The Total Variation-Wasserstein Problem: A New Derivation of the Euler-Lagrange Equations

Antonin Chambolle[1,2], Vincent Duval[1,2], and João Miguel Machado[1,2](✉)

[1] CEREMADE, Université Paris-Dauphine, PSL University, 75016 Paris, France
joao-miguel.machado@ceremade.dauphine.fr
[2] Inria Paris (Mokaplan), 75012 Paris, France

Abstract. In this work we analyze the Total Variation-Wasserstein minimization problem. We propose an alternative form of deriving optimality conditions from the approach of [8], and as result obtain further regularity for the quantities involved. In the sequel we propose an algorithm to solve this problem alongside two numerical experiments.

Keywords: Total variation · Optimal transport · Image analysis

1 Introduction

The Wasserstein gradient flow of the total variation functional has been studied in a series of recent papers [2,4,8], for applications in image processing. In the present paper, we revisit the work of Carlier & Poon [8] and derive Euler-Lagrange equations for the problem: given $\Omega \subset \mathbb{R}^d$ open, bounded and convex, $\tau > 0$ and an absolutely continuous probability measure $\rho_0 \in \mathcal{P}(\Omega)$

$$\inf_{\rho \in \mathcal{P}(\Omega)} \mathrm{TV}(\rho) + \frac{1}{2\tau} W_2^2(\rho_0, \rho), \tag{TV-W}$$

where τ is interpreted as a time discretization parameter for an implicit Euler scheme, as we shall see below.

The *total variation functional* of a Radon measure $\rho \in \mathcal{M}(\Omega)$ is defined as

$$\mathrm{TV}(\rho) = \sup \left\{ \int_\Omega \mathrm{div} z \mathrm{d}\rho : z \in C_c^1\left(\Omega; \mathbb{R}^N\right), \|z\|_\infty \leq 1 \right\}, \tag{TV}$$

which is not to be mistaken in this paper with the *total variation measure* $|\mu|$ of a Radon measure μ or its *total variation norm* $|\mu|(\Omega)$. We call $\mathrm{BV}(\Omega)$ the subspace of functions $u \in L^1(\Omega)$ whose weak derivative Du is a *finite Radon measure*. It can also be characterized as the L^1 functions such that $\mathrm{TV}(u) < \infty$, where $\mathrm{TV}(u)$ should be understood as in (TV) with the measure $u\mathcal{L}^d \llcorner \Omega$, and it holds that $\mathrm{TV}(u) = |Du|(\Omega)$. As $\mathrm{BV}(\mathbb{R}^d) \hookrightarrow L^{\frac{d}{d-1}}(\mathbb{R}^d)$, solutions to (TV-W) are also absolutely continuous w.r.t. the Lebesgue measure. Therefore, w.l.o.g. we

F. Nielsen and F. Barbaresco (Eds.): GSI 2023, LNCS 14071, pp. 610–619, 2023.
https://doi.org/10.1007/978-3-031-38271-0_61

can minimize on $L^{\frac{d}{d-1}}(\Omega)$, which is a reflexive Banach space. In addition, a function ρ will have finite energy only if $\rho \in \mathcal{P}(\Omega)$.

The data term is given by the Wasserstein distance, defined through the value of the optimal transportation problem (see [19])

$$W_2^2(\mu,\nu) \overset{\text{def.}}{=} \min_{\gamma \in \Pi(\mu,\nu)} \int_{\Omega\times\Omega} |x-y|^2 \mathrm{d}\gamma = \sup_{\substack{\varphi,\psi \in C_b(\Omega) \\ \varphi\oplus\psi \leq |x-y|^2}} \int_\Omega \varphi \mathrm{d}\mu + \int_\Omega \psi \mathrm{d}\nu, \tag{1}$$

where the minimum is taken over all the probability measures on $\Omega \times \Omega$ whose marginals are μ and ν. An optimal pair (φ, ψ) for the dual problem is referred to as Kantorovitch potentials.

Using total variation as regularization was suggested in [18] with a L^2 data term for the *Rudin-Osher-Fatemi problem*

$$\inf_{u \in L^2(\Omega)} \mathrm{TV}(u) + \frac{1}{2\lambda} \|u-g\|^2_{L^2(\Omega)}, \tag{ROF}$$

see [6] for an overview. Other data terms were considered to better model the oscillatory behavior of the noise [15,17]. More recently Wasserstein energies have shown success in the imaging community [12], the model (TV-W) being used for image denoising in [2,4].

