Technische Universität München
Fakultät für Mathematik
Lehrstuhl für Dynamics

L^p-Wasserstein and flux-limited gradient flows: Entropic discretization, convergence analysis and numerics

Benjamin Söllner

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende:		Prof. Dr. Christina Kuttler
Prüfer der Dissertation:	1.	Prof. Dr. Daniel Matthes
	2.	Prof. Dr. Guillaume Carlier

Die Dissertation wurde am 06.05.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 28.05.2020 angenommen.

Abstract

This Ph.D. thesis is concerned with the investigation of structure-preserving, spatio-temporal discretizations and temporal semi-discretizations for approximating PDEs with gradient flow structure. We are considering evolution problems on the space of probability measures equipped with some optimal transport distance implied by a family of cost functions that include in particular p-Wasserstein cost $s \mapsto \frac{1}{p}|s|^p$ and flux-limiting cost functions like $s \mapsto \gamma\left(1-\sqrt{1-\left|\frac{s}{\gamma}\right|^2}\right)$ for $|s| \leq \gamma$. We will investigate variational formulations to receive a one-dimensional Lagrangian scheme, a multi-dimensional Eulerian scheme with entropic regularization and a temporal semi-discretization of order two with entropic regularization. We show preservation of qualitative properties of the PDE under discretization. In particular, non-negativity, free energy monotonicity, mass conservation, comparison principles, etc. . In each case, convergence to a solution of the PDE with vanishing discretization parameters is the main result.

Zusammenfassung

Diese Doktorarbeit beschäftigt sich mit strukturerhaltenden Raum-Zeit-Diskretisierungen und Zeit-Diskretisierungen zur Approximation von partiellen Differentialgleichungen die eine Gradientenflußstruktur besitzen. Wir betrachten dabei insbesondere Probleme auf dem Raum der Wahrscheinlichkeitsmaße mit Metriken die durch Probleme optimalen Transports bzgl. einer Familie von Kostenfunktionen gegeben sind die insbesondere p-Wasserstein Kosten $s \mapsto \frac{1}{p}|s|^p$ und flussbegrenzende Kosten $s \mapsto \gamma\left(1-\sqrt{1-\left|\frac{s}{\gamma}\right|^2}\right)$ for $|s| \leq \gamma$ beinhaltet. Wir untersuchen variationelle Formulierungen um ein eindimensionales Lagrangesches Schema, ein mehrdimensionales Eulersches Schema mit entropischer Regularisierung und eine zeitliche Diskretisierung der Ordnung zwei mit entropischer Regularisierung zu erhalten. Wir zeigen, dass einige Eigenschaften von Lösungen der PDE von den Schema erhalten werden. Insbesondere Nichtnegativität, Monotonie bzgl. der freien Energie, Massenerhaltung, Vergleichssätze, usw. . In jedem der drei Fälle ist das Hauptresultat die Konvergenz der Approximation gegen eine Lösung der PDE mit verschwindenden Diskretisierungsparametern.

Contents

Acknowledgements

First of all, I would like to thank my advisor Prof. Dr. Daniel Matthes for the great opportunity to work with him. Without his guidance and support as well as his valuable feedback this thesis would not have been possible. I am particularly grateful for our hours of constructive discussion, for reading my pieces of work, for sharing his mathematical insights into them and for bearing my occasional lack of zeal.

I want to express my gratitude to Prof. Dr. Guillaume Carlier for reviewing this thesis and the kind hospitality at the INRIA MOKAPLAN in Paris and I am particularly thankful for his feedback on my research during my stay. Furthermore I would like to thank Prof. Dr. Oliver Junge for the time we spend on our joint project and Prof. Dr. Jürgen Scheurle for his invaluable support especially during my first year at his chair. I would also like to thank Prof. Dr. Christina Kuttler for taking the chair of the examination committee.

Many thanks to all my friends and colleagues at the faculty for stimulating discussions on a multitude of mathematical and non-mathematical topics and the atmosphere of mutual helpfulness. Out of those friends and colleagues I want to thank Dr. Carl-Friedrich Kreiner in particular for accompanying me from my very first lecture in mathematics up until now with a constantly open door, advice and counsel whenever I needed them. Furthermore I want to thank Dr. Andreas Johann, Dr. Frank Hofmaier and Dr. Hans-Peter Kruse for their collegiality and helpfulness in teaching matters. A special thanks goes to my former office mate Dr. Simon Plazotta who substantially added to the positive work environment and endured my habits in addition to always being ready to share his insight in mathematical and non-mathematical matters.

I especially want to thank my family, for giving me the opportunity to become a mathematician, for their patience with me and their unconditional care for me. Last but not least I want to thank Bianca for her love and general support in every situation of life. It enabled me to go through all ups and downs of my academic career in mathematics.

Chapter 1

Introduction and Main Results

In this thesis will consider schemes for approximation of solutions to the following family of non-linear parabolic equations

$$\partial_t \rho = \operatorname{div}\, \left[\rho \cdot \mathcal{A}\left(\nabla[\mathrm{u}(\rho) + \mathrm{v} + \mathrm{w} * \rho]\right)\right] \tag{1.0.1}$$

where ρ is a curve in the space of probability measures on an open set $\Omega \subset \mathbb{R}^d$. Here $\mathcal{A} : \mathbb{R}^d \to \mathbb{R}^d$ will be a – possibly non-linear – map given by $\mathcal{A} = \nabla(\mathbf{c}^*)$. $\mathbf{c}^*$ denotes the Legendre transform $\mathbf{c}^*(z) = \sup_y \langle z, y\rangle - \mathbf{c}(y)$ of a cost function $\mathbf{c} : \mathbb{R}^d \to [0, \infty]$. u, v and w will be internal energy potentials, external potentials and interaction potentials respectively. These functions are subject to further hypothesis in each chapter.

To motivate the schemes we will be investigating, we will start at the JKO-scheme or minimizing movement scheme (c.f. [23]) which is a well established scheme for temporal discretization of solutions for equations with gradient flow structure. Incidentally, (1.0.1) possesses a gradient flow structure on the set of probability measures on Ω w.r.t. a suitable optimal transport distance $\mathbf{T}_{\mathbf{c},\tau}$, equipped with a time-step parameter τ, for a free energy functional $\mathcal{E}$.

We will give a short sketch of the JKO-scheme for our equation (1.0.1) and introduce $\mathbf{T}_{\mathbf{c},\tau}$ and $\mathcal{E}$ in the process. Assume ρ_* is a solution of this equation, then the JKO-scheme defines, for a time step $\tau > 0$, a sequence of approximations $\rho_\tau^{(n)}$ of $\rho_*(t)$ at times $t = n\tau$ by inductive solution of

$$\rho_\tau^{(n)} \in \arg\min_\rho \tau \mathbf{T}_{\mathbf{c},\tau}(\rho, \rho_\tau^{(n-1)}) + \mathcal{E}(\rho)\ , \tag{1.0.2}$$

where $\mathbf{T}_{\mathbf{c},\tau}(\rho, \mu)$ is an optimal transport distance between two probability measures ρ, μ on Ω for cost $\mathbf{c}$ with parameter τ defined via the minimization problem

$$\mathbf{T}_{\mathbf{c},\tau}(\rho, \mu) = \inf_{\gamma \in \Pi(\rho,\mu)} \iint_{\Omega^2} \mathbf{c}\left(\frac{x-y}{\tau}\right) \mathrm{d}\gamma(x, y)\ . \tag{1.0.3}$$

$\Pi(\rho, \mu)$ denotes the set of all transport plans from μ to ρ i.e. probability measures γ on $\Omega \times \Omega$ with first and second marginal ρ and μ respectively. Furthermore $\mathcal{E}(\rho)$ is a free energy functional of the form

$$\mathcal{E}(\rho) = \begin{cases} \int_\Omega \mathrm{u}(\rho) + \mathrm{v} + \mathrm{w} * \rho \,\mathrm{d}x & \text{if } \rho \text{ is a.c. w.r.t. Lebesgue} \\ +\infty & \text{otherwise} \end{cases} \tag{1.0.4}$$

where

$$\mathrm{u}(s) = \mathrm{u}_m(s) = \begin{cases} s\log(s) & \text{if } m = 1 \\ \frac{1}{m-1} s^m & \text{if } m > 1 \end{cases} \tag{1.0.5}$$

and $s\log(s)$ is assumed to be continuously extended to $[0,\infty)$. Furthermore v and w will be an external and an interaction potential assumed to be $C^2(\mathbb{R}^d)$. Note that (1.0.1) can be rewritten to read

$$\partial_t\rho = \operatorname{div}\left[\rho\cdot\mathcal{A}\left(\frac{\delta\mathcal{E}}{\delta\rho}\right)\right]$$

where $\frac{\delta\mathcal{E}}{\delta\rho}$ denotes the first variation of $\mathcal{E}$.

Our goal is to actually calculate approximate solutions. But applying a naive discretization in space using some constant mesh to arrive at a fully discrete scheme will result in a scheme that is expensive in each step. This can easily be seen by the encapsulated second minimization problem $\mathbf{T}_{\mathbf{c},\tau}$ in the minimization problem that is to be solved in each step. A main goal of this thesis was to consider schemes that, already as temporal semi-discretization, do not suffer from this problem and therefore the corresponding spatio-temporal discretization do not suffer from this problem, too.

One possible way of completely getting rid of the minimization in $\mathbf{T}_{\mathbf{c},\tau}$, which is examined in **Chapter 3**, is to consider equation (1.0.1) in one spatial dimension. Then the inner minimization problem is explicitly evaluable and the infimum amounts to

$$\mathbf{T}_{\mathbf{c},\tau}(\rho,\mu) = \int_\Omega \mathbf{c}\left(\frac{X_\rho(\xi)-X_\mu(\xi)}{\tau}\right)\,\mathrm{d}\xi \tag{1.0.6}$$

where X_ρ, X_μ are the inverse cummulative distribution functions of ρ and μ respectively. This is explained in more detail in **Subsection 2.2.3**. As we can see, this new form does not involve any minimization in the optimal distance any more, leaving our JKO-scheme with one minimization only, which can be calculated in each step with a reasonable amount of effort. This method is utilized in **Chapter 3** to define a Lagrangian scheme for finding approximate solutions of (1.0.1). The family of cost functions considered therein is quite general and includes also particular cost functions that result in (1.0.1) becoming the one-dimensional p-Laplace equation

$$\partial_t\rho = \partial_x\left(|\partial_x\rho|^q\cdot\partial_x\rho\right)\,. \tag{1.0.7}$$

or Rosenau's relativistic heat equation on the line

$$\partial_t\rho = \partial_x\left(\rho\frac{\partial_x\rho}{\sqrt{\rho^2+|\partial_x\rho|^2}}\right)\,. \tag{1.0.8}$$

The second way to handle the difficulty of the two nested minimizations of the basic JKO-scheme is to apply the powerful machinery of entropic regularization. Replace $\mathbf{T}_{\mathbf{c},\tau}$ with the corresponding entropic regularized optimal transport cost, which consists of adding the ε-scaled negative entropy

$$\mathcal{H}(\gamma) := \iint_{\Omega^2}\gamma(x,y)\log\left(\gamma(x,y)\right)\,\mathrm{d}(x,y) \tag{1.0.9}$$

where $\mathcal{H}(\gamma)=+\infty$ if γ does not admit a density w.r.t. Lebesgue, of the transport plans γ in the optimal transport minimization problem:

$$\mathbf{T}_{(\mathbf{c},\tau,\varepsilon)}(\rho,\mu) = \inf_{\gamma\in\Pi(\rho,\mu)}\iint_{\Omega^2}\mathbf{c}\left(\frac{x-y}{\tau}\right)\,\mathrm{d}\gamma(x,y)+\varepsilon\mathcal{H}(\gamma)\,. \tag{1.0.10}$$

Then combine the two minimization problems in (1.0.2) in one minimization problem involving the first and second marginal $(\mathrm{P}_x)_\#\gamma$, $(\mathrm{P}_y)_\#\gamma$ of a probability measure γ on $\Omega\times\Omega$ in the following way

$$\gamma^{(n)} \in \operatorname*{arg\,min}_{\gamma\in\mathscr{P}(\Omega\times\Omega)}\ \Psi^n_{\tau,\varepsilon}(\gamma) \tag{1.0.11}$$

where the argument of the minimization problem is given by

$$\Psi^n_{\tau,\varepsilon}(\gamma) := \tau\varepsilon \iint_{\Omega^2} \gamma(x,y) \log\left(\frac{\gamma(x,y)}{\mathrm{K}_{\varepsilon,\tau}(x,y)}\right) \mathrm{d}(x,y) + \mathcal{E}((\mathrm{P}_x)_\#\gamma) + \mathbf{1}_{\rho^{(n-1)}}((\mathrm{P}_y)_\#\gamma)\ . \tag{1.0.12}$$

Here $\mathrm{K}_{\varepsilon,\tau}(x,y) := \mathrm{e}^{-\frac{\mathbf{c}(\frac{x-y}{\tau})}{\varepsilon}}$ and $\mathbf{1}_\mu(\rho)$ is the indicator-function that amounts to 0 if $\rho = \mu$ and $+\infty$ otherwise. This minimization problem can then be solved very efficiently in terms of Dykstra's projection algorithm and the so found minimizer gives us our sought for $\rho^{(n)}$ as its first marginal. This procedure will be considered in detail in **Chapter 4** where we consider a family of flux-limiting cost functions, explicitly including $\Omega = \mathbb{R}^d$, in particular including Rosenau's relativistic heat equation

$$\partial_t\rho = \operatorname{div}\left(\rho\frac{\nabla\rho}{\sqrt{\rho^2 + \|\nabla\rho^2\|}}\right)\ . \tag{1.0.13}$$

Finally, in **Chapter 5** we want to apply entropic regularization again, this time to a temporal discretization of (1.0.1) by means of the backward differentiation formula 2 (*BDF 2* scheme for short). The JKO scheme can be considered as a BDF 1 scheme (or backward Euler method), and in this spirit, we can define a discretization for approximating solutions to (1.0.1) that can be considered to be a BDF 2 scheme. In finite dimensions, the BDF 2 scheme results in faster convergence of the approximate solutions than the backward Euler method. This is the reason why we consider an analogue of the finite dimensional BDF 2 scheme. We want to illustrate how to arrive at this BDF 2 scheme, coming from the finite dimensional backward Euler method.

We start by considering a gradient flow in $\mathbb{R}^d$

$$\dot{x} = -\nabla E(x)$$

in the energy landscape of $E : \mathbb{R}^d \to \mathbb{R}$, a convex function. Then one possible way to find an approximate solution to a solution of this gradient flow would be the backward Euler method. For some time-step size $\tau > 0$ and initial vector x^0, in each step the equation

$$\frac{x - x^{n-1}}{\tau} = -\nabla E(x)$$

is solved for x to receive x^n as approximation for a solution curve $x_* : [0,T] \to \mathbb{R}^d$ at time $x_*(n\tau)$. This equation can be considered to be the first order condition of a minimization problem. x^n solves the above equation if

$$x^n \in \operatorname*{arg\,min}_{x\in\mathbb{R}^d} \frac{\left\|x - x^{n-1}\right\|^2}{2\tau} + E(x)$$

holds (c.f. (1.0.2)). Now this backward Euler method can be viewed as the first of a family of backward methods called *backward differentiation formulas.* The recursion problem of the second one reads: find x^n such that

$$\frac{3x^n - 4x^{n-1} + x^{n-2}}{2\tau} = -\nabla E(x^n) \tag{1.0.14}$$

holds. Its first order condition is then

$$x^n \in \operatorname*{arg\,min}_{x\in\mathbb{R}^d} \frac{1}{\tau}\left(\left\|x - x^{n-1}\right\|^2 - \frac{1}{4}\left\|x - x^{n-2}\right\|^2\right) + E(x)\ .$$

Now taking this over to our case of probability measures with the optimal transport distance and our functional $\mathcal{E}$ corresponding to (1.0.1), we arrive at the following recursion

$$\rho_\tau^{(n)} \in \arg\min_\rho \tau \left(\mathbf{T}_{\mathbf{c},\tau}(\rho_\tau^{(n)}, \rho_\tau^{(n-1)}) - \frac{1}{4}\mathbf{T}_{\mathbf{c},\tau}(\rho_\tau^{(n-2)}, \rho_\tau^{(n)}) \right) + \mathcal{E}(\rho) \,. \tag{1.0.15}$$

Finally replacing the optimal transport distances again by entropic regularized ones, we arrive at the scheme that we will refer to as *entropic regularized BDF 2 scheme*

$$\rho_\tau^{(n)} \in \arg\min_\rho \tau \left(\mathbf{T}_{(\mathbf{c},\tau,\varepsilon)}(\rho_\tau^{(n)}, \rho_\tau^{(n-1)}) - \frac{1}{4}\mathbf{T}_{(\mathbf{c},\tau,\varepsilon)}(\rho_\tau^{(n-2)}, \rho_\tau^{(n)}) \right) + \mathcal{E}(\rho) \tag{1.0.16}$$

for approximating solutions to (1.0.1). This is the scheme that will be investigated in **Chapter 5**.

In each of the above cases, our main result will be the convergence of a sequence of approximate solutions. Let us describe the plan of each of these chapters and their main contribution as well as the plan of proof a bit more in detail.

Chapter 3

In [32, 34], Matthes and Osberger applied a one-dimensional Lagrangian scheme to the (non-linear) Fokker-Planck equation initially suggested by L. Gosse and G. Toscani in [22]. In the first chapter we pick up this concept and extend the results from quadratic cost functions to a family of quite general cost functions, including p-cost $s \mapsto \frac{1}{p}|s|^p$ and flux-limiting cost $s \mapsto 1 - \sqrt{1-|s|^2}$, and as was already mentioned above, this choice of cost will include the q-Laplace equation and Rosenau's relativistic heat equation in the family of equations that are investigated.

The main result is the convergence of a fully-discrete approximation in terms of inverse cummulative distribution functions to a solution of (1.0.1). Additionally, monotonicity in the free energy functional, non-negativity, mass conservation and comparison principles are shown for the approximate solutions. Numerical experiments and in particular a numerical convergence analysis are also included.

The proof of convergence consists of steps that are standard in the analysis of gradient flows by means of the JKO-scheme. First we establish the existence of a minimizer, then the Euler-Lagrange equation of said minimizer is calculated and by the a priori estimates we arrive at the classical estimates, giving us enough compactness to pass to a limit with our approximate solution by means of an Aubin-Lions argument. Finally, passing to the limit in our Euler-Lagrange equation will show that the limit curve satisfies (1.0.1) in a weak sense. Non-standard, however, is that we will step from the formulation of (1.0.1) in terms of probability measures ρ to the formulation in terms of inverse (cumulative) distribution functions (IDF) and all these steps described above are then taken in this IDF-framework.

Chapter 4

McCann and Puel proved in [36] the applicability of the JKO-scheme to flux-limiting equations including Rosenau's relativistic heat equation. In particular, that an approximate solution obtained by means of a JKO-scheme converges to a solution of a flux-limiting instance of (1.0.1). After Cuturi introduced in [16] efficient solvers for the entropic regularized optimal transport problem, Peyré demonstrated in [42] their applicability to solve efficiently the minimization problem in an entropic regularized JKO-step. We prove the convergence of an approximate solution obtained from an entropic regularized JKO-scheme to an approximate solution obtained by an unregularized JKO-scheme by means of Γ-convergence. A description of the numerical scheme and numerical experiments are presented at the end, in particular involving a non-convex domain.

The plan of proof in this chapter was to show that the functionals $\Psi^n_{\tau,\varepsilon}$ admit a Γ-convergence limit, which turns out to be the functional that is subject to minimization in an unregularized JKO-step

$$\Psi^n_{\tau,0}(\gamma) := \tau \iint_{\Omega^2} \mathbf{c}\left(\frac{x-y}{\tau}\right) \mathrm{d}\gamma(x,y) + \mathcal{E}((\mathrm{P}_x)_\#\gamma) + \mathbf{1}_{\rho^{(n-1)}}((\mathrm{P}_y)_\#\gamma) \; . \tag{1.0.17}$$

Consequently, the minimizers of $\Psi^n_{\tau,\varepsilon}$ converge narrowly to minimizers of $\Psi^n_{\tau,0}$ and their first marginals converge in the same way to minimizers of the unregularized JKO-scheme. The main hurdle in showing the Γ-convergence was the construction of the recovery sequence needed for the „lim sup“-condition. The construction consists of several involved steps and is surprisingly delicate.

Chapter 5

In [33, 43] the BDF 2 scheme introduced above in (1.0.15) was presented and analysed and just as in these articles, we will consider this scheme with cost function $\mathbf{c}(s) = \frac{1}{2}|s|^2$. As was already mentioned above, this scheme can be considered to be a higher order version of the JKO-scheme. As in **Chapter 4** we want to allow for the entropic regularization to be applied to this higher order scheme, so a recursive sequence defined by (1.0.16) is considered. Proof of convergence to a solution of (1.0.1) in the limit of vanishing time-step and entropic regularization is the main result here.

As in **Chapter 3** the proof follows the standard steps for proving convergence in the JKO-case, but this time we have to overcome additional difficulties. In particular, the negative distance in (1.0.16) will on the one hand obscure the fact that the argument of the arg min in (1.0.16) is in fact narrowly lower semi-continuous in ρ. When this lower semi-continuity is established, it will lead to the existence of a minimizer as easily as in the JKO-case. On the other hand additional regularity of the entropic regularized optimal transport distance is required to receive the Euler-Lagrange equation. This is again due to the negative distance appearing in the scheme. Both problems will be solved by decomposition of the entropic regularized optimal transport distance, a decomposition which the author was advised of by Guillaume Carlier.

Chapter 2

Notation and Preliminaries

2.1 Function spaces

In this thesis $\Omega \subset \mathbb{R}^d$ will denote an open, connected domain with Lipschitz-boundary. Further restrictions will be stated in the corresponding chapters.

Let us begin with some basic notation. Let $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ be two functions on some spaces X and Y. Then we will denote with $f \oplus g$ the *direct sum* of the two functions $(f \oplus g) : X \times Y \to \mathbb{R}$ that is given by $(f \oplus g)(x, y) = f(x) + g(y)$. Define $\otimes$ accordingly as the *direct product*. Furthermore the usual definitions of weak- and weak-*-convergences are used. Let us now introduce some function spaces and their notation.

Continuous functions

Let $C(\Omega)$ be the space of continuous functions $f : \Omega \to \mathbb{R}$, $C(\Omega, \mathbb{R}^d)$ the space of continuous vector fields $\xi : \Omega \to \mathbb{R}^d$, $C_b(\Omega)$ the space of continuous, bounded functions, $C_c(\Omega)$ the space of continuous functions with compact support in Ω, $C^k(\Omega)$ the space of k-times continuously differentiable functions and $C^\infty(\Omega) := \bigcap_{k\in\mathbb{N}} C^k(\Omega)$ as well as $C^0(\Omega) := C(\Omega)$. Define $C_b(\Omega, \mathbb{R}^d), C_c(\Omega, \mathbb{R}^d), C^k(\Omega, \mathbb{R}^d)$ and $C^\infty(\Omega, \mathbb{R}^d)$ analogously for vector fields. Combinations are to be defined as suitable intersections, for example $C_c^\infty(\Omega) = C_c(\Omega) \cap C^\infty(\Omega)$. The space of continuous functions and its subspaces are to be equipped with the uniform norm

$$\|f\|_\infty := \sup_{x\in\Omega} |f(x)| \ .$$

The C^k-spaces are equipped with

$$\|f\|_{C^k(\Omega)} := \sum_{|\alpha|\leq k} \|\partial^\alpha f\|_\infty \ .$$

Again the definition for the vector-field spaces is to be made accordingly.

L^p-spaces

Denote by $\mathcal{L}^d$ the d-dimensional Lebesgue-measure on the domain Ω. We will call a function f Lebesgue-measurable iff the $f^{-1}(A)$ is measurable for every Lebesgue-measurable set $A \subset \mathbb{R}$. For $p \in [1, \infty)$ the L^p-norm of a Lebesgue-measurable function f is defined as

$$\|f\|_{L^p(\Omega)}^p := \int_\Omega |f(x)|^p \ \mathrm{d}\mathcal{L}^d(x)$$

where we will usually abbreviate $\mathrm{d}\mathcal{L}^d(x) = \mathrm{d}x$. For $p = \infty$ we introduce the set of essentially bounded functions $L^\infty(\Omega)$ with the norm

$$\|f\|_{L^\infty(\Omega)} := \operatorname*{ess\,sup}_{x\in\Omega} |f(x)| \, .$$

$L^p(\Omega)$ denotes the set of all Lebesgue-measurable functions with finite L^p-norm.

$L^p(\Omega)$ is a Banach space for every $p \in [1,\infty]$ and for $p \in [1,\infty)$ its dual is given as $L^q(\Omega)$ where q is the Hölder-conjugate of p given by the equation $\frac{1}{p} + \frac{1}{q} = 1$. We define the corresponding spaces $L^p(\Omega, \mathbb{R}^d)$ of p-integrable vector fields analogously.

Functions of bounded variation.

For a given open domain Ω a function $f \in L^1(\Omega)$ is called a *function of bounded variation* iff

$$V(f,\Omega) := \sup\Big\{ \int_\Omega f(x) \operatorname{div}(\xi(x)) \,\mathrm{d}x \mid \xi \in C_c^\infty(\Omega, \mathbb{R}^d), \ \|\xi\|_\infty \le 1 \Big\} < \infty$$

(c.f.[21]). The set of all functions of bounded variation is denoted by $BV(\Omega)$ with the norm:

$$\|f\|_{BV(\Omega)} = \|f\|_{L^1(\Omega)} + V(f,\Omega).$$

For open sets $\Omega \subset \mathbb{R}^d$ the set $BV(\Omega)$ is a Banach space.

Sobolev spaces.

We say a Lebesgue-measurable scalar function f is k-times weakly differentiable if for each multi-index α of order k there exists a function $g^\alpha : \Omega \to \mathbb{R}$ lying in $L^1(K)$ for every $K \subset \Omega$ compact such that

$$\int_\Omega f \partial^\alpha \psi(x) \,\mathrm{d}x = (-1)^{|\alpha|} \int_\Omega g^\alpha \psi(x) \,\mathrm{d}x \quad \forall\, \psi \in C_c^\infty(\Omega).$$

In this case g^α is unique and we denote $g^\alpha = \partial^\alpha f$. Then for $k \in \mathbb{N}$ and $p \in [1,\infty]$ the Sobolev space $\mathrm{W}^{k,p}(\Omega)$ is defined as the set of all Lebesgue-measurable functions f such that $\partial^\alpha f$ exists for all multi-indices $|\alpha| \le k$ and

$$\|f\|_{\mathrm{W}^{k,p}(\Omega)} := \Big(\|f\|^p_{L^p(\Omega)} + \sum_{|\alpha|\le k} \|\partial^\alpha f\|^p_{L^p(\Omega)} \Big)^{1/p} < \infty.$$

For any $k \in \mathbb{N}$ and any $p \in [1,\infty]$ the Sobolev space $\mathrm{W}^{k,p}(\Omega)$ is a Banach space with the norm $\|\cdot\|_{W^{k,p}(\Omega)}$.

Bochner spaces and convergence in measure

We will now consider (not necessarily continuous) curves $u : [0,T] \to V$ in a Banach space V with norm $\|\cdot\|_V$. We will call such a curve *simple* if it is piecewise constant on a finite decomposition of $[0,T]$ in measurable subsets (c.f. [29]).

We say that a curve $u : [0,T] \to V$ is measurable iff there is a sequence of simple curves $u_n : [0,T] \to V$ such that $u_n(t) \to u(t)$ in V for a.e. $t \in [0,T]$. The set of measurable curves will be denoted $\mathcal{M}(0,T;V)$.

We say a sequence $(u_n)_{n\in\mathbb{N}}$ in $\mathcal{M}(0,T;V)$ *converges in measure* therein to a limit curve u_* if and only if

$$\lim_{n\to\infty} \mathcal{L}^1(\{t \in [0,T] \mid \|u_n(t) - u(t)\|_V \ge \epsilon\}) = 0 \qquad \text{for all } \epsilon > 0 \, .$$

We will say that a sequence is *relatively compact in* $\mathcal{M}(0,T;V)$ if each subsequence possesses a subsubsequence converging in measure in $\mathcal{M}(0,T;V)$.

Furthermore for $p \in [1,\infty)$ we define the *Bochner space* $L^p(0,T;V)$ as the set of all $u \in \mathcal{M}(0,T;V)$ such that

$$\|u\|_{L^p(0,T;V)} := \begin{cases} \left(\int_0^T \|u(t)\|_V^p \, \mathrm{d}t \right)^{1/p} & \text{if } p \in [1,\infty) \\ \operatorname{ess\,sup}_{t\in[0,T]} \|u(t)\|_V & \text{if } p = \infty \end{cases}$$

is finite.

The space of probability measures

Let $\Omega \subset \mathbb{R}^d$. We define the set $\mathscr{P}(\Omega)$ to be the set of all probability measures on Ω. The subset of probability measures admitting a density w.r.t. Lebesgue measure are denoted $\mathscr{P}^{ac}(\Omega)$. Usually we will not distinguish notationally between a measure $\rho \in \mathscr{P}^{ac}(\Omega)$ and its density w.r.t. Lebesgue, except when the identification is not clear from the context. Let $T : \Omega \to \Omega$ be a measurable map, then we define the push-forward of a measure $\rho \in \mathscr{P}(\Omega)$ as $T_\#\rho$ given by $T_\#\rho(B) = \rho(T^{-1}(B))$ for all measurable sets B. Consider the product space $\Omega \times \Omega = \Omega^2$ and the projections $\mathrm{P}_x, \mathrm{P}_y : \Omega^2 \to \Omega$ given by $\mathrm{P}_x(x,y) = x$ and $\mathrm{P}_y(x,y) = y$. Therefore, given a $\gamma \in \mathscr{P}(\Omega^2)$, the first and second marginal distributions of γ can be written as $(\mathrm{P}_x)_\#\gamma$ and $(\mathrm{P}_y)_\#\gamma$ and we define for $\rho, \mu \in \mathscr{P}(\Omega)$ the set

$$\Pi(\rho,\mu) := \{\gamma \in \mathscr{P}(\Omega^2) \mid (\mathrm{P}_x)_\#\gamma = \rho, (\mathrm{P}_y)_\#\gamma = \mu\} . \tag{2.1.1}$$

Let $\rho, \mu \in \mathscr{P}(\Omega)$ then the product measure $\rho\otimes\mu$ is the measure $(\rho\otimes\mu) \in \mathscr{P}(\Omega^2)$ defined by $(\rho\otimes\mu)(A\times B) = \rho(A)\mu(B)$ for all measurable sets $A, B \subset \Omega$. Note that for ρ, μ with densities r, m this implies that the density of $\rho \otimes \mu$ is given by the direct product of the densities $r \otimes m$, making this notation consistent.

The notion of convergence we will consider in $\mathscr{P}(\Omega)$ will be narrow convergence of measures. That is, for a sequence and a limit $\rho_k, \rho_* \in \mathscr{P}(\Omega)$ we say ρ_k converges narrowly to ρ_*, $\rho_k \rightharpoonup \rho_*$, iff for all bounded continuous functions $\phi \in C_b(\Omega)$

$$\lim_{k\to\infty} \int_\Omega \varphi(x) \, \mathrm{d}\rho_k(x) = \int_\Omega \varphi(x) \, \mathrm{d}\rho_*(x) . \tag{2.1.2}$$

Since for bounded Ω the spaces $C_b(\Omega)$ and $C(\Omega)$ coincide, the notion of narrow convergence can be adjusted in that case.

2.1.1 An extension of the Aubin-Lions theorem

We will specify *Theorem 2* from [45] to our needs. This will be the main tool in **Chapters 3** and **5** to conclude convergence. Let us define two objects first, that will be needed in stating the theorem.

Definition 2.1. *Let V be a Banach space with norm $\|\cdot\|_V$. Then we will call $\mathfrak{F} : V \to [0,\infty]$ a* normal, coercive integrand *if*

1. *$\mathfrak{F}$ is measurable w.r.t. to the Borel subsets of V;*

2. *the map $u \mapsto \mathfrak{F}(u)$ is lower semicontinuous;*

3. *the map $u \mapsto \mathfrak{F}(u)$ has compact sublevels in V.*

We will call $\mathbf{g}$ a pseudo-distance on V w.r.t. $\mathfrak{F}$ if

1. *for all $u, v \in \mathfrak{F}^{-1}([0,\infty))$ the equation $\mathbf{g}(u,v) = 0$ implies $u = v$;*

2. the map $(u,v) \mapsto \mathbf{g}(u,v)$ is joint lower semi-continuous.

Example 2.2. *An example for a pair $\mathfrak{F}$ and $\mathbf{g}$ that satisfies the above definition if Ω is bounded with $V = L^m(\Omega)$, is*

$$\mathfrak{F}(\rho) = \begin{cases} \int_\Omega |\rho(x)|^m + \|\nabla\rho(x)\|^m \, \mathrm{d}x & \text{if } \rho \in W^{1,m}(\Omega) \\ +\infty & \text{otherwise} \end{cases}$$

and $\mathbf{g}(\rho,\mu) = \mathbf{T}_2(\rho,\mu)$.

The result that $\mathfrak{F}$ and $\mathbf{g}$ are indeed a pair of a normal, coercive integrand with a compatible pseudo-distance was shown in [26] [18] and [8].

Now we can adapt *Theorem 2* of [45]

Theorem 2.3 (Extension of Aubin-Lions)**.** *Let V be a Banach space with norm $\|\cdot\|_V$. Let $\mathfrak{F} : V \to [0,\infty]$ be a normal coercive integrand and $\mathbf{g} : V \times V \to [0,\infty]$ be a pseudo-distance (c.f.* Definition 2.1*). Let $(u_\tau(t))_{\tau\in J}$ be a family of measurable curves $u_\tau : [0,T] \to V$ such that*

$$\sup_{\tau\in J} \int_0^T \mathfrak{F}(u_\tau(t)) \, \mathrm{d}t < \infty \tag{2.1.3}$$

and

$$\lim_{h\searrow 0} \sup_{\tau\in J} \int_0^{T-h} \mathbf{g}(u_\tau(t+h), u_\tau(t)) \, \mathrm{d}t = 0 \,. \tag{2.1.4}$$

Then (u_τ) is relatively compact in $\mathcal{M}(0,T;V)$.

2.2 Optimal transport

We want to give a short introduction to concepts from optimal transport we will be dealing with in this thesis. To that end we begin with stating the optimal transport problem of Kantorovich followed by the corresponding dual problem. Then we will consider the transport problem on the line, which will be central in **Chapter 3** followed by the entropic regularization for our OT problems, which will be central in **Chapter 4** and **5**.

2.2.1 The optimal transport problems

This thesis is concerned with different aspects of the optimal transport distance. We want to introduce here the *Monge problem*, its relaxation, the *Kantorovich problem*, state some general results concerning existence and then move on to the properties of these on the real line as well as the entropic regularization of the *Kantorovich problem*

The history of optimal transport began with Gaspard Monge, who proposed a problem in [38]. It reads as follows.

Definition 2.4 (The Monge problem)**.** *Let $\Omega \subset \mathbb{R}^d$ and $\rho,\mu \in \mathscr{P}(\Omega)$. Then the* Monge problem *of transporting μ to ρ with cost $\mathbf{c}$ is to minimize*

$$\int_\Omega \mathbf{c}(x - S(x)) \, \mathrm{d}\mu \tag{2.2.1}$$

among all measurable $S : \Omega \to \Omega$ with the property $S_\# \mu = \rho$. These S are called transport maps.

We stated the problem here in more generality than Monge, who only considered $\mathbf{c} = \|\cdot\|_2$ and dimensions 2 and 3. The Monge problem was long unsolved, which is due to its constraint $S_{\#}\mu = \rho$ which is non-linear and therefore complicates the matter as is for example pointed out in [46].

Up until 1942 little progress was made w.r.t. the optimal transport problem and no progress in the question of the existence of a minimizer at all, when in [25] Leonid Kantorovich looked at the problem from a linear programming point of view. He arrived at the following problem, nowadays called the *Kantorovich problem.*

Definition 2.5 (The Kantorovich problem). *Let* $\Omega \subset \mathbb{R}^d$ *and* $\rho, \mu \in \mathscr{P}(\Omega)$. *Then the* Kantorovich problem *of transporting* μ *to* ρ *with cost* $\mathbf{c}$ *reads as follows.*

$$\mathbf{T_c}(\rho,\mu) := \inf_{\gamma\in\Pi(\rho,\mu)} \iint_{\Omega^2} \mathbf{c}(x-y)\,\mathrm{d}\gamma(x,y)\ . \tag{2.2.2}$$

The set $\Pi(\rho,\mu) \subset \mathscr{P}(\Omega^2)$ *was defined in* (2.1.1) *as the set of probability measures* γ *with first and second marginal* ρ *and* μ *respectively. These* γ *are called* transport plans. *Let us furthermore denote with* $\mathbf{T}_2$ *the particular* $\mathbf{T_c}$ *with* $\mathbf{c} = \frac{1}{2}\|\cdot\|_2^2$ *and an* optimal transport problem *refers to the* Kantorovich problem.

The following is a well known result for the optimal transport problem which shows existence of a minimizing transport plan. A proof can for example be found in [46, Theorem 1.7.].

Proposition 2.6. *Let* $\rho, \mu \in \mathscr{P}(\Omega)$ *and* $\mathbf{c} : \Omega^2 \to [0,\infty]$ *a proper, lower semi-continuous and bounded from below. Then the Kantorovich problem admits a minimizer.*

Sketch of proof. First, $\gamma \mapsto \langle \mathbf{c}, \gamma\rangle$ is l.s.c. w.r.t. narrow convergence since $\mathbf{c}$ is. Furthermore $\Pi(\rho,\mu)$ is tight, implying narrow compactness. The bound from below on $\mathbf{c}$ implies a bound from below on $\gamma \mapsto \langle \mathbf{c}, \gamma\rangle$ and then we can apply the direct method of the calculus of variation to conclude the existence of a minimizer. □

2.2.2 The Dual problem

Let us state the dual problem to our Kantorovich problem above.

Definition 2.7 (The dual problem). *Let* $\mathbf{c}, \rho, \mu$ *as in* Definition 2.5. *Then the* dual problem *is defined as*

$$\mathcal{D}(\rho,\mu) := \sup_{\substack{\phi,\psi\in C_b(\Omega)\\ \phi\oplus\psi\le\mathbf{c}}} \int_\Omega \phi(x)\,\mathrm{d}\rho(x) + \int_\Omega \psi(y)\,\mathrm{d}\mu(y)\ . \tag{2.2.3}$$

We can make this definition plausible by a short calculation. Let us express the constraint in the Kantorovich problem, $\gamma \in \mathscr{P}(\rho,\mu)$, in a variational way:

$$\sup_{\phi,\psi\in C_b(\Omega)} \int_\Omega \phi\,\mathrm{d}\rho + \int_\Omega \phi\,\mathrm{d}\mu - \iint_{\Omega^2} \phi\oplus\psi\,\mathrm{d}\gamma = \begin{cases} 0 & \text{if } \gamma\in\Pi(\rho,\mu) \\ +\infty & \text{otherwise.} \end{cases}$$

Now assuming we can exchange inf and sup, and denoting with $M_+(\Omega^2)$ the set of non-negative Borel measures on Ω^2, we arrive at

$$\begin{aligned}
\mathbf{T_c}(\rho,\mu) &= \inf_{\gamma\in M_+(\Omega^2)} \iint_{\Omega^2} \mathbf{c}(x-y)\,\mathrm{d}\gamma + \sup_{\phi,\psi\in C_b(\Omega)} \left\{ \int_\Omega \phi\,\mathrm{d}\rho + \int_\Omega \psi\,\mathrm{d}\mu - \iint_{\Omega^2} \phi\oplus\psi\,\mathrm{d}\gamma \right\} \\
&= \sup_{\phi,\psi\in C_b(\Omega)} \int_\Omega \phi\,\mathrm{d}\rho + \int_\Omega \psi\,\mathrm{d}\mu + \inf_{\gamma\in M_+(\Omega^2)} \iint_{\Omega^2} \mathbf{c}(x-y) - \phi\oplus\psi\,\mathrm{d}\gamma\ .
\end{aligned}$$

Now as soon as $\phi \oplus \psi > \mathbf{c}$ somewhere, γ could concentrate more and more on a neighbourhood showing that the inf results in $-\infty$ contradicting our sup. Therefore $\phi \oplus \psi \leq \mathbf{c}$ is letting the inf-expression vanish and resulting in our definition above.

Let us finish this part by stating the connection between the Kantorovich and the dual problem which holds under quite general assumptions (c.f. [2, Theorem 6.1.1]).

Proposition 2.8 (Duality)**.** *Let* $\mathbf{c} : \Omega^2 \to [0, \infty]$ *be a proper l.s.c. function. Then*

$$\mathcal{D}(\rho, \mu) = \mathbf{T_c}(\rho, \mu) \; .$$

2.2.3 One-dimensional optimal transport

Considering optimal transport problems in one dimension admits some special features not found in dimensions two and up. In one dimension, the problem of moving, lets say one probability density to another one, can be described procedurally. The optimal way consists in starting from one end of the real line and, as you move along, you pick up the particles of μ and put them, also coming from the same end of the real line, where the next particle of ρ is supposed to be. As it turns out, for the families of cost functions considered in **Chapter 3**, this description always yields the optimal transport plan rendering it independent from the cost function.

Since this part is preparing for **Chapter 3**, we will restrict ourselves here to $I = [a, b]$ a compact interval and probability densities that are bounded from above and away from zero, i.e. $\mu \in \mathscr{P}^{ac}(I)$ such that its density satisfies $\mu + \frac{1}{\mu} \in L^\infty(I)$.

Consider some $\mu \in \mathscr{P}(\Omega)$. Then the *cumulative distribution function* U_μ is defined as $\int_a^x \mathrm{d}\mu(x)$ and the corresponding *inverse (cumulative) distribution function* (IDF for short) as $X_\mu = U_\mu^{-1} : [0, 1] \to I$. Since the density m of our μ is bounded from above and away from zero, X_μ is a.c. and $X'_\mu = \frac{1}{m \circ X_\mu}$ holds a.e. .

Proposition 2.9 (Adapted from Chapter 2 [46])**.** *Let* $\rho, \mu \in \mathscr{P}^{ac}(I)$ *with densities bound from above and away from zero. Let* $\mathbf{c} : \mathbb{R} \to \mathbb{R}$ *continuous and strictly convex. Then* $\mathrm{d}\gamma(x, y) = (X_\rho, X_\mu)_\# \, \mathrm{d}\mathcal{L}(x)$ *is the optimal transport plan of* $\mathbf{T_c}(\rho, \mu)$ *and consequently*

$$\mathbf{T_c}(\rho, \mu) = \int_{[0,1]} \mathbf{c} \left(X_\rho(\xi) - X_\mu(\xi) \right) \mathrm{d}\xi$$

holds.

This in particular means, that in the one-dimensional case, there is no minimization problem to be solved in $\mathbf{T_c}$, since the optimal plan can be computed in terms of the IDFs of ρ, μ. **Chapter 3** will rely on this fact to arrive at a numerical scheme that does not suffers from a minimization problem encapsulated in another one.

2.3 Entropic regularized optimal transport

Adding an entropic regularization penalty to the original OT problem, we arrive at the so called *entropic regularized OT* which will be central in **Chapter 4** and **5**. This regularization admits some very nice properties. In particular, in fully discrete schemes it allows for highly efficient calculation of the minimizer by means of Dykstras projection algorithm with Bregman divergences (c.f. [42]).

The entropic regularized optimal transportation problem stated below is related to the Schrödinger problem as was pointed out by Christian Léonard in [28]. The connection of the entropic regularized problem to the unregularized problem is investigated therein as well.

As a possibility to receive efficient solvers for our JKO-schemes, the entropic regularized JKO-scheme was introduced by Gabriel Peyré in [42] and then analysed in [8].

The entropic regularized distance

Definition 2.10. *The entropic regularized optimal transport distance for some cost function* $\mathbf{c}$ *is defined as*

$$\mathbf{T}_{\mathbf{c},\varepsilon}(\rho,\mu) := \inf_{\gamma\in\Pi(\rho,\mu)} \iint_{\Omega^2} \mathbf{c}(x-y)\,\mathrm{d}\gamma(x,y) + \varepsilon\mathcal{H}(\gamma)\ . \tag{2.3.1}$$

We want to define the relative entropy, also known as Kullback-Leibler divergence, here, too,

$$\mathcal{H}(\nu \mid \eta) = \int_X \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\eta}\right)\mathrm{d}\nu \tag{2.3.2}$$

if $\nu << \eta$ *and* $+\infty$ *else. Here* $\frac{\mathrm{d}\nu}{\mathrm{d}\eta}$ *denotes the Radon-Nikodym derivative of* ν *w.r.t.* η.

As with *Proposition 2.6*, we want to state the existence of a minimizer in (2.3.1) and give a short sketch of the proof.

Proposition 2.11. *The minimization problem in* (2.3.1) *where* $\mathbf{c} : \Omega^2 \to [0,\infty]$ *is proper, lower semi-continuous and bounded from below, admits exactly one minimizer.*

Sketch of proof. The proof is similar to the one of *Proposition 2.6.* Again, we begin with noting that $\gamma \mapsto \langle \mathbf{c}, \gamma\rangle$ is l.s.c. w.r.t. narrow convergence and so is $\mathcal{H}(\gamma) = \iint_{\Omega^2} \mathrm{h}(\gamma)\,\mathrm{d}(x,y)$ by l.s.c. , convexity and superlinearity of h at infinity. Consequently, $\langle \mathbf{c}, \gamma\rangle + \varepsilon\mathcal{H}(\gamma)$ is l.s.c. , too. Furthermore, when Ω is bounded, $\mathcal{H}$ can be bounded from below. Now, $\langle \mathbf{c}, \gamma\rangle$ can be bounded from below, too, $\Pi(\rho,\mu)$ is tight, and consequently the direct method of the calculus of variation can be applied. Finally, noting that $\gamma \mapsto \langle \mathbf{c}, \gamma\rangle + \varepsilon\mathcal{H}(\gamma)$ is strictly convex in γ, since $\langle \mathbf{c}, \gamma\rangle$ is linear and $\mathcal{H}$ is strictly convex, we can conclude uniqueness of the minimizer. ☐

Remark 2.12. *Note that it is, at least formally, just a matter of a simple calculation to see, that for* $\gamma \in \mathscr{P}^{ac}(\Omega^2)$ *with density* G*, i.e.* $\mathrm{d}\gamma = G \cdot \mathrm{d}\mathcal{L}^{2d}$

$$\begin{aligned}\iint_{\Omega^2} \mathbf{c}(x-y)\,\mathrm{d}\gamma(x,y) + \varepsilon\mathcal{H}(\gamma) &= \iint_{\Omega^2} G(x,y)\left(\mathbf{c}(x-y) + \varepsilon\log(G(x,y))\right)\mathrm{d}(x,y)\\ &= \varepsilon\iint_{\Omega^2} G(x,y)\log\left(\frac{G(x,y)}{\mathrm{K}_\varepsilon(x,y)}\right)\mathrm{d}(x,y)\\ &= \varepsilon\mathcal{H}(\gamma \mid \mathrm{K}_\varepsilon)\end{aligned}$$

holds, where

$$\mathrm{K}_\varepsilon(x,y) = \mathrm{e}^{-\frac{\mathbf{c}(x-y)}{\varepsilon}}\ . \tag{2.3.3}$$

With this consideration at hand, it is easy to see an alternative definition of the entropic regularized OT distance as

$$\mathbf{T}^2_{\mathbf{c},\varepsilon}(\rho,\mu) = \inf_{\gamma\in\Pi(\rho,\mu)} \varepsilon\mathcal{H}(\gamma \mid \mathrm{K}_\varepsilon)\ . \tag{2.3.4}$$

Note that we, slightly abusing notation, denoted with K_ε *the measure on* Ω *with density given by* $\mathrm{K}_\varepsilon(x,y)$.

Chapter 3

1D Lagrangian Scheme

3.1 Introduction and preliminary results

This chapter is based on [47], the joint work of the author with Oliver Junge. It is concerned with the convergence of a Lagrangian numerical scheme in one spatial dimension for solution of equation (1.0.1). The scheme itself will consist of a spatio-temporal discretization founded in the formulation of the equation in terms of (pseudo-)inverse (cumulative) distribution functions (a.k.a. quantile function) which we will abbreviate as IDF.

The main result consists of two parts. Both results state that an approximate solution calculated by means of the spatio-temporal discretization will converge to a solution of (1.0.1) for vanishing mesh size. The first convergence result concerns a family of cost functions that resemble p-Wasserstein cost $s \mapsto \frac{1}{p}|s|^p$. The second convergence result will rely on additional assumptions to show that the result then still holds even when flux-limiting cost are considered.

While proving the convergence results we will receive additional properties of this discretization that it shares with the continuous flow, like entropy monotonicity, mass preservation, a minimum/maximum principle and flux-limitation in the case of the corresponding cost.

In the final part of this chapter the discretization is applied to some problems and additionally a numerical convergence analysis is appended.

3.1.1 Restating the problem

Let us restate (1.0.1). We want to find approximate solutions for the non-linear Fokker-Planck equation with no-flux boundary conditions in one spatial dimension

$$\begin{aligned} \partial_t\rho &= \partial_x\left(\rho\cdot\mathcal{A}\left[\partial_x \mathsf{u}'(\rho) + \mathrm{v}' + (\mathrm{w}' * \rho)\right]\right) && \text{in } (0,T)\times[a,b] \\ \rho(0,\cdot) &= \rho_0 && \text{in } [a,b] \\ \partial_x\rho &= 0 && \text{on } [0,T]\times\{a,b\} \end{aligned} \tag{3.1.1}$$

where $\mathcal{A}, \mathsf{u}, \mathrm{v}$ and w are specified now. We begin with the families of cost functions that will lead up to $\mathcal{A}$.

Definition 3.1 (p-Wasserstein like cost)**.** *We will call a cost function* $\mathbf{c}:\mathbb{R}\to[0,\infty)$ p-Wasserstein like *if it is even, strictly convex, with* $\mathbf{c}(0)=0$ *as well as* $\mathbf{c}\in C^1(\mathbb{R})\cap C^3(\mathbb{R}\setminus\{0\})$ *and satisfies the following bounds.*

There are constants $\alpha,\beta>0$ *such that*

$$\alpha\,|s|^p \le \tilde{\mathbf{c}}(s) \le \beta\,|s|^p \tag{3.1.2}$$

holds for every $s \in \mathbb{R}$ and where we abbreviated

$$\tilde{\mathbf{c}}(s) = s\mathbf{c}'(s) . \tag{3.1.3}$$

Definition 3.2 (flux-limiting cost)**.** *We will call a cost function $\mathbf{c} : \mathbb{R} \to [0, \infty]$* flux-limiting *if it is strictly convex and even with $\mathbf{c}(0) = 0$. Furthermore it is assumed to have a proper domain of the form $\mathbf{c}^{-1}(\mathbb{R}) = [-\gamma, \gamma]$ for some $\gamma > 0$ the* flux-limitation *(or „lightspeed"), its derivative is assumed to be diverging at the boundary*

$$\lim_{s \to \pm\gamma} \mathbf{c}'(s) = \pm\infty \tag{3.1.4}$$

*as well as $\mathbf{c} \in C^0([-\gamma, \gamma]) \cap C^3((-\gamma, \gamma))$ and it satisfies the bounds from **Definition 3.1** with $p = 2$, i.e. there are constants $\alpha, \beta > 0$ such that*

$$\alpha |s|^2 \leq \tilde{\mathbf{c}}(s) \leq \beta |s|^2 \tag{3.1.5}$$

holds for every $s \in [-\gamma, \gamma]$.

Before we can specify $\mathcal{A}$, we have to introduce the *Legendre transform.* Though we will only need it as a result for functions on the real line, let us state this definition for functions on $\mathbb{R}^d$.

Definition 3.3. *Let $f : \mathbb{R}^d \to \mathbb{R}_\infty$ be convex. Then the Legendre transformation $f^* : \mathbb{R}^d \to \mathbb{R}$ of f is given as*

$$f^*(y) := \sup_{x \in \mathbb{R}^n} \langle x, y \rangle - f(x) . \tag{3.1.6}$$

Legendre transforms of differentiable convex functions enjoy some useful properties.

Proposition 3.4 (Adapted from Box 1.12. in [46])**.** *Let $f : \mathbb{R}^d \to \mathbb{R}_\infty$ be convex and differentiable. Let ∇f be furthermore a C^1-diffeomorphism. Then $(\nabla f)^{-1} = \nabla(f^*)$ holds and $\nabla(f^*)$ is monotone in the sense that for all $x, y \in \mathrm{dom}(\nabla f)$, the domain of ∇f*

$$\langle \nabla(f^*)(x) - \nabla(f^*)(y), x - y \rangle \geq 0$$

holds.

Now we can go on with the definition of $\mathcal{A}$.

Definition 3.5 (Specifying $\mathcal{A}, \mathsf{u}, \mathrm{v}$ and w)**.** *The maps $\mathcal{A} : \mathbb{R} \to \mathbb{R}$, $\mathsf{u} : [0, \infty) \to [0, \infty)$, $\mathrm{v} : \mathbb{R} \to \mathbb{R}$ and $\mathrm{w} : \mathbb{R} \to \mathbb{R}$ are assumed to have the following properties.*

1. *There is a cost function $\mathbf{c} : \mathbb{R} \to [0, \infty]$ satisfying either* Definition 3.1 *or* 3.2 *such that*

$$\mathcal{A} = (\mathbf{c}^*)' . \tag{3.1.7}$$

2. u *is defined as (c.f. 1.0.5)*

$$\mathsf{u}(s) = \begin{cases} s \log(s) & \text{if } m = 1 \\ \frac{1}{m-1} s^m & \text{if } m > 1. \end{cases} \tag{3.1.8}$$

3. $\mathrm{v} \in C^3(\mathbb{R})$ *is convex.*

4. $\mathrm{w} \in C^3(\mathbb{R})$ *is convex and even.*

Assumption 3.6. *The combination of $\mathbf{c}$ and the two potentials v and w are subject to the following restrictions.*

- *If* v *and* w *are not both constants,* $\mathbf{c}''$ *has to be bounded away from zero i.e. there is* $\underline{\mathbf{c}''} > 0$ *such that*

$$\mathbf{c}''(s) \geq \underline{\mathbf{c}''} \qquad \text{for } s \neq 0 \, . \tag{3.1.9}$$

This restriction is due to $(\mathbf{c}^*)'$ *being no longer Lipschitz in general, if this prerequisite is not met. The problem* (3.1.1) *might not be well-posed as soon as there is a potential (external or interaction) involved.*

- *If* $\mathbf{c}$ *is flux-limiting,* $\partial_x \rho_0$*, the weak spatial derivative of our initial data, has to be bounded.*

Examples 3.7. *We present three PDEs arising from choosing certain parameters and potentials.*

- *Let* $q \in \mathbb{N}$ *with* $q > 1$*. Then choose* $p \in (1,2)$ *such that* $\frac{2-p}{p-1} = q$ *and pick* $m = 3 - p > 1$*. For any constant external & interaction potential, the equation becomes the q-Laplace equation already stated in* (1.0.7)

$$\partial_t \rho = \partial_x (|\partial_x \rho|^q \cdot \partial_x \rho) \, . \tag{3.1.10}$$

- *Let* $p = 2$ *and* $m > 1$*. Then the equation is a Fokker-Plank equation*

$$\partial_t \rho = \frac{1}{m} \partial_x^2 \rho^m + \partial_x(\rho \mathrm{v}') + \partial_x(\rho(\mathrm{w}' * \rho)) \, . \tag{3.1.11}$$

- *Let*

$$\mathbf{c}(s) = \begin{cases} \gamma \left(1 - \sqrt{1 - \left|\frac{s}{\gamma}\right|^2}\right) & \text{for } |s| \leq \gamma \\ +\infty & \text{elsewhere.} \end{cases} \tag{3.1.12}$$

Then we recover, for $m = 1$*;* $\mathrm{v}, \mathrm{w} = 0$ *Rosenau's relativistic heat equation already introduced in* (1.0.8) *(c.f. [44])*

$$\partial_t \rho = \partial_x \left(\rho \frac{\partial_x \rho}{\sqrt{\rho^2 + |\partial_x \rho|^2}} \right) . \tag{3.1.13}$$

Remark 3.8. *The equation* (3.1.1) *can be written as a transport equation*

$$\partial_t \rho = \mathrm{div}\,(\rho \cdot \mathfrak{V}[\rho]) \tag{3.1.14}$$

where the velocity $\mathfrak{V}[\rho] = \mathcal{A}\left[\partial_x \mathsf{u}'(\rho) + \mathrm{v}' + (\mathrm{w}' * \rho)\right]$ *consists of a non-linear diffusion term* u*, an external potential* v *and an interaction potential* w*.*

3.1.2 Variational Formulation and Discretization

We will see that solutions to our PDE can be approximated by minimizing movement sequences in the energy landscape of $\mathcal{E}$ where, for $\rho \in \mathscr{P}^{ac}(\Omega)$ it is defined as

$$\mathcal{E}(\rho) = \mathcal{U}(\rho) + \mathcal{V}(\rho) + \mathcal{W}(\rho) \tag{3.1.15}$$

and where $\mathcal{E}$ is taken to be $+\infty$ otherwise. Here the internal, external and interaction energies, $\mathcal{U}, \mathcal{V}$ and $\mathcal{W}$, are defined as

$$\mathcal{U}(\rho) := \int_{[a,b]} \mathsf{u}(\rho(x))\,\mathrm{d}x; \quad \mathcal{V}(\rho) := \int_{[a,b]} \mathrm{v}(x)\rho(x)\,\mathrm{d}x \quad \mathcal{W}(\rho) := \iint_{[a,b]^2} \rho(x)\mathrm{w}(x-y)\rho(y)\,\mathrm{d}(x,y) \, . \tag{3.1.16}$$

with respect to the optimal transport distance (c.f. *Proposition 2.9*).

$$\mathbf{T}_{\mathbf{c},\tau}(\rho,\mu) = \inf_{\gamma\in\Pi(\rho,\mu)} \int_{[a,b]^2} \mathbf{c}(x-y)\,\mathrm{d}\gamma(x,y) \tag{3.1.17}$$

Now rewriting these functionals w.r.t. the quantile function X_ρ we arrive at the energy expressed in terms of IDF which reads as follows:

$$\mathcal{E}(\rho) = \int\limits_{[0,1]} \mathsf{u}_*(\partial_\xi X_\rho)\,\mathrm{d}\xi + \int\limits_{[0,1]} \mathrm{v}(X_\rho)\,\mathrm{d}\xi + \iint\limits_{[0,1]^2} \mathrm{w}(X_\rho(\xi) - X_\rho(\zeta))\,\mathrm{d}(\xi,\zeta)\;. \tag{3.1.18}$$

Here $\mathsf{u}_*(s) = s\mathsf{u}(\frac{1}{s})$.

