SIAM J. MATH. ANAL.
Vol. 56, No. 1, pp. 1114–1148

AN OPTIMAL TRANSPORT ANALOGUE OF THE RUDIN–OSHER–FATEMI MODEL AND ITS CORRESPONDING MULTISCALE THEORY*

TRISTAN MILNE† AND ADRIAN NACHMAN†‡

Abstract. In the first part of this paper we develop a theory for image restoration with a learned regularizer that is analogous to that of Meyer's geometric characterization of solutions of the classical variational method of Rudin–Osher–Fatemi (ROF). The learned regularizer we use is a Kantorovich potential for an optimal transport problem of mapping a distribution of noisy images onto clean ones, as first proposed by Lunz, Öktem, and Schönlieb. We show that the effect of their restoration method on the distribution of the images is an explicit Euler discretization of a gradient flow on probability space, while our variational problem, dubbed Wasserstein ROF (WROF), is the corresponding implicit Euler discretization. We obtain our geometric characterization of the solution in this learned regularizer setting by first proving a much more general convex analysis theorem for variational problems having solutions characterized by projections. We then use optimal transport arguments to obtain the corresponding theorem for WROF from this general result, as well as a natural decomposition of a transport map into large scale "features" and small scale "details," where scale refers to the magnitude of the transport distance. In the second part of the paper we leverage our theory for restoration with learned regularizers to analyze two algorithms which iterate WROF. We refer to these as iterative regularization and multiscale transport. For the former we obtain a proof of convergence to the clean data. For the latter we produce successive approximations to the target distribution that match it up to finer and finer scales. These two algorithms are in complete analogy to well-known effective methods based on ROF for iterative denoising, respectively hierarchical image decomposition. We also obtain an analogue of the Tadmor–Nezzar–Vese energy identity, which decomposes the Wasserstein 2 distance between two measures into a sum of nonnegative terms that correspond to transport costs at different scales.

Key words. variational image restoration, learned regularizers, optimal transport, multiscale optimal transport

MSC codes. 94A08, 90B06

DOI. 10.1137/23M1564109

1. Introduction.

A well-known classical method for image restoration is the total variation (TV) approach of Rudin, Osher, and Fatemi (ROF) [29]. In this technique, a noisy image $f \in L^2(\mathbb{R}^2)$ is restored by solving the problem

$$\min_{u \in L^2(\mathbb{R}^2)} \frac{1}{2}\|u - f\|^2_{L^2(\mathbb{R}^2)} + \lambda \|u\|_{TV}. \tag{1.1}$$

Here, $\|u\|_{TV}$ is the TV-norm of u, a regularizer known for promoting smoothness while preserving edges. Related to (1.1) is the more recent variational denoising method of [21]. The important novelty of [21] is that it uses a learned regularizer instead of the TV-norm to impose regularity. The motivation for this is that one may be able to obtain a more effective regularizer—and experiments show that this is in fact

*Received by the editors April 14, 2023; accepted for publication July 17, 2023; published electronically January 31, 2024. This work appeared in part in the first author's Ph.D. thesis.
https://doi.org/10.1137/23M1564109
Funding: The work of the authors was supported by NSERC Discovery Grant RGPIN-06329.
†Department of Mathematics, University of Toronto, 40 St. George Street, Toronto M5T 3J1, ON, Canada (tmilne@math.toronto.edu, nachman@math.toronto.edu).
‡Edward S. Rogers, Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto M5S 3G4, ON, Canada.

the case—by learning it from datasets of noisy and clean images rather than using a handcrafted one. The particular learned regularizer proposed in [21] is a Kantorovich potential u_0 for the Wasserstein 1 distance $W_1(\mu,\nu)$, where μ and ν are probability distributions of noisy and clean data, respectively, on a compact and convex domain $\Omega \subset \mathbb{R}^d$. That is, u_0 solves the problem

$$\sup_{u \in 1\text{-Lip}(\Omega)} \int_\Omega u(x) d\mu(x) - \int_\Omega u(y) d\nu(y),$$

where 1-Lip(Ω) is the set of functions with Lipschitz constant 1 on Ω. The solution u_0 is thus incentivized to take large values on the noisy data μ and small values on the real data ν, justifying its role in restoring a noisy image[1] $x_0 \sim \mu$ by solving

$$\min_{x \in \Omega} \frac{1}{2}|x - x_0|^2 + \lambda u_0(x). \tag{1.2}$$

Experiments in [21] show that denoising performance is improved by using this learned regularizer as opposed to the TV-norm.

The ROF model has been intensively studied and has a well-developed and beautiful theory (e.g., [23, 8, 9, 10]). Let us briefly outline some of the results in [23]. The solution u_λ to (1.1) can be described geometrically as the projection of 0 onto a certain norm ball of radius λ centered at f. Moreover, the wavelet coefficients of the residual $f - u_\lambda$ satisfy an ℓ_∞ bound in terms of λ, and an approximate solution to (1.1) can be obtained via soft thresholding of the wavelet coefficients of f. Building on these results, (1.1) can be solved iteratively to obtain iterative denoising (see [4] or section 7.1 of [32]) and the nonlinear hierarchical image decomposition of [33]. The latter can be viewed as nonlinear harmonic analysis of the image into components at finer and finer scale, and the analogy is further strengthened by an elegant corresponding energy equality.

We were motivated by these results for ROF to search for a corresponding theory for a learned regularizer problem related to (1.2). The first part of this paper establishes theorems analogous to those of [23] for a learned regularizer setting. It also includes a decomposition of a certain transport map into large scale "features" and small scale "details"; in this context, scale refers to the magnitude of the transport distance. The second part of the paper leverages our results to analyze two natural iterative optimal transport procedures. We refer to these as iterative regularization and multiscale transport, as they are in correspondence with iterative denoising with ROF and the multiscale image decomposition of [33]. For the former, we prove convergence toward the clean data distribution ν. The latter has a richer structure and modifies ν at each stage to obtain a "sketch" of μ which is indistinguishable from it up to a predefined scale. Our results in this direction also include an energy identity analogous to that of [33] which decomposes the squared Wasserstein 2 distance $W_2^2(\mu,\nu)$ into a sum of nonnegative terms which picks out the scales of transport.

While (1.2) is a pointwise formulation of image restoration, the setting is more global in that u_0 depends on the distribution μ and ν of noisy and clean images. We have thus found it more natural to analyze the measure obtained by modifying μ with the solution map to (1.2). Taking this as a starting point, the main object of study in this paper is

$$\inf_{\rho \in \mathcal{P}(\Omega)} \frac{1}{2} W_2^2(\mu,\rho) + \lambda W_1(\rho,\nu). \tag{WROF}$$

[1]Images are taken as vectors in $\mathbb{R}^d$ here, unlike (1.1), where they are elements of $L^2(\mathbb{R}^2)$.

Here $\mathcal{P}(\Omega)$ is the space of Borel probability measures on Ω, and for $p \geq 1$, $W_p : \mathcal{P}(\Omega) \times \mathcal{P}(\Omega) \to \mathbb{R}$ is the Wasserstein p distance; for more background on optimal transport we refer the reader to [30] or [34]. Given that μ consists of noisy images, and ν is a distribution of clean images, we view $\frac{1}{2}W_2^2(\mu, \rho)$ as a fidelity term while $W_1(\rho, \nu)$ measures regularity. As we will see in Theorems 1.3 and 1.6, this problem has properties which are in exact correspondence with the aforementioned results for ROF. As a consequence we call it Wassertein ROF (or WROF for short).

To motivate the study of (WROF), let us specify its relationship to the image denoising technique of [21]. We will show, in Lemma 3.3, that the measure obtained by pushing μ forward under the solution map of (1.2) is the unique solution to

$$\inf_{\rho \in \mathcal{P}(\Omega)} \frac{1}{2} W_2^2(\rho, \mu) + \lambda \langle u_0, \rho \rangle. \tag{1.3}$$

Since u_0 is a subgradient of the convex functional $\mu \mapsto W_1(\mu, \nu)$, (1.3) can be viewed as an explicit Euler discretization of a gradient flow on the space $\mathbb{W}_2(\Omega)$ of probability distributions metrized by the Wasserstein 2 distance. A step of the implicit Euler discretization of the same flow is (WROF). We focus on (WROF), as opposed to (1.3), because in general the implicit method has better properties than the explicit one. We note, however, that in certain settings the two approaches coincide (see Proposition 3.5). In addition, the implicit Euler approach retains a pointwise reconstruction method; there is a continuous function φ_λ such that the solution ρ_λ to (WROF) is obtained by modifying μ pointwise by the solution map for

$$\inf_{x \in \Omega} \frac{1}{2}|x - x_0|^2 - \varphi_\lambda(x). \tag{1.4}$$

In fact, φ_λ is a Kantorovich potential for the transport from μ to ν under the cost function $c_{2,\lambda}$ defined in (1.7) (see Proposition 1.4). In this sense, the solution to (WROF) is obtained via restoration with a learned regularizer $-\varphi_\lambda$. Moreover, φ_λ can be taken so that $\frac{1}{2}|x|^2 - \varphi_\lambda(x)$ is convex, which implies that the pointwise restoration algorithm (1.4) has the additional benefit of being a convex optimization problem; in this light, (1.4) bears a similarity to the convex learned regularizers of [28]. We also suspect that restoration via (1.4) may be more effective than (1.2), since Proposition 1.8 shows that iterations of this procedure provably converge to the clean image distribution ν.

In the remainder of this section we will summarize our main results, with subsection 1.1 describing our geometric characterization of the solution of (WROF), while subsections 1.2 and 1.3 outline our iterative procedures.

1.1. Geometric characterization of the solution of (WROF). In this section we provide analogues in the setting of a learned regularizer of results giving a geometric characterization of the solution to ROF.

First, we recall some classical results for ROF. In studying this problem, it is helpful to define the dual norm to $\|\cdot\|_{TV}$; for $v \in L^2(\mathbb{R}^2)$, define the $*$-norm as

$$\|v\|_* = \sup \left\{ \int_{\mathbb{R}^2} v u dx \mid \|u\|_{TV} \leq 1 \right\}. \tag{1.5}$$

The following theorem, mentioned in section 1, is a slight reformulation of results from [23] on the solution to (1.1). Specifically, it characterizes the solution as a projection of 0 onto a ball in the $*$-norm centered at f.

TABLE 1
The analogy between (1.1) *and* (WROF). *The decompositions of* f *and* S_0 *are described in* (1.15) *and* (1.13), *respectively.*

	ROF	WROF
Fidelity	$\|u-f\|^2_{L^2(\mathbb{R}^2)}$	$W_2^2(\rho,\mu)$
Regularity	$\|u\|_{TV}$	$W_1(\rho,\nu)$
Projection metric	$\|u\|^2_{L^2(\mathbb{R}^2)}$	$D_\lambda(\nu,\rho)$
Projection set	$\{u \mid \|u-f\|_* \le \lambda\}$	$B_\lambda(\mu)$
Decomposition	$f = v_\lambda + u_\lambda$	$S_0 = T_\lambda^{-1} \circ S_\lambda$

THEOREM 1.1 (Meyer). *For all* $\lambda > 0$, (1.1) *has a unique solution* u_λ, *which can also be expressed as the solution to*

$$\min_{\|u-f\|_* \le \lambda} \|u\|^2_{L^2(\mathbb{R}^2)}. \tag{1.6}$$

Consequently, if $\|f\|_* \le \lambda$, $u_\lambda = 0$. *On the other hand, if* $\|f\|_* > \lambda$, *then* $\|f - u_\lambda\|_* = \lambda$ *and*

$$\int_{\mathbb{R}^2} u_\lambda (f - u_\lambda) dx = \lambda \|u\|_{TV}.$$

Remark 1.2. Theorem 1.1 provides a formal statement of some of the results we have mentioned in section 1. For a statement of further results on ROF, such as the ℓ_∞ bound on the wavelet coefficients of $f - u_\lambda$ or the fact that an approximate solution can be obtained by applying soft thresholding to the wavelet coefficients of f, see [23, Lemma 10, section 1.14].

Our Theorem 1.3 gives analogous results for (WROF). To make the analogy clear, Table 1 gives the correspondence between the key concepts. In this case, the measure ν is projected with respect to a divergence D_λ onto a set of measures $B_\lambda(\mu)$. We will be more precise about D_λ and $B_\lambda(\mu)$ in (5.8) and (5.6). We will see that these notions are natural from the point of view of convex analysis; for now, we describe them in intuitive terms.

A key role will be played by an optimal transport problem that uses a cost function $c_{2,\lambda} : \Omega \times \Omega \to \mathbb{R}$ related to the Huber loss function [13] for robust estimation. It is given by

$$c_{2,\lambda}(x,y) = \begin{cases} \frac{1}{2}|x-y|^2, & |x-y| \le \lambda, \\ \lambda|x-y| - \frac{\lambda^2}{2}, & |x-y| \ge \lambda. \end{cases} \tag{1.7}$$

This can be viewed as a variation on the standard cost function $c_2(x,y) = \frac{1}{2}|x-y|^2$, except with a certain economy of scale; in particular, the cost of transport at distances larger than λ is discounted. This may be advantageous for image restoration since this cost is robust to outliers. The relationship between the solution ρ_λ to (WROF) and an optimal plan transporting μ to ν under the cost $c_{2,\lambda}$ will be made explicit in Proposition 1.4. We also note that the minimum value of (WROF) is the optimal transport cost from μ to ν for the pointwise cost $c_{2,\lambda}$; see Corollary 5.8.

The set $B_\lambda(\mu)$ consists of measures which can be reached from μ with displacement less than λ by an optimal transport plan for the cost $c_{2,\lambda}$. In this sense, measures in $B_\lambda(\mu)$ are indistinguishable from μ up to scale λ.

The divergence $D_\lambda(\nu,\rho)$ is nonnegative and is 0 only when $\rho=\nu$ provided μ is absolutely continuous with respect to Lebesgue measure, which we denote by $\mu \ll \mathcal{L}_d$. Further, we will show that $D_\lambda(\nu,\rho)$ has an interesting economic interpretation. In short, assuming that goods are sold to consumers with distribution ν and purchased from a manufacturer with distribution ρ, $D_\lambda(\nu,\rho)$ represents the total loss of value in a supply chain when the transport cost has an economy of scale and consumers adopt a "buy local" policy. More concretely, at the optimal ρ_λ for (WROF), $D_\lambda(\nu,\rho_\lambda)$ measures the amount of transport between μ and ν at scale larger than λ; our results (specifically Theorem 5.6, together with Corollary 5.8) imply

$$\int_{\Omega^2} \frac{1}{2}(|x-y|-\lambda)_+^2 d\tilde{\gamma}_0 \geq D_\lambda(\nu,\rho_\lambda) \geq \int_{\Omega^2} \frac{1}{2}(|x-y|-\lambda)_+^2 d\gamma_0, \tag{1.8}$$

where $\tilde{\gamma}_0$ and γ_0 are optimal plans for transporting μ to ν under the costs $c_{2,\lambda}$ and c_2, respectively.

Analogously to Theorem 1.1, our first theorem expresses the solution to (WROF) as a projection of ν onto $B_\lambda(\mu)$. We also include an additional result (see (1.11)) which is analogous to the ℓ_∞ bound on the wavelet coefficients of the residual $f-u_\lambda$ mentioned in Remark 1.2.

THEOREM 1.3 (main theorem, part 1). *Let Ω be compact and convex with non-negligible interior, and suppose $\mu \ll \mathcal{L}_d$. For all $\lambda>0$,* (WROF) *has a unique solution ρ_λ, which can also be expressed as the solution to*

$$\min_{\rho \in B_\lambda(\mu)} D_\lambda(\nu,\rho). \tag{1.9}$$

Consequently, if $\nu \in B_\lambda(\mu)$, $\rho_\lambda=\nu$. On the other hand, if $\nu \notin B_\lambda(\mu)$, then there exists φ_λ a Kantorovich potential for $W_2(\mu,\rho_\lambda)$ satisfying $Lip(\varphi_\lambda)=\lambda$ and

$$\int_\Omega \varphi_\lambda(d\nu - d\rho_\lambda) = \lambda W_1(\rho_\lambda,\nu). \tag{1.10}$$

Finally, the optimal transport map T_λ for $W_2(\mu,\rho_\lambda)$ satisfies

$$\|I-T_\lambda\|_{L^\infty(\mu)} \leq \lambda. \tag{1.11}$$

A more detailed version of this result is given in Theorem 5.6. In sections 4 and 5 we will clarify the strong similarities between Theorems 1.1 and 1.3 by proving a general theorem for a class of convex optimization problems of the form (4.1) for which the solution map is a projection. We will show that ROF and (WROF) are included in this class, so that Theorems 1.1 and 1.3 will follow as particular cases.

