Numerical Aspects of Optimal Transport

Hira Abdul Ghani
997557013

August 28, 2014

Abstract

In this project, we formulate the Monge-Kantorovich optimal transport problem and consider possible ways to compute solutions. One approach is using linear programming to solve the problem. Computational examples are also demonstrated with the linear programming approach at a solution. Another approach taken is turning the optimal transport problem into a polar factorization problem. This becomes a matter of solving evolution equations. We also discuss existence and uniqueness of solutions when finding solutions to the evolution equations.

1 Introduction

We begin to describe the optimal mass transport problem by a simple analogy. Suppose we have a pile of sand and a hole that needs to filled up with sand. Define $\Phi(x, y)$ to be the cost to transport one unit of sand from location x to location y. The goal is to transport the sand from the pile to the hole at the minimal cost (Villani, 2003).

1.1 Monge Formulation

Mathematically, the pile and hole will be modeled by probability measures μ_0 on a measure space Ω_0 and μ_1 on a measure space Ω_1 respectively (Villani, 2003). In words, given subsets A, B of Ω_0, Ω_1 respectively, $\mu_0[A]$ gives a measure of how much sand is in A and $\mu_1[B]$ gives a measure of how much sand can be piled in B. The cost function Φ is a measurable map from $X \times Y$ to $\mathbb{R} \cup \{+\infty\}$. As described above, this function gives the cost to transport one unit of mass from location x to location y.

We wish to find a transport map $T : \Omega_0 \to \Omega_1$ such that we minimize the cost $\Phi(x, y)$:

$$M(T) = \int_{\Omega_0} \Phi(x, T(x))\, d\mu_0(x), \tag{1}$$

such that $T(\mu_0) = \mu_1$. In other words, for $B \subseteq \Omega_1$, $\mu_0\left(T^{-1}(B)\right) = \mu_1(B)$. So the μ_0 mass transported to B must equal the μ_1 mass of B. Here, we say that μ_1 is the pushforward of μ_0 by T (Villani, 2003), or that T transports μ_0 onto μ_1, and write

$$\mu_1 = T_{\#}\mu_0.$$

Notice that this formulation implies that all the mass at location x goes to the same location y.

1.2 Kantorovich's Formulation

Kantorovich allows us to relax the optimization problem by allowing the mass at $x \in \Omega_0$ to be split (Villani, 2003). Instead, for $x \in \Omega_0$, we have a measure γ that describes how the mass of x is distributed in Ω_1 (Villani, 2003). In Kantorovich's formulation, we define a measure $\gamma \geq 0$ on $\Omega_0 \times \Omega_1$ to minimize

$$M(\gamma) = \int_{X \times Y} \Phi(x, y)\, d\gamma(x, y), \tag{2}$$

such that μ_0, μ_1 are marginal distributions on Ω_0, Ω_1 respectively, ie. for $A \subseteq \Omega_0, B \subseteq \Omega_1$

$$\gamma(A \times \Omega_1) = \mu_0(A), \tag{3}$$
$$\gamma(\Omega_0 \times B) = \mu_1(B). \tag{4}$$

What (3) means is the γ mass transported from A equals the μ_0 mass of A. Similarly, (4) says the γ mass transported into B equals the μ_1 mass of B (Zemel, 2012). We call γ the transference plan.

1.3 Outline

This paper is organized as follows. We first talk about reducing the optimal transport problem to a linear programming problem by discretizing the domains Ω_0, Ω_1. In the next section, we discuss a method of using partial differential equations to find a solution. We also talk about what defines existence and uniqueness of solutions found using partial differential equations.

2 The Discrete Case

We consider the case of supply and demand to explain the discrete case. Say we have an initial configuration of resources $X = \{x_1, x_2, x_3, x_4\}$ where $x_i \in \mathbb{R}^2$. At every x_i, we produce a quantity $p_{0,i} = p_0(x_i)$. The final configuration is $Y = \{y_1, y_2, y_3\}$. At every y_i, the quantity $p_{1,j} = p_1(y_j)$ is expected. Define $\Phi_{ij} = \Phi(x_i, y_j)$ to be the unitary cost of transporting one unit from point x_i to point y_j (Savare, 2010). Possible transference plans $T_{ij} = T(x_i, y_j)$ that choose the quantities moved from x_i to y_j must satisfy the following constraints:

1. $\gamma(x_i, y_j) \geq 0$,
2. $\sum_{y \in Y} \gamma(x_i, y) = p_0(x_i)$,
3. $\sum_{x \in X} \gamma(x, y_j) = p_1(y_j)$.

The first requirement is intuitive since the quantity transferred cannot be negative. The second requirement tells us that, for each x_i, the quantity transferred, from x_i equals the quantity available at x_i. Similarly, the third requirement tells us that, for each y_j, the quantity transferred to y_j equals the quantity demanded at y_j. We assume the problem is balanced: the total supply available equals the total demand required. The cost of a transference plan γ is then

$$C(\gamma) := \sum_{x \in X, y \in Y} \Phi(x, y)\gamma(x, y).$$

From the set of admissible tranference plans γ, we want to find one that minimizes the cost functional $C(\gamma)$. Hence, we have a structure of a linear program(Savare, 2010):

Given positive coefficients $p_{0,i}, p_{1,j}$, and Φ_{ij}, find γ_{ij} that minimizes

$$C(\gamma) := \sum_{i,j} \Phi_{ij}\gamma_{ij},$$

subject to the constraints

$$\gamma(x_i, y_j) \geq 0, \sum_{y \in Y} \gamma(x_i, y) = p_0(x_i), \sum_{x \in X} \gamma(x, y_j) = p_1(y_j).$$

The yellow dots represent supply and the green dots represent demand. The quantity available at each point in X1 and the supply demanded at each point in X2 can be representated with the following matrices:

$$p_0 = \begin{pmatrix} 0.28701 \\ 0.19786 \\ 0.026992 \\ 0.48808 \end{pmatrix}, p_1 = \begin{pmatrix} 0.28807 \\ 0.30294 \\ 0.40898 \end{pmatrix}.$$

(Notice the problem is balanced: $\sum_{i=1}^{4} p_{0,i} = \sum_{j=1}^{3} p_{1,j}$).
We take the cost function to be the L^2 Wasserstein distance (Vilanni, 2003)

$$\Phi_{i,j} = \|x_i - y_j\|^2.$$

The optimal transport plan, solved using linear programming (Peyre, 2010), is then

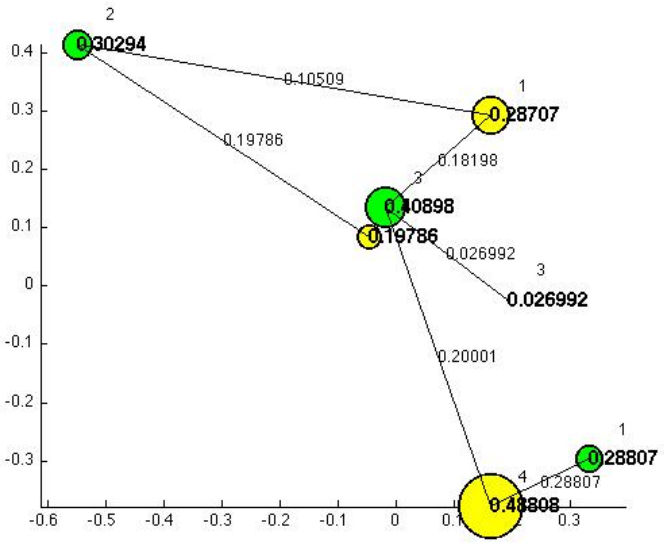

Figure 1: Visual Representation of Example

$$\gamma = \begin{pmatrix} 0 & 0.10509 & 0.18198 \\ 0 & 0.19786 & 0 \\ 0 & 0 & 0.026992 \\ 0.28807 & 0 & 0.20001 \end{pmatrix}.$$

In this example, we have a distribution of only a total of seven points. Let's consider a larger example.

