rho_i - \rho_{i+1}| + \sum_{i\in I_-} h_i^{n_k}\,|\rho_i - \rho_{i-1}|, \tag{5.23}$$

where the sets $I_-$ and $I_+$ consist of all indices in $\{1,\ldots n\}$ where the maximum in $|\rho_i - \breve{U}^{n_k}(x_i)|$ is attained for the leftmost element $\breve{U}^{n_k}(x_i) = \rho_{i-1}$ and rightmost element $\breve{U}^{n_k}(x_i) = \rho_{i+1}$ in the definition of $\breve{U}^{n_k}$, respectively.

Now we just argue for the sum over the index set $I_+$, the other case being analogue. Note that we may invoke Hölder's inequality to infer

$$\sum_{i\in I_+} h_i^{n_k}\,|\rho_i - \rho_{i+1}| = \sum_{i\in I_+}\left(\frac{|\rho_i^m - \rho_{i+1}^m|}{\left((x_{i+1}^{n_k} - x_i^{n_k})\hat\rho_{i+1,i}\right)^{1/2}}\right)\left(h_i^{n_k}\left((x_{i+1}^{n_k} - x_i^{n_k})\hat\rho_{i+1,i}\right)^{1/2}\frac{|\rho_i^m - \rho_{i+1}^m|}{|\rho_i^m - \rho_{i+1}^m|}\right) \le$$

$$\le \left(\sum_{i\in I_+}\frac{|\rho_i^m - \rho_{i+1}^m|^2}{(x_{i+1}^{n_k} - x_i^{n_k})\hat\rho_{i+1,i}}\right)^{1/2}\left(\sum_{i\in I_+}(h_i^{n_k})^2(x_{i+1}^{n_k} - x_i^{n_k})\hat\rho_{i+1,i}\left|\frac{\rho_i - \rho_{i+1}}{\rho_i^m - \rho_{i+1}^m}\right|^2\right)^{1/2}. \tag{5.24}$$

The first term in (5.24) is just the square root of the discrete Fisher information $\mathcal{I}^{n_k}(u^{n_k})$, which is uniformly bounded in $L^2(0,T)$.

On the other hand, we may invoke the bound $x_{i+1}^{n_k} - x_i^{n_k} \le 2h_i^{n_k}$ and expand the definition of $\hat\rho_{i+1,i}$ to estimate the second term in (5.24) as

$$\sum_{i\in I_+}(h_i^{n_k})^2(x_{i+1}^{n_k} - x_i^{n_k})\hat\rho_{i+1,i}\left|\frac{\rho_i - \rho_{i+1}}{\rho_i^m - \rho_{i+1}^m}\right|^2 \le \sum_{i\in I_+} 2\left(h_i^{n_k}\right)^3\,\tilde\theta_m(\rho_i,\rho_{+1}),$$

where $\tilde\theta_m$ is a weight function defined for every $0 < m \le 3/2$ by

$$\tilde\theta_m(s,t) := \frac{m-1}{m}\,\frac{(s-t)^2}{(s^m - t^m)(s^{m-1} - t^{m-1})}.$$

it is easy to check that $\tilde\theta_m$ satisfies (W1) to (W5) in Definition 3.1.1. In particular, we have the bound $\tilde\theta_m(s,t) \le C_m \max\{s,t\}^{3-2m}$ for some constant $C_m > 0$. Therefore, we obtain the estimate

$$\sum_{i\in I_+}(h_i^{n_k})^2(x_{i+1}^{n_k} - x_i^{n_k})\hat\rho_{i+1,i}\left|\frac{\rho_i - \rho_{i+1}}{\rho_i^m - \rho_{i+1}^m}\right|^2 \le \sum_{i\in I_+} 2C_m\left(h_i^{n_k}\right)^3\,\max\{\rho_i,\rho_{i+1}\}^{3-2m}, \tag{5.25}$$

where the right-hand side converges to zero as $(k\to\infty)$. ∎

5.3.3 **Remark** In Assumptions 5.1.2.i the condition $m \le 3/2$ was stated. This upper bound on $m$ was needed for the proof of Lemma 5.3.2 to go through. Indeed, the the right-hand side of the estimate in (5.25) need not be bounded any more for $m > 3/2$. Actually, this turns out to be the only point in this section where we have to make this restriction on $m$. Otherwise, it suffices to assume that $\theta_m$ is a concave weight function which is true for $m \le 2$.

**5.3.4 Proposition** *Let $(U^{n_k})_{k\in\mathbb{N}}$ be a weakly converging sequence as in* Proposition 5.3.1. *Then the limit measure $\mu(t,\mathrm{d}x) = U(t,x)\mathrm{d}x$ satisfies*

$$\liminf_{k\to\infty}\int_0^T \mathcal{R}^{n_k}(u^{n_k},\dot{u}^{n_k}) + \mathcal{I}^n(u^n)\,\mathrm{d}t \geq \frac{1}{2}\int_{\Omega\times[0,T]} U\,|Y|^2 \;+\; \frac{|\nabla(U^m)|^2}{U} + U\,|\nabla\upsilon|^2\,\mathrm{d}(x,t), \tag{5.26}$$

*where $U$ and $Y$ are connected via the continuity equation* (2.9) *in the sense of distributions.*

Before starting with the proof of the result above, we are going to obtain uniform estimates for $\mathcal{F}_m^{n_k}\big(u^{n_k}(0)\big)$ and $\mathcal{V}\big(u^{n_k}(0)\big)$: For the discrete Rényi entropy $\mathcal{F}_m^{n_k}$ we invoke Jensen's inequality to infer

$$\mathcal{F}_m^{n_k}\big(u^{n_k}(0)\big) = F_m\big(U^{n_k}(0)\big) = \frac{1}{(m-1)}\sum_{i=1}^{n_k} h_i^{n_k}\Bigg(\int_{\omega_i^{n_k}}\frac{U_0(x)}{h_i^{n_k}}\,\mathrm{d}x\Bigg)^m \leq \tag{5.27.a}$$

$$\leq \frac{1}{(m-1)}\sum_{i=1}^{n_k}\int_{\omega_i^{n_k}}\big(U_0(x)\big)^m\,\mathrm{d}x = F_m(U_0). \tag{5.27.b}$$

For the discrete potential energy $\mathcal{V}^{n_k}$ we have

$$\mathcal{V}^{n_k}\big(u^{n_k}(0)\big) = \sum_{i=1}^{n_k}\rho(0)\int_{\omega_i^{n_k}}\upsilon\,\mathrm{d}x = \int_{\Omega}\upsilon U^{n_k}(0)\,\mathrm{d}x = V(U_0^{n_k}), \tag{5.28}$$

where $U^{n_k}(0)$ is weakly converging to $U_0$. Since $\upsilon$ is a bounded continuous function on $\Omega$, this means that $\lim_{k\to\infty}\mathcal{V}^{n_k}\big(u^{n_k}(0)\big) = V(U_0)$. In particular, $\mathcal{V}^{n_k}\big(u^{n_k}(0)\big)$ is uniformly bounded for all $k\in\mathbb{N}$.

Proof of Proposition 5.3.4 The following proof is arranged in 4 steps. We start with introducing the velocity fields

$$Y^{n_k} := \frac{Q^{n_k}}{\widetilde{U}^{n_k}} \qquad\text{and}\qquad Z^{n_k} := \frac{G^{n_k}}{\widetilde{U}^{n_k}} + H^{n_k}.$$

Note that (5.16), together with (5.27) and (5.28), implies that

$$\frac{1}{2}\int_{\Omega\times[0,T]}\big(|Y^{n_k}|^2 + |Z^{n_k}|^2\big)\widehat{U}^{n_k}\,\mathrm{d}(x,t) \leq \frac{1}{2}\int_{\Omega\times[0,T]}\frac{|Q^{n_k}|^2}{\widetilde{U}^{n_k}} + \widehat{U}^{n_k}\left|\frac{G^{n_k}}{\widetilde{U}^{n_k}} + H^{n_k}\right|^2\mathrm{d}(x,t) \leq \tag{5.29.a}$$

$$\leq \mathcal{V}^{n_k}\big(u^{n_k}(0)\big) + \mathcal{V}^{n_k}\big(u^{n_k}(0)\big) \leq F_m(U_0) + V(U_0^{n_k}), \tag{5.29.b}$$

which is uniformly bounded for all $k\in\mathbb{N}$.

(i) Extraction of a converging subsequence: We first consider the sequence $(Y^{n_k})_{k\in\mathbb{N}}$.
Consider the push-forward measures $\upsilon_{n_k} := (\mathrm{Id}\times Y^{n_k})_{\#}\widetilde{U}^{n_k}$. We already observed that the projection $\pi^1_{\#}\upsilon_{n_k} = \widetilde{U}^{n_k}$ onto $\overline{\Omega}\times[0,T]$ is weakly convergent. Hence, $(\pi^1_{\#}\upsilon_{n_k})_{n\in\mathbb{N}}$ is uniformly tight by Prokhorov's theorem (Theorem A.1.10). Moreover, due to the uniform bound in (5.29), every projection $\pi^2\upsilon_{n_k}$ onto $\mathbb{R}$ satisfies

$$\sup_{k\in\mathbb{N}}\int_{\mathbb{R}}|y|^2\,\mathrm{d}\pi^2\upsilon^{n_k}(y) = \sup_{k\in\mathbb{N}}\int_{\Omega\times(0,T)}|Y^{n_k}(x,t)|^2\,\widetilde{U}^{n_k}(x,t)\,\mathrm{d}(x,t) < +\infty.$$

Thus, we can apply the results obtained in Facts A.1.9: The sequence $(\pi^2_{\#}\upsilon_{n_k})_{k\in\mathbb{N}}$ is uniformly tight by the integral criterion for tightness; hence, $(\upsilon_{n_k})_{k\in\mathbb{N}}$ is uniformly tight as well. Now we may again invoke Prokhorov's theorem to conclude that there exists a subsequence, also denoted by $(\upsilon_{n_k})_{k\in\mathbb{N}}$, weakly convergent to some Borel probability measure $\upsilon$ on $\overline{\Omega} \times [0,T] \times \mathbb{R}$.

Furthermore, (A.2) implies that the map $(y,x,t) \mapsto |y|$ is uniformly integrable with respect to $(\upsilon_{n_k})_{k\in\mathbb{N}}$ since (5.29) gives us the uniform bound

$$\int_{\mathbb{R}\times\Omega\times[0,T]} |y|^2 \,\mathrm{d}\upsilon_{n_k}(y,x,t) = \int_{\Omega\times[0,T]} |Y^{n_k}(x,t)|^2 \,\breve{U}^{n_k}(x,t)\,\mathrm{d}(x,t) < C \qquad \forall k \in \mathbb{N}.$$

Therefore, we may invoke Lemma A.1.7.i to obtain

$$\lim_{k\to\infty} \int_{\mathbb{R}\times\Omega\times[0,T]} \psi(s,t) y \,\mathrm{d}\upsilon_{n_k}(y,x,t) = \int_{\mathbb{R}\times\Omega\times[0,T]} \psi(s,t) y \,\mathrm{d}\upsilon(y,x,t) \qquad \forall \psi \in C_b\big(\overline{\Omega} \times [0,T]\big). \tag{5.30}$$

Note that Theorem A.2.1 guarantees the existence of disintegration measures $\upsilon_{(x,t)}$ of $\upsilon$ on $\mathbb{R}$ with respect to $U(x,t)\mathrm{d}(x,t)$. In particular, for every non-negative Borel function $g : \mathbb{R}\times\Omega\times(0,T) \to \mathbb{R}_0^+$ we have

$$\int_{\mathbb{R}\times\Omega\times(0,T)} g(y,x,t)\,\mathrm{d}\upsilon(y,x,t) = \int_{\Omega\times(0,T)} \Big(\int_{\mathbb{R}} g(y,x,t)\,\mathrm{d}\upsilon_{(x,t)}(y)\Big)\, U(x,t)\,\mathrm{d}(x,t).$$

Together with (5.30), we arrive at

$$\lim_{k\to\infty} \int_{\Omega\times(0,T)} \psi(x,t) Y^{n_k}(x,t) \breve{U}^{n_k}(x,t)\,\mathrm{d}(x,t) = \int_{\Omega\times(0,T)} \psi(x,t) Y(x,t) U(x,t)\,\mathrm{d}(x,t) \qquad \forall \psi \in C_b\big(\overline{\Omega} \times [0,T]\big),$$

where $Y(x,t) := \int_{\mathbb{R}} y\,\mathrm{d}\upsilon_{(x,t)}(y)$ is the barycentric projection of $\upsilon$.

The argument for $(Z^{n_k})_{k\in\mathbb{N}}$ follows along the same lines considering the respective push-forward measures $\zeta_{n_k} := (\mathrm{Id}\times Z^{n_k})_{\#}\widehat{U}^{n_k}$ instead; thus, extracting a subsequence of $(\zeta_{n_k})_{k\in\mathbb{N}}$, which is weakly convergent to some Borel probability measure $\zeta$ on $\overline{\Omega} \times [0,T] \times \mathbb{R}$. In the spirit of the argument above, one obtains disintegration measures $\zeta_{(x,t)}$ of $\zeta$ with respect to $U(x,t)\mathrm{d}(x,t)$ such that

$$\lim_{k\to\infty} \int_{\Omega\times(0,T)} \psi(x,t) Z^{n_k}(x,t) \widehat{U}^{n_k}(x,t)\,\mathrm{d}(x,t) = \int_{\Omega\times(0,T)} \psi(x,t) Z(x,t) U(x,t)\,\mathrm{d}(x,t) \qquad \forall \psi \in C_b\big(\overline{\Omega} \times [0,T]\big),$$

where $Z(x,t) := \int_{\mathbb{R}} y\,\mathrm{d}\zeta_{(x,t)}(y)$.

(ii) Lower estimate for the limit inferior: We invoke Lemma A.1.7.ii and Jensens' inequality for the convex function $y \mapsto |y|^2$ to obtain

$$\begin{aligned}
\liminf_{k\to\infty} \int_{\Omega\times(0,T)} |Y^{n_k}(x,t)|^2\, \breve{U}^{n_k}(x,t)\,\mathrm{d}(x,t) &= \liminf_{n\to\infty} \int_{\mathbb{R}\times\Omega\times(0,T)} |y|^2 \,\mathrm{d}\eta_{n_k}(y,x,t) \geq \\
\geq \int_{\mathbb{R}\times\Omega\times(0,T)} |y|^2 \,\mathrm{d}\eta(y,x,t) &= \int_{\Omega\times(0,T)} \Big(\int_{\mathbb{R}} |y|^2 \,\mathrm{d}\eta_{(x,t)}(y)\Big)\, U(x,t)\,\mathrm{d}(x,t) \geq \\
\geq \int_{\Omega\times(0,T)} \Big|\int_{\mathbb{R}} y\,\mathrm{d}\eta_{(x,t)}(y)\Big|^2 U(x,t)\,\mathrm{d}(x,t) &= \int_{\Omega\times(0,T)} |Y(x,t)|^2\, U(x,t)\,\mathrm{d}(x,t).
\end{aligned}$$

Since an analogous estimate holds for $Z^{n_k}$ in place of $Y^{n_k}$, we can use the inequalities obtained in Lemma 5.2.3 to arrive at

$$\liminf_{k\to\infty} \int_0^T \mathcal{R}^{n_k}(u^{n_k}, \dot{u}^{n_k}) + \mathcal{F}_m^{n_k}(u^{n_k})\,\mathrm{d}t \geq \frac{1}{2} \int_{\Omega\times(0,T)} \big(\, |Y(x,t)|^2 + |Z(x,t)|^2 \,\big) U(x,t)\,\mathrm{d}(x,t).$$

(iii) Identification of the limit $Y$: Let $\psi \in C_c^\infty\big((0,T)\times\Omega\big)$ be a test function and let $\Psi^{n_k}(t,x) := \psi(t, x_i^{n_k})$ for $x \in \omega_i^{n_k}$ be the corresponding interpolant. Then integration by parts and summation by parts imply

$$\int_0^T\int_\Omega U^{n_k}(t,x)\dot\psi(t,x)\,dx\,\mathrm{d}t = -\int_0^T\sum_{i=1}^{n_k}\psi(t,x_i^{n_k})\big(Q^{n_k}(\sigma_i,t)-Q^{n_k}(\sigma_{i-1},t)\big)\,\mathrm{d}t + \varepsilon_{n_k} = \tag{5.31.a}$$

$$= \int_0^T\sum_{i=1}^{n_k-1} Q^{n_k}(t,\sigma_i)\big(\psi(t,x_i^{n_k})-\psi(t,x_{i+1}^{n_k})\big)\,\mathrm{d}t + \varepsilon_{n_k} = \tag{5.31.b}$$

$$= -\int_0^t\int_\Omega Y^{n_k}(t,x)\nabla_x\psi(t,x)\widetilde{U}^{n_k}(t,x)\,\mathrm{d}x\,\mathrm{d}t + \varepsilon_{n_k} \tag{5.31.c}$$

where the error term $\varepsilon_{n_k} \in \mathcal{O}(\|\dot\psi - \dot\Psi^{n_k}\|_\infty)$ goes to zero as $(n\to\infty)$. Since the sequences $(U^{n_k})_{k\in\mathbb{N}}$, $(\widetilde{U}^{n_k})_{k\in\mathbb{N}}$, and $(Y^{n_k})_{k\in\mathbb{N}}$ are weakly convergent, we may pass to the limit in **(5.31)** as $(k\to\infty)$ and, therefore, arrive at

$$\int_{\Omega\times(0,T)} U(t,x)\dot\psi(t,x)\,\mathrm{d}(x,t) = -\int_{\Omega\times(0,T)} Y(t,x)\nabla_x\psi(t,x)U(t,x)\,\mathrm{d}(x,t),$$

which is precisely the continuity equation introduced in Theorem 2.3.3 relating $U$ to $Y$.