More insight into φ_λ and T_λ from Theorem 1.3 is given in the following proposition.

PROPOSITION 1.4. *Under the notation and assumptions of Theorem* 1.3,

1. *φ_λ is a solution to*

$$\sup_{\varphi \in C(\Omega)} \int_\Omega \varphi^{c_{2,\lambda}} d\mu + \int_\Omega \varphi d\nu,$$

   where $\varphi^{c_{2,\lambda}}(x) = \inf_{y\in\Omega} c_{2,\lambda}(x,y) - \phi(y)$,
2. *T_λ, which by definition satisfies $(T_\lambda)_\# \mu = \rho_\lambda$, is the solution map to* (1.4), *and*

3. *if γ_0 is an optimal transport plan for transporting μ to ν under the cost $c_{2,\lambda}$, and if $(x,y) \in \mathrm{spt}(\gamma_0)$, then*

$$T_\lambda(x) = \begin{cases} y, & |x-y| \le \lambda, \\ \left(1 - \dfrac{\lambda}{|x-y|}\right) x + \dfrac{\lambda}{|x-y|} y, & |x-y| > \lambda. \end{cases} \tag{1.12}$$

Remark 1.5. Proposition 1.4 shows precisely the outcome $T_\lambda(x_0)$ of restoring a noisy image x_0 by solving (1.4) with the learned regularizer φ_λ. The answer is determined by γ_0; if $(x_0, y_0) \in \mathrm{spt}(\gamma_0)$ is such that $|x_0 - y_0| \le \lambda$, T_λ completes the transport from x_0 to y_0. On the other hand, if $|x_0 - y_0| > \lambda$, T_λ takes a step of size λ in the direction of y_0.

Assuming that ν is also absolutely continuous, we further establish in the following theorem that ρ_λ is obtained by applying soft thresholding to an optimal transport map from ν to μ. Recall that the soft thresholding map is given by $s_\lambda : \mathbb{R} \to \mathbb{R}$,

$$s_\lambda(t) := \mathrm{sign}(t)(|t| - \lambda)_+.$$

This provides an analogous result to the soft thresholding property of ROF mentioned in Remark 1.2, except that here we obtain the exact solution rather than an approximate one.

THEOREM 1.6 (main theorem, part 2). *In addition to the hypotheses of Theorem 1.3, assume that $\nu \ll \mathcal{L}_d$. Then*

1. *$\rho_\lambda \ll \mathcal{L}_d$,*
2. *S_0 is an optimal transport map for the cost $c_{2,\lambda}$ sending ν to μ if and only if*

$$S_0 = T_\lambda^{-1} \circ S_\lambda, \tag{1.13}$$

where T_λ^{-1} is a Borel map satisfying $T_\lambda^{-1} \circ T_\lambda(x) = x$ μ almost everywhere, and S_λ is an optimal transport map for $W_1(\nu, \rho_\lambda)$.

3. *For any such S_0, the solution ρ_λ to* (WROF) *is obtained as $\rho_\lambda = (S_\lambda)_\# \nu$, where*

$$S_\lambda(y) := y + s_\lambda(|S_0(y) - y|) \frac{S_0(y) - y}{|S_0(y) - y|}. \tag{1.14}$$

Remark 1.7. This result gives a further interpretation of λ as a scale parameter, in the sense that the solution ρ_λ to (WROF) is obtained from ν by only transporting mass that moves larger than distance λ under S_0. The formula (1.13) also deepens the analogy to ROF. Recall that, writing the residual $f - u_\lambda$ as v_λ, ROF provides a decomposition of the image f into "features" u_λ and "details" v_λ, connected by the formula

$$f = v_\lambda + u_\lambda. \tag{1.15}$$

Equation (1.13) is an optimal transport analogue of this decomposition, the analogy being obtained by replacing addition with composition. Thus, the transport map S_0 is decomposed into S_λ (which we think of as features in the sense that it only involves large scale transport) and details T_λ^{-1} which only involve transport less than distance λ (see (1.11)). This decomposition will be analyzed in detail in section 7.

1.2. Iterative regularization. We now move to a description of the results in the second part of the paper and introduce our first iterative procedure. It is in correspondence with iterated denoising through repeated applications of ROF (see [4] or section 7.1 of [32]). Here we study iterations of the problem (WROF), where at each stage μ is replaced with the previous solution ρ_λ. When μ is a distribution of noisy images and ν is a distribution of clean ones, this represents the iterative regularization of μ. The following proposition is our main result in this direction.

PROPOSITION 1.8. *Let Ω be convex and compact with nonnegligible interior. Let $\mu, \nu \ll \mathcal{L}_d$, and suppose that $(\lambda_n)_{n=0}^{\infty}$ is a sequence of positive step sizes with*

$$\sum_{n=0}^{\infty} \lambda_n = +\infty. \tag{1.16}$$

Given $\mu_0 := \mu$, for each $n \geq 0$ define

$$\mu_{n+1} := \operatorname*{arg\,min}_{\rho \in \mathcal{P}(\Omega)} \frac{1}{2} W_2^2(\rho, \mu_n) + \lambda_n W_1(\rho, \nu). \tag{1.17}$$

Then

$$\lim_{n \to \infty} W_1(\mu_n, \nu) = 0. \tag{1.18}$$

We note that due to statement 1 of Theorem 1.6, if μ and ν are absolutely continuous, then the solution to (WROF) is absolutely continuous as well. In connection with Theorem 1.3, this guarantees that the argmin in (1.17) is unique for each n, establishing that the sequence μ_n is well defined.

1.3. Multiscale transport and a nonlinear energy decomposition. Our second iterative process proceeds in the other direction (i.e., "adding detail" as opposed to denoising) and reveals a richer structure. It is analogous to the hierarchical image decomposition from [33], and so we first briefly recall those results here. This approach leverages ROF to decompose an image f into a hierarchical representation $(u_n)_{n=1}^{\infty}$ of features at different scales by setting

$$u_{n+1} := \operatorname*{arg\,min}_{u \in L^2(\mathbb{R}^2)} \|u - v_n\|_{L^2(\mathbb{R}^2)}^2 + \lambda_{n+1} \|u\|_{TV}, \quad v_n = f - \sum_{i=1}^{n} u_i, \tag{1.19}$$

where $v_0 := f$ and $\lambda_n = 2^{-n+1}\lambda_1$. Thus, at each stage the "detail" component v_n is broken down into smaller scale features u_{n+1} and details v_{n+1}. The following theorem[2] establishes that $(u_n)_{n=1}^{\infty}$ is indeed a decomposition of f and provides a nonlinear harmonic analysis identity for $\|f\|_{L^2(\mathbb{R}^2)}^2$.

THEOREM 1.9 (from [26]). *For $f \in L^2(\mathbb{R}^2)$, the sequence $(u_n)_{n=1}^{\infty}$ defined by* (1.19) *satisfies*

$$f = \sum_{n=1}^{\infty} u_n, \tag{1.20}$$

[2] [33] included this result for the cases $f \in BV(\mathbb{R}^2)$ or f in an intermediate space between $BV(\mathbb{R}^2)$ and $L^2(\mathbb{R}^2)$. A proof requiring only $f \in L^2(\mathbb{R}^2)$ was obtained in [26].

where the convergence holds in the strong sense in $L^2(\mathbb{R}^2)$. Further,

$$\|f\|^2_{L^2(\mathbb{R}^2)} = \sum_{n=1}^{\infty} \|u_n\|^2_{L^2(\mathbb{R}^2)} + \lambda_n \|u_n\|_{TV}. \tag{1.21}$$

More insight on the scale of the decomposition $(u_n)_{n=1}^{\infty}$ can be obtained from Theorem 1.1, which states that

$$\left\| f - \sum_{i=1}^{n} u_i \right\|_* \leq \frac{\lambda_1}{2^n}. \tag{1.22}$$

Thus f and the partial sum $\sum_{i=1}^{n} u_i$ agree up to a term of scale at most $2^{-n}\lambda_1$ in the norm $\|\cdot\|_*$. As we have mentioned, according to [23, Lemma 10, section 1.14] this puts an ℓ^∞ bound on the wavelet coefficients of $f - \sum_{i=1}^{n} u_i$.

By analogy to this approach, our iterative process evolves by leaving μ untouched at each step and replacing ν with the previous iterate, ν_n; the manner in which this is analogous to (1.19) will be made precise in Remark 8.2. We describe this procedure as "adding detail" since by solving (WROF) with a large value of λ we obtain a modification of ν which is a "sketch" of μ, in that the two measures are indistinguishable up to transport at scale λ (see Theorem 1.3). By repeating this process with a smaller value of λ we refine this sketch, obtaining at each stage finer details of μ. Note also that under the additional assumption $\nu \ll \mathcal{L}_d$, Theorem 1.6 implies that we are decomposing a transport map at each stage of this procedure into "features" and "details," as determined by the scale of the transport relative to λ. Due to the soft thresholding (see (1.14)), the latter are untouched, to be resolved at future steps, while the former are partially carried out until the remaining transport becomes a detail.

Finally, we obtain in (1.25) a decomposition of the total energy $W_2^2(\mu, \nu)$ which includes all the scales of transport from ν to μ via (1.8); this is in correspondence with the identity (1.21).

The following proposition summarizes the properties of this multiscale algorithm which are not directly implied by Theorem 1.3 or Theorem 1.6. Note that we do not require $\nu \ll \mathcal{L}_d$ for these results.

THEOREM 1.10. *Let $\Omega \subset \mathbb{R}^d$ be compact and convex with a nonnegligible interior. Take $\mu, \nu \in \mathcal{P}(\Omega)$ with $\mu \ll \mathcal{L}_d$. Suppose λ_0 is given. For each $n \geq 0$, set $\lambda_{n+1} = \lambda_n/2$ and define*

$$\nu_{n+1} := \operatorname*{arg\,min}_{\rho \in \mathcal{P}(\Omega)} \frac{1}{2} W_2^2(\rho, \mu) + \lambda_n W_1(\rho, \nu_n), \tag{1.23}$$

where $\nu_0 := \nu$. We have that

1. *the sequence ν_n converges to μ with rate*

$$\frac{1}{2} W_2^2(\mu, \nu_n) \leq 2^{-2n+1} \lambda_0^2, \tag{1.24}$$

   and

2. *the following energy equality holds:*

$$\frac{1}{2} W_2^2(\nu, \mu) = \sum_{n=0}^{\infty} D_{\lambda_n}(\nu_n, \nu_{n+1}) + \lambda_n W_1(\nu_n, \nu_{n+1}). \tag{1.25}$$

Remark 1.11. If we add the assumption that ν is absolutely continuous, we obtain that the measures ν_n specified in Theorem 1.10 can be written as $(S_{\lambda_{n-1}} \circ \cdots \circ S_{\lambda_0})_{\#}\nu$; see Theorem 1.6. In this way, ν_n is built up from a composition of Wasserstein 1 optimal maps applied to ν. In this sense we are replacing the summation of the decomposition in (1.20) with composition, as was done for a multiscale decomposition of diffeomorphisms in [26].

We now describe the organization of the paper. We provide in section 2 a discussion of related work. Then, in section 3 we elucidate the connection between (WROF) and the restoration via the learned regularizer technique of [21]. In section 4 we introduce a general class of optimization problems (which includes both ROF and (WROF)) and prove Theorem 4.5 characterizing their solution maps as projections. In section 5, we use optimal transport arguments to obtain Theorem 1.3 from Theorem 4.5. In section 6 we prove that the solution to (WROF) is absolutely continuous if μ and ν are, and in section 7 we prove the existence of an optimal transport map from ν to μ under the Huber cost $c_{2,\lambda}$, as well as the soft thresholding formula (1.14). Together, sections 6 and 7 prove Theorem 1.6. Finally, the results for our iterative procedures (i.e., Proposition 1.8 and Theorem 1.10) are proved in section 8.

2. Discussion and related work. There is a connection between our iterative regularization procedure defined in Proposition 1.8 and the JKO scheme [15]. The latter is related to gradient flows in $\mathbb{W}_2(\Omega)$, which are analyzed in more detail in [1] (see also [31]). The JKO algorithm produces a sequence of measures ρ_n by iteratively solving an equation of the type

$$\rho_n := \operatorname*{arg\,min}_{\rho \in \mathcal{P}(\Omega)} \frac{1}{2\lambda} W_2^2(\rho_{n-1}, \rho) + F(\rho), \tag{2.1}$$

where F is a functional. For $F(\rho) = W_1(\rho, \nu)$, this problem is precisely (WROF). For general F, by allowing λ to go to zero and examining the optimality conditions of (2.1), one can obtain convergence of an interpolation of the iterates ρ_n to a curve of measures $\rho(t)$. This curve satisfies a PDE which can be viewed as a gradient flow on F in the metric space $\mathbb{W}_2(\Omega)$. We expect the PDE that corresponds to our iterative denoising algorithm to be of the form

$$\partial_t \rho(t) - \nabla \cdot (\rho(t) \nabla u_0(t)) = 0, \tag{2.2}$$

where for all t, $u_0(t)$ is a Kantorovich potential for $W_1(\rho(t), \nu)$. We leave the rigorous derivation to a separate paper. Note that by analogy to ROF such a flow would be in correspondence with the TV flow in [4].

Other problems of a form similar to (WROF) have been considered in the literature. A notable example is [6], which finds a smoothed version of a probability measure μ while retaining edges by solving

$$\min_{\rho \in \mathcal{P}(\Omega)} \frac{1}{2\lambda} W_2^2(\mu, \rho) + F(\rho), \quad F(\rho) := \begin{cases} \|\rho\|_{TV}, & \rho = \rho(x)dx, \\ +\infty & \text{else.} \end{cases}$$

A related problem is that of [20], which keeps F as the TV-norm of a probability density but replaces the fidelity term $\frac{1}{2\lambda} W_2^2(\mu, \rho)$ with the Kantorovich–Rubinstein norm, a quantity that is closely related to the Wasserstein 1 distance, but is able to handle measures with different mass. To our knowledge the specific problem given in (WROF) has not been treated before in the literature. Given that previous works

have used the TV-norm of a probability density function as a regularity term, we briefly compare this to our approach of using $W_1(\rho,\nu)$ in the particular case of ν as the normalized Lebesgue measure. One might imagine that for this choice of ν, $W_1(\rho,\nu)$ would serve a role similar to the TV-norm, since it is the minimal amount of work required to "smooth out" ρ to the constant function. This is not the case, however. Take $\Omega=[0,1]^2$, with $\rho=\rho_k(x)dx$ given by

$$\rho_k(x_1,x_2)=2(1+\operatorname{sign}(\sin(2\pi k x_1))).$$

As $k\to\infty$, $W_1(\rho_k,\nu)\to 0$, and yet $\|\rho_k\|_{TV}\to+\infty$. So the two regularizers play different roles.

The field of image restoration with learned regularizers is rapidly developing, and there are many interesting approaches (e.g., [12, 17, 18, 21, 27]). We focus on [21] as we found it to be a natural and compelling analogue of ROF. Note that [21] includes several theoretical results, which focus on issues such as stability of the reconstruction method and a geometric formula for the Kantorovich potential u_0 under certain conditions. Let us also note that [27], being related to iterations of the method from [21], forms a parallel approach to our iterated regularization discussed in subsection 1.2.

Last, numerical results for either of the procedures outlined in subsection 1.2 or subsection 1.3 could be obtained using the dual problem (see subsection 5.1),

$$\sup_{\varphi\in\lambda\text{-Lip}(\Omega)}\int_\Omega \varphi^{c_2}d\mu+\int_\Omega\varphi d\nu, \tag{2.3}$$

where λ-Lip(Ω) is the set of Lipschitz continuous functions on Ω with constant λ, and φ^{c_2} is the c_2 transform of φ, defined in Definition 3.1 below. Indeed, Theorem 5.6 shows that the solution ρ_λ to (WROF) can be realized by applying the solution map to (1.4) pointwise to μ, where φ_λ solves (2.3). By analogy to [11], it is natural to obtain such a φ_λ by parametrizing it with a neural network φ_w with weights w and solving the gradient penalty problem

$$\sup_w\int_\Omega \varphi_w^{c_2}(x)d\mu(x)+\int_\Omega\varphi_w(y)d\nu(y)-\frac{\lambda}{2}\int_\Omega(|\nabla\varphi_w|-\lambda)_+^2 d\sigma(x),$$

for large λ, where σ is the sampling distribution from [11]. Optimizing the weights w requires the computation of the c_2-transform of φ_w. A general and efficient numerical algorithm to do so has been introduced in [14], a method specific to neural networks has been given in [22], and a new approach which scales well to high dimensions has recently been proposed in [3].

