

Density Estimation and Smoothing based on Regularised Optimal Transport

Martin Burger* Marzena Franek† Carola-Bibiane Schönlieb‡

March 10, 2011

Abstract

The aim of this paper is to investigate a novel nonparametric approach for estimating and smoothing densities based on a variational regularisation method with the Wasserstein metric as a data fidelity. The approach allows a unified treatment of discrete and continuous probability measures and is hence attractive for various tasks. In particular the variational model for special regularisation functionals yields a natural method for estimating densities and for preserving edges in the case of total variation regularisation.

In order to compute solutions of the variational problems, a regularised optimal transport problem needs to be solved, for which we discuss several possible formulations and provide a detailed analysis. Moreover we compute special self-similar solutions for several standard regularisation functionals in order to yield further insight into the structure of the favoured reconstructions, and we discuss several computational approaches and results.

AMS Subject Classification: 68U10, 65J20, 65K10, 62G07, 35K55

Keywords: Optimal Transport, Density Estimation, Image processing, Total variation methods.

1 Introduction

In various applied problems densities are involved, which have to be estimated from given measurements, see e.g. [55] for a general introduction into density estimation. 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. Recently there is growing interest in multivariate density estimation for data sets having a spatial dependence. 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, e.g., [44, 23]. Recent works in [42] aim to estimate crime probabilities for different districts or localities within a city and in [53] density estimation is used for wildfire predictions. In medical imaging, e.g.,

*Institute for Computational and Applied Mathematics, Westfälische Wilhelms Universität (WWU) Münster, Einsteinstrasse 62, D 48149 Münster, Germany. Email: martin.burger@wwu.de

†Institute for Computational and Applied Mathematics, Westfälische Wilhelms Universität (WWU) Münster, Einsteinstrasse 62, D 48149 Münster, Germany. Email: marzena.franek@math.uni-muenster.de

‡DAMTP, Centre for Mathematical Sciences, Wilberforce Road, Cambridge, CB3 0WA, United Kingdom. Email: C.B.Schoenlieb@damtp.cam.ac.uk

magnetic resonance tomography (MRI), the images of interest are densities restored from undersampled measurements, cf. , e.g., [24]. 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 right density has to involve some additional smoothing, i.e., regularisation that assumes an a-priori information on the density.

State of the art techniques A common method for density estimation are kernel density estimation techniques [54, 55]. Here the probability density is created by interpolating pointwise kernel densities, i.e., Gaussians in every sampling point.

Another popular density estimation method is the so-called Maximum Penalized Likelihood Estimation (MPLE) [22, 31]. In cases where special structure of densities, e.g. discontinuities, are expected or need to be preserved, standard linear density estimation techniques are not appropriate. In this respect variational approaches gained significant attention recently, which consist in solving

$$u^* \in \operatorname{argmin}_u [D(u, \nu) + \epsilon E(u)], \quad (1.1)$$

where D is a distance measure between the density u and the discrete measure ν (a sum of point densities), $\epsilon > 0$ is a regularisation parameter, and E an appropriate regularisation functional. Frequently used examples are squared L^2 -distances (using an expansion without the square term of ν in order to allow point measures, in particular finite sampling). Then, for a function u defined on a domain $\Omega \subset \mathbb{R}^d, d = 1, 2$, we have

$$D(u, \nu) = \frac{1}{2} \int_{\Omega} u^2 dx - \int_{\Omega} u d\nu, \quad (1.2)$$

or the log-likelihood

$$D(u, \nu) = - \int_{\Omega} \log u d\nu. \quad (1.3)$$

Steeming from (1.1) various methods for density estimation have been proposed, which differ in the choice of the regulariser E , see, e.g., [22, 31, 54, 36, 41, 52]. In particular, in [36, 41, 52] the proposed regulariser $E(u)$ is the total variation (TV) [3, 26, 49] 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 ν has a regular density. For this reason some presmoothing or preestimation steps are usually needed. In [43] 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 y_i , defined to be

$$\frac{1}{\text{distance between sampling points}} \approx \text{density},$$

and a piecewise constant initial guess d for the density, they recover the original density u^* by solving

$$u^* = \operatorname{argmin}_u \left\{ \frac{1}{2} \|u - d\|_{L^2}^2 + \epsilon \int |\nabla u| dx \right\}.$$

In [56] the authors further extend the approach in [41] by introducing additional spatial information into the problem. Other density estimation methods are, e.g., the taut string

method restricted to the univariate case [18, 19] or the logspline technique [37]. 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 [41] this additional constraint is enforced by applying Bregman iteration [45] to the iterative solution of the problem.

Density estimation using the Wasserstein distance In this paper we shall investigate the solution of the variational problem

$$\frac{1}{2}W_2(\nu, u\mathcal{L}^d)^2 + \epsilon E(u) \rightarrow \min_u \quad (1.4)$$

where ν is a given probability measure on Ω and u is a probability density. Further, $W_2(\nu, u\mathcal{L}^d)$ is the 2- Wasserstein distance between ν and $u\mathcal{L}^d$ [4, 60, 61, 2] and defines a distance within the space of probability measures. The typical regularisation energy E we consider is the total variation defined by

$$E(u) := |Du|(\Omega) = \sup_{\mathbf{g} \in C_0^\infty(\Omega; \mathbb{R}^d), \|\mathbf{g}\|_\infty \leq 1} \int_{\Omega} u \, \nabla \cdot \mathbf{g} \, dx. \quad (\text{Total variation regularisation}) \quad (1.5)$$

Alternative regularisers are the gradient squared

$$E(u) = \frac{1}{2} \int_{\Omega} |\nabla u|^2 \, dx \quad (\text{Dirichlet regularisation}), \quad (1.6)$$

the squared L^2 -norm

$$E(u) = \frac{1}{2} \int_{\Omega} u^2 \, dx \quad (\text{Squared Lebesgue regularisation}) \quad (1.7)$$

or statistically motivated functionals like the log entropy

$$E(u) = \int_{\Omega} u \ln u \, dx \quad (\text{Log-entropy}) \quad (1.8)$$

whose use in (1.4) results in an isotropic smoothing of u , similar to a Gaussian filter. Moreover the Fisher information

$$E(u) = \int_{\Omega} u |\nabla \ln u|^2 \, dx = \int_{\Omega} \frac{|\nabla u|^2}{u} \, dx \quad (\text{Fisher information}) \quad (1.9)$$

might appear in statistical applications. Note that (formally) the total variation is the exponent one version of the Fisher information (1.9)

$$E(u) = \int_{\Omega} |\nabla u| \, dx = \int_{\Omega} u |\nabla \ln u| \, dx.$$

A particular advantage of the Wasserstein metric is the possibility to obtain a unified distance functional for densities continuous with respect to the Lebesgue measure as well as concentrated densities, since the Wasserstein metric is defined for arbitrary probability densities ν with finite second moment. Thus, the functional can be used for density estimation as well as for smoothing and decomposition of continuous densities. 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. The standard image processing methods of decomposing images, e.g. into cartoon and texture, are based on additive approaches, which is not fitting well to probability measures, since one may argue that the cartoon of a probability measure should be a probability measure as well. 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 preserving.

Besides these nice properties, the variational model can be well motivated from a Bayesian maximum a-posteriori probability (MAP) estimation model. Assume one takes N measurements F_i , where each measurement follows a Gaussian error model with mean x_i and standard deviation σ . Then the MAP estimate is computed from minimising the negative logarithm of the posterior probability density

$$\frac{1}{2\sigma^2} \sum_i (x_i - F_i)^2 + \epsilon E(x_1, \dots, x_N). \quad (1.10)$$

The squared distance between the x_i and F_i can be translated into a multiple of the Wasserstein metric between the empirical measures

$$\nu = \frac{1}{N} \sum_{i=1}^N \delta_{F_i}, \quad u = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}, \quad (1.11)$$

thus we recover a discrete version of our variational model.

Gradient flows From the point of view of gradient flows a solution of the minimisation problem (1.4) can be interpreted as a discrete approximation of a solution of the gradient flow of $E(u)$ with respect to the L^2 Wasserstein metric. More precisely (1.4) represents one timestep of De Giorgi's minimising movement scheme to the functional $E(u)$ with timestep ϵ . In particular, let u^0 be an initial probability density and solve for $k \geq 0$

$$u^{k+1} \in \operatorname{argmin}_u \left(\frac{1}{2} W_2(u^k \mathcal{L}^d, u \mathcal{L}^d)^2 + (t_{k+1} - t_k) E(u) \right). \quad (1.12)$$

This iterative approach is also called JKO scheme, after Jordan, Kinderlehrer and Otto [34] who proposed (1.12) to approximately solve diffusion equations of the type

$$\begin{aligned} \partial_t u &= \nabla \cdot (u \nabla E'(u)), \\ u(0, x) &= u_0(x) \geq 0, \end{aligned} \quad (1.13)$$

with normalized mass $\int_{\Omega} u_0 dx = 1$. The numerical solution of equations (1.13) with schemes that respect its gradient flow structure, e.g., schemes which guarantee monotonic decrease of the corresponding energy functional, have raised growing interest in the last years, cf., e.g., [15, 7, 21, 9, 32] and references therein. In Table 1 we give a list of corresponding diffusion equations (1.13) to our set of regularising functionals (1.5)–(1.9). Note that with our approach in (1.4) we are not directly computing solutions of the gradient flow. In fact, a minimiser of (1.4) represents only one timestep of size ϵ of (1.12). However, solving (1.4) iteratively, i.e., computing a minimizer u_ϵ and subsequently a minimizer $u_{2\epsilon}$ for $\nu = u_\epsilon \mathcal{L}^d$ and so forth, is a possible and robust numerical approximation to the corresponding gradient flow (1.13) with time step size ϵ . Note however that we are interested not only in the limiting case of $\epsilon \rightarrow 0$, which adds additional requirements for our numerical schemes.

Energy	PDE	
(1.8)	$\partial_t u = \Delta u$	Heat equation
(1.7)	$\partial_t u = \Delta u^2$	Porous medium equation
(1.6)	$\partial_t u = -\operatorname{div}(u \nabla \Delta u)$	Thin film equation
(1.9)	$\partial_t u = -\frac{1}{2} \sum_{i,j=1}^2 \partial_{i,j}^2 \left(u \cdot \partial_{i,j}^2 \ln u \right)$	Derrida-Lebowitz-Speer-Spohn (DLSS) -equation
(1.5)	$\partial_t u = -\nabla \cdot \left(u \nabla \nabla \cdot \left(\frac{\nabla u}{ \nabla u } \right) \right)$	Highly nonlinear fourth-order diffusion equation, not studied

Table 1: Different regularising energies and their corresponding gradient flows w.r.t. to the L^2 Wasserstein distance.

Organization of the paper In section 2 we start with the definition of different formulations of (1.4), i.e., Kantorovich’s formulation, its dual and the Benamou-Brenier fluid-dynamics scheme. Then we prove existence and uniqueness results for solutions of our model, followed by stability estimates and some comments about the relation to diffusion filters and random sampling. Moreover, in section 3 we compute explicitly self-similar solutions of (1.4) for energies E as in (1.5)–(1.9). Eventually, in section 4 we present our numerical approaches and corresponding numerical results.

2 Analysis of the Variational Problems

In this section we shall investigate different equivalent formulations for (1.4) and its analytical properties such as existence, uniqueness, and stability of solutions. We shall come across these alternative formulations for (1.4) again when discussing its numerical solution. For theoretical results for (1.4) with non-gradient-dependent energies E we also refer to several recent works, e.g., the ones of Santambrogio et al. [12, 50, 51]. There, the authors derive optimality conditions for (1.4) together with regularity properties of minimisers.

2.1 Preliminaries

2.1.1 The Wasserstein distance, narrow convergence and geodesic convexity

Let (X, d) be a separable metric space and $\mu \in \mathbb{P}(X)$ a probability measure on X . Let further X satisfy the Radon property, i.e.,

$$\forall B \in \mathbb{B}(X), \epsilon > 0 \text{ there exists an } K_\epsilon \subset\subset B : \quad \mu(B \setminus K_\epsilon) \leq \epsilon,$$

and let $p \geq 1$. The $(p - th)$ Wasserstein distance between two probability measures $\mu^1, \mu^2 \in \mathbb{P}_p(X)$ (the space of all probability measures on X with μ -integrable $p - th$ moments) is defined by

$$W_p(\mu^1, \mu^2)^p := \min_{\Pi \in \Gamma(\mu^1, \mu^2)} \int_{X^2} d(x, y)^p d\Pi(x, y).$$

Here $\Gamma(\mu^1, \mu^2)$ denotes the class of all transport maps $\gamma \in \mathbb{P}(X^2)$ such that

$$\pi_{\#}^1 \gamma = \mu^1, \quad \pi_{\#}^2 \gamma = \mu^2,$$

where $\pi^i : X^2 \rightarrow X$, $i = 1, 2$, are the canonical projections on X , and $\pi_{\#}^i \gamma \in \mathbb{P}(X)$ is the push-forward of γ through π^i .

