A FAST ALGORITHM FOR EARTH MOVER'S DISTANCE BASED ON OPTIMAL TRANSPORT AND L_1 TYPE REGULARIZATION

WUCHEN LI, STANLEY OSHER, AND WILFRID GANGBO

Abstract. We propose a new algorithm to approximate the Earth Mover's distance (EMD). Our main idea is motivated by the theory of optimal transport, in which EMD can be reformulated as a familiar L_1 type minimization. We use a regularization which gives us a unique solution for this L_1 type problem. The new regularized minimization is very similar to problems which have been solved in the fields of compressed sensing and image processing, where several fast methods are available. In this paper, we adopt a primal-dual algorithm designed there, which uses very simple updates at each iteration and is shown to converge very rapidly. Several numerical examples are provided.

1. Introduction

In this paper we propose a new algorithm to approximate the Earth Mover's distance (EMD), which is motivated by the theory of optimal transport and methods related to those used in compressed sensing and image processing.

We begin by reviewing some well known facts. EMD, which is also named the Monge problem or the Wasserstein metric, plays a central role in many applications, including image processing, computer vision and statistics e.t.c [13, 16, 20, 25]. The EMD is a particular metric defined on the probability space of a convex, compact set $\Omega \subset \mathbb{R}^d$. Given two probability densities ρ^0, ρ^1 in a probability set $\mathcal{P}(\Omega)$, where

$$\mathcal{P}(\Omega) = \{\rho(x) \in L^1(\Omega) \ : \ \int_\Omega \rho(x)dx = 1, \quad \rho(x) \geq 0\} \ .$$

The EMD deals with following (linear) minimization problem

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

with the constraint that the joint measure (also called the transport function) $\pi(x,y)$ has $\rho^0(x)$ 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. In this paper, we consider the ground metric either be the Euclidean distance (L_2) [4, 5] or the Manhattan

Key words and phrases. Earth Mover's distance; Optimal transport; Compressed sensing; Primal-dual algorithm; L_1 regularization.

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

distance (L_1) [15]. I.e. $d(x,y) := \|x-y\|_2$ or $\|x-y\|_1$. We call (1) with the L_1, L_2 ground metric the EMD-L_1, EMD-L_2.

In recent years, (1) has been well studied by the theory of optimal transport [1, 9, 22, 26]. The theory (remarkably) points out that (1) is equivalent to a new minimization problem

$$EMD(\rho^0,\rho^1) = \inf_m \{ \int_\Omega \|m(x)\| dx \ : \ \nabla \cdot m(x) + \rho^1(x) - \rho^0(x) = 0 \} \ , \tag{2}$$

where $\|\cdot\|$ is 2-norm (L_2 ground metric) or 1-norm (L_1 ground metric) in $\mathbb{R}^d$ and $m: \ \Omega \to \mathbb{R}^d$ is a flux vector satisfying a zero flux condition. I.e. $m(x)\cdot\nu(x) = 0$, where $\nu(x)$ is the unit normal vector for the boundary of Ω.

The minimization (2) has an interesting fluid dynamics interpretation. It turns out that (2) can be viewed as the following optimal control problem

$$\inf_{\rho,m} \{ \int_0^1 \int_\Omega \|m(t,x)\| dx dt \ : \ \frac{\partial \rho}{\partial t} + \nabla \cdot m = 0 \ , \quad \rho(0) = \rho^0, \quad \rho(1) = \rho^1 \ \} \ ,$$

where the minimum is taken among all possible flux functions $m(t,x)$, such that the probability density function is moved continuously in time, from ρ^0 to ρ^1. The optimal control problem has many minimizers. One of them is such that $\rho(t,x) = t\rho^1 + (1-t)\rho^0$. Here the flux function $m(t,x)$ does not depend on the time and $\nabla \cdot m = -\frac{\partial \rho}{\partial t} = \rho^0 - \rho^1$, so that the control problem becomes (2), see [3, 10, 26] for details related to EMD-L_2 and [8, 24] for similar optimal transport problems related to EMD-L_1.

The formulation (2) has two benefits numerically. First, the dimension in (2) is lower than the one in the original problem (1). Suppose we discretize Ω by a grid with N nodes. Since the unknown variable $\pi(x,y)$ in (1) is supported on $\Omega \times \Omega$ and $m(x)$ in (1) only depends on Ω, the number of grid points used in (2) is N while the number needed in (1) is N^2. Second, (2) is an L_1-type minimization problem, which shares its structure with many problems in compressed sensing and image processing. It is possible to borrow a very fast and simple algorithm used there to solve EMD, see e.g. [11, 21, 27].