3. Links between (WROF) and denoising by adversarial regularization. In this section we will study the relationship between (WROF) and the denoising technique of [21]. We will show in subsection 3.1 that the approach of [21] can be viewed as an explicit Euler discretization of the gradient flow on $W_1(\cdot,\nu)$ in the metric space $\mathbb{W}_2(\Omega)$. In contrast, (WROF) can be viewed as an implicit Euler discretization of the same flow on the same metric space. Moreover, we will establish in subsection 3.2 that these techniques produce identical measures under the assumption that the minimal displacement of the ray monotone optimal transport map for $W_1(\mu,\nu)$ (see [2] or section 3.1 of [30]) is larger than λ.

3.1. Explicit and implicit Euler on $\mathbb{W}_2(\Omega)$. We begin with Lemma 3.2, which states that (1.2) has a unique solution for almost all x_0. This is a standard result; we include the proof for completeness. We first recall the following definition.

DEFINITION 3.1. *For a symmetric cost function $c : \Omega \times \Omega \to \mathbb{R}$, and $\phi \in C(\Omega)$, the function*

$$\phi^c(x) = \inf_{y \in \Omega} c(x, y) - \phi(y)$$

is called the c-transform of ϕ. If ϕ is such that there exists a function ψ with $\phi = \psi^c$, then one says that ϕ is c-concave, written $\phi \in c\text{-}conc(\Omega)$.

Throughout this paper we will make use of the well-known fact that $\phi \leq \phi^{cc}$, with equality if and only if ϕ is c-concave (see, e.g., [30, Proposition 1.34]).

LEMMA 3.2. *Let Ω be compact with boundary of Lebesgue measure zero. Let $u_0 : \Omega \to \mathbb{R}$ be lower semicontinuous. Then for almost all $x \in \Omega$, the problem*

$$\min_{y \in \Omega} \frac{1}{2}|x - y|^2 + \lambda u_0(y) \tag{3.1}$$

has a unique solution given by $x - \nabla(-\lambda u_0)^{c_2}(x)$.

Proof. Since Ω is compact and u_0 is lower semicontinuous, (3.1) has a solution for all $x \in \Omega$ and the value of the minimum is finite. Compactness of Ω also implies that $(-\lambda u_0)^{c_2}$ is Lipschitz (see, for example, Box 1.8 of [30]), and thus the set of $x_0 \in \Omega \setminus \partial\Omega$ such that $\nabla(-\lambda u_0)^{c_2}(x_0)$ exists has full Lebesgue measure.

For x_0 selected in this way, let $y_0 \in \Omega$ solve (3.1). By definition, for all $x \in \Omega$,

$$(-\lambda u_0)^{c_2}(x) \leq \frac{1}{2}|x - y_0|^2 + \lambda u_0(y_0) \tag{3.2}$$

with equality at $x = x_0$. Thus, we obtain that the function $x \mapsto \frac{1}{2}|x-y_0|^2 - (-\lambda u_0)^{c_2}(x)$ is minimized at x_0. We therefore have

$$y_0 = x_0 - \nabla(-\lambda u_0)^{c_2}(x_0). \tag{3.3}$$

This expresses the minimizer y_0 of (3.1) for $x = x_0$ explicitly in terms of x_0; the minimizer is therefore unique. ∎

Lemma 3.2 implies that whenever $\mu \ll \mathcal{L}_d$ and u_0 is continuous, (3.1) has a unique solution μ almost everywhere, given by $(I - \nabla(-\lambda u_0)^{c_2})(x_0)$. The following lemma characterizes the measure we obtain if we push μ forward under this solution map.

LEMMA 3.3. *In addition to the assumptions of Lemma 3.2, let $\mu \in \mathcal{P}(\Omega)$ satisfy $\mu \ll \mathcal{L}_d$. Let T be a Borel map which coincides with $I - \nabla(-\lambda u_0)^{c_2}$, μ almost everywhere. Then the measure $T_{\#}\mu$ is the unique solution to the optimization problem*

$$\inf_{\rho \in \mathcal{P}(\Omega)} \frac{1}{2} W_2^2(\rho, \mu) + \lambda \langle u_0, \rho \rangle. \tag{3.4}$$

Proof. First we note that the map

$$\rho \mapsto \frac{1}{2} W_2^2(\rho, \mu) + \lambda \langle u_0, \rho \rangle \tag{3.5}$$

is strictly convex by Theorem 7.19 from [30], which holds since μ is absolutely continuous and $\mathcal{L}_d(\partial\Omega) = 0$. Thus, if a solution ρ_0 to (3.4) exists it is unique. A measure ρ_0 is a minimizer of (3.4) if and only if

$$0 \in \partial \left(\frac{1}{2} W_2^2(\cdot, \mu) + \langle \lambda u_0, \cdot \rangle \right)(\rho_0).$$

Since $\rho \mapsto \langle u_0, \rho \rangle$ is linear, this is equivalent to

$$-\lambda u_0 \in \partial \left(\frac{1}{2} W_2^2(\cdot, \mu) \right)(\rho_0).$$

By Proposition 7.17 of [30], which characterizes the subdifferential of the convex function $\rho \mapsto \frac{1}{2} W_2^2(\rho, \mu)$, we conclude that ρ_0 is a minimizer of (3.4) if and only if

$$\int_\Omega (-\lambda u_0)^{c_2} d\mu + \int_\Omega (-\lambda u_0) d\rho_0 = \frac{1}{2} W_2^2(\mu, \rho_0). \tag{3.6}$$

This equality will be proved in in Lemma 5.4 for $\rho_0 = T_\# \mu$, when T is μ almost everywhere equal to $I - \nabla(-\lambda u_0)^{c_2}$. ∎

Remark 3.4. We note that Lemma 3.3 describes the distribution one obtains by applying the denoising technique from [21] pointwise to an absolutely continuous distribution μ. Indeed, that procedure consists of solving (3.1) given x when u_0 is a Kantorovich potential for $W_1(\mu, \nu)$. It is interesting to observe that while the denoising technique of [21] applied to a specific image x_0 amounts to an implicit Euler scheme on a Kantorovich potential u_0, Lemma 3.3 shows that the distribution one thus obtains on all images is characterized as an explicit Euler step on the functional $W_1(\cdot, \nu)$; this holds since such a u_0 is a subgradient of this functional evaluated at μ. Implicit Euler discretizations are often better behaved, motivating us to replace (1.3) with (WROF).

3.2. Equivalence of (WROF) and denoising by adversarial regularization. Here we will show that that under the assumption that λ is less than the minimal transport length for the ray monotone Wasserstein 1 transport from μ to ν, the solution to (WROF) and the measure obtained via the technique of [21] are actually the same.

PROPOSITION 3.5. *Suppose that $\Omega \subset \mathbb{R}^d$ is compact and convex and that $\mu, \nu \in \mathcal{P}(\Omega)$. Suppose that $\mu \ll \mathcal{L}_d$ and that $\lambda > 0$ satisfies*

$$\operatorname*{ess\,inf}_\mu |x - T_0(x)| > \lambda, \tag{3.7}$$

where T_0 is the unique ray monotone optimal transport map for $W_1(\mu, \nu)$. Take $u_0 \in 1\text{-}Lip(\Omega)$ a Kantorovich potential for $W_1(\mu, \nu)$, and let T be a Borel map equal to $I - \nabla(-\lambda u_0)^{c_2}$ μ almost everywhere. Then $\rho_\lambda := T_\# \mu$ is the unique solution to (WROF).

Proof. Since Ω is convex we immediately obtain that $\mathcal{L}_d(\partial\Omega) = 0$ (see, for example, [19]). Thus, $\mu \ll \mathcal{L}_d$ implies that the functional in (WROF) is strictly convex (via [30, Theorem 7.19] again), and so the solution to (WROF) is unique if it exists. Next, we claim that if $\rho_0 \in \mathcal{P}(\Omega)$ and there exists $\varphi_0 \in \lambda\text{-Lip}(\Omega)$ such that

$$\int_\Omega \varphi_0 d\nu - \int_\Omega \varphi_0 d\rho_0 = \lambda W_1(\rho_0, \nu), \tag{3.8}$$

and

$$\int_\Omega \varphi_0^{c_2} d\mu + \int_\Omega \varphi_0 d\rho_0 = \frac{1}{2} W_2^2(\mu, \rho_0), \tag{3.9}$$

then ρ_0 solves (WROF). Indeed, by Proposition 7.17 from [30], assumptions (3.8) and (3.9) imply that

$$-\varphi_0 \in \partial \left(\lambda W_1(\cdot, \nu) \right)(\rho_0), \quad \varphi_0 \in \partial \left(\frac{1}{2} W_2^2(\cdot, \mu) \right)(\rho_0).$$

As such,

$$\begin{aligned}0 &= \varphi_0 - \varphi_0 \\ &\in \partial\left(\frac{1}{2}W_2^2(\cdot,\mu)\right)(\rho_0) + \partial\left(\lambda W_1(\cdot,\nu)\right)(\rho_0) \\ &\subset \partial\left(\frac{1}{2}W_2^2(\cdot,\mu) + \lambda W_1(\cdot,\nu)\right)(\rho_0),\end{aligned}$$

and thus ρ_0 solves (WROF), proving the claim.

Now we assert that these conditions hold for $\rho_\lambda := T_\# \mu$ and $\varphi_0 := -\lambda u_0$. First, we note that by Proposition 9 from [25] and Lemma 3.2, assumption (3.7) implies that

$$I - \nabla(-\lambda u_0)^{c_2}(x) = I - \lambda \nabla u_0(x) \tag{3.10}$$

μ almost everywhere. Next, observe that convexity of Ω, together with (3.7) and standard properties of Wasserstein 1 Kantorovich potentials, implies that $\rho_\lambda \in \mathcal{P}(\Omega)$. Also, by Theorem 1(i) of [24], u_0 is a Kantorovich potential for $W_1(\rho_\lambda, \nu)$. As such, $\varphi_0 = -\lambda u_0$ satisfies $\varphi_0 \in \lambda\text{-Lip}(\Omega)$ and (3.8). Finally, (3.9) is given by Lemma 5.4, since $\rho_\lambda = T_\# \mu$, and by definition $T = I - (-\lambda \nabla u_0)^{c_2}$ μ almost everywhere. ∎

The link between (WROF) and the denoising method of [21] having been established, we now analyze solutions of (WROF).

4. A class of minimization problems with solutions given by projections. In this section we will prove a general theorem about the minimization of a certain class of convex functions, establishing that the solution map is equivalent to a projection. We will show that ROF (see (1.1)) and (WROF) are examples of this class of problems. Thus, we can apply this general theorem to yield Theorem 1.1 and, with additional arguments from optimal transport, our Theorem 1.3. This puts ROF and (WROF) within a common framework and provides a fruitful analogy in what follows.

Let X be a Hausdorff locally convex topological vector space,[3] and take X^* as its continuous dual; in general we will denote by x and x^* points in X and X^*, respectively. Let $F : X \to \mathbb{R}$ be a proper lower semicontinuous convex functional. Recall that the Legendre dual of such a function is given by $F^* : X^* \to \mathbb{R} \cup \{+\infty\}$,

$$F^*(x^*) := \sup_{x \in X} \langle x, x^* \rangle - F(x),$$

with $\langle \cdot, \cdot \rangle$ denoting the duality pairing, and set

$$\operatorname{dom}(F^*) = \{x^* \in X^* \mid F^*(x^*) < +\infty\}.$$

When studying the subdifferential of F^* we will restrict the dual of X^* to $X \subset X^{**}$, i.e.,

$$\partial F^*(x^*) := \{x \in X \mid \forall y^* \in X^*, F^*(y^*) \geq F^*(x^*) + \langle x, y^* - x^* \rangle\}.$$

We will focus on F which are in fact continuous, and such that F^* is a strictly convex function. Take $K \subset X$ as a closed, convex, nonempty set satisfying $K = -K$ and

[3] We will not need this amount of generality for our applications, but we phrase our theorem in this setting to indicate that nothing more is needed.

let 1_K denote the indicator function of K. For $y_0^* \in X^*$, consider the optimization problem

$$\min_{x^* \in X^*} F^*(x^*) + 1_K^*(x^* - y_0^*). \tag{4.1}$$

To motivate the analysis of such problems, we will now indicate that both ROF and (WROF) are examples.

Example 4.1 (ROF). Take $X = L^2(\mathbb{R}^2)$, and $F : L^2(\mathbb{R}^2) \to \mathbb{R}$ as

$$F(u) = \frac{1}{2}\|u\|^2_{L^2(\mathbb{R}^2)} + \langle f, u\rangle.$$

This functional is obviously continuous and convex. It is a simple exercise to show that its dual is

$$F^*(u) = \frac{1}{2}\|u - f\|^2_{L^2(\mathbb{R}^2)},$$

which is strictly convex. Take the set K as

$$K := \{v \in L^2(\mathbb{R}^2) \mid \|v\|_* \le \lambda\}.$$

It is clear that K is convex, $K = -K$, and K is closed. It is also not difficult to show that

$$1_K^*(u) := \sup_{v \in K} \int_{\mathbb{R}^2} vu dx = \lambda \|u\|_{TV}.$$

Thus, we see that ROF (i.e., (1.1)) is an example of (4.1), with $y_0^* = 0$.

Example 4.2 (WROF). Assume $\Omega \subset \mathbb{R}^d$ is compact and convex (and therefore $\mathcal{L}_d(\partial\Omega) = 0$). Let $X = C(\Omega)$ with the topology induced by the sup norm. Then $X^* = \mathcal{M}(\Omega)$, the set of finite signed Borel measures on Ω. Let $\mu \in \mathcal{P}(\Omega)$ with $\mu \ll \mathcal{L}_d$, and take $F : C(\Omega) \to \mathbb{R}$ as the functional

$$F(\varphi) := -\int_\Omega \varphi^{c_2} d\mu.$$

It is shown in the proof of Proposition 7.17 of [30] that F defined in this way is convex and continuous, and that F^* satisfies, for $\rho \in \mathcal{M}(\Omega)$,

$$F^*(\rho) = \begin{cases} \frac{1}{2}W_2^2(\rho, \mu), & \rho \in \mathcal{P}(\Omega), \\ +\infty & \text{else.} \end{cases}$$

Further, Proposition 7.19 of [30] proves that F^* is strictly convex when $\mu \ll \mathcal{L}_d$.

Now take $K = \lambda\text{-Lip}(\Omega)$. This set is convex and closed in $C(\Omega)$ and satisfies $K = -K$. In addition, for $\nu, \rho \in \mathcal{P}(\Omega)$, we have

$$\begin{aligned} 1_K^*(\rho - \nu) &= \sup_{\varphi \in \lambda\text{-Lip}(\Omega)} \langle \varphi, \rho - \nu\rangle \\ &= \lambda W_1(\rho, \nu). \end{aligned}$$

Thus, (WROF) is of the form (4.1).

For our analysis of (4.1), we find it natural to define the divergence $D : \text{dom}(F^*) \times \text{dom}(F^*) \to \mathbb{R} \cup \{+\infty\}$ by

$$D(y^*, x^*) := F^*(y^*) - F^*(x^*) - \sup_{x \in \partial F^*(x^*) \cap K} \langle x, y^* - x^*\rangle. \tag{4.2}$$

The following lemma shows that D has properties similar to those of a Bregman divergence.

LEMMA 4.3. *For all $y^*, x^* \in dom(F^*)$, the functional D satisfies*

$$D(y^*, x^*) \geq 0.$$

Moreover, if F^* is strictly convex, then $D(y^*, x^*) = 0$ if and only if $\partial F^*(x^*) \cap K \neq \emptyset$ and $y^* = x^*$.