(iv) Identification of the limit $Z$: As before, let $\psi \in C_c^\infty\big((0,T)\times\Omega\big)$ be a test function. We compute

$$\int_0^T\int_\Omega Z^{n_k}(t,x)\psi(t,x)\widetilde{U}^{n_k}(t,x)\,dx\,\mathrm{d}t = \tag{5.32.a}$$

$$= \int_0^T\sum_{i=1}^{n_k-1}\left(\frac{\rho_{i+1}^m-\rho_i^m}{x_{i+1}^{n_k}-x_i^{n_k}}\int_{x_i^{n_k}}^{x_{i+1}^{n_k}}\psi(t,x)\,\mathrm{d}x + \frac{\upsilon_{i+1}^{n_k}-\upsilon_i^{n_k}}{x_{i+1}^{n_k}-x_i^{n_k}}\int_{x_i^{n_k}}^{x_{i+1}^{n_k}}\psi(t,x)\widetilde{U}^{n_k}(t,x)\,\mathrm{d}x\right)\mathrm{d}t. \tag{5.32.b}$$

Note that the Lebesgue differentiation formula implies

$$\lim_{k\to\infty}\left(\psi(t,\sigma_i^{n_k}) - \frac{1}{x_{i+1}^{n_k}-x_i^{n_k}}\int_{x_i^{n_k}}^{x_{i+1}^{n_k}}\psi(t,x)\,\mathrm{d}x\right) = 0. \tag{5.33}$$

Hence, by summation by parts we obtain for the first term in the sum of **(5.32.b)** that

$$\int_0^T\sum_{i=1}^{n_k-1}\frac{\rho_{i+1}^m-\rho_i^m}{x_{i+1}^{n_k}-x_i^{n_k}}\int_{x_i^{n_k}}^{x_{i+1}^{n_k}}\psi(t,x)\,\mathrm{d}x\,\mathrm{d}t = -\int_0^T\sum_{i=1}^{n_k}\rho_i^m(t)\big(\psi(t,\sigma_i^{n_k})-\psi(t,\sigma_{i-1}^{n_k})\big)\,\mathrm{d}t + \varepsilon_{n_k} = \tag{5.34.a}$$

$$= -\int_0^T\int_\Omega\big(U^{n_k}(t,x)\big)^m\nabla_x\psi(t,x)\,\mathrm{d}x\,\mathrm{d}t + \varepsilon_{n_k} \tag{5.34.b}$$

where $\varepsilon_{n_k}$ goes to zero as $(k\to\infty)$.
In addition to **(5.34)**, we also have

$$\lim_{k\to\infty}\big(\upsilon(x_i^{n_k})-\upsilon_i^{n_k}\big) = \lim_{k\to\infty}\left(\upsilon(x_{i+1}^{n_k}) - \frac{1}{h_i^{n_k}}\int_{\omega_i^{n_k}}\upsilon(x)\,\mathrm{d}x\right) = 0.$$

In particular, this implies that

$$\lim_{k\to\infty}\left(\nabla_x\upsilon(\sigma_i^{n_k}) - \frac{\upsilon_{i+1}^{n_k}-\upsilon_i^{n_k}}{x_{i+1}^{n_k}-x_i^{n_k}}\right) = 0.$$

Therefore, the potential energy term in (5.32.b) may be computed as

$$\int_0^T \sum_{i=1}^{n_k-1} \frac{\upsilon_{i+1}^{n_k} - \upsilon_i^{n_k}}{x_{i+1}^{n_k} - x_i^{n_k}} \int_{x_i^{n_k}}^{x_{i+1}^{n_k}} \psi(t,x)\widetilde{U}^{n_k}(t,x)\,\mathrm{d}x\,\mathrm{d}t = \tag{5.35.a}$$

$$= \int_0^T \int_\Omega \nabla_x \upsilon(x)\psi(t,x)\widetilde{U}^{n_k}(t,x)\,\mathrm{d}x\,\mathrm{d}t + \varepsilon_{n_k} \tag{5.35.b}$$

where again $\varepsilon_{n_k}$ goes to zero as $(k \to \infty)$.

Now weak convergence of $(U^{n_k})_{k\in\mathbb{N}}$, $(\widetilde{U}^{n_k})_{k\in\mathbb{N}}$, and $(Z^{n_k})_{k\in\mathbb{N}}$ allows us to pass to the limit in (5.34) and in (5.35) as $(k \to \infty)$; in other words, we arrive at

$$\int_{\Omega\times(0,T)} Z\psi U\,\mathrm{d}(x,t) = -\int_{\Omega\times(0,T)} U^m \nabla_x \psi\,\mathrm{d}(x,t) - \int_{\Omega\times(0,T)} \upsilon\nabla_x\psi\,\mathrm{d}(x,t).$$

This means that we can identify $\widehat{V}$ with $\nabla(U^m)/U + \nabla\upsilon$. ■

With the estimates developed so far, we are finally in the position to proof our main result.

5.3.5 **Theorem** *Let $(u^n)_{n\in\mathbb{N}}$ be a sequence of spatial finite-volume discretisations $(u_i^n)_{1\le i\le n}$ as in* (5.3) *such that* Assumptions 5.1.2 *are satisfied. Let $(U^n)_{n\in\mathbb{N}}$ be the corresponding piecewise constant interpolants given by* (5.12). *Then there exists a subsequence $(U^{n_k})_{k\in\mathbb{N}}$ and a curve $U \in AC(\mathbb{R}_0^+, \mathscr{P}_2(\overline{\Omega}))$, independent of $T > 0$, with initial value $U(0) = U_0$, such that the following statements hold:*

(i) *For all times $t \in (0,T)$ the interpolant $U^{n_k}(t)$ converges weakly to $U(t)$ as $(k \to \infty)$;*

(ii) *the limit curve $U$ is a gradient flow in the EDE sense for $\phi = F_m + V$.*

Proof In this proof we collect all the prerequisites of Proposition 4.1.2 which in turn implies above theorem. To this aim, we may invoke Proposition 5.1.3 for every $T > 0$ to obtain a sequence $(\rho^n)_{n\in\mathbb{N}}$ of continuously differentiable curves $(0,T) \to \mathcal{P}^n$ which satisfy the energy dissipation equality (5.7).

In the next step, we set $X^n = \mathcal{P}^n$ and $X = \mathscr{P}_2(\overline{\Omega})$ and denote by $\iota^n : \mathcal{P}^n \to \mathscr{P}_2(\overline{\Omega})$ the mapping, induced by (5.12), which maps each $\rho^n$ to its corresponding interpolant $U^n$.

Let us verify that the slope $|\partial F_m|$ is a strong upper gradient. Indeed, Proposition 2.4.6.i implies that the Rényi entropy functional $F_m$ is convex along (generalised) geodesics in $(\mathscr{P}_2(\mathbb{R}^n), W_2)$. Hence, we can invoke Proposition 1.2.4.ii to obtain that $|\partial F_m|$ is a strong upper gradient.

The limit curve is constructed by successively invoking Proposition 5.3.1 for increasing times $T > 0$. In this way we can extract a suitable subsequence $U^{n_k}_{k\in\mathbb{N}}$, pointwise weakly convergent to a continuous limit curve $U : \mathbb{R}_0^+ \to \mathscr{P}_2(\Omega)$ such that (4.3) is satisfied. Note that Theorem 2.3.3 implies that $U$ belongs to $AC((0,T), \mathscr{P}_2(\overline{\Omega}))$ and the metric derivative of $U$ satisfies the following estimate

$$|\dot{U}|(t) \le \int_\Omega |V(t,x)|^2\, U(t,x)\,\mathrm{d}x \qquad \mathscr{L}^1\text{-a.e. } t \in (0,T). \tag{5.36}$$

On the other hand, we may identify the metric slope $|\partial F_m|$ with the continuous Fisher information by Lemma 2.4.9, i.e. the following identity holds:

$$|\partial F_m|^2(U) = \int_\Omega \frac{|\nabla(U^m)|}{U}\,\mathrm{d}x. \tag{5.37}$$

Recalling Proposition 5.3.4, both (5.36) and (5.37) imply

$$\frac{1}{2}\int_0^T |\dot{U}|^2 + |\partial F_m|^2 (U)\,\mathrm{d}t \leq \liminf_{k\to\infty} \int_0^T \mathcal{R}^{n_k}(u^{n_k}, \dot{u}^{n_k}) + \mathcal{I}^{n_k}(u^{n_k})\,\mathrm{d}t.$$

Recalling that the metric derivative and the slope in the space int$\mathcal{P}^n$ correspond to the discrete dissipation potential $\mathcal{R}^n$ and the discrete Fisher information $\mathcal{I}^n$, respectively, we have shown (4.5). Finally, to obtain the Γ-lim sup bound (4.4), we simply invoke (5.27). ■

In general, one cannot identify the limit curve $U$ obtained in above theorem as a solution of the corresponding porous medium equation by properties of abstract EDE gradient flows alone. To overcome this issue, one may for instance appeal to convergence results of the underlying finite-volume scheme.

In the context of gradient flows, is also possible to invoke the results of Section 2.4, usually obtained by subdifferential calculus in the 2-Wasserstein space over $\mathbb{R}^n$. Indeed, combining the the statement of Theorem 5.3.5 with Corollary 2.4.13 results directly in the following corollary.

**5.3.6 Corollary** *Let $U$ be the limit curve of Theorem 5.3.5. Then $U$ is the unique EVI gradient flow for $F_m$. Additionally, $U$ is a solution in the sense of distributions to the following porous medium equation with drift and non-flux Neumann boundary condition:*

$$\frac{\mathrm{d}}{\mathrm{d}t}U = \Delta(U^m) + \operatorname{div}(U\nabla\upsilon) \qquad \text{in } \mathbb{R}^+ \times \Omega.$$

## 5.4 Bibliographical Notes

This chapter is heavily inspired by the approach of Disser and Liero in [24] where a gradient flows for the one-dimensional Fokker-Plank equation with drift was approximated by their discrete relative entropy counterparts. To this aim, they use a simple finite-volume discretisation of the Fokker-Plank equation equation

$$\dot{U} = \operatorname{div}(\nabla U + U\nabla\upsilon) \qquad \text{in } \Omega \tag{5.38}$$

to induce a discrete relative entropy gradient flow structure from the Onsager point of view (3.29). Note that the mixed energy functional $\phi = F_1 + V$ corresponding to the Wasserstein gradient flow for (5.38) may be written as relative entropy in the form

$$\phi(\mu) = \int_\Omega U \log U \,\mathrm{d}x + \int_\Omega \upsilon\, U \,\mathrm{d}x = \int_\Omega U \log(U/w)\,\mathrm{d}x, \tag{5.39}$$

whenever $\mu$ is absolutely continuous with density $U$ on $\Omega$. Here $w(x) = e^{-\upsilon(x)}$ is the steady state of (5.38). There are several choices for approximating the flux in the corresponding finite-volume discretisation, which heavily influence the numerical stability of the scheme. See Bessemoulin-Chatard and Filbet [10] for a general finite-volume scheme corresponding to entropy functionals of of type (5.39), or Carrillo, Chertock and Huang [16] for another finite-volume scheme suited for nonlinear evolution equations possessing more general Wasserstein gradient flow structures. A recent survey for an overview on finite-volume methods is provided by Barth and Ohlberger [7].

In the recent article [28], Erbar et al. follow along lines, similar to the ones in this chapter, to investigate the limit of energy gradient flow structures for discrete McKean-Vlasov equations. Essentially the same approach, relying on convergence of discrete gradient structures as in Chapter 3, was taken by Fathi and Simon in [31] to study macroscopic hydrodynamic behaviour of interacting particle systems by passing to the limit in gradient flow structures related to the particle systems.

# 6 Limit Passage of EVI Gradient Flows for Discrete Heat Equations

## 6.1 Finite-Volume Discretisation Revisited

The aim of this chapter is to establish the limit curve of the spatial finite-volume discretisations as gradient flow in the much stronger EVI sense by reinforcing the main result of the previous chapter, Theorem 5.3.5. To actually obtain such a result we will use Proposition 4.2.2 as abstract backbone. To this end, we will need to exploit geodesic convexity of the underlying functional. However, we already saw in Example 3.3.4 that the discrete Rényi entropy functional $\mathcal{F}_n^m$ need not be geodesically $\kappa$-convex for every $m \in (0,2]$. Nevertheless, we will see that geodesic convexity of $\mathcal{F}_n^m$ can be obtained at least in the homogeneous case when $m = 1$.

We begin by recalling the finite-volume scheme introduced in Section 5.1 for $m = 1$ and vanishing potential $\upsilon \equiv 0$. In this case, the porous medium equation (5.1) becomes the homogeneous *heat equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t,x) = \Delta u(t,x) \qquad \forall (t,x) \in (0,T) \times [0,1]. \tag{6.1}$$

Again, we assume that (6.1) is subject to the non-flux Neumann boundary condition

$$\nabla u(t,0) = \nabla_x u(t,1) = 0 \qquad \forall t \in (0,T).$$

For simplicity, we will work with an equidistant spatial finite-volume discretisation in this chapter. Furthermore, it is convenient to swap the rôles of the $x_i^n$ and the $\sigma_i^n$: First we set $\sigma_i^n = i/n$ for $i \in \{0,\dots n\}$ to obtain an equidistant partition of $\Omega := (0,1)$. Then we assume that $x_i^n$ is a midpoint between $\sigma_{i-1}^n$ and $\sigma_i^n$; thus, arriving at

$$x_i^n = \frac{\sigma_{i-1}^n + \sigma_i^n}{2} = \frac{i - 1/2}{n} \qquad \forall i \in \{0,\dots n\}.$$

As usual, $\omega_i^n = [\sigma_{i-1}^n, \sigma_i^n)$ denotes a control volume with length $h_i^n = 1/n$.