In our setting $(X, d) = (\Omega, |\cdot|)$, where Ω is an open and bounded domain in \mathbb{R}^d and $|\cdot|$ is the usual Euclidean distance on \mathbb{R}^d . For this case the above assumptions are trivially fulfilled. Further $p = 2$, i.e., we only consider the quadratic Wasserstein distance $W_2(\cdot, \cdot)$ on the space $\mathbb{P}_2(X)$.

For the lower-semicontinuity of the Wasserstein distance we need a new type of convergence in the space of probability measures.

Definition 1. We say that a sequence $(\mu_n) \subset \mathbb{P}(X)$ is narrowly convergent to $\mu \in \mathbb{P}(X)$ as $n \rightarrow \infty$ if

$$\lim_{n \rightarrow \infty} \int_X f(x) d\mu_n(x) = \int_X f(x) d\mu(x)$$

for every function $f \in C_b^0(X)$, the space of continuous and bounded real functions defined on X .

A useful result is the following:

Proposition 2.1. (Proposition 7.1.3. in [4]) Let (X, d) be a metric space. Let $(\mu_n^1), (\mu_n^2) \subset \mathbb{P}(X)$ be two sequences narrowly converging to μ_1, μ_2 respectively, and let $\Pi_n \in \Gamma(\mu_n^1, \mu_n^2)$ be a sequence of corresponding optimal transport plans with $\int_X d(x_1, x_2)^p \Pi_n$ bounded. Then (Π_n) is narrowly relatively compact in $\mathbb{P}(X^2)$ and any narrow limit point $\Pi \in \Gamma(\mu^1, \mu^2)$ is an optimal transport plan for (μ^1, μ^2) , with

$$W_2(\mu^1, \mu^2) \leq \liminf_{n \rightarrow \infty} \int_{X^2} d(x, y)^p d\Pi^n(x, y) = \liminf_{n \rightarrow \infty} W_2(\mu_n^1, \mu_n^2).$$

Further we recall some general concepts of geodesical convexity in $(\mathbb{P}(X), W_2(\cdot, \cdot))$.

Definition 2. Let X be a separable Hilbert space and let $\varphi : \mathbb{P}(X) \rightarrow (-\infty, +\infty]$. Given $\lambda \in \mathbb{R}$, we say that φ is λ -geodesically convex in $\mathbb{P}(X)$ if for every couple $\mu^1, \mu^2 \in \mathbb{P}(X)$ there exists an optimal transfer plan $\Pi \in \Gamma(\mu^1, \mu^2)$ such that

$$\varphi(\mu_t^{1 \rightarrow 2}) \leq (1-t)\varphi(\mu^1) + t\varphi(\mu^2) - \frac{\lambda}{2}t(1-t)W_2^2(\mu^1, \mu^2) \quad \forall t \in [0, 1],$$

where $\mu_t^{1 \rightarrow 2} = (\pi_t)$

Now, if the space X has dimension greater or equal two we know that there is no constant λ such that the Wasserstein distance is λ -convex along geodesics (cf. [4] Example 9.1.5). Therefore a more general definition of geodesic convexity is introduced, i.e., convexity along generalized geodesics.

Definition 3. A generalized geodesic joining μ^2 to μ^3 (with base μ^1) is a curve of the type

$$\mu_t^{2 \rightarrow 3} = (\pi_t^{2 \rightarrow 3})_{\#} \mu, \quad t \in [0, 1],$$

where $\mu \in \Gamma(\mu^1, \mu^2, \mu^3)$ is a transport plan and $\pi_{\#}^{12} \mu \in \Gamma_0(\mu^1, \mu^2)$, $\pi_{\#}^{13} \mu \in \Gamma_0(\mu^1, \mu^3)$ are optimal transport plans.

Definition 4. Given $\lambda \in \mathbb{R}$, we say that φ is λ -convex along generalized geodesics if for any $\mu^1, \mu^2, \mu^3 \in D(\varphi)$ there exists a generalized geodesic $\mu_t^{2 \rightarrow 3}$ induced by a plan $\mu \in \Gamma(\mu^1, \mu^2, \mu^3)$ satisfying the properties from Definition 3 such that

$$\varphi(\mu_t^{2 \rightarrow 3}) \leq (1-t)\varphi(\mu^2) + t\varphi(\mu^3) - \frac{\lambda}{2}t(1-t)W_\mu^2(\mu^2, \mu^3), \quad \forall t \in [0, 1],$$

where

$$W_\mu^2(\mu^2, \mu^3) := \int_{X^3} |x_3 - x_2|^2 d\mu(x_1, x_2, x_3).$$

2.1.2 The total variation

We recall that for $u \in L_{loc}^1(\Omega)$

$$V(u, \Omega) := \sup \left\{ \int_{\Omega} u \operatorname{div} \varphi \, dx : \varphi \in [C_c^1(\Omega)]^d, \|\varphi\|_{\infty} \leq 1 \right\}$$

is the total variation of u and that $u \in BV(\Omega)$ (the space of bounded variation functions, [3, 26]) if and only if $V(u, \Omega) < \infty$, see [3, 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.

2.2 Formulations of the Minimisation Problem

The classical representation of the Wasserstein distance is the **Kantorovich formulation**, cf. [60, Introduction, Section 1]. Implemented in our variational problem (1.4) it reads

$$\frac{1}{2} \int_{\Omega \times \Omega} |x - y|^2 d\Pi(x, y) + \epsilon E(u) \rightarrow \min_{\Pi, u}, \quad (2.1)$$

subject to

$$\int_{A \times \Omega} d\Pi(x, y) = \int_A d\nu(y) \quad (2.2)$$

$$\int_{\Omega \times A} d\Pi(x, y) = \int_A u(x) dx \quad (2.3)$$

for all $A \subset \Omega$ measurable, where u is a probability density in the domain of E and Π is a probability measure on $\Omega \times \Omega$. Kantorovich introduced also a dual formulation of (2.1)-(2.3), see [60, Chapter 1, Theorem 1], which gives that the above Kantorovich formulation is equivalent to

$$\inf_u \sup_{\varphi, \psi} \Lambda(u, \varphi, \psi) \quad \text{subject to } \varphi(x) + \psi(y) \leq \frac{1}{2}|x - y|^2, \quad (2.4)$$

with

$$\Lambda(u, \varphi, \psi) = \epsilon E(u) + \int_{\Omega} \varphi(x) u(x) \, dx + \int_{\Omega} \psi(y) \, d\nu(y). \quad (2.5)$$

Benamou and Brenier [5] reintroduced the idea of Monge himself, to propose an interpretation of the classical Kantorovich problem within a fluid mechanics framework. Here, the optimal solution is given by a pressureless potential flow, i.e., we have

$$\inf_{(u, \mu, v)} \left\{ \frac{1}{2} \int_0^1 \int_{\Omega} |v|^2 d\mu(x) \, dt + \epsilon E(u) \right\}, \quad (2.6)$$

subject to

$$\partial_t \mu + \nabla \cdot (\mu v) = 0 \quad (2.7)$$

$$\mu(t=0) = \nu \quad (2.8)$$

$$\mu(t=1) = u \mathcal{L}^d \quad (2.9)$$

In [5] the authors used this formulation to numerically compute the Wasserstein distance for two given probability measures. Writing down the Lagrangian for the minimisation problem (2.6) with equality constraint (2.7) and boundary conditions (2.8)–(2.9) we get

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

with complementarity condition $\eta u = 0$, where η denotes the Lagrange multiplier for the condition $u \geq 0$ and $\eta_0 \in \mathbb{R}$. The corresponding optimality conditions follow

$$\partial_t \mu + \nabla \cdot (\mu v) = 0, \quad (2.10)$$

$$\partial_t \lambda + v \cdot \nabla \lambda - \frac{1}{2} |v|^2 = 0, \quad (2.11)$$

$$\mu v - \mu \nabla \lambda = 0, \quad (2.12)$$

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

$$\epsilon \partial E(u) + \lambda(t=1) \ni 0, \quad (2.13)$$

$$\mu(t=1) = u \mathcal{L}^d, \quad (2.14)$$

$$\mu(t=0) = \nu. \quad (2.15)$$

For the numerical solution of (1.4) we shall in particular exploit the Benamou-Brenier optimality conditions (2.10)–(2.15). We will also consider a dual method based on augmenting the constraint at $t=0$, which leads to the Lagrangian

$$\begin{aligned} L(u, \mu, \lambda, \lambda_0, \eta) &= -\frac{1}{2} \int_0^1 \int_{\Omega} |\nabla \lambda|^2 d\mu(x) dt + \epsilon E(u) - \int_0^1 \int_{\Omega} \partial_t \lambda d\mu(x) dt \\ &\quad + \int_{\Omega} \nu \lambda_0 dx - \int_{\Omega} \lambda_0 d\mu(t=0) - \int_{\Omega} \lambda(t=1) d\mu(t=1) \\ &\quad + \int_{\Omega} u \lambda(t=1) dx + \int_{\Omega} \eta u dx + \eta_0 \left(\int_{\Omega} u dx - 1 \right). \end{aligned}$$

The optimality conditions for this problem are (2.10)–(2.13) and

$$\mu(t=0) = \nu, \quad (2.16)$$

$$\lambda(t=0) = \lambda_0, \quad (2.17)$$

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

$$\epsilon E(u) + \int_{\Omega} u \lambda(t=1) dx \rightarrow \min_u. \quad (2.18)$$

2.3 Existence

In this section we prove existence of minimisers of (1.4). For most of the considered functionals, i.e., (1.6)–(1.9) existence of solutions has been already dealt with or can directly be inferred from general results due to displacement convexity [4, 25, 28, 34, 39, 40, 57, 60]. Therefore, we concentrate here on the proof for the total variation case only, i.e., $E(u) = |Du|(\Omega)$, which to the best of our knowledge has not been treated yet in the literature. However, this proof can be easily adapted also to other choices of $E(u)$, and in particular the existence result holds as well for the alternative regularisers (1.6)–(1.9) discussed in this paper.

Let $X = \Omega \subset \mathbb{R}^d$ and open and bounded domain with compact Lipschitz boundary. We consider the following minimisation problem

$$\mathcal{J}(u) = \frac{1}{2}W_2(\nu, u\mathcal{L}^d)^2 + \epsilon |Du|(\Omega) \rightarrow \min_{u\mathcal{L}^d \in \mathbb{P}(X)}, \quad (2.19)$$

where $\nu \in \mathbb{P}(X)$, the set of probability measures on X , u is a probability density, and \mathcal{L}^d is the usual Lebesgue measure on \mathbb{R}^d . We have the following existence result.

Theorem 2.2. *The functional \mathcal{J} in (2.19) has a minimiser $u \in \mathbb{P}(X)$.*

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

$$\mathcal{J}(u^n) \rightarrow \inf_{u\mathcal{L}^d \in \mathbb{P}(X)} \mathcal{J}(u) =: m \quad \text{as } n \rightarrow \infty.$$

Since u^n is a minimising sequence there exists a constant $M > 0$ such that

$$M \geq \mathcal{J}(u^n) = \frac{1}{2}W_2(\nu, u^n\mathcal{L}^d)^2 + \epsilon |Du^n|(\Omega),$$

and hence

$$W_2(\nu, u^n\mathcal{L}^d) \leq M \text{ and } |Du^n|(\Omega) \leq M \quad \forall n \geq 1.$$

Additionally u^n is uniformly bounded in $L^1(X, \mathcal{L}^d)$. Since the Lebesgue measure is the standard measure in Euclidean space, the space $L^p(X, \mathcal{L}^d)$ is the usual $L^p(X)$ and integration with respect to \mathcal{L}^d will be denoted by dx . The uniform boundedness of u^n in L^1 follows from the assumption that $u^n\mathcal{L}^d \in \mathbb{P}(X)$. Namely, since $u^n\mathcal{L}^d \in \mathbb{P}(X)$ 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 $u \in BV(X)$ such that $u^n \rightarrow u$ strongly in $L^1(X)$ and weakly* in $BV(X)$. From the convergence in $L^1(X)$ we have

$$\int_{\Omega} \psi d(u^n\mathcal{L}^d) \rightarrow \int_{\Omega} \psi d(u\mathcal{L}^d), \quad \forall \psi \in L^1(X).$$

Since $C_b^0(X) \subset L^p(X)$ for all $p \geq 1$, it follows

$$\int_{\Omega} \psi d(u^{n_k}\mathcal{L}^d) \rightarrow \int_{\Omega} \psi d(u\mathcal{L}^d), \quad \forall \psi \in C_b^0(X),$$

and therefore $u^n\mathcal{L}^d \subset \mathbb{P}(X)$ narrowly converges to $u\mathcal{L}^d \in \mathbb{P}(X)$.