Suppose now we have two discrete distributions μ_0 and μ_1. μ_0 has n_0 points, μ_1 has n_1 points. Hence, $X_0 = (x_{0,i})_{i=1}^{n_0}$ and $X_1 = (x_{1,i})_{i=1}^{n_1}$ are our point clouds. So,

$$\mu_k = \sum_{i=1}^{n_k} p_{k,i}\delta_{x_k,i} \quad k = 0, 1. \tag{5}$$

δ_x is the dirac at $x \in \mathbb{R}^2$ (Peyre, 2010). Just as before, we consider transference plans such that

$$\sum_j \gamma_{i,j} = p_{0,i}, \tag{6}$$

$$\sum_i \gamma_{i,j} = p_{1,j}. \tag{7}$$

We use the L^2 Wasserstein distance as our cost function

$$\Phi_{i,j} = \|x_{0,i} - x_{1,j}\|^2. \tag{8}$$

Hence, we are minimizing

$$\sum_{i,j} \gamma_{i,j}\Phi_{i,j}. \tag{9}$$

The following figures will show the optimal transport plans for various point clouds. We consider examples where μ_0, μ_1 are translated dilates of each other. Define $t(x) = \lambda x - z$ to be the translation mapping. If μ_1 is the image of μ_0 under t and we have a strictly convex cost function then, as shown in the figures, the translation mapping is the optimal solution(Gangbo and McCann, 1996).

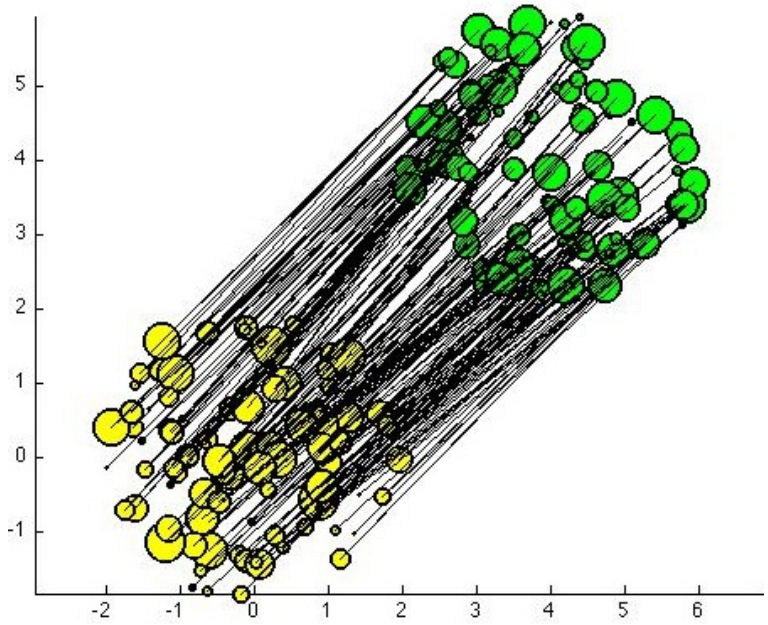

Figure 2: Point cloud X0 is uniformly distributed in a circle of radius 2, centered at (0,0). Point cloud X1 is uniformly distributed in a circle of radius 2, centered at (4,4).

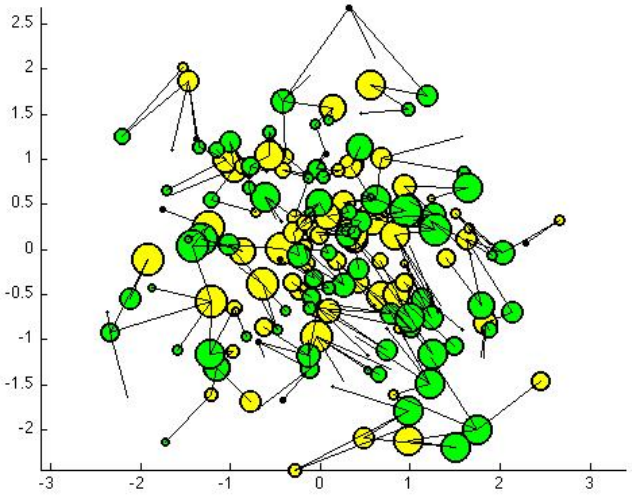

Figure 3: Both point clouds have the same normal distribution.

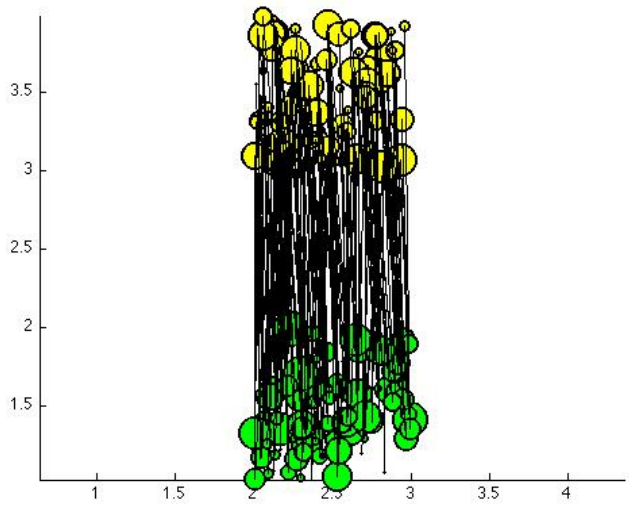

Figure 4: Point cloud X0 is uniformly distributed in a rectangle of length 1, width 1, centered at (2.5,3.5). Point cloud X1 is uniformly distributed in a rectangle of length 1, width 1, centered at (2.5,1.5)

3 Gradient Descent

In this section, we discuss another approach to solve the Monge-Kantorovich problem using gradient descent. We divide this section into two subsections: first, we will introduce the process of gradient descent simply by using domains in $\mathbb{R}^d$, densities on those domains, and defining a mapping from the first domain to the other (Haker, Zhu, Tannenbaum, and Angenent, 2004) . In the second section, we get more abstract, introducing the notion of measures as well as theorems and lemmas about the existence and uniqueness of solutions (Angenent, Haker, and Tannenbaum, 2003). Throughout the section, we denote ∇ as the gradient. Also, a superscript above a function indicates that said function is parameterized by superscript variable.

