

MOTIVATIONS, IDEAS AND APPLICATIONS OF RAMIFIED OPTIMAL TRANSPORTATION *

QINGLAN XIA¹

Abstract. In this survey article, the author summarizes the motivations, key ideas and main applications of ramified optimal transportation that the author has studied in recent years.

Mathematics Subject Classification. 90B10, 49Q10, 49Q20.

Received March 10, 2015.

Published online November 5, 2015.

1. MOTIVATIONS AND THE GENERAL SET UP

1.1. Motivations

In this survey article, the author summarizes the motivations, key ideas and some applications of ramified optimal transportation which the author has studied in recent years. For the author, there are at least three motivations for studying the ramified optimal transportation.

The first motivation arises directly in the field of optimal transportation. The transport problem introduced by Monge in 1781 has been studied in many works (see the books [20, 21] and references therein). In these works, the cost of a transport mapping or a transport plan is usually an integral of some function of the distance. However, in many real applications, the actual cost of the transport procedures is not necessarily determined by just knowing some optimal mapping from the starting position to the target position. For example in shipping two items from nearby cities to the same far away city, it may be less expensive to first bring them to a common location and put them on a single truck for most of the transport. In this case, a “Y shaped” path is preferable to a “V shaped” path. In both cases, the transport mapping from the given two sources to the single target is trivially the same, but the actual transport path naturally gives the total cost. In general, due to economy of scale, optimal transportation along ramified paths are typically more cost efficient than a “linear” structure.

Keywords and phrases. Optimal transportation, transport path, branching network, directed graph, ramified transportation.

* *This work is supported by the NSF grant DMS-1109663.*

¹ University of California at Davis, Department of Mathematics, One Shields Ave, Davis, CA, 95616, USA.
qlxia@math.ucdavis.edu

The second motivation comes from geometric measure theory. One of the main motivations of geometric measure theory comes from the modeling of soap films. The famous Plateau’s problem says: find a surface of the least area that has the given curve as its boundary. The answer to this question depends on the exact definitions of “surface”, “area”, and “boundary curve” (see for instance, the work of J. Douglas, Federer and Fleming, De Giorgi, J. Taylor, and others.) In particular, Almgren introduced the concept of size minimizer: let the given curve be the union of two parallel circles oriented the same way. When the circles are far away, an area minimizer is given by the two parallel disks. When the circles become close enough, pinching the centers of the two disks together, forming a soap film consisting of a (smaller) disk and a catenoid type surface, further minimizes the size of the area. An analogous one dimensional example is shown in Figure 1. These kinds of examples again motivate us to study optimal branching structures.

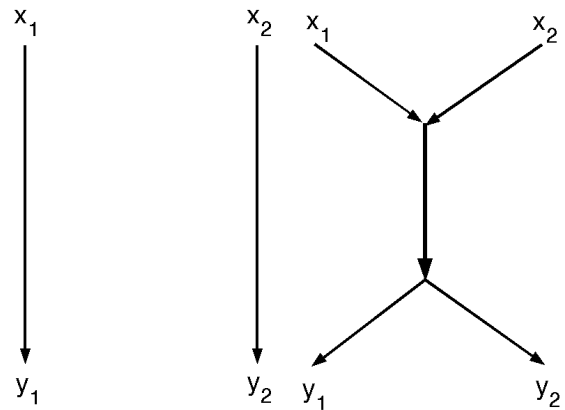


FIGURE 1.

Of course, the beauty of the branching structures found in nature also motivates us to explore the mathematics behind them. Many living systems such as trees, the veins on a leaf, as well as animal cardiovascular/circulatory systems exhibit branching structures, as do many non-living systems such as river channel networks, railways, airline networks, electric power supply and communication networks. Why do nature and engineers both prefer these ramifying structures? What are the advantages of these branching structures over non-branching structures?

Together these considerations motivate the author and many others to study the mathematics behind the ramified optimal transportation.

1.2. General set-up [24]

In general, we may consider: given two probability measures μ^+ and μ^- representing the source and the sink, find an optimal path to transport μ^+ to μ^- .

To solve this problem, one needs to find a suitable category of transport paths as well as a suitable cost functional acting on these paths. Particularly, such a category should be broad enough to give existence of an optimal transport path. Also, an optimal transport path should allow the possibility that some parts overlap in a cost efficient (maybe complicated) fashion but still enjoy some nice regularity properties. If possible, one may hope to visualize such an optimal transport path using numerical analysis and computer graphics.

We first recall some concepts about optimal transport paths between measures as studied in [24]. Let X be a convex compact subset in a Euclidean space \mathbb{R}^d . For any $x \in X$, let δ_x be the Dirac measure centered at x . An atomic measure in X is in the form of

$$\sum_{i=1}^k m_i \delta_{x_i}$$

with distinct points $x_i \in X$, and $m_i > 0$ for each $i = 1, \dots, k$. For any $\Lambda > 0$, let

$$\mathcal{A}_\Lambda(X) \tag{1.1}$$

be the space of all atomic measures on X with total mass Λ .

