Fach Mathematik

Variational Methods using Transport Metrics and Applications

Inaugural-Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften im Fachbereich Mathematik und Informatik der Mathematisch-Naturwissenschaftlichen Fakultät der Westfälischen Wilhelms-Universität Münster

vorgelegt von Marzena Magdalene Franek -2011-
Dekan: Prof. Dr. Matthias Löwe

Erster Gutachter: Prof. Dr. Martin Burger

Zweiter Gutachter: Prof. Dr. Daniel Matthes

Tag der mündlichen Prüfung:

Tag der Promotion:
Abstract

The goal of this thesis is to introduce a new variational approach based on the optimal transport theory. In particular, we want to investigate the solution of an optimisation problem consisting of a data term, which is here the quadratic Wasserstein distance between probability measures and a regularisation term. This novel formulation leads to a method, which can handle unified discrete measures as well as continuous probability measures. Moreover, with a special kind of regularisation, namely the TV-regularisation, discontinuities in the data can be preserved. Further, if we consider the Wasserstein distance in the data term we obtain mass conservation. This is an improvement to existing models, since the mass conservation must be usually implemented as an additional constraint.

The second but equivalent method we derive in this thesis is in the spirit of the fluid dynamic formulation of Benamou and Brenier. By introducing an artificial time variable we obtain a regularised time dependent optimal transport problem. In addition, there is a time interpolant between masses given.

Furthermore, the connection to the gradient flow theory is developed in this thesis and several evolution equations are identified as Wasserstein gradient flow equations. This connection yields a new understanding of the partial differential equations and to an improved study of existence, uniqueness and the asymptotic behaviour of solutions. Especially, we study an energy functional which consists of a nonlocal interaction potential and an internal energy. The associated gradient flow equation is an aggregation equation with diffusion term which has applications in chemotaxis or swarming. We provide a detailed analysis for this aggregation equation. In particular, we give existence and uniqueness results of nontrivial stationary solutions. The existence versus nonexistence of such solutions is ruled by a threshold phenomenon, namely nontrivial stationary solutions exist if and only if the diffusivity constant is strictly smaller than the total mass of the interaction kernel. Furthermore, we prove that nontrivial stationary solutions in one space dimension with fixed mass and center of mass are unique. We also provide numerical results.

The central core of this thesis is the discussion of the variational problem with several standard regularisation functionals, e.g. the logarithmic entropy, the $L^2$-regularisation, the Dirichlet-regularisation, the Fisher information and the TV-regularisation. Besides existence and uniqueness results for each regularisation energy we present numerical
results which illustrate the different impact on the data. We calculate the self-similar solutions for the different regularisation functionals to give an idea of the structure of the solutions. We apply the numerical algorithms to synthetic data as well as on real life data.

**Keywords:** Optimal Transport, Variational Calculus, Gradient Flow Theory, Total Variation, Aggregation Equation
Contents

1 Introduction .......................................................... 7
   1.1 Motivation ....................................................... 8
      1.1.1 Estimation Methods ...................................... 8
      1.1.2 Image Analysis ......................................... 11
      1.1.3 Evolution Equations .................................... 13
   1.2 Organisation of this Thesis .................................. 15
   1.3 Acknowledgements ............................................. 17

2 The Theory of Optimal Transport .................................. 19
   2.1 Two Formulations of the Optimal Transport Problem .......... 19
      2.1.1 Monge’s Optimal Transport Formulation ............... 19
      2.1.2 The relaxed Problem: Kantorovich’s Formulation ....... 21
      2.1.3 Existence and Uniqueness ................................ 24
      2.1.4 The Optimal Transport Problem in one Dimension ...... 26
   2.2 Metrics on the Space of Probability Measures ................. 27
      2.2.1 The Wasserstein Distance ................................ 28
      2.2.2 Kantorovich-Rubinstein distance ....................... 29
2.3 The time-dependent Problem ........................................ 30
   2.3.1 Geodesics, Displacement Interpolant and Convexity ......... 30
   2.3.2 The Eulerian Formulation ..................................... 34

3 The Problem Formulation ............................................. 37
   3.1 The Variational Problem ............................................ 37
      3.1.1 Motivation with Bayesian MAP-Estimation .................. 39
      3.1.2 Stability .................................................... 41
      3.1.3 Random Sampling ............................................. 43
   3.2 The Fluid Dynamic Formulation ..................................... 44
      3.2.1 Fluid dynamic Formulation with Regularisation .......... 47
      3.2.2 Optimality Conditions ....................................... 48
      3.2.3 Rescaling and Diffusion Filtering ............................ 49
      3.2.4 Uniqueness ................................................. 50
      3.2.5 Self-Similar Solutions ...................................... 52

4 Gradient Flows .......................................................... 55
   4.1 The Gradient Flow Formulation ................................... 55
   4.2 Motivation of the Gradient Flow Formulation .................. 58
      4.2.1 The discrete variational scheme ............................ 58
      4.2.2 Gradient Flow Formulation on Riemannian Manifolds .... 61
   4.3 Trend to Equilibrium .............................................. 64
      4.3.1 Entropy Dissipation Method ................................. 64
      4.3.2 Otto’s Approach ............................................. 66
5 The Logarithmic Entropy and the $L^2$-regularization

5.1 The Logarithmic Entropy

5.1.1 The Problem Formulation

5.1.2 Existence and Uniqueness

5.1.3 Stationary Solutions

5.1.4 Self-Similar Solutions

5.1.5 Numerical Results and Discussion

5.2 The $L^2$-regularisation

5.2.1 The Problem Formulation

5.2.2 Existence and Uniqueness

5.2.3 Stationary Solutions

5.2.4 Self-Similar Solutions

5.2.5 Numerical Results and Discussion

6 The Gradient Regularisation

6.1 The Dirichlet Regularisation

6.1.1 The Problem Formulation

6.1.2 Existence and Uniqueness

6.1.3 Stationary Solutions

6.1.4 Self-Similar Solutions

6.1.5 Numerical Results and Discussion

6.2 The Fisher Information

6.2.1 The Problem Formulation

6.2.2 Existence and Uniqueness
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2.3</td>
<td>Stationary Solutions</td>
<td>99</td>
</tr>
<tr>
<td>6.2.4</td>
<td>Self-Similar Solutions</td>
<td>99</td>
</tr>
<tr>
<td>6.2.5</td>
<td>Numerical Results and Discussion</td>
<td>101</td>
</tr>
<tr>
<td>6.3</td>
<td>TV-Regularisation</td>
<td>102</td>
</tr>
<tr>
<td>6.3.1</td>
<td>The Problem Formulation</td>
<td>103</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Existence and Uniqueness</td>
<td>105</td>
</tr>
<tr>
<td>6.3.3</td>
<td>Self-Similar Solutions</td>
<td>107</td>
</tr>
<tr>
<td>6.3.4</td>
<td>Numerical Results and Discussion</td>
<td>109</td>
</tr>
<tr>
<td>7</td>
<td>The Nonlocal Interaction Functional</td>
<td>115</td>
</tr>
<tr>
<td>7.1</td>
<td>The Aggregation Equation</td>
<td>115</td>
</tr>
<tr>
<td>7.2</td>
<td>Existence and Uniqueness</td>
<td>117</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Preliminaries</td>
<td>118</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Stationary Solutions in multiple Dimensions</td>
<td>126</td>
</tr>
<tr>
<td>7.2.3</td>
<td>Stationary Solutions in the one dimensional Case</td>
<td>129</td>
</tr>
<tr>
<td>7.3</td>
<td>Numerical Results</td>
<td>143</td>
</tr>
<tr>
<td>8</td>
<td>Numerical Methods</td>
<td>147</td>
</tr>
<tr>
<td>8.1</td>
<td>Numerical Discretisation</td>
<td>147</td>
</tr>
<tr>
<td>8.2</td>
<td>Convex Minimisation Scheme</td>
<td>148</td>
</tr>
<tr>
<td>8.3</td>
<td>The Fluid Dynamic Schemes</td>
<td>149</td>
</tr>
<tr>
<td>8.4</td>
<td>The Gradient Flow Scheme</td>
<td>154</td>
</tr>
<tr>
<td>9</td>
<td>Conclusion</td>
<td>159</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

This thesis deals with the study of a variational problem based on the optimal transport theory. In particular, we are interested in the minimisation of the variational problem

\[ \frac{1}{2} W_2(\nu, u \mathcal{L}^d)^2 + \epsilon E(u) \rightarrow \min_u, \]  

(1.0.1)

consisting of a fidelity term, which is here the quadratic Wasserstein distance between probability measures \( \nu \) and the probability density \( u \mathcal{L}^d \), i.e. \( u \geq 0 \) and \( \int u = 1 \), and a regularisation term \( E(u) \) multiplied with a parameter \( \epsilon > 0 \). \( \mathcal{L}^d \) denotes the usual Lebesgue measure. This model provides conservation of mass and the ability to work with densities continuous with respect to the Lebesgue measure as well as with concentrated densities. Furthermore, we derive an equivalent formulation of the variational problem (1.0.1) in a fluid dynamic framework, based on the work of Benamou and Brenier [17]. A natural connection is also given to the gradient flow equations. The minimisation of a free energy functional

\[ \text{Minimise } E(u), \]  

(1.0.2)

subject to \( \int u = 1, \ u \geq 0 \),

(1.0.3)

can also be formulated as a gradient flow with respect to the Wasserstein metric [99]. For several energy functionals we provide the corresponding gradient flow equation. The regularisation energies which we consider in this thesis are

(i) The logarithmic entropy \( E(u) = \int_{\Omega} u \ln u \ dx, \)

(ii) The \( L^2 \)-regularisation \( E(u) = \frac{1}{2} \int_{\Omega} u^2 \ dx, \)

(iii) The Dirichlet regularisation \( E(u) = \frac{1}{2} \int_{\Omega} |\nabla u|^2 \ dx, \)
(iv) The Fisher information \( E(u) = \int_{\Omega} u |\nabla \ln u|^2 \, dx = \int_{\Omega} \frac{\left| \nabla u \right|^2}{u} \, dx \),

(v) The TV-regularisation \( E(u) := |Du|(\Omega) = \sup_{g \in C^\infty_0(\Omega; \mathbb{R}^N), \|g\|_{\infty} \leq 1} \int_{\Omega} u \cdot \nabla g \, dx \).

We provide a detailed analysis, i.e. existence and uniqueness results for the introduced models and in addition numerical algorithms and results.

1.1 Motivation

The optimal transport problem was first studied by G. Monge in 1781 [131] and then relaxed by L.V. Kantorovich in 1942 [108]. Since then it stirred growing interest in different fields of mathematic, e.g. in the probability theory and statistics [151], the calculus of variations, partial differential equation, kinetic theory [140, 99, 165], geometry of metric measures and functional analysis [15]. Furthermore, the mass transport problem is also used in image processing, e.g. image registration, restoration and image decomposition [10, 179] and also in the data assimilation in the meteorology [63] and oceanography. For an extensive overview on optimal transport and applications we refer to [150, 175, 176, 74, 9, 7].

In the following, we give a motivation for our variational model (1.0.1) and illustrate the properties and advantages of the formulation compared to existing methods. Since the variational problem is formulated for probability densities as well as for point measures we can apply it to different fields, e.g. regression analysis, density estimation and image processing.

Furthermore, the connection to the gradient flow theory given by Jordan, Kinderlehrer and Otto [99], provides a new understanding of partial differential equations. In particular, special evolution equations can be identified as gradient flow equations for some energy functionals. Especially, we consider in this thesis the aggregation equation with nonlinear diffusion term, which has applications e.g. in swarming and chemotaxis, and is the gradient flow equation for an energy functional consisting of an interaction potential and an internal energy.

1.1.1 Estimation Methods

In terrestrial sciences, in particular for the analysis and prediction of earthquakes, densities of intensities and locations of terrestrial incidents are computed from given data measured over time, see [135, 72]. Recent works in [130] aim to estimate crime probabilities for different districts or localities within a city and in [157] estimation methods are used for wildfire predictions. In [160], the authors propose an estimation
method, which incorporates also the spatial information. In medical imaging, e.g., magnetic resonance tomography (MRI), the images of interest are densities restored from undersampled measurements, cf., e.g., [73]. Moreover, in most of the cases the estimation method has to take into account that the given data might not be perfect, but corrupted by noise or blur. Hence an effective method to recover the true density or regression function has to involve some additional smoothing, i.e., regularisation that assumes an a–priori information.

In the density estimation problem we consider a set of sampling points \( X_1, \ldots, X_n \) and the aim is to estimate the underlying density. Here, the term density means that the unknown function in the problem represents a certain quantity per unit measure and is constrained to have a fixed mass, e.g., a probability density would have mass equal to one.

There exist various methods for density estimation, see [159] for a general introduction. A distinction is drawn between parametric methods and non-parametric estimation methods. The parametric ansatz is based on the idea, that the measurements are drawn from one of a known parametric family of distributions, e.g. the normal distribution which is characterised through the mean \( \mu \) and the variance \( \sigma \). Hence, the aim is to calculate this parameters and to obtain the underlying distribution. Our proposed model (1.0.1) does not consider special parameters, but provides a method for estimating the probability density or regression function. It belongs to the non-parametric methods.

A simple example for a non-parametric estimation ansatz is the histogram estimation model. The idea is to assign to every sampling point \( X_1, \ldots, X_n \) a bin, which has a certain width \( h \). Then the estimated density function is defined as

\[
\hat{f}(x) = \frac{1}{nh} \text{(number of } X_i \text{ in the same bin as } x). \tag{1.1.1}
\]

Although the method is very simple, the lack of differentiability of \( \hat{f} \) is a disadvantage of the histogemm estimator. Further, a common method for density estimation is the kernel density estimator [158, 159, 153]. Here the probability density is created by interpolating pointwise kernel densities, i.e., the normal distribution but also other kernels, in every sampling point

\[
\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h} \right).
\]

Consider the normal distribution as the kernel function \( K \) with \( \int K = 1 \). Hence the estimated density function \( \hat{f} \) is the sum of \( n \) small Gaussian curves centered at the data point. The function \( \hat{f} \) inherits all properties of the kernel function \( K \), i.e. smoothness, differentiability and so on. But the goodness of the estimation depends mostly on
the bandwidth \( h \). The bandwidth is the width of the kernel which is considered. See [159, 158] for an extensive introduction into kernel methods.

Another popular density estimation method is the so-called Maximum Penalized Likelihood Estimation (MPLE) [71, 90]. For a given sample \( u(X_i), i = 1, \ldots, n \) of probability values in \( X_i \) the likelihood is defined as

\[
L(u|X_1, \ldots, X_n) = \prod_{i=1}^{n} u(X_i). \quad (1.1.2)
\]

Since the likelihood function can take the value \( \{+\infty\} \) we have to add a penalisation term \( E(u) \) before maximising (1.1.2) or equivalently minimising the negative logarithm. The wanted density function reads

\[
\hat{u} = \text{argmin}_{u, \int_{\Omega} u \, dx = 1} \left\{ \epsilon E(u) - \sum_{i=1}^{N} \log(u(X_i)) \right\}, \quad (1.1.3)
\]

with a parameter \( \epsilon > 0 \) which steers the strength of the regulariser \( E(u) \) in proportion to the maximum likelihood term. Other density estimation methods are, e.g., the taut string method restricted to the univariate case [64] or the logspline technique [113].

The regression analysis is the estimation of a discretely sampled function. In particular, it consist of estimating a function, which approximate given observations or data points in a best way. Let \( (x_1, y_1), \ldots, (x_n, y_n) \) be given identical and independently distributed observations of the pair of random variables \( (X, Y) \). Regression means to estimate the connection between \( X \) and \( Y \). For the introduction to several regression methods, parametric methods as well as non-parametric methods we refer to the survey articles of Blundell and Duncan [31] and Yatchew [178], as well as to [77, 76, 91].

A classical model is the linear regression model, which reads

\[
y_i = m(x_i) + \epsilon_i, \quad i = 1, \ldots, n.
\]

Where \( m \) denotes the regression function and \( \epsilon_i \) describes the measurement error, which occurs during the observation. The goal is to obtain a relationship between \( X \) and \( Y \).

In cases where special structure of densities, e.g. discontinuities, are expected or need to be preserved, standard linear density estimation techniques or regression methods are not appropriate. In this context variational approaches gained significant attention recently, which consist in solving

\[
\hat{u} \in \text{argmin}_{u} \left[ D(u, \nu) + \epsilon E(u) \right], \quad (1.1.4)
\]

where \( D \) is a distance measure between the density \( u \) and the discrete measure \( \nu \) (a sum of point densities), \( \epsilon > 0 \) is a regularisation parameter, and \( E \) an appropriate
regularisation functional. Frequently used examples are a squared $L^2$-distance, using an expansion without the square term of $\nu$ in order to allow point measures (in particular finite sampling)

$$D(u, \nu) = \frac{1}{2} \int_{\Omega} u^2 dx - \int_{\Omega} u \, d\nu. \quad (1.1.5)$$

Steaming from (1.1.4) various estimation methods have been proposed, which differ in the choice of the regulariser $E$, see, e.g., [71, 90, 158, 112, 129, 156]. In particular, in [112, 129, 156] the proposed regulariser $E(u)$ is the total variation (TV) $[8, 75, 154]$ of the probability density $u$. For regularisation the total variation (TV) has become a standard approach due to its capability to preserve edges in the reconstruction. Unfortunately, due to the low regularity of functions of bounded variation, fidelities are not well-defined in this class unless $\nu$ has a regular density. For this reason some presmoothing or preestimation steps are usually needed. In [134] the authors propose a method to estimate a density in one- and two dimensions via TV regularisation with an $L^2$ data discrepancy.

More precisely, starting with a random sampling of data points $X_i$, defined to be
distance between sampling points $\approx$ density,
and a piecewise constant initial guess $d$ for the density, they recover the original density $u^*$ by solving

$$u^* = \arg\min_u \left\{ \frac{1}{2} \|u - d\|_{L^2}^2 + \epsilon \int_{\Omega} |\nabla u| \, dx \right\},$$

where $\int_{\Omega} |\nabla u| \, dx$ equals the total variation of $u$ for $u \in W^{1,1}(\Omega)$.

In [160] the authors further extend the approach in [129] by introducing additional spatial information into the problem. Note that in all of these methods, the property of $u$ to be a probability density, i.e., $\int_{\Omega} u \, dx = 1$, has to be imposed in the method as an additional constraint on the class of admissible solutions. For instance, in [129] this additional constraint is enforced by applying Bregman iteration [138] to the iterative solution of the problem.

### 1.1.2 Image Analysis

The optimal transport ansatz can also be well applied to image processing tasks, e.g. image restoration, image decomposition and registration. Many approaches in this field take advantage of the automatically given mass conservation. Especially, the Wasserstein metric yields a distance function for densities continuous with respect to the Lebesgue measure as well as for concentrated densities.
Image restoration is a common task in image processing, since mostly images are blurred or noisy. Density images from Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET) or images from satellites are afflicted with a special assimilation error. Therefore effective tools for image restoration are wanted [37, 166, 154, 58, 81].

Another branch is the image decomposition. By decomposition we mean in particular the ability to obtain structural properties of a density, such as the main regions of its support and the mass on these regions, which corresponds to the notion of a cartoon as used in image processing. In particular, we want to decompose a given image \( f = u + v \) into \( u \), which is the piecewise smooth part and represents the structure respectively cartoon part of the image and into \( v \), which represents the oscillatory parts respectively the texture. The common decomposition models are based on the Rudin-Osher-Fatemi model [154],

\[
\min_u J(u) = \int_{\Omega} |\nabla u| \, dx + \epsilon \int_{\Omega} |f - u|^2 \, dx \, dy, \tag{1.1.6}
\]

for a given image \( f \in \Omega \subset \mathbb{R}^2 \) on an open and bounded domain. We consider here the TV-regularisation \( E(u) = \int_{\Omega} |\nabla u| \, dx \) due to the property of preserving discontinuities in the data in one dimension. Note, that this is not in general true in two dimensions [5].

The standard image processing methods of decomposing images are based on additive approaches, which are not fitting well to probability measures, since one may argue that the cartoon of a probability measure should be a probability measure as well [139]. We refer to [139, 174] for an overview on decomposition methods. This property is automatically guaranteed in our variational model. Moreover, using the Wasserstein distance as a fidelity term automatically gives us a mass conservation constraint on the solution and as such is density conserving.

The optimal transport ansatz has been recently applied to image registration and warping problems [10, 179], since it yields automatically mass conservation and a parameter free approach. By image registration we mean the transformation of given related data sets into one coordinate system. In particular, for given two images, i.e. mass densities, we want to compute the intermediate images by generating a deformation grid. The image registration problem occurs in medical applications, e.g. MR images are mostly taken at different time or by different assimilation and have to be registered. There exists various methods which consists in minimising a certain cost functional which penalises the mismatch of images. An extensive review on this topic is given in [167]. The novel approach introduced by [10] is based on the optimal transport theory and also on the Benamou-Brenier fluid dynamic ansatz [17]. They provide a variational scheme consisting of a comparison term measuring the change of intensity in the data, which is exactly the density of the image, and the quadratic Wasserstein distance, which measures the transport costs. With this scheme, they take advantage of the mass conservation and the provided time interpolant.
1.1.3 Evolution Equations

Many evolution equations can be identified as gradient flow equations for an energy functional with respect to the Wasserstein metric \cite{99, 142, 124, 84}. In this thesis we study in detail the energy functional which is the sum of a quadratic free energy and the interaction energy

\[ E(u) := \frac{\varepsilon}{2} \int_{\mathbb{R}^N} u^2(x)dx - \frac{1}{2} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} G(x - y)u(y)u(x)dydx. \]  

(1.1.7)

The corresponding gradient flow equation with respect to the Wasserstein distance is the nonlocal aggregation equation with degenerate diffusion \cite{9, 52, 53, 175, 176}, given by

\[ \partial_t u = \nabla \cdot (u \nabla (\varepsilon u - G * u)), \quad \text{in } [0, T) \times \mathbb{R}^N, \]

(1.1.8)

with initial data

\[ u(0, \cdot) = u_0 \quad \text{in } \mathbb{R}^N, \]

where \( u(t, x) \) represents the density of the particles or individuals at time \( t \geq 0 \) and position \( x \in \mathbb{R}^N \). The function \( a(u) = \varepsilon u \) models local repulsion between particles and \( G * u \) the wide-range attraction.

Phenomena with long-range aggregation and short-range repulsion appear in our daily life. In biology for example, the study of swarming of animal behavior, e.g. bird flocks, insect swarms and fish schools but also the chemotaxis of cells received growing interest in the last years \cite{136, 145, 34, 161}. More specifically, the main interest lies in the behavior of organisms and the interaction with each other and with their environment. A variety of mathematical models, e.g. the Patlak-Keller-Segel model \cite{146, 110, 96, 95}, have been proposed to describe this phenomena. We discuss in this thesis the aggregation equation (1.1.8) which arises in a natural way as the limit of a stochastic interacting particle model with pair interactions (cf. e.g., \cite{132, 137} respectively \cite{89}). A similar model is studied in \cite{111, 16}.

The following three types of kernels are known in literature:

1. The regular interaction potential: A typical example is the local repulsive double-well potential \( G(x) = x^4 - x^2 \) but also the quadratic Morse potential

\[ G(x) = -C_a\varepsilon^{-x^2/l_a} + C_r\varepsilon^{-x^2/l_r}. \]

\( G \) is typically convex. See \cite{152, 79, 78, 39} for more results on regular kernels.

2. The interaction potential with an attractive singularity at \( x = 0 \): The kernel

\[ G(x) = -\frac{1}{2\pi} \log |x| \]

appears e.g. in swarm models and chemotaxis \cite{49, 118, 23}. 

3. The interaction potential with an repulsive singularity at \( x = 0 \): The attractive-repulsive Morse potential

\[
G(x) = -C_a \varepsilon^{-|x|/l_a} + C_r \varepsilon^{-|x|/l_r},
\]

is an example for this kind of kernel \([152, 79, 78]\). It appears e.g. in swarming models.

In this thesis we consider the interaction kernel \( G \) to be radial and attractive, nonnegative and integrable, i.e. \( G(x) = g(|x|) \) with \( g'(r) > 0 \) as \( r > 0 \).

Models of the above form have been investigated with respect to several aspects. The existence and uniqueness using entropy solution techniques was studied in \([38, 27]\). The well-posedness in the context of Wasserstein gradient flows and - in particular in connection with the classical Patlak-Keller-Segel model for chemotaxis \([146, 110]\) - with respect to blow-up versus large-time existence was discussed in \([98, 92, 29, 60, 16, 111]\) for models with linear diffusion and \([45]\) for models with nonlinear diffusion.

In particular, if nonlocal repulsion acts at a smaller scale with respect to nonlocal attractive forces in the large particle limit, then a nonlocal repulsion term can be replaced by a local term with nonlinear diffusion, we refer to \([128, 33, 168, 169, 59, 68, 107, 149, 32, 119]\) for several examples.

An interesting and important question is the characterisation of large-time behavior of solutions to equations of the form \((1.1.8)\), which is related with the possible existence of nontrivial steady states. This issue is solved in detail for purely diffusive equations, in which solutions decay to zero with a prescribed rate for large times. In particular, they behave like the (compactly supported) Barenblatt-Prattle profiles \([14, 148, 173, 142, 57]\). In the purely nonlocal case, namely when \( \varepsilon = 0 \), this asymptotic behavior has been studied extensively in many papers \([120, 39, 118, 24, 23, 49, 25, 97, 79, 78]\), combined with the study of the regularity of solutions compared to the attractive singularity of the interaction kernel. In particular, solutions are known to concentrate to a Dirac delta centered at the initial center of mass (invariant) either in a finite or in an infinite time, depending on the properties of the kernel \( G \) at \( x = 0 \). When the kernel \( G \) is supported on the whole space, the Dirac delta is the unique steady state of \((1.1.8)\) with unit mass and zero center of mass.

The asymptotic behavior in the general case with both nonlinear diffusion and nonlocal interaction has been only partially addressed. A first attempt in this direction was performed in \([39]\), in which the existence of steady states of \((1.1.8)\) for sufficiently small \( \varepsilon \) and the non-existence for large \( \varepsilon \) in the one-dimensional case was proven by means of the pseudo-inverse representation of the Wasserstein distance. More refined results in a similar model derived in \([93, 144]\) with cut-off density have been found in \([40, 117]\).
A key open problem in this context is the uniqueness of stationary solutions under mass and center of mass constraint, its main difficulty being the fact that the functional $E$ is neither convex in the classical sense nor in the displacement convex sense \cite{126}.

In this thesis we prove a threshold phenomenon, namely that stationary solutions of (1.1.8) exist if and only if the diffusivity constant is strictly smaller than the total mass of the interaction kernel, i.e.:

- If $\varepsilon \geq \int G$, then there exists no steady state.
- If $\varepsilon < \int G$, there exists a stationary state.

Furthermore, we investigate a detailed structure of steady states in one space dimension. Especially, we prove that nontrivial stationary solutions in one space dimension with fixed mass and center of mass are unique. If $G'$ only vanishes at zero, then there exists a unique stationary state (up to translation), which is a minimiser of the energy $E$ at fixed mass. The stationary states have compact support, which increases with $\varepsilon$. Moreover, such a steady state is symmetric with a single maximum at the center of mass, and monotone on both sides of the center of mass. If $\varepsilon$ is small enough, the stationary states are concave on their support.

\section*{1.2 Organisation of this Thesis}

The work is organised as follows:
In Chapter \ref{ch2}, we give a brief introduction to the optimal transport theory. We introduce Monge’s and Kantorovich’s optimal transport problem in Section \ref{sec2.1}. Additionally, we recall existence and uniqueness results for both problems and the special case of transport on the real line. In Section \ref{sec2.2}, we are concerned with the notion of the Wasserstein distance in the space of probability measures. The last Section \ref{sec2.3} is devoted to the time-dependent problem. The definition of a constant speed geodesics and of displacement convexity are provided.
In Chapter \ref{ch3} we introduce two different problem formulations which we are interested in. The formulation of the variational problem, based on the Kantorovich formulation, is established in Section \ref{sec3.1}. We give a motivation with Bayes’ maximum-a-posteriori probability estimation model. Furthermore, we discuss stability of the variational model. In Section \ref{sec3.2} we introduce an time-dependent model based on the Benamou-Brenier formulation. We derive optimality conditions and give a uniqueness result for the velocity. Calculations for the self-similar solutions are done in the last Section. The fact that some evolution equations can be formulated as gradient flows for special
energy functionals with respect to a certain metric is known for a long time. In Chapter 4 we provide an insight into the wide field of gradient flows and the connection to the Wasserstein metric. In Section 4.1 we exemplify this idea by identifying some special evolution equations as gradient flows for some energy functionals with respect to the Wasserstein distance. Section 4.2 is concerned with two different approaches to the gradient flow theory. On the one hand we present the minimising movement scheme which can be recast in different metric frameworks. On the other hand we introduce the approach on a Riemannian manifold. In Section 4.3 we introduce two different methods to study the asymptotic behavior of solutions, namely the Entropy dissipation method and the method based on calculations on a Riemannian manifold. The following three chapters are concerned with different energy functionals. In particular, we discuss in Chapter 5 the logarithmic entropy and the \(L^2\)-regularisation, in Chapter 6 the Dirichlet regularisation, the Fisher information and the TV-regularisation and in Chapter 7 a nonlocal interaction potential. For each model, we provide existence and uniqueness results and compute self-similar solutions. Furthermore, we present numerical results. Especially in Chapter 7 we discuss the structure of the stationary solutions for the associated gradient flow equation, which is an aggregation equation with diffusion term. We provide a detailed analysis, i.e. we give existence and uniqueness results of nontrivial stationary solutions in Section 7.2. The existence versus nonexistence of such solutions is ruled by a threshold phenomenon. We prove, that nontrivial stationary solutions exist if and only if the diffusivity constant is strictly smaller than the total mass of the interaction kernel (see Section 7.2.2). Furthermore, in Section 7.2.3 we prove that nontrivial stationary solutions in one space dimension with fixed mass and center of mass are unique. In Section 7.3 we provide numerical results.

Chapter 8 is devoted to the numerical algorithms for the optimisation problems introduced in Chapter 3 and 4. In particular, we introduce an algorithm for the variational problem (3.1.1) in one dimension in Section 8.2. Additionally, we propose in Section 8.3 algorithms for the Benamou-Brenier formulation which can be implemented in two dimensions. Section 8.4 is devoted to the numerical study of the gradient flow equations introduced in Chapter 4.

In the last Chapter 9 we summarise the results and address open problems.

Parts of the results presented in Chapter 3, 5, 6 and 8 are submitted [42]. The results we present in Chapter 7 are also submitted [41].
1.3 Acknowledgements

First of all I want to thank my supervisor Prof. Dr. Martin Burger for his support and the interesting research topic. It was a challenging task, but mostly I loved to work on it.
Also I would like to thank Prof. Dr. Daniel Matthes for being the co-referee of my thesis.
Besides, I would like to thank Carola-Bibiane Schönlieb for her support. She is a guide for me and inspires me a lot. Also, I thank Marco Di Francesco for the joint work and the hospitality in L’Aquila.
I thank my colleagues Christina Stöcker, Martin Benning and Bärbel Schlake - it was a pleasure to share an office with you - and all my colleagues in the Imaging working group. Also I want to thank Kathrin Smetana and Patrick Henning for their support and friendship.
Special thanks go to my family.
Chapter 2

The Theory of Optimal Transport

In this chapter we give a brief introduction to the optimal transport theory. In particular, in Section 2.1 we introduce the optimal transport formulations given by Gaspard Monge in 1781 [131] and by Leonid V. Kantorovich in 1942 [108, 109]. Additionally, existence and uniqueness results as well as the special case of transport on the real line are discussed. The definition of the Wasserstein distance is given in Section 2.2. The last Section 2.3 deals with the time-dependent transport problem and the definition of geodesics and displacement convexity. We refer for an extensive overview to the books of Villani [175, 176] and Ambrosio et. al. [9]. For a probabilistic introduction we refer to [151].

2.1 Two Formulations of the Optimal Transport Problem

This section is concerned with two different formulations of the optimal transport problem, namely Monge’s and Kantorovich’s formulations. The difference between this two approaches is that Monge’s formulation does not allow the splitting of mass, whereas Kantorovich’s approach does.

2.1.1 Monge’s Optimal Transport Formulation

In 1781, the French mathematician Gaspard Monge published in his work “Mémoire sur la théorie des déblais et des remblais” [131] the optimal transport problem, which is nowadays still an interesting and challenging problem with applications in different
mathematical fields. The initial problem is the following: We start with a pile of material (the déblai) and we want to transport it completely into a hole which has the same volume (see Figure 2.1.1). The aim is to transport the mass with minimal cost. Mathematically, we formulate this optimisation problem with probability measures on

![Diagram](image)

**Figure 2.1: Mass transportation problem (see [175]).**

a certain space, here Polish spaces which are complete and separable equipped with a metric. Let \( X \) and \( Y \) be two Polish spaces with associated probability measures \( \mu \) and \( \nu \) modeling the pile of sand, respectively the hole. For \( A \subset X \) and \( B \subset Y \) measurable we obtain the following interpretation

\[
\mu[A] \text{ is the amount of material in } A,
\]

\[
\nu[B] \text{ specifies the amount of mass which fits into } B.
\]

The nonnegative measurable function \( c : X \times Y \to \mathbb{R}_+ \cup \{+\infty\} \) is called cost function and measures the cost which arises during the transport. The transport is modeled by a map, the so called transport map \( T : X \to Y \). The push-forward is defined as follows [175]:

**Definition 2.1.1 (push forward).** Let \( \mu \) and \( \nu \) be Borel probability measures on \( X \) respectively \( Y \) and \( T : X \to Y \), then

\[
\nu = T_#\mu
\]

(2.1.1)

defines the push-forward, or image measure of \( \mu \) by \( T \). We can say that the map \( T \) transports the mass represented by the measure \( \mu \) to the mass represented by the measure \( \nu \).

Then, for any measurable subset \( B \subset Y \) we can follow

\[
\nu[B] = \mu[T^{-1}(B)],
\]

(2.1.2)

or more generally for \( f : Y \to \mathbb{R} \) bounded

\[
\int_X f(T(x))d\mu(x) = \int_Y f(y)dT_#\mu(y).
\]

(2.1.3)

With this preliminaries we can formulate Monge’s optimal transport problem:
Definition 2.1.2 (Monge’s optimal transport problem). ([9, Chapter 6]) Let \( \mu \) and \( \nu \) be Borel probability measures on Polish spaces \( X \) respectively \( Y \), and \( c : X \times Y \to \mathbb{R}_+ \cup \{+\infty\} \) the nonnegative measurable cost function. Let \( T : X \to Y \) be the transport map. Then Monge’s optimal transport problem reads

\[
\inf_{T} \left\{ \int_X c(x, T(x)) \, d\mu(x) \mid T_\# \mu = \nu \right\}.
\]  

(2.1.4)

If a minimiser of (2.1.4) exists, it is called the optimal transport map and the associated cost is called optimal transport cost. Although Monge’s mass transport problem seems to be simple it is not easy to solve and furthermore the existence and uniqueness of solutions cannot be expected. This is due to the highly nonlinear constraint on the transport map \( T \). For \( \mu \) and \( \nu \) absolutely continuous probability measures with respect to the Lebesgue measure and associated density functions \( f \) respectively \( g \) we write

\[ d\mu(x) = f(x)dx, \quad d\nu = g(y)dy. \]

Then with the change of variable formula we can rewrite the push forward definition to

\[ f(x) = g(T(x))|\det \nabla T(x)|, \]  

(2.1.5)

which illustrates the highly nonlinear constraint on \( T \). If \( g \) is positive, (2.1.5) can be rewritten to

\[ \det \nabla T(x) = \frac{f(x)}{g(T(x))}, \]  

(2.1.6)

which is a particular case of the Monge-Ampère equation.

The first analytical discussion of Monge’s optimal transport problem (2.1.4) was done by Sudakov in 1979 [164], but the first correct existence proof was given not until 1996 by Gangbo and Evans (cf. [82]). In general we cannot expect existence or uniqueness. A simple example is the following: Let \( \mu = \delta_0 \) and \( \nu = (\delta_{-1/2} + \delta_{1/2})/2 \), then there exists no transport map, which pushes \( \mu \) to \( \nu \), since Monge’s formulation does not allow the splitting of mass. Furthermore, let \( \mu \) give mass to small sets and \( \mu \) and \( \nu \) are probability measures on \( \mathbb{R}^2 \), concentrated on \( \{(0, 0), (1, 1)\} \) respectively \( \{(0, 1), (1, 0)\} \). Then there exists no unique transport map which pushes \( \mu \) to \( \nu \).

However, under some assumptions on the cost function and on \( \mu \) and \( \nu \), there exists a unique optimal transport map, which is the gradient of a convex function. This result is known as Brenier’s theorem [35] which we will state in the following (see Theorem 2.1.7).

### 2.1.2 The relaxed Problem: Kantorovich’s Formulation

About two hundred years later Leonid V. Kantorovich introduced an optimal transport problem in the context of economical processes (cf. [108, 109]). His research was
awarded a Nobel prize for economics in 1975 and furthermore many mathematician recognised the importance of his theories for their special fields. Kantorovich’s formulation can be seen as a relaxed version of Monge’s optimal transport formulation (2.1.4). By embedding the map $T$ into the space $L^1(X; \mathcal{P}(Y))$ with $T : [x \mapsto \delta_{T(x)}]$, we extend the class of transport maps to the class of transport plans defined on the product space $\mathcal{P}(X \times Y)$. Hence, the nonlinear constraint on $T$ (2.1.5) turns to a linear constraint on the transport plan $\pi$:

**Definition 2.1.3** (Transport plan). A transport plan is a probability measure $\pi$ defined on the product space $X \times Y$ with values in $\mathbb{R}_+$. Then we can formulate the marginal conditions

$$
\int_Y d\pi(x, y) = d\mu(x), \quad \text{all the mass taken from point } x \text{ coincides with } d\mu(x), \quad (2.1.7)
$$

$$
\int_X d\pi(x, y) = d\nu(y), \quad \text{all the mass transferred to } y \text{ coincides with } d\nu(y). \quad (2.1.8)
$$

For all measurable subsets $A \subset X$ and $B \subset Y$ we can also write

$$
\pi[A \times Y] = \mu[A], \quad \pi[X \times B] = \nu[B], \quad (2.1.9)
$$

which is equivalent to

$$
\int_{X \times Y} (\varphi(x) + \psi(y)) d\pi(x, y) = \int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y),
$$

for all test functions $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$. With this preliminaries we can state the Kantorovich problem:

**Definition 2.1.4** (Kantorovich optimal transport problem). Let $\mu$ and $\nu$ be Borel probability measures on Polish spaces $X$ respectively $Y$ and $\pi$ is a probability measure defined on a product space $\mathcal{P}(X \times Y)$ fullfilling the marginal conditions (2.1.9). $c$ is a measurable lower semicontinuous function. Then the Kantorovich problem reads

$$
\inf_{\pi \in \Pi(\mu, \nu)} I[\pi] = \int_{X \times Y} c(x, y) \ d\pi(x, y), \quad (2.1.10)
$$

where $\Pi$ is the nonempty set of admissible transport plans defined as

$$
\Pi(\mu, \nu) = \{ \pi \in \mathcal{P}(X \times Y) | \pi[A \times Y] = \mu[A], \pi[X \times B] = \mu[B] \} . \quad (2.1.11)
$$

$I[\pi]$ is called total transportation cost associated to $\pi$, and $T_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} I[\pi]$ the optimal transportation cost between $\mu$ and $\nu$. If a minimiser $\pi$ exists, it is called
the optimal transport plan. Moreover, we can switch from Kantorovich’s formulation (2.1.10) to Monge’s mass transport problem (2.1.4) by introducing the following relationship

\[ d\pi(x, y) = d\pi_T(x, y) \equiv d\mu(x)\delta[y = T(x)], \]  

(2.1.12)

or equivalently \( \pi_T = (\text{Id} \times T)_#\mu \) and

\[ \int_{X \times Y} \xi(x, y) d\pi_T(x, y) = \int_X \xi(x, T(x)) d\mu(x) \]

for a nonnegative measurable function \( \xi \). Then we say that the transport plan \( \pi \) is induced by the transport map \( T \). The existence for the Kantorovich problem then automatically implies the existence of the Monge problem.

Kantorovich introduced additionally a dual formulation for (2.1.10)

**Definition 2.1.5** (Duality theorem). ([175, Chapter 1]) Let \( X, Y \) be Polish spaces and \( \mu \) and \( \nu \) Borel probability measures on \( X \) resp. \( Y \). Let \( c \) be a lower semicontinuous function. Let \( \pi \in \mathcal{P}(X \times Y) \) and \((\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)\),

\[ I[\pi] = \int_{X \times Y} c(x, y) d\pi(x, y), \quad J(\varphi, \psi) = \int_X \varphi(x) d\mu + \int_Y \psi(y) d\nu, \]  

(2.1.13)

with \( \Pi(\mu, \nu) \) defined as in (2.1.11) and \( \Phi_c \) is the set of all measurable functions \((\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)\) with

\[ \varphi(x) + \psi(y) \leq c(x, y) \]  

(2.1.14)

for \( d\mu \)-almost all \( x \in X \) and \( d\nu \)-almost all \( y \in Y \), then

\[ \inf_{\pi \in \Pi(\mu, \nu)} I[\pi] = \sup_{(\varphi, \psi) \in \Phi_c} J(\varphi, \psi). \]  

(2.1.15)

For the proof we refer to [175].

When the cost function \( c \) is bounded, we can write \( J(\varphi, \psi) \) in terms of \( c \)-concave functions \((\varphi^{cc}, \varphi^c)\):

**Remark 2.1.6.** ([175, Chapter 1, Remark 1.12]) Let \( \varphi \) be bounded. Then the pair of conjugate \( c \)-concave functions \((\varphi^{cc}, \varphi^c)\) is defined as

\[ \varphi^c(y) = \inf_{x \in X} [c(x, y) - \varphi(x)], \quad \varphi^{cc}(x) = \inf_{y \in Y} [c(x, y) - \varphi^c(y)]. \]  

(2.1.16)

Additionally \( \varphi^c \) is measurable due to

\[ \varphi^c = \lim_{l \to \infty} \psi_l(y) = \lim_{l \to \infty} \inf_{x \in X} [c_l(x, y) - \varphi(x)], \]
with each $\psi_l$ uniformly continuous. With the inequalities
\[-\sup \varphi \leq \varphi^c \leq \|c\|_{\infty} - \sup \varphi\]
\[-\sup \varphi^c \leq \varphi = \varphi^{cc} \leq \|c\|_{\infty} - \sup \varphi^c\]
we can write the supremum of $J$ as a function of conjugate $c$-concave functions $(\varphi^{cc}, \varphi^c)$:
\[
\sup \{J(\varphi, \psi) | (\varphi, \psi) \in \Phi_c\} = \sup \{J(\varphi, \psi) | (\varphi, \psi) \in \Phi_c, \ 0 \leq \varphi \leq \|c\|_{\infty}, -\|c\|_{\infty} \leq \psi \leq 0\}
= \sup \{J(\varphi^c, \varphi^{cc}) | \varphi \in L^1(d\mu)\}.
\]

### 2.1.3 Existence and Uniqueness

As mentioned before we cannot expect the existence and uniqueness of an optimal transport map for the Monge optimisation problem (2.1.4). However, under some assumptions on the cost function $c$ and the probability measures $\mu$ and $\nu$ we obtain the following existence theorem (cf. [173, Theorem 2.12]).

**Theorem 2.1.7.** Let $X$ and $Y$ be Polish spaces. We consider here $c(x, y) = |x - y|^p$ with $p = 2$ and $\mu \in \mathcal{P}_2(X)$, $\nu \in \mathcal{P}_2(Y)$ be probability measures with second finite moment.

