

SOLVING LARGE-SCALE OPTIMIZATION PROBLEMS WITH  
A CONVERGENCE RATE INDEPENDENT OF GRID SIZE\*MATT JACOBS<sup>†</sup>, FLAVIEN LÉGER<sup>‡</sup>, WUCHEN LI<sup>‡</sup>, AND STANLEY OSHER<sup>‡</sup>

**Abstract.** We present a primal-dual method to solve  $L^1$ -type nonsmooth optimization problems independently of the grid size. We apply these results to two important problems: the Rudin–Osher–Fatemi image denoising model and the  $L^1$  earth mover’s distance from optimal transport. Crucially, we provide analysis that determines the choice of optimal step sizes and we prove that our method converges independently of the grid size. Our approach allows us to solve these problems on grids as large as  $4096 \times 4096$  in a few minutes without parallelization.

**Key words.** total variation denoising, earth mover’s distance, optimal transport, primal-dual algorithm, grid size independence

**AMS subject classifications.** 49M29, 65K10

**DOI.** 10.1137/18M118640X

**1. Introduction.** In recent years there has been an explosion of interest ([12, 4, 18, 16, 5, 17, 20] and many more) in solving convex optimization problems using first-order algorithms. The primary advantage of first-order algorithms (as compared to, say, Newton’s method) is that one need only evaluate the proximal operator or the gradient of the functional at the current position. As a result, the complexity of each iteration is typically linear in the total number of grid points. This opens the door to solving extremely large problems, which would be infeasible with other methods.

However, such a viewpoint often sweeps under the rug that the convergence rate of first-order methods may depend badly on the size of the problem. This dependence may enter through two competing factors—the distance between the minimizer and the initial point, and the stability of the descent information. These factors are easiest to understand in the context of smooth gradient descent. Indeed, given a smooth convex functional  $F$  with a unique global minimum at  $u^*$ , gradient descent using the inner product  $(\cdot, \cdot)_\mathcal{H}$  has the convergence rate

$$(1.1) \quad F(u_n) \leq F(u^*) + 2L_\mathcal{H} \frac{\|u^* - u_0\|_\mathcal{H}^2}{n+4},$$

where  $u_n$  is the  $n$ th iterate,  $u_0$  is the initial point, and  $L_\mathcal{H}$  is the Lipschitz constant of  $\nabla_\mathcal{H} F$  in the norm  $\|\cdot\|_\mathcal{H}$  [19]. Strengthening the inner product  $(\cdot, \cdot)_\mathcal{H}$  decreases  $L_\mathcal{H}$  at the expense of increasing  $\|u^* - u_0\|_\mathcal{H}$  (and vice versa). In the continuum setting, if  $L_\mathcal{H}$  or  $\|u^* - u_0\|_\mathcal{H}$  is infinite, then on a discrete grid the corresponding quantity will grow as the grid resolution becomes finer. In these cases, each iteration of the first-order

\*Received by the editors May 10, 2018; accepted for publication (in revised form) March 13, 2019; published electronically May 16, 2019.

<http://www.siam.org/journals/sinum/57-3/M118640.html>

**Funding:** The work of the first, second, and fourth authors was supported by National Science Foundation grant DMS-1737770. The work of the first and second authors was supported by DARPA award FA8750-18-2-0066. The work of the third and fourth authors was supported by DOE grant DE SC0018383. The work of the fourth author was supported by STROBE National Science Foundation grant 1554564.

<sup>†</sup>Corresponding author. Department of Mathematics, UCLA, Los Angeles, CA 90095 ([mjaco@umich.edu](mailto:mjaco@umich.edu)).

<sup>‡</sup>Department of Mathematics, UCLA, Los Angeles, CA 90095 ([flavien@ucla.edu](mailto:flavien@ucla.edu), [wcli@math.ucla.edu](mailto:wcli@math.ucla.edu), [sjo@math.ucla.edu](mailto:sjo@math.ucla.edu)).

method is extremely efficient, but the number of required iterations depends on the problem size. This can place a severe restriction on the size of solvable problems.

The situation appears to be particularly dire for pathological problems where at least one of  $L_{\mathcal{H}}$  or  $\|u^* - u_0\|_{\mathcal{H}}$  is infinite for *any* choice of inner product. In this case, (1.1) would suggest that it is not possible to obtain a convergence rate independent of the grid size. Our goal in this paper is to show that this is in fact not the case—even for pathological problems the convergence rates of first-order methods can be made independent of the problem size.

Our approach is inspired by a more powerful convergence rate estimate given by Nesterov in [19]. In addition to an accelerated convergence rate, Nesterov's estimate sequence framework reveals that for smooth convex  $F$  one has

$$(1.2) \quad F(u_n) \leq \min_u \left[ F(u) + 4L_{\mathcal{H}} \frac{\|u - u_0\|_{\mathcal{H}}^2}{(n+2)^2} \right].$$

This estimate gives far more flexibility, as we can attempt to approximate the minimizer  $u^*$  with a sequence  $\{u_R\}_{R>0}$  where each  $u_R$  satisfies  $u_R \in \arg \min_{\|u-u_0\| \leq R} F(u)$ . If we then let  $\delta_F(R) = F(u_R) - F(u^*)$  we see that

$$(1.3) \quad F(u_n) - F(u^*) \leq \delta_F(R) + 4L_{\mathcal{H}} \frac{R^2}{(n+2)^2}.$$

As long as  $\delta_F(R) \rightarrow 0$  as  $R \rightarrow \infty$ , we can choose  $R$  and  $n$  so that the right-hand side of (1.3) is as small as desired. This perspective makes it clear that  $L_{\mathcal{H}} < \infty$  should be prioritized over  $\|u^* - u_0\|_{\mathcal{H}} < \infty$  when choosing an inner product. More importantly, we can see that the convergence rate can be made independent of the problem size.

In this paper, we are interested in  $L^1$ -type problems where the functional  $F$  is not smooth. As such, we must consider methods which can handle nondifferentiable functions. A powerful framework for nonsmooth optimization is given by primal-dual splitting schemes. Primal-dual algorithms convert minimization problems of the form

$$(1.4) \quad F(u) = f(Ku) + g(u)$$

into saddle point problems

$$(1.5) \quad \mathcal{L}(u, p) = (Ku, p)_{\mathcal{Z}} + g(u) - f^*(p),$$

where  $f$  and  $g$  are convex functions,  $K : \mathcal{H} \rightarrow \mathcal{Z}$  is a linear map between Hilbert spaces, and  $f^*$  is the convex dual of  $f$ . If one can easily compute the proximal operators of  $f$  and  $g$ , then there are many efficient algorithms for finding the saddle point of (1.5) such as Douglas–Rachford splitting, augmented Lagrangian, the alternating direction of multipliers method, split Bregman, PDHG, and Nesterov's excessive gap method [7, 15, 11, 10, 12, 17].

In this paper, we work with a modified version of Chambolle and Pock's primal-dual hybrid gradient algorithm (PDHG) [4], which we call G-prox PDHG; see below. (We pause here to note that through various reductions G-prox PDHG can be shown to be equivalent to the well-known Douglas–Rachford splitting algorithm.) We also note that we could have carried out our analysis and results by building upon Nesterov's excessive gap technique [17] instead of using PDHG.

Both PDHG and G-prox PDHG search for the saddle point of (1.5) by alternating proximal updates of the primal and dual variables. The key difference between G-prox PDHG and the original PDHG algorithm is that our  $u$  update equation uses the

---

**Algorithm.** G-prox PDHG.

---

$$\begin{aligned} u_{n+1} &= \arg \min_{u \in \mathcal{H}} g(u) + (Ku, \bar{p}_n)_{\mathcal{Z}} + \frac{1}{2\tau} \|K(u - u_n)\|_{\mathcal{Z}}^2, \\ p_{n+1} &= \arg \max_{p \in \mathcal{Z}} -f^*(p) + (Ku_{n+1}, p)_{\mathcal{Z}} - \frac{1}{2\sigma} \|(p - p_n)\|_{\mathcal{Z}}^2, \\ \bar{p}_{n+1} &= 2p_{n+1} - p_n. \end{aligned}$$


---

generalized prox term  $\|K(u - u_n)\|_{\mathcal{Z}}^2$  as opposed to  $\|u - u_n\|_{\mathcal{Z}}^2$  for the original algorithm. This can be understood as preconditioning the  $u$  descent direction with the operator  $(K^T K)^{-1}$ . As a result, this scheme has a much more satisfactory stability condition. The primal-dual step size parameters  $\tau$  and  $\sigma$  only need to satisfy

$$\tau\sigma < 1$$

compared to  $\tau\sigma < \frac{1}{\|K^T K\|_{\mathcal{H}}}$  for the original algorithm. When  $K$  is an unbounded operator, say,  $K = \nabla$  and  $\mathcal{H} = \mathcal{Z} = L^2$ , the step sizes of the discrete PDHG algorithm must depend on the grid resolution. On the other hand, the step sizes of the discrete G-prox PDHG algorithm will be clearly independent of the grid size. As one might expect from our exposition above, we must pay for the increased stability by increasing the distance between the solution  $u^*$  and the initial point  $u_0$ . Indeed this is the case; the convergence rate will now depend on  $\|K(u^* - u_0)\|_{\mathcal{Z}}$  as opposed to  $\|u^* - u_0\|_{\mathcal{H}}$  for the original PDHG. However, this trade-off is worth it. We shall show in section 3 (cf. Theorem 3.1) that under certain technical conditions the averaged sequence of primal iterates  $u^N = \frac{1}{N} \sum_{n=1}^N u_n$  satisfies

$$(1.6) \quad F(u^N) \leq \min_u \left[ F(u) + \frac{C\|K(u - u_0)\|_{\mathcal{H}}}{N} \right]$$

for some constant  $C < \infty$ . This estimate shows that an approximate solution to the optimization problem can be obtained independently of the grid size as long as

$$\delta_F(R) = \min_{\|K(u - u_0)\|_{\mathcal{Z}} \leq R} F(u) - F(u^*)$$

goes to zero as  $R \rightarrow \infty$ .

In order to obtain the convergence rate given in (1.6) the step sizes  $\tau$  and  $\sigma$  must be chosen optimally. Note this is nontrivial as the stability condition  $\tau\sigma < 1$  has a degree of freedom. As it turns out, the optimal choices of  $\tau$  and  $\sigma$  are highly dependent on the properties of the functional  $F$ , the underlying space  $\mathcal{H}$ , and the primal and dual solutions  $u^*$  and  $p^*$ , respectively. Furthermore, we shall see that the optimal choices of  $\tau$  and  $\sigma$  may depend on the user's desired error tolerance. For example, the optimal step sizes used to find an  $\epsilon$  accurate solution may be different from the optimal step sizes used to find an  $\epsilon/2$  accurate solution!

In the face of such a complication, it seems unlikely that there is an elegant or concise statement which provides the optimal convergence rate and optimal step sizes for general  $F$ . Instead, we focus on two important problems: the Rudin–Osher–Fatemi (ROF) image denoising model and the earth mover's distance (EMD) between two probability measures. Both of these problems can be solved very efficiently with

our method, as the matrix inversion  $(K^T K)^{-1}$  can be carried out in log-linear time using the fast Fourier transform (FFT). For both of these problems, we provide a principled strategy for choosing approximately optimal step sizes  $\tau$  and  $\sigma$  and give an explicit upper bound for the convergence rate in terms of the number of iterations  $N$ . Although other works have considered solving these problems with preconditioned algorithms, their convergence rates have not been independent of the grid size [1]. Thus, to the best of our knowledge, this paper provides the first proof that these problems can be solved with a convergence rate independent of the grid discretization. In addition, our arguments give a blueprint for extending the convergence rate results to other functionals of interest.