In this paper, we propose a new algorithm for EMD, leveraging the structure of the formulation (2). The algorithm mainly uses a finite volume method to discretize the domain Ω and then applies the framework of primal-dual iterations designed in [7, 18]. We overcome the lack of strict convexity of (2) (EMD-L_1) by a regularization. Since the regularized minimization is a perturbation of a homogenous degree one minimization, our algorithm inherits all the benefits of the primal-dual algorithm: First, we use a 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. Thus the complexity of the algorithm is very low and the program is very simple.

EMD has been shown to be effective in applications [17] and many linear programming techniques have been proposed, see e.g. [12, 15] and many references therein. Recently, the authors in [4, 5, 25] 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 EMD-L_1, EMD-L_2.

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

2. Algorithm

The EMD problem, as presented in (2), has similar structure to many homogenous degree one regularized problems. In this section we will use a finite volume discretization to approximate (2). The discretized problem becomes an L_1-type optimization with linear constraints, which allows us to apply the hybrid primal-dual method designed in [7, 18].

We begin with considering EMD-L_2. We shall consider a uniform lattice graph $G = (V, E)$ with spacing Δx to discretize the spatial domain, where V is the vertex set

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

and E is the edge set. Here $i = (i_1, \cdots, i_d) \in V$ 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 Δx. 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 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.

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

where $e_v = (0, \cdots, \Delta x, \cdots, 0)^T$, Δx is at the v-th column. In other words, if we denote $i = (i_1, \cdots, i_d) \in \mathbb{R}^d$ and $m(x) = (m^1(x), \cdots, m^d(x))$, then

$$m_{i+\frac{1}{2}e_v} \approx m^v(i_1, \cdots, i_{v-1}, i_v + \frac{1}{2}\Delta x, i_{v+1}, \cdots, i_d)\Delta x^d \ .$$

We consider a zero flux condition. So if a point $i + \frac{1}{2}e_v$ is outside the domain Ω, we let $m_{i+\frac{1}{2}e_v} = 0$. Based on such a flux m, we define a discrete divergence operator $\mathrm{div}_G(m) := (\mathrm{div}_G(m_i))_{i=1}^N$, where

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

We next introduce the discrete cost functional

$$\|m\| := \sum_{i=1}^{N} \|m_{i+\frac{1}{2}}\|_2 = \sum_{i=1}^{N} \sqrt{\sum_{v=1}^{d} |m_{i+\frac{e_v}{2}}|^2} \ .$$

To summarize, (2) forms an optimization problem

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

which can be written explicitly as

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

We observe that (3) is an optimization problem, which is very similar to some problems in compressed sensing and image processing e.g. [11], whose cost functional is convex and whose constraints are linear. Thus we solve (3) by looking at its saddle point structure. Denote $\Phi = (\Phi_i)_{i=1}^N$ as (3)'s Lagrange multiplier, thus we have

$$\min_m \max_\Phi \quad L(m, \Phi) := \min_m \max_\Phi \quad \|m\| + \Phi^T (\mathrm{div}_G(m) + p^1 - p^0) \ . \tag{4}$$

Saddle point problems, such as (4), are well studied by the first order primal-dual algorithm [7, 18]. The iteration steps are as follows:

$$\begin{cases} m^{k+1} = & \arg\min_m \|m\| + (\Phi^k)^T \mathrm{div}_G(m) + \frac{\|m - m^k\|_2^2}{2\mu} \ ; \\ \Phi^{k+1} = & \arg\max_\Phi \Phi^T \mathrm{div}_G(m^{k+1} + \theta(m^{k+1} - m^k)) - \frac{\|\Phi - \Phi^k\|_2^2}{2\tau} \ , \end{cases} \tag{5}$$

where μ, τ are two small step sizes, $\theta \in [0, 1]$ is a given parameter, $\|m - m^k\|_2^2 = \sum_{i=1}^N \sum_{v=1}^d (m_{i+\frac{1}{2}e_v} - m^k_{i+\frac{1}{2}e_v})^2$ 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 Φ and a gradient descent in the primal variable m.