Definition 1.1. For any $A > 0$, and any two atomic measures

$$\mathbf{a} = \sum_{i=1}^k m_i \delta_{x_i} \text{ and } \mathbf{b} = \sum_{j=1}^{\ell} n_j \delta_{y_j} \tag{1.2}$$

in $\mathcal{A}_A(X)$, a *transport path* from \mathbf{a} to \mathbf{b} is a weighted directed graph G consisting of a vertex set $V(G) \subseteq X$, a directed edge set $E(G)$ and a weight function

$$w : E(G) \rightarrow (0, +\infty)$$

such that $\{x_1, x_2, \dots, x_k\} \cup \{y_1, y_2, \dots, y_l\} \subset V(G)$ and for any vertex $v \in V(G)$,

$$\sum_{\substack{e \in E(G) \\ e^- = v}} w(e) = \sum_{\substack{e \in E(G) \\ e^+ = v}} w(e) + \sum_{\substack{\text{if } v = x_i \text{ for} \\ \text{some } i = 1, \dots, k}} m_i - \sum_{\substack{\text{if } v = y_j \text{ for} \\ \text{some } j = 1, \dots, \ell}} n_j. \tag{1.3}$$

where e^- and e^+ denotes the starting and ending endpoints of each directed edge $e \in E(G)$. In other words, G satisfies the Kirchoff's law at each of its interior vertices.

Note that the balance equation (1.3) simply means the conservation of mass at each vertex. In terms of polyhedral chains, we simply have $\partial G = \mathbf{b} - \mathbf{a}$.

Let

$$\text{Path}(\mathbf{a}, \mathbf{b})$$

be the space of all transport paths from \mathbf{a} to \mathbf{b} .

Remark 1.2. Each transport path $G \in \text{Path}(\mathbf{a}, \mathbf{b})$ determines the vector-valued measure on X

$$G = \sum_{e \in E(G)} w(e) [[e]],$$

where $[[e]]$ is the vector-valued measure $\mathcal{H}^1|_e \vec{e}$ for each edge $e \in E(G)$ with unit directional vector \vec{e} . The above conditions can be simplified to be a single divergence condition on G

$$\text{div}(G) = \mathbf{a} - \mathbf{b},$$

in the sense of distribution.

Among all paths in $\text{Path}(\mathbf{a}, \mathbf{b})$, we want to find an optimal path which allows the possibility that some parts overlap in a cost efficient fashion. To get such a ‘‘Y-shaped’’ optimal path, we define the following cost function on transport paths.

Definition 1.3. For any $\alpha \in [0, 1]$, the M_α cost of a transport path

$$G = \{V(G), E(G), w : E(G) \rightarrow (0, +\infty)\} \in \text{Path}(\mathbf{a}, \mathbf{b})$$

is defined by

$$M_\alpha(G) := \sum_{e \in E(G)} w(e)^\alpha \text{length}(e), \tag{1.4}$$

where $\text{length}(e)$ denotes the Euclidean distance between endpoints e^- and e^+ of e . Given $\mathbf{a}, \mathbf{b} \in \mathcal{A}_A(X)$, an M_α minimizer in $\text{Path}(\mathbf{a}, \mathbf{b})$ is called an α -optimal transport path from \mathbf{a} to \mathbf{b} .

Problem 1.4 (Ramied optimal transport problem: Discrete version). Given two atomic measures $\mathbf{a}, \mathbf{b} \in \mathcal{A}_A(X)$ on X of equal mass and $\alpha \in [0, 1)$, minimize $M_\alpha(G)$ among all $G \in \text{Path}(\mathbf{a}, \mathbf{b})$.

2. PROPERTIES AND TECHNIQUES FOR OPTIMAL TRANSPORT PATH: DISCRETE VERSION

2.1. Basic properties of optimal transport paths [24]

An arbitrary transport path $G \in \text{Path}(\mathbf{a}, \mathbf{b})$ is a weighted directed graph, but not necessarily a directed tree. In other words, G may contain some cycles. Here, a weighted directed graph $G = \{V(G), E(G), w : E(G) \rightarrow (0, 1]\}$ contains a *cycle* if for some $k \geq 3$, there exists a list of distinct vertices $\{v_1, v_2, \dots, v_k\}$ in $V(G)$ such that for each $i = 1, \dots, k$, either the segment $[v_i, v_{i+1}]$ or $[v_{i+1}, v_i]$ is a directed edge in $E(G)$, with the agreement that $v_{k+1} = v_1$. However, the following proposition says G can be modified to be an acyclic graph $\tilde{G} \in \text{Path}(a, b)$ with less \mathbf{M}_α cost.

Proposition 2.1 ([24], Prop. 2.1). *Given \mathbf{a}, \mathbf{b} as in (1.2). For any transport path $G \in \text{Path}(\mathbf{a}, \mathbf{b})$, there exists a transport path $\tilde{G} \in \text{Path}(\mathbf{a}, \mathbf{b})$ such that $V(\tilde{G}) \subseteq V(G)$, $\mathbf{M}_\alpha(\tilde{G}) \leq \mathbf{M}_\alpha(G)$ and \tilde{G} contains no cycles.*

Let

$$\text{Path}_0(\mathbf{a}, \mathbf{b}) = \{G \in \text{Path}(\mathbf{a}, \mathbf{b}) : G \text{ contains no cycles}\}.$$

From the above proposition, we may restrict our transport paths to be acyclic. The following proposition says that the number of all branching vertices of $G \in \text{Path}_0(a, b)$ is bounded above by $k + \ell - 2$.