Given (6.1) in the integrated form

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\omega_i^n} u(t,x)\,\mathrm{d}x = u'(t,\sigma_i^n) - u'(t,\sigma_{i-1}^n),$$

the finite-volume scheme is supposed to approximate the integral $\int_{\omega_i^n} u(t,x)\,\mathrm{d}x$ by $u_i^n(t)$, and the flux $\nabla_x u(t, x\sigma_i^n)$ by

$$\frac{1}{x_{i+1}^n - x_i^n}\left(\frac{u_{i+1}^n}{h_{i+1}^n} - \frac{u_i^n}{h_i^n}\right).$$

Since $h_i^n = h_{i+1}^n = x_{i+1}^n - x_i^n = 1/n$, this means that the rate coefficients take the form

$$\alpha_0 = \beta_0 = 0, \qquad \alpha_i = \beta_i = n^2 \quad \text{for } i \in \{1,\dots n-1\}, \qquad \alpha_n = \beta_n = 0.$$

With this notation, we may write (6.1) as

$$\dot{u}_i^n = \alpha_{i-1}u_{i-1}^n - \big(\alpha_i + \beta_{i-1}\big)u_i^n + \beta_i u_{i+1}^n \qquad \forall i \in \{1,\dots n\}. \tag{6.2}$$

Introducing the infinitesimal generator

$$\mathcal{Q} = n^2 \begin{pmatrix} -1 & 1 & 0 & 0 & \dots & 0 \\ 1 & -2 & 1 & 0 & \dots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \dots & 0 & 1 & -2 & 1 \\ 0 & \dots & 0 & 0 & 1 & -1 \end{pmatrix}, \tag{6.3}$$

the equation in (6.2) may be written in even more compact matrix-vector form $\dot{u}^n = \mathcal{Q}u^n$.

We consider the same piecewise constant interpolant $U^n : \Omega \to \mathbb{R}$ was already introduced in Section 5.2:

$$U^n(x) = \rho_i = \frac{u_i^n}{h_i^n} \qquad \text{for } x \in \omega_i^n. \tag{6.4}$$

Finally, since $m = 1$ and $\upsilon \equiv 0$, the assumptions we required in Section 5.1 simplify to the following conditions on the initial condition for the spatial finite-volume scheme.

**6.1.1 Assumptions** We make the following assumptions on the homogeneous heat equation in (6.1):
Re initial condition: Let $U_0$ be a probability density with respect to the Lebesgue measure on $\Omega$ such that $U_0 \log U_0 \in L^1(\Omega)$. Then the initial value of the finite-volume scheme is given by

$$u_i^n(0) = \int_{\omega_i^n} U_0(x)\,\mathrm{d}x. \tag{6.5}$$

## 6.2 Gromov-Hausdorff Convergence of the Discrete Transportation Metrics

In this section we are going to establish compatibility between the discrete transportation metrics $\mathcal{W}$ and the 2-Wasserstein metric $W_2$ for infinitesimal generators of the specific form (6.3). To this aim, we will use the notion of Gromov-Hausdorff convergence in the form of $\varepsilon$-isometries which were already introduced in Section 4.2. Indeed, it turns out that the piecewise constant discretisation, used to discretise the initial condition (6.5), is an appropriate candidate for an $\varepsilon$-isometry from $\mathscr{P}_2(\overline{\Omega})$ to $\mathcal{P}^n$ as long as the probability measure in $\mathscr{P}_2(\overline{\Omega})$ has a suitable density.

In order to remove such restrictions on the domain of such prospective $\varepsilon$-isometries, we will first regularise any probability measures in $\mathscr{P}_2(\overline{\Omega})$ before applying the piecewise constant discretisation. Regularisation will be provided by convolution with a rescaled version of the heat kernel on $\mathbb{R}$.

Let us recall that that the ***heat kernel*** on the real line is a Gaussian function defined by

$$h_t(x) := \frac{1}{\sqrt{4\pi t}} e^{-x^2/(4t)} \qquad \forall (t, x) \in \mathbb{R}^+ \times \mathbb{R}.$$

The ***heat semigroup*** with respect to a Borel measure $\mu$ with finite total variation on $\mathbb{R}$ is then given by convolution with with the heat kernel, to wit

$$H_t\mu(x) := (h_t * \mu)(x) = \int_{\mathbb{R}} h_t(x - y)\,\mathrm{d}\mu(y).$$

In order to put those definitions into our framework of probability measures on $\overline{\Omega}$, we also introduce the following rescaled variant of the heat semigroup: Every $\mu \in \mathscr{P}_2(\overline{\Omega})$ can be trivially extended to a probability measure (also denoted by $\mu$) on $\mathbb{R}$ with $\operatorname{supp}\mu \subseteq \overline{\Omega}$. Therefore, we may write $\bar{H}_t\mu := \bar{h}_t * \mu$, where $\bar{h}_t := \cancel{c}_{\mu,t} h_t$ is the ***rescaled heat kernel*** with $\cancel{c}_{\mu,t} > 0$ such that $\bar{H}_t\mu(x)\,\mathrm{d}x$ belongs to $\mathscr{P}_2(\overline{\Omega})$. To keep the notation simple, we will identify the density $\bar{H}_t\mu$ with the corresponding probability measure.

Now we are in the position to formulate the main result of this section.

**6.2.1 Theorem** *Let $(\mathcal{X}^n, \mathcal{Q}, \pi)$ be an irreducible continuous-time Markov chain with infinitesimal generator $\mathcal{Q}$ of the form* (6.3)*. Then for $0 < m \leq 2$, the metric spaces $(\mathcal{P}^n, \mathcal{W})$ induced by the weight function $\theta_1$ converge to $(\mathscr{P}_2(\overline{\Omega}), W_2)$ in the following sense:*
*For every $\varepsilon > 0$ there exist $n \in \mathbb{N}$ and $s > 0$ such that the map from $\mathscr{P}_2(\overline{\Omega})$ to $\mathcal{P}^n$ given by*

$$\mu \mapsto \frac{1}{h_i^n} \int\limits_{\omega_i^n} \bar{H}_s \mu(x)\,\mathrm{d}x \tag{6.6}$$

*is an $\varepsilon$-isometry.*

The proof of the result above essentially follows along the lines of the argument in [35]. There Gromov-Hausdorff convergence of the $d$-dimensional discrete torus endowed with corresponding discrete transportation metrics was considered instead of $(\mathcal{P}^n, \mathcal{W})$. Here we will just give a sketch to point out the main alterations needed to make the argument in [35] go through in our case as well.

We start by restating some well known facts about the heat semigroup on the real line for $(\bar{H}_t)_{t\geq 0}$.

**6.2.2 Facts** (Rescaled heat semigroup)

(i) Given a rescaled heat semigroup $(\bar{H}_t\mu)_{t\geq 0}$, the probability measure $\mu \in \mathscr{P}_2(\overline{\Omega})$ can be recovered in the sense that $\bar{H}_t\mu \xrightarrow{w^*} \mu$ as $(t \searrow 0)$. This follows from the fact that $t \mapsto \cent_{\mu,t}$ is a strictly increasing function with $\lim_{t\searrow 0} \cent_{\mu,t} = 1$. In addition, we have the equality $\|\bar{H}_t\mu\|_1 = \|\hbar_t\|_1\,\|\mu\|_{\mathrm{TV}} = \|\hbar_t\|_1 = \cent_{\mu,t}$ with respect to the Lebesgue measure on $\mathbb{R}$.

Note that if we set $\cent_t := 1/\|\hbar_t \mathbb{1}_{[0,1]}\|_1$, then we have found a constant independent of $\mu$ such that $\cent_{\mu,t} \leq \cent_t$ and $\lim\limits_{t\searrow 0} \cent_t = 1$

(ii) For any positive time $t > 0$, the rescaled heat flow $\bar{H}_t\mu$ is nowhere vanishing on $\mathbb{R}$, i.e.

$$\bar{H}_t\mu(x) \geq \inf_{y\in\mathbb{R}} \hbar_t(y) > 0 \qquad \forall (t,x) \in \mathbb{R}^+ \times \mathbb{R}. \tag{6.7}$$

This property is also known as *infinite speed of propagation* of the heat semigroup.

Another property which follows immediately from the definition of $\bar{H}_t\mu$ as convolution is Lipschitz continuity with respect to the spatial variable:

$$|\bar{H}_t\mu(x) - \bar{H}_t\mu(y)| \leq C_t\,|x-y| \qquad \forall x,y \in \mathbb{R} \tag{6.8}$$

with $C_t = \sup\limits_{x\in\mathbb{R}} \nabla_x \hbar_t(x) < +\infty$ for all times $t > 0$.

(iii) The flow curve $t \mapsto \bar{H}_t\mu$ is $1/2$-Hölder continuous with respect to the 2-Wasserstein distance. To see this, one can exploit the convexity of $(\mu,\nu) \mapsto W_2^2(\mu,\nu)$ – a property already noticed in (2.13) – to apply the following version of the Jensen inequality for random elements taking values in $(\mathcal{M}(\overline{\Omega}), \|\cdot\|_{\mathrm{TV}})$, the Banach space of signed measures with finite total variation on $\overline{\Omega}$:

$$f\Big(\int\limits_{\overline{\Omega}} \psi\,\mathrm{d}\mu\Big) \leq \int\limits_{\overline{\Omega}} f \circ \psi\,\mathrm{d}\mu, \tag{6.9}$$

where $\mu \in \mathscr{P}_2(\overline{\Omega})$, $\psi$ is an integrable random element on $(\overline{\Omega}, \mu)$, $\mu$-a.s. taking values in a convex subset $K \subseteq \mathcal{M}(\overline{\Omega}) \times \mathcal{M}(\overline{\Omega})$, and $f : K \to \mathbb{R}$ is a measurable convex function.

Now setting $f = W_2^2$ and $\psi(x) = (\bar{H}_t\delta_x, \delta_x)$ in (6.9) and using the fact that the 2-Wasserstein distance to a Dirac measure can be expressed explicitly, implies the desired inequality

$$W_2^2(\bar{H}_t\mu, \mu) = W_2^2\Big(\int\limits_{\overline{\Omega}} \bar{H}_t\delta_x\,\mathrm{d}\mu(x), \int\limits_{\overline{\Omega}} \delta_x\,\mathrm{d}\mu(x)\Big) \leq \int\limits_{\overline{\Omega}} W_2^2(\bar{H}_t\delta_x, \delta_x)\,\mathrm{d}\mu(x) \leq \tag{6.10.a}$$

$$\leq c_T \int\limits_{\overline{\Omega}} \int\limits_{\mathbb{R}} |y-x|^2\, \hbar_t(y-x)\,\mathrm{d}y\,\mathrm{d}\mu(x) = c_T \int\limits_{\overline{\Omega}} \int\limits_{\mathbb{R}} t z^2 \hbar_1(z)\,\mathrm{d}z\,\mathrm{d}\mu(x) = \widehat{C}_T^2 t \tag{6.10.b}$$

with a constant $\widehat{C}_T^2 := c_T \int_{\mathbb{R}} x^2 \hbar_1(x)\,\mathrm{d}x$ for all $t \in [0,T]$.

(iv) Let $\mu : [0,1] \to \mathscr{P}_2(\overline{\Omega})$ be a geodesic connecting $\mu_0$ to $\mu_1$ with corresponding Borel vector fields $v_t \in \mathcal{T}(\overline{\Omega})$ such that $\int_0^1 \|v_t\|_{L^2(\mu(t))}\,\mathrm{d}t < +\infty$ and the continuity equation **(2.9)** holds in the sense of distributions. Then $(\rho_{s,t}, V_{s,t})_{t\in[0,1]}$ with $\rho_{s,t} = \mathcal{H}_s\mu(t)$ and $V_{s,t} = \mathcal{H}_s(v_t\mu(t))$ provides a solution to the following variant of the continuity equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_{s,t} + \mathrm{div} V_{s,t} = 0 \qquad \forall s \geq 0.$$

This follows directly from

$$\mathrm{div} V_{s,t} = \mathrm{div}\big(\hbar_s * (v_t\mu(t))\big) = \hbar_s * \mathrm{div}(v_t\mu(t)) = \mathcal{H}_s \mathrm{div}(v_t\mu(t))$$

in the sense of distributions.

If in addition the vector fields $v_t$ archive the minimum in the Benamou-Brennier formula **(2.10)**, then we have the following inequality which – despite being less sharp than the corresponding counterpart in **[35]** – is sufficient for our purpose:

$$\int_0^1 \int_{\overline{\Omega}} \frac{V_{s,t}^2(x)}{\rho_{s,t}(x)}\,\mathrm{d}x\,\mathrm{d}t \leq \cent_s W_2^2(\mu_0,\mu_1). \tag{6.11}$$

To see this, we use the convexity of the mapping $f(z,a) := z^2/a$ on $\mathbb{R}\times\mathbb{R}^+$ to invoke the Jensen inequality with respect to the measure $\mathrm{d}\eta_t(y) = \hbar_s(x-y)\,\mathrm{d}\mu_t(y)$ in the computation

$$\begin{aligned}\frac{V_{s,t}^2(x)}{\rho_{s,t}(x)} &= f\Big(\int_{\mathbb{R}} v_t(y)\,\mathrm{d}\eta_t(y), \int_{\mathbb{R}} 1\,\mathrm{d}\eta_t(y)\Big) \leq \\ \leq \int_{\mathbb{R}} f(v_t(y),1)\,\mathrm{d}\eta_t(y) &= \int_{\mathbb{R}} v_t(y)\hbar_s(x-y)\,\mathrm{d}\mu_t(y).\end{aligned}$$

Now an integration with respect to $x$ over $\overline{\Omega}$ in the inequality above, followed by an application of Fubini's theorem leads to the following estimate

$$\int_{\overline{\Omega}} \frac{V_{s,t}^2(x)}{\rho_{s,t}(x)}\,\mathrm{d}x \leq \int_{\mathbb{R}} v_t^2(y) \int_{\overline{\Omega}} \hbar_s(x-y)\,\mathrm{d}x\,\mathrm{d}\mu_t(y) \leq \cent_{\mu_t,s} \int_{\mathbb{R}} v_t^2(y)\,\mathrm{d}\mu_t(y).$$

Finally, another integration with respect to $t$ over $[0,1]$ and using the fact that $v_t$ archives the minimum in the Benamou-Brennier formula **(2.10)** establish the sought inequality.

In addition to the rescaled heat semigroup $(\mathcal{H}_t)_{t\geq 0}$, we also introduce its discrete counterpart on $(\mathcal{P}^n, \mathcal{Q}, h^n)$: The ***discrete heat semigroup*** acting on functions $\zeta : \mathcal{X}^n \to \mathbb{R}$ is defined by

$$\mathcal{H}_t^n \zeta_i := \sum_{j=1}^n \exp(t\mathcal{Q})_{i-j,n}\zeta_j, \tag{6.12}$$

where we assume that $i \mapsto \exp(t\mathcal{Q})_{in}$ is extended periodically to $\mathbb{Z}$. Let us collect some facts about the discrete heat semigroup $(\mathcal{H}_t^n)_{t\geq 0}$.

**6.2.3 Facts** (Discrete heat semigroup)

(i) It can be shown that the eigenvalues and the corresponding eigenvectors of the infinitesimal generator $\mathcal{Q}$ are given by

$$\lambda^k = 2n^2\left(\cos\left(\frac{k\pi}{n+1/2}\right) - 1\right) \qquad \text{and} \qquad w_j^k = \begin{cases} \sin\left(\frac{k(j-1/2)\pi}{n}\right) & \text{if } k < n, \\ 1 & \text{if } k = 1, \end{cases}$$

respectively. As a result, we may express the the matrix exponential in (6.12) more concisely as

$$\exp(t\mathcal{Q})_{in} = \sum_{k=1}^{n} e^{t\lambda_k} w_i^k.$$

Simple consequences of this representation are the Lipschitz estimates

$$\left|\mathcal{H}_t^n \rho_i - \mathcal{H}_t^n \rho_j\right| \leq \frac{\check{C}_t}{n} |i-j| \quad \text{and} \quad \left|\mathcal{H}_t^n \rho_i - \mathcal{H}_t^n \rho_j\right| \leq \left|\rho_i - \rho_j\right| \qquad \forall \rho \in \mathcal{P}^n$$

for a constant $\check{C}_t > 0$.