The rest of the paper is structured as follows. We conclude the introduction with a summary of our contributions. Next, in section 2, we provide background on convex optimization and introduce further notation. In section 3, we prove the main results of the paper. In section 4, we perform various numerical experiments that highlight the need for our rigorous theoretical analysis. Finally, in section 5, we conclude the paper with a brief discussion.

**Contributions.** The following is a summary of the present paper's contributions:

- We conclusively demonstrate that the ROF problem and the  $L^1$  EMD problem can be solved with a convergence rate independent of the grid discretization. Furthermore, our arguments apply to a general class of  $L^1$  functionals. Surprisingly, these results are nontrivial and require a detailed analysis.
- Crucially, our analysis provides the optimal step sizes for splitting algorithms. As a result, we are able to solve these problems orders of magnitude faster than previous works. We can solve problems on grids of size  $2048 \times 2048$  in less than 2 minutes and grids of size  $4096 \times 4096$  in less than 10 minutes without parallelization.

**2. Background and notation.** In this paper we will be interested in convex functionals  $F : X \rightarrow \mathbb{R} \cup \{+\infty\}$ , where  $X$  is a convex subset of the space of functions  $\{u : [0, 1]^d \rightarrow \mathbb{R}\}$ . In order to minimize  $F$ , one typically needs to add additional structure in the form of a distance. The distance is used to control how far the scheme is allowed to move using only local information about  $F$ . In principle, these distances can be very general [6]. Here, we will focus on the case where the distance is induced by an inner product  $(\cdot, \cdot)_{\mathcal{H}}$ . Thus, it will be useful for us to assume that  $F$  is defined on some Hilbert space  $\mathcal{H}$  with the inner product  $(\cdot, \cdot)_{\mathcal{H}}$ .

Typically, there is an enormous amount of freedom in choosing the Hilbert space  $\mathcal{H}$  (it is usually easy to extend  $F$  to a larger space or restrict it to a smaller space). In the introduction, we alluded to the fact that there may not be a single Hilbert space  $\mathcal{H}$  which is the “natural” or “correct” choice. Ideally, one should choose a Hilbert space where  $\nabla F$  is Lipschitz continuous and  $F$  is coercive in the norm  $\|\cdot\|_{\mathcal{H}}$ . This is enough to imply that  $F$  has a minimizer  $u^* \in \mathcal{H}$  [8] and that  $\nabla F$  behaves stably in local neighborhoods. However, for many interesting functionals, no such “natural” Hilbert space exists. For example, there is no Hilbert space where the total variation functional is both Lipschitz continuous and coercive. Thus, when choosing an inner product we must be aware of the trade-offs that such a choice entails.

For nonsmooth functionals, we cannot use gradient descent to search for a minimizer. Instead, we turn to the proximal operator of  $F$

$$(2.1) \quad \text{prox}_{\tau}(F, u) = \arg \min_{u' \in \mathcal{H}} F(u') + \frac{1}{2\tau} \|u' - u\|_{\mathcal{H}}^2,$$

which is well-defined when  $F$  is merely lower semicontinuous and bounded below on  $\mathcal{H}$  [8]. Roughly speaking, the proximal operator searches for the smallest value of  $F$  in a neighborhood of the current point  $u$ . When  $F$  is smooth, we know that  $\text{prox}_\tau(F, u) \approx u - \tau \nabla_{\mathcal{H}} F(u)$  for small  $\tau$ . Thus, the proximal operator generalizes the notion of gradient descent to nonsmooth functionals.

Unfortunately, computing the proximal operator of a nontrivial functional  $F$  is extremely difficult. Indeed, one should expect that computing  $\text{prox}_\tau(F, u)$  is as difficult as finding a minimizer of  $F$ . On the other hand, a large class of nontrivial functionals  $F$  can be written as a sum

$$(2.2) \quad F(u) = f(Ku) + g(u),$$

where the proximal operators of  $f(u)$  and  $g(u)$  can be computed easily. This leads to a large class of closely related algorithms (Douglas–Rachford splitting, augmented Lagrangian, alternating direction of multipliers method, split Bregman, PDHG, Nesterov's excessive gap method) [7, 15, 11, 10, 12, 17], which minimize  $F$  by appropriately combining the proximal operators of  $f$  and  $g$ . For the remainder of this paper we shall focus on the PDHG algorithm.

### 2.1. PDHG.

PDHG converts the minimization problem

$$(2.3) \quad F(u) = f(Ku) + g(u)$$

into the primal-dual saddle point problem

$$(2.4) \quad \mathcal{L}(u, p) = (Ku, p)_Z + g(u) - f^*(p).$$

$f^*$  is the convex dual of  $f$ , which is defined by the Legendre–Fenchel transform:

$$(2.5) \quad f^*(p) = \sup_{v \in Z} (v, p) - f(v).$$

For convex functions  $f$ , the Legendre–Fenchel transform is an involution  $f^{**} = f$ . Therefore,  $F$  can be recovered from  $\mathcal{L}$  by

$$F(u) = \sup_{p \in Z} \mathcal{L}(u, p) - f^*(p).$$

If  $F$  has a unique minimizer  $u^*$  and  $\mathcal{L}$  has a saddle point  $(\hat{u}, \hat{p})$ , then  $\hat{u} = u^*$ . Indeed,

$$F(\hat{u}) = \sup_{p \in Z} \mathcal{L}(\hat{u}, p) = \mathcal{L}(\hat{u}, \hat{p}) \leq \mathcal{L}(u^*, \hat{p}) \leq F(u^*).$$

Therefore, instead of directly minimizing  $F$ , we can achieve the same goal by searching for a saddle point of  $\mathcal{L}$ .

To proceed further, we must know what it means for a point  $(u, p)$  to be close to a saddle point. A notion of closeness can be defined through the primal-dual gap:

$$\mathcal{G}(u, p) = \sup_{p' \in Z} \mathcal{L}(u, p') - \inf_{u' \in H} \mathcal{L}(u', p).$$

By definition,  $\mathcal{G}(u, p) \geq 0$  for all  $(u, p)$ . Furthermore,  $\mathcal{G}(\hat{u}, \hat{p}) = 0$  if and only if  $(\hat{u}, \hat{p})$  is a saddle point. Indeed,  $\mathcal{L}$  is convex in  $u$  for fixed  $p$  and concave in  $p$  for fixed  $u$ , thus the inequalities

$$\sup_{p' \in H} \mathcal{L}(\hat{u}, p') \leq \mathcal{L}(\hat{u}, \hat{p}) \leq \inf_{u' \in H} \mathcal{L}(u', \hat{p})$$

encode the definition of a saddle point. In addition, the primal-dual gap controls how close one is to the minimizer of  $F$ , namely,

$$\mathcal{G}(u, p) = F(u) - \inf_{u' \in \mathcal{H}} \mathcal{L}(u', p) \geq F(u) - \inf_{u' \in \mathcal{H}} F(u').$$

Now we are ready to discuss the PDHG algorithm. PDHG searches for a saddle point of  $\mathcal{L}$  as follows.

---

**Algorithm 1.** PDHG.

---

$$(2.6) \quad u_{n+1} = \arg \min_{u \in \mathcal{H}} g(u) + (u, K^T \bar{p}_n)_\mathcal{H} + \frac{1}{2\tau} \|u - u_n\|_\mathcal{H}^2,$$

$$(2.7) \quad p_{n+1} = \arg \max_{p \in \mathcal{Z}} -f^*(p) + (Ku_{n+1}, p)_\mathcal{Z} - \frac{1}{2\sigma} \|(p - p_n)\|_\mathcal{Z}^2,$$

$$(2.8) \quad \bar{p}_{n+1} = 2p_{n+1} - p_n.$$


---

The main source of instability in the PDHG algorithm is the decoupling of the  $u$  and  $p$  update steps. The scheme is stable if  $\tau\sigma\|K^T K\|_\mathcal{H} < 1$  [4]. However, if  $K$  is an unbounded operator from  $\mathcal{H}$  to  $\mathcal{Z}$ , there are no nonzero step sizes which produce a stable scheme. Thus, we see that the underlying Hilbert spaces  $\mathcal{H}$  and  $\mathcal{Z}$  play a crucial role in the stability of the algorithm.

We conclude the background section with an important result of Chambolle and Pock which provides a convergence rate for the PDHG algorithm. The convergence rate is given in terms of a slightly unusual object, the partial primal-dual gap

$$(2.9) \quad \mathcal{G}_{R_1, R_2}(u, p) = \sup_{\|p' - p_0\|_\mathcal{Z} \leq R_1} \mathcal{L}(u, p') - \inf_{\|u' - u_0\|_\mathcal{H} \leq R_2} \mathcal{L}(u', p),$$

where  $u_0$  and  $p_0$  are the initial iterates of  $u$  and  $p$ . The partial primal-dual gap restricts the search for maximizers  $p'$  and minimizers  $u'$  to balls of finite radius centered at the initial iterates. As a result, it is possible for the partial primal-dual gap to vanish at non saddle points. However, if  $\mathcal{G}_{R_1, R_2}(\hat{u}, \hat{p})$  vanishes and  $\|\hat{p} - p_0\|_\mathcal{Z} < R_1$ ,  $\|\hat{u} - u_0\|_\mathcal{H} < R_2$ , then  $(\hat{u}, \hat{p})$  is a saddle point [4].

**THEOREM 2.1** (Chambolle and Pock [4]). *Suppose that  $K : \mathcal{H} \rightarrow \mathcal{Z}$  is a bounded operator and the step sizes  $\tau$  and  $\sigma$  satisfy  $\tau\sigma\|K^T K\|_\mathcal{H} < 1$ . Let  $u^N = \frac{1}{N} \sum_{n=1}^N u_n$  and  $p^N = \frac{1}{N} \sum_{n=1}^N p_n$ , where  $u_n$  and  $p_n$  are the sequence of iterates produced by Algorithm 1. After  $N$  iterations the partial primal-dual gap satisfies*

$$(2.10) \quad \mathcal{G}_{R_1, R_2}(u^N, p^N) \leq \frac{1}{2N} \left( \frac{R_1^2}{\tau} + \frac{R_2^2}{\sigma} \right).$$

Formula (2.10) is very interesting. The radii  $R_1$  and  $R_2$  play the same role as the distance term  $\|u - u_0\|_\mathcal{H}^2$  in the gradient descent convergence rate formulas (1.1) and (1.2). Similarly, the step size restriction  $\tau\sigma\|K^T K\|_\mathcal{H} < 1$  plays the same role as the Lipschitz constant  $L_\mathcal{H}$ . Thus, we see that the convergence rate of PDHG depends on the inner products  $(\cdot, \cdot)_\mathcal{H}$  and  $(\cdot, \cdot)_\mathcal{Z}$  in the same way as gradient descent. We shall see shortly that we will be able to use these features to convert Theorem 2.1 into our main result.

**3. Main results.** Let us recall that our goal in this paper is to solve optimization problems with a convergence rate independent of the grid size. If  $K : \mathcal{H} \rightarrow \mathcal{Z}$  is an unbounded operator, then PDHG is not well-defined in the continuous setting. In the discrete approximation,  $K$  will be bounded but the operator norm of  $K$  will grow with the grid size. This implies that at least one of the step sizes  $\tau$  or  $\sigma$  must shrink to zero as the grid resolution approaches the continuous limit. In that case, it is clear from formula (2.10) that the convergence rate will depend on the grid size. Thus, our immediate goal is to modify PDHG to ensure that  $K$  is always a bounded operator.

