L_1 Monge-Kantorovich problem with applications

Stanley Osher

UCLA

May 2, 2017

Joint work with Wilfrid Gangbo, Wuchen Li, Yupeng Li, Ernest Ryu, Bao Wang, Penghang Yin and Wotao Yin.

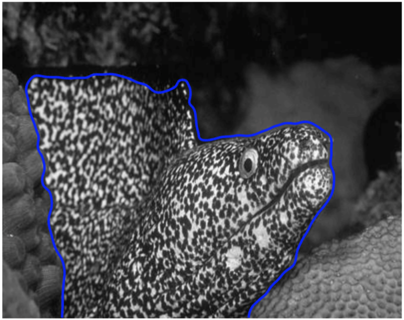
Partially supported by ONR and DOE.

Outline

Motivation

The optimal transport distance between histograms plays a vital role in many applications:

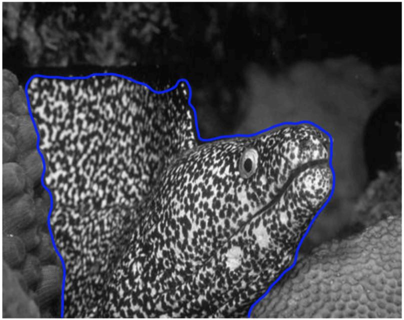
- Image segmentation;



- Statistics; Machine learning ;
- Mean field games .

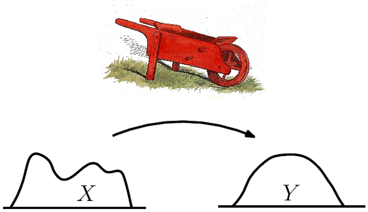
Motivation

It can also be applied to image alignment, which has many applications in computer version, drug design, and robotics:



(a) Some nice movies here.

Earth Mover's distance



What is the optimal way to move (transport) some materials with shape X, density $\rho^0(x)$ to another shape Y with density $\rho^1(y)$?

The question leads to the definition of the Earth Mover's distance (EMD), also called the Wasserstein metric, or the Monge-Kantorovich problem.

Problem statement

Consider

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

s.t.

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

In this talk, we will present fast and simple algorithms for EMD and related applications. Here we focus on two different choices of d, which are homogenous degree one:

$$d(x,y) = \|x-y\|_2 \quad (\text{Euclidean}) \quad \text{or} \quad \|x-y\|_1 \quad (\text{Manhattan})\ .$$

This choice of d was originally proposed by Monge in 1781.

Dynamic formulation

There exists a crucial reformulation of the problem. Since

$$d(x, T(x)) = \inf_{\gamma}\{\int_0^1 \|\dot{\gamma}(t)\| dt \ : \ \gamma(0) = x \ , \gamma(1) = T(x)\} \ ,$$

where $\|\cdot\|$ is 1 or 2-norm, the problem thus can be reformulated into an optimal control setting (Brenier-Benamou 2000):

$$\inf_{m,\rho} \int_0^1 \int_{\Omega} \|m(t,x)\| dx dt$$

where $m(t,x)$ is a flux function satisfying zero flux condition $(m(x) \cdot n(x) = 0$ on $\partial\Omega)$, such that

$$\frac{\partial \rho(t,x)}{\partial t} + \nabla \cdot m(t,x) = 0 \ .$$

Main problem: L_1 minimization

By Jensen's inequality, EMD is equivalent to the following minimal flux problem:

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

This is an L_1 minimization problem, whose minimal value can be obtained by a linear program, and whose minimizer solves a PDE system, known as the Monge-Kantorovich equation:

$$\begin{cases} p(m(x)) = \nabla \Phi(x) \ , \quad \nabla \cdot m(x) + \rho^1(x) - \rho^0(x) = 0 \ , \\ \|\nabla \Phi(x)\| = 1 \ , \end{cases}$$

where p is the sub-gradient operator and Φ is the Lagrange multiplier.

Outline

L_1 minimization

From numerical purposes, the minimal flux formulation has two benefits

- The dimension is much lower, essentially the square root of the dimension in the original linear optimization problem.
- It is an L_1-type minimization problem, which shares structure with problem arising in compressed sensing. We borrow a very fast and simple algorithm used there.