Proposition 2.2. *Suppose $G \in \text{Path}_0(\mathbf{a}, \mathbf{b})$. Then,*

$$|\{v : \text{deg}(v) \geq 3\}| \leq k + \ell - 2.$$

Proof. By calculating the total number of vertices in $V(G)$, we have

$$2|E(G)| = \sum_{v \in V(G)} \text{deg}(v).$$

By means of the Euler number $\chi_G = |V(G)| - |E(G)|$ of G , we have

$$2 \sum_{v \in V(G)} 1 - 2\chi_G = 2|V(G)| - 2\chi_G = 2|E(G)| = \sum_{v \in V(G)} \text{deg}(v).$$

That is,

$$\sum_{v \in V(G), \text{deg}(v)=1} 1 - 2\chi_G = \sum_{v \in V(G), \text{deg}(v) \geq 3} (\text{deg}(v) - 2).$$

Hence,

$$|\{v : \text{deg}(v) \geq 3\}| \leq \sum_{\text{deg}(v) \geq 3} (\text{deg}(v) - 2) = \sum_{\text{deg}(v)=1} 1 - 2\chi_G \leq k + \ell - 2. \quad \square$$

As a result, given k and ℓ , there are finitely many possibilities of topologies of all transport paths G in $\text{Path}_0(\mathbf{a}, \mathbf{b})$. For instance, when $k = 2, \ell = 1$, there are two possibilities: a “V-shape” or a “Y-shape”. It is clear that there exists at least one minimizer in each given topology. As a result, there always exists an M_α -minimizer in $\text{Path}(\mathbf{a}, \mathbf{b})$. That is, there always exists an α -optimal transport path in $\text{Path}_0(\mathbf{a}, \mathbf{b})$ for any $\alpha \in [0, 1)$.

For each $G \in \text{Path}_0(\mathbf{a}, \mathbf{b})$, we have the following trivial but important lemma.

Lemma 2.3 ([24], Lem. 2.1). *Suppose $G \in \text{Path}_0(\mathbf{a}, \mathbf{b})$ with $\mathbf{a}, \mathbf{b} \in \mathcal{A}_\Lambda$ as in (1.1). Then for any edge $e \in E(G)$, we have $0 < w(e) \leq \Lambda$. Moreover, $\frac{\mathbf{M}_\alpha(G)}{\Lambda^\alpha} \geq \frac{\mathbf{M}_1(G)}{\Lambda}$.*

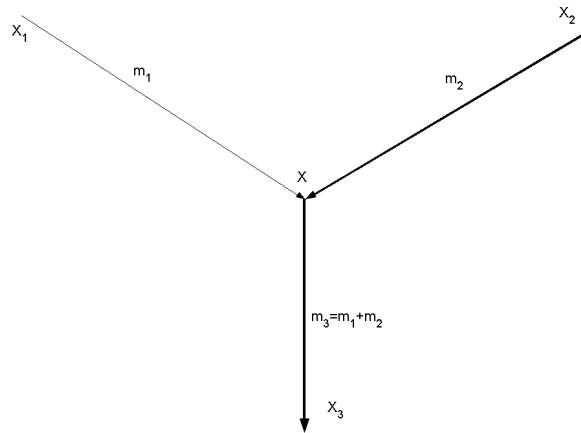
For any atomic measures \mathbf{a} and \mathbf{b} on X of equal mass, define the minimum transportation cost as

$$d_\alpha(\mathbf{a}, \mathbf{b}) := \min \{ \mathbf{M}_\alpha(G) : G \in \text{Path}(\mathbf{a}, \mathbf{b}) \}. \tag{2.1}$$

As shown in Xia [24], d_α is indeed a metric on the space of atomic measures of equal mass. Also, for each $\lambda > 0$, it holds that

$$d_\alpha(\lambda\mathbf{a}, \lambda\mathbf{b}) = \lambda^\alpha d_\alpha(\mathbf{a}, \mathbf{b}).$$

Example: Let $\mathbf{a} = m_1\delta_{x_1} + m_2\delta_{x_2}$ and $\mathbf{b} = m_3\delta_{x_3}$ with $m_3 = m_1 + m_2$. Then, in the non-degenerate case, the optimal transport path from \mathbf{a} to \mathbf{b} under the \mathbf{M}_α cost looks like the “Y shaped” graph.



Here the interior vertex x is determined by a balance formula:

$$m_1^\alpha \vec{n}_1 + m_2^\alpha \vec{n}_2 + m_3^\alpha \vec{n}_3 = \vec{0}, \tag{2.2}$$

where $\vec{n}_i = \frac{x_i - x}{|x_i - x|}$ is the unit vector from x to x_i , $i = 1, 2, 3$. Let θ_i be the angle between \vec{n}_i and $-\vec{n}_3$ for $i = 1, 2$ and $k_1 = \frac{m_1}{m_1 + m_2}$, $k_2 = \frac{m_2}{m_1 + m_2} = 1 - k_1$. Then, it is easy to find that the angles satisfy

$$\cos \theta_1 = \frac{k_1^{2\alpha} + 1 - k_2^{2\alpha}}{2k_1^\alpha} \text{ and } \cos \theta_2 = \frac{k_2^{2\alpha} + 1 - k_1^{2\alpha}}{2k_2^\alpha}.$$

Moreover,

$$\cos(\theta_1 + \theta_2) = \frac{1 - k_1^{2\alpha} - k_2^{2\alpha}}{2k_1^\alpha k_2^\alpha}.$$

In particular, if $m_1 = m_2$, then

$$\theta_1 + \theta_2 = \arccos(2^{2\alpha-1} - 1).$$

Note that for any $m_1 > 0$ and $m_2 > 0$,

- if $\alpha = 0$, then $\theta_1 + \theta_2 = 120^\circ$;
- if $\alpha = 1/2$, then $\theta_1 + \theta_2 = 90^\circ$;
- if $\alpha = 2/3$, then $\theta_1 + \theta_2$ is nearly 75° .