It turns out that iteration (5) can be solved by simple explicit formulae. Since the unknown variable m, Φ is component-wise separable in this problem, each of its components $m_{i+\frac{1}{2}}$, Φ_i can be independently obtained by solving (5).

First, notice

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

where $\nabla_G \Phi^k_{i+\frac{1}{2}} := \frac{1}{\Delta x}(\Phi^k_{i+e_v} - \Phi^k_i)_{v=1}^d$. The first iteration in (5) has an explicit solution, which is:

$$m^{k+1}_{i+\frac{1}{2}} = \mathrm{shrink}_2(m^k_{i+\frac{1}{2}} + \mu \nabla_G \Phi^k_{i+\frac{1}{2}}, \mu) ,$$

where we define the shrink_2 operation

$$\mathrm{shrink}_2(y, \alpha) := \frac{y}{\|y\|_2} \max\{\|y\|_2 - \alpha, 0\} , \quad \text{where } y \in \mathbb{R}^d .$$

Second, consider

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

Thus the second iteration in (5) becomes

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

We are now ready to state our algorithm.

Primal-Dual for EMD-L_2

Input: Discrete probabilities p^0, p^1;
Initial guess of m^0, step size μ, τ, $\theta \in [0, 1]$.

Output: m and EMD value $\|m\|$.

---

1. for $k = 1, 2, \cdots$ Iterates until convergence
2. $\quad m^{k+1}_{i+\frac{1}{2}} = \mathrm{shrink}_2(m^k_{i+\frac{1}{2}} + \mu \nabla_G \Phi^k_{i+\frac{1}{2}}, \mu)$;
3. $\quad \Phi_i^{k+1} = \Phi_i^k + \tau \{ \mathrm{div}_G(m_i^{k+1} + \theta(m_i^{k+1} - m_i^k)) + p_i^1 - p_i^0 \}$;
4. **end**

We next consider EMD-L_1. Similarly, (2) forms the following optimization problem

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

which can be written explicitly as

$$\begin{aligned} \underset{m}{\text{minimize}} \quad & \sum_{i=1}^{N}\sum_{v=1}^{d} |m_{i+\frac{e_v}{2}}| \\ \text{subject to} \quad & \frac{1}{\Delta x}\sum_{v=1}^{d}(m_{i+\frac{e_v}{2}} - m_{i-\frac{e_v}{2}}) + p_i^1 - p_i^0 = 0 \ , \quad i = 1, \cdots, n \ , \ v = 1, \cdots, d \ . \end{aligned} \tag{6}$$

We observe that (6) is an L_1 optimization problem, whose cost function is convex and whose constraints are linear. However, the cost functional in (6) 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:

$$\begin{aligned} \underset{m}{\text{minimize}} \quad & \|m\|_1 + \frac{\epsilon}{2}\|m\|_2^2 \\ \text{subject to} \quad & \text{div}_G(m) + p^1 - p^0 = 0 \ . \end{aligned} \tag{7}$$

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

From now on, we solve (7) by looking at its saddle point structure. Denote $\Phi = (\Phi_i)_{i=1}^N$ as (7)'s Lagrange multiplier, we have

$$\min_m \max_\Phi \quad L(m, \Phi) := \min_m \max_\Phi \ \|m\|_1 + \frac{\epsilon}{2}\|m\|_2^2 + \Phi^T(\text{div}_G(m) + p^1 - p^0) \ . \tag{8}$$

Since $L(\cdot, \Phi)$ is strictly convex which grows quadratically, and $L(m, \cdot)$ is linear, L admits a saddle point solution. Again, we solve (8) by the first order primal-dual algorithm [7, 18]. The iteration steps are as follows:

$$\begin{cases} m^{k+1} = & \arg\min_m \|m\|_1 + \frac{\epsilon}{2}\|m\|_2^2 + (\Phi^k)^T \text{div}_G(m) + \frac{\|m - m^k\|_2^2}{2\mu} \ ; \\ \Phi^{k+1} = & \arg\max_\Phi \Phi^T \text{div}_G(m^{k+1} + \theta(m^{k+1} - m^k)) - \frac{\|\Phi - \Phi^k\|_2^2}{2\tau} \ . \end{cases} \tag{9}$$

As in the computation of EMD-L_2, we use simple exact formulae for (9). First, the update for m^{k+1} has explicit solution, which acts separately on each component $m_{i+\frac{e_v}{2}}^{k+1}$:

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

where $\nabla_G \Phi_{i+\frac{e_v}{2}}^k := \frac{1}{\Delta x}(\Phi_{i+e_v}^k - \Phi_i^k)$. So the first iteration in (9) has an explicit solution:

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

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

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

Second, the update for Φ^{k+1} is same as the one in EMD-L_2. Since the second iteration in (9) is identical to the one in (5).