Proof. The claim $D(y^*, x^*) \geq 0$ clearly holds if $\partial F^*(x^*) \cap K = \emptyset$. On the other hand, if $\partial F^*(x^*) \cap K \neq \emptyset$ the definition of the subdifferential of F^* confirms that $D(y^*, x^*) \geq 0$. Clearly, if $\partial F^*(x^*) \cap K \neq \emptyset$ and $y^* = x^*$ we have $D(y^*, x^*) = 0$. On the other hand, let F^* be strictly convex. If $D(y^*, x^*) = 0$, then take $\epsilon > 0$ and $x_\epsilon \in \partial F^*(x^*) \cap K$ such that

$$\sup_{x \in \partial F^*(x^*) \cap K} \langle x, y^* - x^* \rangle - \epsilon \leq \langle x_\epsilon, y^* - x^* \rangle.$$

Since $D(y^*, x^*) = 0$, we therefore obtain

$$F^*(y^*) \leq F^*(x^*) + \langle x_\epsilon, y^* - x^* \rangle + \epsilon.$$

Hence, for $t \in [0, 1]$,

$$\begin{aligned} F^*((1-t)x^* + ty^*) &\leq (1-t)F^*(x^*) + tF^*(y^*) \\ &\leq (1-t)F^*(x^*) + tF^*(x^*) \\ &\quad + \langle x_\epsilon, (1-t)x^* + ty^* - x^* \rangle + t\epsilon \\ &\leq F^*((1-t)x^* + ty^*) + t\epsilon. \end{aligned}$$

Since ϵ is arbitrary, we obtain that F^* is affine on the segment $[x^*, y^*]$, a contradiction to strict convexity unless $x^* = y^*$. ∎

Example 4.4. Let us determine D is in the context of ROF. Recall that in this case, $F^*(u) = \frac{1}{2}\|u - f\|^2_{L^2(\mathbb{R}^2)}$. Then $\partial F^*(u)$ is a singleton, given by $\{u - f\}$. So $D(v, u) = +\infty$ unless $\|u - f\|_* \leq \lambda$. In that case,

$$\begin{aligned} D(v, u) &= \frac{1}{2}\|v - f\|^2_{L^2(\mathbb{R}^2)} - \frac{1}{2}\|u - f\|^2_{L^2(\mathbb{R}^2)} - \langle u - f, v - u \rangle \\ &= \frac{1}{2}\|u - v\|^2_{L^2(\mathbb{R}^2)}. \end{aligned}$$

The description of D in the context of (WROF) will be given in subsection 5.1.1.

For a nonempty convex set $K \subset X$ satisfying $K = -K$, define the seminorm $\|\cdot\|_K : X \to \mathbb{R} \cup \{+\infty\}$ given by

$$\|x\|_K = \inf \left\{ t > 0 \mid \frac{x}{t} \in K \right\}.$$

We can now state the main result of this section, which provides conditions under which the solution to (4.1), if it exists, can be expressed as a projection in the divergence D onto the set of x^* such that $\partial F^*(x^*) \cap K \neq \emptyset$.

THEOREM 4.5. *Suppose that X is a Hausdorff locally convex topological vector space, with X^* as its dual. Assume $F : X \to \mathbb{R}$ is continuous and convex, and that its dual F^* is strictly convex. Let $K \subset X$ be a closed, convex, nonempty set satisfying $K = -K$. Suppose that $y_0^* \in dom(F^*)$, and the problem*

$$\sup_{x \in K} \langle x, y_0^* \rangle - F(x), \tag{4.3}$$

has a solution x_0. Then

a. (4.1) *has a unique solution x_0^* given by the single element of $\partial F(x_0)$,*

b. *x_0^* is also a solution to*

$$\min_{F^*(x^*)\cap K\neq\emptyset} D(y_0^*, x^*), \tag{4.4}$$

and

c. *the values of* (4.1), (4.3), *and $F^*(y_0^*) - D(y_0^*, x_0^*)$ coincide.*

Given b, *we obtain the following dichotomy:*

1. *If $\partial F^*(y_0^*)\cap K\neq\emptyset$, then $x_0^* = y_0^*$.*
2. *Otherwise, $x_0^*\neq y_0^*$, and any solution x_0 to* (4.3) *satisfies $x_0\in\partial F^*(x_0^*)$, $\|x_0\|_K = 1$, and*

$$\langle x_0, y_0^* - x_0^*\rangle = 1_K^*(x_0^* - y_0^*). \tag{4.5}$$

Remark 4.6. Let the solution map to (4.1) as a function of y_0^* be denoted P. Then the dichotomy presented in Theorem 4.5 confirms that $P(P(y_0^*)) = P(y_0^*)$; i.e., P is a projection.

Before proving Theorem 4.5, we show how it yields Theorem 1.1.

Proof of Theorem 1.1. Recalling Examples 4.1 and 4.4, we have that (1.1) is of the form (4.1) for $y_0^* = 0$, and

$$F(u) = \frac{1}{2}\|u\|^2_{L^2(\mathbb{R}^2)} + \langle f, u\rangle, F^*(u) = \frac{1}{2}\|u - f\|^2_{L^2(\mathbb{R}^2)},$$

$$K := \{v\in L^2(\mathbb{R}^2) \mid \|v\|_* \leq \lambda\},$$

$$D(v,u) = \begin{cases} \frac{1}{2}\|u - v\|^2_{L^2(\mathbb{R}^2)}, & \|u - f\|_* \leq \lambda, \\ +\infty, & \|u - f\|_* > \lambda. \end{cases}$$

The problem (4.3) therefore takes the form

$$\max_{\|v\|_*\leq\lambda} -\frac{1}{2}\|v\|^2_{L^2(\mathbb{R}^2)} - \langle v, f\rangle.$$

This problem has a unique solution $\tilde{v}_\lambda$ since K is convex, nonempty, and closed, and the function $v\mapsto \frac{1}{2}\|v\|^2_{L^2(\mathbb{R}^2)} + \langle v, f\rangle$ is continuous, strictly convex, and coercive on $L^2(\mathbb{R}^2)$. We may therefore apply Theorem 4.5 to obtain that (1.1) has a unique solution given by

$$u_\lambda = f + \tilde{v}_\lambda.$$

Given our calculation for D in Example 4.4, we obtain that u_λ is also a solution of the problem in (4.4), which is

$$\min_{\|u-f\|_*\leq\lambda} \|u\|^2_{L^2(\mathbb{R}^2)}.$$

Thus, if $\|f\|_*\leq\lambda$, it is clear that $u_\lambda = 0$. On the other hand, $\|f\|_* > \lambda$ if and only if $\partial F^*(0)\cap K = \emptyset$. Using Theorem 4.5, we obtain that $\tilde{v}_\lambda\in K$ satisfies $\|\tilde{v}_\lambda\|_K = 1$, and

$$\int_{\mathbb{R}^2} u_\lambda(f - u_\lambda)dx = \langle\tilde{v}_\lambda, -u_\lambda\rangle = \lambda\|u_\lambda\|_{TV}.$$

Finally, we compute $\|v\|_K = \|v\|_*/\lambda$. As such, $\|\tilde{v}_\lambda\|_K = 1$ is equivalent to $\|\tilde{v}_\lambda\|_* = \lambda$, and the proof is complete. ∎

Proof of Theorem 4.5. We start by proving statement a. We note that this result is obtainable using Theorem 2.7.1 of [35], but we provide an elementary proof here. Let x_0 be a solution to (4.3). Then, equivalently, x_0 solves

$$\min_{x\in X} F(x) - \langle x, y_0^*\rangle + 1_K(x).$$

Noting that $x \mapsto F(x) - \langle x, y_0^*\rangle$ and $1_K(x)$ are both proper convex functions, and that the former is finite and continuous, we can apply part (iii) of Theorem 2.8.7 from [35] to conclude that

$$\partial\left(F - \langle\cdot, y_0^*\rangle + 1_K\right) = \partial F - \{y_0^*\} + \partial 1_K. \tag{4.6}$$

Since F^* is strictly convex, $\partial F(x_0)$ contains at most one element. Further, since F is convex, proper, and continuous, Theorem 2.4.9 from [35] shows that $\partial F(x)$ is nonempty for all $x \in X$, and thus $\partial F(x)$ contains a unique element for all $x \in X$. Since x_0 is a solution of (4.3), the unique element $x_0^* \in \partial F(x_0)$ satisfies

$$0 \in x_0^* - y_0^* + \partial 1_K(x_0).$$

Since $K = -K$, we have $\partial 1_K(-x_0) = -\partial 1_K(x_0)$. Thus,

$$x_0^* - y_0^* \in \partial 1_K(-x_0). \tag{4.7}$$

Next, since $H : X \to \mathbb{R}\cup\{+\infty\}$ is proper, convex, and lower semicontinuous, then for all x such that $H(x) < +\infty$ we have the well-known fact that[4]

$$x^* \in \partial H(x) \Leftrightarrow x \in \partial H^*(x^*) \Leftrightarrow H(x) + H^*(x^*) = \langle x, x^*\rangle. \tag{4.8}$$

We apply this to $H = 1_K$, which is proper because K is nonempty, convex because K is convex, and lower semicontinuous because K is closed. Thus, (4.7) yields

$$-x_0 \in \partial 1_K^*(x_0^* - y_0^*).$$

It is an elementary fact that $(\partial H(\cdot - y^*))(x^*) = \partial H(x^* - y^*)$. Recalling that $x_0^* \in \partial F(x_0)$ and using (4.8) again, we get

$$0 \in \partial F^*(x_0^*) + \partial 1_K^*(x_0^* - y_0^*) \subset \partial\left(F^* + 1_K^*(\cdot - y_0^*)\right)(x_0^*),$$

which confirms that x_0^* is a minimizer of (4.1). By the assumed strict convexity of F^*, x_0^* is the unique minimizer. We have shown that if x_0 solves (4.3), then the unique $x_0^* \in \partial F(x_0)$ solves (4.1), which proves statement a.

Statements b and c will be proven together. Regarding the values of (4.1) and (4.3), note that by definition of the Legendre dual, for all $x^* \in X^*$ and $x \in K$,

$$\begin{aligned} F^*(x^*) + 1_K^*(x^* - y_0^*) &\geq \langle x, x^*\rangle - F(x) + \langle -x, x^* - y_0^*\rangle \\ &= \langle x, y_0^*\rangle - F(x). \end{aligned}$$

Hence,

$$\inf_{x^*\in X^*} F^*(x^*) + 1_K^*(x^* - y_0^*) \geq \sup_{x\in K} \langle x, y_0^*\rangle - F(x).$$

[4]See, for example, Theorem 2.4.4 from [35] for a proof of this in our setting.

On the other hand, for x_0 optimal in (4.3) and $x_0^* \in \partial F(x_0)$ optimal in (4.1), (4.8) implies

$$\begin{aligned} F^*(x_0^*) + 1_K^*(x_0^* - y_0^*) &= \langle x_0, x_0^* \rangle - F(x_0) + \langle -x_0, x_0^* - y_0^* \rangle \\ &= \langle x_0, y_0^* \rangle - F(x_0). \end{aligned}$$

This establishes that the values of (4.1) and (4.3) are the same.

Next, we turn to (4.4). Invoking (4.8) again, and recalling that $\partial F(x)$ contains a unique element for all $x \in X$, we obtain that for each $x \in K$, there exists $x^* \in \partial F(x)$ such that $\partial F^*(x^*) \cap K \neq \emptyset$. As such,

$$\begin{aligned} \langle x, y_0^* \rangle - F(x) &= \langle x, y_0^* - x^* \rangle + F^*(x^*) \\ &= F^*(y_0^*) - (F^*(y_0^*) - F^*(x^*) - \langle x, y_0^* - x^* \rangle) \\ &\leq F^*(y_0^*) - (F^*(y_0^*) - F^*(x^*) - \sup_{z \in \partial F^*(x^*) \cap K} \langle z, y_0^* - x^* \rangle. \end{aligned}$$

Thus,

$$\sup_{x \in K} \langle x, y_0^* \rangle - F(x) \leq F(y_0^*) - \inf_{x^* \in X^*} D(y_0^*, x^*). \tag{4.9}$$

On the other hand, for $x^* \in X^*$ with $\partial F^*(x^*) \cap K \neq \emptyset$, let $\epsilon > 0$ and take $x \in \partial F^*(x^*) \cap K$ satisfying

$$\sup_{z \in \partial F^*(x^*) \cap K} \langle z, y_0^* - x^* \rangle - \epsilon \leq \langle x, y_0^* - x^* \rangle.$$

Then

$$\begin{aligned} F(y_0^*) - D(y_0^*, x^*) &\leq F(y_0^*) - (F^*(y_0^*) - F^*(x^*) - \langle x, y_0^* - x^* \rangle - \epsilon) \\ &= \langle x, y_0^* \rangle - F(x) + \epsilon. \end{aligned}$$

Hence,

$$\sup_{x \in K} \langle x, y_0^* \rangle - F(x) + \epsilon \geq F(y_0^*) - \inf_{x^* \in X^*} D(y_0^*, x^*).$$

Since ϵ is arbitrary, we obtain equality of the values of $F(y_0^*) - \inf_{x^*} D(y_0^*, x^*)$ and (4.3). Finally, if x_0 solves (4.3), then we know that $x_0^* \in \partial F(x_0)$ solves (4.1). We also have

$$\begin{aligned} \sup_{x \in K} \langle x, y_0^* \rangle - F(x) &= \langle x_0, y_0^* \rangle - F(x_0) \\ &\leq F^*(y_0^*) - D(y_0^*, x_0^*) \\ &\leq F^*(y_0^*) - \inf_{x^* \in \mathrm{dom}(F^*)} D(y_0^*, x^*) \\ &= \sup_{x \in K} \langle x, y_0^* \rangle - F(x). \end{aligned}$$

Equality of the first expression and the last mean that each inequality is an equality; thus x_0^* solves (4.4) as claimed, which completes the proof of statements b and c.

We now address the dichotomy. If $\partial F^*(y_0^*) \cap K \neq \emptyset$, then by Lemma 4.3 the only minimizer of (4.4) is y_0^*. So suppose $\partial F^*(y_0^*) \cap K = \emptyset$. Then it is clear that $x_0^* \neq y_0^*$, since $D(y_0^*, x_0^*) < +\infty$, and hence $\partial F^*(x_0^*) \cap K \neq \emptyset$. For any solution x_0 to (4.3),

we obtain $x_0 \in \partial F^*(x_0^*)$ by (4.8). We also have (4.7), and thus via the last equality of (4.8),

$$\langle -x_0, x_0^* - y_0^* \rangle = 1_K^*(x_0^* - y_0^*),$$

which is (4.5). Since $x_0 \in K$, we have $\|x_0\|_K \leq 1$. On the other hand, if $\|x_0\|_K < 1$, then since $x_0^* \neq y_0^*$ there is no possibility of (4.5) holding. ∎

5. Proof of Theorem 1.3. In subsection 5.1 we will demonstrate that the hypotheses of Theorem 4.5 hold for (WROF). In addition, we will describe the divergence D in this context. In subsection 5.2 we will use these preliminaries to complete the proof of Theorem 1.3.

5.1. Preliminaries. Recall that in the context of (WROF), $K = \lambda\text{-Lip}(\Omega)$, and $F : C(\Omega) \to \mathbb{R}$ given by

$$F(\varphi) = -\int_\Omega \varphi^{c_2} d\mu.$$

We mentioned in Example 4.2 that F defined in this way is convex and continuous, and that F^* is strictly convex provided $\mu \ll \mathcal{L}_d$ and $\mathcal{L}_d(\partial\Omega) = 0$. Further, $K = \lambda\text{-Lip}(\Omega)$ is closed, convex, and nonempty and satisfies $K = -K$. The only remaining hypothesis of Theorem 4.5 to verify is that (4.3) has a solution. In this setting (4.3) takes the form

$$\sup_{\varphi \in \lambda\text{-Lip}(\Omega)} \int_\Omega \varphi^{c_2} d\mu + \int_\Omega \varphi d\nu. \tag{5.1}$$

The existence of a solution could be proved by standard arguments, but we will do so by rewriting (5.1) as an unconstrained problem in terms of the Huber cost function $c_{2,\lambda}$ (see (1.7) for the definition); this will be useful to us later.

For $\mu, \rho \in \mathcal{P}(\Omega)$, let $\mathcal{I}_{c_{2,\lambda}}(\mu, \rho)$ be the transport cost, i.e.,

$$\mathcal{I}_{c_{2,\lambda}}(\mu, \rho) := \inf_{\gamma \in \Pi(\mu,\rho)} \int_{\Omega\times\Omega} c_{2,\lambda}(x, y) d\gamma(x, y),$$

where $\Pi(\mu, \rho)$ is the set of probability distributions on $\Omega \times \Omega$ with marginal distributions given by μ and ρ. Since $c_2(x, y) \geq c_{2,\lambda}(x, y)$, we have $\frac{1}{2}W_2^2(\mu, \rho) \geq \mathcal{I}_{c_{2,\lambda}}(\mu, \rho)$. We now prove an easy fact about the $c_{2,\lambda}$-transform.