3.1 Mapping Approach

Let Ω_0 and Ω_1 be subdomains of $\mathbb{R}^d$ with smooth boundaries and with positive density functions m_0,m_1 respectively. We assume that the problem is balanced so the total mass associated with Ω_0 is the same as Ω_1:

$$\int_{\Omega_0} m_0(x)dx = \int_{\Omega_1} m_1(x)dx.$$

Note here that $x \in \mathbb{R}^d$ so for $d \geq 2$, x is a vector and dx is the standard Lebesque measure on the domain (Haker et al., 2004) . We will sometimes condense the notation by simply writing

$$\int_\Omega f(x)dx := \int_\Omega f.$$

We say that a diffeomorphism[1] $u : \Omega_0 \to \Omega_1$ satisfies the mass preservation property(MP) if:

$$m_0(x) = |Du(x)| m_1 (u(x)) , \tag{10}$$

(Haker et al., 2004). Here, Du is the jacobian map and $|Du|$ the determinant of the jacobian. We take the cost function to be the L^p Kantorovich-Wasserstein metric. The goal is to find $u(x) \in$ MP such that $u(x)$ is optimal, ie. it minimizes the Monge-functional:

$$M(u) = \int_{\Omega_0} \|u(x) - x\|^p m_0(x) dx.$$

We consider the case for p=2. An important result (Knott and Smith, 1984) that will help us derive a solution is that there exists a unique optimal $\tilde{u} \in MP$ and this $\tilde{u}$ is the gradient of a convex function w, ie. $\tilde{u} = \nabla p$.

Note that for this $\tilde{u}$,

$$D\tilde{u} = D(\nabla p) = Hp,$$

where H is the Hessian matrix and so the MP property becomes what we call the Monge-Ampère equation(Haker et al., 2004):

$$m_0 = |Hp| m_1(\nabla p).$$

Using the fact that the optimal solution has the form $\tilde{u} = \nabla p$, we first find an initial mapping $u^0 : (\Omega_0, m_0) \to (\Omega_1, m_1)$. Then, the polar factorization (Gangbo, 1994) of u^0 with respect to m_0 is

$$u^0 = \nabla p \circ s^t.$$

where p is a convex function and s is an MP mapping $s : (\Omega_0, m_0) \to (\Omega_0, m_0)$. Since we know the solution is of the form $\tilde{u} = \nabla p$,

$$\begin{aligned} \Rightarrow \nabla p &= u^0 \circ (s^t)^{-1} \\ \tilde{u} = \nabla p &= u^0 \circ (s^t)^{-1}. \end{aligned}$$

So, the problem of finding the optimal mapping now becomes the problem of finding the polar factorization of our initial mapping u^0 (Haker et al., 2004) . An important result (Strang and Aarikka, 1986) that will come into play is that for a vector field u, we can decompose u into the sum of a curl-free and a divergence-free vector field, called it's Helmholtz decomposition,

$$u = \nabla p + \xi,$$

where $div(\xi) = 0$. Let's first find the initial mapping.

[1] A diffeomorphism is a differentiable map between manifolds that is bijective and has a differentiable inverse.

Let's consider $\Omega_0 = \Omega_1 = [0,1]^2$ to be rectangles. To find the initial mapping, let $u^0(x,y) = (a(x), b(x,y))$ (Knothe, 1957). We find $a(x)$ by solving

$$\int_0^{a(x)} \int_0^1 m_1(\eta, y) dy d\eta = \int_0^x \int_0^1 m_0(\eta, y) dy d\eta. \tag{11}$$

To find $b(x,y)$, we solve

$$a'(x) \int_0^{b(x,y)} m_1(a(x), \rho) d\rho = \int_0^y m_0(x, \rho) d\rho. \tag{12}$$

We have that $Du = \begin{pmatrix} a_x & a_y \\ b_x & b_y \end{pmatrix}$ and so $|Du| = a_x b_y - b_x a_y$. Note that $a(x)$ is independent of y so $a_y = 0$. Hence, $|Du| = a_x b_y$. We need this fact to show that this u^0 satisfies the MP property. Differentiating (5) with respect to y, we get

$$\begin{gathered} a'(x) b_y(x,y) m_1(a(x), b(x,y)) = m_0(x,y) \\ a_x b_y m_1 = m_0 \\ |Du| m_1 = m_0. \end{gathered}$$

For the cases where the domain is not a rectangle, we use the method of Moser(1965) and Dacorogna and Moser(1990).

Now, take s^t to initially be the identity mapping and set $u^t = u^0 \circ (s^t)^{-1}$. We want the update for u^t to satisfy the MP property and hence, s must also satisfy the MP property. This gives us the evolution equations for s^t and u^t (Haker et al., 2004):

$$\boxed{s_t^t = \left(\frac{1}{m_0}\zeta\right) \circ s,} \tag{13}$$

$$\boxed{u_t^t = -\frac{1}{m_0} Du^t \zeta.} \tag{14}$$

Here, ζ is some vector field on Ω_0 such that $div(\zeta) = 0$ and $\langle \zeta, \vec{n} \rangle = 0$, $\vec{n}$ being the normal to the boundary of Ω_0.

Proof. We omit the proof for the evolution equation of s^t and just show the proof for u_t^t. We know $u^0 = u^t \circ s^t \Rightarrow u^0 = u^t(s^t(x))$ (Remember that $s^t(x)$ gives us a new point $x' \in \Omega_0$). Differentiate with respect to t (Haker et al., 2004):

$$\begin{aligned} \Rightarrow 0 &= u_t^t(s^t(x)) + (Du(s^t(x)))s_t^t, \\ \Rightarrow u_t^t \circ s^t &= -(Du^t \circ s^t)s_t^t, \\ \Rightarrow u_t^t &= -Du^t s_t^t \circ (s^t)^{-1}. \end{aligned} \tag{15}$$

Using the fact that $s_t^t = \left(\frac{1}{m_0}\zeta\right) \circ s^t \Rightarrow s_t^t \circ (s^t)^{-1} = \frac{1}{m_0}\zeta$. Plugging this into (8), we get our evolution equation

$$u_t^t = -\tfrac{1}{m_0} Du^t \zeta.$$

□

Now, the question is, what should we take ζ as? We want to minimize the Monge-Kantorovich functional so let us take the derivative with respect to t. This will lead us to a conclusion of what to ζ should be. We find that (Haker et al., 2004)

$$\frac{-1}{2} M_t = \int_{\Omega_0} \left\langle u^t, \zeta \right\rangle. \tag{16}$$

Proof. Recall that the MK functional is

$$\begin{aligned} M &= \int_{\Omega_0} \|u^t(x) - x\|^2 m_0(x) dx \\ &= \int_{\Omega_0} \|u^t(x)\|^2 m_0(x) dx - 2 \int_{\Omega_0} \left\langle u^t(x), x \right\rangle m_0(x) dx + \int_{\Omega_0} \|x\|^2 m_0(x) dx. \end{aligned}$$

Clearly, the last term is independent of t so the derivative with respect to t will be zero. Consider $\int_{\Omega_0} \|u^t(x)\|^2 m_0(x) dx$. First, we show that $m_0(x)dx = m_0(y)dy$ for $y \in \Omega_0$ (Haker et al., 2004). Since s^t is bijective, we can find y such that $x = s^t(y) \Rightarrow y = (s^t)^{-1}(x)$. So,

$$\begin{aligned} m_0(x)dx &= m_0(s^t(y)) d(s^t(y)), \\ &= m_0(s^t(y)) |Ds^t(y)| dy \\ &= m_0(y) dy. \end{aligned}$$

where we have used the fact that s^t satisfies the MP property so $m_0(s^t(y))|Ds^t(y)| = m_0(y)$. Now, noting that $u^t = u^0 \circ (s^t)^{-1}(x)$ and $(s^t)^{-1}(x) = y$,

$$\begin{aligned} \int_{\Omega_0} \|u^t(x)\|^2 m_0(x) dx &= \int_{\Omega_0} \|u^0 \circ (s^t)^{-1}(x)\|^2 m_0(x) dx \\ &= \int_{\Omega_0} \|u^0(y)\|^2 m_0(y) dy, \end{aligned}$$

and we see that this term is also independent of t. Hence, we are left with

$$\begin{aligned} M &= -2 \int_{\Omega_0} \left\langle u^t(x), x \right\rangle m_0(x) dx \\ \frac{-1}{2} M &= \int_{\Omega_0} \left\langle u^t(x), x \right\rangle m_0(x) dx \\ &= \int_{\Omega_0} \left\langle u^0 \circ (s^t)^{-1}(x), x \right\rangle m_0(x) dx \\ &= \int_{\Omega_0} \left\langle u^0(y), s^t(y) \right\rangle m_0(y) dy. \end{aligned}$$