Assume that the inner product  $(\cdot, \cdot)_\mathcal{Z}$  for the variable  $p$  has already been chosen. The simplest way to ensure that  $K : \mathcal{H} \rightarrow \mathcal{Z}$  will be a bounded operator is to define the inner product  $(\cdot, \cdot)_\mathcal{H}$  by  $(u, v)_\mathcal{H} = (Ku, Kv)_\mathcal{Z}$  (note we can always assume that  $K$  is injective—it is trivial to solve for and eliminate the components of  $u$  which are not coupled to  $p$ ). This simple modification leads us to Algorithm 2, G-prox PDHG.

We again note that G-prox PDHG is equivalent to the Douglas–Rachford splitting algorithm (cf. section 4.2 in [4]). However, we prefer this formulation as it allows us to directly use the convergence result (2.10). G-prox PDHG modifies Algorithm 1 by choosing a specific inner product for the  $u$  update. Thus, G-prox PDHG is a special case of Algorithm 1 where  $K$  is a bounded operator with operator norm  $\|K^T K\|_\mathcal{H} = 1$ . As long as  $\tau\sigma < 1$ , the convergence result, Theorem 2.1, applies to G-prox PDHG.

---

**Algorithm 2.** G-prox PDHG.

---

$$(3.1) \quad u_{n+1} = \arg \min_{u \in \mathcal{H}} g(u) + (Ku, \bar{p}_n)_\mathcal{Z} + \frac{1}{2\tau} \|K(u - u_n)\|_\mathcal{Z}^2,$$

$$(3.2) \quad p_{n+1} = \arg \max_{p \in \mathcal{Z}} -f^*(p) + (Ku_{n+1}, p)_\mathcal{Z} - \frac{1}{2\sigma} \|(p - p_n)\|_\mathcal{Z}^2,$$

$$(3.3) \quad \bar{p}_{n+1} = 2p_{n+1} - p_n.$$


---

Now we still need to address the choice of the Hilbert space  $\mathcal{Z}$  for the dual variable  $p$  and the impact of the distances  $\|K(u - u_0)\|_\mathcal{Z}$  and  $\|p - p_0\|_\mathcal{Z}$  on the convergence rate. The highest priority is to choose  $(\cdot, \cdot)_\mathcal{Z}$  so that the updates (3.1) and (3.2) can be computed efficiently. Indeed, if steps (3.1) and (3.2) cannot be computed in linear or log-linear time (in the number of grid points), then the entire purpose of choosing a first-order method is lost. For this reason, in the remainder of this paper we shall always take  $(\cdot, \cdot)_\mathcal{Z}$  to be the usual  $L^2$  inner product on a domain. Thus, our inner products shall always be given by

Primal inner product:

$$(u, v)_\mathcal{H} = (Ku, Kv)_{L^2},$$

Dual inner product:

$$(p, q)_\mathcal{Z} = (p, q)_{L^2}.$$

Note that there may be other specific problems where a different choice of inner product is more appropriate.

Now we are ready to state and prove Theorem 3.1, which shows that for certain problems the convergence rate of G-prox PDHG is independent of the problem size even when the distance  $\|K(u^* - u_0)\|_\mathcal{Z}$  is infinite.

**THEOREM 3.1** (convergence of G-prox PDHG). *Suppose that  $F$  is a functional on a Hilbert space  $\mathcal{H}$  of the form*

$$F(u) = f(Ku) + g(u),$$

where  $f$  and  $g$  are convex functions. Furthermore suppose that for all  $u \in \mathcal{H}$  the maximizer  $p(u) \in \arg \max_{p \in \mathcal{Z}} (Ku, p)_\mathcal{Z} - f^*(p)$  is uniformly bounded:

$$\sup_{u \in \mathcal{H}} \|p(u)\|_\mathcal{Z} = C < \infty.$$

Let  $u^N = \frac{1}{N} \sum_{n=1}^N u_n$ , where  $u_n$  is the sequence of iterates produced by G-prox PDHG starting from  $u_0$ . Then there is an optimal choice of step sizes  $\tau$  and  $\sigma$ , satisfying  $\sigma\tau < 1$  such that after  $N$  iterations

$$F(u^N) \leq \min_u \left[ F(u) + \frac{C\|K(u - u_0)\|_\mathcal{Z}}{N} \right].$$

Furthermore, if  $\lim_{R \rightarrow \infty} \min_{\|K(u - u_0)\|_\mathcal{Z} \leq R} F(u) = \inf_{u \in \mathcal{H}} F(u) = \bar{F}$ , then  $\lim_{N \rightarrow \infty} F(u^N) = \bar{F}$ .

*Remark 1.* Although the boundedness condition on the dual variable

$$\sup_{u \in \mathcal{H}} \|p(u)\|_\mathcal{Z} = C < \infty$$

appears to be very restrictive, it is trivially satisfied if  $p$  arises from the dual of an  $L^1$  norm. Indeed, if  $f(u) = \|u\|_{L^1}$ , then

$$f^*(p) = \begin{cases} 0 & \text{if } \|p\|_\infty \leq 1, \\ \infty & \text{otherwise.} \end{cases}$$

Thanks to this observation we shall be able to apply this theorem to the two main problems we are interested in.

*Proof.* From (2.10) and the definition of  $C$  we have

$$\mathcal{G}_{C,R}(u^N, p^N) = F(u^N) - \min_{\|Ku\|_\mathcal{Z} \leq R} [g(u) + (Ku, p^N) - f^*(p^N)] \leq \frac{1}{2N} \left( \frac{R^2}{\tau} + \frac{C^2}{\sigma} \right).$$

Now we wish to estimate the second term on the left-hand side with a quantity related to  $F$ . Immediately we can see that

$$\begin{aligned} \min_{\|K(u - u_0)\|_\mathcal{Z} \leq R} [g(u) + (Ku, p^N) - f^*(p^N)] &\leq \min_{\|K(u - u_0)\|_\mathcal{Z} \leq R} \max_{p \in \mathcal{Z}} [g(u) + (Ku, p) - f^*(p)] \\ &= \min_{\|K(u - u_0)\|_\mathcal{Z} \leq R} F(u). \end{aligned}$$

Putting things together we have

$$F(u^N) - \bar{F} \leq \frac{1}{2N} \left( \frac{R^2}{\tau} + \frac{C^2}{\sigma} \right) + \min_{\|K(u - u_0)\|_\mathcal{Z} \leq R} F(u) - \bar{F}.$$

For any fixed  $R$  the best choice of the step sizes  $\tau$  and  $\sigma$  is to take  $\tau = \frac{R}{C}$  and  $\sigma = \frac{C}{R}$ , which gives

$$F(u^N) - \bar{F} \leq \frac{RC}{N} + \min_{\|K(u - u_0)\|_\mathcal{Z} \leq R} F(u) - \bar{F}.$$

Since  $R$  is arbitrary, we can minimize the right-hand side over  $R \geq 0$  to get the first result. For the second result, it is enough to let  $R = R(N)$  grow to infinity with  $N$  such that  $\lim_{N \rightarrow \infty} \frac{RC}{N} = 0$ .  $\square$

By examining the above proof, we see that the rate of convergence is governed by the gap

$$\delta_F(R) = \min_{\|K(u-u_0)\|_2 \leq R} F(u) - \bar{F}.$$

Obtaining explicit upper bounds for  $\delta_F(R)$  is of great practical importance. The behavior of this quantity informs our choice of step sizes and thus directly impacts the performance of the algorithm. We cannot expect to give a general statement on the behavior of  $\delta_F(R)$ , as it is highly dependent on the properties of the functional  $F$ . Instead, we will closely analyze two important problems, total variation denoising and the EMD. For these two problems we shall provide explicit upper bounds on a convergence rate of G-prox PDHG and we shall show how to choose the optimal step sizes  $\sigma$  and  $\tau$ .

**3.1. Total variation denoising of images.** Image processing is a source of many important large-scale problems. Simple consumer devices, such as cell phone cameras, have pixel counts in the tens of millions. More importantly, medical images such as MRI scans may be three-dimensional images of physical objects. The pixel counts of three-dimensional images grow cubically with the resolution; thus even relatively low-resolution three-dimensional images have enormous pixel counts.

Digital images are defined on either two or three dimensional grids. At each grid point, the image takes a value in  $[0, 1]$ , which represents the brightness of the image at that location. In what follows, we shall consider images as discrete approximations to a function  $I : B \rightarrow [0, 1]$ , where  $B$  is a rectangle in the plane or a box in 3-space. By rescaling the side lengths, we shall always assume that  $I$  is defined on the unit cube  $[0, 1]^d$ .

A fundamental problem in image processing is image denoising. The goal of image denoising is to remove pixel errors while preserving as much of the original image information as possible. The most important information is typically contained in the edges of objects and scenery. Mathematically, edges correspond to sharp discontinuities in the image intensity function. Thus, variational models for image denoising must be able to produce discontinuous solutions.

A popular model for image denoising is the ROF model [22]

$$(3.4) \quad F_\lambda(u, I) = \|u\|_{TV} + \frac{\lambda}{2} \|u - I\|_{L^2}^2,$$

where  $I$  is the original image to be denoised and  $\|u\|_{TV}$  is the total variation of  $u$ . For smooth functions,  $\|u\|_{TV} = \|\nabla u\|_{L^1}$ .

We shall consider the saddle point formulation:

$$(3.5) \quad (\nabla u, p)_{L^2} + \frac{\lambda}{2} \|u - I\|_{L^2}^2 - \chi_\infty(p).$$

Here  $\chi_\infty(p)$  is the convex indicator function of the  $L^\infty$  unit ball, i.e.,

$$\chi_\infty(p) = \begin{cases} 0 & \text{if } |p(x)| \leq 1 \text{ for all } x \in [0, 1]^d, \\ \infty & \text{otherwise.} \end{cases}$$

Clearly,  $p$  will always have the  $L^2$  norm bounded by 1; thus the ROF problem satisfies the hypotheses of Theorem 3.1 with  $C = 1$ .

We shall use the prox term  $\|\nabla(u - u^n)\|_{L^2([0,1]^2)}^2$  for the primal updates and  $\|p - p^n\|_{L^2([0,1]^2)}^2$  for the dual updates. Thus, the G-prox PDHG updates for the ROF problem have the form

$$(3.6) \quad u_{n+1} = (\lambda\tau\text{Id} - \Delta)^{-1}(\lambda\tau I + \tau\nabla \cdot \bar{p}_n - \Delta u_n),$$

$$(3.7) \quad p_{n+1}(x) = \frac{p_n(x) + \sigma\nabla u_{n+1}(x)}{\max(1, |p_n(x) + \sigma\nabla u_{n+1}(x)|)},$$

$$(3.8) \quad \bar{p}_{n+1} = 2p_{n+1} - p_n,$$

where the identity matrix  $\text{Id}$  should not be confused with the image to be denoised  $I$ . Note that the  $u$  update can be conveniently expressed as a matrix vector product, whereas it is more convenient to express the  $p$  update in a pointwise fashion.