LEMMA 5.1. *If Ω is convex and $\varphi \in \lambda$-Lip(Ω), then*

$$\varphi^{c_{2,\lambda}}(x) = \varphi^{c_2}(x) \quad \forall x \in \Omega, \tag{5.2}$$

where, recall,

$$\varphi^{c_{2,\lambda}}(x) := \inf_{y\in\Omega} c_{2,\lambda}(x, y) - \varphi(y).$$

Proof. Let $B_\lambda(x)$ be the closed Euclidean ball of radius λ centered at x. We claim that $\varphi \in \lambda\text{-Lip}(\Omega)$ implies that for all $x \in \Omega$,

$$\varphi^{c_2}(x) = \inf_{y\in\Omega\cap B_\lambda(x)} c_2(x, y) - \varphi(y). \tag{5.3}$$

Indeed, for $y \in \Omega \setminus B_\lambda(x)$, let z be the projection of y onto $B_\lambda(x)$ in the Euclidean norm; note that $z \in \Omega$ by convexity of Ω. Since c_2 is convex and differentiable in its second variable, we have

$$c_2(x,y) \geq c_2(x,z) + \langle z - x, y - z\rangle = c_2(x,z) + \lambda|z-y|,$$

so

$$c_2(x,y) - \varphi(y) \geq c_2(x,z) + \lambda|z-y| - \varphi(y) \geq c_2(x,z) - \varphi(z).$$

This proves (5.3). We can also prove, by a nearly identical argument, that the infimum in

$$\varphi^{c_{2,\lambda}}(x) = \inf_{y\in\Omega} c_{2,\lambda}(x,y) - \varphi(y)$$

can also be restricted to $B_\lambda(x) \cap \Omega$. Since $c_{2,\lambda}(x,y) = c_2(x,y)$ when $|x-y| \leq \lambda$, the conclusion follows. ∎

An immediate consequence of the preceding lemma is that we can rewrite (5.1) as an unconstrained problem in terms of the cost $c_{2,\lambda}$.

LEMMA 5.2. *When Ω is convex, the problems* (5.1) *and*

$$\sup_{\varphi\in C(\Omega)} \int_\Omega \varphi^{c_{2,\lambda}} d\mu + \int_\Omega \varphi d\nu \tag{5.4}$$

are equivalent; they have the same value, a solution to (5.1) *is a solution to* (5.4)*, and a $c_{2,\lambda}$-concave solution to* (5.4) *is a solution to* (5.1).

Proof. Let $\varphi \in C(\Omega)$ be a candidate for maximizing (5.4). Without loss of generality we may take φ $c_{2,\lambda}$-concave, and since $c_{2,\lambda}(x,y) = h(|x-y|)$ for $h \in \lambda\text{-Lip}(\mathbb{R}^+)$, we obtain $\varphi \in \lambda\text{-Lip}(\Omega)$ as well. So (5.4) can be rewritten as

$$\sup_{\varphi\in\lambda\text{-Lip}(\Omega)} \int_\Omega \varphi^{c_{2,\lambda}} d\mu + \int_\Omega \varphi d\nu.$$

We have already shown in Lemma 5.1 that when Ω is convex and $\varphi \in \lambda\text{-Lip}(\Omega)$, $\varphi^{c_{2,\lambda}} = \varphi^{c_2}$. This establishes the equivalence of the problems. ∎

The existence of a solution to (5.1) now follows from the existence of a Kantorovich potential for the transport problem $\mathcal{I}_{c_{2,\lambda}}(\mu,\nu)$.

LEMMA 5.3. *Let $\Omega \subset \mathbb{R}^n$ be compact and convex. For $\mu,\nu \in \mathcal{P}(\Omega)$, problem* (5.1) *has a solution.*

Proof. Since the cost $c_{2,\lambda}$ is uniformly continuous and bounded on $\Omega\times\Omega$, we may use Theorem 1.39 of [30] to conclude that there exists a $c_{2,\lambda}$-concave function φ_λ such that

$$\begin{aligned}\mathcal{I}_{c_{2,\lambda}}(\mu,\nu) &= \sup_{\varphi\in C(\Omega)} \int_\Omega \varphi^{c_{2,\lambda}} d\mu + \int_\Omega \varphi d\nu \\ &= \int_\Omega \varphi_\lambda^{c_{2,\lambda}} d\mu + \int_\Omega \varphi_\lambda d\nu.\end{aligned}$$

By Lemma 5.2, φ_λ is a solution of (5.1). ∎

The hypotheses of Theorem 4.5 being validated, we may apply it to (WROF), and we will do so in subsection 5.2. As a preliminary step, however, it will be helpful to specify the subdifferential of F, since the minimizer of (WROF) will be given by $\partial F(\varphi_\lambda)$ if φ_λ solves (5.1).

LEMMA 5.4. *For $\varphi \in C(\Omega)$, $\partial F(\varphi)$ is nonempty, and*

$$\partial F(\varphi) = \left\{ \rho \in \mathcal{P}(\Omega) \mid \int_\Omega \varphi^{c_2} d\mu + \int_\Omega \varphi d\rho = \frac{1}{2} W_2^2(\rho, \mu) \right\}. \tag{5.5}$$

Further, if $\partial\Omega$ has Lebesgue measure 0, $\mu \ll \mathcal{L}_d$, and $T : \Omega \to \Omega$ is any Borel map μ almost everywhere equal to $I - \nabla\varphi^{c_2}$, then

$$\partial F(\varphi) = \{T_\# \mu\}.$$

Proof. Since F is convex, proper, and continuous everywhere, Theorem 2.4.9 from [35] shows that $\partial F(\varphi)$ is nonempty for all φ. Since F is a convex, proper, and continuous function, we invoke (4.8) to state that

$$\rho \in \partial F(\varphi) \Leftrightarrow \varphi \in \partial F^*(\rho) = \partial \left(\frac{1}{2} W_2^2(\cdot, \mu) \right)(\rho).$$

Via Proposition 7.17 from [30], we obtain (5.5).

Thus, $\rho \in \partial F(\varphi)$ means that φ^{c_2} is a Kantorovich potential for $W_2(\mu, \rho)$. Suppose in addition $\partial\Omega$ has Lebesgue measure 0 and $\mu \ll \mathcal{L}_d$. The characterization of the optimal transport map for the cost $c_2(x, y) = \frac{1}{2}|x - y|^2$ in Theorem 1.17 of [30] then confirms that $\rho = T_\# \mu$ for any T μ almost everywhere equal to $I - \nabla\varphi^{c_2}$. ∎

It will also be useful to study the divergence D in the context of (WROF), specifically where it is finite. This is the content of the next subsection.

5.1.1. The divergence D in the context of (WROF). Here we will provide a characterization of the set of measures ρ such that $D(\nu, \rho) < +\infty$. We will also provide an economic interpretation of D on this set.

First, set $B_\lambda(\mu)$ as the set of all measures ρ that are reachable from μ under an optimal plan for the cost $c_{2,\lambda}$ such that no point moves more than distance λ,

$$B_\lambda(\mu) = \{\rho \in \mathcal{P}(\Omega) \mid \exists \gamma_0 \text{ optimal for } \mathcal{I}_{c_{2,\lambda}}(\mu, \rho) \text{ s.t. } \mathrm{spt}(\gamma_0) \subset \{|x - y| \le \lambda\}\}. \tag{5.6}$$

We consider $B_\lambda(\mu)$ because the following lemma shows that it is exactly the set of $\rho \in \mathcal{P}(\Omega)$ such that $\partial F^*(\rho) \cap K \neq \emptyset$, and thus $D(\nu, \rho) < +\infty$. In particular it is the set of measures ρ such that $W_2^2(\rho, \mu)$ has an λ-Lipschitz Kantorovich potential. We also provide a third characterization of $B_\lambda(\mu)$ as the set of all measures which are close enough to μ that there are no savings to be had using the discounted cost $c_{2,\lambda}$.

LEMMA 5.5. *Let Ω be compact and convex. Then the following are equivalent:*

1. *$\rho \in B_\lambda(\mu)$,*
2. *$\partial(\frac{1}{2} W_2^2(\cdot, \mu))(\rho) \cap \lambda\text{-}Lip(\Omega) \neq \emptyset$, and*
3. *$\frac{1}{2} W_2^2(\mu, \rho) = \mathcal{I}_{c_{2,\lambda}}(\mu, \rho)$.*

Proof. We will proceed by proving $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 1$. Let $\rho \in B_\lambda(\mu)$. Let φ be a $c_{2,\lambda}$-concave function such that

$$\mathcal{I}_{c_{2,\lambda}}(\mu, \rho) = \int_\Omega \varphi^{c_{2,\lambda}} d\mu + \int_\Omega \varphi d\rho.$$

Since φ is $c_{2,\lambda}$-concave we obtain that $\varphi \in \lambda\text{-Lip}(\Omega)$. For γ_0 the optimal plan transporting μ to ρ from the definition of $B_\lambda(\mu)$, we have

$$\begin{aligned}\frac{1}{2}W_2^2(\mu,\rho) &\le \frac{1}{2}\int_{\Omega\times\Omega}|x-y|^2 d\gamma_0\\ &= \int_{\Omega\times\Omega} c_{2,\lambda}(x,y)d\gamma_0\\ &= \int_\Omega \varphi^{c_{2,\lambda}}d\mu + \int_\Omega \varphi d\rho\\ &= \int_\Omega \varphi^{c_2}d\mu + \int_\Omega \varphi d\rho\\ &\le \frac{1}{2}W_2^2(\mu,\rho).\end{aligned}$$

In the second to last line we have used Lemma 5.1. Equality of the first and last terms means we have equality throughout, and thus $\varphi \in \partial(\frac{1}{2}W_2^2(\cdot,\mu))(\rho)\cap\lambda\text{-Lip}(\Omega)$.

Second, if $\rho\in\mathcal{P}(\Omega)$ is such that there exists φ satisfying

$$\varphi \in \partial\left(\frac{1}{2}W_2^2(\cdot,\mu)\right)(\rho)\cap\lambda\text{-Lip}(\Omega),$$

then, using Lemma 5.1,

$$\begin{aligned}\frac{1}{2}W_2^2(\mu,\rho) &= \int_\Omega \varphi^{c_{2,\lambda}}d\mu + \int_\Omega \varphi d\rho\\ &\le \mathcal{I}_{c_{2,\lambda}}(\mu,\rho).\end{aligned}$$

Since $\frac{1}{2}W_2^2(\mu,\rho)\ge\mathcal{I}_{c_{2,\lambda}}(\mu,\rho)$ in general, we have $\frac{1}{2}W_2^2(\mu,\rho)=\mathcal{I}_{c_{2,\lambda}}(\mu,\rho)$.

Finally, suppose $\frac{1}{2}W_2^2(\mu,\rho)=\mathcal{I}_{c_{2,\lambda}}(\mu,\rho)$. Since Ω is compact, there exists an optimal plan $\gamma_0\in\Pi(\mu,\rho)$ for $W_2(\mu,\rho)$. We compute

$$\begin{aligned}\frac{1}{2}W_2^2(\mu,\rho) &= \int_{\Omega\times\Omega} c_2(x,y)d\gamma_0(x,y)\\ &\ge \int_{\Omega\times\Omega} c_{2,\lambda}(x,y)d\gamma_0(x,y)\\ &\ge \mathcal{I}_{c_{2,\lambda}}(\mu,\rho)\\ &= \frac{1}{2}W_2^2(\mu,\rho).\end{aligned}\tag{5.7}$$

Equality of the first and last terms means we have equality throughout. This indicates that γ_0 is optimal for $\mathcal{I}_{c_{2,\lambda}}(\mu,\rho)$, and

$$\gamma_0(\{(x,y)\in\Omega \mid |x-y|>\lambda\})=0,$$

otherwise the inequality in (5.7) would be strict. Thus, $\rho\in B_\lambda(\mu)$. ∎

In the context of (WROF), the divergence D previously defined in (4.2) takes the following form:

$$\begin{aligned}D_\lambda(\nu,\rho) &= \frac{1}{2}W_2^2(\nu,\mu)-\frac{1}{2}W_2^2(\rho,\mu)\\ &\quad -\sup\left\{\langle\varphi,\nu-\rho\rangle \mid \varphi\in\partial\left(\frac{1}{2}W_2^2(\cdot,\mu)\right)(\rho)\cap\lambda\text{-Lip}(\Omega)\right\},\end{aligned}\tag{5.8}$$

with $\mu \in \mathcal{P}(\Omega)$ a fixed reference measure. Here we have introduced the notation D_λ to make the dependence of D on the scale λ explicit.

We now detail the economic interpretation of (5.8) that we mentioned in subsection 1.1. Here we assume that goods are manufactured with distribution μ, purchased from the manufacturer with distribution ρ and sold to consumers with distribution ν. In this setting $D_\lambda(\nu, \rho)$ represents the total loss of value in a supply chain when the transport cost has an economy of scale and consumers adopt a "buy local" policy.

Indeed if anyone can move goods from x to y for a transport cost of $c_{2,\lambda}(x, y)$, it is well known[5] that the maximum profit obtainable for transporting μ to ρ while still being competitive with this global shipping rate is $\mathcal{I}_{c_{2,\lambda}}(\mu, \rho)$, and that a potential

$$\varphi \in \partial \left(\mathcal{I}_{c_{2,\lambda}}(\cdot, \mu) \right)(\rho)$$

represents an optimal sale price as a function of location. We suppose that instead of shipping directly to consumers, the manufacturer sells to a retailer, who purchases product with distribution ρ and sells with distribution ν, both at price φ. The profits obtained by the retailer are therefore

$$\langle \varphi, \nu - \rho \rangle.$$

Given μ and ρ, there may be several optimal prices φ, and since all of them result in the same benefit for the manufacturer, they allow the retailer to choose one that maximizes their profit. However, the manufacturer specifies that $\varphi \in \lambda\text{-Lip}(\Omega)$; otherwise the retailer may be able to exploit an arbitrage against the global shipping cost $c_{2,\lambda}$. The profits of the retailer are then

$$\sup\{ \langle \varphi, \nu - \rho \rangle \mid \varphi \in \partial \left(\mathcal{I}_{c_{2,\lambda}}(\cdot, \mu) \right)(\rho) \cap \lambda\text{-Lip}(\Omega) \}.$$

We now suppose that the consumers impose a "buy local" policy, in the sense that they will not tolerate goods being shipped more than distance λ to retailers. The retailer must modify ρ to compensate for this, and by definition the only admissible distributions are those in $B_\lambda(\mu)$. If $\rho \in B_\lambda(\mu)$, however, Lemmas 5.1 and 5.5 show that

$$\varphi \in \partial \left(\mathcal{I}_{c_{2,\lambda}}(\cdot, \mu) \right)(\rho) \cap \lambda\text{-Lip}(\Omega) \Leftrightarrow \varphi \in \partial \left(\frac{1}{2} W_2^2(\cdot, \mu) \right)(\rho) \cap \lambda\text{-Lip}(\Omega).$$

Since $\mathcal{I}_{c_{2,\lambda}}(\mu, \rho) = \frac{1}{2} W_2^2(\mu, \rho)$ for $\rho \in B_\lambda(\mu)$, the total profits for both manufacturer and retailer are

$$\frac{1}{2} W_2^2(\mu, \rho) + \sup \left\{ \langle \varphi, \nu - \rho \rangle \mid \varphi \in \partial \left(\frac{1}{2} W_2^2(\cdot, \mu) \right)(\rho) \cap \lambda\text{-Lip}(\Omega) \right\}.$$

Subtracting this from the baseline $\frac{1}{2} W_2^2(\mu, \nu)$, we see that $D_\lambda(\nu, \rho)$ is indeed the total loss of value when a product is purchased by retailers at distribution ρ and sold at distribution ν under a buy local policy for consumers and when transportation over scale λ is discounted.

5.2. Applying Theorem 4.5 to (WROF). With $B_\lambda(\mu)$ and D_λ defined, we can finally apply Theorem 4.5 to characterize ρ_λ, the unique minimizer of (WROF), as a projection of ν onto $B_\lambda(\mu)$ with respect to the divergence D_λ. The following result is a more detailed version of Theorem 1.3 from section 1.

[5] See, for example, [34, p. 65].

THEOREM 5.6. *Let Ω be compact and convex with nonnegligible interior, and suppose $\mu \ll \mathcal{L}_d$. Then*

a. (WROF) *has a unique solution $\rho_\lambda = (T_\lambda)_\# \mu$, where $T_\lambda = I - \nabla \varphi_\lambda^{c_2}$ almost everywhere and φ_λ solves* (5.1),

b. *ρ_λ is also a solution to*

$$\min_{\rho \in B_\lambda(\mu)} D_\lambda(\nu, \rho), \tag{5.9}$$

and

c. *the values of* (WROF), (5.1), *and $\frac{1}{2}W_2^2(\nu,\mu) - D_\lambda(\nu,\rho_\lambda)$ coincide.*

Given statement b, *we have the following dichotomy.*

1. *If $\nu \in B_\lambda(\mu)$, then $\rho_\lambda = \nu$.*
2. *Otherwise, $\rho_\lambda \neq \nu$. Furthermore, any solution φ_λ to* (5.1) *satisfies $\varphi_\lambda \in \partial(\frac{1}{2}W_2^2(\cdot,\mu))(\rho_\lambda)$, Lip$(\varphi_\lambda) = \lambda$, and*

$$\langle \varphi_\lambda, \nu - \rho_\lambda \rangle = \lambda W_1(\rho_\lambda, \nu). \tag{5.10}$$

Finally, T_λ is the unique optimal transport map for $W_2(\mu, \rho_\lambda)$ and satisfies (1.11).