Now we can apply the Proposition 2.1. In particular, in our setting, we have $\nu \in \mathbb{P}(X)$, $u^n\mathcal{L}^d \subset \mathbb{P}(X)$, and $u^n\mathcal{L}^d$ narrowly converges to $u\mathcal{L}^d$. Let further be Π^n be a sequence of

optimal transport plans. Since $W_2(\nu, u^n \mathcal{L}^d) := \int_{X^2} |x - y|^2 d\Pi^n(x, y) < \infty$, we can apply Proposition 2.1 and get: Π^n is narrowly relatively compact in $\mathbb{P}(X^2)$ and any narrow limit point Π is an optimal transport plan for $(\nu, u \mathcal{L}^d)$, with

$$W_2(\nu, u \mathcal{L}^d) \leq \liminf_{n \rightarrow \infty} \int_{X^2} |x - y|^2 d\Pi^n(x, y) = \liminf_{n \rightarrow \infty} W_2(\nu, u^n \mathcal{L}^d). \quad (2.20)$$

Moreover from the lower-semicontinuity of $|Du|(\Omega)$ in $L^1(X)$ we get

$$\liminf_{n \rightarrow \infty} \mathcal{J}(u^n) \geq \frac{1}{2} W_2(\nu, u \mathcal{L}^d)^2 + \epsilon |Du|(\Omega) = \mathcal{J}(u).$$

It follows that $\mathcal{J}(u)$ has at least one minimiser $u \in BV(\Omega)$ and $u \mathcal{L}^d \in \mathbb{P}(X)$. \square

2.4 Uniqueness

In this subsection we shall investigate the uniqueness of both the variational problems (1.4) and the Benamou–Brenier optimality conditions (2.10)–(2.15).

In order to prove uniqueness of a minimiser of (1.4) we use the dual formulation introduced above. For fixed u the maximisation of the dual variables φ and ψ results in the optimality conditions

$$u(x) dx = \int_{\Omega} d\Pi(x, y) \quad (2.21)$$

$$d\nu(x) dx = \int_{\Omega} d\Pi(y, x) \quad (2.22)$$

and the complementarity condition

$$\lambda \geq 0, \quad (\varphi(x) + \psi(y) - \frac{1}{2}|x - y|^2) d\Pi(x, y) = 0, \quad (2.23)$$

where actually Π is the solution in the primal Kantorovich formulation.

Note that the optimality condition for u implies

$$\epsilon \langle E'(u), v - u \rangle \leq - \int_{\Omega} \varphi(v - u) dx \quad (2.24)$$

for all admissible v .

Theorem 2.3. *Let E be differentiable and strictly convex and let $\epsilon > 0$. Then there exists at most one minimiser of (1.4).*

Proof. Assume that u_1 and u_2 are minimisers of (1.4) with associated product measures Π and dual variables $\varphi_i, \psi_i, i = 1, 2$. Then we conclude from (2.24)

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

Inserting the property of u_i being a marginal density of Π_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

$$\begin{aligned}
-\int_{\Omega \times \Omega} (\varphi_1(x) - \varphi_2(x)) d\Pi(x, y) &= -\int_{\Omega \times \Omega} \varphi_1(x) d\Pi_1(x, y) - \int_{\Omega \times \Omega} \varphi_2(x) d\Pi_2(x, y) \\
&\quad + \int_{\Omega \times \Omega} \varphi_1(x) d\Pi_2(x, y) + \int_{\Omega \times \Omega} \varphi_2(x) d\Pi_1(x, y) \\
&\leq \int_{\Omega} (\psi_1(y) - \psi_2(y)) d\nu(y) - \int_{\Omega} (\psi_1(y) - \psi_2(y)) d\nu(y) = 0.
\end{aligned}$$

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

Remark 2.4. Another possible strategy to prove the uniqueness of minimisers of (1.4) is to show that the regularising functionals $E(u)$ are (λ) -geodesically convex, i.e., convex on the space of probability measures endowed with the Wasserstein distance $(\mathbb{P}(X), W_2(\cdot, \cdot))$. A class of functionals among which we can find examples which are convex along generalized geodesics is given by the internal energy [4, Proposition 9.3.9]. This class contains the squared Lebesgue norm (1.7) and the log-entropy (1.8). Now for the other functionals under consideration, i.e., the ones involving derivatives of u , we either do not know if they are geodesically convex or we actually know that they are not. Now consider first-order functionals of the form

$$E_\alpha(u) := \frac{1}{2\alpha} \int_{\mathbb{R}^d} |Du^\alpha|^2 dx.$$

Among this family of functionals we have two prominent candidates, that is the classical Fisher information (1.9) when $\alpha = 1/2$ and the Dirichlet energy (1.6) for $\alpha = 1$. Even for these cases E_α are not geodesically convex, cf. [16]. For the total variation regularised optimal transport, no results are known to the authors.

In the following we investigate the uniqueness of a solution of the Benamou Brenier optimality conditions (2.10)-(2.15).

Theorem 2.5. *Let E be differentiable and strictly convex, let $\epsilon > 0$ and let (μ, λ, u) be a solution of (2.10)-(2.15). Then the transport velocity $v = \nabla \lambda$ is unique on the support of μ .*

Proof. To prove the claim we basically follow the strategy to prove uniqueness for mean-field games introduced in [38]. Let us first rewrite the equations. Inserting (2.12), i.e., $v = \nabla \lambda$, into (2.10) and (2.11) we get

$$\begin{aligned}
\mu_t + \nabla \cdot (\mu \cdot \nabla \lambda) &= 0 \\
\lambda_t + \frac{1}{2} |\nabla \lambda|^2 &= 0.
\end{aligned} \tag{2.25}$$

Let (μ_1, λ_1, u_1) and (μ_2, λ_2, u_2) be two solutions to (2.10)-(2.15). Writing the equations in (2.25) for both solutions and subtracting the corresponding ones we get

$$\begin{aligned}
(\mu_1 - \mu_2)_t + \nabla \cdot (\mu_1 \nabla \lambda_1 - \mu_2 \nabla \lambda_2) &= 0 \\
(\lambda_1 - \lambda_2)_t + \frac{1}{2} (|\nabla \lambda_1|^2 - |\nabla \lambda_2|^2) &= 0.
\end{aligned}$$

By multiplying the first equation with $(\lambda_1 - \lambda_2)$, the second equation with $(\mu_1 - \mu_2)$, integrating both equations over $\Omega \times (0, 1)$ and adding them we get

$$\begin{aligned}
& \int_{\Omega \times (0,1)} (\mu_1 - \mu_2)_t (\lambda_1 - \lambda_2) + (\mu_1 - \mu_2) (\lambda_1 - \lambda_2)_t \, dxdt \quad (= I) \\
& + \int_{\Omega \times (0,1)} \nabla \cdot (\mu_1 \nabla \lambda_1 - \mu_2 \nabla \lambda_2) (\lambda_1 - \lambda_2) \, dxdt \quad (= II) \\
& + \frac{1}{2} \int_{\Omega \times (0,1)} (|\nabla \lambda_1|^2 - |\nabla \lambda_2|^2) (\mu_1 - \mu_2) \, dxdt \quad (= III) \\
& = 0.
\end{aligned}$$

Then

$$\begin{aligned}
I &= \int_{\Omega \times (0,1)} ((\mu_1 - \mu_2) \cdot (\lambda_1 - \lambda_2))_t \, dxdt \\
&= \int_{\Omega} (\mu_1 - \mu_2) \cdot (\lambda_1 - \lambda_2) \Big|_0^1 \, dx \\
&= \int_{\Omega} (\mu_1(t=1) - \mu_2(t=1)) \cdot (\lambda_1(t=1) - \lambda_2(t=1)) \, dx \\
&= - \int_{\Omega} (u_1 \mathcal{L}^d - u_2 \mathcal{L}^d) \cdot (E'(u_1) - E'(u_2)) \, dx.
\end{aligned}$$

Since E is strictly convex $I \leq 0$. 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

$$\begin{aligned}
II &\leq \frac{1}{2} \int_{\Omega \times (0,1)} \mu_1 (|\nabla \lambda_2|^2 - |\nabla \lambda_1|^2) + \mu_2 (|\nabla \lambda_1|^2 - |\nabla \lambda_2|^2) \, dxdt \\
&= -\frac{1}{2} \int_{\Omega \times (0,1)} (\mu_1 - \mu_2) (|\nabla \lambda_1|^2 - |\nabla \lambda_2|^2) \, dxdt \\
&= -III.
\end{aligned}$$

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

$$\begin{aligned}
& - \int_{\Omega \times (0,1)} (\mu_1 \nabla \lambda_1 - \mu_2 \nabla \lambda_2) (\nabla \lambda_1 - \nabla \lambda_2) + \frac{1}{2} (|\nabla \lambda_1|^2 - |\nabla \lambda_2|^2) (\mu_1 - \mu_2) \, dxdt = 0 \\
& - \int_{\Omega \times (0,1)} (\nabla \lambda_1 - \nabla \lambda_2)^2 (\mu_1 + \mu_2) \, dxdt = 0,
\end{aligned}$$

and hence either $\nabla \lambda_1 - \nabla \lambda_2 = 0$ 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., $\nabla \lambda_1 = \nabla \lambda_2$. \square

Remark 2.6. From Theorem 2.5 we have that the Benamou-Brenier optimality conditions deliver a unique transport in the sense that the velocity $v = \nabla \lambda$ and the endpoint $u \mathcal{L}^d = \mu(t=1)$ of the transport are unique on the support of μ . Since we cannot guarantee sufficient regularity of the transport velocity v [1] this is not enough to deduce uniqueness of μ and λ on the whole time interval. In particular the solution may be non-unique outside the support of μ , cf. also section 3 for examples of such self-similar solutions.

2.5 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 ν are close in the Wasserstein metric. It has been demonstrated by many authors (cf. [10, 11, 47, 6]) that Bregman distances related to the regularisation

functional E are natural error measures for variational regularisation methods with E convex. In particular Pöschl [47] 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 ν_1 and ν_2 are two given probability measures with corresponding minimisers u_1 and u_2 of (1.4). Let further be Π_i , $i = 1, 2$, the associated product measures and φ_i and ψ_i , $i = 1, 2$, the variables in the general dual formulation of the Kantorovich formulation.

Theorem 2.7. *Let E be differentiable and strictly convex, let $\epsilon > 0$, and let u_i , $i = 1, 2$ be minimisers of (1.4) for ν_i , $i = 1, 2$, 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. As before in the proof of Theorem 2.3 we use inequality (2.24) 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 in terms of c-concave functions (which are the function and its Legendre transform in the current case of the quadratic Wasserstein distance). In fact, the supremum in (2.4) 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. [60, Proof of Theorem 4 in Chapter 1]. Substituting φ_i^{cc} for φ_i and φ_i^c for ψ_i , $i = 1, 2$, and following the steps in the proof of Theorem 2.3 we obtain

$$\begin{aligned} \epsilon \langle E'(u_1) - E'(u_2), u_1 - u_2 \rangle &\leq - \int_{\Omega} (\varphi_1^{cc} - \varphi_2^{cc})(u_1 - u_2) dx \\ &\leq \int_{\Omega} (\varphi_1^c(y) - \varphi_2^c(y)) d(\nu_1 - \nu_2)(y). \end{aligned}$$

The functions φ_i^c , $i = 1, 2$ satisfy

$$\begin{aligned} \varphi_i^c(y) - \varphi_i^c(y') &= \sup_x \left\{ \varphi_i(x) - \frac{1}{2} |x - y|^2 \right\} + \inf_x \left\{ \frac{1}{2} |x - y|^2 - \varphi_i(x) \right\} \\ &\leq \sup_x \left\{ \varphi_i(x) - \frac{1}{2} |x - y|^2 \right\} + \sup_x \left\{ \frac{1}{2} |x - y|^2 - \varphi_i(x) \right\} \\ &= \frac{1}{2} \sup_x \{ |x - y|^2 - |x - y'|^2 \} \quad \underbrace{\leq}_{\text{triangle inequality}} \quad \frac{1}{2} |y - y'|^2, \end{aligned}$$

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

$$\begin{aligned} \epsilon \langle E'(u_1) - E'(u_2), u_1 - u_2 \rangle &\leq \int_{\Omega} (\varphi_1^c(y) - \varphi_2^c(y)) d(\nu_1 - \nu_2)(y) \\ &\leq \sup_{\varphi} \left\{ \int_{\Omega} \varphi(y) d(\nu_1 - \nu_2)(y); \varphi \in \bigcap L^1(d|\nu_1 - \nu_2|), \|\varphi\|_{Lip} \leq 1 \right\}. \quad (2.26) \end{aligned}$$

We recall the following 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 (1.4) we have $d(x, y) = |x - y|^2$.

Theorem 2.8. (*Kantorovich-Rubinstein theorem*)

Let $X = Y$ be a complete separable metric space and let d be a lower semi-continuous metric on X . Let \mathcal{T}_d be the cost of optimal transportation for the cost $c(x, y) = d(x, y)$,

$$\mathcal{T}_d(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times X} d(x, y) d\pi(x, y).$$

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

$$\mathcal{T}_d(\mu, \nu) = \sup \left\{ \int_X \varphi d(\mu - \nu); \varphi \in \bigcap L^1(d|\mu - \nu|), \|\varphi\|_{\text{Lip}} \leq 1 \right\}.$$

Using Theorem 2.8 the last expression in (2.26) is equal to the cost of the optimal transport between ν_1 and ν_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 total variation) a completely analogous proof holds to obtain an estimate for the generalized Bregman distance

$$D_E^{p_1, p_2}(u_1, u_2) := \langle p_1 - p_2, u_1 - u_2 \rangle \quad p_i \in \partial E(u_i), \quad (2.27)$$

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. [10, 11, 45]).