Current methods

Linear programming

P: Many tools;

C: Involves quadratic number of variables and does not use the structure of L_1 minimization.

Alternating direction method of multipliers (ADMM) [1]

P: Fewer iterations;

C: Solves an elliptic equation at each iteration; Not easy to parallelize.

In this talk, we apply the Primal-Dual method of Chambolle and Pock.

[1] (Benamou et.al, 2014), (Benamou et.al, 2016), (Solomon et.al, 2014)

Settings

Introduce a uniform grid $G = (V, E)$ with spacing Δx to discretize the spatial domain, where V is the vertex set and E is the edge set. $i = (i_1, \cdots, i_d) \in V$ represents a point in $\mathbb{R}^d$.

Consider a discrete probability set supported on all vertices:

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

and a discrete flux function defined on the edge of G :

$$m_{i+\frac{1}{2}} = (m_{i+\frac{1}{2}e_v})_{v=1}^d \ ,$$

where $m_{i+\frac{1}{2}e_v}$ represents a value on the edge $(i, i+e_v) \in E$, $e_v = (0, \cdots, \Delta x, \cdots, 0)^T$, Δx is at the v-th column.

Minimization: Euclidean distance

We first consider EMD with the Euclidean distance. The discretized problem becomes

$$\begin{aligned} \underset{m}{\text{minimize}} \quad & \|m\|_{1,2} \\ \text{subject to} \quad & \operatorname{div}(m) + p^1 - p^0 = 0 \ , \end{aligned}$$

which can be formulated explicitly

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

Chambolle-Pock Primal-dual algorithm

We solve the minimization problem by looking at its saddle point structure. Denote $\Phi = (\Phi_i)_{i=1}^N$ as a Lagrange multiplier:

$$\min_m \max_\Phi \quad \|m\| + \Phi^T(\mathrm{div}(m) + p^1 - p^0) \ .$$

The iteration steps are as follows:

$$\begin{cases} m^{k+1} = & \arg\min_m \|m\| + (\Phi^k)^T \mathrm{div}(m) + \frac{\|m - m^k\|_2^2}{2\mu} \ ; \\ \Phi^{k+1} = & \arg\max_\Phi \Phi^T \mathrm{div}(2m^{k+1} - m^k + p^1 - p^0) - \frac{\|\Phi - \Phi^k\|_2^2}{2\tau} \ , \end{cases}$$

where μ, τ are two small step sizes. These steps are alternating a *gradient ascent* in the dual variable Φ and a *gradient descent* in the primal variable m.

Algorithm: 2 line codes

Primal-dual method for EMD-Euclidean metric

1. For $k = 1, 2, \cdots$ $\quad$ Iterates until convergence
2. $\quad m^{k+1}_{i+\frac{1}{2}} = \mathrm{shrink}_2(m^k_{i+\frac{1}{2}} + \mu \nabla \Phi^k_{i+\frac{1}{2}}, \mu)$;
3. $\quad \Phi^{k+1}_i = \Phi^k_i + \tau\{\mathrm{div}(2m^{k+1}_i - m^k_i) + p^1_i - p^0_i\}$;
4. End

Here the shrink_2 operator for the Euclidean metric is

$$\mathrm{shrink}_2(y, \alpha) := \frac{y}{\|y\|_2} \max\{\|y\|_2 - \alpha, 0\} \ , \quad \text{where } y \in \mathbb{R}^2 \ .$$

Minimization: Manhattan distance

Similarly, the discretized problem becomes

$$\begin{aligned} \underset{m}{\text{minimize}} \quad & \|m\|_{1,1} + \frac{\epsilon}{2}\|m\|_2^2 = \sum |m_{i+\frac{1}{2}}| + \frac{\epsilon}{2}\sum |m_{i+\frac{1}{2}}|^2 \\ \text{subject to} \quad & \operatorname{div}(m) + p^1 - p^0 = 0 \ . \end{aligned}$$