Taking derivative with respect to t,

$$\begin{aligned}\frac{-1}{2}M_t &= \int_{\Omega_0} \left\langle u^0(y), s_t^t(y) \right\rangle m_0(y)dy \\ &= \int_{\Omega_0} \left\langle u^t \circ s^t(y), \frac{1}{m_0}\zeta \circ s^t(y) \right\rangle m_0(y)dy \\ &= \int_{\Omega_0} \left\langle u^t(x), \frac{1}{m_0(x)}\zeta(x) \right\rangle m_0(x)dx \\ &= \int_{\Omega_0} \left\langle u^t(x), \zeta(x) \right\rangle dx \\ &= \int_{\Omega_0} \left\langle u^t, \zeta \right\rangle .\end{aligned}$$

□

Once we find the Helmholtz Decomposition of u^t, $u^t = \nabla p + \chi$, we find that we take $\zeta = \chi$ (Haker et al., 2004).

Proof. Assuming we have found the Helmholtz Decomposition, we have that $u^t = \nabla p + \chi$ where $div(\chi) = 0$. Plugging this into (9), we get

$$\begin{aligned}\frac{-1}{2}M_t &= \int_{\Omega_0} \langle \nabla p + \chi, \zeta \rangle \\ &= \int_{\Omega_0} \langle \nabla p, \zeta \rangle + \int_{\Omega_0} \langle \chi, \zeta \rangle .\end{aligned}$$

Using the property $div(p\zeta) = \langle \nabla p, \zeta \rangle + pdiv(\zeta)$, we then get

$$\begin{aligned}\frac{-1}{2}M_t &= \int_{\Omega_0} (div(p\zeta) - pdiv(\zeta)) + \int_{\Omega_0} \langle \chi, \zeta \rangle \\ &= \int_{\Omega_0} div(p\zeta) + \int_{\Omega_0} \langle \chi, \zeta \rangle .\end{aligned}$$

Using the divergence theorem,

$$\begin{aligned}\frac{-1}{2}M_t &= \int_{\delta\Omega_0} (p\zeta) \cdot \vec{n} + \int_{\Omega_0} \langle \chi, \zeta \rangle \\ &= \int_{\delta\Omega_0} p \langle \zeta, \vec{n} \rangle + \int_{\Omega_0} \langle \chi, \zeta \rangle \\ &= \int_{\Omega_0} \langle \chi, \zeta \rangle ,\end{aligned}$$

since $\langle \zeta, \vec{n} \rangle = 0$ on $\delta\Omega_0$. □

From this, the easy logical choice is $\zeta = \chi$. Now all that remains is finding the Helmholtz Decomposition of u^t.

Helmholtz Decomposition The Helmholtz Decomposition of u can be found by setting up a Neumann boundary problem (Haker et al., 2004). Knowing that $div(\zeta) = 0$ and $\langle \zeta, \vec{n} \rangle = 0$ on $\delta\Omega_0$,

$$\begin{aligned} u^t &= \nabla p + \zeta \\ \Rightarrow div(u^t) &= div(\nabla p) + div(\zeta), \\ div(u^t) &= \Delta p. \end{aligned}$$

Hence, the Neumann Boundary problem that we solve is

$$\begin{aligned} \Delta p &= div(u^t), \\ \langle \nabla p, \vec{n} \rangle &= \left\langle u^t, \vec{n} \right\rangle \quad \text{on } \delta\Omega_0. \end{aligned} \tag{17}$$

and we set $\chi = u^t - \nabla p$. Consider $u^t = div^{-1}(\nabla p)$ from (17). So, $\chi = div^{-1}(\nabla p) - \nabla p$. Plug this into our evolution equation (7) to get:

$$\boxed{u_t^t = -\frac{1}{m_0} Du^t(u^t - \nabla \Delta^{-1} div(u^t)).} \tag{18}$$

Note that, if $\tilde{u}$ is the optimal mapping, it is given by $\tilde{u} = \nabla p$, and we have

$$\begin{aligned} \tilde{u} - \nabla \Delta^{-1} div(\tilde{u}) &= \nabla p - \nabla \Delta^{-1} div(\nabla p) \\ &= \nabla p - \nabla \Delta^{-1} \Delta p \\ &= \nabla p - \nabla p \\ &= 0. \end{aligned}$$

So, $\tilde{u_t} = 0$ and we say that the process has achieved steady state (Haker et al., 2004).

Algorithm Summary Find the initial mapping u^0 using the process described and let s initially be the identity mapping. Let $u^t = u^0 \circ (s^t)^{-1}$. Decompose u^t into it's Helmholtz decomposition, $u^t = \nabla p + \chi$ by solving (17). Use the χ found to update u^t according to $u_t^t = \frac{-1}{m_0} Du^t \chi$.

3.2 Reallocation Measures Approach

In this section, we generalize the notion of a mapping from $u : \Omega_0 \to \Omega_1$ by identifying the measure preserving map u by it's Borel measure γ_u on $\Omega_0 \times \Omega_1$ (Angenent et al., 2003).

Suppose we have two borel measures μ_0, μ_1 on Ω_0, Ω_1 with density functions m_0 and m_1 respectively, ie.

$$\begin{aligned} d\mu_0 &= m_0(x)dx, \\ d\mu_1 &= m_1(x)dx. \end{aligned}$$

We want the same total mass associated with Ω_0 and Ω_1, so

$$\mu_0(\Omega_0) = \mu_1(\Omega_1),$$

or

$$\int_{\Omega_0} m_0(x)dx = \int_{\Omega_1} m_1(x)dx.$$

Given a map $u : (\Omega_0, \mu_0) \to (\Omega_1, \mu_1)$ that satisfies the MP property (4), we introduce the Borel measure γ_u on $\Omega_0 \times \Omega_1$ which is defined by

$$\gamma_u(E) = \mu_0\left(\{x \in \Omega_0 : (x, u(x)) \in E\}\right).$$

This measure is the pushforward of μ_0 under the map $(id \times u)$ so we write $\gamma_u = (id \times u)_{\#}\mu_0$ (Angenent et al., 2003).
Since we want the map u to be measure preserving, we consider

$$\Xi = \{ \text{ Borel measures } \gamma \geq 0 \text{ on } \Omega_0 \times \Omega_1 | (p_j)_{\#}\gamma = \mu_j, j = 0, 1\},$$

where $p_j : \Omega_0 \times \Omega_1 \to \Omega_j$ is the canonical projection. Then, if the measure γ has a density m, ie. $d\gamma(x, y) = m(x, y)dxdy$, $\gamma \in \Xi$ when

$$\int_{\Omega_0} m(x, y)dx = m_1(y),$$
$$\int_{\Omega_1} m(x, y)dy = m_0(x).$$

Finally, all measures $\gamma \in \Xi$ have total mass

$$\gamma(\Omega_0 \times \Omega_1) = \mu_0(\Omega_0) = \mu_1(\Omega_1).$$

With respect to this measure, the MK functional is given by

$$M(\gamma) = \langle \gamma, \Phi \rangle = \int_{\Omega_0 \times \Omega_1} \Phi(x, y)d\gamma(x, y), \tag{19}$$

where $\Phi : \bar{\Omega_0} \times \bar{\Omega_1} \to \mathbb{R}$ is a positive C^1 cost function.
Also, we will need the following: for any bounded Borel measurable function $\phi : \Omega_0 \times \Omega_1 \to \mathbb{R}$, we write

$$\mathbb{E}\left(\phi(x, y)|x\right) = \int_{\Omega_1} \phi(x, y)dP_x(y)$$

for the expectation of $\phi(x, \cdot)$ with respect to the probability measure P_x (Angenent et al., 2003).

