

A Primal-Dual Approach for a Total Variation Wasserstein Flow

Martin Benning¹, Luca Calatroni², Bertram Düring³,
and Carola-Bibiane Schönlieb⁴

¹ Magnetic Resonance Research Centre, University of Cambridge, UK

² Cambridge Centre for Analysis, University of Cambridge, UK

³ Department of Mathematics, University of Sussex, UK

⁴ Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, UK

Abstract. We consider a nonlinear fourth-order diffusion equation that arises in denoising of image densities. We propose an implicit time-stepping scheme that employs a primal-dual method for computing the subgradient of the total variation seminorm. The constraint on the dual variable is relaxed by adding a *penalty term*, depending on a parameter that determines the weight of the penalisation. The paper is furnished with some numerical examples showing the denoising properties of the model considered.

1 Introduction

For an open and bounded domain $\Omega \subset \mathbb{R}^2$ with Lipschitz boundary, we consider the following equation for $u = u(t, x)$

$$\begin{aligned} u_t &= \nabla \cdot (u \nabla q), \quad q \in \partial \mathcal{E}(u), \quad \text{in } \Omega \times (0, T), \\ u(0, x) &= u_0(x) \geq 0 \quad \text{in } \Omega, \end{aligned} \tag{1.1}$$

with normalised mass $\int_{\Omega} u_0 \, dx = 1$ and where the total variation (TV) functional \mathcal{E} (see [2,20]) is defined by

$$\mathcal{E}(u) := |Du|(\Omega) = \sup_{\mathbf{g} \in C_0^\infty(\Omega; \mathbb{R}^d), \|\mathbf{g}\|_\infty \leq 1} \int_{\Omega} u \nabla \cdot \mathbf{g} \, dx, \tag{1.2}$$

$d = 1, 2$. Equation (1.1) can be formally derived as the L^2 -Wasserstein gradient flow of the TV functional \mathcal{E} in (1.2) and constitutes a nonlinear fourth-order diffusion equation. In this paper we study this equation as a regularising procedure for u_0 being a noisy image. Motivated by previous contributions in higher-order regularisation (see [8,6,16,24,25]), this approach promises to maintain the desirable properties of TV regularisation, such as preservation of edges in the image, while at the same time reducing well-known artifacts of TV such as *stair-casing*.

Originally, (1.1) has been proposed in [11] for density estimation and smoothing. Therein, the authors propose to compute a smoothed version u of a given probability density u_0 as a minimiser of

$$\frac{1}{2}W_2(u_0\mathcal{L}^d, u\mathcal{L}^d)^2 + \alpha\mathcal{E}(u). \quad (1.3)$$

Here, $W_2(u_0\mathcal{L}^d, u\mathcal{L}^d)$ is the L^2 -Wasserstein distance between $u_0\mathcal{L}^d$ and $u\mathcal{L}^d$ (\mathcal{L}^d denotes the Lebesgue measure in \mathbb{R}^d , $d = 1, 2$) and defines a distance within the space of probability measures [3,27,28,1,22]. This minimisation problem can be interpreted as a discrete approximation of a solution of the gradient flow (1.1) of $\mathcal{E}(u)$ with respect to the L^2 -Wasserstein metric. More precisely, the minimisation of (1.3) represents one timestep of De Giorgi's minimising movement scheme (see, e.g. [3,21]) to the functional $\mathcal{E}(u)$ with timestep α . By construction the regularisation method (1.3) proposed in [11] is non-smooth, i.e., edge preserving, and conserves mass, i.e., is density preserving. In [11] the numerical solution of (1.3) has been accomplished by a combination of the Benamou-Brenier formulation [4] for the Wasserstein distance, an augmented Lagrangian method, and an operator splitting technique [9]. This numerical procedure is in the flavour of several recently proposed numerical schemes for equations with gradient flow structure, cf., e.g., [13,7,18,10] and references therein.