Concerning the optimal transport distance (c.f. *Proposition 2.9*)

$$\mathbf{T}_{\mathbf{c},\tau}(\rho,\mu) = \int_{[0,1]} \mathbf{c}(X_\rho(\xi) - X_\mu(\xi))\,\mathrm{d}\xi\;. \tag{3.1.19}$$

Finally we discretize the X piecewise constant on a equidistant mesh $\{0 = \xi_0, \dots, \xi_k = 1\} \subset [0,1]$ of strictly increasing values $\xi_i = \frac{i}{k}$ receiving strictly increasing vectors $\mathbf{x} = (a = x_0, x_1, \dots, x_k = b)$ representing a discrete X in the following way

$$\mathbf{x} \mapsto X(\xi) = \sum_{i=0}^{k-1} x_i^{(n)} \mathbf{1}_{(\xi_i,\xi_{i+1})}(\xi) \tag{3.1.20}$$

where $\mathbf{x}$ has to lie in the set $\mathcal{X}_k([a,b])$ defined as

$$\mathcal{X}_k([a,b]) := \{\mathbf{x} = (x_0, x_1, \dots, x_k) \in \mathbb{R}^{k+1} \mid a = x_0 < x_1 < \dots < x_k = b\}\;. \tag{3.1.21}$$

furthermore we also represent the derivative of X by a piecewise constant function

$$\mathbf{x} \mapsto \delta_\xi X(\xi) = \sum_{i=0}^{k-1} \delta x_i^{(n)} \mathbf{1}_{(\xi_i,\xi_{i+1})}(\xi)\;.$$

Applying this discretization to our functionals we arrive at the discrete minimizing movement scheme, which is our fully discrete scheme.

$$\text{minimize } \mathbf{x} \mapsto \Phi(\tau; \mathbf{x}^{(n-1)}, \mathbf{x}) \tag{3.1.22}$$

where

$$\begin{aligned}\Phi(\tau;\mathbf{x}^{(n-1)},\mathbf{x}) &:= \tau\mathbf{T}^{\mathbf{x}}_{\mathbf{c},\tau}(\mathbf{x}^{(n)},\mathbf{x}^{(n-1)}) + \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(n)}) \\ &= \tau\mathbf{T}^{\mathbf{x}}_{\mathbf{c},\tau}(\mathbf{x}^{(n)},\mathbf{x}^{(n-1)}) + \mathcal{U}^{\mathbf{x}}(\rho) + \mathcal{V}^{\mathbf{x}}(\rho) + \mathcal{W}^{\mathbf{x}}(\rho)\end{aligned}$$

with

$$\mathcal{U}^{\mathbf{x}}(\mathbf{x}^{(n)}) := h_k \sum_{i=0}^{k-1} \mathsf{u}_*(\delta x_{i-1}^{(n)});\quad \mathcal{V}^{\mathbf{x}}(\mathbf{x}^{(n)}) := h_k \sum_{i=0}^{k} \mathrm{v}(x_i^{(n)});\quad \mathcal{W}^{\mathbf{x}}(\mathbf{x}^{(n)}) := h_k \sum_{i=0}^{k} h_k \sum_{j=0}^{k} \mathrm{w}(x_i^{(n)} - x_j^{(n)})$$

and

$$\mathbf{T}^{\mathbf{x}}_{\mathbf{c},\tau}(\mathbf{x}^{(n)},\mathbf{x}^{(n-1)}) := h_k \sum_{i=0}^{k} \mathbf{c}\left(\frac{x_i^{(n)} - x_i^{(n-1)}}{\tau}\right)\;.$$

We note the following result concerning $\mathcal{V}^{\mathbf{x}}$ and $\mathcal{W}^{\mathbf{x}}$.

Remark 3.9. *The convexity assumption of* v *and* w *imply the the following convexity for the potentials.*

$$\langle \nabla \mathcal{F}(\mathbf{x}), \mathbf{x} - \mathbf{y} \rangle \leq \mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{y})$$

for $\mathcal{F}$ *either* $\mathcal{V}^{\mathbf{x}}$ *or* $\mathcal{W}^{\mathbf{x}}$ *specified in* (3.1.18).

Remark 3.10. *A short remark on the choice of piecewise constant spatial discretization opposed to piecewise affine discretization is advisable. Usually the IDF* X *is discretized in a piecewise affine way. This way the corresponding densities are piecewise constant on* (x_i, x_{i+1}) *with the same mass on each of these intervals. Anticipating the Euler-Lagrange equation, we see that the r.h.s. and the l.h.s. would not fit together by non-linearity of* $(\mathbf{c}^*)'$ *when considering the discretized IDF. One way to circumvent this problem is to consider piecewise constant discretized IDF, which is exactly what we do.*

To simplify the expressions appearing in the following claims and proofs, we abbreviate the finite forward difference quotient of our vectors $\mathbf{x}^{(n)} = (x_0^{(n)}, \ldots, x_k^{(n)}) \in \mathcal{X}_k([a, b])$ as

$$\delta \mathbf{x}^{(n)} = (\delta x_0^{(n)}, \ldots, \delta x_{k-1}^{(n)}) \quad \text{where} \quad \delta x_i^{(n)} = \frac{x_{i+1}^{(n)} - x_i^{(n)}}{h_k} \quad \text{for} \quad i = 0, \ldots, k-1 \,.$$

Furthermore we now define the second symmetric difference quotient as

$$\delta^2 \mathbf{x}^{(n)} = (\delta^2 x_1^{(n)}, \ldots, \delta^2 x_{k-1}^{(n)}) \quad \text{where} \quad \delta^2 x_i^{(n)} := \frac{x_{i+1}^{(n)} - 2x_i^{(n)} + x_{i-1}^{(n)}}{h_k^2} \quad \text{for } i = 1, \ldots, k-1 \,.$$

This way $\delta[\delta x_{i-1}^{(n)}] = \delta^2 x_i^{(n)}$ holds.

We state a result allowing the approximation of the weak derivative of a function with the just introduced finite differences. It is adapted from [20, Chapter 5.8.2].

Proposition 3.11 (Finite differences and the weak derivative)**.** *Let* $\delta_z = \delta_{z,k}$ *be the finite forward difference operator in direction of space* $z = \xi$ *or time* $z = t$ *w.r.t. to the step size in space* h_k *or time* τ_k *respectively.*

Let furthermore $f_k : (0, T) \times (0, 1) \to \mathbb{R}$ *be a sequence in* $L^p((0, T) \times (0, 1))$ *with* $p \in (1, \infty)$ *such that* $f_k \rightharpoonup f_*$ *in* $L^p((0, T) \times (0, 1))$ *and* $\delta_z f_k$ *is bounded w.r.t.* $\|\cdot\|_{L^p((0,T)\times(0,1))}$*. Then* $\delta_z f_k \rightharpoonup \partial_z f_* \in L^p((0, T) \times (0, 1))$*.*

If $\delta_x f_k \rightharpoonup \partial_x f_*$ *and* $\delta_t f_k \rightharpoonup \partial_t f_*$ *hold, then* $f_k \to f_*$ *strongly w.r.t.* $L^p((0, T) \times (0, 1))$*.*

Proof. We will prove the claim for $z = \xi$. The proof for $z = t$ is virtually the same.

By Banach-Alaoglu theorem and reflexivity, for each subsequence of $\delta_{\xi,k} f_k$ there is a subsubsequence converging weakly to some g_* in $L^p((0, T) \times (0, 1))$. Now let $\varepsilon > 0$. Then for said (unrelabelled) subsubsequence and all $\phi \in C_c^\infty((0, T) \times (0, 1))$ with $\operatorname{supp} \phi \subset [0, T] \times [\varepsilon, 1 - \varepsilon]$ we can calculate

$$\begin{aligned} \int_\Omega g_* \phi &= \lim_{k \to \infty} \int_\Omega \delta_{\xi,k} f_k \phi \\ &= - \lim_{k \to \infty} \int_\Omega f_k \delta_{\xi,k}^- \phi \\ &= - \int_\Omega f_* \partial_\xi \phi \end{aligned}$$

where we denoted the backwards finite difference operator as $\delta_{\xi,k}^-$, used that g_* is the weak limit of $\delta_{\xi,k} f_k$, integration by parts for the finite difference operator and assuming $h_k < \varepsilon$ to be able to neglect the boundary-terms of the integration by parts.

This shows that $g_* = \partial_\xi f_*$ on $[0,T] \times (\varepsilon, 1-\varepsilon)$ holds and by uniqueness of the weak derivative, the original sequence $\delta_{\xi,k} f_k$ converges to $\partial_\xi f_*$ on $[0,T] \times (\varepsilon, 1-\varepsilon)$. Since $\varepsilon > 0$ was arbitrary, we can conclude that these results hold on $(0,T) \times (0,1)$.

Finally, if $\delta_\xi f_k \rightharpoonup \partial_\xi f_*$ and $\delta_t f_k \rightharpoonup \partial_t f_*$ hold, as well as the assumed $f_k \rightharpoonup f_*$, then f_k converges weakly in $W^{1,p}((0,T) \times (0,1))$ to f_* and the Rellich–Kondrachov theorem implies the strong convergence we stated above. □

3.1.3 Main Theorem

The main result concerning the sequence X_k approximating the IDF is given in the following theorem. It states that the limit of our sequence X_k is a weak solution to (3.1.1) rewritten in terms of inverse distribution functions X:

$$\begin{aligned} \partial_t X(t,\xi) &= (\mathbf{c}^*)' \left[\partial_\xi \mathrm{u}_*'(\partial_\xi X(t,\xi)) + \mathrm{v}'(X(t,\xi)) + \int_0^1 \mathrm{w}'(X(t,\xi) - X(t,\zeta))\,\mathrm{d}\zeta \right] \text{ on } (0,T) \times (0,1) \\ X(0,\xi) &= X^{(0)} \quad \text{on} \quad (0,T) \end{aligned} \tag{3.1.23}$$

where we additionally assume that $X(t,\cdot)$ is a IDF a.e. on $[0,1]$.

Theorem 1. *Let $k \in \mathbb{N}$, $\tau_k > 0$ be a sequence monotonically converging to zero and $\mathbf{c}$ a cost function satisfying* Definition 3.1. *Let $\mathbf{x}_k^{(0)}$ be a sequence in k with*

1. *$\mathbf{x}_k^{(0)} \in \mathcal{X}_k(I)$ (where $\mathcal{X}_k(I)$ is defined in* (3.1.21)*);*
2. *The piecewise affine interpolation of $\mathbf{x}_k^{(0)}$ as defined in* (3.1.20) *converges pointwise to an IDF $X^{(0)}$;*
3. *The energy $\mathcal{E}^{\mathbf{x}}(\mathbf{x}_k^{(0)})$ is uniformly bounded in k;*
4. *There are upper and lower bounds $0 < \underline{\delta \mathbf{x}^{(0)}} < \overline{\delta \mathbf{x}^{(0)}}$ such that $\underline{\delta \mathbf{x}^{(0)}} < \delta \mathbf{x}_k^{(0)} < \overline{\delta \mathbf{x}^{(0)}}$ holds for all k.*

Let finally the sequence $X_k : [0,T] \times [0,1] \to \mathbb{R}$ be the piecewise constant interpolation of the $\mathbf{x}_k^{(n)}$ created with initial data $\mathbf{x}_k^{(0)}$ and the recursion (3.1.22).

Then the sequence X_k has the following properties. There is an unrelabled subsequence such that X_k converges in $L^q([0,T] \times [0,1])$ to $X_* \in W^{1,\infty}((0,T) \times (0,1)) \cap C^{0,\alpha}(0,T; L^1([0,1]))$ for every $q \in [1,\infty)$, where $\alpha = 1/p$, $\partial_\xi X_* \in L^1(0,T; W^{1,p'}((0,1)))$ and $p' = \frac{1}{1-1/p}$ denotes the Hölder conjugate. The limit X_* solves the following weak formulation of (3.1.23)*:*

$$\begin{aligned} &-\int_0^T \int_0^1 X_*(t,\xi) \partial_t \varphi(t,\xi)\,\mathrm{d}\xi\,\mathrm{d}t \\ &\quad = \int_0^T \int_0^1 (\mathbf{c}^*)' \left(\partial_\xi \mathrm{u}_*'(\partial_\xi X_*(t,\xi)) - \mathrm{v}'(X_*(t,\xi)) - \int_0^1 \mathrm{w}'(X_*(t,\xi) - X_*(t,\zeta))\,\mathrm{d}\zeta \right) \varphi(t,\xi)\,\mathrm{d}\xi\,\mathrm{d}t \end{aligned}$$

holds for all $\varphi \in C_c^\infty((0,T) \times (0,1))$. Additionally the initial data are assumed continuously $\lim_{t \searrow 0} X_*(t) = X^{(0)}$ *in $L^p([0,1])$ and $X_*(t,\cdot)$ is an IDF for a.e. $t \in [0,1]$.*

As already mentioned, a similar theorem can be stated for the family of flux-limiting cost.

Theorem 2. *Let the prerequisites of Theorem 1 hold, except $\mathbf{c}$ does this time satisfy* Definition 3.2 *instead of* 3.1*. Additionally we assume finite bounds*

$$-\infty < \underline{\delta^2 \mathbf{x}^{(0)}} < \overline{\delta^2 \mathbf{x}^{(0)}} < \infty \quad \text{such that} \quad \underline{\delta^2 \mathbf{x}^{(0)}} < \delta^2 \mathbf{x}_k^{(0)} < \overline{\delta^2 \mathbf{x}^{(0)}} \quad \text{holds for all } k$$

Then convergence to a weak solution of (3.1.23) *holds in the sense of Theorem 1 with $p = 2$ and the initial data are again assumed continuously.*

3.2 Variational properties of the minimization problem

In this section we want to lay the foundation for the successive section, which will show the crucial convergences that are needed for the main theorem. First of all, we will show that the sequences $\mathbf{x}^{(n)}$ are indeed well defined. This is followed by the Euler-Lagrange equation for the minimization problem .

Finally we will show two estimates. The first estimate is a maximum-/minimum principle for the forward differences $\frac{x_{i+1}^{(n)}-x_i^{(n)}}{h_k}$ giving us control over $\mathsf{u}_*\left(\frac{x_{i+1}^{(n)}-x_i^{(n)}}{h_k}\right)$ and $\mathsf{u}'_*\left(\frac{x_{i+1}^{(n)}-x_i^{(n)}}{h_k}\right)$. This bound itself corresponds to bounding the probability density ρ itself from above and away from zero.

The second estimate will be a second comparison principle bounding the second order central differences $\delta^2 x_i^{(n)} = \frac{x_{i+1}^{(n)}-2x_i^{(n)}+x_{i-1}^{(n)}}{h_k}$ which in turn corresponds to a bound for the spatial derivative of ρ.

The second one is derived by a descent-in-energy argument in combination with displacement convexity of our internal energy term $\mathcal{U}$ which corresponds in our IDF terms to convexity of $\mathcal{U}^{\mathbf{x}}$. This estimate yields in combination with properties of v and w a bound for the r.h.s. of the Euler-Lagrange equation and will guarantee convergence of the corresponding sequence of functions to some limit.

3.2.1 Existence and uniqueness

Lemma 3.12. *The minimization problem has a unique minimizer*

$$\mathbf{x} = (x_0, \dots, x_k) \in \mathcal{X}_k([a,b]) .$$

Proof. The set $\mathcal{X}_k([a,b])$ is bounded in $\mathbb{R}^{k+1}$ and $\Phi(\tau;\mathbf{x}^{(n-1)},\mathbf{x})$ is continuous in $\mathcal{X}([a,b])$ and lower semi-continuous in $\overline{\mathcal{X}([a,b])}$. Furthermore $\overline{\mathcal{X}([a,b])}$ is compact, so there exists a minimizer $\mathbf{x}_{min} = (\underline{x_0}, \dots, \underline{x_k})$. $\mathbf{x} = \mathbf{x}^{(n-1)}$ generates a finite value in Φ, so at each minimizer $\Phi(\tau;\mathbf{x}^{(n-1)},\cdot)$ is finite. Additionally, the functional is strictly convex, since all summands are and $\tau > 0$, so said minimizer is unique.

Note that u_* is monotonously decreasing and furthermore

$$\inf_{\mathbf{x}\in\mathcal{X}_k([a,b])} \mathcal{V}^{\mathbf{x}}(\mathbf{x}) =: \underline{\mathcal{V}^{\mathbf{x}}} > -\infty$$

as well as

$$\inf_{\mathbf{x}\in\mathcal{X}_k([a,b])} \mathcal{W}^{\mathbf{x}}(\mathbf{x}) =: \underline{\mathcal{W}^{\mathbf{x}}} > -\infty$$

holds by properties of $\mathcal{X}_k([a,b])$, v and w.

This way we have for $j = 0, \dots, k-1$:

$$\begin{aligned} C &\geq \Phi(\tau;\mathbf{x}^{(n-1)},\mathbf{x}_{min}) \\ &\geq \tau \mathbf{T}^{\mathbf{x}}_{\mathbf{c},\tau}(\mathbf{x}_{min},\mathbf{x}^{(n-1)}) + \mathcal{E}^{\mathbf{x}}(\mathbf{x}_{min}) \\ &\geq h_k \sum_{i=0}^{k-1} \mathsf{u}_*\left(\frac{x_{i+1}^{(n)}-x_i^{(n)}}{h_k}\right) + \underline{\mathcal{V}^{\mathbf{x}}} + \underline{\mathcal{W}^{\mathbf{x}}} \\ &\geq h_k \mathsf{u}_*\left(\frac{x_{i+1}^{(n)}-x_i^{(n)}}{h_k}\right) + h_k(k-1)\mathsf{u}_*\left(\frac{b-a}{h_k}\right) + \underline{\mathcal{V}^{\mathbf{x}}} + \underline{\mathcal{W}^{\mathbf{x}}} \end{aligned} \tag{3.2.1}$$

This yields a bound from above for $\frac{x_{i+1}^{(n)}-x_i^{(n)}}{h_k}$ for every i since $\mathsf{u}_*(s)$ diverges to $+\infty$ for $s \searrow 0$. □

Now that we know that the minimization is well defined, we can establish the following result, which is the standard a priori estimate based on $\mathbf{x}^{(n)}$ being a descending sequence in the energy landscape of $\mathcal{E}^{\mathbf{x}}$ showing that our sequence is indeed monotonously descending in $\mathcal{E}^{\mathbf{x}}$.

Corollary 3.13. *Let $\mathbf{x}^{(n-1)} \in \mathcal{X}_k([a,b])$ and $\mathbf{x}^{(n)}$ the minimizer of our minimization problem. Then*

$$\mathbf{T}^{\mathbf{x}}_{\mathbf{c},\tau}(\mathbf{x}^{(n)},\mathbf{x}^{(n-1)}) \leq \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(n-1)}) - \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(n)}) \tag{3.2.2}$$

holds.

Proof. Plugging in the feasible $\mathbf{x} = \mathbf{x}^{(n-1)}$ in the minimization problem combining it with $\mathbf{x}^{(n)}$ being a minimizer and with $\mathbf{T}^{\mathbf{x}}_{\mathbf{c},\tau}(\mathbf{x}^{(n-1)},\mathbf{x}^{(n-1)}) = 0$ yields the result after rearranging. □

This result in particular shows, that our sequence $\mathbf{x}^{(n)}$ is monotonously descending in the energy landscape of $\mathcal{E}$.

3.2.2 The Euler-Lagrange equation

The Euler-Lagrange equation of our minimization problem will form a substantial part of the foundation for our calculations. In the case of p-Wasserstein cost, the proof is straight forward. In the case of flux-limiting cost, however, we have to make sure the minimization problem in some sense does not see the discontinuity. Unfortunately this cannot be done a posteriori with the second maximum-/minimum principle since the proof of the principle does rely on the Euler-Lagrange equation.

Lemma 3.14. *In the case of flux-limiting cost, the minimizer $\mathbf{x}^{(n)}$ lies in*

$$\left\{\mathbf{x}\middle|\mathbf{x} \in \mathcal{X}_k([a,b]), \left|x_j - x_j^{(n-1)}\right| < \gamma\tau \text{ for } j = 0,\dots,k\right\} .$$

Proof. Let $P := \left\{i \in \{0,\dots,k\} \mid \left|x_j^{(n)} - x_j^{(n-1)}\right| = \gamma\tau\right\}$. We can omit the cases $\left|x_j^{(n)} - x_j^{(n-1)}\right| > \gamma\tau$, since they don't occur at a minimizer. Let $P \subset \{0,1,\dots,k\}$ and define a partial convex combination of $\mathbf{x}^{(n)}$ and $\mathbf{x}^{(n-1)}$ w.r.t. P as

$$\mathbf{x}^{\varepsilon} := \begin{cases} (1-\varepsilon)x_i^{(n)} + \varepsilon x_i^{(n-1)} & \text{if } i \in P \\ x_i^{(n)} & \text{if } i \notin P . \end{cases}$$

We will now show that if $P \neq \emptyset$, then for a suitable ε the partial convex combination $\mathbf{x}^{\varepsilon}$ is a feasible candidate with $\Phi(\tau,\mathbf{x}^{(n-1)},\mathbf{x}^{\varepsilon}) < \Phi(\tau,\mathbf{x}^{(n-1)},\mathbf{x}^{(n)})$ contradicting $\mathbf{x}^{(n)}$ being a minimizer in the first place.

We note that $\mathbf{x}^{\varepsilon} \in \mathcal{X}_k([a,b])$ for ε small enough, since $\mathbf{x}^{(n)} \in \mathcal{X}_k([a,b])$ and $\mathcal{X}_k([a,b])$ is open w.r.t. $\{a\} \times \mathbb{R}^{k-1} \times \{b\}$.

Furthermore recall the behaviour of $\mathbf{c}'$ at the boundary (c.f. *Definition 3.2*).

Define two index sets as follows.

$$P^+ := \{i \in \{0,\dots,k\} \mid i \notin P, i+1 \in P\} \quad \text{and} \quad P^- := \{i \in \{0,\dots,k\} \mid i \in P, i+1 \notin P\} .$$

We will consider $\Phi(\tau;\mathbf{x}^{(n-1)},\mathbf{x}^{(n)}) - \Phi(\tau;\mathbf{x}^{(n-1)},\mathbf{x}^{\varepsilon})$ and we can see that, since $\mathbf{x}^{\varepsilon} \to \mathbf{x}^{(n)}$ for $\varepsilon \searrow 0$ and since v and w are both continuous, that the corresponding expressions are bounded.

Recall $\mathbf{c}$ and u_* are strictly convex functions in $C^1(I)$, which allows us to calculate now only for the remaining expressions

$$
\begin{aligned}
&h_k \sum_{i=0}^{k} \mathbf{c}\left(\frac{1}{\tau}(x_i^{(n)} - x_i^{(n-1)})\right) - \mathbf{c}\left(\frac{1}{\tau}(x_i^{\varepsilon} - x_i^{(n-1)})\right) \\
&+ h_k \sum_{i=0}^{k-1} \mathsf{u}_*\left(\frac{x_{i+1}^{(n)} - x_i^{(n)}}{h_k}\right) - \mathsf{u}_*\left(\frac{x_{i+1}^{\varepsilon} - x_i^{\varepsilon}}{h_k}\right) \\
&> \varepsilon h_k \left(\sum_{i\in P} \mathbf{c}'\left(\frac{1}{\tau}(x_i^{\varepsilon} - x_i^{(n-1)})\right) \frac{1}{\tau}\left(x_i^{(n)} - x_i^{(n-i)}\right) + h_k \sum_{i\in P\setminus P^-} \mathsf{u}_*\left(\frac{x_{i+1}^{\varepsilon} - x_i^{\varepsilon}}{h_k}\right)\left(\delta x_i^{(n)} - \delta x_{i-1}^{(n)}\right) \right. \\
&\left. + h_k \sum_{i\in P^-} \mathsf{u}_*\left(\frac{x_{i+1}^{(n)} - x_i^{\varepsilon}}{h_k}\right) k\left(x_i^{(n-1)} - x_i^{(n)}\right) + h_k \sum_{i\in P^+} \mathsf{u}_*\left(\frac{x_{i+1}^{\varepsilon} - x_i^{(n)}}{h_k}\right) k\left(x_i^{(n)} - x_i^{(n-1)}\right) \right) .
\end{aligned}
$$

By continuity, the expressions $\mathsf{u}_*\left(\frac{x_{i+1}^{\varepsilon} - x_i^{\varepsilon}}{h_k}\right)$, $\mathsf{u}_*\left(\frac{x_{i+1}^{(n)} - x_i^{\varepsilon}}{h_k}\right)$ $\mathsf{u}_*\left(\frac{x_{i+1}^{\varepsilon} - x_i^{(n)}}{h_k}\right)$ and $\mathsf{u}_*\left(\frac{x_{i+1}^{\varepsilon} - x_i^{(n)}}{h_k}\right)$ are bounded in ε for $\varepsilon \searrow 0$. On the other hand,

$$
\mathbf{c}'\left(\frac{1}{\tau}(x_i^{\varepsilon} - x_i^{(n-1)})\right)\left(x_i^{(n)} - x_i^{(n-1)}\right) \to \infty \quad \text{for } \varepsilon \searrow 0 .
$$

Note that by continuity of v and w, the differences $\mathcal{V}^{\mathbf{x}}(\mathbf{x}^{(n)}) - \mathcal{V}^{\mathbf{x}}(\mathbf{x}^{\varepsilon})$ and $\mathcal{W}^{\mathbf{x}}(\mathbf{x}^{(n)}) - \mathcal{W}^{\mathbf{x}}(\mathbf{x}^{\varepsilon})$ are bounded too.

So we know that, for ε small enough, $\Phi(\tau; \mathbf{x}^{(n-1)}, \mathbf{x}^{(n)}) - \Phi(\tau; \mathbf{x}^{(n-1)}, \mathbf{x}^{\varepsilon})$ will be positive which is our sought for contradiction. □

Lemma 3.15 (Euler-Lagrange equation)**.** *Let $\overline{M}, \underline{M} \in (0,\infty)$ and $\mathbf{x}^{(n-1)} \in \mathcal{X}_k([a,b])$. Let $\mathbf{x}^{(n)}$ be the minimizer of our minimization problem. Then it satisfies the system of equations*

$$
\frac{x_i^{(n)} - x_i^{(n-1)}}{\tau_k} = (\mathbf{c}^*)' \left[\left(\nabla \mathcal{U}^{\mathbf{x}}(\mathbf{x}^{(n)})\right)_i - \left(\nabla \mathcal{V}^{\mathbf{x}}(\mathbf{x}^{(n)})\right)_i - \left(\nabla \mathcal{W}^{\mathbf{x}}(\mathbf{x}^{(n)})\right)_i \right] \tag{3.2.3}
$$

where

$$
\begin{aligned}
\left(\nabla \mathcal{U}^{\mathbf{x}}(\mathbf{x}^{(n)})\right)_i &= \frac{\mathsf{u}_*'\left(\frac{x_{i+1}^{(n)} - x_i^{(n)}}{h_k}\right) - \mathsf{u}_*'\left(\frac{x_i^{(n)} - x_{i-1}^{(n)}}{h_k}\right)}{h_k} \\
\left(\nabla \mathcal{V}^{\mathbf{x}}(\mathbf{x}^{(n)})\right)_i &= \mathrm{v}'(x_i^{(n)}) \\
\left(\nabla \mathcal{W}^{\mathbf{x}}(\mathbf{x}^{(n)})\right)_i &= 2h_k \sum_{j=0}^{k} \mathrm{w}'(x_i^{(n)} - x_j^{(n)})
\end{aligned}
$$

for each $i = 1, \ldots, k-1$. Note that the above equation reduces to $\frac{x_i^{(n)} - x_i^{(n-1)}}{\tau_k} = 0$ for $i \in \{0, k\}$.

Remark 3.16. *We will abbreviate the argument of $(\mathbf{c}^*)'$ above:*

$$
a_{i,k}^{(n)} := \frac{\mathsf{u}_*'\left(\delta x_i^{(n)}\right) - \mathsf{u}_*'\left(\delta x_{i-1}^{(n)}\right)}{h_k} - \mathrm{v}'(x_i^{(n)}) - 2h_k \sum_{j=0}^{k} \mathrm{w}'(x_i^{(n)} - x_j^{(n)}) . \tag{3.2.4}
$$

Proof of Lemma 3.15. The functional $\Phi(\tau_k; \mathbf{x}^{(n-1)}, \cdot)$ is continuously differentiable on $\mathcal{X}_k([a,b]) = \{\mathbf{x} \mid a = x_0 < x_1 < \cdots < x_k = b\}$ with gradient

$$\begin{aligned}
&\partial_{x_j} \Phi(\tau; \mathbf{x}^{(n-1)}, \mathbf{x}) \\
&\quad = h_k \mathbf{c}' \left(\frac{1}{\tau}(x_j - x_{j-1}^{(n)}) \right) + \mathsf{u}_*'(\delta x_{j-1}) - \mathsf{u}_*'(\delta x_j) + h_k \mathrm{v}'(x_j) \\
&\quad + h_k^2 \sum_{j=0}^{k} \mathrm{w}'(x_i^{(n)} - x_j^{(n)}) - \mathrm{w}'(x_j^{(n)} - x_i^{(n)}) \, .
\end{aligned} \tag{3.2.5}$$

At our minimizer $\mathbf{x} = \mathbf{x}^{(n)}$ this has to be equal to zero, and rearranging the resulting equation yields (3.2.3), using that the Legendre transform satisfies $(\mathbf{c}^*)' = (\mathbf{c}')^{-1}$.

Note that we use that w is an even function, making w' an odd function which allows us to combine $\mathrm{w}'(x_i^{(n)} - x_j^{(n)}) - \mathrm{w}'(x_j^{(n)} - x_i^{(n)}) = 2\mathrm{w}'(x_i^{(n)} - x_j^{(n)})$ to achieve the final result. □

3.2.3 The discrete minimum/maximum principle

The minimum/maximum principles shown in this section hold in a more general way than we need. Indeed we know that v'' and w'' are non-negative since v and w are assumed to be convex. But for this section we drop this assumption and show the minimum/maximum principles only for the assumption that $\mathrm{v}, \mathrm{w} \in C^2(\mathbb{R})$.

We will prove the first discrete minimum/maximum principle next. It bounds the forward difference quotient of $\mathbf{x}_k^{(n)}$ uniformly from above and away from zero, if we have initial data as described in *Theorem 1 (4)*. This initial condition corresponds to ρ_0 being bound from above and away from zero.

Lemma 3.17. *Let* $\mathbf{x}^{(n-1)} \in \mathcal{X}_k([a,b])$. *Let* $\mathbf{x}^{(n)}$ *be the minimizer of* $\Phi(\tau; \mathbf{x}^{(n-1)}, \mathbf{x})$.

1. The first maximum-/minimum principle.
   The following inequality chain holds for every $j \in \{0, \ldots, k-1\}$

$$\mathrm{e}^{-\underline{\kappa}_1 \tau} \min_i \delta x_i^{(n-1)} \leq \delta x_j^{(n)} \leq \mathrm{e}^{\overline{\kappa}_1 \tau} \max_i \delta x_i^{(n-1)}$$

   where

$$\overline{\kappa}_1 := \begin{cases} 0 & \text{if } \mathrm{v}'' + \mathrm{w}'' \geq 0 \\ \underline{(\mathbf{c}^*)''}(\underline{\mathrm{v}''} + \underline{\mathrm{w}''}) & \text{if } \mathrm{v}'' + \mathrm{w}'' < 0 \, . \end{cases}$$

$$\underline{\kappa}_1 := \begin{cases} 0 & \text{if } \mathrm{v}'' + \mathrm{w}'' \leq 0 \\ \underline{(\mathbf{c}^*)''}(\overline{\mathrm{v}''} + \overline{\mathrm{w}''}) & \text{if } \mathrm{v}'' + \mathrm{w}'' > 0 \, . \end{cases}$$

2. The second maximum-/minimum principle.
   Let

$$\begin{aligned}
\overline{\kappa}_2 &:= \overline{(\mathbf{c}^*)''} \cdot (\underline{\mathrm{v}''} + \underline{\mathrm{w}''}) \\
\underline{\kappa}_2 &:= \underline{(\mathbf{c}^*)''} \cdot (\overline{\mathrm{v}''} + \overline{\mathrm{w}''}) \\
\overline{\kappa}_3 &:= (\underline{[(\mathbf{c}^*)''\mathrm{v}'']'} + \underline{[(\mathbf{c}^*)''\mathrm{w}'']'}\ ) \\
\underline{\kappa}_3 &:= (\overline{[(\mathbf{c}^*)''\mathrm{v}'']'} + \overline{[(\mathbf{c}^*)''\mathrm{w}'']'}\ ) \, .
\end{aligned}$$

   with the modulation that for $j = 2, 3$ $\overline{\kappa}_j$ *is replaced by* 0 *if it is positive and the same holds for* $\underline{\kappa}_j$ *if it is negative. The following holds for every* $j \in \{1, \ldots, k-1\}$.

- *Either* $\delta^2 x_j^{(n)} \leq 0$ *or*

$$(1+\tau\overline{\kappa}_2)\delta^2 x_j^{(n)} + \tau\overline{\kappa}_3 \delta x_j^{(n)} \leq \delta^2 x_j^{(n-1)}$$

holds.

- *Either* $\delta^2 x_j^{(n)} \geq 0$ *or*

$$(1-\tau\underline{\kappa}_2)\delta^2 x_j^{(n)} + \tau\underline{\kappa}_3 \delta x_j^{(n)} \geq \delta^2 x_j^{(n-1)}$$

holds.

Remark 3.18 (The global maximum-/minimum principle). *The first max/min principle has a global form. Inductively applying the inequality-chain we receive*

$$\mathrm{e}^{-\underline{\kappa}_1 T} \min_i \delta x_i^{(0)} \leq \delta x_j^{(n)} \leq \mathrm{e}^{\underline{\kappa}_1 T} \max_i \delta x_i^{(0)} \ .$$

This clearly bounds $\delta x_i^{(n)}$ *from above and away from zero uniformly in* i, n, τ, h_k *as soon as the initial data are bounded from above and away from zero.*

Remark 3.19 (The second global maximum-/minimum principle). *The second maximum-/minimum principle implies a global-in-time bound as follows. Inductively , we receive*

$$\begin{aligned} (1+\tau\overline{\kappa}_2)\delta^2 x_i^{(n)} + \tau\overline{\kappa}_3 \delta x_i^{(n)} &\leq \delta^2 x_i^{(n-1)} \\ (1+\tau\overline{\kappa}_2)^2\delta^2 x_i^{(n)} + \tau\overline{\kappa}_3(1+\tau\overline{\kappa}_2)\delta x_i^{(n)} + \tau\overline{\kappa}_3 \delta x_i^{(n-1)} &\leq \delta^2 x_i^{(n-2)} \\ &\vdots \\ (1+\tau\overline{\kappa}_2)^n\delta^2 x_i^{(n)} + \tau\overline{\kappa}_3 \sum_{l=0}^{n-1}(1+\tau\overline{\kappa}_2)^l \delta x_i^{(l+1)} &\leq \delta^2 x_i^{(0)} \end{aligned}$$

Now by the first max./min. principle we can estimate $\delta x_i^{(l+1)} \leq \max_i \delta x_i^{(0)}$. *Then we note that the remaining sum* $\sum_{l=0}^{n-1}(1+\tau\overline{\kappa}_2)^l$ *is part of a geometric series, at least for* $\tau < \frac{1}{\overline{\kappa}_2}$ *since* $\overline{\kappa}_2 \leq 0$ *by definition. So* $\left|\sum_{l=0}^{n-1}(1+\tau\overline{\kappa}_2)^l\right| \leq \frac{1}{1-\tau\overline{\kappa}_2} \to 1$ *for* $\tau \searrow 0$. *So we can assume* $\sum_{l=0}^{n-1}(1+\tau\overline{\kappa}_2)^l \leq 2$, *therefore arriving at*

$$\delta^2 x_i^{(n)} \leq \max\left\{\mathrm{e}^{-T\overline{\kappa}_2} \max_i \delta^2 x_i^{(0)} + 2\overline{\kappa}_3 \max_i \delta x_i^{(0)}, 0\right\}$$

and

$$\delta^2 x_i^{(n)} \geq \min\left\{\mathrm{e}^{T\underline{\kappa}_2} \min_i \delta^2 x_i^{(0)} + 2\underline{\kappa}_3 \min_i \delta x_i^{(0)}, 0\right\} \ .$$

These two bounds, as in the first case, $\delta^2 x_i^{(n)}$ *from above and below uniformly in* i, n, τ, h_k *as soon as the initial data are bounded from above and away from zero.*

Proof of Lemma 3.17. We will only consider the bounds from above, since the bounds from below can be achieved by the same calculations, mutatis mutandis.

We begin with a general result concerning the finite forward difference of the r.h.s. of the Euler-Lagrange equation. Taylor's formula gives us some vector $(\zeta_0, \ldots, \zeta_{k-1})$ such that

$$\mathbf{D}\delta(\mathbf{c}^*)'(a_{i,k}^{(n)}) = \mathbf{D}\left[(\mathbf{c}^*)''(\zeta_i)\ \delta a_{i,k}^{(n)}\right] \ . \tag{3.2.6}$$

Here $\mathbf{D}$ denotes either Id for the first, and δ applied at $i-1$ for the second principle. ζ_i is some point between $a_{i,k}^{(n)}$ and $a_{i+1,k}^{(n)}$.

The proof will now rely upon assuming that we are at i such that $\delta x_i^{(n)} \geq \delta x_j^{(n)}$ for all j (or $\delta^2 x_i^{(n)} \geq \delta^2 x_j^{(n)}$ for all j), proving a suitable upper bound for $\delta x_i^{(n)}$ which then carries over to all $x_j^{(n)}$.

The first maximum-/minimum principle is achieved by the calculation (remember, δ is the finite forward difference w.r.t. i)

$$\begin{aligned}
\delta a_{i,k}^{(n)} &= \delta \left[\delta \mathsf{u}_*'(\delta x_{i-1}^{(n)}) - \mathrm{v}'(x_i^{(n)}) - h_k \sum_{j=0}^{k} \mathrm{w}'(x_i^{(n)} - x_j^{(n)}) \right] \\
&= \delta^2 \mathsf{u}_*'(\delta x_i^{(n)}) - \delta \mathrm{v}'(x_i^{(n)}) - h_k 2 \sum_{j=0}^{k} \delta \mathrm{w}'(x_i^{(n)} - x_j^{(n)}) \\
&\leq 0 - \underline{\mathrm{v}''} \delta x_i^{(n)} - \underline{\mathrm{w}''} \delta x_i^{(n)} \\
&= -(\underline{\mathrm{v}''} + \underline{\mathrm{w}''}) \delta x_i^{(n)}
\end{aligned}$$

where we used the monotonicity of u_*' which translates the maximum of $\delta x_i^{(n)}$ in i to a maximum of $\mathsf{u}_*'(\delta x_i^{(n)})$ in i and therefore its second symmetric finite difference is not positive. Furthermore the positivity of $\delta x_i^{(n)}$ was used to estimate the terms involving v' and w'.

Let's consider (3.2.6) again and we see, that if $\underline{\mathrm{v}''}+\underline{\mathrm{w}''} \geq 0$ we can estimate $-(\mathbf{c}^*)''(\zeta)(\underline{\mathrm{v}''}+\underline{\mathrm{w}''})\delta x_i^{(n)} \leq 0$ and if it is not, then we arrive only at $-(\mathbf{c}^*)''(\zeta)(\underline{\mathrm{v}''}+\underline{\mathrm{w}''})\delta x_i^{(n)} \leq -\overline{(\mathbf{c}^*)''}(\underline{\mathrm{v}''}+\underline{\mathrm{w}''})\delta x_i^{(n)}$. Recall, that $(\mathbf{c}^*)''$ is only finite, since we have bounded $\mathbf{c}''$ away from zero in *Assumption 3.6.* So let us define accordingly

$$\overline{\kappa}_1 := \begin{cases} 0 & \text{if } \mathrm{v}'' + \mathrm{w}'' \geq 0 \\ \overline{(\mathbf{c}^*)''}(\underline{\mathrm{v}''} + \underline{\mathrm{w}''}) & \text{if } \mathrm{v}'' + \mathrm{w}'' < 0 \,. \end{cases}$$

Plugging this back in (3.2.6) and using $0 < (\mathbf{c}^*)'' \leq \overline{(\mathbf{c}^*)''} < \infty$, we arrive at

$$\frac{\delta x_i^{(n)} - \delta x_i^{(n-1)}}{\tau} \leq -\overline{\kappa}_1 \delta x_i^{(n)}$$

and therefore, for every $j \in \{0, \ldots, k-1\}$ we have

$$\mathrm{e}^{\overline{\kappa}_1 \tau} \delta x_j^{(n)} \leq \mathrm{e}^{\overline{\kappa}_1 \tau} \delta x_i^{(n)} \leq (1 + \overline{\kappa} \tau) \delta x_i^{(n)} \leq x_i^{(n-1)} \leq \max_j x_j^{(n-1)}$$

which is the estimate we wanted to show.

The second maximum-/minimum principle can be shown similarly. But the terms will cause more work which is why we will split them up in this calculation.

Again we begin with assuming that i is such that $\delta^2 x_i^{(n)} \geq \delta^2 x_j^{(n)}$ for all j and, since otherwise nothing is left to show, we assume $\delta^2 x_i^{(n)} > 0$. By linearity, we have three terms to consider.

The internal pressure term can be dealt with as follows. Consider

$$\frac{(\mathbf{c}^*)''(\zeta_i) \left(\delta[\mathsf{u}_*'(\delta x_i^{(n)})] - \delta[\mathsf{u}_*'(\delta x_{i-1}^{(n)})] \right) + (\mathbf{c}^*)''(\zeta_{i-1}) \left(\delta[\mathsf{u}_*'(\delta x_{i-1}^{(n)})] - \delta[\mathsf{u}_*'(\delta x_{i-2}^{(n)})] \right)}{h_k^2} \,.$$

We will show that the terms in the big brackets are smaller than zero and we will carry out the calculation representatively for the left one.

By the assumptions on $\delta^2 x_i^{(n)}$ we know $\delta x_i^{(n)} > \delta x_{i-1}^{(n)}$. Consider the following two cases

If $\delta x_{i+1}^{(n)} > \delta x_i^{(n)}$, then $\delta^2 x_{i+1}^{(n)} > 0$. We can then apply Taylors theorem to receive

$$\delta[\mathsf{u}_*'(\delta x_i^{(n)})] - \delta[\mathsf{u}_*'(\delta x_{i-1}^{(n)})] = \mathsf{u}_*''(\hat{\zeta}_i)\delta^2 x_{i+1}^{(n)} - \mathsf{u}_*''(\hat{\zeta}_{i-1})\delta^2 x_i^{(n)}$$

where $\hat{\xi}_i > x_i^{(n)} > \hat{\xi}_{i-1}$ and therefore, by monotonicity, $\mathsf{u}_*''(\hat{\zeta}_i) < \mathsf{u}_*''(\hat{\zeta}_{i-1})$. Consequently

$$\mathsf{u}_*''(\hat{\zeta}_i)\delta^2 x_{i+1}^{(n)} - \mathsf{u}_*''(\hat{\zeta}_{i-1})\delta^2 x_i^{(n)} < \mathsf{u}_*''(\hat{\zeta}_{i-1})(\delta^2 x_{i+1}^{(n)} - \delta^2 x_i^{(n)}) \leq 0$$

by our assumptions on $\delta^2 x_i^{(n)}$ and $\mathsf{u}_*'' > 0$.

Now if $\delta x_{i+1}^{(n)} < \delta x_i^{(n)}$, this implies together with $\delta x_{i-1}^{(n)} < \delta x_i^{(n)}$ and by monotonicity of u_*'

$$\delta[\mathsf{u}_*'(\delta x_i^{(n)})] - \delta[\mathsf{u}_*'(\delta x_{i-1}^{(n)})] = \frac{\mathsf{u}_*'(\delta x_{i+1}^{(n)}) - 2\mathsf{u}_*'(x_i^{(n)}) + \mathsf{u}_*'(x_{i-1}^{(n)})}{h_k} \leq 0$$

right away.

Since $(\mathbf{c}^*)'' > 0$ the internal pressure term is taken care of.

Concerning the external potential term, we have to make use of a discrete version of the product rule of differentiation. We receive

$$\begin{aligned}
\delta\left[(\mathbf{c}^*)''(\zeta_{i-1})\delta \mathrm{v}'(x_{i-1}^{(n)})\right] &= \delta\left[(\mathbf{c}^*)''(\zeta_{i-1})\mathrm{v}''(\hat{\zeta}_{i-1})\delta x_{i-1}^{(n)}\right] \\
&= \delta\left[(\mathbf{c}^*)''(\zeta_{i-1})\mathrm{v}''(\hat{\zeta}_{i-1})\right]\delta x_i^{(n)} \\
&\quad + (\mathbf{c}^*)''(\zeta_{i-1})\mathrm{v}''(\hat{\zeta}_{i-1})\delta^2 x_i^{(n)} \\
&\geq \underline{[(\mathbf{c}^*)''\mathrm{v}'']'}\,\delta x_i^{(n)} + \overline{(\mathbf{c}^*)''}\cdot\underline{\mathrm{v}''}\delta^2 x_i^{(n)}\ .
\end{aligned}$$

Here we used the lower bound on $(\mathbf{c}^*)'''$ which leads to the finiteness of $\underline{[(\mathbf{c}^*)''\mathrm{v}'']'}$ by means of product rule and the bounds from above and below for $(\mathbf{c}^*)''$ and v'' respectively where we assume for the moment that $\mathrm{v}'' \leq 0$ which will be explained later on. Especially the usage of the latter bounds depend on our assumption $\delta^2 x_i^{(n)} \geq 0$.

Finally, basically the same calculation as the one above leads to the bounds

$$\delta\left[(\mathbf{c}^*)''(\zeta_{i-1})\delta h_k \sum_{j=0}^{k} \mathrm{w}'(x_{i-1}^{(n)} - x_{j-1}^{(n)})\right] \geq \underline{[(\mathbf{c}^*)''\mathrm{w}'']'}\,\delta x_i^{(n)} + \overline{(\mathbf{c}^*)''}\cdot\underline{\mathrm{w}''}\delta^2 x_i^{(n)}\ .$$

The main difference between the calculation concerning v and w is that we have to pull out h_k and the sum by linearity first and note that the $x_{j-1}^{(n)}$ are removed by the estimates on w as well as h_k and the sum cancel out by definition of h_k.

We define $\overline{\kappa}_2$ and $\overline{\kappa}_3$ as

$$\begin{aligned}
\overline{\kappa}_2 &:= \overline{(\mathbf{c}^*)''}\cdot(\underline{\mathrm{v}''} + \underline{\mathrm{w}''}) \\
\overline{\kappa}_3 &:= (\underline{[(\mathbf{c}^*)''\mathrm{v}'']'} + \underline{[(\mathbf{c}^*)''\mathrm{w}'']'}\ )\ .
\end{aligned}$$

with the modulation that for $j = 2, 3$ $\overline{\kappa}_j$ is replaced by 0 if it is positive.

This explains the assumption on v'' above, since we can carry out the calculations above for v and w together and then the particular part of v'' above will be played by $\mathrm{v}'' + \mathrm{w}''$.

We plug these results together to arrive at

$$\frac{\delta^2 x_i^{(n)} - \delta^2 x_i^{(n-1)}}{\tau} \leq -\overline{\kappa}_2 \delta^2 x_i^{(n)} - \overline{\kappa}_3 \delta x_i^{(n)}$$

which can be rewritten as

$$(1+\tau\overline{\kappa}_2)\delta^2 x_i^{(n)} + \tau\overline{\kappa}_3 \delta x_i^{(n)} \leq \delta^2 x_i^{(n-1)} \ . \qquad \square$$

3.3 A priori estimates

Note that the calculations below are more general than needed in the case of $\mathbf{c} \in C^3(\mathbb{R})$. For these results $\mathbf{c} \in C^2(\mathbb{R})$ suffices. In that case, we could simply retreat to the known consequence of the optimality of $\Phi(\tau; \mathbf{x}^{(n-1)}, \mathbf{x}^{(n)})$ in comparison to $\Phi(\tau; \mathbf{x}^{(n-1)}, \mathbf{x}^{(n-1)})$ which leads directly to

$$\tau \mathbf{T}^{\mathbf{x}}_{\mathbf{c},\tau}(\mathbf{x}^{(n)}, \mathbf{x}^{(n-1)}) \leq \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(n-1)}) - \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(n)}) \ . \tag{3.3.1}$$

With this we can achieve basically the same Hölder type inequality as below. Furthermore the second minimum/maximum principle implies uniform bounds on $\left|a_{i,k}^{(n)}\right|$ and $\left|(\mathbf{c}^*)'(a_{i,k}^{(n)})\right|$ independent of N, k as in *Lemma 3.23* with $p, p' = \infty$.

3.3.1 One time-step

The special structure of $\mathbf{x}^{(n)}$ as a minimizer of the functional $\mathcal{E}^{\mathbf{x}}$ penalized with some distance allows us to estimate the distance covered in a step $\mathbf{x}^{(n-1)}$ to $\mathbf{x}^{(n)}$ in terms of the descent in energy $\mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(n-1)})$ to $\mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(n)})$.

Proposition 3.20. *Let $\mathbf{x}^{(n-1)} \in \mathcal{X}_k([a,b])$ and $\mathbf{x}^{(n)}$ the unique minimizer of $\Phi(\tau; \mathbf{x}^{(n-1)}, \cdot)$. Let us furthermore treat the flux-limiting cost as $p = 2$. Then*

$$\tau h_k \sum_{i=0}^{k} \tilde{\mathbf{c}} \left(\frac{x_i^{(n)} - x_i^{(n-1)}}{\tau} \right) \leq \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(n-1)}) - \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(n)}) \ . \tag{3.3.2}$$

holds.

Proof. We begin with the system of equations we received as our Euler-Lagrange equations in *Lemma 3.15*

$$h_k \mathbf{c}' \left(\frac{x_i^{(n)} - x_i^{(n-1)}}{\tau} \right) = \left(\nabla \mathcal{U}^{\mathbf{x}}(\mathbf{x}^{(n)}) \right)_i - \left(\nabla \mathcal{V}^{\mathbf{x}}(\mathbf{x}^{(n)}) \right)_i - \left(\nabla \mathcal{W}^{\mathbf{x}}(\mathbf{x}^{(n)}) \right)_i$$

multiply it by $(x_i^{(n-1)} - x_i^{(n)})$ and sum it up over i to receive by the convexity of $\mathcal{E}^{\mathbf{x}}$ which was established in *Remark 3.9*

$$\tau h_k \sum_{i=0}^{k} \tilde{\mathbf{c}} \left(\frac{x_i^{(n)} - x_i^{(n-1)}}{\tau} \right) = \left\langle \nabla \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(n)}), \mathbf{x}^{(n)} - \mathbf{x}^{(n-1)} \right\rangle \leq \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(n-1)}) - \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(n)}) \ . \qquad \square$$

3.3.2 Multiple steps and the Hölder type inequality

A corollary of *Proposition 3.20* is the corresponding inequality with several time-steps.

Corollary 3.21. *Let $m_1 < m_2$ be two integers in $\{0, \dots, N\}$. Let $\mathbf{x}^{(m_1)} \in \mathcal{X}_k([a,b])$ and $\mathbf{x}^{(n)}$ for $n \in \{m_1, \dots, m_2\}$ be a sequence of recursively defined minimizers of Φ. Under the assumptions of Proposition 3.20 on τ the inequality*

$$\tau h_k \sum_{n=m_1+1}^{m_2} \sum_{i=0}^{k} \tilde{\mathbf{c}} \left(\frac{x_i^{(n)} - x_i^{(n-1)}}{\tau} \right) \leq \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(m_1)}) - \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(m_2)}) \tag{3.3.3}$$

holds.

Proof. This is a direct consequence of summing up the inequality in *Proposition 3.20*, minding the telescopic sum on the r.h.s. . □

With this estimate at hand, we can show the following estimate that will ultimately give us in-time compactness.

Proposition 3.22 (Hölder-type estimate)**.** *Let m_1, m_2 and τ be as in the corollary above. Let furthermore $t_1 = m_1\tau$ and $t_2 = m_2\tau$. Then*

$$h_k \left\| \mathbf{x}^{(m-1)} - \mathbf{x}^{(m-2)} \right\|_p \leq (t_2 - t_1 + \tau)^{1/p'} \frac{1}{\sqrt[p]{\alpha}} \left(\mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(0)}) - \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(N)}) \right)^{1/p} . \tag{3.3.4}$$

Note that the expression $\mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(0)}) - \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(N)})$ is positive can be bounded from above uniformly in τ and h_k by

$$\sup_k \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(0)}) - \underline{\mathcal{E}^{\mathbf{x}}}$$

since $\sup_k \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(0)})$ is assumed to be bounded and $\mathcal{E}^{\mathbf{x}}$ is bounded from below.

Proof. The proof relies on the Hölder-inequality as well as the *multiple-steps-inequality* above. We receive

$$\begin{aligned} h_k \left\| \mathbf{x}^{(m-1)} - \mathbf{x}^{(m-2)} \right\|_p &\leq \tau h_k \sum_{n=m_1+1}^{m_2} \left\| \frac{1}{\tau}(\mathbf{x}^{(n)} - \mathbf{x}^{(n-1)}) \right\|_p \\ &\leq \left(\sum_{n=m_1+1}^{m_2} \tau \right)^{1/p'} \left(\tau h_k \sum_{n=m_1+1}^{m_2} \left\| \frac{1}{\tau}(\mathbf{x}^{(n)} - \mathbf{x}^{(n-1)}) \right\|_p^p \right)^{1/p} \\ &\leq (t_2 - t_1 + \tau)^{1/p'} \left(\frac{1}{\alpha} \tau h_k \sum_{n=m_1+1}^{m_2} \sum_{i=0}^{k} \tilde{\mathbf{c}} \left(\frac{x_i^{(n)} - x_i^{(n-1)}}{\tau} \right) \right)^{1/p} \\ &\leq (t_2 - t_1 + \tau)^{1/p'} \frac{2}{\sqrt[p]{\alpha}} \left(\mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(0)}) - \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(N)}) \right)^{1/p} . \end{aligned}$$

□

Finally we want to bound the r.h.s. $(\mathbf{c}^*)'(a_{i,k}^{(n)})$ and its argument $a_{i,k}^{(n)}$ of our Euler-Lagrange equation.

Lemma 3.23. *Let τ, N be constants and $\mathbf{x}^{(n)}$ be a sequence as in Corollary 3.21. Then*

$$\tau h_k \sum_{n=1}^{N} \sum_{i=0}^{k-1} \left| a_{i,k}^{(n)} \right|^{p'} \leq \sqrt[p-1]{\beta} \left(\sup_k \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(0)}) - \underline{\mathcal{E}^{\mathbf{x}}} \right)$$

$$\tau h_k \sum_{n=1}^{N} \sum_{i=0}^{k-1} \left| (\mathbf{c}^*)'(a_{i,k}^{(n)}) \right|^{p} \leq \frac{1}{\alpha} \left(\sup_k \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(0)}) - \underline{\mathcal{E}^{\mathbf{x}}} \right)$$

holds.

Proof. Since by assumption $\tilde{\mathbf{c}}(s) \geq \alpha\,|s|^p$ we receive by Euler-Lagrange equation the bound

$$\tau h_k \sum_{n=0}^{N}\sum_{i=0}^{k} \alpha \left|(\mathbf{c}^*)'(a_{i,k}^{(n)})\right|^p \leq \tau h_k \sum_{n=0}^{N}\sum_{i=0}^{k} \tilde{\mathbf{c}}\left(\frac{x_i^{(n)} - x_i^{(n-1)}}{\tau}\right) \leq \sup_k \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(0)}) - \underline{\mathcal{E}^{\mathbf{x}}}$$

right away.

On the other hand, using the Euler-Lagrange equations and $\tilde{\mathbf{c}}(s) = s\mathbf{c}'(s)$ we see

$$a_{i,k}^{(n)}(\mathbf{c}^*)'(a_{i,k}^{(n)}) = \mathbf{c}'\left(\frac{x_i^{(n)} - x_i^{(n-1)}}{\tau}\right)\frac{x_i^{(n)} - x_i^{(n-1)}}{\tau} = \tilde{\mathbf{c}}\left(\frac{x_i^{(n)} - x_i^{(n-1)}}{\tau}\right)$$

which bounds

$$\tau h_k \sum_{n=1}^{N}\sum_{i=0}^{k-1} a_{i,k}^{(n)}(\mathbf{c}^*)'(a_{i,k}^{(n)}) \leq \sup_k \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(0)}) - \underline{\mathcal{E}^{\mathbf{x}}}\ .$$

The assumption $s\mathbf{c}'(s) \leq \beta\,|s|^p$ implies the upper bound $s(\mathbf{c}^*)'(s) \geq \frac{1}{\sqrt[p-1]{\beta}}\,|s|^{p'}$. Applied to our case this yields

$$\tau h_k \sum_{n=1}^{N}\sum_{i=0}^{k-1} \left|a_{i,k}^{(n)}\right|^{p'} \leq \sqrt[p-1]{\beta}\left(\sup_k \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(0)}) - \underline{\mathcal{E}^{\mathbf{x}}}\right)$$

□

3.4 Convergence of the approximate solution

The limits that are ultimately important for us are the limit of the two sides of the Euler-Lagrange equation for our approximations, stated in (3.2.5). The l.h.s. is an easy task, whereas receiving the limit of the r.h.s. will be most of the remaining chapter. The reasons for this are first $\mathfrak{V}_k$ is a non-linear function applied to A_k and we do not have enough compactness to pass with the limit of A_k through this non-linearity. This section will only find the limit of A_k and state that $\mathfrak{V}_k$ has some limit $\mathfrak{V}_*$ which will identified in a subsequent section by means of the so-called Browder-Minty trick, a monotonicity argument.