(ii) It is possible to obtain a discrete counterpart to the estimate in (6.11). For a probability density $\rho \in \mathcal{P}^n$ and a *momentum vector field* $V^n : \mathcal{X}^n \times \mathcal{X}^n \to \mathbb{R}$ define the following *action functional:*

$$\mathcal{A}^n(\rho, V^n) := \sum_{i,j} \frac{\left|V_{ij}^n\right|^2}{\hat{\rho}_{ij}} \mathcal{Q}_{ij} h_i^n. \tag{6.13}$$

Now using how the semigroup $(\mathcal{H}_t^n)_{t \geq 0}$ acts on $V^n$, namely

$$\mathcal{H}_t^n V_{i,i\pm 1}^n := \sum_{k=1}^{n} \exp(t\mathcal{Q})_{i-k,n} V_{k,k\pm 1}^n,$$

yields the inequality

$$\mathcal{A}^n(\mathcal{H}_t^n \rho, \mathcal{H}_t^n V^n) \leq \mathcal{A}^n(\rho, V^n) \qquad \forall t \geq 0. \tag{6.14}$$

Indeed, the infinitesimal generator $\mathcal{Q}_{ij}$ vanishes except for $i = j \pm 1$. Therefore, the Jensen inequality for the 1-homogeneous convex mapping $(x, a, b) \mapsto x^2/\theta_1(a, b)$ on $\mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+$ implies

$$\begin{aligned} \Big|\sum_k \exp(t\mathcal{Q})_{i-k,n} V_{k,k\pm 1}^n\Big|^2 \theta_1\Big(\sum_k \exp(t\mathcal{Q})_{i-k,n}\rho_k, 1/n \sum_k \exp(t\mathcal{Q})_{i\pm 1-k,n}\rho_k\Big)^{-1} &\leq \\ \leq \sum_k \frac{\left|V_{k,k\pm 1}^N\right|^2}{\theta_1(\rho_k, \rho_{k\pm 1})} \exp(t\mathcal{Q})_{i-k,n} & \end{aligned}$$

Then, using that the Markov semigroup satisfies $\sum_i \exp(t\mathcal{Q})_{i-k,n} = 1$, the inequality in (6.14) is immanent.

The facts stated above provide all the tools necessary to carry out the proof of Theorem 6.2.1 in our setting adapted from the original argument for convergence of transportation metrics on the discrete torus. We will give only a condensed sketch thereafter; the reader may consult [35] for further details.

Sketch of proof of Theorem 6.2.1 To be in line with the notation used in [35], we write

$$P_n(\mu)_i := n\mu(\omega_i^n)$$

for the discrete approximation of a probability measure $\mu \in \mathscr{P}^n(\overline{\Omega})$, whereas

$$Q_n(\rho)(x) := \rho_i \qquad \text{for } x \in \omega_i^n$$

denotes the piecewise constant interpolant of a discrete probability density $\rho \in \mathcal{P}^n$.
In order to proof convergence of the spaces $(\mathcal{P}^n, \mathcal{W})$ to the 2-Wasserstein space over $\overline{\Omega}$ in the sense of Gromov-Hausdorff, we will compare the 2-Wasserstein distance with the discrete transportation

distance $\mathcal{W}$ on $\mathcal{P}^n$. To this aim, we will collect some useful estimates in the first part of the sketch, before we are going to show that the mapping (6.6) is an $\varepsilon$-isometry in the second part.

(i) Estimates for the transportation distances: First, we need a suitable upper bound for $\mathcal{W}$ in terms of the 2-Wasserstein distance. Indeed, a rather straightforward argument shows that there exists a constant $C_s > 0$ such that

$$\mathcal{W}\big(P_n(\mathcal{H}_s\mu_0), P_n(\mathcal{H}_s\mu_1)\big) \leq ¢_s W_2(\mu_0, \mu_1) + \frac{C_s}{\sqrt{n}} \qquad \forall \mu_0, \mu_1 \in \mathscr{P}_2(\overline{\Omega}), \tag{6.15}$$

with the constant $¢_t > 0$ coming from (6.11). In fact, the most involving part of the proof is to establish a converse bound of $W_2$ by means of $\mathcal{W}$. To this aim, it is convenient to work instead with the *harmonic mean*

$$\theta_{-1}(s,t) = \frac{2st}{s+t},$$

being related to $\theta_1$ by the elementary inequalities

$$\theta_{-1}(s,t) \leq \theta_1(s,t) \qquad \text{and} \qquad 1 - \frac{\theta_{-1}(s,t)}{\theta_1(s,t)} \leq \frac{|s-t|^2}{st}.$$

Using these relations between those two weight functions, one can establish the following estimate as an intermediate step: For all densities $\rho_0, \rho_1$ in

$$\mathcal{P}_\delta^n := \Big\{\rho \in \mathcal{P}^n : \rho_i \geq \delta \text{ and } \big|\rho_i - \rho_j\big| = (\delta n)^{-1}\,|i-j| \quad \forall i,j \in \mathcal{X}^n\Big\}$$

we have the inequality

$$W_2\big(Q_n(\rho_0), Q_n(\rho_1)\big) \leq \Big(1 - \frac{1}{\delta^4 n^2}\Big)^{-1/2} \mathcal{W}_\delta(\rho_0, \rho_1). \tag{6.16}$$

Here the distance function $\mathcal{W}_\delta$ on $\mathcal{P}_\delta^n$ is defined by the Benamou-Brenier formula

$$\mathcal{W}_\delta^2(\rho_0, \rho_1) := \inf\Big\{\int_0^1 \mathcal{A}^n(\rho(t), V^n(t))\,\mathrm{d}t\Big\}$$

with the infimum being taken amongst all pairs of continuous curves $\rho : [0,1] \to \mathcal{P}_\delta^n$ connecting $\rho_0$ to $\rho_1$ and integrable momentum fields $t \mapsto V^n(t) := \nabla\psi_{ij}(t)/\hat{\rho}_{ij}(t)$, being distributional solutions to the discrete continuity equation in the form of

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_i(t) + \sum_j V_{ij}^n(t)\mathcal{Q}_{ij} = 0 \qquad \forall t \in (0,1). \tag{6.17}$$

Finally, a comparison of the distances $\mathcal{W}_\delta$ and $\mathcal{W}_n$ – the most involved part of the argument – yields that given $\varepsilon_0, \delta > 0$, there exists $\overline{\delta} > 0$ such that

$$\mathcal{W}_{\overline{\delta}}(\rho_0, \rho_1) \leq \mathcal{W}(\rho_0, \rho_1) + \varepsilon_0. \tag{6.18}$$

(ii) Constructing the $\varepsilon$-isometry: Having collected all the estimates necessary, we are in the position to sketch that the mapping given in (6.6), namely $\mu \mapsto P_n(\mathcal{H}_s\mu)$ is an $\varepsilon$-isometry from $\mathscr{P}_2(\overline{\Omega})$ to $\mathcal{P}^n$: Note that we may combine the inequalities in (6.16) and (6.18) to obtain

$$W_2\big(Q_n(\rho_0), Q_n(\rho_1)\big) \leq \Big(1 - \frac{1}{\overline{\delta}(\varepsilon_0, s)^4 n^2}\Big)^{-1/2}\big(\mathcal{W}(\rho_0, \rho_1) + \varepsilon_0\big), \tag{6.19}$$

which is valid only for probability densities in $\rho_0, \rho_1 \in \mathcal{P}_\delta^n$. Here the regularising properties of the rescaled heat semigroup $(\mathcal{H}_t)_{t\geq 0}$ come into play. Moreover, both (6.7) and (6.8) are preserved by the

discretisation mappings $P_n$; thus, ensuring that the probability densities $\rho_0 = P_n(H_s\mu_0)$ and $\rho_1 = P_n(H_s\mu_1)$ belong to $\mathcal{P}^n_\delta$ without need for further assumptions on the measures $\mu_0, \mu_1 \in \mathscr{P}_2(\overline{\Omega})$.
In addition, Proposition 2.2.4.ii allows us to control the error of the approximation $Q_n \circ P_n$ in terms of the 2-Wasserstein distance, to wit

$$W_2^2\big(Q_n \circ P_n(H_s\mu_k), H_s\mu_k\big) \le \sum_{i=1}^n H_s\mu_k(\omega_i^n)\,\mathrm{diam}\,\omega_i^n \le \frac{1}{N^2} \qquad \forall k \in \{0,1\}, \tag{6.20}$$

where we used that $|H_s\mu_i - \mu|(\omega_i^n) \le H_s\mu_i(\omega_i^n)$, due to the fact that both measures agree for each control volume $\omega_i^n$.
Combining this inequality with (6.19) and the Hölder estimate in (6.10) for our particular choices for $\rho_0$ and $\rho_1$, we arrive at

$$W_2(\mu_0,\mu_1) \le \left(1 - \frac{1}{\overline{\delta}(\varepsilon_0,s)^4 n^2}\right)^{-1/2}\big(\mathcal{W}\big(P_n(H_s\mu_0),P_n(H_s\mu_1)\big)+\varepsilon_0\big)+2\widehat{C}_T\sqrt{s}+\frac{2}{n} \quad \forall \mu_0,\mu_1 \in \mathscr{P}_2(\overline{\Omega}). \tag{6.21}$$

Now using the fact that the diameter of the spaces $(\mathcal{P}^n, \mathcal{W})$ is uniformly bounded for all $n \in \mathbb{N}$, both (6.15) and (6.21) yield for every $\varepsilon > 0$ the desired estimate

$$\big|W_2(\mu),\mu_1) - \mathcal{W}\big(P_n(H_s\mu_0),P_n(H_s\mu_1)\big)\big| \le \varepsilon \qquad \forall \mu_0,\mu_1 \in \mathscr{P}_2(\overline{\Omega})$$

for $n \in \mathbb{N} > 0$ sufficiently large and $\varepsilon_0 > 0$ sufficiently small.

It remains to show that given $\varepsilon > 0$, for every $\rho \in \mathcal{P}^n$ there exists $\mu \in \mathscr{P}_2(\overline{\Omega})$ such that

$$\mathcal{W}\big(\rho, P_n(H_s\mu)\big) \le \varepsilon.$$

In order to see this, set $\mu = Q_n(\rho)$. Then we may use the fact that $P_n \circ Q_n = \mathrm{Id}$ on $\mathcal{P}^n$ together with Proposition 3.1.8 to obtain the following bound in terms of the 2-Wasserstein distance with respect to $|\cdot|$ on $\mathcal{X}^n$ for some constant $C > 0$:

$$\mathcal{W}\big(\rho,P_n(H_s\mu)\big) = \mathcal{W}\big(P_n \circ Q_n(\rho), P_n(H_s\mu)\big) \le \frac{C}{n} W_2^{\mathrm{gra}}\big(P_n \circ Q_n(\rho), P_n(H_s\mu)\big).$$

To proceed, we need to compare the right-hand side of this estimate with the 2-Wasserstein distance on $\mathscr{P}_2(\overline{\Omega})$: Define a mapping $T_n : \overline{\Omega} \to \mathcal{X}^n$ by setting $T_n(x) := i$ whenever $x \in \omega_i^n$. Then we have the obvious inequality

$$|T_n(x) - T_n(y)| \le n\,|x-y| + 1 \qquad \forall x,y \in \overline{\Omega}.$$

Now we may fix an optimal plan $\sigma \in \Pi_{\mathrm{opt}}(Q_n(\rho), H_s\mu)$ with respect to the 2-Wasserstein distance on $\mathscr{P}_2(\overline{\Omega})$ to infer that $(T_n,T_n)_\#\sigma$ is an admissible plan in $\Pi\big(P_n \circ Q_n(\rho), P_n(H_s\mu)\big)$. As a consequence, the triangle inequality in $L^2(\sigma)$ implies

$$\begin{aligned} W_2^{\mathrm{gra}}\big(P_n \circ Q_n(\rho), P_n(H_s\mu)\big)^2 &\le \int_{\overline{\Omega}\times\overline{\Omega}} |T_n(x) - T_n(y)|^2 \,\mathrm{d}\sigma(x,y) \le \\ &\le n^2 \int_{\overline{\Omega}\times\overline{\Omega}} |x-y|^2 \,\mathrm{d}\sigma(x,y) + 1 = n^2 W_2(Q_n(\rho), H_s\mu)^2 + 1. \end{aligned}$$

Using this estimate and (6.10), we conclude that

$$\mathcal{W}\big(\rho,P_n(H_s\mu)\big) \le C W_2(Q_n(\rho), H_s\mu) + \frac{C}{n} \le C\left(\widehat{C}_T\sqrt{s} + \frac{1}{n}\right) \le \varepsilon \tag{6.22}$$

for $n \in \mathbb{N} > 0$ sufficiently large and $s > 0$ sufficiently small. ∎

Note that the maps in (6.6) are defined from $\mathcal{P}_2(\overline{\Omega})$ to $\mathcal{P}^n$, whereas in order to apply Proposition 4.2.2, we require $\varepsilon$-isometries from $\mathcal{P}^n$ to $\mathcal{P}_2(\overline{\Omega})$. This leads to the following concept of *almost inverse* mappings.

**6.2.4 Definition** Let $\iota : Y \to X$ be an $\varepsilon$-isometry between metric spaces $(Y, d_Y)$ and $(X, d_X)$. Then a mapping $\iota' : X \to Y$ is called an ***ε-inverse*** of $\iota$ if

$$d_X\big(\iota \circ \iota'(x), x\big) \le \varepsilon \quad \forall x \in X \qquad \text{and} \qquad d_Y\big(\iota' \circ \iota(y), y\big) \le 3\varepsilon \quad \forall y \in Y.$$

It follows directly from the definitions above that every $\varepsilon$-inverse is a $4\varepsilon$-isometry as well.

The next corollary shows that it is straightforward to obtain an $\varepsilon$-inverse of (6.6) with Theorem 6.2.1 already being established.

**6.2.5 Corollary** *Let $(\mathcal{X}^n, \mathcal{Q}, \pi)$ be an irreducible continuous-time Markov chain with infinitesimal generator $\mathcal{Q}$ of the form* (6.3) *and weight function $\theta_m$ for $0 < m \le 2$. Then there exist $n \in \mathbb{N}$ and $s > 0$ such that the map*

$$(\rho_i)_{i \le n} \mapsto U^n, \tag{6.23}$$

*where $U^n : \Omega \to \mathbb{R}$ is the piecewise constant interpolant of $\rho_i$ as in* (5.12)*, defines an $\varepsilon$-inverse of* (6.6)*. In particular,* (6.23) *is a $4\varepsilon$-isometry from $\mathcal{P}^n$ to $\mathcal{P}_2(\overline{\Omega})$.*

Proof For every given $\varepsilon > 0$, we have to verify that both inequalities in Definition 6.2.4 hold: The first inequality takes the form

$$\mathcal{W}\big(P_n(\mathcal{H}_s Q_n(\rho)), \rho\big) \le \varepsilon \qquad \forall \rho \in \mathcal{P}^n$$

for $n \in \mathbb{N} > 0$ sufficiently large and $s > 0$ sufficiently small, which means that we can just follow along the same lines we used to show (6.22).