Existence and uniqueness of solutions for (TV-W) follow from the direct method in the calculus of variations, and the strict convexity of $W_2^2(\rho_0,\cdot)$ whenever ρ_0 is absolutely continuous, see [19, Prop. 7.19]. However, it is not easy to compute the subdifferential of the sum, which makes the derivation of the Euler-Lagrange equations not trivial.

In [8], the authors studied the gradient flow scheme defined by the successive iterations of (TV-W), and following the seminal work [14] they showed that, in dimension 1 as the parameter $\tau \to 0$, the discrete scheme converges to the solution of a fourth order PDE. They used an entropic regularization approach, followed by a Γ-convergence argument, to derive an Euler-Lagrange equation, which states that there exists a Kantorovitch potential ψ_1 coinciding with some $\operatorname{div} z \in \partial\,\mathrm{TV}(\rho_1)$ in the set $\{\rho_1 > 0\}$. On $\{\rho_1 = 0\}$, these quantities are related through a bounded Lagrange multiplier β associated with the nonnegativity constraint $\rho_1 \geq 0$.

In this work we propose an alternative way to derive the Euler-Lagrange equations which relies on the well established properties of solutions of (ROF) and shows further regularity of the quantities $\operatorname{div} z, \beta$.

Theorem 1. *For any $\rho_0 \in L^1(\Omega) \cap \mathcal{P}(\Omega)$, let ρ_1 be the unique minimizer of* (TV-W)*. The following hold.*

1. *There is a vector field $z \in L^\infty(\Omega;\mathbb{R}^d)$ with $\operatorname{div} z \in L^\infty(\Omega)$ and a bounded Lagrange multiplier $\beta \geq 0$ such that*

$$\begin{cases} \operatorname{div} z + \dfrac{\psi_1}{\tau} = \beta, & a.e.\ in\ \Omega \\ z \cdot \nu = 0, & on\ \partial\Omega \\ \beta\rho_1 = 0, & a.e.\ in\ \Omega \\ z \cdot D\rho_1 = |D\rho_1|, \ \|z\|_\infty \leq 1, \end{cases} \tag{TVW-EL}$$

where ψ_1 is a Kantorovitch potential associated with ρ_1.

2. *The Lagrange multiplier β is the unique solution to (ROF) with $\lambda = 1$ and $g = \psi_1/\tau$.*
3. *The functions $\operatorname{div} z, \psi_1$ and β are Lipschitz continuous.*

2 The Euler-Lagrange Equation

Let X and $X^\star$ be duality-paired spaces and $f : X \to \mathbb{R} \cup \{\infty\}$ be a convex function, the subdifferential of f on X is given by

$$\partial_X f(u) \stackrel{\text{def.}}{=} \{p \in X^\star : f(v) \geq f(u) + \langle p, v - u\rangle\,,\ \text{for all } v \in X\}\,. \tag{2}$$

In order to derive optimality conditions for (TV-W) we will need some properties of the subdifferential of TV and of (ROF).

Proposition 1. *[3, 6, 16] If $u \in \mathrm{BV}(\Omega) \cap L^2(\Omega)$, then the subdifferential of* TV *in $L^2(\Omega)$ at u assumes the form*

$$\partial_{L^2} \mathrm{TV}(u) = \left\{ p \in L^2(\Omega) : \begin{array}{c} p = -\operatorname{div} z,\ \ z \in H_0^1(\operatorname{div}; \Omega), \\ \|z\|_\infty \leq 1,\ \ |Du| = z \cdot Du \end{array} \right\}$$

If in addition u solves (ROF), then

1. *u^+ solves (ROF) with the constraint $u \geq 0$;*
2. *it holds that*
$$0 \in \frac{u - g}{\lambda} + \partial_{L^2} \mathrm{TV}(u), \tag{3}$$
and conversely, if u satisfies (3), u minimizes (ROF);
3. *for Ω convex, if g is uniformly continuous with modulus of continuity ω, then u has the same modulus of continuity.*

In the previous proposition, we recall that $H_0^1(\operatorname{div}; \Omega)$ denotes the closure of $C_c^\infty(\Omega; \mathbb{R}^d)$ with respect to the norm $\|z\|^2_{H^1(\operatorname{div})} = \|z\|^2_{L^2(\Omega)} + \|\operatorname{div} z\|^2_{L^2(\Omega)}$.