In the more general case, we have the following results for angles between edges:

Proposition 2.4 ([29], Prop. 2.1). *Let $G \in \text{Path}_0(\mathbf{a}, \mathbf{b})$ be any α -optimal transport path with $\alpha \in [0, 1)$. Let $v \in V(G)$ be an interior vertex of G , i.e. not one of the boundary vertices $\{x_1, \dots, x_k, y_1, \dots, y_\ell\}$. Let $\{e_i\}_{i=1}^{\text{deg}(v)}$ be the edges in G with v being one of its endpoints. Let $w(e_i)$ be the corresponding weight on e_i , and \vec{e}_i be the*

unit directional vector of the edge e_i from v to the other endpoint of e_i . Then,

(1) There exists a balance equation at v :

$$\sum_{i=1}^{\deg(v)} [w(e_i)]^\alpha \vec{e}_i = \vec{0}. \tag{2.3}$$

(2) The minimum angle between any two edges in $\{e_i\}_{i=1}^{\deg(v)}$ is uniformly bounded below by

$$\theta_\alpha := \begin{cases} \frac{\pi}{2}, & \text{if } 0 < \alpha \leq \frac{1}{2}. \\ \arccos(2^{2\alpha-1} - 1), & \text{if } \frac{1}{2} < \alpha < 1 \text{ or } \alpha = 0. \end{cases} \tag{2.4}$$

(3) The degree $\deg(v)$ of v is bounded above by some constant

$$D(\alpha, m), \tag{2.5}$$

depending only on α and the dimension m of \mathbb{R}^m .

The following proposition provides an interesting formula of $\mathbf{M}_\alpha(G)$ of an optimal transport path G in terms of boundary values.

Proposition 2.5 ([36], Prop. 6.4). *For any $G \in \text{Path}(\mathbf{a}, \mathbf{b})$, it holds that*

$$\mathbf{M}_\alpha(G) = \sum_{v \in V(G)} \vec{\mathbf{m}}_\alpha(v) \cdot \mathbf{v}, \tag{2.6}$$

where \mathbf{v} is the corresponding position vector of v in \mathbb{R}^m and

$$\vec{\mathbf{m}}_\alpha(v) = \sum_{e^+=v} w(e)^\alpha \vec{e} - \sum_{e^-=v} w(e)^\alpha \vec{e}.$$

If G is an α -optimal transport path, then a reformulation of the balance equation (2.3) gives $\vec{\mathbf{m}}_\alpha(v) = 0$ for any vertex $v \in V(G) \setminus \{x_1, \dots, x_k, y_1, \dots, y_\ell\}$ and thus

$$\mathbf{M}_\alpha(G) = \sum_{v \in \{x_1, x_2, \dots, x_k, y_1, \dots, y_\ell\}} \vec{\mathbf{m}}_\alpha(v) \cdot \mathbf{v}. \tag{2.7}$$

Using (2.7) as well as the property that $\mathbf{M}_\alpha(G)$ is translational invariant, one gets a necessary condition for a transport path being optimal:

Corollary 2.6 ([36], Cor. 6.5). *Suppose G is an α -optimal transport path from \mathbf{a} to \mathbf{b} . Then it holds that*

$$\sum_{v \in \{x_1, x_2, \dots, x_k, y_1, \dots, y_\ell\}} \vec{\mathbf{m}}_\alpha(v) = 0.$$

2.2. Decomposition techniques [29]

In this subsection, we describe the decomposition of a signed measure as well as a transport path into a dominant part and a remainder part.

An infinite atomic measure on X is a signed measure μ on X of the form

$$\mu = \sum_{i=1}^{\infty} a_i \delta_{x_i}, \tag{2.8}$$

where $\{a_i\}$ is a sequence of real numbers with $\{|a_i|\}$ decreasing, $\sum_i |a_i| < +\infty$, and $\{x_i\}$ is a sequence of points in X .

Proposition 2.7 ([29], Cor. 3.3). *Suppose μ is an infinite atomic measure on X in the form of (2.8) with $\sum_{i=1}^{\infty} |a_i|^\alpha \leq C$ for some constant $C > 0$ and $\alpha \in [0, 1)$. For any $\epsilon > 0$, one can decompose μ as the sum of*

$$\mu = \mu_P + \mu_R$$

such that

$$\mu_P = \sum_{i=1}^N a_i \delta_{x_i} \text{ and the mass } \|\mu_R\| := \sum_{i=N+1}^{\infty} |a_i| \leq \epsilon$$

where N is the least integer satisfying $N \geq C^{\frac{1}{1-\alpha}} \epsilon^{\frac{-\alpha}{1-\alpha}}$.

Similarly, we have a decomposition result for transport paths.