(i) **Knott-Smith optimality criterion**

A transport plan $\pi \in \Pi(\mu, \nu)$ is optimal for the Kantorovich problem (2.1.10), if and only if there exists a convex lower semi-continuous function $\varphi$ such that
\[
\text{Supp}(\pi) \subset \text{Graph}(\partial \varphi),
\]
or equivalently: for $d\mu$-almost all $(x, y)$, $y \in \partial \varphi(x)$.
Moreover, the pair $(\varphi, \varphi^*)$ has to be minimiser of
\[
\inf \left\{ \int_{\mathbb{R}^N} \varphi d\mu + \int_{\mathbb{R}^N} \psi d\nu ; \forall (x, y), \ x \cdot y \leq \varphi(x) + \psi(y) \right\}.
\]

(ii) **Brenier’s Theorem (cf. [35])**

If $\mu$ does not give mass to small sets, then there is a unique optimal transport plan $\pi$, which is
\[
d\pi(x, y) = d\pi_T(x, y) = d\mu(x)\delta[y = \nabla \varphi],
\]
or equivalently $\pi = (\text{Id} \times \nabla \varphi)_\# \mu$ and $\nabla \varphi$ is the unique (i.e. uniquely determined $d\mu$-almost everywhere) gradient of a convex function $\varphi$ with
\[
\nabla \varphi_\# \mu = \nu.
\]
Moreover: $\text{Supp}(\nu) = \overline{\nabla \varphi(\text{Supp}(\mu))}$. 

(iii) With (ii) we get: $\nabla \varphi$ is the unique solution to Monge's optimal transport problem, i.e.:

$$\int_{\mathbb{R}^N} |x - \nabla \varphi(x)|^2 d\mu(x) = \inf_T \int_{\mathbb{R}^N} |x - T(x)|^2 d\mu(x).$$

(iv) If $\nu$ also does not give mass to small sets, then for $d\mu$ almost all $x$ and $d\nu$-almost all $y$

$$\nabla \varphi^* \circ \nabla \varphi(x) = x, \quad \nabla \varphi \circ \nabla \varphi^*(y) = y,$$

and $\nabla \varphi^*$ is the unique gradient of a convex function with $\nabla \varphi^* \# \nu = \mu$.

For $\nabla \varphi$ smooth and one-to-one (2.1.5) reads

$$f(x) = g(\nabla \varphi(x)) \det D^2 \varphi(x),$$

and with $g$ positive

$$\det D^2 \varphi(x) = \frac{f(x)}{g(\nabla \varphi(x))}.$$

**Proof.** For an explicit proof of Theorem 2.1.7 we refer to [175]. In the following we only recover the ideas to prove the existence of a minimiser for the Kantorovich problem (2.1.10).

Let $X = Y = \mathbb{R}^n$ and $c(x, y) = \frac{|x-y|^2}{2}$ be lower semicontinuous and $\mu \in \mathcal{P}_2(X)$ and $\nu \in \mathcal{P}_2(X)$. Then

$$M_2 = \int_{\mathbb{R}^N} \frac{|x|^2}{2} d\mu(x) + \int_{\mathbb{R}^N} \frac{|y|^2}{2} d\nu(y) < +\infty,$$

and $I[\pi]$ is always finite on $\Pi(\mu, \nu)$, indeed

$$I[\pi] = \int_{\mathbb{R}^N \times \mathbb{R}^N} \frac{|x - y|^2}{2} d\pi(x, y),$$

$$\leq \int_{\mathbb{R}^N} |x|^2 d\mu(x) + \int_{\mathbb{R}^N} |y|^2 d\nu(y) := 2M_2.$$

Furthermore, for $\mu$ and $\nu$ tight, i.e. for $\delta > 0$, $K \subset X, L \subset Y$, such that $\mu(X \setminus K] \leq \delta, \nu[Y \setminus L] \leq \delta$, we conclude for an admissible transport plan $\pi \in \Pi(\mu, \nu)$ the tightness of $\Pi$:

$$\pi[(X \times Y) \setminus (K \times L)] \leq \pi[X \times (Y \setminus L)] + \pi[(X \setminus K) \times L] \leq 2\delta.$$

Using Prokhorov’s lemma we deduce the relative compactness of the set $\Pi$ with respect to the weak topology. Hence, there exists a minimising sequence of transport plans $\pi_k$.
and a cluster point $\pi_* \in \Pi$. Let $(c_l)_{l \in \mathbb{N}}$ be a nondecreasing sequence of bounded and continuous functions. Using the minimising property of $(\pi_k)$ and $c_l \leq c$ we conclude

$$I(\pi_*) = \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y) d\pi_*(x, y) = \lim_{l \to \infty} \int c_l(x, y) d\pi_*(x, y) \leq \lim_{l \to \infty} \limsup_{k \to \infty} \int_{\mathbb{R}^n} c_l(x, y) d\pi_k(x, y) \leq \limsup_{k \to \infty} \int_{\mathbb{R}^n} c(x, y) d\pi_k(x, y) = \inf_{\pi \in \Pi(\mu, \nu)} I(\pi).$$

Eventually we proved that $\pi_*$ is a minimiser of $I$. \hfill \Box

In general, the existence of a solution depends on the cost function, on the structure of the space as well as on the regularity of $\mu$ and $\nu$. For strictly convex cost function $c$ and $\mu, \nu$ being absolutely continuous with respect to the Lebesgue measure the existence of an unique optimal transport plan can be proved. For $c(x, y) = |x - y|$, which was the cost function Monge considered originally, we cannot ensure existence. In general no solution exists for $c(x, y) = |x - y|^p$ if $p < 1$. However, Cotar, Friesecke and Klüppelberg proved recently the uniqueness of an optimal map for $p = -1$ (cf. [61]).

### 2.1.4 The Optimal Transport Problem in one Dimension

The restriction to one dimension, i.e. $X = Y = \mathbb{R}$ simplifies the optimal transport problem, since we can write the probability measures $\mu$ and $\nu$ using cumulative distribution functions, i.e.

$$F(x) = \int_{-\infty}^{x} d\mu(x) = \mu([-\infty, x]),$$

where $F : \mathbb{R} \to [0, 1]$ is a right continuous, nondecreasing function with $F(-\infty) = 0, F(\infty) = 1$. Furthermore we can define the pseudo inverse of $F$ by

$$u(t) = F^{-1}(t) = \inf \{x \in \mathbb{R} | F(x) > t\}. \quad (2.1.21)$$

$F^{-1}$ is also right-continuous and satisfies

$$F^{-1}(F(x)) \geq x \ \forall x \in \mathbb{R}; \quad F(F^{-1}(t)) \geq t \ \forall t \in [0, 1]. \quad (2.1.22)$$

Transport plans can be respresented by joint two dimensional cumulative distribution functions

$$H(x_0, y_0) = \int_{R(x_0, y_0)} d\pi = \pi[R(x_0, y_0)]. \quad (2.1.23)$$

and $R(x_0, y_0) = \{(x, y) \in \mathbb{R}^2 | x \leq x_0, y \leq y_0\}$ denotes a rectangle. Then we can formulate the following theorem:
Theorem 2.1.8. ([175, Theorem 2.18]) Let $\mu_1, \mu_2$ be two probability measures on $\mathbb{R}$ with cumulative distribution functions $F_1$ and $F_2$. Let $\pi$ be the probability measure on $\mathbb{R}^2$ with joint two-dimensional cumulative distribution function

$$H(x, y) = \min(F_1(x), F_2(y)).$$

(2.1.24)

Then $\pi$ belongs to $\Pi(\mu_1, \mu_2)$ and is optimal in the Kantorovich transportation problem between $\mu_1$ and $\mu_2$ for the quadratic cost $c(x, y) = |x - y|^2$.

Theorem 2.1.9. ([9, Theorem 6.0.]) Let $\mu_1, \mu_2$ be two probability measures on $\mathbb{R}$ with cumulative distribution functions $F_1$ and $F_2$. Let $\pi$ be the probability measure on $\mathbb{R}^2$ and $c(x, y) = h(x - y)$, with $h \geq 0$ convex and with $p$ growth.

(i) If $\mu$ has no atom, i.e. $F_1$ is continuous, then $F_2^{-1} \circ F_1$ is an optimal transport map. It is the unique optimal transport map if $h$ is strictly convex.

(ii) We have

$$\min \left\{ \int_{\mathbb{R}^2} c(x, y)d\pi \mid \pi \in \Pi(\mu_1, \mu_2) \right\} = \int_0^1 c(F_1^{-1}(t), F_2^{-1}(t))dt.$$  

Proof. We refer for the proof to [9].

For the special case of quadratic Euclidean distance $c(x, y) = |x - y|^2$ we obtain

$$T_2(\mu_1, \mu_2) = \int_0^1 |F_1^{-1}(t) - F_2^{-1}(t)|^2dt.$$  

(2.1.25)

The latter equality results from the following computation

$$T_2(\mu_1, \mu_2) = \int_{\mathbb{R}} |x - T(x)|^2d\mu_1(x) = \int_0^1 |F_1^{-1}(t) - T(F_1^{-1}(t))|^2dt,$$

$$= \int_0^1 |F_1^{-1}(t) - F_2^{-1}(t)|^2dt \left( = \int_0^1 |u_1(t) - u_2(t)|^2dt \right).$$

2.2 Metrics on the Space of Probability Measures

In the next section we introduce the definition of the Wasserstein distance, which yields a way to measure the distance between probability measures. In computer sciences the Wasserstein distance is called earth mover’s distance, since it can be interpreted as the minimal cost which is necessary to transport the pile of material to the hole. We refer to the book of Rachev and Rüschendorf for a detailed introduction [151, 155].
2.2.1 The Wasserstein Distance

The Wasserstein distance, introduced in 1969 by the Russian mathematician Leonid N. Vasershtein, is defined in the following way:

**Definition 2.2.1** (Wasserstein distance). ([9, Chapter 7]) Let \( X = Y \) be a Polish space induced with a metric \( d \). Let \( \mu \) and \( \nu \) be probability measures on \( X \). Then the Wasserstein distance of order \( p \) with \( p \geq 1 \) between \( \mu \) and \( \nu \) is defined as

\[
W_p(\mu, \nu) = \left( \inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times X} d(x, y)^p d\pi(x, y) \right)^{\frac{1}{p}}.
\]  

(2.2.1)

where \( \Pi(\mu, \nu) \) is the set of admissible transport plans (see (2.1.11)).

The Wasserstein distance \( W_p \) is really a metric on \( \mathcal{P}(X) \), since it satisfies the conditions

(i) Let \( \mu, \nu \in \mathcal{P}(X) \), if \( W_2(\mu, \nu) = 0 \) then \( \mu = \nu \).

(ii) \( W_p(\mu, \nu) = W_p(\nu, \mu) \)

(iii) Let \( \mu^i \in \mathcal{P}(X), \ i = 1, 2, 3 \), then \( W_p(\mu^1, \mu^3) \leq W_p(\mu^1, \mu^2) + W_p(\mu^2, \mu^3) \).

See [9, Chapter 7] for the proof.

The Wasserstein distance defines a distance on the space \( \mathcal{P}_p(X) \), which is the space of probability measures with finite \( p \)-th moment

\[
\mathcal{P}_p(X) := \left\{ \mu \in \mathcal{P}(X) \mid \int_X d(x_0, x)^p d\mu(x) < +\infty \right\},
\]  

(2.2.2)

for any \( x_0 \in X \). If \( d \) is bounded, then \( \mathcal{P}_p(X) \) coincides with \( \mathcal{P}(X) \). A more probabilistic interpretation of the Wasserstein distance is the following:

\[
W_p(\mu, \nu)^p = \inf E[|X - Y|^p],
\]  

(2.2.3)

where \( E[Z] \) denotes the expected value of a random variable \( Z \). The infimum is taken over all simultaneous distributions of the random variables \( X \) and \( Y \), and the marginal conditions must be satisfied.

An important proposition, which we will use in the sequel to prove existence is the following result on the stability of optimality and narrow lower semicontinuity of the Wasserstein distance:
Proposition 2.2.2. (Proposition 7.1.3. in [9]) Let \((X,d)\) be a metric space. Let \((\mu_1^n, \mu_2^n)\) ⊂ \(P(X)\) be two sequences narrowly converging to \(\mu_1, \mu_2\) respectively, and let \(\pi_n \in \Pi(\mu_1^n, \mu_2^n)\) be a sequence of corresponding optimal transport plans with \(\int_X d(x_1, x_2) \ d\pi_n(x, y)\) bounded. Then \((\pi_n)\) is narrowly relatively compact in \(P(X^2)\) and any narrow limit point \(\pi \in \Pi(\mu_1, \mu_2)\) is an optimal transport plan for \((\mu_1, \mu_2)\), with

\[
W_p(\mu_1^1, \mu_2^2) \leq \liminf_{n \to \infty} \int_{X^2} d(x, y)^p \ d\pi_n(x, y) \tag{2.2.4}
= \liminf_{n \to \infty} W_p(\mu_1^n, \mu_2^n). \tag{2.2.5}
\]

Proposition 2.2.3. (9 Proposition 7.1.5]) \(P_p(X)\) endowed with the \(p\)-Wasserstein distance is a separable metric space which is complete if \(X\) is complete. A set \(K \subset P_p(X)\) is relatively compact if and only if it is \(p\)-uniformly integrable and tight. In particular, for a given sequence \((\mu_n) \subset P_p(X)\) we have the equivalence

(i) \(W_p(\mu_n, \mu) \to 0\) for \(n \to \infty\).

(ii) \(\mu_n \to \mu\) narrowly for \(n \to \infty\) and there is a convergence of the moment of order \(p\)

\[
\int_X d(x_0, x)^p d\mu_n(x) \to \int_X d(x_0, x)^p d\mu(x).
\]

2.2.2 Kantorovich-Rubinstein distance

For the special case \(p = 1\) in (2.2.1) and \(d\) denoting a metric on \(X\), we obtain the so called Kantorovich-Rubinstein distance:

Theorem 2.2.4. (Kantorovich-Rubinstein theorem)([175, Theorem 1.14]) Let \(X = Y\) be a Polish space induced with a metric \(d\). Let \(\mu\) and \(\nu\) be probability measures on \(X\). Then the Kantorovich-Rubinstein distance between \(\mu\) and \(\nu\) is defined as

\[
W_1(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times X} d(x, y) d\pi(x, y). \tag{2.2.6}
\]

Let \(\text{Lip}(X)\) denote the space of all Lipschitz functions on \(X\), and

\[
\|\varphi\|_{\text{Lip}} \equiv \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x, y)},
\]

then we can rewrite (2.2.6) to

\[
W_1(\mu, \nu) = \sup \left\{ \int_X \varphi d(\mu - \nu) \mid \varphi \in \cap L^1(d|\mu - \nu|), \|\varphi\|_{\text{Lip}} \leq 1 \right\}. \tag{2.2.7}
\]
In this special case, the Kantorovich-Rubinstein theorem implies that the total cost only depends on the difference \((\mu - \nu)\). Moreover we can define a norm on \(M_1(X)\), which is the vector space generated by \(P_1(X)\):

\[
\|\sigma\|_{KR} = \sup \left\{ \int_X \varphi d\sigma; \varphi \in \cap L^1(d|\sigma|); \|\varphi\|_{Lip} \leq 1 \right\}
\]

and hence

\[
T_d(\mu, \nu) = \|\mu - \nu\|_{KR}
\]

for all probability measures \(\mu, \nu\) in \(P_1(X)\).

### 2.3 The time-dependent Problem

This section is concerned with the introduction of the time-dependent optimal transport problem. Monge’s optimal transport problem (2.1.4) is a distance problem, dependent on the initial and final data, i.e. it does not incorporate a kind of transport history. Although Monge mentioned the notion of trajectories between given data in his work (cf. [131]), the first time-dependent optimal transport problem was not formulated until 1999 by J.D. Benamou and Y. Brenier (cf. [17]). We discuss in the following the notion of geodesics and the property of convexity along these (cf. [126]). Note that this reformulation links the optimal transport theory and the gradient flow theory (see Chapter 4).

#### 2.3.1 Geodesics, Displacement Interpolant and Convexity

To reformulate (2.1.4) as a time-dependent optimal transport problem, we introduce a time interval \([0, 1]\) and consider a trajectory for each particle \(x\), i.e. \((T_t(x))_{0 \leq t \leq 1}\), which describes the path of a particle. The time-dependent minimisation problem reads

\[
\inf \left\{ \int_X C[(T_t(x))_{0 \leq t \leq 1}] d\mu(x); \ T_0 = \text{Id}, \ (T_1)_\#\mu = \nu \right\}, \tag{2.3.1}
\]

where the infimum is taken over all trajectories \((T_t(x))\). We assume in the sequel that the map \(t \mapsto T_t(x)\) is continuous and piecewise \(C^1\) for \(d\mu\)-almost all \(x\).

The connection to the non-time-dependent optimal transport formulation is the following: if an optimal trajectory \(T_t\) for (2.3.1) exists, then it follows that \(T = T_1\) is an optimal transport map for the Monge problem (2.1.4) and the transport costs of both problems coincide. Furthermore, we can reformulate \(C[(z_t)]\) by using the derivative formulation \(C(z_t) = \int_0^1 c(\dot{z}_t) dt\).
Proposition 2.3.1. (\cite{175} Proposition 5.2) If \(c\) is a convex function on \(\mathbb{R}^N\), then

\[
\inf \left\{ \int_0^1 c(\dot{z}_t) dt; \ z_0 = x, \ z_1 = y \right\} = c(y - x), \tag{2.3.2}
\]

\[
\inf \left\{ \int_0^T c(\dot{z}_t) dt; \ z_0 = x, \ z_1 = y \right\} = Tc \left( \frac{y - x}{T} \right). \tag{2.3.3}
\]

Moreover, if \(c\) is strictly convex, then the unique optimal geodesic is a straight line

\[z_t = x + t(y - x).\]

In the case of \(c(x, y) = |x - y|^2\) the optimal trajectory \(T_t(x)\) of the time dependent minimisation problem (2.3.1) coincides with the notion of a displacement interpolant (cf. \cite{126})

Definition 2.3.2. (Displacement interpolant) Let \(\mu, \nu \in \mathcal{P}_2(X)\) be two probability measures which do not give mass to small sets. With Theorem \[2.1.7\] we know that there exists a gradient of a convex function \(\varphi\), such that \(\nabla \varphi \# \mu = \nu\). Then the displacement interpolant between \(\mu\) and \(\nu\) for \(t \in [0, 1]\) reads

\[
\rho_t := [\mu, \nu]_t = [(1 - t)\text{Id} + t\nabla \varphi] \# \mu. \tag{2.3.4}
\]

with \(\rho_0 = [\mu, \nu]_0 = \mu, \ \rho_1 = [\mu, \nu]_1 = \nu\), and the family of probability measures \((\rho_t)_{0 \leq t \leq 1}\) interpolates between \(\mu\) and \(\nu\).

Since \([1 - t)\text{Id} + t\nabla \varphi = \nabla [\frac{1}{2}(1 - t)|\cdot|^2 + t\varphi]\) is the gradient of a convex function this yields

\[
W_2(\mu, \rho_t)^2 = \int_{\mathbb{R}^N} |x - ((1 - t)x + t\nabla \varphi)|^2 d\mu(x)
\]

\[
= \int_{\mathbb{R}^N} |tx - t\nabla \varphi|^2 d\mu(x) = tW_2(\mu, \nu)^2.
\]

Definition 2.3.3 (Constant speed geodesic). (\cite{9} Chapter 7.2) A curve \(\rho_t \in \mathcal{P}_\mu(X), t \in [0, 1]\) is a constant speed geodesic, if

\[
W_2(\rho_s, \rho_t)^2 = (t - s)W_2(\mu, \nu)^2 \quad \forall 0 \leq s \leq t \leq 1.
\]

For a strictly convex \(c\) the displacement interpolant can be written with a \(c\)-concave function \(\psi\):

\[
\rho_t := [\mu, \nu]_t = [\text{Id} - t\nabla c^*(\nabla \psi)] \# \mu. \tag{2.3.5}
\]

McCann introduced also the notion of displacement convexity, and therefore a method to study the convexity of a function along the geodesics (cf. \cite{126}). Here we introduce the notion in the space \(\mathcal{P}_2(\mathbb{R}^N)\):
Definition 2.3.4 (Displacement convexity). ([175, Definition 5.12]) Let $\sigma_t : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}^N$ be defined by $\sigma_t(x, y) = (1 - t)x + ty$.

(i) A subset $\mathcal{P}$ of $\mathcal{P}_2(\mathbb{R}^N)$ is said to be displacement convex if for all $\rho_0, \rho_1 \in \mathcal{P}$, for all $\pi$ optimal in the Kantorovich problem with probability measures $\rho_0, \rho_1$ and quadratic cost, and for all $t \in [0, 1]$, $\rho_t := \sigma_t \# \pi$ lies in $\mathcal{P}$.

(ii) A function $E$ defined on $\mathcal{P} \subset \mathcal{P}_2(\mathbb{R}^N)$ is said to be displacement convex with values in $\mathbb{R} \cup \{+\infty\}$ if for all $\rho_0 = \mu$ and $\rho_1 = \nu$

$$t \mapsto E(\rho_t)$$

is convex on $[0, 1]$.

Additionally we have the notions of strict displacement convexity, $\lambda$- uniformly displacement convexity and semi-displacement convexity:

Definition 2.3.5. ([175, Definition 5.14]) With the same notation as in Definition 2.3.4:

(i) A functional $E$ is said to be strictly displacement convex on $\mathcal{P}$ if for all $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^N)$ and $\mu \neq \nu$,

$$t \mapsto E(\rho_t)$$

is strictly convex on $[0, 1]$.

(ii) $E$ is said to be $\lambda$-uniformly displacement convex on $\mathcal{P}$ for some $\lambda > 0$ if for all $\mu, \nu \in \mathcal{P}$,

$$\frac{d^2}{dt^2} E(\rho_t) \geq \lambda W_2(\mu, \nu)^2$$

on $[0, 1]$.

(iii) $E : \mathcal{P} \rightarrow \mathbb{R} \cup \{+\infty\}$ is said to be semi-displacement convex on $\mathcal{P}$, with constant $C \geq 0$, if for all $\mu, \nu \in \mathcal{P}$,

$$\frac{d^2}{dt^2} E(\rho_t) \geq -CW_2(\mu, \nu)^2$$

on $[0, 1]$. 
Examples for displacement convex functionals

McCann introduced the notion of displacement convexity and investigated three main classes of displacement convex functionals

(i) Internal energy: \( U(\rho) = \int_{\mathbb{R}^N} U(\rho) \, dx \), \( U \) measurable \( \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\} \);

(ii) Potential energy: \( V(\rho) = \int_{\mathbb{R}^N} \rho V(\rho) \, dx \), \( V \) measurable \( \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\} \);

(iii) Interaction energy: \( G(\rho) = \int_{\mathbb{R}^N \times \mathbb{R}^N} G(x-y) \rho(x) \rho(y) \, dx \, dy \), \( G \) measurable \( \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\} \);

Under some appropriate conditions, which we will discuss in the following, these functionals are displacement convex.

**Theorem 2.3.6.** ([175] Theorem 5.15) Let \( \mathcal{P} \) be a displacement convex subset of \( \mathcal{P}_{ac}(\mathbb{R}^N) \) and the internal energy \( U(\rho) = \int_{\mathbb{R}^N} U(\rho(x)) \, dx \) is well defined with values in \( \mathbb{R} \cup \{+\infty\} \), i.e. let \( U : [0, +\infty) \to (-\infty, +\infty] \) be a proper lower semicontinuous convex function such that

\[
U(0) = 0, \quad \liminf_{s \to 0} \frac{U(r)}{r^\alpha} > -\infty \text{ for some } \alpha > \frac{N}{N+2}.
\]

Then if \( U \) satisfies

\[
\psi : r \mapsto r^n U(r^{-n}) \text{ is convex, nonincreasing on } (0, +\infty), \tag{2.3.6}
\]

then \( U \) is displacement convex on \( \mathcal{P} \).

The displacement convexity of functionals is important, since it guides the way to existence and uniqueness of a minimiser (cf. [126]):

**Theorem 2.3.7.** Let \( \rho \) be an absolute continuous probability measure on \( \mathbb{R}^N \). The energy functional

\[
\mathcal{F}(\rho) = \int_{\mathbb{R}^N} U(\rho) \, dx + \int_{\mathbb{R}^N} V \rho \, dx + \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} W(x-y) \rho(x) \rho(y) \, dx, \tag{2.3.7}
\]

has a unique minimiser \( \rho \in \mathcal{P}_{ac}(\mathbb{R}^N) \) if \( U \) satisfies the condition (2.3.6), \( \inf V > -\infty \) and additionally \( V \) and \( W \) are convex.

**Proof.** Let \( \rho_1, \rho_2 \) be two minimisers of \( \mathcal{F} \) with \( \rho = [\rho_1, \rho_2]_{\frac{1}{2}} \). We know that \( \mathcal{F} \) is strictly displacement convex and hence \( t \mapsto \mathcal{F}(\rho(t)) \) is strictly convex. This implies \( \mathcal{F}(\rho) < \frac{1}{2} (\mathcal{F}(\rho_0) + \mathcal{F}(\rho_1)) \). But this is a contradiction to \( \rho_0, \rho_1 \) being minimisers. \( \Box \)

A class of functionals among which we can find examples that are convex along geodesics is given by the internal energy
the entropy functionals $\mathcal{F}(\rho) = \int_{\mathbb{R}^N} \rho \log \rho \, dx$,

and $\mathcal{F}(\rho) = \int_{\mathbb{R}^N} \rho (\log \rho)^2 \, dx$

the $L^2$-norm $\mathcal{F}(\rho) = \int_{\mathbb{R}^N} \rho^2 \, dx$ for all positive dimensions $d$.

For other standard functionals involving derivatives of $u$, we either do not know if they are geodesically convex or we actually know that they are not. Let us consider first-order functionals of the form

$$\mathcal{F}_\alpha(\rho) := \frac{1}{2 \alpha} \int_{\mathbb{R}^d} |D \rho^\alpha|^2 \, dx.$$  

Among this family of functionals we have two prominent candidates, that is the classical Fisher information when $\alpha = 1/2$ and the Dirichlet energy for $\alpha = 1$. Even for these cases $E_\alpha$ is not geodesically convex, cf. [55]. For the total variation functional

$$\mathcal{F}(\rho) = \int_{\mathbb{R}^N} |\nabla \rho| \, dx$$

no results are known.

Furthermore we state the following result, which is used in Chapter 7:

**Theorem 2.3.8.** ([175, Theorem 5.15]) Let the interaction energy $\mathcal{G}$ be well defined on $\mathcal{P}(\mathbb{R}^N)$ with values in $\mathbb{R} \cup \{+\infty\}$, i.e. it is bounded below by some number. Then if $\mathcal{G}$ is convex (respectively semi-displacement convex with constant $C$), then $\mathcal{G}$ is displacement convex (respectively semi-displacement convex with constant $C$). If $\mathcal{G}$ is strictly convex (respectively $\lambda$-uniformly convex), then $\mathcal{G}$ is strictly displacement convex (respectively $\lambda$-uniformly-displacement convex) on the subspace $\mathcal{P}_m$ for all $m \in \mathbb{R}$. If $\mathcal{G}$ is displacement convex, then $\mathcal{G}$ is convex.

2.3.2 The Eulerian Formulation

There are two different ways to describe the particle flow in a fluid dynamic. The Lagrangian point of view follows an individual particle through time and space along the trajectory, whereas the Eulerian point of view considers the velocity field of the particles. We are able to switch between this two coordinates using the formula

$$v(t, g(t, x)) = \frac{d}{dt} g(t, x),$$

where $v$ denotes the time dependent velocity field which moves the particles around and $g(t, x)$ denotes the family of trajectories, $g(t, x) = T_t(x)$. When we consider the
Euclidean distance for the cost function $c$, the path is a diffeomorphism. The connection between $g$ and $T_t$ yields
\[
\frac{\partial}{\partial t} g(t, x) = \frac{\partial}{\partial t} T_t(x) = \frac{\partial}{\partial t}(x - t \nabla c^*(\nabla \psi)) = -\nabla c^*(\nabla \psi).
\]
Therefore
\[
v(0) = v(0, g(0, x)) = -\nabla c^*(\nabla \psi).
\]
Let $(T_t)_{0 \leq t \leq 1}$ be the solution of (2.3.1), then $\rho_t = (T_t)_#\mu$ denotes the intermediate probability measure at time $t$. A very important theorem is the following

**Theorem 2.3.9.** ([175, Theorem 5.34]) Let $X = \mathbb{R}^N$, or more generally a smooth complete manifold. $(T_t)_{0 \leq t \leq T^*}$ be the solution of (2.3.1) a locally Lipschitz family of diffeomorphism in $X$ with $T_0 = \text{Id}$ and $T_1 = T$. Let $v = v(t, x)$ be the velocity field associated with trajectories $T_t$. Let $\mu$ be a probability measure on $X$ and $\rho_t = (T_t)_#\mu$ the intermediate measure. Then $\rho_t = \rho(t, \cdot)$ is the unique solution of the linear transport equation (continuity equation)
\[
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0, \quad \text{for } 0 < t < T^*, \text{ with } \rho_0 = (T_0)_#\mu = \mu. \tag{2.3.8}
\]
in $C([0, T]; \mathcal{P}(X))$, where $\mathcal{P}(X)$ is equipped with the weak topology.

**Proof.** The idea of the proof is to show, that for all $\varphi \in C_c^\infty$ the mapping $t \mapsto \int_{\mathbb{R}^N} \varphi d\rho_t$ is Lipschitz on $(0, T)$, with derivative
\[
\frac{d}{dt} \int_{\mathbb{R}^N} \varphi d\rho_t = \int_{\mathbb{R}^N} (\nabla \varphi v_t) d\rho_t.
\]
With $\rho_t = (T_t)_#\mu$ we can write
\[
\int_{\mathbb{R}^N} \varphi d\rho_t = \int_{\mathbb{R}^N} (\varphi \circ T_t) d\mu.
\]
Further, we know, that $\varphi$ is compactly supported and $T_t^{-1}$ is continuous, then the function $\varphi \circ T_t$ is supported on a compact set and additionally it is Lipschitz with derivative
\[
\frac{\partial}{\partial t} (\varphi \circ T_t) = (\nabla \varphi \circ T_t) \cdot \frac{\partial T_t}{\partial t} = (\nabla \varphi \circ T_t) \cdot (v_t \circ T_t).
\]
With $h > 0$ we can write
\[
\frac{1}{h} \left( \int_{\mathbb{R}^N} \varphi d\rho_{t+h} - \int_{\mathbb{R}^N} \varphi d\rho_t \right) = \int_{\mathbb{R}^N} \left( \rho \circ T_{t+h} - \rho \circ T_t \right) d\mu.
\]
The right term is uniformly bounded on \([0, T - h] \times \mathbb{R}^N\) and it converges to \((\nabla \varphi \circ T_t)v_t\) as \(h \to 0\). With the Lebesgue dominated convergence theorem we deduce that the map \(t \mapsto \int \varphi d\rho_t\) is differentiable for almost all \(t\), and
\[
\frac{d}{dt} \int \nabla \varphi \circ T_t \cdot (v_t \circ T_t) d\mu = \int_{\mathbb{R}^N} \nabla \varphi \cdot v d\rho_t.
\]
We proved that \(\rho_t\) solves the continuity equation. For the uniqueness proof we refer to [175].

**Proposition 2.3.10.** ([175], Proposition 5.38]) Let \(v_0 : \mathbb{R}^N \to \mathbb{R}^N\) be a continuous function on \(\mathbb{R}^N\), differentiable everywhere and \(T_t(x) = x - tv_0\) be a family of trajectories. Each particle moves with constant velocity. Then the velocity field \(v_t = T_t^{-1} \circ \frac{d}{dt} T_t\) satisfies the equation
\[
\frac{\partial v}{\partial t} + v \cdot \nabla v = 0.
\]

The Eulerian system of optimal time-dependent mass transportation problem then looks like
\[
\begin{align*}
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) &= 0, \quad 0 < t < T^*, \quad \rho_0 = (T_0)_\# \mu = \mu, \quad \text{(2.3.9)} \\
\frac{\partial v}{\partial t} + v \cdot \nabla v &= 0. \quad \text{(2.3.10)}
\end{align*}
\]
Chapter 3

The Problem Formulation

In the following chapter we introduce the variational problem and give a motivation with Bayesians MAP estimation (Section 3.1). Furthermore, we derive in Section 3.2 an equivalent formulation, which is inspired by the work of J.D. Benamou and Y. Brenier [17]. We study the Lagrange functional and optimality conditions.

3.1 The Variational Problem

Let $X = \Omega$ be an open and bounded subset of $\mathbb{R}^N$ and $\nu$ a probability measure on $\Omega$. Furthermore, let $u$ be a probability density, i.e. $u \geq 0$ and $\int_{\Omega} u \, dx = 1$, $\mathcal{L}^N$ is the usual Lebesgue measure. In this thesis we want to investigate the solution of the variational problem

$$\frac{1}{2} W_2(\nu, u \mathcal{L}^N)^2 + \epsilon E(u) \rightarrow \min_u . \quad (3.1.1)$$

The functional consists of a fidelity term, which is the quadratic Wasserstein distance $W_2(\nu, u \mathcal{L}^N)$ between $\nu$ and $u \mathcal{L}^N$ (2.2.1), and of a regularisation term $E(u)$ multiplied by a parameter $\epsilon > 0$. The regularisation parameter $\epsilon$ steers the impact of the regularisation, for which we consider e.g. the logarithmic entropy and the $L^2$-regularisation of $u$, as well as functionals which involves derivatives of $u$, e.g. the Dirichlet regularisation, the Fisher information and the TV-regularisation. A particular advantage of the Wasserstein metric is the possibility to obtain a unified distance functional for probability densities continuous with respect to the Lebesgue measure as well as concentrated densities, since the Wasserstein metric is defined for arbitrary probability measures with finite second moment. Moreover, using the Wasserstein distance as a fidelity term automatically gives us a mass conservation constraint on the solution and is as such density preserving.

37
In the sequel we consider the quadratic Euclidean distance $d(x, y)^2 = |x - y|^2$ for $d$ in the Wasserstein distance \[2.2.1\]. Then the classical representation of the Wasserstein distance is the Kantorovich problem \[2.1.10\], in particular

\[ I(u) = \frac{1}{2} \int_{\Omega \times \Omega} |x - y|^2 d\pi(x, y) + \epsilon E(u) \to \min_{\pi, u}, \]  

subject to

\[ \int_{A \times \Omega} d\pi(x, y) = \int_A d\nu(y), \]  

\[ \int_{\Omega \times A} d\pi(x, y) = \int_A u(x) dx, \]  

for all $A \subset \Omega$ measurable, where \[3.1.3\] and \[3.1.4\] are the marginal conditions for $\nu$ and $u \mathcal{L}^N$, and $\pi$ is a probability measure on $\Omega \times \Omega$. Similar to the Kantorovich’s dual theory (see \[2.1.13\]) we state here the dual formulation for \[3.1.2\]-\[3.1.4\]

\[ \inf_u \sup_{\phi, \psi \in \Phi} J(u, \phi, \psi) = \epsilon E(u) + \int_{\Omega} \phi(x)u(x) dx + \int_{\Omega} \psi(y) d\nu(y), \]  

subject to \[ \phi(x) + \psi(y) \leq \frac{1}{2}|x - y|^2. \]  

For fixed $u$ the maximisation of the dual variables $\phi$ and $\psi$ results in the optimality conditions

\[ u(x) dx = \int_{\Omega \times \Omega} d\pi(x, y), \]  

\[ d\nu(x) dx = \int_{\Omega \times A} d\pi(y, x) \]  

and the complementarity condition

\[ \lambda \geq 0, \quad (\phi(x) + \psi(y) - \frac{1}{2}|x - y|^2) d\pi(x, y) = 0, \]  

where actually $\pi$ is the solution in the primal Kantorovich formulation.

Note that the optimality condition for $u$ implies

\[ \epsilon \langle E'(u), v - u \rangle \leq - \int_{\Omega} \phi(v - u) dx \]  

for all admissible $v$. With this preliminaries we can state the following uniqueness result:

**Theorem 3.1.1.** Let $E$ be differentiable and strictly convex and $\epsilon > 0$. Then there exists at most one minimiser of \[3.1.1\].
Proof. Assume that \( u_1 \) and \( u_2 \) are minimisers of \( (3.1.1) \) with associated product measures \( \pi \) and dual variables \( \varphi_i, \psi_i, i = 1, 2 \). Then we conclude from \( (3.1.10) \)

\[
\epsilon \langle E'(u_1) - E'(u_2), u_1 - u_2 \rangle \leq -\int_{\Omega} (\varphi_1 - \varphi_2)(u_1 - u_2) \, dx.
\]

Inserting the property of \( u_i \) being a marginal density of \( \pi_i \) we further have

\[
\epsilon \langle E'(u_1) - E'(u_2), u_1 - u_2 \rangle \leq -\int_{\Omega \times \Omega} (\varphi_1(x) - \varphi_2(x)) \, d(\pi_1 - \pi_2)(x,y).
\]

The last term can be manipulated using the complementarity conditions as well as the inequality constraint to obtain

\[
-\int_{\Omega \times \Omega} (\varphi_1(x) - \varphi_2(x)) \, d\pi(x,y) = -\int_{\Omega \times \Omega} \varphi_1(x) \, d\pi_1(x,y) - \int_{\Omega \times \Omega} \varphi_2(x) \, d\pi_2(x,y)
+ \int_{\Omega \times \Omega} \varphi_1(x) \, d\pi_2(x,y) + \int_{\Omega \times \Omega} \varphi_2(x) \, d\pi_1(x,y)
\leq \int_{\Omega} (\psi_1(y) - \psi_2(y)) \, d\nu(y) - \int_{\Omega} (\psi_1(y) - \psi_2(y)) \, d\nu(y) = 0.
\]

Thus,

\[
\langle E'(u_1) - E'(u_2), u_1 - u_2 \rangle \leq 0
\]

and the strict convexity of \( E \) implies \( u_1 = u_2 \). \( \square \)

3.1.1 Motivation with Bayesian MAP-Estimation

In the following, we want to verify that the variational problem \( (3.1.1) \), in particular the discrete variational problem, can be well motivated from the Bayesian maximum a-posteriori probability estimation (MAP-estimation). Furthermore, the a-priori information about the wanted quantity is incorporated as a regularisation term in the minimisation problem.

Let \( F \) be a measurement or observation which follows a Gaussian error model with mean \( x \) and variance \( \sigma^2 \). Let \( f \) be the unknown sampling distribution of \( F \) and \( \Theta = (x, \sigma^2) \), then \( f(F|\Theta) \) is the probability density of \( F \) under parameter \( \Theta \). The likelihood function and the maximum likelihood function are defined as follows:

\[
\Theta \mapsto f(F|\Theta), \quad \text{(Likelihood Function),} \quad (3.1.11)
\]

\[
\hat{\Theta}_{ML} = \arg\max_{\Theta} f(F|\Theta), \quad \text{(Maximum Likelihood Estimate).} \quad (3.1.12)
\]
Then, we can compute an estimate for $\Theta$, which is close to the true value, by maximising the a-posteriori probability density $f(\Theta|F)$, or equivalent by minimising the negative logarithmic of $f(\Theta|F)$, since the logarithmic function is strictly increasing:

$$\bar{\Theta}_{\text{MAP}} = \arg\max_{\Theta} f(\Theta|F) \quad \text{(MAP Estimate)} \quad (3.1.13)$$

$$= \arg\min_{\Theta} (-\log f(\Theta|F)). \quad (3.1.14)$$

With the formula of Bayes we follow

$$f(\Theta|F) = \frac{f(F|\Theta)g(\Theta)}{\int_{\Theta'\in\Phi} f(F|\Theta')g(\Theta')d\Theta'}, \quad (3.1.15)$$

where $g(\Theta)$ denotes the a-priori probability for $\Theta$ and $f(F) = \int_{\Theta'\in\Phi} f(F|\Theta')g(\Theta')d\Theta'$ the prior probability for $F$. Since the denominator does not depend on $\Theta$ it has no influence on the result in the minimisation. We eventually obtain for (3.1.14)

$$\bar{\Theta}_{\text{MAP}} = \arg\min_{\Theta} (-\log(f(F|\Theta)g(\Theta))) \quad (3.1.16)$$

$$= \arg\min_{\Theta} (-\log(f(F|\Theta)) - \log(g(\Theta))). \quad (3.1.17)$$

Now we extend this model to $M$ independent identically distributed measurements $F_1, ..., F_M$. Each measurement follows a Gaussian error model with mean $x_i$ and variance $\sigma^2$, for $i = 1, ..., M$. Then the joint distribution function reads

$$f(F|\Theta) = f(F_1|(x_1, \sigma^2))f(F_2|(x_2, \sigma^2))...f(F_M|(x_M, \sigma^2)). \quad (3.1.18)$$

Each distribution function can be identified with the probability density of the Gaussian noise

$$f(F_i|(x_i, \sigma^2)) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(F_i - x_i)^2\right), \quad (3.1.19)$$

$$\Rightarrow f(F|\Theta) = \prod_{i=1}^{M} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(F_i - x_i)^2\right). \quad (3.1.20)$$

The prior informations $g(x, \sigma^2)$ are mostly modelled with Gibb’s functions, i.e.

$$g(x, \sigma^2) = \exp(-\epsilon E(x)). \quad (3.1.21)$$

Then, with (3.1.17) we conclude

$$\bar{\Theta}_{\text{MAP}} = \arg\min_{\Theta} \left(-\log \left(\prod_{i=1}^{M} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(F_i - x_i)^2\right)\right) - \log(g(\Theta))\right), \quad (3.1.22)$$

$$= \arg\min_{\Theta} \left(\frac{1}{2\sigma^2} \sum_{i=1}^{M} (F_i - x_i)^2 + \epsilon E(x_1, ..., x_M)\right). \quad (3.1.23)$$
Note that the constants can be neglected. The squared distance can be translated into a multiple of the Wasserstein metric between the empirical measures

$$\nu = \frac{1}{M} \sum_{i=1}^{M} \delta_{F_i} \quad \text{and} \quad u = \frac{1}{M} \sum_{i=1}^{M} \delta_{x_i}, \quad (3.1.24)$$

and the second term coincides with the regularisation term in the variational model. Thus we recover a discrete version of the variational model (3.1.1) and give a motivation to the non-parametric regression estimation.

### 3.1.2 Stability

In the following, we discuss a stability estimate for solutions with different input measures, since a stable variational approach should yield reconstructed densities $u$ close in a stronger error measure if the measures $\nu$ are close in the Wasserstein metric. It has been demonstrated by many authors (cf. [13, 44, 147, 19]) that Bregman distances related to the regularisation functional $E$ are natural error measures for variational regularisation methods with $E$ convex. In particular Pöschl [147] has derived estimates for variational regularisation methods with powers of metrics, which can be translated to the case with the Wasserstein metric we consider here. However, since for our special problem we can derive estimates with explicit constants, we provide a different proof here.

Our setup is as follows: Let $\nu_1$ and $\nu_2$ are two given probability measures with corresponding minimisers $u_1$ and $u_2$ of (3.1.1). Let further be $\pi_i$, $i = 1, 2$, the associated product measures and $\varphi_i$ and $\psi_i$, $i = 1, 2$, the variables in the general dual formulation of the Kantorovich formulation.