Minimizers of the ROF model are functions of bounded variation (BV). BV functions may have discontinuities along curves; thus the model will preserve the edges of  $I$  for  $\lambda$  sufficiently large but finite. For example, if  $I$  is the characteristic function of a disc of radius  $r$  and  $\lambda > \frac{2}{r}$ , then the solution is the still discontinuous function  $u^*(x) = (1 - \frac{2}{\lambda r})I(x) + \frac{2\pi r}{\lambda(1-\pi r^2)}$  [3]. As a result, the unique minimizer  $u^*$  of the ROF model is not in general an element of the Hilbert space  $H^1([0, 1]^d) = \{u \in L^2([0, 1]^d) : \nabla u \in L^2([0, 1]^d)\}$ . Thus, ROF is not coercive in the  $H^1$  norm and we shall have to compute the gap  $\min_{\|\nabla(u-u_0)\|_{L^2} \leq R} \text{ROF}_\lambda(u, I) - \text{ROF}_\lambda(u^*, I)$  to obtain a convergence rate for G-prox PDHG.

**PROPOSITION 3.2.** *Given an image  $I$  taking values in  $[0, 1]$ , the decay of the ROF gap is bounded by*

$$(3.9) \quad \min_{\|\nabla u\|_{L^2} \leq R} \text{ROF}_\lambda(u, I) - \text{ROF}_\lambda(u^*, I) \leq \frac{3\lambda d^2}{2R^2} \|u^*\|_{TV}^2$$

when  $u_0 = 0$ .

*Proof.* We estimate the gap by constructing approximate minimizers  $u_\delta$  of the ROF functional such that  $u_\delta$  has finite  $H^1$  norm. Our trick is to take the solution  $u^*$  and convolve it with the Gaussian approximation to the identity  $G_\delta(z) = \delta^{-d} e^{-\pi(z/\delta)^2}$  (note that convolutions can be appropriately defined on  $[0, 1]^d$ ; see the appendix for details).

If we let  $u_\delta = G_\delta * u^*$ , then  $u_\delta$  is a  $C^\infty$  function and thus an element of  $H^1$ . Adding and subtracting  $u_\delta$  into the  $L^2$  term we get

$$\text{ROF}_\lambda(u_\delta, I) = \|u_\delta\|_{TV} + \frac{\lambda}{2} \|u^* - I\|_{L^2}^2 + \frac{\lambda}{2} \|u_\delta - u^*\|_{L^2}^2 + \lambda(u^* - I, u_\delta - u^*)_{L^2}.$$

Using Jensen's inequality we can pull the convolution out of the  $TV$  norm and get  $\|u_\delta\|_{TV} \leq \|u^*\|_{TV}$ . If we then apply Holder's inequality to  $\lambda(u^* - I, u_\delta - u^*)$  we get

$$\text{ROF}_\lambda(u_\delta, I) \leq \text{ROF}_\lambda(u^*, I) + \frac{\lambda}{2} \|u_\delta - u^*\|_{L^2}^2 + \lambda \|u^* - I\|_{L^\infty} \|u_\delta - u^*\|_{L^1}.$$

The minimizer of the ROF problem satisfies a maximum principle; therefore we know that  $u^*(x) \in [0, 1]$  for all  $x$  [2]. It only remains to estimate the decay of  $\|u_\delta - u^*\|_{L^q}^q$  and the growth of  $\|\nabla u_\delta\|_{L^2}^2$ . See the appendix for the details on these computations.  $\square$

With Proposition 3.2 in hand, we can now give an upper bound on the convergence rate of G-prox PDHG applied to the ROF model.

**THEOREM 3.3.** *An  $\epsilon$  approximate solution to the ROF model may be obtained in at most*

$$N = \frac{d\sqrt{3\lambda} \|u^*\|_{TV}}{\epsilon^{3/2}}$$

*iterations of G-prox PDHG using the step sizes*

$$\tau = \frac{d\sqrt{3\lambda} \|u^*\|_{TV}}{\epsilon^{1/2}}, \sigma = \tau^{-1}.$$

Note that the step size in Theorem 3.3 depends on  $\|u^*\|_{TV}$ , which is unknown until the problem is solved. We can remedy this by providing a simple estimate for  $\|u^*\|_{TV}$  in terms of  $I$ . Let  $I_0$  be the average value of  $I$  over the domain. By evaluating the functional at either  $I$  or the constant  $I_0$ , we know that  $\text{ROF}_\lambda(u^*, I) \leq \min(\|I\|_{TV}, \frac{\lambda}{2}\|I_0 - I\|_{L^2}^2)$ . This bound immediately implies  $\|u^*\|_{TV} \leq \min(\|I\|_{TV}, \frac{\lambda}{2}\|I_0 - I\|_{L^2}^2)$ . Thus, we have a strategy for choosing the step sizes using only quantities available at the start of computation.

Finally, we conclude this section with a convergence result for the infinite dimensional ROF problem.

**COROLLARY 3.4.** *Let  $u^N = \frac{1}{N} \sum_{n=1}^N u_n$ , where  $u_n$  is the sequence of primal variables produced by G-prox PDHG. Then  $u^N$  converges to the minimizer  $u^* \in BV$  of the ROF problem strongly in  $L^2([0, 1]^d)$ .*

*Proof.* The  $\text{ROF}_\lambda$  functional is  $\lambda$ -strongly convex with respect to the  $L^2$  distance. Therefore,

$$\text{ROF}_\lambda(u, I) - \text{ROF}_\lambda(u^*, I) \geq \frac{\lambda}{2} \|u - u^*\|_{L^2}^2.$$

The convergence

$$\lim_{N \rightarrow \infty} \text{ROF}_\lambda(u^N, I) - \text{ROF}_\lambda(u^*, I) = 0$$

then gives the result.  $\square$

**3.2. Earth mover's distance.** The EMD is a statistical distance on probability measures. Given probability measures  $\rho^1$  and  $\rho^0$  defined on a space  $\Omega$ , the EMD measures the minimal cost required to move the distribution of  $\rho^0$  onto the distribution of  $\rho^1$ . The cost is measured according to a predetermined function  $c(x, y)$ , which gives the expense of transporting a unit of mass at location  $x \in \Omega$  to location  $y \in \Omega$ . Nowadays, the EMD plays important roles in machine learning, image retrieval, and image segmentation [14, 23, 21, 24]. This widespread usage is due to the fact that the EMD incorporates the geometry of the underlying space  $\Omega$  (via the cost function).

We shall concentrate on the (important) special case where  $\Omega = [0, 1]^d$  and the cost function is the usual Euclidean norm,  $c(x, y) = |x - y|$ . We shall assume that the probability measures  $\rho^1, \rho^0$  are elements of the dual space  $C([0, 1]^d)^*$ . Furthermore we assume that there exists a compact set  $K \subset (0, 1)^d$  such that  $\rho^1(K) = \rho^0(K) = 1$ . In this setting, the EMD coincides with the following convex optimization problem:

$$(3.10) \quad \text{EMD}(\rho^1, \rho^0) = \min_{\nabla \cdot m = \rho^1 - \rho^0} \int_{[0, 1]^d} |m|,$$

where  $m$  is a  $d$ -dimensional vector-valued measure satisfying  $m \cdot n = 0$  on the boundary and  $|\cdot|$  is the 2 norm on  $d$ -dimensional vectors.

Using the Hodge decomposition we can decompose  $m = u + \nabla\psi$ , where  $u$  is a divergence-free vector field and  $\nabla\psi$  is a gradient field. Now we see that  $\nabla \cdot m = \Delta\psi = \rho^1 - \rho^0$ . If we let  $\psi$  solve the Poisson equation  $\Delta\psi = \rho^1 - \rho^0$  (with zero Neumann boundary conditions) we can rewrite the problem as

$$(3.11) \quad \text{EMD}(\rho^1, \rho^0) = \min_u F(u, \psi) = \min_u \int_{[0,1]^d} |u + \nabla\psi| + \chi_{\nabla^\perp}(u),$$

where  $\chi_{\nabla^\perp}(u)$  is the convex indicator function encoding the constraints  $\nabla \cdot u = 0$  and  $u \cdot n = 0$ . If  $\rho^1, \rho^0$  are singular measures, then  $\psi$  solves the Poisson equation in a weak sense only. Thus,  $\psi$  does not satisfy the usual regularity properties enjoyed by solutions to the Poisson equation. Nonetheless, the Hardy–Littlewood–Sobolev lemma implies that  $\nabla\psi \in L^r([0,1]^d)$  for any  $r < \frac{d}{d-1}$  [13]. Therefore, the right-hand side of (3.11) is well-defined. In general, one can find a vector-valued measure  $u^*$  which minimizes  $F$ ; however, the minimizer need not be unique [9].

Let us briefly note that the EMD problem is closely related to the ROF model when  $d = 2$ . In two dimensions, divergence-free vector fields  $u$  can be written in the form  $u = \nabla^\perp h$ , where  $h$  is a scalar function and  $\nabla^\perp h = (\partial_y h, -\partial_x h)$ . In this case, the EMD distance can be written as the unconstrained minimization problem

$$(3.12) \quad \text{EMD}(\rho^1, \rho^0) = \min_h \int_{[0,1]^2} |\nabla^\perp h + \nabla\psi|.$$

Now we can see that the EMD problem has the same structure as the ROF model, where we are minimizing the Euclidean norm of a differential operator applied to a function.

Returning to (3.11), the saddle point formulation of the EMD problem has the form

$$(3.13) \quad (u + \nabla\psi, p)_{L^2} + \chi_{\nabla^\perp}(u) - \chi_\infty(p),$$

where  $\chi_\infty$  is defined as in (3.1).

The G-prox PDHG updates for the EMD problem have the form

$$(3.14) \quad u_{n+1} = u_n - \tau \mathbb{P}_{\nabla^\perp}(\bar{p}_n),$$

$$(3.15) \quad p_{n+1}(x) = \frac{p_n(x) + \sigma(u_{n+1} + \nabla\psi(x))}{\max(1, |p_n(x) + \sigma(u_{n+1} + \nabla\psi(x))|)},$$

$$(3.16) \quad \bar{p}_{n+1} = 2p_{n+1} - p_n,$$

where  $\mathbb{P}_{\nabla^\perp}$  is the Leray projection onto divergence-free vector fields,

$$(3.17) \quad \mathbb{P}_{\nabla^\perp}(p) = p - \nabla\Delta^{-1}\nabla \cdot p.$$

Again, the  $u$  update can be conveniently expressed as a matrix vector product, whereas it is more convenient to express the  $p$  update in a pointwise fashion.

It is clear from the saddle point formulation that the EMD problem satisfies the hypotheses of Theorem 3.1 with  $C = 1$ . Since minimizers of the EMD functional

are measures, we should not expect the minimizers to have finite  $L^2$  norm. Thus, to obtain a convergence rate for the EMD problem we shall need to estimate the gap

$$\min_{\|u\|_{L^2} \leq R} F(u, \psi) - F(u^*, \psi).$$

Estimating the gap is more difficult than the ROF problem. The regularity of  $u^*$  is dependent on the regularity of the measures  $\rho^1$  and  $\rho^0$ , but also on more complicated geometric properties of  $\rho^1$  and  $\rho^0$ . We shall estimate the gap assuming only that  $\int_{[0,1]^d} |\rho^1 - \rho^0| \leq 2$ . As a result, we will have an upper bound on the gap which is valid for any probability measures but may be too pessimistic when  $\rho^1$  and  $\rho^0$  are “nice.”

Once again, we shall turn to convolutions. Given a minimizer  $u^*$ , we construct approximate minimizers via convolution with the Gaussian kernel  $u_\delta = G_\delta * u^*$ . The convolution takes vector-valued measures to smooth functions, thus we have  $u_\delta \in L^2([0, 1]^d)$ . Here convolutions are an especially important tool as they preserve the divergence-free constraint.