On the other hand, in order to prove the second inequality

$$W_2\big(Q_n \circ P_n(\mathcal{H}_s \mu), \mu\big) \le \varepsilon \qquad \forall \mu \in \mathcal{P}_2(\overline{\Omega}),$$

it is enough to invoke (6.20) and (6.10). ■

## 6.3 The Limit Passage in the EVI Case

**6.3.1 Proposition** (Geodesic convexity for tridiagonal generators) *Provided that the discrete transportation metric $\mathcal{W}$ is induced by the weight function $\theta_1$ with an infinitesimal generator $\mathcal{Q}$ of the form* (6.3)*, the discrete Shannon entropy $\mathcal{F}_1^n$ is geodesically convex.*

Proof The strategy of the proof is to show that $\operatorname{Hess} \mathcal{F}_1^n(\rho)$ is positive semidefinite for every point $\rho$ in the interior of $\mathcal{P}^n$. Then we can invoke Proposition 3.3.2 to deduce that $\mathcal{F}_1^n$ is geodesically convex in $(\mathcal{P}^n, \mathcal{W})$. To this aim, we make use of the explicit expression of the Hessian of $\mathcal{F}_1^n$ in form of the quadratic form $\mathcal{B}_1$ as given in Lemma 3.3.1.

To compute $\mathcal{B}_1$, we use the symmetry of $\mathcal{Q}$ together with Fact 3.1.2.vii to arrive at

$$\mathcal{B}_1(\rho, \nabla\psi) = \frac{1}{n}\sum_{i=1}^{n-1} (\psi_i - \psi_{i+1})^2 \left(\partial_1\theta_1(\rho_{i+1}, \rho_i)\frac{\rho_i + \rho_{i+2}}{2} + \partial_2\theta_1(\rho_{i+1}, \rho_i)\frac{\rho_{i-1} + \rho_{i+1}}{2}\right) + \tag{6.24.a}$$

$$+ \frac{1}{n}\sum_{i=1}^{n-1} \hat{\rho}_{i+1,i}\big((\psi_i - \psi_{i+1})^2 + (\psi_i - \psi_{i+1})(\psi_{i+2} - \psi_{i+1} + \psi_i - \psi_{i-1})\big), \tag{6.24.b}$$

where we have to set $\rho_0 = \rho_1,\ \rho_{n+1} = \rho_n$ and $\psi_0 = \psi_1,\ \psi_{n+1} = \psi_n$ to obtain correct boundary terms.

It remains to show that $\mathcal{B}_1(\rho, \nabla\psi)$ is non-negative for all $\rho \in \operatorname{int}\mathcal{P}^n$ and $\nabla\psi \in T_\rho\mathcal{P}^n$: The only possibly negative terms are the mixed products of discrete gradients in (6.24.b), which can be estimated as

$$(\psi_i - \psi_{i+1})(\psi_{i+2} - \psi_{i+1} + \psi_i - \psi_{i-1}) \geq -(\psi_i - \psi_{i+1})^2 - \frac{1}{2}(\psi_{i+2} - \psi_{i+1})^2 - \frac{1}{2}(\psi_i - \psi_{i-1})^2 \tag{6.25}$$

by means of the AM-GM inequality. The first term on the right-hand side of this inequality cancels out with the corresponding non-negative term in (6.24.b). To control the other two terms, we may invoke Fact 3.1.2.viii to infer

$$\partial_1\theta_1(\rho_{i+1}, \rho_i)\rho_i + \partial_2\theta_1(\rho_{i+1}, \rho_i)\rho_{i-1} \geq \hat{\rho}_{i,i-1}$$

and

$$\partial_1\theta_1(\rho_{i+1}, \rho_i)\rho_{i+2} + \partial_2\theta_1(\rho_{i+1}, \rho_i)\rho_{i+1} \geq \hat{\rho}_{i+1,i+2}$$

which in turn, after some index shifts, imply for the terms in (6.24.a) the estimates

$$(\psi_{i+1} - \psi_{i+2})^2\left(\partial_1\theta_1(\rho_{i+2}, \rho_{i+1})\frac{\rho_{i+1}}{2} + \partial_2\theta_1(\rho_{i+2}, \rho_{i+1})\frac{\rho_i}{2}\right) \geq \frac{1}{2}(\psi_{i+1} - \psi_{i+2})^2\hat{\rho}_{i+1,i} \tag{6.26.a}$$

and

$$(\psi_i - \psi_{i-1})^2\left(\partial_1\theta_1(\rho_i, \rho_{i-1})\frac{\rho_{i+1}}{2} + \partial_2\theta_1(\rho_i, \rho_{i-1})\frac{\rho_i}{2}\right) \geq \frac{1}{2}(\psi_i - \psi_{i-1})^2\hat{\rho}_{i+1,i} \tag{6.26.b}$$

respectively. For the right-hand sides of (6.26) cancelling out with the remaining two terms on the right-hand side of (6.25), we conclude. ■

In Section 5.3 we already exploited Γ-convergence of the discrete Rényi entropy to some extend. More precisely, we showed that $\mathcal{F}_m^n$ satisfies the lim inf-bound (5.17) for a specific sequence of discrete measures $u^n$ given by the spatial finite-volume scheme. Indeed, it is not hard to establish full-fledged Γ-convergence of this functional as the following result shows.

**6.3.2 Proposition** (Sequential Γ-convergence of the functionals) *The discrete Shannon entropy $\mathcal{F}_1^n$ is sequentially Γ-convergent to the continuous counterpart $F_1$ as $(n \to \infty)$.*

Proof In Proposition 2.4.3 we we showed that the functional $F_1$ is lower semicontinuous with respect to the 2-Wasserstein distance; whereas we noticed in Fact 5.2.2.i that the interpolants are chosen in such a way that $\mathcal{F}_m^n(u^n) = F_1(U^n)$. Combining those two results immediately gives the lim inf-bound in Definition 4.1.1.

On the other hand, by definition of the the continuous functional $F_1$, it does not pose any restriction to assume that a given probability measure on $\Omega$ is absolutely continuous with density $\acute{U}$. Therefore, we can use the elementary estimates (5.27) and (5.28) to obtain the bound

$$\limsup_{n\to\infty} \mathcal{F}_1^n(\acute{u}^n) \leq F_m(\acute{U})$$

for discretisations $\acute{u}_i^n := \int_{\omega_i^n} \acute{U}(x)\,\mathrm{d}x$ which are weakly convergent to $\acute{U}$. ■

**6.3.3 Theorem** *Let $(u^n)_{n\in\mathbb{N}}$ be a sequence of spatial finite-volume discretisations $(u_i^n)_{1\leq i\leq n}$ as in (6.2) such that Assumptions 6.1.1 are satisfied. Let $(U^n)_{n\in\mathbb{N}}$ be the corresponding piecewise constant interpolants given by (6.4). Then there exists a subsequence $(U^{n_k})_{k\in\mathbb{N}}$ and a curve $U \in AC(\mathbb{R}_0^+, \mathcal{P}_2(\overline{\Omega}))$, independent of $T > 0$, with initial value $U(0) = U_0$, such that the following statements hold:*

(i) *For all times $t \in (0,T)$ the interpolant $U^{n_k}(t)$ converges weakly to $U(t)$ as $(k \to \infty)$;*
(ii) *for $\kappa = 0$, the limit curve $U$ is the unique gradient flow in the EVI sense with respect to the Shannon entropy functional $F_1$;*
(iii) *the limit curve $U$ is a solution in the distributional sense to the linear heat equation with non-flux Neumann boundary condition:*

$$\frac{\mathrm{d}}{\mathrm{d}t}U = \Delta U \qquad \text{in } \mathbb{R}^+ \times \Omega. \tag{6.27}$$

Proof In the first part of the proof we show that the density $\rho$ of each finite-volume discretisation $(u_i^n)_{1\le i\le n}$ forms an EVI gradient flow in the space $(\mathcal{P}^n, \mathcal{W})$: We already observed in Proposition 5.1.3 that the curve $\rho$ is a gradient flow in the Riemannian sense for the discrete functional $\mathcal{F}_q^n$. Now is enough to recall that we established geodesic convexity of $\mathcal{F}_q^n$ in Proposition 6.3.1. Therefore, Proposition 1.2.5 implies that $\rho$ is a gradient flow in the EVI sense as well.

The existence of the limit curve $U$ follows along the same lines of the previous chapter: We invoke Proposition 5.1.3 for increasing times $T > 0$ to extract a suitable subsequence $(U^{n_k})_{k\in\mathbb{N}}$ being pointwise weakly convergent to a continuous limit curve $U : \mathbb{R}_0^+ \to \mathcal{P}_2(\Omega)$. Note that pointwise convergence of $(U^{n_k})_{k\in\mathbb{N}}$ to $U$ can also be expressed in terms of the 2-Wasserstein distance since $\Omega$ is bounded.

To check that the limit curve $U$ is a gradient flow in the EVI sense for $\kappa = 0$, we apply Proposition 4.2.2 with the following prerequisites: Corollary 6.2.5 ensures that the interpolation mappings defined by (6.23) form $4\varepsilon$-isometries from $(\mathcal{P}^n, \mathcal{W})$ to $(\mathcal{P}_2(\overline{\Omega}), W_2)$. Moreover, Proposition 6.3.2 guarantees that the sequence $(\mathcal{F}_1^{n_k})_{k\in\mathbb{N}}$ is sequentially $\Gamma$-convergent to the continuous Shannon entropy $F_1$.

Finally, we already observed in Corollary 2.4.13 that the limit curve $U$ may be identified as distributional solution to the linear heat equation (6.27), in case $m = 1$. ∎

## 6.4 Bibliographical Notes

Section 6.2, discussing the Gromov-Hausdorff convergence of the discrete transportation metrics for Markov chains with nearest-neighbour transitions, is based on the work [35] of Gigli and Maas, where a periodic setting is considered. Note that in this article the discrete Laplacian is induced by a Markov kernel $K_n$ instead which is related to the infinitesimal generator in this thesis by the identity $\mathcal{Q} = 2n^2(K_n - \mathrm{Id})$. indexinequality+Jensen+for locally convex real TVS The variant of the Jensen inequality used in Fact 6.2.2.iii is due to Perlman [57] and holds in locally convex real topological vector spaces. The spectral decomposition of the generator used in Fact 6.2.3.i was obtained by Yueh [72].

Actually, the geodesic convexity of the discrete Shannon entropy $\mathcal{F}_m^n$ obtained Proposition 6.3.1 holds for more general tridiagonal generators than the Toeplitz-like structure in (6.3). See Mielke's article [50] for several results in this direction.

In order to extend the scope of this chapter a Fokker-Plank equation with drift, one may consider a relative entropy functional of the form (5.39) as done by Disser and Liero [24]. With this approach, the resulting infinitesimal generator is still of tridiagonal form but looses the structure of a Toeplitz matrix needed to make the argument in [35] go through.

Fairly recently, Al Reda and Maury proposed in [1] a generalisation of the entropic gradient flow structures induced by a finite-volume discretisation for Fokker-Plank equations to higher dimensions. Indeed, this is in line with the Gromov-Hausdorff convergence result in [35] which considered higher dimensions as well. However, a convergence result still needs to be obtained for this generalised setting.

# Appendix A

In this appendix we summarise some topics of probability theory which are needed in the main text. The first section is devoted to results about weak convergence of Borel measures on metric spaces. Essentially, we follow along the lines of section 5.1 of [4] by Ambrosio, Gigli, Savaré. A more detailed account on this topic may be found in Billingsley's classic text book [11]. For a modern approach of weak convergence of measures on general topological spaces see for instance chapter 8 in Bogachev's text book [12].

In the second section we take a brief look at the disintegration of probability measures, where we follow the brief overview given in section 5.3 of [4]. For the relation to conditional expectations and conditional measures see sections 10.6 and 10.10(ii) in [12].

## A.1 Weak Convergence of Borel Measures

A topological spaces carries a natural $\sigma$-algebra which is closely related to the topology of the space.

**A.1.1 Definition** The ***Borel $\sigma$-algebra*** $\mathcal{B}(X)$ of a Hausdorff space $(X, \mathcal{T})$ is the smallest $\sigma$-algebra which contains all open sets of $X$. The elements of $\mathcal{B}(X)$ are called the ***Borel sets*** in $X$.

A mapping $f : X \to Y$ between topological spaces $(X, \mathcal{T})$ and $(Y, \mathcal{O})$ is called ***Borel measurable*** or simply ***Borel*** if $f^{-1}(\mathcal{B}(Y)) \subseteq \mathcal{B}(X)$. For instance, every continuous function $f : X \to Y$ is Borel.

A countably additive signed measure on the Borel $\sigma$-algebra $\mathcal{B}(X)$ is called a ***Borel measure*** on $X$.

From now on we restrict ourselves to the case where the Borel $\sigma$-algebra $\mathcal{B}(X)$ is generated by a metric space $(X, d)$. Recall that the *weak-* topology* on the continuous dual space $C_b'(X)$ is the *weak topology* $\sigma(C_b'(X), C_b(X))$, i.e. the initial topology on $C_b'(X)$ with respect to $\iota(C_b(X))$, where $\iota : C_b(X) \to C_b^{**}(X)$ is the canonical embedding of $C_b(X)$ into the bidual space $C_b^{**}(X)$.

Denote by $\mathcal{M}_\sigma(X)$ the linear space of all Borel measures on $(X, \mathcal{T})$ with finite variation. Then we can identify $\mathcal{M}_\sigma(X)$ with a subspace of $C_b'(X)$ by means of functionals of the form

$$L(f) := \int_X f \, \mathrm{d}\mu \qquad \forall f \in C_b(X)$$

for all $\mu \in \mathcal{M}_\sigma(X)$. Hence, we may consider convergence of Borel measures on $X$ with respect to $\sigma(C_b'(X), C_b(X))$.

**A.1.2 Definition** Let $(X, d)$ be a metric space. A sequence $(\mu_n)_{n\in\mathbb{N}}$ of Borel measures in $\mathcal{M}_\sigma(X)$ is called ***weakly convergent*** to a measure $\mu \in \mathcal{M}_\sigma(X)$ if $(\mu_i)_{i\in I}$ converges to $\mu$ with respect to the weak-* topology $\sigma(C_b'(X), C_b(X))$. In this case we write $\mu_n \xrightarrow{w^*} \mu$.

**A.1.3 Facts** Let $(X, d)$ be a metric space and consider a sequence $(\mu_n)_{n\in\mathbb{N}}$ of Borel measures in $\mathcal{M}_\sigma(X)$.

(i) A more concise characterisation of weak convergence can be given as follows: For every $\mu \in \mathcal{M}_\sigma(X)$ we have

$$\mu_n \xrightarrow{w^*} \mu \qquad \text{iff} \qquad \lim_{n\to\infty} \int_X f \, \mathrm{d}\mu_n = \int_X f \, \mathrm{d}\mu \quad \forall f \in C_b(X)\,. \tag{A.1}$$

(ii) A stronger notion of convergence of measures is given by the total variation $|\mu|(X)$ of a measure $\mu$ on $X$. Indeed, the space of all signed measures on $X$ with finite variation forms a Banach space with respect to $|\cdot|(X)$.

If $(\mu_n)_{n\in\mathbb{N}}$ converges in the total variation to a measure $\mu \in \mathcal{M}_\sigma(X)$, then $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to $\mu$ as well.

The following elementary example shows that weak convergence is actually a weaker notion than convergence in total variation.

**A.1.4 Example** Let $p \in L^1(\mathbb{R},\mathbb{R})$ be a probability density and define probability measures $\nu_n$ with densities $p_n := np(nt)$, $n \in \mathbb{N}$. Then we infer that $(\nu_n)_\mathbb{N}$ is weakly convergent to the Dirac measure $\delta_0$ by means of *dominated convergence* applied to

$$\lim_{n\to\infty} \int_\mathbb{R} f(t)p_n(t)\,\mathrm{d}\mathcal{L}(t) = \lim_{n\to\infty} \int_\mathbb{R} f(s/t)p(s)\,\mathrm{d}\mathcal{L}(s) = f(0) = \int_\mathbb{R} f\,\mathrm{d}\delta_0 \qquad \forall f \in C_\mathrm{b}(\mathbb{R})\,.$$

On the other hand, we have $|\nu_n - \delta_0|(\mathbb{R}) = 2$ for all $n \in \mathbb{N}$. Hence, $(\nu_n)_{n\in\mathbb{N}}$ does not converge in total variation. ◀

In case, a given function on $X$ is only semicontinuous, (A.1) need not hold any more. However, we have the following result for weakly converging nets of probability measures.