Equation (1.1) has been further investigated in [19], where the authors numerically study the scale space properties and high-contrasting effects of the equation by solving it with a dimensional *alternating direction implicit* (ADI) operator splitting approach. There, the subgradient q of the TV seminorm in (1.1) is approximated by an ϵ -regularisation of the form

$$q \approx \nabla \cdot \left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \epsilon}} \right), \quad 1 \gg \epsilon > 0. \quad (1.4)$$

From a computational point of view, finding numerical schemes that solve higher-order equations like (1.1) is a challenging problem. Dealing with an evolutionary nonlinear fourth-order partial differential equation, we aim to find an efficient and reliable method avoiding a naive explicit discretisation in time that might present time step size restrictions (compare [26]) and, because of the strong nonlinearity of the subgradients of TV, additionally add constraints to the stability condition of the discrete time stepping scheme, compare [15,17,12].

In this paper we propose a formulation of (1.1) which characterises the elements of the subdifferential of TV in an alternative way. Instead of considering a characterisation of the type (1.4) for these elements, we use the approach proposed in [5,23] and deal with a relaxed primal-dual formulation of (1.1). In [12] the authors consider such a kind of approach applied to a similar fourth-order PDE as well as a directional splitting strategy that has proposed there as a direction for future research.

Notation. We denote by u the solution of the continuous equations and by U the solution of the time discrete numerical schemes we are going to present. We

write U_n to indicate the approximation of $u(n\Delta t, \cdot)$, $n \geq 1$, where Δt is the time step size. We will typically consider a rectangular domain $\Omega = [a, b] \times [c, d]$ and approximate it by a finite grid $\{a = x_1 < \dots < x_N = b\} \times \{c = y_1 < \dots < y_M = d\}$ with equidistant step-size $h = (b - a)/N = (d - c)/M$. For vectors, we will indicate their components using the superscripts notation: $\mathbf{y} = (y^1 \ y^2)^\top$. With a little abuse of notation, we will still use the notations ∇ and $\nabla \cdot$ to indicate the differential operators discretised with either forward or backward finite differences applied to the discretised quantities (see Section 3).

2 Primal-Dual Formulation of the TV-Wasserstein Flow

We aim to characterise the elements in the subdifferential of the TV seminorm (1.2) by primal-dual iterations, as suggested, for instance, in [5,23] when dealing with the classical second-order TV-denoising model. In what follows we discuss such a strategy combined with an implicit time stepping method for solving equation (1.1). By definition of the subdifferential, the property $q \in \partial|Du|(\Omega)$ means:

$$q \in \partial|Du|(\Omega) \iff |Du|(\Omega) - \int_{\Omega} qu \, dx \leq |Dv|(\Omega) - \int_{\Omega} qv \, dx, \quad \forall v \in L^2(\Omega). \tag{2.1}$$

Equivalently, if $u \in BV(\Omega) \subset L^2(\Omega)$ achieves the minimum of the following variational problem

$$\min_{u \in BV(\Omega)} \left\{ |Du|(\Omega) - \int_{\Omega} qu \, dx \right\}, \tag{2.2}$$

then, by definition of being minimum, (2.1) is fulfilled and then $q \in \partial|Du|(\Omega)$. Inserting the definition of the total variation seminorm (1.2) into (2.2) we receive

$$\min_{u \in BV(\Omega)} \left\{ \sup_{\mathbf{p} \in C_0^\infty(\Omega; \mathbb{R}^2), \|\mathbf{p}\|_\infty \leq 1} \int_{\Omega} u \nabla \cdot \mathbf{p} \, dx - \int_{\Omega} qu \, dx \right\} \tag{2.3}$$

which is typically known as the *primal-dual* formulation of the problem (1.1). The constraint on \mathbf{p} appearing in (2.3) can be relaxed, for instance, by a penalty method. To this end we remove the constraint from the minimisation in (2.3) and instead add a term to the functional that penalises it if $\|\mathbf{p}\|_\infty > 1$. A typical example for such a penalty term F is

$$F(s) = \frac{1}{2} \|\max\{s, 0\}\|_2^2.$$

With these considerations we reformulate (2.3) into the following minimisation problem

$$\min_{u \in BV(\Omega)} \sup_{\mathbf{p} \in C_0^\infty(\Omega; \mathbb{R}^2)} \left\{ \int_{\Omega} u \nabla \cdot \mathbf{p} \, dx - \frac{1}{\varepsilon} F(|\mathbf{p}| - 1) - \int_{\Omega} qu \, dx \right\}, \tag{2.4}$$

where the parameter $1 \gg \varepsilon > 0$ is small and measures the weight of our penalisation. We can then find the optimality conditions for both \mathbf{p} and u in (2.4) which, merged with the original equation (1.1), allow us to consider the following, alternative formulation of the TV-Wasserstein model:

$$\begin{cases} u_t = \nabla \cdot (u \nabla q), \\ q = \nabla \cdot \mathbf{p}, \\ 0 = -\nabla u - \frac{1}{\varepsilon} H(\mathbf{p}). \end{cases} \tag{2.5}$$