As before, given an initial map u^0, let $u^t = u^0 \circ (s^t)^{-1}$ for a family of diffeomorphisms $s^t : \Omega_0 \to \Omega_0$. Note that since u^0 sends μ_0 to μ_1 and s^t preserves the measure μ_0 then $u^t = u^0 \circ (s^t)^{-1}$ also sends μ_0 to μ_1 (Angenent et al., 2003).

This can be extended to an action on Ξ by

$$s \cdot \gamma = (s \times id_{\Omega_1})_{\#}\gamma.$$

The family of diffeomorphisms $s^t : \Omega_0 \to \Omega_0$ will be determined by it's velocity defined by $\partial_t\ s^t = v^t \circ s^t$ (Angenent et al., 2003). We choose the velocity field such that we minimize the MK functional (19).

Assume we have a single parameter family of μ_0 preserving C^1 diffeomorphisms $s^t : \Omega_0 \to \Omega_0$ with velocity field v^t and $\gamma^t = s^t \cdot \gamma$ for some $\gamma \in \Xi$ (Angenent et al., 2003). Then,

$$\begin{aligned}
\frac{dM(\gamma^t)}{dt} &= \frac{d}{dt}\int_{\Omega_0\times\Omega_1} \Phi(x,y) d(s^t \times id)_{\#}\gamma(x,y) \\
&= \frac{d}{dt}\int_{\Omega_0\times\Omega_1} \Phi(s^t(x),y) d\gamma(x,y) \\
&= \int_{\Omega_0\times\Omega_1} \partial_t s^t \cdot \Phi_x(s^t(x),y) d\gamma(x,y) \\
&= \int_{\Omega_0\times\Omega_1} v^t(s^t(x)) \cdot \Phi_x(s^t(x),y) d\gamma(x,y) \\
&= \int_{\Omega_0\times\Omega_1} v^t(x) \cdot \Phi_x(x,y) d\gamma^t(x,y). \qquad (20)
\end{aligned}$$

We need the following lemma to proceed.

Lemma 1. *(Angenent et al., 2003) For any bounded measurable function $F : \Omega_0 \times \Omega_1 \to \mathbb{R}$ there exists a bounded measurable function $\tilde{F} : \Omega_0 \to \mathbb{R}$ for which*

$$\int_{\Omega_0\times\Omega_1} \phi(x)F(x,y)d\gamma(x,y) = \int_{\Omega_0} \phi(x)\tilde{F}(x)d\mu_0(x) \qquad (21)$$

holds for all $\phi \in L^1(\Omega_0; d\mu_0)$.

We denote $\tilde{F}$ by

$$\tilde{F}(x) = \mathbb{E}_\gamma(F(x,y)|x).$$

If γ has density $m(x,y)$ then

$$\tilde{F}(x) = \mathbb{E}_\gamma\left(F|x\right) = \int_{\Omega_1} F(x,y)\frac{m(x,y)}{m_0(x)}dy. \qquad (22)$$

So, directly from (21), (20) becomes

$$\frac{dM(\gamma^t)}{dt} = \int_{\Omega_0} v^t(x) \cdot W^t(x) d\mu_0(x), \qquad (23)$$

where

$$W^t(x) = \mathbb{E}_{\gamma^t}(\Phi_x(x,y)|x), \qquad (24)$$

(Angenent et al., 2003). We use the notation $\mathcal{P}$ to denote the helmholtz projection. It just means that $div(\mathcal{P}(\chi)) = 0$ for any vector field χ. We split W^t by using the Helmholtz decomposition to get

$$W^t = \nabla p^t + \mathcal{P}(W^t),$$

where $div(\mathcal{P}(W^t)) = 0$ (Angenent et al., 2003). Plugging this into (23), we get

$$\frac{dM(\gamma^t)}{dt} = \int_{\Omega_0} m_0(x) v^t(x) \mathcal{P}(W^t) dx. \tag{25}$$

The velocity field we choose (Angenent et al., 2003) is

$$v^t = \frac{-1}{m_0(x)} \mathcal{P}\mathcal{A}^2\mathcal{P}(W^t), \tag{26}$$

where $\mathcal{A}$ is an operator on the Hilbert space

$$\hbar = L^2(\Omega_0) \otimes \mathbb{R}^d.$$

See appendix for more details on operators and tensor products. We assume that $\mathcal{A}$ is a bounded, symmetric, and injective operator on $\hbar$ (Angenent et al., 2003). We now state the evolution equations for the measure γ^t, map u^t, and s^t.

Evolution Equations Let $\gamma^t = (s^t \times id)_{\#}\gamma^0$ for some initial $\gamma^0 \in \Xi$. The evolution equation for γ^t (Angenent et al., 2003) is

$$\boxed{\frac{\partial \gamma^t}{\partial t} = \nabla_x \cdot \left(\frac{1}{m_0(x)} \mathcal{P}\mathcal{A}^2\mathcal{P} \left(\mathbb{E}_{\gamma^t}(\Phi_x | x) \right) \gamma^t \right).} \tag{27}$$

To derive the evolution equation of s^t, we assume that the operator $\mathcal{P}\mathcal{A}^2\mathcal{P}$ can be represented as an integral operator with kernel $K(x, \xi)$ (Angenent et al., 2003). So, for a vector field $W \in L^2(\Omega_0; \mathbb{R}^d)$, we can write

$$\left(\mathcal{P}\mathcal{A}^2\mathcal{P}W\right)(x) = \textstyle\int_{\Omega_0} K(x, \xi) W(\xi) d\xi,$$

where dy is the Lebesgue measure, $K(x, y)$ is an $n \times n$ matrix and $K(x, y) \cdot W(y)$ is pointwise matrix multiplication. Rewriting the velocity field by replacing $\mathcal{P}\mathcal{A}^2\mathcal{P}$ with the integral kernel, we get

$$\begin{aligned} v^t(x) &= \frac{-1}{m_0(x)} \int_{\Omega_0} K(x, \xi) \mathbb{E}_{\gamma^t} \left(\Phi_x(\xi, \eta) | \xi \right) d\xi \\ &= \frac{-1}{m_0(x)} \int_{\Omega_0 \times \Omega_1} K(x, \xi) \cdot \Phi_x(\xi, \eta) \frac{d\gamma^t(\xi, \eta)}{m_0(\xi)} \quad \text{from Lemma 1,} \\ &= -\int_{\Omega_0 \times \Omega_1} \frac{K(x, \xi)}{m_0(x) m_0(\xi)} \cdot \Phi_x(\xi, \eta) d\gamma^t(\xi, \eta). \end{aligned}$$

Recalling that $\partial_t s^t = v^t \circ s^t$ and combining it with this velocity field, we get

$$\frac{\partial s^t}{\partial t} = -\int_{\Omega_0 \times \Omega_1} \frac{K(s^t(x), \xi)}{m_0(s^t(x)) m_0(\xi)} \cdot \Phi_x(\xi, \eta) d\gamma^t(\xi, \eta). \tag{28}$$