Primal-dual method for EMD $-$ L_1

Input: Discrete probabilities p^0, p^1;
Initial guess of m^0, parameter $\epsilon>0$, step size μ, τ, $\theta\in[0,1]$.
Output: m and EMD value $\|m\|_1$.

---

1. for $k=1,2,\cdots$ Iterates until convergence
2. $m^{k+1}_{i+\frac{e_y}{2}} = \frac{1}{1+\epsilon\mu}\text{shrink}(m^k_{i+\frac{e_y}{2}}+\mu\nabla_G\Phi^k_{i+\frac{e_y}{2}},\mu)$;
3. $\Phi^{k+1}_i = \Phi^k_i+\tau\{\text{div}_G(m^{k+1}_i+\theta(m^{k+1}_i-m^k_i))+p^1_i-p^0_i\}$;
4. **end**

Remark 1. Here we use the conventional shrink operator for EMD-L_1, while we apply what we call a shrink$_2$ operator for EMD-L_2.

3. Numerical analysis

In this section, we prove that the primal-dual algorithm converges to the minimizer of our discretized minimization (3) or (6). For illustration, we prove the result for EMD-L_2.

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

$$Km = (div_G(m_i))_{i=1}^N\ ,$$

and a saddle point of L *in* (4) *as* (m^*,Φ^*). *Choose* $\theta=1$, $\tau\mu\|K\|_\infty^2<1$. *Then* m^k, Φ^k *in iteration* (5) *converges to* m^*, Φ^*. *Moreover,* Φ^* *satisfies*

$$\|\nabla_G\Phi^*_{i+\frac12}\|_2 = 1\ ,\quad \textit{if}\quad \|m^*_{i+\frac12}\|_2>0\ . \tag{10}$$

Proof. First, we only need to show that saddle point problem L satisfies the condition of Theorem 1 in [7, 18]. We rewrite L as

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

where $G(m)=\|m\|$, $Km=(\text{div}_G(m_i))_{i=1}^N$, and $F(\Phi)=\sum_{i=1}^N\Phi_i(p^0_i-p^1_i)$. 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 $\|m^*_{i+\frac12}\|_2>0$, we have

$$0=\frac{\partial L}{\partial m_{i+\frac12}}|_{(m^*,\Phi^*)} = \frac{m^*_{i+\frac12}}{\|m^*_{i+\frac12}\|_2}-\nabla_G\Phi^*_{i+\frac12}\ .$$

Thus

$$1 = \| \frac{m^*_{i+\frac{1}{2}}}{\|m^*_{i+\frac{1}{2}}\|_2} \|_2 = \|\nabla_G \Phi^*_{i+\frac{1}{2}}\|_2 \ .$$

We have proven Φ^* satisfies (10). □

Remark 2. Theorem 1 holds similar for EMD-L_1 (6). Denote the saddle point of (6) as (m^*, Φ^*). Then the scalar components of Φ^* satisfy

$$|\nabla_G \Phi^*_{i+\frac{e_v}{2}}| = 1 \ , \quad \text{if} \quad |m^*_{i+\frac{e_v}{2}}| > 0 \ .$$

Since if $|m_{i+\frac{e_v}{2}}| > 0$, we have

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

Thus

$$1 = |\frac{m^*_{i+\frac{e_v}{2}}}{|m^*_{i+\frac{e_v}{2}}|}| = |\nabla_G \Phi^*_{i+\frac{e_v}{2}}| \ .$$

From above convergence results, we are ready to show that the computational complexity for this primal-dual algorithm is $O(NM)$, where M is the iteration number for a given error. It is known in [7], 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 good performance, see the next section. This is because of the well known performance of shrink operations as observed in compressed sensing and image processing calculations.