**A.1.5 Proposition** *Let $(\mu_i)_{n\in\mathbb{N}}$ be a sequence of Borel probability measures on a metric space $(X,d)$, weakly converging to a Borel probability measure $\mu$. Then for every bounded function $f : X \to \mathbb{R}$ the following statements hold:*

(i) *If $f$ is upper semicontinuous, then*

$$\limsup_{n\to\infty} \int_X f\,\mathrm{d}\mu_n \le \int_X f\,\mathrm{d}\mu.$$

(ii) *If $f$ is lower semicontinuous, then*

$$\liminf_{n\to\infty} \int_X f\,\mathrm{d}\mu_n \ge \int_X f\,\mathrm{d}\mu.$$

Supposed that a net $(\mu_n)_{n\in\mathbb{N}}$ of Borel measures is weakly convergent to some limit measure $\mu$, can we find a suitable class of functions $f$ such that the net $(f \cdot \mu_n)_{n\in\mathbb{N}}$ is weakly convergent to $f \cdot \mu$? The answer is given by the notion of uniform integrability.

**A.1.6 Definition** Let let $\mathcal{N}$ be a family of Borel probability measures on a metric space $(X,d)$. Then we say that a Borel function $f : X \to \mathbb{R}_0^+ \cup \{= \infty\}$ is ***uniformly integrable*** with respect to $\mathcal{N}$ if for every $\varepsilon > 0$ there exists an integer $N \in \mathbb{N}$ such that

$$\int_{[f\ge N]} f\,\mathrm{d}\mu < \varepsilon \qquad \forall \mu \in \mathcal{N}.$$

In particular, $f$ is uniformly integrable with respect to $\mathcal{N}$ when there exists some $p > 1$ such that the norm $\|f\|_{L^p(\mu)}$ is uniformly bounded for all $\mu \in \mathcal{N}$. Indeed, this is immediately implied by integration of the elementary inequality $f(x)N^{p-1} \le f^p(x)$ for all $x \in [f \ge N]$, which results in

$$\int_{[f\ge N]} f\,\mathrm{d}\mu < N^{1-p} \int_X f^p\,\mathrm{d}\mu \qquad \forall \mu \in \mathcal{N}. \tag{A.2}$$

**A.1.7 Lemma** *For a sequence $(\mu_n)_{n\in\mathbb{N}}$ of Borel probability measures on a metric space $(X,d)$, converging weakly to a Borel probability measure $\mu$, the following statements hold:*

(i) *If $f : X \to \mathbb{R}$ is a continuous function such that $|f|$ is uniformly integrable with respect to $(\mu_n)_{n\in\mathbb{N}}$, then*

$$\lim_{n\to\infty} \int_X f\,\mathrm{d}\mu_n = \int_X f\,\mathrm{d}\mu.$$

(ii) *If $g : X \to \mathbb{R} \cup \{+\infty\}$ is a lower semicontinuous function such that $(g)^-$ is uniformly integrable with respect to $(\mu_n)_{n\in\mathbb{N}}$, then*

$$\liminf_{n\to\infty} \int_X g \, \mathrm{d}\mu_n \geq \int_X g \, \mathrm{d}\mu > -\infty.$$

Now we turn to one of the central results in the theory of weakly convergent measures, which relates weak convergence of measures to the concept of uniformly tightness of measures.

**A.1.8 Definition** A family $\mathcal{N}$ of finite Borel measures on a metric space $(X, \mathcal{T})$ is called ***uniformly tight*** if for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon$ such that

$$|\mu|(X \setminus K_\varepsilon) < \varepsilon \qquad \forall \mu \in \mathcal{N}.$$

In particular, every finite family $\mathcal{N}$ of Borel measures is uniformly tight since the finite union of compact sets is again compact in $X$.

**A.1.9 Facts** Let $\mathcal{N}$ be a family of finite Borel measures on a metric space $(X, d)$.

(i) The family $\mathcal{N}$ is uniformly tight, precisely, when there exists a nonnegative function $\varphi : X \to \mathbb{R}_0^+ \cup \{+\infty\}$ such that all sublevel sets $\{x \in X : \varphi(x) \leq c\}$ are compact in $X$ and

$$\sup_{\mu\in\mathcal{N}} \int_X \varphi \, \mathrm{d}\mu < +\infty. \tag{A.3}$$

Indeed, if $\mathcal{N}$ is uniformly tight, then there exists an exhaustion $(K_n)_{n\in\mathbb{N}}$ by compact sets $K_n$ such that $\mu(X \setminus K_n) \leq 2^{-n}$ for all $n \in \mathbb{N}$ and all $\mu \in \mathcal{N}$. This means that the function

$$\varphi(x) := \inf\{n \in \mathbb{N} : x \in K_n\} = \sum_{n=1}^{\infty} \mathbb{1}_{X\setminus K_n}(x)$$

has compact sublevel sets and satisfies (A.3) by monotone convergence since we have the estimate

$$\int_X \varphi \, \mathrm{d}\mu = \sum_{n=1}^{\infty} \mu(X \setminus K_n) \leq 2.$$

Conversely, if there exists a nonnegative function $\varphi$ such that (A.3) holds, then Chebyshev's inequality implies

$$\mu[\varphi > c] \leq \frac{1}{c} \int_X \varphi \, \mathrm{d}\mu \qquad \forall \mu \in \mathcal{N}.$$

Thus, the family $\mathcal{N}$ is uniformly tight, provided that all sublevel sets of $\varphi$ are compact.

(ii) Assume that $X$ is a product space of the form $X = X_1 \times X_2$ where $X_1$ and $X_2$ are metric spaces. Then for all compact sets $K_1 \subseteq X_1$ and$K_2 \subseteq X_2$ we have the estimate

$$\mu\big((X_1 \times X_2) \setminus (K_1 \times K_2)\big) \leq \pi^1_\#\mu(X_1 \setminus K_1) + \pi^2_\#\mu(X_2 \setminus K_2) \qquad \forall \mu \in \mathcal{N}.$$

This means if both the marginal families $(\pi^1_\#\mu)_{\mu\in\mathcal{N}}$ and $(\pi^2_\#\mu)_{\mu\in\mathcal{N}}$ are uniformly tight, then $\mathcal{N}$ is uniformly tight as well.

Now we are ready to formulate the aforementioned result which relates weak convergence to uniform tightness.

**A.1.10 Theorem** (Prokhorov) *Let let $\mathcal{N}$ be a family of Borel measures on a Polish metric space $(X, d)$. Then the following conditions are equivalent:*

(i) *Every sequence $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{N}$ contains a weakly convergent subsequence.*

(ii) *The family $\mathcal{N}$ is uniformly tight and uniformly bounded in total variation norm $|\cdot|(X)$.*

## A.2 Disintegration of Probability Measures

Let $Y$ be a Polish space and let $(\mu_y)_{y\in Y}$ be a family of Borel probability measures on another Polish space $X$. Let us assume that the mapping $y \mapsto \mu_y(A)$ is Borel measurable for each Borel set $A \subseteq X$. Then every Borel probability measure $\eta$ on $Y$ induces a Borel probability measure $\mu$ on $X$ by the formula

$$\mu(A) = \int_Y \mu_y(A)\,\mathrm{d}\eta(y) \qquad \forall A \in \mathcal{B}(X). \tag{A.4}$$

The following theorem shows that for every Borel probability measure $\mu$ on $X$, there exists a family $(\mu_y)_{y\in Y}$ such that $\mu$ is represented by (A.4), as long as $\eta$ is a pushforward measure of $\mu$.

**A.2.1 Theorem** (Disintegration) *Let $\pi : X \to Y$ be a Borel mapping between Polish spaces $X$ and $Y$. For every Borel probability measure $\mu$ on $X$, there exists a $\pi_\#\mu$-a.e. uniquely determined family $(\mu_y)_{y\in Y}$ of Borel probability measures $\mu_y$ on $X$ such that the following statements hold:*

(i) *The function $y \mapsto \mu_y(A)$ is Borel measurable for each Borel set $A \subseteq X$.*

(ii) *$\pi_\#\mu$-almost surely the measure $\mu_y$ is concentrated on the level set $\pi^{-1}(y)$, i.e.*

$$\mu_y\big(X \setminus \pi^{-1}(y)\big) = 0 \qquad \pi_\#\mu\text{-a.e. } y \in Y.$$

(iii) *For every nonnegative Borel function $f : Y \to \mathbb{R}_0^+ \cup \{+\infty\}$ the following identity holds:*

$$\int_X f\,\mathrm{d}\mu = \int_X \Big(\int_{[\pi=y]} f(x)\,\mathrm{d}\mu_y(x)\Big)\,\mathrm{d}\pi_\#\mu(y).$$

*In particular, the measure $\mu$ is given by*

$$\mu(A) = \int_Y \mu_y(A)\,\mathrm{d}\pi_\#\mu(y) \qquad \forall A \in \mathcal{B}(X). \tag{A.5}$$

The measures $\mu_y$ are called ***disintegration measures*** of $\mu$ with respect to $\pi_\#\mu$.

Typically, above theorem is applied to the case where $X$ is a product space of the form $X = X_1 \times X_2$ and the the Borel mapping is given by the projection $\pi^1 : X_1 \times X_2 \to X_1$. Then on the generator $\{A \times B : A \in \mathcal{B}(X),\, B \in \mathcal{B}(Y)\}$ of the product $\sigma$-algebra $\mathcal{B}(X) \otimes \mathcal{B}(Y)$, (A.5) takes the form

$$\mu(A \times B) = \int_B \mu(A|\pi^1 = y)\,\mathrm{d}\pi^1_\#\mu(y) \qquad \forall A \in \mathcal{B}(X),\, B \in \mathcal{B}(Y),$$

where $\mu(\,\cdot\,|\pi^1 = y)$ is the conditional measure of $\mu$ under $\pi^1 = y$.

# Appendix B

Here we provide some background material on smooth and Riemannian manifolds. In the first section we recall the basic framework used in the main text. Amongst the abundance of literature on manifold theory we just mention the classic text books [42] and [58] by J.M. Lee and Petersen, respectively. More on the notion of *smooth manifolds with corners* may be found for instance in J.M. Lee [43].

In the second section we summarise some existence results regarding flows on smooth manifolds. We refer to Chapter 9 of [43] for a more detailed treatment of this topic.

The final section of this appendix is devoted to geodesics Riemannian manifolds. The results stated there may be found in any standard reference on Riemannian geometry, a.e. Jost [37] or the already mentioned reference [58].

## B.1 Topological, Smooth, and Riemannian Manifolds

We start with definitions for various types of manifolds.

**B.1.1 Definition** An $n$-dimensional ***topological manifold*** is a second-countable Hausdorff space $M$ together with homeomorphisms $\varphi_j : U_j \to V_j$ for open sets $V_j \subseteq \mathbb{R}^n$ such that all $U_j$ provide an open cover of $M$. The pairs $(U_j, \varphi_j)$ are called ***local coordinate charts*** for $M$.

The definition of an $n$-dimensional ***topological manifold with boundary*** follows along the same lines with the charts $\varphi$ taking values in the *$n$-dimensional closed upper half-space*

$$\mathbb{H}^n := \{(x_1, \dots x_n) \in \mathbb{R}^n : x_n \geq 0\}$$

instead. In this context, we call $(U_j, \varphi_j)$ an ***interior chart*** if $\varphi_j(U_j)$ is an open subset of $\mathbb{R}^n$ or a ***boundary chart*** if $\varphi_j(U_j)$ is an open subset of $\mathbb{H}^n$ such that $\varphi_j(U_j) \cap \partial\mathbb{H}^n \neq \emptyset$. We say that a point $x \in M$ belongs to the *boundary* $\partial M$ if there exists a chart $(U_j, \varphi_j)$ such that $x \in U_j$ and $\varphi_j$ sends $x$ to $\partial\mathbb{H}^n$. Otherwise, we say that the point $x$ belong to the ***interior*** $\operatorname{int} M$. One can show that $\partial M$ and $\operatorname{int} M$ are disjoint sets whose union is $M$.

We say that a topological manifold (with or without boundary) is ***smooth***, provided that all the mappings of the form $\varphi_j \circ \varphi_i^{-1}$ are $C^\infty$-smooth.

Geometric objects like simplices are topological manifolds with boundary which do not admit a smooth structure, due to 'having corners'. A remedy is provided by restricting the codomain of the boundary charts to the subset

$$\overline{\mathbb{R}}^n_+ := \{(x_1, \dots x_n) \in \mathbb{R}^n : x_1 \geq 0, \dots x_n \geq 0\}.$$

Note that $\overline{\mathbb{R}}^n_+$ is homeomorphic but not diffeomorphic to $\mathbb{H}^n$. This leads to the following more general notion: A ***chart with corners*** for a topological manifold with boundary $M$ is a pair $(U_j, \varphi_j)$ such that $\varphi_j : U_j \to V_j$ is an homeomorphism for some relative open set $V_j \subseteq \overline{\mathbb{R}}^n_+$. Now a ***smooth manifolds with corners*** is a topological manifold with boundary $M$ together with a collection of interior and boundary charts $(U_j, \varphi_j)$ such that all composite mappings $\varphi_j \circ \varphi_i^{-1}$ are $C^\infty$-smooth.

Finally, we call a smooth symmetric covariant 2-tensor field $g : T_pM \times T_pM \to \mathbb{R}$ ***Riemannian metric*** for a smooth manifold $M$, provided that $g$ is positive definite at each point. Then the pair $(M, g)$ is called ***Riemannian manifold***.

## B.2 Flows on Smooth Manifolds

In this section we give a brief review about flows on smooth manifolds. At the beginning we consider only smooth manifolds without boundary.

**B.2.1 Definition** Denote by $I$ an interval in $\mathbb{R}$. Given a vector field $V$ on $M$, we call a differentiable curve $\gamma : I \to M$ an ***integral curve*** of $V$ if the velocity field of $\gamma$ satisfies $\dot{\gamma}(t) = V_{\gamma(t)}$ for all $t \in I$.

The nice thing about integrals is that they always exist at least locally.

**B.2.2 Proposition** *Let $V$ be a smooth vector field on a smooth manifold $M$. Then for each point $p \in M$, there exists a smooth curve $\gamma : (-\varepsilon, \varepsilon) \to M$ that is an integral curve of $V$ starting at $\gamma(0) = p$.*

Let us turn to flows on manifolds which is a concept closely related to integral curves.

**B.2.3 Definition** Let $M$ be a smooth manifold. A ***flow domain*** for $M$ is an open subset $\mathcal{D} \subseteq \mathbb{R} \times M$ such that for each point $p \in M$, the set $\mathcal{D}^{(p)} := \{t \in \mathbb{R} : (t, p) \in \mathcal{D}\}$ is an interval containing 0. A ***(local) flow*** on $M$ is a continuous mapping $\varkappa : \mathcal{D} \to M$ such that $\mathcal{D}$ is a flow domain and $\varkappa_t(\kappa) := \varkappa(t, \kappa)$ satisfies the following group laws for all points $p \in M$:

$$\varkappa_0(p) = p \qquad \text{and} \qquad \varkappa_t(\varkappa_s(p)) = \varkappa_{s+t}(p) \quad \forall s \in \mathcal{D}^{(p)},\, t \in \mathcal{D}^{(\varkappa_s(p))} : s + t \in \mathcal{D}^{p}. \tag{B.1}$$

Provided that $\mathcal{D}^{(p)} = \mathbb{R}$ for all $p \in M$, $\kappa$ is called ***global flow*** on $M$.