Using $\gamma^t = (s^t \times id)_{\#}\gamma^0$, we get

$$\boxed{\frac{\partial s^t}{\partial t} = -\int_{\Omega_0 \times \Omega_1} \frac{K(s^t(x), s^t(\xi))}{m_0(s^t(x))m_0(s^t(\xi))} \cdot \Phi_x(s^t(\xi), \eta) d\gamma^0(\xi, \eta).} \tag{29}$$

Finally, if the measure γ^t is given by $\gamma^t = \gamma_{u^t}$ for $u^t : (\Omega_0, \mu_0) \to (\Omega_1, \mu_1)$, then $u^0 = u^t \circ s^t$. The family of diffeomorphisms u^t satisfy the transport equation (Angenent et al., 2003)

$$\frac{\partial u^t}{\partial t} + v^t \cdot \nabla u^t = 0. \tag{30}$$

The velocity field is given by (26). However, for $\gamma^t = \gamma_{u^t}$, we can take $W^t = \Phi_x(x, u^t(x))$ (Angenent et al., 2003) so we have

$$v^t = \frac{-1}{m_0(x)} \mathcal{P}\mathcal{A}^2\mathcal{P}\{\Phi_x(x, u^t(x))\}. \tag{31}$$

Plugging this into (30), we get our equation

$$\boxed{\frac{\partial u^t}{\partial t} - \frac{1}{m_0(x)} \mathcal{P}\mathcal{A}^2\mathcal{P}\{\Phi_x(x, u^t(x))\} \cdot \nabla u^t = 0.} \tag{32}$$

Now that we have our evolution equations set up, let's discuss the notion of weak solutions.

Weak Solutions Let $\gamma^t = (s^t \times id)_{\#}\gamma^0$ for a smooth family of diffeomorphisms $s^t : \bar{\Omega_0} \to \bar{\Omega_0}$ whose velocity field is (26).

Definition 1. *(Angenent et al., 2003) A weak solution to (27) is a map $t \in [0, T) \longmapsto \gamma^t \in \Xi$ which is weak continuous and which satisfies*

$$\int_{t_0}^{t_1} \left(\mathcal{A}\mathcal{P}\mathbb{E}_{\gamma^t}\left(\varphi_x | x\right), \mathcal{A}\mathcal{P}\mathbb{E}_{\gamma^t}\left(\Phi_x | x\right)\right)_{\hbar} \tag{33}$$

for all test functions $\varphi \in C^1(\bar{\Omega_0} \times \bar{\Omega_1})$ and for all $0 \leq t_0 < t_1 < T$.

Definition 2. *(Angenent et al., 2003) A classical solution is a family of measures $\{\gamma^t, t \in [0, T)\}$ of the form $\gamma^t = (s^t \times id)_{\#}\gamma^0$ for some family of C^1 diffeomorphisms $s^t : \Omega_0 \to \Omega_0$ whose velocity field $v^t = (\partial_t s^t) \circ (s^t)^{-1}$ satisfies $v^t = \frac{-1}{m_0}\mathcal{P}\mathcal{A}^2\mathcal{P}\{\mathbb{E}_{\gamma^t}(\Phi_x | x)\}$.*

From (33), if we let $\varphi = \Phi$, we get the energy identity:

Lemma 2. *(Angenent et al., 2003)(Energy Identity) For any weak solution $\{\gamma^t, t \in [0, T)\}$ and any $0 \leq t_0 < t_1 < T$, we have*

$$\int_{t_0}^{t_1} \|\mathcal{A}\mathcal{P}\mathbb{E}_{\gamma^t}\left(\Phi_x | x\right)\|_{\hbar}^2 dt = M(\gamma^{t_0}) - M(\gamma^{t_1}). \tag{34}$$

From this, we get the result

Lemma 3. *(Angenent et al., 2003) For any weak solution $\{\gamma^t, t \in [0,T)\}$ the Monge-Kantorovich cost functional is nonincreasing. It is constant if and only if $\mathcal{P}\mathbb{E}_{\gamma^t}(\Phi_x|x) = 0$ for almost all $t \in [0,T)$.*

Proof. To show nonincreasing is simple: from the energy identity (34), $M(\gamma^{t_0}) - M(\gamma^{t_1}) \geq 0$ with $0 \leq t_0 < t_1 < T$. So, $M(\gamma^t)$ is nonincreasing.
We now prove the if and only if statement.

$\boxed{\Rightarrow}$ Assume the MK functional is constant. So, $M(\gamma^{t_0}) = M(\gamma^{t_1})$, $t_0 < t_1$.
Then $\mathcal{A}\mathcal{P}\mathbb{E}_{\gamma^t}(\Phi_x|x) = 0$ from the energy identity.
But, $\mathcal{A}$ is injective so $\mathcal{P}\mathbb{E}_{\gamma^t}(\Phi_x|x) = 0$.

$\boxed{\Leftarrow}$ This direction is trivial:
$0 = M(\gamma^{t_0}) - M(\gamma^{t_1})$ from the energy identity (34) for any $0 \leq t_0 < t_1 < T$.
$M(\gamma^{t_0}) = M(\gamma^{t_1})$.
Hence, $M(\gamma^t)$ is constant. □

Notice that when $\mathcal{P}\mathbb{E}_{\gamma^t}(\Phi_x|x) = 0$, we get that $\mathbb{E}_{\gamma^t}(\Phi_x|x) = \nabla p^t$ when we break $\mathbb{E}_{\gamma^t}(\Phi_x|x)$ into it's Helmholtz decomposition. This is relevant because a critical point is a measure $\gamma \in \Xi$ for which $\mathbb{E}_{\gamma}(\Phi_x|x) = \nabla p$ for a lipschitz continuous function $p : \Omega_0 \to \mathbb{R}$ (Angenent et al., 2003).

Existence and Uniqueness of Solutions Recall the evolution equation of s^t. We can construct classical solutions $\gamma^t = (s^t \times id)_{\#}\gamma^0$ (Angenent et al., 2003) by adding an initial condition (Angenent et al., 2003) and solving the initial value problem

$$\frac{\partial s^t}{\partial t} = -\int_{\Omega_0 \times \Omega_1} \frac{K(s^t(x), s^t(\xi))}{m_0(s^t(x))m_0(s^t(\xi))} \cdot \Phi_x(s^t(\xi), \eta) d\gamma^0(\xi, \eta), \tag{35}$$

$$s^0(x) = x, \quad x \in \Omega_0. \tag{36}$$

We will write

$$F(s, \sigma; \xi, \eta) = \frac{K(s^t(x), s^t(\xi))}{m_0(s^t(x))m_0(s^t(\xi))} \cdot \Phi_x(s^t(\xi), \eta). \tag{37}$$

So, F is a map from $\Omega_0 \times \Omega_0 \times \Omega_0 \times \mathbb{R}^d$ to $\mathbb{R}^d$ (Angenent et al., 2003).

Theorem 1. *(Angenent et al., 2003) Let the cost function Φ be C^1. Assume the integral kernel K of the smoothing operator $\mathcal{A}$ is $C^{1,\infty}$. Then, for any initial measure $\gamma^0 \in \Xi$ the initial value problem (35) has a solution $\{s^t \in C^{1,\infty}(\bar{\Omega_0}; \bar{\Omega_0}) : 0 < t < \infty\}$.*

If the cost function Φ is C^2 then the solution $\{s^t : t \geq 0\}$ is unique.

