

# FAST ALGORITHMS FOR EARTH MOVER DISTANCE BASED ON OPTIMAL TRANSPORT AND $L_1$ REGULARIZATION II

WUCHEN LI, STANLEY OSHER, AND WILFRID GANGBO

ABSTRACT. We modify a fast algorithm which we designed in [15] for computing the Earth mover's distance (EMD), whose cost is a Manhattan metric. From the theory of optimal transport, the EMD problem can be reformulated as a familiar  $L_1$  minimization. We use a regularization which gives us a unique solution for this non strictly convex  $L_1$  problem. We adopt a primal-dual algorithm for the regularized problem, which uses very simple updates at each iteration and converges very rapidly. Several numerical examples are presented.

## 1. INTRODUCTION

The Earth Mover's distance (EMD), also named the Monge problem, plays a central role in many applications, including image processing, computer vision and statistics etc. [13, 17, 20, 24].

The EMD is a metric defined on the probability space of a convex, compact set  $\Omega \subset \mathbb{R}^d$ . Given two probability densities  $\rho^0, \rho^1$  on  $\Omega$ , the EMD deals with following minimization problem

$$\text{EMD}(\rho^0, \rho^1) := \min_{\pi} \int_{\Omega \times \Omega} d(x, y) \pi(x, y) dx dy \quad (1)$$

with the constraint that the joint probabilities (also called the transport plan)  $\pi(x, y)$  have  $\rho^0(x)$  and  $\rho^1(y)$  as marginals, i.e.

$$\int_{\Omega} \pi(x, y) dy = \rho^0(x), \quad \int_{\Omega} \pi(x, y) dx = \rho^1(y), \quad \pi(x, y) \geq 0.$$

Here  $d$  is a distance function on  $\mathbb{R}^d$  which we call the ground metric. Originally the ground metric was chosen as an Euclidean distance [4, 5, 15]. In this paper we consider  $d$  as a Manhattan distance or  $L_1$  distance. I.e.  $d(x, y) := \|x - y\|_1 = \sum_{v=1}^d |x_v - y_v|$ , for any  $x = (x_v)_{v=1}^d, y = (y_v)_{v=1}^d \in \Omega$ . We call (1) with the  $L_1$  ground metric the EMD- $L_1$  problem. We will propose a fast algorithm to approximate it.

The EMD in formulation (1) is well studied by the theory of optimal transport [1, 9, 22, 25]. The theory (remarkably) points out that (1) is equivalent to a new minimization

---

*Key words and phrases.* Earth Mover distance; Optimal transport; Compressed sensing; Primal-dual algorithm;  $\ell_1$  minimization.

This work is partially supported by ONR grants N000141410683, N000141210838 and DOE grant DE-SC00183838.

problem:

$$\text{EMD} - L_1(\rho^0, \rho^1) = \inf_m \left\{ \int_{\Omega} \|m(x)\|_1 dx : \nabla \cdot m(x) + \rho^1(x) - \rho^0(x) = 0 \right\}, \quad (2)$$

where  $m(x) = (m_v(x))_{v=1}^d$  is a flux vector satisfying a zero flux condition:  $m(x) \cdot \nu(x) = 0$ , where  $\nu(x)$  is the unit normal vector for the boundary of  $\Omega$ . The minimization EMD- $L_1$  (2) can be obtained as a limit of systems of equations which have an interesting fluid dynamics interpretation. A similar transport problem has been considered by [8] in the context of symplectic form, see also [23].

The formulation (2) has two benefits numerically. First, the dimension for computing (2) is much lower than the dimension in (1). Second, (2) is exactly an  $L_1$  minimization problem with linear constraints, which shares its structure with many problems in compressed sensing and image processing, see e.g. [11, 21, 26].

In this paper, we apply a modification of the algorithm designed in [15] to approximate EMD- $L_1$ , leveraging the structure of the formulation (2). The algorithm uses a finite volume method to discretize the domain  $\Omega$  and then applies the framework of primal-dual iterations, see e.g. [7, 19]. We overcome the lack of strict convexity of EMD- $L_1$  by a regularization. Since the regularized minimization is a perturbation of a homogenous degree one minimization, our algorithm is very simple and fast: First, we use a one dimensional shrink operator at each step, which handles the sparsity easily, see e.g. [11]; Second, the algorithm converges rapidly and each step involves very simple formulae. Our algorithm is roughly 10 times faster for EMD with an  $L_1$  ground metric than that with a Euclidean ground metric. This is because of the expense in computing square roots for the Euclidean metric.