In the system above H denotes the derivative of the penalty term $F(|\mathbf{p}| - 1)$, i.e.

$$H(\mathbf{p}) = \mathbb{1}_{\{|\mathbf{p}| \geq 1\}} \operatorname{sgn}(\mathbf{p})(|\mathbf{p}| - 1),$$

which we linearise via its first-order Taylor approximation

$$H(\mathbf{p}) \approx H(\tilde{\mathbf{p}}) + H'(\tilde{\mathbf{p}})(\mathbf{p} - \tilde{\mathbf{p}}),$$

where with H' we indicate the Jacobian of H . In order to guarantee the invertibility of the now linear operator that defines the system, we add an additional damping term in \mathbf{p} , as suggested, for instance, in [23]. Collecting everything, we propose the following numerical scheme for solving (2.5),

$$\begin{cases} \frac{U_{n+1}^{(k)} - U_n}{\Delta t} = \nabla \cdot (U_n \nabla Q_{n+1}^{(k)}) \\ Q_{n+1}^{(k)} = \nabla \cdot \mathbf{P}_{n+1}^{(k)}, \\ 0 = -\nabla U_{n+1}^{(k)} - \frac{1}{\varepsilon} H(\mathbf{P}_{n+1}^{(k-1)}) \\ \quad - \frac{1}{\varepsilon} H'(\mathbf{P}_{n+1}^{(k-1)})(\mathbf{P}_{n+1}^{(k)} - \mathbf{P}_{n+1}^{(k-1)}) - \tau^k (\mathbf{P}_{n+1}^{(k)} - \mathbf{P}_{n+1}^{(k-1)}). \end{cases} \tag{2.6}$$

We apply Newton’s method to solve (2.6). The scheme consists of two nested iterations. The subscripts n are related to the outer time step evolution of the process evolving U . At each time step an implicit approximation of the quantities U_{n+1} , Q_{n+1} and \mathbf{P}_{n+1} is obtained by the application of an inner damped Newton process that runs depending on the superscript k . The sequence of parameters τ^k controls the damping of the Newton iterations: it starts from a large value τ^0 and then decreases, thus ensuring faster convergence. System (2.6) could now be discretised in space as described in Section 3. For computational simplicity we consider a slightly different penalty term F (see [23]):

$$F(\mathbf{p}) = F(p^1, p^2) = \frac{1}{2} \|\max\{|p^1| - 1, 0\}\|_2^2 + \frac{1}{2} \|\max\{|p^2| - 1, 0\}\|_2^2,$$

that results into an anisotropic TV term. Whence

$$H(p^1, p^2) = \begin{pmatrix} \operatorname{sgn}(p^1)(|p^1| - 1) \mathbb{1}_{\{|p^1| \geq 1\}} \\ \operatorname{sgn}(p^2)(|p^2| - 1) \mathbb{1}_{\{|p^2| \geq 1\}} \end{pmatrix}, \quad H'(p^1, p^2) = \begin{pmatrix} \mathbb{1}_{\{|p^1| \geq 1\}} & 0 \\ 0 & \mathbb{1}_{\{|p^2| \geq 1\}} \end{pmatrix}.$$

3 Numerical Results

We discretise the gradient and the divergence operators appearing in the system above using standard forward and backward finite differences, thus preserving the adjointness properties of the operators as described in [14] and used also in [5,23]. We then build up the matrices representing system (2.6). In each step of Newton’s method the block-structure of the Jacobian matrix is exploited by inverting it with a Schur complement strategy. For all the following tests we use the following stopping criterion: $\|U_{n+1}^{(k)} - U_{n+1}^{(k-1)}\| / \|U_{n+1}^{(k)}\| \leq \epsilon_{tol}$, where $\|\cdot\|$ is the ℓ^2 -norm and ϵ_{tol} the tolerance.