Here a quadratic modification is considered with a small $\epsilon > 0$. This is to overcome the non strict convexity and hence possible non uniqueness of minimizers in the original problem

Algorithm: 2 line codes

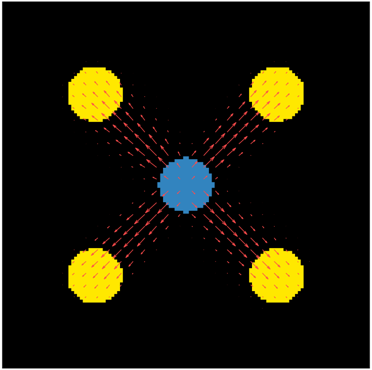
Primal-dual method for EMD-Manhattan distance

1. For $k = 1, 2, \cdots$ $\quad$ Iterates until convergence
2. $\quad m^{k+1}_{i+\frac{e_v}{2}} = \frac{1}{1+\epsilon\mu}\text{shrink}(m^k_{i+\frac{e_v}{2}} + \mu\nabla\Phi^k_{i+\frac{e_v}{2}}, \mu)$;
3. $\quad \Phi^{k+1}_i = \Phi^k_i + \tau\{\text{div}(2m^{k+1} - m^k)_i + p^1_i - p^0_i\}$;
4. End

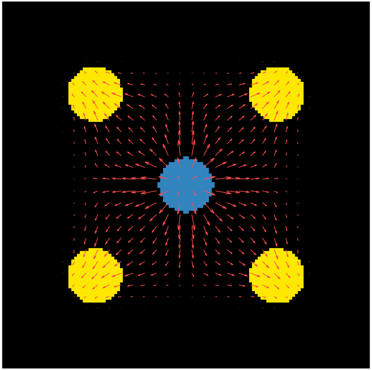
Here the shrink operator for the Manhattan metric is

$$\text{shrink}(y, \alpha) := \frac{y}{|y|}\max\{|y| - \alpha, 0\} \ , \quad \text{where } y \in \mathbb{R}^1 \ .$$

Optimal flux I

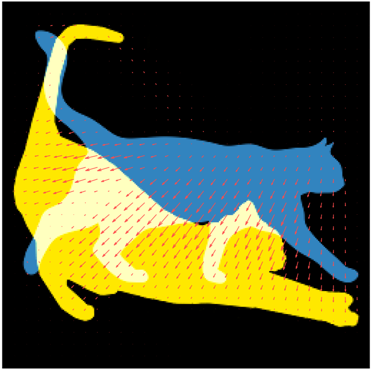


(b) EMD with Euclidean distance.

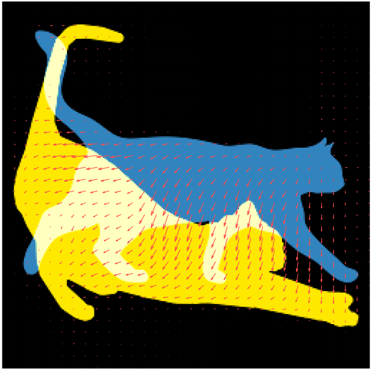


(c) EMD with Manhattan distance.

Optimal flux II



(d) EMD with Euclidean distance.



(e) EMD with Manhattan distance.

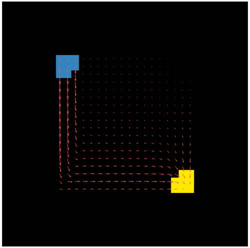
Manhattan vs Euclidean

Grids number (N)	Time (s) Manhattan	Time (s) in Euclidean
100	0.0162	0.1362
400	0.07529	1.645
1600	0.90	12.265
6400	22.38	130.37

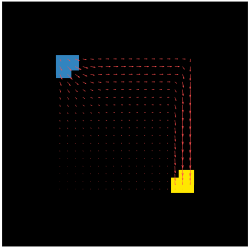
Table: We compute an example for Earth Mover's distance with respect to Manhattan or Euclidean distance.

This is result by using Matlab in a serial computer. In a parallel code using CUDA, it takes around 1 second to find a solution in a 256×256 grid for the Euclidean metric. It speeds up roughly 10^4 times.