Unless otherwise stated, we consider in the sequel $X = L^{\frac{d}{d-1}}(\Omega)$, $X^\star = L^d(\Omega)$ and we drop the index X in the notation ∂_X. Under certain regularity conditions, one can see the Kantorovitch potentials as the first variation of the Wasserstein distance, [19]. As a consequence, Fermat's rule $0 \in \partial\left(W_2^2(\rho_0, \cdot) + \mathrm{TV}(\cdot)\right)(\rho_1)$ assumes the following form.

Lemma 1. *Let ρ_1 be the unique minimizer of (TV-W), then there exists a Kantorovitch potential ψ_1 associated to ρ_1 such that*

$$-\frac{\psi_1}{\tau} \in \partial\left(\mathrm{TV} + \chi_{\mathcal{P}(\Omega)}\right)(\rho_1). \tag{4}$$

Proof. For simplicity, we assume $\tau = 1$. Take $\rho \in \mathrm{BV}(\Omega) \cap \mathcal{P}(\Omega)$ and define $\rho_t \stackrel{\text{def.}}{=} \rho + t(\rho_1 - \rho)$. Since $\overline{\Omega}$ is compact, the sup in (1) admits a maximizer [19, Prop. 1.11]. Let φ_t, ψ_t denote a pair of Kantorovitch potentials between ρ_0 and ρ_t. From the optimality of ρ_1 it follows

$$\begin{aligned}
&\frac{1}{2}W_2^2(\rho_0,\rho_1) + \mathrm{TV}(\rho_1) \le \int_\Omega \varphi_t \mathrm{d}\rho_0 + \int_\Omega \psi_t \mathrm{d}\rho_t + \mathrm{TV}(\rho_t)\\
&\le \int_\Omega \varphi_t \mathrm{d}\rho_0 + \int_\Omega \psi_t \mathrm{d}\rho_1 + \mathrm{TV}(\rho_1) + (1-t)\left(\int_\Omega \psi_t \mathrm{d}(\rho - \rho_1) + \mathrm{TV}(\rho) - \mathrm{TV}(\rho_1)\right)\\
&\le \frac{1}{2}W_2^2(\rho_0,\rho_1) + \mathrm{TV}(\rho_1) + (1-t)\left(\int_\Omega \psi_t \mathrm{d}(\rho - \rho_1) + \mathrm{TV}(\rho) - \mathrm{TV}(\rho_1)\right).
\end{aligned}$$

Hence, $-\psi_t \in \partial\left(\mathrm{TV} + \chi_{\mathcal{P}(\Omega)}\right)(\rho_1)$ for all $t \in (0,1)$. Notice that as the optimal transport map from ρ_0 to ρ_t is given by $T_t = \mathrm{id} - \nabla\psi_t$ and assumes values in the bounded set Ω, the family $(\psi_t)_{t\in[0,1]}$ is uniformly Lipschitz so that by Arzelà-Ascoli's Theorem ψ_t converges uniformly to ψ_1 as t goes to 1 (see also [19, Thm. 1.52]). Therefore, $-\psi_1 \in \partial\left(\mathrm{TV} + \chi_{\mathcal{P}(\Omega)}\right)(\rho_1)$. ■

With these results we can prove Theorem 1.

Proof (of Theorem 1). Here, to simplify, we still assume $\tau = 1$. The subdifferential inclusion (4) is conceptually the Euler-Lagrange equation for (TV-W), however it can be difficult to verify the conditions for direct sum between subdifferentials and give a full characterization. Therefore, for some arbitrary $\rho \in \mathcal{M}_+(\Omega)$ and $t > 0$, set

$$\rho_t = \frac{\rho_1 + t(\rho - \rho_1)}{1 + t\alpha}, \text{ where } \alpha = \int_\Omega \mathrm{d}(\rho - \rho_1).$$