**PROPOSITION 3.5.** *For any probability measures  $\rho^1$  and  $\rho^0$  the EMD gap satisfies*

$$\min_{\|u\|_{L^2} \leq R} F(u, \psi) - F(u^*, \psi) \leq C_d \left( \frac{\text{EMD}(\rho^1, \rho^0)}{R^2} \right)^{\frac{1}{d-1}} \log \left( \frac{R^2}{\text{EMD}(\rho^1, \rho^0)} \right),$$

where  $C_d$  is a constant that depends on the dimension only.

*Proof.* Let  $u^*$  be a minimizer of  $F(u, \psi)$ . Let  $u_\delta = G_\delta * u^*$  and  $\psi_\delta = G_\delta * \psi$ . By the triangle inequality, we have

$$F(u_\delta, \psi) \leq F(u_\delta, \psi_\delta) + \|\nabla \psi_\delta - \nabla \psi\|_{L^1}.$$

Using Jensen’s inequality to pull the convolution out of  $F(u_\delta, \psi_\delta)$  we have

$$F(u_\delta, \psi) \leq F(u^*, \psi) + \|\nabla \psi_\delta - \nabla \psi\|_{L^1}.$$

Thus, we only need to estimate the decay of  $\|\nabla \psi_\delta - \nabla \psi\|_{L^1}$  and the growth of  $\|u_\delta\|_{L^2}$ . In the appendix we show that

$$\|\nabla \psi_\delta - \nabla \psi\|_{L^1} \leq \delta (|\log(\delta)| + 1) C'_d \int_{[0,1]^d} |\rho^1 - \rho^0|$$

and

$$\|u_\delta\|_{L^2} \leq \left( (2\delta)^{1-d} \text{EMD}(\rho^1, \rho^0) \right)^{1/2},$$

where  $C'_d$  is a constant which depends on the dimension only. By using  $\int_{[0,1]^d} |\rho^1 - \rho^0| \leq 2$ , and assuming  $\delta < 1/2$ , we can simplify the bound for  $\|\nabla \psi_\delta - \nabla \psi\|_{L^1}$  to

$$\delta |\log(\delta)| C_d.$$

Putting everything together we get the result.  $\square$

Now we can give an upper bound on the convergence rate of G-prox PDHG applied to the EMD problem.

**THEOREM 3.6.** *Suppose that  $\rho^1$  and  $\rho^0$  are probability measures on  $[0, 1]^d$ . Then  $\text{EMD}(\rho^1, \rho^0)$  can be computed with error at most  $\epsilon$  in*

$$N = \frac{C_d}{\epsilon} \left( \frac{\text{EMD}(\rho^1, \rho^0)^{1/2} \log(1/\epsilon)}{\epsilon} \right)^{(d-1)/2}$$

iterations of G-prox PDHG with the step sizes  $\tau = C_d \left( \frac{\text{EMD}(\rho^1, \rho^0)^{1/2} \log(1/\epsilon)}{\epsilon} \right)^{(d-1)/2}$  and  $\sigma = \tau^{-1}$ , where  $C_d$  is a constant depending on the dimension only.

Note that we do not know the value of  $\text{EMD}(\rho^1, \rho^0)$  until the problem is solved. This is easily dealt with, as we have the estimate  $\text{EMD}(\rho^1, \rho^0) \leq \int_{[0,1]^d} |\rho^1 - \rho^0|$ .

We conclude this section with a convergence result.

**COROLLARY 3.7.** *Let  $u^N = \frac{1}{N} \sum_{n=1}^N u_n$ , where  $u_n$  is the sequence of primal variables produced by G-prox PDHG. Then there exists a subsequence that weakly converges to a measure  $u^*$  which is a minimizer of the EMD functional.*

*Proof.* The sequence  $u^N$  has uniformly bounded total variation. By the Banach-Alaoglu theorem, the sequence has a weak cluster point  $u^*$ . We have established the convergence  $\lim_{N \rightarrow \infty} F(u^N, \psi) - \inf_{u \in L^2} F(u, \psi) = 0$ . The EMD functional is weakly lower semicontinuous, thus

$$F(u^*, \psi) \leq \inf_{u \in L^2} F(u, \psi). \quad \square$$

**4. Numerical experiments.** All numerical algorithms were coded in C and executed on a single 1.6 GHz core with 8 GB RAM. Inversion of the Laplace operator was performed using the FFT. All FFTs were calculated using the free FFTW C library. The code used in this paper is available on GitHub at [https://github.com/majacoma/G\\_prox\\_pdhg](https://github.com/majacoma/G_prox_pdhg).

In this work we do not consider parallelization; however, our approach is still amenable to parallelization. The only step of our method which is nontrivial to parallelize is the computation of the FFT. This is not an insurmountable hurdle, as many modern parallel computing platforms, such as CUDA, have built-in FFT subroutines.

For all of our experiments, given an error tolerance  $\epsilon$ , we run the algorithms until the condition  $F(u_n) - \inf_{u \in \mathcal{H}} F(u) < \epsilon$  is satisfied for the  $n$ th iterate  $u_n$ . If we do not know  $\inf_{u \in \mathcal{H}} F(u)$  in advance, we precompute it by running G-prox PDHG for an extremely large number of iterations to obtain a “ground truth” value. To ensure that this ground truth is sufficiently accurate, we use the primal-dual gap to check that we are within  $\frac{\epsilon}{10}$  of the exact value. Note that the ground truth value may depend on the grid discretization. As a result, we must compute a ground truth value for each grid size that we test.

We will run G-Prox PDHG using a constant multiple of the step sizes from our analysis with one caveat. On a discrete  $d$ -dimensional grid with cell length  $\Delta x$  and  $M = (\frac{1}{\Delta x})^d$  points, all  $L^p$  norms are equivalent up to a factor depending upon  $M$ . Clearly,  $\|u\|_{L^\infty} \leq M \|u\|_{L^1}$  and it then follows that  $\|u\|_{L^p} \leq M^{1-1/p} \|u\|_{L^1}$ . For the ROF problem we have  $\|\nabla_M u\|_{L^2} \leq M^{1/2} \|u\|_{TV}$ , and for the EMD problem we have  $\|u\|_{L^2} \leq M^{1/2} \|u\|_{L^1}$ . This means that the gap  $\delta_F(R)$  will vanish at a finite value  $R_M$ , even though one must take  $R \rightarrow \infty$  in the continuum. Let  $R_\epsilon$  be the optimal choice of  $R$  for solving the continuum problem with error at most  $\epsilon$ . Then in the discrete case, we will choose the step size  $\tau = \min(R_M, R_\epsilon)/C$  (as opposed to the choice  $\tau = R_\epsilon/C$ ). Calculating  $R_M$  exactly is difficult; however, we shall give some simple upper bounds in what follows below.

**4.1. Total variation denoising.** For the total variation denoising problem, we will consider a simple two-dimensional example where the image  $I : [0, 1]^2 \rightarrow [0, 1]$  is the characteristic function of a disc of radius  $1/4$  centered at  $(1/2, 1/2)$ . This allows us to easily create a test image with any desired resolution. For  $\lambda > 8$  the

TABLE 4.1  
*G-Prox PDHG*  $\lambda = 10$ .

Grid size	Error $10^{-2}$		Error $10^{-3}$	
	Iterations	Time (s)	Iterations	Time (s)
$512 \times 512$	33	1.1	61	1.80
$1024 \times 1024$	34	5.0	89	10.3
$2048 \times 2048$	34	21.2	124	58.4
$4096 \times 4096$	34	86.7	168	348.4

TABLE 4.2  
*G-Prox PDHG*  $\lambda = 20$ .

Grid size	Error $10^{-2}$		Error $10^{-3}$	
	Iterations	Time (s)	Iterations	Time (s)
$512 \times 512$	51	1.5	209	4.5
$1024 \times 1024$	66	8.2	232	24.1
$2048 \times 2048$	83	42.4	265	112.6
$4096 \times 4096$	87	186.8	308	586.3

TABLE 4.3  
*G-Prox PDHG*  $\lambda = 20$  *discretization-independent step sizes.*

Grid size	Error $10^{-2}$		Error $10^{-3}$	
	Iterations	Time (s)	Iterations	Time (s)
$512 \times 512$	79	2.0	505	10.3
$1024 \times 1024$	81	9.5	412	44.1
$2048 \times 2048$	83	41.2	396	172.2
$4096 \times 4096$	86	176.9	401	794.5

continuous solution is given by  $u^*(x) = (1 - \frac{8}{\lambda})I(x) + \frac{8\pi}{\lambda(16-\pi)}$ . On a discrete grid with  $M$  points, the minimizer  $u_M^*$  is different (and needs to be calculated) but approaches the continuum solution  $u^*$  as the grid size grows [28]. For our experiments, we take  $\lambda = 10$  and  $20$  and  $\epsilon = 10^{-2}$  and  $10^{-3}$  as error cutoffs.

For our step sizes we will choose  $\tau = \min(\frac{\sqrt{\lambda}\|I\|_{TV}}{\epsilon^{1/2}}, \|\nabla I\|_{L^2})$ . This choice depends only on quantities that are easily estimated at the start of computation. Note that  $\|\nabla I\|_{L^2}$  gives a reasonable upper bound for the discrete grid quantity  $R_M = \|\nabla u_M^*\|_{L^2}$ , and we have dropped the dimensionality constants from Theorem 3.3.

In Tables 4.1 and 4.2, we present the results of G-prox PDHG when  $\lambda = 10$  and  $\lambda = 20$ , respectively. The algorithm converges faster for  $\lambda = 10$ , since less fidelity to the original image is required. This behavior is predicted in our convergence rate analysis, Theorem 3.3, where the rate depends on  $\sqrt{\lambda}$ . When  $\lambda = 10$  and  $\epsilon = 10^{-2}$ , the value of  $\frac{\sqrt{\lambda}\|I\|_{TV}}{\epsilon^{1/2}}$  is smaller than  $\|\nabla I\|_{L^2}$ , for every grid size. As a result, the iteration count is the same for all grid sizes. In the other experiments,  $\frac{\sqrt{\lambda}\|I\|_{TV}}{\epsilon^{1/2}}$  is larger than  $\|\nabla I\|_{L^2}$  on the smaller grids, thus the algorithm converges faster on the smaller grids.

In Table 4.3, we rerun the  $\lambda = 20$  experiment where we do not allow the step sizes  $\tau$  to depend on the grid discretization. In other words we take  $\tau = \frac{\sqrt{\lambda}\|I\|_{TV}}{\epsilon^{1/2}}$  for every grid size. In this experiment, the iteration count stays relatively uniform as the grid size changes. This demonstrates that the algorithm has a convergence rate which is truly independent of the grid size, but when  $\epsilon$  is small one can get faster convergence by taking into account the grid discretization.

TABLE 4.4  
CP2 (from [4]) ROF disc  $\lambda = 10$ .

Grid size	Error $10^{-2}$		Error $10^{-3}$	
	Iterations	Time (s)	Iterations	Time (s)
$512 \times 512$	2196	15.6	4473	31.8
$1024 \times 1024$	4415	156.2	8471	296.2
$2048 \times 2048$	8855	1305.2	17724	2576.3
$4096 \times 4096$	—	—	—	—

TABLE 4.5  
CP2 (from [4]) ROF disc  $\lambda = 20$ .