3.4.1 The approximate solution

The approximate solutions of our PDE (3.1.1) will be introduced next. To that end, let $\mathcal{I}_{i,k}^{(n)} : [0,T]\times[0,1] \to \mathbb{R}$ with $\mathcal{I}_{i,k}^{(n)}(t,\xi) = \mathbf{1}_{((n-1)\tau,n\tau])}(t)\mathbf{1}_{(\xi_i,\xi_{i+1})}(\xi)$.

Let $\tau_k, h_k > 0$, $\mathbf{x}^{(0)}$ feasible initial data and $\mathbf{x} = (x_i^{(n)})_{\substack{n\in\{0,\dots,N\}\\ i\in\{0,\dots,k\}}}$ the sequence recursively defined as minimizers of $\Phi(\tau;\mathbf{x}^{(n-1)},\cdot)$.

Though it is notational heavy, it will abbreviate some definitions, so let us furthermore define the map $\mathfrak{P}$ mapping a sequence $\mathbf{y} = (y_i^{(n)})_{\substack{n\in\{0,\dots,N\}\\ i\in\{0,\dots,k\}}}$ to the corresponding quantile functions

$$X_{\mathbf{y}}(t,\xi) = \mathfrak{P}[\mathbf{y}](t,\xi) := \sum_{i=0}^{k-1}\sum_{n=0}^{N} y_i^{(n)}\mathcal{I}_{i,k}^{(n)}(t,\xi)\ .$$

Then we define the approximate solution to our (3.1.1) in terms of quantile functions as

$$X_k(t,\xi) = \mathfrak{P}[\mathbf{x}](t,\xi) = \sum_{i=0}^{k-1}\sum_{n=0}^{N} x_i^{(n)}\mathcal{I}_{i,k}^{(n)}(t,\xi)\ . \tag{3.4.1}$$

Furthermore let us define the approximate first and second spatial derivative of X_k as well as the approximate first temporal derivative and the argument of the r.h.s. and its argument of the Euler-Lagrange equation

$$\begin{aligned} \delta_\xi X_k(t,\xi) &= \mathfrak{P}[\delta \mathbf{x}](t,\xi) \\ \delta_\xi^2 X_k(t,\xi) &= \mathfrak{P}[\delta^2 \mathbf{x}](t,\xi) \\ \delta_\tau X_k(t,\xi) &= \mathfrak{P}[\delta_\tau \mathbf{x}](t,\xi) \\ A_k(t,\xi) &= \mathfrak{P}[\mathbf{A}_k](t,\xi) \\ \mathfrak{V}_k(t,\xi) &= (\mathbf{c}^*)'(A_k(t,\xi)) \ . \end{aligned} \tag{3.4.2}$$

where $\mathbf{A}_k := (a_{i,k}^{(n)})_{\substack{n\in\{0,\ldots,N\}\\ i\in\{0,\ldots,k\}}}$, c.f. (3.2.4) and $\mathfrak{V}_k$ is the approximation of the velocity $\mathfrak{V}$ in our transport equation reformulation (3.1.14), hence the notation.

If the vector applied to $\mathfrak{P}$ is too short, as is for instance $\delta\mathbf{x}$, then the value at the missind indices is to be taken as zero.

3.4.2 Restating the results

We will recap some results from **Section 3.2 & 3.3** in terms of our approximate solution X_k.

The Euler-Lagrange equation reads as follows

$$\delta_\tau X_k(t,\xi) = (\mathbf{c}^*)'(A_k(t,\xi)) \ . \tag{3.4.3}$$

Anticipating the convergence results below, we can say, that every subsequence possesses a subsubsequence together with limits of the left and right hand side $\partial_t X_*$ and $\mathfrak{V}_*$ such that in the limit

$$\partial_t X_* = \mathfrak{V}_* \tag{3.4.4}$$

holds on $(0,T)\times(0,1)$.

Furthermore, the maximum/minimum principles imply the bounds

$$0 < \underline{M}_1 \le \delta_\xi X_k \le \overline{M}_1 < \infty \tag{3.4.5}$$

$$-\infty \le \underline{M}_2 \le \delta_\xi^2 X_k \le \overline{M}_2 < \infty \tag{3.4.6}$$

where the different M do not depend on k.

X_k is approximately Hölder continuous, that is to say

$$\|X_k(t_1,\cdot) - X_k(t_2,\cdot)\|_1 \le (|t_1-t_2| + \tau_k)^{1/p'} \sqrt[p]{\frac{2}{\alpha}} \left(\mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(0)}) - \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(N)})\right)^{1/p} \tag{3.4.7}$$

holds for all $t_1, t_2 \in [0,T]$. Note that $\sqrt[p]{\frac{2}{\alpha}} \left(\mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(0)}) - \mathcal{E}^{\mathbf{x}}(\mathbf{x}^{(N)})\right)^{1/p}$ does not depend on k.

3.4.3 Convergence results

Some convergence results can be received right away.

Lemma 3.24. *We have, up to the extraction of the same subsequence in each case, the existence of an increasing $X_* : [0,1] \to \mathbb{R}$ and the following convergences for $k\to\infty$ where $q = p$ or $q \in (1,\infty)$, depending on the regularity assumption of* $\mathbf{c}$.

1. *$X_k \to X_*$ w.r.t. strong $L^p((0,T)\times(0,1))$-topology and $X_* \in W^{1,q}((0,T)\times(0,1))$.*

2. *$\delta_\tau X_k \rightharpoonup \partial_t X_*$ and $\delta_\xi X_k \rightharpoonup \partial_\xi X_*$ w.r.t. $L^q((0,T)\times(0,1))$.*

3. *A_k lies in a subset of $L^{p'}([0,T]\times[0,1])$ that is compact w.r.t. the weak topology.*

4. *$\mathfrak{V}_k$ lies in a subset of $L^p([0,T]\times[0,1])$ that is compact w.r.t. the weak topology.*

Proof.

1. X_k is uniformly bounded in $L^q([0,T]\times[0,1])$ for every $q\in[1,\infty]$, it has a subsequence that weakly converges to X_*. By the result stated in *2.* together with *Proposition 3.11*, we arrive at strong convergence in $L^p((0,T)\times(0,1))$. In turn we can use a pointwise convergent subsequence together with the uniform bounds to receive strong convergence in $L^{\overline{p}}((0,T)\times(0,1))$ for each $\overline{p}\in[1,\infty]$, as well as $X_*\in W^{1,q}((0,T)\times(0,1))$.

2. By the Euler-Lagrange equation above (3.4.3) and with the uniform $L^p([0,T]\times[0,1])$ bound on $\mathfrak{V}_k$ from *Lemma 3.23*, we see that $\delta_\tau X_k$ lies in a set that is compact w.r.t. weak $L^p([0,T]\times[0,1])$-topology.

   The sequence $\delta_\xi X_k$, on the other hand, is uniformly bounded in $L^q([0,T]\times[0,1])$ for every $q\in[1,\infty]$ by the maximum-/minimum principle, implying compactness w.r.t. to weak $L^q([0,T]\times[0,1])$ convergence.

   Now using *Proposition 3.11* we receive that the limits of $\delta_\xi X_k$ and $\delta_\tau X_k$ are in fact the corresponding weak derivatives of X_*.

3 & 4 These are both direct consequences of the uniform bounds found in *Lemma 3.23* together with

$$\|A_k\|^{p'}_{L^{p'}((0,T)\times(0,1))} = \tau h_k \sum_{n=1}^{N}\sum_{i=0}^{k-1}\left|a^{(n)}_{i,k}\right|^{p'}$$
$$\|\mathfrak{V}_k\|^{p}_{L^{p}((0,T)\times(0,1))} = \tau h_k \sum_{n=1}^{N}\sum_{i=0}^{k-1}\left|(\mathbf{c}^*)'(a^{(n)}_{i,k})\right|^{p}\ .$$

Indeed since we have these uniform bounds and $p,p'\neq 1,\infty$, Banach-Alaoglu yields the results. □

Strong convergence of $\delta_\xi X_k$

Lacking pointwise convergence in $\delta_\xi X_k$, we cannot conclude strong convergence of this sequence immediately. It will be the result of this subsection to establish this convergence.

3.4.4 Strong convergence of $\delta_\xi X_k$

To achieve strong convergence of $\delta_\xi X_k$ on $[0,T]\times[0,1]$ we will make use of *Theorem 2.3*, which will basically consist in three steps. First we will choose suitable $\mathfrak{F}$ and $\mathbf{g}$ such that *Definition 2.1* is met and then we show that the preliminaries of *Theorem 2.3* are fulfilled. Finally we conclude compactness of our piecewise constant curve and then improve the compactness by the estimates we have already established.

We will use the definition of the set of functions of bounded variation on an open set Ω, $BV(\Omega)$ and of the total variation $\mathrm{Var}(f,\Omega)$ of such a function as was introduced in [17].

Tightness w.r.t. $\mathfrak{F}$

Consider the functional $\mathfrak{F} : L^1([0,1]) \to [0,\infty]$ defined as

$$\mathfrak{F}(Y) := \begin{cases} \mathrm{Var}(Y,(0,1)) & \text{if } Y \in BV((0,1)) \\ +\infty & \text{elsewhere.} \end{cases} \tag{3.4.8}$$

Lemma 3.25. *The functional $\mathfrak{F}$ defined in* (3.4.8) *is normal and coercive in the sense of [45, (1.7 a-c)]*

Proof.

1. *normal:* $\mathrm{Var}(\cdot,(0,1))$ is lower semicontinuous w.r.t. L^1-convergence [21, Thm. 1.9].

2. *coercive:* Consider $A_c := \mathfrak{F}^{-1}((-\infty, c])$. By definition of $\mathfrak{F}$, $A_c \subset BV((0,1))$ and the BV-norm of elements of A_c is uniformly bounded by c. Consequently, by [21, Thm. 1.19], A_c is $L^1((0,1))$-strongly compact. ☐

The above lemma makes sure that $\mathfrak{F}$ is suitable. Now we show that $\delta_\xi X_k$ is tight w.r.t. $\mathfrak{F}$.

Lemma 3.26. *Our sequence $\delta_\xi X_k$ is tight w.r.t. $\mathfrak{F}$, that is to say*

$$\sup_{k\in\mathbb{N}} \int_0^T \mathfrak{F}(\delta_\xi X_k(t,\cdot))\,\mathrm{d}t < \infty\ . \tag{3.4.9}$$

Proof. We begin with

$$\begin{aligned}\int_0^T \mathfrak{F}(\delta_\xi X_k(t,\cdot))\,\mathrm{d}t &= \int_0^T \mathrm{Var}(\delta_\xi X_k(t,\cdot),(0,1))\,\mathrm{d}t \\ &= \tau_k \sum_{n=1}^{N_k} \mathrm{Var}(\delta_\xi X_k(n\tau_k,\cdot),(0,1))\ .\end{aligned}$$

Now note that for every $t \in [0,T]$ we have that $X_k(t,\cdot) : [0,1] \to \mathbb{R}$ is a piecewise constant function and therefore its total variation can be calculated as the sum over the modulus of the jumps. This gives us

$$\begin{aligned}\tau_k \sum_{n=1}^{N_k} \mathrm{Var}(\delta_\xi X_k(n\tau_k,\cdot),(0,1)) &= \tau_k \sum_{n=1}^{N_k} \sum_{i=0}^{k-1} \left|\delta x_{i+1}^{(n)} - \delta x_i^{(n)}\right| \\ &= \tau_k \sum_{n=1}^{N_k} h_k \sum_{i=0}^{k-1} \left|\frac{\delta x_{i+1}^{(n)} - \delta x_i^{(n)}}{h_k}\right|\ .\end{aligned}$$

Now our goal is to utilize the estimates concerning $a_{i,k}^{(n)}$ from *Lemma 3.23.* To that end we see that by maximum-/minimum principle, positivity and monotony of u_*'' we receive a $D > 0$ such that we have the lower bound

$$\left|\delta x_{i+1}^{(n)} - \delta x_i^{(n)}\right| \le D \left|\mathsf{u}_*'(\delta x_{i+1}^{(n)}) - \mathsf{u}_*'(\delta x_i^{(n)})\right|\ .$$

Incorporating this estimate we arrive at

$$\begin{aligned}\int_0^T \mathfrak{F}(\delta_\xi X_k(t,\cdot))\,\mathrm{d}t &= \tau_k \sum_{n=1}^{N_k} h_k \sum_{i=0}^{k-2} \left| \frac{\delta x_{i+1}^{(n)} - \delta x_i^{(n)}}{h_k} \right| \\ &\le D\tau_k \sum_{n=1}^{N_k} h_k \sum_{i=1}^{k-1} \left| \frac{\mathsf{u}'_*(\delta x_i^{(n)}) - \mathsf{u}'_*(\delta x_{i-1}^{(n)})}{h_k} \right| \\ &\le D\tau_k \sum_{n=1}^{N_k} h_k \sum_{i=1}^{k-1} \left| a_{i,k}^{(n)} \right| + D\tau_k \sum_{n=1}^{N_k} h_k \sum_{i=0}^{k-1} \left| \mathrm{v}'(x_i^{(n)}) \right| \\ &\quad + D\tau_k \sum_{n=1}^{N_k} h_k \sum_{i=0}^{k-1} h_k \sum_{j=0}^{k} 2 \left| \mathrm{w}'(x_i^{(n)} - x_j^{(n)}) \right| \ .\end{aligned}$$

Now the first sums are bounded by *Lemma 3.23* and the second sums are by continuity of v' as well as w' together with the bounds on $x_i^{(n)}$.

This shows the sought for bound and therefore $\delta_\xi X_k$ is tight w.r.t. $\mathfrak{F}$. □

Equiintegrability w.r.t. the pairing g.

Now let $Y \in L^1([0,1])$. Then we can define $I[Y](\xi) := a + \int_0^\xi Y(\zeta)\,\mathrm{d}\zeta$ where $I[Y] \in W^{1,1}((0,1))$.

Next we consider the functional $\mathbf{g} : L^p([0,1]) \times L^p([0,1]) \to [0,\infty)$ defined as

$$\mathbf{g}(Y,Z) := \|I[Y] - I[Z]\|_{L^1([0,1])} \ . \tag{3.4.10}$$

Then the following lemma holds

Lemma 3.27. $\mathbf{g}$ *is lower semicontinuous w.r.t. strong $L^p([0,1]) \times L^p([0,1])$ topology.*

Proof. Let $Y_k \xrightarrow{L^p([0,1])} Y$ and let $\hat{Y} = I[Y_k]$. Then $\hat{Y}_k(0) = 0$ and $\partial_\xi \hat{Y}_k = Y_k$. This already implies that $\hat{Y}_k \xrightarrow{L^p([0,1])} \hat{Y}$ where $\hat{Y} \in W^{1,p}((0,1))$ with $\partial_\xi \hat{Y} = Y$.

Now consider two sequences Y_k, Z_k converging in $L^p([0,1])$ to Y and Z respectively. Then $I[Y_k] - I[Z_k] = I[Y_k - Z_k] - a$ converges in $L^p([0,1])$ to $I[Y-Z]-a$ and the $L^1([0,1])$-norm is lower semicontinuous w.r.t. $L^p([0,1])$ convergence. □

Finally we show the equiintegrability w.r.t. $\mathbf{g}$.

Lemma 3.28. *The sequence $\delta_\xi X_k$ is equiintegrable w.r.t.* $\mathbf{g}$*, that is to say*

$$\lim_{r \searrow 0} \sup_{k\in\mathbb{N}} \int_0^{T-r} \mathbf{g}\left(\delta_\xi X_k(t+r), \delta_\xi X_k(t)\right)\,\mathrm{d}t = 0 \ . \tag{3.4.11}$$

Proof. First we establish an estimate concerning $I[\delta_\xi X_k]$ and X_k. Let $t \in [0, T-r]$ and $j(\xi) = \lfloor \xi/d_k \rfloor$

$$\begin{aligned}
&\|I[\delta_\xi X_k(t+r)] - I[\delta_\xi X_k(t)]\|_{L^1([0,1])} \\
&\quad = \int_0^1 \left| \int_0^\xi \delta_\xi X_k(t+r,\zeta) - \delta_\xi X_k(t,\zeta)\,\mathrm{d}\zeta \right| \mathrm{d}\xi \\
&\quad = \int_0^1 \left| X_k(t+r, \tfrac{j(\xi)}{k}) - X_k(t, \tfrac{j(\xi)}{k}) + \int_{j(\xi)/k}^\xi \delta_\xi X_k(t+r,\zeta) - \delta_\xi X_k(t,\zeta)\,\mathrm{d}\zeta \right| \mathrm{d}\xi \\
&\quad \leq \int_0^1 \left| X_k(t+r, \tfrac{j(\xi)}{k}) - X_k(t, \tfrac{j(\xi)}{k}) \right| \mathrm{d}\xi \\
&\qquad + \frac{1}{h_k} \int_0^1 \int_{j(\xi)/k}^\xi \left| X_k(t+r, \tfrac{j(\xi)+1}{k}) - X_k(t+r, \tfrac{j(\xi)}{k}) - (X_k(t, \tfrac{j(\xi)+1}{k}) - X_k(t, \tfrac{j(\xi)}{k}))\,\mathrm{d}\zeta \right| \mathrm{d}\xi \\
&\quad \leq \|X_k(t+r) - X_k(t)\|_{L^1(J)} \\
&\qquad + \left(\sum_{j=0}^{k-1} \left| X_k(t+r, \tfrac{j+1}{k}) - X_k(t, \tfrac{j+1}{k}) \right| + \left| (X_k(t+r, \tfrac{j}{k}) - X_k(t, \tfrac{j}{k})) \right| \right) \int_{jd_k}^{(j+1)d_k} \frac{\xi - \frac{j}{k}}{h_k}\,\mathrm{d}\zeta \\
&\leq 3\,\|X_k(t+r) - X_k(t)\|_{L^1([0,1])} \ .
\end{aligned}$$

Together with the Hölder-estimate from *Proposition 3.22*, this yields, for some constant C such that

$$\begin{aligned}
\lim_{r\searrow 0} \sup_{k\in\mathbb{N}} \int_0^{T-r} \mathbf{g}\left(\delta_\xi X_k(t+r), \delta_\xi X_k(t)\right)\,\mathrm{d}t &= \lim_{r\searrow 0} \sup_{k\in\mathbb{N}} \int_0^{T-r} \|I[\delta_\xi X_k(t+r)] - I[\delta_\xi X_k(t)]\|_{L^1([0,1])}\,\mathrm{d}t \\
&= 3 \lim_{r\searrow 0} \sup_{k\in\mathbb{N}} \int_0^{T-r} \|X_k(t+r) - X_k(t)\|_{L^1([0,1])}\,\mathrm{d}t \\
&\leq 3 \lim_{r\searrow 0} \sup_{k\in\mathbb{N}} \int_0^{T-r} r^{1/p} C\,\mathrm{d}t \\
&= 3 \lim_{r\searrow 0} \sup_{k\in\mathbb{N}} r^{1/p}(T-r)C \\
&= 0
\end{aligned}$$

holds, proving the claim. □

Strong convergence

Proposition 3.29. *The sequence $\delta_\xi X_k$ is sequentially compact w.r.t. strong $L^{p'}((0,T) \times (0,1))$ convergence.*

Proof. This will be an application of *Theorem 2.3*. Note that $t \mapsto \delta_\xi X_k(t)$ is a sequence of $L^p([0,1])$-valued functions. The functional $\mathfrak{F}$ is normal and coercive in the sense of *Definition 2.1* as was shown in *Lemma 3.25* and *Lemma 3.27* shows that the substitute distance $\mathbf{g}$ satisfies the joint lower semi-continuity. Additionally, our sequence is tight w.r.t. $\mathfrak{F}$ and it is equiintegrable w.r.t. $\mathbf{g}$ which was shown in *Lemma 3.26* and *Lemma 3.28* respectively and therefore the preliminaries of *Theorem 2.3* are met. All that is left to show is the compatibility of $\mathbf{g}$ with $\mathfrak{F}$.

To show the compatibility, let $Y, Z \in L^1([0,1])$ with $\mathfrak{F}(Y), \mathfrak{F}(Z) < \infty$. Then $\|I[Y] - I[Z]\|_{L^1([0,1])} = 0$ implies that $I[Y] = I[Z]$ a.e. and consequently $Y = Z$. So *Theorem 2.3* is applicable and we receive a subsequence of $\delta_\xi X_k$ that converges in measure as a curve $\delta_\xi X_k : [0,T] \to L^1([0,1])$ to some limit curve Y.

Furthermore, since $\delta_\xi X_k$ is dominated on $[0,T]$ by the maximum-/minimum principle, we can enhance this result, possibly by passing to a subsubsequence, to strong convergence in $L^{p'}([0,T]\times[0,1])$. □

Finally we have to establish a connection between the cluster points of $\delta_\xi X_k$ and X_*, the limit of X_k.

Corollary 3.30. *$\delta_\xi X_k$ converges strongly w.r.t. $L^{p'}([0,T]\times(0,1))$ to $\partial_\xi X_*$, the weak spatial derivative of X_*.*

Proof. We already know that this sequence converges weakly to this limit, so the uniqueness of the limit proves the claim. □

Since we will need it later on, we state here, that the strong convergence of $\delta_\xi X_k$ can be carried over the non-linearity of u_* and u_*' to receive the following result.

Corollary 3.31. *Every subsequence of $\delta_\xi X_k$ has a (unrelabeled) subsubsequence such that $\mathsf{u}_*(\delta_\xi X_k)\to\mathsf{u}_*(\partial_\xi X_*)$ and $\mathsf{u}_*'(\delta_\xi X_k)\to\mathsf{u}_*'(\partial_\xi X_*)$ in $L^p([0,T]\times[0,1])$.*

Proof. Since $\delta_\xi X_k$ converges strongly, it has a subsequence that converges pointwise and by continuity, this pointwise convergence carries over through u_* and u_*'. Now we see that the dominated convergence theorem yields our claim. □

3.4.5 The limit of A_k

In this part, we will show that the limit of A_k is indeed unique and equal to $\partial_\xi\mathsf{u}_*'(\partial_\xi X_*(t,\xi))-\mathrm{v}'(X_*(t,\xi))-2\int_0^1\mathrm{w}'(X_*(t,\xi)-X_*(t,\zeta))\,\mathrm{d}\zeta$.

By the consideration above we will consider the u_*' summands and the summands including v' and w' separately, by means of

$$\begin{aligned}A_k(t,\xi) &= \frac{\mathsf{u}_*'(\delta_\xi X_k(t,\xi+h_k)-\mathsf{u}_*'(\delta_\xi X_k(t,\xi)}{h_k}\\&\quad-\left(\mathrm{v}'(X_k(t,\xi))+2\int_0^1\mathrm{w}'(X_k(t,\xi)-X_k(t,\zeta))\,\mathrm{d}\zeta\right)\\&=:\overline{A}_k(t,\xi)-\underline{A}_k(t,\xi)\ .\end{aligned}$$

To avoid ambiguity: $\overline{A}_k$ corresponds to the quotient of u_*' and $\underline{A}_k$ to the brackets with v' and w'.

Proposition 3.32. *We have the following convergences*

1. *$\underline{A}_k$ converges w.r.t. $L^q((0,T)\times(0,1))$-norm, where $q\in[1,\infty)$, to its limit*

$$\underline{A}_*(t,\xi):=\mathrm{v}'(X_*(t,\xi))+2\int_0^1\mathrm{w}'(X_*(t,\xi)-X_*(t,\zeta))\,\mathrm{d}\zeta\ .$$

2. *$\overline{A}_k$ converges w.r.t. to weak $L^{p'}((0,T)\times(0,1))$ topology to the limit*

$$\overline{A}_*:=\partial_\xi\mathsf{u}_*'(\partial_\xi X_*)\ .$$

Proof. To abbreviate $\mathrm{w}'(X_k(t,\xi)-X_k(t,\zeta))$ in this proof, we will simply write $\mathrm{w}'(X_k-X_k)$.

1. By a standard argument, $\mathrm{v}'(X_k)$ and $\mathrm{w}'(X_k - X_k)$ converge strongly in $L^q((0,T)\times(0,1))$.

   Indeed since X_k converges strongly in $L^1((0,T)\times(0,1))$ as was shown in *Lemma 3.24* it converges pointwise to X_* and consequently, by continuity of v' and w', $\mathrm{v}'(X_k)$ and $\mathrm{w}'(X_k - X_k)$ converge pointwise to $\mathrm{v}'(X_*)$ and $\mathrm{w}'(X_k - X_k)$ respectively.

   Combining this with the dominated convergence theorem, the uniform bounds on X_k and the continuity of v' and w' we arrive at the sought for strong convergence for v'.

   We also know by strong convergence that we have pointwise convergence of the map

$$\xi \mapsto \int_0^1 \mathrm{w}'(X_k(t,\xi) - X_k(t,\zeta))\,\mathrm{d}\zeta$$

   to the corresponding limit.

   Now using the dominated convergence theorem again, we arrive at the strong convergence for the w'-part of $\underline{A}_k$, too.

2. Since both, A_k and $\underline{A}_k$ converge at least weakly w.r.t. $L^{p'}((0,T)\times(0,1))$-topology, $\overline{A}_k = A_k - \underline{A}_k$ converges in the same way, too, and is therefore bounded in $L^{p'}((0,T)\times(0,1))$.

   So we have on the one hand, by Banach-Alaoglu, the weak convergence of the finite forward difference of $\mathsf{u}'_*(\delta_\xi X_k)$ and on the other hand, we already know $\mathsf{u}'_*(\delta_\xi X_k)$ converges to $\mathsf{u}'_*(\delta_\xi X_*)$ by *Corollary 3.31*. Applying *Proposition 3.11* now takes care of identifying the limit. □

3.5 Identification of the non-linear limit

In the preceeding section we have established some convergence results for X_k, $\delta_\xi X_k$, A_k and $\mathfrak{V}_k$, some only up to subsequences. In this section we assume that X_k etc. are already non-relabelled subsequences, such that all of the above convergence results hold. In particular, since $\mathfrak{V}_k$ lies in a compact set w.r.t. weak $L^p((0,T)\times(0,1))$ topology, we can assume it to converge to a limit $\mathfrak{V}_*$.

So far we have shown that our Euler-Lagrange equation admits a limit, but the limit of the r.h.s. is, by nonlinearity of $(\mathbf{c}^*)'$ still not identified. To identify this limit and therefore receive our IDF $X_*(t)$ as a weak solution of (3.1.1) in terms of IDF we will make use of a Browder-Minty argument.

We want to sketch the argument first. Let $x_k \to x_*$ and let f some nonlinear, monotone function. We suspect $f(x_k) \to f(x_*)$ but the convergence of x_k is too weak to receive this result right away. Instead we only know $f(x_k)$ converges to some y. Now the Browder-Minty argument considers

$$0 \le \langle f(x_k) - f(x_* - \varepsilon z), x_k - (x_* - \varepsilon z)\rangle \tag{3.5.1}$$

which holds by monotony of f for every $\varepsilon > 0$ and z in some set Z. If we can show

$$\limsup_{k\to\infty} \langle f(x_k) - f(x_* - \varepsilon z), x_k - (x_* - \varepsilon z)\rangle \le \langle y - f(x_* - \varepsilon z), \varepsilon z\rangle \tag{3.5.2}$$

we can establish $y = f(x_*)$ as soon as εz converges for $\varepsilon \searrow 0$ nicely enough to imply $f(x_* - \varepsilon z) \to f(x_*)$. Indeed we can in that case divide the inequality

$$0 \le \langle y - f(x_* - \varepsilon z), \varepsilon z\rangle$$

by ε and let $\varepsilon \searrow 0$ to receive $0 \le \langle y - f(x_*), z\rangle$ which already implies

$$0 = y - f(x_*)$$

if Z is big enough.

In our case the role of f, x_k and y will be played by $(\mathbf{c}^*)'$, A_k and $\mathfrak{V}_*$ respectively. Z will be specified later on.

The monotonicity inequality we want to consider will be

$$0 \le ((\mathbf{c}^*)'(A_k(t,\xi)) - (\mathbf{c}^*)'(A_*(t,\xi) - \varepsilon\psi(\xi)))(A_k(t,\xi) - (A_*(t,\xi) - \varepsilon\psi(\xi))) \tag{3.5.3}$$

where $\varepsilon > 0$ and $\psi \in C^\infty([0,1])$.

Since the above inequality holds for every $(t,\xi) \in (0,T) \times (0,1)$ we can integrate it weighted in time with some non-negative $u \in C_c^\infty((0,T))$ and making use of the abbreviation $\mathfrak{V}_k$ to arrive at

$$0 \le \int_0^T \int_0^1 \big(\mathfrak{V}_k(t,\xi) - (\mathbf{c}^*)'(A_*(t,\xi) - \varepsilon\psi(\xi))\big)\big(A_k(t,\xi) - (A_*(t,\xi) - \varepsilon\psi(\xi))\big)u(t)\,\mathrm{d}\xi\,\mathrm{d}t \tag{3.5.4}$$

which will be our specific instance of (3.5.1).

In order to prove the limit (3.5.2), we will expand the expression. Expressed in terms of the sketch of the argument, we will show $\limsup_k \langle f(x_k), x_k\rangle \le \langle y, x_*\rangle$, $\liminf_k \langle f(x_* - \varepsilon\psi), x_k - (x_* - \varepsilon\psi)\rangle \ge \langle f(x_* - \varepsilon\psi), x_* - (x_* - \varepsilon\psi)\rangle$ and $\liminf_k \langle f(x_k), x_* - \varepsilon\psi\rangle \le \langle y, x_* - \varepsilon\psi\rangle$ in this order. Plugging together these results and using subadditivity of the lim sup then shows the sought for estimate (3.5.2) or in our particular instance

$$\begin{aligned}\limsup_{k\to\infty} \int_0^T \int_0^1 &(\mathfrak{V}_k(t,\xi) - (\mathbf{c}^*)'(A_*(t,\xi) - \varepsilon\psi(\xi)))(A_k(t,\xi) - (A_*(t,\xi) - \varepsilon\psi(\xi)))u(t)\,\mathrm{d}\xi\,\mathrm{d}t \\ &\le \int_0^T \int_0^1 [\mathfrak{V}_*(t,\xi) - (\mathbf{c}^*)'(A_*(t,\xi) - \varepsilon\psi(\xi))]\,\varepsilon\psi(\xi)u(t)\,\mathrm{d}\xi\,\mathrm{d}t\,.\end{aligned} \tag{3.5.5}$$

3.5.1 $\limsup_k \langle f(x_k), x_k\rangle \le \langle y, x_*\rangle$

This part will be singled out, since it includes the most work.

We begin with splitting up $A_k = \overline{A}_k - \underline{A}_k$

$$\begin{aligned}\int_0^T \int_0^1 &\mathfrak{V}_k(t,\xi)A_k(t,\xi)u(t)\,\mathrm{d}\xi\,\mathrm{d}t \\ &= \int_0^T \int_0^1 \mathfrak{V}_k(t,\xi)\overline{A}_k(t,\xi)u(t)\,\mathrm{d}\xi\,\mathrm{d}t - \int_0^T \int_0^1 \mathfrak{V}_k(t,\xi)\underline{A}_k(t,\xi)u(t)\,\mathrm{d}\xi\,\mathrm{d}t\,.\end{aligned}$$

The second integral can be treated easily, since $\underline{A}_k \to \underline{A}_*$ strongly w.r.t. $L^{p'}([0,T]\times[0,1])$ by *Proposotion 3.32*. Together with the convergence of $\mathfrak{V}_k$ from *Lemma 3.24* we receive a subsequence with

$$\lim_{k\to\infty} \int_0^T \int_0^1 \mathfrak{V}_k(t,\xi)\underline{A}_k(t,\xi)u(t)\,\mathrm{d}\xi\,\mathrm{d}t = \int_0^T \int_0^1 \mathfrak{V}_*(t,\xi)\underline{A}_*(t,\xi)u(t)\,\mathrm{d}\xi\,\mathrm{d}t$$

right away.

To deal with the remaining integral will require more work. Indeed we will have a product of two weakly converging sequences, which, without further structure, can not be expected to converge to a limit.

We begin by applying the Euler-Lagrange equation (3.4.3) to the integral to receive

$$\int_0^T \int_0^1 \mathfrak{V}_k(t,\xi)\overline{A}_k(t,\xi)u(t)\,\mathrm{d}\xi\,\mathrm{d}t = \int_0^T \int_0^1 \delta_\tau X_k(t,\xi)\overline{A}_k(t,\xi)u(t)\,\mathrm{d}\xi\,\mathrm{d}t\,.$$

From here on, we will rewrite this integral by using its piecewise constant structure. The result will then be subject to summation by parts, a convexity estimate and again summation by parts and we will arrive at a integral which has a limit we can receive by the convergences we attained in the last section.

We begin with exploiting the piecewise constant structure

$$\begin{aligned}
&\int_0^T \int_0^1 \delta_\tau X_k(t,\xi)\overline{A}_k(t,\xi)u(t)\,\mathrm{d}\xi\,\mathrm{d}t \\
&\quad = \int_0^T \int_0^1 \sum_{n=1}^{N_k}\sum_{i=1}^{k-1} \frac{x_i^{(n)} - x_i^{(n-1)}}{\tau_k} \frac{\mathsf{u}_*'(\delta x_i^{(n)}) - \mathsf{u}_*'(\delta x_{i-1}^{(n)})}{h_k} \mathcal{I}_{i,k}^{(n)}(t,\xi)u(t)\,\mathrm{d}\xi\,\mathrm{d}t \\
&\quad = h_k \sum_{n=1}^{N_k}\sum_{i=1}^{k-1} \frac{x_i^{(n)} - x_i^{(n-1)}}{\tau_k} \frac{\mathsf{u}_*'(\delta x_i^{(n)}) - \mathsf{u}_*'(\delta x_{i-1}^{(n)})}{h_k} \int_{(n-1)\tau_k}^{n\tau_k} u(t)\,\mathrm{d}t\ .
\end{aligned}$$

We will abbreviate the integral in the last line as $Z_n^u := \int_{(n-1)\tau_k}^{n\tau_k} u(t)\,\mathrm{d}t$.

Now we apply summation by parts w.r.t. i, a convexity estimate and summation by parts w.r.t. n to receive

$$\begin{aligned}
&\sum_{n=1}^{N_k}\sum_{i=1}^{k-1} \frac{\mathsf{u}_*'(\delta x_i^{(n)}) - \mathsf{u}_*'(\delta x_{i-1}^{(n)})}{h_k} \frac{x_i^{(n)} - x_i^{(n-1)}}{\tau_k} h_k Z_n^u \\
&\quad = -\sum_{n=1}^{N_k}\sum_{i=1}^{k-1} \mathsf{u}_*'(\delta x_i^{(n)}) \frac{\delta x_i^{(n)} - \delta x_i^{(n-1)}}{\tau_k} h_k Z_n^u \\
&\quad \le -\sum_{n=1}^{N_k}\sum_{i=1}^{k-1} \frac{\mathsf{u}_*(\delta x_i^{(n)}) - \mathsf{u}_*(\delta x_i^{(n-1)})}{\tau_k} h_k Z_n^u \\
&\quad = \sum_{n=1}^{N_k}\sum_{i=1}^{k-1} \mathsf{u}_*(\delta x_i^{(n)}) h_k \frac{Z_{n+1}^u - Z_n^u}{\tau_k} \\
&\quad = \int_0^T \int_0^1 \sum_{n=1}^{N_k}\sum_{i=1}^{k-1} \mathsf{u}_*(\delta x_i^{(n)}) \mathcal{I}_{i,k}^{(n)}(t,\xi) \frac{u(t+\tau_k) - u(t)}{\tau_k}\,\mathrm{d}\xi\,\mathrm{d}t \\
&\quad = \int_0^T \int_0^1 \mathsf{u}_*(\delta_\xi X_k(t,\xi)) \frac{u(t+\tau_k) - u(t)}{\tau_k}\,\mathrm{d}\xi\,\mathrm{d}t
\end{aligned}$$

where the inequality is justified by the convexity of u_*.

Now as was shown, $\mathsf{u}_*(\delta_\xi X_k)$ converges strongly to $\mathsf{u}_*(\partial_\xi X_*)$ and since $u \in C_c^\infty((0,T))$ the difference quotient converges uniformly in k. Consequently we receive in the limit

$$\lim_{k\to\infty} \int_0^T \int_0^1 \mathsf{u}_*(\delta_\xi X_k(t,\xi)) \frac{u(t+\tau_k) - u(t)}{\tau_k}\,\mathrm{d}\xi\,\mathrm{d}t = \int_0^T \int_0^1 \mathsf{u}_*(\partial_\xi X_*(t,\xi))\partial_t u(t)\,\mathrm{d}\xi\,\mathrm{d}t\ .$$

Now we would want to get the partial differentiation w.r.t. t back to the u_* again to receive by chain rule a possibility to get our $\mathfrak{V}_*$ back. Unfortunately the expression $\mathsf{u}_*(\delta_\xi X_*)$ does not have enough regularity to allow for that.

But we can help ourselves by undoing the limit of just the difference quotient again. We receive

$$\begin{aligned}\int_0^T \int_0^1 \mathsf{u}_*(\delta_\xi X_*(t,\xi))\partial_t u(t)\,\mathrm{d}\xi\,\mathrm{d}t &= \lim_{k\to\infty} \int_0^T \int_0^1 \mathsf{u}_*(\delta_\xi X_*(t,\xi))\frac{u(t)-u(t-\tau_k)}{\tau_k}\,\mathrm{d}\xi\,\mathrm{d}t \\ &= \lim_{k\to\infty} -\int_0^T \int_0^1 \frac{\mathsf{u}_*(\delta_\xi X_*(t+\tau_k,\xi)) - \mathsf{u}_*(\delta_\xi X_*(t,\xi))}{\tau_k} u(t)\,\mathrm{d}\xi\,\mathrm{d}t \\ &\le \lim_{k\to\infty} -\int_0^T \int_0^1 \mathsf{u}_*'(\delta_\xi X_*(t,\xi))\frac{\delta_\xi X_*(t+\tau_k,\xi)-\delta_\xi X_*(t,\xi)}{\tau_k} u(t)\,\mathrm{d}\xi\,\mathrm{d}t\end{aligned}$$

where we used summation by parts w.r.t. n disguised in $t+\tau_k$ and convexity of u_*.

Now since $\mathsf{u}_*'(\delta_\xi X_*(t,\xi))$ is weakly differentiable w.r.t. ξ for every $t \in (0,T)$ we can integrate by parts to receive

$$\begin{aligned}-\lim_{k\to\infty} \int_0^T \int_0^1 \mathsf{u}_*'(\delta_\xi X_*(t,\xi))\frac{\delta_\xi X_*(t+\tau_k,\xi)-\delta_\xi X_*(t,\xi)}{\tau_k} u(t)\,\mathrm{d}\xi\,\mathrm{d}t \\ = \lim_{k\to\infty} \int_0^T \int_0^1 \partial_\xi \mathsf{u}_*'(\delta_\xi X_*(t,\xi))\frac{X_*(t+\tau_k,\xi)-X_*(t,\xi)}{\tau_k} u(t)\,\mathrm{d}\xi\,\mathrm{d}t\ .\end{aligned}$$

Note that $X_*(t+\tau_k,0) - X_*(t,0) = X_*(t+\tau_k,1) - X_*(t,1) = 0$ for every t by construction of X_k and the pointwise limit.

Finally we recall that $\partial_\xi \mathsf{u}_*'(\delta_\xi X_*(t,\xi))u(t) \in L^{p'}([0,T]\times[0,1])$ and $X_* \in W^{1,p}((0,T)\times(0,1))$, so the temporal difference quotient of X_* converges at least weakly in $L^p((0,T)\times(0,1))$ to the equation $\partial_t X_* = \mathfrak{V}_*$ which holds in the sense of $L^p((0,T)\times(0,1))$ and which legitimises the last step

$$\begin{aligned}\lim_{k\to\infty} \int_0^T \int_0^1 \partial_\xi \mathsf{u}_*'(\delta_\xi X_*(t,\xi))\frac{X_*(t+\tau_k,\xi)-X_*(t,\xi)}{\tau_k} u(t)\,\mathrm{d}\xi\,\mathrm{d}t \\ = \int_0^T \int_0^1 \partial_\xi \mathsf{u}_*'(\delta_\xi X_*(t,\xi))\mathfrak{V}_*(t,\xi)u(t)\,\mathrm{d}\xi\,\mathrm{d}t\ .\end{aligned}$$

3.5.2 $\liminf \langle f(x_* - \varepsilon\psi), x_k - (x_* - \varepsilon\psi)\rangle \ge \langle f(x_* - \varepsilon\psi), \varepsilon\psi\rangle$ & $\liminf \langle f(x_k), x_* - \varepsilon\psi\rangle \ge \langle y, x_* - \varepsilon\psi\rangle$

The inequality we want to show next is

$$\begin{aligned}\liminf_{k\to\infty} \int_0^T \int_0^1 (\mathbf{c}^*)'\big(A(t,\xi) - \varepsilon\psi(\xi)\big)\big(A_k(t,\xi) - (A(t,\xi)-\varepsilon\psi(\xi))\big)\,\mathrm{d}\xi\,\mathrm{d}t \\ \ge \int_0^T \int_0^1 (\mathbf{c}^*)'\big(A(t,\xi) - \varepsilon\psi(\xi)\big)\varepsilon\psi(\xi)\,\mathrm{d}\xi\,\mathrm{d}t\ .\end{aligned}$$

Basically all of the work has already been done in the preceding sections, so all that is left to do is plug them together to receive the above inequality.

Note that we receive the sought for result not only in the lim inf but actually in the limit. Recall A_k converges w.r.t. $L^{p'}([0,T]\times[0,1])$ weakly to A_* (*Proposition 3.32*). So we have established the sought for inequality as equation for the limit as soon as we can show $(\mathbf{c}^*)'\big(A_* - \varepsilon\psi\big) \in L^p([0,T]\times[0,1])$.

The assumption $\alpha\,|s|^p \le s\mathbf{c}'(s)$ implies $\frac{1}{\sqrt[p-1]{\alpha}}\,|r|^{\frac{1}{p-1}} \ge |(\mathbf{c}^*)'(r)|$ by means of $(\mathbf{c}'(s))^{-1} = (\mathbf{c}^*)'(s)$. With this estimate we receive

$$\begin{aligned}\int_0^T \int_0^1 \big|(\mathbf{c}^*)'\big(A_*(t,\xi) - \varepsilon\psi(\xi)\big)\big|^p\,\mathrm{d}\xi\,\mathrm{d}t &\le \frac{1}{\sqrt[p-1]{\alpha}} \int_0^T \int_0^1 |A_*(t,\xi) - \varepsilon\psi(\xi)|^{p'}\,\mathrm{d}\xi\,\mathrm{d}t \\ &= \frac{1}{\sqrt[p-1]{\alpha}}\,\|A_* - \varepsilon\psi\|^{p'}_{L^{p'}([0,T]\times[0,1])}\ .\end{aligned}$$

Now since ψ is a test-function, $\|A_* - \varepsilon\psi\|^{p'}_{L^{p'}([0,T]\times[0,1])} < \infty$ and consequently $(\mathbf{c}^*)'\big(A_* - \varepsilon\psi\big) \in L^p([0,T]\times[0,1])$.

The final inequality we will show to identify the limit $\mathfrak{V}_*$ of $\mathfrak{V}_k$ is

$$\liminf_{k\to\infty} \int_0^T \int_0^1 \mathfrak{V}_k(t,\xi)(A_*(t,\xi) - \varepsilon\psi(\xi))\, \mathrm{d}\xi\, \mathrm{d}t \geq \int_0^T \int_0^1 \mathfrak{V}_*(t,\xi)(A_*(t,\xi) - \varepsilon\psi(\xi))\, \mathrm{d}\xi\, \mathrm{d}t\ .$$

Lemma 3.24 which states $\mathfrak{V}_k \to \mathfrak{V}_*$ weakly in $L^p([0,T]\times[0,1])$. Again we receive the sought for limit inferior estimate as an equation for the limit since we already confirmed that $A_* - \varepsilon\psi \in L^{p'}([0,T]\times[0,1])$.

3.5.3 Identification of the limit $\mathfrak{V}_*$

Plugging together the above calculations we arrive at (3.5.5). Combining this with (3.5.4) we receive

$$0 \leq \int_0^T \int_0^1 [\mathfrak{V}_*(t,\xi) - (\mathbf{c}^*)'(A_*(t,\xi) - \varepsilon\psi(\xi))]\, \varepsilon\psi(\xi)u(t)\, \mathrm{d}\xi\, \mathrm{d}t$$

which holds for every $\varepsilon > 0$, $u \in C_c^\infty((0,T))$ with $u \geq 0$ and $\psi \in C_c^\infty((0,1))$. Dividing by $\varepsilon > 0$ and exchanging $\psi \leftrightarrow -\psi$ we receive the equation

$$0 = \int_0^T \int_0^1 [\mathfrak{V}_*(t,\xi) - (\mathbf{c}^*)'(A_*(t,\xi) - \varepsilon\psi(\xi))]\psi(\xi)u(t)\, \mathrm{d}\xi\, \mathrm{d}t\ .$$

Finally we send $\varepsilon \searrow 0$ and receive, since $\varepsilon\psi$ converges uniformly to zero for $\varepsilon \searrow 0$,

$$0 = \int_0^T \int_0^1 [\mathfrak{V}_*(t,\xi) - (\mathbf{c}^*)'(A_*(t,\xi))]\psi(\xi)u(t)\, \mathrm{d}\xi\, \mathrm{d}t$$

which then implies

$$\mathfrak{V}_* = (\mathbf{c}^*)' \circ A_* \quad \text{a.e. .}$$

3.6 Solution of our PDE

We will consider, as with the cost functions, two cases. The one corresponding to cost defined as in *Definition 3.1*, which will be dealt with first, and cost defined as in *Definition 3.2* which will need a regularization argument made possible by our second maximum-/minimum principle will be dealt with thereafter.

3.6.1 Solution of our PDE

We will now show that the limit X_* is indeed a solution to our PDE (3.1.1) in terms of IDF, that is to say it satisfies a weak formulation of (3.1.23).

To that end we have to show that $X_*(t)$ is an inverse distribution function for every $t \in [0,T]$, that is to say that it is non-decreasing and $X_*(t,0) = a$ as well as $X_*(t,1) = b$ holds for all $t \in [0,T]$.

Consider the sequence $X_k(t)$ for some $t \in (0,T)$. By construction of $X_k(t)$, we know $X_k(t)$ is non-decreasing and that $X_k(t,0) = a$, $X_k(t,1) = b$ holds, so $X_k(t)$ converges pointwise to X_* and the limit is again non-decreasing. Furthermore, $X_*(t,0) = a$ as well as $X_*(t,1) = b$ holds, too.

We have to show that the initial data are assumed $X_*(t) \to X_*^{(0)}$ for $t \to 0$. To that end, we will show that X_* is Hölder continuous as a curve in $L^1([0,1])$. This will be a consequence of *Corollary 3.22*.

Indeed, written in terms of X_k, this inequality reads, for every $t_1, t_2 \in [0,T]$ as

$$\|X_k(t_1) - X_k(t_2)\|_{L^p([0,T])} \leq |t_2 - t_1 + \tau_k|^{1/p'} \sqrt[p]{\frac{2}{\alpha}} \left(\mathcal{E}^X(X_k^{(0)}) - \underline{\mathcal{E}^{\mathbf{x}}}\right)^{1/p} .$$

Now we know on the one hand, that $X_k(t)$ converges strongly in $L^1([0,1])$, and that $\mathcal{E}^X(X_k^{(0)})$ is bounded. This implies in the limit the following inequality for every $t_1, t_2 \in [0,T]$

$$\|X_*(t_1) - X_*(t_2)\|_{L^p([0,T])} \leq C\,|t_2 - t_1|^{1/p'}$$

for some $C > 0$. So our limit curve X_* is $\frac{1}{p'} = \frac{p-1}{p}$-Hölder continuous as a curve in $L^p([0,1])$.

In particular this implies $\lim_{t\searrow 0} X_*(t) = X^{(0)}$ in the $L^p([0,1])$ sense.

Finally we have already shown that X_* is a weak solution of the PDE corresponding to (**??**) in terms of IDF, or more explicitly we have shown that for every $\varphi \in C_c^\infty((0,T)\times(0,1))$ the equation

$$\begin{aligned} &-\int_0^T \int_0^1 X_*(t,\xi)\partial_t\varphi(t,\xi)\,\mathrm{d}\xi\,\mathrm{d}t \\ &\quad= \int_0^T \int_0^1 (\mathbf{c}^*)' \left(\partial_\xi \mathsf{u}_*'(\partial_\xi X_*(t,\xi)) - \mathrm{v}'(X_*(t,\xi)) - \int_0^1 \mathrm{w}'(X_*(t,\xi) - X_*(t,\zeta))\,\mathrm{d}\zeta\right)\varphi(t,\xi)\,\mathrm{d}\xi\,\mathrm{d}t \end{aligned}$$

holds.

3.6.2 The flux-limiting case

To prove *Theorem 2*, we will show that, given the prerequisites of *Theorem 2*, the flux-limiting cost functions can be regularized to be actually p-Wasserstein cost without changing the minimizers of our algorithm steps. The bound on $\delta^2\mathbf{x}_k^{(n)}$ we receive from *Lemma 3.17* will play the central role.

Let us assume $\mathbf{c}$, $\mathbf{x}_k^{(0)}$ are as in *Lemma 3.17*. Then by the very same result, the bounds for $\delta^2\mathbf{x}_k^{(0)}$ hold for all $\delta^2\mathbf{x}_k^{(0)}$, too. As a first consequence, this yields finite bounds from above and below for $a_{i,k}^{(n)}$. Indeed, by the properties of u_* we receive

$$a_{i,k}^{(n)} \in \left[\underline{\delta^2\mathbf{x}^{(0)}} \cdot \mathsf{u}_*''\left(\overline{\delta\mathbf{x}^{(0)}}\right), \overline{\delta^2\mathbf{x}^{(0)}} \cdot \mathsf{u}_*''\left(\underline{\delta\mathbf{x}^{(0)}}\right)\right]$$

if $\underline{\delta^2\mathbf{x}^{(0)}} < 0 < \overline{\delta^2\mathbf{x}^{(0)}}$. If it is the case that $\underline{\delta^2\mathbf{x}^{(0)}}$ and $\overline{\delta^2\mathbf{x}^{(0)}}$ lie on the same side of zero, one of the bounds in the interval has to be replaced with zero.

Now since $(\mathbf{c}^*)'$ is monotonously increasing, we receive bounds for the discrete temporal backward difference

$$\frac{x^{(n)}n_i - x^{(n)}n-1_i}{\tau_k} = (\mathbf{c}^*)'(a_{i,k}^{(n)}) \in \left[(\mathbf{c}^*)'\left(\mathsf{u}_*''(\overline{\delta\mathbf{x}^{(0)}})\underline{\delta^2\mathbf{x}^{(0)}}\right), (\mathbf{c}^*)'\left(\mathsf{u}_*''(\underline{\delta\mathbf{x}^{(0)}})\overline{\delta^2\mathbf{x}^{(0)}}\right)\right]$$

again with the appropriate corrections if $\underline{\delta^2\mathbf{x}^{(0)}}$ and $\overline{\delta^2\mathbf{x}^{(0)}}$ lie on the same side of zero. So to summarize, there are uniform bounds

$$\underline{C} := (\mathbf{c}^*)'\left(\mathsf{u}_*''(\overline{\delta\mathbf{x}^{(0)}})\underline{\delta^2\mathbf{x}^{(0)}}\right) > -\gamma; \quad \text{and} \quad \overline{C} := (\mathbf{c}^*)'\left(\mathsf{u}_*''(\underline{\delta\mathbf{x}^{(0)}})\overline{\delta^2\mathbf{x}^{(0)}}\right) < \gamma$$

such that for every k, n our minimization problem in the algorithm can be narrowed down to a minimization over $\mathbf{x}$ such that $\frac{x^{(n)}n_i - x^{(n)}n-1_i}{\tau_k} \in [\underline{C}, \overline{C}]$. This allows us in particular to regularize the flux-limiting cost $\mathbf{c}$ outside of $[\frac{\underline{C}-\gamma}{2}, \frac{\overline{C}+\gamma}{2}]$ to satisfy *Definition 3.1*

This shows that the above results hold for flux-limiting cost as well, as soon as we have initial data that satisfies the additional regularity assumption from *Theorem 2*.

3.7 Numerical experiments

3.7.1 Implementation

We perform the minimization of $F : \mathbf{x} \mapsto \Phi(\tau; \mathbf{x}^{(n-1)}, \mathbf{x})$ by a damped Newton scheme

$$\begin{aligned} p_j &= -HF(\mathbf{x}_j)^{-1}\nabla F(\mathbf{x}_j) \\ \mathbf{x}_{j+1} &= \mathbf{x}_j + h_j p_j, \quad j = 0, 1, \dots \end{aligned}$$

for the gradient of F. The choice of the step size h_j in each step is realized by an Armijo-type heuristic, i.e. we choose h_j as the largest value from the sequence $\{1, \frac{1}{2}, \frac{1}{4}, \dots\}$ for which

$$\mathbf{x}_j + h_j p_j \in \{\mathbf{x} \in \mathbb{R}^{k+1} \mid a = x_0 < x_1 < \dots < x_k = b, |x_i - x_i^{(n-1)}| \le \tau\},$$

i.e. such that the next iterate $\mathbf{x}_{j+1}$ is still an IDF and has a well defined optimal transport distance in the flux limited case.

A Matlab code for the following experiments is given in the Appendix.

3.7.2 Linear diffusion

We start with the case $m = 1$, i.e. the case of the Boltzmann entropy for the internal energy potential. All experiments have been carried out with $k = 1000$ grid points and time step $\tau = 0.01$. Figs. 3.1 and 3.2 show the evolution of an initial distribution with localized support over the time interval $[0, 2]$ for Wasserstein costs with $p = 7$, while Fig. 3.3 shows the same evolution for the flux limited case with c given by (3.1.12), $\gamma = 1$.

In Fig. 3.4 (left), we depict the L_1-error of the computed density in dependence of the mesh size. The error is estimated by computing the exact L_1 difference to a reference solution on a grid with 10000 points, the same initial condition as for Fig. 3.1 has been used. The experiments suggest that the error decreases linearly with the grid size. To the right in this figure, the L_1-error of the computed density (on a grid of 1000 points) in dependence of the time step is plotted. Again, we estimate this error by comparing to a reference solution, here with time step $\tau = 0.001$. The result clearly suggests a linear dependence of this error on the time step.

3.7.3 Porous medium equation

As a second experiment, we consider the case of nonlinear diffusion with $m = \frac{5}{3}$. We choose $p = \frac{4}{3}$ so that we obtain the q-Laplace equation with $q = \frac{2-p}{p-1} = 2$. Fig. 3.5 shows the evolution of the densities and the associated characteristics.

3.8 Auxiliary convergence results

3.8.1 Difference quotients and weak derivatives

(c.f. [20, Chapter 5.8.2])

Let $f_k : \Omega \to \mathbb{R}$ be a sequence of real valued functions on a open rectangle $\Omega \in \mathbb{R}^2$, $r_k \searrow 0$ a sequence and let $p \in (1, \infty)$ as well as $\mathsf{e}_1, \mathsf{e}_2$ be the canonical basis vectors. If f_k is uniformly bounded w.r.t. L^p and, for some $i = 1, 2$, $\delta_i f_k(x) := \frac{f_k(x + r_k \mathsf{e}_i) - f_k(x)}{r_k} \mathbf{1}_{\Omega_{r_k,i}}(x)$ with $\Omega_{\varepsilon,i} := \{x \in \Omega \mid x \pm \varepsilon \mathsf{e}_i \in \Omega\}$ is uniformly bounded w.r.t. $L^p(\Omega)$, then $f_k \to f_*$ in $L^p(I)$ and $\delta_i f_k \xrightarrow{weakly} \partial_{x_i} f_*$ on Ω.

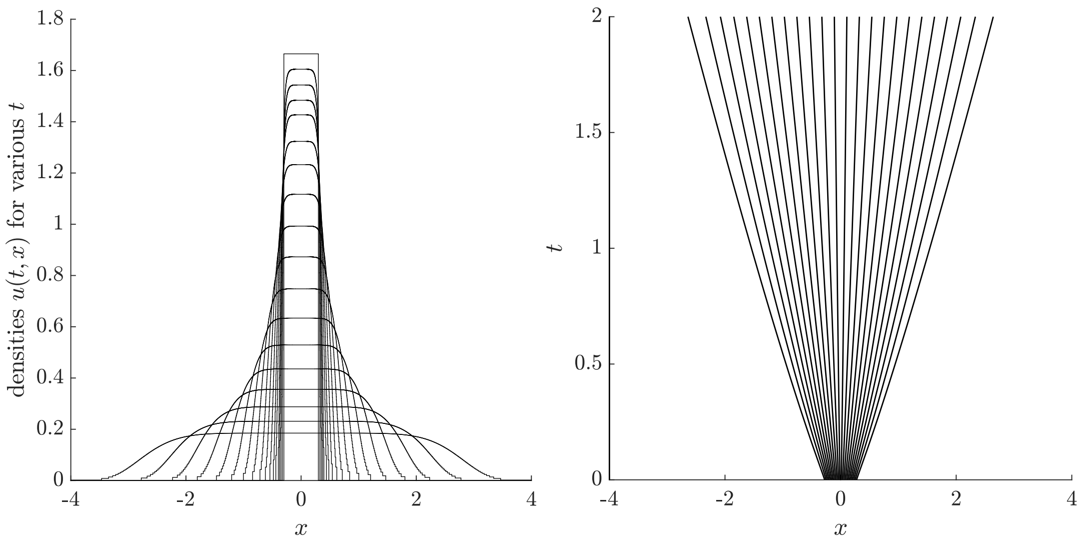


Figure 3.1: Experiment: p-Wasserstein cost, linear diffusion. Left: Approximate densities $u(t,\cdot)$ at $t = 0.01 \cdot 10^k$, $k = 0, 0.12, 0.24, \ldots, \log_{10}(200)$, initial mass uniformly distributed on $[-0.3, 0.3]$. Right: the corresponding characteristics.