**B.2.4 Proposition** *If the flow $\varkappa : \mathcal{D} \to M$ is smooth, then each curve $\kappa(\cdot, p)$ is an integral curve of the smooth vector field*

$$\dot{\kappa}_0(p) := \frac{\mathrm{d}}{\mathrm{d}t}\kappa_t(p)\Big|_{t=0}.$$

The vector field $\dot{\varkappa}_0$ is called the ***infinitesimal generator*** of $\varkappa$.

The next theorem represents the central result of this section. To formulate the statement, we introduce the following notion: An integral curve is called ***maximal*** if it does not admit any extension to an integral curve on a larger open interval. Likewise, a flow is ***maximal*** if there does not exist any extension to a flow on a larger flow domain.

**B.2.5 Theorem** (Fundamental theorem on flows) *Let $V$ be a smooth vector field on a smooth manifold $M$. There exists a unique smooth maximal flow $\kappa : \mathcal{D} \to M$ whose infinitesimal generator is given by $V$. This flow satisfies the following properties:*

(i) *For each point $p \in M$ the curve $\kappa(\cdot, p) : \mathcal{D}^{(p)} \to M$ is the unique maximal integral curve of $V$ starting at $\kappa_0(p) = p$;*

(ii) *if $s \in \mathcal{D}^{(p)}$, then $\mathcal{D}^{(\kappa_s(p))}$ is the shifted interval $\mathcal{D}^{(p)} - s$;*

(iii) *for each $t \in \mathbb{R}$, the set*

$$M_t := \{p \in M : (t, p) \in \mathcal{D}\}$$

*is open in $M$ and the mapping $\varkappa_t : M_t \to M_{-t}$ is a diffeomorphism with inverse given by $\varkappa_{-t}$.*

Above theorem does not give answer to the question whether a vector field gives rise to a global flow; such vector fields are called ***complete***. In case the smooth manifold (without boundary) is compact, the answer is positive.

**B.2.6 Proposition** *On a compact smooth manifold every smooth vector field is complete. In particular, each of its maximal integral curves is defined for all times $t \in \mathbb{R}$.*

For smooth manifolds with corners the existence of global flows is a more delicate issue. Indeed, if a smooth vector field is not tangent to the boundary, then some integral curves starting at boundary points may not exist at all.

However, it is still possible to state some existence results for the weaker notion of semiflows.

**B.2.7** **Definition** Let $M$ be a smooth manifold with corners. A ***global semiflow*** on $M$ is a continuous mapping $\varkappa : \mathbb{R}_0^+ \times M \to M$ with the properties that the group laws in (B.1) are satisfied for all $p \in M$ and $\mathscr{D}^{(p)} = \mathbb{R}_0^+$.

Contrary to flows on manifolds without boundary, it is not enough to assume that a smooth manifold with corners is compact, in order to ensure existence of a global semiflow. In addition, we need to make sure that the vector field is nowhere outward pointing on the boundary. Recall that for a smooth manifold with corners $M$ and a point $p \in \partial M$, a vector $v \in T_pM \setminus T_p\partial M$ is said to be ***outward pointing*** if there exists a smooth curve $\gamma : (-\varepsilon, 0] \to M$ such that $\gamma(0) = p$ and $\dot\gamma(0) = v$.

**B.2.8** **Proposition** *Let $M$ be a compact smooth manifold with corners and let $V$ be a smooth vector field on $M$ that is nowhere outward pointing on the boundary $\partial M$. Then there exists a unique smooth global semiflow on $M$, whose infinitesimal generator is given by $V$. In particular, each of its integral curves is defined for all times $t \in \mathbb{R}_0^+$.*

## B.3 Geodesics on Riemannian manifolds

In this section we will only consider a Riemannian manifold $(M, g)$ without boundary. With the *Levi-Civita connection* $\nabla$ at hand, we may introduce geodesics in a notion slightly different from the the definition of metric geometry given in Section 1.2.

**B.3.1** **Definition** A smooth curve $\rho$ on a Riemannian manifold $(M, g)$ is called ***geodesic*** if $\nabla_{\dot\rho}\dot\rho = 0$ at each point along the curve $\rho$. In local coordinates the condition is equivalent to the system of ***geodesic equations*** (cf. a.e. Theorem 2.2.3 in [37])

$$\dot\rho^i - \sum_j g^{ij} p_j = 0,$$

$$\dot p_i + \frac{1}{2}\sum_{j,k} \frac{\partial}{\partial x^i} g^{jk} p_j p_k = 0,$$

where $g^{ij}$ denote the local coordinates of the inverse metric of $g$.

We remark that a geodesic in this sense need not be be a shortest path; i.e. a geodesic joining points $p, q \in M$ may have length more than $d(p,q)$. Here the metric $d_g(p,q)$, called ***Riemannian distance function*** between points $p, q \in M$, is given by the infimum of the length functional

$$L(\gamma) := \int_0^1 |\dot\gamma(t)|_g \,\mathrm{d}t$$

over all piecewise smooth curves $\gamma : [0,1] \to M$ connecting $p$ to $q$. Nevertheless, sufficiently short curve segments of geodesics are minimisers for the length functional $L$. Equivalently, one may consider the following energy functional instead of $L$, in order to overcome smoothness issues:

$$E(\gamma) := \frac{1}{2}\int_0^1 |\dot\gamma(t)|_g^2 \,\mathrm{d}t.$$

Let $\gamma : (-\varepsilon, \varepsilon) \times [0,1] \to M$ be a smooth mapping and introduce for $\gamma_t(s) := \gamma(t,s)$ the notation

$$\dot\gamma_t(s) := \frac{\partial}{\partial s}\gamma_t(s) \qquad \text{and} \qquad \gamma_t'(s) := \frac{\partial}{\partial t}\gamma_t(s).$$

Then the energy functional $E$ satisfies the ***first variation formula***

$$\frac{\mathrm{d}}{\mathrm{d}t}E(\gamma_t) = -\int_0^1 \left\langle \nabla_{\dot{\gamma}_t(s)}\dot{\gamma}_t(s), \gamma_t'(s)\right\rangle_g \mathrm{d}s + \left\langle \dot{\gamma}_t(1), \gamma_t'(1)\right\rangle_g - \left\langle \dot{\gamma}_t(0), \gamma_t'(0)\right\rangle_g. \tag{B.2}$$

One can show that for any two points $p, q \in M$, every local minimum of the energy functional $E$ (or equivalently of the length functional $L$) over all smooth curves $\gamma : [0,1] \to M$ between $p$ and $q$ is a geodesic in $M$. Indeed, this follows from the fact that any stationary point $\gamma_0$ of $E$, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}E(\gamma_t)\Big|_{t=0} = 0$$

for some smooth mapping $\gamma : (-\varepsilon, \varepsilon) \times [0,1] \to M$ with $\gamma_t(0) = p$ and $\gamma_t(1) = q$, turns out to be a geodesic connecting $p$ to $q$.

The following result shows that a geodesic are already uniquely determined by an initial point together with an initial tangent vector.

**B.3.2 Proposition** *Let $(M, g)$ be a Riemannian manifold. Then for every choice of $p \in M$ and $v \in T_pM$, there exists a unique geodesic $\rho : [0, \varepsilon] \to M$ with $\rho(0) = p$ and $\dot{\rho}(0) = v$. This curve $\rho$ depends smoothly on $p$ and $v$.*

This proposition justifies the following definition.

**B.3.3 Definition** For every tangent vector $v \in T_pM$, let $\rho_v$ the unique geodesic $\rho_v(0) = p$ and $\dot{\rho}_v(0) = v$. Denote by $O_p \subseteq T_pM$ be the set such that $\rho_v$ is defined on at least the interval $[0,1]$. Now the ***exponential map*** at $p$ is a function $\exp_p : O_p \to M$ defined by $\exp_p(v) := \rho_v(1)$.

Note that the uniqueness of $\rho_v$ implies the following homogeneity property for the exponential map:

$$\exp_p(tv) = c_v(t) \qquad \forall t > 0 : tv \in O_p.$$

Moreover, it can be shown that the exponential map is a local diffeomorphism around the origin, i.e. $\exp_p$ maps a neighbourhood of $0 \in T_pM$ diffeomorphically onto a neighbourhood of $p \in M$.

At the end of this brief section, we state one of the foundational centrepieces of Riemannian geometry.

**B.3.4 Theorem** (Hopf-Rinow) *For a connected Riemannian manifold $(M, g)$ the following statements a equivalent:*

(i) *$M$ is a complete metric space;*

(ii) *$M$ satisfies the Heine-Borel property, i.e. every closed bounded set in $M$ is compact;*

(iii) *$M$ is geodesically compete, i.e. for every point $p \in M$, $\exp_p$ is defined on the entire tangent space $T_pM$;*

(iv) *there exists a point $p \in M$ where $\exp_p$ is defined on the entire tangent space $T_pM$.*

*In particular, any of the above statements implies that $(M, d_g)$ is a geodesic space, i.e. any two points $p, q \in M$ can be joined by a geodesic of length $d_g(p, q)$.*

# References


[1] F. Al Reda and B. Maury, *Interpretation of finite volume discretization schemes for the fokker planck equation as gradient flows for the discrete wasserstein distance* (2016)

[2] A. Alexandrov, *Über eine verallgemeinerung der riemannschen geometrie*, Schr. Forschungsinst. Math. Berlin **1** (1957), 33–84

[3] L. Ambrosio and N. Gigli, A user's guide to optimal transport, in *Modelling and Optimisation of Flows on Networks* (Springer), p. 1–155.

[4] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures.* (Birkhäuser Verlag, Basel, 2008), Second ed.

[5] L. Ambrosio, N. Gigli, G. Savaré, and others, *Metric measure spaces with riemannian Ricci curvature bounded from below*, Duke Mathematical Journal **163** (2014), no. 7, 1405–1490

[6] L. Ambrosio and G. Savaré, *Gradient flows of probability measures*, Handbook of differential equations: evolutionary equations **3** (2007), 1–136

[7] T. Barth and M. Ohlberger, *Finite volume methods: foundation and analysis*, Encyclopedia of computational mechanics

[8] J.-D. Benamou and Y. Brenier, A numerical method for the optimal time-continuous mass transport problem and related problems, in *Monge Ampère equation: applications to geometry and optimization (Deerfield Beach, FL, 1997)* , number 226 in Contemp. Math. (Amer. Math. Soc., Providence, RI), p. 1–11.

[9] P. Bénilan, *Solutions intégrales d'équations d'évolution dans un espace de Banach*, C. R. Acad. Sci. Paris Sér. A-B **274** (1972), A47–A50

[10] M. Bessemoulin-Chatard and F. Filbet, *A finite volume scheme for nonlinear degenerate parabolic equations*, SIAM J. Sci. Comput. **34** (2012), no. 5, B559–B583

[11] P. Billingsley, *Convergence of probability measures.* (John Wiley & Sons, 1999).

[12] V. Bogachev, *Measure Theory*, volume Vol. 1–2 of *Measure Theory.* (Springer, 2006).

[13] V. I. Bogachev and A. V. Kolesnikov, *The Monge-Kantorovich problem: achievements, connections, and perspectives*, Russian Mathematical Surveys **67** (2012), no. 5, 785

[14] A. Braides, *A handbook of $\gamma$-convergence*, Handbook of Differential Equations: stationary partial differential equations **3** (2006), 101–213

[15] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics.* (American Mathematical Society, Providence, RI, 2001).

[16] J. A. Carrillo, A. Chertock, and Y. Huang, *A finite-volume method for nonlinear nonlocal equations with a gradient flow structure*, Commun. Comput. Phys. **17** (2015), no. 1, 233–258

[17] S.-N. Chow, W. Huang, Y. Li, and H. Zhou, *Fokker-Planck equations for a free energy functional or Markov process on a graph*, Arch. Ration. Mech. Anal. **203** (2012), no. 3, 969–1008

[18] P. Clément and W. Desch, *Some remarks on the equivalence between metric formulations of gradient flows*, Boll. Unione Mat. Ital. (9) **3** (2010), no. 3, 583–588

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

[20] S. Daneri and G. Savaré, Lecture notes on gradient flows and optimal transport, in *Optimal transportation* , number 413 in London Math. Soc. Lecture Note Ser. (Cambridge Univ. Press, Cambridge), p. 100–144.

[21] E. De Giorgi, New problems on minimizing movements, in *Boundary value problems for partial differential equations and applications* , number 29 in RMA Res. Notes Appl. Math. (Masson, Paris), p. 81–98.

[22] E. De Giorgi, M. Degiovanni, A. Marino, and M. Tosques, *Evolution equations for a class of nonlinear operators*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **75** (1983), no. 1-2, 1–8 (1984)

[23] E. De Giorgi, A. Marino, and M. Tosques, *Problems of evolution in metric spaces and maximal decreasing curve*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **68** (1980), no. 3, 180–187

[24] K. Disser and M. Liero, *On gradient structures for Markov chains and the passage to Wasserstein gradient flows*, Netw. Heterog. Media **10** (2015), no. 2, 233–253

[25] R. L. Dobrushin, *Prescribing a system of random variables by conditional distributions*, Theory of Probability & Its Applications **15** (1970), no. 3, 458–486

[26] D. A. Edwards, *A simple proof in Monge-Kantorovich duality theory*, Studia Math. **200** (2010), no. 1, 67–77

[27] D. A. Edwards, *On the Kantorovich-Rubinstein theorem*, Expo. Math. **29** (2011), no. 4, 387–398

[28] M. Erbar, M. Fathi, V. Laschos, and A. Schlichting, *Gradient flow structure for McKean-Vlasov equations on discrete spaces*, arXiv preprint arXiv:1601.08098 (2016)

[29] M. Erbar and J. Maas, *Ricci curvature of finite Markov chains via convexity of the entropy*, Arch. Ration. Mech. Anal. **206** (2012), no. 3, 997–1038

[30] M. Erbar and J. Maas, *Gradient flow structures for discrete porous medium equations*, Discrete Contin. Dyn. Syst. **34** (2014), no. 4, 1355–1374

[31] M. Fathi and M. Simon, *The gradient flow approach to hydrodynamic limits for the simple exclusion process*, arXiv preprint arXiv:1507.06489 (2015)

[32] L. C. Ferreira, M. C. Santos, and J. C. Valencia-Guevara, *Minimizing movement for a fractional porous medium equation in a periodic setting*, arXiv preprint arXiv:1603.03393 (2016)

[33] D. L. Forkert, *Gradient flows in general metric and Wasserstein spaces* (2014). Diploma thesis, Vienna University of Technology.

[34] N. Gigli, *On the heat flow on metric measure spaces: existence, uniqueness and stability*, Calc. Var. Partial Differential Equations **39** (2010), no. 1-2, 101–120

[35] N. Gigli and J. Maas, *Gromov-Hausdorff convergence of discrete transportation metrics*, SIAM J. Math. Anal. **45** (2013), no. 2, 879–899

[36] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal. **29** (1998), no. 1, 1–17

[37] J. Jost, *Riemannian geometry and geometric analysis*. (Springer, Heidelberg, 2011), Sixth ed.

[38] L. V. Kantorovich *On the translocation of masses*, in *Dokl. Akad. Nauk SSSR* (1942), pp. 199–201.

[39] L. V. Kantorovich, *On a problem of Monge*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **312** (2004b), no. Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 11, 15–16

[40] L. V. Kantorovich, *On mass transportation*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **312** (2004a), no. Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 11, 11–14

[41] N. Q. Le, *A gamma-convergence approach to the Cahn-Hilliard equation*, Calc. Var. Partial Differential Equations **32** (2008), no. 4, 499–522

[42] J. M. Lee, *Manifolds and differential geometry*, volume 107 of *Graduate Studies in Mathematics*. (American Mathematical Society, Providence, RI, 2009).

[43] J. M. Lee, *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics*. (Springer, New York, 2013), Second ed.