Remark 5.7. This result, together with Lemma 5.2, provides a proof of statement 1 of Proposition 1.4. Moreover, recalling Lemma 3.2, we observe that $I - \nabla \varphi_0^{c_2}$ is the solution map to (1.4). Thus, Theorem 5.6 also proves statement 2 of Proposition 1.4.

Proof. We have already described how $F(\varphi) = -\int_\Omega \varphi^{c_2} d\mu$ and $K = \lambda\text{-Lip}(\Omega)$ satisfy the hypotheses of Theorem 4.5; in particular, $\mu \ll \mathcal{L}_d$ and $\mathcal{L}_d(\partial\Omega) = 0$ guarantee strict convexity of F^*. Further, Lemma 5.3 guarantees the existence of a solution to (5.1). Statements a, b, and c then follow immediately from Theorem 4.5 and Lemma 5.4.

Since we have shown that $\rho \in B_\lambda(\mu)$ is equivalent to $\partial F^*(\rho) \cap K \neq \emptyset$ in Lemma 5.5, we see that the condition of the dichotomies in this proposition and Theorem 4.5 correspond. The only part of statements 1 and 2 in Theorem 5.6 that is not an immediate implication of Theorem 4.5 is that $\text{Lip}(\varphi_\lambda) = \lambda$, but this comes from determining that $\|\varphi_\lambda\|_K = \text{Lip}(\varphi_\lambda)/\lambda$. Further, T_λ is optimal for $W_2(\mu, \rho_\lambda)$ since $T_\lambda = I - \nabla \varphi_\lambda^{c_2}$ almost everywhere and $\varphi_\lambda \in \partial(\frac{1}{2}W_2^2(\cdot,\mu))(\rho_\lambda)$. Finally, (1.11) holds since $\text{Lip}(\varphi_\lambda^{c_2}) \leq \lambda$ by Lemma 5.1. ∎

This result, together with Lemma 5.2, furnishes an additional description of the value of (WROF) which is useful in proving the interpretation of $D_\lambda(\nu, \rho_\lambda)$ in (1.8).

COROLLARY 5.8. *Under the hypotheses of Theorem* 5.6, *the minimal value of* (WROF) *is equal to*

$$\mathcal{I}_{c_{2,\lambda}}(\mu,\nu) = \inf\left\{ \int_{\Omega\times\Omega} c_{2,\lambda}(x,y) d\gamma \mid \gamma \in \Pi(\mu,\nu) \right\}. \tag{5.11}$$

Proof. Observe that (5.4) is a standard Kantorovich potential problem, and thus via Theorem 5.6, Lemma 5.2, and Theorem 1.39 of [30] we get that the value of (WROF) coincides with (5.11). ∎

We now turn to the proof of Theorem 1.6. A crucial role is played by the absolute continuity of ρ_λ, and the proof of this property is the focus of the following section.

6. Absolute continuity of $\boldsymbol{\rho_\lambda}$. The following proposition provides conditions under which ρ_λ is guaranteed to be absolutely continuous, and proves statement 1 from Theorem 1.6.

PROPOSITION 6.1. *Suppose that $\Omega \subset \mathbb{R}^d$ is compact and convex, with a nonempty interior. If μ and ν are absolutely continuous with respect to Lebesgue measure, then ρ_λ, the unique solution to* (WROF)*, is absolutely continuous as well.*

The rest of this section is devoted to proving Proposition 6.1, and the plan is as follows. First, in subsection 6.1 we use the alternate expression for the dual problem (5.1) furnished by (5.4) to obtain a better understanding of how ρ_λ relates to μ and ν. Namely, there is an optimal transport plan γ_0 for (5.11), and ρ_λ is obtained by completing all transport in this plan that moves less than distance λ, as well as progressing all transport that moves more than distance λ as much as possible while retaining $\rho_\lambda \in B_\lambda(\mu)$. We use this understanding to decompose ρ_λ into a sum of two measures, and by proving that each of these is absolutely continuous, we will obtain that ρ_λ is absolutely continuous as well.

6.1. Consequences of Lemma 5.2 for a minimizer of (WROF). Recall Corollary 5.8, which says that the value of (WROF) coincides with that of (5.11). By Theorem 1.4 of [30], an optimal plan for the latter exists since Ω is compact; throughout this section we will refer to this plan by the notation γ_0. Let us also fix φ_λ as a solution of (5.1) which is $c_{2,\lambda}$-concave; such a φ_λ exists by Lemma 5.2. The following simple result characterizes $\nabla\varphi_\lambda^{c_2}$, and thus the solution of (WROF) (see Theorem 5.6), in terms of γ_0.

LEMMA 6.2. *Let Ω be compact and convex with nonnegligible interior. Let γ_0 be optimal in* (5.11)*. If $(x,y) \in \mathrm{spt}(\gamma_0)$, with x in the interior of Ω and a differentiable point of $\varphi_\lambda^{c_2}$, then*

$$\nabla\varphi_\lambda^{c_2}(x) = \begin{cases} x - y, & |x-y| \le \lambda, \\ \lambda \dfrac{x-y}{|x-y|}, & |x-y| \ge \lambda. \end{cases}$$

Thus, there is at most one $y \in B_\lambda(x)$ such that $(x,y) \in \mathrm{spt}(\gamma_0)$ and in that case $x - \nabla\varphi_0^{c_2}(x) = y$.

Proof. Since φ_λ solves (5.1), Lemma 5.2 implies that φ_λ also solves (5.4), and Lemma 5.1 gives that $\varphi^{c_2} = \varphi^{c_{2,\lambda}}$. Since φ_λ is optimal potential in (5.4) we have

$$\varphi_\lambda^{c_2}(x) + \varphi_\lambda(y) \le c_{2,\lambda}(x,y)$$

with equality on the support of γ_0. Thus, if $(x,y) \in \mathrm{spt}(\gamma_0)$, the minimum of

$$\inf_z c_{2,\lambda}(z,y) - \varphi_\lambda^{c_2}(z)$$

is obtained at x. If x is interior to Ω and a differentiable point of $\varphi_\lambda^{c_2}$, then

$$0 = \nabla_x c_{2,\lambda}(x,y) - \nabla\varphi_\lambda^{c_2}(x).$$

Computing the derivative of $c_{2,\lambda}$, we obtain the claim. ∎

We note that since $T_\lambda(x) = x - \nabla\varphi_\lambda^{c_2}(x)$ almost everywhere, Lemma 6.2 proves statement 3 of Proposition 1.4.

6.2. A decomposition of $\boldsymbol{\rho_\lambda}$. Define the Borel measures

$$\mu^a = (\pi_x)_\# \gamma_0|_{\{|x-y|\le\lambda\}}, \quad \mu^b = (\pi_x)_\# \gamma_0|_{\{|x-y|>\lambda\}}, \tag{6.1}$$

where π_x and π_y are the canonical projections. Let T_λ be a Borel map which is almost everywhere equal to $I - \nabla\varphi_\lambda^{c_2}$. Recalling that T_λ is optimal for the transport between μ and ρ_λ for the cost c_2 (see Theorem 5.6), define

$$\rho_\lambda^a = (T_\lambda)_\# \mu^a, \quad \rho_\lambda^b = (T_\lambda)_\# \mu^b. \tag{6.2}$$

It is clear that $\mu = \mu^a + \mu^b$, and from this we obtain $\rho_\lambda = \rho_\lambda^a + \rho_\lambda^b$. We will prove that $\rho_\lambda \ll \mathcal{L}_d$ by showing the same for ρ_λ^a and ρ_λ^b.

It is easier to prove that $\rho_\lambda^a \ll \mathcal{L}_d$, and that is the content of the following lemma. We will actually prove the stronger result that $\rho_\lambda^a(E) \le \nu(E)$ for all Borel E. This inequality should be expected given the discussion following Lemma 6.2, which says that the map $I - \nabla\varphi_\lambda^{c_2}$ completes all transport in γ_0 that moves less than distance λ. Since the mass that moves less than distance λ under γ_0 is precisely μ^a, and γ_0 transports μ to ν, that $\rho_\lambda^a \le \nu$ is not surprising.

LEMMA 6.3. *If $\mu \ll \mathcal{L}_d$, then for all $E \subset \Omega$ Borel we have*

$$\rho_\lambda^a(E) \le \nu(E). \tag{6.3}$$

As such, if $\nu \ll \mathcal{L}_d$, we have $\rho_\lambda^a \ll \mathcal{L}_d$ as well.

Proof. Observe that if

$$\gamma_0|_{\{|x-y|\le\lambda\}} = (I, T_\lambda)_\# \mu^a, \tag{6.4}$$

then we are done, since then for $E \subset \Omega$ Borel,

$$\begin{aligned}
\rho_\lambda^a(E) &= (\pi_y)_\# (I, T_\lambda)_\# \mu^a(E) \\
&= (\pi_y)_\# \gamma_0|_{\{|x-y|\le\lambda\}}(E) \\
&= \gamma_0(\Omega \times E \cap \{|x-y| \le \lambda\}) \\
&\le \gamma_0(\Omega \times E) \\
&= \nu(E).
\end{aligned}$$

So, we focus on proving (6.4). Note first that if $\gamma_0|_{\{|x-y\}\le\lambda}$ is the zero measure, then (6.4) automatically holds. We therefore proceed assuming that

$$\gamma_0|_{\{|x-y|\le\lambda\}}(\Omega \times \Omega) > 0.$$

Recall the potential φ_λ, optimal in (5.1). Since $\varphi_\lambda^{c_2}$ is Lipschitz, it is differentiable almost everywhere. Thus, $\mathcal{L}_d(\partial\Omega) = 0$ implies that there exists a Borel measurable set $G \subset \Omega \setminus \partial\Omega$ such that $\varphi_\lambda^{c_2}$ is differentiable on G, $T_\lambda(x) = x - \nabla\varphi_\lambda^{c_2}(x)$ on G, and $\mathcal{L}_d(G^c) = 0$. We therefore have, for $E_1, E_2 \subset \Omega$ Borel,

$$\begin{aligned}
\gamma_0|_{\{|x-y|\le\lambda\}}(E_1 \times E_2) &= \gamma_0(E_1 \times E_2 \cap \{|x-y| \le \lambda\}) \\
&= \gamma_0(E_1 \times E_2 \cap \{|x-y| \le \lambda\} \cap \operatorname{spt}(\gamma_0) \cap G \times \Omega).
\end{aligned}$$

Here, the second equality holds since $\mu \ll \mathcal{L}_d$. Next, we claim that

$$\{|x-y| \le \lambda\} \cap \operatorname{spt}(\gamma_0) \cap G \times \Omega \subset \Gamma_{T_\lambda}(G),$$

the latter being the graph of the map T_λ over G. Indeed, if (x, y) is in the set on the left-hand side, then according to Lemma 6.2 we get $y = x - \nabla\varphi_0^{c_2}(x) = T_\lambda(x)$, which proves the claim. We observe, however, that

$$(E_1 \times E_2) \cap \Gamma_{T_\lambda}(G) = (E_1 \cap T_\lambda^{-1}(E_2) \times \Omega) \cap \Gamma_{T_\lambda}(G).$$

As such,

$$\begin{aligned}\gamma_0|_{\{|x-y|\le\lambda\}}(E_1\times E_2) &= \gamma_0(E_1\times E_2\cap\{|x-y|\le\lambda\}\cap \operatorname{spt}(\gamma_0)\cap G\times\Omega)\\ &= \gamma_0(E_1\cap T_\lambda^{-1}(E_2)\times\Omega\cap\{|x-y|\le\lambda\})\\ &= \gamma_0|_{\{|x-y|\le\lambda\}}(E_1\cap T_\lambda^{-1}(E_2)\times\Omega)\\ &= \mu^a(E_1\cap T_\lambda^{-1}(E_2))\\ &= (I,T_\lambda)_\#\mu^a(E_1\times E_2).\end{aligned}$$

Thus, $\gamma_0|_{\{|x-y|\le\lambda\}}$ and $(I,T_\lambda)_\#\mu^a$ agree on all measurable rectangles $E_1\times E_2$. Since $\mu^a(\Omega)=\gamma_0|_{\{|x-y|\le\lambda\}}(\Omega\times\Omega)$, we can multiply $\gamma_0|_{\{|x-y|\le\lambda\}}$ and $(I,T_\lambda)_\#\mu^a$ by the same constant to obtain probability measures. These probability measures agree on all measurable rectangles, and hence by Theorem 3.3 of [5] they are equal. This implies (6.4), completing the proof. ∎

Remark 6.4. We note that Lemma 6.3 implies that absolute continuity of μ is not enough to obtain $\rho_\lambda\ll\mathcal{L}_d$. Indeed, if ν and $\mathcal{L}_d$ are singular and μ^a is nonzero, then ρ_λ^a is nonzero and Lemma 6.3 implies that ρ_λ^a and $\mathcal{L}_d$ are singular. Thus, for singular ν, ρ_λ may have a nonzero singular component with respect to Lebesgue measure.

6.3. Proof that $\boldsymbol{\rho_\lambda^b\ll\mathcal{L}_d}$. The general idea of the argument is to take $E\subset\Omega$ Borel with measure 0 and write

$$\begin{aligned}\rho_\lambda^b(E) &= (\pi_x)_\#\gamma_0|_{|x-y|>\lambda}(T_\lambda^{-1}(E))\\ &= \gamma_0(T_\lambda^{-1}(E)\times\Omega\cap\{|x-y|>\lambda\}\cap\operatorname{spt}(\gamma_0)).\end{aligned}$$

We will show that the set in the preceding line is contained in a set of the form $A\times\Omega$ for A Borel with measure 0, which will guarantee that $\rho_\lambda^b(E)=0$ since $\mu\ll\mathcal{L}_d$.

We will start with some simple observations about the set of (x,y) inside the support of γ_0 with $|x-y|>\lambda$. We will use the notion of the transport rays of a 1-Lipschitz function (see, for example, Definition 3.7 of [30]).

LEMMA 6.5. *If $(x,y)\in\operatorname{spt}(\gamma_0)$ with $|x-y|>\lambda$, then x and y are in transport rays of $\varphi_\lambda^{c_2}/\lambda$ and $-\varphi_\lambda/\lambda$, respectively. If $\varphi_\lambda^{c_2}$ is differentiable at x and x is an interior point of Ω, then the increasing directions of both rays are parallel to $\nabla\varphi_\lambda^{c_2}(x)$. Further, $x-\nabla\varphi_\lambda^{c_2}(x)$ is in the same ray as y, and this is the unique transport ray of $-\varphi_\lambda/\lambda$ containing $x-\nabla\varphi_\lambda^{c_2}(x)$.*

Proof. Since $(x,y)\in\operatorname{spt}(\gamma_0)$, Kantorovich duality gives us that

$$\varphi_\lambda^{c_2}(x)+\varphi_\lambda(y)=c_{2,\lambda}(x,y).$$

By the equality $\varphi_\lambda^{c_2}=\varphi_\lambda^{c_{2,\lambda}}$,

$$\varphi_\lambda^{c_2}(x)=\inf_{z\in\Omega}c_{2,\lambda}(x,z)-\varphi_\lambda(z),$$

and so we know the infimum is obtained at y. Note that since $|x-y|>\lambda$, by traversing the segment $[x,y]$ starting at y we obtain a rate of decrease of λ per unit distance for $c_{2,\lambda}(x,\cdot)$. Since $-\varphi_\lambda$ is λ-Lipschitz, we must therefore have that the infimum is also obtained at every point $z\in[x,y]$ with $|x-z|\ge\lambda$. This is possible only if $-\varphi_\lambda$ increases at maximal rate along this nontrivial segment, and thus $[x+\lambda\frac{y-x}{|y-x|},y]$, and therefore y, is contained in a transport ray of $-\varphi_\lambda/\lambda$. This transport ray has increasing direction parallel to $x-y$, which will be needed later.