Proof. The uniform bounds of f_k and $\delta_i f_k$ in L^p imply weak convergences of an (unrelabeled) subsequence of f_k, $\delta_i f_k$ to some limits $f_* : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ respectively, in L^p.

Furthermore, we receive for every $\varepsilon > 0$, $\phi \in C_c^\infty(\Omega_{\varepsilon,i})$ and k big enough such that $r_k < \varepsilon$

$$\begin{aligned}\int_{\Omega_{\varepsilon,i}} g(x)\phi(x)\,\mathrm{d}x &= \lim_{k\to\infty} \int_{\Omega_{\varepsilon,i}} \frac{f_k(x + r_k \mathsf{e}_i) - f_k(x)}{r_k} \phi(x)\,\mathrm{d}x \\ &= -\lim_{k\to\infty} \int_{\Omega_{\varepsilon,i}} f_k(x) \frac{\phi(x - r_k \mathsf{e}_i) - \phi(x)}{r_k}\,\mathrm{d}x \\ &= -\int_{\Omega_{\varepsilon,i}} f_k(x)\partial_{x_i}\phi(x)\,\mathrm{d}x\,.\end{aligned}$$

This argument holds for every $\varepsilon > 0$ so the claim follows by uniqueness of the weak derivative. □

3.8.2 Strong convergence and Lipschitz-functions

Let $f_k : \Omega \to W$ be a sequence of real valued functions on an open $\Omega \in \mathbb{R}^n$ with values in some closed W converging strongly w.r.t. $L^p(\Omega)$ for some $p \in (1, \infty)$ to $f_* : \Omega \to \mathbb{R}$. Let furthermore $g : \mathbb{R} \to \mathbb{R}$ be Lipschitz on W. Then $g(f_k) \to g(f_*)$ in $L^p(\Omega)$.

Proof. Strong convergence of f_k implies uniform boundedness for every subsequence a subsubsequence converging pointwise to some limit. By continuity of g this carries over to $g(f_k)$ and said subsubsequences thereof. On the other hand, f_k being bounded in L^p implies, by Lipschitz-continuity $g(f_k)$ being bounded in L^p and therefore, by the dominated convergence theorem, the pointwise convergence of said subsubsequences of $g(f_k)$ is actually a convergence w.r.t. L^p. Finally, since f_k converges to f_*. all the cluster points of $g(f_k)$ are $g(f_*)$, which implies the claim. □

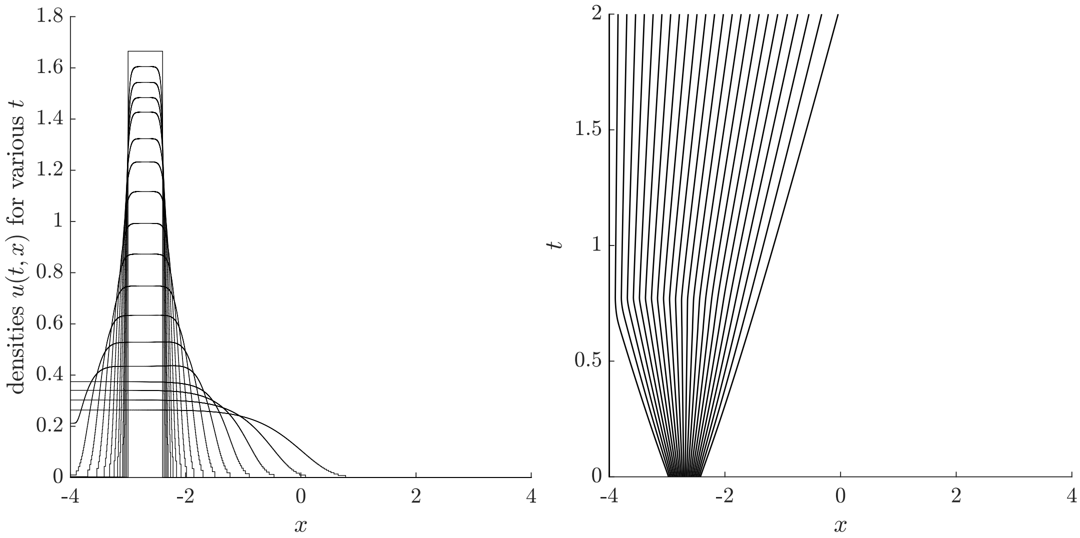


Figure 3.2: Experiment: p-Wasserstein cost, linear diffusion. Left: Approximate densities $u(t,\cdot)$ at $t = 0.01 \cdot 10^k$, $k = 0, 0.12, 0.24, \ldots, \log_{10}(200)$, initial mass uniformly distributed on $[-3, -2.4]$. Right: the corresponding characteristics.

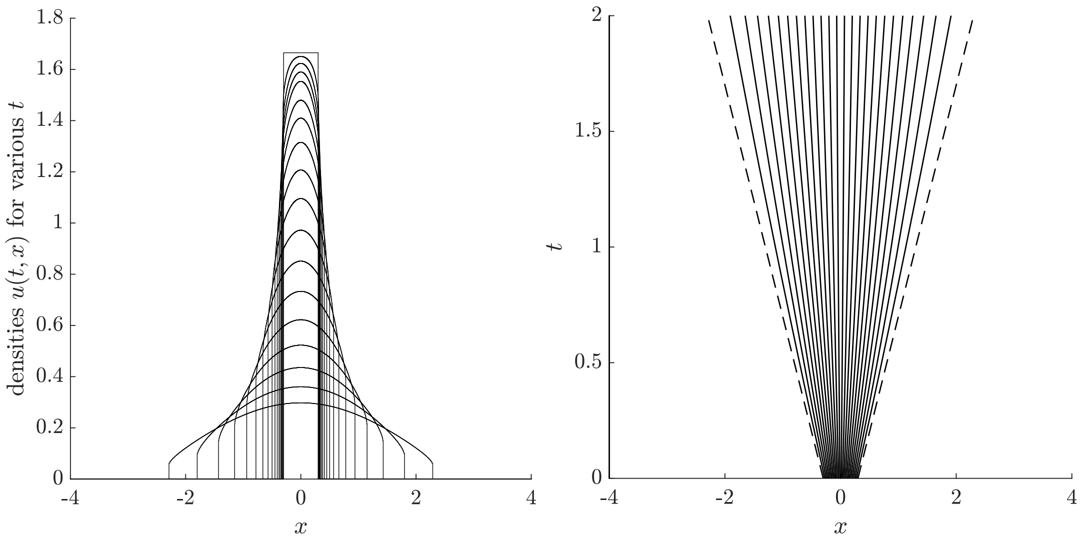


Figure 3.3: Experiment: relativistic cost, linear diffusion. Left: Approximate densities $u(t,\cdot)$ for $t = 0.01 \cdot 10^k$, $k = 0, 0.12, 0.24, \ldots, \log_{10}(200)$, initial mass uniformly distributed on $[-0.3, 0.3]$. Right: the corresponding characteristics (dashed: speed of light).

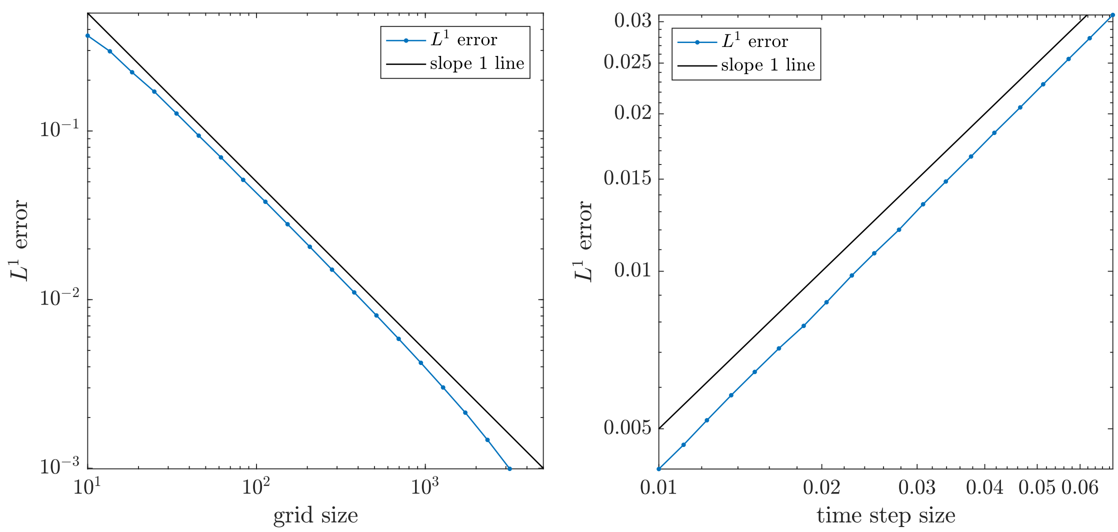


Figure 3.4: Convergence analysis: relativistic cost, linear diffusion. L^1-error of the inverse distribution function in dependence of the grid size (left), and in dependence of the time step (right).

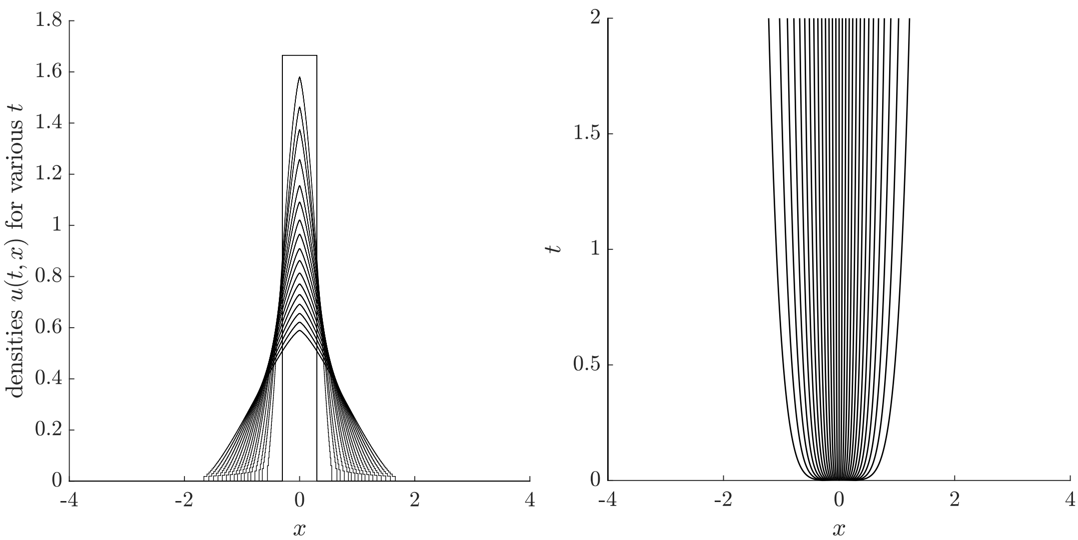


Figure 3.5: Experiment: q-Laplace ($p = \frac{4}{3}, m = \frac{5}{3}$). Left: Approximate densities $u(t, \cdot)$ for $t = 0.01 \cdot 10^k$, $k = 0, 0.12, 0.24, \ldots, \log_{10}(200)$, initial mass uniformly distributed on $[-0.3, 0.3]$. Right: the corresponding characteristics.

Chapter 4

Entropic regularized relativistic heat equation

4.1 Introduction

This chapter follows very closely the joint work of the author with Daniel Matthes [35]. The main difference to [35] consists of the generelization of the free energy functional $\mathcal{E}$ and therefore the family of equations (4.1.1).

4.1.1 General idea

In the field of numerical solution of transportation problems — like estimation of Wasserstein distances, computation of barycenters, or parameter estimation — entropic regularization has been proven a versatile and impressively efficient tool. Based on Cuturi's adaptation of the Sinkhorn algorithm for "lightspeed computation of optimal transport" [16], a huge variety of highly efficient methods for various current applications of transport theory have been developed, see the recent book [41] for an overview. The focus has been mainly on image and data science, but the ideas have been applied for numerical approximation of gradient flows as well, see e.g. [42, 8]. Here, we develop this approach further to define an efficient scheme for approximation of solutions to flux-limited equations of the type

$$\partial_t \rho + \nabla \cdot \big[\rho\, a\big(\nabla\, [h'(\rho) + \mathrm{v} + (\mathrm{w} * \rho)]\,\big)\big] = 0, \quad \rho(0,\cdot) = \rho^0. \tag{4.1.1}$$

In that problem, the sought solution ρ is a time-dependent probability density, either on $\Omega = \mathbb{R}^d$ with finite second moment, or on a bounded domain $\Omega \subset \mathbb{R}^d$ with no-flux boundary conditions. The given function $h : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is convex and super-linear, the external and interaction potentials v and w are positive, bounded and in $C^2(\mathbb{R}^d)$ while w is symmetric in the sense that $\mathrm{w}(-z) = \mathrm{w}(z)$ for all $z \in \mathbb{R}^d$ and $a : \mathbb{R}^d \to \overline{\mathbb{B}}$ is a monotone map into the closed unit ball $\overline{\mathbb{B}}$ of $\mathbb{R}^d$. This implies the aforementioned flux limitation, since (4.1.1) can be considered as a transport equation with velocities $a(\nabla h'(\rho))$ of modulus less than one. The expression $(\mathrm{w} * \rho)$ denotes the convolution of w with ρ

$$(\mathrm{w} * \rho)(x) = \int_\Omega \mathrm{w}(x-y)\rho(y)\,\mathrm{d}y\ .$$

Our primary example will be Rosenau's relativistic heat equation [44],

$$\partial_t \rho = \nabla \cdot \left(\rho\, \frac{\nabla \rho}{\sqrt{\rho^2 + |\nabla \rho|^2}}\right), \tag{4.1.2}$$

which is (4.1.1) with $h(r) = r(\log r - 1)$, $\mathrm{v} = \text{const.} = \mathrm{w}$ and $a(p) = (1 + |p|^2)^{-1/2}p$. This equation has been analyzed in great detail, mostly by Caselles and collaborators, see [13, 3, 4] and references therein. Schemes for numerical solution of (4.1.2) have been developed as well, see e.g. [9], however, these are very different from the approach taken here.

In the definition of the entropic regularization of (4.1.1), its discretization in space and time, and the efficient numerical implementation, we closely follow the blueprint laid out in [42] for gradient flows in the L^2-Wasserstein metric. In order to make that variational approach feasible, we require a special structure of a, namely that it can be written in the form

$$a(p) = \nabla \mathbf{C}^*(-p), \tag{4.1.3}$$

where $\mathbf{C}^*$ is the Legendre transform of a convex cost function $\mathbf{C} : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$. The flux limitation is implemented by requiring further that $\mathbf{C}$ is continuous on the closed unit ball $\overline{\mathbb{B}}$, equal to one on the boundary $\partial\mathbb{B}$, and is $+\infty$ outside of $\overline{\mathbb{B}}$. As observed by Brenier [7], the relativistic heat equation (4.1.2) fits into that framework, by choosing $\mathbf{C}(v) = 1 - \sqrt{1 - |v|^2}$ for $v \in \overline{\mathbb{B}}$.

4.1.2 Gradient flow structure

With the assumption (4.1.3) on a, (4.1.1) can be considered as a gradient flow on the space $\mathcal{P}(\Omega)$ of probability measures on Ω, at least formally. We briefly recall the basic idea in a language that is suitable for formulation of our approximation later. We refer e.g. to [2, 1, 7, 37] for further details on the variational structure of (4.1.1).

The potential of that gradient flow is the energy functional

$$\mathcal{E}(\rho) := \int_\Omega h(\rho) + \rho(\mathrm{v} + \tfrac{1}{2}(\mathrm{w} * \rho))\,\mathrm{d}x. \tag{4.1.4}$$

And the respective dissipation $\mathfrak{D}(\rho; q)$ for a given "tangential vector" q at $\rho \in \mathcal{P}(\Omega)$ — that is, $q \in L^1(\Omega)$ is of zero average — is defined by

$$\mathfrak{D}(\rho; q) := \inf_{q = \nabla\cdot(\rho v)} \int_\Omega \mathbf{C}(v)\rho\,\mathrm{d}x. \tag{4.1.5}$$

Here the infimum runs over all vector fields $v : \Omega \to \mathbb{R}^d$ for which $q = \nabla \cdot (\rho v)$, and equals infinity if there is no such v. The integral in (4.1.5) represents the friction resulting from the infinitesimal motion of all mass elements in ρ along the vector field v; taking the infimum over v's means that the infinitesimal mass elements move in the least dissipative way to realize the macroscopic change determined by q.

A curve $\rho : \mathbb{R}_{\geq 0} \to \mathcal{P}(\Omega)$ is of *steepest descent in $\mathcal{E}$'s landscape with respect to $\mathfrak{D}$* if at each instance $t_0 > 0$, the derivative $\partial_t\rho(t_0)$ is such that the sum

$$\mathfrak{D}\big(\rho(t_0); \partial_t\rho(t_0)\big) + \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=t_0} \mathcal{E}\big(\rho(t)\big) \tag{4.1.6}$$

is minimized, i.e., the decrease in energy is optimal with respect to the induced dissipation. Assuming that $\rho(t_0)$ is smooth and positive everywhere, then a straight-forward calculation shows that the minimizing $\partial_t\rho(t_0) = \nabla \cdot (\rho(t_0)v(t_0))$ is determined by the vector field $v(t_0)$ that minimizes

$$v \mapsto \int_\Omega \big[\mathbf{C}(v)\rho(t_0) + \big[h'\big(\rho(t_0)\big) + \mathrm{v} + (\mathrm{w} * \rho(t_0))\big]\,\nabla \cdot \big(\rho(t_0)v\big)\big]\,\mathrm{d}x.$$

In view of (4.1.3), this produces the evolution equation (4.1.1).

4.1.3 Discretization and regularization

To connect to the variational framework of optimal transport, we perform a time-discrete approximation of (4.1.6) in the spirit of the minimizing movement scheme [2], which is often refered to as JKO method [23] in the context of optimal transport. For a given time step $\tau > 0$, a sequence $(\rho^n)_{n=0}^{\infty}$ is constructed inductively: given an approximation ρ^{n-1} of $\rho((n-1)\tau)$, i.e., the solution ρ to (4.1.1) at time $t = (n-1)\tau$, choose as approximation ρ^n of $\rho(n\tau)$ the minimizer of

$$\rho \mapsto \inf_{\gamma} \iint_{\Omega\times\Omega} \mathbf{c}_\tau(x,y)\,\mathrm{d}\gamma(x,y) + \frac{1}{\tau}\big[\mathcal{E}(\rho) - \mathcal{E}(\rho^{n-1})\big]. \tag{4.1.7}$$

Above, the infimum runs over all probability measures $\gamma \in \mathcal{P}(\Omega\times\Omega)$ on the product space $\Omega\times\Omega$ whose first and second marginal, denoted by $X\#\gamma$ and $Y\#\gamma$, respectively, equal to ρ^{n-1} and ρ. Further, $\mathbf{c}_\tau(x,y)$ is the $\mathbf{C}$-induced cost of the transport from x to y in time τ; if Ω is convex, then simply $\mathbf{c}_\tau(x,y) = \mathbf{C}(\frac{y-x}{\tau})$, i.e., $\mathbf{c}_\tau(x,y)$ is the average dissipation induced by the motion of a unit mass element with constant velocity $v = \frac{y-x}{\tau}$. The general definition of $\mathbf{c}_\tau$ is given in Section 4.2.4. In the language of optimal transport, γ is a transport plan from ρ^{n-1} to ρ^n: roughly speaking, $\gamma(x,y)$ determines the amount of ρ^{n-1}'s mass at position x to be moved to ρ^n's mass at position y. The double integral in (4.1.7) is visibly an approximation of the integral in (4.1.6).

The difficulty in the numerical implementation of (4.1.7) is to calculate the infimum of the integral for given ρ^{n-1} and ρ, and its variation with respect to ρ. A common approach is to go to the Lagrangian formulation, using that the optimal γ is typically concentrated on the graph of a transport map $T : \Omega \to \Omega$. This is extremely efficient in one space dimension [6, 32, 31], but becomes significantly more cumbersome — and difficult to analyze — in multiple dimensions [5, 11, 10, 24]. Various alternatives to the Lagrangian approach are available, including finite volume methods [30], blob methods [12] etc.

Here, we use the "lightspeed computation" of the optimal plan γ by employing entropic regularization to the minimization problem. Recall that γ's negative entropy is

$$\mathcal{H}(\gamma) = \iint_{\Omega\times\Omega} G(x,y)\log G(x,y)\,\mathrm{d}(x,y) \tag{4.1.8}$$

if $\gamma = G\mathcal{L}^d$ is absolutely continuous, and $\mathcal{H}(\gamma) = +\infty$ otherwise. Adding this as a regularization inside the dissipation term in (4.1.7), we arrive at the new minimization problem

$$\rho \mapsto \inf_{\gamma} \left[\iint_{\Omega\times\Omega} \mathbf{c}_\tau(x,y)\,\mathrm{d}\gamma(x,y) + \varepsilon\mathcal{H}(\gamma)\right] + \frac{1}{\tau}\big[\mathcal{E}(\rho) - \mathcal{E}(\rho_\tau^{n-1})\big], \tag{4.1.9}$$

$\varepsilon \geq 0$ being the parameter of the regularization. Finally, we discretize the problem (4.1.9) in space by restricting minimization to $\mathcal{P}_\delta(\Omega)$, the set of absolutely continuous ρ's whose densities are piecewise constant on the cells Q of a given tesselation $\mathcal{Q}_\delta$ of Ω; here $\delta > 0$ parametrizes the size of the cells Q, and $\delta \to 0$ is the continuous limit. It is further admissible to approximate $\mathbf{c}_\tau$ by a more convient cost function $\mathbf{c}_{\tau,\delta}$. E.g., in the actual numerical experiments, we use a $\mathbf{c}_{\tau,\delta}$ that is piecewise constant on the products $Q\times Q'$ of cells $Q, Q' \in \mathcal{Q}_\delta$; this makes the minimization feasible in practice since it then suffices to consider only absolutely continuous γ's that are piecewise constant on $Q \times Q'$.

In summary, for given $\varepsilon \geq 0$ and $\delta \geq 0$ — corresponding to a tesselation $\mathcal{Q}_\delta$ and a cost function $\mathbf{c}_{\tau,\delta}$ — a time-discrete approximation $(\rho^n)_{n=0}^{\infty}$ of a solution to (4.1.1) is defined inductively by

$$\rho^n := Y\#\gamma^n, \quad \text{with} \quad \gamma^n := \arg\min \mathcal{E}_{\tau,\varepsilon,\delta}\big(\gamma\big|\rho^{n-1}\big), \tag{4.1.10}$$

where, using the indicator functional ι_Q that is zero if Q is true, and $+\infty$ otherwise,

$$\mathcal{E}_{\tau,\varepsilon,\delta}\big(\gamma\big|\bar\rho\big) = \iint_{\Omega\times\Omega} \mathbf{c}_{\tau,\delta}(x,y)\,\mathrm{d}\gamma(x,y) + \varepsilon\mathcal{H}(\gamma) + \frac{1}{\tau}\mathcal{E}(Y\#\gamma) + \iota_{X\#\gamma=\bar\rho} + \iota_{Y\#\gamma\in\mathcal{P}_\delta(\Omega)}. \tag{4.1.11}$$

4.1.4 Convergence result

Our analytical result concerns the joint limit of infinitely refined spatial discretization $\delta \to 0$ and vanishing entropic regularization $\varepsilon \to 0$.

Theorem 3. *Assume $\Omega = \mathbb{R}^d$, and that ρ^0 has finite second moment. Assume further that $h(r) = r^m$ with some $m > 1$.*

Fix a time step $\tau > 0$, and non-negative sequences (ε_k) and (δ_k) of entropic regularizations and spatial discretizations, respectively, that converge to zero. Under hypotheses on the tesselations $\mathcal{Q}_{\delta_k}$ and cost functions $\mathbf{c}_{\tau_k,\delta_k}$ that are detailed in Section 4.2.5 below, the inductive scheme in (4.1.10)*, with $\varepsilon = \varepsilon_k$ and $\delta = \delta_k$, is well-defined and produces time-discrete approximations $(\rho_k^n)_{n=0}^\infty$ for each k. Moreover, $\rho_k^n \to \rho^n$ narrowly and weakly in $L^m(\mathbb{R}^d)$ as $k \to \infty$, for each n, and $(\rho^n)_{n=0}^\infty$ is a sequence of minimizers in* (4.1.7)*.*

We emphasize that the special cases $\varepsilon_k \equiv 0$ (spatial discretization without entropic regularization) and $\delta_k \equiv 0$ (entropic regularization without spatial discretization) are included. Further, we remark that the choice $\Omega = \mathbb{R}^d$ is mainly made for definiteness; the proof is actually slightly more difficult than in the case of bounded Ω. Also, $h(r) = r^m$ has been chosen to simplify the presentation; the method of proof would apply to any convex $h : [0, \infty) \to \mathbb{R}$ with $h(0) = 0$ that has superlinear growth at infinity.

The proof is based on the Γ-convergence of the functional in (4.1.11) to the one in (4.1.7) without $\mathcal{E}(\rho^{n-1})$, which is made precise in Proposition 1 below. That Γ-limit would be fairly easy to obtain in the situation of regular cost functions, i.e., when $\mathbf{C}$ is a continuous and strictly convex function on all of $\mathbb{R}^d$. In the flux limited situation that we consider here, the construction of the recovery sequence is surprisingly delicate.

We emphasize that we do not consider the passage $\tau \to 0$ from the JKO method (4.1.7) to a solution of the PDE (4.1.1). That kind of limit has been studied extensively, albeit rarely in the flux-limited case. Particularly for L^2-Wasserstein gradient flows, corresponding to $\mathbf{C}(v) = \frac{1}{2}|v|^2$ and to $a(p) = p$, the existing literature is huge, and also covers much more general nonlinearities in (4.1.1) than just $h'(\rho)$. The JKO method has been used to construct solutions to linear and non-linear Fokker-Planck equations [40], to degenerate fourth order parabolic equations [39], to PDEs with non-local terms [6], to coupled systems [27], and many more. There are fewer results on a JKO-like variational approximation of (4.1.1) with a non-linear power functions $a(\xi) = |\xi|^{p-2}\xi$, with $p \neq 2$; this includes in particular the p-Laplace-equations. The corresponding theory of gradient flows in the L^q-Wasserstein metric with $\mathbf{C}(v) = \frac{1}{q}|v|^q$ (with $q = p' \neq 2$) has been developed in [2, 1]. Finally, concerning the situation of interest here, which is (4.1.1) with flux-limitation: the analysis is significantly more challenging in that situation, but still, the limit $\tau \to 0$ has been carried out successfully on the JKO-like variational approximation of the relativistic heat equation in a work of McCann and Puel [37]. The techniques developed therein should apply to the more general class (4.1.1) considered here.

To the best of our knowledge, our result is the first one that rigorously shows the stability of the minimizers in the JKO scheme under entropic regularization. In a related problem, namely for (4.1.1) with $a(\xi) = \xi$, i.e., in the L^2-Wasserstein case, the combined limit of $\tau \to 0$ and $\varepsilon \to 0$ (without spatial discretization, $\delta = 0$) has been carried out by Carlier et al [8]. Also there, the Γ-limit of an entropically regularized transport is studied, however in a different sense, namely for fixed marginals, and for quadratic costs, both of which makes the analysis much easier. We remark further that a joint limit of spatio-temporal refinement has been performed recently [47] for a structurally different fully discrete approximation of the relativistic heat equation in one space dimension, using Lagrangian maps.

4.2 Notations and general hypotheses

Below, we summarize several basic notations and hypotheses, most of which have been mentioned in the introduction in an informal way.

4.2.1 Domains and measures

In the proof of Theorem 3, $\Omega = \mathbb{R}^d$. In the numerical experiments, $\Omega \subset \mathbb{R}^d$ is an open, bounded and connected set with Lipschitz boundary. $\mathcal{L}^d$ is the d-dimensional Lebesgue measure on Ω.

For a measurable subset M of an euclidean space $\mathbb{R}^m$, we denote by $\mathcal{P}(M)$ the affine space of probability measures on M that have finite second moment (which is irrelevant if M is bounded). By abuse of notation, we shall frequently identify absolutely continuous $\mu = \rho\mathcal{L}^d \in \mathcal{P}(M)$ and their Lebesgue-densities $\rho \in L^1(M)$.

For a measurable map $T : M \to M'$, the push-forward $T\#\mu \in \mathcal{P}(M')$ of $\mu \in \mathcal{P}(M)$ is defined via $T\#\mu[A] = \mu[T^{-1}(A)]$ for all measurable sets $A \subset M$. Canonical projections $X, Y : M \times M \to M$ are given by $X(x, y) = x$ and $Y(x, y) = y$. With these notations, the two marginals of $\gamma \in \mathcal{P}(\Omega \times \Omega)$ are given by $X\#\gamma, Y\#\gamma \in \mathcal{P}(\Omega)$, respectively.

The natural notion of convergence in $\mathcal{P}(M)$ is narrow convergence, that is weak convergence as measures in duality to bounded continuous functions $\varphi \in C_b(M)$. For $M = \mathbb{R}^m$, we shall occasionally use a slightly stronger kind of convergence, namely convergence in $\mathbf{W}_2$ (the Wasserstein distance is recalled below), which means narrow convergence plus convergence of the second moment.

4.2.2 Wasserstein distance

The L^2-Wasserstein distance between $\rho_0, \rho_1 \in \mathcal{P}(M)$ is given by

$$\mathbf{W}_2(\rho_0, \rho_1) = \left(\inf_{\gamma \in \mathcal{P}(M \times M)} \left[\iint_{M \times M} |x - y|^2 \,\mathrm{d}\gamma(x, y) + \iota_{X\#\gamma=\rho_0} + \iota_{Y\#\gamma=\rho_1} \right] \right)^{1/2}.$$

The infimum above is actually a minimum, and minimizers γ are called *optimal plans* for the transport from ρ_0 to ρ_1. We use the following fact: if ρ_0 is absolutely continuous, then there exists a measurable $T : M \to M$, called an *optimal map*, such that $T\#\rho_0 = \rho_1$, and

$$\mathbf{W}_2(\rho_0, \rho_1) = \left(\int_M |T(x) - x|^2 \rho_0(x) \,\mathrm{d}x \right)^{1/2}.$$

$\mathbf{W}_2$ is a genuine metric on $\mathcal{P}(M)$. Convergence in $\mathbf{W}_2$ is equivalent to narrow convergence and convergence of the second moment.

4.2.3 Energy functional

By abuse of notation, the definition of $\mathcal{E} : \mathcal{P}(\Omega) \to \mathbb{R} \cup \{+\infty\}$ in (4.1.4) has to be understood in the sense that if $\mu = \rho\mathcal{L}^d$ is absolutely continuous, then $\mathcal{E}(\mu) = \mathcal{E}(\rho)$ is given by the integral, and $\mathcal{E}(\mu) = +\infty$ otherwise. Since h is convex, l.s.c. and super-linear at infinity, the map $\rho \mapsto \int_\Omega h(\rho)\,\mathrm{d}x$ is is lower semi-continuous with respect to narrow convergence.

By $\mathrm{v}, \mathrm{w} \in C_b(\Omega)$ the map $\rho \mapsto \int_\Omega \rho(\mathrm{v} + (\mathrm{w} * \rho))\,\mathrm{d}x$ is continuous w.r.t. narrow convergence. Indeed the part involving v is directly by definition.

For the part involving w we note the map $\gamma \mapsto \iint_{\Omega^2} \mathrm{w}(x-y)\,\mathrm{d}\gamma(x, y)$ is by definition continuous w.r.t. to narrow convergence in $\mathscr{P}(\Omega^2)$ and the map $\rho \mapsto \rho \otimes \rho$ is a map that is continuous as a map from $\mathscr{P}(\Omega)$ to $\mathscr{P}(\Omega^2)$. Here $\rho \otimes \mu \in \mathscr{P}(\Omega^2)$ is the measure defined by $\rho \otimes \mu(A \times B) = \rho(A)\mu(B)$ for all Borel-sets $A, B \subset \Omega$. Consequently $\mathcal{E}$ is lower semi-continuous with respect to narrow convergence.

The methods we present are suited to study general energy functionals of the form (4.1.4) with a smooth and convex function h of superlinerar growth at infinity. In the proof of Theorem 3, we restrict ourselves to $h(r) = r^m$ with $m > 1$ to facilitate readability. In the numerical experiments, we additionally use $h(r) = r(\log r - 1)$.

4.2.4 Derived cost function

We assume that $\mathbf{C} : \mathbb{R}^d \to [0, \infty]$ is strictly convex, continuous and bounded on $\overline{\mathbb{B}}$, and $+\infty$ outside of $\overline{\mathbb{B}}$, with unique minimum $\mathbf{C}(0) = 0$. For technical reasons, we further assume that $\mathbf{C} \equiv 1$ on $\partial\overline{\mathbb{B}}$. Then the gradient of the Legendre dual $\mathbf{C}^\star$ lies in $\overline{\mathbb{B}}$.

The cost function $\mathbf{c} : \Omega \times \Omega \to [0, \infty]$ is derived from $\mathbf{C}$ via

$$\mathbf{c}_\tau(x, y) = \inf\left\{ \frac{1}{\tau}\int_0^\tau \mathbf{C}(\dot{z}(t))\,\mathrm{d}t \,\middle|\, z : [0, \tau] \to \Omega \text{ differentiable, } z(0) = x,\ z(\tau) = y \right\}. \tag{4.2.1}$$

If Ω is convex (e.g., $\Omega = \mathbb{R}^d$), then thanks to the convexity of $\mathbf{C}$,

$$\mathbf{c}_\tau(x, y) = \mathbf{C}\left(\frac{y - x}{\tau}\right). \tag{4.2.2}$$

4.2.5 Spatial discretization

We assume that for each $\delta > 0$, a tesselation $\mathcal{Q}_\delta$ of Ω is given. That is, $\mathcal{Q}_\delta$ consists of finitely (if Ω bounded) or infinite-countably (if $\Omega = \mathbb{R}^d$) many open non-overlapping cells Q such that the union of their closures $\overline{Q}$ cover Ω. We further require that there is a constant $\underline{r} > 0$ such that

$$\operatorname{diam}(Q) \le \sqrt{d}\delta \quad \text{and} \quad |Q| := \mathcal{L}^d(Q) \ge (\underline{r}\delta)^d \quad \text{for all } Q \in \mathcal{Q}_\delta. \tag{4.2.3}$$

A canonical example for $\Omega = \mathbb{R}^d$ is — setting $\underline{r} := 1$ —

$$\mathcal{Q}_\delta = \left\{\delta(\{\mathbf{j}\} + K) \mid \mathbf{j} \in \mathbb{Z}^d\right\} \qquad \text{where} \qquad K := (-\tfrac{1}{2}, \tfrac{1}{2})^d .$$

Accordingly, we define $\mathcal{P}_\delta(\Omega)$ as the space of those $\rho\mathcal{L}^d \in \mathcal{P}(\Omega)$ for which ρ is constant on each $Q_i \in \mathcal{Q}_\delta$. Further, $\mathcal{P}_\delta(\Omega \times \Omega)$ consists of those $\gamma \in \mathcal{P}(\Omega \times \Omega)$ for which $Y\#\mathcal{P} \in \mathcal{P}_\delta(\Omega)$. We emphasize that the condition is only on the y-marginal $Y\#\gamma$, not on the x-marginal $X\#\gamma$, which does not even need to be absolutely continuous. For convenience, we set $\mathcal{P}_0(\Omega) := \mathcal{P}(\Omega)$.

For a probability density $\bar{\rho} \in L^1(\Omega)$, let

$$\Gamma_\delta(\bar{\rho}) = \left\{\gamma \in \mathcal{P}_\delta(\Omega \times \Omega)\,;\, X\#\gamma = \bar{\rho}\mathcal{L}^d\right\}$$

be the subset of measures with $\bar{\rho}\mathcal{L}^d$ as first marginal.

Moreover, we assume that for each $\delta > 0$, a function $\mathbf{c}_{\tau,\delta} : \Omega \times \Omega \to [0, \infty]$ is given that approximates the cost function $\mathbf{c}_\tau$ as follows: there are $\alpha_{\tau,\delta} \in (0, 1)$ with $\alpha_{\tau,\delta} \to 0$ as $\delta \to 0$ for fixed $\tau > 0$, such that

$$|\mathbf{c}_{\tau,\delta}(x, y) - \mathbf{c}_\tau(x, y)| \le \alpha_{\tau,\delta} \quad \text{for } |x - y| \le \tau, \quad \text{and} \tag{4.2.4}$$

$$\mathbf{c}_{\tau,\delta}(x, y) \ge 1 - \alpha_{\tau,\delta} + \frac{1}{\alpha_{\tau,\delta}}\big(|y - x| - \tau\big)^2 \quad \text{for } |x - y| > \tau. \tag{4.2.5}$$

Naturally, one can always take $\mathbf{c}_{\tau,\delta} \equiv \mathbf{c}_\tau$. Note that any $\mathbf{c}_{\tau,\delta}$ with $\mathbf{c}_{\tau,\delta} = +\infty$ on $|x-y| > \tau$ automatically satisfies (4.2.5).

For brevity, we write $\mathbf{c}_k$ for $\mathbf{c}_{\tau,\delta_k}$, and accordingly α_k for the constants α_{τ,δ_k} appearing in (4.2.4)&(4.2.5).

4.3 Proof of Theorem 3

The proof of Theorem 3 immediately follows from a Γ-convergence result that we formulate below.

Proposition 1. *In addition to the hypotheses of Theorem 3, let a sequence $(\rho_k)_{k=1}^\infty$ of densities $\rho_k \in \mathcal{P}_2(\mathbb{R}^d)$ be given such that ρ_k converges in $\mathbf{W}_2$ to some $\rho_* \in \mathcal{P}_2(\mathbb{R}^d)$, and $\sup_k \mathcal{E}(\rho_k) < \infty$. Let furthermore $\delta_k > 0$ be a sequence tending to zero slowly enough such that*

$$\varepsilon_k \log(\delta_k^{-1}) \to 0 \tag{4.3.1}$$

holds. Then the sequence of functionals $\mathcal{E}_k^\tau : \mathcal{P}(\Omega \times \Omega) \to [0, +\infty]$ with, c.f. (4.1.11),

$$\mathcal{E}_k^\tau(\gamma) := \mathcal{E}_{\varepsilon_k,\delta_k,\mathbf{c}_k}^\tau(\gamma|\rho_k)$$

Γ-converges in the narrow topology to $\mathcal{E}_*^\tau : \mathcal{P}(\Omega \times \Omega) \to [0, +\infty]$ with

$$\mathcal{E}_*^\tau(\gamma) = \iint_{\Omega\times\Omega} \mathbf{c}_{\tau,\delta}(x, y)\,\mathrm{d}\gamma(x, y) + \frac{1}{\tau}\mathcal{E}(Y\#\gamma) + \iota_{X\#\gamma=\rho_*}.$$

Moreover, each $\mathcal{E}_k^\tau$ possesses a (unique, if $\varepsilon_k > 0$) minimizer $\hat{\gamma}_k \in \Gamma_{\delta_k}(\rho_k)$, and a subsequence of these minimizers converges in $\mathbf{W}_2$ to a minimizer $\hat{\gamma} \in \Gamma(\rho)$ of $\mathcal{E}^\tau(\cdot; \rho_*)$.

Remark 1. *Note that* (4.3.1)*, which is needed for* 4.4.3. **Limsup condition***, imposes no additional restriction if the tesselation $\mathcal{Q}_\delta$ is made of identical cubes, since then $\Gamma_{\delta'}(\tilde{\rho}) \subseteq \Gamma_\delta(\tilde{\rho})$ if $\delta' > 0$ is an integer multiple of $\delta > 0$, or is arbitrary if $\delta = 0$, — recall that the additional condition induced by δ is only on the Y-marginal, not on the X-marginal — and we can replace (δ_k) by a sequence (δ'_k) with $\delta'_k \geq \delta_k$ that still goes to zero and satisfies* (4.3.1)*, and the recovery sequence $\gamma_k \in \Gamma_{\delta'_k}(\rho_k)$ that we obtain is clearly also a recovery sequence with $\gamma_k \in \Gamma_{\delta_k}(\rho_k)$.*

It is now easy to conclude Theorem 3 by induction on n. Trivially, $\rho_k^0 = \rho^0$ converges to $\rho_*^0 = \rho^0$. Assume that for some $n = 1, 2, \ldots$, there is a (non-relabeled) subsequence $(\rho_k^{n-1})_{k=1}^\infty$ that converges in $\mathbf{W}_2$ and weakly in $L^m(\mathbb{R}^d)$ to a limit ρ_*^{n-1}. That sequence $(\rho_k^{n-1})_{k=1}^\infty$ satisfies the hypotheses of Proposition 1, since weak convergence in $L^m(\mathbb{R}^d)$ implies that $\mathcal{E}(\rho_k^{n-1}) = \|\rho_k^{n-1}\|_{L^m}^m$ remains bounded. Hence the respective functionals $\mathcal{E}_k^\tau$ with $\rho_k := \rho_k^{n-1}$ Γ-converge narrowly to $\mathcal{E}_*^\tau$, with ρ_*^{n-1} in place of ρ_*, and a (non-relabeled) subsubsequence $(\gamma_k^n)_{k=1}^\infty$ of the minimizers converges to a limit γ_*^n in $\mathbf{W}_2$. It is obvious that $\rho_*^n := Y\#\gamma_*^n$ is a minimizer in (4.1.7). It is further obvious that for the subsubsequence under consideration, the convergence of γ_k^n in $\mathbf{W}_2$ is inherited by the marginal ρ_{k-1}^n. Finally, to conclude the weak convergence in $L^m(\mathbb{R}^d)$, possibly after passing to yet another subsequence, observe that the γ_k^{n-1} are minimizers of the respective $\mathcal{E}_k^\tau$, that $\|\rho_k\|_{L^m}^m = \mathcal{E}(\rho_k) \leq \mathcal{E}_k^\tau(\gamma_k^{n-1})$ by definition of $\mathcal{E}_{\varepsilon,\delta,\mathbf{c}}^\tau$, and that $\mathcal{E}_k^\tau$ Γ-converges to $\mathcal{E}_*^\tau$. Alaoglou's theorem allows us to select a subsequence that converges weakly in $L^m(\mathbb{R}^d)$.

The rest of the analytical part of this paper is devoted to proving Proposition 1.

4.4 Proof of Proposition 1

Throughout the proof, let a sequence $(\rho_k)_{k=1}^\infty$ be fixed that satisfies the hypotheses of Proposition 1, i.e., $\rho_k \in \mathcal{P}_2(\mathbb{R}^d)$, $\sup_k \mathcal{E}(\rho_k) < \infty$, and $\rho_k \to \rho_*$ in $\mathbf{W}_2$.

The proof is divided into three steps. First, we prove the *liminf-condition* for Γ-convergence: if $\gamma_k \in \Gamma_{\delta_k}(\rho_k)$ converges to $\gamma_* \in \Gamma(\rho_*)$ narrowly, then

$$\mathcal{E}_*^\tau(\gamma_*) \leq \liminf_{k\to\infty} \mathcal{E}_k^\tau(\gamma_k). \tag{4.4.1}$$

Second, and by far more difficult, is the construction of a *recovery sequence*: if $\gamma_* \in \Gamma(\rho_*)$ is such that $\mathcal{E}_*^\tau(\gamma_*) < \infty$, then there are $\gamma_k \in \Gamma_{\delta_k}(\rho_k)$ such that $\gamma_k \to \gamma_*$ narrowly, and

$$\mathcal{E}_*^\tau(\gamma_*) \geq \limsup_{k\to\infty} \mathcal{E}_k^\tau(\gamma_k). \tag{4.4.2}$$

These two steps together verify the Γ-convergence of the $\mathcal{E}_k^\tau$. In particular, it follows that if $\hat{\gamma}_k$ are minimizers of the $\mathcal{E}_k^\tau$ which converge to $\hat{\gamma} \in \Gamma(\rho_*)$, then $\hat{\gamma}$ is a minimizer of $\mathcal{E}_*^\tau$. Now, in the final step, we verify that each $\mathcal{E}_k^\tau$ actually possesses a minimizer $\hat{\gamma}_k \in \Gamma_{\delta_k}(\rho_k)$, and that a subsequence of those converges narrowly to a limit $\hat{\gamma} \in \Gamma(\rho_*)$, which then is necessarily a minimizer of $\mathcal{E}^\tau(\cdot|\rho_*)$.

4.4.1 Preliminary results

Before starting with the core of the proof, we draw two immediate conclusions from the hypotheses stated above.

Lemma 1. *The ρ_k have k-uniformly bounded second moments, and $\int_{\mathbb{R}^d} \rho_k(x)\log\rho_k(x)\,\mathrm{d}x$ is k-uniformly bounded from above and below.*

Proof. By hypothesis, ρ_k converges to ρ_* in $\mathbf{W}_2$, which implies in particular the convergence of ρ_k's second moment to the one of ρ_*. Boundedness of the integral is obtained by means of a classical estimate: first, observe that $r\log r \geq -\frac{d+1}{e} r^{\frac{d}{d+1}}$ for all $r > 0$. By Hölder's inequality, it follows that

$$\begin{aligned}
\int_{\mathbb{R}^d} \rho(x)\log\rho(x)\,\mathrm{d}x &\geq -\frac{d+1}{e}\int_{\mathbb{R}^d} \rho(x)^{\frac{d}{d+1}}\,\mathrm{d}x \\
&\geq -\frac{d+1}{e}\left(\int_{\mathbb{R}^d} \frac{\mathrm{d}x}{\big(1+|x|^2\big)^d}\right)^{\frac{1}{d+1}} \left(\int_{\mathbb{R}^d} \rho(x)\big(1+|x|^2\big)\,\mathrm{d}x\right)^{\frac{d}{d+1}},
\end{aligned}$$

which yields a finite lower bound that only depends on the second moment of ρ_k. An upper bound easily follows from the k-uniform boundedness of $\mathcal{E}(\rho_k)$ and the fact that $r\log r \leq \frac{1}{(m-1)e} r^m$ for all $r > 0$. □

For the next result, recall that $\alpha_k = \alpha_{\tau,\delta_k}$ are the quantities that appear in conditions (4.2.4)&(4.2.5).

Lemma 2. *There is a constant C such that — uniformly for all k large enough — the second moment of each $\gamma \in \Gamma(\rho_k)$ is controlled via*

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \big(|x|^2 + |y|^2\big)\,\mathrm{d}\gamma(x,y) \leq C\big(1 + \alpha_k \mathcal{E}_k^\tau(\gamma)\big), \tag{4.4.3}$$

and $\mathcal{E}_k^\tau$ is bounded from below as follows,

$$\mathcal{E}_k^\tau(\gamma) \geq (\tau - C\alpha_k\varepsilon_k)\iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{c}_k\,\mathrm{d}\gamma + \mathcal{E}(Y\#\gamma). \tag{4.4.4}$$

In particular, $\mathcal{E}_k^\tau$ is non-negative for all sufficiently large k such that $C\alpha_k\varepsilon_k \leq \tau$.

Proof. On the one hand, with the same idea as in the proof of Lemma 1 above, we find for every $\gamma = G\mathcal{L}^d \otimes \mathcal{L}^d$ that

$$\mathcal{H}(\gamma) \geq -C_d\left(1 + \iint_{\mathbb{R}^d\times\mathbb{R}^d} \big(|x|^2 + |y|^2\big)\,\mathrm{d}\gamma(x,y)\right),$$

where

$$C_d := \frac{2d+1}{e}\left(\iint_{\mathbb{R}^d\times\mathbb{R}^d} \frac{\mathrm{d}(x,y)}{\left(1+|x|^2+|y|^2\right)^{2d}}\right)^{\frac{1}{2d+1}}$$

is a finite constant that only depends on d. On the other hand, using hypothesis (4.2.5) on $\mathbf{c}_k$, it follows that

$$\begin{aligned}\iint_{\mathbb{R}^d\times\mathbb{R}^d} |y|^2\,\mathrm{d}\gamma(x,y) &\le \iint_{\mathbb{R}^d\times\mathbb{R}^d} \left[|x| + \left(|y-x|-\tau\right)\mathbf{1}_{|y-x|\ge\tau} + \tau\right]^2 \mathrm{d}\gamma(x,y)\\ &\le 2\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x|^2\,\mathrm{d}\gamma(x,y) + 4\tau^2 \iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathrm{d}\gamma(x,y) + 4\iint_{|y-x|\ge\tau} \left(|x-y|-\tau\right)^2 \mathrm{d}\gamma(x,y)\\ &\le 2\int_{\mathbb{R}^d} |x|^2\rho_k(x)\,\mathrm{d}x + 4\tau^2 + 4\alpha_k \iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{c}_k\,\mathrm{d}\gamma,\end{aligned}$$

which yields

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \left(|x|^2+|y|^2\right)\mathrm{d}\gamma(x,y) \le 4\left[\tau^2 + \int_{\mathbb{R}^d} |x|^2\rho_k(x)\,\mathrm{d}x + \alpha_k \iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{c}_k\,\mathrm{d}\gamma\right]. \tag{4.4.5}$$

In view of Lemma 1 above, the second moment of ρ_k is uniformly controlled, and therefore

$$\mathcal{H}(\gamma) \ge -C\left(1 + \alpha_k \iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{c}_k\,\mathrm{d}\gamma\right), \tag{4.4.6}$$

with a k-independent C. This induces the bound (4.4.4). The other bound (4.4.3) follows for all k such that, say, $C\alpha_k\varepsilon_k \le \tau/2$, by re-inserting (4.4.4) into (4.4.5) and using once again the uniform bound on ρ_k's second moment. □

4.4.2 Liminf condition

Proposition 2. *Assume that a sequence of measures $\gamma_k \in \Gamma(\rho_k)$ converges narrowly to $\gamma_* \in \Gamma(\rho)$ Then (4.4.1) holds.*

Proof. Recall from Lemma 2 that $\mathcal{E}_k^\tau$ is non-negative for k large enough. And if $\mathcal{E}_k^\tau(\gamma_k) \to +\infty$, there is nothing to prove. Hence, it suffices to consider a sequence (γ_k) such that $\mathcal{E}_k^\tau(\gamma_k)$ converges to a finite value. From (4.4.4), one directly concludes k-uniform boundedness of $\iint \mathbf{c}_k\,\mathrm{d}\gamma_k$. Thanks to the bound (4.2.5) on $\mathbf{c}_k$, it follows for every $t>0$ that γ_k's mass in $|x-y| \ge \tau + t$ goes to zero as $k\to\infty$. Thus, γ_* is supported in $|x-y|\le\tau$.

Define the continuous function $\tilde{\mathbf{c}} : \mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}$ by $\hat{\mathbf{c}}(x,y) = \mathbf{c}(x,y)$ for $|x-y|\le\tau$, and $\hat{\mathbf{c}} \equiv 1$ otherwise. From (4.2.4) and (4.2.5) it is clear that $\mathbf{c}_k \ge \hat{\mathbf{c}} - \alpha_k$, and so

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{c}_k\,\mathrm{d}\gamma_k \ge \iint_{\mathbb{R}^d\times\mathbb{R}^d} \left(\hat{\mathbf{c}} - \alpha_k\right)\mathrm{d}\gamma_k = \iint_{\mathbb{R}^d\times\mathbb{R}^d} \hat{\mathbf{c}}\,\mathrm{d}\gamma_k - \alpha_k \overset{k\to\infty}{\longrightarrow} \iint_{\mathbb{R}^d\times\mathbb{R}^d} \hat{\mathbf{c}}\,\mathrm{d}\gamma_* = \iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{c}\,\mathrm{d}\gamma_*.$$

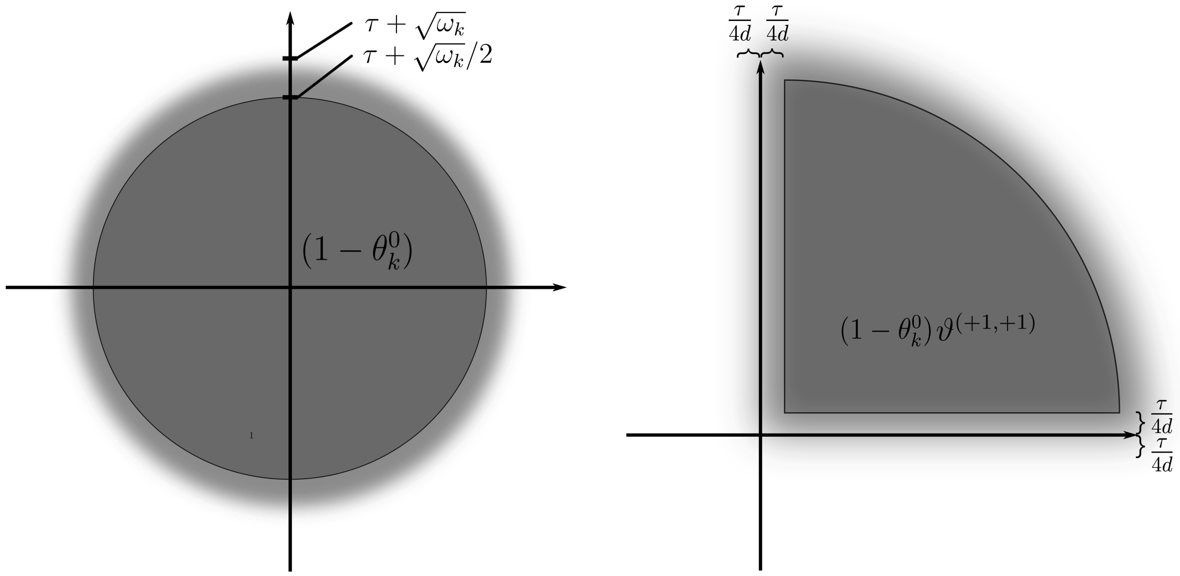
So, by (4.4.4),

$$\liminf_{k\to\infty} \mathcal{E}_k^\tau(\gamma_k) \ge \tau \iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{c}\,\mathrm{d}\gamma_* + \liminf_{k\to\infty} \mathcal{E}(Y\#\gamma_k). \tag{4.4.7}$$

Finally, since the projection Y is a continuous map, the push-forwarded measure $Y\#\gamma_k$ converges narrowly to $Y\#\gamma_*$, and since $r\mapsto r^m$ is a convex function and v, w are bounded continuous functions, it follows that

$$\liminf_{k\to\infty} \mathcal{E}(Y\#\gamma_k) \ge \mathcal{E}(Y\#\gamma_*),$$

so the sum on the right-hand side of (4.4.7) is greater or equal to $\mathcal{E}^\tau(\gamma_*|\rho_*)$. □



(a) The density of $1-\theta_k^0$ in $\mathbb{R}^2$.

(b) The density of $\vartheta^\beta(1-\theta_k^0)$ in $\mathbb{R}^2$ for $\beta=(+1,+1)$.

Figure 4.1: The two smoothed indicator functions used in Step 2 to cup up the transport map $\gamma_k^{(1)}$ displayed for $d=2$. Note that the set, on which both have density 1 has been emphasized by an additional black border and a small step in the grayscale.

4.4.3 Limsup condition

Proposition 3. *For every $\gamma_*\in\Gamma(\rho_*)$ with $\mathcal{E}^\tau(\gamma_*|\rho_*)<\infty$, there exists a sequence of $\gamma_k\in\Gamma_{\delta_k}(\rho_k)$ such that $\gamma_k\to\gamma_*$ narrowly, and* (4.4.2) *holds.*

For future reference, define $\eta_*:=Y\#\gamma_*$. From our hypothesis $\mathcal{E}(Y\#\gamma_*)<\infty$ together with v, w being bounded it follows that $\eta_*\in L^m(\mathbb{R}^d)$.

Construction of the recovery sequence

In the following, let $k=1,2,\dots$ be fixed. We are going to construct $\gamma_k\in\Gamma_{\delta_k}(\rho_k)$ in several steps.

Step 1: *Modify γ_* into $\gamma^{(1)}$ such that $X\#\gamma_k^{(1)}=\rho_k\mathcal{L}^d$ and $Y\#\gamma_k^{(1)}=\eta_*\mathcal{L}^d$.*

To that end, let $T_k:\mathbb{R}^d\to\mathbb{R}^d$ be an optimal map for the transport from ρ_* to ρ_k in $\mathbf{W}_2$; such a map exists since ρ_* is a probability density, and both ρ_k and ρ_* have finite second moment. Then $\gamma_k^{(1)}:=(T_k\circ X,Y)\#\gamma_*$ has the desired marginals. For later use, define

$$\omega_k:=\left(\int_{\mathbb{R}^d}|T_k(x)-x|^2\rho_*(x)\,\mathrm{d}x\right)^{\frac{1}{2}}=\mathbf{W}_2(\rho_*,\rho_k), \tag{4.4.8}$$

which goes to zero by our hypothesis that ρ_k converges to ρ_* in $\mathbf{W}_2$.

Step 2: *Decompose $\gamma_k^{(1)}$ into the sum of 2^d non-negative measures $\gamma_k^{(2,\beta)}$ — each of which fits into the cylinder $|x-y|\le\tau$ after proper translation — and a remainder $\gamma_k^{(2,0)}$ of small mass.*

This is done with the help of several cut-off functions that we define now: for each $\beta\in\{+1,-1\}^d$, choose $\vartheta^\beta\in C^\infty(\mathbb{R}^d)$ such that

- $0 \le \vartheta^\beta \le 1$ and
$$\sum_\beta \vartheta^\beta = 1 \quad \text{on } \mathbb{R}^d,$$
- ϑ^β is supported on the set where $\beta_j x_j \ge -\frac{\tau}{4d}$ for all $j = 1, \dots, d$.