EMD- $L_1$  has been shown to be effective in applications [18] and many linear programming techniques have been proposed, see e.g. [12, 16] and many references therein. Recently, the authors in [4, 5, 24] used the Alternating Direction Method of Multipliers (ADMM), which can also solve the EMD with a general Finsler ground metric. We are using the primal-dual algorithm in [22] rather than ADMM. So we do not need to solve an elliptic problem (i.e. the inverting of a Laplacian) and every iteration is explicit. Our updates are quite simple while ADMM might need fewer iterations. In addition, it is clear that the ADMM is difficult to parallelize while ours is quite easy. This means, with a parallel computer, the algorithm we propose can be made much faster. The primal-dual algorithm has been used before in optimal transport. The authors in [6] use it to compute the stationary solution of mean field games. However, the problem we consider is totally different. The emphasis of this paper is that we design simple and fast algorithms for a Monge problem of a particular but important shape.

The outline of this paper is as follows. In section 2, we propose a primal-dual algorithm for (2) on a uniform grid. Then we analyze the algorithm in section 3. Several numerical examples are demonstrated in section 4.

## 2. ALGORITHM

The EMD- $L_1$  in formulation (2) shares a similar structure with many problems in the field of compressed sensing and image processing. In this section, following ideas in [15], we use a finite volume discretization to reformulate (2). The discretized problem becomes an  $\ell_1$  optimization with linear constraints, which allows us to apply the primal-dual method designed in [7, 19].

We shall consider a uniform grid  $G = (V, E)$  with spacing  $\Delta x$  to discretize the spatial domain, where  $V$  is the vertex set

$$V = \{1, 2, \dots, N\},$$

and  $E$  is the edge set. Here  $i = (i_1, \dots, i_d) \in V$  represents a point in  $\mathbb{R}^d$ .

We consider a discrete probability set supported on all vertices:

$$\mathcal{P}(G) = \{(p_i)_{i=1}^N \in \mathbb{R}^n \mid \sum_{i=1}^N p_i = 1, p_i \geq 0, i \in V\},$$

where  $p_i$  represents a probability at node  $i$ , i.e.  $p_i = \int_{C_i} \rho(x) dx$ ,  $C_i$  is a cube centered at  $i$  with length  $\Delta x$ . So  $\rho^0(x)$ ,  $\rho^1(x)$  is approximated by  $p^0 = (p_i^0)_{i=1}^N$  and  $p^1 = (p_i^1)_{i=1}^N$ .

We use two steps to consider the EMD- $L_1$  on  $\mathcal{P}(G)$ . We first define a flux on a lattice. Denote a matrix  $m = (m_{i+\frac{1}{2}})_{i=1}^N \in \mathbb{R}^{N \times d}$ , where each component  $m_{i+\frac{1}{2}}$  is a row vector in  $\mathbb{R}^d$ , i.e.

$$m_{i+\frac{1}{2}} := (m_{i+\frac{e_v}{2}})_{v=1}^d := \left( \int_{C_{i+\frac{e_v}{2}}} m(x) dx \right)_{v=1}^d,$$

where  $e_v = (0, \dots, \Delta x, \dots, 0)^T$ ,  $\Delta x$  is at the  $v$ -th column. We consider a zero flux condition. So if a point  $i + \frac{e_v}{2}$  is outside the domain  $\Omega$ , we let  $m_{i+\frac{e_v}{2}} = 0$ . Based on such a flux  $m$ , we define a discrete divergence operator  $\text{div}_G(m) := (\text{div}_G(m_i))_{i=1}^N$ , where

$$\text{div}_G(m_i) := \frac{1}{\Delta x} \sum_{v=1}^d (m_{i+\frac{e_v}{2}} - m_{i-\frac{e_v}{2}}).$$

To summarize, (2) forms an optimization problem

$$\begin{aligned} & \underset{m}{\text{minimize}} && \|m\|_1 \\ & \text{subject to} && \text{div}_G(m) + p^1 - p^0 = 0, \end{aligned}$$

which can be written explicitly as

$$\begin{aligned} & \underset{m}{\text{minimize}} && \sum_{i=1}^N \sum_{v=1}^d |m_{i+\frac{e_v}{2}}| \\ & \text{subject to} && \frac{1}{\Delta x} \sum_{v=1}^d (m_{i+\frac{e_v}{2}} - m_{i-\frac{e_v}{2}}) + p_i^1 - p_i^0 = 0, \quad i = 1, \dots, n. \end{aligned} \tag{3}$$

We observe that (3) is an  $\ell_1$  optimization problem, whose cost function is convex and whose constraints are linear. However, the cost functional  $\|m\|_1$  in (3) is not strictly

convex, which often implies the existence of multiple minimizers. To deal with this issue, we consider a small quadratic perturbation, through which we pick up a unique solution for a modified problem:

$$\begin{aligned} & \underset{m}{\text{minimize}} && \|m\|_1 + \frac{\epsilon}{2}\|m\|_2^2 \\ & \text{subject to} && \text{div}_G(m) + p^1 - p^0 = 0 . \end{aligned} \quad (4)$$

Here  $\|m\|_2^2 = \sum_{i=1}^N \sum_{v=1}^d m_{i+\frac{ev}{2}}^2$ , and  $\epsilon$  is a positive scalar.