2.6 Random Sampling

In the following we analyze the consistency of the estimate obtained from our variational approach in the classical case of available samples from a given probability measure ν . For this sake we introduce the empirical measure ν^N , which is a random variable depending on the samples X_1, \dots, X_N :

$$\nu^N = \frac{1}{N} \sum_{j=1}^N \epsilon_{X_j} \quad (2.28)$$

where ϵ_X is the point measure at X .

We denote by $u^{\epsilon, N}$ a minimiser of the variational problem with given measure ν^N . Note that also u^N is a random variable depending on the samples and thus also subsequent estimates

are to be interpreted for any realization of the random variable. Assuming that ν is regular enough, i.e. there exists a density \hat{u} such that

$$\nu = \hat{u}\mathcal{L}^d, \quad E(\hat{u}) < \infty, \quad (2.29)$$

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

$$W_2(u^{\epsilon, N}, \nu^N)^2 \leq W_2(u^{\epsilon, N}, \nu^N)^2 + 2\epsilon E(u^{\epsilon, N}) \leq W_2(\nu, \nu^N)^2 + 2\epsilon E(\hat{u}), \quad (2.30)$$

thus by the triangle inequality

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

In particular we obtain for the expected value

$$\mathbf{E}(W_2(u^{\epsilon, N}\mathcal{L}^d, \nu)) \leq 2\mathbf{E}(W_2(\nu^N, \nu)) + \sqrt{2\epsilon E(\hat{u})}. \quad (2.31)$$

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 [48] gives a rate of order $N^{-1/d}$ for distributions with uniformly bounded moments, while Horowitz et. al. [33] even provide a rate of order $N^{-2/(d+4)}$ if the $(d+5)$ -th moment is finite. The second term is obviously converging to zero as $\epsilon \rightarrow 0$, we thus obtain:

Theorem 2.9. *Let ν be such that (2.29) is satisfied, then the variational model (1.4) is consistent in the sense of narrow convergence, which is metrized by the Wasserstein metric, more precisely*

$$\mathbf{E}(W_2(u^{\epsilon, N}\mathcal{L}^d, \nu)) \rightarrow 0 \quad (2.32)$$

as $N \rightarrow \infty$ and $\epsilon \rightarrow 0$.

Note that from the arguments above we even obtain a rate for the Wasserstein metric in terms of ϵ and N under additional moment conditions, respectively a rate in terms of N if ϵ is chosen in dependence of N , 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 . Again the Bregman distance, more precisely $\mathbf{E}(D_E(u^{\epsilon, N}, \hat{u}))$ appears to be a correct error measure. Note that from Theorem 2.7 we directly obtain an estimate for the reconstructed densities

$$\mathbf{E}(D_E(u^{\epsilon, N}, u)) \leq \frac{1}{\epsilon} \mathbf{E}(W_2(\nu^N, \nu)^2). \quad (2.33)$$

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 [47]. 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 one can directly combine the techniques from Theorem 2.7 and [47] to obtain an estimate of the form

$$\mathbf{E}(D_E(u^{\epsilon, N}, \hat{u})) \leq \frac{1}{\epsilon} F\left(\frac{1}{N}\right) + G(\epsilon) \quad (2.34)$$

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. Since the computations are rather lengthy and not the central scope of the paper, we omit further details here.

2.7 Rescaling and Diffusion Filtering

In the Benamou-Brenier formulation we can perform standard rescaling of the time variable and associated rescaling of density, velocity, and dual variable to

$$\hat{t} = \epsilon t, \quad \hat{\mu}(dx, \hat{t}) = \mu(dx, t), \quad \hat{v}(x, \hat{t}) = \frac{1}{\epsilon} v(x, t), \quad \hat{\lambda}(x, \hat{t}) = \lambda(x, t). \quad (2.35)$$

The optimisation problem (2.6)-(2.9) in rescaled variables becomes

$$E(u) + \frac{1}{2} \int_0^\epsilon \int_\Omega |\hat{v}|^2 d\hat{\mu}(x) dt, \quad (2.36)$$

subject to

$$\partial_{\hat{t}} \hat{\mu} + \nabla \cdot (\hat{\mu} \hat{v}) = 0 \quad (2.37)$$

$$\hat{\mu}(t=0) = \nu \quad (2.38)$$

$$\hat{\mu}(t=1) = u \mathcal{L}^d \quad (2.39)$$

and the optimality condition with dual variable $\hat{\lambda}$ is again $\hat{v} = \frac{1}{\epsilon} \nabla \hat{\lambda}$ and

$$\partial_{\hat{t}} \hat{\lambda} + \hat{v} \cdot \nabla \hat{\lambda} - \frac{1}{2} |\hat{v}|^2 = 0. \quad (2.40)$$

The rescaling is useful in the numerical solution, since it guides the choice of the time discretization in ϵ . Moreover, the relation to diffusion filtering can be understood by the asymptotic $\epsilon \rightarrow 0$. The limit $\epsilon \rightarrow 0$ of a variational model as (1.4), respectively a sequence thereof, has been well studied via the method of minimising movements approximating a metric gradient flow with energy E (cf. [4]). Then the gradient flow is given by (1.13). For the Wasserstein metric this relation has been first studied in [34] for the linear Fokker-Planck equation. The corresponding discrete time stepping strategy for the construction of solutions of (1.13) is known under the name Jordan-Kinderlehrer-Otto (JKO)-scheme [34]. 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., [46, 27, 13, 17, 57, 58, 14, 20, 28, 39]. For the energies E considered in (1.5)–(1.9) the corresponding gradient flows (1.13) are listed in Table 1

3 Self-Similar Solutions

In the following we compute an interesting special class of solutions of self-similar structure. The basis of our computations is the characterization of the optimal transport between a measure and its rescaled version, derived in [5]. 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

$$d\nu = \frac{1}{\delta^d} u\left(\frac{x}{\delta}\right) dx. \quad (3.1)$$

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.

In the Benamou-Brenier formulation it is well-known that optimal transport between two scaled versions of the same density is performed with density

$$\rho(x, t) = \frac{1}{(at + b)^d} u\left(\frac{x}{at + b}\right), \quad (3.2)$$

velocity and dual variable

$$v(x, t) = \frac{ax}{at + b}, \quad \lambda(x, t) = \frac{a|x|^2}{2(at + b)} \quad (3.3)$$

with nonnegative constants a and b such that $a + b = 1$.

With this particular λ we examine the optimality condition (2.13). Note that for (2.13) the functional E needs to be defined as $+\infty$ if u violates nonnegativity and integral equals one. We reformulate these with the help of Lagrange multipliers to obtain

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

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

$$E'(u) = -\frac{|x|^2}{2\sigma^2} + \frac{\eta(x)}{\epsilon} - \frac{\eta_0}{\epsilon}, \quad (3.4)$$

where still η satisfies complementarity. We compute such solutions in the most relevant cases, that means for different regularisation energies $E(u)$.

3.1 Logarithmic Entropy

In the case of the logarithmic entropy $E(u) = \int_{\Omega} u \log u \, dx$ (1.8) we obtain for (3.4)

$$u(x) = \exp\left(-\frac{|x|^2}{2\sigma^2} + \frac{\eta(x)}{\epsilon} - \frac{\eta_0}{\epsilon} - 1\right). \quad (3.5)$$

Obviously u is positive everywhere and therefore $\eta u = 0$ yields to $\eta \equiv 0$. We obtain η_0 by solving

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

After some computations we obtain $C = (2\sigma^2\pi)^{-d/2}$ and finally

$$u(x) = \frac{1}{(2\sigma^2\pi)^{d/2}} \exp\left(-\frac{|x|^2}{2\sigma^2}\right). \quad (3.6)$$

We show this solutions for different ϵ in Figure 1(a). 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 ν is a point measure centered at zero.

3.2 Squared Lebesgue Regularisation

In the case of the squared Lebesgue norm $E(u) = \frac{1}{2} \int_{\Omega} u^2 dx$ (1.7) the condition (3.4) becomes

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

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

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

In one dimension the constant η_0 can be easily computed for a u symmetrically centered around zero

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

After computing the zeros x_0 we conclude

$$u(x) = \left(-\frac{|x|^2}{2\sigma^2} + \frac{\sqrt[3]{9}}{(\sigma 4\sqrt{2})^{2/3}} \right)_+. \quad (3.7)$$

We display this results in Figure 1(b).

3.3 Dirichlet Regularisation

In the following we to compute the self-similar solution for the Dirichlet regularisation $E(u) = \frac{1}{2} \int_{\Omega} |\nabla u(x)|^2 dx$ (1.6), for which (3.4) yields

$$-\Delta u = -\frac{|x|^2}{2\sigma^2} + \frac{\eta(x)}{\epsilon} - \frac{\eta_0}{\epsilon}, \quad (3.8)$$

and in one-dimension for $u(x) > 0$ in a neighborhood around zero we can simplify (3.8) to

$$u''(x) = \left(\frac{x^2}{2\sigma^2} + \frac{\eta_0}{\epsilon} \right)_+, \quad (3.9)$$

$$u(x) = \left(\frac{x^4}{24\sigma^2} + \frac{C}{2}x^2 + \tilde{C}x + \bar{C} \right)_+, \quad \text{with } C = \frac{\eta_0}{\epsilon}, \tilde{C}, \bar{C} \in \mathbb{R}. \quad (3.10)$$

We can compute the constants using the property, that our solution is centered at zero and symmetric, e.g. $u(-x_0) = u(x_0) = 0$, $u'(-x_0) = u'(x_0) = 0$.

After some computations we obtain $\tilde{C} = 0$, $\bar{C} = \frac{3}{2\epsilon}\eta_0^2\sigma^2$. We obtain $C = -\frac{1}{\epsilon}\sqrt[5]{\frac{25a^3\epsilon^2}{384}}$ as a result of

$$\int_{-x_0}^{x_0} u(x) dx = \int_{-x_0}^{x_0} \left(\frac{x^4}{24\sigma^2} + C\frac{x^2}{2} + \bar{C} \right)_+ dx = 1.$$

Finally the self-similar solution for the Dirichlet regularisation is

$$u(x) = \left(\frac{x^4}{24\sigma^2} - \frac{1}{2\epsilon}\sqrt[5]{\frac{25a^3\epsilon^2}{384}}x^2 + \frac{3}{2\epsilon a}\eta_0^2 \right)_+. \quad (3.11)$$

We refer to Figure 1(c) for an illustration of (3.11).

3.4 Total Variation

In case of the total variation (1.5), 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 characterized by

$$\partial E(u) = \left\{ \nabla \cdot g : g \in C_0^\infty(\Omega, \mathbb{R}^d), \|g\|_\infty \leq 1 \right\}.$$

Hence, from the optimality condition (3.4) in one dimension we have

$$g'(x) = -\frac{x^2}{2\sigma^2} + \frac{\eta}{\epsilon} - \frac{\eta_0}{\epsilon}, \text{ with } g' \in \partial E(u). \quad (3.12)$$

Assuming $u(x) \geq 0$ in neighbourhood around zero we have that $\eta(x)$ vanishes within this neighbourhood and hence

$$g'(x) = -\frac{x^2}{2\sigma^2} - C, \text{ with } C = \eta_0/\epsilon \in \mathbb{R}$$

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

$$2\alpha = |Du| = \int u g' = \int_{-x_0}^{x_0} \alpha g' = \alpha(g(x_0) - g(-x_0)) = 2\alpha g(x_0),$$

and hence $g(x_0) = 1$. Note that this also reassembles the characterization 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 α by using the optimality condition (3.12) and the mass constraint on u respectively, i.e., which yields $x_0 = \sqrt{-2\tilde{C}\sigma}$, and $\alpha = 1/(2x_0) = 1/(2\sqrt{-2\tilde{C}\sigma})$.

The constant $C = \eta_0/\epsilon$ is obtained from $1 = g(x_0)$. Finally we have computed a self similar solution of the form

$$u(x) = \begin{cases} 1/(2x_0) & (-x_0, x_0), \\ 0 & \text{otherwise.} \end{cases}$$

with, for $\tilde{C} = 0$, $x_0 = \sqrt[6]{9\frac{\epsilon}{a}}$. Please compare Figure 1(d) for a sketch of such solutions.

3.5 Fisher Information

Computing the self-similar solution for the Fisher information $E(u) = \int_\Omega u |\nabla \ln u|^2 dx = \int_\Omega \frac{|\nabla u|^2}{u} dx$ (1.9) means to solve

$$-(\nabla \ln u)^2 - 2\Delta \ln u = -\frac{|x|^2}{2\sigma^2} + \frac{\eta(x)}{\epsilon} - \frac{\eta_0}{\epsilon}. \quad (3.13)$$

As before, we consider a positive solution $u(x) > 0$ in a neighborhood of zero and the one dimensional case for simplicity:

$$-((\ln u)')^2 - 2(\ln u)'' = -\frac{|x|^2}{2\sigma^2} - \frac{\eta_0}{\epsilon}. \quad (3.14)$$

In the next step we substitute $u = e^y$

$$-(y')^2 - 2y'' = -\frac{|x|^2}{2\sigma^2} - \frac{\eta_0}{\epsilon},$$

and obtain the solution $y = -\frac{x^2}{2\sqrt{2}\sigma} + C$ with $C \in \mathbb{R}$. The constant C follows from

$$\int_{\Omega} u(x) dx = \int_{\Omega} e^{-\frac{x^2}{2\sqrt{2}\sigma}} e^C dx = 1$$

Finally we obtain

$$u(x) = \frac{1}{(2\sqrt{2}\sigma\pi)^{1/2}} e^{-\frac{x^2}{2\sqrt{2}\sigma}} \quad (3.15)$$

and illustrate the result in Figure 1(e).