Thus, each ϑ^β is essentially a smoothed indicator function for one of the 2^d orthants in $\mathbb{R}^d$. The label β corresponds the signs of the d coordinates in the respective orthant. Next, let $\theta_k^0 \in C^\infty(\mathbb{R}^d)$ be a smoothed indicator function of the complement of the closed ball $\overline{\mathbb{B}}_\tau$ of radius τ with the following properties:

- $0 \le \theta_k^0 \le 1$ and $|\nabla \theta_k^0| \le 3\omega_k^{-1/2}$,
- θ_k^0 vanishes on $\overline{\mathbb{B}}_{\tau+\sqrt{\omega_k}/2}$, and is identical to one on the complement of $\overline{\mathbb{B}}_{\tau+\sqrt{\omega_k}}$.

Now define $\theta_k^\beta := \vartheta^\beta(1-\theta_k^0)$ for all $\beta \in \{-1,+1\}^d$, which are smoothed indicator functions of the sectors of the ball $\overline{\mathbb{B}}_\tau$ corresponding to the respective β-orthant (c.f. *Figure* **??** *(b)*). Note that

$$\theta_k^0 + \sum_\beta \theta_k^\beta = 1 \quad \text{on } \mathbb{R}^d. \tag{4.4.9}$$

For brevity, introduce further $\Theta_k^\beta(x,y) = \theta_k^\beta(x-y)$ as well as $\Theta_k^0(x,y) = \theta_k^0(x-y)$, and define

$$\gamma_k^{(2,\beta)} := \Theta_k^\beta \gamma_k^{(1)}, \quad \gamma^{(2,0)} := \Theta_k^0 \gamma_k^{(1)}.$$

From (4.4.9), it follows that

$$\gamma_k^{(2,0)} + \sum_\beta \gamma_k^{(2,\beta)} = \gamma_k^{(1)}. \tag{4.4.10}$$

Roughly speaking, $\gamma^{(2,0)}$ contains the part of $\gamma^{(1)}$ corresponding to transport with speed that exceeds — by $\sqrt{\omega_k}/\tau$ or more — the limit set by the flux limitation. The part $\gamma^{(2,\beta)}$ corresponds to transport that either respects the flux limitation, or violates it by — no more than $\sqrt{\omega_k}/\tau$ — in the β-directions.

Step 3a: *Translate each of the* $\gamma_k^{(2,\beta)}$ *in* y*-direction to obtain a* $\gamma^{(3,\beta)}$ *that fits in the cylinder* $|x-y| \le \tau - \delta_k$.

With

$$\sigma_k := 12\big(\delta_k + \sqrt{\omega_k}\big),$$

we define $\gamma_k^{(3,\beta)} := (X, Y - \sigma_k\beta)\#\gamma_k^{(2,\beta)}$. The fact that $\gamma_k^{(3,\beta)}$ is supported in the aforementioned cylinder is not completely obvious, and is verified in Lemma 5 below.

Step 3b: *From the remainder* $\gamma_k^{(2,0)}$*, define a measure* $\gamma_k^{(3,0)}$*, which has the same first marginal as* $\gamma_k^{(2,0)}$ *and a smooth second marginal, and is supported in the cylinder* $|x-y| \le \tau/2$.

Let λ be a some smooth probability density on $\mathbb{R}^d$ with support in $\overline{\mathbb{B}}_{\tau/2}$. Consider the product measure $\gamma_k^{(2,0)} \otimes \lambda$ on $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$. With $(X, X+Z)$ being the map $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \ni (x,y,z) \mapsto (x, x+z) \in \mathbb{R}^d \times \mathbb{R}^d$, one easily sees that $\gamma_k^{(3,0)} := (X, X+Z)\#(\gamma_k^{(2,0)} \otimes \lambda)$ has the desired properties. Intuitively, on each vertical fiber $\{x\} \times \mathbb{R}^d$, one redistributes the disintegrated mass of $\gamma^{(2,0)}$ in a smooth way around the point $y = x$.

In summary of Steps 1–3, define

$$\gamma_k^{(3)} := \gamma_k^{(3,0)} + \sum_\beta \gamma_k^{(3,\beta)}.$$

Step 4: *Project $\gamma_k^{(3)} \in \Gamma(\rho_k)$ onto a $\gamma_k \in \Gamma_{\delta_k}(\rho_k)$.*

For each $Q \in \mathcal{Q}_{\delta_k}$, consider the Borel measure γ_k^Q on $\mathbb{R}^d$ defined by $\gamma_k^Q(A) := \gamma_k^{(3)}(A \times Q)$ for each measurable $A \subset \mathbb{R}^d$. Since

$$\sum_{Q \in \mathcal{Q}_{\delta_k}} \gamma_k^Q = X\#\gamma_k^{(3)} = \rho_k \mathcal{L}^d, \tag{4.4.11}$$

it follows that γ_k^Q possesses a non-negative Lebesgue density $g_k^Q \in L^1(\mathbb{R}^d)$. From the g_k^Q, we define a probability density function $G_k \in L^1(\mathbb{R}^d \times \mathbb{R}^d)$ via

$$G_k(x, y) := \frac{g_k^Q(x, y)}{|Q|} \quad \text{where } Q \in \mathcal{Q}_{\delta_k} \text{ is chosen such that } y \in Q. \tag{4.4.12}$$

Our final definition of the recovery sequence is $\gamma_k := G_k \mathcal{L}^d \otimes \mathcal{L}^d$.

Properties of the recovery sequence

We prove various properties of the sequence (γ_k) that eventually allow to conclude (4.4.2).

Lemma 3. $\gamma_k \in \Gamma_{\delta_k}(\rho_k)$. *Moreover, its second moment is k-uniformly bounded.*

Proof. This is essentially clear from the construction.

First, γ_k is a probability measure since the construction is a combination of push-forwards (Steps 1 and 3), decomposition into a finite sum of non-negative measures (Step 2), re-arrangement of these components (Step 3), and finally a projection (Step 4), each of which is easily checked to preserve non-negativity and total mass of the measure.

Second, the X-marginal of γ_k is $\rho_k \mathcal{L}^d$, since Step 1 is made such that $X\#\gamma^{(1)} = T_k\#(X\#\gamma_*) = T_k\#(\rho_* \mathcal{L}^d) = \rho_k \mathcal{L}^d$, and all further steps keep the X-marginal fixed.

Third, γ_k has finite and, in fact, even k-uniformly bounded second moment. Indeed, since γ_k is supported in $|x - y| \le \tau$ (which follows from the purely geometric considerations in Lemma 5 below), one has γ_k-a.e. that

$$|y|^2 = |x + (y - x)|^2 \le 2|x|^2 + 2\tau^2$$

and therefore, recalling that γ_k has X-marginal $\rho_k \mathcal{L}^d$,

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} \left(|x|^2 + |y|^2\right) \mathrm{d}\gamma_k \le \iint_{\mathbb{R}^d \times \mathbb{R}^d} \left(3|x|^2 + 2\tau^2\right) \mathrm{d}\gamma_k = 2\tau^2 + 3 \int_{\mathbb{R}^d} |x|^2 \rho_k(x)\, \mathrm{d}x.$$

The last expression is finite, and is even k-uniformly bounded since the same is true for ρ_k's second moment, see Lemma 1. □

Lemma 4. *There is a constant C such that*

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} G_k(x, y) \log G_k(x, y)\, \mathrm{d}(x, y) \le C + d \log(\delta_k^{-1}).$$

Consequently, $\varepsilon_k \mathcal{H}(\gamma_k) \to 0$.

Proof. By definition of G_k,

$$\begin{aligned}\iint_{\mathbb{R}^d\times\mathbb{R}^d} G_k(x,y)\log G_k(x,y)\,\mathrm{d}(x,y) &= \sum_{Q\in\mathcal{Q}_{\delta_k}} \iint_{\mathbb{R}^d\times Q} \left(\frac{g_k^Q(x)}{|Q|}\right)\log\left(\frac{g_k^Q(x)}{|Q|}\right)\mathrm{d}(x,y)\\ &= \sum_{Q\in\mathcal{Q}_{\delta_k}}\left[\int_{\mathbb{R}^d} g_k^Q(x)\log g_k^Q(x)\,\mathrm{d}x - \log(|Q|)\int_{\mathbb{R}^d} g_k^Q(x)\,\mathrm{d}x\right]\\ &\leq \int_{\mathbb{R}^d}\rho_k(x)\log\rho_k(x)\,\mathrm{d}x - d\log(\delta_k)\int_{\mathbb{R}^d}\rho_k(x)\,\mathrm{d}x,\end{aligned}$$

where we have estimated $|Q|\geq\delta_k^d$ on grounds of (4.2.3), and have used (4.4.11) in combination with the superadditivity of the function $s\mapsto s\log s$, that is,

$$a\log a + b\log b \leq (a+b)\log(a+b) \quad \text{for arbitrary } a,b\geq 0.$$

The latter is an immediate consequence of the monotonicity of the logarithm. Recalling Lemma 1 and our assumption (4.3.1), the convergence follows. □

Lemma 5. *For all k large enough, the γ_k are supported in $|x-y|\leq\tau$.*

Proof. The main step is to show that the measures $\gamma_k^{(3)}$ are supported in $|x-y|\leq\tau-\delta_k$. The function θ_k^β is supported on the set

$$S^\beta := \left\{y\in\mathbb{R}^d\,;\,\beta_j y_j\geq -\frac{\tau}{4d}\ \text{for all } j,\ |y|\leq\tau+\sqrt{\omega_k}\right\}.$$

We show that the translate $S^\beta-\sigma_k\beta$ is a subset of $\overline{\mathbb{B}}_{\tau-\delta_k}$. Observe that S^β is the convex hull of the point $o^\beta := -\frac{\tau}{4d}\beta$ and the spherical cap

$$\mathfrak{S}^\beta = \left\{y\in\mathbb{R}^d\,;\,\beta_j y_j\geq -\frac{\tau}{4d}\ \text{for all } j,\ |y|=\tau+\sqrt{\omega_k}\right\}.$$

Since $\overline{\mathbb{B}}_{\tau-\sqrt{\omega_k}}$ is convex, it thus suffices to verify that the translate of o^β, i.e., the point $-\left(\frac{\tau}{4d}+\sigma_k\right)\beta$, and the translate of the cap, i.e., $\mathfrak{S}^\beta-\sigma_k\beta$, belong to $\overline{\mathbb{B}}_{\tau-\delta_k}$. For all k large enough so that $\sigma_k\leq\frac{\tau}{4d}$, the claim $-\left(\frac{\tau}{4d}+\sigma_k\right)\beta\in\overline{\mathbb{B}}_{\tau-\delta_k}$ is obvious. To prove that also $\mathfrak{S}^\beta\subset\overline{\mathbb{B}}_{\tau-\delta_k}$, consider an arbitrary point $x\in\mathfrak{S}^\beta$. Observing that

$$\beta\cdot x = \sum_j\beta_j x_j \geq \sum_j\left(|x_j|-\frac{\tau}{2d}\right) = \sum_j|x_j| - \frac{\tau}{2} \geq \tau+\sqrt{\omega_k}-\frac{\tau}{2}\geq\frac{\tau}{2},$$

it follows that

$$|x-\sigma_k\beta|^2 = |x|^2+\sigma_k^2|\beta|^2-2\sigma_k\beta\cdot x \leq \left(\tau+\sqrt{\omega_k}\right)^2 + d\sigma_k^2-\tau\sigma_k.$$

Recall that k is large enough such that $\sigma_k\leq\frac{\tau}{4d}$; on the one hand, this yields that

$$d\sigma_k^2-\tau\sigma_k\leq -\frac{3}{4}\tau\sigma_k,$$

and on the other hand, we obtain

$$\left(\tau+\sqrt{\omega_k}\right)^2-\left(\tau-\delta_k\right)^2 = \left(2\tau+\sqrt{\omega_k}-\delta_k\right)\left(\sqrt{\omega_k}+\delta_k\right)\leq 3\tau\left(\sqrt{\omega_k}+\delta_k\right)\leq\frac{\tau}{4}\sigma_k.$$

In summary, we conclude that

$$|x - \sigma_k \beta|^2 \leq \big(\tau - \delta_k\big)^2,$$

which verifies that $\mathfrak{S}^\beta - \sigma_k\beta \subset \overline{\mathbb{B}}_{\tau-\delta_k}$.

By definition, $\gamma_k^{(2,\beta)}$ is supported in the region where $y - x \in S^\beta$. Its translate $\gamma_k^{(3,\beta)} = (X, Y - \sigma_k\beta)\#\gamma_k^{(2,\beta)}$ is therefore supported where $y - x \in S^\beta - \sigma_k\beta \subset \overline{\mathbb{B}}_{\tau-\delta_k}$, where the inclusion is a consequence of the considerations above.

This proves that each $\gamma_k^{(3)}$ is supported in $|x-y| \leq \tau - \delta_k$. From the construction of γ_k it is clear that $\operatorname{supp}\gamma_k$ intersects $\{x\} \times Q$ for some $x \in \mathbb{R}^d$ and $Q \in \mathcal{Q}_{\delta_k}$ only if $\operatorname{supp}\gamma_k^{(3)}$ intersects $\{x\} \times Q$. Since the distance of two points in Q is less than δ_k by (4.2.3), it follows that γ_k is supported in $|x - y| \leq \tau$. □

Lemma 6. *$\gamma_k^{(2,0)}$'s total mass does not exceed $4\omega_k$.*

Proof. Recall that $|x-y| \leq \tau$ for γ_*-a.e. (x, y). Hence $|T_k(x) - y| \geq \tau + \sqrt{\omega_k}/2$ implies that $|T_k(x) - x| \geq \sqrt{\omega_k}/2$ for γ_*-a.e. (x, y). Consequently, recalling that $\gamma_k^{(2,0)} = \Theta_k^0\,(T_k \circ X, Y)\#\gamma_*$:

$$\begin{aligned}
\gamma_k^{(2,0)}[\mathbb{R}^d \times \mathbb{R}^d] &= \iint_{\mathbb{R}^d\times\mathbb{R}^d} \theta_k^0(T_k(x) - y)\,\mathrm{d}\gamma_*(x,y) \\
&\leq \iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{1}_{|T_k(x)-y|\geq\tau+\sqrt{\omega_k}/2}\,\mathrm{d}\gamma_*(x,y) \\
&\leq \iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{1}_{|T_k(x)-x|\geq\sqrt{\omega_k}/2}\,\mathrm{d}\gamma_*(x,y) \\
&= \int_{\mathbb{R}^d} \mathbf{1}_{|T_k(x)-x|^2\geq\omega_k/4}\,\rho_k(x)\,\mathrm{d}x \\
&\leq \frac{4}{\omega_k}\int_{\mathbb{R}^d} |T_k(x) - x|^2\rho_k(x)\,\mathrm{d}x = 4\omega_k,
\end{aligned}$$

where we have used the definition (4.4.8) of ω_k in the last step. □

Lemma 7. *γ_k converges narrowly to γ_*, and moreover,*

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{c}_k\,\mathrm{d}\gamma_k \to \iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{c}\,\mathrm{d}\gamma_*. \tag{4.4.13}$$

Proof. To start with, we show that $\gamma_k^{(1)}$ converges to γ_* narrowly. Since both each $\gamma_k^{(1)}$ and the proposed limit γ_* are probability measures, it suffices to show convergence in distribution, i.e., for all test functions $\psi \in C_c^\infty(\mathbb{R}^d \times \mathbb{R}^d)$. Since $\omega_k \to 0$ in (4.4.8), it follows that T_k converges to the identity map in measure with respect to ρ_*, and hence also $(T_k \circ X, Y)$ converges to (X, Y) in measure with respect to γ_*. And — ψ being smooth and compactly supported — $\psi(T_k \circ X, Y)$ converges to ψ in measure with respect to γ_*. By the dominated convergence theorem,

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi\,\mathrm{d}\gamma_k^{(1)} = \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(T_k \circ X, Y)\,\mathrm{d}\gamma_* \to \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi\,\mathrm{d}\gamma_*.$$

Next, we show that also $\gamma_k^{(3)}$ converges to γ_*:

$$\begin{aligned}\iint_{\mathbb{R}^d\times\mathbb{R}^d}\psi(x,y)\,\mathrm{d}\gamma_k^{(3)}(x,y) &= \iint_{\mathbb{R}^d\times\mathbb{R}^d}\psi(x,y)\,\mathrm{d}\gamma_k^{(3,0)}(x,y) + \sum_\beta\iint_{\mathbb{R}^d\times\mathbb{R}^d}\psi(x,y)\,\mathrm{d}\gamma_k^{(3,\beta)}(x,y)\\
&= \iint_{\mathbb{R}^d\times\mathbb{R}^d}\left[\int_{\mathbb{R}^d}\psi(x,x+z)\lambda(z)\,\mathrm{d}z\right]\mathrm{d}\gamma_k^{(2,0)}(x,y)\\
&\quad + \sum_\beta\iint_{\mathbb{R}^d\times\mathbb{R}^d}\psi(x,y-\sigma_k\beta)\,\mathrm{d}\gamma_k^{(2,\beta)}(x,y)\\
&= \iint_{\mathbb{R}^d\times\mathbb{R}^d}\left[\int_{\mathbb{R}^d}\psi(x,x+z)\lambda(z)\,\mathrm{d}z - \sum_\beta\psi(x,y-\sigma_k\beta)\vartheta^\beta(x-y)\right]\mathrm{d}\gamma_k^{(2,0)}(x,y) \qquad (4.4.14)\\
&\quad + \sum_\beta\iint_{\mathbb{R}^d\times\mathbb{R}^d}\psi(x,y-\sigma_k\beta)\vartheta^\beta(x-y)\,\mathrm{d}\gamma_k^{(1)}(x,y). \qquad (4.4.15)\end{aligned}$$

Here we have used that, by definition of $\gamma^{(2,\beta)}$ from $\gamma^{(1)}$ in Step 2,

$$\mathrm{d}\gamma_k^{(2,\beta)} = \vartheta^\beta(x-y)\big(1-\theta_k^0(x-y)\big)\,\mathrm{d}\gamma_k^{(1)}(x,y) = \vartheta^\beta(x-y)\,\mathrm{d}\gamma_k^{(1)}(x,y) - \vartheta^\beta(x-y)\,\mathrm{d}\gamma_k^{(2,0)}(x,y).$$

The integral in (4.4.14) converges to zero thanks to Lemma 6; observe that the expression inside the square brackets is a continuous function that is bounded independently of k. Concerning the sum in (4.4.15), observe that $\psi(x,y-\sigma_k\beta)\to\psi(x,y)$ uniformly in (x,y) since ψ is compactly supported, and recall from above that $\gamma_k^{(1)}$ converges to γ_* narrowly. This suffices to conclude that

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d}\psi(x,y)\,\mathrm{d}\gamma_k^{(3)}(x,y)\to\sum_\beta\iint_{\mathbb{R}^d\times\mathbb{R}^d}\psi(x,y)\vartheta^\beta(x-y)\,\mathrm{d}\gamma_*(x,y) = \iint_{\mathbb{R}^d\times\mathbb{R}^d}\psi(x,y)\,\mathrm{d}\gamma_*(x,y),$$

where we have used that the smooth expressions $\vartheta^\beta(x-y)$ sum up to unity on the support of γ_*.

As the last step, we show that γ_k converges to γ_* as well. For each $Q\in\mathcal{Q}_{\delta_k}$, define $\Psi_k^Q\in C_c(\mathbb{R}^d)$ by

$$\Psi_k^Q(x) = \frac{1}{|Q|}\int_Q\psi(x,y)\,\mathrm{d}y.$$

Note that there is one common compact set on which all the Ψ_k^Q are supported. From the definition of γ_k, it follows that

$$\begin{aligned}\iint_{\mathbb{R}^d\times\mathbb{R}^d}\psi(x,y)\,\mathrm{d}\gamma_k(x,y) &= \sum_{Q\in\mathcal{Q}_{\delta_k}}\int\Psi^Q(x)g_k^Q(x)\,\mathrm{d}x = \sum_{Q\in\mathcal{Q}_{\delta_k}}\iint_{\mathbb{R}^d\times Q}\Psi^Q(x)\,\mathrm{d}\gamma_k^{(3)}(x,y)\\
&= \iint_{\mathbb{R}^d\times\mathbb{R}^d}\psi(x,y)\,\mathrm{d}\gamma_k^{(3)}(x,y) + \sum_{Q\in\mathcal{Q}_{\delta_k}}\iint_{\mathbb{R}^d\times Q}\big[\Psi^Q(x)-\psi(x,y)\big]\,\mathrm{d}\gamma_k^{(3)}(x,y).\end{aligned}$$

Now since the term in square brackets converges uniformly to zero as the mesh is refined, and since $\gamma_k^{(3)}$ converges to γ_* narrowly, distributional — and subsequently narrow — convergence of γ_k to γ_* follows.

Finally, in combination with the fact that — thanks to Lemma 5 — all the γ_k are supported inside $|x-y|\le\tau$, where $\mathbf{c}_k$ converges to $\mathbf{c}$ uniformly by hypothesis (4.2.4), the claimed convergence (4.4.13) is proven. ☐

Lemma 8. *$Y\#\gamma_k^{(3)}$ has a Lebesgue density $\eta_k^{(3)}\in L^m(\mathbb{R}^d)$, and $\eta_k^{(3)}\to\eta_*$ in $L^m(\mathbb{R}^d)$.*

Proof. By Step 3b, $Y\#\gamma_k^{(3,0)} = \eta_k^{(3,0)}\mathcal{L}^d$ for a smooth density $\eta_k^{(3,0)} \in L^1 \cap L^\infty(\mathbb{R}^d)$. Moreover, for each $\beta \in \{-1,+1\}^d$, the marginal $Y\#\gamma_k^{(3,\beta)}$ is a translate of $Y\#\gamma^{(2,\beta)}$, and from (4.4.10) it follows that

$$Y\#\gamma_k^{(2,0)} + \sum_\beta Y\#\gamma_k^{(2,\beta)} = Y\#\gamma_k^{(1)} = \eta_*\mathcal{L}^d,$$

hence $Y\#\gamma^{(3,\beta)}$ has a Lebesgue density

$$\eta_k^{(3,\beta)} \leq \eta_*(\cdot + \sigma_k\beta) \in L^1 \cap L^m(\mathbb{R}^d). \tag{4.4.16}$$

Define further η_*^β as the density of $Y\#(\Theta_k^\beta\gamma_*)$; this definition is independent of the index k, since γ_* is supported in the region $|x-y| \leq \tau$ where $\theta_k^0(x-y)$ vanishes. Obviously

$$\eta_* = \sum_\beta \eta_*^\beta, \quad \eta_k^{(3)} = \eta_k^{(3,0)} + \sum_\beta \eta_k^{(3,\beta)}. \tag{4.4.17}$$

In the convergence proof that follows, we use the dual representation of the norm on $L^q(\mathbb{R}^d)$:

$$\|f\|_{L^q} = \sup\left\{\int \psi(x)f(x)\,\mathrm{d}x\,;\ \psi \in C_c(\mathbb{R}^d),\ \|\psi\|_{L^{q'}} \leq 1\right\},$$

where $q' = \frac{q}{q-1}$ is the Hölder conjugate exponent of $q > 1$.

To begin with, observe that $\eta_k^{(3,0)}$ converges to zero in $L^m(\mathbb{R}^d)$. For that, let $\psi \in C(\mathbb{R}^d)$ with $\|\psi\|_{L^{m'}} \leq 1$. Then, with the help of Hölder's inequality and Lemma 6 above,

$$\begin{aligned}\iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(y)\,\mathrm{d}\gamma_k^{(3,0)}(x,y) &= \iint_{\mathbb{R}^d\times\mathbb{R}^d} \left[\int_{\mathbb{R}^d} \psi(x+z)\lambda(z)\,\mathrm{d}z\right] \mathrm{d}\gamma_k^{(2,0)}(x,y) \\ &\leq \iint_{\mathbb{R}^d\times\mathbb{R}^d} \|\psi\|_{L^{m'}}\|\lambda\|_{L^m}\,\mathrm{d}\gamma_k^{(2.0)} \leq 4\|\lambda\|_{L^m}\omega_k.\end{aligned}$$

Next, we show that $\eta_k^{(3,\beta)} \to \eta_*^\beta$ in $L^q(\mathbb{R}^d)$, for each β, where $q := \frac{2m}{m+1} < m$; note that $q' = 2m'$. For $\psi \in C(\mathbb{R}^d)$ with $\|\psi\|_{L^{q'}} \leq 1$, we have

$$\begin{aligned}&\int_{\mathbb{R}^d} \psi(y)\big[\eta_k^{(3,\beta)}(y) - \eta_*^{(3,\beta)}(y)\big]\,\mathrm{d}y = \iint_{\mathbb{R}^d\times\mathbb{R}^d} \Big[\psi(y-\sigma_k\beta)\Theta_k^\beta(T_k(x),y) - \psi(y)\Theta_k^\beta(x,y)\Big]\,\mathrm{d}\gamma_*(x,y) \\ &= \iint_{\mathbb{R}^d\times\mathbb{R}^d} [\psi(y-\sigma_k\beta) - \psi(y)]\Theta_k^\beta(x,y)\,\mathrm{d}\gamma_*(x,y) \\ &\quad + \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(y-\sigma_k\beta)\big[\theta_k^\beta(T_k(x)-y) - \theta_k^\beta(x-y)\big]\,\mathrm{d}\gamma_*(x,y) \\ &\leq \int_{\mathbb{R}^d} [\psi(y-\sigma_k\beta) - \psi(y)]\eta_*^\beta(y)\,\mathrm{d}y \\ &\quad + \left(\int_{\mathbb{R}^d} |\psi(y-\sigma_k\beta)|^2\eta_*^\beta(y)\,\mathrm{d}y\right)^{\frac12}\left(\|\nabla\theta_k^\beta\|_{L^\infty}^2 \int_{\mathbb{R}^d} |T_k(x)-x|^2\rho_*(x)\,\mathrm{d}x\right)^{\frac12} \\ &\leq \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(y)\big[\eta_*^\beta(y+\sigma_k\beta) - \eta_*^\beta(y)\big]\,\mathrm{d}y + \|\psi\|_{L^{2m'}}^{\frac12}\|\eta_*\|_{L^m}^{\frac12}\|\nabla\theta_k^\beta\|_{L^\infty}\omega_k \\ &\leq \big\|(\mathrm{id}-\sigma_k\beta)\#\eta_*^\beta - \eta_*^\beta\big\|_{L^q} + 3\|\eta_*\|_{L^m}^{\frac12}\omega_k^{\frac12}.\end{aligned}$$

In the last step, we have used that $\nabla\theta_k^\beta = (1-\theta_k^0)\nabla\vartheta^\beta - \vartheta^\beta\nabla\theta_k^0$, and hence $\|\nabla\theta_k^\beta\|_{L^\infty} \le 4\omega_k^{-1/2}$ by our hypotheses on θ_k^0 and ϑ^β, at least for all sufficiently large k. The first term of the final sum above goes to zero, since $\sigma_k \to 0$, and the translation semi-group is continuous in $L^q(\mathbb{R}^d)$; the second term goes to zero since $\omega_k \to 0$.

From this, we conclude convergence of $\eta_k^{(3,\beta)}$ to η_*^β in $L^q(\mathbb{R}^d)$, and in particular also in measure. Further, from the bound (4.4.16), it follows that $\eta_k^{(3,\beta)}$ is equi-integrable in $L^m(\mathbb{R}^d)$. Hence $\eta_k^{(3,\beta)} \to \eta_*^\beta$ also in $L^m(\mathbb{R}^d)$. In view of (4.4.17), this verifies the claim. □

Lemma 9. *Define η_k by $\eta_k(y) = \int_{\mathbb{R}^d} G_k(x,y)\,\mathrm{d}x$, with G_k from* (4.4.12)*. Then $\eta_k \in L^m(\mathbb{R}^d)$, and $\eta_k \to \eta$ in $L^m(\mathbb{R}^d)$. Consequently, $\mathcal{E}(Y\#\gamma_k) \to \mathcal{E}(Y\#\gamma_*)$.*

Proof. First, we recall two properties of the linear projection operator $\Pi_\delta : L^m(\mathbb{R}^d) \to L^m(\mathbb{R}^d)$ given by

$$\Pi_\delta[f](y) = \fint_Q f(y')\,\mathrm{d}y' \qquad \text{where } Q \in \mathcal{Q}_\delta \text{ is such that } y \in Q.$$

Namely,

(a) $\|\Pi_\delta[f] - \Pi_\delta[g]\|_{L^m(\mathbb{R}^d)} \le \|f-g\|_{L^m(\mathbb{R}^d)}$ for all $f, g \in L^m(\mathbb{R}^d)$;

(b) $\Pi_\delta[f] \to f$ in $L^m(\mathbb{R}^d)$ for each $f \in L^m(\mathbb{R}^d)$ as $\delta \searrow 0$.

Indeed, claim (a) is an easy consequence of Jensen's inequality:

$$\begin{aligned}\big\|\Pi_\delta[f] - \Pi_\delta[g]\big\|_{L^m(\mathbb{R}^d)}^m &= \sum_{Q\in\mathcal{Q}_{\delta_k}} \big\|\Pi_\delta[f] - \Pi_\delta[g]\big\|_{L^m(Q)}^m = \sum_{Q\in\mathcal{Q}_{\delta_k}} \int_Q \left|\fint_Q \big[f(y') - g(y')\big]\mathrm{d}y'\right|^m \mathrm{d}y \\ &\le \sum_{Q\in\mathcal{Q}_{\delta_k}} \int_Q \left[\fint_Q \big|f(y') - g(y')\big|^m \mathrm{d}y'\right] \mathrm{d}y = \sum_{Q\in\mathcal{Q}_{\delta_k}} \|f-g\|_{L^m(Q)}^m = \|f-g\|_{L^m(\mathbb{R}^d)}^m.\end{aligned}$$

Concerning claim (b), we use that thanks to hypothesis (4.2.3), arbitrary $y' \in Q$ lie in a ball of radius δ_k around any given $y \in Q$

$$\begin{aligned}\big\|\Pi_\delta[f] - f\big\|_{L^m(\mathbb{R}^d)}^m &= \sum_{Q\in\mathcal{Q}_{\delta_k}} \big\|\Pi_\delta[f] - f\big\|_{L^m(Q)}^m = \sum_{Q\in\mathcal{Q}_{\delta_k}} \int_Q \left|\fint_Q \big[f(y') - f(y)\big]\,\mathrm{d}y'\right|^m \mathrm{d}y \\ &\le \sum_{Q\in\mathcal{Q}_{\delta_k}} \int_Q \left[\fint_Q \big|f(y') - f(y)\big|^m \mathrm{d}y'\right] \mathrm{d}y \le \int_{\overline{\mathbb{B}}} \big\|f - f(\cdot + \delta_k z)\big\|_{L^m(\mathbb{R}^d)}^m \,\mathrm{d}z.\end{aligned}$$

The norm inside the final integral goes to zero as $\delta_k \to 0$, since $f(\cdot + \delta_k z) \to f$ in $L^m(\mathbb{R}^d)$, uniformly with respect to $z \in \overline{\mathbb{B}}$.

To connect this auxiliary result to the claim of the Lemma, recall that $\eta_k^{(3)} \to \eta_*$ in $L^m(\mathbb{R}^d)$ by Lemma 8 above, and observe that $\eta_k = \Pi_{\delta_k}[\eta_k^{(3)}]$. Therefore,

$$\|\eta_k - \eta_*\|_{L^m} \le \|\Pi_{\delta_k}[\eta_k^{(3)}] - \Pi_{\delta_k}[\eta_*]\|_{L^m} + \|\Pi_{\delta_k}[\eta_*] - \eta_*\|_{L^m} \le \|\eta_k^{(3)} - \eta_*\|_{L^m} + \|\Pi_{\delta_k}[\eta_*] - \eta_*\|_{L^m}$$

tends to zero.

From *Lemma 7* we have the narrow convergence of γ_k and consequently of $Y_\#\gamma_k = \eta_k$ which passes through $\int_\Omega \eta_k(\mathrm{v} + (\mathrm{w} * \eta_k)\,\mathrm{d}x$ and concludes this proof. □

4.4.4 Existence and convergence of minimizers

Lemma 10. *For each k large enough, $\mathcal{E}_k^\tau$ has a (unique if $\varepsilon_k > 0$) minimizer $\hat{\gamma}_k \in \Gamma_{\delta_k}(\rho_k)$.*

Proof. We use the estimates from Lemma 2: thanks to (4.4.4), the $\mathcal{E}_k^\tau$ are bounded below for all sufficiently small k. And thanks to (4.4.3), the γ's in the sublevels of $\mathcal{E}_k^\tau$ have uniformly bounded second moment, hence are relatively compact with respect to narrow convergence. Moreover, it is easily seen that $\mathcal{E}_k^\tau$ is the sum of three convex (in the sense of convex combinations of measures) functionals, and thus is lower semi-continuous with respect to narrow convergence. Moreover, $\mathcal{H}$ is a strictly convex functional on $\Gamma(\rho_k)$, so $\mathcal{E}_k^\tau$ is strictly convex if $\varepsilon_k > 0$. This together allows to invoke the direct methods from the calculus of variations and conclude the existence of a minimizer, which is unique if $\varepsilon_k > 0$. □

Lemma 11. *Let $\hat{\gamma}_k \in \Gamma_{\delta_k}(\rho_k)$ be minimizers of the respective $\mathcal{E}_k^\tau$. Then a subsequence of $(\hat{\gamma}_k)$ converges in $\mathbf{W}_2$ to a minimizer of $\mathcal{E}^\tau(\cdot|\rho_*)$.*

Proof. We begin by showing that the second momenta of the $\hat{\gamma}_k$ are k-uniformly bounded. In view of estimate (4.4.3), it suffices to show that $\mathcal{E}_k^\tau(\hat{\gamma}_k)$ is k-uniformly bounded. But this is a consequence of Γ-convergence: since $\mathcal{E}^\tau(\cdot|\rho_*)$ is not identically $+\infty$ — for instance, $\mathcal{E}^\tau((X,X)\#\rho_*\mathcal{L}^d|\rho_*) = \mathcal{E}(\rho_*) < \infty$ — there is a recovery sequence γ_k such that $\mathcal{E}_k^\tau(\gamma_k)$ is bounded, and hence also $\mathcal{E}_k^\tau(\hat{\gamma}_k)$ is bounded.

Consequently, there is a subsequence that converges narrowly to a limit $\hat{\gamma}_*$. Since $X\#\gamma_k = \rho_k\mathcal{L}^d \to \rho_*\mathcal{L}^d$ narrowly by hypothesis, and since the projection X is continuous, it follows that $\gamma^* \in \Gamma(\rho_*)$. Thus, by the fundamental properties of Γ-convergence, γ_* is a minimizer of $\mathcal{E}^\tau(\cdot|\rho_*)$.

It remains to be shown that actually $\hat{\gamma}_k \to \hat{\gamma}_*$ in $\mathbf{W}_2$. It suffices to verify that $\hat{\gamma}_k$'s second moment converges to that of $\hat{\gamma}_*$. The second moment of $\hat{\gamma}_k$ amounts to

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \left(|x|^2+|y|^2\right) \mathrm{d}\hat{\gamma}_k = 2\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x|^2\,\mathrm{d}\hat{\gamma}_k + \iint_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^2\,\mathrm{d}\hat{\gamma}_k + 2\iint_{\mathbb{R}^d\times\mathbb{R}^d} x\cdot(y-x)\,\mathrm{d}\hat{\gamma}_k. \tag{4.4.18}$$

Thanks to Lemma 1,

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x|^2\,\mathrm{d}\gamma_k = \int_{\mathbb{R}^d} |x|^2\rho_k(x)\,\mathrm{d}x \to \int_{\mathbb{R}^d} |x|^2\rho_*(x)\,\mathrm{d}x = \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x|^2\,\mathrm{d}\gamma_*.$$

Further, recalling the lower bound (4.2.5) on $\mathbf{c}_k$ and estimate (4.4.4), we obtain for all sufficiently large k that

$$\iint_{|y-x|\geq 2\tau} |y-x|^2\,\mathrm{d}\hat{\gamma}_k \leq 4\iint_{|y-x|\geq 2\tau} \left(|y-x|-\tau\right)^2 \mathrm{d}\hat{\gamma}_k \leq 4\alpha_k \iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{c}_k\,\mathrm{d}\hat{\gamma}_k \leq \frac{8\alpha_k}{\tau}\mathcal{E}_k^\tau(\hat{\gamma}_k),$$

which converges to zero as $k\to\infty$ since $\mathcal{E}_k^\tau(\hat{\gamma}_k)$ is bounded. In the same spirit, also

$$\left|\iint_{|y-x|\geq 2\tau} x\cdot(y-x)\,\mathrm{d}\hat{\gamma}_k\right| \leq \frac{\sqrt{\alpha_k}}{2}\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x|^2\,\mathrm{d}\hat{\gamma}_k + \frac{1}{2\sqrt{\alpha_k}}\iint_{|y-x|\geq 2\tau} |y-x|^2\,\mathrm{d}\hat{\gamma}_k$$

converges to zero. The continuous function $|y-x|^2$ is bounded on the set where $|y-x|\leq 2\tau$, so narrow convergence $\hat{\gamma}_k \to \hat{\gamma}_*$ implies

$$\iint_{|y-x|\leq 2\tau} |y-x|^2\,\mathrm{d}\hat{\gamma}_k \to \iint_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^2\,\mathrm{d}\hat{\gamma}_*.$$

Finally, for $|y-x| \leq 2\tau$, the function $x \cdot (y-x)$ is bounded in modulus by $2\tau|x|$. Since the $\hat{\gamma}_k$ have k-uniformly bounded second momenta, Prokhorov's theorem yields

$$\iint_{|y-x|\leq 2\tau} x \cdot (y-x)\, \mathrm{d}\hat{\gamma}_k \to \iint_{\mathbb{R}^d \times \mathbb{R}^d} x \cdot (y-x)\, \mathrm{d}\hat{\gamma}_*.$$

In summary, we can pass to the limit $k \to \infty$ in each term on the right-hand side of (4.4.18), obtaining the second moment of $\hat{\gamma}_*$. ☐

4.5 Numerical scheme

4.5.1 Formulation of the minimization problem

Throughout this section, we assume that the following are fixed: a bounded domain $\Omega \subset \mathbb{R}^d$, a tesselation $\mathcal{Q}_\delta$ of Ω with cells of diameter at most $\delta > 0$, see (4.2.3), an entropic regularization parameter $\varepsilon > 0$, a time step $\tau > 0$, and an approximation $\tilde{\mathbf{c}} := \mathbf{c}_{\tau,\delta}$ of the distance cost function $\mathbf{c}$, which is such that $\tilde{\mathbf{c}}$ is constant (possibly $+\infty$) on each $Q \times Q'$ where $Q, Q' \in \mathcal{Q}_\delta$, and such that $\tilde{\mathbf{c}}(x,y) < \infty$ at each (x,y) with $|x-y| \leq \tau$. We assume that the elements Q_i of $\mathcal{Q}_\delta$ are enumerated with an index $i \in I$, where I is a finite index set, and for each $i \in I$, a point $x_i \in Q_i$ is given.

We need to fix some further notations: indexed quanties $u = (u_i)_{i\in I}$ are considered as (column) vectors, quantities $g = (g_{i,j})_{i,j\in I}$ with double index as matrices. Below, we use $\odot$ to denote the entry-wise products of vectors and matrices, $[u \odot v]_j = u_j v_j$ and $[g \odot h]_{i,j} = g_{i,j} h_{i,j}$, respectively. In the same spirit, $\frac{u}{v}$ and $\frac{g}{h}$ denote entry-wise division. Further, for a vector u, we denote by $\operatorname{diag}(u)$ the diagonal matrix with the vector u on the diagonal $[\operatorname{diag}(u)]_{i,j} = u_i \delta_{i,j}$ where $\delta_{i,j}$ denotes the Kronecker delta. For the sake of disambiguation, the usual matrix-vector product is written as $g \cdot u$, i.e., $[g \cdot u]_i = \sum_j g_{i,j} u_j$, and $u \otimes v$ denotes outer product of the vectors u and v, that is $[u \otimes v]_{i,j} = u_i v_j$.

Remark 2. *With the x_i at hand, a practical choice for $\tilde{\mathbf{c}}$ that conforms with* (4.2.4) *and* (4.2.5) *is the following:*

$$\tilde{\mathbf{c}}(x,y) = \tilde{\mathbf{c}}_{i,j} := \mathbf{C}\left(\frac{|x_i - y_j|}{\tau + \delta}\right) \text{ for all } x \in Q_i,\ y \in Q_j, \tag{4.5.1}$$

and extend $\tilde{\mathbf{c}}$ by lower semi-continuity to all of $\mathbb{R}^d \times \mathbb{R}^d$. The modified denominator $\tau + \delta$ has been chosen such that $\tilde{\mathbf{c}}$ is finite on each $2d$-cube $Q_i \times Q_j$ that intersects the region $|x - y| \leq \tau$.

A density $\rho \in \mathcal{P}_\delta(\Omega)$ is the conveniently identified with the vector $r = (r_i)$, where r_i is the constant density on Q_i. Now, if $\rho \in \mathcal{P}_\delta(\Omega)$, and if $\gamma = G\mathcal{L}^d \otimes \mathcal{L}^d$ is a minimizer of $\mathcal{E}^\tau_{\varepsilon,\delta,\tilde{\mathbf{c}}}(\cdot|\rho)$ on $\Gamma_\delta(\rho)$, then G is constant on each $2d$-cube $Q_i \times Q_j$; this follows by Jensen's inequality and strict convexity of $\mathcal{H}$. Accordingly, the set of all possible minimizers γ can be parametrized by matrices g, where $g_{i,j}$ is the constant value of γ's density on $Q_i \times Q_j$.

For notational simplicity, introduce the vector $\mathbb{I}_\delta$ with $[\mathbb{I}_\delta]_j = |Q_j|$ for all j, so that

$$[\mathbb{I}_\delta^T \cdot g]_j = \sum_i |Q_i| g_{i,j}, \quad [g \cdot \mathbb{I}_\delta]_i = \sum_j |Q_j| g_{i,j}.$$

In this notation, the constraint $X\#\gamma = \rho\mathcal{L}^d$ then becomes $g \cdot \mathbb{I}_\delta = r$, and we have

$$\mathcal{H}(\gamma) = \sum_{i,j} |Q_i||Q_j|\big[g_{i,j}\log g_{i,j} - g_{i,j}\big], \quad \mathcal{E}(Y\#\gamma) = \sum_j \left[|Q_j| h\left(\sum_i |Q_i| g_{i,j}\right)\right].$$

In terms of the notations introduced above, the variational problem (4.1.10) turns into

$$
\begin{aligned}
g^n = \underset{g=(g_{i,j})}{\arg\min} \Bigg(& \varepsilon \sum_{i,j} \Big[|Q_i||Q_j| \Big(\frac{\tau}{\varepsilon} \tilde{\mathbf{c}}_{i,j} + \log g_{i,j} \Big) g_{i,j} - g_{i,j} \Big] \\
& + \sum_j \left[|Q_j| h \left(\sum_i |Q_i| g_{i,j} \right) \right] + \iota_{(\bar{r}^{(n-1)} - g \mathbb{I}_\delta)} \Bigg),
\end{aligned}
\tag{4.5.2}
$$

where $\bar{r}^{n-1} = g^{n-1} \cdot \mathbb{I}_\delta$ encodes the datum from the previous step.

4.5.2 Excursion: Dykstra's algorithm

In this section, we briefly summarize the concept of the generalized Dykstra algorithm that is the basis for the efficient numerical approximation of Wasserstein gradient flows in the spirit of [42].

Let $F : X \to \mathbb{R}$ be a convex differentiable function defined on a Hilbert space X, and let F^* be its Legendre dual. Below, we identify at each $x \in X$ the differentials $F'(x), (F^*)'(x) \in X'$ by their respective Riesz duals in X. The *Bregman divergence* $D_F(x, y)$ of $x \in X$ relative to $y \in X$ is defined by

$$
D_F(x|y) = F(x) - F(y) - \langle F'(y), x - y \rangle. \tag{4.5.3}
$$

By convexity, $D_F(x|y) \geq 0$. Further, let $\phi_1, \phi_2 : X \to \mathbb{R} \cup \{+\infty\}$ be two proper, convex and lower semi-continuous functionals on X, and consider, for a given $y \in X$, the variational problem

$$
D_F(x|y) + \phi_1(x) + \phi_2(x) \longrightarrow \min. \tag{4.5.4}
$$

In this setting, the generalized Dykstra algorithm for approximation of a minimizer $x^* \in X$ is the following. Let $x^{(0)} := y$ and $q^{(0)} := q^{(-1)} := 0$, and define for $k = 0, 1, 2, \dots$ inductively:

$$
\begin{aligned}
x^{(k+1)} &:= \underset{x \in X}{\arg\min} \left[D_F\big(x \big| (F^*)'(F'(x^{(k)}) + q^{(k-1)})\big) + \phi_{[k]}(x) \right], \\
q^{(k+1)} &:= F'(x^{(k)}) + q^{(k-1)} - F'(x^{(k+1)}),
\end{aligned}
\tag{4.5.5}
$$

where $[k] = 1$ if k is even, and $[k] = 2$ if k is odd. In the special case that $F(x) = \frac{1}{2}\langle x, x\rangle$ and ϕ_1, ϕ_2 are the indicator functions of two convex sets with non-empty intersection, then (4.5.5) reduces to the original Dykstra projection algorithm.

Under certain hypotheses (for instance, if X is finite-dimensional), it can be proven that $x^{(k)}$ converges to a minimizer x^* of (4.5.4) in X. The core idea of the convergence proof is to study the dual problem for (4.5.4), for which the iteration (4.5.5) attains a considerably easier form. We refer to [42, 8, 14] for further discussion of the algorithm, including questions of well-posedness and convergence, in the context of fully discrete approximation of gradient flows.

4.5.3 From the minimization problem to the iteration

In this section, we follow once again closely [42] with the goal is to rewrite (4.5.2) in the form (4.5.4), and then to apply the algorithm (4.5.5) to its solution. The Hilbert space is that of matrices $g = (g_{i,j})_{i,j \in I}$ endowed with the scalar product

$$
\langle g, g' \rangle = \sum_{i,j} |Q_i||Q_j| g_{i,j} g'_{i,j},
$$

and we shall choose F in (4.5.3) as

$$F(g) = \sum_{i,j} |Q_i||Q_j| g_{i,j} \log g_{i,j},$$

with the convention that $0 \log 0 = 0$ and $r \log r = +\infty$ for any $r < 0$, which has Legendre dual

$$F^*(\omega) = \sum_{i,j} |Q_i||Q_j| e^{\omega_{i,j}},$$

and respective derivatives — recall that we identify the functional $F'(g)$ with its Riesz dual —

$$\big[F'(g)\big]_{i,j} = \log g_{i,j}, \quad \big[(F^*)'(\omega)\big]_{i,j} = \exp \omega_{i,j}.$$

The corresponding Bregman distance is the Kullback-Leibler divergence,

$$\mathrm{KL}(g|\omega) := D_F(g|\omega) = \sum_{i,j} |Q_i||Q_j| \big[g_{i,j}(\log g_{i,j} - \log \omega_{i,j}) - g_{i,j} + \omega_{i,j}\big],$$

which is defined for matrices g and ω with non-negative entries. The correct interpretation of the logarithmic terms is the following: if $\omega_{i,j} = 0$, then the entire term in square brackets is $+\infty$ unless $g_{i,j} = 0$ as well, in which case this term is zero.

Next, we rewrite our minimization problem (4.5.2) in the form (4.5.4). As the reference density $\xi = (\xi_{i,j})$ for the divergence, we choose

$$\xi_{i,j} = \begin{cases} \exp\left(-\frac{\tau}{\varepsilon}\tilde{\mathbf{c}}_{i,j}\right) & \text{if } \tilde{\mathbf{c}}_{i,j} \text{ is finite,} \\ 0 & \text{if } \tilde{\mathbf{c}}_{i,j} = +\infty. \end{cases}$$

Thus $\tau\tilde{\mathbf{c}}_{i,j} g_{i,j} = -\varepsilon g_{i,j} \log \xi_{i,j}$, with the convention that $0 \log 0 = 0$, but $(-a)\log(-a) = +\infty$ and $-a \log 0 = +\infty$ for any $a > 0$. The sum of the first two terms in the variational functional (4.5.2) takes the convenient form

$$\begin{aligned} \sum_{i,j} \Big[|Q_i||Q_j| \Big(\frac{\tau}{\varepsilon}\tilde{\mathbf{c}}_{i,j} + \log g_{i,j} - 1\Big) g_{i,j}\Big] &= \sum_{i,j} |Q_i||Q_j| g_{i,j}\big(\log g_{i,j} - \log \xi_{i,j} - 1\big) \\ &= \mathrm{KL}(g|\xi) - \sum_{i,j} |Q_i||Q_j| \xi_{i,j}. \end{aligned}$$

Recall that $\mathrm{KL}(g|\xi) \geq 0$ by construction, and that $\mathrm{KL}(g|\xi) = +\infty$ unless $g_{i,j} = 0$ for all (i,j) with $\tilde{\mathbf{c}}_{i,j} = +\infty$. Neglecting irrelevant factors and constants, the minimization problem (4.5.2) attains the form

$$g^n = \arg\min_g \big[\varepsilon \mathrm{KL}(g|\xi) + \phi_1(\mathbb{I}_\delta^T \cdot g) + \phi_2^n(g \cdot \mathbb{I}_\delta)\big], \tag{4.5.6}$$

where

$$\phi_1(s) = \mathcal{E}_\delta(s) = \sum_j |Q_j| h(s_j), \quad \phi_2^n(r) = \iota_{(\bar{r}^{n-1} - r)} = \begin{cases} 0 & \text{if } r = \bar{r}^{(n-1)}, \\ +\infty & \text{otherwise.} \end{cases}.$$

Using that for our choice of F,

$$\big[(F^*)'\big(F'(x) + q\big)\big]_{i,j} = \exp\big(\log x_{i,j} + q_{i,j}\big) = (x \odot s)_{i,j}, \quad \text{with } s_{i,j} := e^{q_{i,j}},$$

Dykstra's algorithm (4.5.5) translates into the following: from $g^{(0)} = \xi$ and $s^{(0)} = s^{(-1)} \equiv 1$, define inductively

$$g^{(k+1)} = \Phi_{[k]}(g^{(k)} \odot s^{(k-1)}), \quad s^{(k+1)} = \frac{g^{(k)} \odot s^{(k-1)}}{g^{(k+1)}}, \tag{4.5.7}$$

again with $[k] = 1$ for even k, and $[k] = 2$ for odd k, where $\Phi_1(\omega)$ and $\Phi_2(\omega)$ are, respectively, the solutions to the minimization problems

$$\varepsilon \mathrm{KL}(g|\omega) + \phi_1(\mathbb{I}_\delta^T \cdot g) \to \min \quad \text{and} \quad \varepsilon \mathrm{KL}(g|\omega) + \phi_2^n(g \cdot \mathbb{I}_\delta) \to \min\ . \tag{4.5.8}$$

These minimization problems can be solved almost explicitly. Their respective Euler-Lagrange equations are, at each (i,j) with $\omega_{i,j} > 0$,

$$0 = \varepsilon \log \frac{g_{i,j}}{\omega_{i,j}} + h'\left(\sum_i |Q_i| g_{i,j}\right), \quad \text{and} \quad 0 = \varepsilon \log \frac{g_{i,j}}{\omega_{i,j}} + \lambda_i,$$

where the λ_i are Lagrange multipliers to realize the constraint $g \cdot \mathbb{I}_\delta = \bar{r}^{n-1}$. After dividing these equations by ε, taking the exponential, and evaluation of the marginals, one obtains in a straight-forward way the following representation of the minimizers in (4.5.8):

$$\Phi_1(\omega) = \omega \cdot \mathrm{diag}\left(\frac{H_\varepsilon^{-1}(\mathbb{I}_\delta^T \cdot \omega)}{\mathbb{I}_\delta^T \cdot \omega}\right), \quad \text{and} \quad \Phi_2(\omega) = \mathrm{diag}\left(\frac{\bar{r}^{n-1}}{\omega \cdot \mathbb{I}_\delta}\right) \cdot \omega,$$

where $[H_\varepsilon^{-1}(\eta)]_j$ for given $\eta_j \geq 0$ is the solution z to the nonlinear relation

$$H_\varepsilon(z) = z\mathrm{exp}\left(\frac{h'(z)}{\varepsilon}\right) = \eta_j;$$

note that the equations for the components of $H_\varepsilon^{-1}(\eta)$ are decoupled.

Finally, a significant reduction in the computational complexity of the algorithm is achieved by taking advantage of the Dyadic structure of g and s that is inherited from each iteration to the next: at each stage k, there are vectors $\alpha^{(k)}$, $\beta^{(k)}$ and $u^{(k)}$, $v^{(k)}$ such that

$$g^{(k)} = \left(\alpha^{(k)} \otimes \beta^{(k)}\right) \odot \xi, \quad s^{(k)} = u^{(k)} \otimes v^{(k)}. \tag{4.5.9}$$

Inserting this special form into (4.5.7), one obtains iteration rules for $\alpha^{(k)}$, $\beta^{(k)}$ and $u^{(k)}$, $v^{(k)}$, that are summarized below.

Proposition 4. *Initialize $\alpha_i^{(0)} = \beta_j^{(0)} = 1$ and $u_i^{(0)} = u_i^{(-1)} = v_j^{(0)} = v_j^{(-1)} = 1$ for all i, j, and calculate inductively $\alpha^{(k)}$, $\beta^{(k)}$ and $u^{(k)}$, $v^{(k)}$ for $k = 1, 2, \dots$ from*

$$\alpha^{(k+1)} = \begin{cases} \alpha^{(k)} \odot u^{(k-1)} & \text{if } k \text{ odd,} \\ \frac{\bar{r}^{n-1}}{\xi \cdot (\beta^{(k+1)} \odot \mathbb{I}_\delta)} & \text{if } k \text{ even,} \end{cases} \quad \beta^{(k+1)} = \begin{cases} \frac{H_\varepsilon^{-1}\left((\xi^T \cdot (\alpha^{(k+1)} \odot \mathbb{I}_\delta)) \odot \beta^{(k)} \odot v^{(k-1)}\right)}{\xi^T \cdot (\alpha^{(k+1)} \odot \mathbb{I}_\delta)} & \text{if } k \text{ odd,} \\ \beta^{(k)} \odot v^{(k-1)} & \text{if } k \text{ even,} \end{cases}$$

$$u^{(k+1)} = \frac{\alpha^{(k)} \odot u^{(k-1)}}{\alpha^{(k+1)}}, \quad v^{(k+1)} = \frac{\beta^{(k)} \odot v^{(k-1)}}{\beta^{(k+1)}},$$

with the understanding that for odd k, one calculates $\alpha^{(k+1)}$ first and $\beta^{(k+1)}$ next, and the other way around for even k. Further, the quotient $\frac{0}{0}$ is interpreted as 0.

Then (4.5.9) produces the iterates $g^{(k)}$ and $s^{(k)}$ of (4.5.7).

Proof. We assume that $g^{(\ell)} = (\alpha^{(\ell)} \otimes \beta^{(\ell)}) \odot \xi$ and $s^{(\ell)} = u^{(\ell)} \otimes v^{(\ell)}$ are in the form (4.5.9) for all $\ell = 0, 1, 2, \dots, k$; we show that with $\alpha^{(k+1)}$, $\beta^{(k+1)}$ and $u^{(k+1)}$, $v^{(k+1)}$ defined as above, $g^{(k+1)} = (\alpha^{(k+1)} \otimes \beta^{(k+1)}) \odot \xi$ and $s^{(k+1)} = u^{(k+1)} \otimes v^{(k+1)}$ satisfy the original induction formula (4.5.7).