Now ρ_t is admissible for the subdifferential inequality and using the positive homogeneity of TV we can write

$$\mathrm{TV}(\rho_1) - \int_\Omega \psi_1 \mathrm{d}\left(\rho_t - \rho_1\right) \le \frac{\mathrm{TV}(\rho_1) + t\left(\mathrm{TV}(\rho) - \mathrm{TV}(\rho_1)\right)}{1 + t\alpha}.$$

After a few computations we arrive at $\mathrm{TV}(\rho) \ge \mathrm{TV}(\rho_1) + \int_\Omega (C - \psi_1)\mathrm{d}(\rho - \rho_1)$, where $C = \mathrm{TV}(\rho_1) + \int_\Omega \psi_1 \mathrm{d}\rho_1$. Notice that $(\phi + C, \psi - C)$ remains an optimal potential. So we can replace ψ_1 by $\psi_1 - C$, and obtain that for all $\rho \ge 0$ the following holds

$$\mathrm{TV}(\rho) \ge \mathrm{TV}(\rho_1) + \int_\Omega -\psi_1 \mathrm{d}(\rho - \rho_1), \text{ with } \mathrm{TV}(\rho_1) = \int_\Omega -\psi_1 \mathrm{d}\rho_1. \tag{5}$$

In particular, this means $-\psi_1 \in \partial\left(\mathrm{TV} + \chi_{\mathcal{M}_+(\Omega)}\right)(\rho_1)$ and ρ_1 is optimal for

$$\inf_{\rho \ge 0} \mathcal{E}(\rho) := \mathrm{TV}(\rho) + \int_\Omega \psi_1(x)\rho(x)\mathrm{d}x. \tag{6}$$

This suggests a penalization with an L^2 term *e.g.*

$$\inf_{u \in L^2(\Omega)} \mathcal{E}_t(u) := \mathrm{TV}(u) + \int_\Omega \psi_1(x)u(x)\mathrm{d}x + \frac{1}{2t}\int_\Omega |u - \rho_1|^2 \mathrm{d}x \tag{7}$$

which is a variation of (ROF) with $g = \rho_1 - t\psi_1$. In order for (7) to make sense, we need $\rho_1 \in L^2(\Omega)$, which is true if ρ_0 is L^∞ since then [8, Thm. 4.2] implies $\rho_1 \in L^\infty$. Suppose for now that ρ_0 is a bounded function.

Let u_t denote the solution of (7), from Prop. 1 if u_t solves (7), then u_t^+ solves the same problem with the additional constraint that $u \geq 0$, see [5, Lemma A.1]. As $\rho_1 \geq 0$ we can compare the energies of u_t^+ and ρ_1 and obtain the following inequalities

$$\mathcal{E}(\rho_1) \leq \mathcal{E}(u_t^+) \text{ and } \mathcal{E}_t(u_t^+) \leq \mathcal{E}_t(\rho_1).$$

Summing both inequalities yields

$$\int_\Omega |u_t^+ - \rho_1|^2 \mathrm{d}x \leq 0, \text{ therefore } u_t^+ = \rho_1 \text{ a.e. on } \Omega. \tag{8}$$

In particular, we also have that $u_t \leq \rho_1$. But as u_t solves a (ROF) problem, the optimality conditions from Prop. 1 give

$$\beta_t - \psi_1 \in \partial_{L^2} \mathrm{TV}(u_t), \text{ where } \beta_t \overset{\text{def.}}{=} \frac{\rho_1 - u_t}{t} \geq 0. \tag{9}$$

Notice from the characterization of $\partial_{L^2} \mathrm{TV}(\cdot)$ that $\partial_{L^2} \mathrm{TV}(u) \subset \partial_{L^2} \mathrm{TV}(u^+)$. Since $u_t^+ = \rho_1$, we have that

$$\beta_t - \psi_1 \in \partial_{L^2} \mathrm{TV}(\rho_1), \tag{10}$$

which proves (TVW-EL).

Now we move on to study the family $(\beta_t)_{t>0}$. Since $\rho_1 = u_t^+$, by definition $\beta_t = u_t^-/t$ and using the fact that $\partial_{L^2} \mathrm{TV}(u) \subset \partial_{L^2} \mathrm{TV}(u^-)$ in conjunction with Eq. (9), it holds that

$$\psi_1 - \beta_t \in \partial_{L^2} \mathrm{TV}(\beta_t). \tag{11}$$

But then, from Prop. 1, β_t solves (ROF) with $g = \psi_1$ and $\lambda = 1$. As this problem has a unique solution, the family $\{\beta_t\}_{t>0} = \{\beta\}$ is a singleton.