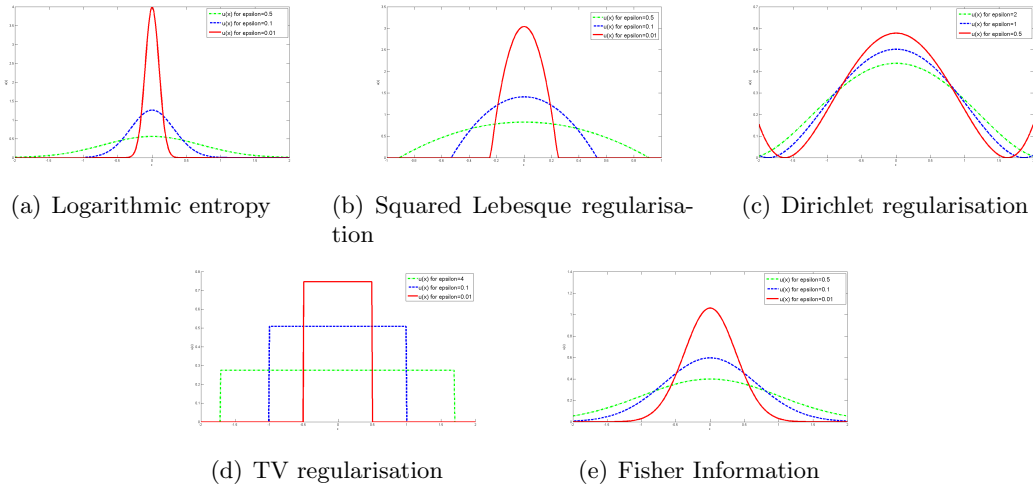


Figure 1: Self-similar Solutions for different Regularisation functionals (1.5)-(1.9) for different ϵ .

4 Numerical Methods and Results

In this section we present numerical results for the approaches to solve (1.4) introduced in the previous sections. In particular, we discuss solutions for the Kantorovich optimal transport problem (2.1)-(2.3). This problem is only efficiently solvable in one spatial dimension. Therefore we additionally consider the Benamou-Brenier formulation (2.6)-(2.9), for which we present numerical results in one and two dimensions. Since the solution of the optimisation problem (2.1)-(2.3) can be interpreted as a discrete approximation of a solution of the

gradient flow of $E(u)$ with respect to the L^2 Wasserstein metric we additionally solve the diffusion equation (1.13), and compare the numerical results with the solutions of (2.1)-(2.3). In all three numerical approaches we consider the regularisation energies (1.5)-(1.9).

4.1 Discretization

For the numerical implementation we need to approximate the derivatives in the differential equations involved in our models. We consider finite difference schemes for this behalf. Let $\Omega = [a, b]^d$, $d = 1, 2$ be the space we consider for our problem. We partitionate the space-time-cylinder $[a, b]^d \times [0, T]$ into N equidistant cells $[x_i, x_{i+1}]^d$, $i = 0, \dots, N-1$ with $x_0 = a$, $x_N = b$ and step-size $\Delta x = (b-a)/N$ and M time intervalls $[t_j, t_{j+1}]$ for $j = 0, \dots, M-1$ and time-steps $\Delta t = \frac{T}{M}$. Let u be an arbitrary function defined on Ω , then we denote with $u_i^l = u(x_i, t_l)$ the approximation of u in the node x_i at time t_l . In the one-dimensional case we approximate the spatial derivative of u in x_i at time t_l with forward-difference quotients $((D_x^+ u)_i^l)$ or backward-difference quotients $((D_x^- u)_i^l)$

$$u_x(x_i, t_l) \approx \left\{ (D_x^+ u)_i = \begin{cases} \frac{u_{i+1} - u_i}{\Delta x} & i = 0, \dots, N-1 \\ \frac{u_N - u_{N-1}}{\Delta x} & i = N \end{cases}, (D_x^- u)_i = \begin{cases} \frac{u_i - u_{i-1}}{\Delta x} & i = 1, \dots, N \\ \frac{u_1 - u_0}{\Delta x} & i = 0 \end{cases} \right\}. \quad (4.1)$$

For the second spatial derivative we formulate the discrete Laplace-operator in the following way:

$$(D_{\Delta} u)_i = \frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2}. \quad (4.2)$$

For two-dimensional computations we discretize ∇u by $(\nabla u(x_i, y_j, t_l)) \approx ((D_x^+ u)_{i,j}^l, (D_y^+ u)_{i,j}^l)$

$$(D_x^+ u)_{i,j}^l = \begin{cases} \frac{u_{i+1,j}^l - u_{i,j}^l}{\Delta x} & i = 0, \dots, N-1 \\ \frac{u_{N,j}^l - u_{N-1,j}^l}{\Delta x} & i = N \end{cases}, (D_y^+ u)_{i,j}^l = \begin{cases} \frac{u_{i,j+1}^l - u_{i,j}^l}{\Delta y} & j = 0, \dots, N-1 \\ \frac{u_{i,N}^l - u_{i,N-1}^l}{\Delta y} & j = N \end{cases}. \quad (4.3)$$

The discrete Laplace-operator and divergence-operator in two dimensions can be written as:

$$(D_{\Delta} u)_{i,j}^l = \frac{u_{i+1,j}^l - 4u_{i,j}^l + u_{i-1,j}^l + u_{i,j+1}^l + u_{i,j-1}^l}{\Delta x \Delta y}, \quad (4.4)$$

$$(\operatorname{div} u)_{i,j}^l = (D_x^+ u)_{i,j}^l + (D_y^+ u)_{i,j}^l. \quad (4.5)$$

4.2 Convex Minimisation Schemes

In the following we present some numerical results for the variational model (2.1)-(2.3). Let $X = \{x_1, \dots, x_N\} \subset \mathbb{R}^d$ and $Y = \{y_1, \dots, y_M\} \subset \mathbb{R}^d$ be two discrete density distributions. Then

it is possible to reformulate (2.1)-(2.3) as a linear program:

$$\min_{p,u} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M |x_i - y_j| p_{i,j} + \epsilon E(u_i), \quad (4.6)$$

$$\text{s.t.} \quad \sum_{i=1}^N p_{i,j} = a_j \quad j = 1, \dots, M, \quad \sum_{j=1}^M p_{i,j} = u_i \quad i = 1, \dots, N, \quad (4.7)$$

$$0 \leq p_{i,j}, \quad 0 \leq u_i, \quad i = 1, \dots, N, \quad j = 1, \dots, M, \quad (4.8)$$

$$\sum_{i=1}^N \sum_{j=1}^M p_{i,j} = 1, \quad \sum_{i=1}^N u_i = \sum_{j=1}^M a_j = 1, \quad (4.9)$$

with given initial condition a_j and matrix $p_{i,j}$ (see [35]). In the case of E being convex, we obtain a convex linear problem. We solve the linear problem directly, i.e. we discretize (2.1)-(2.3) and optimize with the solver IPOPT. Without going into detail we mention here, that IPOPT uses an interior-point-method to solve linear and nonlinear constrained optimisation problems. See [62] for more details. In Figure 2 we present numerical results for the linear program (4.6)-(4.9) in one dimension. We consider different regularisation functionals (1.5)-(1.9). Note, that because of numerical difficulties we consider an approximation of the TV regularisation, namely $E(u) = \int \sqrt{|\nabla u|^2 + \delta}$ with small $\delta > 0$. Furthermore, for the computations with the Fisher information (1.9) we shift the initial data to 0.1 to avoid numerical problems arising from a division by zero.

Our initial condition is a Gaussian curve with mean zero and variance one (colored in blue). The disadvantage of the approach (2.1)-(2.3) this algorithm is that two dimensional applications are not efficiently solvable due to doubling the spatial dimension. Therefore, we neglect two dimensional for this approach and deal with that using (2.6)-(2.9) later.

4.3 Gradient Flow Ansatz

Following the theory of gradient flows we know that a solution of the minimisation problem (1.4) can be interpreted as a discrete approximation of a solution of the gradient flow of $E(u)$ with respect to the L^2 -Wasserstein metric. This motivates a comparison of the numerical results we get for (1.4) in one-dimension with discrete solutions of the equations (1.13) introduced in Table 1. Hence, in the following we present these discrete numerical solutions for the one-dimensional case.

Note here, that we are not trying to compete with existing numerical schemes for these equations, cf. section 1 for references. We solely aim for a comparison with solutions obtained from (1.4).

In Figure 3(a) we show discrete solutions for the heat equation

$$\partial_t u = \Delta u, \quad \forall x \in \Omega, t > 0, \quad (4.10)$$

$$u(x, 0) = u_0(x) > 0, \quad \forall x \in \Omega. \quad (4.11)$$

We discretize the time derivative using an explicit Euler method and a second order finite-difference-quotient (4.2) for the Laplace-operator with $\Delta x = 0.1$ and $\Delta t = 5 \cdot 10^{-4}$.

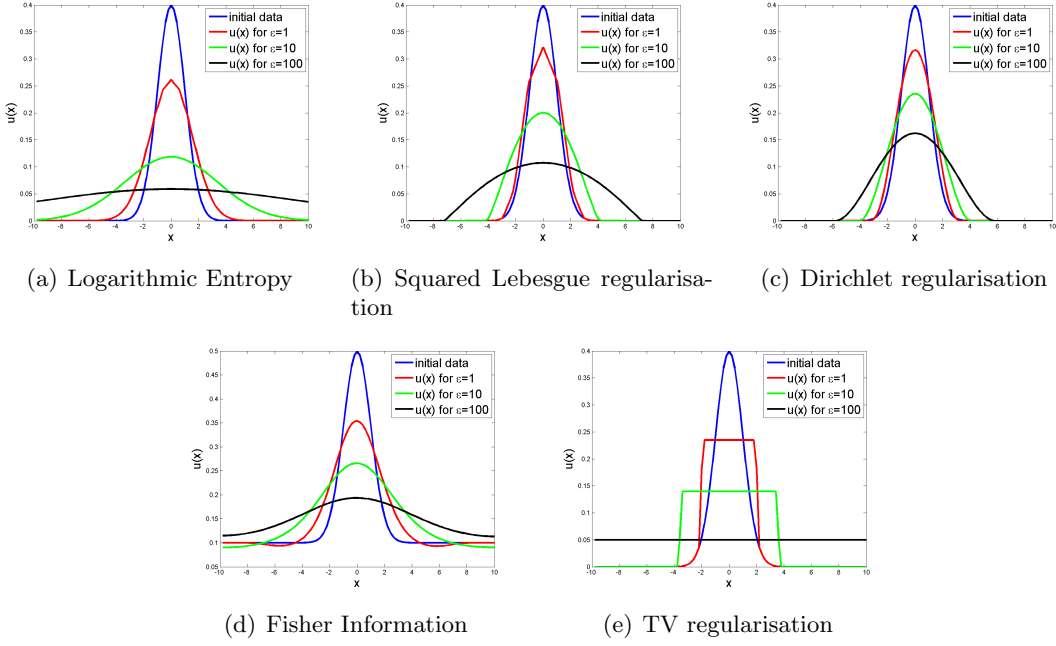


Figure 2: Numerical results for (2.1)-(2.3).

The squared Lebesgue regularisation yields, as mentioned in Table 1, the porous medium equation

$$\partial_t u = \Delta u^2, \quad \forall x \in \Omega, t > 0, \quad (4.12)$$

$$u(x, 0) = u_0(x) > 0, \quad \forall x \in \Omega. \quad (4.13)$$

Using the same numerical approximation as for (4.10)-(4.11) with $\Delta x = 0.2$ and $\Delta t = 5 \cdot 10^{-4}$ we obtain the discrete solutions which are presented in Figure 3(b) and are approximations of the in the literature well known Barenblatt-Pattle solutions [59]. Furthermore, the thin film equation

$$\partial_t u = -\operatorname{div}(u \nabla \Delta u), \quad \forall x \in \Omega, t > 0, \quad (4.14)$$

$$u(x, 0) = u_0(x) > 0, \quad \forall x \in \Omega, \quad (4.15)$$

is the associated gradient flow for the Dirichlet regularisation (1.7) with respect to the Wasserstein metric. After reformulating

$$\partial_t u = -\nabla \cdot (u \nabla \Delta u) = -\nabla u \nabla \Delta u - u \Delta \Delta u, \quad (4.16)$$

we discretize

$$(D_t^+ u)_i^l = -(D_x^+ u)_i^l D_x^- ((D_{\Delta} u)_i^l)_i^l - (u_i^l) D_{\Delta} (((D_{\Delta} u)_i^l)_i^l).$$

For the thin film equation the robust maximum principle is not guaranteed, and therefore the solution of (4.14) is not necessary bounded through the maximum or minimum of the initial data. To avoid numerical problems we have to guarantee that $u \geq 0$. The discrete solutions for $\Delta x = 0.2$ and time-stepsize $\Delta t = 5 \cdot 10^{-4}$ are presented in Figure 3(c).