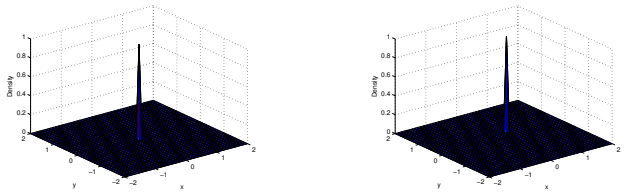
It is also worth mentioning that if $\epsilon = 0$, the cost functional $\|m\|_1$ in EMD-L_1 is not strictly convex, so there may exist multiple minimizers for problem (6). If $\epsilon > 0$, the modified cost functional $\|m\|_1 + \frac{\epsilon}{2}\|m\|_2^2$ is strongly convex, and we pick up a unique solution for the perturbed problem (6). In next section, we use numerical examples to demonstrate that such a unique solution approximates a particular minimizer of (6) when ϵ is sufficient small.

We do not claim that our discrete approximation converges as $\Delta x \to 0$ to the solution of (2). However, if as $\Delta x \to 0$ the family m stays uniformly bounded, then it is easy to show that m converges to a weak solution of (2).

4. Examples

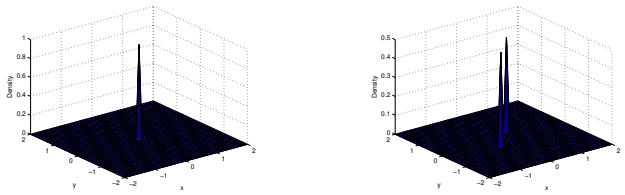
In this section, we demonstrate several numerical results on a square $[-2, 2] \times [-2, 2]$. Our discretization used in the Figure 1-5 is a uniform 40×40 lattice. The parameters are chosen as $\mu = \tau = 0.025$, $\theta = 1$. The initial flux m and Φ are chosen as all zeros. We use the stopping criteria

$$\frac{1}{N} \sum_{i=1}^{N} |\mathrm{div}_G(m^k_i) + p^1_i - p^0_i| \leq 10^{-5} \ .$$



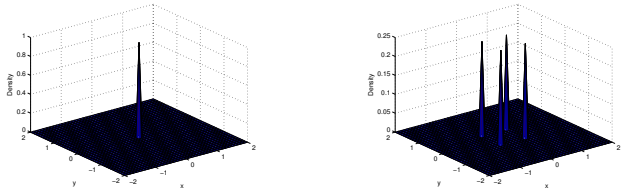
(A) ρ^0. (B) ρ^1.

FIGURE 1. Here ρ^0 and ρ^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 EMD-$L_1$, EMD-$L_2$ is 0.7981, 0.6232



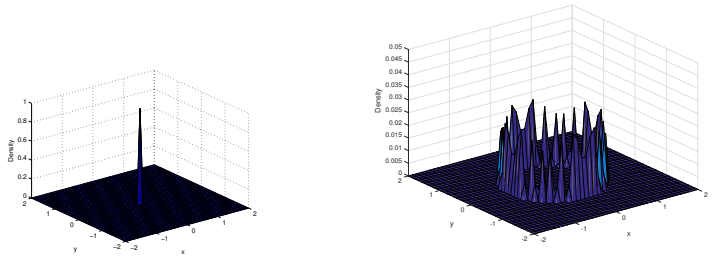
(A) ρ^0. (B) ρ^1.

FIGURE 2. Here ρ^0 is concentrated at $(0,0)$, ρ^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 EMD-$L_1$, EMD-$L_2$ is 0.8016, 0.6232.



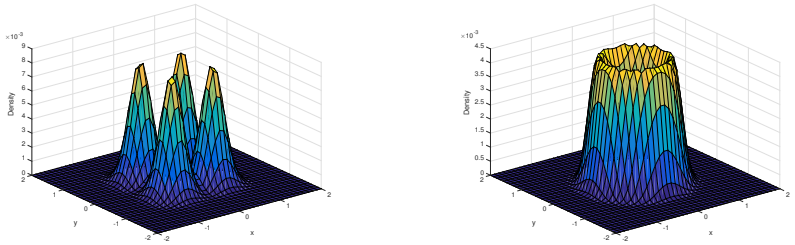
(A) ρ^0. (B) ρ^1.

FIGURE 3. Here ρ^0 is concentrated at $(0,0)$, ρ^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 EMD-$L_1$, EMD-$L_2$ is 0.8002, 0.5882.



(A) ρ^0. (B) ρ^1.

FIGURE 4. Here ρ^0 is concentrated at $(0,0)$, ρ^1 is a measure supported on a circle, i.e. $\rho^0 = \delta_{(0,0)}$, $\rho^1 = \frac{1}{K}\big(e^{\frac{x^2+y^2}{\sigma}} - \frac{(x^2+y^2)^2}{\sigma}\big)$, where K is a normalized constants and $\sigma = 10^{-3}$. The computed EMD-L_1, EMD-L_2 is 0.8794, 0.6943.