**Theorem 3.1.2.** Let $E$ be differentiable and strictly convex and $\epsilon > 0$, then the symmetric Bregman distance is bounded by

$$D_E(u_1, u_2) = \langle E'(u_1) - E'(u_2), u_1 - u_2 \rangle \leq \frac{1}{\epsilon} W_2(\nu_1, \nu_2).$$

**Proof.** Note that the optimality condition for the dual Kantorovich problem (3.1.5)-(3.1.6) with respect to $u$ implies

$$\epsilon \langle E'(u), v - u \rangle \leq - \int_{\Omega} \varphi(v - u) \, dx \quad (3.1.25)$$

for all admissible $v$. We use inequality (3.1.25) to conclude

$$\epsilon \langle E'(u_1) - E'(u_2), u_1 - u_2 \rangle \leq - \int_{\Omega} (\varphi_1 - \varphi_2)(u_1 - u_2) \, dx.$$
Now, we recall the formulation of the dual problem (3.1.5)-(3.1.6) in terms of c–concave functions (see Remark 2.1.6), which are the function and its Legendre transform in the current case of the quadratic Wasserstein distance. In fact, the supremum in (3.1.5) can be restricted to those admissible pairs \((\varphi_{cc}, \varphi_c)\), where

\[
\varphi_c(y) = \inf_{x \in \Omega} \left\{ \frac{1}{2} |x - y|^2 - \varphi(x) \right\}, \quad \varphi_{cc}(x) = \inf_{y \in \Omega} \left\{ \frac{1}{2} |x - y|^2 - \varphi_c(y) \right\},
\]

cf. [175, Proof of Theorem 4 in Chapter 1]. Substituting \(\varphi_{cc}^i\) for \(\varphi_i\) and \(\varphi_c^i\) for \(\psi_i\), \(i = 1, 2\), we obtain

\[
\epsilon \langle E'(u_1) - E'(u_2), u_1 - u_2 \rangle \leq - \int_{\Omega} (\varphi_{cc}^1 - \varphi_{cc}^2)(u_1 - u_2) \, dx
\]
\[
\leq \int_{\Omega} (\varphi_1^c(y) - \varphi_2^c(y)) \, d(\nu_1 - \nu_2)(y).
\]

The functions \(\varphi_i^c, i = 1, 2\) satisfy

\[
\varphi_i^c(y) - \varphi_i^c(y') = \sup_x \left\{ \varphi_i(x) - \frac{1}{2} |x - y'|^2 \right\} + \inf_x \left\{ \frac{1}{2} |x - y|^2 - \varphi_i(x) \right\}
\]
\[
\leq \sup_x \left\{ \varphi_i(x) - \frac{1}{2} |x - y'|^2 \right\} + \sup_x \left\{ \frac{1}{2} |x - y|^2 - \varphi_i(x) \right\}
\]
\[
= \frac{1}{2} \sup_x \{|x - y|^2 - |x - y'|^2\} \leq \frac{1}{2} |y - y'|^2,
\]

and hence are 1–Lipschitz continuous. Implementing this into our inequality we have

\[
\epsilon \langle E'(u_1) - E'(u_2), u_1 - u_2 \rangle \leq \int_{\Omega} (\varphi_1^c(y) - \varphi_2^c(y)) \, d(\nu_1 - \nu_2)(y)
\]
\[
\leq \sup_{\varphi} \left\{ \int_{\Omega} \varphi(y) \, d(\nu_1 - \nu_2)(y); \varphi \in \bigcap L^1(d|\nu_1 - \nu_2|), \|\varphi\|_{Lip} \leq 1 \right\}. \tag{3.1.26}
\]

Here we use the Kantorovich-Rubinstein duality theorem for the case when the cost function \(c(x, y)\) in the Kantorovich formulation is a metric \(d(x, y)\). In the case of the proposed optimisation problem (3.1.1) we have \(d(x, y) = |x - y|^2\). Hence, the last expression in (3.1.26) is equal to the cost of the optimal transport between \(\nu_1\) and \(\nu_2\) and we eventually get

\[
\epsilon \langle E'(u_1) - E'(u_2), u_1 - u_2 \rangle \leq W_2(\nu_1, \nu_2).
\]

We finally mention that in the case of \(E\) being convex but not differentiable (such as
for the total variation) a completely analogous proof holds to obtain an estimate for the generalised Bregman distance

$$D^{p_1, p_2}_E(u_1, u_2) := \langle p_1 - p_2, u_1 - u_2 \rangle \quad p_i \in \partial E(u_i),$$

(3.1.27)

where the $p_i$ are the subgradients arising in the respective optimality conditions. In the case of total variation, the Bregman distance is not a strict distance anymore, it can vanish also for $u_1 \neq u_2$. However, it provides natural information about discontinuity sets (edges in the image setting, cf. [43, 44, 138]).

### 3.1.3 Random Sampling

In the following, we analyse the consistency of the estimate obtained from our variational approach in the classical case of available samples from a given probability measure $\nu$. For this sake we introduce the empirical measure $\nu^M$, which is a random variable depending on the samples $F_1, \ldots, F_M$:

$$\nu^M = \frac{1}{M} \sum_{j=1}^M \epsilon_{F_j},$$

(3.1.28)

where $\epsilon_F$ is the point measure at $F$.

We denote by $u^{\epsilon, M}$ a minimiser of the variational problem with given measure $\nu^M$. Note that also $u^M$ is a random variable depending on the samples and thus also subsequent estimates are to be interpreted for any realisation of the random variable. Assuming that $\nu$ is regular enough, i.e. there exists a density $\hat{u}$ such that

$$\nu = \hat{u} \mathcal{L}^M, \quad E(\hat{u}) < \infty,$$

(3.1.29)

we immediately conclude from the definition of $u^{\epsilon, M}$

$$W_2(u^{\epsilon, M}, \nu^M)^2 \leq W_2(u^{\epsilon, M}, \nu^M)^2 + 2\epsilon E(u^{\epsilon, M})$$

$$\leq W_2(\nu, \nu^M)^2 + 2\epsilon E(\hat{u}),$$

thus by the triangle inequality and Minkowski’s inequality we obtain

$$W_2(u^{\epsilon, M}, \nu) \leq W_2(u^{\epsilon, M}, \nu^M) + W_2(\nu, \nu^M) \leq 2W_2(\nu, \nu^M) + \sqrt{2\epsilon E(\hat{u})}.$$ 

In particular, we obtain for the expected value

$$E(W_2(u^{\epsilon, M} \mathcal{L}^M, \nu)) \leq 2E(W_2(\nu, \nu^M)) + \sqrt{2\epsilon E(\hat{u})}.$$

(3.1.30)

The convergence of the first term on the right-hand side is a well-known problem (clearly related to the law of large numbers), even rates are available, e.g. Rachev
gives a rate of order $M^{-1/N}$ for distributions with uniformly bounded moments, while Horowitz et al. [94] even provide a rate of order $M^{-2/(N+4)}$ if the $(N+5)$-th moment is finite. The second term is obviously converging to zero as $\epsilon \to 0$, we thus obtain:

**Theorem 3.1.3.** Let $\nu$ be such that (3.1.29) is satisfied, then the variational model (3.1.1) is consistent in the sense of narrow convergence, which is metrised by the Wasserstein metric, more precisely

$$E(W_2(u^{\epsilon,M}L^M, \nu)) \to 0$$

(3.1.31)
as $M \to \infty$ and $\epsilon \to 0$.

Note that from the arguments above we even obtain a rate for the Wasserstein metric in terms of $\epsilon$ and $M$ under additional moment conditions, respectively a rate in terms of $N$ if $\epsilon$ is chosen in dependence of $M$, which is the standard case in practice. Since we are measuring fidelities in the Wasserstein metric, this is not surprising and one would rather expect a stronger convergence or, under additional smoothness assumptions on $\hat{u}$, convergence rates in a stronger error measure related to the regularisation functional $E$. The Bregman distance, more precisely $E(D_E(u^{\epsilon,M}, \hat{u}))$ appears to be a correct error measure. Note that from Theorem 3.1.2 we directly obtain an estimate for the reconstructed densities

$$E(D_E(u^{\epsilon,M}, u)) \leq \frac{1}{\epsilon} E(W_2(\nu^M, \nu)^2).$$

(3.1.32)
The remaining effort is basically to perform an estimate for the error $D_E(u, \hat{u})$, which is purely deterministic and already studied under additional smoothness conditions (so-called source conditions in regularisation theory) in [147]. In cases where some power of the Bregman distance satisfies a triangle inequality these estimates can be combined directly (e.g. for $L^2$ and $H^1$ regularisations the Bregman distance is just the square of the norm), in other cases we can directly combine the techniques from Theorem 3.1.2 and [147] to obtain an estimate of the form

$$E(D_E(u^{\epsilon,M}, \hat{u})) \leq \frac{1}{\epsilon} F \left( \frac{1}{M} \right) + G(\epsilon)$$

(3.1.33)
with decreasing functions $F$ and $G$ such that $F(0, 0) = G(0) = 0$. The exact form of $F$ and $G$ depends on the specific assumptions used and also on the chosen regularisation functional.

### 3.2 The Fluid Dynamic Formulation

J.D. Benamou and Y. Brenier introduced in 2000 a reformulation of the original Kantorovich problem (2.1.10), which sets the optimal transport problem in a fluid mechanic
framework (see [17]).
To be compatible with the notations in fluid mechanics, we denote here the initial
density with \( \rho_0 \) and the final density with \( \rho_T \). Moreover we consider densities with
normalised mass to one
\[
m(\Omega) = \int_{\Omega} \rho_0 \, dx = \int_{\Omega} \rho_T \, dx = 1.
\] (3.2.1)
Hence, the optimal transport problem can be reformulated by introducing a time inter-
val \([0, T] \), a time-dependent density field \( \rho(t, x) \geq 0 \) and a velocity field \( v(t, x) \in \mathbb{R}^N \),
which fulfill:
\[
\partial_t \rho + \nabla \cdot (\rho v) = 0, \quad \forall \ t \in (0, T), \quad \text{(continuity equation)} \quad (3.2.2)
\]
\[
\rho(0, \cdot) = \rho_0, \quad \rho(T, \cdot) = \rho_T, \quad \text{(initial conditions).} \quad (3.2.3)
\]
with \( \rho \in C([0, T], \mathcal{P}_{ac}(\Omega)) \), and \( \mathcal{P}_{ac}(\Omega) \) is endowed with the weak topology and \( v \in L^2(d\rho(t, x)dt) \). Additionally \( \cup_{0 \leq t \leq T} \text{Supp}(\rho(t, x)) \) is bounded. Benamou and Brenier
stated the following equivalence:

**Proposition 3.2.1.** The quadratic Wasserstein distance is equal to the infimum of
\[
A(\rho, v) = T \int_{\Omega} \int_{0}^{T} \rho(t, x) |v(t, x)|^2 \, dx \, dt,
\] (3.2.4)
where the infimum is taken over all \((\rho, v)\) satisfying (3.2.2)-(3.2.3).

**Proof.** We just want to give an idea of the proof and refer to [17, 175] for more details.
Let \( \rho \) and \( v \) be smooth, bounded and of class \( C^1 \).
First we proof that the lower bound holds
\[
\inf \{ A(\rho, v); \, (\rho, v) \in V(\rho_0, \rho_1) \} \geq W_2(\rho_0, \rho_1)^2,
\] (3.2.5)
and \( V(\rho_0, \rho_1) \) is the set of all \((\rho, v) = (\rho_t, v_t)_{0 \leq t \leq T} \) satisfying (3.2.2)-(3.2.3). If \( \rho \) is
continuous, it can be identified with its Lebesgue measure \( d\rho(x) = \rho(x)dx \) and it follows
\[
W_2(\rho_0, \rho_1)^2 = \inf \left\{ \int_{\Omega} \rho_0 |x - T(x)|^2 \, dx; \, T_{\#} \rho_0 = \rho_1 \right\},
\]
\[
= \inf_{\Omega} \left\{ \int_{\Omega} \rho_0 |x - T(x)|^2 \, dx; \, T_{\#} \rho_0 = \rho_1 \right\}.
\]
Since we know that \((\rho, v)\) is smooth, we can consider the associated trajectory \( T_t(x) \)
with \( \frac{d}{dt} T_t(x) = v(t, T_t(x)) \), \( T_0(x) = x \) and the displacement interpolant \( T_t(x)_{\#} \rho_0 = \rho_t \).
The family of trajectories \((T_t(x))_{0 \leq t \leq T}\) is locally Lipschitz and a diffeomorphism on \( \mathbb{R}^N \).
With Theorem 2.3.9 we can conclude that \( \rho_t \) is a solution of the continuity equation in the weak sense, and this yields
\[
\int_\Omega \rho_t(x)|v(t, x)|^2dx = \int_\Omega \rho_0(x)|v(t, T_t(x))|^2dx \\
= \int_\Omega \rho_0(x)\frac{d}{dt}T_t(x)^2dx.
\]

With integration over \([0, T]\) and Jensen’s inequality it follows
\[
A(\rho, v) \geq \int_\Omega \int_0^T \rho_0(x)|\frac{d}{dt}T_t(x)|^2dxdt = \int_\Omega \rho_0(x) \left( \int_0^T |\frac{d}{dt}T_t(x)|^2dt \right) dx \\
= \int_\Omega \rho_0(x)|T_1(x) - T_0(x)|^2dx \\
= \int_\Omega \rho_0(x)|T_1(x) - x|^2dx \\
= \int_\Omega \rho_0(x)|T(x) - x|^2dx \\
\geq W_2(\rho_0, \rho_1)^2.
\]

In the second step we prove the existence of \((\rho, v) \in V(\rho_0, \rho_1)\) such that \(A(\rho, v) = W_2(\rho_0, \rho_1)^2\). Let \(T = \nabla \psi\) be the optimal map with \(\psi\) convex and
\[
T_t(x) = (1 - x)t + tT(x) = \nabla \psi_t(x),
\]
and displacement interpolant \(\rho_t = (T_t)_#\rho_0\). let \(\nabla \psi_t^*\) be the inverse of \(\nabla \psi\). Then we obtain with
\[
\frac{d}{dt}T_t(x) = v(t, T_t(x)),
\]
the following result for the velocity:
\[
v(t, x) = \frac{d}{dt}T_t(x) \circ T_t^{-1}.
\]

\(v_t\) is bounded, in view of our assumptions on \(\rho_0\) and \(\rho_1\). We can prove that \((\rho_t, v_t)\) are weak solutions to the continuity equation. With a nonnegative measurable function \(\Phi\) we can follow
\[
\int_\Omega \rho(t, x)\Phi(v(t, x))dx = \int_\Omega \rho_0(x)\Phi(v(t, T_t(x)))dx \\
= \int_\Omega \rho_0(x)\Phi(\frac{d}{dt}T_t(x))dx \\
= \int_\Omega \rho_0(x)\Phi(T(x) - x)dx.
\]
With $\Phi(z) = |z|^2$ and with integration over the time we obtain

$$A(\rho, v) = W_2(\rho_0, \rho_1)^2.$$  \hfill (3.2.7)

In Physics (3.2.4) is known as the action functional, respectively the time integral of the kinetic energy. It describes the total effort which is necessary to move particles around according to the velocity field $v(t, x)$. Due to the investigation of an artificial time variable $t$, $\rho(t, x)$ can be interpreted as an interpolant between $\rho_0$ and $\rho_T$. Besides, we obtain a velocity field $v(t, x)$ which gives us information about the velocity of each particle at time $t \in (0, T)$. Moreover, by replacing $j = \rho v$, the Benamou-Brenier problem becomes a convex space-time-minimisation problem with respect to $(\rho, j)$ on $[0, T] \times \Omega$:

$$\inf_{\rho,j} \int_0^T \int_{\Omega} \frac{|j(t,x)|^2}{\rho(t,x)} dx \, dt,$$  \hfill (3.2.8)

subject to

$$\partial_t \rho + \nabla \cdot j = 0,$$

$$\rho(0, \cdot) = \rho_0, \quad \rho(T, \cdot) = \rho_T.$$  \hfill (3.2.9), (3.2.10)

### 3.2.1 Fluid dynamic Formulation with Regularisation

The time dependent optimal transport formulation given by Benamou and Brenier inspires our model. Let $[0, 1]$ be a time interval, $\mu(t, x)$ a probability measure with $t \in [0, 1]$ and $v(t, x)$ the velocity field. With given initial data $\nu \in \mathcal{P}(\Omega)$, we search for an optimal $\mu, v$ and a final probability density $u$, i.e.

$$u \geq 0, \quad \int_{\Omega} u dx = 1,$$

which solves the following problem:

**Minimise** $\frac{1}{2} \int_0^1 \int_{\Omega} |v|^2 d\mu(x) \, dt + \epsilon E(u)$, \hfill (3.2.11)

**subject to**

$$\partial_t \mu + \nabla \cdot (\mu v) = 0, \quad \text{(continuity equation)}$$

$$\mu(t = 0) = \nu, \quad \text{(initial condition)}$$

$$\mu(t = 1) = u\mathcal{L}^N, \quad \text{(final condition)}$$ \hfill (3.2.12), (3.2.13), (3.2.14)
where we minimise over \((u\mathcal{L}^N, \mu, v)\). We obtain (3.2.11)-(3.2.14) from the original formulation (3.2.4) by adding a regularisation energy \(E(u)\) to the action functional, which is multiplied by a parameter \(\epsilon > 0\). In contrast to the Benamou-Brenier formulation (3.2.4), where the Wasserstein distance is computed for two given probability measures, we minimise here over the density \(u\), which is the density at time \(t = 1\) (see (3.2.14)). Note that if \(\mu\) is absolutely continuous, it can be identified with its density \(d\mu(t,x) = \rho(t,x)dx\). This is not always the case, since we allow \(\nu\) to be also a point measure.

In the following we investigate the Lagrange functional and the optimality conditions for (3.2.11)-(3.2.14), which we exploit for the numerical solution.

### 3.2.2 Optimality Conditions

Writing down the Lagrangian for the minimisation problem (3.2.11) with constraint (3.2.12) and boundary conditions (3.2.13)-(3.2.14) yields

\[
L(u, \mu, v, \lambda, \eta) = \epsilon E(u) + \frac{1}{2} \int_0^1 \int_\Omega |v|^2 \, d\mu \, dt + \int_0^1 \int_\Omega (-\partial_t \lambda - v \cdot \nabla \lambda) \, d\mu \, dt \\
- \int_\Omega \lambda (t = 1) \, d\mu (t = 1) + \int_\Omega u \lambda (t = 1) \, dx - \int_\Omega \lambda (t = 0) \, d\mu (t = 0) \\
+ \int_\Omega \nu \lambda (t = 0) \, dx + \int_\Omega \eta u \, dx + \eta_0 \left( \int_\Omega u \, dx - 1 \right),
\]

with complementarity condition \(\eta u = 0\) and primal variables \(u, \mu, v\) and dual variables \(\lambda\) and \(\eta\). We compute the Karush-Kuhn-Tucker conditions, which are the necessary conditions for a minimum:

\[
\frac{\partial L}{\partial \lambda} = 0 : \quad \partial_t \mu + \nabla \cdot (\mu v) = 0, \quad (3.2.15)
\]

\[
\frac{\partial L}{\partial \mu} = 0 : \quad \partial_t \lambda + v \cdot \nabla \lambda - \frac{1}{2} |v|^2 = 0, \quad (3.2.16)
\]

\[
\frac{\partial L}{\partial v} = 0 : \quad \mu v - \mu \nabla \lambda = 0, \quad (3.2.17)
\]

in \((0, 1) \times \Omega\) with

\[
\frac{\partial L}{\partial u} = 0 : \quad \epsilon \partial E(u) + \lambda(t = 1) \geq 0, \quad (3.2.18)
\]

\[
\frac{\partial L}{\partial \lambda} = 0 : \quad \mu(t = 1) = u\mathcal{L}^N, \quad (3.2.19)
\]

\[
\frac{\partial L}{\partial \nu} = 0 : \quad \mu(t = 0) = \nu. \quad (3.2.20)
\]

For the numerical solution we shall in particular exploit the Benamou-Brenier optimality conditions (3.2.15)-(3.2.20). We also consider a dual method based on augmenting
the constraint at $t = 0$. Replacing $v = \nabla \lambda$ leads to the Lagrangian

$$L(u, \mu, \lambda, \eta) = -\frac{1}{2} \int_0^1 \int_{\Omega} |\nabla \lambda|^2 \, d\mu(x) \, dt + \epsilon E(u) - \int_0^1 \int_{\Omega} \partial_t \lambda \, d\mu(x) \, dt$$

$$+ \int_{\Omega} \nu \lambda(t = 0) \, dx - \int_{\Omega} \lambda(t = 0) \, d\mu(t = 0) - \int_{\Omega} \lambda(t = 1) \, d\mu(t = 1)$$

$$+ \int_{\Omega} u \lambda(t = 1) \, dx + \int_{\Omega} \eta u \, dx + \eta_0 \left( \int_{\Omega} u \, dx - 1 \right).$$

Again, the variations of the Lagrange functional with respect to all variables are

\[
\frac{\partial L}{\partial \lambda} = 0 : \quad \partial_t \mu + \nabla \cdot (\mu \nabla \lambda) = 0, \quad (3.2.21) \\
\frac{\partial L}{\partial \mu} = 0 : \quad \partial_t \lambda + \frac{1}{2} |\nabla \lambda|^2 = 0, \quad (3.2.22)
\]

in $(0, 1) \times \Omega$ with

$$\lambda(t = 0) = \lambda_0,$$

$$\mu(t = 0) = \nu,$$  

(3.2.23)  

(3.2.24)

where $\epsilon E'(u) + \lambda(t = 1) \ni 0$ is the optimality condition for the optimisation problem

$$\epsilon E(u) + \int_{\Omega} u \lambda(t = 1) \, dx \to \min_{u \in \mathcal{L}}.$$  

(3.2.25)

We mention here the second formulation because of the condition $(3.2.18)$, which is equal to zero only in the case when $u > 0$. Indeed $(3.2.18)$ reads

$$\epsilon \partial E(u) + \lambda(t = 1) - \eta(x) + \eta_0 = 0,$$

with complementarity condition $\eta u = 0$ and constant $\eta_0$ which is chosen such that $\int u = 1$. Hence, if $u = 0$ the Lagrange multiplier $\eta$ can take arbitrary values and $(3.2.18)$ is not necessary equal to zero. The reformulation $(3.2.25)$ avoids this problem. However, we obtain a unique solution for $(3.2.25)$ only for strictly convex functionals $E(u)$. For the numerical solution we will exploit the Benamou-Brenier optimality conditions $(3.2.15)-(3.2.20)$ and $(3.2.21)-(3.2.24)$.

### 3.2.3 Rescaling and Diffusion Filtering

In the formulation $(3.2.11)-(3.2.14)$ we can perform standard rescaling of the time variable and associated rescaling of density, velocity and dual variable to

$$\hat{t} = \epsilon t, \quad \hat{\mu}(\hat{t}, dx) = \mu(t, dx), \quad \hat{v}(\hat{t}, x) = \frac{1}{\epsilon} v(t, x), \quad \hat{\lambda}(\hat{t}, x) = \lambda(t, x).$$  

(3.2.26)
The optimisation problem \((3.2.11)-(3.2.14)\) in rescaled variables becomes

\[
E(u) + \frac{1}{2} \int_0^\epsilon \int_\Omega |\hat{\nu}|^2 d\hat{\mu}(x) \, dt,
\]

subject to

\[
\partial_t \hat{\mu} + \nabla \cdot (\hat{\mu} \hat{\nu}) = 0 \tag{3.2.28}
\]

\[
\hat{\mu}(\hat{t} = 0) = \nu \tag{3.2.29}
\]

\[
\hat{\mu}(\hat{t} = \epsilon) = u\mathcal{L}^N \tag{3.2.30}
\]

and the optimality condition with dual variable \(\hat{\lambda}\) is again \(\hat{\nu} = \frac{1}{\epsilon} \nabla \hat{\lambda}\) and

\[
\partial_t \hat{\lambda} + \hat{\nu} \cdot \nabla \hat{\lambda} - \frac{1}{2} |\hat{\nu}|^2 = 0. \tag{3.2.31}
\]

The rescaling is useful in the numerical solution, since it guides the choice of the time discretisation in \(\epsilon\). Moreover, the relation to diffusion filtering can be understood by the asymptotic \(\epsilon \to 0\). The limit \(\epsilon \to 0\) of a variational model as \((3.1.1)\), respectively a sequence thereof, has been well studied via the method of minimising movements approximating a metric gradient flow with energy \(E\) (cf. [9]). We execute this statement in Chapter 4.

### 3.2.4 Uniqueness

In the following, we investigate an uniqueness result for the velocity \(v\):

**Theorem 3.2.2.** Let \(E\) be a differentiable and strictly convex energy functional. Let \(\epsilon > 0\) and \((\mu, \lambda, u)\) be a solution of \((3.2.15)-(3.2.20)\). Then the transport velocity \(v = \nabla \lambda\) is unique on the support of \(\mu\).

**Proof.** To prove the claim we basically follow the strategy to prove uniqueness for mean-field games introduced in [116]. Let us first rewrite the equations.

Inserting \((3.2.17)\), i.e., \(v = \nabla \lambda\), into \((3.2.15)\) and \((3.2.16)\) we get

\[
\partial_t \mu + \nabla \cdot (\mu \cdot \nabla \lambda) = 0 \tag{3.2.32}
\]

\[
\partial_t \lambda + \frac{1}{2} |\nabla \lambda|^2 = 0.
\]

Let \((\mu_1, \lambda_1, u_1)\) and \((\mu_2, \lambda_2, u_2)\) be two solutions to \((3.2.15)-(3.2.20)\). Writing the equations in \((3.2.32)\) for both solutions and subtracting the corresponding ones we get

\[
\partial_t (\mu_1 - \mu_2) + \nabla \cdot (\mu_1 \nabla \lambda_1 - \mu_2 \nabla \lambda_2) = 0 \tag{3.2.33}
\]

\[
\partial_t (\lambda_1 - \lambda_2) + \frac{1}{2} (|\nabla \lambda_1|^2 - |\nabla \lambda_2|^2) = 0. \tag{3.2.34}
\]
By multiplying the first equation with \((\lambda_1 - \lambda_2)\), the second equation with \((\mu_1 - \mu_2)\), integrating both equations over \((0, 1) \times \Omega\) and adding them we get

\[
\int_{(0,1) \times \Omega} \partial_t (\mu_1 - \mu_2) (\lambda_1 - \lambda_2) + (\mu_1 - \mu_2) \partial_t (\lambda_1 - \lambda_2) \, dx dt \quad (= I)
\]

\[
+ \int_{(0,1) \times \Omega} \nabla \cdot (\mu_1 \nabla \lambda_1 - \mu_2 \nabla \lambda_2) (\lambda_1 - \lambda_2) \, dx dt \quad (= II)
\]

\[
+ \frac{1}{2} \int_{(0,1) \times \Omega} \left( |\nabla \lambda_1|^2 - |\nabla \lambda_2|^2 \right) (\mu_1 - \mu_2) \, dx dt = 0 \quad (= III)
\]

Then

\[
I = \int_{(0,1) \times \Omega} \partial_t ((\mu_1 - \mu_2) \cdot (\lambda_1 - \lambda_2)) \, dx dt
\]

\[
= \int_{\Omega} (\mu_1 - \mu_2) \cdot (\lambda_1 - \lambda_2)|_0^1 \, dx
\]

\[
= \int_{\Omega} (\mu_1(t = 1) - \mu_2(t = 1)) \cdot (\lambda_1(t = 1) - \lambda_2(t = 1)) \, dx
\]

\[
= - \int_{\Omega} (u_1 \mathcal{L}^d - u_2 \mathcal{L}^d) \cdot (E'(u_1) - E'(u_2)) \, dx.
\]

Since \(E\) is strictly convex we conclude \(I \leq 0\).

Moreover, the function \(H(p) = \frac{1}{2}p^2\) is strictly convex with derivative \(H'(p) = p\). Hence for \(p_1\) and \(p_2\) we have

\[
H'(p_1)(p_2 - p_1) \leq H(p_2) - H(p_1).
\]

Implementing this property into the integral (II) with \(p_1 = \nabla \lambda_1\) and \(p_2 = \nabla \lambda_2\) we receive

\[
II = \int_{(0,1) \times \Omega} \mu_1 \nabla \lambda_1 (\nabla \lambda_2 - \nabla \lambda_1 + \mu_2 \nabla \lambda_2)(\nabla \lambda_1 - \nabla \lambda_2) \, dx dt
\]

\[
\leq \frac{1}{2} \int_{(0,1) \times \Omega} \mu_1 (|\nabla \lambda_2|^2 - |\nabla \lambda_1|^2) + \mu_2 (|\nabla \lambda_1|^2 - |\nabla \lambda_2|^2) \, dx dt
\]

\[
= - \frac{1}{2} \int_{(0,1) \times \Omega} (\mu_1 - \mu_2)(|\nabla \lambda_1|^2 - |\nabla \lambda_2|^2) \, dx dt
\]

\[
= -III.
\]

Therefore, it follows that \(II + III \leq 0\). Since additionally \(I + II + III = 0\) and \(I \leq 0\), we have that \(I = 0\) and \(II + III = 0\). From \(I = 0\) and the strict convexity of \(E\), necessarily \(u_1 = u_2 =: u\). Moreover \(II + III = 0\) implies that

\[
- \int_{(0,1) \times \Omega} (\mu_1 \nabla \lambda_1 - \mu_2 \nabla \lambda_2)(\nabla \lambda_1 - \nabla \lambda_2) + \frac{1}{2} (|\nabla \lambda_1|^2 - |\nabla \lambda_2|^2)(\mu_1 - \mu_2) \, dx dt = 0,
\]
and therefore
\[- \int_{\Omega \times (0,1)} (\nabla \lambda_1 - \nabla \lambda_2)^2 (\mu_1 + \mu_2) \, dxdt = 0,\]
and hence either
\[\nabla \lambda_1 - \nabla \lambda_2 = 0 \text{ or } \mu_1 = -\mu_2.\]
Since \(\mu_i \geq 0, \ i = 1, 2\) for all times, the second condition is fulfilled only if \(\mu_1 = \mu_2 = 0\), which is the trivial case of no transport. If \(\mu_i \neq 0, \ i = 1, 2\), then
\[\nabla \lambda_1 - \nabla \lambda_2 = 0, \ i.e., \ \nabla \lambda_1 = \nabla \lambda_2.\]

Remark 3.2.3. From Theorem [3.2.2] we know that the Benamou-Brenier optimality conditions deliver a unique transport in the sense that the velocity \(v = \nabla \lambda\) and the endpoint \(u\mathcal{L}^N = \mu(t = 1)\) of the transport are unique on the support of \(\mu\). Since we cannot guarantee sufficient regularity of the transport velocity \(v\) (cf. [6]) this is not enough to deduce uniqueness of \(\mu\) and \(\lambda\) on the whole time interval. In particular, the solution may be non-unique outside the support of \(\mu\).

3.2.5 Self-Similar Solutions

In the following, we compute an interesting special class of solutions of self-similar structure. The basis of our computation is the characterisation of the optimal transport between a measure and its rescaled version, derived in [17]. Since we are looking for smoothed measures with density, in this case also the initial measure needs to have density. Hence we look for solutions of the form
\[d\nu = \frac{1}{\delta^N} \ u \left( \frac{x}{\delta} \right) \ dx. \tag{3.2.35}\]
In the Benamou-Brenier formulation it is well-known that optimal transport between two scaled versions of the same density is performed with density
\[\rho(t, x) = \frac{1}{(at + b)^N} u \left( \frac{x}{at + b} \right), \tag{3.2.36}\]
which is a translation and dilation of the initial data with nonnegative constants \(a\) and \(b\) such that \(a + b = 1\). In this approach we consequently look for solutions centered at zero, but by a simple shift of the \(x\) variable we obtain the same solutions centered at different points. To compute the special form for the velocity vector \(v\) and the dual variable \(\lambda\) we consider the continuity equation
\[\partial_t \rho + \nabla \cdot (\rho v) = \partial_t \rho + \nabla \rho \cdot v + \rho \nabla \cdot v = 0. \tag{3.2.37}\]
Hence, with simple calculations we follow

$$
\partial_t \rho = \frac{-aN}{(at+b)^{N+1}} u \left( \frac{x}{at+b} \right) + \frac{1}{(at+b)^N} \nabla u \left( \frac{x}{at+b} \right) \frac{-ax}{at+b},
$$

$$
\nabla \rho = \frac{1}{(at+b)^N} \nabla u \left( \frac{x}{at+b} \right) \frac{1}{at+b}.
$$

Inserting this into (3.2.37) we obtain

$$
\frac{-aN}{(at+b)^{N+1}} u + \frac{1}{(at+b)^N} \nabla u \left( \frac{x}{at+b} \right) \frac{-ax}{at+b} = \frac{1}{(at+b)^N} \nabla u \frac{1}{at+b} v - \frac{1}{(at+b)^N} u \nabla v.
$$

Therefore we obtain for the velocity

$$
v(t, x) = \frac{ax}{at+b},
$$

(3.2.38)

and with $v = \nabla \lambda$ for the dual variable

$$
\lambda(t, x) = \frac{a|x|^2}{2(at+b)}.
$$

(3.2.39)

With this particular $\lambda$ we examine the optimality condition (3.2.18)

$$
\epsilon E'(u) + \lambda(1, x) - \eta(x) + \eta_0 = 0,
$$

(3.2.40)

where $\eta \geq 0$ satisfies the complementarity condition $\eta u = 0$ almost everywhere. With the above form of $\lambda$ and rescaling the dual variables with $\epsilon$, and $\sigma = \sqrt{\frac{\epsilon}{a}}$ we rewrite (3.2.40)

$$
E'(u) = -\frac{|x|^2}{2\sigma^2} + \frac{\eta(x)}{\epsilon} - \frac{\eta_0}{\epsilon},
$$

(3.2.41)

where $\eta$ still satisfies complementarity. In Chapter 5 and 6 we compute such solutions in the most relevant cases, that means for different regularisation energies $E(u)$. The goal is to get more insight into the structure of the solutions.
Chapter 4

Gradient Flows

The following chapter deals with the connection of the variational problem (3.1.1) to certain gradient flow equations. The fact that some evolution equations can be formulated as gradient flows for special energy functionals with respect to a certain metric is known for a long time, in particular since De Giorgi introduced the minimising movement scheme [86, 85]. With respect to the Wasserstein metric this relation has been first studied by Jordan, Kinderlehrer and Otto in [99]. This pioneering work has resulted in a number of valuable contributions for the study of non-linear evolution equations in terms of Wasserstein gradient flows, e.g., [4, 142, 83, 52, 48, 162, 163, 53, 65, 84, 125]. In Section 4.1 we exemplify this idea by identifying some special evolution equations as gradient flow equations for some energy functionals. Section 4.2 is concerned with two different approaches to the gradient flow theory. On the one hand we present the minimising movement scheme which can be recast in different metric frameworks. On the other hand we introduce the approach on a Riemannian manifold given by Otto [142].

4.1 The Gradient Flow Formulation

The identification of some evolution equations as gradient flows for special energy functionals with respect to the Wasserstein metric yields a new understanding of several nonlinear partial differential equations. Especially the gradient flow formulation can be well used to prove existence and uniqueness in an easy way. But it also leads to new contraction and energy estimates, which can be used to study the asymptotic behavior of solutions. In the following, we first introduce the idea of the gradient flow scheme in a general framework and afterwards we state the gradient flow equation in the space of probability measures.
Let $X$ be in $\mathbb{R}^N$ and $E : \mathbb{R}^N \to \mathbb{R}$ be a smooth and differentiable energy functional. The gradient flow equation
\[
\frac{dX}{dt} = -\text{grad}E(X)
\]  
(4.1.1)
describes the relation between the time derivative of the state and the gradient of the energy functional $E$.

To exemplify (4.1.1), we state some gradient flow equations for special energy functionals with respect to the $L^2$ metric:

Logarithmic entropy: 
\[
E(X) = \int_{\mathbb{R}^N} X \log X \, dx \\
\partial_t X = -(\log X + 1)
\]

$L^2$ – regularisation: 
\[
E(X) = \frac{1}{2} \int_{\mathbb{R}^N} X^2 \, dx \\
\partial_t X = -X
\]

Dirichlet regularisation: 
\[
E(X) = \frac{1}{2} \int_{\mathbb{R}^N} |\nabla X|^2 dx \\
\partial_t X = \Delta X
\]

Fisher information: 
\[
E(X) = \frac{1}{2} \int_{\mathbb{R}^N} X |\nabla \log X|^2 dx \\
\partial_t X = 2 \frac{\Delta \sqrt{X}}{\sqrt{X}}
\]

TV regularisation: 
\[
E(X) = \int_{\mathbb{R}^N} |\nabla X| \, dx \\
\partial_t X = \nabla \cdot \left( \frac{\nabla X}{|\nabla X|} \right)
\]

In the sequel we are interested in the Wasserstein gradient flow. Let $u \mathcal{L}^N \in \mathcal{P}_p(\mathbb{R}^N)$ and $E : \mathcal{P}_p(\mathbb{R}^N) \to \mathbb{R} \cup \{+\infty\}$. We consider the variational integral
\[
\mathcal{E}(\mu) := \begin{cases} 
E(u) = \int_{\mathbb{R}^N} f(x, u(x), \nabla u(x)) \, dx & \text{if } \mu = u \cdot \mathcal{L}^N \text{ with } u \in C^1(\mathbb{R}^N), \\
+\infty & \text{otherwise},
\end{cases}
\]

with a $C^2$-function $f(x, z, p) : \mathbb{R}^N \times [0, \infty] \times \mathbb{R}^N \to \mathbb{R}$, with $f(x, 0, p) = 0$ for every $x, p \in \mathbb{R}^N$.

For smooth functions we write the gradient flow with respect to the Wasserstein metric in the following way [175, 9]:
\[
\text{grad}_W E(u) = -\nabla \cdot \left( u \nabla \frac{\delta E}{\delta u} \right).
\]  
(4.1.2)

We conclude that the gradient flow equation in the space of probability measures endowed with the Wasserstein distance reads as follows:
\[
\frac{\partial u}{\partial t} = \nabla \cdot \left( u \nabla \frac{\delta E}{\delta u} \right), \quad \text{in } (0, +\infty) \times \mathbb{R}^N,
\]  
(4.1.3)

where the first variation of $E$ is defined as
\[
\frac{\delta E}{\delta u} = -\nabla \cdot f_p(x, u, \nabla u) + f_z(x, u, \nabla u),
\]  
(4.1.4)
with respect to the $L^2$-Euclidean structure.

Following the example given by McCann [126], we consider a general energy functional

$$E(u) = \mathcal{U}(u) + \mathcal{V}(u) + \mathcal{G}(u)$$

with internal, potential and interaction energy defined as in Section 2.3. Then, the gradient flow equation (4.1.3) for $E$ with respect to the Wasserstein distance reads

$$\frac{\partial u}{\partial t} = \nabla \cdot \left( u(\nabla U'(u) + \nabla V + (\nabla G * u)) \right).$$

In the last years, many partial differential equations have been identified as gradient flows for special energy functionals with respect to the Wasserstein distance. Since we can write (4.1.3) in the following gradient flow scheme:

$$\begin{align*}
\partial_t u + \nabla \cdot (uv) &= 0 & \text{(continuity equation)} \\
uvw &= u\nabla \psi & \text{(gradient condition)} \\
\psi &= -\frac{\delta E}{\delta u} & \text{(nonlinear relation)},
\end{align*}$$

we can identify the heat equation as the gradient flow for the logarithmic entropy with respect to the Wasserstein distance:

$$\begin{align*}
\partial_t u - \Delta u &= 0 & \text{(continuity equation)} \\
uvw &= -\frac{\nabla u}{u} & \text{(gradient condition)} \\
\psi &= -\log u - 1,
\end{align*}$$

and (4.1.8) implies $E(u) = \int_{\mathbb{R}^N} u \log u \, dx$.

Furthermore, the gradient flow equation for the $L^2$-regularisation energy with respect to the Wasserstein metric is the well known porous medium equation. This connection and the asymptotic analysis was investigated by Otto in 2001 (cf. [142]). The thin film equation constitutes the gradient flow for the Dirichlet regularisation with respect to the Wasserstein metric (cf. [124]), and the Derrida-Lebowitz-Speer-Spohn equation (DLSS), also known as drift diffusion equation was identified as the gradient flow for the Fisher information with respect to the Wasserstein metric by Gianazza, Savaré and Toscani (cf. [84]). Because we consider in the sequel also the TV-regularisation, we state here the corresponding gradient flow equation with respect to the Wasserstein distance, which is a highly non-linear fourth order diffusion equation.
Logarithmic entropy: \( E(u) = \int_{\mathbb{R}^N} u \log u \, dx \), \( \partial_t u = \Delta u \),

\( L^2 \) – regularisation: \( E(u) = \int_{\mathbb{R}^N} u^2 \, dx \), \( \partial_t u = \Delta u^2 \),

Dirichlet regularisation: \( E(u) = \frac{1}{2} \int_{\mathbb{R}^N} |\nabla u|^2 \, dx \), \( \partial_t u = -\nabla \cdot (u \nabla \Delta u) \),

Fisher information: \( E(u) = \int_{\mathbb{R}^N} |\nabla \ln u|^2 \, dx \), \( \partial_t u = -2 \nabla \left( u \nabla \left( \frac{\Delta \sqrt{u}}{\sqrt{u}} \right) \right) \),

TV regularisation: \( E(u) = \int_{\mathbb{R}^N} |\nabla u| \, dx \), \( \partial_t u = -\nabla \cdot \left( u \nabla \nabla \cdot \left( \frac{\nabla u}{|\nabla u|} \right) \right) \).

We refer for the analytical discussion of these energy functionals and the corresponding gradient flow equations to Chapter 5|6.

4.2 Motivation of the Gradient Flow Formulation

The gradient flow formulation (4.1.3) can be motivated in two ways. The first approach is based on the concept of minimising movements by De Giorgi [86, 85]. Especially, the idea that the solution of a gradient flow equation can be approximated by solving a variational scheme, with respect to the Wasserstein metric, was motivated by the work of Jordan, Kinderlehrer and Otto in 1999 [99]. Hence, the discrete time stepping strategy for the construction of solutions of (4.1.3) is known under the name JKO-scheme.

The minimising movement scheme can be rewritten in different metric frameworks. In contrast, the second approach formulates the gradient flow on a smooth Riemannian manifold by using the notions of tangent spaces and velocity vector fields. Furthermore an appropriate definition of the subdifferential of the energy is necessary [9]. We explain these two approaches in the following.

4.2.1 The discrete variational scheme

The idea that the gradient flow formulation suggests a general variational scheme, in particular that the solution of the gradient flow equation can be approximated by a recursive scheme with respect to a certain distance is based on the minimising movement approach by De Giorgi [86 85]. This concept can be implemented in several metric frameworks, e.g. the investigation to the optimal transport structure was done
by Otto et. al [99], for the Fokker-Planck equation, which is also based on the ideas in [140].

First we introduce the minimising movement scheme in a simple manner, i.e. in a general framework. Let $E : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ be a smooth and differentiable energy functional. We consider an equidistant partition of the time interval $(0, +\infty]$

$$P_{\tau} := \{0 = t_0^\tau < t_1^\tau < \cdots t_n^\tau < \cdots\}, \quad I_n^\tau := (t_n^\tau, t_{n+1}^\tau],$$

with time step size $\tau = t_n^\tau - t_{n-1}^\tau > 0$. Let $X_0$ be the initial data and $X_n^\tau \approx X(t_n^\tau)$ for $n = 1, \ldots$ the approximation of the function $X$ at time $t_n^\tau$. Then the discretisation of (4.1.1) with the implicit Euler method in the Euclidean context yields

$$\frac{X_n^\tau - X_{n-1}^\tau}{\tau} = -\nabla E(X_n^\tau). \quad (4.2.1)$$

Obviously, (4.2.1) is the Euler equation for the optimisation problem

$$\frac{1}{2\tau}|V - X_{n-1}^\tau|^2 + E(V) \to \min_V. \quad (4.2.2)$$

For a general distance function $d$ (4.2.2) reads

$$\frac{1}{2\tau}d(V, X_n^\tau)^2 + E(V) \to \min_V.$$

Out of this simple idea we can conclude, that the solution of the gradient flow equation (4.1.1) at time $t \approx n\tau$ can be approximated by the discrete solution $X_n^\tau$ obtained by a recursive minimisation of a variational problem.