Since Ω is convex, and we know that the Kantorovitch potentials are Lipschitz continuous, cf. [19], so β, as a solution of (ROF) with Lipschitz data $g = \psi_1$, is also Lipschitz continuous with the same constant, following [16, Theo. 3.1].

But from (10) and the characterization of the subdifferential of TV, there is a vector field z such that $z \cdot D\rho_1 = |D\rho_1|$ such that

$$\beta - \psi_1 = \operatorname{div} z,$$

and as a consequence $\operatorname{div} z$ is also Lipschitz continuous, with constant at most twice the constant of ψ_1.

In the general case of $\rho_0 \in L^1(\Omega)$, define $\rho_{0,N} \overset{\text{def.}}{=} c_N(\rho_0 \wedge N)$ for $N \in \mathbb{N}$, where c_N is a renormalizing constant. Then $\rho_{0,N} \in L^\infty(\Omega)$ and $\rho_{0,N} \xrightarrow[N\to\infty]{L^1} \rho_0$.

Let $\rho_{1,N}$ denote the unique minimizer of (TV-W) with data term $\rho_{0,N}$, we can assume that $\rho_{1,N}$ w-$\star$ converges to some $\tilde{\rho}$. Then for any $\rho \in \mathcal{P}(\Omega)$ we have

$$\mathrm{TV}(\rho_{1,N}) + \frac{1}{2\tau} W_2^2(\rho_{0,N}, \rho_{1,N}) \leq \mathrm{TV}(\rho) + \frac{1}{2\tau} W_2^2(\rho_{0,N}, \rho).$$

Passing to the limit on $N \to \infty$ we have that $\tilde{\rho}$ is a minimizer and from uniqueness it must hold that $\tilde{\rho} = \rho_1$.

Hence, consider the functions $z_N, \psi_{1,N}, \beta_N$ that satisfy (TVW-EL) for $\rho_{1,N}$. Up to a subsequence, we may assume that z_N converges weakly-$\star$ to some $z \in L^\infty(\Omega; \mathbb{R}^d)$. Since $\psi_{1,N}$, β_N and $\operatorname{div} z_N$ are Lipschitz continuous with the same Lipschitz constant for all N, by Arzelà-Ascoli, we can assume that $\psi_{1,N}, \beta_N$ and $\operatorname{div} z_N$ converge uniformly to Lipschitz functions $\psi_1, \beta, \operatorname{div} z = \beta - \psi_1$. In addition, passing to the limit in (11), we find that β solves (ROF) for $\lambda = 1$ and $g = \psi_1$.

Since β_N converges uniformly and $\rho_{1,N}$ converges w-$\star$ we have

$$0 = \lim_{N \to \infty} \int_\Omega \beta_N \rho_{1,N} \mathrm{d}x = \int_\Omega \beta \rho_1 \mathrm{d}x,$$

and hence $\beta\rho_1 = 0$ a.e. in Ω since both are nonnegative. In addition, ψ_1 is a Kantorovich potential associated to ρ_1 from the stability of optimal transport (see [19, Thm. 1.52]). From the optimality of $\rho_{1,N}$ it holds that

$$\mathrm{TV}(\rho_{1,N}) + \frac{1}{2\tau} W_2^2(\rho_{0,N}, \rho_{1,N}) \leq \mathrm{TV}(\rho_1) + \frac{1}{2\tau} W_2^2(\rho_{0,N}, \rho),$$

so that $\lim \mathrm{TV}(\rho_{1,N}) \leq \mathrm{TV}(\rho_1)$. Changing the roles of ρ_1 and $\rho_{1,N}$ we get an equality. So it follows that

$$\int_\Omega (\beta - \psi_1)\rho_1 \mathrm{d}x = \lim_{N \to \infty} \int_\Omega (\beta_N - \psi_{1,N})\rho_{1,N} \mathrm{d}x = \lim_{N \to \infty} \mathrm{TV}(\rho_{1,N}) = \mathrm{TV}(\rho_1),$$

Since TV is 1-homogeneous we conclude that $\beta - \psi_1 \in \partial \, \mathrm{TV}(\rho_1)$. ■

We say E is a set of finite perimeter if the indicator function $\mathbb{1}_E$ is a BV function, and we set $\mathrm{Per}(E) = \mathrm{TV}(\mathbb{1}_E)$. As a byproduct of the previous proof we conclude that the level sets $\{\rho_1 > s\}$ are all solutions to the same prescribed curvature problem.