We first prove the existence of a solution. To do this, we set up a fixed point argument (Angenent et al., 2003). In order to use the argument, we will require to extend F. Then,

we can define a solution to the initial value problem by a mapping with will have certain properties that will allow us to use the Ascoli-Arzela theorem. Once we have all that, we can apply the fixed point theorem. Let's begin.

Theorem 2. *(McOwen, 2003)* ***Formulation of Fixed Point Theorem*** *Let A be a closed convex set in a Banach space X and $T : A \to A$ be continuous such that T(A) is compact in X. Then T has a fixed point.*

In order to use the argument, we need the space of maps $\{s^t : \Omega_0 \to \Omega_0, 0 \leq t < T\}$ to be linear. So, we instead regard the space of maps $s^t : \Omega_0 \to \mathbb{R}^d$. Hence, we extend $F(s, \sigma, \xi, \eta)$ to include all $(s, \sigma, \xi, \eta) \in \mathbb{R}^d \times \mathbb{R}^d \times \Omega_0 \times \Omega_1$ (Angenent et al., 2003).

The following lemma states some important properties of our extended F, denoted F_*.

Lemma 4. *(Angenent et al., 2003) The extension $F_* : \mathbb{R}^d \times \mathbb{R}^d \times \Omega_0 \times \Omega_1 \to \mathbb{R}^d$ of F is continuous in (s, σ), and uniformly Lipschitz in $s \in \mathbb{R}^d$,ie. $\forall \sigma \in \mathbb{R}^d$ and $\eta \in \Omega_0, \eta \in \Omega_1$, we have*

$$|F_*(s, \sigma, \xi, \eta) - F_*(s', \sigma, \xi, \eta)| \leq M|s - s'|, \tag{38}$$

for some $M < \infty$. Also, F_* is uniformly bounded,

$$|F_*(s, \sigma, \xi, \eta)| \leq M', \tag{39}$$

for some $M' < \infty$ and $\forall s, \sigma \in \mathbb{R}^d and \xi \in \Omega_0, \eta \in \Omega_1$.

If the cost function Φ is C^2 then $F_*(s, \sigma, \xi, \eta)$ is uniformly Lipschitz in $(s, \sigma) \in \mathbb{R}^d$, ie.,

$$|F_*(s, \sigma, \xi, \eta) - F_*(s', \sigma', \xi, \eta)| \leq M''\{|s - s'| + |\sigma - \sigma'|\}, \tag{40}$$

for $M'' < \infty, \forall s, s', \sigma, \sigma' \in \mathbb{R}^d$ and $\xi \in \Omega_0$,$\eta \in \Omega_1$.

Before we can prove existence, we need the following definition in order to state the Ascoli-Arzela theorem.

Definition 3. *(Scheidemann, 2005) Let (X, d) be a compact metric space, $(E, \|\cdot\|_E)$ a Banach space, $U \subset X$ an open subset and $\mathcal{F}$ a family of functions $f : U \to E$. The family $\mathcal{F}$ is called equicontinuous if for every $\epsilon > 0$ $\exists \delta = \delta_\epsilon > 0$ such that $\|f(x) - f(y)\| < \epsilon$ $\forall f \in \mathcal{F}$ and $x, y \in U$ satisfying $d(x, y) < \delta$*

Now we can state the Ascoli-Arzela theorem which will we will need in order to use the fixed point argument.

Theorem 3. *(Schlumprecht)****(Generalized Ascoli-Arzela Theorem)*** *Assume $\mathcal{F} \subset C(X, Y) = \{f : X \to Y continuous\}$. Then $\mathcal{F}$ is compact if and only if*

- *$\bigcup_{f \in \mathcal{F}} f(X)$ is totally bounded,*
- *$\mathcal{F}$ is closed,*
- *$\mathcal{F}$ is equicontinuous.*

We are now ready to set up our fixed point problem. Given F_*, let $\mathcal{C}_T$ (Angenent et al., 2003) be the Banach space

$$\mathcal{C}_T = C^0\left([0,T]\times\Omega_0;\mathbb{R}^d\right).$$

Lemma 5. *(Angenent et al., 2003) Let $\sigma\in\mathcal{C}_T$ be given. Define $s=\mathcal{F}(\sigma)$ to be the solution of*

$$\frac{\partial s^t}{\partial t} = \int_{\Omega_0} F_*(s^t(x),\sigma^t(\xi);\xi,\eta)d\gamma^0(\xi,\eta), \quad s^0(x)=x. \tag{41}$$

Then,

$$\left|\frac{\partial s^t}{\partial t}\right| \le M'|\Omega_0|, \tag{42}$$

and

$$\left|s^t(x)-s^t(x')\right| \le e^{M|\Omega_0|t}|x-x'|. \tag{43}$$

Note that (42),(43) imply that $\mathcal{F}(\sigma)$ is compact by Ascoli-Arzela. Hence, $\mathcal{F}$ maps all of $\mathcal{C}_T$ into a compact subset of $\mathcal{C}_T$. So, by the fixed point theorem, there exists a fixed point $s_T\in\mathcal{C}_T$ for $\mathcal{F}$ (Angenent et al., 2003). Therefore, the initial value problem (33) has a solution on a finite time interval $0\le t<T$.

Finally, we can discuss uniqueness.

Theorem 4. *(Angenent et al., 2003) If the cost function Φ is C^2 then there is only one solution.*

Proof. Let $s^t,\bar{s^t}\in\mathcal{C}_T$ be any two solutions to the IVP and consider $w^t(x)=s^t(x)-\bar{s^t}(x)$ (Angenent et al., 2003). Let

$$v^t(s) = \int_{\Omega_0} F_*(s,\sigma^t(\xi);\xi,\eta)d\gamma^0(\xi,\eta). \tag{44}$$

So, $\partial_t s^t = v^t(s^t)$ and $\partial_t\bar{s}^t = v^t(\bar{s}^t)$. Then, by (40), it follows that

$$|v^t(s)-v^t(\bar{s})| \le M|\Omega_0|\sup_{\xi\in\Omega_0}|s(\xi)-\bar{s}(\xi)|. \tag{45}$$

Since $|v^t(s)-v^t(\bar{s})| = |\partial_t w^t|$ and $|s(\xi)-\bar{s}(\xi)| = w^t(\xi)$, we get

$$|\partial_t w^t| \le M|\Omega_0|\sup_{\xi\in\Omega_0}|w^t(\xi)|. \tag{46}$$

Then, by standard theorems for ODEs (Angenent et al., 2003), (46) implies

$$\sup|w^t| \le e^{M''|\Omega_0|t} sup|w^0|. \tag{47}$$

Since $w^0 = s^0-\bar{s}^0 = x-x = 0$ (since we have the same initial condition), we have that $w^t=0$. Hence, $s^t=\bar{s}^t$. □

4 Conclusion

The optimal transport problem has many applications in areas such as image registration, fluid dynamics, and economics (Angenent et al., 2003). We discussed two possible methods of finding a solution to the optimal transport problem. We start by discretizing the domain and so the transport problem becomes a matching problem. The brute force method of using linear programming becomes computationally complex once we get into larger domains (Haker et al., 2004).

We then introduce the approach of solving partial differential equations. This becomes the equivalent of finding the polar factorization of a measure preserving mapping. Compared to the linear programming approach, this method is a lot more efficient in terms of speed of finding a solution (Haker et al., 2004).