Proposition 2.8 ([29], Prop. 3.4). *Let \mathbf{a} and \mathbf{b} be any two atomic measures in $\mathcal{A}_\Lambda(X)$ in the form of (1.2), and $\lambda > 0$. Suppose there exist natural numbers $N_1 \leq k$ and $N_2 \leq \ell$ such that*

$$\sum_{i=N_1+1}^k m_i + \sum_{j=N_2+1}^{\ell} n_j < \lambda.$$

Then, for each transport path $G \in \text{Path}_0(\mathbf{a}, \mathbf{b})$, there exist decompositions of \mathbf{a}, \mathbf{b} and G (see Fig. 2):

$$\mathbf{a} = \mathbf{a}_P + \mathbf{a}_R, \mathbf{b} = \mathbf{b}_P + \mathbf{b}_R, \text{ and } G = P + R \tag{2.9}$$

as atomic measures and polyhedral 1-chain such that

- (1) \mathbf{a}_P is an atomic measure supported on $\{x_1, \dots, x_{N_1}\}$, \mathbf{b}_P is another atomic measure supported on $\{y_1, \dots, y_{N_2}\}$ with the same mass as that of \mathbf{a}_P and $P \in \text{Path}_0(\mathbf{a}_P, \mathbf{b}_P)$ is a transport path.
- (2) \mathbf{a}_R is an atomic measure supported on $\{x_1, \dots, x_k\}$, \mathbf{b}_R is an atomic measure supported on $\{y_1, \dots, y_\ell\}$ and $R \in \text{Path}_0(\mathbf{a}_R, \mathbf{b}_R)$ is a transport path. Also, the mass $\|\mathbf{a}_R\| = \|\mathbf{b}_R\| < \lambda$.
- (3) Moreover, the λ -superlevel set G_λ of G is contained in the support of P , where G is viewed as a polyhedral 1-chain.

Sometimes, we call P in (2.9) the λ -dominant part of G .

If in addition, $G \in \text{Path}(\mathbf{a}, \mathbf{b})$ is an α -optimal transport path for some $0 \leq \alpha < 1$, then we can get more information about the dominant part P of G . From Proposition 2.8, the dominant part P of G is still a transport path containing no cycles. A vertex v of P is *removable* if there exists only one edge (*i.e.* the line segment induced from G) of P that flows into v and only one edge of P that flows out of v . From now on, we will only consider non-removable vertices of P with the agreement that any edge of P is a connected polyhedral curve between non-removable vertices of P , not necessarily a line segment. *e.g.* in Figure 2b, the path P has only three edges. Let $\{\Gamma_i\}_{i=1}^K$ be the set of all these “topological” edges of P for some $K \in \mathbb{N}$. Then, as polyhedral chains, P can be expressed as

$$P = \sum_{i=1}^K m_i \Gamma_i$$

for some positive numbers m_i . The following proposition says that each edge Γ_i of P is a bi-Lipschitz curve.

Proposition 2.9 ([29], Prop. 3.5). *Suppose $\mathbf{a}, \mathbf{b} \in \mathcal{A}_\Lambda(X)$ and $G \in \text{Path}_0(\mathbf{a}, \mathbf{b})$ is an α -optimal transport path for some $0 \leq \alpha < 1$. Let P be the dominant part of G as in (2.9) for some λ . For each edge Γ_i of P , let ϕ_i be the arc parametrization of Γ_i . Then ϕ_i is bi-Lipschitz with*

$$\text{Lip}(\phi_i^{-1}) \leq \frac{(m_i)^\alpha}{\Lambda^\alpha - (\Lambda - m_i)^\alpha}. \tag{2.10}$$

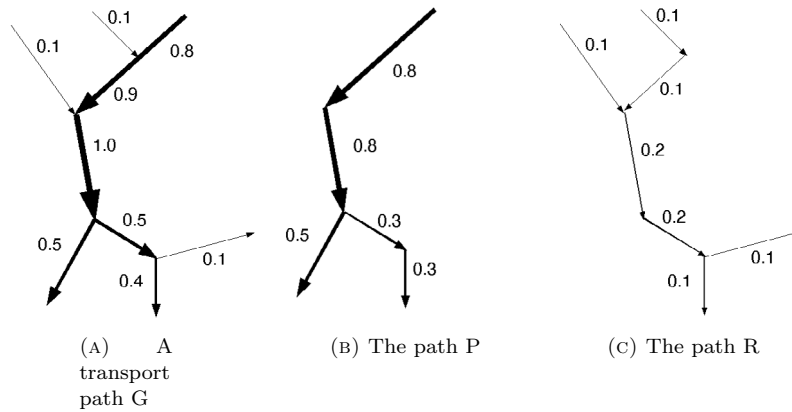


FIGURE 2. Decomposition of a transport path G as sums of a dominant part P and a remainder part R with $\lambda = 0.35$.

The proof of this proposition is based on the following idea: By the optimality of G , for any two points x, y on Γ_i , redirect the flow from x to y via the line segment $[x, y]$ will not be able to reduce the transportation cost. This idea leads to a comparison argument which eventually implies (2.10). We refer to ([29], Prop. 3.5) for more details.