Considering the gradient flow for the Fisher information

$$E(u) = \frac{1}{2} \int_{\Omega} u |\nabla \ln u|^2 dx = \frac{1}{2} \int_{\Omega} \frac{|\nabla u|^2}{u} dx = 2 \int_{\Omega} (\nabla \sqrt{u})^2 dx, \quad (4.17)$$

yields a non-linear fourth order equation, known as the Derrida-Lebowitz-Speer-Spohn (DLSS) equation

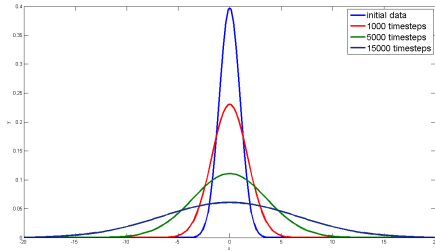
$$\partial_t u = -\Delta(u\Delta(\ln u)) = -2\nabla \left(u \nabla \left(\frac{\Delta \sqrt{u}}{\sqrt{u}} \right) \right), \quad \forall x \in \Omega, t > 0, \quad (4.18)$$

$$u(x, 0) = u_0(x) > 0, \quad \forall x \in \Omega. \quad (4.19)$$

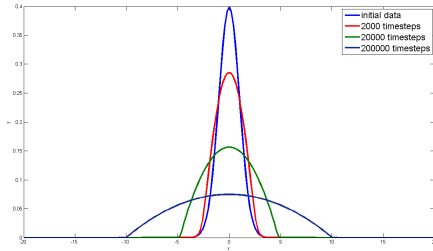
This result was proven by Gianazza, Savaré, Toscani [28]. Furthermore, the numerical solution of the DLSS equation is discussed in detail in [21]. We consider a semi-implicit discretization of (4.18) with $\Delta x = 0.4$ and time-stepsize $\Delta t = 5 \cdot 10^{-4}$

$$(D_t^+ u)_i^l = -D_{\Delta} \left(u_i^{l+1} D_{\Delta}(\ln u)_i^l \right) \quad (4.20)$$

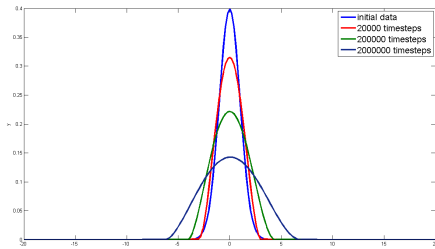
and present results in Figure 3(d).



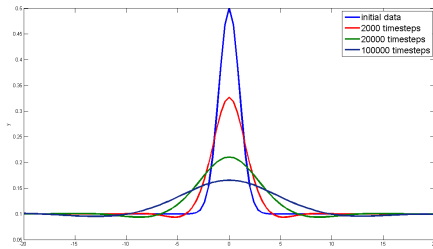
(a) Heat equation



(b) Porous Medium equation



(c) Thin Film equation



(d) DLSS

Figure 3: Discrete Solutions of the equations introduced in Table 1 after different timesteps.

4.4 Fluid Dynamic Schemes

In the following we present numerical results for the optimisation problem (2.6)-(2.9) based on the Benamou-Brenier formulation. As mentioned before, the convex minimisation scheme (4.6)-(4.9) is not appropriate to solve two dimensional problems efficiently. This yields the

motivation to introduce three iterative algorithms for (2.6)-(2.9), namely a gradient descent scheme, a dual ascent scheme and an augmented Lagrangian scheme, for the total variation regularisation. With these schemes also two dimensional models are efficiently treatable.

Algorithm: (GDS) Gradient Descent Scheme for (2.10)-(2.15)

Solve iteratively for $k = 0$ until $\|u^{k+1} - u^k\|_{L^2} \leq \delta$:

1. Initialisation: $\mu^{k+1}(t = 0) = \nu$.
2. Solve the state-equation (forward in time): $\partial_t \mu^{k+1} + \nabla \cdot (\mu^{k+1} v^k) = 0$.
3. Evaluate: $u^{k+1} = \mu^{k+1}(t = T)$, $\lambda^{k+1}(t = T) + \epsilon \partial E(u^{k+1}) \ni 0$.
4. Solve the adjoint-equation (backward in time): $\partial_t \lambda^{k+1} + v^k \nabla \lambda^{k+1} = \frac{1}{2} |v^k|^2$.
5. Update ($\tau > 0$): $(1 + \tau) v^{k+1} = \nabla \lambda^{k+1} + \tau v^k$.

The gradient descent scheme (GDS) iteratively finds a local minimum of (2.10)-(2.15) by going into the direction of the negative gradient. Starting with the initial data ν we first solve the state equation forward in time and then evaluate (3). (3) yields the initial condition for step (4), which is to solve the adjoint equation backward in time. In the last step we update the velocity v .

There is one difficulty that arises due to the condition (2.13). Only for $u > 0$ the complementary condition $\eta u = 0$ yields $\eta = 0$ and therefore $\epsilon \partial E(u) + \lambda(t = T) = 0$. That is a hard restriction, by reason that we can only expect $u > 0$ for the logarithmic entropy (3.6). This motivates the consideration of another iterative scheme for (2.10)-(2.15), the dual ascent scheme circumvent this problem. Starting with an initial condition for $\lambda(t = 0)$ we first solve the adjoint equation forward in time and afterwards the optimisation problem $\epsilon E(u) + \int_{\Omega} u \lambda^{k+1}(t = T) dx \rightarrow \min_u$. The solution, which is unique for strict convex $E(u)$, yields the initial condition for the state equation, which is solved backward in time (4). In the last step we update λ_0 .

Algorithm: (DAS) Dual Ascent Scheme for (2.10)-(2.13) and (2.16)-(2.17)

Solve iteratively 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 $\epsilon \partial E(u^{k+1}) + \lambda^{k+1}(t = T) \ni 0$, and furthermore $\mu^{k+1}(t = T) = u^{k+1}$.
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 (1.5) we introduce an algorithm based on an augmented Lagrangian ansatz [29], due to the non-differentiability and nonlinearity of $E(u)$. Inspired by the Bregman split approach [30] we modify (2.10)-(2.15) 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 [8]):

$$\begin{aligned} \inf_{\mu, u, z} \sup_{\xi, \lambda, \lambda_0} L(\mu, u, z, \xi, \lambda, \lambda_0) &= -\frac{1}{2} \int_0^T \int_{\Omega} |\nabla \lambda|^2 d\mu(x) dt + \epsilon \int_{\Omega} |z| dx - \int_0^T \int_{\Omega} \partial_t \lambda d\mu(x) dt \\ &\quad - \int_{\Omega} \nu \lambda_0 dx + \int_{\Omega} \lambda_0 d\mu(t=0) - \int_{\Omega} \lambda(t=T) d\mu(t=T) \\ &\quad + \int_{\Omega} u \lambda(t=T) dx + \int_{\Omega} (z - \nabla u) \xi dx + \frac{r}{2} \int_{\Omega} |z - \nabla u|^2 dx, \end{aligned}$$

with parameter $r \in \mathbb{R}$. The optimality conditions for the above problem are

$$\partial_t \lambda + \frac{1}{2} |\nabla \lambda|^2 = 0, \quad (4.21)$$

$$\partial_t \mu + \nabla \cdot (\mu \nabla \lambda) = 0, \quad (4.22)$$

$$z^{k+1} = \arg \min_z \left\{ \epsilon \int_{\Omega} |z| dx + \int_{\Omega} \xi^k z dx + \frac{r}{2} \int_{\Omega} |z - \nabla u^{k+1}|^2 dx \right\}, \quad (4.23)$$

$$= \arg \min_z \left\{ \epsilon \int_{\Omega} |rz| dx + \frac{1}{2} \int_{\Omega} |rz - (r \nabla u^{k+1} - \xi^{k+1})|^2 dx \right\} \quad (4.24)$$

$$\xi^{k+1} = \arg \max_{\xi} \left\{ \int_{\Omega} (z^{k+1} - \nabla u^{k+1}) \xi dx \right\}. \quad (4.25)$$

The solution for (4.23) with $w = \nabla u - \frac{\xi}{r}$ is explicitly given by the shrinkage operation

$$z = \begin{cases} \left(1 - \frac{\epsilon}{r|w(x,y)|}\right) (w(x,y)), & |w(x,y)| > 1, \\ 0, & |w(x,y)| \leq 1. \end{cases} \quad (4.26)$$

Finally we can write down the following iterative algorithm:

Algorithm: (ALS) Augmented Lagrangian Scheme

Solve iteratively 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 $\Delta u^{k+1} = (\frac{1}{r} \lambda^{k+1}(t=T) + \frac{1}{r} \nabla \cdot \xi^k + \nabla \cdot z^k)$, $u^{k+1} = \mu^{k+1}(t=T)$.
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 ($\tau_1, \tau_2 > 0$): $\lambda_0^{k+1} = \lambda_0^k + \tau_1 (\mu^{k+1}(t=0) - \nu)$,

$$z^{k+1} = (4.26),$$

$$\xi^{k+1} = \xi^k + \tau_2 (z^{k+1} - \nabla u^{k+1}).$$

To discretize the equations which appear in our algorithms we consider the upwind-method, using the difference quotients defined in Section 4.1. In one dimension the discretization of the state-equation $\partial_t \mu = -\nabla \cdot (\mu v) = -\nabla \mu v - \mu \nabla \cdot v$ is given by:

$$(D_t^+ \mu)_i^l = -(\max(v_i^l, 0)(D_x^- \mu)_i^l + \min(v_i^l, 0)(D_x^+ \mu)_i^l) - \mu_i^l (D_x^+ v)_i^l. \quad (4.27)$$

We use the explicit Euler method to approximate the time-derivative and an upwind scheme for the derivatives in space. In a similar way we discretize the adjoint-equation

$$(D_t^+ \lambda)_i^l = -\frac{1}{2} |v_i^l|^2 + (\max(v_i^l, 0)(D_x^+ \lambda)_i^l + \min(v_i^l, 0)(D_x^- \lambda)_i^l). \quad (4.28)$$

In two dimensions the discretization of the state-equation using an upwind method is given by:

$$\begin{aligned} (D_t^+ \mu)_{i,j}^l &= - \left(\max((v_1)_{i,j}^l, 0)(D_x^- \mu)_{i,j}^l + \min((v_1)_{i,j}^l, 0)(D_x^+ \mu)_{i,j}^l \right) \\ &\quad - \left(\max((v_2)_{i,j}^l, 0)(D_y^- \mu)_{i,j}^l + \min((v_2)_{i,j}^l, 0)(D_y^+ \mu)_{i,j}^l \right) \\ &\quad - \mu_{i,j}^l \left((D_x^+ v_1)_{i,j}^l + (D_y^+ v_2)_{i,j}^l \right), \end{aligned}$$

where v_1 denotes the velocity in x -direction and v_2 in y -direction. In a similar way we discretize the adjoint-equation.

We start our numerical simulations by considering the Gaussian curve with mean zero and variance one as initial data ν . For the logarithmic entropy and the Fisher information we use the gradient descent scheme (GDS). Note that computing the Fisher information means in general dealing with numerical problems, which can be avoided by shifting the initial data to 0.1. The squared Lebesgue regularisation and the Dirichlet regularisation are computed with the dual ascent scheme (DAS) and the TV regularisation with the augmented Lagrangian scheme (ALS). Numerical results in one dimension are illustrated in Figure 4 for the energies (1.5)-(1.9). Two dimensional results for (2.6)-(2.9) are presented in Figure 5. In particular, the results for the TV regularisation in two dimensions are presented in Figure 5(d)-(i), where we differ between the anisotropic TV regularisation $E(u) = \int |\nabla_x u| + |\nabla_y u|$ and the isotropic TV regularisation $E(u) = \int \sqrt{|\nabla_x u|^2 + |\nabla_y u|^2}$. Isotropic diffusion behaves perpendicular to edges, i.e. smoothing at edges is inhibited. Instead, diffusion along edges is achieved using the anisotropic TV regularisation.

Next we consider real data. In Figure 6 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 is provided by the Istituto Nazionale di Geofisica e Vulcanologia ¹. The values are zero in distributional regions, without aftershock events and otherwise one. The smoothed result computed with the Dirichlet regularisation with $\epsilon = 10$ is illustrated in Figure 6(b). The letter constitutes an estimate of the continuous distribution of aftershocks based on the few measured samples in Figure 6(a).

Additionally we consider real MRI - data provided from the MPI-Göttingen (see Figure 7 a). For comparison, we did numerical simulations on a 150×150 -detail of the scan using the Dirichlet regularisation and the TV-regularisation (see Figure 7(c)-(h)) with different regularisation parameters.