From now on, we solve (4) by looking at its saddle point structure. Denote  $\Phi = (\Phi_i)_{i=1}^N$  as (4)'s Lagrange multiplier, we have

$$\min_m \max_{\Phi} L(m, \Phi) := \min_m \max_{\Phi} \|m\|_1 + \frac{\epsilon}{2}\|m\|_2^2 + \Phi^T(\text{div}_G(m) + p^1 - p^0) . \quad (5)$$

Since  $L(\cdot, \Phi)$  is strictly convex which grows quadratically, and  $L(m, \cdot)$  is linear,  $L$  admits a saddle point solution. Saddle point problems, such as (5), are well studied by the first order primal-dual algorithm [7, 19, 15]. The iteration steps are as follows:

$$\begin{cases} m^{k+1} = \arg \min_m \|m\|_1 + \frac{\epsilon}{2}\|m\|_2^2 + \Phi^T \text{div}_G(m) + \frac{\|m - m^k\|_2^2}{2\mu} ; \\ \Phi^{k+1} = \arg \max_{\Phi} \Phi^T \text{div}_G(m^{k+1} + \theta(m^{k+1} - m^k)) - \frac{\|\Phi - \Phi^k\|_2^2}{2\tau} , \end{cases} \quad (6)$$

where  $\mu, \tau$  are two small step sizes,  $\theta \in [0, 1]$  is a given parameter,  $\|m - m^k\|_2^2 = \sum_{i=1}^N \sum_{v=1}^d (m_{i+\frac{ev}{2}} - m_{i+\frac{ev}{2}}^k)^2$  and  $\|\Phi - \Phi^k\|_2^2 = \sum_{i=1}^N (\Phi_i - \Phi_i^k)^2$ . These steps are alternating a gradient ascent in the dual variable  $\Phi$  and a gradient descent in the primal variable  $m$ .

In our updates for (6), we use simple exact formulae. Since the unknown variables  $m, \Phi$  are component-wise separable in this problem, each of its components  $m_{i+\frac{1}{2}}, \Phi_i$  can be independently obtained by solving (6).

Firstly, notice

$$\begin{aligned} & \min_m \|m\|_1 + \frac{\epsilon}{2}\|m\|_2^2 + \Phi^T \text{div}_G(m) + \frac{\|m - m^k\|_2^2}{2\mu} \\ & = \min_m \sum_{i=1}^N \sum_{v=1}^d \left\{ |m_{i+\frac{ev}{2}}| + \frac{\epsilon}{2}m_{i+\frac{ev}{2}}^2 + \frac{1}{\Delta x} \Phi_i (m_{i+\frac{ev}{2}} - m_{i-\frac{ev}{2}}) + \frac{(m_{i+\frac{ev}{2}} - m_{i+\frac{ev}{2}}^k)^2}{2\mu} \right\} \\ & = \sum_{i=1}^N \sum_{v=1}^d \min_{m_{i+\frac{ev}{2}}} \left\{ |m_{i+\frac{ev}{2}}| + \frac{\epsilon}{2}m_{i+\frac{ev}{2}}^2 - \nabla_G \Phi_{i+\frac{ev}{2}}^T m_{i+\frac{1}{2}} + \frac{1}{2\mu} (m_{i+\frac{ev}{2}} - m_{i+\frac{ev}{2}}^k)^2 \right\} , \end{aligned}$$

where  $\nabla_G \Phi_{i+\frac{ev}{2}} := \frac{1}{\Delta x} (\Phi_{i+ev} - \Phi_i)$ . So the first iteration in (6) has an explicit solution, which acts separately on each component of the vectors  $m$  and  $\nabla_G \Phi_{i+\frac{ev}{2}}$ :

$$m_{i+\frac{ev}{2}}^{k+1} = \frac{1}{1 + \epsilon\mu} \text{shrink}(m_{i+\frac{ev}{2}}^k + \mu \nabla_G \Phi_{i+\frac{ev}{2}}, \mu) ,$$

where we define a shrink operation in  $\mathbb{R}^1$

$$\text{shrink}(y, \alpha) := \text{sign}(y) \max\{|y| - \alpha, 0\} = \begin{cases} y - \alpha & \text{if } y > \alpha ; \\ 0 & \text{if } -\alpha \leq y \leq \alpha ; \\ y + \alpha & \text{if } y < -\alpha . \end{cases}$$