First, note that

$$g^{(k)} \odot s^{(k-1)} = (\alpha^{(k)} \otimes \beta^{(k)}) \odot \xi \odot (u^{(k-1)} \otimes v^{(k-1)}) = \big((\alpha^{(k)} \odot u^{(k-1)}) \otimes (\beta^{(k)} \odot v^{(k-1)})\big) \odot \xi.$$

Further, we shall use the rule that for arbitrary vectors p, q and x, and matrices h,

$$[(p \otimes q) \odot h] \cdot x = p \odot [h \cdot (q \odot x)].$$

Now, if k is odd, then

$$\begin{aligned}
\alpha^{(k+1)} \otimes \beta^{(k+1)} &= \frac{\bar{r}^{n-1}}{\xi \cdot (\beta^{(k+1)} \odot \mathbb{I}_\delta)} \otimes \beta^{(k+1)} \\
&= \left(\frac{r^{n-1}}{\alpha^{(k)} \odot u^{(k-1)} \odot (\xi \cdot (\beta^{(k)} \odot v^{(k-1)} \odot \mathbb{I}_\delta))} \odot \alpha^{(k)} \odot v^{(k-1)} \right) \otimes (\beta^{(k)} \odot v^{(k-1)}) \\
&= \operatorname{diag}\left(\frac{r^{n-1}}{(g^{(k)} \odot s^{(k-1)}) \cdot \mathbb{I}_\delta} \right) \cdot \frac{g^{(k)} \odot s^{(k-1)}}{\xi} = \frac{\Phi_1(g^{(k)} \odot s^{(k-1)})}{\xi} = \frac{g^{(k+1)}}{\xi}.
\end{aligned}$$

In the same spirit, for k even, one shows that

$$\begin{aligned}
\alpha^{(k+1)} \otimes \beta^{(k+1)} &= (\alpha^{(k)} \odot u^{(k-1)}) \otimes \left(\beta^{(k)} \odot v^{(k-1)} \odot \frac{H_\varepsilon^{-1}\big((\xi^T \cdot (\alpha^{(k+1)} \odot \mathbb{I}_\delta)) \odot \beta^{(k)} \odot v^{(k-1)}\big)}{(\xi \cdot (\alpha^{(k+1)} \odot \mathbb{I}_\delta)) \odot \beta^{(k)} \odot v^{(k-1)}} \right) \\
&= \frac{\Phi_2(g^{(k)} \odot s^{(k-1)})}{\xi} = \frac{g^{(k+1)}}{\xi}.
\end{aligned}$$

Finally,

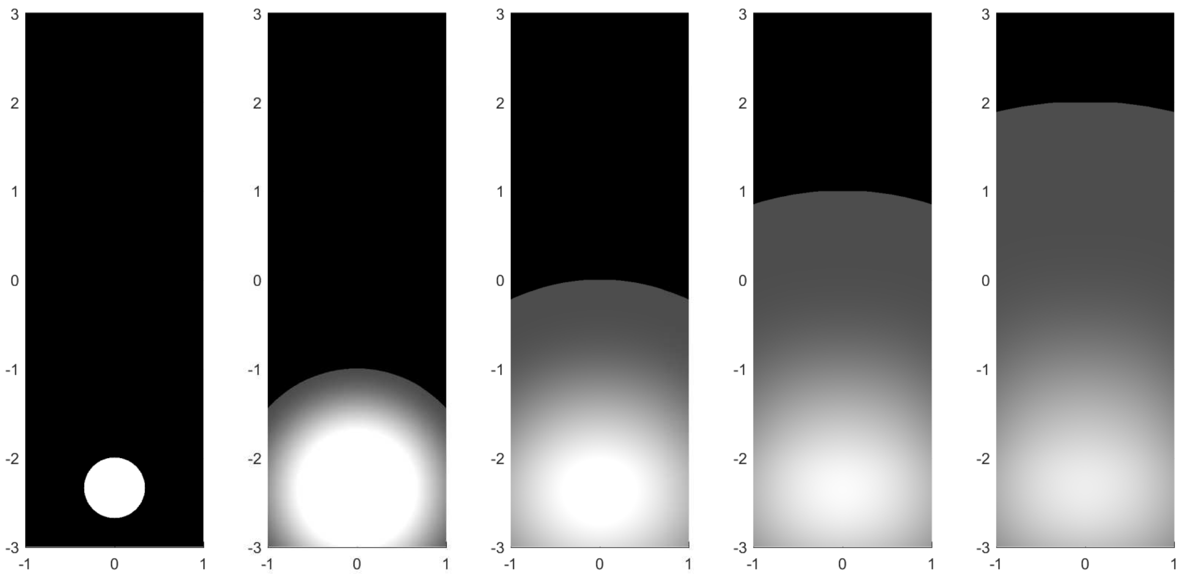
$$\begin{aligned}
u^{(k+1)} \otimes v^{(k+1)} &= \frac{\alpha^{(k)} \odot u^{(k-1)}}{\alpha^{(k+1)}} \otimes \frac{\beta^{(k)} \odot v^{(k-1)}}{\beta^{(k+1)}} \\
&= \frac{(\alpha^{(k)} \otimes \beta^{(k)}) \odot \xi \odot (u^{(k-1)} \otimes v^{(k-1)})}{(\alpha^{(k+1)} \otimes \beta^{(k+1)}) \odot \xi} = \frac{g^{(k)} \odot s^{(k-1)}}{g^{(k+1)}} = s^{(k+1)}.
\end{aligned}$$

□

4.5.4 Implementation

Based on the discussion above, we introduce a numerical scheme for approximate solution of the initial value problem for (4.1.1) as follows. Choose a spatial mesh width $\delta > 0$ and an entropic regularization parameter $\varepsilon > 0$. Further, define a suitable approximation $\tilde{\mathbf{c}}$ of the cost function $\mathbf{c}$ that is constant on cubes $Q_i \times Q_j$, for instance as in (4.5.1), and an approximation r^0 of the initial condition, for instance $r_i^0 = \fint_{Q_i} \rho^0(x)\,\mathrm{d}x$.

From a given r^{n-1}, the next iterate r^n is obtained as second marginal, $r_j^n = \sum_i |Q_i| g_{i,j}^n$, of the minimizer g^n to the variational problem (4.5.2) or, equivalently, (4.5.6). To calculate g^n from r^{n-1}, we use Dykstra's algorithm (4.5.7) as shown in Proposition 4 above. That is, we calculate alternatingly the scaling factors $\alpha^{(k)}$, $\beta^{(k)}$, and the auxiliary vectors $u^{(k)}$ and $v^{(k)}$, using the iteration from Proposition 4 with $\bar{r} := r^{n-1}$. The updates of $\alpha^{(k+1)}$, $u^{(k+1)}$ and $v^{(k+1)}$ are obviously very efficient. To calculate the term involving H_ε^{-1} in the update for $\beta^{(k+1)}$, we use a Newton iteration, which converges in few steps. The iteration in k is repeated until the changes in α and β from one iteration to the next meets a smallness condition. Then $g_{i,j}^n := \alpha_i^{(k)} \xi_{i,j} \beta_j^{(k)}$.



(a) The initial probability density $\rho^{(0)}$. (b) The density after the first iteration $\rho^{(1)}$. (c) The density after the second iteration $\rho^{(2)}$. (d) The density after the third iteration $\rho^{(3)}$. (e) The density after the fourth iteration $\rho^{(4)}$.

Figure 4.2: A support of the density propagates at most with "light speed". The greyscale possesses a step from black (representing density 0) to the darkest displayed gray (representing the smallest double-precision floating-point number greater than 0) in order to illustrate the support of ρ moving with finite speed. The Iteration was performed on a grid of 400×1200 uniformly distributed gridpoints on $[-1,1] \times [-3,3] \subset \mathbb{R}^2$ with parameters $\tau = 1$, $\varepsilon = 0.5$, $m = 2$ and lightspeed 1. As initial distribution we used $\rho^{(0)}$ with its mass equally distributed on its support, a ball with radius 0.8 centered at $(0, -2.8)$. This way, the uppermost points in the support of $\rho^{(0)}$ have ordinate $y = -2$ and the propagation with lightspeed can be observed over the displayed plots.

4.5.5 Numerical experiments

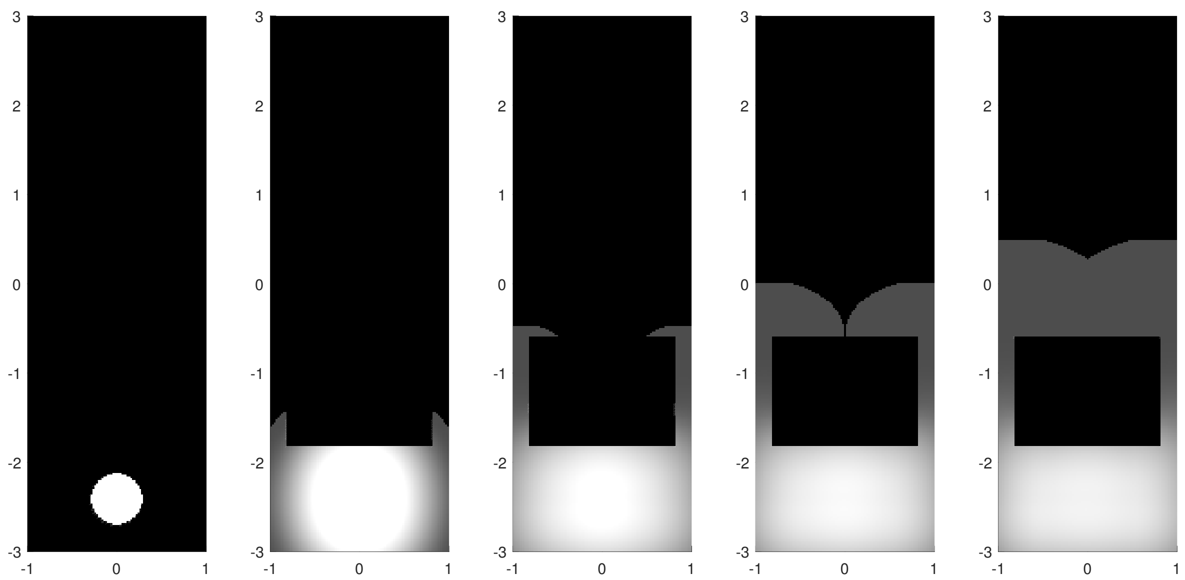
In our expriments, we study the application of our discretization method to the equation

$$\partial_t \rho = \nabla \cdot \left[\rho \, \frac{\nabla \rho}{\sqrt{1 + |\nabla \rho|^2}} \right],$$

which is (4.1.1) with the relativistic cost $\mathbf{C}(v) = 1 - \sqrt{1 - |v|^2}$ and the energy from (4.1.4) with $h(r) = r^2/2$. Naturally, all experiments are carried out on finite domains Ω, which are either of dimension $d = 1$ or $d = 2$.

Finite speed of propagation

In the first experiment, we study how the flux limitation becomes manifest numerically. We consider the rectangular box $\Omega = (-1, 1) \times (-3, 3)$ in $\mathbb{R}^2$, and a discretization by squares of edge length 0.005. Our



(a) The initial probability density ρ^0. (b) The density after two iterations ρ^2. (c) The density after four iterations ρ^4. (d) The density after five iterations ρ^5. (e) The density after six iterations ρ^6.

Figure 4.3: Evolution of a density around a obstacle. Grayscale as in *Figure* **??** . The iteration was performed on the remaining part of a 100×300, equidistant, quadrilateral grid on $[-1, 1] \times [-3, 3]$, after the obstacle points were removed. The parameters were $\tau = 0.5$, $\varepsilon = 0.1$, $m = 2$ and again, lightspeed set to 1. As initial distribution we used ρ^0 with its mass equally distributed over a small ball with center $(x, y) = (2, 1.2)$.

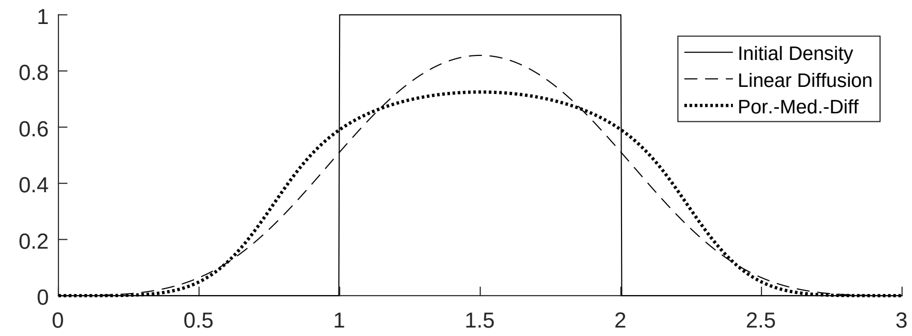
time step is $\tau = 1$. The chosen discrete approximation $\tilde{\mathbf{c}}$ to the cost function $\mathbf{c}$ is of the type (4.5.1), so in particular we set $\xi_{i,j} = 0$ if $|x_i - x_j| > 1$. We chose a (discontinuous) initial condition $\rho^{(0)}$ that is a uniformly distributed on a ball.

Figure **??** shows (from left to right) the initial density, and then the solution at $t = \tau, 2\tau, 3\tau$ and $t = 4\tau$. In order to make the finite speed of propagation of the support visible, we set the grayscale to black for $\rho(x) = 0$, and to a gray visibly lighter than black as soon as $\rho(x) > 0$. Additionally, the support of the initial density is chosen as a ball, positioned at $(0, -2.8)$ and with radius 0.8. This way, $\rho^{(0)}$ is supported in $y \leq -2$ and the propagation of the support with lightspeed can easily be observed as the support increases its radius by 1 in each step.

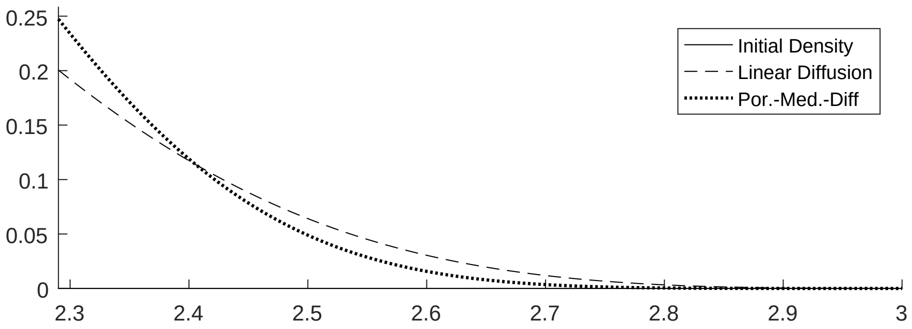
Motion around obstacles

The algorithm we used here allows for an easy implementation of impentrable obstacles in the domain. The only thing that has to be altered is the matrix ξ. There the columns and rows corresponding to a point lying within the obstacle have to be set to zero and components of ξ corresponding to a pair of points whose connecting line segment crosses the obstacle have to be recalculated (c.f. (4.2.1)).

In *Figure* **??** we have realized a impenetrable box and a density flowing around it. Again we have used the step in the grayscale to illustrate the support of ρ and again we can observe the finite speed of



(a) Comparison between Linear diffusion and porous medium diffusion with parameter $m = 5$.



(b) Magnification of the comparison.

Figure 4.4: The iteration for the same discontinuous initial data depicted by a solid line. The Iteration is performed on an equidistand grid with 1000 grid points with time-step $\tau = 0.02$ and time horizon $T = 1$ and entropic regularization parameter $\varepsilon = 0.04$.

propagation.

Comparison: Linear diffusion and porous medium diffusion

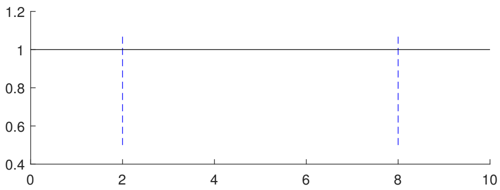
The iteration can be carried out with porous medium as well as with linear diffusion. In *Figure* **??** some features of the two different diffusions can be compared. The figure shows the result of iterating both with the same initial data. Note that the iteration is already advanced enough that the fronts that can be expected with flux-limitation and such discontinuous initial data are already dispersed.

Porous medium diffusion disperses the mass faster than linear diffusion where there is a high density and is slower when there is low density which results in the lower density for our porous medium evolution around $x = 0$ compared to linear diffusion. On the other hand, as can clearly be seen in the magnification, linear diffusion disperses the mass faster for densities close to zero.

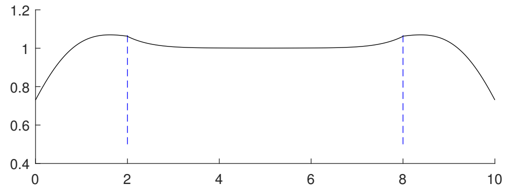
Finally, though it can not be observed easily in the plots, the support of both, the linear diffusion evolution and the porous medium evolution, expands with the same velocity, which is our lightspeed.

Edge effect

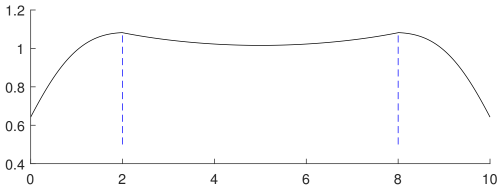
Our last experiment is posed on a one-dimensional interval $[0, 10]$, which is discretized with 1000 intervals of equal length. The result in Figure **??** highlights an undesired effect at the edges: although we initialize



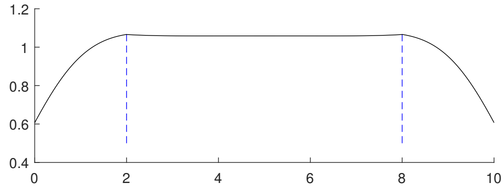
(a) The initial probability density $\rho^{(0)}$.



(b) The density after the first iteration $\rho^{(1)}$.



(c) The density after the fourth iteration $\rho^{(4)}$.



(d) The steady state corresponding to this initial vector.

Figure 4.5: The edge effect caused by the blurring with the Gibbs kernel. Iteration performed on a 1000 grid points equidistantly distributed on $[0, 10]$ with parameters $\tau = 2$, $\varepsilon = 2$, $m = 2$. As initial distribution we used $\rho_i^0 = 1$. The horizontal, dotted lines are drawn at $x = 0 + \tau$ and $10 - \tau$ and mark the width of the edge effect.

with a uniform distribution (which corresponds to a stationary solution), the density becomes non-homogeneous near the boundary points very quickly. In first order, the solution represents the second marginal of the matrix ξ; since the matrix is "cut off" at the boundary, there is a lack of mass near the end points. The energy introduces a second order effect, which tries to compensate the primary effect by transporting mass from the bulk to the edges.

This effect is the stronger, the larger the entropic regularization parameter $\varepsilon > 0$ is; the pictures have been produced for a "huge" value $\varepsilon = 2$.

Chapter 5

Entropic regularized BDF-2 Scheme

5.1 Introduction and preliminary results

5.1.1 Introduction

This chapter is concerned with well-posedness and convergence of an entropic regularized BDF 2 (Backwards Differentiation Formula 2) scheme, a semi-discretization in time which was, as an unregularized scheme, introduced and analysed in [33, 43]. We motivated the derivation of this scheme in the introduction of this thesis, also introducing the entropic regularized BDF 2 scheme (1.0.16), which we will be considering in this chapter.

Our goal is to apply this scheme to the non-linear Fokker-Planck equation, i.e. (1.0.1) with $\mathcal{A} = \mathrm{Id}$.

$$\partial_t \rho = \Delta(p(\rho)) + \operatorname{div}\left(\rho \nabla [V + (W * \rho)]\right) \qquad \text{on } (0,\infty)\times\Omega \tag{5.1.1}$$

$$\mathbf{n}\cdot\nabla\rho = 0 \qquad \text{on } (0,\infty)\times\partial\Omega \tag{5.1.2}$$

$$\rho(0,\cdot) = \rho^0 \qquad \text{on } \Omega \tag{5.1.3}$$

on some open, convex, bounded set $\Omega \subset \mathbb{R}^d$ with Lipschitz boundary, where we denote with $\mathbf{n}$ the outward pointing unit normal vectorfield to $\partial\Omega$ and $\rho^0 \in \mathscr{P}^{ac}(\Omega)$ with finite entropy $\mathcal{H}(\rho^0) < \infty$ the initial datum. Furthermore the connection of p to u is given by $p(s) = s\mathsf{u}'(s) - \mathsf{u}(s)$, so $p(s) = s^m$. u is again defined as in (1.0.5) where $\mathsf{u} : [0,\infty) \to \mathbb{R}$ with

$$\mathsf{u}(s) = \mathsf{u}_m(s) = \begin{cases} \mathrm{h}(s) & \text{if } m = 1 \\ \frac{1}{m-1}s^m & \text{if } m > 1 \end{cases}$$

and h is given as

$$\mathrm{h}(s) := \begin{cases} s\log(s) & \text{for } s > 0 \\ 0 & \text{for } s = 0\,. \end{cases} \tag{5.1.4}$$

In this chapter we assume $\mathrm{v} \in C^2(\Omega)$ and $\mathrm{w} \in C^2(\mathbb{R}^d)$. Furthermore w is to be even, implying $\nabla\mathrm{w}(-z) = -\nabla\mathrm{w}(z)$. We will only be considering one cost function, that is $\mathbf{c}(s) = \frac{1}{2}\|s\|^2$. The notation for the corresponding optimal transport distance and its entropic regularization will be $\mathbf{T}_{2,\tau}$ and $\mathbf{T}_{(2,\tau,\varepsilon)}$ respectively. Let us adapt (1.0.3)

$$\mathbf{T}_{2,\tau}(\rho,\mu) = \inf_{\gamma\in\Pi(\rho,\mu)} \iint_{\Omega^2} \frac{1}{2}\left\|\frac{x-y}{\tau}\right\|_2^2 \mathrm{d}\gamma(x,y)\,. \tag{5.1.5}$$

and (1.0.10)

$$\mathbf{T}_{(2,\tau,\varepsilon)}(\rho,\mu) = \inf_{\gamma\in\Pi(\rho,\mu)} \iint_{\Omega^2} \frac{1}{2}\left\|\frac{x-y}{\tau}\right\|_2^2 \mathrm{d}\gamma(x,y) + \varepsilon\mathcal{H}(\gamma) \tag{5.1.6}$$

to our specific cost $\mathbf{c}(s) = \frac{1}{2}\|s\|_2^2$. Recall the definition of the (negative) entropy functional $\mathcal{H}$ from (1.0.9): Let $\gamma \in \mathscr{P}(\Omega^2)$ be some probability measure on Ω^2 then

$$\mathcal{H}(\gamma) = \iint_{\Omega^2} \gamma(x,y)\log(\gamma(x,y))\,\mathrm{d}(x,y)$$

if γ admits a density w.r.t. Lebesgue measure and $+\infty$ otherwise and define $\mathcal{H}(\rho)$ analogously for ρ, a probability measures in $\mathscr{P}(\Omega)$.

Note that, as opposed to general cost functions, we can extract the time-step parameter τ completely from the entropic regularized optimal transport, after absorbing τ^2 in ε (c.f. (5.1.12) which assumes $\varepsilon \le C\tau^2$), by

$$\begin{aligned}\tau\mathbf{T}_{(2,\tau,\varepsilon)}(\rho,\mu) &= \tau\left(\inf_{\gamma\in\Pi(\rho,\mu)} \iint_{\Omega^2} \frac{1}{2}\left\|\frac{x-y}{\tau}\right\|_2^2 \mathrm{d}\gamma(x,y) + \varepsilon\mathcal{H}(\gamma)\right)\\ &= \frac{1}{\tau}\left(\inf_{\gamma\in\Pi(\rho,\mu)} \iint_{\Omega^2} \frac{1}{2}\|x-y\|_2^2\,\mathrm{d}\gamma(x,y) + \varepsilon\mathcal{H}(\gamma)\right) =: \frac{1}{\tau}\mathbf{T}_\varepsilon^2(\rho,\mu)\;.\end{aligned}$$

Analogously

$$\tau\mathbf{T}_{2,\tau}(\rho,\mu) = \frac{1}{\tau}\iint_{\Omega^2} \frac{1}{2}\|x-y\|_2^2\,\mathrm{d}\gamma(x,y) =: \frac{1}{\tau}\mathbf{T}_2^2(\rho,\mu)\;.$$

Let us recall the recursion that was introduced in (1.0.16) to receive approximations $\rho_{\tau,\varepsilon}^{(n)}$ for a solution $\rho(n\tau)$ and apply the just found possibility to extract τ from the optimal transport distances. We then arrive at the scheme we will be investigating in this chapter. Given a pair of initial data $\rho_{\tau,\varepsilon}^{(-1)}, \rho_{\tau,\varepsilon}^{(0)}$, minimize recursively the functional

$$\Phi_{\tau,\varepsilon}^n(\rho) := \Phi_{\tau,\varepsilon}(\tau;\rho_{\tau,\varepsilon}^{(n-1)},\rho_{\tau,\varepsilon}^{(n-2)},\rho) := \frac{1}{\tau}\left(\mathbf{T}_\varepsilon^2(\rho,\rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\mathbf{T}_\varepsilon^2(\rho,\rho_{\tau,\varepsilon}^{(n-2)})\right) + \mathcal{E}(\rho)$$

in ρ over $\mathscr{P}(\Omega)$ to receive $\rho_{\tau,\varepsilon}^{(n)}$. Here $\mathcal{E}$ is again given as

$$\mathcal{E}(\rho) = \mathcal{U}(\rho) + \mathcal{V}(\rho) + \mathcal{W}(\rho)$$

if ρ is absolutely continuous w.r.t. Lebesgue measure and $+\infty$ if it is not. We recall the definition of the functionals

$$\mathcal{U}(\rho) = \int_\Omega \mathrm{u}(\rho(x))\,\mathrm{d}x\;,\qquad \mathcal{V}(\rho) = \int_\Omega \mathrm{v}(x)\rho(x)\,\mathrm{d}x \quad\text{and}\quad \mathcal{W}(\rho) = \int_{\Omega^2} \rho(x)\mathrm{w}(x-y)\rho(y)\,\mathrm{d}y\,\mathrm{d}x$$

and the relative entropy from (2.3.2) which reads for $\gamma,\eta \in \mathscr{P}(\Omega^2)$

$$\mathcal{H}(\gamma \mid \eta) = \iint_{\Omega^2} \log\left(\frac{\mathrm{d}\gamma}{\mathrm{d}\eta}\right)\mathrm{d}\gamma(x,y) \tag{5.1.7}$$

if $\gamma << \eta$ and $+\infty$ otherwise. Here $\frac{\mathrm{d}\gamma}{\mathrm{d}\eta}$ denotes the Radon-Nikodym derivative of γ w.r.t. η. Again, this is definined for probability measures in $\mathscr{P}(\Omega)$ analogously.

Finally, with $\mathcal{L}^d$, the d-dimensional Lebesgue measure, we note

$$\mathcal{H}(\gamma) = \mathcal{H}(\gamma \mid \mathcal{L}^d)\;.$$

5.1.2 The main result

Let us specify the recursion's initial data, $\rho_{\tau,\varepsilon}^{(0)}$ and $\rho_{\tau,\varepsilon}^{(-1)}$. They are assumed to be given as absolutely continuous probability measures approximating ρ^0 in the Wasserstein-2-distance with order τ, that is to say

$$\mathbf{T}_2^2(\rho_{\tau,\varepsilon}^{(-1)}, \rho^0), \mathbf{T}_2^2(\rho_{\tau,\varepsilon}^{(0)}, \rho^0) = \mathcal{O}(\tau) \quad \text{for } \tau \searrow 0 \, . \tag{5.1.8}$$

Furthermore we assume both to have uniform bounded entropy, and energy i.e.

$$\mathcal{H}(\rho_{\tau,\varepsilon}^{(-1)}), \mathcal{H}(\rho_{\tau,\varepsilon}^{(0)}), \mathcal{E}(\rho_{\tau,\varepsilon}^{(-1)}), \mathcal{E}(\rho_{\tau,\varepsilon}^{(0)}) \leq C < \infty \, . \tag{5.1.9}$$

Definition 5.1 (The in-time approximation)**.** *Let $\tau, \varepsilon > 0$ be the time-step and entropic regularization parameter, $T > 0$ be a time horizon and let $N := \lceil \frac{T}{\tau} \rceil$. Let $\rho_{\tau,\varepsilon}^{(-1)}, \rho_{\tau,\varepsilon}^{(0)}$ be given initial data with the properties described above. Let furthermore $\tau, \varepsilon > 0$ be the time-step and entropic regularization parameter. Define the sequence $\rho_{\tau,\varepsilon}^{(n)}$, $n = -1, 0, 1, \ldots, N$ with the initial values given above and the recursion*

$$\rho_{\tau,\varepsilon}^{(n)} \in \operatorname*{arg\,min}_{\rho \in \mathscr{P}(\Omega)} \Phi_{\tau,\varepsilon}^n(\rho) \tag{5.1.10}$$

inducing an in-time approximation $\rho_{\tau,\varepsilon} : [0,T] \to \mathscr{P}(\Omega)$ for a solution of (5.1.1) *by means of piecewise constant interpolation*

$$\rho_{\tau,\varepsilon} : [0,T] \to \mathscr{P}^{ac}(\Omega) \quad \textit{with} \quad \rho_{\tau,\varepsilon}(t) = \rho_{\tau,\varepsilon}^{(n)} \qquad \textit{for } t \in ((n-1,\tau, n\tau] \, . \tag{5.1.11}$$

Let us state the main result of this chapter.

Theorem 4. *Let C be some positive constant. Let $\tau, \varepsilon > 0$ go to zero in a way that satisfies*

$$0 \leq \varepsilon, \varepsilon \, |\log(\varepsilon)| \leq C\tau^2 \, . \tag{5.1.12}$$

Let furthermore $\rho_{\tau,\varepsilon}^{(-1)}, \rho_{\tau,\varepsilon}^{(0)}$ be sequences of suitable approximate initial data in particular satisfying (5.1.8) *and* (5.1.9)*.*

Then the $(\rho_{\tau,\varepsilon}^{(n)})_n$ spawned by the entropic regularized BDF 2 scheme and $\rho_{\tau,\varepsilon}$ defined in Definition 5.1 *have the following properties.*

1. *The sequence $\rho_{\tau,\varepsilon}^{(n)}$ is well defined, that is to say $\Phi_{\tau,\varepsilon}^n(\rho)$ in* (5.1.10) *has a minimizer for each n.*
2. *The sequence of approximate solutions $\rho_{\tau,\varepsilon}$ possesses a subsequence converging w.r.t. strong $L^m((0,T)\times\Omega)$-topology to a limit curve $\rho_* \in L^m((0,T)\times\Omega)$.*
3. *Furthermore ρ_* satisfies a weak formulation of* (5.1.1)*: For every $\varphi \in C_c^\infty((0,T)\times\Omega)$, ρ_* satisfies*

$$\int_0^T \int_\Omega \rho_*(t,x) \partial_t \varphi(t,x) \, \mathrm{d}x \, \mathrm{d}t$$
$$= - \int_0^T \int_\Omega p(\rho_*(t,x)) \Delta\varphi(t,x) \, \mathrm{d}x \, \mathrm{d}t + \int_0^T \int_\Omega \langle \rho_*(t,x) \nabla[\mathrm{v}(x) + (\mathrm{w} * \rho_*)(t,x)], \nabla\varphi(t,x) \rangle \, \mathrm{d}x \, \mathrm{d}t \, .$$

4. *Finally ρ_* is 1/2-Hölder-continuous in time w.r.t. the quadratic optimal transport distance $\mathbf{T}_2$ and assumes the initial value ρ^0, the limit of $\rho_{\tau,\varepsilon}^{(-1)}$ and $\rho_{\tau,\varepsilon}^{(0)}$ given above, continuously w.r.t. $\mathbf{T}_2$, i.e.*

$$\lim_{t \searrow 0} \mathbf{T}_2(\rho_*(t), \rho^0) = 0 \, .$$

The rest of this chapter will consist of proving *Theorem 4* showing the well-definedness of the recursion (5.1.10) in **Section 5.3** followed by some properties of the sequence $\rho_{\tau,\varepsilon}^{(n)}$ in **Section 5.4** and **Section 5.5**. Finally we establish convergence of $\rho_{\tau,\varepsilon}$ and show that the limit curve is indeed a solution of (5.1.1) in **Section 5.6**. So the first claim of the above theorem is shown in **Section 5.3** . The last three are shown in **Section 5.6** .

5.2 Preliminary results

We will show some properties of $\mathbf{T}_\varepsilon$ now. First of all, taking the infimum in the definition is actually a minimization, as was shown for example in [8, Proposition 2.3] and recalled in *Proposition* **??** in the preliminaries of this thesis. Let us restate

Proposition 5.2 (Adapted from [8]). *Let $\rho, \mu \in \mathscr{P}^{ac}(\Omega)$ with $\mathcal{H}(\rho), \mathcal{H}(\mu) < \infty$. Then there is exactly one minimizing γ_* such that*

$$\mathbf{T}_\varepsilon^2(\rho,\mu) = \iint_{\Omega^2} \|x-y\|_2^2 \, \mathrm{d}\gamma_*(x,y) + \varepsilon \mathcal{H}(\gamma_*) \ .$$

5.2.1 Bounding $\mathbf{T}_\varepsilon$ with $\mathbf{T}$.

This part is concerned with bounds for $\mathbf{T}_\varepsilon$. We begin with two bounds that can be received by brief calculations. Note that we will sometimes use the abbreviation

$$\langle \mathbf{c}, \gamma \rangle := \iint_{\Omega^2} \|x-y\|_2^2 \, \mathrm{d}\gamma(x,y) \ .$$

Lemma 5.3. *Let $\rho, \mu \in \mathscr{P}^{ac}(\Omega)$ with finite entropy $\mathcal{H}(\rho), \mathcal{H}(\mu) < \infty$. Then we can bound $\mathbf{T}_\varepsilon^2(\rho,\mu)$ from below by*

$$\mathbf{T}_\varepsilon^2(\rho,\mu) \geq \mathbf{T}^2(\rho,\mu) + \varepsilon\mathcal{H}(\gamma_\varepsilon) \geq -\varepsilon \log(\left|\Omega^2\right|) \tag{5.2.1}$$

where γ_ε denotes the optimal transport plan in $\mathbf{T}_\varepsilon(\rho,\mu)$.

Proof. Let us denote the optimal transport plan in the unregularized OT problem $\mathbf{T}_2^2(\rho,\mu)$ as γ_*. The inequality

$$\mathbf{T}_\varepsilon^2(\rho,\mu) \geq \mathbf{T}^2(\rho,\mu) + \varepsilon\mathcal{H}(\gamma_\varepsilon)$$

can then be seen immediately when considering that $\langle \mathbf{c}, \gamma_* \rangle \leq \langle \mathbf{c}, \gamma_\varepsilon \rangle$ by the optimality of γ_* and since γ_ε is a competitor in the minimization of $\mathbf{T}_2^2(\rho,\mu)$.

To arrive at the second inequality we will invoke $\mathbf{T}^2(\rho,\mu) \geq 0$ and Jensen's inequality together with h being convex. Indeed

$$\begin{aligned}
\mathbf{T}^2(\rho,\mu) + \varepsilon\mathcal{H}(\gamma_\varepsilon) &\geq \varepsilon \left|\Omega^2\right| \iint_{\Omega^2} \mathrm{h}(\gamma(x,y)) \frac{1}{|\Omega^2|} \, \mathrm{d}(x,y) \\
&\geq \varepsilon \left|\Omega^2\right| \mathrm{h}\left(\iint_{\Omega^2} \gamma(x,y) \frac{1}{|\Omega^2|} \, \mathrm{d}(x,y) \right) \\
&= -\varepsilon \log\left(\left|\Omega^2\right|\right) \ .
\end{aligned}$$

holds. In particular we note

$$\mathcal{H}(\rho) \geq -\log(|\Omega|) \tag{5.2.2}$$

holds for every $\rho \in \mathscr{P}(\Omega)$. □

Bounding $\mathbf{T}_\varepsilon$ with the block approximation

Bounding $\mathbf{T}_\varepsilon$ from above with $\mathbf{T}_2$ will need more work than the bound from below and a specific construction. We will establish the following proposition.

Proposition 5.4. *Let $\rho, \mu \in \mathscr{P}^{ac}(\Omega) \cap L\log L(\Omega)$. Then*

$$\mathbf{T}_\varepsilon^2(\rho,\mu) \leq \varepsilon\hat{C} + 2d\varepsilon\log\varepsilon + \mathbf{T}_2^2(\rho,\mu) \tag{5.2.3}$$

holds.

Proof. Since we cannot use the unregularized transport plan γ_* as a competitor on $\mathbf{T}_\varepsilon$ to arrive at suitable estimates because it has infinite entropy, we will, by *block approximation* construct a specific γ_δ such that, on the one hand, we can control $\varepsilon\mathcal{H}(\gamma_\delta)$ and on the other hand admits the limit $\langle \mathbf{c}, \gamma_\delta\rangle \to \langle \mathbf{c}, \gamma_*\rangle$ holds in a suitable order of δ. This construction has already been used in [8, Definition 2.9], where it received its name „block approximation“, and was used in the proof of Γ-convergence to construct a recovery sequence. We want to mention beforehand that ultimately we will choose $\delta = \varepsilon$, but since this construction can be made general, we will carry it out for a general $\delta > 0$.

The calculations of this proof, up until the estimate for the entropy of γ_δ, follow along the lines of section 2 of [8].

We begin with considering the tessellation $\mathcal{Q}_\delta^\Omega$ of Ω given as

$$\mathcal{Q}_\delta^\Omega = \left\{\delta(\{\mathbf{j}\} + K) \cap \Omega \mid \mathbf{j} \in \mathbb{Z}^d\right\} \quad \text{where} \quad K := [-\tfrac{1}{2}, \tfrac{1}{2})^d$$

and $\mathcal{Q}_\delta^{\Omega^2} := \{Q \times P \mid Q, P \in \mathcal{Q}_\delta^\Omega\}$.

Now let γ_* be the optimal transport plan of $\mathbf{T}_2(\rho,\mu)$. Then we can reshape γ_* on each $Q \in \mathcal{Q}_\delta^{\Omega^2}$ in such a way that the marginals remain unchanged while the entropy will have a finite value.

For readability and to ease calculations later on, let the tessellation of Ω be enumerated $\mathcal{Q}_\delta^\Omega = \{Q_i\}_i$ where i lies in a suitable subset of $\mathbb{N}$ and let $\mathcal{Q}_\delta^{\Omega^2} = \{Q_{i,j}\}_{i,j}$ be enumerated accordingly.

The construction on one $Q_{i,j} \in \mathcal{Q}_\delta^{\Omega^2}$ is carried out here representatively. We take the restrictions of ρ and μ to Q_i and Q_j respectively and rescale them to be probability measures again. Then we take their product measure, a probability measure on Ω^2 with support in the closure of $Q_i \times Q_j$, and finally rescale it to have mass $\gamma_*(Q_i \times Q_j)$. Repeating this procedure on each $Q_i \times Q_j$ in $\mathcal{Q}_\delta^{\Omega^2}$ we arrive at γ_δ which is given as

$$\gamma_\delta(x,y) = \gamma_*(Q_i \times Q_j)\frac{\rho(x)\mu(y)}{\rho(Q_i)\mu(Q_j)} \quad \text{if } (x,y) \in Q_i \times Q_j\ .$$

Furthermore we set $\gamma_\delta = 0$ on $Q_i \times Q_j$ if either $\rho(Q_i)$ or $\mu(Q_j)$ equals 0. Let

$$\mathcal{J}_\rho := \{i \mid Q_i \in \mathcal{Q}_\delta^\Omega \text{ and } \rho(Q_i) \neq 0\}$$

and $\mathcal{J}_\mu$ accordingly, as well as $\mathcal{J}_{\rho,\mu}^2 := \mathcal{J}_\rho \times \mathcal{J}_\mu$. This way we can write

$$\gamma_\delta(x,y) = \sum_{(i,j)\in\mathcal{J}_{\rho,\mu}^2} \gamma_*(Q_i \times Q_j)\frac{\rho(x)\mu(y)}{\rho(Q_i)\mu(Q_j)}\mathbf{1}_{Q_i\times Q_j}(x,y)$$

We want to show some properties of γ_δ now. First and foremost, γ_δ lies in $\mathscr{P}^{ac}(\Omega)$ and is a competitor in $\mathbf{T}_\varepsilon(\rho,\mu)$, that is to say its marginals equal ρ and μ respectively and it has finite entropy.

We can see that the marginal constrained is satisfied since for some $x \in \Omega$ either $\rho(Q_i) = 0$ on the corresponding $Q_i \ni x$ and consequently $\int_\Omega \gamma_\delta(x,y)\,\mathrm{d}y = 0 = \rho(x)$ by construction of γ_δ or, if $\rho(Q_i) > 0$

we can calculate

$$\begin{aligned}\int_\Omega \gamma_\delta(x,y)\,\mathrm{d}y &= \sum_{j\in\mathcal{J}_\mu} \frac{\gamma_*(Q_i\times Q_j)}{\rho(Q_i)}\rho(x)\int_{Q_j}\frac{\mu(y)}{\mu(Q_j)}\,\mathrm{d}y\\ &= \sum_{j\in\mathcal{J}_\mu} \frac{\gamma_*(Q_i\times Q_j)}{\rho(Q_i)}\rho(x)\\ &= \frac{\gamma_*(Q_i\times\Omega)}{\rho(Q_i)}\rho(x)\\ &= \rho(x)\end{aligned}$$

and analogously the second marginal of $(\mathrm{P}_y)_\#\gamma_\delta = \mu$ is recovered.

We will formulate an intermediate result for the entropy next.

Lemma 5.5. *For the γ_δ constructed above*

$$\mathcal{H}(\gamma_\delta) \le \mathcal{H}(\rho) + \mathcal{H}(\mu) + 2\log(|\Omega|) + 2d\log(\delta) \tag{5.2.4}$$

holds.

Proof of Lemma 5.5. To show this, we have to take several steps.

Let us begin with decomposing

$$\mathrm{h}\left(\gamma_*(Q_i\times Q_j)\frac{\rho(x)\mu(y)}{\rho(Q_i)\mu(Q_j)}\right) = \gamma_*(Q_i\times Q_j)\frac{\rho(x)\mu(y)}{\rho(Q_i)\mu(Q_j)}\left(\log(\gamma_*(Q_i\times Q_j)) + \log\left(\frac{\rho(x)\mu(y)}{\rho(Q_i)\mu(Q_j)}\right)\right)$$

which is justified by $\mathrm{h}(sr) = sr\log(sr) = sr(\log(s)+\log(r))$. As integrand, this yields

$$\begin{aligned}\mathcal{H}(\gamma_\delta) &= \iint_{\Omega^2} \mathrm{h}(\gamma_\delta(x,y))\,\mathrm{d}(x,y)\\ &= \sum_{(i,j)\in\mathcal{J}^2_{\rho,\mu}} \iint_{Q_i\times Q_j} \gamma_*(Q_i\times Q_j)\frac{\rho(x)\mu(y)}{\rho(Q_i)\mu(Q_j)}\left(\log(\gamma_*(Q_i\times Q_j)) + \log\left(\frac{\rho(x)\mu(y)}{\rho(Q_i)\mu(Q_j)}\right)\right)\mathrm{d}(x,y)\\ &= \sum_{(i,j)\in\mathcal{J}^2_{\rho,\mu}} \gamma_*(Q_i\times Q_j)\log(\gamma_*(Q_i\times Q_j))\frac{\iint_{Q_i\times Q_j}\rho(x)\mu(y)\,\mathrm{d}(x,y)}{\rho(Q_i)\mu(Q_j)}\\ &\quad+ \sum_{(i,j)\in\mathcal{J}^2_{\rho,\mu}} \gamma_*(Q_i\times Q_j)\iint_{Q_i\times Q_j}\frac{\rho(x)\mu(y)}{\rho(Q_i)\mu(Q_j)}\log\left(\frac{\rho(x)\mu(y)}{\rho(Q_i)\mu(Q_j)}\right)\mathrm{d}(x,y)\,.\end{aligned}$$

For the first sum we can easily see

$$\sum_{(i,j)\in\mathcal{J}^2_{\rho,\mu}} \gamma_*(Q_i\times Q_j)\log(\gamma_*(Q_i\times Q_j))\frac{\iint_{Q_i\times Q_j}\rho(x)\mu(y)\,\mathrm{d}(x,y)}{\rho(Q_i)\mu(Q_j)} = \sum_{(i,j)\in\mathcal{J}^2_{\rho,\mu}} \gamma_*(Q_i\times Q_j)\log(\gamma_*(Q_i\times Q_j)) \le 0$$

since $\log(\gamma_*(Q_i\times Q_j)) \le 0$.

The argument of the second sum can be split up one more time

$$\begin{aligned}
&\sum_{(i,j)\in\mathcal{J}^2_{\rho,\mu}} \gamma_*(Q_i\times Q_j)\iint_{Q_i\times Q_j}\frac{\rho(x)\mu(y)}{\rho(Q_i)\mu(Q_j)}\log\left(\frac{\rho(x)\mu(y)}{\rho(Q_i)\mu(Q_j)}\right)\mathrm{d}(x,y)\\
&\quad=\sum_{(i,j)\in\mathcal{J}^2_{\rho,\mu}} \gamma_*(Q_i\times Q_j)\Bigg(\int_{Q_j}\frac{\mu(y)}{\mu(Q_j)}\int_{Q_i}\frac{\rho(x)}{\rho(Q_i)}\log\left(\frac{\rho(x)}{\rho(Q_i)}\right)\mathrm{d}x\,\mathrm{d}y\\
&\qquad+\int_{Q_i}\frac{\rho(x)}{\rho(Q_i)}\int_{Q_j}\frac{\mu(y)}{\mu(Q_j)}\log\left(\frac{\mu(y)}{\mu(Q_j)}\right)\mathrm{d}y\,\mathrm{d}x\Bigg)\\
&\quad=\sum_{(i,j)\in\mathcal{J}^2_{\rho,\mu}} \gamma_*(Q_i\times Q_j)\Bigg(\int_{Q_i}\frac{\rho(x)}{\rho(Q_i)}\log\left(\frac{\rho(x)}{\rho(Q_i)}\right)\mathrm{d}x+\int_{Q_j}\frac{\mu(y)}{\mu(Q_j)}\log\left(\frac{\mu(y)}{\mu(Q_j)}\right)\mathrm{d}y\Bigg)
\end{aligned}$$

Let us take a closer look at

$$\begin{aligned}
&\sum_{(i,j)\in\mathcal{J}^2_{\rho,\mu}} \gamma_*(Q_i\times Q_j)\int_{Q_i}\frac{\rho(x)}{\rho(Q_i)}\log\left(\frac{\rho(x)}{\rho(Q_i)}\right)\mathrm{d}x\\
&\qquad=\sum_{i\in\mathcal{J}_\rho}\sum_{j\in\mathcal{J}_\mu}\frac{\gamma_*(Q_i\times Q_j)}{\rho(Q_i)}\int_{Q_i}\rho(x)\log(\rho(x))\,\mathrm{d}x-\gamma_*(Q_i\times Q_j)\log(\rho(Q_i))\\
&\qquad=\sum_{i\in\mathcal{J}_\rho}\int_{Q_i}\rho(x)\log(\rho(x))\,\mathrm{d}x-\rho(Q_i)\log(\rho(Q_i))\\
&\qquad=\mathcal{H}(\rho)-\sum_{i\in\mathcal{J}_\rho}\rho(Q_i)\log(\rho(Q_i))\ .
\end{aligned}$$

Therefore, plugging everything back together, we arrive at

$$\begin{aligned}
&\sum_{(i,j)\in\mathcal{J}^2_{\rho,\mu}} \gamma_*(Q_i\times Q_j)\iint_{Q_i\times Q_j}\frac{\rho(x)\mu(y)}{\rho(Q_i)\mu(Q_j)}\log\left(\frac{\rho(x)\mu(y)}{\rho(Q_i)\mu(Q_j)}\right)\mathrm{d}(x,y)\\
&\qquad=\mathcal{H}(\rho)+\mathcal{H}(\mu)-\sum_{i\in\mathcal{J}_\rho}\rho(Q_i)\log(\rho(Q_i))-\sum_{j\in\mathcal{J}_\mu}\mu(Q_j)\log(\mu(Q_j))
\end{aligned}$$

In a final step, in order to establish the sought for estimate on $\mathcal{H}(\gamma_\delta)$, we want to bound $\sum_{i\in\mathcal{J}_\rho}\rho(Q_i)\log(\rho(Q_i))$ and $\sum_{j\in\mathcal{J}_\mu}\mu(Q_j)\log(\mu(Q_j))$ from below to arrive at the sought for inequality. To that end, replace ρ by a piecewise constant approximation on $\mathcal{Q}^\Omega_\delta$ defined as $\overline{\rho}(x)=\sum_{i\in\mathcal{J}}\rho(Q_i)\frac{1}{\delta^d}\mathbf{1}_{Q_i}(x)$ and then use $\overline{\rho}(Q_i)=\rho(Q_i)$ by construction and the bound from below for $\mathcal{H}$ (c.f. (5.2.2)), to arrive at

$$\begin{aligned}
\sum_{i\in\mathcal{J}_\rho}\rho(Q_i)\log(\rho(Q_i))&=\sum_{i\in\mathcal{J}_\rho}\overline{\rho}(Q_i)\log(\overline{\rho}(Q_i))\\
&=\int_\Omega\overline{\rho}(x)\log\left(\overline{\rho}(x)\frac{1}{\delta^d}\right)\mathrm{d}x\\
&=\int_\Omega\overline{\rho}(x)\log(\overline{\rho}(x))\,\mathrm{d}x+\int_\Omega\overline{\rho}(x)\log\left(\frac{1}{\delta^d}\right)\mathrm{d}x\\
&\ge-\log(|\Omega|)-d\log(\delta)\ .
\end{aligned}$$

Analogously we receive

$$\sum_{j\in\mathcal{J}_\mu} \mu(Q_j)\log(\mu(Q_j)) \geq -\log(|\Omega|) - d\log(\delta)$$

and plugging everything back together we arrive at

$$\mathcal{H}(\gamma_\delta) \leq \mathcal{H}(\rho) + \mathcal{H}(\mu) + 2\log(|\Omega|) + 2d\log(\delta)$$

which is the estimate (5.2.4) we wanted to show. □

Finally, we want to compare $\langle \mathbf{c}, \gamma_\delta\rangle$ with the optimal $\mathbf{T}_2^2(\rho,\mu) = \langle \mathbf{c}, \gamma_*\rangle$. To that end, let us consider some $Q_i \times Q_j$ where $(i,j) \in \mathcal{J}^2_{\rho,\mu}$ and the optimal transport problem $\mathbf{T}_2^2(\gamma_{(*,i,j)}, \gamma_{(\delta,i,j)})$ where $\gamma_{(\cdot,i,j)} = \gamma_{(\cdot)}\ |_{Q_i\times Q_j}$. In other words, we consider the problem to transport the mass on $Q_i \times Q_j$ distributed according to γ_* to the same mass distributed according to γ_δ.

Since $\gamma_{(\delta,i,j)}$ has by construction a density w.r.t. Lebesgue measure, i.e. it does not give mass to small sets, so the Monge problem of transporting the mass of $\mathbf{T}_2^2(\gamma_{(*,i,j)}, \gamma_{(\delta,i,j)})$ has a solution and we receive a measurable map $S : Q_i \times Q_j \to Q_i \times Q_j$ pushing $\gamma_{(\delta,i,j)}$ forward to $\gamma_{(*,i,j)}$, that is to say $S_\#\gamma_{(\delta,i,j)} = \gamma_{(*,i,j)}$. That this map is indeed realizing a optimal transport is secondary for us.

By the domain and image of $S(x,y) = (S_1(x,y), S_2(x,y))^T \in Q_i \times Q_j$, by $Q_i \times Q_j \subset \Omega^2$ and by $\|\cdot\|_2^2$ being Lipschitz with constant $L > 0$ on the bounded domain, we receive

$$\left|\|S_1(x,y) - S_2(x,y)\|_2^2 - \|x-y\|_2^2\right| \leq L\,\|(S_1(x,y)-x) - (S_2(x,y)-y)\|_2 \leq L2\sqrt{d}\delta$$

and with it we can calculate

$$\begin{aligned}
&\left|\left\langle \mathbf{c}, \gamma_{(\delta,i,j)}\right\rangle - \left\langle \mathbf{c}, \gamma_{(*,i,j)}\right\rangle\right| \\
&\qquad = \left|\iint_{Q_i\times Q_j} \|x-y\|_2^2 \ \mathrm{d}[S_\#\gamma_{(*,i,j)}](x,y) - \left\langle \mathbf{c}, \gamma_{(*,i,j)}\right\rangle\right| \\
&\qquad \leq \iint_{Q_i\times Q_j} \left|\|S_1(x,y) - S_2(x-y)\|_2^2 - \|x-y\|_2^2\right| \mathrm{d}\gamma_{(*,i,j)}(x,y) \\
&\qquad \leq 2\sqrt{d}L\delta\,|Q_i|\,|Q_j| \ .
\end{aligned}$$

Summing up to receive a result for γ_* and γ_δ we arrive at

$$|\langle \mathbf{c}, \gamma_\delta\rangle - \langle \mathbf{c}, \gamma_*\rangle| \leq \sum_{(i,j)\in\mathcal{J}^2_{\rho,\mu}} 2\sqrt{d}L\delta\,|Q_i|\,|Q_j| \leq 2\sqrt{d}L\,|\Omega|^2\,\delta \ .$$

We will combine the above results now in an estimate for $\mathbf{T}_\varepsilon(\rho,\mu)$ from above by $\mathbf{T}_2(\rho,\mu)$.

Let γ_ε denote the optimal entropic regularized transport plan and γ_* the optimal transport plan w.r.t. to quadratic distance. Let furthermore γ_δ be the block approximation of γ_* as above. Then we can calculate, using the minimizing property of γ_ε together with $\gamma_\delta \in \Pi(\rho,\mu)$

$$\begin{aligned}
\mathbf{T}_\varepsilon^2(\rho,\mu) &= \langle \mathbf{c}, \gamma_\varepsilon\rangle + \varepsilon\mathcal{H}(\gamma_\varepsilon) \\
&\leq \langle \mathbf{c}, \gamma_\delta\rangle + \varepsilon\mathcal{H}(\gamma_\delta) \\
&= \varepsilon\mathcal{H}(\gamma_\delta) + \langle \mathbf{c}, \gamma_\delta\rangle - \langle \mathbf{c}, \gamma_*\rangle + \mathbf{T}_2^2(\rho,\mu) \\
&\leq \varepsilon\left[\mathcal{H}(\rho) + \mathcal{H}(\mu) + 2\log(|\Omega|) + 2d\log(\delta)\right] \\
&\qquad + 2\sqrt{d}L\,|\Omega|^2\,\delta + \mathbf{T}_2^2(\rho,\mu) \ .
\end{aligned}$$

Since δ is at our disposal, choose $\delta = \varepsilon$ and abbreviate

$$\hat{C} := \mathcal{H}(\rho) + \mathcal{H}(\mu) + 2\log(|\Omega|) + 2\sqrt{d}L\,|\Omega|^2$$

to arrive at

$$\mathbf{T}_\varepsilon^2(\rho,\mu) \le \varepsilon\hat{C} + 2d\varepsilon\log\varepsilon + \mathbf{T}_2^2(\rho,\mu)\;. \qquad \square$$

Bounding $\mathbf{T}_\varepsilon$ with the tensor-product plan.

The *tensor-product plan* $\rho\otimes\mu \in \mathscr{P}(\rho,\mu)$, minimizes the negative entropy $\mathcal{H}(\gamma)$ among the set of transport plans $\gamma \in \mathscr{P}(\rho,\mu)$. Indeed, as was shown in [15, Theorem 1.2], the entropic regularized OT plans converge narrowly to the tensor-product plan $\rho\otimes\mu$ for $\varepsilon\to\infty$.

Since it is on the one hand the minimizer of $\mathcal{H}$ on the set of transport plans from ρ to μ and on the other hand a feasible candidate in the minimization problem $\mathbf{T}_\varepsilon^2(\rho,\mu)$, we arrive at the following estimates.

Lemma 5.6. *Let $\rho,\mu \in \mathscr{P}^{ac}(\Omega)$ with finite entropy, $\varepsilon > 0$ and γ_ε the optimal transport plan of $\mathbf{T}_\varepsilon^2(\rho,\mu)$. Then*

$$\langle \mathbf{c},\gamma_\varepsilon\rangle + \varepsilon\left(\mathcal{H}(\rho)+\mathcal{H}(\mu)\right) \le \mathbf{T}_\varepsilon^2(\rho,\mu) \le \langle \mathbf{c},\rho\otimes\mu\rangle + \varepsilon\left(\mathcal{H}(\rho)+\mathcal{H}(\mu)\right)$$

holds.

Proof. This is merely a matter of plugging in the tensor-product plan as a minimizer of $\mathcal{H}$, to receive the bound from below, and as a feasible candidate in $\Pi(\rho,\mu)$ for minimization in $\mathbf{T}_\varepsilon$ to receive the bound from above. In both cases we can see

$$\begin{aligned}\mathcal{H}(\rho\otimes\mu) &= \iint_{\Omega^2} \rho\otimes\mu\log(\rho\otimes\mu)\,\mathrm{d}(x,y)\\ &= \int_\Omega \rho\log(\rho)\left(\int_\Omega \mu\,\mathrm{d}y\right)\mathrm{d}x + \int_\Omega \mu\log(\mu)\left(\int_\Omega \rho\,\mathrm{d}x\right)\mathrm{d}y\\ &= \mathcal{H}(\rho)+\mathcal{H}(\mu)\;.\end{aligned} \qquad \square$$

5.2.2 A stability result for the Kantorovich potentials

The entropic regularized OT distance $\mathbf{T}_\varepsilon^2(\rho,\mu)$ admits a decomposition which will be needed in the subsequent proofs of the existence of a minimizer and in establishing the Euler-Lagrange equation. The decomposition consists of rewriting $\mathbf{T}_\varepsilon^2$ as the sum of a „distance“ $\mathbf{S}_\varepsilon^2$ similar to $\mathbf{T}_\varepsilon^2$ and the entropy of the marginals ρ,μ. This new $\mathbf{S}_\varepsilon^2$ will furthermore be shown to be continuous w.r.t. narrow convergence. In the proof of continuity we will make extensively use of properties of $\mathbf{S}_\varepsilon^2$ established in [19].

We begin with recalling *Remark 2.12* which stated

$$\mathbf{T}_\varepsilon^2(\rho,\mu) = \inf_{\gamma\in\Pi(\rho,\mu)} \varepsilon\mathcal{H}(\gamma \mid \mathrm{K}_\varepsilon)\;. \tag{5.2.5}$$

Now we define a similar expression.

Definition 5.7. *Let $\rho,\mu \in \mathscr{P}^{ac}(\Omega)$ with finite entropy. Define*

$$\mathbf{S}_\varepsilon^2(\rho,\mu) := \inf_{\gamma\in\Pi(\rho,\mu)} \varepsilon\mathcal{H}(\gamma \mid \mathrm{K}_\varepsilon\rho\otimes\mu) \tag{5.2.6}$$

(c.f. (5.2.5)*) where*

$$\mathcal{H}(\gamma \mid \mathrm{K}_\varepsilon \rho \otimes \mu) = \iint_{\Omega^2} G(x,y) \log\left(\frac{G(x,y)}{\mathrm{K}_\varepsilon(x,y)}\right) \mathrm{d}(\rho \otimes \mu)$$

holds due to $\gamma = G\rho \otimes \mu$.