To confirm that the investigated functionals can also be used for image decomposition we present in Figure 8 some numerical experiments. Note here that we do not want to compare our results with existing numerical methods for image decomposition, but we want to mention, that in our case 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, where texture parts are combined with nontexture parts. In the

¹INGV: <http://bollettinosismico.rm.ingv.it>

middle row of Figure 8 we show the structure or non-texture part of the initial images, which is the minimizer of (2.6)-(2.9) with TV-regularisation solved with the augmented Lagrangian scheme. The texture images, presented in the last row follow from subtracting the structure image from the initial image.

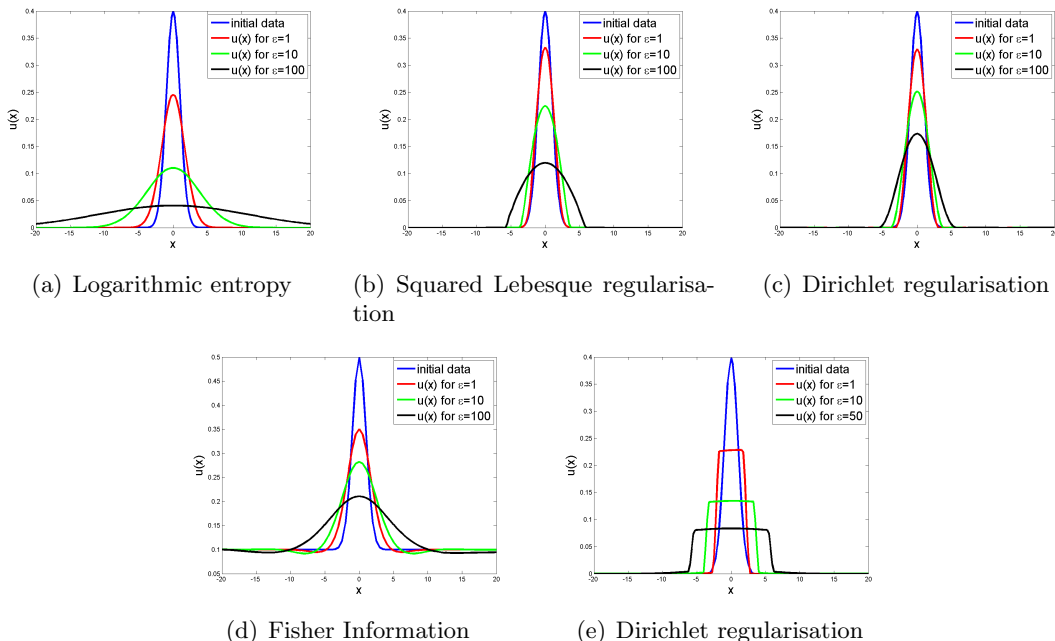


Figure 4: Fluid dynamic scheme (2.6)-(2.9) 1D: (a) Logarithmic entropy (GDS) $\Delta x = 0.25, \Delta t = 0.003, \tau = 1000$; (b) Squared Lebesgue regularisation (DAS) $\Delta x = 0.25, \Delta t = 0.3, \tau = 0.002$; (c) Dirichlet regularisation (DAS) $\Delta x = 0.25, \Delta t = 0.3, \tau = 0.002$; (d) Fisher information (GDS) $\Delta x = 0.5, \Delta t = 0.003, \tau = 1000$; (e) TV regularisation (ALS), $\Delta x = 0.25, \Delta t = 0.3, \tau = 0.1$.

4.5 Discussion

In this section we presented several numerical results in one dimension as well as in two dimension for the Kantorovich formulation, the Benamou Brenier formulation and the gradient flow formulation. The main aspect was to introduce the algorithms, discuss the regularisation energies (1.5)-(1.9), illustrate their different impact on the data and to compare the results. Because the different formulations are equivalent to each other, we obtain quite similar solutions for the same initial data and regularisation parameter, which also confirms the correctness of our models. For example, the logarithmic entropy causes isotropic smoothing, similar to a Gaussian filter. This kind of uniform melting for increasing ϵ is illustrated in Figure 2(a), 3(a) and 4(a). Here we recognize, that the results are similar to each other for the same ϵ and furthermore coincide with the self-similar solution (3.6). Computing the optimisation problem with the squared Lebesgue regularisation yields a solution, which has compact support (see Figure 2(b), Figure 3(b) and Figure 4(b)). Note here, that solutions for the porous medium equation are the well known Barenblatt-Pattle solutions.

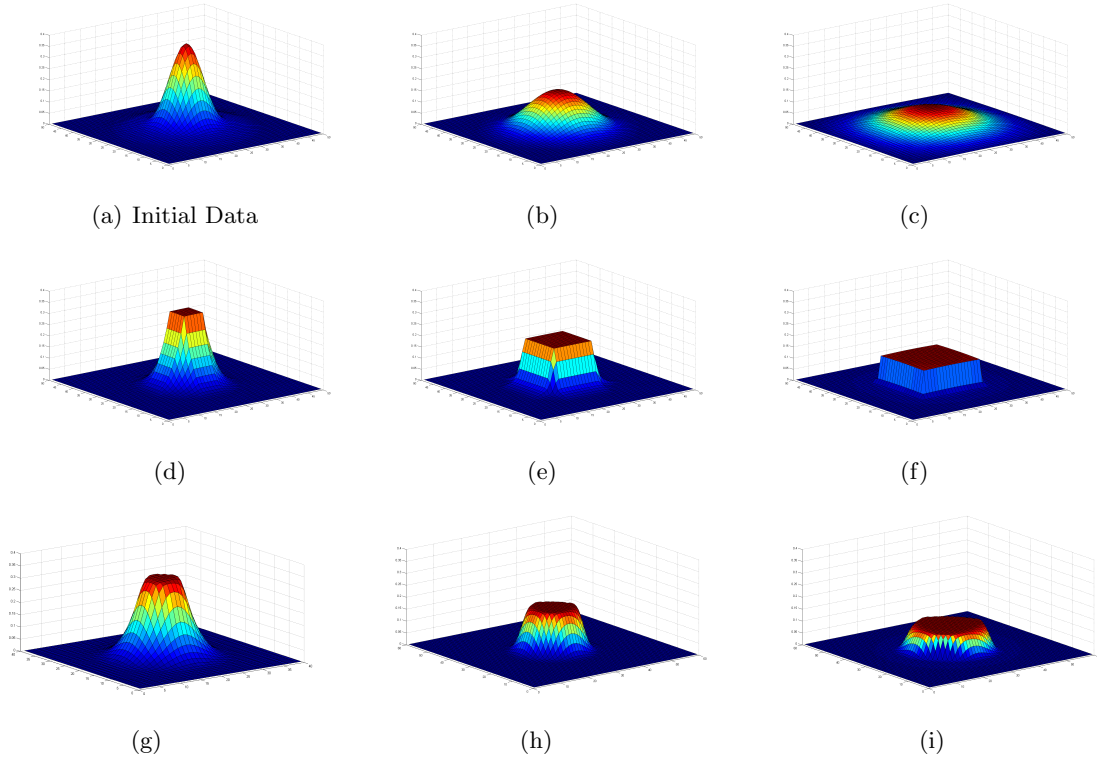


Figure 5: Two dimensional Benamou-Brenier formulation (2.6)-(2.9): (a) Initial data; (b)-(c) Dirichlet regularisation (DAS) $\epsilon = 5, 50$; (d)-(f) Anisotrop TV regularisation (ALS), $\epsilon = 1, 10, 50$; (g)-(i) Isotrop TV regularisation (ALS) , $\epsilon = 1, 10, 50$.

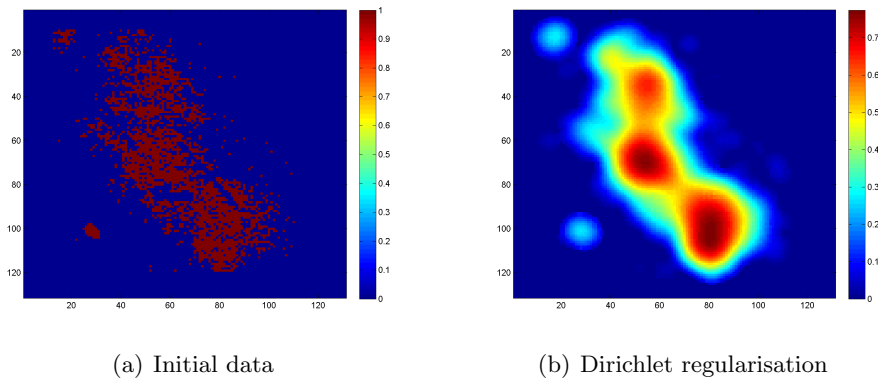


Figure 6: Density estimates for earthquake data; (a) Initial data: Epicentral distribution of earthquakes in the time interval of aftershock sequence from 6 April 2009 to 30 June 2009. (source INGV: <http://bollettinosismico.rm.ingv.it>); (b)Results with the Dirichlet regularisation (GDS) for $\epsilon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

In Figure 6 we presented numerical simulations with real initial data given by earthquake

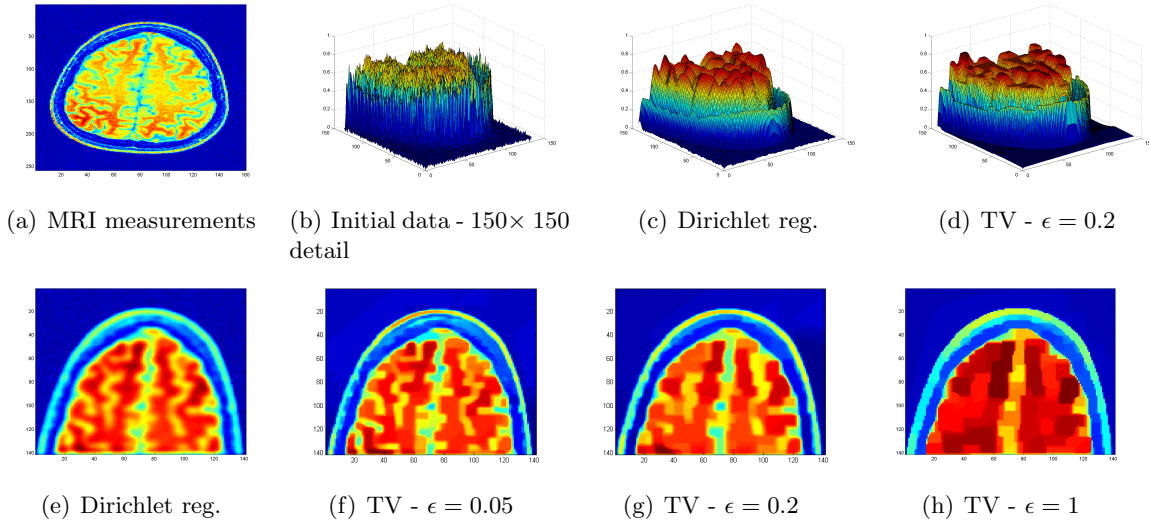


Figure 7: (a) MRI-measurements provided by the MPI for biophysical chemistry Göttingen; (b) Initial data, 150×150 detail; (c) Results with Dirichlet regularisation (GDS); (d) Results with the TV-regularisation (ALS) for $\epsilon = 0.2$; (e) Dirichlet regularisation (GDS); The results for the isotrop TV-regularisation (ALS) for $\epsilon = 0.05, 0.2, 1$ are shown in (f)-(h).

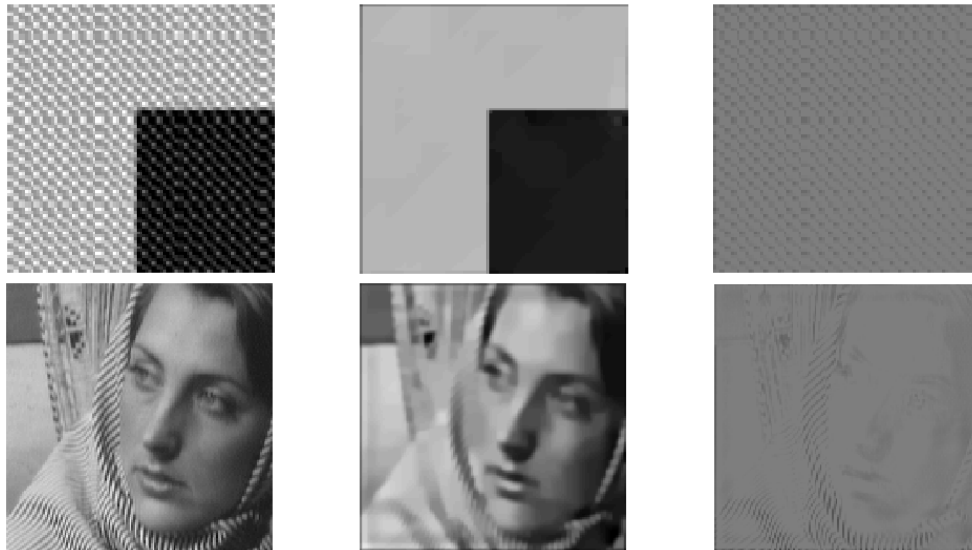


Figure 8: Image decomposition using the TV-regularisation (ALS); From left to right: Initial image, cartoon image , texture image.

measurements. Applying the gradient descent with Dirichlet regularisation we obtain a relative probability for an earthquake event in a certain region (see Figure 6(b)). For this example the probability for an earthquake event in the red colored region is greater than in the blue colored region.