Corollary 1. *The following properties of the level sets of ρ_1 hold.*

1. *For $s > 0$ and ψ_1 in (TVW-EL)*

$$\{\rho_1 > s\} \in \operatorname*{argmin}_{E \subset \Omega} \mathrm{Per}(E; \Omega) + \frac{1}{\tau} \int_E \psi_1 \mathrm{d}x$$

2. *$\partial\{\rho_1 > s\} \setminus \partial^*\{\rho_1 > s\}$ is a closed set of Hausdorff dimension at most $d - 8$, where ∂^* denotes the reduced boundary of a set, see [1]. In addition, $\partial^*\{\rho_1 > s\}$ is locally the graph of a function of class $W^{2,q}$ for all $q < +\infty$.*

Proof. For simplicity take $\tau = 1$. Inside the set $\{\rho_1 > s\}$, for $s > 0$, we have $-\psi_1 = \operatorname{div} z$, so from the definition of the perimeter we have

$$\int_{\{\rho_1 > s\}} -\psi_1 \mathrm{d}x = \int_{\{\rho_1 > s\}} \operatorname{div} z \mathrm{d}x \leq \operatorname{Per}(\{\rho_1 > s\}).$$

So using the fact that $\operatorname{TV}(\rho_1) = \int_\Omega -\psi_1 \mathrm{d}x$, the coarea formula and Fubini's Theorem give

$$\int_0^{+\infty} \operatorname{Per}(\mathbb{1}_{\{\rho_1 > s\}}) \mathrm{d}s = \int_\Omega -\psi_1 \int_0^{\rho_1(x)} \mathrm{d}s \mathrm{d}x = \int_0^{+\infty} \int_{\{\rho_1 > s\}} -\psi_1 \mathrm{d}x \mathrm{d}s.$$

Hence, $\operatorname{Per}(\{\rho_1 > s\}) = \int_{\{\rho_1 > s\}} -\psi_1 \mathrm{d}x$ for *a.e.* $s > 0$. But as $\beta\psi_1 = 0$ a.e., we have $-\psi_1 = \operatorname{div} z$ in $\{\rho_1 > s\}$, so that $-\psi_1 \in \partial \operatorname{TV}(\mathbb{1}_{\{\rho_1 > s\}})$ for *a.e.* $s > 0$; and by a continuity argument, for all $s > 0$. The subdifferential inequality with $\mathbb{1}_E$ gives

$$\{\rho_1 > s\} \in \underset{E \subset \Omega}{\operatorname{argmin}} \operatorname{Per}(E) + \int_E \psi_1(x) \mathrm{d}x. \tag{12}$$

Item (2) follows directly from the properties of (ROF), see [6], since $\rho_1 = u^+$, where u solves a problem (ROF).

3 Numerical Experiments

We solve (TV-W) for an image denoising application using a Douglas-Rachford algorithm [9] with Halpern acceleration [11], see Table 1. For this we need subroutines to compute the prox operators defined, for a given $\lambda > 0$, as

$$\operatorname{prox}_{\lambda \operatorname{TV}}(\bar\rho) \overset{\text{def.}}{=} \underset{\rho \in L^2(\Omega)}{\operatorname{argmin}} \operatorname{TV}(\rho) + \frac{1}{2\lambda} \|\rho - \bar\rho\|^2_{L^2(\Omega)}, \tag{13}$$

$$\operatorname{prox}_{\lambda W_2^2}(\bar\rho) \overset{\text{def.}}{=} \underset{\rho \in L^2(\Omega)}{\operatorname{argmin}} \frac{1}{2\tau} W_2^2(\rho_0, \rho) + \frac{1}{2\lambda} \|\rho - \bar\rho\|^2_{L^2(\Omega)}. \tag{14}$$