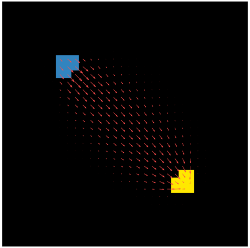
Importance of ϵ



(f) $\epsilon = 0$.



(g) $\epsilon = 0$.



(h) ϵ small.

Two different minimizers above on left are for $\epsilon = 0$.

PDEs behind ϵ

It is worth mentioning that the minimizer of the ϵ regularized problem

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

satisfies a nice (formal) system

$$\begin{cases} m(x) = \frac{1}{\epsilon}(\nabla\Phi(x) - \frac{\nabla\Phi(x)}{|\nabla\Phi(x)|}) \;, \\ \frac{1}{\epsilon}(\Delta\Phi(x) - \nabla \cdot \frac{\nabla\Phi(x)}{|\nabla\Phi(x)|}) = \rho^0(x) - \rho^1(x) \;, \end{cases}$$

where the second equation holds when $|\nabla\Phi| \geq 1$.

Notice that the term $\nabla \cdot \frac{\nabla\Phi(x)}{|\nabla\Phi(x)|}$ is the mean curvature formula.

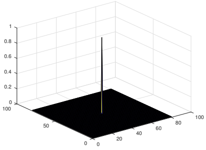
Outline

Unbalanced optimal transport

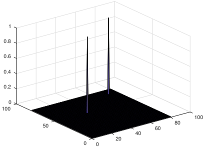
The original problem assumes that the total mass of given densities should be equal, which often does not hold in practice. E.g. the intensities of two images can be different.

Partial optimal transport seeks optimal plans between two measures ρ^0, ρ^1 with unbalanced masses, i.e.

$$\int_\Omega \rho^0(x)dx \neq \int_\Omega \rho^1(y)dy \ .$$



(i) ρ^0.



(j) ρ^1.

Unbalanced optimal transport

A particular example is the weighted average of Earth Mover's metric and L_1 metric, known as Kantorovich-Rubinstein norm. One important formulation is

$$\inf_{u,m} \ \{\int_\Omega \|m(x)\|dx \ : \ \nabla \cdot m(x) + \rho^0(x) - u(x) = 0 \ , \quad 0 \le u(x) \le \rho^1(x)\} \ .$$

Our method can solve the problem by 3 line codes.

Algorithm: 3 lines code

Primal-dual method for Partial optimal transport

Input: Discrete probabilities p^0, p^1;

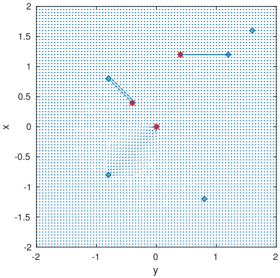
Initial guess of m^0, parameter $\epsilon > 0$, step size μ, τ, $\theta \in [0, 1]$.

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

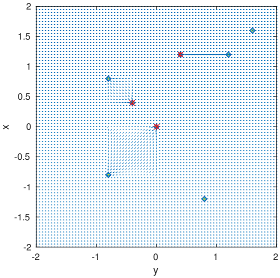
---

1. for $k = 1, 2, \cdots$ $\quad$ Iterates until convergence
2. $\quad m^{k+1}_{i+\frac{e_v}{2}} = \frac{1}{1+\epsilon\mu}\mathrm{shrink}(m^k_{i+\frac{e_v}{2}} + \mu\nabla\Phi^k_{i+\frac{e_v}{2}}, \mu)$;
3. $\quad u^{k+1}_i = \mathrm{Proj}_{C_i}(u^k_i - \mu\Phi^k_i)$;
4. $\quad \Phi^{k+1}_i = \Phi^k_i + \tau\big\{\mathrm{div}(2m^{k+1} - m^k)_i + 2u^{k+1}_i - u^k_i) - p^0_i\big\}$;
5. End

Partial optimal flux



(k) Euclidean distance.



(l) Manhattan distance.

Figure: Unbalanced transportation from three delta measures concentrated at two points (red) to five delta measures (blue).