2.3. Compatibility between transport plan and Path

Suppose \mathbf{a} and \mathbf{b} are two atomic measures on X as in (1.2). Recall that a *transport plan* from \mathbf{a} to \mathbf{b} is an atomic measure

$$q = \sum_{i=1}^k \sum_{j=1}^{\ell} q_{ij} \delta_{(x_i, y_j)} \tag{2.11}$$

on the product space $X \times X$ such that for each i and j , $q_{ij} \geq 0$,

$$\sum_{i=1}^k q_{ij} = n_j \text{ and } \sum_{j=1}^{\ell} q_{ij} = m_i. \tag{2.12}$$

Denote $\text{Plan}(\mathbf{a}, \mathbf{b})$ as the space of all transport plans from \mathbf{a} to \mathbf{b} .

Now, as in Section 7.1 of Xia [24], we want to consider the compatibility between a transport path and a transport plan. Let G be a given transport path in $\text{Path}(\mathbf{a}, \mathbf{b})$. We assume that for each x_i and y_j ,

$$\text{there exists at most one directed polyhedral curve } g_{ij} \text{ from } x_i \text{ to } y_j. \tag{2.13}$$

In other words, there exists a list of distinct vertices

$$V(g_{ij}) := \{v_{i_1}, v_{i_2}, \dots, v_{i_h}\} \tag{2.14}$$

in $V(G)$ with $x_i = v_{i_1}$, $y_j = v_{i_h}$, and each $[v_{i_t}, v_{i_{t+1}}]$ is a directed edge in $E(G)$ for each $t = 1, 2, \dots, h - 1$. The Assumption (2.13) clearly holds when G contains no cycles.

For some pairs of (i, j) , such a curve g_{ij} from x_i to y_j may fail to exist, due to reasons like geographical barriers, law restrictions, etc. If such curve does not exist, we set $g_{ij} = 0$ to denote the *empty* directed polyhedral curve. By doing so, we construct a matrix

$$g = (g_{ij})_{k \times \ell} \tag{2.15}$$

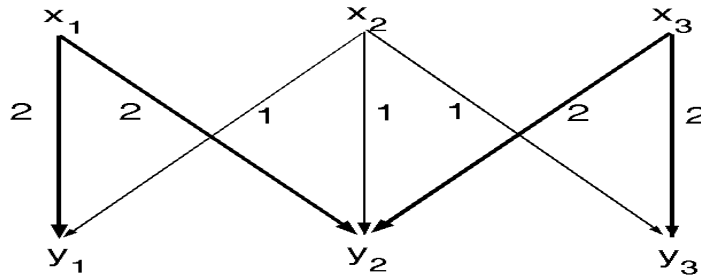


FIGURE 3. A transport path from $4\delta_{x_1} + 3\delta_{x_2} + 4\delta_{x_3}$ to $3\delta_{y_1} + 5\delta_{y_2} + 3\delta_{y_3}$ with $g_{13} = 0, g_{31} = 0$.

with each element of g being a polyhedral curve. A very simple example satisfying these conditions is illustrated in Figure 3. For any transport path $G \in \text{Path}(\mathbf{a}, \mathbf{b})$ satisfying (2.13), such a matrix $g = (g_{ij})$ is uniquely determined.

Definition 2.10. Let $G \in \text{Path}(\mathbf{a}, \mathbf{b})$ be a transport path satisfying (2.13) and $q \in \text{Plan}(\mathbf{a}, \mathbf{b})$ be a transport plan. The pair (G, q) is compatible if $q_{ij} = 0$ whenever $g_{ij} = 0$ and

$$G = q \cdot g. \tag{2.16}$$

Here, equation (2.16) means that as polyhedral chains,

$$G = \sum_{i=1}^k \sum_{j=1}^{\ell} q_{ij} \cdot g_{ij},$$

where the product $q_{ij} \cdot g_{ij}$ denotes that an amount q_{ij} of commodity is moved along the polyhedral curve g_{ij} from factory i to household j . In terms of edges, it says that for each edge $e \in E(G)$, we have

$$\sum_{e \subseteq g_{ij}} q_{ij} = w(e).$$

For instance, the transport path in Figure 3 can be expressed as

$$G = 2g_{11} + 2g_{12} + g_{21} + g_{22} + g_{23} + 2g_{32} + 2g_{33}, \tag{2.17}$$

which means that the transport plan

$$q = 2\delta_{(1,1)} + 2\delta_{(1,2)} + \delta_{(2,1)} + \delta_{(2,2)} + \delta_{(2,3)} + 2\delta_{(3,2)} + 2\delta_{(3,3)}$$

is compatible with G in (2.17).

Roughly speaking, the compatibility conditions check whether a transport plan is realizable by a transport path. Given a transport plan, the planner must design a transport path which can support this plan. To see the concept more precisely, let

$$\mathbf{a} = \frac{1}{4}\delta_{x_1} + \frac{3}{4}\delta_{x_2} \text{ and } \mathbf{b} = \frac{5}{8}\delta_{y_1} + \frac{3}{8}\delta_{y_2},$$

and consider a transport plan

$$q = \frac{1}{8}\delta_{(x_1,y_1)} + \frac{1}{8}\delta_{(x_1,y_2)} + \frac{1}{2}\delta_{(x_2,y_1)} + \frac{1}{4}\delta_{(x_2,y_2)} \in \text{Plan}(\mathbf{a}, \mathbf{b}). \tag{2.18}$$

It is straight forward to see from Figure 4 that q is compatible with G_1 but not G_2 . This is because there is no directed curve g_{12} from factory 1 to household 2 in G_2 .