**Theorem 4.2.1.** The iterative scheme, based on the minimising movement by De Giorgi [85] reads:

1. Initialisation: $P_\tau$ given, $\tau > 0$, $X_0^\tau \approx X^0 \in \mathbb{R}^N$ initial data.

2. Given $X_n^\tau$. Find $X_{n+1}^\tau$ as a solution of

$$X_{n+1}^\tau = \arg\min_X \left(E(X) + \frac{d^2(X, X_n^\tau)}{2\tau}\right). \quad (4.2.3)$$

3. Define $X_\tau > 0$ as the piecewise constant function (discrete solution):

$$X_\tau(t) := X_n^\tau \quad \text{if} \quad t \in (t_{n-1}^\tau, t_n^\tau].$$

4. Pass to the limit: $\tau \to 0$.

That means, that the solution of the partial differential equation is a limiting curve of a variational scheme when the time steps $\tau$ tend to zero.
Let us reformulate this idea for optimal transport problems. Let $\Omega \subset \mathbb{R}^N$ be open and bounded and $u \in \mathcal{P}_p(\Omega)$ with $p \geq 1$ a probability density and $E : \mathcal{P}_p(\Omega) \to \mathbb{R} \cup \{+\infty\}$ be a smooth and differentiable functional and lower semicontinuous. Furthermore,

$$\nu \mapsto \Phi(\tau, \mu; \nu) = \frac{1}{p} W_p(\mu, \nu)^2 + \tau^{p-1} E(\nu)$$

(4.2.4)

admits a minimum point $\mu_\tau$ for all $\tau \in (0, \tau^*)$ and $\mu \in \mathcal{P}_p(\Omega)$. Then we can state the minimising movement scheme (see [9], Section 11.1.3):

**Theorem 4.2.2** (Minimising movement scheme in $\mathcal{P}_p(\Omega)$). Consider a uniform partition of the time interval $(0, +\infty]$

$$P_\tau := \{ 0 = t_0^\tau < t_1^\tau < \cdots t_n^\tau < \cdots \}, \quad I_n^\tau := (t_n^\tau, t_{n+1}^\tau],$$

with time step size $\tau = t_n^\tau - t_{n-1}^\tau > 0$ and a family of intial data $M_0^\tau$, such that

$$M_\tau^0 \to \mu_0, \quad \text{in } \mathcal{P}_p(\Omega), \quad E(M_\tau^0) \to E(\mu_0) \quad \text{as } \tau \downarrow 0,$$

For every $\tau \in (0, \tau^*)$ a corresponding family of sequences $(M_n^\tau)_{n \in \mathbb{N}}$ recursively defined as

$$M_n^\tau \text{ minimises } \mu \mapsto \Phi(\tau, M_{n-1}^\tau; \mu)$$

(4.2.5)

always exists. The piecewise constant interpolant

$$\overline{M}_\tau(t) := M_n^\tau \text{ if } t \in (t_{n-1}^\tau, t_n^\tau],$$

is called the discrete solution. A curve $\mu_t$ is a generalised minimising movement $GMM(\Phi; \mu_0)$, if there exists a sequence $(\tau_k) \downarrow 0$ such that

$$\overline{M}_{\tau_k}(t) \to \mu_t$$

narrowly in $\mathcal{P}(\Omega)$ for every $t > 0$, as $k \to \infty$.

The convergence to a solution of the gradient flow equation is not trivial. Some suitable estimates have to be fulfilled to ensure relative compactness of the family $(M_\tau)$ and furthermore the computation for $\tau \to 0$ is extensive. The uniqueness follows from the convexity of $E$ and the strict convexity of the variational scheme. In conclusion we can say, that from the point of view of gradient flows, a solution of the minimisation problem

$$\frac{1}{2} W_2(\nu, u\mathcal{L}^N)^2 + \epsilon E(u) \to \min_{u\mathcal{L}^N}$$

(4.2.6)

with $\nu \in \mathcal{P}_2(\Omega)$ and probability density $u\mathcal{L}^N$, can be interpreted as a discrete approximation of a solution of the gradient flow of $E(u)$ with respect to the quadratic
Wasserstein metric. More precisely, \( (4.2.6) \) represents one timestep of De Giorgi’s minimising movement scheme to the functional \( E(u) \) with timestep \( \epsilon > 0 \). In this case the variational scheme reads:

Let \( u^0 \) be an initial probability density and \((u^k)_{k \geq 1}\) a sequence of solutions for the variational scheme. Then, \( u^{k+1} \) is the solution for the minimisation problem

\[
 u^{k+1} \in \arg\min_u \left( \frac{1}{2} W_2^2(u^k \mathcal{L}^d, u \mathcal{L}^d)^2 + \tau E(u) \right).
\]  
(4.2.7)

When the time step \( \tau \) tends to zero, we hope to obtain convergence to the solution of the adapted gradient flow equation.

Note that we are not directly computing solutions of the gradient flow equation. In fact, a minimiser of \( (4.2.6) \) represents only one timestep of size \( \epsilon \) of \( (4.2.7) \). However, solving \( (4.2.6) \) iteratively, i.e., computing a minimiser \( u_\epsilon \) and subsequently a minimiser \( u_2 \) for \( \nu = u_\epsilon \mathcal{L}^d \) and so forth, is a possible strategy to numerically approximate the corresponding gradient flow \( (4.1.3) \) with time step size \( \epsilon \).

### 4.2.2 Gradient Flow Formulation on Riemannian Manifolds

The Wasserstein gradient flow \( (4.1.3) \) can also be motivated in a Riemannian setting. In particular, we justify, that the space of probability measures \( \mathcal{P}(\mathbb{R}^N) \) is endowed with a Riemannian metric \[142\].

Therefore, we introduce the notion of a tangent space \( T_\mu \mathcal{P} \) as

\[
 T_\mu \mathcal{P} = \left\{ s : \mathbb{R}^N \to \mathbb{R} | \int s = 0 \right\}.
\]  
(4.2.8)

Furthermore we need a metric tensor \( \langle \cdot, \cdot \rangle_W \) on each tangent space \( T_\mu \mathcal{P} \) which is necessary, to convert the co-tangent vector field of \( E \) into a tangent vector field, where \( E : \mathcal{P}(\mathbb{R}^N) \to \mathbb{R} \) be a smooth, differentiable functional, i.e.

\[
 \langle \text{grad}_W E, s \rangle_W = \text{diff}E.s, \quad \text{for all vector fields } s.
\]

This metric tensor shall define a norm \( ||\cdot||_p \) on each tangent space, so that we can write for the Benamou and Brenier formulation

\[
 W_2^2(\mu_0, \mu_1) = \inf \left\{ \int_0^1 \left\| \frac{d\mu}{dt} \right\|_{\mu(t)}^2 dt ; \mu(0) = \mu_0, \mu(1) = \mu_1 \right\}
\]

where the infimum is taken over all paths connecting \( \mu_0 \) and \( \mu_1 \) with

\[
 \left\| \frac{d\mu}{dt} \right\|_{\mu(t)}^2 = \inf_v \left\{ \int |v|^2 d\mu(x) | \frac{\partial\mu}{\partial t} + \nabla \cdot (\mu v) = 0 \right\}.
\]
The tangent vectors in $T_\mu \mathcal{P}$ are time derivatives at time zero of some trajectory $t \mapsto \mu(t)$ with $\mu(0) = \mu$. From the fluid dynamic point of view, $\mu$ describes the mass of the density and therefore fulfills the conservation of mass principle, i.e. the continuity equation

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

Therefore the tangent vectors shall have the form $-\nabla \cdot (\mu v)$. Additionally we make assumptions on the velocity $v$, since it shall be the minimiser of the kinetic energy: Assume that $v_0$ is a minimiser of the kinetic energy and $v_0 + \epsilon \frac{w}{\mu}$ with $w$ be a vector field with zero divergence, fulfills the continuity equation. Since

$$\int |v_0|^2 d\mu(x) \leq \int |v_0 + \epsilon \frac{w}{\mu}|^2 d\mu(x),$$

$$\leq \int \left(v_0^2 + 2v_0 \left(\frac{\epsilon w}{\mu}\right) + \left(\frac{\epsilon w}{\mu}\right)^2\right) d\mu(x),$$

we conclude, after neglecting the terms without $\epsilon$, dividing through $\epsilon$ and $\epsilon \to 0$,

$$\int v_0 w = 0.$$

Hence, $v_0 = \nabla u$ must be a gradient and we obtain

$$-\nabla \cdot (\mu \nabla u) = s.$$

With this underlying Riemannian structure we can compute the scalar product of two tangent vectors

$$\langle s_1, s_2 \rangle_W = \int \mu \langle \nabla u_1, \nabla u_2 \rangle$$

where $u_1$ and $u_2$ satisfy

$$-\nabla \cdot (\mu \nabla u_1) = s_1, \quad -\nabla \cdot (\mu \nabla u_2) = s_2.$$

We can expand (4.1.1) along the tangent vector $s$, using the above definitions.

$$\left\langle \frac{\partial \mu}{\partial t}, s \right\rangle_W + \langle \text{grad}_W E, s \rangle_W = \left\langle \frac{\partial \mu}{\partial t}, s \right\rangle_W + \text{diff} E, s = 0.$$ (4.2.11)

After simple computations we obtain

$$\text{grad}_W E(u) = -\nabla \cdot \left(u \nabla \frac{\partial E}{\partial u}\right).$$ (4.2.12)
Porous medium equation

In the following we want to recapitulate the proof, that the porous medium equation constitutes a gradient flow for the energy functional $E(u) = \frac{1}{m(m-1)} \int_\Omega u^m dx$ for $m > 1$ with respect to the Wasserstein metric $[142]$.

$$\text{diff} E(u).s = \frac{1}{m-1} \int_\Omega u^{m-1}s \, dx$$

with $s = -\nabla \cdot (u \nabla v)$ we obtain with integration by parts

$$\text{diff} E(u).s = -\frac{1}{m-1} \int_\Omega u^{m-1} \nabla \cdot (u \nabla v) dx$$

$$= \int_\Omega u^{m-1} \nabla u \nabla v \, dx = -\frac{1}{m} \int_\Omega \Delta u^m v \, dx$$

and with

$$\left\langle \frac{\partial u}{\partial t}, s \right\rangle_W = \int_\Omega \frac{\partial u}{\partial t} v \, dx$$

we conclude

$$\int_\Omega \frac{\partial u}{\partial t} v dx - \frac{1}{m} \int_\Omega \Delta u^m v \, dx = 0$$

and eventually the porous medium equation $\frac{\partial u}{\partial t} - \frac{1}{m} \Delta u^m = 0$.

Heat equation

A similar computation identifies the heat equation as the gradient flow for the logarithmic entropy $E(u) = \int_\Omega u \log u \, dx$ with respect to the Wasserstein metric:

$$\text{diff} E(u).s = \int_\Omega (\log u + 1)s \, dx$$

with $s = -\nabla \cdot (u \nabla v)$ we obtain with integration by parts

$$\text{diff} E(u).s = -\int_\Omega (\log u + 1) \nabla \cdot (u \nabla v) \, dx$$

$$= \int_\Omega \nabla (\log u + 1) u \nabla v dx = -\int_\Omega \nabla (\nabla u) v \, dx$$

Then, we conclude

$$\int_\Omega \frac{\partial u}{\partial t} v dx - \int_\Omega \Delta u v \, dx = 0$$

and eventually the heat equation $\frac{\partial u}{\partial t} - \Delta u = 0$. This concludes the proof, that the heat equation is the gradient flow for the logarithmic entropy with respect to the Wasserstein distance.
4.3 Trend to Equilibrium

The discussion of partial differential equations of the form (4.1.1) goes along with the study of long time asymptotic behavior. We present in the following two different approaches to this topic. On the one hand, we introduce the entropy dissipation method, proposed by Carrillo et. al. [50], which considers the convergence estimate in terms of the relative entropy. On the other hand, we give a brief introduction to Otto’s approach, who investigated convergence inequalities by simple Riemannian computation [142].

4.3.1 Entropy Dissipation Method

The idea of the entropy dissipation method introduced in [50] is to compute the convergence rate in terms of an energy functional $E$. Since the stationary solution $u_*$ is the minimum of the energy functional $E$, the unique minimum if $E$ is displacement convex [2.3.7], the convergence discussion for the partial differential equation can be recast in terms of $E$, i.e.

$$E(u(t)) \to E(u_*) \quad \text{for } t \to \infty.$$  

Convergence estimates are usually stated in terms of the relative entropy given by

$$E(u|u_*) = E(u) - E(u_*), \quad (4.3.1)$$

and the entropy production functional $D$

$$\frac{d}{dt} E(u(t)) = -D(u(t)). \quad (4.3.2)$$

$E$ is a Lyapunov functional if $D$ is positive, i.e. if $E$ is non-increasing in time along the solution. Moreover, we compute the second time derivative of $D$. In the following we derive the exponential convergence for a special energy functional $E$: If the second derivative of $E$ with respect to the time has the form

$$\frac{dD(u(t))}{dt} = -\lambda D(u(t)) - R(t)$$

for $\lambda \leq 0$ and $R(t) \geq 0$ on $\mathbb{R}^N$. Then we can evaluate

$$-\lambda D(u(t)) \geq \frac{dD}{dt}(u(t)).$$

Using Gronwall’s Lemma we conclude

$$D(u(t)) \leq \exp(-\lambda t)D(u(0)).$$
The integration over time yields

\[ E(u(0)) - E(u_*) \leq \frac{1}{\lambda} D(u(0)), \]

and therefore

\[ \frac{d(E(u(t)) - E(u_*))}{dt} \leq -\lambda(E(u(t)) - E(u_*)). \]

It is easy to see that the last inequality yields the following log sobolev inequality

\[ E(u(t)|u_*) \leq \frac{1}{\lambda} D(u(t)|u_*), \tag{4.3.3} \]

and we can conclude the exponential convergence of the energy functional \( E \)

\[ E(u|u_*) \leq \exp(-2\lambda t) E(u_0|u_*). \]

Mostly we obtain only an entropy-entropy production inequality:

\[ D(u) \geq \theta(E(u|u_*)), \tag{4.3.4} \]

where \( H \to \theta(H) \) is some continuous function and strictly increasing and strictly positive when \( H > 0 \). For \( \theta(H) = \lambda H \) we can conclude, as seen above, the exponential convergence to the equilibrium, and for \( \theta(H) = KH^{1+\alpha} \) for \( \alpha > 0 \) the polynomial rate of convergence. Mostly the convergence in the entropy sense implies the convergence in the \( L^1 \)-sense of \( u(t) \) to \( u_* \), which is known as the Csiszar-Kullback type inequality.

**Theorem 4.3.1.** ([175, Theorem 9.2]) Let \( E \) be a lower semi-continuous functional on \( \mathcal{P}(\mathbb{R}^N) \), equipped with the quadratic Wasserstein distance \( W_2^2 \). Then If \( E \) is uniformly displacement convex with constant \( \lambda > 0 \), then it admits a unique minimiser \( u_* \) in \( \mathcal{P}_2(\mathbb{R}^N) \) and

\[ \forall u \in \mathcal{P}_2(\mathbb{R}^N), \quad D(u) \geq 2\lambda E(u|u_*). \]

Furthermore, we state the following inequality for the relative entropy:

**Corollary 4.3.2.** ([175, Corollary 9.2]) If \( E \) is \( \lambda \)-uniformly displacement convex for some \( \lambda > 0 \) then one has for the equation the following estimates of trend to equilibrium

\[ E(u(t)|u_*) \leq E(u_0|u_*) \exp(-2\lambda t). \]

We refer to [11] [12] [143] for more details on sobolev inequalities and entropy methods.
Entropy dissipation method for the logarithmic entropy

To exemplify the above theory, we consider here the entropy dissipation method for the logarithmic entropy $E(u) = \int_\Omega u \log u \, dx$. First of all we compute the time derivative of $E$:

$$\frac{dE}{dt} = \int_\Omega \partial_t u (\log u + 1) \, dx = \int_\Omega \Delta u (\log u + 1) \, dx$$
$$= - \int_\Omega \nabla u \nabla (\log u + 1) \, dx = - \int_\Omega \frac{1}{u} (\nabla u)^2 \, dx.$$

The entropy production functional $D(u)$ is positive if $u \geq 0$. Furthermore we can compute the second time derivative

$$\frac{dD}{dt} = \int_\Omega -\frac{1}{u^2} \partial_t u (\nabla u)^2 \, dx - 2 \int_\Omega \nabla \left( \frac{\nabla u}{u} \right) \partial_t u \, dx.$$
$$= \int_\Omega -\frac{1}{u^2} \Delta u (\nabla u)^2 \, dx - 2 \int_\Omega \nabla \left( \frac{\nabla u}{u} \right) \Delta u \, dx.$$
$$= \int_\Omega -\frac{1}{u^2} \Delta u (\nabla u)^2 \, dx + 2 \int_\Omega \nabla \left( \frac{\nabla u}{u^2} \right) \Delta u \, dx - \int_\Omega \frac{2}{u} (\Delta u)^2 \, dx$$
$$= \int_\Omega \frac{1}{u^2} \Delta u (\nabla u)^2 \, dx - \int_\Omega \frac{2}{u} (\Delta u)^2 \, dx$$
$$= \int_\Omega \frac{1}{u} \Delta u D(u(t)) \, dx - R(t).$$

We can conclude the exponential decay.

4.3.2 Otto’s Approach

In the following, we introduce a different way to study the asymptotic behavior introduced by Otto [142]. Furthermore, the displacement convexity plays an important role in this approach.

Consider a partial differential equation, which allows for a self-similar solution of the form with $\alpha > 0$

$$u_*(t,x) = \frac{1}{t^\alpha} \hat{u}_* \left( \frac{x}{t^\alpha} \right).$$

Then we rescale the time and the space according to

$$x = t^\alpha, \quad t = \exp(\tau).$$

Then we can follow

$$u(t,x) = \frac{1}{t^\alpha} \hat{u} \left( \ln t, \frac{x}{t^\alpha} \right).$$
Let \( F(\hat{u}) = E(\hat{u}) + \alpha M_2(\hat{u}) \) be the augmented functional with second moment
\[
M_2(u) = \int_{\Omega} |x|^2u(x)dx.
\]
The idea in Otto’s approach is that solutions \( \hat{u} \) behave in the following way:

(i) \( \hat{u}_* \) satisfies the gradient flow equation
\[
\frac{\partial \hat{u}}{\partial \tau} = -\text{grad} F|_{\hat{u}}.
\]

(ii) \( \hat{u}_* \) is a minimiser of the energy functional \( EF \), i.e.
\[
F(\hat{u}) - F(\hat{u}_*) \geq 0, \forall \hat{u} \in \mathcal{M}.
\]
Furthermore it is a stationary solution \( 0 = -\text{grad} F|_{\hat{u}_*} \).

(iii) And furthermore
\[
\text{Hess} F|_{\hat{u}} \geq \alpha \text{Id}, \forall \hat{u} \in \mathcal{M}.
\]

Out of this assumptions, Otto derived the following inequalities out of simple Riemannian calculations [142],
\[
\frac{d}{d\tau} \left( \exp(2\alpha \tau)|\text{grad} F|^2|_{\hat{u}} \right) \leq 0, \quad (4.3.5)
\]
\[
\frac{d}{d\tau} \left( \exp(2\alpha \tau)(F(\hat{u}) - F(\hat{u}_*)) \right) \leq 0, \quad (4.3.6)
\]
\[
\frac{d}{d\tau} \left( \exp(2\alpha \tau)(W_2(\hat{u}, \hat{u}_*)^2) \right) \leq 0, \quad (4.3.7)
\]
Each inequality means, that the rate of convergence towards the stationary solution is exponential fast with rate \( \alpha \). In particular, the first inequality measures how far \( \hat{u} \) is from being a stationary point, the second measures how far \( \hat{u} \) is away to be the minimiser of \( F \) and the third inequality measures how far \( \hat{u} \) is from the stationary point \( \hat{u}_* \).
Chapter 5

The Logarithmic Entropy and the $L^2$-Regularisation

This chapter is concerned with two different entropy functionals, namely the logarithmic entropy (Section 5.1) and the $L^2$-regularisation (Section 5.2). We implement these functionals into the variational problem (3.1.1) and into the fluid dynamic formulation (3.2.11)-(3.2.14) and discuss existence and uniqueness. Additionally, we name the associated gradient flow equations and study the asymptotic behaviour. Finally, we compute self-similar solutions and provide numerical results.

5.1 The Logarithmic Entropy

In this section, we consider the logarithmic entropy functional

$$E(u) = \int_{\Omega} u \log u \, dx. \quad (5.1.1)$$

It has applications in the information theory, where it is known under the name Shannon entropy. There it measures the quantity of information contained in $u$, in the sense of the volume of possible microstates corresponding to the macrostate $u$. But furthermore, in thermodynamics the entropy is the H-functional for the Boltzmann equation, i.e. it decreases in time along solutions.

The derivation of the functional is based on stochastic theory. Assume we have $n$ boxes and $P$ balls, which are distributed over the boxes, so that every box contains at least one ball. The frequency is described by

$$u_i = \frac{A_i}{P} \quad \text{for } i = 1, \ldots, n,$$
where $A_i$ is the number of balls lying in the $i$–th box. We differ between a microscopic description, where the states of all particles are collected, and a macroscopic description, where the particles or balls are distinguished by their state. With elementary combinatorics, the microscopic configuration reads

$$W = \frac{P!}{A_1! \cdot \ldots \cdot A_n!},$$

which is the number of possible combinations of outcomes for the set of events $P$. The entropy is given through

$$e(u) = \log W.$$

After a simple calculation we obtain

$$e(u) = \log \left( \frac{P!}{A_1! \cdot \ldots \cdot A_n!} \right),$$

$$= \log(P!) - \log(A_1!) - \ldots - \log(A_n!)$$

$$= \sum_{i=1}^{P} \log i - \sum_{i=1}^{A_1} \log i - \ldots - \sum_{i=1}^{A_n} \log i.$$

We approximate the summations by integrals

$$e(u) = \int_{1}^{P} \log x \, dx - \int_{1}^{A_1} \log x \, dx - \ldots - \int_{1}^{A_n} \log x \, dx$$

$$= P \log P - P + 1 - (A_1 \log A_1 - A_1 + 1) - \ldots - (A_n \log A_n - A_n + 1)$$

$$= P \log P + 1 - \sum_{k=1}^{n} A_k \log A_k + (1 - n),$$

and divide through $P$. After neglecting the constant terms we obtain the discrete logarithmic entropy:

$$E(u) = - \sum_{k=1}^{n} \frac{A_k}{P} \log \frac{A_k}{P} = - \sum_{k=1}^{n} u_k \log u_k.$$

### 5.1.1 The Problem Formulation

Let $X = \Omega$ be an open and bounded subset of $\mathbb{R}^N$, $\nu \in \mathcal{P}_2(\Omega)$ a probability measure with finite second moment and $u$ be a probability density, i.e. $u \geq 0$ and $\int_{\Omega} u \, dx = 1$. We want to investigate the solution of the
Variational problem

\[ W_2^2(\nu, u L^N) + \epsilon \int_{\Omega} u \log u \, dx \rightarrow \min_{u L^N \in P(\Omega)}, \quad (5.1.2) \]

with \( L^N \) the usual Lebesgue measure. We consider here the logarithmic entropy regularisation \((5.1.1)\) multiplied by a parameter \( \epsilon > 0 \). In the sequel, we always consider the quadratic Euclidean distance \( c(x, y) = |x - y|^2 \), i.e. the optimisation problem:

\[ I(u) = \frac{1}{2} \int_{\Omega \times \Omega} |x - y|^2 d\pi(x, y) + \epsilon \int_{\Omega} u \log u \, dx \rightarrow \min_{\pi, u L^N}, \quad (5.1.3) \]

subject to

\[ \int_{A \times \Omega} d\pi(x, y) = \int_{A} d\nu(y), \quad (5.1.4) \]
\[ \int_{\Omega \times A} d\pi(x, y) = \int_{A} u(x) dx, \quad (5.1.5) \]

for all \( A \subset \Omega \) measurable. This optimisation problem can be transferred into a fluid dynamic framework (see Section 3.2):

Benamou-Brenier formulation

\[ \inf_{(u, \mu, v)} \left\{ \frac{1}{2} \int_{0}^{1} \int_{\Omega} |v|^2 d\mu(x) \, dt + \epsilon \int_{\Omega} u \log u \, dx \right\}, \quad (5.1.6) \]

subject to

\[ \partial_t \mu + \nabla \cdot (\mu v) = 0, \quad \text{(continuity equation)} \quad (5.1.7) \]
\[ \mu(t = 0) = \nu, \quad \text{(initial condition)} \quad (5.1.8) \]
\[ \mu(t = 1) = u L^N. \quad \text{(final condition)} \quad (5.1.9) \]

Furthermore, the heat equation constitutes the gradient flow for the logarithmic entropy with respect to the Wasserstein metric. This result was proven by Otto et al \cite{99}, following the idea of De Giorgi’s minimising movement scheme, which we introduced in Section 4.2.

Heat equation

\[ \partial_t u = \Delta u = \nabla \cdot (u \nabla (\log u + 1)), \quad \forall t > 0, x \in \Omega \quad (5.1.10) \]
\[ u(0, x) = u_0 > 0, \quad x \in \Omega. \]

Note again that from the point of view of gradient flow, the solution of the variational scheme is a discrete approximation of the solution of the gradient flow of \( E(u) \) with respect to the Wasserstein metric.
5.1.2 Existence and Uniqueness

The existence and uniqueness of solutions of the variational problem (5.1.2) are discussed extensively in [99]. We recall here the main ideas of the proof:

**Theorem 5.1.1.** Let \( X = \Omega \subset \mathbb{R}^N \) be an open and bounded domain with compact Lipschitz boundary. The probability measure \( \nu \in \mathcal{P}_2(\Omega) \) has second finite moment and \( u \) is a probability density on \( \Omega \). \( \mathcal{L}^N \) is the usual Lebesgue measure on \( \mathbb{R}^N \) and \( c(x, y) = |x - y|^2 \). Then, the optimisation problem

\[
\mathcal{J}(u) = \frac{1}{2} W_2(\nu, u \mathcal{L}^N)^2 + \epsilon(E(u) + M_2(u)) \to \min_{u \mathcal{L}^N \in \mathcal{P}_2(\Omega)},
\]

(5.1.11)

with \( E(u) = \int_{\Omega} u \log u \, dx \) and finite second moment \( M_2(u) = \int_{\Omega} |x|^2 u \, dx < \infty \) admits a unique minimiser \( u \mathcal{L}^N \in \mathcal{P}_2(\Omega) \) and \( u \in L^1(\Omega) \).

**Proof.** Let \( K \) be the set of admissible probability densities with finite second moment

\[
K = \left\{ u : \mathbb{R}^N \to [0, \infty) \text{ measurable} \mid u \geq 0, \int_{\Omega} u = 1, M_2(u) < \infty \right\}.
\]

The energy functional \( E(u) = \int_{\Omega} u \log u \, dx \) is well defined on \( K \) with values in \((-\infty, \infty]\), since

\[
\int_{\Omega} \min \{u \log u, 0\} = \int_{\Omega_0} u \log u \, dx + \int_{\Omega_1} u \log u \, dx \\
\leq C \int_{\Omega} \exp \left(-\frac{|x|}{2}\right) \, dx + \int_{\Omega} |x| \, u \, dx \\
\leq C \int_{\Omega} \exp \left(-\frac{|x|}{2}\right) \, dx + \alpha \int_{\Omega} |x|^2 \, u \, dx + \frac{1}{4\alpha} \int_{\Omega} u \, dx \\
\leq C \int_{\Omega} \exp \left(-\frac{|x|}{2}\right) \, dx + \alpha M_2(u)
\]

with \( \Omega_0 = \Omega \cap \{u \leq \exp(-|x|)\} \) and \( \Omega_1 = \Omega \cap \{\exp(-|x|) < u \leq 1\} \) and \( |z| \log |z| \leq C \sqrt{z} \) for \( C \in \mathbb{R} \) and \( |x| \leq \alpha |x|^2 + \frac{1}{4\alpha} \).

Furthermore, the functional \( \mathcal{J}(u) \) is bounded below, indeed

\[
\mathcal{J}(u) = \frac{1}{2} W_2(\nu, u \mathcal{L}^N)^2 + \epsilon(E(u) + M_2(u)) \geq \frac{1}{4} M_2(u) - \frac{1}{2} M_2(\nu) + \epsilon(E(u) + M_2(u)) \\
\geq \frac{1}{4} M_2(u) - \frac{1}{2} M_2(\nu) + C \int_{\Omega} \exp \left(-\frac{|x|}{2}\right) \, dx + c_1 M_2(u), \quad \text{with } c_1 \in \mathbb{R}.
\]

Let \((u^n)_{n \in \mathbb{N}}\) be a minimising sequence for \( \mathcal{J}(u) \), i.e.,

\[
\mathcal{J}(u^n) \to \inf_{u \mathcal{L}^N \in \mathcal{P}_2(\Omega)} \mathcal{J}(u) =: m \quad \text{as } n \to \infty,
\]
with $u^n\mathcal{L}^N \subset \mathcal{P}_2(\Omega)$. In particular, there exists a constant $M > 0$ such that
\[
M \geq J(u^n) = \frac{1}{2} W_2(\nu, u^n \mathcal{L}^N)^2 + \epsilon (E(u^n) + M_2(u^n)),
\]
and hence
\[
W_2(\nu, u^n \mathcal{L}^N) \leq M \text{ and } E(u^n) \leq M, \forall n \geq 1.
\]
Obviously, we have $E(u^n)$ bounded above and $M_2(u^n)$ bounded. As $z \mapsto \max\{z \log z, 0\}$ has superlinear growth and $E(u^n)$ is bounded above, this yields the existence of a subsequence, still denoted by $u^n$, and a $u \in K$, such that
\[
u^n \rightharpoonup u \text{ in } L^1(\Omega).
\]
From the convergence in $L^1(\Omega)$ we follow
\[
\int_\Omega u^n \psi \, dx \to \int_\Omega u \psi \, dx, \quad \forall \psi \in L^1(\Omega)
\]
\[
\iff \int_\Omega \psi \, du^n \mathcal{L}^N \to \int_\Omega \psi \, du \mathcal{L}^N, \quad \forall \psi \in L^1(\Omega).
\]
Since $C^0_b(\Omega) \subset L^p(\Omega)$ for all $p \geq 1$, it follows
\[
\int_\Omega \psi \, du^n \mathcal{L}^N \to \int_\Omega \psi \, du \mathcal{L}^N, \quad \forall \psi \in C^0_b(\Omega),
\]
and therefore $u^n \mathcal{L}^N \subset \mathcal{P}_2(\Omega)$ is narrowly convergent to $u \mathcal{L}^N \in \mathcal{P}_2(\Omega)$. We apply Proposition 2.2.2 and conclude the lower semicontinuity of the Wasserstein distance
\[
W_2(\nu, u \mathcal{L}^N)^2 \leq \liminf_{n \to \infty} \int_{\Omega^2} |x - y|^2 \, d\pi^n(x, y)
\]
\[
= \liminf_{n \to \infty} W_2(\nu, u^n \mathcal{L}^N)^2.
\]
Since $E(u)$ is lower-semicontinuous in $L^1(\Omega)$ we conclude
\[
\liminf_{n \to \infty} J(u^n) = \liminf_{n \to \infty} (W_2(\nu, u^n \mathcal{L}^N)^2 + E(u^n) + M_2(u^n))
\]
\[
\geq \frac{1}{2} W_2(\nu, u \mathcal{L}^N)^2 + \epsilon (E(u) + M_2(u)) = J(u).
\]
It follows that $J(u)$ has at least one minimiser $u \in L^1(\Omega)$ and $u \mathcal{L}^N \in \mathcal{P}_2(\Omega)$. The uniqueness of a minimiser follows from the convexity of $K$ and the strict convexity of $J(u)$. We can also apply Theorem 3.1.1 to prove uniqueness.
In addition, the uniqueness follows from the displacement convexity of the logarithmic entropy (cf. e.g. [9, 175] and Section 2.3.1), which can be proved with a simple calculation along the geodesics parametrised with $s$:

$$
\frac{dE}{ds}(u(s, x)) = \int_\Omega u_s(\log u + 1)dx = \int_\Omega \nabla \cdot (uv)(\log u + 1)dx
$$

$$
= -\int_\Omega uv\nabla(\log u + 1)dx = -\int_\Omega v\nabla u dx.
$$

And the second derivative of $E$ along the geodesic reads:

$$
\frac{d^2E}{ds}(u(s, x)) = \int_\Omega u_s\nabla \cdot v dx + \int_\Omega u\nabla \cdot (v_s)dx = \int_\Omega \nabla \cdot (uv)\nabla \cdot vdx + \int_\Omega u\Delta(\frac{1}{2}|v|^2)dx
$$

$$
= \int_\Omega u\nabla uv\nabla \cdot vdx + \int_\Omega u(\nabla \cdot v)^2dx + \int_\Omega u\Delta(\frac{1}{2}|v|^2)dx
$$

$$
= -\int_\Omega u\nabla \cdot (v\nabla \cdot v)dx + \int_\Omega u\nabla \cdot v)^2dx + \int_\Omega u\Delta(\frac{1}{2}|v|^2)dx
$$

$$
= -\int_\Omega uv\nabla(\nabla \cdot v)dx + \int_\Omega u\Delta(\frac{1}{2}|v|^2)dx
$$

$$
= \int_\Omega u(-v\nabla(\nabla \cdot v)dx + \Delta(\frac{1}{2}|v|^2))dx
$$

$$
= \int_\Omega utr(Dv)^2dx \geq \int_\Omega \frac{1}{n}u(\nabla \cdot v)^2dx,
$$

where we used Bochner’s formula for a space with Ricci curvature zero in the end

$$
\Delta \frac{1}{2}|v|^2 - v\nabla(\nabla \cdot v) = \text{tr}(Dv)^2. \quad (5.1.12)
$$

The positivity of the second derivative and therefore the displacement convexity follow from $\frac{1}{n}u \geq 0$.

5.1.3 Stationary Solutions

The literature on the long time behaviour of solutions for the heat equation (5.1.10) is vast (cf., e.g. [56, 127]). The heat equation allows for a self-similar solution of the form

$$
u_*(t, x) = \frac{1}{t^{\alpha}}\hat{\nu}_*\left(\frac{x}{t^{\alpha}}\right)
$$

with $\alpha = \frac{1}{2}$ where $\hat{\nu}_*$ is implicitly given through

$$
e'(\hat{\nu}_*(y)) = \ln \hat{\nu}_*(y) + 1 = \lambda - \frac{\alpha}{2}|y|^2,
$$
which implies
\[ \hat{u}_*(y) = C \exp \left( -\frac{\alpha}{2} |y|^2 \right). \] (5.1.13)
The constant \( C \) is computed such that \( \int \hat{u}_* = 1 \). The fundamental solution \( 5.1.13 \) was investigated by Barenblatt and Prattle \( 14, 148 \). Furthermore, the exponential convergence to the stationary solution is a consequence of Corollary \( 4.3.2 \) (see also \( 142 \)).

5.1.4 Self-Similar Solutions

Based on the calculations derived in Section \( 3.2.5 \) we compute here self-similar solutions for the logarithmic entropy \( 5.1.1 \). The goal is to provide a quick insight into the behaviour and structure of solutions. As we mentioned in Section \( 3.2.5 \) the solution \( u \) must fulfill the following equation \( 3.2.41 \)

\[ E'(u) = -\frac{|x|^2}{2\sigma^2} + \frac{\eta(x)}{\epsilon} - \frac{\eta_0}{\epsilon} \]

with \( \sigma = \sqrt{\frac{\epsilon}{a}} \). If we consider the logarithmic entropy \( E(u) = \int_\Omega u \log u \, dx \) as regularisation we obtain

\[ \log u(x) + 1 = -\frac{|x|^2}{2\sigma^2} + \frac{\eta(x)}{\epsilon} - \frac{\eta_0}{\epsilon}, \] (5.1.14)

and hence

\[ u(x) = \exp \left( -\frac{|x|^2}{2\sigma^2} + \frac{\eta(x)}{\epsilon} - \frac{\eta_0}{\epsilon} - 1 \right), \] (5.1.15)
Obviously, $u$ is positive everywhere, and because of the complementary condition $\eta u = 0$ this yields $\eta \equiv 0$ almost everywhere. The constant $\eta_0$ is the result of

$$\int_{\Omega} u(x) \, dx = C \int_{\Omega} \exp \left( -\frac{|x|^2}{2\sigma^2} \right) \, dx = 1, \quad \text{with } C = \exp \left( -\frac{\eta_0}{\epsilon} - 1 \right).$$

We substitute $|z|^2 = \frac{|x|^2}{2\sigma^2}$ and get

$$C \int_0^\pi \exp \left( -\frac{|z|^2}{2\sigma^2} \right) (2\sigma^2)^{N/2} \, dz = C (2\sigma^2 \pi)^{N/2} = 1.$$

We conclude $C = (2\sigma^2 \pi)^{-N/2}$ where $N$ denotes the dimension. Eventually, we follow the self-similar solution

$$u(x) = \frac{1}{(2\sigma^2 \pi)^{N/2}} \exp \left( -\frac{|x|^2}{2\sigma^2} \right), \quad \text{with } \sigma = \sqrt{\frac{\epsilon}{a}}, \quad (5.1.16)$$

which is a Gaussian curve with mean zero and variance $\sigma$. We illustrate the self-similar solution (5.1.16) for $a = 1$ in Figure 5.1. Noticing that $0 \leq a \leq 1$, we see that any Gaussian measure with variance greater equal $\sqrt{\epsilon}$ is obtained as a self-similar solution if the original measure is a Gaussian with variance $\sigma_0 = 1$. In the extreme case $a = 1$ the given measure $\nu$ is a point measure centered at zero.

**Example:**

For given initial data we can compute analytically the exact solution for different values for $\epsilon$. Let us exemplify this statement. We consider a the Gaussian curve with mean zero and variance one as initial data $\nu$. Since the solution $u$ results from dilation and translation of $\nu$ (see Section 3.2.5) it follows

$$\nu = \rho(0, x) = \frac{1}{b} u \left( \frac{x}{b} \right),$$

and here

$$\frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right) = \frac{1}{b} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \frac{a}{b^2} \right). \quad (5.1.17)$$

This equation is only true if $\frac{b^2 \epsilon}{a} = 1$. Together with $a + b = 1$ we obtain the following equation dependend only on $a$ and $\epsilon$:

$$a = 1 + \frac{1}{2\epsilon} \pm \frac{1}{2\epsilon} \sqrt{4\epsilon + 1}.$$

Hence, we can compute the value for $a$ for different $\epsilon$, i.e. the variance of the Gaussian curve. We illustrate the solutions for $\epsilon = 1, 10, 100$ in Figure 5.2(a). In addition, if we consider the same initial data for the numerical simulations, we can compare them with the analytical solutions to determine the correctness.
5.1.5 Numerical Results and Discussion

Figure 5.2: Numerical results for the logarithmic entropy: (a) Self-similar solution; (b) Solutions for the variational problem (5.1.2) computed with the algorithm (8.2.1)-(8.2.5); (c) Solutions for the Benamou-Brenier formulation (5.1.6)-(5.1.9) computed with the gradient descent algorithm (GDS) (8.3.7) with $\Delta t = 0.003, \tau = 1000$; (d) Solutions for the heat equation after certain time.

In Figure 5.2(b) we present numerical results solved with the linear algorithm (8.2.1)-(8.2.5) for the variational problem (5.1.2) in one dimension. We consider the Gaussian curve with mean zero and variance one as initial data, i.e. $\int u_0 = 1$, which is coloured in blue in Figure 5.2(b). The solutions for different regularisation parameters $\epsilon > 0$ are coloured in red, green and black. In addition, we present here numerical results for the Benamou-Brenier formulation (5.1.6)-(5.1.9) (see Figure 5.2(c)) and solutions for the heat equation (5.1.10) after certain timesteps (see Figure 5.2(d)). The Benamou-Brenier formulation is solved with a gradient descent scheme (GDS) (8.3.7) with $\Delta t = 0.005$, $\Delta x = 0.02$ and $\tau = 10000$. The heat equation is discretised with an explicit
Euler scheme and finite-difference quotients (see Section 8.4). Two dimensional results for the Benamou-Brenier formulation are presented in Figure 5.3. The regularisation parameter steers the influence of the regularisation term on the data, i.e. the larger $\epsilon$ is the larger is the effect of regularisation as we see in Figure 5.2. With the logarithmic entropy as regularisation term, we obtain isotropic smoothing of the data, similar to a Gaussian filter. Moreover, the results for the heat equation, see Figure 5.2(d), illustrate also the uniform melting respectively smoothing of the initial data. Furthermore, the presented numerical results coincide with the analytical computed solution presented in Figure 5.2(a), which justify the correctness of the algorithm.

![Figure 5.3: Two dimensional numerical results for the Benamou-Brenier formulation with logarithmic entropy $\rho(x) = -\log(x)$ solved with the gradient descent scheme (GDS) (8.3.7).](image)

(a) Initial data  
(b) $\epsilon = 5$  
(c) $\epsilon = 10$  
(d) $\epsilon = 50$
5.2 The $L^2$-regularisation

In addition, we consider in this chapter the $L^2$-regularisation

$$E(u) = \frac{1}{2} \int_\Omega u^2 dx,$$  \hspace{1cm} (5.2.1)

which is mostly used in inverse problems for image restoration and in statistics for regression analysis.

5.2.1 The Problem Formulation

In the following, we apply the $L^2$-regularisation to the optimisation problems already introduced in Chapter $\Box$. In particular, we want to investigate the solution of the regularised optimisation problem

$$W^2_2(\nu, u^{L^N}) + \epsilon \int_\Omega u^2 dx \to \min_{u^{L^N} \in \mathcal{P}(\Omega)},$$  \hspace{1cm} (5.2.2)

and with the quadratic Euclidean distance $c(x, y) = |x - y|^2$ this yields:

**Variational problem:**

$$I(u) = \frac{1}{2} \int_{\Omega \times \Omega} |x - y|^2 d\pi(x, y) + \epsilon \int_\Omega u^2 dx \to \min_{\pi, u},$$  \hspace{1cm} (5.2.3)

subject to

$$\int_{A \times \Omega} d\pi(x, y) = \int_A d\nu(y),$$  \hspace{1cm} (5.2.4)
$$\int_{\Omega \times A} d\pi(x, y) = \int_A u(x) dx,$$  \hspace{1cm} (5.2.5)

for all $A \subset \Omega$ measurable, where the parameter $\epsilon$ steers the impact of the regularisation term. Furthermore, the $L^2$-regularisation implemented in the fluid dynamic formulation reads:

**Benamou-Brenier formulation**

Minimise $\frac{1}{2} \int_0^1 \int_\Omega |v|^2 d\mu(x) \ dt + \epsilon \int_\Omega u^2 dx,$

where the minimum is taken over $(u, \mu, v)$, subject to

$$\partial_t \mu + \nabla \cdot (\mu v) = 0, \hspace{1cm} \text{(continuity equation)}$$  \hspace{1cm} (5.2.7)
$$\mu(t = 0) = \nu, \hspace{1cm} \text{(initial condition)}$$  \hspace{1cm} (5.2.8)
$$\mu(t = 1) = u^{L^N}. \hspace{1cm} \text{(final condition)}$$  \hspace{1cm} (5.2.9)
As a third formulation we state here the porous medium equation which constitutes the gradient flow for the $L^2$-regularisation with respect to the Wasserstein distance. For the proof we refer to [142] and to Section 4.2.2.

Porous medium equation:

$$
\partial_t u = \Delta u^2 = \nabla \cdot (2u\nabla u), \quad \forall t > 0, x \in \Omega \quad (5.2.10)
$$

$$
u(0, x) = u_0 > 0, \quad x \in \Omega.
$$

5.2.2 Existence and Uniqueness

Remark 5.2.1. To state the existence of a minimiser for (5.2.2) we can adapt the proof given in Section 5.1.2 for the logarithmic entropy. With the same arguments it follows that (5.2.2) admits a minimiser $u \in L^2(\Omega)$ and $u\mathcal{L}^N \in \mathcal{P}(\Omega)$. Furthermore the uniqueness of a minimiser can be deduced from the convexity of (5.2.2) and of the set of admissible solutions, as well as from Theorem 3.1.1 since $E$ is strictly convex.