We implemented the prox of TV with the algorithm from [10], modified to account for Dirichlet boundary conditions. From [7, Theo. 2.4] it is consistent with the continuous total variation. The prox of W_2^2 is computed by expanding the L^2 data term as

$$\begin{aligned}\operatorname{prox}_{\lambda W_2^2}(\bar\rho) &= \underset{\rho \in L^2(\Omega)}{\operatorname{argmin}} \frac{1}{2\tau} W_2^2(\rho_0, \rho) + \frac{1}{2\lambda} \int_\Omega \rho^2 \mathrm{d}x + \int_\Omega \rho \underbrace{\left(-\frac{\bar\rho}{\lambda}\right)}_{=V} \mathrm{d}x + \underbrace{\frac{1}{2\lambda} \bar\rho^2 \mathrm{d}x}_{cst} \\ &= \underset{\rho \in L^2(\Omega)}{\operatorname{argmin}} \frac{1}{2\tau} W_2^2(\rho_0, \rho) + \frac{1}{2\lambda} \int_\Omega \rho^2 \mathrm{d}x + \int_\Omega \rho V \mathrm{d}x,\end{aligned}$$

which is one step of the Wasserstein gradient flow of the porous medium equation $\partial_t \rho_t = \lambda^{-1} \Delta(\rho_t^2) + \operatorname{div}(\rho_t \nabla V)$, where the potential is $V = -\bar\rho/\lambda$, see [13,19]. To compute it we have used the back-n-forth algorithm from [13].

Algorithm 1. Halpern accelerated Douglas-Rachford algorithm

$\beta_0 \leftarrow 0$
$x_0 \leftarrow$ Initial Image
while $n \geq 0$ **do**
  $y_n \leftarrow \mathrm{prox}_{\lambda\,\mathrm{TV}}(x_n)$
  $\lambda_n \in [\varepsilon, 2-\varepsilon]$
  $z_n \leftarrow x_n + \lambda_n \left(\mathrm{prox}_{\lambda W_2^2}(2y_n - x_n) - y_n\right)$
  $\beta_n \leftarrow \frac{1}{2}\left(1+\beta_{n-1}^2\right)$ ▷ Optimal constants for Halpern acceleration from [11]
  $x_{n+1} \leftarrow (1-\beta_n)x_0 + \beta_n z_n$
end while

3.1 Evolution of Balls

Following [8], in dimension 1, whenever the initial measure is uniformly distributed over a ball, the solutions remain balls. In $\mathbb{R}^d$, one can prove this remains true. If ρ_0 is uniformly distributed over a ball of radius r_0, then the solution to (TV-W) is uniformly distributed in a ball of radius r_1 solving the following polynomial equation for r_1

$$r_1^2(r_1 - r_0) = r_0^2(d+2)\tau.$$

This theoretical predictions are corroborated by the numerical experiments found in Fig. 1.

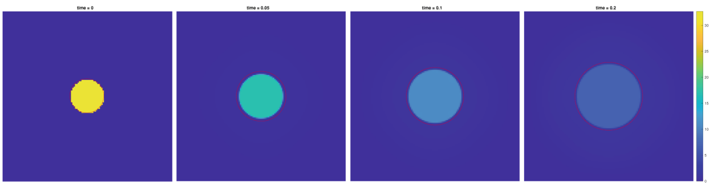

Fig. 1. Evolution of circles: from left to right initial condition and solutions for $\tau = 0.05, 0.1, 0.2$. The red circles correspond to the theoretical radius. (Color figure online)

3.2 Reconstruction of Dithered Images

In this experiment we use model (TV-W) to reconstruct dithered images. In $\mathcal{P}(\mathbb{R}^2)$ the dithered image is a sum of Dirac masses, so the model (TV-W) outputs a new image which is close in the Wasserstein topology, but with small total variation. In Fig. 2 below, we compared the result with the reconstruction given by (ROF), both with a parameter $\tau = 0.2$. Although the classical (ROF) model was able to create complex textures, these remain granulated, whereas the (TV-W) model is able to generate both smooth and complex textures.