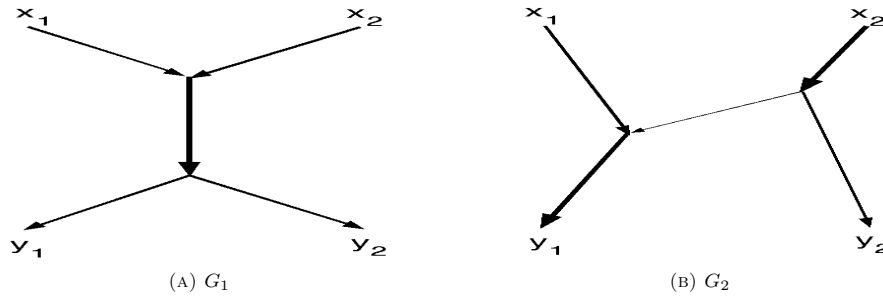


FIGURE 4. Compatibility between transport plan and transport path.

3. RAMIFIED OPTIMAL TRANSPORTATION BETWEEN RADON MEASURES

3.1. Set-up and existence result [24]

Let $\mathcal{M}_\Lambda(X)$ be the space of Radon measures μ on X with total mass $\mu(X) = \Lambda$. When $\Lambda = 1$, we sometimes also denote $\mathcal{M}_1(X)$ by $\mathcal{P}(X)$, the space of all probability measures on X .

Recall that in Remark 1.2, each transport path between atomic measures can be viewed as a vector-valued measure on X . This enables us to consider transport paths between two Radon measures as follows: Let $\mu^+, \mu^- \in \mathcal{M}_\Lambda(X)$ be any two Radon measures on X with equal total mass Λ . We say a vector-valued measure $T \in \mathcal{M}^m(X)$ is a transport path from μ^+ to μ^- if there exist two sequences $\{\mathbf{a}_i\}, \{\mathbf{b}_i\}$ of atomic measures in $\mathcal{A}_\Lambda(X)$ with a corresponding sequence of transport paths $G_i \in \text{Path}(\mathbf{a}_i, \mathbf{b}_i)$ such that

$$\mathbf{a}_i \rightharpoonup \mu^+, \quad \mathbf{b}_i \rightharpoonup \mu^-, \quad G_i \rightharpoonup T$$

weakly as Radon measures and vector-valued measures.

The sequence of triples $\{\mathbf{a}_i, \mathbf{b}_i, G_i\}$ is called an approximating graph sequence for T . Note that for any such T , $\text{div}(T) = \mu^+ - \mu^-$ in the sense of distributions. Let

$$\text{Path}(\mu^+, \mu^-) \subset \mathcal{M}^m(X)$$

be the space of all transport paths from μ^+ to μ^- . Also, given any $\alpha \in [0, 1]$, for any $T \in \text{Path}(\mu^+, \mu^-)$, we define its \mathbf{M}_α cost to be

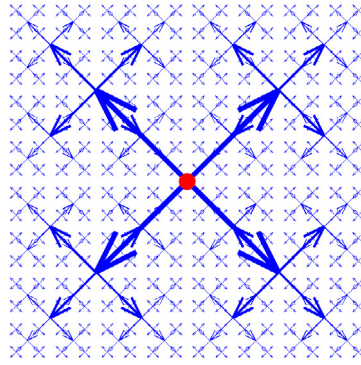
$$\mathbf{M}_\alpha(T) := \inf\{\liminf_{i \rightarrow \infty} \mathbf{M}_\alpha(G_i)\},$$

where the infimum is over the set of all possible approximating graph sequence $\{a_i, b_i, G_i\}$ of T .

Problem 3.1 (Ramified optimal transport problem: Continuous version). Given two Radon measures $\mu^+, \mu^- \in \mathcal{M}_\Lambda(X)$ on $X \subset \mathbb{R}^m$ and $\alpha \in [0, 1)$, minimize $\mathbf{M}_\alpha(T)$ among all $T \in \text{Path}(\mu^+, \mu^-)$.

Suppose $\mu \in \mathcal{M}_1(X)$ is supported in a cube $Q \subset \mathbb{R}^m$ whose center is denoted by c_Q . By using a sequence of dyadic approximations of μ , one can construct a transport path T from μ to δ_{c_Q} such that for any $\alpha \in (1 - \frac{1}{m}, 1]$,

$$\mathbf{M}_\alpha(T) \leq \frac{\text{diam}(Q)}{2^{1-m(1-\alpha)} - 1} \frac{\sqrt{m}}{2}.$$



Using this result and the direct method of calculus of variations, we have the following existence result proven in ([24], Thm. 3.1) for Problem 3.1:

Theorem 3.2 (Existence theorem). *Given two Radon measures $\mu^+, \mu^- \in \mathcal{M}_\Lambda(X)$ on $X \subset \mathbb{R}^m$ and $\alpha \in (1 - \frac{1}{m}, 1]$, there exists an optimal transport path S with least \mathbf{M}_α cost among all transport paths in the family $\text{Path}(\mu^+, \mu^-)$. Moreover*

$$\mathbf{M}_\alpha(S) \leq \frac{\Lambda^\alpha}{2^{1-m(1-\alpha)} - 1} \frac{\sqrt{md}}{2},$$

where d is the diameter of the convex hull of supports of μ^+ and μ^- .