Grid size	Error $10^{-2}$		Error $10^{-3}$	
	Iterations	Time (s)	Iterations	Time (s)
$512 \times 512$	1252	8.7	2439	17.4
$1024 \times 1024$	2497	80.4	4063	132.4
$2048 \times 2048$	5005	711.0	8044	1133.8
$4096 \times 4096$	10103	6980.2	—	—

Next, we compare G-prox PDHG to an accelerated version of PDHG (Algorithm 2 from [4]), which we will refer to as CP2. CP2 has two advantages over G-prox PDHG. CP2 has extremely simple updates which do not require solving any linear systems. As a result, each iteration of CP2 is faster than G-prox PDHG, which needs to compute an FFT to invert  $(\lambda\tau\text{Id} - \Delta)$ . Second, the primal-dual formulation of the ROF problem

$$\mathcal{L}(u, p) = (\nabla u, p)_{L^2} + \frac{\lambda}{2} \|u - I\|_{L^2}^2 - \chi_\infty(p)$$

is  $L^2$  strongly convex in  $u$ . CP2 computes the  $u$  update in the  $L^2$  norm, thus the  $L^2$  strong convexity can be exploited to accelerate the algorithm. As a result, CP2 has quadratic convergence rate (i.e., the restricted primal-dual gap after  $N$  iterations has decay  $O(1/N^2)$ ). However, these advantages are offset by the fact that the convergence rate of CP2 depends heavily on the grid size.

In Tables 4.4 and 4.5 we present the results of CP2 on the same experimental setup. CP2 accelerates PDHG by changing the step sizes  $\tau = \tau_n$  and  $\sigma = \sigma_n$  at each iteration  $n$  according to a special update rule. One only needs to choose initial values  $\tau_0$  and  $\sigma_0$  satisfying  $\tau_0\sigma_0\|\Delta_M\|_{L^2} \leq 1$  where  $\|\Delta_M\|_{L^2}$  is the  $L^2$  norm of the discrete Laplace operator  $\Delta_M$ . In [4], it is suggested to take  $\tau_0$  extremely large. We seemed to obtain the best results by choosing  $\tau_0 = \sigma_0 = \frac{1}{\sqrt{\|\Delta_M\|_{L^2}}}$ , thus we report results with this choice. Comparing Tables 4.4 and 4.5 to Tables 4.1 and 4.2, we see that CP2 is slower in both time and iterations in every case. In some of the cases, CP2 was unable to complete the computation in the allotted time (2 hours).

**4.2. EMD.** For the EMD, we will consider two different two-dimensional problems. In the first problem,  $\rho^1$  and  $\rho^0$  are both measures supported on a disc of radius  $1/4$  where  $\rho^1$  is centered at  $(5/8, 5/8)$  and  $\rho^0$  is centered at  $(3/8, 3/8)$ . In the second problem,  $\rho^1$  and  $\rho^0$  are delta measures at the points  $(5/8, 5/8)$  and  $(3/8, 3/8)$ , respectively. In both cases,  $\rho^1$  is the translation of  $\rho^0$  by the vector  $(1/4, 1/4)$ .

When two measures differ by a translation, it is possible to determine a minimizer  $m^* = u^* + \nabla\psi$  explicitly [26]. Suppose that  $\rho^1$  is given by translating  $\rho^0$  by the

TABLE 4.6  
EMD discs

Grid size	Error $10^{-3}$		Error $10^{-4}$	
	Iterations	Time (s)	Iterations	Time (s)
$512 \times 512$	64	1.7	163	3.6
$1024 \times 1024$	64	7.5	167	16.0
$2048 \times 2048$	64	30.6	168	67.4
$4096 \times 4096$	65	126.4	168	294.6

vector  $v$ . Then given a continuous, vector-valued test function  $p$  on  $[0, 1]^d$ , there exists a minimizer  $m^*$  such that

$$(4.1) \quad (m^*, p) = \int_0^1 \int_{[0,1]^d} v \cdot p(x - tv) d\rho^0(x) dt.$$

It then follows that the EMD distance must be equal to  $|v|$ . Thus, the EMD distance for both experiments is  $\frac{1}{\sqrt{8}}$ . Due to grid anisotropy, the solution on a discrete grid will be different, but it will approach  $\frac{1}{\sqrt{8}}$  as the grid becomes finer.

In the case of the two discs, the densities  $\rho^1$  and  $\rho^0$  are  $L^\infty$  functions. From formula (4.1) we can deduce that  $u^*$  must have bounded  $L^2$  norm. In fact,  $\|m^*\|_{L^2}^2 = \|u^*\|_{L^2}^2 + \|\nabla \psi\|_{L^2}^2$ , and thus  $\|u^*\|_{L^2} \leq \|m^*\|_{L^2} = 8^{-1/4}$ . Thus, we do not need to estimate the gap  $\delta_F(R)$ —it is already zero when  $R > 8^{-1/4}$ . This suggests that we can simply choose step sizes  $\tau = \sigma = 1$ . The performance of G-prox PDHG on the disc experiment is presented in Table 4.6. The convergence rate is clearly independent of the grid size for both error tolerances  $10^{-3}$  and  $10^{-4}$ .

The case of two delta measures is different. From formula (4.1) we see that  $m^*$  is a singular measure which concentrates on a one-dimensional line segment. We know that  $\nabla \psi \in L^q$  for  $q < 2$ . Therefore,  $u^* = m^* - \nabla \psi$  is also a singular measure and does not have finite  $L^2$  norm. As such, we will need to use Theorem 3.6 to help choose the step sizes. On a grid with  $M$  points, we will take  $\tau = \min(\sqrt{\frac{1}{\epsilon |\log \epsilon|}}, 2M^{1/4})$ , which again consists only of quantities that are easily calculated at the start of computation. Note that here we have dropped the dimensionality constant in Theorem 3.6 and used the trivial estimate  $\text{EMD}(\rho^1, \rho^0) \leq 1$  to get  $\sqrt{\frac{1}{\epsilon |\log \epsilon|}}$ . To get the second term  $2M^{1/4}$  we first use the inequality  $\|u_M^*\|_{L^2} \leq \|m_M^*\|_{L^2}$ . Then we note that  $m_M^*$  is approximately supported on a one-dimensional line segment and thus

$$\|m_M^*\|_{L^2} \leq 2M^{1/4} \|m_M^*\|_{L^1}^{1/2} \leq 2M^{1/4}.$$

In Table 4.7 we present the results of G-prox PDHG on the delta measure experiment. When  $\epsilon = 10^{-2}$ ,  $\sqrt{\frac{1}{\epsilon |\log \epsilon|}}$  is smaller than  $2M^{1/4}$  for every grid size. Therefore, the optimal step size  $\tau$  is the same for every grid, and the number of iterations needed to reach the error cutoff is always 30. When  $\epsilon = 10^{-3}$ ,  $\sqrt{\frac{1}{\epsilon |\log \epsilon|}}$  is larger than  $2M^{1/4}$  on the  $512 \times 512$  grid, approximately equal on the  $1024 \times 1024$  grid, and smaller on the  $2048 \times 2048$  and  $4096 \times 4096$  grids. As a result, the  $2048 \times 2048$  and  $4096 \times 4096$  grids have nearly identical iteration counts, while the algorithm converges faster on the  $512 \times 512$  and  $1024 \times 1024$  grids. When  $\epsilon = 10^{-4}$ ,  $2M^{1/4}$  is smaller than  $\sqrt{\frac{1}{\epsilon |\log \epsilon|}}$  on every tested grid size. Therefore, the algorithm converges faster on the smaller grids. Once again, if we did not allow  $\tau$  to depend on the grid discretization we would get nearly identical iteration counts for each grid, but with slower convergence.

TABLE 4.7  
*EMD delta measures*

Grid size	Error $10^{-2}$		Error $10^{-3}$		Error $10^{-4}$	
	Iterations	Time (s)	Iterations	Time (s)	Iterations	Time (s)
$512 \times 512$	30	1.2	56	1.6	121	2.8
$1024 \times 1024$	30	5.5	81	9.1	149	14.6
$2048 \times 2048$	30	22.8	98	45.6	185	75.3
$4096 \times 4096$	30	83.3	101	201.9	236	417.8

TABLE 4.8  
*EMD delta measures non-optimal step sizes*

Grid size	Error $10^{-2}$		Error $10^{-3}$		Error $10^{-4}$	
	Iterations	Time (s)	Iterations	Time (s)	Iterations	Time (s)
$512 \times 512$	93	2.4	611	12.0	1864	35.3
$1024 \times 1024$	112	11.9	1055	90.7	3835	322.9
$2048 \times 2048$	128	58.2	1747	675.6	8413	2985.2
$4096 \times 4096$	137	253.3	2530	4149.5	—	—

In [24] the authors solve the  $L^1$  EMD problem with a completely equivalent ADMM method. However, [24] does not consider how to choose optimal step sizes. For singular measures this can lead to a significant slowdown. In Table 4.8 we repeat the delta measure experiment with nonoptimal step sizes  $\tau = \sigma = 1$ . Comparing Tables 4.7 and 4.8 we see that the nonoptimal step sizes lead to runtimes that are up to 50 times slower. These results highlight the need for our careful theoretical analysis.

Finally, let us note that the EMD functional is not strongly convex in  $L^2$ . Without strong convexity, it is not possible to use the accelerated algorithms from [4]. As a result, G-prox PDHG will be orders of magnitude faster than PDHG type algorithms which do not use preconditioning. We can verify this by comparing our algorithm to the state-of-the-art results for the EMD problem presented in [23]. In [23], the authors approach the EMD minimization problem

$$\min_{\nabla \cdot m = \rho^1 - \rho^0} \int_{[0,1]^d} |m|$$

by converting it into a different unconstrained saddle point problem

$$\min_m \max_p \int_{[0,1]^d} |m| + (\nabla \cdot m + \rho^0 - \rho^1, p)$$

and searching for the saddle point using PDHG. Since the divergence operator  $\nabla \cdot$  is not preconditioned by PDHG, the convergence rate of the algorithm depends on the grid size. Due to this dependence, [23] only considers grids of size  $256 \times 256$  and less and requires parallelization for efficient computation. Notably, our serial algorithm on grids of size  $512 \times 512$  appears to be faster than their parallel algorithm on grids of size  $256 \times 256$ .

**5. Conclusion.** In this paper we have shown that G-prox PDHG, a variant of the PDHG algorithm, can be used to solve large-scale optimization problems with a convergence rate independent of the grid size. We have demonstrated our results both theoretically and numerically for two important optimization problems, the ROF denoising model and the EMD between probability measures. Our method is able to

solve these problems on grids as large as  $4096 \times 4096$  in a few minutes—a benchmark which seems to be out of reach for previous approaches.

In future works we hope to further extend our analysis and numerical results to other large-scale problems of interest. Furthermore, we hope to use our approach to more efficiently simulate the dynamics of stiff differential equations involving total variation.

**Appendix A. The domain  $[0, 1]^d$ .** Functions  $u : [0, 1]^d \rightarrow \mathbb{R}$  have a natural extension  $\tilde{u}$  to the larger domain  $[-1, 1]^d$  via even reflections. We can define  $\tilde{u}$  explicitly by taking  $\tilde{u}(x_1, \dots, x_d) = u(|x_1|, \dots, |x_d|)$ .  $u$  takes the same value on opposite boundaries, thus we can glue opposite boundaries together and identify  $[-1, 1]^d$  with the torus  $\mathbb{T}^d$ . This allows us to perform Fourier analysis on  $[0, 1]^d$ , as we can manipulate the Fourier series of  $\tilde{u}$  and then restrict the result back to  $[0, 1]^d$ . Therefore, any Fourier multiplier type operator, such as convolution, can be defined on  $[0, 1]^d$  (these operators can also be defined in physical space through the translation invariance of  $[-1, 1]^d$ ).