Define the dual problem $\mathcal{D}_\varepsilon(\rho,\mu)$ *of* $\mathbf{S}^2_\varepsilon(\rho,\mu)$ *as*

$$\mathcal{D}_\varepsilon(\rho,\mu) := \sup_{\phi,\psi \in L^\infty(\Omega)} D_\varepsilon(\phi,\psi,\rho,\mu) \tag{5.2.7}$$

where the argument of the supremum is defined as

$$D_\varepsilon(\phi,\psi,\rho,\mu) := \int_\Omega \phi(x)\rho(x)\,\mathrm{d}x + \int_\Omega \psi(y)\mu(y)\,\mathrm{d}y - \varepsilon \iint_{\Omega^2} \mathrm{e}^{\frac{\phi(x)+\psi(y)}{\varepsilon}} \mathrm{K}_\varepsilon(x,y)\,\mathrm{d}(\rho\otimes\mu)(x,y)\ .$$

For $(\phi,\psi) \in L^\infty(\Omega)^2$ *define the* Sinkhorn iteration map $\mathbb{S}_\varepsilon$ *by*

$$\mathbb{S}_\varepsilon(\phi,\psi;\rho,\mu) = (\hat{\phi},\hat{\psi})$$

where $\hat{\phi},\hat{\psi}$ *are calculated as follows:*

$$\begin{aligned}\hat{\phi} &:= -\varepsilon \log\left(\int_\Omega \mathrm{e}^{\frac{\psi(y)-\mathbf{c}(x-y)}{\varepsilon}}\,\mathrm{d}\mu(y)\right)\\ \hat{\psi} &:= -\varepsilon \log\left(\int_\Omega \mathrm{e}^{\frac{\hat{\phi}(x)-\mathbf{c}(x-y)}{\varepsilon}}\,\mathrm{d}\rho(x)\right)\ .\end{aligned}$$

Proposition 5.8. *Let* $\rho,\mu \in \mathscr{P}^{ac}(\Omega)$ *with finite entropy. Then the following holds.*

1. $\mathbf{S}_\varepsilon$ *and* $\mathbf{T}^2_\varepsilon$ *are connected by*

   $$\mathbf{T}^2_\varepsilon(\rho,\mu) = \mathbf{S}^2_\varepsilon(\rho,\mu) + \varepsilon\mathcal{H}(\rho) + \varepsilon\mathcal{H}(\mu)$$

   (c.f. Lemma 1.5. *in [19]).*

2. $\mathbf{S}_\varepsilon$ *and* $\mathcal{D}_\varepsilon$ *admit optimal elements* $\gamma_{opt} \in \mathscr{P}^{ac}(\Omega^2) \cap \Pi(\rho,\mu)$ *and* $(\phi^*,\psi^*) \in L^\infty(\Omega)^2$ *where* γ_{opt} *is unique and* (ϕ^*,ψ^*) *is unique up to recalibration* $(\phi^*,\psi^*) \to (\phi^*+a,\psi^*-a)$. *The recalibration can always be chosen such that* $\|\phi^*\|_\infty, \|\psi^*\|_\infty \le \frac{3}{2}\|\mathbf{c}\|_\infty$. *(c.f.* Section 2.*, in particular* Lemma 2.7., Theorem 2.8. *and* Lemma 2.10. *in [19]).*

3. *Duality between* $\mathbf{S}_\varepsilon$ *and* $\mathcal{D}_\varepsilon$ *holds in the following way*

   $$\mathbf{S}^2_\varepsilon(\rho,\mu) = \mathcal{D}_\varepsilon(\rho,\mu)$$

   and, let γ_{opt}, (ϕ^*,ψ^*) *be optimal in* $\mathbf{S}_\varepsilon, \mathcal{D}_\varepsilon$, *then*

   $$\mathbb{S}_\varepsilon(\phi^*,\psi^*,\rho,\mu) = (\phi^*,\psi^*)$$

   holds and this is a sufficient condition for (ϕ^*,ψ^*) *to be optimal in* $\mathcal{D}_\varepsilon(\rho,\mu)$. *Furthermore*

   $$\gamma_{opt}(x,y) = \mathrm{e}^{\phi^*(x)/\varepsilon}\mathrm{K}_\varepsilon(x,y)\mathrm{e}^{\psi^*(y)/\varepsilon}(\rho\otimes\mu)(x,y)$$

   holds (c.f. Lemma 2.10. *and* Proposition 2.11. *in [19]).*

Remark 5.9. *We want to make* Proposition 5.8.1. *plausible with a short formal calculation while skipping the measure theoretic considerations. Let $\gamma \in \mathscr{P}^{ac}(\Omega^2)$ with density $\tilde{G}$ w.r.t. Lebesgue measure and G w.r.t. $\rho \otimes \mu$. By $\rho, \mu \in \mathscr{P}^{ac}(\Omega)$, this entails $\tilde{G}(x,y) = G(x,y)r(x)m(y)$ where r, m denote the densities of ρ and μ respectively. We can now calculate*

$$\begin{aligned}
\mathcal{H}(\gamma \mid \mathrm{K}_\varepsilon) &= \iint_{\Omega^2} \tilde{G}(x,y) \log\left(\frac{\tilde{G}(x,y)}{\mathrm{K}_\varepsilon(x,y)}\right) \mathrm{d}(x,y) \\
&= \iint_{\Omega^2} G(x,y) \log\left(\frac{G(x,y)r(x)m(y)}{\mathrm{K}_\varepsilon(x,y)}\right) \mathrm{d}(\rho\otimes\mu)(x,y) \\
&= \iint_{\Omega^2} G(x,y) \log\left(\frac{G(x,y)}{\mathrm{K}_\varepsilon(x,y)}\right) \mathrm{d}(\rho\otimes\mu)(x,y) + \int_\Omega r(x)\log(r(x))\,\mathrm{d}x + \int_\Omega m(y)\log(m(y))\,\mathrm{d}y \\
&= \mathcal{H}(\gamma \mid \mathrm{K}_\varepsilon(\rho\otimes\mu)) + \mathcal{H}(\rho) + \mathcal{H}(\mu)\,.
\end{aligned}$$

The definitions of $\mathbf{T}^2_\varepsilon$ and $\mathbf{S}^2_\varepsilon$ now imply the result Proposition 5.8.1. .

By a short formal calculation, the duality result in Proposition 5.8.3. *can be made plausible, too. Indeed, let $\gamma \in \Pi(\rho,\mu) \cap \mathscr{P}^{ac}(\Omega^2)$ and let $\phi, \psi \in L^\infty(\Omega)$. Then*

$$\begin{aligned}
&\varepsilon\mathcal{H}(\gamma \mid \mathrm{K}_\varepsilon(\rho\otimes\mu)) \\
&= \iint_{\Omega^2} \Big(\|x-y\|_2^2 + \varepsilon\log(G(x,y)) - \phi(x) - \psi(y)\Big) G(x,y)\,\mathrm{d}(\rho\otimes\mu)(x,y) + \int_\Omega \phi\,\mathrm{d}\rho + \int_\Omega \psi\,\mathrm{d}\mu \\
&\geq -\varepsilon \iint_{\Omega^2} \mathrm{e}^{\frac{\phi(x)+\psi(y)-\|x-y\|_2^2}{\varepsilon}-1}\,\mathrm{d}(\rho\otimes\mu) + \int_\Omega \phi\,\mathrm{d}\rho + \int_\Omega \psi\,\mathrm{d}\mu \\
&= D_\varepsilon(\phi,\psi,\rho,\mu)
\end{aligned}$$

where we absorbed the -1 as ε in one of the two Kantorovich potentials. The inequality in this calculation is justified by noting that $t \mapsto (s+\varepsilon\log(t))t$ is convex for all s and has its unique minimizer at $t = -\mathrm{e}^{-\frac{s}{\varepsilon}-1}$. So identifying $s = \|x-y\|_2^2 - \phi - \psi$ and $t = G \geq 0$ we arrive at the inequality in the calculation above and see furthermore that $\varepsilon\mathcal{H}(\gamma \mid \mathrm{K}_\varepsilon(\rho\otimes\mu)) = D_\varepsilon(\phi,\psi,\rho,\mu)$ iff $\mathrm{d}\gamma = \mathrm{e}^{\frac{\phi+\psi}{\varepsilon}}\mathrm{K}_\varepsilon\,\mathrm{d}(\rho\otimes\mu)$.

With these results at hand, we can state the first consequence towards our stability result.

Lemma 5.10. *Let $\mu, \rho_*, \rho_k \in \mathscr{P}^{ac}(\Omega)$ with uniformly bounded entropy. Let furthermore $\rho_k \rightharpoonup \rho_*$ narrowly in $\mathscr{P}(\Omega)$ and denote a sequence of corresponding optimal Kantorovich potentials (ϕ_k, ψ_k) of $\mathcal{D}_\varepsilon(\rho_k, \mu)$. Note that we assume the Kantorovich potentials to be chosen such that in particular the estimate in* Proposition 5.8.2. *holds.*

Then we can recalibrate the sequence $(\phi_k,\psi_k) \to (\phi_k + a_k, \psi_k - a_k)$ with a sequence $\sup_k |a_k| \leq 3\|\mathbf{c}\|_\infty$ in such a way that the components of the recalibrated sequence converge

$$\lim_{k\to\infty} \phi_k = \phi_* \qquad \textit{and} \qquad \lim_{k\to\infty}\psi_k = \psi_*$$

uniformly. Additionally, the limit (ϕ_*,ψ_*) is an optimal pair in $\mathcal{D}_\varepsilon(\rho_*,\mu)$ such that a relaxed version of the estimate in Proposition 5.8.2. *holds for (ϕ^*,ψ^*), too.*

Proof. We will establish, in addition to $\|\phi_k\|_\infty, \|\psi_k\|_\infty \leq \frac{3}{2}\|\mathbf{c}\|_\infty$, that these sequences are, as sequences of continuous functions on the bounded Ω, uniformly equicontinuous. That will be shown by establishing a uniform Lipschitz constant for both sequences. Consequently, by Arzela-Ascoli, these sequences are compact w.r.t. uniform convergence.

To establish this Lipschitz constant we will follow along the lines of the proof of [19, Proposition 2.4.1].

We can establish, by boundedness of Ω, the Lipschitz estimate

$$\left| \|x_1 - y\|_2^2 - \|x_2 - y\|_2^2 \right| \leq 2\,\mathrm{diam}(\Omega)\,\|x_1 - x_2\|_2$$

where the upper bound can for example be seen by $z_1 = x_1 - y$, $z_2 = x_2 - y$

$$\|z_1\|_2^2 - \|z_2\|_2^2 = (\|z_1\|_2 - \|z_2\|_2)(\|z_1\|_2 + \|z_2\|_2) \leq 2\,\mathrm{diam}(\Omega)\,\|z_1 - z_2\|_2$$

and we used the binomial theorem and the inverse triangle inequality.

By *Proposition 5.8.3.* we have that (ϕ_k, ψ_k) is a fixed point of $\mathbb{S}_\varepsilon(\phi, \psi, \rho_k, \mu)$, and so for our Kantorovich potential ϕ_k we can write

$$\begin{aligned}
\phi_k(x_1) - \phi_k(x_2) &= -\varepsilon\left(\log\left(\int_\Omega \mathrm{e}^{\frac{\psi_k(y) - \mathbf{c}(x_1 - y)}{\varepsilon}}\,\mathrm{d}\mu(y)\right) - \log\left(\int_\Omega \mathrm{e}^{\frac{\psi_k(y) - \mathbf{c}(x_2 - y)}{\varepsilon}}\,\mathrm{d}\mu(y)\right)\right)\\
&= -\varepsilon\left(\log\left(\int_\Omega \mathrm{e}^{\frac{\psi_k(y) - \mathbf{c}(x_2 - y) + \mathbf{c}(x_2 - y) - \mathbf{c}(x_1 - y)}{\varepsilon}}\,\mathrm{d}\mu(y)\right) - \log\left(\int_\Omega \mathrm{e}^{\frac{\psi_k(y) - \mathbf{c}(x_2 - y)}{\varepsilon}}\,\mathrm{d}\mu(y)\right)\right)\\
&\leq 2\,\mathrm{diam}(\Omega)\,\|x_1 - x_2\|_2 - \varepsilon\left(\log\left(\int_\Omega \mathrm{e}^{\frac{\psi_k(y) - \mathbf{c}(x_2 - y)}{\varepsilon}}\,\mathrm{d}\mu(y)\right) - \log\left(\int_\Omega \mathrm{e}^{\frac{\psi_k(y) - \mathbf{c}(x_2 - y)}{\varepsilon}}\,\mathrm{d}\mu(y)\right)\right)\\
&= 2\,\mathrm{diam}(\Omega)\,\|x_1 - x_2\|_2
\end{aligned}$$

and the other way round

$$\phi_k(x_1) - \phi_k(x_2) \geq -2\,\mathrm{diam}(\Omega)\,\|x_1 - x_2\|_2$$

by virtually the same calculation, giving us a Lipschitz constant $L = 2\,\mathrm{diam}(\Omega)$ for ϕ_k not depending on ρ_k, μ or k.

The same calculation holds for ψ_k and so does the estimate. We can therefore conclude that the two sequences ϕ_k, ψ_k are on the one hand uniformly bounded in $\|\cdot\|_\infty$ and on the other hand are uniformly equicontinuous, allowing us to apply Arzela-Ascoli to receive that these sequences are compact w.r.t. uniform convergence. That is to say, that in particular each subsequence of (ϕ_k, ψ_k) has a subsubsequence such that its components converge uniformly to some ϕ_∞, ψ_∞.

We want to show that $(\phi_\infty, \psi_\infty)$ is an optimal pair for $\mathcal{D}_\varepsilon(\rho_*, \mu)$. By *Proposition 5.8.3.* this is equivalent to showing that the limit is a fixed point of $\mathbb{S}_\varepsilon(\cdot, \cdot, \rho_*, \mu)$. We have

$$\begin{aligned}
\phi_\infty &= \lim_{k\to\infty} \phi_k = \lim_{k\to\infty} -\varepsilon\log\left(\int_\Omega \mathrm{e}^{\frac{\psi_k(y) - \mathbf{c}(x-y)}{\varepsilon}}\,\mathrm{d}\mu(y)\right)\\
\psi_\infty &= \lim_{k\to\infty} \psi_k = \lim_{k\to\infty} -\varepsilon\log\left(\int_\Omega \mathrm{e}^{\frac{\phi_k(x) - \mathbf{c}(x-y)}{\varepsilon}}\,\mathrm{d}\rho_k(x)\right)\ .
\end{aligned}$$

So by the integrals being bounded away from 0 and by continuity of log, what we have to show is

$$\begin{aligned}
\lim_{k\to\infty}\int_\Omega \mathrm{e}^{\frac{\psi_k(y) - \mathbf{c}(x-y)}{\varepsilon}}\,\mathrm{d}\mu(y) &= \int_\Omega \mathrm{e}^{\frac{\psi_\infty(y) - \mathbf{c}(x-y)}{\varepsilon}}\,\mathrm{d}\mu(y)\\
\lim_{k\to\infty}\int_\Omega \mathrm{e}^{\frac{\phi_k(x) - \mathbf{c}(x-y)}{\varepsilon}}\,\mathrm{d}\rho_k(x) &= \int_\Omega \mathrm{e}^{\frac{\phi_\infty(x) - \mathbf{c}(x-y)}{\varepsilon}}\,\mathrm{d}\rho_*(x)\ .
\end{aligned}$$

We will representatively show the second limit, since the sequence ρ_k makes this one a little harder than the first one.

We have the uniform bound $\|\phi_k - \mathbf{c}\|_{L^\infty(\Omega^2)} \leq \frac{5}{3}\|\mathbf{c}\|_{L^\infty(\Omega^2)}$ by *Proposition 5.8.2.* and triangle inequality, so we can assume e to be uniform continuous in our case which then implies that the uniform convergence of ϕ_k carries over to $\mathrm{e}^{\frac{\phi_k(x)-\mathbf{c}(x-y)}{\varepsilon}}$. Finally, since ρ_k converges narrowly, we receive for every $y \in \Omega$

$$\lim_{k\to\infty} \int_\Omega \mathrm{e}^{\frac{\phi_k(x)-\mathbf{c}(x-y)}{\varepsilon}} \,\mathrm{d}\rho_k(x) = \int_\Omega \mathrm{e}^{\frac{\phi_\infty(x)-\mathbf{c}(x-y)}{\varepsilon}} \,\mathrm{d}\rho_*(x)\ .$$

For ψ_∞ we receive the same result, showing that ϕ_∞, ψ_∞ are actually optimal Kantorovich potentials in $\mathcal{D}_\varepsilon(\rho_*, \mu)$.

Now let us assume that there are two subsequences converging to two different optimal pairs (ϕ_*, ψ_*) and $(\phi_* + a, \psi_* - a)$. Then we can assume the two subsequences to be disjoint and simply recalibrate the latter subsequence by a, perhaps weakening the estimate $\|\phi_*\|_\infty, \|\psi_*\|_\infty \le \frac{3}{2}\|\mathbf{c}\|_\infty$ but not enough to break our argument above, and receive that this recalibrated subsequence actually converges to (ϕ_*, ψ_*). Finally, since limit points of the initial sequence cannot lie further apart than $\frac{3}{2}\|\mathbf{c}\|_\infty$ in the $L^\infty(\Omega)$-norm, we can always choose the $a \in [-3\|\mathbf{c}\|_\infty, 3\|\mathbf{c}\|_\infty]$ (where the 3 is chosen for good measure) and can apply this to all remaining disjoint subsequences converging to another optimal pair to receive a sequence actually having (ϕ_*, ψ_*) as the limit. □

Corollary 5.11. *Let $\rho_k, \rho_*, \mu \in \mathscr{P}^{ac}(\Omega)$ with finite entropy and let $\rho_k \rightharpoonup \rho_*$ narrowly in $\mathscr{P}(\Omega)$. Let furthermore γ_k be the sequence of optimal transport plans in $\mathbf{S}^2_\varepsilon(\rho_k, \mu)$ and G_k their density w.r.t. $\rho_k \otimes \mu$. Then the G_k are continuous and converge uniformly to G_*, the density w.r.t. $\rho_* \otimes \mu$ of the optimal transport plan γ_* of $\mathbf{S}^2_\varepsilon(\rho_*, \mu)$.*

In particular

$$\lim_{k\to\infty} \mathbf{S}^2_\varepsilon(\rho_k, \mu) = \mathbf{S}^2_\varepsilon(\rho_*, \mu)\ . \tag{5.2.8}$$

Proof. We consider the sequence (ϕ_k, ψ_k) of optimal potentials established in *Lemma 5.10.* Then ϕ_k, ψ_k are Lipschitz and uniformly bounded: $\|\phi_k\|_\infty, \|\psi_k\|_\infty \le \frac{3}{2}\|\mathbf{c}\|_\infty$. The structure of $G_k(x,y) = \mathrm{e}^{\phi_k(x)+\psi_k(y)/\varepsilon}\mathrm{K}_\varepsilon(x,y)$ allows us to conclude that G_k converges uniformly and its limit has the form

$$G_*(x,y) = \mathrm{e}^{\frac{\phi_*(x)+\psi_*(y)}{\varepsilon}}\mathrm{K}_\varepsilon(x,y).$$

The fact that the limit G_* is indeed optimal is a direct consequence (ϕ_*, ψ_*) being optimal potentials and *Proposition 5.8.3.* .

Let us proceed with G_k. By boundedness of Ω there is a uniform upper bound on $\mathbf{c}$. In combination, this gives us a uniform bound away from zero for G_k and G_*, showing the uniform convergence of G_k passes on to $G_k \log\left(\frac{G_k}{\mathrm{K}_\varepsilon}\right)$.

Finally, we can calculate

$$\begin{aligned}
\lim_{k\to\infty} \mathbf{S}_\varepsilon(\rho_k, \mu) &= \lim_{k\to\infty} \varepsilon \iint_{\Omega^2} G_k(x,y)\log\left(\frac{G_k(x,y)}{\mathrm{K}_\varepsilon(x,y)}\right) \mathrm{d}(\rho_k\otimes\mu)(x,y) \\
&= \varepsilon \iint_{\Omega^2} G_*(x,y)\log\left(\frac{G_*(x,y)}{\mathrm{K}_\varepsilon(x,y)}\right) \mathrm{d}(\rho_*\otimes\mu)(x,y) \\
&= \mathbf{S}_\varepsilon(\rho_*, \mu)
\end{aligned}$$

where, as we have seen $G_k(x,y)\log\left(\frac{G_k(x,y)}{\mathrm{K}_\varepsilon(x,y)}\right)$ converges uniformly to $G_*(x,y)\log\left(\frac{G_*(x,y)}{\mathrm{K}_\varepsilon(x,y)}\right)$ and $\rho_k \otimes \mu$ narrowly to $\rho_* \otimes \mu$. □

5.3 Existence of a minimizer

The existence of a minimizer in each recursion step is usually a result that can be shown more or less effortlessly in the JKO-case, for there the summands of the functional that is the subject of minimization are all themselves l.s.c. and therefore the direct method of the calculus of variation can be applied almost immediately.

In the BDF2 case on the other hand, the negative term $-\frac{1}{4}\mathbf{T}_\varepsilon^2(\rho,\rho_{\tau,\varepsilon}^{(n-2)})$ provides a substantial hurdle, since the lower semicontinuity is no longer immediate to obtain. But with the results from the preceeding section, we can see $\mathbf{T}_\varepsilon^2(\rho,\rho_{\tau,\varepsilon}^{(n-1)})-\frac{1}{4}\mathbf{T}_\varepsilon^2(\rho,\rho_{\tau,\varepsilon}^{(n-2)})$ is semi-continuous in ρ w.r.t. to narrow convergence.

We will begin with showing that $\Phi_{\tau,\varepsilon}(\tau;\rho_{\tau,\varepsilon}^{(n-1)},\rho_{\tau,\varepsilon}^{(n-2)},\rho)$ is bounded from below in ρ. Then we will show n.l.s.c. of the energy functional $\mathcal{E}(\rho)$. Finally we will establish n.l.s.c. of the kinetic term and conclude with plugging these results together to the existence of the minimizer.

5.3.1 Bound from below for $\Phi_{\tau,\varepsilon}$.

Now for the bound from below, let us for a moment only consider $\mathbf{T}_\varepsilon^2(\rho,\rho_{\tau,\varepsilon}^{(n-1)})-\frac{1}{4}\mathbf{T}_\varepsilon^2(\rho,\rho_{\tau,\varepsilon}^{(n-2)})$. Then we can use *Lemma 5.6* to receive the estimate

$$\begin{aligned}
&\mathbf{T}_\varepsilon^2(\rho,\rho_{\tau,\varepsilon}^{(n-1)})-\frac{1}{4}\mathbf{T}_\varepsilon^2(\rho,\rho_{\tau,\varepsilon}^{(n-2)})\\
&\quad\geq \left\langle \mathbf{c},\gamma_{\tau,\varepsilon}^{(n-1)}\right\rangle+\varepsilon\left(\mathcal{H}(\rho)+\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-1)})\right)-\frac{1}{4}\left(\langle\mathbf{c},\rho\otimes\mu\rangle+\varepsilon\left(\mathcal{H}(\rho)+\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-2)})\right)\right]\\
&\quad\geq \frac{3}{4}\varepsilon\mathcal{H}(\rho)-\frac{1}{4}\|\mathbf{c}\|_{L^\infty(\Omega^2)}-\varepsilon\left(\log(|\Omega|)+\frac{1}{4}\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-2)})\right)\\
&\quad\geq -\frac{1}{4}\|\mathbf{c}\|_{L^\infty(\Omega^2)}-\varepsilon\left(\frac{7}{4}\log(|\Omega|)+\frac{1}{4}\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-2)})\right)
\end{aligned}$$

where we used $0\leq\langle\mathbf{c},\gamma\rangle\leq\|\mathbf{c}\|_{L^\infty(\Omega^2)}$ as well as (5.2.2).

As an intermediate result we receive, from the next to the last line, the bound

$$\frac{3}{4}\varepsilon\mathcal{H}(\rho)\leq\mathbf{T}_\varepsilon^2(\rho,\rho_{\tau,\varepsilon}^{(n-1)})-\frac{1}{4}\mathbf{T}_\varepsilon^2(\rho,\rho_{\tau,\varepsilon}^{(n-2)})+\frac{1}{4}\|\mathbf{c}\|_{L^\infty(\Omega^2)}+\varepsilon\left(\log(|\Omega|)+\frac{1}{4}\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-2)})\right)\ . \tag{5.3.1}$$

The energy term can be estimated from below as well by means of Jensen's inequality

$$\mathcal{U}(\mu)=\int_\Omega \mathsf{u}(\mu(x))\,\mathrm{d}x\geq|\Omega|\,\mathsf{u}\left(\frac{1}{|\Omega|}\right)$$

and by the properties of v and w the values $\underline{\mathrm{v}}:=\min_\Omega \mathrm{v}(x)$ and $\underline{\mathrm{w}}:=\min_{x,y\in\Omega}\mathrm{w}(x-y)$ exist and are finite so we receive

$$\mathcal{V}(\mu)+\mathcal{W}(\mu)\geq\underline{\mathrm{v}}+\underline{\mathrm{w}}$$

for every $\mu\in\mathscr{P}(\Omega)$ right away. Plugging everything together and abbreviating $\underline{\mathcal{E}}:=|\Omega|\,\mathsf{u}\left(\frac{1}{|\Omega|}\right)+\underline{\mathrm{v}}+\underline{\mathrm{w}}$ we arrive at

$$\Phi_{\tau,\varepsilon}(\tau;\rho_{\tau,\varepsilon}^{(n-1)},\rho_{\tau,\varepsilon}^{(n-2)},\rho)\geq-\frac{1}{4\tau}\left(\|\mathbf{c}\|_{L^\infty(\Omega^2)}+7\varepsilon\log(|\Omega|)+\varepsilon\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-2)})\right)+\underline{\mathcal{E}}\ .$$

So there is a bound from below for $\Phi_{\tau,\varepsilon}$ that only depends on $\tau,\varepsilon,\mathsf{u},\mathrm{v},\mathrm{w},\mathbf{c}$ and $\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-2)})$ and in particular not on ρ.

Picking up the intermediate result (5.3.1) again we receive the following result.

Lemma 5.12. *Let $\tau, \varepsilon > 0$ and let $\rho_k \in \mathscr{P}^{ac}(\Omega)$ be a minimizing sequence of $\rho \mapsto \Phi_{\tau,\varepsilon}(\tau; \rho_{\tau,\varepsilon}^{(n-1)}, \rho_{\tau,\varepsilon}^{(n-2)}, \rho)$. Let furthermore the initial data $\rho_{\tau,\varepsilon}^{(n-1)}, \rho_{\tau,\varepsilon}^{(n-2)}$ have finite entropy and energy*

$$\mathcal{H}(\rho_{\tau,\varepsilon}^{(-1)}), \mathcal{E}(\rho_{\tau,\varepsilon}^{(-1)}), \mathcal{H}(\rho_{\tau,\varepsilon}^{(0)}), \mathcal{E}(\rho_{\tau,\varepsilon}^{(0)}) \leq C < \infty \ .$$

Then there is a subsequence ρ_{k_l} with uniformly bounded entropy and energy $\mathcal{H}(\rho_{k_l}), \mathcal{E}(\rho_{k_l}) \leq C < \infty$.

Proof. Since $\rho_{\tau,\varepsilon}^{(n-2)}$ is a feasible candidate in the minimization problem, we know that for k big enough we have

$$\Phi_{\tau,\varepsilon}(\tau; \rho_{\tau,\varepsilon}^{(n-1)}, \rho_{\tau,\varepsilon}^{(n-2)}, \rho_k) \leq \Phi_{\tau,\varepsilon}(\tau; \rho_{\tau,\varepsilon}^{(n-1)}, \rho_{\tau,\varepsilon}^{(n-2)}, \rho_{\tau,\varepsilon}^{(n-2)}) + 1$$

where the $+1$ has to be there in case $\rho_{\tau,\varepsilon}^{(n-2)}$ is itself a minimizer. Now on the one hand we have by (5.3.1)

$$\begin{aligned}\Phi_{\tau,\varepsilon}(\tau; \rho_{\tau,\varepsilon}^{(n-1)}, \rho_{\tau,\varepsilon}^{(n-2)}, \rho_k) &= \frac{1}{\tau}\left(\mathbf{T}_\varepsilon^2(\rho_k, \rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\mathbf{T}_\varepsilon^2(\rho_k, \rho_{\tau,\varepsilon}^{(n-2)})\right) + \mathcal{E}(\rho_k) \\ &\geq \frac{1}{\tau}\left(\frac{3}{4}\varepsilon\mathcal{H}(\rho_k) - \frac{1}{4}\|\mathbf{c}\|_{L^\infty(\Omega^2)} - \varepsilon\left(\log(|\Omega|) + \frac{1}{4}\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-2)})\right)\right) + \mathcal{E}(\rho_k) \\ &= \frac{3\varepsilon}{4\tau}\mathcal{H}(\rho_k) + \mathcal{E}(\rho_k) - \frac{\varepsilon}{4\tau}\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-2)}) + C_{\mathbf{c},\tau,\varepsilon,\Omega}\end{aligned}$$

and on the other hand we have with *Lemma 5.6*

$$\begin{aligned}\Phi_{\tau,\varepsilon}(\tau; \rho_{\tau,\varepsilon}^{(n-1)}, \rho_{\tau,\varepsilon}^{(n-2)}, \rho_{\tau,\varepsilon}^{(n-2)}) &= \frac{1}{\tau}\left(\mathbf{T}_\varepsilon^2(\rho_{\tau,\varepsilon}^{(n-2)}, \rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\mathbf{T}_\varepsilon^2(\rho_{\tau,\varepsilon}^{(n-2)}, \rho_{\tau,\varepsilon}^{(n-2)})\right) + \mathcal{E}(\rho_{\tau,\varepsilon}^{(n-2)}) \\ &\leq \frac{1}{\tau}\left(\left\langle \mathbf{c}, \rho_{\tau,\varepsilon}^{(n-2)} \otimes \rho_{\tau,\varepsilon}^{(n-1)}\right\rangle + \varepsilon\left(\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-2)}) + \mathcal{H}(\rho_{\tau,\varepsilon}^{(n-1)})\right) - \frac{\varepsilon}{2}\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-2)})\right) + \mathcal{E}(\rho_{\tau,\varepsilon}^{(n-2)}) \\ &\leq \frac{\varepsilon}{2\tau}\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-2)}) + \mathcal{E}(\rho_{\tau,\varepsilon}^{(n-2)}) + \frac{\varepsilon}{\tau}\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-1)}) + \frac{1}{\tau}\|\mathbf{c}\|_\infty \ .\end{aligned}$$

Plugging this together yields

$$\frac{3\varepsilon}{4\tau}\mathcal{H}(\rho_k) + \mathcal{E}(\rho_k) \leq \frac{3\varepsilon}{4\tau}\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-2)}) + \mathcal{E}(\rho_{\tau,\varepsilon}^{(n-2)}) + \frac{\varepsilon}{\tau}\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-1)}) + C_{\mathbf{c},\tau,\varepsilon,\Omega} + 1 \ .$$

Iteratively this yields a bound for $\frac{3\varepsilon}{4\tau}\mathcal{H}(\rho_k) + \mathcal{E}(\rho_k)$ that depends on the energy and entropy of $\rho_{\tau,\varepsilon}^{(n-1)}, \rho_{\tau,\varepsilon}^{(n-2)}$ and $\mathbf{c}, \tau, \varepsilon, \Omega$ and n, making this estimate uniform in k.

Now since $\mathcal{H}$ and $\mathcal{E}$ are both bounded from below, this estimate shows that $\mathcal{H}(\rho_k)$ and $\mathcal{E}(\rho_k)$ are each on its own uniformly bounded from above. □

Note that this bound might decay with τ, ε, but we only need it to establish existence of a minimizer, which is done for fixed $\tau, \varepsilon > 0$.

5.3.2 Weak lower semi-continuity of the kinetic term and $\mathcal{E}$.

As was shown for example in [46, Proposition 7.1 & 7.2], $\mathcal{V}$ and $\mathcal{W}$ are continuous w.r.t. to narrow convergence of probability measures. Furthermore [46, Proposition 7.7] implies that $\mathcal{U}$ and $\mathcal{H}$ are lower semicontinuous w.r.t. narrow convergence as well since h and u are both continuous, convex and superlinear at infinity.

Lemma 5.13 (Adaption from Section 7 in [46])**.** *We have that*

$$\rho \mapsto \mathcal{E}(\rho) \quad \textit{and} \quad \rho \mapsto \mathcal{H}(\rho) \tag{5.3.2}$$

are n.l.s.c. .

Showing that the kinetic term, $\mathbf{T}_\varepsilon^2(\rho, \rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\mathbf{T}_\varepsilon^2(\rho, \rho_{\tau,\varepsilon}^{(n-2)})$, will take the detour over $\mathbf{S}_\varepsilon$.
We begin by using *Proposition 5.8.1.* to rewrite our kinetic term.

$$\begin{aligned}\mathbf{T}_\varepsilon^2(\rho, \rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\mathbf{T}_\varepsilon^2(\rho, \rho_{\tau,\varepsilon}^{(n-2)}) \\ = \mathbf{S}_\varepsilon^2(\rho, \rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\mathbf{S}_\varepsilon^2(\rho, \rho_{\tau,\varepsilon}^{(n-2)}) + \frac{3}{4}\varepsilon\mathcal{H}(\rho) + \varepsilon\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\varepsilon\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-2)})\end{aligned} \tag{5.3.3}$$

where we used that for $\mu \in \mathscr{P}^{ac}(\Omega)$ we have $\mathcal{H}(\mu \mid \mathcal{L}^d) = \mathcal{H}(\mu)$.

Now we can see, the last two terms are constant w.r.t. ρ. The first two have been shown to be continuous in ρ w.r.t. to narrow convergence in $\mathscr{P}^{ac}(\Omega)$ with uniform bounded entropy in *Corollary 5.11.* And finally the third term, the negative entropy of ρ is n.l.s.c. by *Lemma 5.13.*

In conclusion, we have established that $\Phi_{\tau,\varepsilon}^n(\rho)$ consists of summands that are all n.l.s.c., which implies that $\Phi_{\tau,\varepsilon}^n(\rho)$ is n.l.s.c. in ρ.

5.3.3 Existence of a minimizer.

Showing the existence of a minimizer follows now the standard way by the direct method of the calculus of variation.

We have shown that $\rho \mapsto \Phi_{\tau,\varepsilon}(\tau; \rho_{\tau,\varepsilon}^{(n-1)}, \rho_{\tau,\varepsilon}^{(n-2)}, \rho)$ is bounded from below and lower semicontinuous w.r.t narrow convergence of ρ. So let us abbreviate $\Phi_{\tau,\varepsilon}^n(\rho) := \Phi_{\tau,\varepsilon}(\tau; \rho_{\tau,\varepsilon}^{(n-1)}, \rho_{\tau,\varepsilon}^{(n-2)}, \rho)$ and let us assume a minimizing sequence ρ_k. Then, since $\mathscr{P}(\Omega)$ is narrowly compact by boundedness of Ω, we can assume ρ_k to be narrowly converging to some ρ_*. By *Lemma 5.12* we can assume this sequence to lie in $\mathscr{P}^{ac}(\Omega)$ and have uniformly bounded entropy. We receive for every $\rho \in \mathscr{P}(\Omega)$

$$\Phi_{\tau,\varepsilon}^n(\rho) \geq \lim_{k\to\infty} \Phi_{\tau,\varepsilon}^n(\rho_k) = \liminf_{k\to\infty} \Phi_{\tau,\varepsilon}^n(\rho_k) \geq \Phi_{\tau,\varepsilon}^n(\rho_*)$$

implying that ϕ_* is a minimizer of $\Phi_{\tau,\varepsilon}^n$.

The first inequality is here implied by ρ_k being a minimizing sequence and the second inequality is derived from the n.l.s.c. of $\Phi_{\tau,\varepsilon}^n$.

This shows part *1.* of **Theorem 4** .

5.4 The Euler-Lagrange equation

To establish the Euler-Lagrange equation is the next step we take towards showing convergence of our approximate sequence to a solution of the initial equation.

Establishing the equation will consist of three distinct steps. First we will restate a well known result concerning the variation of our energy functional $\mathcal{E}$ along solutions of the transport equation $\partial_\lambda \rho_\lambda = \operatorname{div}(\rho_\lambda \xi)$.

Then we will establish, a similar result for the kinetic term $\rho \mapsto \frac{1}{\tau}\left(\mathbf{T}_\varepsilon^2(\rho, \rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\mathbf{T}_\varepsilon^2(\rho, \rho_{\tau,\varepsilon}^{(n-2)})\right)$. This step is a little more involved and consists of rewriting the kinetic term first w.r.t. $\mathbf{S}_\varepsilon$ as we did in (5.3.3), then shift to the dual problem $\mathcal{D}_\varepsilon$ of $\mathbf{S}_\varepsilon$ and finally using the stability result *Lemma 5.10* and *Corollary 5.11* for the corresponding Kantorovich potentials and the density of the optimal transport plans. Plugging these results together we will arrive at an estimate for the difference quotient of the variation along solutions of the transport equation.

Finally, by a standard argument, we establish the Euler-Lagrange equation utilizing the minimizing property of $\rho_{\tau,\varepsilon}^{(n)}$.

5.4.1 The variation of the energy functional and $\mathcal{H}$

To achieve these equations, we will vary our optimal $\rho_{\tau,\varepsilon}^{(n)}$ along the transport equation

$$\partial_\lambda \mu_\lambda = \operatorname{div}(\mu_\lambda \nabla\zeta)$$

with initial value $\mu_0 = \rho_{\tau,\varepsilon}^{(n)}$. It is well known (e.g. [2, Chapter 6] or [46, Chapter 4]) that the solution curve μ_λ can be written as a push-forward of $\rho_{\tau,\varepsilon}^{(n)}$ in the sense that $\mu_\lambda = (Z_\lambda)_\# \rho_{\tau,\varepsilon}^{(n)}$ and Z_λ is given as a solution of

$$Z_0 = \mathrm{id}_\Omega; \quad \partial_\lambda Z_\lambda = \nabla\zeta(Z_\lambda)\ .$$

Lemma 5.14. *Let $\rho_{\tau,\varepsilon}^{(n)} \in \mathscr{P}^{ac}(\Omega) \cap L^m(\Omega)$ and $\zeta \in C_c^\infty(\Omega, \mathbb{R})$. Consider μ_λ as given above. Then the first variation of $\mathcal{E}$ along the curve μ_λ is given as*

$$\partial_\lambda\Big|_{\lambda=0} \mathcal{E}(\mu_\lambda) = -\int_\Omega p(\rho_{\tau,\varepsilon}^{(n)}) \operatorname{div}(\nabla\zeta)\,\mathrm{d}x - \int_\Omega \rho_{\tau,\varepsilon}^{(n)}(x)(\nabla \mathrm{v} + 2(\nabla \mathrm{w} * \rho_{\tau,\varepsilon}^{(n)}))\nabla\zeta(x)\,\mathrm{d}x$$

where $p(s) = s\mathsf{u}'(s) - \mathsf{u}'(s)$.

As a direct consequence is the variation of the entropy along solutions of the transport equations

$$\partial_\lambda\Big|_{\lambda=0} \mathcal{H}(\mu_\lambda) = -\int_\Omega \rho_{\tau,\varepsilon}^{(n)} \operatorname{div}(\nabla\zeta)\,\mathrm{d}x\ . \tag{5.4.1}$$

Proof. Writing $\mu_\lambda = (Z_\lambda)_\# \rho_{\tau,\varepsilon}^{(n)}$ and minding the regularity, we receive the representation $\mu_\lambda = \frac{\rho_{\tau,\varepsilon}^{(n)}}{\det(DZ_\lambda)} \circ Z_\lambda^{-1}$ which is well defined for small λ. Here $DZ_\lambda \in C(\Omega, \mathbb{R}^{d\times d})$ is the Jacobian of Z_λ.

Now we can write

$$\mathcal{U}(\mu_\lambda) = \int_\Omega \mathsf{u}\left(\frac{\rho_{\tau,\varepsilon}^{(n)}}{\det(DZ_\lambda)}\right) \det(DZ_\lambda)\,\mathrm{d}x$$

and for deriving the determinant, there is Jacobi's formula yielding

$$\partial_\lambda \det(DZ_\lambda) = \operatorname{div}(\nabla\zeta(Z_\lambda)) \det(DZ_\lambda)$$

so we arrive, with u being a monomial of order m and the dominated convergence theorem, at

$$\partial_\lambda\Big|_{\lambda=0} \mathcal{U}(\mu_\lambda) = \int_\Omega p(\rho_{\tau,\varepsilon}^{(n)}) \Delta\zeta\,\mathrm{d}x$$

where p abbreviates the product rule struck u by $p(s) = s\mathsf{u}'(s) - \mathsf{u}(s)$.

The remaining parts $\mathcal{V}$ and $\mathcal{W}$ are easier to vary. Indeed, by linearity and the regularity of v as well as ζ having compact support in the open Ω, we receive

$$\begin{aligned}\partial_\lambda\Big|_{\lambda=0} \mathcal{V}(\mu_\lambda) &= \lim_{\lambda\searrow 0} \int_\Omega \mathrm{v}\ \frac{\mu_\lambda - \mu_0}{\lambda}\,\mathrm{d}x \\ &= \int_\Omega \mathrm{v}\ \operatorname{div}(\rho_{\tau,\varepsilon}^{(n)} \nabla\zeta)\,\mathrm{d}x \\ &= \int_\Omega \operatorname{div}(\rho_{\tau,\varepsilon}^{(n)} \nabla \mathrm{v})\zeta\,\mathrm{d}x\ .\end{aligned}$$

The calculation for $\mathcal{W}$ is very similar, after expanding

$$\iint_{\Omega^2} \frac{\mu_\lambda(x)\mathrm{w}(x-y)\mu_\lambda(y) - \mu_0(x)\mathrm{w}(x-y)\mu_0(y)}{\lambda}\,\mathrm{d}(x,y)$$
$$= \iint_{\Omega^2} \frac{\mu_\lambda(x)-\mu_0(x)}{\lambda}\mathrm{w}(x-y)\mu_\lambda(y)\,\mathrm{d}(x,y) + \iint_{\Omega^2} \mu_0(x)\mathrm{w}(x-y)\frac{\mu_\lambda(y)-\mu_0(y)}{\lambda}\,\mathrm{d}(x,y)\ .$$

Now by nearly the same arguments as above, we receive, after incorporating the symmetry and regularity of w,

$$\begin{aligned}\partial_\lambda\Big|_{\lambda=0}\mathcal{W}(\mu_\lambda) &= 2\int_\Omega \operatorname{div}(\rho_{\tau,\varepsilon}^{(n)}\nabla\zeta)\mathrm{w} * \rho_{\tau,\varepsilon}^{(n)}\,\mathrm{d}x \\ &= 2\int_\Omega \operatorname{div}(\rho_{\tau,\varepsilon}^{(n)}(\nabla\mathrm{w} * \rho_{\tau,\varepsilon}^{(n)}))\zeta\,\mathrm{d}x\ .\end{aligned}$$ □

Finally, when choosing $\mathsf{u} = \mathrm{h}$, $\mathrm{v} = \mathrm{w} = 0$ in $\mathcal{E}$ we see that $\mathcal{E}(\rho) = \mathcal{H}(\rho)$ for all $\rho \in \mathscr{P}(\Omega)$. Therefore the variation we just calculated holds for $\mathcal{H}$, too, and minding $p(s) = s$ in the case of $\mathsf{u}(s) = s\log(s)$, we arrive at (5.4.1).

5.4.2 The variation of the kinetic term

Showing a suitable estimate for the variation of the kinetic term requires a little more work. The plan consists of rewriting the kinetic term again as

$$\begin{aligned}&\mathbf{T}_\varepsilon(\rho,\rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\mathbf{T}_\varepsilon(\rho,\rho_{\tau,\varepsilon}^{(n-2)}) \\ &\quad= \mathbf{S}_\varepsilon(\rho,\rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\mathbf{S}_\varepsilon(\rho,\rho_{\tau,\varepsilon}^{(n-2)}) + \varepsilon\frac{3}{4}\mathcal{H}(\rho) + \varepsilon\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\varepsilon\mathcal{H}(\rho_{\tau,\varepsilon}^{(n-2)})\end{aligned}$$

and splitting this up in terms that can be handled one by one. First we note that the entropies of $\rho_{\tau,\varepsilon}^{(n-1)}$ and $\rho_{\tau,\varepsilon}^{(n-2)}$ are constant w.r.t. ρ and therefore play no role in the variation. Furthermore, the first two summands can, by similarity, be handled by similar arguments. Concerning the remaining part, $\mathbf{S}_\varepsilon(\rho,\rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\mathbf{S}_\varepsilon(\rho,\rho_{\tau,\varepsilon}^{(n-2)})$, we will use the stability results and the dual formulation from *Lemma 5.8* and *Corollary 5.11*. Finally, the entropy of ρ has already been taken care of by *Lemma 5.14.* as we will point out right away.

Lemma 5.15. *Let $\rho_{\tau,\varepsilon}^{(n-1)}, \rho_{\tau,\varepsilon}^{(n-2)} \in \mathscr{P}^{ac}(\Omega)$ with finite entropy and $\rho_{\tau,\varepsilon}^{(n)}$ the minimizer of $\Phi_{\tau,\varepsilon}^n(\cdot)$. Let $\zeta \in C_c^\infty(\Omega,\mathbb{R})$ and consider μ_λ the solution of $\partial_\lambda\mu_\lambda = \operatorname{div}(\mu_\lambda\nabla\zeta)$ with initial value $\mu_0 = \rho_{\tau,\varepsilon}^{(n)}$. Then an estimate of the first variation of $\rho \mapsto \mathbf{S}_\varepsilon(\rho,\rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\mathbf{S}_\varepsilon(\rho,\rho_{\tau,\varepsilon}^{(n-2)})$ at $\rho = \rho_{\tau,\varepsilon}^{(n)}$ is given as*

$$\begin{aligned}&\limsup_{\lambda\searrow 0}\frac{1}{\lambda}\left[\frac{1}{\tau}\left(\mathbf{S}_\varepsilon(\mu_\lambda,\rho_{\tau,\varepsilon}^{(n-2)}) - \frac{1}{4}\mathbf{S}_\varepsilon(\mu_\lambda,\rho_{\tau,\varepsilon}^{(n-2)})\right) - \frac{1}{\tau}\left(\mathbf{S}_\varepsilon(\mu_0,\rho_{\tau,\varepsilon}^{(n-2)}) - \frac{1}{4}\mathbf{S}_\varepsilon(\mu_0,\rho_{\tau,\varepsilon}^{(n-2)})\right)\right] \\ &\quad\le \frac{1}{\tau}\left(-\iint_{\Omega^2} 2\langle x-y,\nabla\zeta(x)\rangle\,\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y) + \frac{1}{4}\iint_{\Omega^2} 2\langle x-z,\nabla\zeta(x)\rangle\,\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-2)}(x,z)\right)\ .\end{aligned}$$

Proof. Let us begin with the variation of $\mathbf{S}_\varepsilon(\rho,\rho_{\tau,\varepsilon}^{(n-1)})$. Recall the dual problem $\mathcal{D}_\varepsilon$ and its functional D_ε defined in *Definition 5.7* and let us denote by $\phi_\lambda,\psi_\lambda$ the Kantorovich potentials of $\mathbf{S}_\varepsilon(\mu_\lambda,\rho_{\tau,\varepsilon}^{(n-1)})$, $\phi,\psi := \phi_0,\psi_0$ and $G = \mathrm{e}^{\phi_0+\psi_0/\varepsilon}\mathrm{K}_\varepsilon$ the density w.r.t. $\mu_0\otimes\rho_{\tau,\varepsilon}^{(n-1)}$ of the optimal transport plan γ_* of

$\mathbf{S}_\varepsilon(\mu_0, \rho_{\tau,\varepsilon}^{(n-1)})$. We receive for the difference quotient for $\lambda > 0$ small enough the estimate

$$\begin{aligned}
&\frac{1}{\lambda}\left(\mathbf{S}_\varepsilon(\mu_\lambda, \rho_{\tau,\varepsilon}^{(n-1)}) - \mathbf{S}_\varepsilon(\mu_0, \rho_{\tau,\varepsilon}^{(n-1)})\right) \\
&\quad= \frac{1}{\lambda}\left(\mathcal{D}_\varepsilon(\mu_\lambda, \rho_{\tau,\varepsilon}^{(n-1)}) - \mathcal{D}_\varepsilon(\mu_0, \rho_{\tau,\varepsilon}^{(n-1)})\right) \\
&\quad= \frac{1}{\lambda}\left(D_\varepsilon(\phi_\lambda, \psi_\lambda, \mu_\lambda, \rho_{\tau,\varepsilon}^{(n-1)}) - D_\varepsilon(\phi_0, \psi_0, \mu_0, \rho_{\tau,\varepsilon}^{(n-1)})\right) \\
&\quad\le \frac{1}{\lambda}\left(D_\varepsilon(\phi_\lambda, \psi_\lambda) - D_\varepsilon(\phi_\lambda, \psi_\lambda)\right) \\
&\quad= \int_\Omega \phi_\lambda \frac{\mu_\lambda - \mu_0}{\lambda}\,\mathrm{d}x - \varepsilon \iint_{\Omega^2} \mathrm{e}^{\phi_\lambda(x)+\psi_\lambda(y)/\varepsilon}\mathrm{K}_\varepsilon(x,y)\,\mathrm{d}[(\frac{\mu_\lambda - \mu_0}{\lambda}) \otimes \rho_{\tau,\varepsilon}^{(n-1)}](x,y) \\
&\quad= \int_\Omega \phi_\lambda \frac{\mu_\lambda - \mu_0}{\lambda}\,\mathrm{d}x - \varepsilon \iint_{\Omega^2} G_\lambda(x,y)\,\mathrm{d}[(\frac{\mu_\lambda - \mu_0}{\lambda}) \otimes \rho_{\tau,\varepsilon}^{(n-1)}](x,y)\ .
\end{aligned}$$

This is exactly where the stability result *Lemma 5.10* and the subsequent corollary is needed, since we have to pass to the limit with the argument of the integral and the measures against which we are integrating. Consequently we receive in the limit $\lambda \searrow 0$ by the uniform convergence of ϕ_λ and G_λ

$$\begin{aligned}
&\limsup_{\lambda\searrow 0}\frac{1}{\lambda}\left(\mathbf{S}_\varepsilon(\mu_\lambda, \rho_{\tau,\varepsilon}^{(n-1)}) - \mathbf{S}_\varepsilon(\mu_0, \rho_{\tau,\varepsilon}^{(n-1)})\right) \\
&\qquad\le \int_\Omega \phi\,\mathrm{div}\,(\rho_{\tau,\varepsilon}^{(n)}\nabla\zeta)\,\mathrm{d}x - \varepsilon\iint_{\Omega^2} G(x,y)\mathrm{div}\,(\rho_{\tau,\varepsilon}^{(n)}(x)\nabla\zeta(x))\rho_{\tau,\varepsilon}^{(n-1)}(y)\,\mathrm{d}(x,y)\ .
\end{aligned}$$

Now let us consider for a moment the marginal constraints of the optimal transport plan $\gamma_* = G\cdot\mathcal{L}^d$. We have

$$\gamma_*(x,y) = G(x,y)\rho_{\tau,\varepsilon}^{(n)}(x)\rho_{\tau,\varepsilon}^{(n-1)}(y)$$

since $\rho_{\tau,\varepsilon}^{(n)}, \rho_{\tau,\varepsilon}^{(n-1)} \in \mathscr{P}^{ac}(\Omega)$ and consequently

$$\rho_{\tau,\varepsilon}^{(n)}(x) = \rho_{\tau,\varepsilon}^{(n)}(x)\int_\Omega G(x,y)\rho_{\tau,\varepsilon}^{(n-1)}(y)\,\mathrm{d}y$$

implying on $\mathrm{supp}(\rho_{\tau,\varepsilon}^{(n)})$

$$1 = \int_\Omega G(x,y)\rho_{\tau,\varepsilon}^{(n-1)}(y)\,\mathrm{d}y \quad \text{a.e.}\ . \tag{5.4.2}$$

Anticipating the calculation to follow, we calculate the gradient of G w.r.t. x now, which exists since G and ϕ are both Lipschitz. Minding $\nabla_x\|x-y\|_2^2 = 2(x-y)$ we receive

$$\int_\Omega \nabla_x G(x,y)\rho_{\tau,\varepsilon}^{(n-1)}(y)\,\mathrm{d}y = \frac{1}{\varepsilon}\left(\nabla\phi(x)\int_\Omega G(x,y)\rho_{\tau,\varepsilon}^{(n-1)}(y)\,\mathrm{d}y - \int_\Omega 2(x-y)G(x,y)\rho_{\tau,\varepsilon}^{(n-1)}(y)\,\mathrm{d}y\right)\ .$$

By the very same regularity, we can integrate by parts, minding $\zeta \in C_c^\infty(\Omega)$, to arrive on the one hand at

$$\int_\Omega \phi\,\mathrm{div}\,(\rho_{\tau,\varepsilon}^{(n)}\nabla\zeta)\,\mathrm{d}x = -\int_\Omega \nabla\phi\rho_{\tau,\varepsilon}^{(n)}\nabla\zeta\,\mathrm{d}x$$

and on the other hand at

$$
\begin{aligned}
-\varepsilon \iint_{\Omega^2} G(x,y)\rho_{\tau,\varepsilon}^{(n-1)}(y)\mathrm{div}\,(\rho_{\tau,\varepsilon}^{(n)}(x)\nabla\zeta(x))\,\mathrm{d}(x,y) &= \varepsilon \int_\Omega \left(\int_\Omega \nabla_x G(x,y)\rho_{\tau,\varepsilon}^{(n-1)}(y)\,\mathrm{d}y\right)\rho_{\tau,\varepsilon}^{(n)}(x)\nabla\zeta(x)\,\mathrm{d}x \\
&= \int_\Omega \left(\nabla\phi(x)\int_\Omega G(x,y)\rho_{\tau,\varepsilon}^{(n-1)}(y)\,\mathrm{d}y - \int_\Omega 2(x-y)G(x,y)\rho_{\tau,\varepsilon}^{(n-1)}(y)\,\mathrm{d}y\right)\rho_{\tau,\varepsilon}^{(n)}(x)\nabla\zeta(x)\,\mathrm{d}x \\
&= \int_\Omega \nabla\phi\rho_{\tau,\varepsilon}^{(n)}\nabla\zeta\,\mathrm{d}x - \iint_{\Omega^2} 2\langle x-y,\nabla\zeta(x)\rangle\,\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y)\,.
\end{aligned}
$$

Plugging these two terms together we receive

$$
\begin{aligned}
\int_\Omega \phi\,\mathrm{div}\,(\rho_{\tau,\varepsilon}^{(n)}\nabla\zeta)\,\mathrm{d}x - \varepsilon\iint_{\Omega^2} G(x,y)\rho_{\tau,\varepsilon}^{(n-1)}(y)\mathrm{div}\,(\rho_{\tau,\varepsilon}^{(n)}(x)\nabla\zeta(x))\,\mathrm{d}(x,y)& \\
= -\iint_{\Omega^2} 2\langle x-y,\nabla\zeta(x)\rangle\,\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y)&
\end{aligned}
$$

where we denote with $\gamma_{\tau,\varepsilon}^{(n-1)}$ the optimal transport plan of $\mathbf{S}_\varepsilon(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-1)})$.

Note that in the next to the last equality we used (5.4.2) to cancel out the first and last integral.

Establishing the variation of $-\frac{1}{4}\mathbf{S}_\varepsilon(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-2)})$ follows virtually the same lines of argument with one difference. The first estimate has to be made in the other direction, due to the negative sign, giving us

$$
\begin{aligned}
-\frac{1}{\lambda}&\left(\mathbf{S}_\varepsilon(\mu_\lambda,\rho_{\tau,\varepsilon}^{(n-2)}) - \mathbf{S}_\varepsilon(\mu_0,\rho_{\tau,\varepsilon}^{(n-2)})\right) \\
&\le -\frac{1}{\lambda}\left(D_\varepsilon(\phi_0,\psi_0,\mu_\lambda,\rho_{\tau,\varepsilon}^{(n-1)}) - D_\varepsilon(\phi_0,\psi_0,\mu_0,\rho_{\tau,\varepsilon}^{(n-1)})\right) \\
&= -\left(\int_\Omega \phi\frac{\mu_\lambda-\mu_0}{\lambda}\,\mathrm{d}x - \varepsilon\iint_{\Omega^2} \mathrm{e}^{\phi_0(x)+\psi_0(y)/\varepsilon}\mathrm{K}_\varepsilon(x,y)\,\mathrm{d}[(\frac{\mu_\lambda-\mu_0}{\lambda})\otimes\rho_{\tau,\varepsilon}^{(n-1)}](x,y)\right)\,.
\end{aligned}
$$

So from here on, the calculations are actually easier than before, since we only have to deal with arguments in our integrals that are constant w.r.t. λ.

This being the only difference in the considerations of $\mathbf{S}_\varepsilon(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-1)})$ and $\mathbf{S}_\varepsilon(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-2)})$, we can establish

$$
\limsup_{\lambda\searrow 0} -\mathbf{S}_\varepsilon(\mu_\lambda,\rho_{\tau,\varepsilon}^{(n-2)}) \ge \iint_{\Omega^2} 2\langle x-z,\nabla\zeta(x)\rangle\,\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-2)}(x,z)
$$

where $\gamma_{\tau,\varepsilon}^{(n-2)}$ denotes the optimal transport plan of $\mathbf{S}_\varepsilon(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-2)})$ and we renamed $y\to z$ to avoid ambiguity.