Second, consider

$$\begin{aligned} & \max_{\Phi} \Phi^T \text{div}_G(m^{k+1} + \theta(m^{k+1} - m^k)) - \frac{\|\Phi - \Phi^k\|_2^2}{2\tau} \\ &= \sum_{i=1}^N \max_{\Phi_i} \left\{ \Phi_i (\text{div}_G(m_i^{k+1} + \theta(m_i^{k+1} - m_i^k))) + p_i^1 - p_i^0 - \frac{(\Phi_i - \Phi_i^k)^2}{2\tau} \right\} . \end{aligned}$$

Thus the second iteration in (6) becomes

$$\Phi_i^{k+1} = \Phi_i^k + \tau \{ \text{div}_G(m_i^{k+1} + \theta(m_i^{k+1} - m_i^k)) + p_i^1 - p_i^0 \} .$$

**2.1. Algorithm.** We are now ready to state our algorithm.

---

#### Prime-dual method for EMD – $L_1$

**Input:** Discrete probabilities  $\rho^0, \rho^1$ ;

Initial guess of  $m^0$ , parameter  $\epsilon > 0$ , step size  $\mu, \tau, \theta \in [0, 1]$ .

**Output:** Optimal plan  $m$  and Earth Mover's metric- $L_1$  value  $\|m\|_1$ .

---

1. for  $k = 1, 2, \dots$  Iterates until convergence
  2.  $m_{i+\frac{\epsilon v}{2}}^{k+1} = \frac{1}{1+\epsilon\mu} \text{shrink}(m_{i+\frac{\epsilon v}{2}}^k + \mu \nabla_G \Phi_{i+\frac{\epsilon v}{2}}, \mu)$  ;
  3.  $\Phi_i^{k+1} = \Phi_i^k + \tau \{ \text{div}_G(m_i^{k+1} + \theta(m_i^{k+1} - m_i^k)) + p_i^1 - p_i^0 \}$  ;
  4. **end**
- 

*Remark 1.* Here we use the conventional shrink operator since the ground metric function is  $\ell_1$ . In [15] we apply what we call a shrink<sub>2</sub> operator, where the ground metric of EMD is  $\ell_2$ .

### 3. NUMERICAL ANALYSIS

In this section, we show several numerical properties of the algorithm (6). We prove that the primal-dual algorithm (6) converges to the minimizer of discretized minimization (4).

**Theorem 1.** Denote a linear operator  $K : \mathbb{R}^{N \times d} \rightarrow \mathbb{R}^N$ , such that

$$Km = (\text{div}_G(m_i))_{i=1}^N ,$$

and a saddle point of  $L$  in (5) as  $(m^*, \Phi^*)$ . Choose  $\theta = 1$ ,  $\tau\mu\|K\|_\infty^2 < 1$ . Then  $m^k, \Phi^k$  in iteration (6) converges to  $m^*, \Phi^*$ . Moreover, if  $\epsilon = 0$ , then the scalar components of  $\Phi^*$  satisfy the following equation

$$|\nabla_G \Phi_{i+\frac{\epsilon v}{2}}^*| = 1 , \quad \text{for any } i, v \text{ with } |m_{i+\frac{\epsilon v}{2}}^*| > 0 . \quad (7)$$

*Proof.* First, we only need to show that saddle point problem  $L$  satisfies the condition of Theorem 1 in [7, 19]. We rewrite  $L$  as

$$L(m, \Phi) = G(m) + \Phi^T K m - F(\Phi) ,$$

where  $G(m) = \|m\|_1$ ,  $K m = (\operatorname{div}_G(m_i))_{i=1}^N$ , and  $F(\Phi) = \sum_{i=1}^N \Phi_i(p_i^0 - p_i^1)$ . It is easy to observe that  $G, F$  is a convex continuous function and  $K$  is a linear operator. From Theorem 1 in [7], we prove the convergence result.

Second, since if  $|m_{i+\frac{\epsilon v}{2}}| > 0$ , we have the equation for each component:

$$0 = \frac{\partial L}{\partial m_{i+\frac{\epsilon v}{2}}} \Big|_{(m^*, \Phi^*)} = \frac{m_{i+\frac{\epsilon v}{2}}^*}{|m_{i+\frac{\epsilon v}{2}}^*|} - \nabla_G \Phi_{i+\frac{\epsilon v}{2}}^* .$$

Thus

$$1 = \left| \frac{m_{i+\frac{\epsilon v}{2}}^*}{|m_{i+\frac{\epsilon v}{2}}^*|} \right| = |\nabla_G \Phi_{i+\frac{\epsilon v}{2}}^*| .$$

We have proven that  $\Phi^*$  satisfies (7). □

As in [15], the computational complexity for this algorithm is  $O(NM)$ , where  $M$  is the iteration number for a given error. It was shown in [7] that the algorithm converges with the rate of  $O(\frac{1}{M})$ . We also have the fact that each iteration has simple updates, which only need  $O(N)$  operations. So our method requires overall  $O(N) \times O(M)$  computations. In practice, we observe roughly 10 times better performance here than the problem in [15], see the next section.