In the following we recall the calculations for the displacement convexity for the $L^2$-regularisation (5.2.1) which is also discussed in Section 2.3 and [9, 175]. We compute the first derivative of $E$ along the geodesic parametrised with $s$:

$$
\frac{dE}{ds} u(s, x) = \int_{\Omega} u_s u \, dx = \int_{\Omega} \nabla \cdot (uv) u \, dx \\
= - \int_{\Omega} u \nabla uv \, dx.
$$
With \( f' = u \) we compute the second derivative
\[
\frac{d^2 E}{ds} = \int f'u_s \nabla \cdot v + \int f \nabla \cdot (v_s) = \int f' \nabla \cdot (uv) \nabla \cdot v + \int f \Delta \left( \frac{1}{2} |v|^2 \right)
\]
\[= \int \nabla f v \nabla \cdot v + \int f'u(\nabla \cdot v)^2 + \int f \Delta \left( \frac{1}{2} |v|^2 \right) \]
\[= - \int f \nabla \cdot (v \nabla \cdot v) + \int f'u(\nabla \cdot v)^2 + \int f \Delta \left( \frac{1}{2} |v|^2 \right) \]
\[= \int (f' - f)(\nabla \cdot v)^2 + \int f (-v \Delta v + \Delta \left( \frac{1}{2} |v|^2 \right)) \]
\[= \int (f' - f)(\nabla \cdot v)^2 + \int f \text{tr}(Dv)^2 \]
\[= \int (u^2 - \frac{1}{2} u^2)(\nabla \cdot v)^2 + \int \frac{1}{2} u^2 \text{tr}(Dv)^2 \]
\[= \int \frac{1}{2} u^2(\nabla \cdot v)^2 + \int \frac{1}{2} u^2 \text{tr}(Dv)^2 \]

In the end we used Bochner’s formula
\[\Delta \left( \frac{1}{2} |v|^2 \right) - v \nabla \cdot (\nabla \cdot v) = \text{tr}(Dv)^2.\]  
(5.2.11)

With \((\nabla \cdot v)^2 \leq n \text{tr}(Dv)^2\) the displacement convexity follows from \(1 + \frac{1}{n} u^2 \geq 0\).

### 5.2.3 Stationary Solutions

That the porous medium equation can be seen as a gradient flow of the \(L^2\)-regularisation with respect to the Wasserstein distance was proven by Otto in 2001 [142]. This gradient flow formulation yields on the one hand to new asymptotic results and on the other hand we can use the numerical results for comparison.

In a general form the porous medium equation reads
\[
\frac{\partial u}{\partial t} = \frac{1}{m-1} \Delta u^m, \quad \forall t \geq 0, x \in \Omega. \tag{5.2.12}
\]

Note that for the special case \( m = 1 \) we obtain the heat equation and for \( m < 1 \) the fast diffusion equation. There exists vast literature for the porous medium equation. We refer for a detailed introduction to the books of Vázquez [173, 170, 171, 172]. The porous medium equation allows for a self-similar solution of the form
\[
u_*(t,x) = \frac{1}{t^\alpha} \hat{u}_* \left( \frac{x}{t^\alpha} \right),
\]
where \( \hat{u}_* \) is implicitly given for \( m > 1 \) through

\[
e'(\hat{u}_*(y)) = \frac{m}{m-1} \hat{u}_*(y)^{m-1} = \max \left\{ \lambda - \frac{1}{2} \|y\|^2, 0 \right\},
\]

and hence

\[
\hat{u}_*(y) = \left( C - \frac{\alpha(m-1)}{2m} \|y\|^2 \right)^{1/(m-1)}
\]

(5.2.13)

with \( \alpha = \frac{1}{N(m-1)+2} \). The constant \( C \) is calculated such that \( \int \hat{u}_* = 1 \). The profiles \( \hat{u}_* \) are known as the Barenblatt-Prattle solutions, since they were first introduced by Barenblatt and Prattle in the 50’s [138, 14]. Hence, the long time asymptotic for the porous medium equation is described by the Barenblatt-Prattle solutions and furthermore it was proven that the profiles converge exponentially fast to the stationary solution (cf. [57]). This is also the statement of the following theorem:

**Theorem 5.2.2.** [142, Section 5.2] Let \( m \) satisfy

1. \( m > \frac{N}{N+2} \): (which guarantees the finitness of the second moment.)
2. \( m \geq 1 - \frac{1}{N} \): (which guarantees the displacement convexity of the equation.)

Let \( u \) be a weak solution of the porous medium equation with initial data \( \rho_0 \). Additionally let

\[
\int \rho_0 = 1, \quad \text{and} \quad \int \rho_0 \frac{1}{2} \|x\|^2 < \infty.
\]

We consider a function \( u \) on \( (-\infty, \infty) \times \mathbb{R}^N \) given by rescaled time and space according to

\[
x = t^\alpha, \quad t = \exp(\tau),
\]

\[
u(t, x) = \frac{1}{t^N \alpha} \hat{u} \left( \ln t, \frac{x}{t^\alpha} \right),
\]

where \( \alpha = \frac{1}{(m-1)N+2} \). Then, in a distributional sense \( \hat{u} \) fullfills the inequalities (4.3.5)- (4.3.7), which states the exponential decay to a stationary solution.

### 5.2.4 Self-Similar Solutions

We derive in the following the self-similar solutions for the \( L^2 \)-regularisation \( E(u) = \frac{1}{2} \int_{\Omega} |u|^2 dx \). In this case, the condition \( (3.2.41) \) just becomes

\[
u(x) = -\frac{|x|^2}{2\sigma^2} + \frac{\eta(x)}{\epsilon} - \frac{\eta_0}{\epsilon} \quad \text{with} \quad \sigma = \sqrt{\frac{\epsilon}{a}}.
\]
Figure 5.4: Self-similar solution for the $L^2$-regularisation (5.2.1) with $a = 1$ in (5.2.17).

We restrict our computations to positive solutions $u(x) \geq 0$ and therefore $\eta = 0$ in the neighborhood of zero

$$u(x) = \left(-\frac{|x|^2}{2\sigma^2} - \frac{\eta_0}{\epsilon}\right)_+.$$  

The constant $\eta_0$ can be easily computed in one dimension for a solution $u$ symmetrically centered around zero using the condition $\int_\Omega u \, dx = 1$

$$\int_\Omega u(x) \, dx = \int_{-x_0}^{x_0} \left(-\frac{|x|^2}{2\sigma^2} - \frac{\eta_0}{\epsilon}\right)_+ \, dx = -\frac{1}{2\sigma^2} \int_{-x_0}^{x_0} |x|^2 \, dx - \frac{\eta_0}{\epsilon} \int_{-x_0}^{x_0} 1 \, dx = -\frac{1}{2\sigma^2} \left[\frac{1}{3} x^3\right]_{-x_0}^{x_0} - \frac{\eta_0}{\epsilon} \left[x\right]_{-x_0}^{x_0} = -\frac{1}{3\sigma^2} x_0^3 - \frac{2\eta_0}{\epsilon} x_0 = 1.$$  

(5.2.14)

Further, $u$ is a polynomial of order 2, so it has two zeros at most. We assume $u$ to be symmetrically centered around zero, and hence we only need to compute

$$u(x_0) = -\frac{x_0^2}{2\sigma^2} - \frac{\eta_0}{\epsilon} = 0,$$

which yields $x_0 = \sqrt{-2\sigma^2 \frac{\eta_0}{\epsilon}}$. Then, (5.2.14) reads

$$-\frac{1}{3\sigma^2} \left(\sqrt{-2\sigma^2 \frac{\eta_0}{\epsilon}}\right)^3 - \frac{2\eta_0}{\epsilon} \sqrt{-2\sigma^2 \frac{\eta_0}{\epsilon}} = 1,$$

(5.2.15)

and we obtain after a simple calculation

$$\eta_0 = -\frac{\epsilon \sqrt{3}}{(\sigma 4\sqrt{2})^{2/3}}.$$  

(5.2.16)
Then, we conclude for the self-similar solution

\[ u(x) = \left( -\frac{|x|^2}{2\sigma^2} + \frac{\sqrt{9}}{(\sigma\sqrt{2})^{2/3}} \right)_+. \tag{5.2.17} \]

Hence, the solution behaves like the Barenblatt-Prattle solution. We illustrate (5.2.17) in Figure 5.4 for \( a = 1 \) and different \( \epsilon \).

### 5.2.5 Numerical Results and Discussion

![Numerical Results](image)

Figure 5.5: Numerical results for the \( L^2 \)-regularisation solved with the Kantorovich algorithm (8.2.1)-(8.2.5); \( u_0 \)-initial data, \( u \)-solution, \( \epsilon \)-regularisation parameter.

In the following, we show numerical results for the optimisation problems introduced in Section 5.2.1. We refer to Chapter 8 for the explanation of numerical algorithms. In Figure 5.5 we present results for the one dimensional linear program (8.2.1)-(8.2.5) with the \( L^2 \)-regularisation for different values for \( \epsilon \). We consider the Gaussian curve with mean zero and variance one as initial data, which is coloured in blue in Figure 5.5. The solutions for different regularisation parameters \( \epsilon \) are coloured in red. In Figure 5.6(a) we present numerical results obtained with the dual ascent scheme (DAS) (8.3.13) for the \( L^2 \)-regularisation in one dimension. To justify the correctness of the results we present in Figure 5.6(b) the results for the porous medium equation after different timesteps. Two dimensional results of the Benamou-Brenier problem are
presented in Figure 5.7 and 5.8. We consider a Gauss-curve as well as a rectangle as initial data. In addition, we contrast results obtained with the logarithmic entropy and the $L^2$-regularisation in Figure 5.8.

The solutions have the well known Barenblatt-Prattle profile, especially they have compact support. For different regularisation parameters we obtain different degrees of regularisation. The difference between the impact of the logarithmic entropy and the $L^2$-regularisation is obvious (see Figure 5.8). Whereas the numerical results for the logarithmic entropy are bell shaped the solutions for the $L^2$-regularisation are compact supported.

Figure 5.6: Numerical results for the $L^2$-regularisation: (a) Solutions for the Benamou-Brenier formulation solved with the dual ascent scheme (DAS) (8.3.13) with $\Delta x = 0.25, \Delta t = 0.3, \tau = 0.002$; (b) Solutions for the porous medium equation after certain timesteps.
Figure 5.7: Two dimensional numerical results for the logarithmic entropy (5.1.1) with (GDS) (8.3.7).

Figure 5.8: Two dimensional numerical results for the Benamou-Brenier formulation computed with (GDS) (8.3.7). First row: Initial data; Two results for the logarithmic entropy with \( \epsilon = 10, 50 \). Second row: Initial data; Two results for the \( L^2 \)-regularisation with \( \epsilon = 10, 50 \).
Chapter 6

The Gradient Regularisation

This chapter is devoted to the discussion of the regularisation energies that involve derivatives, e.g., the Dirichlet regularisation in Section 6.1, the Fisher information in Section 6.2, and the TV-regularisation in Section 6.3. In particular, we implement these regularisation energies into the variational problem (3.1.1) and into the Benamou-Brenier formulation (3.2.11)-(3.2.14) and examine existence and uniqueness as well as self-similar solutions. Additionally, we consider the associated gradient flow equations and study the asymptotic behaviour. Finally, we provide numerical results.

6.1 The Dirichlet Regularisation

In this section we consider the Dirichlet functional

$$E(u) = \int_{\Omega} |\nabla u|^2 \, dx,$$

(6.1.1)

as a regularisation term in the variational problem (3.1.1).

6.1.1 The Problem Formulation

Let $X = \Omega$ be an open and bounded subset of $\mathbb{R}^N$, $\nu \in \mathcal{P}_2(\Omega)$ a probability measure and $u$ be a probability density in $\Omega$. The aim is to find the solution of the

Variational problem

$$W_2(\nu, u\mathcal{L}^N)^2 + \frac{\epsilon}{2} \int_{\Omega} |\nabla u|^2 \, dx \rightarrow \min_{u\mathcal{L}^N \in \mathcal{P}(\Omega)}$$

(6.1.2)
Again, we consider here the quadratic Euclidean distance $c(x, y) = |x - y|^2$ and obtain:

$$I(u) = \frac{1}{2} \int_{\Omega \times \Omega} |x - y|^2 d\pi(x, y) + \frac{\epsilon}{2} \int_{\Omega} |\nabla u|^2 dx \to \min_{\pi, u} \quad (6.1.3)$$

subject to

$$\int_{A \times \Omega} d\pi(x, y) = \int_{A} d\nu(y), \quad (6.1.4)$$

$$\int_{\Omega \times A} d\pi(x, y) = \int_{A} u(x) dx, \quad (6.1.5)$$

for all $A \subset \Omega$ measurable. This optimisation problem can also be transferred into a fluid dynamic framework (see Section 3.2):

**Benamou-Brenier formulation**

Minimise $\frac{1}{2} \int_{0}^{1} \int_{\Omega} |v|^2 d\mu(x) \ dt + \frac{\epsilon}{2} \int_{\Omega} |\nabla u|^2 dx$, \quad (6.1.6)

where the minimum is taken over $(u, \mu, v)$, subject to

$$\partial_t \mu + \nabla \cdot (\mu v) = 0, \quad \text{(continuity equation)} \quad (6.1.7)$$

$$\mu(t = 0) = \nu, \quad \text{(initial condition)} \quad (6.1.8)$$

$$\mu(t = 1) = u L^N. \quad \text{(final condition)} \quad (6.1.9)$$

In addition, the associated Wasserstein gradient flow equation is the thin film equation: Let $X = \Omega$ be an open and bounded subset of $\mathbb{R}^N$, $\nu \in \mathcal{P}_2(\Omega)$ a probability measure and $u$ be a probability density in $\Omega$. The aim is to find the solution of the

**Thin film equation:**

$$\partial_t u = -\nabla \cdot (u \nabla \Delta u) = 0, \quad \forall x \in \Omega, \forall t > 0, \quad (6.1.10)$$

$$u(0, \cdot) = u_0 \geq 0. \quad \forall x \in \Omega, \quad (6.1.11)$$

which was proven by Matthes, McCann and Savaré [125] by following the idea of the minimising movement scheme introduced in Chapter 4. Indeed, the thin film equation (6.1.10) can formally be written as a Wasserstein gradient flow system (4.1.5)-(4.1.7)

$$\partial_t u + \nabla \cdot (u \nabla \Delta u) = 0,$$

$$uv = -u \nabla \Delta u,$$

$$\psi = -\Delta u,$$

where the first variation of $E$ is given by

$$\frac{\delta E}{\delta u} = -\Delta u,$$

and this implies $E(u) = \frac{1}{2} \int_{\Omega} |\nabla u|^2 dx$. 
6.1.2 Existence and Uniqueness

In the following, we state the proof for the existence of the variational problem (6.1.2).

**Theorem 6.1.1.** Let \( X = \Omega \subset \mathbb{R}^N \) be an open and bounded domain with compact Lipschitz boundary. The probability measure \( \nu \in \mathcal{P}_2(\Omega) \) has second finite moment and \( u \) is a probability density on \( \Omega \). \( \mathcal{L}^N \) is the usual Lebesgue measure on \( \mathbb{R}^N \) and \( c(x, y) = |x - y|^2 \). Then, the optimisation problem

\[
J(u) = \frac{1}{2} W_2(\nu, u \mathcal{L}^N)^2 + \epsilon (E(u) + M_2(u)) \to \min_{u \mathcal{L}^N \in \mathcal{P}_2(\Omega)}, \tag{6.1.12}
\]

with \( E(u) = \int_\Omega |\nabla u|^2 \, dx \) and finite second moment \( M_2(u) = \int_\Omega |x|^2 u \, dx < \infty \) admits a unique minimiser \( u \mathcal{L}^N \in \mathcal{P}_2(\Omega) \) and \( u \in H^1(\Omega) \).

**Proof.** Let \( K \) be the set of admissible probability densities with finite second moment

\[
K = \left\{ u \in H^1(\Omega) \left| \int_\Omega u = 1, u \geq 0, M_2(u) < \infty \right. \right\}.
\]

Then the energy functional \( E(u) \) is well defined on \( K \) and also bounded below. Let \((u^n)_{n \in \mathbb{N}}\) be a minimising sequence for \( J(u) \), i.e.,

\[
J(u^n) \to \inf_{u \mathcal{L}^N \in \mathcal{P}_2(\Omega)} J(u) =: m \quad \text{as} \quad n \to \infty,
\]

with \( u^n \mathcal{L}^N \subset \mathcal{P}_2(\Omega) \). In particular there exists a constant \( M > 0 \) such that

\[
M \geq J(u^n) = \frac{1}{2} W_2(\nu, u^n \mathcal{L}^N)^2 + \epsilon (E(u^n) + M_2(u^n)),
\]

and hence

\[
W_2(\nu, u^n \mathcal{L}^N) \leq M \quad \text{and} \quad E(u^n) \leq M, \quad \forall n \geq 1.
\]

Then we deduce that the minimising sequences are bounded in \( H^1(\Omega) \) and this yields to the existence of a subsequence, still denoted by \( u^n \), and a \( u \in K \), such that

\[
u^n \rightharpoonup \nu \quad \text{and strongly in} \quad L^2(\Omega) \quad \text{and in} \quad L^1(\Omega).
\]

Then we can apply Proposition 2.2.2 and continue with the lower semicontinuity of the Wasserstein distance

\[
W_2(\nu, u \mathcal{L}^N)^2 \leq \liminf_{n \to \infty} \int_{\Omega^2} |x - y|^2 \, d\pi_n(x, y)
\]

\[
= \liminf_{n \to \infty} W_2(\nu, u^n \mathcal{L}^N)^2.
\]
Since $E(u)$ is lower-semicontinuous on $H^1(\Omega)$ we conclude
\[\liminf_{n \to \infty} J(u^n) \geq \frac{1}{2} W_2(\nu, u \mathcal{L}^N)^2 + \epsilon (E(u) + M_2(u)) = J(u).\]
It follows that $J(u)$ has at least one minimiser $u \in H^1(\Omega)$ and $u \mathcal{L}^N \in \mathcal{P}_2(\Omega)$. The uniqueness of a minimiser follows from the strict convexity of $J(u)$.

In the following, we discuss the stability of the evolution equation along the geodesic. In general, we follow the linear stability from $E'' > 0$. Since we are here on a manifold, we compute the stability along geodesics. We state the following theorem:

**Theorem 6.1.2.** The unique minimiser $u_\infty$ of the Dirichlet functional
\[ E(u) = \frac{1}{2} \int_\Omega |\nabla u|^2 + \frac{1}{2} \int_\Omega |x|^2 u \, dx, \]
is a stable critical point for the energy functional.

**Proof.** To obtain the stability result, we compute the second derivative of $E$ along the geodesics $\partial_s u = \nabla \cdot (uv)$ for an arbitrary vector field or perturbation $v$ and evaluate it at the point $u_\infty$:
\[
\frac{dE}{ds} = \int_\Omega \nabla u_s \nabla u \, dx + \frac{1}{2} \int_\Omega |x|^2 u_s \, dx = \int_\Omega \nabla (\nabla \cdot (uv)) \nabla u \, dx + \frac{1}{2} \int_\Omega |x|^2 u_s \, dx
\[
= -\int_\Omega \nabla u \nabla \cdot (uv) \, dx - N \int_\Omega v \cdot xu \, dx.
\]
where $N$ denotes the spatial dimension. The second derivative with $v_s = \frac{1}{2} \nabla |v|^2$ reads:
\[
\frac{d^2 E}{ds^2} = -\int_\Omega \Delta u_s \nabla \cdot (uv) \, dx - \int_\Omega \Delta u \nabla \cdot (u_s v) \, dx - \int_\Omega \Delta u \nabla \cdot (uv_s) \, dx
\]
\[- \int_\Omega N (u_s v + uv_s) \, dx
\]
\[
= -\int_\Omega \Delta (\nabla \cdot (uv)) \nabla \cdot (uv) \, dx - \int_\Omega \Delta u \nabla \cdot (\nabla \cdot (uv)) \, dx - \int_\Omega \Delta u \nabla \cdot (u \frac{1}{2} \nabla |v|^2) \, dx
\]
\[- N \int_\Omega \nabla \cdot (uv) v \cdot x dx - N \int_\Omega (\frac{1}{2} \nabla |v|^2) u x \, dx
\]
\[
= \int_\Omega \nabla (\nabla \cdot (uv)) \cdot \nabla (\nabla \cdot (uv)) \, dx + \int_\Omega v \cdot \nabla (\Delta u) \nabla \cdot (uv) \, dx + \int_\Omega (\nabla (\Delta u) \cdot v) u \nabla \cdot v \, dx
\]
\[+ N \int_\Omega (\nabla (x \cdot v) \cdot v) \, dx - N \int_\Omega x \cdot vu \nabla \cdot v \, dx
\]
\[
= \int_\Omega \nabla (\nabla \cdot (uv)) \cdot \nabla (\nabla \cdot (uv)) \, dx - \int_\Omega v \cdot \nabla (\nabla \Delta u) \, uv \, dx + N^2 \int_\Omega u |v|^2 \, dx
\]
Hence, the second derivative evaluated in $u_\infty$ reads
\[
\frac{d^2E}{ds^2} = \int_\Omega \nabla(\nabla \cdot (u_\infty v)) \cdot \nabla(\nabla \cdot (u_\infty v)) \, dx - \int_\Omega v \cdot \nabla(\nabla \Delta u_\infty) u_\infty v \, dx + N^2 \int_\Omega u_\infty |v|^2 \, dx.
\]
The first term as well as the third term are positive, so we have only to determine the value of the second term. Therefore, we use the property that $u_\infty$ is the minimiser of the energy $E$ and must fulfill
\[
-\Delta u_\infty + \frac{1}{2} |x|^2 = 0.
\]
Hence, we compute
\[
\nabla(\nabla \Delta u_\infty) = N^2 \text{Id},
\]
and conclude
\[
\frac{d^2E}{ds^2} > 0
\]
which confirms the stability of the evolution equation. \qed

### 6.1.3 Stationary Solutions

As mentioned above, the thin film equation
\[
\begin{align*}
\partial_t u &= -\nabla \cdot (u \nabla \Delta u) = 0, \quad \forall x \in \Omega, \forall t > 0, \quad (6.1.13) \\
u(0, .) &= u_0 \geq 0, \quad \forall x \in \Omega \quad (6.1.14)
\end{align*}
\]
is a gradient flow equation for the Dirichlet regularisation $\text{(6.1.1)}$ with respect to the Wasserstein metric \[125\]. Furthermore, computing the time derivative of $E$ along regular solutions $u \geq 0$ yields
\[
\frac{dE}{dt} u(t, x) = \int_\Omega \nabla u \nabla \partial_t u \, dx = -\int_\Omega \nabla u \nabla \cdot (u \nabla \Delta u) \, dx \\
= \int_\Omega \Delta u (\nabla \cdot (u \nabla \Delta u)) \, dx = -\int_\Omega (\nabla \Delta u)^2 u \, dx \leq 0. \quad (6.1.15)
\]
Hence, the energy $E$ is nonincreasing along solutions in time. The thin film equation was first studied by Friedmann and Bernis \[21\] and since then a lot of research has been done on this topic. However, the connection to the Wasserstein gradient flow was first given in \[65, 125\] and in the one dimensional case, where the equation generates the Hele Shaw flow, in \[141\]. We refer for more results to \[133\]. Furthermore, existence is studied by Beretta, Bertsch and Dal Passo \[20\] and also by \[101, 105, 46\]. In particular, the long time behaviour and the finite time blow up are discussed in \[20, 26, 22\].
This highly nonlinear fourth order equation does not fulfill a maximum principle or a comparison principle, i.e. we cannot assume unique nonnegative solutions, although the initial data is nonnegative. In particular, the proof of nonnegative solutions is still an open problem. Furthermore the solution is not bounded by the maximum or the minimum of the initial data.

However, the thin film equation allows for a self-similar solution. In particular, the long time behaviour is described by the Barenblatt-Pattle profiles \([14, 148, 65]\)

\[
\hat{u}_*(x) = \left( C - \frac{1}{2\sqrt{2(N+2)|x^2|}} \right)^2 .
\] (6.1.17)

### 6.1.4 Self-Similar Solutions

![Self-similar solution u(x) for squared Sobolev regularization](image)

Figure 6.1: Self-similar solution for the Dirichlet-regularisation (6.1.1) with \(a = 1\) in (6.1.21).

In the following, we compute the self-similar solution for the optimal transport problem with Dirichlet regularisation \(E(u) = \frac{1}{2} \int_\Omega |\nabla u(x)|^2 dx\). Implementing the first variation of \(E\) into (3.2.41) yields

\[
- \Delta u = -\frac{|x|^2}{2\sigma^2} + \frac{\eta(x)}{\epsilon} - \frac{\eta_0}{\epsilon} .
\] (6.1.18)

For the one-dimensional case and for \(u(x) > 0\) almost everywhere we simplify (6.1.18)
to
\[ u''(x) = \frac{x^2}{2\sigma^2} + \frac{\eta_0}{\epsilon}, \]
\[ u'(x) = \frac{x^3}{6\sigma^2} + Cx + \tilde{C}, \]
\[ u(x) = \frac{x^4}{24\sigma^2} + \frac{C}{2}x^2 + \tilde{C}x + \bar{C}, \]
with constants \( C = \frac{\eta_0}{\epsilon}, \tilde{C} \) and \( \bar{C} \). We compute the constants using the property that the solution is centered at zero and symmetric, i.e.
\[ u(-x_0) = u(x_0) = 0, \quad u'(-x_0) = u'(x_0) = 0. \]

With \( u(x_0) - u(-x_0) = 2\tilde{C}x_0 = 0 \) we get \( \tilde{C} = 0 \). Further,
\[ u'(x_0) = \frac{x_0^3}{6\sigma^2} + Cx_0 = 0, \]
and because of \( x_0 \neq 0 \) we obtain \( x_0 = \sqrt{-\frac{\eta_0}{\epsilon}6\sigma^2} \). With
\[ u(x_0) = \frac{x_0^4}{24\sigma^2} + \frac{C}{2}x_0^2 + \bar{C} = 0, \quad (6.1.19) \]
we conclude
\[ \bar{C} = \frac{3}{2\epsilon}\eta_0^2\sigma^2. \quad (6.1.20) \]
The constant \( C \) results from \( \int_{-x_0}^{x_0} u(x)dx = 1 \):
\[
\int_{-x_0}^{x_0} u(x)dx = \int_{-x_0}^{x_0} \left( \frac{x^4}{24\sigma^2} + \frac{C}{2}x^2 + \bar{C} \right) dx
\]
\[ = \frac{1}{120\sigma^2} \left[ x^5 \right]_{-x_0}^{x_0} + \frac{C}{6} \left[ x^3 \right]_{-x_0}^{x_0} + \bar{C} \left[ x \right]_{-x_0}^{x_0}
\]
\[ = \frac{1}{60\sigma^2}x_0^5 + \frac{C}{3}x_0^3 + 2\bar{C}x_0 = 1. \]
Hence,
\[ C = -\frac{1}{\epsilon}\sqrt{\frac{25a^3\epsilon^2}{384}}, \]
and eventually we obtain the self-similar solution
\[ u(x) = \left( \frac{x^4}{24\sigma^2} - \frac{1}{2\epsilon}\sqrt{\frac{25a^3\epsilon^2}{384}}x^2 + \frac{3}{2\epsilon a}\eta_0^2 \right). \quad (6.1.21) \]
We illustrate the result for \( a = 1 \) and different \( \epsilon \) in Figure 6.1.
6.1.5 Numerical Results and Discussion

In the following present numerical results for the variational problem (6.1.2) and the formulations introduced above in one dimension as well as in two dimensions. We consider the Gaussian curve with mean zero and variance one as initial data $\nu$, i.e. $\int_{\Omega} \nu dx = 1$. In Figure 6.2 we show one dimensional results for the variational problem (6.1.2) with Dirichlet regularisation solved with the algorithm (8.2.1)-(8.2.5). We consider different regularisation parameters $\epsilon$ which results in different degrees of regularisation. To confirm the correctness of the numerical results, we present also the solutions for the Benamou-Brenier formulation (6.1.6)-(6.1.9) in one dimension (see Figure 6.3(a)). We solved the Benamou-Brenier problem with the dual ascent scheme (DAS) (8.3.13). Moreover, we present solutions for the thin film equation (6.1.10) after certain time steps in Figure 6.3(b).

In addition, we provide numerical simulations on real life data, which justifies that our model can be applied to density estimation problems in a simple way. In particular, in Figure 6.4 we present numerical results for the epicentral distribution of earthquakes in L’Aquila (Italy) in the time interval of aftershock sequence from 6 April 2009 to 30 June 2009. The data are provided by the Istituto Nazionale di Geofisica e Vulcanologia [3]. The values are zero in distributional regions, without aftershock events and oth-
Figure 6.3: Numerical results for the Dirichlet-regularisation: (a) Computed with (DAS) (8.3.13) with $\Delta x = 0.25, \Delta t = 0.3, \tau = 0.002$; (b) Time discrete solutions for the thin film equation.

Figure 6.4: (a) Initial data: Epicentral distribution of earthquakes in the time interval of aftershock sequence from 6 April 2009 to 30 June 2009. (source INGV: http://bollettinosismico.rm.ingv.it); (b) Results for (3.1.1) with Dirichlet regularisation solved with (GDS) (8.3.7), $\epsilon = 10$. 

otherwise one. The result computed for $\epsilon = 10$ is illustrated in Figure 6.4(b). The latter constitutes an estimate of the continuous distribution of aftershocks based on the few measured samples in Figure 6.4(a).
6.2 The Fisher Information

This section is concerned with the Fisher information

\[ E(u) = \frac{1}{2} \int_{\Omega} u \left| \nabla \log u \right|^2 dx, \tag{6.2.1} \]

implemented in the variational problem (3.1.1) and equivalent formulations. Under some regularity assumptions on \( u \) the energy (6.2.1) can also be written as (see [84])

\[ E(u) = 2 \int_{\Omega} |\nabla \sqrt{u}|^2 dx = \frac{1}{2} \int_{\Omega} \frac{|\nabla u|^2}{u} dx = -\frac{1}{2} \int_{\Omega} u \Delta (\log u) dx. \]

Further equivalent formulations can be found in [84].

The Fisher information was introduced by the statistician R. Fisher [80] and has applications in the mathematical statistic and information theory.

6.2.1 The Problem Formulation

Let \( \Omega \) be an open and bounded subset of \( \mathbb{R}^N \), \( \nu \in \mathcal{P}_2(\Omega) \) a probability measure and \( u \) a probability density in \( \Omega \). In this section, we want to investigate the solution of the variational problem (3.1.1) with the Fisher information (6.2.1), which reads

\[
\text{Variational problem:}
W_2(\nu, u \mathcal{L}^N)^2 + \frac{\epsilon}{2} \int_{\Omega} u |\nabla \log u|^2 dx \to \min_{u \mathcal{L}^N \in \mathcal{P}(\Omega)}, \tag{6.2.2}
\]

with quadratic Euclidean distance \( c(x, y) = |x - y|^2 \). This optimisation problem can also be transferred into a fluid dynamic framework (see Section 3.2):

\[
\text{Benamou-Brenier formulation}
\begin{align*}
\text{Minimise} & \quad \frac{1}{2} \int_0^1 \int_{\Omega} |v|^2 d\mu(x) \, dt + \frac{\epsilon}{2} \int_{\Omega} u |\nabla \log u|^2 dx, \tag{6.2.3} \\
\text{subject to} & \quad \partial_t \mu + \nabla \cdot (\mu v) = 0, \quad \text{(continuity equation)} \\
& \quad \mu(t = 0) = \nu, \quad \text{(initial condition)} \\
& \quad \mu(t = 1) = u \mathcal{L}^N, \quad \text{(final condition)} \quad \tag{6.2.4} 
\end{align*}
\]
and the minimum is taken over \((u, \mu, \nu)\). It was proven by Gianazza, Savaré and Toscani in 2009, that the drift diffusion equation, or also called

Derrida-Lebowitz-Speer-Spohn (DLSS) equation

\[
\begin{align*}
\partial_t u &= -\Delta(u\Delta \ln u) = -2\nabla \cdot \left( u\nabla \left( \frac{\Delta \sqrt{u}}{\sqrt{u}} \right) \right), \quad \forall \, x \in \Omega, \, \forall \, t \geq 0, \quad (6.2.7) \\
u(0, x) &= u_0(x) > 0, \quad \forall \, x \in \Omega. \quad (6.2.8)
\end{align*}
\]

with appropriate boundary conditions, constitutes a gradient flow for the Fisher information with respect to the Wasserstein distance in the space \(\mathcal{P}_2(\Omega)\) \[^{84}\]. Based on the theory of generalised minimising movement introduced in Chapter \[4\] the authors proved, that the limit curve of a variational scheme converges to the solution of (6.2.7).

To formally justify the gradient flow structure (4.1.5)-(4.1.7), we reformulate (6.2.7) as a system of three equations:

\[
\begin{align*}
\partial_t u + \nabla \cdot (uv) &= 0, \quad (6.2.9) \\
v &= \nabla \psi, \quad (6.2.10) \\
\psi &= -\frac{\delta E(u)}{\delta u}, \quad (6.2.11)
\end{align*}
\]

where the Euler-Lagrange first variation of (6.2.1) is given by

\[
\frac{\delta E(u)}{\delta u} = -2 \left( \frac{\Delta \sqrt{u}}{\sqrt{u}} \right).
\]

Indeed, \(E\) can be written as an integral function,

\[
E(u) = \int_{\Omega} f(x, u(x), \nabla u(x)) \, dx
\]

with \(f(x, z, p) : \Omega \times [0, \infty) \times \Omega \to \mathbb{R}\), here \(f(x, z, p) = \frac{1}{2} \frac{|p|^2}{z}\).

Then we conclude for positive regular functions

\[
\frac{\delta E(u)}{\delta u} = f_z(x, u, \nabla u) - \nabla \cdot f_p(x, u, \nabla u)
= -\frac{2 |\nabla u|^2 + |\nabla u|^2 - \Delta u}{2u^2} = -2 \left( \frac{\Delta \sqrt{u}}{\sqrt{u}} \right).
\]
6.2.2 Existence and Uniqueness

In the following we state the existence result for the variational problem (6.2.2) [84].

**Theorem 6.2.1.** Let $X = \Omega \subset \mathbb{R}^N$ be an open and bounded domain with compact Lipschitz boundary. $\nu \in \mathcal{P}_2(\Omega)$ be a probability measure on $\Omega$ and $u$ is a probability density and $\mathcal{L}^N$ the usual Lebesgue measure on $\mathcal{R}^N$. We consider the cost function $c(x, y) = |x - y|^2$. Then, the optimisation problem

$$J(u) = \frac{1}{2} W_2(\nu, u\mathcal{L}^N)^2 + \epsilon E(u) \to \min_{u\mathcal{L}^N \in \mathcal{P}(\Omega)},$$

with $E(u) = \int_{\Omega} u |\nabla \log u|^2$ and $\sqrt{u} \in W^{1,2}(\Omega)$ admits a minimiser $u\mathcal{L}^N \in \mathcal{P}(\Omega)$.

**Proof.** The existence proof is similar to the foregoing proofs. We use here some properties of the Fisher information stated in [84, Lemma 2.4] which yields the strong convergence of $u$ in $L^1(\Omega)$.

Let $u^n$ be a minimising sequence satisfying

$$\sup_n (M_2(u^n) + E(u^n)) < +\infty.$$ 

Then, there exists a $u\mathcal{L}^N \in \mathcal{P}_2(\Omega)$ and a subsequence, still denoted by $u^n$ such that

$$u^n \mathcal{L}^N \rightharpoonup u \mathcal{L}^N \text{ in } \mathcal{P}_2(\Omega),$$

and

$$u^n \to u \text{ strongly in } L^1(\Omega)$$

Moreover, we have the strong convergence of $\sqrt{u}$:

$$\sqrt{u^n} \to \sqrt{u} \text{ strongly in } L^2(\Omega).$$

In particular

$$u^n \to u \text{ strongly in } \begin{cases} L^r_{\text{loc}}(\overline{\Omega}), & 1 \leq r \leq \frac{N}{N-2} \text{ if } 2 < N, \\ L^r_{\text{loc}}(\overline{\Omega}), & 1 \leq r < +\infty \text{ if } 2 = N, \\ L^\infty_{\text{loc}}(\overline{\Omega}), & \text{ if } N < 2. \end{cases}$$

and

$$\nabla \sqrt{u^n} \to \nabla \sqrt{u}, \text{ weakly in } L^2(\Omega).$$

With the strong convergence in $L^1(\Omega)$ and narrow convergence in $\mathcal{P}_2(\Omega)$ as well as with the Proposition (2.2.2) we conclude the existence of a minimiser $\sqrt{u} \in W^{1,2}(\Omega)$ and $u\mathcal{L}^N \in \mathcal{P}_2(\Omega)$. Since the Fisher information is not strict convex nor displacement convex [9], we cannot follow the uniqueness. \qed
6.2.3 Stationary Solutions

The partial differential equation (6.2.7) is known as the Derrida-Lebowitz-Speer-Spohn equation (DLSS), since it was first investigated by B. Derrida et. al. in the context of interface fluctuations between regions of predominantly positive and negative spins in the anchored Toom-model (cf. [67, 66]). Besides this, in the multi-dimensional case it finds applications in the quantum semiconductor physic where it is also known as the quantum drift diffusion equation [100, 104]. Here, the solution $u$ describes the density of electrons in a quantum plasma (semiconductor device) and satisfies the quantum drift diffusion model:

$$\partial_t u = \nabla \cot \left( T \nabla u + u \nabla \left( V_{el} - \frac{\epsilon^2 \Delta \sqrt{u}}{6 \sqrt{u}} \right) \right),$$

(6.2.14)

with temperature $T > 0$ and a potential, splitted into an electric potential $V_{el}$ and the Bohm-potential. Then the DLSS-equation (6.2.7) results from $V_{el} = 0$ and $T \to 0$.

The DLSS equation is a highly nonlinear fourth order equation which does not fulfill a maximum principle. Hence, even though $u_0 > 0$ we cannot assume that $u(t) > 0$, $\forall t > 0$, and the solution is not necessarily bounded by the minimum and maximum of the initial data. Until now, there is no proof that the function is bounded below from zero. There exist various literature on the DLSS-equation. We refer for a detailed analysis on existence and uniqueness to [30, 103, 84, 102]. Furthermore, the DLSS equation is discussed in detail in [70].

The numerical study of the long time asymptotics, i.e. the proof of exponential convergence to equilibrium, can be found in [51]. Moreover, the DLSS equation admits a self similar solution which is a Gaussian curve [102]

$$\hat{u}_*(t,x) = (8\pi^2 t)^{-N/4} \exp \left( -\frac{|x|^2}{\sqrt{8t}} \right).$$

(6.2.15)

6.2.4 Self-Similar Solutions

For the Fisher information $E(u) = \int_{\Omega} u |\nabla \ln u|^2 dx = \int_{\Omega} \frac{|\nabla u|^2}{u} dx$, the first variation of $E$ reads:

$$E'(u) = - (\nabla \ln u)^2 - 2\Delta \ln u.$$

(6.2.16)

Then the optimality condition (3.2.41) just becomes

$$- (\nabla \ln u)^2 - 2\Delta \ln u = -\frac{|x|^2}{2\sigma^2} + \frac{\eta(x)}{\epsilon} - \frac{\eta_0}{\epsilon} \quad \text{with} \quad \sigma = \sqrt{\frac{\epsilon}{a}}.$$

(6.2.17)
Figure 6.5: Self-similar solution for the Fisher Information (6.2.1) with $a = 1$.

From (6.2.16) we deduce that $u$ is positive everywhere and hence $\eta \equiv 0$. Considering the one-dimensional case we have to compute the solution $u(x)$ from

$$-(\ln u)'^2 - 2(\ln u)'' = -\frac{|x|^2}{2\sigma^2} - \frac{\eta_0}{\epsilon}.$$  

We substitute $u = e^y$ and obtain

$$-(y')^2 - 2y'' = -\frac{|x|^2}{2\sigma^2} - \frac{\eta_0}{\epsilon},$$

and after transforming into a system of equations of first order

$$y' = \psi$$

$$\psi' = \frac{|x|^2}{4\sigma^2} + \frac{\eta_0}{2\epsilon} - \frac{1}{2} \psi^2,$$

we conclude $\psi = \pm \frac{x}{\sqrt{2}\sigma}$ for a special $\frac{\eta_0}{2\epsilon} = -\frac{1}{\sqrt{2}\sigma}$ if $\psi$ is negative. Then we get $y = -\frac{x^2}{2\sqrt{2}\sigma} + C$ with $C \in \mathbb{R}$. Knowing that $\int_{\Omega} u(x)dx = 1$ we calculate the constant $C$:

$$\int_{\Omega} u(x)dx = \int_{\Omega} e^{-\frac{x^2}{2\sqrt{2}\sigma}} e^C dx,$$

$$= e^C \int_{\Omega} e^{-z^2} (2\sqrt{2}\sigma)^{N/2} dz,$$

$$= e^C (2\sqrt{2}\sigma\pi)^{N/2} = 1,$$

and we conclude

$$e^C = \frac{1}{(2\sqrt{2}\sigma\pi)^{1/2}}.$$

Eventually we obtain the self-similar solution for the Fisher-information

$$u(x) = \frac{1}{(2\sqrt{2}\sigma\pi)^{1/2}} e^{-\frac{x^2}{2\sqrt{2}\sigma}},$$

which behaves like a Gaussian curve. We illustrate the result in Figure 6.5.
6.2.5 Numerical Results and Discussion

Figure 6.6: Numerical results for the Fisher information computed with the Kantorovich algorithm (8.2.1)-(8.2.5): \( u_0 \)–initial data, \( u \)–solution, \( \epsilon \)–regularisation parameter.

In Figure 6.6 we present numerical results for the variational problem (6.2.2) with Fisher information as regularisation term in one dimension. The initial data, which is coloured in blue, is a Gaussian curve with mean zero and variance one, i.e. the mass is normalised to one. Note that we consider here positive initial data to avoid numerical problems, which can occur when the solution touches zero or becomes negative. We discuss different values for the regularisation parameter \( \epsilon \), which yields to different numerical results, since \( \epsilon \) controls the influence of the regularisation term. Hence, the larger \( \epsilon \) is the larger is the effect of regularisation. In addition, we present numerical results for the Benamou-Brenier formulation in one dimension in Figure 6.7(a) solved with the gradient descent scheme (GDS) (8.3.7). These results coincide with the solutions presented in Figure 6.6 and also with the time discrete solutions of the DLSS-equation presented in Figure 6.7(b). The difficulty in the numerical implementation is the fact, that the positivity of the solution must be preserved and furthermore the conservation of mass. We also refer to the work of Düring, Matthes and Milisic [70] for an efficient implementation of the DLSS equation. In particular, they introduced an algorithm in dimension one based on the pseudo inverse formulation of the Wasserstein distance.
6.3 TV-Regularisation

This section deals with the TV-regularisation

$$E(u) := |Du|(\Omega) = \sup_{g \in C_0^\infty(\Omega; \mathbb{R}^N), \|g\|_{\infty} \leq 1} \int_{\Omega} u \nabla \cdot g \, dx. \quad (6.3.1)$$

When $u \in C^1(\Omega)$ we can also write the TV-regularisation in the following form

$$E(u) := \int_{\Omega} |\nabla u| \, dx.$$ 

We recall that for $\Omega \subset \mathbb{R}^N$ for $(N = 1, 2)$ bounded open set with Lipschitz boundary, $u \in L^1_{i\text{oc}}(\Omega)$

$$V(u, \Omega) := \sup \left\{ \int_{\Omega} u \nabla \cdot \varphi \, dx : \varphi \in [C^1_c(\Omega)]^N, \|\varphi\|_{\infty} \leq 1 \right\}$$

is the variation of $u$ and that $u \in BV(\Omega)$ (the space of bounded variation functions, [8, 75]) if and only if $V(u, \Omega) < \infty$, see [8, Proposition 3.6]. In such a case, $|D(u)|(\Omega) = V(u, \Omega)$, where $|D(u)|(\Omega)$ is the total variation of the finite Radon measure $Du$, the derivative of $u$ in the sense of distributions. Then $u \mapsto |Du|(\Omega)$ is lower semicontinuous in $BV$ with respect to the $L^1_{i\text{oc}}$ topology.

The TV-regularisation is well known in image processing, e.g. Rudin, Osher and Fatemi introduced in 1992 a variational ansatz with TV-regularisation for image denoising [154]. In the last years the applications in signal and image processing increased, since the TV-regularisation provides some nice properties, e.g. it preserves
discontinuities and edges in reconstructions in one dimension. Moreover, in statistics the TV-regularisation has become a standard approach for density estimation [112, 129, 156, 134].