(A) ρ^0. (B) ρ^1.

FIGURE 5. Here $\rho^0 = \frac{1}{K_1}e^{-\frac{x^2+y^2-|x|-|y|}{\sigma}}$, $\rho^1 = \frac{1}{K_2}\big(e^{\frac{x^2+y^2}{\sigma}} - \frac{(x^2+y^2)^2}{\sigma}\big)$, where K_1, K_2 are normalized constants and $\sigma = 0.2$. The computed EMD-L_1, EMD-L_2 is 0.1778, 0.1259.

Table 1, 2 reports the time under different grids for computing EMD-L_1, EMD-L_2 in Figure 1, 2, 3. The implementation is done in MATLAB 2016, on a 2.40 GHZ Intel Xeon processor with 4GB RAM.

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 and 2.

Example 1:

Grids number (N)	Time (s)	Iteration	Relative Error
100	0.0411	58	0.2071
400	0.4015	167	0.1495
1600	4.78	524	0.1021
6400	64.81	1802	0.0607

Example 2:

Grids number (N)	Time (s)	Iteration	Relative Error
100	0.089	133	0.2072
400	1.44	597	0.1497
1600	8.41	901	0.1014
6400	118.8	3001	0.0596

Example 3:

Grids number (N)	Time (s)	Iteration	Relative Error
100	0.1334	210	0.0641
400	1.644	689	0.0536
1600	12.27	1347	0.0386
6400	130.37	3590	0.0199

TABLE 1. We compute EMD-L_2 for Figure 1, 2, 3. Time is in seconds. The relative error is defined by $\frac{|\|m\|-0.4\sqrt{2}|}{0.4\sqrt{2}}$, where m is the computed minimizer of (3) and $0.4\sqrt{2}$ is the analytical solution of EMD-L_2 (Euclidean 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 3. For Figure 3, we compare the computation time for EMD-L_1 and EMD-L_2. Here we use the same stopping criteria: $\frac{1}{N}\sum_{i=1}^{N} |\mathrm{div}_G(m_i^k)+p_i^1-p_i^0| \leq 10^{-5}$.

We also observe that we get approximately 10 times faster speed for EMD-L_1 than EMD-L_2 in table 3. This is expected, since it is expensive to compute square roots for the Euclidean ground metric.

Example 1:

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

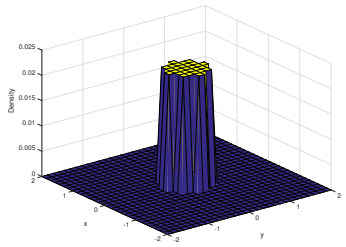
Example 2:

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

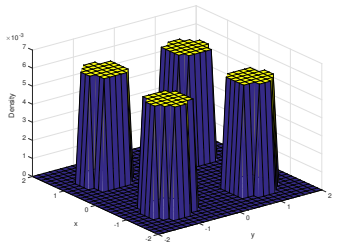
Example 3:

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

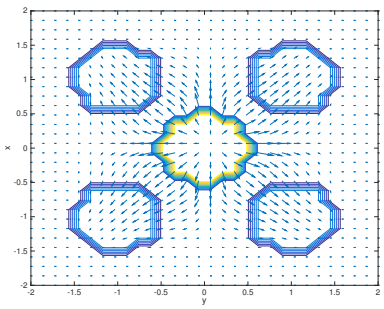
TABLE 2. We compute EMD-L_1 with $\epsilon = 0.01$ for Figure 1, 2, 3. Here the stopping criteria is $\frac{1}{N}\sum_{i=1}^{N} |\mathrm{div}_G(m_i^k) + p_i^1 - p_i^0| \leq 10^{-9}$. The relative error is defined by $\frac{\|m^\epsilon\|_1 + \epsilon\|m^\epsilon\|_2^2 - 0.8}{0.8}$, where m^ϵ is the computed minimizer of (7) and 0.8 is the analytical solution of EMD-L_1 (Manhattan distance between $(0.4, 0.4)$, $(0, 0)$).