It is also worth mentioning that if  $\epsilon = 0$ , the cost functional  $|m|_1$  is not convex, so there may exist multiple minimizers for problem (3). If  $\epsilon > 0$ , the modified cost functional  $|m|_1 + \frac{\epsilon}{2} \|m\|_2^2$  is strongly convex, and we pick up a unique solution for the perturbed problem (3). In next section, we use numerical examples to demonstrate that such a unique solution approximates a particular minimizer of (3) when  $\epsilon$  is sufficient small.

#### 4. EXAMPLES

In this section, we demonstrate numerical results on a square  $[-2, 2] \times [-2, 2]$ . Our discretization is a uniform  $80 \times 80$  lattice. The parameters are chosen as  $\mu = \tau = 0.00625$ ,  $\theta = 1$ . The initial flux is chosen as all zeros. In following examples, we apply the stopping criteria as

$$\frac{1}{N} \sum_{i=1}^N |\operatorname{div}_G(m_i^k) + p_i^1 - p_i^0| \leq 10^{-9} .$$

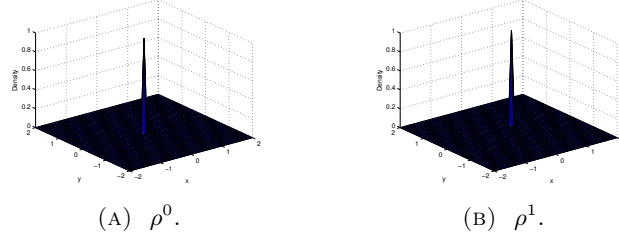


FIGURE 1. Here  $\rho^0$  and  $\rho^1$  are concentrated at  $(0, 0)$ ,  $(0.4, 0.4)$ , i.e.  $\rho^0 = \delta_{(0,0)}$ ,  $\rho^1 = \delta_{(0.4,0.4)}$ . The computed earth mover's metric is 0.7999

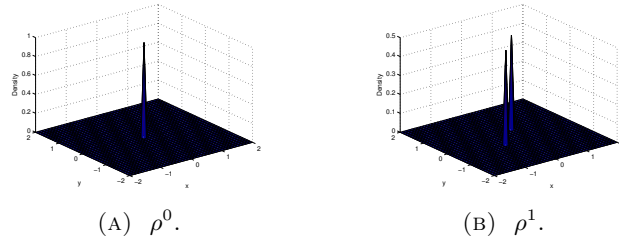


FIGURE 2. Here  $\rho^0$  is concentrated at  $(0, 0)$ ,  $\rho^1$  is concentrated at two positions,  $(0.4, 0.4)$  and  $(-0.4, -0.4)$ , i.e.  $\rho^0 = \delta_{(0,0)}$ ,  $\rho^1 = \frac{1}{2}(\delta_{(0.4,0.4)} + \delta_{(-0.4,-0.4)})$ . The computed earth mover's metric is 0.80078.

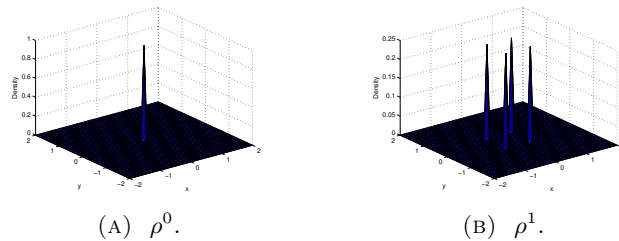


FIGURE 3. Here  $\rho^0$  is concentrated at  $(0, 0)$ ,  $\rho^1$  is concentrated at four positions,  $(0.4, 0.4)$ ,  $(0.4, -0.4)$ ,  $(-0.4, 0.4)$ ,  $(-0.4, -0.4)$ , i.e.  $\rho^0 = \delta_{(0,0)}$ ,  $\rho^1 = \frac{1}{4}(\delta_{(0.4,0.4)} + \delta_{(0.4,-0.4)} + \delta_{(-0.4,0.4)} + \delta_{(-0.4,-0.4)})$ . The computed earth mover's metric is 0.80042.