Image segmentation

Given a grey-value image $I\colon \Omega \to \mathbb{R}$. The problem is to find two regions Ω_1, Ω_2, such that $\Omega_1 \cup \Omega_2 = \Omega$, $\quad \Omega_1 \cap \Omega_2 = \emptyset$.

Idea of Mumford-Shah model:

$$\min_{\Omega_1, \Omega_2} \quad \lambda \mathrm{Per}(\Omega_1, \Omega_2) + \mathrm{Dist}(\Omega_1, a) + \mathrm{Dist}(\Omega_2, b) \ .$$

where a, b are some given references generated by the image $I(x)$, known as the supervised terms, and Dist is some functional which estimates the closeness between region and references. There are some famous models, such as Mumford-Shah, Chan-Vese, Chan, Ni et al. 2007, Rabin et al. 2017.

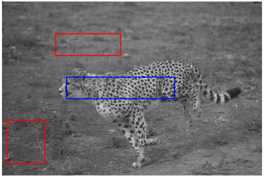
Orignal Monge-Kantorovich+ Segmentation

- It avoids overfitting of features (Swoboda and Schnorr (2003));
- It is L_1 minimization, which is great for computations.

Given intensity $I(x)$, the proposed model is:

$$\min_u \quad \lambda \int_\Omega |\nabla u(x)| dx + \mathrm{EMD}(H_I u, a) + \mathrm{Dist}(H_I(1-u), b) \ ,$$

where u is the indicator function of region, H_I is a linear operator depending on I which changes u into histograms, a, b are histograms in the selected red or blue regions:



Problem formulation

$$\inf_{u,m_1,m_2} \lambda \int_\Omega \|\nabla_x u(x)\| dx + \int_{\mathcal{F}} \|m_1(y)\| dy + \int_{\mathcal{F}} \|m_2(y)\| dy \ ,$$

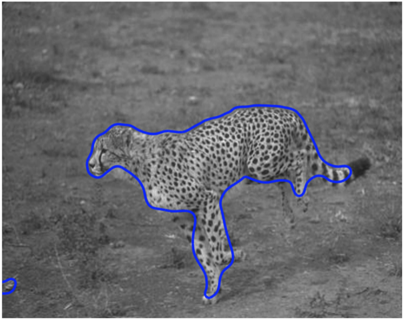
where the infimum is taken among $u(x)$ and flux functions $m_1(y)$, $m_2(y)$ satisfying

$$\begin{cases} 0 \leq u(x) \leq 1 \\ \nabla_y \cdot m_1(y) + H_I(u)(y) - a(y) \int_{\mathcal{F}} H_I(u)(y) dy = 0 \\ \nabla_y \cdot m_2(y) + H_I(1-u)(y) - b(y) \int_{\mathcal{F}} H_I(1-u)(y) dy = 0 \ . \end{cases}$$

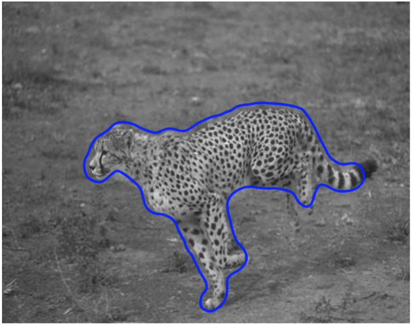
Here $x \in \Omega$, $y \in \mathcal{F}$, $H_I : \mathrm{BV}(\Omega) \to \mathrm{Measure}(\mathcal{F})$ is a linear operator.

Our algorithm can be easily used into this area. It contains only 6 simple and explicit iterations using the Chamolle-Pock primal dual method.