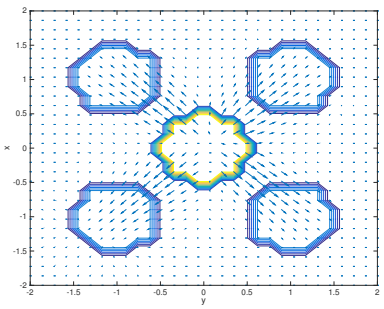
(A) ρ^0.



(B) ρ^1.



(C) Manhattan distance.



(D) Euclidean distance.

FIGURE 6. Comparison of minimizers $m(x)$ for EMD-L_1 and EMD-L_2. 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 (7) is necessary. We design the following two numerical results to demonstrate this.

(i) In table 4, we show that if we set $\epsilon = 0$ in (7), the relative error can be enlarged if the total number of grids N increases;
(ii) In table 5, we demonstrate that the minimizer of perturbed problem (7) approximates one particular minimizer of (3) when ϵ approaches 0.

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

TABLE 4. We compute EMD-L_1 in Figure 3 with $\epsilon = 0$ and different meshes. The terminal condition is $\frac{1}{N}\sum_{i=1}^{N} |\mathrm{div}_G(m_i^k) + p_i^1 - p_i^0| \leq 10^{-6}$. The relative error is computed by $\frac{|\|m^0\|_1 - 0.8|}{0.8}$, where m^0 is the computed minimizer of (7).

ϵ	Relative error
0.1	9.1×10^{-4}
0.01	9.4×10^{-5}
0.001	1.6×10^{-5}
0.0001	6.0×10^{-6}

TABLE 5. We compute EMD-L_1 in Figure 3 with a fixed mesh and different values of ϵ. The number of grid points is $N = 1600$ and the terminal condition is $\frac{1}{N}\sum_{i=1}^{N} |\mathrm{div}_G(m_i^k) + p_i^1 - p_i^0| \leq 10^{-6}$. The relative error is computed by $\frac{|\|m^\epsilon\|_1 + \frac{\epsilon}{2}\|m^\epsilon\|_2^2 - 0.8|}{0.8}$, where m^ϵ is the computed minimizer of (7).

5. Conclusions

To summarize, we applied a primal-dual algorithm to solve EMD with the L_1, L_2 ground metric. The algorithm inherits both key ideas in optimal transport theory and homogenous degree one regularized problems. Compared to current methods, our algorithm has following advantages:

- First, it leverages the structure of optimal transport, which transfers EMD into a L_1-type 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) 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

$$\inf_m \{ \int_\Omega \|m(x)\| + \frac{\epsilon}{2}\|m\|_2^2 dx \ : \ \nabla \cdot m(x) + \rho^1(x) - \rho^0(x) = 0 \} \ , \tag{11}$$

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 (11), especially the relation between m^ϵ and m when ϵ 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] M. Beckmann. A continuous model of transportation, *Econometrica* 20, 643660, 1952.

[3] 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.

[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 Hatchi. 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] Mikhail Feldman and Robert McCann. Monges transport problem on a Riemannian manifold. *Transactions of the American Mathematical Society*, 354 (4): 1667–1697, 2002.

[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. 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] H. Ling and K. Okada. An Efficient Earth Movers Distance Algorithm for Robust Histogram Comparison. *PAMI*, 2007.

[16] 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.

[17] Ofir Pele and Michael Werman. Fast and robust earth mover's distances. *2009 IEEE 12th International Conference on Computer Vision*, 460–467, 2009.

[18] 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.

[19] Yossi Rubner, Carlo Tomasi and Leonidas Guibas. A metric for distributions with applications to image databases. *Computer Vision, 1998. Sixth International Conference on*, 59–66, IEEE, 1998.

[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] Sameer Shirdhonkar and David Jacobs. Approximate earth movers distance in linear time. *Computer Vision and Pattern Recognition IEEE conference*, 2008.

[24] Gilbert Strang. L_1 and L_∞ approximation of vector fields in the plane. *North-Holland Mathematics Studies*, 81, 273–288, 1983.

[25] Justin Solomon, Raif Rustamov, Leonidas Guibas and Adrian Butscher. Earth mover's distances on discrete surfaces. *ACM Transactions on Graphics (TOG)*, 33(4), 2014.

[26] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

[27] Wotao Yin, Stanley Osher, Donald Goldfarb and Jerome Darbon. Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing, *SIAM Journal on Imaging sciences*, 1(1): 143–168, 2008.

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

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

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

Department of Mathematics, University of California, Los Angeles.