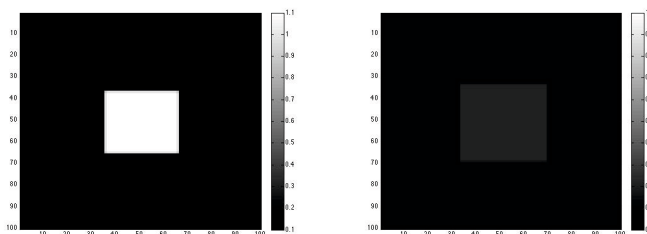


Fig. 1. Initial condition (l.) and solution of the *TV*-Wasserstein gradient flow after 10000 iterations (r.). $\epsilon = 10^{-3}, \tau^0 = 1$.

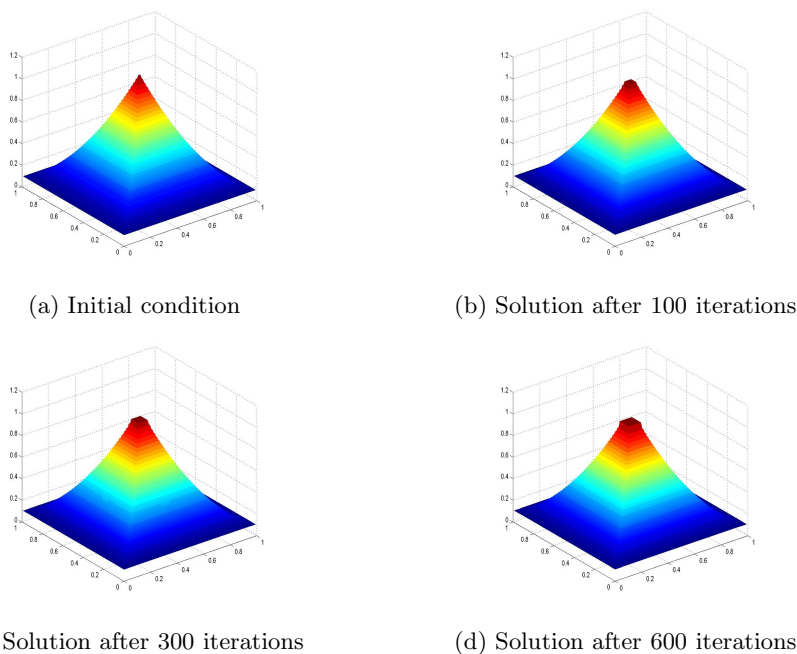


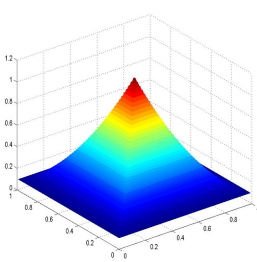
Fig. 2. Time evolution of (2.6) with $\epsilon = 10^{-7}, \tau^0 = 1$

Figure 1 shows the solution of (1.1) computed via (2.6) on a 100×100 pixels initial condition of a square. The result shows the difference between the TV -Wasserstein approach and the pure TV one. The latter, in fact, would just decrease the intensity of the square and increase the intensity of the background without changing its support. Because of the mass preservation properties the solution of (1.1) enlarge their support, as pointed out in [11] where the authors show the self-similarity of the solutions.

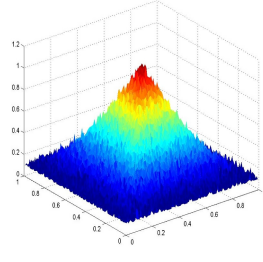
Figure 2 shows the time evolution of the method (2.6) applied to a pyramidal initial condition defined on the domain $\Omega = [0, 1] \times [0, 1]$.

We now apply our method to a denoising problem. Figure 3 shows the pyramid and a noisy version of it obtained by adding Gaussian noise with variance 0.001. The denoised version is obtained both with the primal-dual TV method with penalty term described in [23] and with our method. We observe that while the simple application of the TV model creates staircaising, the use of higher-order models reduces artifacts and preserves structures.