6.3.1 The Problem Formulation

Let $X = \Omega$ be an open and bounded subset of $\mathbb{R}^N$ and $\nu \in \mathcal{P}_2(\Omega)$ a probability measure. Furthermore, let $u$ be a probability density, i.e. $u \geq 0$ and $\int_\Omega u \, dx = 1$, $\mathcal{L}^N$ is the usual Lebesgue measure. We want to investigate the solution of the variational problem

Variational problem: \[
W_2^2(\nu, u\mathcal{L}^N) + \epsilon |Du|(\Omega), \tag{6.3.2}
\]

which results by implementing the TV-regularisation into (3.1.1). In the sequel we consider the quadratic Euclidean distance $c(x, y) = |x - y|^2$ and obtain:

\[
I(u) = \frac{1}{2} \int_{\Omega \times \Omega} |x - y|^2 d\pi(x, y) + \epsilon |Du|(\Omega) \rightarrow \min_{\pi, u}, \tag{6.3.3}
\]

subject to

\[
\int_A \times \Omega \ d\pi(x, y) = \int_A d\nu(y), \tag{6.3.4}
\]

\[
\int_{\Omega \times A} d\pi(x, y) = \int_A u(x) dx, \tag{6.3.5}
\]

for all $A \subset \Omega$ measurable. This optimisation problem can also be transferred into a fluid dynamic framework (see Section 3.2):

Benamou-Brenier formulation

Minimise \[
\frac{1}{2} \int_0^1 \int_{\Omega} |v|^2 d\mu(x) \ dt + \epsilon |Du|(\Omega), \tag{6.3.6}
\]

where the minimisation is taken over $(u, \mu, v)$, subject to

\[
\partial_t \mu + \nabla \cdot (\mu v) = 0, \quad \text{(continuity equation)} \tag{6.3.7}
\]

\[
\mu(t = 0) = \nu, \quad \text{(initial condition)} \tag{6.3.8}
\]

\[
\mu(t = 1) = u\mathcal{L}^N, \quad \text{(final condition)} \tag{6.3.9}
\]

We also want to derive an associated gradient flow equation. We obtain for the TV-regularisation a highly nonlinear fourth order equation:
Gradient flow equation for the TV-regularisation:

\[ \partial_t u = -\nabla \cdot \left( u \nabla \cdot \left( \frac{\nabla u}{|\nabla u|} \right) \right) \quad \forall x \in \Omega, \forall t > 0, \quad (6.3.10) \]

\[ u(0, x) = u_0 \geq 0 \quad \forall x \in \Omega. \quad (6.3.11) \]

with appropriate boundary conditions. To formally justify the gradient flow structure (see (4.1.5)-(4.1.7)), we reformulate (6.3.10) in the following way:

\[ \partial_t u + \nabla \cdot (uv) = 0, \quad (6.3.12) \]

\[ v = \nabla \psi, \quad (6.3.13) \]

\[ \psi = -\frac{\delta E(u)}{\delta u}, \quad (6.3.14) \]

where the Euler-Lagrange first variation of (6.3.1) with \( u \in C^1(\Omega) \) is given by

\[ \frac{\delta E(u)}{\delta u} = -\nabla \cdot \left( \frac{\nabla u}{|\nabla u|} \right). \]

Indeed, \( E \) can be written as an integral function

\[ E(u) = \int_{\Omega} f(x, u(x), \nabla u(x)) dx \]

with \( f(x, z, p) : \Omega \times [0, \infty] \times \Omega \rightarrow \mathbb{R} \), here \( f(x, z, p) = |p| \).

Then we conclude for positive regular functions

\[ \frac{\delta E(u)}{\delta u} = f_z(x, u, \nabla u) - \nabla \cdot f_p(x, u, \nabla u) = -\nabla \cdot \left( \frac{\nabla u}{|\nabla u|} \right). \]

We also state here the following computations for regular functions:

\[ \text{diff} E(u).s = \int_{\Omega} \frac{\nabla u}{|\nabla u|} s dx \]

with \( s = -\nabla \cdot (u \nabla v) \) it follows

\[ \text{diff} E(u).s = -\int_{\Omega} \nabla \cdot \left( \frac{\nabla u}{|\nabla u|} \right) \nabla \cdot (u \nabla v) dx \]

\[ = \int_{\Omega} \nabla \nabla \cdot \left( \frac{\nabla u}{|\nabla u|} \right) u \nabla v dx \]

\[ = -\int_{\Omega} \nabla \left( u \nabla \nabla \cdot \left( \frac{\nabla u}{|\nabla u|} \right) \right) v dx \]

And with

\[ \left\langle \frac{\partial u}{\partial t}, s \right\rangle_W = \int \frac{\partial u}{\partial t} v dx, \]

eventually

\[ \partial_t u = \nabla \cdot \left( u \nabla \nabla \cdot \left( \frac{\nabla u}{|\nabla u|} \right) \right). \]
6.3.2 Existence and Uniqueness

In this section we prove the existence of a minimiser of (6.3.2).

**Theorem 6.3.1.** Let \( X = \Omega \subset \mathbb{R}^N \) be an open and bounded domain with compact Lipschitz boundary. We consider the following minimisation problem

\[
J(u) = \frac{1}{2} W_2(\nu, u \mathcal{L}^N)^2 + \epsilon |Du| (\Omega) \rightarrow \min_{u \mathcal{L}^N \in \mathcal{P}(\Omega)},
\]

(6.3.15)

where \( \nu \in \mathcal{P}(\Omega) \), the set of probability measures on \( \Omega \), \( u \) is a probability density, and \( \mathcal{L}^N \) is the usual Lebesgue measure on \( \mathbb{R}^N \). Then, the functional \( J \) in (6.3.15) has a minimiser \( u \in \mathcal{P}(\Omega) \).

**Proof.** Let \((u^n)_{n \in \mathbb{N}} \) be a minimising sequence for \( J(u) \), i.e.,

\[
J(u^n) \rightarrow \inf_{u \mathcal{L}^N \in \mathcal{P}(\Omega)} J(u) =: m \quad \text{as} \quad n \rightarrow \infty.
\]

Since \( u^n \) is a minimising sequence \( J(u^n) \) is bounded, thus \( u^n \mathcal{L}^N \in \mathcal{P}(\Omega) \) and \( u^n \) uniformly bounded. In particular there exists a constant \( M > 0 \) such that

\[
M \geq J(u^n) = \frac{1}{2} W_2(\nu, u^n \mathcal{L}^N)^2 + \epsilon |Du^n| (\Omega),
\]

and hence

\[
W_2(\nu, u^n \mathcal{L}^N) \leq M \quad \text{and} \quad |Du^n| (\Omega) \leq M \quad \forall n \geq 1.
\]

Additionally \( u^n \) is uniformly bounded in \( L^1(\Omega, \mathcal{L}^N) \). Since the Lebesgue measure is the standard measure in Euclidean space, the space \( L^p(\Omega, \mathcal{L}^N) \) is the usual space \( L^p(\Omega) \) and integration with respect to \( \mathcal{L}^N \) will be denoted by \( dx \). The uniform boundedness of \( u^n \) in \( L^1 \) follows from the assumption that \( u^n \mathcal{L}^N \in \mathcal{P}(\Omega) \). Namely, since \( u^n \mathcal{L}^N \in \mathcal{P}(\Omega) \) is a probability measure, it follows that \( \int_{\Omega} |u^n| \, dx = \int_{\Omega} u^n \, dx = 1 \).

From the last result and the boundedness of \( |Du^n| \) we have that there exists a subsequence, still denoted by \( u^n \), and a \( u \in BV(\Omega) \) such that \( u^n \rightarrow u \) strongly in \( L^1(\Omega) \) and weakly* in \( BV(\Omega) \). From the convergence in \( L^1(\Omega) \) we have

\[
\int_{\Omega} u^n \psi \, dx \rightarrow \int_{\Omega} u \psi \, dx, \quad \forall \psi \in L^1(X)
\]

\[
\iff
\]

\[
\int_{\Omega} \psi \, du^n \mathcal{L}^N \rightarrow \int_{\Omega} \psi \, du \mathcal{L}^N, \quad \forall \psi \in L^1(\Omega).
\]
Since \( C_0^0(\Omega) \subset L^p(\Omega) \) for all \( p \geq 1 \), it follows
\[
\int_{\Omega} \psi \, du^n L^N \to \int_{\Omega} \psi \, du L^N, \quad \forall \psi \in C_0^0(\Omega),
\]
and therefore \( u^n L^N \subset P(\Omega) \) converges narrowly to \( u L^N \in P(\Omega) \). With Proposition 2.2.2, here with \( \nu \in P(\Omega) \), \( u^n L^N \subset P(\Omega) \) we obtain
\[
W_2(\nu, u L^N) \leq \liminf_{n \to \infty} \int_{\Omega} |x - y|^2 \, d\pi^n(x, y) = \liminf_{n \to \infty} W_2(\nu, u^n L^N).
\]
Moreover from the lower-semicontinuity of \( |Du| (\Omega) \) in \( L^1(\Omega) \) we get
\[
\liminf_{n \to \infty} J(u^n) \geq \frac{1}{2} W_2(\nu, u L^N)^2 + \epsilon |Du| (\Omega) = J(u).
\]
It follows that \( J(u) \) has at least one minimiser \( u \in BV(\Omega) \) and \( u L^N \in P(\Omega) \).

Similar as for the Dirichlet regularisation we discuss here the stability of the evolution equation. Therefore we calculate the first derivative of \( E \) along the geodesic:
\[
\frac{dE}{ds} = \int_{\Omega} \frac{\nabla u}{|\nabla u|} \nabla u_s dx + \int_{\Omega} \frac{1}{2} |x|^2 u_s dx
\]
\[
= \int_{\Omega} \frac{\nabla u}{|\nabla u|} \nabla (\nabla \cdot (uv)) dx - 2N \int_{\Omega} uv \cdot x \, dx.
\]
And the second derivative reads
\[
\frac{d^2 E}{ds^2} = \int_{\Omega} \left( -\frac{\nabla u \nabla u^T}{|\nabla u|^3} \nabla u_s + \frac{\nabla u_s}{|\nabla u|} \right) \nabla (\nabla \cdot (uv)) dx + \int_{\Omega} \frac{\nabla u}{|\nabla u|} \nabla (\nabla \cdot (u_s v) + \nabla \cdot (uv_s))
\]
\[
+ N^2 \int_{\Omega} u|v|^2 dx
\]
\[
= \int_{\Omega} \left( -\frac{\nabla u \nabla u^T}{|\nabla u|^3} + \frac{1}{|\nabla u|} \right) (\nabla (\nabla \cdot (uv))) \cdot (\nabla (\nabla \cdot (uv))) dx
\]
\[
+ \int_{\Omega} \frac{\nabla u}{|\nabla u|} \nabla (\nabla \cdot (uv)) dx
\]
\[
+ \int_{\Omega} \frac{\nabla u}{|\nabla u|} \nabla (\nabla \cdot (uv) v) dx + \nabla \cdot (u \frac{1}{2} \nabla |v|^2) dx + N^2 \int_{\Omega} u|v|^2 dx
\]
\[
= \int_{\Omega} \left( -\frac{\nabla u \nabla u^T}{|\nabla u|^3} + \frac{1}{|\nabla u|} \right) (\nabla (\nabla \cdot (uv))) \cdot (\nabla (\nabla \cdot (uv))) dx
\]
\[
+ \int_{\Omega} \nabla \nabla \cdot \left( \frac{\nabla u}{|\nabla u|^2} \right) (\nabla (\nabla \cdot (uv)) + uv \nabla \cdot v) dx + N^2 \int_{\Omega} u|v|^2 dx
\]
\[
= \int_{\Omega} \frac{1}{|\nabla u|} \left( -\frac{\nabla u \nabla u^T}{|\nabla u|^2} + \text{Id} \right) (\nabla (\nabla \cdot (uv))) \cdot (\nabla (\nabla \cdot (uv))) dx
\]
\[
- \int_{\Omega} v \cdot \left( \nabla \left( \nabla \cdot \left( \frac{\nabla u}{|\nabla u|^2} \right) \right) \right) uv dx + N^2 \int_{\Omega} u|v|^2 dx.
\]
To prove stability we have to evaluate the second derivative in the critical point \( u_\infty \) and show the strict positivity. Since the first term is positive we have to determine only the value for the second term. The minimiser \( u_\infty \) of \( E \) must fulfill the following condition
\[
\nabla \cdot \left( \frac{\nabla u_\infty}{|\nabla u_\infty|} \right) = \frac{1}{2} |x|^2.
\]
So we can conclude
\[
\nabla \left( \nabla \nabla \cdot \left( \frac{\nabla u_\infty}{|\nabla u_\infty|} \right) \right) = N^2 \text{Id},
\]
and the second and third term vanish. Since the matrix
\[
\left( \text{Id} - \frac{\nabla u_\infty \nabla u_\infty^T}{|\nabla u|^2} \right)
\]
is positive semi-definite, we conclude
\[
\frac{d^2 E}{ds^2} \geq 0.
\]

### 6.3.3 Self-Similar Solutions

![Self-similar solution for TV regularization](image)

Figure 6.8: Self-similar solution for the TV-regularisation (6.3.1) with \( a = 1 \).

In case of the total variation, the energy \( E(u) \) is not differentiable. Hence instead of the gradient, we consider a subgradient of \( E \), i.e., an element in the subdifferential \( \partial E(u) \). Elements in the subdifferential of \( E \) are characterised by
\[
\partial E(u) = \left\{ \nabla \cdot g : g \in C_0^\infty(\Omega, \mathbb{R}^N), \|g\|_\infty \leq 1 \right\}.
\]
From the optimality condition (3.2.41) we have
\[ \nabla \cdot g = -\left| \frac{x^2}{2\sigma^2} + \frac{\eta}{\epsilon} - \frac{\eta_0}{\epsilon} \right|, \]
for a \( g \) with \( \nabla \cdot g \in \partial E(u) \). In the following we shall compute a self-similar solution \( u \) for dimension \( N = 1 \). Hence we have
\[ g'(x) = -\left| \frac{x^2}{2\sigma^2} + \frac{\eta}{\epsilon} - \frac{\eta_0}{\epsilon} \right|. \]
(6.3.18)

Since \( u \) is symmetrically centered around zero (radial invariance assumption), \( u(x) > 0 \) within a neighborhood of zero.

**Remark 6.3.2.** Since the weak derivative \( Du(0) = 0 \) we have that \( u \) attains a maximum or minimum value in this point. Since \( \int u = 1 \) this value has to be either strictly greater or strictly less than zero. Further, an element \( g' \) of the subdifferential of \( |Du| \) can be characterised to be equal to \( g' = \frac{Du}{|Du|}' \) for points where \( Du \neq 0 \), which is a weighted second derivative of \( u \). From the optimality condition (6.3.18) above we know that \( g' \) is a polynomial of order 2, and hence has two zeros at most (these are the points where \( g' \) changes sign). In addition we know from (6.3.18) that \( g' > 0 \) for \( |x| \) large enough, which means that \( u \) attains its minimum values for large \( |x| \). Hence the maximum must be attained at zero and thus \( u(x) > 0 \) within a neighborhood of zero.

Eventually we have that \( \eta(x) \) vanishes within this neighborhood and hence
\[ g'(x) = -\frac{|x^2|}{2\sigma^2} - C, \]
within a neighborhood around zero and for a constant \( C = \eta_0/\epsilon \). By integration we get
\[ g(x) = -\frac{x^3}{6\sigma^2} - Cx + \tilde{C}. \]

We make the following ansatz for \( u \):
\[ u(x) = \begin{cases} 
\alpha & (-x_0, x_0) \\
0 & \text{otherwise,} 
\end{cases} \]
with \( \alpha > 0 \) is a positive constant. For such a \( u \) the total variation gives
\[ 2\alpha = |Du| = \int_{-x_0}^{x_0} ug' dx = \int_{-x_0}^{x_0} \alpha g' dx = \alpha (g(x_0) - g(-x_0)) = 2\alpha g(x_0), \]
and hence \( g(x_0) = 1 \).

**Remark 6.3.3.** Note that this also reassembles the characterisation of \( g = \frac{Du}{|Du|} = \pm 1 \) in points where \( Du \neq 0 \).
Additionally we can compute the points \(-x_0, x_0\) and the constant \(\alpha\) by using the optimality condition (6.3.18) and the mass constraint on \(u\) respectively, i.e.,

\[
0 = g'(x_0) = -\frac{|x_0|^2}{2\sigma^2} - C,
\]
gives \(x_0 = \sqrt{-2C\sigma}\), and

\[
1 = \int_{-x_0}^{x_0} u \, dx = 2\alpha x_0,
\]
gives \(\alpha = 1/(2x_0) = 1/(2\sqrt{-2C\sigma})\). It remains to compute the constant \(C = \eta_0/\epsilon\) which we obtain from the following computation

\[
1 = g(x_0) = (-C)^{3/2} \frac{2\sqrt{2\sigma}}{6} + \tilde{C},
\]
and hence

\[
C = -\frac{1}{2} \sqrt{\frac{9(1 - \tilde{C})^2}{\sigma^2}}, \quad x_0 = \sqrt[3]{\frac{9(1 - \tilde{C})^2 \sigma}{6}}.
\]
Finally we have computed a self-similar solution of the form

\[
u(x) = \begin{cases} 1/(2x_0) & (-x_0, x_0), \\ 0 & \text{otherwise.} \end{cases}
\]

For a special \(\tilde{C} = 0\) we obtain the point \(x_0 = \sqrt[3]{9\sigma}\).

Please compare Figure 6.8 for a sketch of such solutions.

### 6.3.4 Numerical Results and Discussion

We present in the following numerical results for the variational problem (6.3.2) with TV-regularisation. In Figure 6.9 results with the Gaussian curve as initial data \(\nu\) are presented. Especially, we consider a Gaussian curve with mean zero and variance one, i.e. \(\int_\Omega \nu = 1\). For the variational formulation (6.3.2) we use the algorithm (8.2.1)-(8.2.5) to compute the solutions. In this direct approach, i.e. we first discretise and then optimise, we have to consider an approximation of the TV regularisation, namely \(E(u) = \int \sqrt{\nabla u}^2 + \delta\) with small \(\delta > 0\). This is due to numerical difficulties which can occur when \(\nabla u\) is equal to zero. In addition, we present in Figure 6.9(b) numerical results for the Benamou-Brenier formulation, which we solved with the augmented Lagrangian scheme (ALS) (8.3.19). Since the numerical results for both are identical, this confirms the correctness of our models and algorithms. The numerical results
Figure 6.9: Numerical results for the TV-regularisation in one dimension: (a) Solutions for the variational problem (6.3.2) solved with the algorithm (8.2.1)-(8.2.5) for different regularisation parameter $\epsilon$; (b) Solutions for the Benamou-Brenier formulation (6.3.6)-(6.3.9) computed with the augmented Lagrangian scheme (ALS) ((8.3.19)) with $\Delta x = 0.25$, $\Delta t = 0.3$, $\tau = 0.1$.

illustrate the property of the TV-regularisation, namely the construction of edges. Note that this is not in general true in two dimensions [5].

We also provide numerical simulations in two dimensions with the Gaussian curve as initial data, as well as with real life data, e.g. MRI (Magnetic Resonance Imaging) measurements and grayvalue images. We can switch between the

**Anisotropic TV regularisation**

$$E(u) = \int_\Omega |\nabla_x u| + |\nabla_y u| \, dx$$

and the

**Isotropic TV regularisation**

$$E(u) = \int_\Omega \sqrt{|\nabla_x u|^2 + |\nabla_y u|^2} \, dx.$$ 

In Figure 6.10 we present numerical results with a Gaussian initial data (computed with the ALS-algorithm (8.3.19)), which demonstrates the different impact of these two terms. The edges for the isotropic regularisation term is more rounded than for the anisotropic numerical results. In addition, we consider MRI-data provided from the Max-Planck-Institut for biochemical chemistry in Göttingen [2] (see Figure 6.11). For comparison, we did numerical simulations on a 140 $\times$ 140-detail of the MRI-scan.
using the Dirichlet regularisation (6.1.1) and additionally the TV-regularisation (6.3.1). We present these numerical results in Figure 6.11(c)-(h). In Figure 6.11(b) we show the contour-plot of the initial data, and in (c) respectively (d) the contour-plot for the solution computed with Dirichlet regularization, respectively with anisotropic TV-regularisation. In Figure 6.11(f)-(h) different degrees of regularisation with TV are presented, which demonstrates the property of the TV-regularisation to preserve and construct edges in the image.

In order to confirm that the investigated functionals can also be used for image decomposition we present in Figure 6.12 some numerical experiments on grayvalue images. By image decomposition we mean, that a given image \( f = u + v \) is decomposed into \( u \), which is the piecewise smooth part and represents the structure resp. cartoon part of the image and into \( v \), which represents the oscillatory parts resp. the texture. Note here that we do not want to compare our results with existing numerical methods for image decomposition [154, 139, 174]. However, a nice advantage of our model is, that the resulting cartoon image is a probability measure since the initial data is a probability measure. We consider two different initial data, a simple synthetic image and a detail of the Barbara-image [1], where texture parts are combined with nontexture parts. In the middle row of the Figure 6.12 we show the structure-, or non-texture part of the initial images, which is the minimiser of (6.3.6)-(6.3.9) with TV-regularisation solved with the augmented Lagrangian scheme (ALS) (8.3.19). The texture images, presented in the last row results from subtracting the structure image from the initial image.
Figure 6.10: Numerical results in 2D computed with the augmented Lagrangian algorithm (ALS) (8.3.19): First column: Anisotropic TV regularisation, $\epsilon = 1, 10, 50$; Second column: Isotropic TV regularisation, $\epsilon = 1, 10, 50$. 
Figure 6.11: (a) MRI-measurements provided by the MPI for biophysical chemistry Göttingen [2]; (b) Initial data, $140 \times 140$ detail; (c) Dirichlet regularisation (GDS); (d) TV-regularisation (ALS) for $\epsilon = 0.2$; (e) Dirichlet regularisation (GDS); (f)-(h) Isotropic TV-regularisation (ALS) for $\epsilon = 0.05, 0.2, 1$. 
Figure 6.12: Image decomposition into a cartoon and texture part. Results from the augmented Lagrangian scheme (ALS) with TV-regularisation (see Section 8.3): (a) Left: Initial $50 \times 50$ Barbara-detail; Middle: Cartoon part of the image; Right: Texture part of the image. (b) Left: Synthetic image; Middle: Cartoon part of the image; Right: Texture part of the image.
Chapter 7

The Nonlocal Interaction Functional

In this chapter we discuss a nonlocal interaction energy functional. The associated gradient flow equation with respect to the Wasserstein distance is an aggregation equation with quadratic diffusion term (see Section 7.1). Additionally, we give existence and uniqueness results of nontrivial stationary solutions in Section 7.2. We prove a threshold phenomenon, namely that stationary solutions exist if and only if the diffusivity constant is strictly smaller than the total mass of the interaction kernel (see Section 7.2.2). Furthermore, in Section 7.2.3 we prove that nontrivial stationary solutions in one space dimension with fixed mass and center of mass are unique. In Section 7.3 we will provide numerical results.

7.1 The Aggregation Equation

In the following, we are concerned with the aggregation equation

\[
\partial_t u = \nabla \cdot (u \nabla (\varepsilon u - G * u)) , \quad x \in \mathbb{R}^N, \quad t \geq 0, \quad (7.1.1)
\]
\[
u(0, \cdot) = u_0 \geq 0. \quad (7.1.2)
\]

The aim is to investigate the stationary solutions for (7.1.1) which satisfy the equation

\[
0 = \nabla \cdot (u \nabla (\varepsilon u - G * u)) \quad (7.1.3)
\]

posed on the whole space \(\mathbb{R}^N\) with \(\varepsilon \geq 0\). The convolution \(G * u\) is carried out with a kernel \(G : \mathbb{R} \to \mathbb{R}\) which fulfills the following assumptions

\begin{itemize}
  \item[(i)] \( G \geq 0 \), and \( \text{supp}(G) = \mathbb{R}^N \),
  \item[(ii)] \( G \in W^{1,1}(\mathbb{R}^N) \cap L^\infty(\mathbb{R}^N) \cap C^2(\mathbb{R}^N) \),
  \item[(iii)] \( G(x) = g(|x|) \) for all \( x \in \mathbb{R}^N \),
  \item[(iv)] \( g'(r) < 0 \) for all \( r > 0 \),
  \item[(v)] \( g''(0) < 0 \),
  \item[(vi)] \( \lim_{r \to +\infty} g(r) = 0 \).
\end{itemize}

Moreover, we assume for simplicity that the norm of \( G \) is normalised to one
\[
\|G\|_{L^1} = \int_{\mathbb{R}^N} G(x) dx = 1,
\]
since the kernel \( G \) can always be normalised by modifying the diffusion constant and the time scale as follows
\[
\tau = \|G\|_{L^1(\mathbb{R}^N)} \quad \tilde{G} = G/\|G\|_{L^1(\mathbb{R}^N)}, \quad \varepsilon' = \varepsilon/\|G\|_{L^1(\mathbb{R}^N)}
\]
\[
\partial_\tau u = \text{div} \left( u\nabla (\varepsilon' u - \tilde{G} * u) \right)
\]

Also, we assume \( u \) to have normalised mass one and \( u \geq 0 \). Let us recall that the equation (7.1.1) preserves

\begin{itemize}
  \item[(i)] the total mass \( M = \int_{\mathbb{R}} u(t,x) dx \), \( \forall x \in \mathbb{R}, \ t \geq 0 \),
  \item[(ii)] and the center of mass \( \text{CM}[u(t)] := \int_{\mathbb{R}} xu(t,x) dx \), \( \forall x \in \mathbb{R}, \ t \geq 0 \).
\end{itemize}

Indeed, let \( x \) be a test function, then
\[
\frac{d}{dt} \int_{\mathbb{R}} u(t,x) x dx = \int_{\mathbb{R}} \partial_1 u(t,x) x \ dx = \int_{\mathbb{R}} \partial_2 (u\partial_2(\varepsilon u - G * u)) x \ dx
\]
\[
= - \int_{\mathbb{R}} u\partial_2(\varepsilon u - G * u)dx = \int_{\mathbb{R}} \partial_2 \left( \frac{\varepsilon u^2}{2} - uG' * u \right) dx,
\]
Then the first term as well as the second term vanish, which is due to the symmetry of $G$:
\[
\int_{\mathbb{R}} uG' * u \, dx = - \int_{\mathbb{R}} uG' * u \, dy.
\]

The equation (7.1.1) can be formulated as a gradient flow in the Wasserstein metric (cf. [126, 18, 52, 53]) of the energy functional (see Chapter 4)
\[
E(u) := \varepsilon \int_{\mathbb{R}^N} u^2(x) \, dx - \frac{1}{2} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} G(x - y)u(y)u(x) \, dy \, dx.
\tag{7.1.4}
\]

Indeed, (7.1.4) fulfills the gradient flow scheme (4.1.5)-(4.1.7) since the first variation of $E$ reads
\[
\frac{\delta E}{\delta u} = \varepsilon u - G * u.
\]

### 7.2 Existence and Uniqueness

The aggregation equation (7.1.1) is well studied in literature. Existence results for aggregation equations with quadratic diffusion have been posed e.g. by [38, 27] in the context of entropy solutions. Furthermore, the aggregation equation is discussed in connection with the classical Patlak-Keller-Segel model for chemotaxis [146, 110] with respect to blow-up vs. large-time existence in [98, 92, 29, 60] for models with linear diffusion and [114, 45] for models with nonlinear diffusion. Moreover, the existence of stationary solutions in the purely aggregative case, i.e. for $a \equiv 0$ in one dimension is studied in [39] by means of the pseudo-inverse representation of the Wasserstein distance. And also the large-time behavior of solutions to equations of the form (7.1.1) in the purely nonlocal case, namely when $\varepsilon = 0$ has been studied extensively in many papers [120, 39, 118, 24, 23, 49, 25, 97, 79, 78], combined with the study of the regularity of solutions compared to the attractive singularity of the interaction kernel.

We know from [39, 16], that for the aggregative case with diffusion the existence of a solution depend on a certain threshold value:

- If $\varepsilon \geq \int G$, then there exists no stationary solution.
- If $\varepsilon < \int G$, there exists a stationary solution.

If $\varepsilon \geq \int G$ the solutions decay to zero with a prescribed rate for large times and behave like the (compactly supported) Barenblatt profiles (see Section 5.2) (cf. [173, 142, 57]). More refined results in a similar model derived in [93, 144] with cut-off density have been found in [40, 117]. In [16] a sharp result of existence of minimisers for the energy...
\(E(u)\) in a multi dimensional framework has been proven. An open problem in this context is the uniqueness of steady states under mass and center of mass constraint, since the functional \(E\) is neither convex in the classical sense nor in the displacement convex sense [120], (except when \(G\) is concave on \(\mathbb{R}^N\), see [52]). In the following, we discuss the characterisation of stationary solutions in the multi- and one-dimensional case.

7.2.1 Preliminaries

In general we cannot expect that the energy functional

\[
E(u) := \frac{\varepsilon}{2} \int_{\mathbb{R}^N} u^2(x)dx - \frac{1}{2} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} G(x - y)u(y)u(x)dydx,
\]

is convex or displacement convex, although the first term fulfills this property (see (2.3.6)). In the following we discuss the displacement convexity of the second term

\[
\mathcal{G} = \frac{1}{2} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} G(x - y)u(y)u(x)dydx
\]

along the geodesic which is parametrised with \(s\). Therefore we compute the first derivative of \(E\) along the geodesic using integration by parts

\[
\frac{d\mathcal{G}}{ds} = -\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} G(x - y)(u_s(y, s)u(x, s) + u(y, s)u_s(x, s))dydx
\]

\[
= -\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} G(x - y)\text{div}(u(y, s)v(y, s))u(x, s)dydx
\]

\[
- \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} G(x - y)\text{div}(u(x, s)v(x, s))u(y, s)dydx
\]

\[
= \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_y(G(x - y)u(x, s))u(y, s)v(y, s)dydx
\]

\[
+ \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x(G(x - y)u(y, s))u(x, s)v(x, s)dydx
\]

\[
= \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_xG(x - y)u(x, s)u(y, s)(v(x, s) - v(y, s))dydx.
\]
And the second derivative along the geodesic reads

\[
\frac{d^2 G}{ds^2} = \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u_s(x, s) u(y, s) v(x, s) dy dx + A \\
+ \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u(x, s) u_s(y, s) v(x, s) dy dx + B \\
+ \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u(x, s) u(y, s) v_s(x, s) dy dx + C \\
- \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u_s(x, s) u(y, s) v(y, s) dy dx + D \\
- \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u(x, s) u_s(y, s) v(y, s) dy dx + E \\
- \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u(x, s) u(y, s) v_s(y, s) dy dx + F.
\]

With \( u_s = \text{div}(uv) \) and integration by parts we obtain for \( A \)

\[
A = \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u_s(x, s) u(y, s) v(x, s) dy dx \\
= \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) \text{div}(u(x, s) v(x, s)) u(y, s) v(x, s) dy dx \\
= -\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x (\nabla_x G(x - y) u(y, s) v(x, s)) u(x, s) v(x, s) dy dx \\
= -\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \langle D^2 G(x - y) v(x, s), v(x, s) \rangle u(x, s) u(y, s) dy dx \\
- \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u(y, s) \nabla v(x, s) u(x, s) v(x, s) dy dx.
\]
Furthermore, we make similar calculations for $B$

\[
B = +\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u(x, s) u_s(y, s) v(x, s) dy dx \\
= \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) \text{div}(u(y, s) v(y, s)) v(x, s) dy dx \\
= -\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_y (\nabla_x G(x - y) u(x, s) v(x, s)) u(y, s) v(y, s) dy dx \\
= +\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \langle D^2 G(x - y) v(x, s), v(y, s) \rangle u(y, s) u(x, s) dy dx.
\]

For the term $C$ we use the property $v_s = \frac{1}{2} \nabla |v|^2$

\[
C = +\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u(x, s) u(y, s) v_s(x, s) dy dx \\
= \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u(x, s) u(y, s) \frac{1}{2} \nabla |v(x, s)|^2 dy dx \\
= \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u(x, s) u(y, s) v(x, s) \nabla v(x, s) dy dx.
\]

The computations for $D$, $E$ and $F$ are similar to the above calculations:

\[
D = -\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u_s(x, s) u(y, s) v(y, s) dy dx \\
= -\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) \text{div}(u(x, s) v(x, s)) u(y, s) v(y, s) dy dx \\
= \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x (\nabla_x G(x - y) u(y, s) v(y, s)) u(x, s) v(x, s) dy dx \\
= \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \langle D^2 G(x - y) v(x, s), v(y, s) \rangle u(y, s) u(x, s) dy dx
\]

\[
E = -\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u(x, s) u_s(y, s) v(y, s) dy dx \\
= -\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u(x, s) \text{div}(u(y, s) v(y, s)) v(y, s) dy dx \\
= \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_y (\nabla_x G(x - y) u(x, s) v(y, s)) u(y, s) v(y, s) dy dx \\
= -\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \langle D^2 G(x - y) v(y, s), v(y, s) \rangle u(x, s) u(y, s) dy dx \\
+ \frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y) u(x, s) \nabla v(y, s) u(y, s) v(y, s) dy dx
\]
\[ F = -\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y)u(x, s)u(y, s)v_s(y, s)dydx \]
\[ = -\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y)u(x, s)u(y, s)\frac{1}{2}\nabla |v(y, s)|^2 dydx \]
\[ = -\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \nabla_x G(x - y)u(x, s)u(y, s)v(y, s)\nabla v(y, s)dydx \]

Eventually, we obtain
\[ \frac{d^2 G}{ds^2} = -\frac{1}{2} \int_{\mathbb{R}^N \times \mathbb{R}^N} \langle D^2 G(x - y)(v(x, s) - v(y, s)), (v(x, s) - v(y, s)) \rangle u(x, s)u(y, s)dx dy \] (7.2.1)

We conclude that \( G \) is displacement convex if \( G \) is concave.

Let us recall the following results on the existence and uniqueness of gradient flow solutions to (7.1.1), which follows from the theory developed in [9].

**Theorem 7.2.1** ([9]). Let \( u_0 \in L^2 \cap \mathcal{P} \) be such that \( u \geq 0 \) and
\[ E(u_0) < +\infty. \]

Let \( G \) satisfy the above assumptions. Then there exists a unique weak solution \( u \) to (7.1.1) with

- \( E(u(t)) < +\infty \) for all \( t \geq 0 \),
- \( \sqrt{u} \nabla (\epsilon u - G * u) \in L^2([0, T] \times \mathbb{R}^2) \) for all \( T > 0 \),

such that the following energy identity is satisfied
\[ E(u(t)) + \int_0^T \int_{\mathbb{R}^N} u |\nabla (\epsilon u - G * u)|^2 dxdt = E(u_0). \] (7.2.2)

We have the following regularity property:

**Lemma 7.2.2** (Regularity of \( L^2 \cap \mathcal{P} \) steady states). Let \( u_0 \in L^2 \cap \mathcal{P} \). Then, the corresponding solution \( u(t) \) to (7.1.1) satisfies
\[ \int u |\nabla u|^2 dx < +\infty, \] (7.2.3)
for almost every \( t > 0 \). In particular, let \( u \) be an \( L^2 \cap \mathcal{P} \) steady state to (7.1.1), then \( u \) satisfies (7.2.3) and \( u \in C^2 \) on supp\((u)\).
Proof. Due to (7.2.2), the quantity
\[ \epsilon^2 \int u |\nabla u|^2 \, dx - 2\epsilon \int u \nabla u \cdot \nabla G * u \, dx + \int u |\nabla G * u|^2 \, dx \]
is finite for almost every \( t > 0 \), and therefore, in view of Cauchy-Schwarz inequality, we have
\[ \frac{\epsilon}{2} \int u |\nabla u|^2 \, dx - C(\epsilon) \int u |\nabla G * u|^2 \, dx + \int u |\nabla G * u|^2 \, dx < +\infty. \]
Because \( G \) fulfills some smoothness assumptions stated above we obtain (7.2.3). Let \( u \) be a steady state, then \( u \) satisfies (7.2.3), too. This implies in particular that \( \nabla u \) is almost everywhere finite on \( \mathbb{R}^2 \). The energy identity (7.2.2) implies then
\[ u |\nabla (\epsilon u - G * u)|^2 = 0 \]
for almost every \( x \in \mathbb{R}^2 \). This means that
\[ \epsilon u - G * u = \text{constant} \]
almost everywhere on every connected component of the support of \( u \). By convolution with standard mollifiers, one can easily see that \( \epsilon u - G * u = C \) for a given \( C \) depending on the connected component of \( \text{supp}(u) \). Since \( G \) is \( C^2 \), this easily implies \( u \in C^2 \) on \( \text{supp}(u) \).

**Corollary 7.2.3** (One dimensional regularity). Let \( u \) be an \( L^2 \cap \mathcal{P} \) solution to (7.1.3) in one space dimension. Then \( u \) is continuous on \( \mathbb{R} \).

**Proof.** Apply the result in Lemma 7.2.2 to the case \( N = 1 \). Since \( \frac{d}{dx} u^{3/2} \in L^2 \), the one dimensional Sobolev embedding implies that \( u^{3/2} \) is continuous. 

First, we start by focusing on the interplay between the solutions to (7.2.6) and the variational calculus on the energy functional
\[ E(u) := \epsilon \int_{\mathbb{R}^N} u^2(x) \, dx - \frac{1}{2} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} G(x - y) u(y) u(x) \, dy \, dx. \]
In the next proposition we prove that being a minimum for the entropy functional is a sufficient condition for being a solution to (7.2.6).

**Proposition 7.2.4** (Stationary solutions via energy minimisation). Let \( u \in L^2(\mathbb{R}^N) \) be a minimiser for the energy functional
\[ E(u) := \frac{1}{2} \int_{\mathbb{R}^N} u (\epsilon u - G * u) \, dx \]
on \( \mathcal{P} \). Then
\[ u \nabla (\epsilon u - G * u) = 0 \quad \text{a.e. in } \mathbb{R}^N. \]
Proof. Let $V \in C^1_c(\mathbb{R}^N)$ be an arbitrary vector field and let $\rho(s, x)$ be a local solution to the continuity equation

$$\partial_s \rho(s, x) + \nabla \cdot (\rho(s, x)V(x)) = 0,$$

with initial datum

$$\rho(0, x) = u(x),$$

with $u$ being the minimiser for $E$. Such a $u$ can be constructed by solving the characteristic differential equation

$$\frac{d}{ds} X(s, x) = V(X(s, x)),$$

coupled with the initial datum

$$X(0, x) = x,$$

locally in $s = 0$, with the local solution $X(s, x)$ being $C^1$, and by taking $\rho(s, x) := [(X(\cdot, s))_{#} u](s, x)$, i.e. $\rho(s, x)$ is defined via

$$\int_{\mathbb{R}^N} \phi(x) \rho(s, x) dx = \int_{\mathbb{R}^N} \phi(X(s, x)) u(x) dx, \quad \text{for all } \phi \in C^1_c(\mathbb{R}^N)$$

(cf. [9, Chapter 8, Proposition 8.1.8]). For all $s$ in the interval of existence of $\rho$ we have

$$\int_{\mathbb{R}^N} \rho(s, x) dx = 1, \quad \rho(s, x) \geq 0, \quad \text{a.e.}$$

and therefore the map $s \mapsto E[\rho(s, \cdot)]$ has a local minimum at $s = 0$. Hence

$$0 \leq \frac{d}{ds} E[\rho(s, \cdot)]|_{s=0} = \int_{\mathbb{R}^N} (\varepsilon \rho - G \ast \rho) \partial_s \rho \ dx|_{s=0}$$

$$= -\int_{\mathbb{R}^N} (\varepsilon \rho - G \ast \rho) \nabla \cdot (\rho V) \ dx|_{s=0} = \int_{\mathbb{R}^N} u \nabla (\varepsilon u - G \ast u) \cdot V dx$$

and replacing $V$ with $-V$ we obtain

$$0 \geq \int_{\mathbb{R}^N} u \nabla (\varepsilon u - G \ast u) \cdot V dx$$

and therefore

$$\int_{\mathbb{R}^N} u \nabla (\varepsilon u - G \ast u) \cdot V dx = 0, \quad \text{for an arbitrary } V \in C^1_c(\mathbb{R}^N)$$

which is the desired assertion. \hfill \square

Let us now compute the first and the second order Gâteaux derivatives of $E$. 
Lemma 7.2.5. Let \( u \in L^2 \cap \mathcal{P} \) be a solution to (7.2.6). Then, \( u \) is a stationary point for the energy functional \( E \). Moreover, the second order Gâteaux derivative of \( E \) on \( u \) satisfies
\[
\frac{d^2}{d\delta^2} E(u + \delta v)|_{\delta = 0} = \epsilon \int_{\mathbb{R}^N} v^2(x) dx - \int_{\mathbb{R}^N} v(x) G * v(x) dx,
\]
for all \( v = \text{div}(uV) \) and \( V \in C^1_c(\mathbb{R}^N) \).

Proof.

Suppose \( u \in L^2 \cap \mathcal{P} \) satisfies (7.2.6). Let us compute
\[
\lim_{\delta \to 0} \frac{1}{\delta} (E(u + \delta v) - E(u))
\]
with \( v = \nabla \cdot (uV) \) for an arbitrary vector field \( V \in C^1_c \), which implies \( \int_{\mathbb{R}^N} v(x) dx = 0 \). We obtain
\[
\frac{1}{\delta} (E(u + \delta v) - E(u))
= \frac{\epsilon}{2\delta} \int_{\text{supp}(u+\delta v)} (u + \delta v)^2 dx - \frac{\epsilon}{2\delta} \int_{\text{supp}(u)} u^2 dx
- \frac{1}{2\delta} \int_{\text{supp}(u+\delta v)} (u + \delta v) G * (u + \delta v) dx + \frac{1}{2\delta} \int_{\text{supp}(u)} u G * u dx.
\]

Therefore we easily get
\[
\lim_{\delta \to 0} \frac{1}{\delta} (E(u + \delta v) - E(u)) = \int_{\mathbb{R}^N} v(\epsilon u - G * u) dx
= \int_{\mathbb{R}^N} \nabla \cdot (uV)(\epsilon u - G * u) dx = - \int_{\mathbb{R}^N} uV \cdot \nabla (\epsilon u - G * u) dx.
\]

Therefore, \( u \) is a stationary point for \( E \) under the constraint \( \int_{\mathbb{R}^N} u dx = 1 \). The computation of the second derivative of \( E \) on \( u \) yields
\[
\frac{d^2}{d\delta^2} E(u + \delta v) = \frac{d^2}{d\delta^2} \frac{\epsilon}{2} \int_{\text{supp}(u+\delta v)} (u + \delta v)^2 dx
- \frac{d^2}{d\delta^2} \frac{1}{2} \int_{\text{supp}(u+\delta v)} (u + \delta v) G * (u + \delta v) dx
= \epsilon \int v^2 dx - \int v G * v dx
\]
which is independent on \( \delta \) and therefore it is valid also on \( \delta = 0 \). 

Now we introduce a very simple technical lemma which will be very useful in the sequel.
Lemma 7.2.6. Suppose \( u \in L^2 \cap \mathcal{P} \) is a solution to (7.2.6) having connected support. Then

\[
\varepsilon u(x) = \int_{\text{supp}(u)} G(x - y)u(y)dy + C
\]

for all \( x \in \text{supp}(u) \) with \( C = 2E(u) \). Moreover, in case \( \text{supp}(u) \) has infinite measure, then \( C = E(u) = 0 \).

Proof. It is immediate from (7.2.6) that

\[
\varepsilon u(x) = \int_{\text{supp}(u)} G(x - y)u(y)dy + C \tag{7.2.5}
\]

for all \( x \in \text{supp}(u) \) for a certain constant \( C \). Then, we multiply (7.2.5) by \( u(x) \) and integrate over \( \text{supp}(u) \) to obtain

\[
\varepsilon \int_{\text{supp}(u)} u^2(x)dx = \int_{\text{supp}(u)} \int_{\text{supp}(u)} G(x - y)u(y)u(x)dydx + C,
\]

where we have used that \( u \) has unit mass. It is therefore clear that \( C = 2E(u) \). Suppose now that \( \text{supp}(u) \) has infinite measure. Suppose by contradiction that \( C \neq 0 \). Let \( \{x_k\} \subset \text{supp}(u) \) be a sequence of points such that \( |x_k| \to +\infty \). We have, for all \( k \),

\[
\varepsilon u(x_k) - \int_{\text{supp}(u)} G(x_k - y)u(y)dy = C
\]

and therefore the same expression should hold in the limit \( k \to +\infty \). Now, the assumptions on \( G \) imply that the integral

\[
\int_{\text{supp}(u)} G(x_k - y)u(y)dy
\]

converges to zero as \( k \to +\infty \). This is due to Lebesgue’s dominated convergence Theorem. Therefore, the term \( u(x_k) \) has a limit \( C \) as \( k \to +\infty \). Such limit is the same for all diverging sequences of points \( \{x_k\} \subset \text{supp}(u) \), which means

\[
\lim_{x \in \text{supp}(u), |x| \to +\infty} u(x) = C.
\]

Now, since \( \text{supp}(u) \) has infinite measure, then \( C \neq 0 \) implies that \( u \) is not integrable, which is a contradiction. Therefore \( C = 0 \). \( \square \)
7.2.2 Stationary Solutions in multiple Dimensions

In this section we state the necessary and sufficient conditions on \( \varepsilon \) and \( \| G \|_{L^1} \) such that there exists non trivial steady states

\[
u \nabla (\varepsilon u - G \ast u) = 0
\]  

(7.2.6)

in the set \( L^2 \cap \mathcal{P} \). We refer to [16] for similar results based on ideas and strategies developed in [121] and [123].