Plugging these results together we arrive at the following estiamte of the variation of the kinetic term

$$
\begin{aligned}
\limsup_{\lambda\searrow 0}\frac{1}{\lambda}&\left[\frac{1}{\tau}\left(\mathbf{S}_\varepsilon(\mu_\lambda,\rho_{\tau,\varepsilon}^{(n-2)}) - \frac{1}{4}\mathbf{S}_\varepsilon(\mu_\lambda,\rho_{\tau,\varepsilon}^{(n-2)})\right) - \frac{1}{\tau}\left(\mathbf{S}_\varepsilon(\mu_0,\rho_{\tau,\varepsilon}^{(n-2)}) - \frac{1}{4}\mathbf{S}_\varepsilon(\mu_0,\rho_{\tau,\varepsilon}^{(n-2)})\right)\right] \\
&\le \frac{1}{\tau}\left(-\iint_{\Omega^2} 2\langle x-y,\nabla\zeta(x)\rangle\,\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y) + \frac{1}{4}\iint_{\Omega^2} 2\langle x-z,\nabla\zeta(x)\rangle\,\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-2)}(x,z)\right)\,. \qquad \square
\end{aligned}
$$

Proposition 5.16. *The Euler-Lagrange equation of the minimization problem $\Phi_{\tau,\varepsilon}^n(\rho)$, varied in direction of the transport equation $\partial_\lambda\mu_\lambda = \mathrm{div}\,(\mu_\lambda\nabla\zeta)$ with initial value $\mu_0 = \rho_{\tau,\varepsilon}^{(n)}$, is given as*

$$
\begin{aligned}
0 = \frac{1}{\tau}&\left(\iint_{\Omega^2} 2\langle x-y,\nabla\zeta(x)\rangle\,\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y) - \frac{1}{4}\iint_{\Omega^2} 2\langle x-z,\nabla\zeta(x)\rangle\,\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-2)}(x,z)\right) \qquad (5.4.3)\\
&- \frac{3\varepsilon}{4\tau}\int_\Omega \rho_{\tau,\varepsilon}^{(n)}\mathrm{div}\,(\nabla\zeta)\,\mathrm{d}x - \int_\Omega p(\rho_{\tau,\varepsilon}^{(n)})\mathrm{div}\,(\nabla\zeta)\,\mathrm{d}x - \int_\Omega \rho_{\tau,\varepsilon}^{(n)}(x)(\nabla\mathrm{v} + 2(\nabla\mathrm{w}*\rho_{\tau,\varepsilon}^{(n)}))\nabla\zeta(x)\,\mathrm{d}x \qquad (5.4.4)
\end{aligned}
$$

Proof. Plugging the results of this section together, we receive for the variation an estimate from above and together with the minimizing property of $\rho^{(n)}_{\tau,\varepsilon}$ in $\Phi^n_{\tau,\varepsilon}$, we receive

$$
\begin{aligned}
0 &\leq \limsup_{\lambda\searrow 0} \frac{1}{\lambda}\left(\Phi^n_{\tau,\varepsilon}(\mu_\lambda) - \Phi^n_{\tau,\varepsilon}(\mu_0)\right) \\
&\leq \frac{1}{\tau}\left(\iint_{\Omega^2} 2\left\langle x-y, \nabla\zeta(x)\right\rangle \,\mathrm{d}\gamma^{(n-1)}_{\tau,\varepsilon}(x,y) - \frac{1}{4}\iint_{\Omega^2} 2\left\langle x-z, \nabla\zeta(x)\right\rangle \,\mathrm{d}\gamma^{(n-2)}_{\tau,\varepsilon}(x,z)\right) \\
&\quad - \frac{3\varepsilon}{4\tau}\int_\Omega \rho^{(n)}_{\tau,\varepsilon}\mathrm{div}\,(\nabla\zeta)\,\mathrm{d}x + \int_\Omega p(\rho^{(n)}_{\tau,\varepsilon})\mathrm{div}\,(\nabla\zeta)\,\mathrm{d}x - \int_\Omega \rho^{(n)}_{\tau,\varepsilon}(x)(\nabla \mathrm{v} + 2(\nabla \mathrm{w} * \rho^{(n)}_{\tau,\varepsilon}))\nabla\zeta(x)\,\mathrm{d}x
\end{aligned}
$$

Now this expression is linear in ζ and flipping $\zeta \to -\zeta$ now shows the claim. □

5.5 A priori estimates

In this section we will assume $\tau, \varepsilon > 0$ and $\rho^{(n)}_{\tau,\varepsilon}$ to be a sequence spawned by initial data $\rho^{(-1)}_{\tau,\varepsilon}, \rho^{(0)}_{\tau,\varepsilon} \in \mathscr{P}^{ac}(\Omega)$ with finite entropy and the recursion

$$
\rho^{(n)}_{\tau,\varepsilon} \in \operatorname*{arg\,min}_{\rho\in\mathscr{P}(\Omega)} \Phi^n_{\tau,\varepsilon}(\rho)
$$

which has been shown to spawn a sequence of $\rho^{(n)}_{\tau,\varepsilon} \in \mathscr{P}^{ac}(\Omega)$ with uniformly bounded entropy.

We want to establish an analogue to the energy diminishing property of the JKO scheme next, though, as was pointed out in [33, Section 4.], we cannot hope for our sequence to be truly energy diminishing. We will however receive a almost-energy-diminishing property of our sequence alike the one in [33].

The entropic regularized distance has one major disadvantage which incidentally is one of the reasons why it is not an actual distance: $\mathbf{T}_\varepsilon(\rho,\rho) \neq 0$ in general. This fact also avoids us from using the standard technique to achieve the classical a priori estimates. To circumvent this problem we will make use of *Proposition 5.4* which we achieved with the block approximation.

To arrive at the classical estimate, the minimizing property of $\rho^{(n)}_{\tau,\varepsilon}$ is used to compare it to $\rho^{(n-1)}_{\tau,\varepsilon}$ which, when $\rho^{(n)}_{\tau,\varepsilon}$ is replaced by it, sets (one of the) kinetic terms to zero and the desired estimate follows. Since transport plans $(\mathrm{Id},\mathrm{Id})_\# \rho^{(n-1)}_{\tau,\varepsilon}$ have infinite entropy, we cannot use this approach but have to retreat to our results from (5.2.1) and *Proposition 5.4*, stating that we can bound $\mathbf{T}^2_\varepsilon$ from above and below by $\mathbf{T}^2$ in the following way:

$$
\mathbf{T}^2(\rho,\mu) + \varepsilon\mathcal{H}(\gamma_\varepsilon) \leq \mathbf{T}^2_\varepsilon(\rho,\mu) \leq \varepsilon\hat{C} + 2d\varepsilon\log\varepsilon + \mathbf{T}^2(\rho,\mu)\ . \tag{5.5.1}
$$

With this estimate at hand, we can establish the following result.

Lemma 5.17 (One step)**.** *Fix some $C > 0$. Let $\tau, \varepsilon > 0$ such that* (5.1.12) *holds which reads*

$$
0 < \varepsilon, \varepsilon\left|\log\varepsilon\right| \leq C\tau^2\ .
$$

Then we have

$$
\frac{1}{2}\mathbf{T}^2(\rho^{(n)}_{\tau,\varepsilon}, \rho^{(n-1)}_{\tau,\varepsilon}) - \frac{1}{4}\mathbf{T}^2(\rho^{(n-1)}_{\tau,\varepsilon}, \rho^{(n-2)}_{\tau,\varepsilon}) \leq \tau^2 C + \tau\left(\mathcal{E}(\rho^{(n-1)}_{\tau,\varepsilon}) - \mathcal{E}(\rho^{(n)}_{\tau,\varepsilon})\right)\ . \tag{5.5.2}
$$

Proof. By the minimizing property of $\rho^{(n)}_{\tau,\varepsilon}$ we have

$$
\mathbf{T}^2_\varepsilon(\rho^{(n)}_{\tau,\varepsilon}, \rho^{(n-1)}_{\tau,\varepsilon}) \leq \mathbf{T}^2_\varepsilon(\rho^{(n-1)}_{\tau,\varepsilon}, \rho^{(n-1)}_{\tau,\varepsilon}) - \frac{1}{4}\left(\mathbf{T}^2_\varepsilon(\rho^{(n-1)}_{\tau,\varepsilon}, \rho^{(n-2)}_{\tau,\varepsilon}) - \mathbf{T}^2_\varepsilon(\rho^{(n)}_{\tau,\varepsilon}, \rho^{(n-2)}_{\tau,\varepsilon})\right) + \tau\left(\mathcal{E}(\rho^{(n-1)}_{\tau,\varepsilon}) - \mathcal{E}(\rho^{(n)}_{\tau,\varepsilon})\right)
$$

and consequently, by (5.5.1) we receive for the l.h.s.

$$\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-1)}) - 2\varepsilon\log(|\Omega|) \le \mathbf{T}_\varepsilon^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-1)})$$

and for the first summands of the r.h.s.

$$\begin{aligned}
&\mathbf{T}_\varepsilon^2(\rho_{\tau,\varepsilon}^{(n-1)},\rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\left(\mathbf{T}_\varepsilon^2(\rho_{\tau,\varepsilon}^{(n-1)},\rho_{\tau,\varepsilon}^{(n-2)}) - \mathbf{T}_\varepsilon^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-2)})\right)\\
&\quad\le (\varepsilon+\varepsilon\,|\log(\varepsilon)|)C - \frac{1}{4}\left(\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n-1)},\rho_{\tau,\varepsilon}^{(n-2)}) + (\varepsilon+\varepsilon\,|\log\varepsilon|)C - \mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-2)}) - 2\varepsilon\log|\Omega|\right)\\
&\quad\le (\varepsilon+\varepsilon\,|\log(\varepsilon)|)C - \frac{1}{2}\varepsilon\log|\Omega| + \frac{1}{2}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-1)}) + \frac{1}{4}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n-1)},\rho_{\tau,\varepsilon}^{(n-2)})\ .
\end{aligned}$$

where we used the triangle inequality for the squared Wasserstein-2-distance

$$\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-2)}) \le 2\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-1)}) + 2\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n-1)},\rho_{\tau,\varepsilon}^{(n-2)})$$

which is founded in the usual triangle inequality, the binomial theorem and youngs inequality for products with $p,q=2$.

Plugging these together we arrive, after moving $\frac{1}{2}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-1)}) + \frac{1}{4}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n-1)},\rho_{\tau,\varepsilon}^{(n-2)})$ to the l.h.s. and $-2\varepsilon\log(|\Omega|)$ to the r.h.s.

$$\frac{1}{2}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n-1)},\rho_{\tau,\varepsilon}^{(n-2)}) \le (\varepsilon+\varepsilon\,|\log(\varepsilon)|)C + \frac{3}{2}\varepsilon\log|\Omega| + \tau\left(\mathcal{E}(\rho_{\tau,\varepsilon}^{(n-1)}) - \mathcal{E}(\rho_{\tau,\varepsilon}^{(n)})\right)\ .$$

Finally including the estimates from (5.1.12), we arrive at

$$\frac{1}{2}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n-1)},\rho_{\tau,\varepsilon}^{(n-2)}) \le \tau^2 C + \tau\left(\mathcal{E}(\rho_{\tau,\varepsilon}^{(n-1)}) - \mathcal{E}(\rho_{\tau,\varepsilon}^{(n)})\right)\ . \qquad \square$$

So we see our sequence is almost energy diminishing, having a small defect $\tau^2 C + \frac{1}{4}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n-1)},\rho_{\tau,\varepsilon}^{(n-2)})$ in each step. The $\frac{1}{4}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n-1)},\rho_{\tau,\varepsilon}^{(n-2)})$ will not play a major role as can be seen in the following corollary, summing up over the individual steps.

Corollary 5.18 (Classical estimate, several steps)**.** *Let $N = \lfloor \frac{T}{\tau} \rfloor + 1$. Then with* (5.1.8)*,*

$$\frac{1}{4}\sum_{n=1}^{N}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-1)}) \le \tau(T+\tau)C + \tau\left(\mathcal{E}(\rho_{\tau,\varepsilon}^{(0)}) - \mathcal{E}(\rho_{\tau,\varepsilon}^{(N)})\right) \tag{5.5.3}$$

holds.

Proof. Summing up the l.h.s. of (5.5.2) over n, we receive, when minding the telescopic sum,

$$\begin{aligned}
&\sum_{n=1}^{N}\frac{1}{2}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-1)}) - \frac{1}{4}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n-1)},\rho_{\tau,\varepsilon}^{(n-2)})\\
&\qquad = \frac{1}{4}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(N)},\rho_{\tau,\varepsilon}^{(N-1)}) - \frac{1}{4}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(-1)},\rho_{\tau,\varepsilon}^{(0)}) + \frac{1}{4}\sum_{n=1}^{N}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-1)})\\
&\qquad \ge -\frac{1}{4}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(-1)},\rho_{\tau,\varepsilon}^{(0)}) + \frac{1}{4}\sum_{n=1}^{N}\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-1)})
\end{aligned}$$

the sum of the r.h.s. then yields

$$\sum_{n=1}^{N} \tau^2 C + \tau \left(\mathcal{E}(\rho_{\tau,\varepsilon}^{(n-1)}) - \mathcal{E}(\rho_{\tau,\varepsilon}^{(n)}) \right) \leq \tau(T+\tau)C + \tau \left(\mathcal{E}(\rho_{\tau,\varepsilon}^{(0)}) - \mathcal{E}(\rho_{\tau,\varepsilon}^{(N)}) \right) .$$

plugging in (5.1.8), which implies by triangle inequality

$$\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(-1)}, \rho_{\tau,\varepsilon}^{(0)}) \leq C\tau$$

after moving $\frac{1}{4}\mathbf{T}_2^2(\rho_{\tau,\varepsilon}^{(-1)}, \rho_{\tau,\varepsilon}^{(0)})$ to the r.h.s. we arrive at the claim. □

Lemma 5.19 (Hölder-type estimate)**.** *Let* $m_1, m_2 \in \mathbb{N}$ *with* $m_2 > m_1$*. Then there is a* $C < \infty$ *not depending on* τ, ε, m_1 *or* m_2 *such that*

$$\mathbf{T}(\rho_{\tau,\varepsilon}^{(m_2)}, \rho_{\tau,\varepsilon}^{(m_1)}) \leq (\tau(m_2 - m_1))^{1/2} C$$

holds.

Proof. We can calculate by triangle and Hölder type estimate

$$\begin{aligned}
\mathbf{T}(\rho_{\tau,\varepsilon}^{(m_2)}, \rho_{\tau,\varepsilon}^{(m_1)}) &\leq \sum_{n=m_1+1}^{m_2} \mathbf{T}(\rho_{\tau,\varepsilon}^{(n)}, \rho_{\tau,\varepsilon}^{(n-1)}) \\
&\leq \left(\sum_{n=m_1+1}^{m_2} \tau \right)^{1/2} \left(\frac{1}{\tau} \sum_{n=m_1+1}^{m_2} \mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n)}, \rho_{\tau,\varepsilon}^{(n-1)}) \right)^{1/2} \\
&\leq (\tau(m_2 - m_1))^{1/2} 2 \left((T+\tau)C + \left(\mathcal{E}(\rho_{\tau,\varepsilon}^{(0)}) - \mathcal{E}(\rho_{\tau,\varepsilon}^{(N)}) \right) \right)^{1/2} .
\end{aligned}$$

Since $\mathcal{E}$ is bounded from below and since $\mathcal{E}(\rho_{\tau,\varepsilon}^{(0)})$ is bounded from above, we can find a $C < \infty$ such that $2\left((T+\tau)C + \left(\mathcal{E}(\rho_{\tau,\varepsilon}^{(0)}) - \mathcal{E}(\rho_{\tau,\varepsilon}^{(N)}) \right) \right)^{1/2} \leq C$. □

5.5.1 The interpolated solution

In this section we will define our approximate solution $\rho_{\tau,\varepsilon}$ to our equation (5.1.1) by means of the sequence $\rho_{\tau,\varepsilon}^{(n)}$ and a piecewise constant in time interpolation.

The entropic regularized BDF2 scheme spawns for every pair $\tau, \varepsilon > 0$ and initial data $\rho^{(0)}, \rho^{(-1)} \in \mathscr{P}^{ac}(\Omega)$ with finite entropy a sequence $(\rho_{\tau,\varepsilon}^{(n)})_{n \in [N]}$ in $\mathscr{P}^{ac}(\Omega)$, as we have seen.

Definition 5.20 (Approximate solution)**.** *Let* $\tau, \varepsilon > 0$*. We will denote the joint convergence minding* (5.1.12) *from now on by* $(\tau, \varepsilon) \searrow 0$*. Analogously,* $\sup_{(\tau,\varepsilon)}$ *assumes a sequence* (τ_k, ε_k) *minding* (5.1.12) *and takes the supremum over its elements. Let* $\rho_{\tau,\varepsilon}^{(n)}$ *the corresponding sequence. Then the piecewise constant in time interpolation* $\rho_{\tau,\varepsilon} : [0,T] \times \Omega \to \mathbb{R}$ *is defined as*

$$\rho_{\tau,\varepsilon}(t,x) := \rho_{\tau,\varepsilon}^{(n)}(x); \qquad \textit{if } t \in ((n-1)\tau, n\tau] .$$

5.5.2 Compactness of the interpolated solution

Anticipating the use of *Theorem 2.3* we show that the assumptions of *Theorem 2.3* hold. Recall *Example 2.2* where $\mathfrak{F}$ and $\mathbf{g}$ were defined as

$$\mathfrak{F}(\rho) = \begin{cases} \int_\Omega |\rho(x)|^m + \|\nabla \rho(x)\|^m \,\mathrm{d}x & \text{if } \rho \in W^{1,m}(\Omega) \\ +\infty & \text{otherwise} \end{cases}$$

and $\mathbf{g}(\rho, \mu) = \mathbf{T}_2(\rho, \mu)$.

Lemma 5.21. *Let $\rho_{\tau,\varepsilon}$ be an approximate solution as stated above. Then*

1. *$\rho_{\tau,\varepsilon}$ satisfies the following equiintegrablility condition*

$$\lim_{h\searrow 0}\sup_{(\tau,\varepsilon)}\int_0^{T-h}\mathbf{T}_2^2(\rho_{\tau,\varepsilon}(t+h),\rho_{\tau,\varepsilon}(t))\,\mathrm{d}t=0$$

where the supremum runs over feasible (τ,ε) for our family of approximate solutions $\rho_{\tau,\varepsilon}$.

2. *$\rho^m_{\tau,\varepsilon}\in L^1(0,T;BV(\Omega))$ and it is bounded therein uniformly w.r.t. (τ,ε), that is to say*

$$\sup_{(\tau,\varepsilon)}\int_0^T\mathfrak{F}(\rho_{\tau,\varepsilon}(t))\,\mathrm{d}t<\infty\ .$$

Proof. 1. This will be a consequence of *Lemma 5.19.* Indeed let $h>0$ and $t\in[0,T-h]$. Then we have $m_1,m_2\in\mathbb{N}$ such that $m_1=\lfloor\frac{t}{\tau}\rfloor$ and $m_2=\lfloor\frac{t+h}{\tau}\rfloor$ and

$$\mathbf{T}_2^2(\rho_{\tau,\varepsilon}(t+h),\rho_{\tau,\varepsilon}(t))=\mathbf{T}_2^2(\rho^{(m_2)}_{\tau,\varepsilon},\rho^{(m_1)}_{\tau,\varepsilon})\le(\tau(m_2-m_1))^{1/2}C\ .$$

With this estimate at hand, an argument analogue the the one in the proof of [18, Proposition 4.8] shows the claim.

2. Let us rearrange our Euler-Lagrange equation (5.4.3)

$$\begin{aligned}&\int_\Omega\mu^{(n)}_{\tau,\varepsilon}\mathrm{div}\,(\nabla\zeta)\,\mathrm{d}x\\&\quad=\frac{1}{\tau}\left(\iint_{\Omega^2}2\langle x-y,\nabla\zeta(x)\rangle\,\mathrm{d}\gamma^{(n-1)}_{\tau,\varepsilon}(x,y)-\frac{1}{4}\iint_{\Omega^2}2\langle x-z,\nabla\zeta(x)\rangle\,\mathrm{d}\gamma^{(n-2)}_{\tau,\varepsilon}(x,z)\right)\\&\qquad-\int_\Omega\rho^{(n)}_{\tau,\varepsilon}\left(\nabla\mathrm{v}+2(\nabla\mathrm{w}*\rho^{(n)}_{\tau,\varepsilon})\right)\nabla\zeta\,\mathrm{d}x\end{aligned}\tag{5.5.4}$$

where we defined $\mu^{(n)}_{\tau,\varepsilon}(x):=\frac{3\varepsilon}{4\tau}\rho^{(n)}_{\tau,\varepsilon}(x)+p(\rho^{(n)}_{\tau,\varepsilon}(x))$.

The plan will be to show that $\mu^{(n)}_{\tau,\varepsilon}$ is bounded in $L^1(\Omega)$ uniformly w.r.t. τ,ε and n and then showing a bound on $\left|\int_\Omega\mu^{(n)}_{\tau,\varepsilon}\mathrm{div}\,(\xi)\,\mathrm{d}x\right|\le C\,\|\xi\|_\infty$ for all $\xi\in C^\infty_c(\Omega,\mathbb{R}^d)$ where $C>0$ does not depend on τ,ε nor n and showing on the other hand, that $\int_\Omega\mu^{(n)}_{\tau,\varepsilon}\mathrm{div}\,(\zeta)=\int_\Omega\theta\zeta$ for some $\theta\in L^1(\Omega)$. Consequently the distributional derivative of $\mu^{(n)}_{\tau,\varepsilon}$ is actually an element of $L^1(\Omega)$. This way we have on the one hand shown that $\mu^{(n)}_{\tau,\varepsilon}\in W^{1,1}(\Omega)$ for every n and on the other hand have the estimates at hand to show that the regularity carries over to $\rho_{\tau,\varepsilon}$ which turns out to be bounded in $L^1((0,T);W^{1,1}(\Omega))$, which is actually a stronger result than our claim. This is due to us having to pass through $\mathsf{f}_{\tau,\varepsilon}$ with our result to pass it from $\mu^{(n)}_{\tau,\varepsilon}$ to $\rho_{\tau,\varepsilon}$.

First we see, that since $\rho_{\tau,\varepsilon}$ is uniformly bounded in $\mathcal{E}$, we have that $\mu^{(n)}_{\tau,\varepsilon}$ is uniformly bounded in $L^1(\Omega)$ right away. Let us prepare the inequality we want to show next.

First we note that

$$\begin{aligned}\sum_{n=1}^{N}\left|\iint_{\Omega^2}\langle x-y,\xi(x)\rangle\ \mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y)\right| &\leq \sum_{n=1}^{N}\iint_{\Omega^2}\|x-y\|_2\ \mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y)\ \|\xi(x)\|_{L^\infty(\Omega)}\\ &\leq \sum_{n=1}^{N}\sqrt{d}\left(\iint_{\Omega^2}\|x-y\|_2^2\ \mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y)\right)^{1/2}\|\xi(x)\|_{L^\infty(\Omega)}\\ &= \sqrt{d}\sum_{n=1}^{N}\tau^{1/2}\left(\frac{1}{\tau}\left\langle \mathbf{c},\gamma_{\tau,\varepsilon}^{(n-1)}\right\rangle\right)^{1/2}\|\xi(x)\|_{L^\infty(\Omega)}\\ &\leq \sqrt{d}\left(\sum_{n=1}^{N}\tau\right)^{1/2}\left(\sum_{n=1}^{N}\frac{1}{\tau}\left\langle \mathbf{c},\gamma_{\tau,\varepsilon}^{(n-1)}\right\rangle\right)^{1/2}\|\xi(x)\|_{L^\infty(\Omega)}\\ &\leq \sqrt{d}T^{1/2}C\,\|\xi(x)\|_{L^\infty(\Omega)}\end{aligned}$$

Where this constant C is derived on the one hand from $\mathbf{T}_\varepsilon^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-1)}) = \left\langle \mathbf{c},\gamma_{\tau,\varepsilon}^{(n-1)}\right\rangle + \varepsilon\mathcal{H}(\gamma_{\tau,\varepsilon}^{(n-1)})$ which in combination with *Proposition 5.4*, (5.1.12) and (5.5.3) yields

$$\frac{1}{\tau}\left\langle \mathbf{c},\gamma_{\tau,\varepsilon}^{(n-1)}\right\rangle \leq \tau C + \frac{1}{\tau}\mathbf{T}_2^2(\rho_{\tau,\varepsilon}^{(n)},\rho_{\tau,\varepsilon}^{(n-1)}) \leq \tau C + D$$

for some C, D not depending on τ,ε nor n. And in the same way we can estimate

$$\sum_{n=1}^{N}\left|\iint_{\Omega^2}\langle x-z,\xi(x)\rangle\ \mathrm{d}\gamma_{\tau,\varepsilon}^{(n-2)}(x,z)\right| \leq T^{1/2}C\,\|\xi(x)\|_{L^\infty(\Omega)}\ .$$

By regularity of v and w we see

$$\int_\Omega\left|\rho_{\tau,\varepsilon}^{(n)}\left(\nabla \mathrm{v} + 2(\nabla \mathrm{w} * \rho_{\tau,\varepsilon}^{(n)})\right)\right|\ \mathrm{d}x \leq \|\nabla \mathrm{v}\|_{L^\infty(\Omega)} + 2\,\|\nabla \mathrm{w}\|_{L^\infty(\Omega)} < \infty\ .$$

Plugging these estimates together we arrive, with (5.5.4) at

$$\begin{aligned}&\int_0^T\left|\int_\Omega \mu_{\tau,\varepsilon}^{(n)}\mathrm{div}\,(\xi(x))\,\mathrm{d}x\right|\ \mathrm{d}t\\ &\quad\leq \int_0^T \frac{1}{\tau}\left|\iint_{\Omega^2}\langle x-y,\nabla\zeta(x)\rangle\ \mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y)\right| + \frac{1}{4\tau}\left|\iint_{\Omega^2}\langle x-z,\nabla\zeta(x)\rangle\ \mathrm{d}\gamma_{\tau,\varepsilon}^{(n-2)}(x,z)\right|\\ &\qquad + \left|\int_\Omega \rho_{\tau,\varepsilon}^{(n)}\left(\nabla \mathrm{v} + 2(\nabla \mathrm{w} * \rho_{\tau,\varepsilon}^{(n)})\right)\nabla\zeta\,\mathrm{d}x\right|\ \mathrm{d}t\\ &\quad= \sum_{n=1}^{N}\left|\iint_{\Omega^2}\langle x-y,\xi(x)\rangle\ \mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y)\right| + \frac{1}{4}\sum_{n=1}^{N}\left|\iint_{\Omega^2}\langle x-z,\xi(x)\rangle\ \mathrm{d}\gamma_{\tau,\varepsilon}^{(n-2)}(x,z)\right|\\ &\qquad + \sum_{n=1}^{N}\tau\int_\Omega\left|\rho_{\tau,\varepsilon}^{(n)}\left(\nabla \mathrm{v} + 2(\nabla \mathrm{w} * \rho_{\tau,\varepsilon}^{(n)})\right)\right|\ \mathrm{d}x\\ &\quad\leq \left(2T^{1/2}C + T(\|\nabla \mathrm{v}\|_{L^\infty(\Omega)} + 2\,\|\nabla \mathrm{w}\|_{L^\infty(\Omega)})\right)\|\xi(x)\|_{L^\infty(\Omega)}\ .\end{aligned}$$

Now to show that $\int_\Omega \mu_{\tau,\varepsilon}^{(n)} \mathrm{div}\,(\xi)$ is actually just a $L^1(\Omega)$-function integrated against ξ, we rewrite by use of the disintegration theorem as well as the fact that $\gamma_{\tau,\varepsilon}^{(n-1)}$ and ρ are a.c. w.r.t. Lebesgue

$$\iint_\Omega \langle x-y,\xi(x)\rangle \;\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y) = \int_\Omega \left\langle \int_\Omega (x-y)\,\mathrm{d}(\gamma_{\tau,\varepsilon}^{(n-1)})_x(y), \xi(x) \right\rangle \rho(x)\,\mathrm{d}x\;.$$

so this expression is ξ integrated against $x \mapsto \int_\Omega (x-y)\,\mathrm{d}(\gamma_{\tau,\varepsilon}^{(n-1)})_x(y)\rho(x)$. To see that this is indeed a $L^1(\Omega)$ function, we calculate

$$\begin{aligned}
\int_\Omega \left\| \int_\Omega (x-y)\,\mathrm{d}(\gamma_{\tau,\varepsilon}^{(n-1)})_x(y)\rho(x) \right\|_2 \mathrm{d}x &\le \iint_{\Omega^2} \|x-y\|_2 \;\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y) \\
&\le \left(\iint_{\Omega^2} \|x-y\|_2^2 \;\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y) \right)^{1/2} \\
&\le C
\end{aligned}$$

as we have seen above. So $\iint_\Omega \langle x-y,\xi(x)\rangle \;\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y)$ is can be seen as ξ integrated against a $L^1(\Omega)$-function. The same holds for $\iint_\Omega \langle x-z,\xi(x)\rangle \;\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-2)}(x,z)$. Finally, since v and w are C^2, the same holds for $\int_\Omega \rho_{\tau,\varepsilon}^{(n)} \left(\nabla \mathrm{v} + 2(\nabla \mathrm{w} * \rho_{\tau,\varepsilon}^{(n)})\right) \xi\,\mathrm{d}x$, which shows by (5.5.4) our claim that $\int_\Omega \mu_{\tau,\varepsilon}^{(n)} \mathrm{div}\,(\xi)$ is actually ξ integrated against a $L^1(\Omega)$ function implying that $\mu_{\tau,\varepsilon}^{(n)} \in W^{1,1}(\Omega)$ for all τ,ε and n. Note however, that we have not shown that this is a uniformly bounded sequence in $W^{1,1}(\Omega)$, since there is still the factor $\frac{1}{\tau}$ in front of the kinetic terms.

Next we want to show that these results hold for $\rho_{\tau,\varepsilon}^m$, too. To that end, consider the map $\mathsf{f}_{\tau,\varepsilon}(s) := \frac{3\varepsilon}{4\tau}s + p(s)$. Note that $\mathsf{f}_{\tau,\varepsilon}(\rho_{\tau,\varepsilon}^{(n)}) = \mu_{\tau,\varepsilon}^{(n)}$ and furthermore, $p(s) = s^m$ for any $m \in [1,\infty)$.

This shows that $\rho_{\tau,\varepsilon}^{(n)} \in L^m(\Omega)$ is uniformly bounded therein and consequently $\rho_{\tau,\varepsilon}^m \in L^1((0,T)\times\Omega)$. Additionally we see that $\mathsf{f}_{\tau,\varepsilon}$ is a smooth diffeomorphism on $[0,\infty)$ and furthermore its inverse is Lipschitz with constant $L = \frac{4\tau}{3\varepsilon}$ and has the derivative

$$\mathsf{f}'_{\tau,\varepsilon}(s) = \frac{1}{\frac{3\varepsilon}{4\tau} + m(\mathsf{f}_{\tau,\varepsilon}^{-1}(s))^{m-1}}\;.$$

Now we arrive at

$$\left|\nabla \rho_{\tau,\varepsilon}^{(n)}(x)\right| = \left|(\mathsf{f}_{\tau,\varepsilon}^{-1})'(\mu_{\tau,\varepsilon}^{(n)}(x))\nabla \mu_{\tau,\varepsilon}^{(n)}(x)\right| \le \frac{\left|\nabla \mu_{\tau,\varepsilon}^{(n)}(x)\right|}{m(\rho_{\tau,\varepsilon}^{(n)}(x))^{m-1}}$$

which rearranges by means of $m(\rho_{\tau,\varepsilon}^{(n)})^{m-1}\nabla \rho_{\tau,\varepsilon}^{(n)} = \nabla((\rho_{\tau,\varepsilon}^{(n)})^m)$ to

$$\left|\nabla(\rho_{\tau,\varepsilon}^{(n)}(x)^m)\right| \le \frac{1}{m}\left|\nabla \mu_{\tau,\varepsilon}^{(n)}(x)\right|\;.$$

Consequently,

$$\int_0^T \int_\Omega \left|\nabla(\rho_{\tau,\varepsilon}^m(t,x))\right| \mathrm{d}x\,\mathrm{d}t \le \tau \sum_{n=1}^N \int_\Omega \left|\nabla(\rho_{\tau,\varepsilon}^{(n)}(x))^m\right| \mathrm{d}x \le \frac{1}{m}\tau \sum_{n=1}^N \left\|\nabla \mu_{\tau,\varepsilon}^{(n)}\right\|_{L^1(\Omega)} \le \frac{1}{m}T\tilde{C}$$

where $\tilde{C} = 2T^{1/2}C + T(\|\nabla \mathrm{v}\|_{L^\infty(\Omega)} + 2\|\nabla \mathrm{w}\|_{L^\infty(\Omega)})$. In combination we have achieved

$$\rho_{\tau,\varepsilon} \in L^1(0,T;W^{1,m}(\Omega))\;.$$

□

5.6 Convergence to the PDE

5.6.1 Convergence

In *Lemma 5.21* we have shown some of the prerequisites of *Theorem 2.3*. We want to make the use of this theorem now rigorous. Let us define $\mathfrak{F} : \mathscr{P}(\Omega) \to [0, \infty]$, $\mathbf{g}$ as in *Example 2.2*, that is to say

$$\mathfrak{F}(\rho) = \begin{cases} \int_\Omega \|\rho\|_{BV(\Omega)} & \text{if } \rho^m \in BV(\Omega) \text{ and } \rho \in \mathscr{P}^{ac}(\Omega) \\ +\infty & \text{otherwise} \end{cases}$$

and $\mathbf{g}(\rho, \mu) = \mathbf{T}_2(\rho, \mu)$. Then by *Example 2.2*, $\mathfrak{F}$ is a normal, coercive integrand and $\mathbf{g}$ is a pseudo-distance, both in the sense of *Definition 2.1*. Furthermore by *Lemma 5.21* the remaining prerequisites of *Theorem 2.3* are met, so we can conclude the following proposition.

Proposition 5.22. *Let $\rho_{\tau,\varepsilon}$ be a sequence of approximate solutions as defined above for feasible inital data and ε, τ such that* (5.1.12) *is satisfied. Then, up to a subsequence, $\rho_{\tau,\varepsilon}$ converges to some ρ_* w.r.t. strong $L^m((0,T) \times \Omega)$ topology.*

Proof. As already explained, all prerequisites of *Theorem 2.3* are met and consequently, $\rho_{\tau,\varepsilon}$ is compact w.r.t. $\mathcal{M}(0, T; L^m(\Omega))$ and uniformly bounded in $L^m((0,T) \times \Omega)$. By dominated convergence theorem, this yields our claim. □

5.6.2 The limit is a solution to the PDE

From here on we denote without relabelling with $\rho_{\tau,\varepsilon}$ the converging subsequence.

The first form of our Euler-Lagrange equation will now be the key to show that this converging subsequence actually converges to a solution of our PDE.

Let us restate the Euler-Lagrange equation (5.4.3) but with some modifications. First we receive for the kinetic integrals

$$\begin{aligned} \iint_{\Omega^2} \langle x - y, \nabla\zeta(x)\rangle \, \mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y) &= \iint_{\Omega^2} (\zeta(x) - \zeta(y)) \, \mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y) + I_1 \\ &= \int_\Omega (\rho_{\tau,\varepsilon}^{(n)}(x) - \rho_{\tau,\varepsilon}^{(n-1)}(x))\zeta(x) \, \mathrm{d}x + I_1 \end{aligned}$$

where

$$I_1 = \iint_{\Omega^2} \frac{1}{2} \nabla^2\zeta(\lambda_{x,y})(x-y)^2 \, \mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y)$$

with $\lambda_{x,y}$ a feasible point on the line segment from x to y.

Analogously we receive

$$\iint_{\Omega^2} \langle x - z, \nabla\zeta(x)\rangle \, \mathrm{d}\gamma_{\tau,\varepsilon}^{(n-2)}(x,z) = \int_\Omega (\rho_{\tau,\varepsilon}^{(n)}(x) - \rho_{\tau,\varepsilon}^{(n-2)}(x))\zeta(x) \, \mathrm{d}x + I_2$$

with

$$I_2 = \iint_{\Omega^2} \frac{1}{2} \nabla^2\zeta(\lambda_{x,z})(x-z)^2 \, \mathrm{d}\gamma_{\tau,\varepsilon}^{(n-2)}(x,z) \; .$$

Now plugging these two results together we arrive, for the kinetic term, at

$$
\begin{aligned}
&\frac{1}{\tau}\left(\iint_{\Omega^2} \langle x-y, \nabla\zeta(x)\rangle \,\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y) - \frac{1}{4}\iint_{\Omega^2} \langle x-z, \nabla\zeta(x)\rangle \,\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-2)}(x,z)\right) \\
&\quad = \frac{1}{\tau}\left(\int_\Omega (\rho_{\tau,\varepsilon}^{(n)}(x) - \rho_{\tau,\varepsilon}^{(n-1)}(x))\zeta(x)\,\mathrm{d}x - \frac{1}{4}\int_\Omega (\rho_{\tau,\varepsilon}^{(n)}(x) - \rho_{\tau,\varepsilon}^{(n-2)}(x))\zeta(x)\,\mathrm{d}x + I_1 - \frac{1}{4}I_2\right) \\
&\quad = \int_\Omega \left(\frac{3}{4}\frac{\rho_{\tau,\varepsilon}^{(n)} - \rho_{\tau,\varepsilon}^{(n-1)}}{\tau} - \frac{1}{4}\frac{\rho_{\tau,\varepsilon}^{(n-1)} - \rho_{\tau,\varepsilon}^{(n-2)}}{\tau}\right)\zeta\,\mathrm{d}x + \frac{1}{\tau}I_1 - \frac{1}{4\tau}I_2\;.
\end{aligned}
$$

Furthermore we want to modify the integral incorporating u. We rewrite

$$
\int_\Omega p(\rho_{\tau,\varepsilon}^{(n)})\mathrm{div}\,(\nabla\zeta)\,\mathrm{d}x = -\int_\Omega \rho_{\tau,\varepsilon}^{(n)}(x)\nabla\mathsf{u}'(\rho_{\tau,\varepsilon}^{(n)})\nabla\zeta(x)\,\mathrm{d}x\;. \tag{5.6.1}
$$

The modified Euler-Lagrange equation then reads as follows. For every $\zeta \in C_c^\infty(\Omega)$ we have

$$
\begin{aligned}
0 = &\int_\Omega \left(\frac{3}{4}\frac{\rho_{\tau,\varepsilon}^{(n)}(x) - \rho_{\tau,\varepsilon}^{(n-1)}(x)}{\tau} - \frac{1}{4}\frac{\rho_{\tau,\varepsilon}^{(n-1)}(x) - \rho_{\tau,\varepsilon}^{(n-2)}(x)}{\tau}\right)\zeta(x)\,\mathrm{d}x \\
&- \int_\Omega \rho_{\tau,\varepsilon}^{(n)}(x)\left(\nabla[\mathsf{u}'(\rho_{\tau,\varepsilon}^{(n)}) + \mathrm{v} + 2(\mathrm{w} * \rho_{\tau,\varepsilon}^{(n)})]\right)\nabla\zeta(x)\,\mathrm{d}x \\
&+ \frac{1}{\tau}I_1 - \frac{1}{4\tau}I_2 - J
\end{aligned}
$$

holds, where

$$
J := \frac{3\varepsilon}{4\tau}\int_\Omega \rho_{\tau,\varepsilon}^{(n)}\mathrm{div}\,(\nabla\zeta)\,\mathrm{d}x = \frac{3\varepsilon}{4\tau}\int_\Omega \rho_{\tau,\varepsilon}^{(n)}\Delta\zeta\,\mathrm{d}x\;.
$$

Our aim is to show that the Euler-Lagrange equation converges pointwise in the space of test functions ζ and ξ, which will suffice to conclude that the limit of the Euler-Lagrange equation holds for the limit curve ρ_*, showing part *3* of *Theorem 4* .

We will now give estimates on the three integral expressions I_1, I_2 and J first, which show that they will vanish in the limit $(\tau,\varepsilon) \searrow 0$.

First we can bound $|J|$, since $\zeta \in C_c^\infty(\Omega)$ with Ω bounded, so

$$
|J| \le \frac{3\varepsilon}{4\tau}\,\|\Delta\zeta\|_{L^\infty(\Omega)}\;.
$$

Concerning I_1 we can estimate with $\zeta \in C^\infty(\Omega)$ and our a priori estimate

$$
\begin{aligned}
|I_1| &\le \left\|\nabla^2\zeta\right\|_{L^\infty(\Omega)} \iint_{\Omega^2} \|x-y\|^2 \,\mathrm{d}\gamma_{\tau,\varepsilon}^{(n-1)}(x,y) \\
&= \left\|\nabla^2\zeta\right\|_{L^\infty(\Omega)} \mathbf{T}_\varepsilon^2(\rho_{\tau,\varepsilon}^{(n)}, \rho_{\tau,\varepsilon}^{(n-1)}) \\
&\le \left\|\nabla^2\zeta\right\|_{L^\infty(\Omega)} \left(\mathbf{T}_2^2(\rho_{\tau,\varepsilon}^{(n)}, \rho_{\tau,\varepsilon}^{(n-1)}) + \varepsilon C + \varepsilon\,|\log\varepsilon|\,C\right) \\
&\le \left\|\nabla^2\zeta\right\|_{L^\infty(\Omega)} \left(\mathbf{T}_2^2(\rho_{\tau,\varepsilon}^{(n)}, \rho_{\tau,\varepsilon}^{(n-1)}) + \tau^2 C\right)
\end{aligned}
$$

where the last inequality is derived from (5.1.12). We receive a similar estimate for $|I_2|$ as well.

Now let us include the temporal aspect as well. Let $\xi \in C_c^\infty((0,T))$ and let τ be small enough such that $\mathrm{supp}(\xi) \subset (2\tau, N-2\tau)$. Then we receive from our Euler-Lagrange equation

$$\int_0^T \int_\Omega \left(\frac{3}{4} \frac{\rho_{\tau,\varepsilon}(t,x) - \rho_{\tau,\varepsilon}(t-\tau x)}{\tau} - \frac{1}{2} \frac{\rho_{\tau,\varepsilon}(t-\tau,x) - \rho_{\tau,\varepsilon}(t-2\tau x)}{\tau} \right) \xi(t)\zeta(x)\,\mathrm{d}x$$
$$+ \int_0^T \int_\Omega \rho_{\tau,\varepsilon}(t,x) \left(\nabla[\mathsf{u}'(\rho_{\tau,\varepsilon}(t,x)) + \mathrm{v} + 2(\mathrm{w} * \rho_{\tau,\varepsilon})(t,x)]\right) \nabla\zeta(x)\xi(t)\,\mathrm{d}x\,\mathrm{d}t$$
$$= \sum_{n=1}^N \tau \overline{\xi}^n \left(\frac{1}{\tau} I_1 - \frac{1}{4\tau} I_2 - J \right)$$

where $\overline{\xi}^n = \int_{(n-1)\tau}^{n\tau} \xi(t)\,\mathrm{d}t$.

First off,

$$\tau \sum_{n=1}^N \overline{\xi}^n \left(\frac{1}{\tau} I_1 - \frac{1}{4\tau} I_2 - J \right)$$
$$\le \|\xi\|_{L^\infty([0,T])} \left(\sum_{n=2}^{N-2} \left[\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n)}, \rho_{\tau,\varepsilon}^{(n-1)}) + \tau^2 C \right] - \frac{1}{4} \sum_{n=2}^{N-2} \left[\mathbf{T}^2(\rho_{\tau,\varepsilon}^{(n-1)}, \rho_{\tau,\varepsilon}^{(n-2)}) + \tau^2 C \right] + \sum_{n=2}^{N-2} \frac{3\varepsilon}{4} \|\Delta\zeta\|_{L^\infty(\Omega)} \right)$$

where we can see that each sum in the large brackets goes to zero with order τ for $\tau \searrow 0$.

We take a look at the l.h.s. of the equation. We can rearrange, using that $\mathrm{supp}\,\xi \subset (2\tau, T-2\tau)$, to arrive at

$$\int_0^T \int_\Omega \left(\frac{3}{4} \frac{\rho_{\tau,\varepsilon}(t,x) - \rho_{\tau,\varepsilon}(t-\tau x)}{\tau} - \frac{1}{2} \frac{\rho_{\tau,\varepsilon}(t-\tau,x) - \rho_{\tau,\varepsilon}(t-2\tau x)}{\tau} \right) \xi(t)\zeta(x)\,\mathrm{d}x$$
$$= \int_{2\tau}^{T-2\tau} \int_\Omega \rho_{\tau,\varepsilon}(t,x) \left[\frac{3}{4} \frac{\xi(t) - \xi(t+\tau)}{\tau} - \frac{1}{4} \frac{\xi(t+\tau) - \xi(t+2\tau)}{\tau} \right] \zeta(x)\,\mathrm{d}x\,\mathrm{d}t$$

Minding that ρ_τ converges strongly in $L^m((0,T)\times\Omega)$ for $(\tau,\varepsilon) \searrow 0$ and recalling that the brackets converge uniformly in τ to $\partial_t \xi$ we see that

$$\lim_{(\tau,\varepsilon)\searrow(0,0)} \int_0^T \int_\Omega \left(\frac{3}{4} \frac{\rho_\tau(t,x) - \rho_\tau(t-\tau x)}{\tau} - \frac{1}{2} \frac{\rho_\tau(t-\tau,x) - \rho_\tau(t-2\tau x)}{\tau} \right) \xi(t)\zeta(x)\,\mathrm{d}x$$
$$= \int_0^T \int_\Omega \rho_*(t,x)\partial_t\xi(t)\zeta(x)\,\mathrm{d}x\,\mathrm{d}t$$

holds for every $\xi \in C_c^\infty((0,T))$ and $\zeta \in C_c^\infty(\Omega)$.

Finally, let's consider the integral corresponding to the free energy, i.e. the integral on the r.h.s. . We recall that $\rho_{\tau,\varepsilon}^m$ converges strongly in $L^1((0,T)\times\Omega)$ to ρ_*^m and receive, after switching back the integral incorporating u' to its representation in terms of $p(\rho_{\tau,\varepsilon})$ with (5.6.1),

$$\lim_{(\tau,\varepsilon)\searrow(0,0)} \int_0^T \xi(t) \int_\Omega p(\rho_{\tau,\varepsilon}(t,x))\mathrm{div}\,(\nabla\zeta(x)) - \rho_{\tau,\varepsilon}(t,x) \left\langle \nabla\mathrm{v}(x) + (\nabla\mathrm{w} * \rho_{\tau,\varepsilon})(t,x), \nabla\zeta(x) \right\rangle \,\mathrm{d}x\,\mathrm{d}t$$
$$= \int_0^T \xi(t) \int_\Omega p(\rho_*(t,x))\mathrm{div}\,(\nabla\zeta(x)) - \rho_*(t,x) \left\langle \nabla\mathrm{v}(x) + (\nabla\mathrm{w} * \rho_{\tau,\varepsilon})(t,x), \nabla\zeta(x) \right\rangle \,\mathrm{d}x\,\mathrm{d}t\,.$$

Plugging these results together, we see that for every $\xi \in C_c^\infty((0,T))$ and $\zeta \in C_c^\infty(\Omega)$ the limit ρ_* satisfies

$$-\int_0^T \int_\Omega \rho_*(t,x)\partial_t\xi(t)\zeta(x)\,\mathrm{d}x\,\mathrm{d}t$$
$$= \int_0^T \int_\Omega p(\rho_*(t,x))\Delta\zeta(x) - \rho_*(t,x)\left\langle \nabla \mathrm{v}(x) + (\nabla \mathrm{w} * \rho_{\tau,\varepsilon})(t,x), \nabla\zeta(x)\right\rangle \xi(t)\,\mathrm{d}x\,\mathrm{d}t$$

which is the weak formulation of our PDE on $(0,T)\times\Omega$.

As already mentioned, this shows part *3.* of *Theorem 4* .

5.6.3 The initial data are assumed continuously

Finally we will show that the initial value is assumed continuously, which will be a consequence of the Hölder type estimate *Lemma 5.19.* Indeed for $\rho_{\tau,\varepsilon}$ at two distinct points in time $t_1, t_2 \in [0,T]$ we have the estimate

$$\mathbf{T}_2(\rho_{\tau,\varepsilon}(t_1), \rho_{\tau,\varepsilon}(t_2)) \le (|t_1 - t_2| + \tau)^{1/2} C$$

a direct consequence of said lemma. [2, Prop. 7.1.5.] in combination with the $L^m((0,T)\times\Omega)$ convergence of $\rho_{\tau,\varepsilon}$ and the continuity of $\mathbf{T}_2$ w.r.t. to the weaker narrow convergence implies the estimate

$$\mathbf{T}_2(\rho_*(t_1), \rho_*(t_2)) = \lim_{\tau,\varepsilon\searrow 0} \mathbf{T}_2(\rho_{\tau,\varepsilon}(t_1), \rho_{\tau,\varepsilon}(t_2)) \le \lim_{\tau,\varepsilon\searrow 0} (|t_1 - t_2| + \tau)^{1/2} C = |t_1 - t_2|^{1/2} C$$

showing that the limit curve ρ_* is 1/2-Hölder-continuous, and since this estimate holds for $t_1 = 0$, too, and $\rho_{\tau,\varepsilon}(0) = \rho_{\tau,\varepsilon}^{(0)}$ converges w.r.t. $\mathbf{T}$ to the initial value ρ^0, we arrive at the result, that the initial value ρ^0 is assumed continuously by the curve ρ_* w.r.t. $\mathbf{T}$-distance in the space $L^m(\Omega)$.

Finally this shows part *4.* of *Theorem 4* and concludes this chapter.

Bibliography

[1] Martial Agueh et al. “Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory”. In: *Advances in Differential Equations* 10.3 (2005), pp. 309–360.

[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Second. Lectures in Mathematics ETH Zürich. Basel: Birkhäuser Verlag, 2008, pp. x+334. ISBN: 978-3-7643-8721-1.

[3] F Andreu, V Caselles, and JM Mazón. “Some regularity results on the ‘relativistic’heat equation”. In: *Journal of Differential Equations* 245.12 (2008), pp. 3639–3663.

[4] Fuensanta Andreu et al. “The Dirichlet problem associated to the relativistic heat equation”. In: *Mathematische Annalen* 347.1 (2010), pp. 135–199.

[5] Jean-David Benamou et al. “Discretization of functionals involving the Monge–Ampère operator”. In: *Numerische Mathematik* 134.3 (2016), pp. 611–636.

[6] Adrien Blanchet, Vincent Calvez, and José A. Carrillo. “Convergence of the mass-transport steepest descent scheme for the subcritical Patlak-Keller-Segel model”. In: *SIAM J. Numer. Anal.* 46.2 (2008), pp. 691–721. ISSN: 0036-1429. DOI: 10.1137/070683337. URL: https://doi.org/10.1137/070683337.

[7] Yann Brenier. “Extended Monge-Kantorovich theory”. In: *Optimal transportation and applications*. Springer, 2003, pp. 91–121.

[8] Guillaume Carlier et al. “Convergence of entropic schemes for optimal transport and gradient flows”. In: *SIAM Journal on Mathematical Analysis* 49.2 (2017), pp. 1385–1418.

[9] J. A. Carrillo, V. Caselles, and S. Moll. “On the relativistic heat equation in one space dimension”. In: *Proc. Lond. Math. Soc. (3)* 107.6 (2013), pp. 1395–1423. ISSN: 0024-6115. URL: https://doi.org/10.1112/plms/pdt015.

[10] J. A. Carrillo and J. S. Moll. “Numerical simulation of diffusive and aggregation phenomena in nonlinear continuity equations by evolving diffeomorphisms”. In: *SIAM J. Sci. Comput.* 31.6 (2009/10), pp. 4305–4329. ISSN: 1064-8275. URL: https://doi.org/10.1137/080739574.

[11] José A Carrillo et al. “A lagrangian scheme for the solution of nonlinear diffusion equations using moving simplex meshes”. In: *Journal of Scientific Computing* (2017), pp. 1–37.

[12] José Antonio Carrillo, Katy Craig, and Francesco S Patacchini. “A blob method for diffusion”. In: *arXiv preprint arXiv:1709.09195* (2017).

[13] Vicent Caselles. “Convergence of the’relativistic’heat equation to the heat equation as c$\to\infty$”. In: *Publicacions Matemàtiques* (2007), pp. 121–142.

[14] Lenaic Chizat et al. “Scaling algorithms for unbalanced optimal transport problems”. In: *Mathematics of Computation* 87.314 (2018), pp. 2563–2609.

[15] Giovanni Conforti and Luca Tamanini. “A formula for the time derivative of the entropic cost and applications”. In: *arXiv preprint arXiv:1912.10555* (2019).

[16] Marco Cuturi. “Sinkhorn distances: Lightspeed computation of optimal transport”. In: *Advances in neural information processing systems*. 2013, pp. 2292–2300.

[17] Virginia De Cicco, Nicola Fusco, and Anna Verde. “On L1-lower semicontinuity in BV”. In: *J. Convex Anal* 12.1 (2005), pp. 173–185.

[18] Marco Di Francesco and Daniel Matthes. “Curves of steepest descent are entropy solutions for a class of degenerate convection–diffusion equations”. In: *Calculus of Variations and Partial Differential Equations* 50.1-2 (2014), pp. 199–230.

[19] Simone Di Marino and Augusto Gerolin. “An Optimal Transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm”. In: *arXiv preprint arXiv:1911.06850* (2019).

[20] Lawrence C. Evans. *Partial differential equations*. Providence, R.I.: American Mathematical Society, 2010. ISBN: 9780821849743 0821849743.

[21] Enrico Giusti and Graham Hale Williams. *Minimal surfaces and functions of bounded variation*. Vol. 2. 3. Springer, 1984.

[22] Laurent Gosse and Giuseppe Toscani. “Identification of asymptotic decay to self-similarity for one-dimensional filtration equations”. In: *SIAM Journal on Numerical Analysis* 43.6 (2006), pp. 2590–2606.

[23] Richard Jordan, David Kinderlehrer, and Felix Otto. “The variational formulation of the Fokker-Planck equation”. In: *SIAM J. Math. Anal.* 29.1 (1998), pp. 1–17. ISSN: 0036-1410. URL: https://doi.org/10.1137/S0036141096303359.

[24] Oliver Junge, Daniel Matthes, and Horst Osberger. “A fully discrete variational scheme for solving nonlinear Fokker-Planck equations in multiple space dimensions”. In: *SIAM J. Numer. Anal.* 55.1 (2017), pp. 419–443. ISSN: 0036-1429. URL: https://doi.org/10.1137/16M1056560.

[25] L Kantorovich. “On the transfer of masses. Dokl. Acad. Nauk. USSR 37, 78 (1942) 56. Keller, EF, Segel, LA: Initiation of slide mold aggregation viewed as an instability”. In: *J. Theor. Biol* 26 (1970), p. 399415.

[26] Maxime Laborde. “12 On some nonlinear evolution systems which are perturbations of Wasserstein gradient flows”. In: *Topological Optimization and Optimal Transport: In the Applied Sciences* 17 (2017), p. 304.

[27] Philippe Laurençot and Bogdan-Vasile Matioc. “A gradient flow approach to a thin film approximation of the Muskat problem”. In: *Calc. Var. Partial Differential Equations* 47.1-2 (2013), pp. 319–341. ISSN: 0944-2669. DOI: 10.1007/s00526-012-0520-5. URL: https://doi.org/10.1007/s00526-012-0520-5.

[28] Christian Léonard. “From the Schr\” odinger problem to the Monge-Kantorovich problem”. In: *arXiv preprint arXiv:1011.2564* (2010).

[29] Elliott H Lieb and Michael Loss. *Analysis*. Vol. 14. American Mathematical Soc., 2001.

[30] Jan Maas and Daniel Matthes. “Long-time behavior of a finite volume discretization for a fourth order diffusion equation”. In: *Nonlinearity* 29.7 (2016), pp. 1992–2023. ISSN: 0951-7715. URL: https://doi.org/10.1088/0951-7715/29/7/1992.

[31] Daniel Matthes and Horst Osberger. “A convergent Lagrangian discretization for a nonlinear fourth-order equation”. In: *Found. Comput. Math.* 17.1 (2017), pp. 73–126. ISSN: 1615-3375. URL: https://doi.org/10.1007/s10208-015-9284-6.

[32] Daniel Matthes and Horst Osberger. "Convergence of a variational Lagrangian scheme for a nonlinear drift diffusion equation". In: *ESAIM Math. Model. Numer. Anal.* 48.3 (2014), pp. 697–726. ISSN: 0764-583X. URL: `https://doi.org/10.1051/m2an/2013126`.

[33] Daniel Matthes and Simon Plazotta. "A variational formulation of the BDF2 method for metric gradient flows". In: *arXiv preprint arXiv:1711.02935* (2017).

[34] Daniel Matthes and Benjamin Söllner. "Convergent Lagrangian discretization for drift-diffusion with nonlocal aggregation". In: *Innovative Algorithms and Analysis*. Springer, 2017, pp. 313–351.

[35] Daniel Matthes and Benjamin Söllner. "Discretization of flux-limited gradient flows: Γ-convergence and numerical schemes". In: *Mathematics of Computation* 89.323 (2020), pp. 1027–1057.

[36] Robert J. McCann. "A convexity principle for interacting gases". In: *Adv. Math.* 128.1 (1997), pp. 153–179. ISSN: 0001-8708. DOI: `10.1006/aima.1997.1634`. URL: `http://dx.doi.org/10.1006/aima.1997.1634`.

[37] Robert J. McCann and Marjolaine Puel. "Constructing a relativistic heat flow by transport time steps". In: *Ann. Inst. H. Poincaré Anal. Non Linéaire* 26.6 (2009), pp. 2539–2580. ISSN: 0294-1449. URL: `https://doi.org/10.1016/j.anihpc.2009.06.006`.

[38] Gaspard Monge. "Mémoire sur la théorie des déblais et des remblais". In: *Histoire de l'Académie Royale des Sciences de Paris* (1781).

[39] Felix Otto. "Lubrication approximation with prescribed nonzero contact angle". In: *Comm. Partial Differential Equations* 23.11-12 (1998), pp. 2077–2164. ISSN: 0360-5302. DOI: `10.1080/03605309808821411`. URL: `https://doi.org/10.1080/03605309808821411`.

[40] Felix Otto. "The geometry of dissipative evolution equations: the porous medium equation". In: *Comm. Partial Differential Equations* 26.1-2 (2001), pp. 101–174. ISSN: 0360-5302. DOI: `10.1081/PDE-100002243`. URL: `https://doi.org/10.1081/PDE-100002243`.

[41] G. Peyré and M. Cuturi. "Computational Optimal Transport". In: *to appear in Foundations and Trends in Machine Learning* (2018). URL: `https://arxiv.org/abs/1803.00567`.

[42] Gabriel Peyré. "Entropic approximation of Wasserstein gradient flows". In: *SIAM J. Imaging Sci.* 8.4 (2015), pp. 2323–2351. ISSN: 1936-4954. URL: `https://doi.org/10.1137/15M1010087`.

[43] Simon Plazotta. "A BDF2-approach for the non-linear Fokker-Planck equation". In: *arXiv preprint arXiv:1801.09603* (2018).

[44] Philip Rosenau. "Tempered diffusion: A transport process with propagating fronts and inertial delay". In: *Physical Review A* 46.12 (1992), R7371.

[45] Riccarda Rossi and Giuseppe Savaré. "Tightness, integral equicontinuity and compactness for evolution problems in Banach spaces". In: *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze-Serie V* 2.2 (2003), p. 395.

[46] Filippo Santambrogio. "Optimal transport for applied mathematicians". In: *Birkäuser, NY* (2015), pp. 99–102.

[47] Benjamin Söllner and Oliver Junge. "A convergent Lagrangian discretization for p-Wasserstein and flux-limited diffusion equations". In: *arXiv preprint arXiv:1906.01321* (2019).