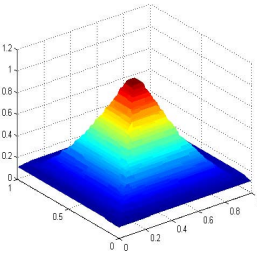
We consider in Figure 4 a real-world image of a LEGO man. The dimension of the image is 200×200 pixels. We add a Gaussian noise with zero mean and variance equal to 0.005 and we show the result of time evolution of the process



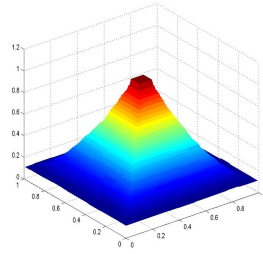
(a) Initial condition



(b) Noisy pyramid with Gaussian noise of variance 0.001



(c) Denoised pyramid with TV primal-dual model



(d) Denoised pyramid with TV -Wasserstein primal-dual model (2.6)

Fig. 3. Denoising with TV and TV -Wasserstein primal-dual methods. $\varepsilon = 10^{-5}$, $\tau^0 = 1$.

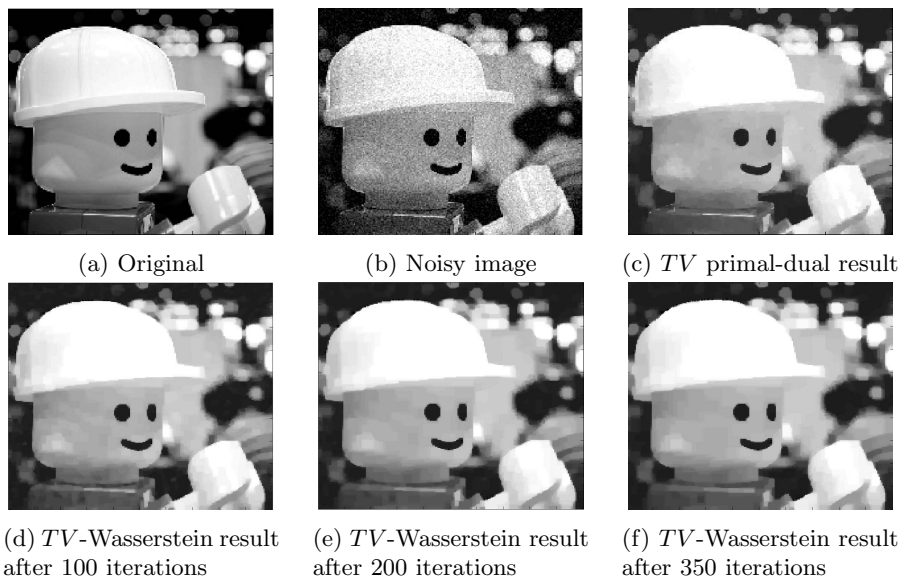


Fig. 4. Comparison with *TV* primal-dual denoising model and *TV*-Wasserstein. $\varepsilon = 10^{-7}$, $\tau^0 = 1$.

after some time iterations. A result with the application of the *TV* primal-dual method is given for comparison as well.

Conclusions. We proposed a numerical method to solve (1.1). Our strategy consists in a relaxation of the characterisation of the subgradients of the total variation via the addition of a *penalty term*. The optimality conditions of the primal-dual formulation form a system of equations that can be solved via a damped Newton's method. The scheme shows good smoothing properties and reduced artifacts in comparison to the pure *TV* method when applied to denoising problems.