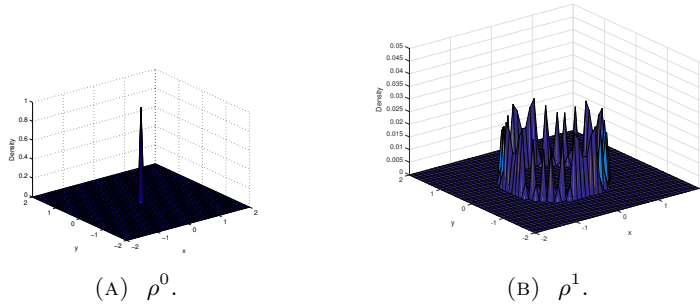


FIGURE 4. Here  $\rho^0$  is concentrated at  $(0,0)$ ,  $\rho^1$  is a measure supported on a circle, i.e.  $\rho^0 = \delta_{(0,0)}$ ,  $\rho^1 = \frac{1}{K}(e^{\frac{x^2}{\sigma} - \frac{x^4}{\sigma}})$ , where  $K$  is a normalized constants and  $\sigma = 10^{-3}$ . The computed earth mover's metric is 0.8920.

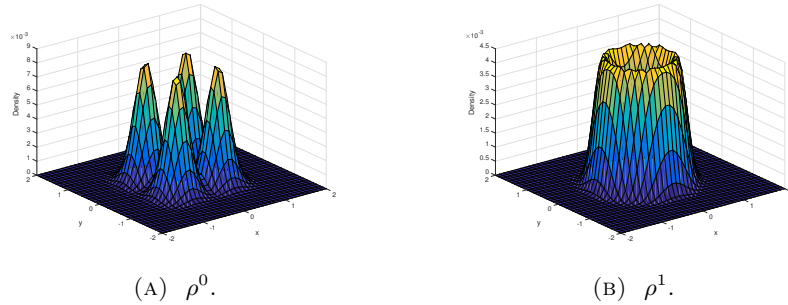


FIGURE 5. Here  $\rho^0 = \frac{1}{K_1}e^{-\frac{x^2 - |x_1 + x_2|}{\sigma}}$ ,  $\rho^1 = \frac{1}{K_2}(e^{\frac{x^2}{2\sigma} - \frac{x^4}{2\sigma}})$ , where  $K_1, K_2$  are normalized constants,  $\sigma = 0.1$ . The computed earth mover's metric is 0.1428.

Table 1 reports the time (in seconds) under different grids for computing EMD in Figure 1, 2, 3. The implementation is done in MATLAB 2016, on a 2.40GHZ Intel Xeon processor with 4GB RAM.



**Example 1:**

Grids number (N)	Time (s)	Iteration	Relative Error
100	0.031	198	$2.0 \times 10^{-3}$
400	0.197	356	$7.5 \times 10^{-4}$
1600	1.669	786	$2.7 \times 10^{-4}$
6400	25.178	3057	$9.1 \times 10^{-5}$

**Example 2:**

Grids number (N)	Time (s)	Iteration	Relative Error
100	0.047	156	$1.0 \times 10^{-3}$
400	0.204	347	$3.8 \times 10^{-4}$
1600	1.814	850	$1.3 \times 10^{-4}$
6400	28.53	3483	$4.6 \times 10^{-5}$

**Example 3:**

Grids number (N)	Time (s)	Iteration	Relative Error
100	0.039	142	$7.5 \times 10^{-4}$
400	0.171	261	$2.6 \times 10^{-4}$
1600	1.678	803	$8.9 \times 10^{-5}$
6400	31.13	3792	$2.9 \times 10^{-5}$

TABLE 1. We solve (4) with  $\epsilon = 0.01$  for Figure 1, 2, 3. Here the stopping criteria is  $\frac{1}{N} \sum_{i=1}^N |\text{div}_G(m_i^k) + p_i^1 - p_i^0| \leq 10^{-9}$ . The relative error is defined by  $\frac{\|m^\epsilon\|_1 + \epsilon \|m^\epsilon\|_2^{2-0.8}}{0.8}$ , where  $m^\epsilon$  is the computed minimizer of (4) and 0.8 is the analytical solution of EMD- $L_1$  (Manhattan distance between  $(0.4, 0.4)$ ,  $(0, 0)$ ).

Grids number (N)	Time (s) EMD- $L_1$	Time (s) in EMD- $L_2$
100	0.0162	0.1362
400	0.07529	1.645
1600	0.90	12.265
6400	22.38	130.37

TABLE 2. This result is for Figure 3. EMD- $L_1$ , EMD- $L_2$  represents an Earth Mover’s distance with Manhattan, Euclidean distance respectively. In this comparison, we use the same stopping criteria:  $\frac{1}{N} \sum_{i=1}^N |\text{div}_G(m_i^k) + p_i^1 - p_i^0| \leq 10^{-5}$ .

As in [15], we observe that the number of iterations is roughly  $O(N)$  (sometimes less). So we claim that the complexity of our algorithm is at most  $O(N^2)$ , which roughly matches the result of the computation time in table 1. Also we have observed a much better time performance than the results reported in [15], see table 2. We get approximately 10 times faster speed for EMD- $L_1$  than EMD with Euclidean ground metric. This is expected, since it is expensive to compute square roots for the Euclidean ground metric.