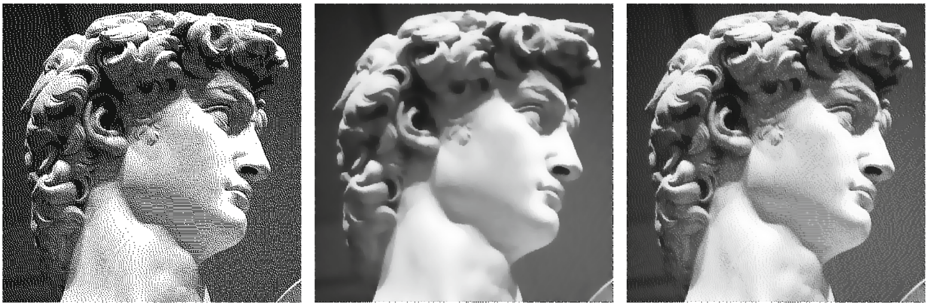

Fig. 2. Dithering reconstruction problem. From left to right: Dithered image, TV-Wasserstein and ROF results.

4 Conclusion

In this work we revisited the TV-Wasserstein problem. We showed how it can be related to the classical (ROF) problem and how to exploit this to derive the Euler-Lagrange equations, obtaining further regularity. We proposed a Douglas-Rachford algorithm to solve it and presented two numerical experiments: the first one being coherent with theoretical predictions and the second being an application to the reconstruction of dithered images.

Acknowledgements. The second author acknowledges support from the ANR CIPRESSI project, grant 19-CE48-0017-01 of the French Agence Nationale de la Recherche. The third author acknowledges the financial support from FMJH concerning his master thesis, when this work took place.

References

1. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems, 2nd edn. Oxford University Press, New York (2000)
2. Benning, M., Calatroni, L., Düring, B., Schönlieb, C.-B.: A primal-dual approach for a total variation wasserstein flow. In: Nielsen, F., Barbaresco, F. (eds.) GSI 2013. LNCS, vol. 8085, pp. 413–421. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40020-9_45
3. Bredies, K., Holler, M.: A pointwise characterization of the subdifferential of the total variation functional. arXiv preprint arXiv:1609.08918 (2016)
4. Burger, M., Franek, M., Schonlieb, C.B.: Regularized regression and density estimation based on optimal transport. Appl. Math. Res. eXpress **2012**(2), 209–253 (2012)
5. Chambolle, A.: An algorithm for mean curvature motion. Interf. Free Bound. **6**(195–218), 2 (2004)
6. Chambolle, A., Caselles, V., Cremers, D., Novaga, M., Pock, T.: An introduction to total variation for image analysis. Theor. Found. Numer. Methods Sparse Rec. **9**(263–340), 227 (2010)

7. Chambolle, A., Pock, T.: Learning consistent discretizations of the total variation. SIAM J. Imaging Sci. **14**(2), 778–813 (2021)
8. Carlier, G., Poon, C.: On the total variation Wasserstein gradient flow and the TV-JKO scheme. ESAIM: COCV **25**(41) (2019)
9. Combettes, P.L., Pesquet, J.C.: Proximal splitting methods in signal processing. In: Bauschke, H., Burachik, R., Combettes, P., Elser, V., Luke, D., Wolkowicz, H. (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer Optimization and Its Applications, vol. 49. Springer, New York (2011). https://doi.org/10.1007/978-1-4419-9569-8_10
10. Condat, L.: Discrete total variation: new definition and minimization. SIAM J. Imaging Sci. **10**(3), 1258–1290 (2017)
11. Contreras, J.P., Cominetti, R.: Optimal error bounds for non-expansive fixed-point iterations in normed spaces. Math. Program. **199**, 343–374 (2022)
12. Cuturi, M., Peyré, G.: Semidual regularized optimal transport. SIAM Rev. **60**(4), 941–965 (2018)
13. Jacobs, M., Lee, W., Léger, F.: The back-and-forth method for Wasserstein gradient flows. ESAIM Control Optim. Calc. Var. **27**, 28 (2021)
14. Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker-Planck equation. SIAM J. Math. Anal. **29**(1), 1–17 (1998)
15. Lieu, L.H., Vese, L.A.: Image restoration and decomposition via bounded total variation and negative hilbert-sobolev spaces. Appl. Math. Optim. **58**, 167–193 (2008)
16. Mercier, G.: Continuity results for TV-minimizers. Indiana Univ. Math. J., 1499–1545 (2018)
17. Meyer, Y.: Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures, vol. 22. American Mathematical Society (2001)
18. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena **60**(4), 259–268 (1992)
19. Santambrogio, F.: Optimal Transport for Applied Mathematicians, 1st edn. Birkhauser, New York (2015)