Non existence of nontrivial steady states for \( \varepsilon > 1 \)

We start by covering the case \( \varepsilon > 1 \). Here, there exist no nontrivial steady states, as it follows from the following simple lemma.

Lemma 7.2.7. Let \( \varepsilon > 1 \). Then, there exists no stationary solutions to (7.2.6) in the space \( L^2 \cap \mathcal{P} \).

Proof. We first prove that there exists no minimiser for \( E(u) \) under the mass constraint \( \int_{\mathbb{R}^N} u = 1 \) and \( u \geq 0 \). To see this, we use Young inequality for convolutions as follows

\[
E(u) = \frac{\varepsilon}{2} \int_{\mathbb{R}^N} u^2 \, dx - \frac{1}{2} \int_{\mathbb{R}^N} u G \ast u \, dx
\]

(7.2.7)

\[
\geq \frac{\varepsilon}{2} \int_{\mathbb{R}^N} u^2 \, dx - \frac{\| G \|_{L^1}}{2} \int_{\mathbb{R}^N} u^2 \, dx = \frac{\varepsilon - 1}{2} \int_{\mathbb{R}^N} u^2 \, dx
\]

(7.2.8)

with \( \varepsilon - 1 > 0 \). Moreover, we have the simple estimate \( E(u) \leq C\| u \|_{L^2}^2 \). Take a family of functions \( u_\lambda(x) \geq 0 \) such that \( \int_{\mathbb{R}^N} u_\lambda(x) \, dx = 1 \) and \( \int_{\mathbb{R}^N} u_\lambda^2(x) \, dx \to 0 \) as \( \lambda \to +\infty \). To construct such a family, we just take a fixed \( L^2_+(\mathbb{R}^N) \) function \( u \neq 0 \) and rescale it by \( u_\lambda(x) = \lambda^{-d} u(\lambda^{-1} x) \). For such a family we therefore have

\[
E(u_\lambda) \to 0, \quad \text{as} \quad \lambda \to \infty.
\]

Therefore, it is impossible to have a minimiser \( u_\infty \) for \( E(u) \) in the set \( \{ u \in L^1_+ : \int_{\mathbb{R}^N} u = 1 \} \) because (7.2.7) would imply that \( E(u_\infty) > 0 \) and we would necessarily have \( 0 < E(u_\lambda) < E(u_\infty) \) for \( \lambda \) large enough.

Now we prove that there exist no steady states. Suppose by contradiction that \( u \) is a steady state. Then, due to Lemma 7.2.5 \( u \) is a stationary point for \( E \). Moreover, the formula (7.2.4) implies that the functional \( E \) is strongly convex, and therefore admits only one stationary point, which coincides with its global minimiser. But this
contradicts the non existence of a global minimiser proven above. Now we prove that there exist non steady states. Suppose $u \in L^2_+$ satisfies (7.2.6) with $\varepsilon > 1$. With $\lim_{\delta \to 0} \frac{1}{\delta} (E(u + \delta v) - E(u)) = \int_{\mathbb{R}^N} \nabla \cdot (uV)(\varepsilon u - G \ast u)dx = -\int_{\mathbb{R}^N} uV \cdot \nabla(\varepsilon u - G \ast u)dx.$

with $v = \nabla \cdot (uV)$ for an arbitrary vector field $V \in C^1_c$, which implies $\int_{\mathbb{R}^N} v(x)dx = 0$, we obtain, that $u$ is a stationary point for $E$ under the constraint $\int udx = 1$. Now, the computation of the second derivative of $E$ on $u$ yields

$$\frac{d^2}{d\delta^2} E(u + \delta v) = \varepsilon \int v^2dx - \int vG \ast vdx \geq (\varepsilon - 1) \int v^2dx$$

and therefore $E$ is convex and it has its global minimum (under the constraint of unit mass) as the only stationary point. Which implies that $u$ is a minimiser for $E$, which contradicts the first step of the proof, since we have proven that $E$ cannot have a minimiser in the set $\int udx = 1$. \hfill \Box

The critical case $\varepsilon = 1$

We aim to solve

$$0 = \nabla \cdot (u \nabla (u - G \ast u)). \quad (7.2.9)$$

We shall prove that no $L^2 \cap \mathcal{P}$ steady states exist in this case.

**Theorem 7.2.8** (Non-existence of nontrivial steady states for $\varepsilon = 1$). There exists no solutions to (7.2.9) in $L^2 \cap \mathcal{P}$.

**Proof.** From the Cauchy–Schwarz inequality we know that

$$\int_{\mathbb{R}^N} uG \ast udx \leq \|u\|_{L^2(\mathbb{R}^N)} \|G \ast u\|_{L^2(\mathbb{R}^N)}$$

and the equality in the above formula holds if and only if $u$ and $G \ast u$ are proportional. In terms of the functional $E$ this means that

$$E(u) \geq 0 \quad \text{for all} \quad u \in L^2(\mathbb{R}^N) \cap \mathcal{P}(\mathbb{R}^N).$$

As in Lemma 7.2.7 we have the estimate $E(u) \leq C\|u\|_{L^2}^2$, and using once again the family $u_\lambda$ of Lemma 7.2.7 we see that $\inf_{u \in L^2 \cap \mathcal{P}} E(u) = 0$. Assume by contradiction
that there exists a stationary solution \( u_\infty \). Then, due to the result in Lemma 7.2.5 and in view of Cauchy–Schwarz inequality, the second order derivative of \( E \) is nonnegative everywhere. Hence, the functional \( E \) is convex and therefore \( u_\infty \) is a global minimiser for \( E \) under the constraint \( u \in L^2 \cap \mathcal{P} \). Then, we must have \( E(u_\infty) = 0 \), which means that \( u \) and \( G \ast u \) are proportional, i.e. there exists a constant \( \lambda \in \mathbb{R}_+ \) such that

\[
u_\infty(x) = \lambda G \ast u_\infty(x) \quad \text{(7.2.10)}
\]

almost everywhere on \( \mathbb{R}^N \). Integrating (7.2.10) over \( \mathbb{R}^N \) yields

\[1 = \lambda \|G\|_{L^1(\mathbb{R}^N)} = \lambda\]

and hence

\[
u_\infty(x) = G \ast u_\infty(x) \quad \text{(7.2.11)}
\]

almost everywhere on \( \mathbb{R}^N \). We can then apply the Fourier transform

\[\hat{f}(\xi) = \int_{\mathbb{R}^N} e^{-2\pi i x \cdot \xi} f(x) dx\]

to both members of the equation (7.2.11) to obtain

\[\hat{\nu}_\infty(\xi) = \hat{G}(\xi) \hat{u}_\infty(\xi), \quad \xi \in \mathbb{R}^N.\]

We have

\[|\hat{G}(\xi)| \leq \int_{\mathbb{R}^N} |G(x)| dx = 1.\]

Moreover, since \( G \) is even, then \( \hat{G}(\xi) < 1 \) for all \( \xi \neq 0 \). In order to see that, write

\[\hat{G}(\xi) = \int_{\mathbb{R}^N} \prod_{k=1}^d e^{-2\pi i x_k \xi_k} G(x) dx = \int_{\mathbb{R}^N} \prod_{k=1}^d (\cos(2\pi x_k \xi_k) - i \sin(2\pi x_k \xi_k)) G(x) dx,\]

then \( G \) being even easily implies that only real valued contributions survive in the above integral; such real valued contributions are of the form

\[\int_{\mathbb{R}^N} f_{h,k}(x, \xi) G(x) dx\]

where the functions \( f_{h,k} \) are such that \( |f_{h,k}(x, \xi)| \leq 1 \) and \( |f_{h,k}(x, \xi)| < 1 \) for \( x \) ranging on a set of positive measure. Therefore, we have proven that

\[\hat{\nu}_\infty(\xi) = 0 \quad \text{for all} \quad \xi \neq 0\]

and \( \hat{\nu}_\infty(0) = 1 \). This implies that \( u(x) = 0 \) almost everywhere, which contradicts the fact that \( u \) has unit mass. \( \square \)
Stationary solutions for $\varepsilon < 1$

Let us now provide a minimiser for the entropy functional in the case $\varepsilon < 1$, which implies the existence of a nontrivial $L^2 \cap P$ steady state for (7.2.6) in view of Proposition 7.2.4. Such result is proven rigorously in [16, Theorem 1], which we recall here.

Theorem 7.2.9 (Existence of minimisers, [16]). Let $\varepsilon < 1$. Then, there exists a radially symmetric non-increasing minimiser $u \in P \cap L^2(\mathbb{R}^N)$ for the entropy functional $E$ restricted to $P$ with $u \neq 0$.

We refer to [16] for the details of the proof, which is based on a sort of subadditivity property needed to provide suitable compactness of the minimising sequence, cf. [16, Lemma 2]. For the sake of clarity, we shall still provide the simple proof of the fact that global minima of $E$ under mass constraint are strictly negative, which forces the minimiser to be non zero.

Lemma 7.2.10. Let $\varepsilon < 1$ and $u \in P \cap L^2(\mathbb{R}^N)$. Then, $\inf_u E(u) < 0$.

Proof. We consider the family $\sigma_\lambda \in L^2 \cap P$ in one dimension, with

$$\sigma_\lambda(x) = \frac{1}{2\lambda} \chi_{[-\lambda,\lambda]}(x).$$

For $\varepsilon < 1$ we have

$$E[\sigma_\lambda] = \frac{\varepsilon}{4\lambda} - \frac{1}{8\lambda^2} \int_{-\lambda}^{\lambda} \int_{-\lambda}^{\lambda} G(x-y)dydx$$

$$= \frac{\varepsilon}{4\lambda} - \frac{1}{4\lambda} \int_{-\lambda}^{\lambda} G(z)dz = \frac{1}{4\lambda} \left( \varepsilon - \int_{-\lambda}^{\lambda} G(z)dz \right)$$

and since

$$\int_{-\lambda}^{\lambda} G(z)dz \to 1 \quad \text{as} \quad \lambda \to +\infty,$$

we easily obtain that there exists a $\lambda$ such that $E[\sigma_\lambda] < 0$. \hfill $\Box$

7.2.3 Stationary Solutions in the one dimensional Case

In this section we prove the main result, namely that nontrivial stationary solutions (which always exist in the case $\varepsilon < 1$) in one space dimension with fixed mass and center of mass are unique. First we provide certain sufficient and necessary conditions on the steady states and then we prove that they are unique under such conditions. We start with a necessary condition on the steady states which deals with a property of their support.
Lemma 7.2.11 (Steady states have connected support). Let \( u \) be a stationary solution to (7.2.6) in one space dimension, namely
\[
 u \partial_x (\varepsilon u - G * u) = 0 \quad \text{a.e. on } \mathbb{R}.
\] (7.2.12)
Then, \( \text{supp}(u) \) is a connected set.

Proof. Let \( u \) solve (7.2.12). Let us first assume that \( u \) is compactly supported. Suppose that \( \text{supp}(u) \) is not connected. Accordingly, let \([a, b]\) be a non trivial interval such that
\[
\begin{align*}
u(x) &
eq 0 \quad \text{if } x < a, \\
u(x) &= 0 \quad \text{if } a \leq x \leq b, \\
u(x) &
eq 0 \quad \text{if } x > b.
\end{align*}
\] (7.2.13)
Let us introduce the velocity field
\[
V(x) := \begin{cases} -1 & \text{if } x \in (-\infty, a) \\ 1 & \text{if } x \in (b, +\infty) \end{cases}
\]
and let \( V \in C^1(\mathbb{R}) \). Let \( \rho(s, x) \) be a local solution to the Cauchy problem for the continuity equation
\[
\begin{align*}
\partial_s \rho + \partial_x (\rho V) &= 0 \\
\rho(0, x) &= \rho(x).
\end{align*}
\]
Let us compute the evolution of the energy \( E \) along \( \rho \) at the time \( s = 0 \):
\[
\frac{d}{ds} E[\rho(s)]|_{s=0} = \int \rho_s(\varepsilon \rho(s, x) - G \ast \rho(s, x))dx|_{s=0} = \int u V \partial_x (\varepsilon u - G \ast u)dx = 0.
\]
Then, by definition of \( V \) we have
\[
\varepsilon \int_{\mathbb{R}} u V \partial_x u = \frac{\varepsilon}{2} \int_{-\infty}^{a} \partial_x u^2 dx + \frac{\varepsilon}{2} \int_{a}^{b} V \partial_x u^2 dx - \frac{\varepsilon}{2} \int_{b}^{+\infty} \partial_x u^2 dx = 0,
\] (7.2.14)
because \( \partial_x u = 0 \) on \([a, b]\) and \( u = 0 \) on \( x = a, b \) and at \( \pm \infty \). Hence, we have
\[
0 = \int_{-\infty}^{+\infty} u V \partial_x G \ast u dx = - \int_{-\infty}^{a} u G' \ast u dx + \int_{b}^{+\infty} u G' \ast u dx.
\] (7.2.15)
We compute
\[
\int_{-\infty}^{a} u G' \ast u dx = \int_{-\infty}^{a} \int_{-\infty}^{a} u(x) G'(x-y)u(y)dydx + \int_{-\infty}^{a} \int_{b}^{+\infty} u(x) G'(x-y)u(y)dydx,
\]
the first term on the above right-hand side is zero since $G'$ is odd and the integration
domain is symmetric in $x$ and $y$. Since $G'(z) \geq 0$ as $z \leq 0$, we have for the second
term
\[ u(x)G'(x - y)u(y) \geq 0 \text{ on } (x, y) \in (-\infty, a) \times (b, +\infty). \]
In a similar way one can prove that
\[ \int_{b}^{+\infty} uG' \ast u dx = \int_{b}^{+\infty} dx \int_{-\infty}^{a} dyu(x)G'(x - y)u(y) \]
with the integrand $u(x)G'(x - y)u(y) \leq 0$ on the integration domain. Therefore,
(7.2.15) implies that
\[ u(x)u(y) \equiv 0 \text{ on } \{ x < a \} \cap \{ y > b \}. \quad (7.2.16) \]
We have thus proven that, whenever (7.2.13) holds, then (7.2.16) has to be necessarily
satisfied. Let $A, B$ be two nonempty connected components of $\text{supp}(u)$ and let $[\alpha, \beta]$
be the maximal interval such that
\[ a < b, \text{ for all } a \in A, \ b \in B \]
\[ u \equiv 0 \text{ on } [\alpha, \beta] \]
\[ \alpha \geq a, \text{ for all } a \in A \]
\[ \beta \leq b, \text{ for all } b \in B. \]
Then, $u(x)u(y) = 0$ for all $(x, y)$ such that
\[ x < \alpha, \ y > \beta, \]
which implies that either $A$ or $B$ cannot be in the support of $u$, and that is a contradiction.
In order to generalise the proof to a stationary solution $u$ which is not compactly
supported, one can cutoff $u$ to have compact support in such a way that the $L^2$ norm of
the compactly supported approximation is arbitrarily close to the $L^2$ norm of $u$. Then,
the estimate
\[ \varepsilon \int |u\partial_x u| dx \leq \|G\|_{L^1} \|u\|_{L^2}^2 \]
implies that the integrals in (7.2.14) converge at infinity, therefore all the above com-
putations are valid up to an arbitrary difference which vanishes in the limit.
\[ \square \]
Remark 7.2.12. In the case $\text{supp}(G) = [-R, R]$ one can use the same strategy as in
Lemma 7.2.11 to prove that, given two connected components $A, B$ of $\text{supp}(u)$ one has
$\text{dist}(A, B) > 2R$.

We now exploit a standard symmetric rearrangement technique to prove that the
minimisers of the energy are symmetric and monotonically decreasing on $x > 0$ under
the constraint of zero center of mass, cf. [121] [122].
Proposition 7.2.13. Let \( u_\infty \) be a minimiser for the energy

\[
E(u) = \frac{\varepsilon}{2} \int u^2(x) dx - \frac{1}{2} \int \int G(x - y) u(x) u(y) dy dx
\]

under the constraint that the center of mass is zero. Then, \( u_\infty \) is symmetric and monotonically decreasing on \( x > 0 \).

Proof. We have to prove, that the energy (7.2.2) increases, when a function \( u \) is replaced by a symmetric rearrangement

\[
u^*(x) = \sup \{ t \geq 0 : \{ u > t \} > 2 |x| \}.
\]

For every exponent \( p \geq 1 \) the following holds:

\[
\int_R (u^*)^p dx = \int_R (u)^p dx. \tag{7.2.18}
\]

Furthermore we know, that the rearrangement does not change the measure, which means that for a \( p \in \mathbb{R} \) and \( a, b \in \mathbb{R} \) with \( a < b \) we get

\[
\{ u > p \} = (a, b) \Rightarrow \{ u^* > p \} = \left( \frac{a - b}{2}, \frac{b - a}{2} \right) = (-c, c). \tag{7.2.19}
\]

Suppose that \( \sup_x u(x) = M, \forall x \in \mathbb{R} \) for a positive constant \( M \). Then we can rewrite the energy functional:

\[
E(u) = \frac{\varepsilon}{2} \int u^2 dx - \frac{1}{2} \int \int G(x - y)(M - u(x))(M - u(y)) dx dy + C,
\]

\[
= \frac{\varepsilon}{2} \int u^2 dx - \frac{1}{2} \int \int G(x - y) \int_{u(x)}^M dp \int_{u(y)}^M dq dx dy + C,
\]

\[
= \frac{\varepsilon}{2} \int u^2 dx - \frac{1}{2} \int_0^M \int_0^M \int_{\{ u > p \}} \int_{\{ u > q \}} G(x - y) dp dq dx dy + C.
\]

Using (7.2.18), we have to prove that the following inequality holds:

\[
\int_0^M \int_0^M \int_{\{ u > p \}} \int_{\{ u > q \}} G(x - y) dp dq dx dy > \int_0^M \int_0^M \int_{\{ u > p \}} \int_{\{ u > q \}} G(x - y) dp dq dx dy.
\]

For \( p > q \) we get \( \{ u > p \} \subset \{ u > q \} \) and \( \{ u^* > p \} \subset \{ u^* > q \} \). Furthermore, there exists a \( z \in \mathbb{R} \) such that \( \{ u > p \} - z = \{ u^* > p \} \). So we can change the variables \( x \to x - z \) and \( y \to y - z \) to obtain

\[
\int_{\{ u > p \}} \int_{\{ u > q \}} G(x - y) dx dy = \int_{\{ u^* > p \}} \int_{\{ u^* > q \} - z} G(x - y) dx dy.
\]
Due to the fact that
\[
\int_{\{u > p\}} \int_{\{u > p\}} G(x - y) \, dx \, dy = \int_{\{u^* > p\}} \int_{\{u^* > p\}} G(x - y) \, dx \, dy.
\]
it is left to prove that
\[
\int_{\{u^* > p\}} \int_{(\{u > q\} - z) \cap \{u^* > p\}} G(x - y) \, dx \, dy < \int_{\{u^* > p\}} \int_{\{u^* > q\} \cap \{u^* > p\}} G(x - y) \, dx \, dy. \tag{7.2.20}
\]
we can write (7.2.20) in the following way
\[
\int_{-c}^{c} \int_{-d}^{-c} G(x - y) \, dx \, dy + \int_{-c}^{c} \int_{c}^{e} G(x - y) \, dx \, dy
\]
\[
< \int_{-c}^{c} \int_{\frac{e-d}{2}}^{c} G(x - y) \, dx \, dy + \int_{-c}^{c} \int_{\frac{e-d}{2}}^{e} G(x - y) \, dx \, dy
\]
\[
\Rightarrow \int_{-c}^{c} \int_{\frac{e-d}{2}}^{e} G(x - y) \, dx \, dy < \int_{-c}^{c} \int_{\frac{d-c}{2}}^{d} G(x - y) \, dx \, dy
\]

Eventually, we get
\[
\int_{-c}^{c} \left( \int_{\frac{e-d}{2}}^{e} G(x - y) \, dx - \int_{\frac{e-d}{2}}^{-d} G(x - y) \, dx \right) \, dy < 0
\]
\[
\int_{-c}^{c} \left( \int_{\frac{e-d}{2}}^{e} \left( G(x - y) - G(x - \frac{e + d}{2} - y) \right) \, dx \right) \, dy < 0
\]

We know, that \( x - y > 0 \) and \( x - y > x - y - \frac{e+d}{2} \), because \( e > |d| \). Furthermore, \( x - y - \frac{e+d}{2} > 0 \), because
\[
\frac{e - d}{2} - c - \frac{e + d}{2} > 0
\]
\[-d - c > 0
\]
with \(-d > c\). Therefore the inequality
\[
\int_{-c}^{c} \left( \int_{\frac{e-d}{2}}^{e} \left( G(x - y) - G(x - \frac{e + d}{2} - y) \right) \, dx \right) \, dy < 0
\]
is true.

Let us rephrase Lemma 7.2.6 in the one-dimensional case.
Lemma 7.2.14. Let $u$ be a steady state in one dimension, i.e.
\[ \varepsilon u = G * u + C \quad \text{on} \ supp(u) \]
for some $C \in \mathbb{R}$. Then, $C = 2E(u)$.

Proof. The support of $u$ is connected in view of Lemma 7.2.11, therefore Lemma 7.2.6 applies.

Lemma 7.2.15. Let $u \in \mathcal{P} \cap L^2$ and let $x_0 \in \mathbb{R}$. Let $u_{x_0}$ be defined by
\[ u_{x_0}(x) := u(x + x_0). \]
Then, $E(u_{x_0}) = E(u)$.

Proof.

Let us compute
\[
E(u_{x_0}) = \frac{\varepsilon}{2} \int u^2(x + x_0)dx - \frac{1}{2} \int \int G(x - y)u(x + x_0)u(y + x_0)dydx
\]
\[
= \frac{\varepsilon}{2} \int u^2(x)dx - \frac{1}{2} \int \int G((x - x_0) - (y - x_0))u(x)u(y)dydx
\]
\[
= \frac{\varepsilon}{2} \int u^2(x)dx - \frac{1}{2} \int \int G(x - y)u(x)u(y)dydx = E(u).
\]

Lemma 7.2.16. Let $u$ be a steady state with $\varepsilon < 1$. Then, the support of $u$ is compact.

Proof. We know from Lemma 7.2.11 that the support of $u$ is a connected set. Suppose that supp$(u)$ is not bounded. That means that supp$(u)$ is of the form $(-\infty, b)$ (b possibly $+\infty$) or $(a, +\infty)$ (a possibly $-\infty$). Assume first supp$(u) = (a, +\infty)$. Then, Lemma 7.2.14 implies
\[
2E(u) = \varepsilon u(x) - \int_a^{+\infty} G(x - y)u(y)dy = 0
\]
for all $x \in (a, +\infty)$. Now, there are two possibilities: either $a = -\infty$ or $a > -\infty$. In the latter case, evaluation on $x = a$ implies
\[ 0 = \varepsilon u(a) = \int_a^{+\infty} G(a - y)u(y)dy \]
which is a contradiction because the integral on the right hand side is strictly positive in view of \( \text{supp}(G) = \mathbb{R} \). In the former case \( a = -\infty \) we have then \( \text{supp}(u) = \mathbb{R} \), which implies

\[
\varepsilon u(x) = \int_{-\infty}^{+\infty} G(x-y)u(y)dy
\]

for all \( x \in \mathbb{R} \). We can therefore integrate over \( \mathbb{R} \) to obtain

\[
\varepsilon = \|G\|_{L^1} = 1
\]

which is a contradiction. The same proof can be produced in the case \( \text{supp}(u) = (-\infty, b) \). Therefore, the support of \( u \) can only be a bounded interval.

**Lemma 7.2.17.** Let \( u \) be a steady state. Then there exists a symmetric steady state \( \tilde{u} \) such that

\[
E(\tilde{u}) = E(u).
\]

**Proof.** From Lemma 7.2.11 and Lemma 7.2.16 we know that \( \text{supp}(u) = (a, b) \) for some \( a, b \in \mathbb{R} \). For a given \( x \in (a, b) \) we have

\[
\varepsilon u(x) = G \ast u(x) + C
\]

for some \( C \in \mathbb{R} \). Evaluation on \( x = a \) and \( x = b \) gives

\[
C = -\int_a^b G(a-y)u(y)dy = -\int_a^b G(b-y)u(y)dy.
\]

Let \( \bar{u}(x) = u(x + x_0) \) with \( x_0 = (a+b)/2 \). Then \( \bar{u} \) is still a steady state and it satisfies \( E(\bar{u}) = E(u) \) thanks to Lemma 7.2.15. Moreover, the support of \( \bar{u} \) is symmetric. Let us introduce

\[
\tilde{u}(x) := \frac{1}{2}(\bar{u}(x) + \bar{u}(-x)).
\]

Clearly, \( \text{supp}[\bar{u}] = \text{supp}[\tilde{u}] \) and we have, for all \( x \in \text{supp}[\tilde{u}] \),

\[
\varepsilon \tilde{u}(x) = \frac{\varepsilon}{2}(\bar{u}(x) + \bar{u}(-x)) = \frac{\varepsilon}{2}(u(x + x_0) + u(-x + x_0))
\]

\[
= \frac{1}{2} \int_a^b G(x + x_0 - y)u(y)dy + \frac{1}{2} \int_a^b G(-x + x_0 - y)u(y)dy + C
\]

\[
= \frac{1}{2} \int_{(a-b)/2}^{(b-a)/2} G(x-z)\bar{u}(z)dz + \frac{1}{2} \int_{(a-b)/2}^{(b-a)/2} G(-x-z)\bar{u}(z)dz + C
\]

\[
= \frac{1}{2} \int_{(a-b)/2}^{(b-a)/2} G(x-z)\bar{u}(z)dz + \frac{1}{2} \int_{(a-b)/2}^{(b-a)/2} G(x-z)\bar{u}(z)dz + C
\]

\[
= \int_{(a-b)/2}^{(b-a)/2} G(x-z) \frac{1}{2} (\bar{u}(z) + \bar{u}(-z)) dz + C = \int_{(a-b)/2}^{(b-a)/2} G(x-z)\bar{u}(z)dz + C
\]

where we have used the symmetry of \( G \). The above computation shows that \( \tilde{u} \) has the same energy as \( u \) in view of the results in Lemma 7.2.15 and Lemma 7.2.14. \( \square \)
Lemma 7.2.18 (support of a minimiser). Let \( u_\infty \) be a global minimiser to \( E \). Let \( u \) be a steady state such that 
\[
\text{meas}(\text{supp}[u_\infty]) \leq \text{meas}(\text{supp}(u)).
\]
Then \( u \) is also a minimiser.

Proof. Assume first that we are in the special case \( \text{supp}[u_\infty] \subseteq \text{supp}(u) \). Let us compute the second variation of \( E \) around the minimiser \( u_\infty \) along the direction \( u_\infty - u \).

\[
\frac{d^2}{d\delta^2} E(u_\infty + \delta(u - u_\infty))|_{\delta=0} = \varepsilon \int (u - u_\infty)^2 dx - \int \int G(x - y)(u(x) - u_\infty(x))(u(y) - u_\infty(y))dydx
= 2E(u) + 2E(u_\infty) - 2 \int_{\text{supp}[u_\infty]} u_\infty(\varepsilon u - G * u)dx
= 2E(u) + 2E(u_\infty) - 4E(u)
\]
where the last step is justified by the fact that \( \text{supp}(u) \subseteq \text{supp}[u_\infty] \). Therefore, since \( u_\infty \) is a minimiser, the second derivative above is nonnegative, i. e.
\[
0 \leq \frac{d^2}{d\delta^2} E(u_\infty + \delta(u - u_\infty))|_{\delta=0} = 2(E(u_\infty) - E(u)),
\]
which yields \( E(u) \leq E(u_\infty) \). Since \( u_\infty \) is a minimiser, then so is \( u \). In the general case in which \( \text{supp}[u_\infty] \not\subseteq \text{supp}(u) \), consider a translation \( u_{x_0}(x) = u(x - x_0) \) in such a way that the support of \( u \) contains the support of \( u_\infty \). Since the energy is invariant after translation in view of Lemma 7.2.15, the assertion is proven.

Let us state the Krein–Rutman Theorem (see e. g. [69] and the references therein).

Theorem 7.2.19 (Krein–Rutman Theorem). Let \( X \) be a Banach space, let \( K \subset X \) be a total cone, i. e. such that \( \lambda K \subseteq K \) for all \( \lambda \geq 0 \) and such that the set \( \{ u - v, \ u, v \in K \} \) is dense in \( X \). Let \( T \) be a compact linear operator such that \( T(K) \subset K \) with positive spectral radius \( r(T) \). Then \( r(T) \) is an eigenvalue for \( T \) with an eigenvector \( u \in K \setminus \{0\} \).

An important consequence [69] of the Krein–Rutman theorem, which will be extremely useful in the sequel, is the following

Theorem 7.2.20 (Krein–Rutman Theorem, strong version). Let \( X \) be a Banach space, \( K \subset X \) a solid cone, i. e. such that \( \lambda K \subseteq K \) for all \( \lambda \geq 0 \) and such that \( K \) has a nonempty interior \( K_0 \). Let \( T \) be a compact linear operator which is strongly positive with respect to \( K \), i. e. such that \( T[u] \in K_0 \) if \( u \in K \). Then,
(i) The spectral radius $r(T)$ is strictly positive and $r(T)$ is a simple eigenvalue with an eigenvector $v \in K_0$. There is no other eigenvalue with a corresponding eigenvector $v \in K$.

(ii) $|\lambda| < r(T)$ for all other eigenvalues $\lambda \neq r(T)$.

We shall now prove the uniqueness of symmetric steady states with unit mass which are monotonically decreasing on the positive semi-axis in the case $\epsilon < 1$. We already know that under the above assumptions we can write, for $x \in [-L, L] = \text{supp}(u)$,

$$
\epsilon u(x) = \int_{-L}^{L} G(x - y)u(y)dy + C, \quad C = 2E(u). \quad (7.2.22)
$$

Taking the derivative w.r.t $x \in [-L, L]$ we obtain

$$
\epsilon u'(x) = \frac{d}{dx} \int_{-L}^{L} G(x - y)u(y)dy = \frac{d}{dx} G \ast u(x) = \int_{-L}^{L} G(x - y)u'(y)dy.
$$

The symmetry of $u$ and $G$ implies, for $x \in [0, L]$,

$$
\epsilon u'(x) = -\int_{0}^{L} G(x+y)u'(y)dy + \int_{0}^{L} G(x-y)u'(y)dy = \int_{0}^{L} [G(x - y) - G(x + y)] u'(y)dy.
$$

Assuming that $u \in C^1([-L, L])$, finding a steady state with the above assumptions is equivalent to find $u$ on $[0, L]$ such that

- $u(L) = 0$,
- $-\rho'(x) = u(x), \quad x \in [0, L]$,
- $\rho \geq 0$, and $\rho$ solves $\epsilon \rho = \int_{0}^{L} H(x, y)\rho(y)dy$,
- $H(x, y) = G(x - y) - G(x + y)$.

To convince ourselves about that, integrate

$$
-\epsilon u'(x) = -\int_{0}^{L} (G(x - y) - G(x + y)) u'(y)dy
$$

over the interval $[\xi, L]$ for some $\xi \in [0, L)$. Then $u(L) = 0$ and integration by parts
imply
\[-\varepsilon u(\xi) = -\int_{\xi}^{L} dx \int_{0}^{L} (G(x - y) - G(x + y)) u'(y) dy = -\int_{\xi}^{L} dx \left[ (G(x - L) - G(x + L)) u(L) - (G(x) - G(x)) u(0) \right] + \int_{\xi}^{L} dx \int_{0}^{L} \left( -G'(x - y) - G'(x + y) \right) u(y) dy \]
\[= \int_{0}^{L} u(y) dy \int_{\xi}^{L} \left( -G'(x - y) - G'(x + y) \right) dx = \int_{0}^{L} u(y) \left[ -G(L - y) - G(L + y) + G(\xi - y) + G(\xi + y) \right] \]
which implies, by the symmetry of \(G\),
\[\varepsilon u(x) = \int_{-L}^{L} G(x - y) u(y) dy + C, \quad C = -\int_{0}^{L} (G(L - y) + G(L + y)) u(y) dy.\]
For further reference, we introduce the operator
\[G_{L}(u)(x) := \int_{0}^{L} \left[ G(x - y) + G(x + y) - G(L - y) - G(L + y) \right] u(y) dy \tag{7.2.23}\]
on the Banach space
\[Y_{L} := \{ u \in C([0, L]) : u(L) = 0 \}.\]
In order to simplify the notation, we also define the following operator
\[H_{L}[\rho](x) := \int_{0}^{L} H(x, y) \rho(y) dy = \int_{0}^{L} \left( G(x - y) - G(x + y) \right) \rho(y) dy.\]

**Proposition 7.2.21.** For a fixed \(L > 0\) there exists a unique symmetric function \(u \in C^{2}([-L, L])\) with unit mass and with \(u'(x) \leq 0\) on \(x \geq 0\) such that \(u\) solves (7.2.22) for some \(\varepsilon = \varepsilon(L) > 0\). Such function \(u\) also satisfies \(u''(0) < 0\). Moreover, \(\varepsilon(L)\) is the largest eigenvalue of the compact operator \(G_{L}\) on the space Banach \(Y_{L}\) and any other eigenfunction of \(G_{L}\) on \(Y_{L}\) with unit mass has the corresponding eigenvalue \(\varepsilon'\) satisfying \(|\varepsilon'| < \varepsilon(L)\).

**Proof.** Since \(G\) is decreasing on the half-line \([0, +\infty)\) we get
\[H(x, y) = G(x - y) - G(x + y) \geq 0, \quad \text{on} \quad x, y \geq 0.\]
Consider now the Banach space
\[X_{L} = \{ f \in C^{1}([0, L]) : f(0) = 0 \}\]
endowed with the $C^1$ norm
\[ \|f\|_{X_L} = \|f\|_{L^\infty([0,L])} + \|f'\|_{L^\infty([0,L])}. \]

It can be easily seen that the set
\[ K := \{ f \in X : f \geq 0 \} \]
is a solid cone in $X$. Indeed, any function $f \in K$ with $f'(0) > 0$ is in the interior of $K$. Moreover, for a given $\rho \in K$, we have
\[ H_L[\rho](x) = \int_0^L H(x,y)\rho(y)dy \geq 0 \]
for all $x \in [0,L]$ and
\[ H_L[\rho](0) = \int_0^L H(0,y)\rho(y)dy = \int_0^L (G(-y) - G(y))\rho(y)dy = 0. \]
Therefore $H$ is a positive operator in the sense provided by the definition of the cone $K$. Indeed, we can prove that $H$ is strongly positive, i.e., for a given $\rho \in K$, $H[\rho]$ belongs to the interior of $K$. In order to see that, for a $\rho \in K \setminus \{0\}$ compute
\[ (H_L[\rho])'(0) = \int_0^L (G'(-y) - G'(y))\rho(y)dy = -2 \int_0^L G'(y)\rho(y)dy > 0, \]
and therefore $H_L[\rho]$ belongs to the interior of $K$. Hence, we can apply the stronger version of Krein–Rutman theorem 7.2.20, which implies the existence of a simple eigenvalue $\epsilon > 0$ equal to the spectral radius of $H_L$. More precisely, there exists a family of solutions $u$ to
\[ \epsilon \rho = H_L[\rho] \]
generated by one given nontrivial element $\bar{\rho}$ in the interior of $K$. This implies that the corresponding set of symmetric and monotone $u$ solving (7.2.22) satisfies
\[ u(x) = u(x) - u(L) = -\int_x^L u'(y)dy = \int_x^L u(y)dy = \alpha \int_x^L \bar{u}(y)dy, \]
for $\alpha > 0$. We choose $\alpha$ as follows
\[ \alpha = \left(2 \int_0^L \int_x^L \bar{\rho}(y)dydx\right)^{-1}, \]
and we obtain that $u$ has unit mass on $[-L,L]$. It is clear that $u'(x) \leq 0$ for $x \geq 0$, $u'(0) = 0$, and $u''(0) < 0$. In view of the statement (i) of theorem 7.2.20 there exists no other eigenvalues to $H_L$ with eigenvectors in $K$ besides the one $\bar{\epsilon}$ with eigenfunction $\bar{u}$, and all other eigenvalues $\epsilon'$ with eigenfunctions in $X_L$ satisfy $|\epsilon'| < \epsilon$. □

The eigenvalue $\epsilon$ (which coincides with the spectral radius of $H_L$) can be considered as a function of $L$, namely $\epsilon = \epsilon(L)$. The behavior of such function is established in the next proposition.
Proposition 7.2.22 (Behavior of the function $\epsilon(L)$). The simple eigenvalue $\epsilon(L)$ found in Proposition 7.2.21 is uniquely determined as a function of $L$ with the following properties

(i) $\epsilon(L)$ is strictly increasing with respect to $L$

(ii) $\lim_{L \to +\infty} \epsilon(L) = 1$

(iii) $\epsilon(0) = 0$.

Proof. In order to prove (i), let us consider the equation

$$\epsilon(L) \rho_L(x) = \mathcal{H}_L[\rho_L](x) = \int_0^L H(x, y) \rho_L(y) dy, \quad x \in [0, L],$$

where $\rho_L$ is the unique eigenfunction obtained in Proposition 7.2.21. We multiply the above equation by $\rho_L(x)$ and integrate over $[0, L]$ to obtain

$$\epsilon(L) \int_0^L \rho_L^2(x) dx = \int_0^L \mathcal{H}_L[\rho_L](x) \rho(x) dx.$$

Recall that the eigenvalue $\rho_L$ satisfies $\rho_L(0) = 0$ and, for $x \in (0, L]$,

$$\rho_L(x) = \frac{1}{\epsilon(L)} \int_0^L H(x, y) \rho(y) dy > 0$$

since $H(x, y) = G(x - y) - G(x + y) > 0$ for all $y \in [0, L]$ under the assumption $x > 0$ in view of the strict decreasing monotonicity of $G$ on $x > 0$. For a general $L \in (0, +\infty)$ and a $\delta > 0$ (small enough) we have

$$I_1 := \epsilon(L + \delta) \int_0^{L+\delta} \rho_{L+\delta}^2(x) dx - \epsilon(L) \int_0^L \rho_L^2(x) dx$$

$$= \int_0^{L+\delta} \mathcal{H}_{L+\delta}[\rho_{L+\delta}](x) \rho_{L+\delta}(x) dx - \int_0^L \mathcal{H}_L[\rho_L](x) \rho_L(x) dx =: I_2. \quad (7.2.24)$$

We analyze the two terms $I_1$ and $I_2$ separately. $I_1$ can be expanded as follows:

$$I_1 = (\epsilon(L + \delta) - \epsilon(L)) \int_0^{L+\delta} \rho_{L+\delta}^2(x) dx$$

$$+ \epsilon(L) \int_0^{L+\delta} (\rho_{L+\delta}(x) - \rho_L(x)) (\rho_{L+\delta}(x) + \rho_L(x)) dx + \epsilon(L) \int_L^{L+\delta} \rho_L^2(x) dx.$$
$I_2$ is given by

$$I_2 = \int_0^{L+\delta} (\mathcal{H}_{L+\delta}[\rho_{L+\delta}](x) - \mathcal{H}_L[\rho_L](x)) \rho_{L+\delta}(x) dx$$

$$+ \int_L^{L+\delta} \mathcal{H}_L[\rho_L](x) \rho_L(x) dx + \int_0^{L+\delta} \mathcal{H}_L[\rho_L](x) (\rho_{L+\delta}(x) - \rho_L(x)) dx$$

$$= \int_0^{L+\delta} (\mathcal{H}_{L+\delta}[\rho_{L+\delta}](x) - \mathcal{H}_L[\rho_L](x)) \rho_{L+\delta}(x) dx$$

$$+ \varepsilon(L) \int_L^{L+\delta} \rho_L^2(x) dx + \varepsilon(L) \int_0^{L+\delta} \rho_L(x) (\rho_{L+\delta}(x) - \rho_L(x)) dx$$

and on substituting $I_1$ and $I_2$ in (7.2.24) we can cancel some terms and obtain

$$(\varepsilon(L + \delta) - \varepsilon(L)) \int_0^{L+\delta} \rho_{L+\delta}^2(x) dx + \varepsilon(L) \int_0^{L+\delta} (\rho_{L+\delta}(x) - \rho_L(x)) \rho_{L+\delta}(x) dx$$

$$= \int_0^{L+\delta} (\mathcal{H}_{L+\delta}[\rho_{L+\delta}](x) - \mathcal{H}_L[\rho_L](x)) \rho_{L+\delta}(x) dx$$

$$= \int_0^{L+\delta} \int_L^{L+\delta} H(x,y) \rho_L(y) \rho_{L+\delta}(x) dy dx$$

$$+ \int_0^{L+\delta} \int_0^{L+\delta} H(x,y) (\rho_{L+\delta}(y) - \rho_L(y)) \rho_{L+\delta}(x) dy dx$$

$$= \int_0^{L+\delta} \int_{L+\delta}^{L+\delta} H(x,y) \rho_L(y) \rho_{L+\delta}(x) dy dx$$

$$+ \varepsilon(L + \delta) \int_0^{L+\delta} \rho_{L+\delta}(x) (\rho_{L+\delta}(x) - \rho_L(x)) dx$$

where we have used the definition of $\mathcal{H}_L$ and the property $H(x,y) = H(y,x)$. By suitably expanding the term on the left hand side in the above identity, we obtain

$$(\varepsilon(L + \delta) - \varepsilon(L)) \int_0^{L+\delta} \rho_L(x) \rho_{L+\delta}(x) dx = \int_0^{L+\delta} \int_L^{L+\delta} H(x,y) \rho_L(y) \rho_{L+\delta}(x) dy dx$$

$$= \varepsilon(L + \delta) \int_L^{L+\delta} \rho_{L+\delta}(y) \rho_L(y) dy$$

and the positivity property of $\rho_L$ implies that

$$\varepsilon(L + \delta) > \varepsilon(L),$$

which proves (i).

Let us now prove (ii). Assume by contradiction that

$$\lim_{L \to +\infty} \varepsilon(L) = \varepsilon_0 < 1.$$
Let $\varepsilon \in (\varepsilon_0, 1)$. We know from Theorem 7.2.9 that there exists a minimiser $u_\varepsilon$ for the energy $E$ with zero center of mass. We also know that the support of $u_\varepsilon$ is compact from Lemma 7.2.16. From Proposition 7.2.13 we know that $u_\varepsilon$ is symmetric and monotonically decreasing on $x > 0$. Therefore, $u_\varepsilon$ is the unique eigenfunction with unit mass provided by Proposition 7.2.21, and the support of $u_\varepsilon$ is $[-L, L]$ for some $L > 0$. Therefore, the corresponding eigenvalue should be $\varepsilon(L) < \varepsilon_0$, which is a contradiction since $\varepsilon$ and $\varepsilon_0$ are two different eigenvalues with the same eigenfunction.

Let us prove (iii). By letting $L \searrow 0$ one has that the operator $\mathcal{H}_L$ is the zero operator, and therefore $\varepsilon(0)$ should be the eigenvalue of the zero operator, which can only be zero.