5 Appendix

Tensor Products The tensor product of two vector spaces V and W, denoted $V \otimes W$, is analogous to multiplication of integers. For example, $\mathbb{R}^n \otimes \mathbb{R}^k = \mathbb{R}^{nk}$. Let $\{\vec{e_1}, ..., \vec{e_n}\}$ be the basis for the V and $\{\vec{f_1}, ..., \vec{f_n}\}$ be the basis for W. Then, the basis for $V \otimes W$ is given by $\vec{e_i} \otimes \vec{f_j}$ for $i = 1, ..., n$,$j = 1, ..., m$ (Murayama).

The vector space $V \otimes W$ is spanned by elements of the form $V \otimes W$ (Rowland) such that for $v_1, v_2 \in V$, $w_1, w_2 \in W$, and any scalar λ

- $(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w$,
- $v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2$,
- $\alpha(v \otimes w) = (\alpha v) \otimes w = v \otimes (\alpha w)$.

Note that every element of $V \otimes W$ can be written as

$$\sum a_{ij} v_i \otimes w_j,$$

where a_{ij} are scalars.

Let's give a simple example. Let V_1, V_2 be independent spaces of real valued quadratic functions on $[0, 1]$ (McCullagh,1987). So,

$$V_1 = span\{1, x_1, x_1^2\},$$
$$V_2 = span\{1, x_2, x_2^2\}.$$

The tensor product $W = V_1 \otimes V_2$ is the space of functions on the unit square:

$$W = V_1 \otimes V_2 = span\{1, x_1, x_1^2, x_2, x_1x_2, x_1^2x_2, x_2^2, x_1x_2^2, x_1^2x_2^2\}.$$

Smoothing Operators For the topic in this paper, the smoothing operators $\mathcal{A}$ have to satisfy the assumptions made previously (Angenent et al., 2003). Here, we state some possible smoothing operators as well as properties of the operators.

Lemma 6. *(Angenent et al., 2003) Let Ω_0 be a domain with $C^{1,\alpha}$ boundary and let A_ϵ be the operator*

$$\mathcal{A}_\epsilon = e^{\epsilon\Delta_N}, \text{ ie. } \mathcal{A}_\epsilon w = (e^{\epsilon\Delta_N} w_1, ..., e^{\epsilon\Delta_N} w_d),$$

where Δ_N is the Neumann Laplacian on Ω_0.

The operator $\mathcal{P}\mathcal{A}^2\mathcal{P}$ is bounded from $\hbar$ to $C^{1,\alpha}(\Omega;\mathbb{R}^d)$ for any $0 < \alpha < 1$.

The operator $\mathcal{P}\mathcal{A}^2\mathcal{P}$ has an integral kernel $K \in C^{1,\alpha}(\overline{\Omega_0} \times \overline{\Omega_0})$.

When Ω_0 is a rectangle, so $\Omega_0 = [0,\pi]^d$, we can define another smoothing operator.

Lemma 7. *(Angenent et al., 2003) Let A_ϵ be the operator*

$$\mathcal{A}_\epsilon w = (e^{\epsilon\Delta_1} w_1, ..., e^{\epsilon\Delta_d} w_d),$$

where Δ_j is the laplacian with Neumann boundary conditions on the sides $x_j = 0$ and $x_j = \pi$ and Dirichlet boundary conditions on the other sides of the rectangle Ω_0.

The operator $\mathcal{P}\mathcal{A}^2\mathcal{P}$ is bounded from $\hbar$ to $C^{1,\alpha}(\Omega;\mathbb{R}^d)$ for any $0 < \alpha < 1$.

The operator $\mathcal{P}\mathcal{A}^2\mathcal{P}$ has an integral kernel $K \in C^{1,\alpha}(\overline{\Omega_0} \times \overline{\Omega_0})$.

References

[1] Angenent, S., Haker, S., and Tannenbaum, A. (2003). Minimizing Flows for the Monge-Kantorovich Problem. *SIAM journal on mathematical analysis*, 35(1), 61-97.

[2] Benamou, J. D., Froese, B. D., and Oberman, A. M. (2014). Numerical solution of the Optimal Transportation problem using the Monge–Ampère equation. *Journal of Computational Physics, 260*, 107-126.

[3] Dacorogna, B., and Moser, J. (1990). On a partial differential equation involving the Jacobian determinant. In *Annales de l'Institut Henri Poincaré. Analyse non linéaire* (Vol. 7, No. 1, pp. 1-26). Elsevier.

[4] Evans, L. C. (1997). Partial differential equations and Monge-Kantorovich mass transfer. *Current developments in mathematics, 1999*, 65-126.

[5] Gangbo, W. (1994). An elementary proof of the polar factorization of vector-valued functions. *Archive for rational mechanics and analysis, 128(4)*, 381-399.

[6] Gangbo, W., and McCann, R. J. (1996). The geometry of optimal transportation. *Acta Mathematica, 177(2)*, 113-161.

[7] Haker, S., Zhu, L., Tannenbaum, A., and Angenent, S. (2004). Optimal mass transport for registration and warping. *International Journal of Computer Vision, 60(3)*, 225-240.

[8] Knothe, H. (1957). Contributions to the theory of convex bodies. *Michigan Math J, 4*, 39-52.

[9] Knott, M., and Smith, C. S. (1984). On the optimal mapping of distributions. *Journal of Optimization Theory and Applications, 43(1)*, 39-49.

[10] McCann, R. J., and Guillen, N. (2011). Five lectures on optimal transportation: geometry, regularity and applications. *Analysis and Geometry of Metric Measure Spaces: Lecture Notes of the Séminaire de Mathématiques Supérieure (SMS) Montréal*, 145-180.

[11] McCullagh, P. (1987). *Tensor methods in Statistics*. London: Chapman and Hall.

[12] McOwen, R. (2003). *Partial Differential Equations: Methods and Applications*. Pearson Education, 2003.

[13] Moser, J. (1965). On the volume elements on a manifold. *Transactions of the American Mathematical Society*, 286-294.

[14] Murayama, H. *221A Lecture Notes: Notes on Tensor Product*. Retrieved from University of California, Berkeley.

[15] Peyre, G. (2010). *Optimal Transport with Linear Programming*. Retrieved from https://www.ceremade.dauphine.fr/∼peyre/numerical-tour/tours/optimaltransp_1_linprog/

[16] Rowland, T. *Vector Space Tensor Product*. Retrieved from http://mathworld.wolfram.com/VectorSpaceTensorProduct.html

[17] Savare, G. (2010). *A quick introduction to Optimal Transport*. Retrieved from University of Pavia.

[18] Scheidemann, V. (2005). *Introduction to complex analysis in several variables*. Basel: Birkhäuser Verlag.

[19] Schlumprecht, T. *Math608 Solutions: Problems in Real Variables*. Retrieved from http://www.math.tamu.edu/∼thomas.schlumprecht/hw3_608_02_10a_sol.pdf

[20] Strang, G., and Aarikka, K. (1986). *Introduction to applied mathematics* (Vol. 16). Wellesley, MA: Wellesley-Cambridge Press.

[21] Villani, C. (2003). Topics in optimal transportation (No. 58). American Mathematical Soc..

[22] Wang, W., Slepčev, D., Basu, S., Ozolek, J. A., and Rohde, G. K. (2013). A linear optimal transportation framework for quantifying and visualizing variations in sets of images. *International journal of computer vision, 101(2)*, 254-269.

[23] Zemel, Y. (2012). Optimal Transportation: Continuous and Discrete.