Other extensions to  $[-1, 1]^d$  are possible; however, even reflections are most natural for our purposes. Since  $\tilde{u}$  is even on  $[-1, 1]^d$ , its Fourier series expansion is a cosine series. Assuming  $\nabla \tilde{u}$  exists, it should have a sine series expansion. As a result,  $\nabla \tilde{u} \cdot n = 0$  on the boundary of  $[0, 1]^d$ . Thus, we see that the extension by even reflections can be used to automatically solve the Poisson equation on  $[0, 1]^d$  with zero Neumann boundary conditions.

### Appendix B. ROF proofs.

LEMMA B.1. *Suppose that  $u : [0, 1]^d \rightarrow [a, b]$  is a function of BV. Let  $G_\delta(z) = \delta^{-d} e^{-\pi(z/\delta)^2}$  be the Gaussian kernel and let  $u = G_\delta * u$ . Then we have the following inequalities:*

$$(B.1) \quad \|u_\delta - u\|_{L^q}^q \leq \delta \left( \frac{d}{2\pi} \right)^{\frac{1}{2}} (b-a)^{q-1} \|u\|_{TV}$$

and

$$\|\nabla u_\delta\|_{L^2}^2 \leq \frac{1}{\delta} \sqrt{2\pi d^3} (b-a) \|u\|_{TV}.$$

*Proof.*

$$\|u_\delta - u\|_{L^q}^q = \int_{[0,1]^d} \left| \int_{\mathbb{R}^d} G(z)(u(x + \delta z) - u(x)) dz \right|^q dx.$$

Using Jensen's inequality and the fact that  $u$  maps to the bounded interval  $[a, b]$ , we bound the above by

$$\leq (b-a)^{q-1} \int_{\mathbb{R}^d} G(z) \int_{[0,1]^d} |u(x + \delta z) - u(x)| dx dz.$$

Next we use the fact that BV functions satisfy a global Lipschitz property to get

$$\leq \delta \|u\|_{TV} \int_{\mathbb{R}^d} |z| G(z) = \delta \|u\|_{TV} A_{d-1} \int_0^\infty r^d e^{-\pi r^2} dr,$$

where  $A_{d-1}$  is the surface area of the sphere  $S^{d-1}$ . The first result follows from the inequality

$$A_{d-1} \int_0^\infty r^d e^{-\pi r^2} \leq \sqrt{\frac{d}{2\pi}}.$$

Now we turn to estimating the  $H^1$  norm of  $u_\delta$ . We have

$$\|\nabla u_\delta\|_{L^2}^2 = \int_{[0,1]^d} \left| \int_{\mathbb{R}^d} \nabla G_\delta(z) u(x+z) dz \right|^2 dx = \frac{1}{\delta^2} \int_{[0,1]^d} \left| \int_{\mathbb{R}^d} \nabla G(z) u(x+\delta z) dz \right|^2 dx.$$

Since  $\int_{\mathbb{R}^d} \nabla G(z) dz = 0$ , we may replace the above with

$$\frac{1}{\delta^2} \int_{[0,1]^d} \left| \int_{\mathbb{R}^d} \nabla G(z) (u(x+\delta z) - u(x)) dz \right|^2 dx.$$

We have the estimate  $\|\nabla G\|_{L^1} \leq \sqrt{2\pi d}$ ; thus the above is

$$\leq \frac{\sqrt{2\pi d}}{\delta^2} (b-a) \int_{\mathbb{R}^d} |\nabla G(z)| \int_{[0,1]^d} |u(x+\delta z) - u(x)| dx dz.$$

Again applying the global Lipschitz property of BV functions we get

$$\leq \frac{\sqrt{2\pi d}}{\delta} (b-a) \|u\|_{TV} \int_{\mathbb{R}^d} |z| |\nabla G(z)|.$$

Finally,  $\int_{\mathbb{R}^d} |z| |\nabla G(z)| = d$ .  $\square$

### Appendix C. EMD proofs.

**LEMMA C.1.** *Let  $\rho^1, \rho^0$  be probability measures and suppose that  $\psi$  solves the Poisson equation  $\Delta\psi = \rho^1 - \rho^0$  on  $[0, 1]^d$  with zero Neumann boundary conditions. Let  $G_\delta$  be the Gaussian kernel with width  $\delta > 0$  and  $\nabla\psi_\delta = \nabla\psi * G_\delta$ . Then*

$$\|\nabla\psi_\delta - \nabla\psi\|_{L^1} \leq C_d \delta (|\log(\delta)| + 1) \int_{[0,1]^d} |\rho^1 - \rho^0|,$$

where  $C_d$  is a constant which depends on the dimension only.

*Proof.* Let us define a linear operator  $T_\delta$

$$T_\delta h = \nabla\Delta^{-1}(G_\delta * h - h).$$

The current proposition is equivalent to

$$\|T_\delta(\rho^1 - \rho^0)\|_{L^1} \leq C_d \delta (|\log(\delta)| + 1) \int_{[0,1]^d} |\rho^1 - \rho^0|.$$

Let  $\rho_k^i = G_{1/k} * \rho^i$ . Then  $\rho_k^i$  is a smooth  $L^1$  function. If we assume that  $T_\delta$  is a bounded operator on  $L^1$ , we can use lower semicontinuity to obtain

$$\|T_\delta(\rho^1 - \rho^0)\|_{L^1} \leq \liminf_{k \rightarrow \infty} \|T_\delta(\rho_k^1 - \rho_k^0)\|_{L^1} \leq \|T_\delta\|_{L^1} \liminf_{k \rightarrow \infty} \|\rho_k^1 - \rho_k^0\|_{L^1}.$$

Applying Jensen's inequality to the last term we have

$$\liminf_{k \rightarrow \infty} \|\rho_k^1 - \rho_k^0\|_{L^1} \leq \int_{[0,1]^d} |\rho^1 - \rho^0|.$$

Thus, it is enough to show that for smooth functions  $h$  the operator  $T_\delta$  satisfies

$$\|T_\delta h\|_{L^1} \leq C_d \delta (|\log(\delta)| + 1) \|h\|_{L^1}.$$

For smooth  $h$  we have

$$\begin{aligned} T_\delta h(x) &= \int_{\mathbb{R}^d} G(z) \Delta^{-1} (\nabla h(x + \delta z) - \nabla h(x)) dz \\ &= \int_{\mathbb{R}^d} G(z) \Delta^{-1} \int_0^\delta z^T D^2 h(x + tz) dt dz. \end{aligned}$$

All of the operators applied to  $h$  commute, so we have

$$\|T_\delta h\|_{L^1} \leq \int_0^\delta \|D^2 \Delta^{-1} h * \tilde{G}_t\|_{L^1} dt,$$

where  $\tilde{G}(z) = z^T G(z)$ . Suppose that  $q(t) \in (1, 2]$  for each  $t \in (0, \delta]$ . Then

$$\int_0^\delta \|D^2 \Delta^{-1} h * \tilde{G}_t\|_{L^1} dt \leq \int_0^\delta \|D^2 \Delta^{-1} h * \tilde{G}_t\|_{L^{q(t)}} dt.$$

Now we use the fact that  $D^2 \Delta^{-1}$  is a bounded operator on  $L^q$  for  $q \in (1, \infty)$ . Moreover, for  $q \in (1, 2]$ , we have the operator norm bound

$$\|D^2 \Delta^{-1}\|_{L^q \rightarrow L^q} \leq \frac{C_d''}{q-1},$$

where  $C_d''$  is a constant depending on the dimension only [25]. Using the above bound and then Young's convolution inequality we get

$$\int_0^\delta \|D^2 \Delta^{-1} h * \tilde{G}_t\|_{L^{q(t)}} dt \leq \int_0^\delta \frac{C_d''}{q(t)-1} \|h * \tilde{G}_t\|_{L^{q(t)}} dt \leq \|h\|_{L^1} \int_0^\delta \frac{C_d'' \|\tilde{G}_t\|_{L^{q(t)}}}{q(t)-1} dt.$$

The  $L^q$  norm of  $\tilde{G}_t$  satisfies

$$\|\tilde{G}_t\|_{L^q} \leq C'_d t^{d(1-q)/q}$$

for some new constant  $C'_d$ . Now we shall make the choice  $q(t) = 1 + \frac{1}{d|\log(t)|}$ . This gives us

$$\|h\|_{L^1} \int_0^\delta \frac{C_d'' \|\tilde{G}_t\|_{L^{q(t)}}}{q(t)-1} dt \leq C_d \|h\|_{L^1} \int_0^\delta |\log(t)| dt,$$

where  $C_d$  is again a new constant. The inequality  $\int_0^\delta |\log(t)| dt \leq \delta |\log(\delta)| + \delta$  finishes the proof.  $\square$

**LEMMA C.2.** *Let  $G_\delta(z) = \delta^{-d} e^{-\pi(z/\delta)^2}$  be the Gaussian kernel. Then there exists a minimizer  $u^*$  of the EMD functional such that  $u_\delta = G_\delta * u^*$  has finite  $L^2$  norm bounded by*

$$\|u_\delta\|_2 \leq \left( (2\delta)^{(1-d)} \text{EMD}(\rho^1, \rho^0) \right)^{1/2}.$$

*Proof.* Young's convolution inequality automatically gives  $\|u_\delta\|_{L^2} \leq \|G_\delta\|_{L^2} \|u^*\|_1 = \delta^{-d/2} \|u^*\|_{L^1}$ . However, this does not take into account the structure of the EMD problem, and better results are possible.

Consider first the case where  $\rho^1$  and  $\rho^0$  are delta measures at locations  $x_1, x_0 \in (0, 1)^d$ , respectively. It is then known ([26]) that the solution  $m^* = u^* + \nabla \psi$  is given by

$$(p, m^*) = (x_1 - x_0) \cdot \int_0^1 p(t(x_1 - x_0) + x_0) dt.$$

The simplicity of the solution allows us to express  $\|G_\delta * m^*\|_2^2$  explicitly in the Fourier domain. This will allow us to bound  $\|G_\delta * u^*\|_2^2$  since  $m^* = u^* + \nabla\psi$  and  $u^*$  and  $\nabla\psi$  are orthogonal in the  $L^2$  inner product.

$$\|G_\delta * m^*\|_2^2 = \frac{\|x_1 - x_0\|_2^2}{2^{d-1}\pi^2} \sum_{n \in \mathbb{Z}^d} \frac{|\sin(\pi(n, x_1)) - \sin(\pi(n, x_0))|^2}{|(n, x_1 - x_0)|^2} e^{-\pi\delta^2|n|^2}.$$

The inner sum can be bounded by

$$1 + \int_{\mathbb{R}^d} \frac{|\sin(\pi(\xi, x_1)) - \sin(\pi(\xi, x_0))|^2}{|(\xi, x_1 - x_0)|^2} e^{-\pi\delta^2\|\xi\|_2^2}.$$

Using the inequality

$$|\sin(\pi(\xi, x_1)) - \sin(\pi(\xi, x_0))| \leq |\sin(\pi(\xi, x_1 - x_0))|$$

we may bound the integral by

$$\int_{\mathbb{R}^d} \frac{\sin(\pi(\xi, x_1 - x_0))^2}{|(\xi, x_1 - x_0)|^2} e^{-\pi\delta^2\|\xi\|_2^2}.$$