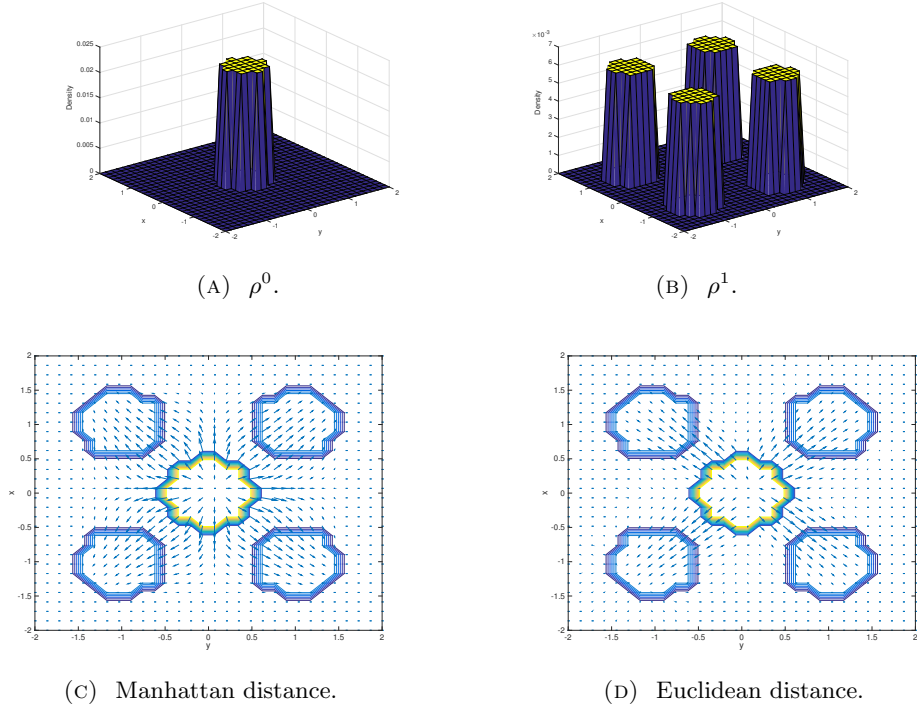


FIGURE 6. Comparison of minimizers  $m(x)$  for EMD under Manhattan or Euclidean distance. Here the initial measure is a uniform measure supported on a disk while the terminal measure is a uniform measure supported on four disjoint disks.

It is also worth mentioning that the quadratic perturbation in modified problem (4) is necessary. We design the following two numerical results to demonstrate this.

- (i) In table 3, we show that if we set  $\epsilon = 0$  in (4), the relative error can be enlarged if the total number of grids  $N$  increases;
- (ii) In table 4, we demonstrate that the minimizer of perturbed problem (4) approximates one particular minimizer of (3) when  $\epsilon$  approaches 0.

Grids number $N$	Relative error
400	$5.1 \times 10^{-5}$
1600	$6.7 \times 10^{-5}$
6400	$5.3 \times 10^{-4}$

TABLE 3. We compute EMD- $L_1$  in Figure 3 with  $\epsilon = 0$  and different meshes. The terminal condition is  $\frac{1}{N} \sum_{i=1}^N |\operatorname{div}_G(m_i^k) + p_i^1 - p_i^0| \leq 10^{-6}$ . The relative error is computed by  $\frac{\|m^{\epsilon=0.8}\|_1}{0.8}$ .

$\epsilon$	Relative error
0.1	$9.1 \times 10^{-4}$
0.01	$9.4 \times 10^{-5}$
0.001	$1.6 \times 10^{-5}$
0.0001	$6.0 \times 10^{-6}$

TABLE 4. We compute EMD- $L_1$  in Figure 3 with a fixed mesh and different values of  $\epsilon$ . The number of grid points is  $N = 1600$  and the terminal condition is  $\frac{1}{N} \sum_{i=1}^N |\operatorname{div}_G(m_i^k) + p_i^1 - p_i^0| \leq 10^{-6}$ . The relative error is computed by  $\frac{\|m^\epsilon - 0.8\|_1}{0.8}$ .

## 5. CONCLUSIONS

To summarize, we applied a primal-dual algorithm to solve EMD with the  $L_1$  ground metric. The algorithm inherits both key ideas in optimal transport theory and homogeneous degree one regularized problems. Compared to current methods, our algorithm has following advantages:

- First, it leverages the structure of optimal transport, which transfers EMD- $L_1$  into a  $L_1$  minimization. The new minimization contains only  $N$  variables, which is much less than the original  $N^2$  linear programming problem;
- Second, it uses simple exact formulas at each iteration (including the shrink operator in  $\mathbb{R}^1$ ) and converges to a minimizer.
- Third, it will be very easy to parallelize and thus speed up the algorithm considerably.