[44] M. Liero and A. Mielke, *Gradient structures and geodesic convexity for reaction-diffusion systems*, Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **371** (2013), no. 2005, 20120346, 28

[45] J. Lott and C. Villani, *Weak curvature conditions and functional inequalities*, J. Funct. Anal. **245** (2007), no. 1, 311–333

[46] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2) **169** (2009), no. 3, 903–991

[47] J. Maas, *Gradient flows of the entropy for finite Markov chains*, J. Funct. Anal. **261** (2011), no. 8, 2250–2292

[48] U. F. Mayer, *Gradient flows on nonpositively curved metric spaces and harmonic maps*, Comm. Anal. Geom. **6** (1998), no. 2, 199–253

[49] R. J. McCann, *A convexity principle for interacting gases*, Adv. Math. **128** (1997), no. 1, 153–179

[50] A. Mielke, *A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems*, Nonlinearity **24** (2011), no. 4, 1329–1346

[51] A. Mielke, *Geodesic convexity of the relative entropy in reversible Markov chains*, Calc. Var. Partial Differential Equations **48** (2013), no. 1-2, 1–31

[52] G. Monge, *Mémoire sur la théorie des déblais et des remblais*. (De l'Imprimerie Royale, 1781).

[53] C. Ortner, *Two variational techniques for the approximation of curves of maximal slope* (2005)

[54] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174

[55] H. Pajot, Y. Ollivier, and C. Villani, *Optimal Transportation: Theory and Applications*, volume 413. (Cambridge University Press, 2014).

[56] A. Papadopoulos, *Metric spaces, convexity and non-positive curvature*, volume 6 of *IRMA Lectures in Mathematics and Theoretical Physics*. (European Mathematical Society (EMS), Zürich, 2014), Second ed.

[57] M. D. Perlman, *Jensen's inequality for a convex vector-valued function on an infinite-dimensional space*, J. Multivariate Anal. **4** (1974), 52–65

[58] P. Petersen, *Riemannian geometry*, volume 171 of *Graduate Texts in Mathematics*. (Springer, New York, 2016), Third ed.

[59] S. T. Rachev and L. Rüschendorf, *Mass transportation problems. Vol. I and II*. (Springer-Verlag, New York, 1998). Theory.

[60] E. Sandier and S. Serfaty, *Gamma-convergence of gradient flows with applications to Ginzburg-Landau*, Comm. Pure Appl. Math. **57** (2004), no. 12, 1627–1672

[61] F. Santambrogio, *Optimal transport for applied mathematicians*. (Birkhäuser/Springer, Cham, 2015). Calculus of variations, PDEs, and modeling.

[62] G. Savaré, *Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds*, C. R. Math. Acad. Sci. Paris **345** (2007), no. 3, 151–154

[63] S. Serfaty, *Gamma-convergence of gradient flows on Hilbert and metric spaces and applications*, Discrete Contin. Dyn. Syst. **31** (2011), no. 4, 1427–1451

[64] J. Solomon, R. Rustamov, L. Guibas, and A. Butscher, *Continuous-Flow Graph Transportation Distances*, ArXiv e-prints (2016)

[65] K.-T. Sturm, *On the geometry of metric measure spaces. I*, Acta Math. **196** (2006a), no. 1, 65–131

[66] K.-T. Sturm, *On the geometry of metric measure spaces. II*, Acta Math. **196** (2006b), no. 1, 133–177

[67] K.-T. Sturm, *Gradient flows for semiconvex functions on metric measure spaces-existence, uniqueness and Lipschitz continuity*, arXiv preprint arXiv:1410.3966 (2014)

[68] L. N. Vasershtein, *Markov processes over denumerable products of spaces describing large system of automata*, Problemy Peredai Informacii **5** (1969), no. 3, 64–72

[69] L. N. Vaserstein, *Markov processes over denumerable products of spaces, describing large systems of automata*, Problemy Peredachi Informatsii **5** (1969), no. 3, 64–72

[70] C. Villani, *Topics in optimal transportation*, volume 58. (American Mathematical Soc., 2003).

[71] C. Villani, *Optimal transport: old and new*, volume 338. (Springer, 2008).

[72] W.-C. Yueh, *Eigenvalues of several tridiagonal matrices*, Appl. Math. E-Notes **5** (2005), 66–74

# Symbol Glossary

The page numbers on the right indicate the first time the symbol is defined or used.

| | |
|---|---|
| $\mathbb{N}$ | natural numbers $1, 2, 3, 4, \ldots$ |
| $\mathbb{Z}$ | integers |
| $\mathbb{Q}$ | rational numbers |
| $\mathbb{R}$ | real numbers |
| $\mathbb{C}$ | complex numbers |
| | |
| $s \wedge t := \inf\{s, t\}$ | |
| $s \vee t := \sup\{s, t\}$ | |
| $\mu_n \xrightarrow{w^*} \mu$ | weak convergence of $(\mu_n)_{n\in\mathbb{N}}$ to $\mu$ **83** |
| $f_\# \mu := \mu \circ f^{-1}$ | pushforward of a measure $\mu$ with respect to $f$ **23** |
| $AC((a, b), X)$ | space of absolutely continuous curves **11** |
| $AC^p((a, b), X)$ | **11** |
| $AC_{\text{loc}}((a, b), X)$ | space of locally absolutely continuous functions **11** |
| $AC^p_{\text{loc}}((a, b), X)$ | **11** |
| $\mathcal{A}_1 \otimes \mathcal{A}_2$, $\bigotimes_{i\in I} \mathcal{A}_i$ | tensor-product $\sigma$-algebra **23** |
| $\mathcal{A}^n$ | action functional **77** |
| $\alpha_i$, $\beta_i$ | rate coefficients **59** |
| $\mathcal{B}_m(\rho, \cdot)$ | quadratic form of the Hessian $\operatorname{Hess} \mathcal{F}^n_m(\rho)$ **44** |
| $\mathscr{B}_{\text{b}}(X)$ | space of bounded Borel measurable functions on $X$ **24** |
| $\mathscr{B}(X)$ | Borel $\sigma$-algebra on $(X, \mathcal{T})$ **83** |
| $\text{C}_0(X)$ | space of continuous functions on $(X, \mathcal{T})$ which vanish at infinity |
| $\text{C}_{\text{b}}(X)$ | space of bounded continuous functions on $(X, \mathcal{T})$ |
| $\varphi^\star(x) := \sup_{y\in X}\{\langle x, y\rangle - \varphi(y)\}$ | convex conjugate of a functional $\varphi$ on $X$ **24** |
| $\chi_A$ | characteristic function of a set $A$ in the sense of convex analysis **33** |
| $¢_{\mu,t}$ | rescaling factor for the heat kernel **74** |
| $¢_t$ | upper bound for the rescaling factor $¢_{\mu,t}$ **75** |
| $\frac{\mathrm{d}^+}{\mathrm{d}t} f(t) := \limsup\limits_{h\searrow 0} \frac{f(t+h)-f(t)}{h}$ | upper right-hand Dini derivative of $f$ at $t$ **14** |
| $\frac{\mathrm{d}^-}{\mathrm{d}t} f(t) := \limsup\limits_{h\nearrow 0} \frac{f(t+h)-f(t)}{h}$ | upper left-hand Dini derivative of $f$ at $t$ **14** |
| $\mathcal{D}$ | flow domain **88** |
| $\mathcal{D}^{(p)} := \{t \in \mathbb{R} : (t, p) \in \mathcal{D}\}$ | **88** |
| $\Delta\psi_i := \sum_{j=1}^n \mathcal{Q}_{ij}(\psi_j - \psi_i)$ | discrete Laplacian for a function $\psi : \mathcal{X}^n \to \mathbb{R}$ **42** |
| $\delta_x$ | Dirac measure with mass at the point $x$ |
| $d_g$ | Riemannian distance function **89** |
| $\frac{\mathrm{d}}{\mathrm{d}t} f(t+) := \lim\limits_{h\searrow t0} \frac{f(t+h)-f(t)}{h}$ | right-hand derivative of $f$ at $t$ **17** |
| $E^{**}$ | lower semicontinuous envelope of $E^0$ **52** |
| $E^\varepsilon$ | Allen-Cahn energy functional **52** |
| $\operatorname{Exp}(t)(x, v) := \exp_x(tv)$ | **27** |
| $L(\rho) := \int_0^1 g\big(\dot\rho(t), \dot\rho(t)\big)\,\mathrm{d}t$ | energy of a smooth curve $\rho : [0, 1] \to M$ **39** |
| $F$ | internal energy functional **29** |
| $F_m$ | continuous Rényi entropy functional **32** |
| $\mathcal{F}^n_1(\rho) := \sum_{i=1}^n f_m(\rho_i)\pi_i$ | discrete Rényi entropy functional **42** |
| $F_1$ | continuous Shannon entropy functional **32** |
| $\mathcal{F}^n_1(\rho) := \sum_{i=1}^n \rho_i \pi_i \log \rho_i$ | discrete Shannon entropy functional **42** |
| $f_m(x) := \frac{1}{m-1} x^m$ | integrand associated to the Rényi entropy functional **32** |
| $\nabla\psi_{ij} := \psi_i - \psi_j$ | discrete gradient of a function $\psi : \mathcal{X}^n \to \mathbb{R}$ **35** |
| $\nabla f$ | Euclidean gradient of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ |

| | | |
|---|---|---|
| $(H_t)_{t\geq 0}$ | heat semigroup | **74** |
| $(\mathscr{H}_t^n)_{t\geq 0}$ | discrete heat semigroup | **76** |
| $(\bar{H}_t)_{t\geq 0}$ | rescaled heat semigroup | **74** |
| $\mathbb{H}^n$ | $n$-dimensional closed upper half-space | **87** |
| $H^p(\Omega)$ | Sobolev space over $\Omega \subseteq \mathbb{R}^n$ | |
| $H^p(\Omega)$ | homogeneous Sobolev space over $\Omega \subseteq \mathbb{R}^n$ | |
| $h_t(x) := \frac{1}{\sqrt{4\pi t}} e^{-x^2/(4t)}$ | heat kernel on $\mathbb{R}$ | **74** |
| $\bar{h}_t$ | rescaled heat kernel | **74** |
| $\mathcal{I}^n(u^n)$ | discrete Fisher information | **62** |
| $I_\tau^n := \big((n-1)\tau, n\tau\big], n \in \mathbb{N}$ | | **19** |
| $\mathcal{L}$ | Lebesgue measure on $\mathbb{R}$ | |
| $\mathcal{L}^n$ | $n$-dimensional Lebesgue measure on $\mathbb{R}^n$ | |
| $L_f(z) := zf''(z) - f(z)$ | | **31** |
| $Lip_L(X)$ | space of Lipschitz continuous functions with Lipschitz constant at most $L$ on $X$ | |
| $L(\rho) := \int_0^1 \sqrt{g\big(\dot\rho(t), \dot\rho(t)\big)}\,\mathrm{d}t$ | length of a smooth curve $\rho : [0,1] \to M$ | **39** |
| $\lvert\dot v\rvert(t+)$ | right-hand derivative of a curve $v : (a,b) \to X$ | **17** |
| $\lvert\dot v\rvert$ | metric differential of a curve $v : (a,b) \to X$ | **11** |
| $\mathcal{M}(X)$ | space of signed measures with finite total variation on $X$ | **75** |
| $\mathcal{M}_\sigma(X)$ | space of Borel measures with finite variation on $(X, \mathcal{T})$ | **83** |
| $\overline{M}_\tau$ | discrete solution | **19** |
| $M_\tau^0$ | discrete initial datum | **19** |
| $M_\tau^n$ | $\tau$-discrete minimizing movement | **19** |
| $\nabla\cdot\psi_i := \frac{1}{2}\sum_{j=1}^n \mathcal{Q}_{ij}(\psi_{ij} - \psi_{ji})$ | discrete divergence for a function $\psi : \mathcal{X}^n \times \mathcal{X}^n \to \mathbb{R}$ | **42** |
| $\mathcal{O}(n)$ | big O notation | |
| $\Omega := (0,1)$ | spatial domain of the finite volume scheme | **59** |
| $(f)^+ := \max\{f, 0\}$ | positive part of $f$ | **13** |
| $(f)^- := -\min\{f, 0\}$ | negative part of $f$ | **13** |
| $\Phi_h$ | | **23** |
| $\Pi(\mu_1, \mu_2)$ | class of admissible plans | **23** |
| $\Pi_{\mathrm{opt}}(\mu_1, \mu_2$ | class of optimal plans | **23** |
| $\mathcal{P}_\delta^n$ | | **78** |
| $\mathcal{P}_p(X)$ | | **25** |
| $\mathcal{P}^n$ | class of all discrete probability densities with respect to a probability measure $\pi$ on $\mathcal{X}^n$ | **37** |
| $\Psi_h$ | | **24** |
| $P_\tau$ | uniform partition of $(0, +\infty)$ with size $\tau > 0$ | **19** |
| $\pi^i$, $\pi^{i,j}$ | projection | **23** |
| $\mathcal{R}^n(u^n, \dot u^n) := \frac{1}{2}\langle \dot u^n, \dot u^n\rangle_g$ | discrete dissipation potential | **61** |
| $\overline{\mathbb{R}}_+^n$ | | **87** |
| $\hat\rho_{ij} := \theta_m(\rho_i, \rho_j)$ | | **37** |
| $\lvert\partial\phi\rvert$ | slope of $\phi$ | **13** |
| $\sigma(X, Y)$ | weak topology on $X$ with respect to $Y \subseteq X'$ | |
| $\sigma_i^n := \frac{x_{i+1}^n + x_i^n}{2}$ | midpoints of the partition $0 = x_1^n < x_2^n < \ldots < x_n^n = 1$ | **59** |

| | |
|---|---|
| $\mathrm{T}(\mu_1, \mu_2)$ | class of admissible transport maps **24** |
| $\mathrm{T}_{\mathrm{opt}}(\mu_1, \mu_2)$ | class of all optimal transport maps **24** |
| $\vartheta$ | weight function **35** |
| $\theta_m(s,t) := \frac{m-1}{m}\frac{s^m-t^m}{s^{m-1}-t^{m-1}}$ | **35** |
| $\mathcal{T}_Y := \{O \cap Y : O \in \mathcal{T}\}$ | subspace topology on $Y \subseteq X$ with respect to $(X, \mathcal{T})$ |
| $\Vert\cdot\Vert_{\mathrm{TV}}$ | total variation norm of a signed measure |
| $\tilde{\theta}_m$ | pseudo-weight function **66** |
| $\mathfrak{U}(x)$ | neighborhood filter at point $x$ of a topological space $(X, \mathcal{T})$ |
| $u_i^n : (0,T) \to \mathbb{R}$ | finite volume discretisation **60** |
| $V$ | potential energy functional **29** |
| $\mathcal{V}(\rho) := \sum_{i=1}^n \upsilon_i \rho_i \pi_i$ | discrete potential energy functional **42** |
| $v_i^n = \int_{\omega_i^n} \upsilon(x)\,\mathrm{d}x$ | discretisation of drift potential $\upsilon$ **60** |
| $W^{**}$ | convex envelope of $W$ **52** |
| $W_p^{\mathrm{gra}}$ | $p$-Wasserstein distance on $\mathcal{P}^n$ with respect to the graph distance induced by $\mathcal{Q}$ **41** |
| $\mathcal{W}_\delta$ | **78** |
| $W_p(\mu, \nu)$ | Wasserstein distance of order $p$ **25** |
| $\omega_i^n := [\sigma_{i-1}^n, \sigma_i^n)$ | control volume **59** |
| $X^*$ | algebraic dual space of $X$ |
| $X^{**}$ | double dual space of $X$ |
| $X'$ | topological dual space of $X$ |
| $\mathcal{X}^n \simeq \{1, 2, \ldots n\}$ | set of cardinality $n \in \mathbb{N}$ **35** |

# Index