Since $\varphi_\lambda^{c_2} \in \lambda\text{-Lip}(\Omega)$ as well, we can prove that x is in a transport ray of $\varphi_\lambda^{c_2}/\lambda$ with a nearly identical argument, starting from the equality

$$\varphi_\lambda(y) = \inf_{z\in\Omega} c_{2,\lambda}(y,z) - \varphi_\lambda^{c_2}(z),$$

which holds since φ_λ is $c_{2,\lambda}$-concave.

If φ^{c_2} is differentiable at x, then it is clear that $\nabla\varphi^{c_2}(x)$ is parallel to the increasing direction of the transport ray of φ^{c_2}/λ that x is in. On the other hand, the increasing direction of the transport ray of $-\varphi_\lambda/\lambda$ containing y is parallel to $x-y$, which is parallel to $\nabla\varphi_\lambda^{c_2}$ by Lemma 6.2. By the same lemma we have

$$x + \lambda\frac{y-x}{|y-x|} = x - \nabla\varphi_\lambda^{c_2}(x),$$

which verifies that $x - \nabla\varphi_\lambda^{c_2}(x)$ is in the same transport ray of $-\varphi_\lambda/\lambda$ as y.

To see that the transport ray containing $x - \nabla\varphi_\lambda^{c_2}(x)$ is unique, suppose $x - \nabla\varphi_\lambda^{c_2}(x)$ is contained in two transport rays of $-\varphi_\lambda/\lambda$. As we have shown, one of these has decreasing direction parallel to $-\nabla\varphi_\lambda^{c_2}(x)$ and, since $|x-y| > \lambda$, nonzero length in this direction. Noting that two transport rays can only collide at a point which is the upper (or lower) endpoint of both rays (see, for example, Lemma 10 of [7]), we get that if $x - \nabla\varphi_\lambda^{c_2}(x)$ is in a second ray, it must be at the upper endpoint of that ray. Let the decreasing direction of the other ray be given by the unit vector v. We compute

$$\begin{aligned}\frac{d}{dt}|_{t=0} c_{2,\lambda}(x, x-\nabla\varphi_\lambda^{c_2}(x)+tv) - \varphi_\lambda(x-\nabla\varphi_\lambda^{c_2}(x)+tv) &= \lambda\left\langle v, -\frac{\nabla\varphi_\lambda^{c_2}(x)}{|\nabla\varphi_\lambda^{c_2}(x)|}\right\rangle - \lambda \\ &= \lambda\left(\left\langle v, \frac{y-x}{|y-x|}\right\rangle - 1\right) \\ &< 0,\end{aligned}$$

the final inequality coming from the fact that $v \neq \frac{y-x}{|y-x|}$, and both have unit norm. As such, $t \mapsto c_{2,\lambda}(x, x-\nabla\varphi_\lambda^{c_2}(x)+tv) - \varphi_\lambda(x-\nabla\varphi_\lambda^{c_2}(x)+tv)$ is strictly decreasing for $t \in (0,\epsilon)$ for some ϵ, contradicting the fact that the infimum of $c_{2,\lambda}(x,y) - \varphi_\lambda(y)$ is obtained at $x - \nabla\varphi_\lambda^{c_2}(x)$. ∎

We can now prove that the upper endpoints of the transport rays of $-\varphi_\lambda/\lambda$ correspond to the upper endpoints of the transport rays of $\varphi_\lambda^{c_2}/\lambda$.

LEMMA 6.6. *Suppose $(x,y) \in \text{spt}(\gamma_0)$ with $|x-y| > \lambda$, and suppose $\varphi_\lambda^{c_2}$ is differentiable at x and x is an interior point of Ω. If $x - \nabla\varphi_\lambda^{c_2}(x)$ is at the upper endpoint of its transport ray of $-\varphi_\lambda/\lambda$ then x is at the upper endpoint of a transport ray of $\varphi_\lambda^{c_2}/\lambda$.*

Proof. From Lemma 6.5 we have that x is in a transport ray of $\varphi_\lambda^{c_2}/\lambda$. Suppose it is not the upper endpoint. Then there exists w on the same transport ray obtaining a strictly larger value of $\varphi_\lambda^{c_2}$. As such

$$\begin{aligned}\varphi_\lambda(y) &= c_{2,\lambda}(x,y) - \varphi_\lambda^{c_2}(x) \\ &= c_{2,\lambda}(x,y) - \varphi_\lambda^{c_2}(w) + \lambda|w-x| \\ &= c_{2,\lambda}(w,y) - \varphi_\lambda^{c_2}(w).\end{aligned}$$

Here the last line holds because the transport ray that x is in is parallel to the segment $[x,y]$. Since $\varphi_\lambda^{c_2} = \varphi_\lambda^{c_{2,\lambda}}$,

$$\varphi_\lambda^{c_2}(w) = \inf_{z\in\Omega} c_{2,\lambda}(w,z) - \varphi_\lambda(z),$$

and the infimum is obtained at y. Since $|w-y| > \lambda + |w-x|$, we obtain that all points on the segment $[w + \lambda\frac{y-w}{|y-w|}, y]$ are on a transport ray of $-\varphi_\lambda/\lambda$. The point $x - \nabla\varphi_\lambda^{c_2}(x)$ is in the interior of this ray and thus is not the upper endpoint. ∎

We can now prove that ρ_λ^b is absolutely continuous with respect to Lebesgue measure. The general argument is the following. Take E Borel negligible, $x \in T_\lambda^{-1}(E)$, and suppose that there exists y such that $(x,y) \in \mathrm{spt}(\gamma_0)$ with $|x-y| > \lambda$. Then, ignoring x at the start of transport rays of $\varphi_\lambda^{c_2}/\lambda$ (which is a Borel negligible set anyway), we can show that $x = z - \nabla\varphi_\lambda(z)$ for $z \in E$. Since x is not at the start of its transport ray, z cannot be at the start of its transport ray. Away from the endpoints of transport rays the map $z \mapsto z - \nabla\varphi_\lambda(z)$ is Lipschitz,[6] allowing us to conclude that our set of x is Lebesgue negligible.

PROPOSITION 6.7. *The measure ρ_λ^b satisfies $\rho_\lambda^b \ll \mathcal{L}_d$.*

Proof. Let $E \subset \Omega$ be Borel negligible. Then

$$\begin{aligned}\rho_\lambda^b(E) &= (\pi_x)_\# \gamma_0|_{|x-y|>\lambda}(T_\lambda^{-1}(E)) \\ &= \gamma_0((T_\lambda^{-1}(E) \cap \mathcal{E}^c \cap G) \times \Omega \cap \{|x-y| > \lambda\} \cap \mathrm{spt}(\gamma_0)),\end{aligned} \tag{6.5}$$

where $\mathcal{E}^c$ is the complement of the set of ray endpoints of $\varphi_\lambda^{c_2}/\lambda$, and G is as before. Both sets are Borel and have full Lebesgue measure,[7] justifying the equality (6.5). If (x,y) is in the set appearing in (6.5), then by Lemma 6.5, $x - \nabla\varphi_\lambda^{c_2}(x)$ is in a unique transport ray of $-\varphi_\lambda/\lambda$, and by Lemma 6.6 $x - \nabla\varphi_\lambda^{c_2}(x)$ is not at the upper endpoint of that ray. Since $|x-y| > \lambda$, $x - \nabla\varphi_\lambda^{c_2}(x)$ is also not at the lower endpoint of that ray. By Lemma 3.6 of [30], $-\varphi_\lambda/\lambda$ is differentiable at $x - \nabla\varphi_\lambda^{c_2}(x)$, and Lemma 6.5 implies that

$$x = x - \nabla\varphi_\lambda^{c_2}(x) - \nabla\varphi_\lambda(x - \nabla\varphi_\lambda^{c_2}(x)) = T_\lambda(x) - \nabla\varphi_\lambda(T_\lambda(x)).$$

Thus, if (x,y) is in the set appearing in (6.5), then $x = z - \nabla\varphi_\lambda(z)$ for some $z \in E$ and in the interior of a transport ray of $-\varphi_\lambda/\lambda$. As in Proposition 6 of [25], for each $j \in \{1,2,\ldots\}$ set A_j as the set of points z that are on a transport ray of $-\varphi_\lambda/\lambda$ and more than distance $1/j$ from either endpoint, and recall that by Lemma 22 of [7], $-\nabla\varphi_\lambda$ is a Lipschitz function on A_j. We therefore obtain that if (x,y) is in the set appearing in (6.5), then

$$x \in \bigcup_{j=1}^{\infty} (I - \nabla\varphi_\lambda)(E \cap A_j).$$

Since E is Borel negligible, and $\nabla\varphi_\lambda$ is a Lipschitz map on A_j, we obtain that the set $(I - \nabla\varphi_\lambda)(E \cap A_j)$ is Lebesgue measurable for all j and has measure 0. By regularity of Lebesgue measure, there exists for each j a Borel set U_j containing $(I - \nabla\varphi_\lambda^{c_2})(E \cap A_j)$ with zero Lebesgue measure. As such,

$$\rho_\lambda^b(E) \le \sum_{j=1}^{\infty} \gamma_0(U_j \times \Omega) = \sum_{j=1}^{\infty} \mu(U_j) = 0$$

because $\mu \ll \mathcal{L}_d$. ∎

We have therefore proven Proposition 6.1, and thus statement 1 of Theorem 1.6, by proving that ρ_λ^b and ρ_λ^a are absolutely continuous (Lemma 6.3 and Proposition 6.7).

[6] See the proof of Lemma 22 of [7] or Proposition 6 of [25].

[7] For a proof that $\mathcal{L}_d(\mathcal{E}) = 0$, see Lemma 25 of [7] or Lemma 3.1.8 of [16].

7. Characterization of an optimal map for the Huber cost. In this section we will prove statements 2 and 3 of Theorem 1.6. The essential result is the characterization of an optimal map transporting ν to μ for the Huber cost $c_{2,\lambda}$ as a composition of a Wasserstein 2 optimal map with a Wasserstein 1 optimal map. We note that the existence of an optimal map for the cost $c_{2,\lambda}$ does not follow trivially from standard results in the optimal transport literature (e.g., Theorem 1.17 of [30]) since the cost $c_{2,\lambda}(x,y)$ is not a strictly convex function of $|x-y|$.

The following lemma proves that the gradient of φ_λ is ν almost surely unchanged by applying an optimal transport map for $W_1(\nu,\rho_\lambda)$, and will be useful in proving the existence of an optimal transport map for the Huber cost. Throughout this section we tacitly assume the hypotheses of Theorem 1.6.

LEMMA 7.1. *Let S_λ be an optimal transport map for $W_1(\nu,\rho_\lambda)$, which exists since $\nu \ll \mathcal{L}_d$. Let φ_λ be a $c_{2,\lambda}$-concave solution to* (5.4)*. Then ν almost everywhere $\nabla\varphi_\lambda(y)$ and $\nabla\varphi_\lambda(S_\lambda(y))$ exist. Further if $S_\lambda(y) \neq y$, they satisfy*

$$\nabla\varphi_\lambda(y) = \nabla\varphi_\lambda(S_\lambda(y)) = \lambda\frac{y - S_\lambda(y)}{|y - S_\lambda(y)|}. \tag{7.1}$$

Proof. The potential φ_λ is $c_{2,\lambda}$-concave, and thus $\varphi_\lambda \in \lambda\text{-Lip}(\Omega)$. Since $\nu \ll \mathcal{L}_d$, φ_λ is therefore differentiable ν almost everywhere. Further,

$$\begin{aligned}\nu(\{y \mid \nabla\varphi_\lambda(S_\lambda(y)) \text{ exists}\}) &= \nu(S_\lambda^{-1}(\{z \mid \nabla\varphi_\lambda(z) \text{ exists}\}) \\ &= \rho_\lambda(\{z \mid \nabla\varphi_\lambda(z) \text{ exists}\}) \\ &= 1,\end{aligned}$$

since $\rho_\lambda \ll \mathcal{L}_d$ (Proposition 6.1). So $\nabla\varphi_\lambda(S_\lambda(y))$ exists ν almost everywhere as well. Since φ_λ solves (5.1) via Lemma 5.2, we obtain via Theorem 5.6 that φ_λ/λ is a Kantorovich potential for $W_1(\nu,\rho_\lambda)$. Since S_λ is an optimal transport map for $W_1(\nu,\rho_\lambda)$ we obtain that for ν almost all $y \in \Omega$,

$$\lambda|S_\lambda(y) - y| = \varphi_\lambda(y) - \varphi_\lambda(S_\lambda(y)).$$

Thus, if $S_\lambda(y) \neq y$, we obtain that $[y, S_\lambda(y)]$ is in a transport ray of φ_λ/λ. Via Lemma 3.6 of [30] we obtain (7.1) whenever φ_λ is differentiable at both y and $S_\lambda(y)$. ∎

Now we can prove the existence of an optimal transport map S_0 from ν to μ under the cost $c_{2,\lambda}$ by composing a Wasserstein 1 optimal map from ν to ρ_λ with a Wasserstein 2 optimal map from ρ_λ to μ. Conversely, we will prove that all such optimal S_0 can be written in this way. The following result proves statement 2 of Theorem 1.6.

LEMMA 7.2. *Let φ_λ be a $c_{2,\lambda}$-concave solution to* (5.4)*. If R_λ is a Borel map almost everywhere equal to $I - \nabla\varphi_\lambda$, and T_λ is a Borel map almost everywhere equal to $I - \nabla\varphi_\lambda^{c_2}$, then*

$$R_\lambda \circ T_\lambda(x) = x \tag{7.2}$$

μ almost everywhere, so we write R_λ as T_λ^{-1}. A map S_0 is an optimal transport map for transporting ν to μ under the cost $c_{2,\lambda}$ if and only if S_0 can be written as $S_0 = T_\lambda^{-1} \circ S_\lambda$, where S_λ is an optimal map for $W_1(\nu,\rho_\lambda)$.

Proof. The claim (7.2) is well known, and it follows since φ_λ is a Kantorovich potential for $\frac{1}{2}W_2^2(\rho_\lambda,\mu)$ and T_λ is an optimal transport map for $\frac{1}{2}W_2^2(\mu,\rho_\lambda)$, and thus

$$\varphi_\lambda^{c_2}(x)+\varphi_\lambda(T_\lambda(x))=\frac{1}{2}|x-T_\lambda(x)|^2,$$

μ almost everywhere. Using $\rho_\lambda \ll \mathcal{L}_d$, one can then easily show (7.2). Let S_0 be given by $S_0 = T_\lambda^{-1}\circ S_\lambda$. This same equality also implies that $(T_\lambda^{-1})_\# \rho_\lambda = \mu$, and so $(S_0)_\#\nu = \mu$. We now wish to prove that for ν almost all y,

$$c_{2,\lambda}(S_0(y),y) = c_{2,\lambda}(S_0(y),S_\lambda(y)) + \lambda|y-S_\lambda(y)|.$$

This is clear if $S_\lambda(y)=y$. If $S_\lambda(y)\neq y$, then this equality is an immediate consequence of Lemma 7.1. We therefore compute

$$\begin{aligned}\int_\Omega c_{2,\lambda}(S_0(y),y)d\nu(y) &= \int_\Omega c_{2,\lambda}(S_0(y),S_\lambda(y))d\nu(y)+\lambda\int_\Omega |y-S_\lambda(y)|d\nu(y)\\ &= \int_\Omega c_2(T_0(z),z)d\rho_\lambda(z)+\lambda W_1(\rho_\lambda,\nu)\\ &= \frac{1}{2}W_2^2(\mu,\rho_\lambda)+\lambda W_1(\rho_\lambda,\nu)\\ &= \mathcal{I}_{c_{2,\lambda}}(\mu,\nu),\end{aligned}$$

where the last line follows from Corollary 5.8. This verifies that S_0 is optimal for transporting ν to μ with the pointwise cost $c_{2,\lambda}$.

Conversely, suppose S_0 is optimal for transporting ν to μ under this cost. If we can prove that $T_\lambda\circ S_0$ is optimal for $W_1(\nu,\rho_\lambda)$, we will be done. Clearly, $(T_\lambda\circ S_0)_\#\nu=\rho_\lambda$, and since φ_λ/λ is a Kantorovich potential for $W_1(\nu,\rho_\lambda)$ via Theorem 5.6, optimality of this map will be proved if we can show that

$$\lambda|y-T_\lambda(S_0(y))| = \varphi_\lambda(y)-\varphi_\lambda(T_\lambda(S_0(y))), \tag{7.3}$$

ν almost everywhere. To see this, observe that $(S_0,I)_\#\nu$ is an optimal plan for (5.11), and $(S_0(y),y)$ is in the support of this plan for ν almost all y. Conditioning ν on $|S_0(y)-y|\leq\lambda$, we obtain $y=T_\lambda(S_0(y))$ with probability 1 via Lemma 6.2, and thus (7.3) holds trivially. Conditioning on $|S_0(y)-y|>\lambda$, we may use Lemma 6.5 to obtain that $[T_\lambda(S_0(y)),y]$ is in a transport ray of φ_λ/λ ν almost surely, which proves (7.3) in this case. ∎

The following result proves statement 3 of Theorem 1.6 by demonstrating that by applying the soft thresholding operator (1.14) to the map S_0, one recovers S_λ.