**Theorem 7.2.23.** Let $\varepsilon < 1$. Then, there exists a unique $u \in L^2$ solution to

$$u \partial_x (\varepsilon u - G \ast u) = 0,$$

with unit mass and zero center of mass. Moreover,

- $u$ is symmetric and monotonically decreasing on $x > 0$,
- $u \in C^2(\text{supp}(u))$,
- $\text{supp}(u)$ is a bounded interval in $\mathbb{R}$,
- $u$ has a global maximum at $x = 0$ and $u''(0) < 0$,
- $u$ is the global minimiser of the energy $E(u) = \frac{\varepsilon}{2} \int u^2 dx - \frac{1}{2} \int u G \ast u dx$.

**Proof.** We know from Theorem 7.2.9 that there exists a minimiser $u_\infty$ with unit mass and zero center of mass, which is symmetric and monotonically decreasing on $x > 0$ in view of Proposition 7.2.13 and compactly supported on a certain $[-L, L]$ in view of Lemma 7.2.16. From the results in Propositions 7.2.21 and 7.2.22, we know that there exists a unique steady state with such properties, because the correspondence $\varepsilon = \varepsilon(L)$ is one-to-one. So, the only possibility to violate uniqueness of steady states with unit mass and zero center of mass is to have a steady state which violates either the monotonicity property or the symmetry. Suppose first that there exists a steady state with zero center of mass which is not symmetric, it is not restrictive to assume the support of $u$ is $[-L', L']$. Then, we know from Lemma 7.2.17 that it is possible to construct a symmetric steady state $u_t$ with the same energy of $u$ and with the same support of $u$. Now, there are two possibilities: either $u_t$ is a minimiser or not. In the former case $u_t$ is also a minimiser and this is a contradiction (a minimiser is symmetric). In the latter case, the support of $u_t$ is strictly contained in the support of $u_\infty$ in view
of Lemma [7.2.18] and $ut$ is not monotonically decreasing on $x > 0$ because otherwise it would be the unique minimiser provided before. Therefore, with the notation of Proposition [7.2.21], $-ut'$ is an eigenfunction for $\mathcal{H}_{L'}$ in the space $X_{L'}$, which is not belonging to the solid cone $K$. Therefore, the stronger version of Krein-Rutman’s Theorem [7.2.20] and the fact that $\varepsilon(L)$ is increasing imply that $L' > L$, since $u$ is an eigenfunction outside the solid cone $K$, and it therefore should have an eigenvalue strictly less than $\varepsilon(L')$. This implies that $\varepsilon(L) < \varepsilon(L')$ and therefore $L < L'$. Now this is a clearly a contradiction because we said before that the support or $ut$ is strictly contained in the support of $u_\infty$, so $L > L'$. The case in which $u$ is symmetric but not monotone on $x > 0$ can be covered by repeating the same argument above (assume $u = ut$).

**Corollary 7.2.24** (Concavity of $u$ for small $\varepsilon$). There exists a value $\varepsilon_0 \in (0, 1)$ such that, for all $\varepsilon \in (0, \varepsilon_0)$ the corresponding stationary solution provided in Theorem [7.2.23] is concave on the whole interval $[0, L]$.

**Proof.** We can differentiate twice w.r.t $x$ in

$$\varepsilon u(x) = \int_{-L}^{L} G(x - y) u(y) dy + C$$

to obtain

$$\varepsilon u''(x) = \int_{-L}^{L} G''(x - y) u(y) dy$$

for all $x \in [-L, L]$. Therefore, $G''$ is evaluated on the interval $[-2L, 2L]$ in the above integral. We know from Proposition [7.2.22] that $L$ is a monotonically increasing function of $\varepsilon$ with $\lim_{\varepsilon \to 0} L(\varepsilon) = 0$. Since $G''(0) < 0$, and $G \in C^2$, then there exists $L_0 > 0$ such that $G'' < 0$ on $[-2L_0, 2L_0]$. Let $\varepsilon_0$ be the eigenvalue in $K$ corresponding to $L = L_0$. Then, the eigenfunction $u$ is concave on its support.

### 7.3 Numerical Results

In the following, we present numerical results for the aggregation equation (7.1.1). We discretise the equation using an explicit Euler scheme and finite difference methods (see Section 8.1). In one dimension, we partition the domain $\Omega = [a, b]$ with an equidistant grid with $n + 1$ grid points $a = x_0 < x_1 < \ldots < x_n = b$ and step size $h = (b - a)/(n + 1)$. Furthermore, the discretised equation reads as follows:

$$\frac{u_{i+1}^j - u_i^j}{dt} = D_+^x (u_i^j D_+^x (\varepsilon u_i^j - G * u_i^j)).$$
The time step size $dt$ has to be chosen appropriately, to guarantee stability. For the numerical simulation we first consider the following interaction potential:

$$G(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right),$$

with mean $\mu = 0$ and variance $\sigma = 1$, i.e. a Gaussian curve. $G$ fulfills the conditions $(1)-(8)$, $\|G\|_{L^1} = 1$ and models a wide range attraction. We present in Figure 7.1 the solutions for the stationary equation $(7.1.3)$, i.e. the largest eigenvalues and corresponding eigenvectors of the operator $G_L$ defined in $(7.2.23)$ for different $L$. The curve for the largest eigenvalues $\epsilon(L)$ is strictly increasing with respect to $L$ and furthermore $\lim_{L \to \infty} \epsilon(L) = 1$ (compare with Proposition 7.2.22). The corresponding eigenfunctions with unit mass are presented in Figure 7.1 (c) and (d). In particular, in Figure 7.1 (c) we present the eigenfunctions for $L \in (0, 1]$, i.e. for $\epsilon < 0.05$ and in Figure 7.1 (c) the eigenfunctions for $L \in [1, 8]$. The difference in the structure of the solution is obvious. For a certain $\epsilon$, which depends on the concavity of the kernel $G$, the solution is not fully concave on its support any more, but bell shaped (compare with Corollary 7.2.24). In addition, we present in Figure 7.1 (b) the stationary solutions of the evolution equation $(7.1.1)$ for $\epsilon \in (0, 1)$. We consider a compactly supported initial datum $u(0, x) = u_0$ with unit mass $\int_{\Omega} u_0 = 1$. For $\epsilon \geq \int G = 1$ we do not obtain stationary solutions for the evolution equation. In this case we expect that the solutions behave like the Barenblatt-Prattle profiles [14, 148]. For $\epsilon = 0$ we obtain a unique stationary solution (with zero center of mass), which is a Dirac-$\delta$-distribution with unit mass centered at zero.

As a second example we consider the following interaction kernel

$$G(x) = \frac{1}{2} \exp(-|x|).$$

The kernel (see Figure 7.2 (a)) has a Lipschitz singularity at the point zero. We present the eigenfunctions for the corresponding operator $G_L$ in Figure 7.2 (b),(c).
Figure 7.1: Largest eigenvalues and corresponding eigenfunctions of the operator $G_L$.

(a) Largest eigenvalues $\epsilon = \epsilon(L)$ of $G_L$ on $L = (0, 20]$; (b) Stationary solutions for (7.1.1) with $\epsilon(0, 1)$; (c) Corresponding eigenfunctions for $\epsilon(L)$ with $L \in [0, 1]$; (d) Corresponding eigenfunctions for $\epsilon(L)$ with $L \in [1, 8]$.

Figure 7.2: (a) Kernel $G(x) = \frac{1}{2} \exp(-|x|)$; (b) Eigenfunctions for $\epsilon(L)$ with $L \in [0, 1]$ resp. $L \in [1, 8]$. 
Chapter 8

Numerical Methods

In the following, we present numerical algorithms for the optimisation problems introduced in Chapter 3 and 4. In particular, we introduce an algorithm for the variational problem (3.1.1) in one dimension in Section 8.2. Due to expensive computational cost by doubling the spatial dimension we neglect here the two-dimensional implementation of the variational problem. In contrast, we propose in Section 8.3 algorithms for the optimal transport problem, which are inspired by the Benamou and Brenier formulation (3.2.15)–(3.2.20) and can be efficiently implemented in two dimensions. The last section is devoted to the numerical study of the gradient flow equations introduced in Chapter 4. Since we know about the relationship between gradient flow equations and variational problems, we use the numerical results for comparison.

8.1 Numerical Discretisation

The underlying discretisation method is based on finite difference quotients. The time derivative is discretised with an explicit respectively implicit Euler method. The spatial derivate is approximated by forward-, backward-, or central difference quotients which we introduce in the following.

First, we have to partition the domain $\Omega \subset \mathbb{R}^N$ and the time interval $[0,T]$. In particular, we partition the space-time-cylinder $[0,T] \times [a,b]^N$ into $n$ equidistant cells $[x_i,x_{i+1}]^N$, $i = 0,...,n-1$ with $x_0 = a$ and $x_n = b$ with step-size $\Delta x = (b-a)/n$ and $m$ time-intervals $[t_j,t_{j+1}]$ for $j = 0,...,m-1$ with time-steps $\Delta t = T/m$.

Let $u$ be an arbitrary function defined on $[0,T] \times \Omega$, then we denote with $u^i_t = u(t_t,x_i)$ the approximation of $u$ in the node $x_i$ at time $t_t$. The spatial derivative of $u$ in $x_i$ can be approximated with the forward $(D_x^+ u)_i$ or backward-difference-quotient $(D_x^- u)_i$.
defined in the following way:

\[ u'(x_i) \approx (D^+_x u)_i = \frac{u_{i+1} - u_i}{\Delta x} \quad i = 0, \ldots, n - 1, \]  
\[ (D^-_x u)_i = \frac{u_i - u_{i-1}}{\Delta x} \quad i = 1, \ldots, n. \]  

The discretisation of the second spatial derivative of \( u \) reads

\[ u''(x_i) \approx (D_\Delta u)_i = \frac{u_{i+1} - 2u_i + u_{i-1}}{(\Delta x)^2}. \]  

Since we are also interested in two-dimensional numerical simulations we additionally introduce the finite difference scheme for \( \Omega \subset \mathbb{R}^2 \), i.e. the forward- and backward-difference quotients in two dimensions.

\[ (D^+_x u)_{i,j} = \frac{u_{i+1,j} - u_{i,j}}{\Delta x} \quad i = 0, \ldots, n - 1, \]  
\[ (D^+_y u)_{i,j} = \frac{u_{i,j+1} - u_{i,j}}{\Delta y} \quad j = 0, \ldots, n - 1, \]  
\[ (D^-_x u)_{i,j} = \frac{u_{i,j} - u_{i-1,j}}{\Delta x} \quad i = 1, \ldots, n, \]  
\[ (D^-_y u)_{i,j} = \frac{u_{i,j} - u_{i,j-1}}{\Delta y} \quad j = 1, \ldots, n. \]

In the sequel, we consider equal spatial discretisation steps, i.e. \( \Delta x = \Delta y \). The discretisation of the second spatial derivative of \( u \) reads

\[ (D_\Delta u)_{i,j} = \frac{u_{i+1,j} - 4u_{i,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}}{\Delta x \Delta y}. \]  

The divergence operator \( (\nabla \cdot \cdot) \) is approximated with forward- or backward-quotients.

### 8.2 Convex Minimisation Scheme

In the following, we introduce an algorithm for the variational problem \( (3.1.2)-(3.1.4) \) in one dimension, i.e. a transport problem in the plane. For the cost function we consider the quadratic Euclidean distance \( c(x, y) = |x - y|^2 \). Given \( n \) points \( X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^N \), which can be interpreted as the \( n \) sources in the optimal transport setting, and \( m \) "destinations" \( Y = \{y_1, \ldots, y_m\} \subset \mathbb{R}^N \). We denote with \( a \) the amount of goods at source \( x \). Then, \( c_{i,j} = |x_i - y_j|^2 \) measures the cost, which is necessary to transport the goods from source \( x_i \) to destination \( y_j \). \( P = \{p_{i,j} \mid i = 1, \ldots, n, j = 1, \ldots, m\} \) is a doubly-stochastic matrix, i.e. it is positive for every component and the sum of the
rows, respectively columns is equal to one. Moreover, $p$ is the set of transport plans, for which conditions $\sum_{i=1}^{n} p_{i,j} = a_j$, $\sum_{j=1}^{m} p_{i,j} = u_i$ and $p_{i,j} \geq 0$ are satisfied. Then the discretised variational problem (3.1.2)-(3.1.4) reads:

$$\min_{p,u} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} |x_i - y_j|^2 p_{i,j} + \epsilon E(u_i),$$  \hspace{1cm} (8.2.1)

s.t. \hspace{1cm} \sum_{i=1}^{n} p_{i,j} = a_j \hspace{1cm} j = 1,..,m, \hspace{1cm} (8.2.2)

\hspace{1cm} \sum_{j=1}^{m} p_{i,j} = u_i \hspace{1cm} i = 1,..,n, \hspace{1cm} (8.2.3)

\hspace{1cm} 0 \leq p_{i,j}, \hspace{1cm} 0 \leq u_i, \hspace{1cm} i = 1,..,n, \hspace{1cm} j = 1,..,m, \hspace{1cm} (8.2.4)

\hspace{1cm} \sum_{i=1}^{n} \sum_{j=1}^{m} p_{i,j} = 1, \hspace{1cm} \sum_{i=1}^{n} u_i = \sum_{j=1}^{m} a_j = 1, \hspace{1cm} (8.2.5)

with given initial condition $a_j$. The minimum is taken over $u$ and the transport plans $p \in \mathbb{R}^{n \times m}$. There exist various algorithms for such linear problems [106], e.g. a nearly optimal algorithm $O(n^2 \log n)$ can be found in [13]. Two dimensional implementations are mostly neglected, due to expensive computational cost by doubling the spatial dimension. Indeed, if we consider an image of size $128 \times 128$, this yields to a $128^2 \times 128^2$ matrix $p$. Hence, here we only investigate the one-dimensional implementation.

The ansatz which we use to solve (3.1.2)-(3.1.4) is a direct one, i.e. we discretise the equations before optimising them. For this purpose we apply an optimisation solver, here IPOPT, which implements a primal-dual interior point method. With the interface AMPL (see www.ampl.com), a modeling language for mathematical programming, we pass the discretised problem in an appropriate syntax to the solver IPOPT. For an extensive introduction we refer to [177].

### 8.3 The Fluid Dynamic Schemes

In this section we are concerned with the implementation of the Benamou-Brenier optimality conditions introduced in Section 3.2 in particular,

$$\partial_t \mu + \nabla \cdot (\mu v) = 0, \hspace{1cm} (8.3.1)$$

$$\partial_t \lambda + v \cdot \nabla \lambda - \frac{1}{2} |v|^2 = 0, \hspace{1cm} (8.3.2)$$

$$\mu v - \mu \nabla \lambda = 0, \hspace{1cm} (8.3.3)$$
in \((0,1) \times \Omega\) with
\[
\begin{align*}
\epsilon \partial E(u) + \lambda(t = 1) & \geq 0, \\
\mu(t = 1) & = u\mathcal{L}^N, \\
\mu(t = 0) & = \nu.
\end{align*}
\] (8.3.4)

The interest in providing an algorithm for (8.3.1)-(8.3.6) is the possibility to implement it in two dimensions. As mentioned in Section 8.2, the minimisation scheme (8.2.1)-(8.2.5) is not appropriate to solve two-dimensional problems efficiently. Therefore, we introduce in the following two iterative algorithms for the optimality conditions (8.3.1)-(8.3.6), namely a gradient descent scheme and a dual ascent scheme. Both algorithms can be implemented in two dimensions. Additionally we introduce an augmented Lagrangian scheme for the TV regularisation. The gradient descent scheme for the coupled system (8.3.1)-(8.3.6) reads:

\[\text{Algorithm: (GDS) Gradient Descent Scheme for (8.3.1)-(8.3.6)}\]

\[
\begin{array}{l}
\text{for } k = 0 \text{ until } \|u^{k+1} - u^k\|_{L^2} \leq \delta. \\
1. \text{Initialisation (8.3.6): } \mu^{k+1}(t = 0) = \nu. \\
2. \text{Solve the state-equation (8.3.1) (forward in time):} \\
   \partial_t \mu^{k+1} + \nabla \cdot (\mu^{k+1} v^k) = 0. \\
3. \text{Evaluate (8.3.4)-(8.3.5):} \\
   u^{k+1} = \mu^{k+1}(t = 1), \quad \lambda^{k+1}(t = 1) = -\epsilon \partial E(u^{k+1}) + \eta(x) - \eta_0 \]
4. \text{Solve the adjoint-equation (8.3.2) (backward in time):} \\
   \partial_t \lambda^{k+1} + v^k \cdot \nabla \lambda^{k+1} - \frac{1}{2} |v^k|^2 = 0. \\
5. \text{Update } (\tau > 0): \quad (1 + \tau)v^{k+1} = \nabla \lambda^{k+1} + \tau v^k.
\end{array}
\]

The gradient descent scheme (GDS) finds iteratively a local minimum by going into the direction of the negative gradient. We start with the initialisation and then solve the
state equation forward in time. The evaluation in step (3) yields the initial condition for the adjoint equation, which is solved backward in time. In the last step we update the velocity \( v \). We iterate as long as a certain stopping criteria is fullfilled, in particular we stop, when the new iterate does not change any more.

The reason why we consider a gradient descent scheme and additionally a dual ascent scheme is due to the condition (8.3.4). Only for the special case \( u > 0 \) the complementarity condition \( \eta u = 0 \) yields \( \eta = 0 \) and therefore \( \epsilon \partial E(u) + \lambda(t = 1) = 0 \). This is a hard restriction because we can only expect \( u > 0 \) for the logarithmic entropy (compare with the self-similar solution (5.1.16)). The dual ascent scheme circumvents this problem. To derive the dual ascent scheme we recall here the optimality conditions for the dual method (see Section 3.2), in particular

\[
\begin{align*}
\partial_t \mu + \nabla \cdot (\mu \nabla \lambda) &= 0, \\
\partial_t \lambda + \frac{1}{2} |\nabla \lambda|^2 &= 0,
\end{align*}
\]

in \((0,1) \times \Omega\) with

\[
\begin{align*}
\lambda(t = 0) &= \lambda_0, \\
\mu(t = 0) &= \nu,
\end{align*}
\]

and \( \epsilon E'(u) + \lambda(t = 1) \geq 0 \) is the optimality condition for the optimisation problem

\[
\epsilon E(u) + \int_{\Omega} u\lambda(t = 1) \ dx \rightarrow \min_{u \in \mathcal{L}}.
\]

Note that we replace here \( v = \nabla \lambda \) as discussed in Section 3.2. We start with the initialisation of \( \lambda(t = 0) \) and solve then the adjoint equation forward in time. Afterwards we solve the optimisation problem \( \epsilon E(u) + \int_{\Omega} u\lambda^{k+1}(t = 1) dx \rightarrow \min_u \), which has a unique solution if \( E(u) \) is strictly convex. This yields the initial condition for the state equation, which is solved backward in time in step (4). In the last step we update \( \lambda_0 \).
Algorithm: (DAS) Dual Ascent Scheme for \([3.2.21]-[3.2.24]\) (8.3.13)

for \(k = 0\) until \(\|u^{k+1} - u^k\|_{L^2} \leq \delta\).

1. Initialisation: \(\lambda^{k+1}(t = 0) = \lambda_0\).

2. Solve the adjoint-equation (forward in time):

\[
\partial_t \lambda^{k+1} + \frac{1}{2} |\nabla \lambda^{k+1}|^2 = 0.
\]

3. Evaluate

\[
\epsilon \partial E(u^{k+1}) + \lambda^{k+1}(t = 1) \ni 0,
\]

and furthermore \(\mu^{k+1}(t = 1) = u^{k+1}\).

4. Solve the state-equation (backward in time):

\[
\partial_t \mu^{k+1} + \nabla \cdot (\mu^{k+1} \nabla \lambda^{k+1}) = 0.
\]

5. Update \((\tau > 0)\):

\[
\lambda_0^{k+1} = \lambda_0^k + \tau (\mu^{k+1}(t = 0) - \nu).
\]

For the TV regularisation \((6.3.1)\), we introduce an algorithm based on an augmented Lagrangian ansatz \([87]\). Inspired by the split Bregman approach \([88]\) we modify \((8.3.1)-(8.3.6)\) by inserting an auxiliary variable \(z = \nabla u\) and adding the constraint \(z - \nabla u = 0\). The augmented Lagrangian ansatz yields a saddle point problem (see also \([36]\)):

\[
\inf_{\mu,u,z} \sup_{\xi,\lambda,\lambda_0} L(\mu, u, z, \xi, \lambda, \lambda_0) = -\frac{1}{2} \int_0^1 \int_\Omega |\nabla \lambda|^2 d\mu(x)dt + \epsilon \int_\Omega |z| dx - \int_0^1 \int_\Omega \partial_t \lambda d\mu(x)dt
\]

\[
- \int_\Omega \nu \lambda_0 dx + \int_\Omega \lambda_0 d\mu(t = 0) - \int_\Omega \lambda(t = 1) d\mu(t = 1)
\]

\[
+ \int_\Omega u \lambda(t = 1) dx + \int_\Omega (z - \nabla u) \xi dx + \frac{r}{2} \int_\Omega |z - \nabla u|^2 dx,
\]

with \(r \in \mathbb{R}\) and primal variables \(\mu, u, z\) and dual variables \(\xi, \lambda, \lambda_0\). The optimality conditions read

\[
\partial_t \lambda + \frac{1}{2} |\nabla \lambda|^2 = 0, \quad (8.3.14)
\]

\[
\partial_t \mu + \nabla \cdot (\mu \nabla \lambda) = 0, \quad (8.3.15)
\]
and
\[ \hat{z}_* = \text{argmin}_z \left\{ \epsilon \int_\Omega |z| \, dx + \int_\Omega \xi z \, dx + \frac{r}{2} \int_\Omega |z - \nabla u|^2 \, dx \right\}, \quad (8.3.16) \]
\[ \hat{\xi}_* = \text{argmax}_\xi \left\{ \int_\Omega (z - \nabla u) \xi \, dx \right\}. \quad (8.3.17) \]
We reformulate (8.3.16) to
\[ \hat{z}_* = \text{argmin}_z \left\{ \epsilon \int_\Omega |rz| \, dx + \frac{1}{2} \int_\Omega |rz - (r\nabla u - \xi)|^2 \, dx \right\}, \]
where the solution is given through
\[ z = \begin{cases} \left(1 - \frac{\epsilon}{r|w(x,y)|}\right)(w(x,y)), & |w(x,y)| > 1, \\ 0, & |w(x,y)| \leq 1, \end{cases} \quad (8.3.18) \]
with \( w = \nabla u - \xi \).

Using the optimality conditions we can write down the following iterative algorithm:

---

**Algorithm: (ALS) Augmented Lagrangian Scheme** \( (8.3.19) \)

for \( k = 0 \) until \( \|u^{k+1} - u^k\|_{L^2} \leq \delta \).

1. Initialisation: \( \lambda^{k+1}(t = 0) = \lambda_0, \ r > 0 \).
2. Solve the adjoint-equation (forward in time):
\[ \partial_t \lambda^{k+1} + \frac{1}{2} |\nabla \lambda^{k+1}|^2 = 0. \]
3. Evaluate
\[ \Delta u^{k+1} = \left(\frac{1}{r}\lambda^{k+1}(t = 1) + \frac{1}{r} \nabla \xi^k + \nabla \cdot z^k\right), \quad u^{k+1} = \mu^{k+1}(t = 1). \]
4. Solve the state-equation (backward in time):
\[ \partial_t \mu^{k+1} + \nabla \cdot (\mu^{k+1} \nabla \lambda^{k+1}) = 0. \]
5. Update \( (\tau_1, \tau_2 > 0) \):
\[ \lambda^{k+1}_0 = \lambda^k_0 + \tau_1 (\mu^{k+1}(t = 0) - \nu), \]
\[ z^{k+1} = (8.3.18), \]
\[ \xi^{k+1} = \xi^k + \tau_2 (z^{k+1} - \nabla u^{k+1}). \]
Discretisation

For the discretisation of the partial differential equations that appear in the algorithms, we consider the upwind-method and the difference quotients defined in Section 8.1. In one dimension the discretisation of the state-equation
\[ \partial_t \mu = -\nabla \cdot (\mu v) = -\nabla \mu v - \mu \nabla \cdot v \]
is given by:
\[ (D^+_t \mu)_i^l = -(\max(v^l_1, 0)(D^-_x \mu)_i^l + \min(v^l_1, 0)(D^+_x \mu)_i^l) - \mu_i^l(D^+_x v)_i^l. \] (8.3.20)

We consider the explicit Euler method to approximate the time-derivative and an upwind scheme for the spatial derivatives. In a similar way we discretise the adjoint-equation
\[ \partial_t \lambda + v \cdot \nabla \lambda - \frac{1}{2} |v|^2 = 0 \]
which yields
\[ (D^+_t \lambda)_i^l = -\frac{1}{2} |v^l_1|^2 + (\max(v^l_1, 0)(D^+_x \lambda)_i^l + \min(v^l_1, 0)(D^-_x \lambda)_i^l). \] (8.3.21)

Note here, that we choose the discretisation operator which we consider for the Hamilton-Jacobi equation conjugate to the discretisation operator which we consider in the continuity equation. This symmetric discretisation yields a more stabil algorithm.

In two dimensions the discretisation of the state-equation with an upwind method is given by:
\[ (D^+_t \mu)_{i,j}^l = - \left( \max((v_1^l)_i,j, 0)(D^-_x \mu)_{i,j}^l + \min((v_1^l)_i,j, 0)(D^+_x \mu)_{i,j}^l \right) \]
\[ - \left( \max((v_2^l)_i,j, 0)(D^-_y \mu)_{i,j}^l + \min((v_2^l)_i,j, 0)(D^+_y \mu)_{i,j}^l \right) \]
\[ - \mu_{i,j}^l \left( (D^+_x v_1)_{i,j}^l + (D^+_y v_2)_{i,j}^l \right), \]
where \( v_1 \) is the velocity in \( x \)-direction and \( v_2 \) in \( y \)-direction. In a similar way we discretise the adjoint-equation. The discretisation of the other equations is straightforward.

8.4 The Gradient Flow Scheme

We know that a solution of the minimisation problem (3.1.1) can be interpreted as a discrete approximation of a solution of the gradient flow of \( E(u) \) with respect to the quadratic Wasserstein distance. This motivates the idea to compare the numerical results we obtain for the variational problem (3.1.1) with discrete solutions of the
gradient flow equations.

Note here, that we are not trying to compete with existing numerical schemes for these equations, e.g. [54, 28, 70]. We solely aim for a comparison with solutions obtained from (3.1.1).

In Figure (8.1a), we present discrete solutions, this means solutions after a fixed number of time-steps, for the heat equation

\[
\partial_t u = \Delta u, \quad \forall \, x \in \Omega, \, t > 0, \quad (8.4.1)
\]

\[
u(0, x) = u_0(x) > 0, \quad \forall \, x \in \Omega, \quad (8.4.2)
\]

which is the associated gradient flow for the logarithmic entropy (5.1.1) with respect to the Wasserstein distance. For the time derivative we consider the explicit Euler method and a second order finite difference-quotient (8.1.3) for the Laplace-operator. We choose here the space step-size \(\Delta x = 0.1\) and \(\Delta t = 5 \cdot 10^{-4}\) for the time-steps.

The \(L^2\)-regularisation yields the porous medium equation (see Section 5.2)

\[
\partial_t u = \Delta u^2, \quad \forall \, x \in \Omega, \, t > 0, \quad (8.4.3)
\]

\[
u(0, x) = u_0(x) > 0, \quad \forall \, x \in \Omega. \quad (8.4.4)
\]

After discretisation with \(\Delta x = 0.2\) and \(\Delta t = 5 \cdot 10^{-4}\), we obtain the solutions which are presented in Figure (8.1b). The solutions are approximations of the in the literature well known Barenblatt-Prattle solutions [173, 148, 14] .

Furthermore, the thin film equation

\[
\partial_t u = -\text{div}(u \nabla \Delta u), \quad \forall \, x \in \Omega, \, t > 0, \quad (8.4.5)
\]

\[
u(0, x) = u_0(x) > 0, \quad \forall \, x \in \Omega. \quad (8.4.6)
\]

is the associated gradient flow for the Dirichlet regularisation (6.1.1) with respect to the Wasserstein metric. We obtain after reformulation

\[
\partial_t u = -\nabla \cdot (u \nabla \Delta u) = -\nabla u \nabla \Delta u - u \Delta \Delta u, \quad (8.4.7)
\]

the discrete problem

\[
(D_t^+ u)_i^l = -(D_x^+ u)_i^l D_x^-((D \Delta u)_i^l) - (u_i^l) D \Delta ((D \Delta u)_i^l),
\]

For the thin film equation the robust maximum principle is not guaranteed, and therefore the solution of (8.4.5) is not necessary bounded through the maximum or minimum of the initial data. To avoid numerical problems we have to guarantee in every iteration step the positivity of \(u\). The discrete solutions for \(\Delta x = 0.2\) and time-steps \(\Delta t = 5 \cdot 10^{-4}\) are presented in Figure (8.1c).
Considering the gradient flow for the Fisher information (6.2.1)

\[ E(u) = \frac{1}{2} \int_{\Omega} u |\nabla \ln u|^2 dx = \frac{1}{2} \int_{\Omega} \frac{|\nabla u|^2}{u} dx = 2 \int_{\Omega} (\nabla \sqrt{u})^2 dx, \]  

(8.4.8)
yields a non-linear fourth order equation, known as the Derrida-Lebowitz-Speer-Spohn (DLSS) equation

\[ \partial_t u = -\Delta (u \Delta (\ln u)) = -2 \nabla \left( u \nabla \left( \frac{\Delta \sqrt{u}}{\sqrt{u}} \right) \right), \quad \forall \ x \in \Omega, t > 0, \]  

(8.4.9)

\[ u(0, x) = u_0(x) > 0, \quad \forall \ x \in \Omega. \]  

(8.4.10)

This result was proved by Gianazza, Savaré and Toscani [84]. The numerical solution is a hard task, since an explicit time discretisation of a fourth order evolution equation yields a strong restriction on the time-step \( \Delta t \), i.e. \( \Delta t = O(\Delta x^4) \). For a stabil discretisation scheme we have to consider this bound on the time discretisation. Furthermore, due to the lack of a maximum principle we have to guarantee that \( u(t) > 0 \) in every time step. Here we consider a semi-implicit discretisation of (8.4.9) with \( \Delta x = 0.4 \) and time-stepsize \( \Delta t = 5 \cdot 10^{-4} \)

\[ (D^+_t u)_i^t = -D_\Delta (u^{i+1}_i D_\Delta (\ln u)_i^t). \]  

(8.4.11)

Düring, Matthes and Milisic investigated an algorithm in terms of the pseudo inverse in one dimension which is appropriate for such equations [70]. In general, the numerical solution of such partial differential equations with schemes that respect its gradient flow structure, e.g., schemes which guaranteeing monotonic decrease of the corresponding energy functional, have raised growing interest in the last years, cf., e.g., [54, 28, 70] and references therein.
Figure 8.1: Solutions (after a certain time $t$) of the gradient flow equations.
Chapter 9

Conclusion

This thesis is mainly concerned with a variational problem based on the optimal transport theory. In particular, we introduced a minimisation problem which consist of a fidelity term, which is here the Wasserstein metric, and of a regularisation term. The provided ansatz can treat discrete measures as well as continuous probability measures.

In addition, we derived an equivalent formulation in a fluid dynamic framework based on the theory of Benamou and Brenier and the optimality conditions, which we exploited for the numerical solution. Furthermore we pointed out the relationship between certain evolution equations and their gradient flow formulation with respect to the Wasserstein distance and gave examples.

Especially, we gave a detailed analysis for the nonlocal interaction energy functional which corresponds to the aggregation equation. We discussed the structure of the stationary solutions for the aggregation equation with diffusion term and gave existence and uniqueness results of nontrivial stationary solutions.

We provided numerical algorithms for the introduced optimisation problems and computed numerical results. For comparison we computed the solutions after certain timesteps for the associated gradient flow equations. We discussed several different regularisation energies, e.g. the logarithmic entropy, the $L^2$-regularisation, the Dirichlet regularisation, the Fisher information and the TV regularisation and illustrated the impact on the data. We justified that our introduced method can be applied to many different fields.

In the following we point out some open problems:

We mentioned in this thesis that the gradient flow equation for the TV-regularisation is a highly nonlinear-fourth order equation. However, the fast numerical solution of this
partial differential equation is still an open problem. Moreover, we neglected in this thesis the comparison of our variational problem with existing methods. Especially, a comparison with other density estimation methods or image decomposition methods would be interesting.

Moreover, in this thesis we discussed several regularisation energies. It would also interesting to discuss other energy functionals, e.g. the local repulsive double-well potential \( G(x) = x^4 - x^2 \) or \( E(u) = \epsilon_1 u^2 + \epsilon_2 \nabla u^2 \). G. Carlier and J. Salomon introduced an optimisation problem with a potential energy functional with applications in crowd motion based on the Benamou-Brenier formulation \([47]\). Similar as in our model, the minimum is also taken over the density at the end time, but in contrast the optimisation problem is governed by the Fokker-Planck equation. It would also be interesting to compare this two approaches.
Notation and Symbols

Function Spaces and Norms

$\Omega \subset \mathbb{R}^N$ be open and bounded. $N$ denotes the dimension.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signification</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C(\Omega)$</td>
<td>Space of continuous functions on $\Omega$.</td>
</tr>
<tr>
<td>$C_b(\Omega)$</td>
<td>Space of bounded continuous functions on $\Omega$.</td>
</tr>
<tr>
<td>$C^k(\Omega)$</td>
<td>Space of functions on $\Omega$ which are differentiable up to order $k \in \mathbb{N}$.</td>
</tr>
<tr>
<td>$L^p(\Omega)$</td>
<td>With $1 \leq p &lt; \infty$: Space of $p$-integrable real-valued functions $f$, such that $\int_{\Omega}</td>
</tr>
<tr>
<td>$L^\infty(\Omega)$</td>
<td>Space of Lebesgue measurable functions $f$, such that there exists a $C \in \mathbb{R}$ with $</td>
</tr>
<tr>
<td>$W^{k,p}(\Omega)$</td>
<td>With $1 \leq p, k \leq \infty$: Sobolev space of real-valued functions with $p$-integrable derivative up to order $k$. The corresponding norm is $|f|<em>{W^{k,p}(\Omega)} = \left( \sum</em>{i=1}^{k} \int_{\Omega}</td>
</tr>
<tr>
<td>$H^k(\Omega)$</td>
<td>Sobolev space $W^{k,2}(\Omega)$. This is a Hilbert space.</td>
</tr>
</tbody>
</table>
Let $X$ be a Banach space with norm $\|\cdot\|_X$ and $v : (0, T) \to X$:

$C^k(0, T; X)$ With $k \geq 0$, $0 < T < \infty$: Space of $k$-times continuously differentiable functions from $[0, T]$ with values in $X$.

$L^p(0, T; X)$ With $1 \leq p < \infty$: Space of functions $v \to v(t)$ measurable on $(0, T)$ for the measure $dt$. It is a Banach space with corresponding norm

$$\|v\|_{L^p(0,T;X)} = \left( \int_0^T \|v(t)\|_X^p \, dt \right)^{1/p} < \infty.$$  

Let $X$ be a metric space, then we denote:

$\mathcal{P}(X)$ The set of all probability measures on $X$.

$\mathcal{M}(X)$ The set of all finite signed measures on $X$, equipped with the norm of total variation,

$$\|\mu\|_{TV} = \inf \{\mu_+[X] + \mu_-[X]\}.$$  

$\mathcal{M}^+_\text{loc}(X)$ The space of all real Radon measures defined in $X$.

$\mathcal{P}_p(X)$ The set of all probability measures with finite $p$-th moment (2.2.2).

$\mathcal{P}_{ac}(X)$ The set of all absolute continuous probability measures.

$L^p(d\mu)$ The Lebesgue space of order $p$ for the reference measure $\mu$.

$T_{\#}\mu$ Push forward of the measure $\mu$ through the map $T$ (2.1.1).

$\text{Supp}(\mu)$ Support of measure $\mu$ on $X$, where $\mu[X\setminus F] = 0$ with $F$ smallest closed set $F \subset X$.

$\mathcal{B}(X)$ Borel sets in a separable metric space $X$.

$M_2(\mu)$ The second moment of $\mu$ (9.0.4).

$W_p(\mu, \nu)$ $p$-th-Wasserstein distance between $\mu$ and $\nu$ (2.2.1).
## Calculus Symbols

For a function $f : \Omega \subset \mathbb{R}^N \to \mathbb{R}$ and a sequence of functions $(f^n)_{n \in \mathbb{N}}$ belonging to a Banach space $X$ we denote:

<table>
<thead>
<tr>
<th>Name</th>
<th>Signification</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{df}{dt}$</td>
<td>Total derivative of a function $f$ with respect to $t$.</td>
</tr>
<tr>
<td>$\frac{\partial f}{\partial t}, \partial_t f$</td>
<td>Partial derivative of a function $f$ with respect to $t$.</td>
</tr>
<tr>
<td>$\nabla f$</td>
<td>Gradient of $f$.</td>
</tr>
<tr>
<td>$\nabla \cdot f$</td>
<td>Divergence of a vector-valued function $f$, i.e., $\nabla \cdot f = \sum_{i=1}^N \frac{\partial f}{\partial x_i}$.</td>
</tr>
<tr>
<td>$\Delta f$</td>
<td>Laplacian operator of $f$, i.e., $\Delta f = \sum_{i=1}^N \frac{\partial^2 f}{\partial x_i^2}$.</td>
</tr>
<tr>
<td>$Df$</td>
<td>Distributional derivative of $f$.</td>
</tr>
<tr>
<td>$\int_{\Omega} f , dx$</td>
<td>Volume integral of a function $f$ over $\Omega$.</td>
</tr>
<tr>
<td>$\int_{\partial\Omega} f , ds$</td>
<td>Surface integral of a function $f$ over $\partial\Omega$.</td>
</tr>
<tr>
<td>$\int_{\partial\Omega} f , ds$</td>
<td>Surface integral of a function $f$ over $\partial\Omega$.</td>
</tr>
<tr>
<td>$f^n \to f$ in $X$</td>
<td>The sequence $(f^n)$ converges strongly to $f$ in $X$.</td>
</tr>
<tr>
<td>$f^n \rightharpoonup f$ in $X$</td>
<td>The sequence $(f^n)$ converges weakly to $f$ in $X$.</td>
</tr>
<tr>
<td>$f^n \rightharpoonup^* f$ in $X$</td>
<td>The sequence $(f^n)$ converges to $f$ in the weak$^*$ topology of $X$.</td>
</tr>
<tr>
<td>$X^*$</td>
<td>Dual of a Banach space $X$.</td>
</tr>
<tr>
<td>$\langle \cdot, \cdot \rangle_X$</td>
<td>Dual product on $X \times X^*$ respectively inner product for a Hilbert space $X$.</td>
</tr>
<tr>
<td>$| \cdot |_X$</td>
<td>Norm in a vector space $X$.</td>
</tr>
<tr>
<td>$</td>
<td>A</td>
</tr>
<tr>
<td>$|A|$</td>
<td>Operator norm of $A$ with $|A| = \sup {</td>
</tr>
</tbody>
</table>
**Miscellaneous Notation**

Let $A$ be bounded and open set in $\mathbb{R}^N$.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signification</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{H}^{N-1}$</td>
<td>The Hausdorff $(N - 1)$-dimensional measure.</td>
</tr>
<tr>
<td>Small sets</td>
<td>Sets of Hausdorff dimension $(N - 1)$.</td>
</tr>
<tr>
<td>$\delta_x$</td>
<td>Dirac mass at a point $x$, i.e., $\delta_x(A) = \begin{cases} 1 &amp; \text{if } x \in A, \ 0 &amp; \text{otherwise.} \end{cases}$</td>
</tr>
<tr>
<td>Id</td>
<td>Identity map</td>
</tr>
</tbody>
</table>
Measure Theory

Radon property
Let $X \subset \mathbb{R}^N$ be a Polish space, a complete separable metric space, and $\mu$ a probability measure on $X$. $X$ satisfies the **Radon property** if
\[ \forall B \in B(X), \ \epsilon > 0 \text{ there exists an } K_\epsilon \subset B : \ \mu(B \setminus K_\epsilon) \leq \epsilon. \] (9.0.1)

A positive Borel measure $\mu$ on $X$ is Radon, if it is finite on the compact subset $X$, if $\mu(\Omega) < +\infty$ in $\mathcal{M}_{\text{loc}}^+$. If $\mu(X) = 1$ then $\mu$ is a probability measure in $\mathcal{P}(X)$.

$X$ Polish space is a Radon space if every Borel probability measure $\mu \in \mathcal{P}(X)$ satisfy the Radon property.

Regular Borel measures
A Borel measures $\mu$ on a Polish space is regular, that means for any Borel set $A$
\[ \mu[A] = \sup \{ \mu[K] ; K \text{ compact}, \ K \subset A \} \]
\[ = \inf \{ \mu[O] ; O \text{ open}, \ A \subset O \} \] (9.0.2)

The set of finite second moment
The set of all probability measures with finite second moment is defined by
\[ \mathcal{P}_2(X) := \{ \mu \in \mathcal{P}(X) \mid M_2(\mu) = \int_X |x|^2 d\mu(x) < +\infty \}. \] (9.0.4)

Narrow Convergence
We say that a sequence $(\mu_n) \subset \mathcal{P}(X)$ is narrowly convergent to $\mu \in \mathcal{P}(X)$ as $n \to \infty$ if
\[ \lim_{n \to \infty} \int_X f(x) \ d\mu_n(x) = \int_X f(x) \ d\mu(x) \]
for every function $f \in C_0^b(X)$, the space of continuous and bounded real functions defined on $X$. In this case $\mu \in \mathcal{P}(X)$.

Prokhorov’s theorem:
Let $K \subset \mathcal{P}(X)$ be tight, then it is relatively compact in $\mathcal{P}(X)$.
That means for any sequence $(\mu_k)$ in $\mathcal{P}$ one can extract a subsequence, still denoted by $(\mu_k)$, and a probability measure $\mu^*$ on $X$, such that fo any $\phi \in C_b(X)$
\[ \lim_{k \to \infty} \int_X \phi d\mu_k = \int_X \phi d\mu^* . \]
An easy application of Prokhorov’s theorem reads:

\[ \mu_n \in \mathcal{P}(\mathbb{R}^N), \sup_n M_2(\mu_n) < +\infty \Rightarrow \exists \mu_{nk}, \mu : \mu_{nk} \rightharpoonup \mu \text{ in } \mathcal{P}(\mathbb{R}^N). \] (9.0.5)

A family \((\mu_k)\) of nonnegative measures on \(X\) is said to be tight, if for all \(\epsilon > 0\) there exists a compact set \(K_\epsilon\) with

\[ \sup_{k \geq 0} \mu_k[X \setminus K_\epsilon] \leq \epsilon. \]

Is \(X\) a Polish space, then every relatively compact subset of \(\mathcal{P}(X)\) is tight. In a Polish space \(X\) each measure \(\mu \in \mathcal{P}(X)\) is tight. In a Radon space, every narrowly converging sequence \((\mu_n) \subset \mathcal{P}(X)\) is tight.

**Lower Semicontinuous**
A function \(F : X \rightarrow \mathbb{R}\) is lower semicontinuous, if it satisfies: For any \(x \in X\)

\[ F(x) \leq \liminf_{y \to x} F(y). \]
Bibliography


Lebenslauf

Marzena Magdalene Franek

Persönliche Daten

Geburtsdatum 18.08.1983
Geburtsort Loslau
Familienstand ledig
Staatsangehörigkeit deutsch
Eltern Mutter: Mariola Franek, geb. Schulz
Vater: Marian Franek

Studium

Seit 01/2008 Beginn der Promotion am Institut für Numerische und Angewandte Mathematik an der Westfälischen-Wilhelms-Universität (WWU) in Münster unter der Betreuung von Prof. Dr. Martin Burger
29.10.2007 Diplom in Mathematik, WWU Münster
10/2002-10/2007 Studium der Diplom Mathematik mit Nebenfach Informatik an der WWU Münster

Schulbildung

08/1995 – 06/2002 Gymnasium der Stadt Meschede
Abschluss: Abitur am 19.06.2002

Berufserfahrung

2008-2011 Wissenschaftliche Mitarbeiterin am Institut für Numerische und Angewandte Mathematik an der WWU Münster
2004-2007 Studentische Hilfskraft am Institut für Numerische und Angewandte Mathematik an der WWU Münster