Acknowledgements. Carola-Bibiane Schönlieb acknowledges financial support provided by the Cambridge Centre for Analysis (CCA), the Royal Society International Exchanges Award IE110314 for the project *High-order Compressed Sensing for Medical Imaging*, the EPSRC first grant Nr. EP/J009539/1 *Sparse & Higher-order Image Restoration*, and the EPSRC / Isaac Newton Trust Small Grant on *Non-smooth geometric reconstruction for high resolution MRI imaging of fluid transport in bed reactors*. 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. Ambrosio, L., Caffarelli, L.A., Brenier, Y.: *Optimal Transportation and Applications*. Lecture Notes in Mathematics. Springer, Heidelberg (2003)
2. Ambrosio, L., Fusco, N., Pallara, D.: *Functions of Bounded Variation and Free Discontinuity problems*. Oxford Mathematical Monographs, p. xviii. Clarendon Press, Oxford (2000)
3. Ambrosio, L., Gigli, N., Savare, G.: *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel (2005)
4. Benamou, J.D., Brenier, Y.: A computational fluid mechanics solution to the Monge- Kantorovich mass transfer problem. *Numer. Math.* 84, 375–393 (2000)
5. Benning, M.: *Singular regularization of inverse problems*, PhD thesis, University of Münster (2011)
6. Benning, M., Brune, C., Burger, M., Müller, J.: Higher-Order TV Methods Enhancement via Bregman Iteration. *Journal of Scientific Computing* 54, 269–310 (2013)
7. Blanchet, A., Calvez, V., Carrillo, J.A.: Convergence of the mass-transport steepest descent scheme for the sub-critical Patlak-Keller-Segel model. *SIAM J. Numer. Anal.* 46, 691–721 (2008)
8. Bredies, K., Kunisch, K., Pock, T.: *Total Generalized Variation*. *SIAM Journal on Imaging Sciences* 3, 1–42 (2008)
9. Brune, C.: *4D Imaging in Tomography and Optical Nanoscopy*. PhD Thesis, University of Münster (July 2010)
10. Burger, M., Carrillo, J.A., Wolfram, M.-T.: A Mixed Finite Element Method for Nonlinear Diffusion Equations. *Kinetic and Related Models* 3, 59–83 (2010)
11. Burger, M., Franek, M., Schönlieb, C.-B.: Regularised Regression and Density estimation based on Optimal Transport. *Appl. Math. Res. Express (AMRX)* 2012(2), 209–253 (2012)
12. Calatroni, L., Düring, B., Schönlieb, C.-B.: ADI splitting schemes for a fourth-order nonlinear partial differential equation from image processing (2012) (preprint)
13. Carrillo, J.A., Moll, J.S.: Numerical simulation of diffusive and aggregation phenomena in nonlinear continuity equations by evolving diffeomorphisms. *SIAM J. Sci. Comput.* 31, 4305–4329 (2009)
14. Chambolle, A.: An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* 20(1-2), 89–97 (2004)
15. Chan, T.F., Mulet, P.: On the Convergence of the Lagged Diffusivity Fixed Point Method in Total Variation Image Restoration. *SIAM Journal on Numerical Analysis* 36(2), 354–367 (1999)
16. Didas, S., Weickert, J., Burgeth, B.: Properties of higher order nonlinear diffusion filtering. *Journal of mathematical imaging and vision* 35, 208–226 (2009)
17. Didas, S., Weickert, J., Burgeth, B.: Stability and local feature enhancement of higher order nonlinear diffusion filtering. In: Kropatsch, W.G., Sablatnig, R., Hanbury, A. (eds.) *DAGM 2005*. LNCS, vol. 3663, pp. 451–458. Springer, Heidelberg (2005)
18. Düring, B., Matthes, D., Milišić, J.-P.: A gradient flow scheme for nonlinear fourth order equations. *Discrete Contin. Dyn. Syst. Ser. B* 14(3), 935–959 (2010)
19. Düring, B., Schönlieb, C.-B.: A high-contrast fourth-order PDE from imaging: numerical solution by ADI splitting. In: Ammari, H., et al. (eds.) *Multi-scale and High-Contrast Partial Differential Equations*. Contemporary Mathematics, vol. 577, pp. 93–103. American Mathematical Society, Providence (2012)

20. Evans, L.C., Gariepy, R.F.: Measure Theory and Fine Properties of Functions. CRC Press (1992)
21. Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* 29, 1–17 (1999)
22. Kantorovich, L.V.: On a problem of Monge. *Uspekhi Mat. Nauk.* 3, 225–226 (1948)
23. Müller, J.: Parallel total variation minimization, Diploma Thesis, University of Münster, WWU (2008)
24. Papafitsoros, K., Schönlieb, C.-B.: A Combined First and Second Order Variational Approach for Image Reconstruction. *Journal of Mathematical Imaging and Vision* (2013)
25. Setzer, S., Steidl, G.: Variational methods with higher order derivatives in image processing. In: *Approximation XII*, pp. 360–386. Nashboro Press, Brentwood (2008)
26. Smereka, P.: Semi-implicit level set methods for curvature and surface diffusion motion. Special issue in honor of the sixtieth birthday of Stanley Osher. *J. Sci. Comput.* 19(1-3), 439–456 (2003)
27. Villani, C.: Topics in optimal transportation. Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence (2003)
28. Villani, C.: Optimal transport. Old and new, *Grundlehren der Mathematischen Wissenschaften*, vol. 338. Springer, Berlin (2009)