PROPOSITION 7.3. *Let $S_0 = T_\lambda^{-1}\circ S_\lambda$ be an optimal transport map from ν to μ for the cost $c_{2,\lambda}$ as obtained in Lemma 7.2. Then ν almost everywhere,*

$$S_\lambda(y) = y + s_\lambda(|S_0(y)-y|)\frac{S_0(y)-y}{|S_0(y)-y|},$$

where $s_\lambda(|S_0(y)-y|)\frac{S_0(y)-y}{|S_0(y)-y|}=0$ if $S_0(y)=y$.

Proof. Take φ_λ and T_λ^{-1} as in Lemma 7.2, and set

$$E := S_\lambda^{-1}(\{z \mid T_\lambda^{-1}(z) = z - \nabla\varphi_\lambda(z)\}).$$

Then E has full ν measure. If $y \in E$ and $S_\lambda(y) = y$, then $S_0(y) = T_\lambda^{-1}(y)$. Since $\varphi_\lambda \in \lambda\text{-Lip}(\Omega)$, we obtain that

$$y + s_\lambda(|S_0(y) - y|)\frac{S_0(y) - y}{|S_0(y) - y|} = y = S_\lambda(y).$$

If $y \in E$ and $S_\lambda(y) \neq y$, then by Lemma 7.1, $|S_0(y) - y| > \lambda$ ν almost surely. Thus,

$$\begin{aligned} y + s_\lambda(|S_0(y) - y|)\frac{S_0(y) - y}{|S_0(y) - y|} &= y + (|S_0(y) - y| - \lambda))\frac{S_\lambda(y) - y}{|S_\lambda(y) - y|} \\ &= y + |S_\lambda(y) - y|\frac{S_\lambda(y) - y}{|S_\lambda(y) - y|} \\ &= S_\lambda(y). \end{aligned}$$

∎

8. Iterative procedures involving (WROF). In this section we study the iterative procedures described in subsections 1.2 and 1.3. The main content is a proof of Proposition 1.8 and Theorem 1.10.

8.1. Iterative regularization. Here we will prove our iterative regularization result Proposition 1.8. Recall the setting; we take $\mu, \nu \ll \mathcal{L}_d$, and $(\lambda_n)_{n=0}^\infty$ a sequence of positive step sizes with sum converging to $+\infty$. Set $\mu_0 := \mu$, and for $n \geq 0$ define

$$\mu_{n+1} := \operatorname*{arg\,min}_{\rho \in \mathcal{P}(\Omega)} \frac{1}{2}W_2^2(\rho, \mu_n) + \lambda_n W_1(\rho, \nu).$$

We note that $(\mu_n)_{n=1}^\infty$ is well defined given Lemma 5.3, Theorem 5.6, and Proposition 6.1. The first two results establish the existence of a unique solution to the minimization problem in (1.17) when $\mu \ll \mathcal{L}_d$, and the latter guarantees that this solution will be absolutely continuous as well.

Before we analyze the convergence of the sequence $(\mu_n)_{n=1}^\infty$, we establish a simple estimate on $W_1(\mu_n, \nu)$.

LEMMA 8.1. *Let Ω be convex and compact with nonnegligible interior. Take $\mu \ll \mathcal{L}_d$, and let ρ_λ solve* (WROF)*. Denoting an arbitrary optimal transport plan for the cost $c_{2,\lambda}$ from μ to ν as γ_0, define μ^a and μ^b as in* (6.1)*. Then*

$$W_1(\rho_\lambda, \nu) \leq \mu^b(\Omega)\operatorname{diam}(\Omega), \tag{8.1}$$

where $\operatorname{diam}(\Omega) = \sup\{|x - y| \mid x, y \in \Omega\}$.

Proof. Recall the definitions of ρ_λ^a and ρ_λ^b from (6.2). Note that if $\rho_\lambda^b(\Omega) = 0$, we obtain via Lemma 6.3 that $\nu = \rho_\lambda^a = \rho_\lambda$, and thus (8.1) holds. We therefore proceed assuming that $\rho_\lambda^b(\Omega) \neq 0$. We have

$$\begin{aligned} W_1(\rho_\lambda, \nu) &= \sup_{u \in 1\text{-Lip}(\Omega)} \langle u, \rho_\lambda - \nu\rangle \\ &= \sup_{u \in 1\text{-Lip}(\Omega)} \langle u, \rho_\lambda^a + \rho_\lambda^b - \nu\rangle \\ &= \sup_{u \in 1\text{-Lip}(\Omega)} \langle u, \rho_\lambda^b - (\nu - \rho_\lambda^a)\rangle. \end{aligned}$$

Via Lemma 6.3 we get that $\nu - \rho_\lambda^a$ is a nonnegative measure. Moreover, it has the same total mass as ρ_λ^b. As such

$$\begin{aligned} W_1(\rho_\lambda, \nu) &= \rho_\lambda^b(\Omega) \sup_{u \in 1\text{-Lip}(\Omega)} \left\langle u, \frac{\rho_\lambda^b}{\rho_\lambda^b(\Omega)} - \frac{\nu - \rho_\lambda^a}{(\nu - \rho_\lambda^a)(\Omega)} \right\rangle \\ &= \mu^b(\Omega) W_1\left(\frac{\rho_\lambda^b}{\rho_\lambda^b(\Omega)}, \frac{\nu - \rho_\lambda^a}{(\nu - \rho_\lambda^a)(\Omega)} \right) \\ &\leq \mu^b(\Omega) \operatorname{diam}(\Omega), \end{aligned}$$

as claimed. ∎

We can now prove our convergence result for $(\mu_n)_{n=1}^\infty$, which relies on Lemma 8.1.

Proof of Proposition 1.8. We first establish that $W_1(\mu_n, \nu)$ is monotonically decreasing in n. Indeed, by definition of μ_n,

$$W_1(\mu_n, \nu) \leq W_1(\mu_{n-1}, \nu) - \frac{1}{2\lambda_{n-1}} W_2^2(\mu_n, \mu_{n-1}) \leq W_1(\mu_{n-1}, \nu).$$

Iterating the first inequality, we also obtain that

$$W_1(\mu_n, \nu) \leq W_1(\mu, \nu) - \sum_{i=0}^{n-1} \frac{1}{2\lambda_i} W_2^2(\mu_{i+1}, \mu_i).$$

Thus,

$$\sum_{i=0}^{\infty} \frac{1}{2\lambda_i} W_2^2(\mu_{i+1}, \mu_i) < \infty. \tag{8.2}$$

For each i, let γ_i be an optimal plan for the transport from μ_i to ν under the cost c_{2,λ_i}, and define

$$\mu_i^b := (\pi_x)_\#(\gamma_i|_{|x-y|>\lambda_i}).$$

Let φ_i be a solution to (5.1) with μ replaced by μ_i and λ replaced by λ_i. Since $I - \nabla\varphi_i^{c_2}$ is almost everywhere equal to an optimal transport map from μ_i to μ_{i+1} (see Theorem 5.6), and using Lemma 6.2, we obtain

$$\frac{1}{2\lambda_i} W_2^2(\mu_i, \mu_{i+1}) \geq \frac{1}{2\lambda_i} \lambda_i^2 \mu_i^b(\Omega) = \frac{1}{2} \lambda_i \mu_i^b(\Omega).$$

As such, (8.2) implies

$$\sum_{i=1}^{\infty} \lambda_i \mu_i^b(\Omega) < \infty. \tag{8.3}$$

By (1.16), we obtain that $\liminf_i \mu_i^b(\Omega) = 0$. Lemma 8.1 implies that

$$W_1(\mu_{i+1}, \nu) \leq \mu_i^b(\Omega) \operatorname{diam}(\Omega).$$

Since $\liminf_i \mu_i^b(\Omega) = 0$, we therefore obtain

$$\liminf_i W_1(\mu_i, \nu) = 0, \tag{8.4}$$

as well. But $W_1(\mu_i, \nu)$ is monotonically decreasing in i, and so (8.4) implies (1.18). ∎

8.2. Multiscale transport and a nonlinear energy decomposition. In this section we prove Theorem 1.10. We already have most of the necessary ingredients.

Let us recall the setting of this procedure. We assume $\mu \ll \mathcal{L}_d$ and suppose λ_0 is given. For each $n \geq 0$, set $\lambda_{n+1} = \lambda_n/2$ and define

$$\nu_{n+1} := \operatorname*{arg\,min}_{\rho \in \mathcal{P}(\Omega)} \frac{1}{2} W_2^2(\rho, \mu) + \lambda_n W_1(\rho, \nu_n), \tag{8.5}$$

where $\nu_0 := \nu$.

Remark 8.2. This procedure consists of iteratively solving (4.1), starting with y_0^* as ν_0 and replacing it at each stage by ν_n, as well as halving the scale parameter. If the same is done in the context of ROF, starting with $y_0^* = 0$, one obtains a sequence of functions $(w_n)_{n=1}^{\infty}$ which are the partial sums of the multiscale decomposition in Theorem 1.9. In this light (8.5) is analogous to (1.19).

Proof of Theorem 1.10. The assumption $\mu \ll \mathcal{L}_d$, together with Lemma 5.3 and Theorem 5.6, guarantees for all n that the argmin in (1.23) exists and is unique. To prove statement 1, we note that by (1.11)

$$\frac{1}{2} W_2^2(\mu, \nu_n) \leq \frac{1}{2} \lambda_{n-1}^2 = 2^{-2n+1} \lambda_0^2,$$

which proves (1.24). To obtain the energy equality (1.25), we observe that

$$\begin{aligned} \frac{1}{2} W_2^2(\mu, \nu) &= \frac{1}{2} W_2^2(\mu, \nu_1) + \frac{1}{2} W_2^2(\mu, \nu_0) - \frac{1}{2} W_2^2(\mu, \nu_1) - \lambda_0 W_1(\nu_0, \nu_1) \\ &\quad + \lambda_0 W_1(\nu_0, \nu_1) \\ &= \frac{1}{2} W_2^2(\mu, \nu_1) + D_{\lambda_0}(\nu_0, \nu_1) + \lambda_0 W_1(\nu_0, \nu_1), \end{aligned}$$

where in the second line we have used the equality of (WROF) and (5.9), proven in Theorem 5.6. Iterating this equality, we obtain

$$\frac{1}{2} W_2^2(\mu, \nu) = \frac{1}{2} W_2^2(\mu, \nu_k) + \sum_{n=0}^{k-1} D_{\lambda_n}(\nu_n, \nu_{n+1}) + \lambda_n W_1(\nu_n, \nu_{n+1}).$$

Letting k go to infinity and using (1.24), we obtain (1.25). ∎

REFERENCES

[1] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*, Springer, New York, 2005.

[2] L. Ambrosio and A. Pratelli, *Existence and stability results in the L1 theory of optimal transportation*, in Optimal Transportation and Applications, Springer, New York, 2003, pp. 123–160.

[3] B. Amos, *On Amortizing Convex Conjugates for Optimal Transport*, preprint, arXiv: 2210.12153, 2022.

[4] P. Athavale, R. Xu, P. Radau, A. Nachman, and G. A. Wright, *Multiscale properties of weighted total variation flow with applications to denoising and registration*, Med. Image Anal., 23 (2015), pp. 28–42.

[5] P. Billingsley, *Probability and Measure*, John Wiley & Sons, New York, 2008.

[6] M. Burger, M. Franek, and C.-B. Schönlieb, *Regularized regression and density estimation based on optimal transport*, Appl. Math. Res. Express AMRX, 2012 (2012), pp. 209–253.

[7] L. Caffarelli, M. Feldman, and R. McCann, *Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs*, J. Amer. Math. Soc., 15 (2002), pp. 1–26.

[8] V. Caselles, A. Chambolle, and M. Novaga, *The discontinuity set of solutions of the TV denoising problem and some extensions*, Multiscale Model. Simul., 6 (2007), pp. 879–894.

[9] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.
[10] A. Chambolle, V. Duval, G. Peyré, and C. Poon, *Geometric properties of solutions to the total variation denoising problem*, Inverse Problems, 33 (2016), 015002.
[11] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, *Improved training of Wasserstein GANs*, in Advances in Neural Information Processing Systems, 2017, pp. 5767–5777.
[12] H. Heaton, S. W. Fung, A. T. Lin, S. Osher, and W. Yin, *Wasserstein-based projections with applications to inverse problems*, SIAM J. Math. Data Sci., 4 (2022), pp. 581–603.
[13] P. J. Huber, *Robust estimation of a location parameter*, Ann. Math. Statist., 35 (1964), pp. 73–101.
[14] M. Jacobs and F. Léger, *A fast approach to optimal transport: The back-and-forth method*, Numer. Math., 146 (2020), pp. 513–544.
[15] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker–Planck equation*, SIAM J. Math. Anal., 29 (1998), pp. 1–17.
[16] B. Klartag, *Needle Decompositions in Riemannian Geometry*, Mem. Amer. Math. Soc. 249, AMS, Providence, RI, 2017.
[17] E. Kobler, A. Effland, K. Kunisch, and T. Pock, *Total deep variation for linear inverse problems*, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 7549–7558.
[18] E. Kobler, T. Klatzer, K. Hammernik, and T. Pock, *Variational networks: Connecting variational methods and deep learning*, in German Conference on Pattern Recognition, Springer, New York, 2017, pp. 281–293.
[19] R. Lang, *A note on the measurability of convex sets*, Arch. Math., 47 (1986), pp. 90–92.
[20] J. Lellmann, D. A. Lorenz, C. Schönlieb, and T. Valkonen, *Imaging with Kantorovich–Rubinstein discrepancy*, SIAM J. Imaging Sci., 7 (2014), pp. 2833–2859.
[21] S. Lunz, O. Öktem, and C.-B. Schönlieb, *Adversarial regularizers in inverse problems*, in Advances in Neural Information Processing Systems, 2018, pp. 8507–8516.
[22] A. Makkuva, A. Taghvaei, S. Oh, and J. Lee, *Optimal transport mapping via input convex neural networks*, in International Conference on Machine Learning, PMLR, 2020, pp. 6672–6681.
[23] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, Univ. Lecture Ser. 22, AMS, Providence, RI, 2001.
[24] T. Milne, É. Bilocq, and A. Nachman, *Trust the Critics: Generatorless and Multipurpose WGANs with Initial Convergence Guarantees*, preprint, arXiv:2111.15099, 2021.
[25] T. Milne, É. Bilocq, and A. Nachman, *A New Method for Determining Wasserstein* 1 *Optimal Transport Maps From Kantorovich Potentials, with Deep Learning Applications*, preprint, arXiv:2211.00820, 2022.
[26] K. Modin, A. Nachman, and L. Rondi, *A multiscale theory for image registration and nonlinear inverse problems*, Adv. Math., 346 (2019), pp. 1009–1066.
[27] S. Mukherjee, M. Carioni, O. Öktem, and C.-B. Schönlieb, *End-to-end reconstruction meets data-driven regularization for inverse problems*, Adv. Neural Inform. Process. Syst., 34 (2021), pp. 21413–21425.
[28] S. Mukherjee, S. Dittmer, Z. Shumaylov, S. Lunz, O. Öktem, and C.-B. Schönlieb, *Learned Convex Regularizers for Inverse Problems*, preprint, arXiv:2008.02839, 2020.
[29] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[30] F. Santambrogio, *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, Progr. Nonlinear Differential Equations Appl. 87, Birkhäuser, Basel, 2015.
[31] F. Santambrogio, {*Euclidean, metric, and Wasserstein*} *gradient flows: An overview*, Bull. Math. Sci., 7 (2017), pp. 87–154.
[32] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational Methods in Imaging*, Springer, New York, 2009.
[33] E. Tadmor, S. Nezzar, and L. Vese, *A multiscale image representation using hierarchical* (BV, L^2) *decompositions*, Multiscale Model. Simul., 2 (2004), pp. 554–579.
[34] C. Villani, *Optimal Transport: Old and New*, Grundlehren Math. Wiss. 338, Springer, New York, 2009.
[35] C. Zalinescu, *Convex Analysis in General Vector Spaces*, World Scientific, River Edge, NJ, 2002.