By rotating, we may assume that  $x_1 - x_0$  is parallel to the first standard basis vector  $e_1$ . Thus, the integral simplifies to

$$\int_{\mathbb{R}^d} \frac{\sin(\pi\xi_1\|x_1 - x_0\|_2)^2}{\xi_1^2\|x_1 - x_0\|_2^2} e^{-\pi\delta^2\|\xi\|_2^2} \leq \delta^{1-d} \int_{\mathbb{R}} \frac{\sin(\pi\xi_1\|x_1 - x_0\|_2)^2}{\xi_1^2\|x_1 - x_0\|_2^2} d\xi_1.$$

The integral on the right-hand side can be computed explicitly and has value  $\frac{\pi^2}{2\|x_1 - x_0\|_2}$ . Thus, we may conclude that

$$\|G_\delta * u^*\|_2^2 \leq \|G_\delta * m^*\|_2^2 \leq (2\delta)^{1-d} \|x_1 - x_0\|_2.$$

Next we extend our result to sums of delta measures. Suppose that  $\rho^1 = \frac{1}{k} \sum_{j=1}^k \delta_{x_j}$  and  $\rho^0 = \frac{1}{k} \sum_{j=1}^k \delta_{y_j}$  for some (potentially repeated) points  $x_1, \dots, x_k, y_1, \dots, y_k \in (0, 1)^d$ . Then there exists a minimizer  $m^*$  with the form

$$(p, m^*) = \frac{1}{k} \sum_{j=1}^k (x_j - y_{\pi(j)}) \cdot \int_0^1 p(y_{\pi(j)} + t(x_j - y_{\pi(j)})) dt,$$

where  $\pi$  is a permutation of  $\{1, \dots, k\}$ , which solves the assignment problem

$$\pi \in \arg \min_{\sigma \in S_k} \frac{1}{k} \sum_{j=1}^k \|x_j - y_{\sigma(j)}\|_2.$$

Using the triangle inequality and then Jensen's inequality, we have

$$\|G_\delta * m^*\|_2 \leq \frac{1}{k} \sum_{j=1}^k ((2\delta)^{1-d} \|x_j - y_{\pi(j)}\|_2)^{1/2} \leq \left( (2\delta)^{1-d} \text{EMD}(\rho^1, \rho^0) \right)^{1/2}.$$

Finally, we wish to extend this result to general probability measures  $\rho^1, \rho^0$ . Let  $P_k$  be the set of all probability measures of the form  $\mu = \frac{1}{k} \sum_{j=1}^k \delta_{x_j}$  for any list of

(potentially repeated) points  $x_1, \dots, x_k \in (0, 1)^d$ . Sums of delta measures are dense in the space of probability measures with the EMD topology [27]; therefore there exist sequences  $\rho_k^1$  and  $\rho_k^0$  such that  $\rho_k^1, \rho_k^0 \in P_k$ ,  $\text{EMD}(\rho_k^1, \rho_k^1) \rightarrow 0$  and  $\text{EMD}(\rho_k^0, \rho_k^0) \rightarrow 0$ . Using these sequences, we can choose for each  $k$

$$u_k \in \arg \min_{\nabla \cdot u = 0} \int_{[0,1]^d} |u + \nabla \psi_k|,$$

where  $\psi_k$  is the solution of the Poisson equation  $\Delta \psi_k = \rho_k^1 - \rho_k^0$ , with zero Neumann boundary conditions.

The solutions  $u_k$  have finite total variation bounded by

$$2\|\nabla \psi_k\|_1 \leq 2C_d \int_{[0,1]^d} |\rho_k^1 - \rho_k^0| \leq 2C_d,$$

where  $C_d$  is the operator norm of  $\|\nabla \Delta^{-1}\|_{L^1 \rightarrow L^{\frac{d}{d-1}, w}}$ . Thus, by the Banach–Alaoglu theorem, there exists a subsequence  $u_{k_n}$  and a vector-valued measure  $\tilde{u}$  such that for every bounded continuous test function  $p$  we have

$$\lim_{n \rightarrow \infty} (u_{k_n} - \tilde{u}, p) = 0.$$

Without loss of generality, we shall assume that this property holds for the full sequence  $u_k$ . Lower semicontinuity gives  $\|G_\delta * \tilde{u}\|_2 \leq \liminf_{k \rightarrow \infty} \|G_\delta * u_k\|_2$ . Thus, if we can show  $\tilde{u} \in \arg \min_{\nabla \cdot u = 0} \int_{[0,1]^d} |u + \nabla \psi|$  we will be done.

To that end, we note that

$$\begin{aligned} \liminf_{k \rightarrow \infty} \text{EMD}(\rho_k^1, \rho_k^0) &= \liminf_{k \rightarrow \infty} \int_{[0,1]^d} |u_k + \nabla \psi_k| \\ &\geq \sup_{\|\varphi\|_\infty = 1} \liminf_{k \rightarrow \infty} (u_k + \nabla \psi_k, \varphi) = \int_{[0,1]^d} |\tilde{u} + \nabla \psi|. \end{aligned}$$

Next by the triangle inequality, we have

$$\text{EMD}(\rho_k^1, \rho_k^0) \leq \text{EMD}(\rho_k^1, \rho_k^1) + \text{EMD}(\rho_k^0, \rho_k^0) + \text{EMD}(\rho_k^1, \rho_k^0),$$

thus  $\limsup_{k \rightarrow \infty} \text{EMD}(\rho_k^1, \rho_k^0) \leq \text{EMD}(\rho^1, \rho^0) = \inf_{\nabla \cdot u = 0} \int_{[0,1]^d} |u + \nabla \psi|$ . Putting everything together, we get the chain of inequalities

$$\limsup_{k \rightarrow \infty} \text{EMD}(\rho_k^1, \rho_k^0) \leq \inf_{\nabla \cdot u = 0} \int_{[0,1]^d} |u + \nabla \psi| \leq \int_{[0,1]^d} |\tilde{u} + \nabla \psi| \leq \liminf_{k \rightarrow \infty} \text{EMD}(\rho_k^1, \rho_k^0),$$

which completes the proof.  $\square$

**Acknowledgments.** The authors are grateful to Wilfrid Gangbo for helpful discussions. The first author is grateful to Selim Esedoḡlu for stimulating an interest in optimization techniques.

## REFERENCES

- [1] S. BARTELS, *Broken Sobolev space iteration for total variation regularized minimization problems*, IMA J. Numer. Anal., 36 (2015), pp. 493–502.
- [2] V. CASELLS, 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, <https://doi.org/10.1137/070683003>.

- [3] A. CHAMBOLLE, M. NOVAGA, D. CREMERS, AND T. POCK, *An introduction to total variation for image analysis*, in Theoretical Foundations and Numerical Methods for Sparse Recovery, De Gruyter, Berlin, 2010.
- [4] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145, <https://doi.org/10.1007/s10851-010-0251-1>.
- [5] A. CHAMBOLLE AND T. POCK, *An introduction to continuous optimization for imaging*, Acta Numer., 25 (2016), pp. 161–319, <https://doi.org/10.1017/S096249291600009X>.
- [6] A. CHAMBOLLE AND T. POCK, *On the ergodic convergence rates of a first-order primal-dual algorithm*, Math. Program., 159 (2016), pp. 253–287, <https://doi.org/10.1007/s10107-015-0957-3>.
- [7] J. DOUGLAS AND H. H. RACHFORD, *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439, <https://doi.org/10.1090/S0002-9947-1956-0084194-4>.
- [8] Z. E., *Nonlinear Functional Analysis and its Applications. III. Variational Methods and Optimization*, Springer, New York, 1985.
- [9] L. C. EVANS AND W. GANGBO, *Differential equations methods for the Monge–Kantorovich mass transfer problem*, Mem. Amer. Math. Soc., 137 (1999), <https://cds.cern.ch/record/2122679>.
- [10] D. GABAY AND B. MERCIER, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Comput. Math. Appl., 2 (1976), pp. 17–40, [https://doi.org/10.1016/0898-1221\(76\)90003-1](https://doi.org/10.1016/0898-1221(76)90003-1).
- [11] R. GLOWINSKI AND A. MARROCO, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires*, ESAIM Math. Model. and Numer. Anal., 9 (1975), pp. 41–76, <https://eudml.org/doc/193269>.
- [12] T. GOLDSTEIN AND S. OSHER, *The split Bregman method for  $L_1$ -regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343, <https://doi.org/10.1137/080725891>.
- [13] L. GRAFAKOS, *Classical Fourier Analysis*, Grad. Texts in Math., Springer, New York, 2014, <https://books.google.com/books?id=94FxBQAAQBAJ>.
- [14] W. LI, E. K. RYU, S. OSHER, W. YIN, AND W. GANGBO, *A parallel method for earth mover's distance*, J. Sci. Comput., 75 (2018), pp. 182–197.
- [15] P. L. LIONS AND B. MERCIER, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979, <https://doi.org/10.1137/0716071>.
- [16] A. NEMIROVSKI, *Prox-method with rate of convergence  $O(1/t)$  for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems*, SIAM J. Optim., 15 (2004), pp. 229–251, <https://doi.org/10.1137/S1052623403425629>.
- [17] Y. NESTEROV, *Excessive gap technique in nonsmooth convex minimization*, SIAM J. Optim., 16 (2005), pp. 235–249.
- [18] Y. NESTEROV, *Gradient methods for minimizing composite functions*, Math. Program., 140 (2013), pp. 125–161, <https://doi.org/10.1007/s10107-012-0629-5>.
- [19] Y. NESTEROV, *Introductory Lectures on Convex Optimization: A Basic Course*, Appl. Optim., Springer, New York, 2013, <https://books.google.com/books?id=2-ElBQAAQBAJ>.
- [20] Y. NESTEROV AND V. SHIKHMAN, *Quasi-monotone subgradient methods for nonsmooth convex minimization*, J. Optim. Theory App., 165 (2015), pp. 917–940.
- [21] G. PEYRÉ AND M. CUTURI, *Computational Optimal Transport*, arXiv:1803.00567, 2018.
- [22] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268, [https://doi.org/10.1016/0167-2789\(92\)90242-F](https://doi.org/10.1016/0167-2789(92)90242-F).
- [23] E. K. RYU, W. LI, P. YIN, AND S. OSHER, *Unbalanced and partial  $L_1$  Monge–Kantorovich problem: A scalable parallel first-order method*, J. Sci. Comput., 75 (2018), pp. 1596–1613.
- [24] J. SOLOMON, R. RUSTAMOV, L. GUIBAS, AND A. BUTSCHER, *Earth mover's distances on discrete surfaces*, ACM Trans. Graph., 33 (2014), pp. 67:1–67:12, <https://doi.acm.org/10.1145/2601097.2601175>.
- [25] E. M. STEIN, *Singular Integrals and Differentiability Properties of Functions*, Princeton Math. Ser. 30, Princeton University Press, Princeton, NJ, 1970.
- [26] C. VILLANI, *Topics in Optimal Transportation*, Grad. Stud. Math. 58, AMS, Providence, RI, 2003, <https://dx.doi.org/10.1007/b12016>.
- [27] C. VILLANI, *Optimal Transport: Old and New*, Grundlehren Math. Wiss. 338, Springer, Berlin, 2009, <https://dx.doi.org/10.1007/978-3-540-71050-9>.
- [28] J. WANG AND B. J. LUCIER, *Error bounds for finite-difference methods for Rudin–Osher–Fatemi image smoothing*, SIAM J. Numer. Anal., 49 (2011), pp. 845–868, <https://doi.org/10.1137/090769594>.