Segmentation with multiple dimensional features



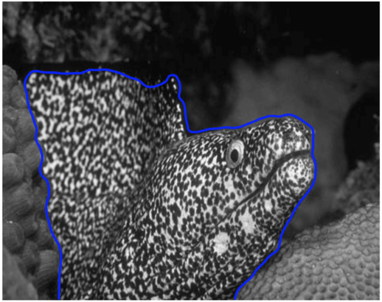
(a) Histogram of intensity, Mean



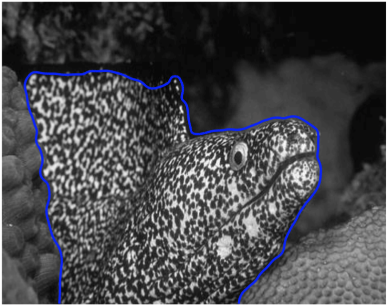
(b) Histogram of intensity, Mean, Texture

We take $\lambda = 1$, the mean and texture (Sochen et. al) are values chosen in 3×3 patches near each pixel.

Segmentation with multiple dimensional features

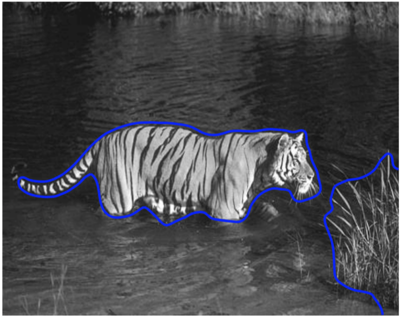


(c) Histogram of intensity, Mean

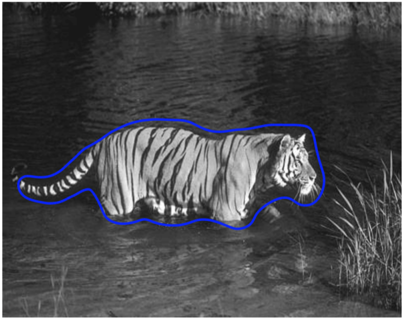


(d) Histogram of intensity, Mean, Texture

Segmentation with multiple dimensional features



(e) Histogram of intensity, Mean



(f) Histogram of intensity, Mean, Texture

PDEs behind segmentation

$$\begin{cases} \lambda \nabla \cdot \dfrac{\nabla u(x)}{\|\nabla u(x)\|} = \displaystyle\int_{\mathcal{F}} h(x,y)\big(\Phi_1(y) - \Phi_2(y)\big)dy - \big(a(y) - b(y)\big) \int_{\Omega \times \mathcal{F}} h(x,y)dxdy \\ \dfrac{1}{\epsilon}\left(\Delta \Phi_1(y) - \nabla \cdot \dfrac{\nabla \Phi_1(y)}{|\nabla \Phi_1(y)|}\right) = \displaystyle\int_{\Omega} h(x,y)u(y)dx - a(y) \int_{\Omega \times \mathcal{F}} h(x,y)u(x)dxdy \\ \dfrac{1}{\epsilon}\left(\Delta \Phi_2(y) - \nabla \cdot \dfrac{\nabla \Phi_2(y)}{|\nabla \Phi_2(y)|}\right) = \displaystyle\int_{\Omega} h(x,y)\big(1 - u(x)\big)dx - b(y) \int_{\Omega \times \mathcal{F}} h(x,y)\big(1 - u(x)\big)dxdy \ , \end{cases}$$

It is interesting to observe that there are three mean curvature formulas in both spatial and feature domains.

Primal-Dual method avoids solving nonlinear PDEs directly!!!

Image alignment via Monge-Kantorovich problem

Discussion

Our method for solving L_1 Monge-Kantorovich related problems

- handles the sparsity of histograms;
- has simple updates and is easy to parallelize;
- introduces a novel PDE system (Mean curvature formula in Monge Kantorovich equation).

It has been successfully used in partial optimal transport, image segmentation, image alignment and elsewhere.

Main references

Wuchen Li, Ernest Ryu, Stanley Osher, Wotao Yin and Wilfrid Gangbo.
A parallel method for Earth Mover's distance, 2016.

Wuchen Li, Penghang Yin and Stanley Osher.
A Fast algorithm for unbalanced L_1 Monge-Kantorovich problem, 2016.

Yupeng Li, Wuchen Li and Stanley Osher.
Image segmentation via original Monge-Kantorovich problem, in prepration.

Penghang Yin, Bao Wang, Wuchen Li and Stanley Osher.
A Fast algorithm for unbalanced L_1 Monge-Kantorovich problem, in preparation.

Thanks!