Furthermore we present numerical results for the TV regularisation with the augmented La-

grangian scheme (ALS) in one dimension (see Figure 2(e), 4(e)) and in two dimension (see Figure 5 (d)-(i)). The difference between the isotropic and anisotropic TV-regularisation is pointed out in Figure 5. The isotropic TV-regularisation favors smoothing along edges (see (g)-(i)). In contrast, using the anisotropic TV-regularisation yields to smoothing along edges and to the result shown in Figure 5(d)-(f). The edge preserving property of the TV-regularisation becomes more clear in Figure 7 (f)-(h). By increasing the regularisation parameter, the number and intensity of edges increases too. For $\epsilon = 1$ the result is oversmoothed and consists of many homogeneous regions. Details are not recognizable anymore (see (h)). For comparison we show in Figure 7 (e) results obtained with the Dirichlet regularisation (GDS). Additionally a nice property of the TV-regularisation is the image decomposition feature presented in Figure 8. In the first row a synthetic image with high presence of texture is satisfactory decomposed into a complete non-texture part, the cartoon image, and a texture part. The edges are kept in the cartoon image and the texture is separated (last row). Moreover, the decomposition of the Barbara-image yields a texture image, where not only the texture of the scarf is kept, but also small details like the eyebrows.

5 Conclusions

In this paper, we investigate a variational regularisation method with the Wasserstein metric as a data fidelity for estimation, smoothing and decomposition of densities. We provide different formulations for this problem such as the Benamou-Brenier formulation and the gradient flow formulation. We gave a detailed analysis of this approaches and presented schemes for their numerical solution. Finally we present numerical simulations on synthetic and real data for different regularisation energies, namely the TV regularisation, the Dirichlet regularisation, the squared Lebesgue regularisation, the logarithmic entropy and the Fisher information. With our models we are able to work with densities continuous with respect to the Lebesgue measure as well as concentrated densities. Moreover we have the advantage of mass conservation.

6 Acknowledgment

The work of MB has been supported by the German Science Foundation (DFG) through project *Regularization with Singular Energies*. CBS acknowledges the financial support provided by the Cambridge Centre for Analysis (CCA), the DFG Graduiertenkolleg 1023 *Identification in Mathematical Models: Synergy of Stochastic and Numerical Methods* and the project WWTF Five senses-Call 2006, *Mathematical Methods for Image Analysis and Processing in the Visual Arts*. Further, this publication is based on work supported by Award No. KUK-I1-007-43 , made by King Abdullah University of Science and Technology (KAUST).

References

- [1] L. Ambrosio, *Transport equation and cauchy problem for non-smooth vector fields*, Calculus of Variations and Nonlinear Partial Differential Equations, Lecture Notes in Mathematics, 2008, Volume 1927/2008, 1-41.

- [2] L.Ambrosio, L.A.Caffarelli, Y.Brenier, *Optimal Transportation and Applications* Lecture Notes in Mathematics, Springer-Verlag Berlin Heidelberg, 2003.
- [3] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity problems*, Oxford Mathematical Monographs. Oxford: Clarendon Press. xviii, 2000.
- [4] L.Ambrosio, N.Gigli, G.Savare, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics ETH Zrich, Birkhäuser Verlag, Basel, 2005.
- [5] J.D.Benamou, Y.Brenier, *A computational fluid mechanics solution to the Monge- Kantorovich mass transfer problem*, Numer. Math. 84 (2000), 375-393.
- [6] M.Benning, M.Burger, Error estimates for general fidelities, Electronic transactions on numerical analysis, to appear 2011.
- [7] A. Blanchet, V. Calvez, J. A. Carrillo, *Convergence of the mass-transport steepest descent scheme for the sub-critical Patlak-Keller-Segel model*, SIAM J. Numer. Anal. 46, pp. 691–721, 2008.
- [8] C. Brune, *4D Imaging in Tomography and Optical Nanoscopy*, PhD Thesis, University of Münster, July 2010.
- [9] M. Burger, J. A. Carrillo, M.-T. Wolfram, *A Mixed Finite Element Method for Nonlinear Diffusion Equations*, Kinetic and Related Models 3, 59–83, 2010.
- [10] M.Burger, S.Osher, Convergence rates of convex variational regularisation, Inverse Problems 20 (2004), 1411-1421.
- [11] M.Burger, E.Resmerita, L.He, Error estimation for Bregman iterations and inverse scale space methods in image restoration, Computing 81 (2007), 109-135
- [12] G. Buttazzo and F. Santambrogio, *A Model for the Optimal Planning of an Urban Area*, SIAM J. Math. Anal., Vol. 37, Nr. 2, pp. 514–530, 2005.
- [13] J.A. Carrillo, R.J. McCann, C. Villani, *Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates*, Revista Matematica Iberoamericana 19, pp. 1-48, 2003.
- [14] J.A. Carrillo, R.J. McCann, C. Villani, *Contractions in the 2-Wasserstein length space and thermalization of granular media*, Archive for Rational Mechanics and Analysis 179, pp. 217-263, 2006.
- [15] J. A. Carrillo, J. S. Moll, *Numerical simulation of diffusive and aggregation phenomena in nonlinear continuity equations by evolving diffeomorphisms*, SIAM J. Sci. Comput. 31, pp. 4305-4329, 2009.
- [16] J. A. Carrillo, and D. Slepcev, *Example of a displacement convex functional of first order*, Calculus of Variations and PDES 36, pp. 547-564, 2009.
- [17] J.A. Carrillo, G. Toscani, *Wasserstein metric and large-time asymptotics of nonlinear diffusion equations*, New Trends in Mathematical Physics, (In Honour of the Salvatore Rionero 70th Birthday), pp. 234-244, 2005.

- [18] P. L. Davies, A. Kovac, *Densities, spectral densities and modality*. Annals of Statistics 32, pp. 1093–1136, 2004.
- [19] P. L. Davies, A. Kovac, and M. Meise, *Nonparametric regression, confidence regions and regularisation*. Annals of Statistics, Annals of Statistics 37, pp. 2597–2625, 2009.
- [20] J. Denzler, R. J. McCann, *Nonlinear diffusion from a delocalized source: affine self-similarity, time reversal, & nonradial focusing geometries*, Ann. Inst. H. Poincaré Anal. Non Linéaire. 25, pp.865888, 2008.
- [21] B. Düring, D. Matthes, and J.-P. Milisic, *A gradient flow scheme for nonlinear fourth order equations*, Discrete Contin. Dyn. Syst. Ser. B 14(3), 2010, pp.935-959.
- [22] P. P. B. Eggermont and V. N. LaRiccia, *Maximum penalized likelihood estimation*. Springer, 2001.
- [23] J. J. Egozcue , A. Barbat, J. A. Canas, J. Miquel and E. Banda, *A method to estimate intensity occurrence probabilities in low seismic activity regions*. Earthquake Engineering & Structural Dynamics 20 (1), pp. 43–60, 2006.
- [24] C. L. Epstein, *Introduction to the Mathematics of Medical Imaging*, SIAM, Second Edition, 2008.
- [25] L.C. Evans, W. Gangbo, O. Savin, *Diffeomorphisms and nonlinear heat flows*, SIAM J. Math. Anal. 37, no. 3, pp. 737–751, 2005.
- [26] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions.*, CRC Press, 1992.
- [27] L. Giacomelli, F. Otto, *Variational formulation for the lubrication approximation of the Hele-Shaw flow*, Calc. Var. Partial Differential Equations, 13, no. 3, 377403, 2001.
- [28] U. Gianazza, G. Savaré, G. Toscani, *The Wasserstein gradient flow of the Fisher information*, Arch. Ration. Mech. Anal. 194, 133–220, 2009.
- [29] R. Glowinski, P. Le Tallec, *Augmented Lagrangians and operator-splitting methods in nonlinear mechanics.*,SIAM, Philadelphia, pp. 224–231, 1989.
- [30] T. Goldstein and S. Osher, *The Split Bregman Method for L1 Regularized Problems*. SIAM Journal on Imaging Sciences 2(2), pp. 323–343, 2009.
- [31] I. J. Good and R. A. Gaskins, *Nonparametric roughness penalties for probability densities*. Biometrika 58 (2), pp. 255–277, 1971.
- [32] L. Gosse, G. Toscani, *Lagrangian numerical approximations to one-dimensional convolution-diffusion equations*, SIAM J. Sci. Comput., 28 (4) (2006) pp. 1203-1227
- [33] J.Horowitz, R.L.Karandikar, *Mean rates of convergence of empirical measures in the Wasserstein metric*, J. Comp. Appl. Math. 55 (1994), 261-273.
- [34] R. Jordan, D. Kinderlehrer and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal 29, pp. 1–17, 1999.

- [35] L. V. Kantorovich, *On a problem of Monge*. Uspekhi Mat. Nauk. 3, 225-226, (1948).
- [36] R. Koenker and I. Mizera, *Density estimation by total variation regularisation*. Advances in Statistical Modeling and Inference, Essays in Honor of Kjell A. Doksum, World Scientific, pp. 613–634, 2007.
- [37] C. Kooperberg and C. J. Stone, *Logspline density estimation with free knots*. Computational Statistics and Data Analysis 12, pp. 327–347, 2002.
- [38] J.-M. Lasry, and P.-L. Lions, *Mean field games*, Japanese Journal of Mathematics, 2(1), Mar. 2007.
- [39] D. Matthes, R. J. McCann, G. Savaré, *A Family of Nonlinear Fourth Order Equations of Gradient Flow Type*, Communications in Partial Differential Equations 34, pp. 1352–1397, 2009.
- [40] R.J.McCann, *A convexity principle for interacting gases*, Adv. Math. 128, 153-179, 1997
- [41] G. Mohler, A. Bertozzi, T. Goldstein and S. Osher, *Fast TV regularisation for 2D Maximum Penalized Likelihood Estimation*, accepted in J. Comp. Graph. Stat. 2010
- [42] G. Mohler and M. Short, *Geographic profiling from kinetic models of criminal behavior*, submitted.
- [43] A. Obereder , O. Scherzer and A. Kovac, *Bivariate density estimation using BV regularisation*. Computational Statistics and Data Analysis. 51., pp. 5622–5634, 2007.
- [44] Y. Ogata, *Space-time point process models for earthquake occurrences*, Ann. Inst. Statist. Math. 50 (2), pp. 379-402, 1998.
- [45] S. Osher, M. Burger, D. Goldfarb, J. Xu and W. Yin, *An iterative regularisation method for total variation-based image restoration*. Simul. 4, pp. 460–489, 2005.
- [46] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26(1-2), pp. 101-174, 2001.
- [47] C.Pöschl, *Tikhonov regularisation with General Residual Term* (PhD-Thesis, University Innsbruck, 2008).
- [48] S.T.Rachev, *Probability Metrics and the Stability of Stochastic Models* (Wiley, New York, 1991).
- [49] L.I.Rudin, S.J.Osher, E.Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D **60** (1992), 259-268.
- [50] F. Santambrogio, *Transport and concentration problems with interaction effects*, J. Global Optim. 38, no. 1, pp. 129141, (2007).
- [51] F. Santambrogio, *Variational problems in transport theory with mass concentration*, PhD Thesis, Edizioni della Normale, Birkhäuser, 2007.
- [52] S. Sardy and P. Tseng, *Density estimation by total variation penalized likelihood driven by the sparsity ℓ_1 information criterion*, to appear in Scandinavian Journal of Statistics, 2010.

- [53] F. P. Schoenberg, C. Chang, J. Keeley, J. Pompa, J. Woods and H. Xu, *A critical assessment of the burning index in Los Angeles county, California*. International Journal of Wildland Fire 16(4), pp. 473–483, 2007.
- [54] B. W. Silverman, *Kernel density estimation using the fast Fourier transform*. Statistical Algorithm, AS176, Appl. Statistics 31, pp. 93–97, 1982.
- [55] B. W. Silverman, *Density estimation for statistics and data analysis*, Chapman and Hall, London, 1986.
- [56] L. Smith, M. Keegan, T. Wittman, G. Mohler and A Bertozzi, *Improving Density Estimation By Incorporating Spatial Information*, EURASIP J. on Advances in Signal Processing, Article ID 265631, doi:10.1155/2010/265631, Volume 2010, 12 pages, special issue on Advanced Image Processing for Defense and Security Applications, 2010.
- [57] K.-T. Sturm, *Convex functionals of probability measures and nonlinear diffusions on manifolds*, J. Math. Pures Appl. 84, 149168, 2005.
- [58] K.-T. Sturm, and M.-K von Renesse, *Transport inequalities, gradient estimates, entropy and Ricci curvature*, Comm. Pure Appl. Math. 58, 923-940, 2005.
- [59] J. L. Vázquez, *The Porous Medium Equation*, Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, Oxford, 2007. Mathematical theory.
- [60] C. Villani, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.
- [61] C. Villani, *Optimal transport. Old and new*, Grundlehren der Mathematischen Wissenschaften, 338. Springer-Verlag, Berlin, 2009.
- [62] A. Wächter, L.T. Biegler, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming.*, Mathematical Programming, 106(1),2006, S.25-27.