In addition, we consider a novel perturbed minimization

$$\inf_m \left\{ \int_{\Omega} \|m(x)\| + \frac{\epsilon}{2} \|m\|_2^2 dx : \nabla \cdot m(x) + \rho^1(x) - \rho^0(x) = 0 \right\}, \quad (8)$$

to approximate EMD problem. Here  $\|\cdot\|$  can be either the 1-norm or the 2-norm. In future work, we will study several theoretical properties of (8), especially the relation between  $m^\epsilon$  and  $m$  when  $\epsilon$  goes to 0.

## REFERENCES

- [1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2006.
- [2] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik* 84(3): 375–393, 2000.
- [3] M. Beckmann. A continuous model of transportation, *Econometrica* 20, 643660, 1952.
- [4] Jean-David Benamou and Guillaume Carlier. Augmented Lagrangian methods for transport optimization, mean field games and degenerate elliptic equations. *Journal of Optimization Theory and Applications*, 167(1): 1–26, 2015.
- [5] Jean-David Benamou, Guillaume Carlier and Roméo Hachi. A numerical solution to Monge's problem with a Finsler distance as cost. *M2AN*, 2016.
- [6] L.M. Briceo-Arias, D. Kalise and F.J. Silva Proximal methods for stationary Mean Field Games with local couplings. *arXiv:1608.07701*, 2016.

- [7] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 120–145, 2011.
- [8] B. Dacorogna and J. Moser. On a partial differential equation involving the Jacobian determinant. *Annales de l’IHP Analyse non linéaire*, 7(1), 1–26, 1990.
- [9] Lawrence Evans and Wilfrid Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem. *Memoirs of AMS*, no 653, vol. 137, 1999.
- [10] Sameer Shirdhonkar and David Jacobs. Approximate earth movers distance in linear time. *Computer Vision and Pattern Recognition IEEE conference*, 2008.
- [11] Tom Goldstein and Stanley Osher. The split Bregman method for L1-regularized problems. *SIAM journal on imaging sciences*, 2(2): 323–343, 2009.
- [12] J. Gudmundsson, O. Klein, C. Knauer, and M. Smid. Small. Manhattan Networks and Algorithmic Applications for the Earth Movers Distance. *In EWCG*, 2007.
- [13] E. Levina and P. Bickel. The earth mover’s distance is the Mallows distance: some insights from statistics. *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on* (2):251–256, 2001.
- [14] Wuchen Li. A study of stochastic differential equations and Fokker-Planck equations with applications. *PhD thesis*, 2016. Georgia Institute of Technology.
- [15] Wuchen Li, Stanley Osher and Wilfrid Gangbo. Fast algorithms for earth mover’s distance based optimal transport and homogeneous degree 1 regularization I. *arXiv:1609.07092*, 2016.
- [16] H. Ling and K. Okada. An Efficient Earth Movers Distance Algorithm for Robust Histogram Comparison. *PAMI*, 2007.
- [17] L. Métivier, R. Brossier, Q. Mérigot, E. Oudet and J. Virieux. Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion. *Geophysical Journal International*, (205) 1: 345–377, 2016.
- [18] Ofir Pele and Michael Werman Fast and robust earth mover’s distances. *2009 IEEE 12th International Conference on Computer Vision*, 460–467, 2009.
- [19] Thomas Pock and Antonin Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization, *2011 International Conference on Computer Vision*, 1762–1769, IEEE.
- [20] Yossi Rubner, Carlo Tomasi and Leonidas Guibas. The earth mover’s distance as a metric for image retrieval. *International journal of computer vision*, 40(2): 99–121, 2000.
- [21] Leonid Rudin, Stanley Osher and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, (60)1: 259–268, 1992.
- [22] Filippo Santambrogio. Absolute continuity and summability of transport densities: simpler proofs and new estimates. *Calculus of Variations and Partial Differential Equations*, 36 (3): 343–354, 2009,
- [23] Gilbert Strang.  $L_1$  and  $L_\infty$  approximation of vector fields in the plane. *North-Holland Mathematics Studies*, 81, 273–288, 1983.
- [24] Justin Solomon, Raif Rustamov, Leonidas Guibas and Adrian Butscher. Earth mover’s distances on discrete surfaces. *ACM Transactions on Graphics (TOG)*, 33(4), 2014.
- [25] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.
- [26] Wotao Yin, Stanley Osher, Donald Goldfarb and Jerome Darbon. Bregman iterative algorithms for  $\ell_1$ -minimization with applications to compressed sensing, *SIAM Journal on Imaging sciences*, 1(1): 143–168, 2008.

*E-mail address:* wcli@math.ucla.edu

*E-mail address:* sjo@math.ucla.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES.

*E-mail address:* wgangbo@math.ucla.edu