Now, for any $\alpha \in (1 - \frac{1}{m}, 1]$, we define

$$d_\alpha(\mu^+, \mu^-) := \min \{ \mathbf{M}_\alpha(T) : T \in \text{Path}(\mu^+, \mu^-) \},$$

for any two Radon measures $\mu^+, \mu^- \in \mathcal{M}_\Lambda(X)$. Note that for any $\Lambda > 0$ and any $\mu^+, \mu^- \in \mathcal{M}_\Lambda(X)$,

$$d_\alpha(\mu^+, \mu^-) = \Lambda^\alpha d_\alpha\left(\frac{\mu^+}{\Lambda}, \frac{\mu^-}{\Lambda}\right).$$

Thus, we may assume $\Lambda = 1$. Some key results about d_α is stated in the following theorem:

Theorem 3.3 ([24], Thms. 4.1, 4.2, 5.1). *d_α is a metric on $\mathcal{M}_1(X)$ and metrizes the weak * topology of $\mathcal{M}_1(X)$. Moreover, the space $(\mathcal{M}_1(X), d_\alpha)$ is a length space in the sense that for any $\mu^+, \mu^- \in \mathcal{M}_1(X)$, each α -optimal transport path T corresponds to a continuous map*

$$\psi : [0, d_\alpha(\mu^+, \mu^-)] \rightarrow \mathcal{M}_1(X)$$

such that $\psi(0) = \mu^+, \psi(d_\alpha(\mu^+, \mu^-)) = \mu^-$ and for any $0 \leq s_1 < s_2 \leq d_\alpha(\mu^+, \mu^-)$,

$$d_\alpha(\psi(s_1), \psi(s_2)) = s_2 - s_1.$$

Suppose $\{\mathbf{a}_i, \mathbf{b}_i, G_i\}$ is any approximating graph sequence of some transport path $T \in \text{Path}(\mu^+, \mu^-)$. One can show that if G_i is optimal in $\text{Path}(\mathbf{a}_i, \mathbf{b}_i)$ for each i , then T is also optimal. Based on this result, to get an optimal transport path $T \in \text{Path}(\mu^+, \mu^-)$, one has the freedom to choose his favorite approximating measures $\{\mathbf{a}_i, \mathbf{b}_i\}$ of (μ^+, μ^-) . An obvious approximation of a Radon measure μ is given by its dyadic approximation $\{A_n(\mu)\}$. Direct calculations yield

$$d_\alpha(\mu, A_n(\mu)) \leq C\beta^n$$

for some constant $C > 0$ and $0 < \beta = 2^{m(1-\alpha)-1} < 1$. This result implies that atomic probability measures are dense in $(\mathcal{M}_1(X), d_\alpha)$.

3.2. Regularity theorems [25, 29]

3.2.1. Some terminology of geometric measure theory

We first recall some terminology about rectifiable 1-currents as in [8] or [18].

Let $\Omega \subseteq \mathbb{R}^m$ be an open subset and $\mathcal{D}^1(\Omega)$ be the set of all C^∞ differential 1-forms in Ω with compact support with the usual Fréchet topology [8]. An 1-dimensional current S in Ω is a continuous linear functional on $\mathcal{D}^1(\Omega)$. Let $\mathcal{D}_1(\Omega)$ denote the set of all 1-dimensional currents in Ω . Motivated by Stokes’s theorem, the boundary of a current $S \in \mathcal{D}_1(\Omega)$ is the distribution defined by $\partial S(\psi) := S(d\psi)$ for any C^∞ test function ψ in Ω with compact support. A sequence of currents $S_i \in \mathcal{D}_1(\Omega)$ is said to be weakly convergent to another current $S \in \mathcal{D}_1(\Omega)$, denoted by $S_i \rightharpoonup S$, if $S_i(\psi) \rightarrow S(\psi)$ for any $\psi \in \mathcal{D}^1(\Omega)$.

As in [18], a subset $M \subseteq \mathbb{R}^m$ is called (countably) 1-rectifiable if $M = \bigcup_{i=0}^\infty M_i$, where $\mathcal{H}^1(M_0) = 0$ under the 1-dimensional Hausdorff measure \mathcal{H}^1 and each M_i , for $i = 1, 2, \dots$, is a subset of an 1-dimensional C^1 submanifold in \mathbb{R}^m . A rectifiable current S is a current coming from an oriented rectifiable set with multiplicities. More precisely, $S \in \mathcal{D}_1(\Omega)$ is a rectifiable current if it can be expressed as

$$S(\omega) = \int_M \langle \omega(x), \xi(x) \rangle \theta(x) \, d\mathcal{H}^1(x), \quad \forall \omega \in \mathcal{D}^1(\Omega)$$

where

- M is a \mathcal{H}^1 measurable and 1-rectifiable subset of Ω ;
- θ is a $\mathcal{H}^1 \llcorner M$ integrable positive function, called the multiplicity function of S ;
- $\xi : M \rightarrow \Lambda_1(\mathbb{R}^m)$ is a \mathcal{H}^1 measurable unit tangent vector field on M , called the orientation of S .

The rectifiable current S described as above is often denoted by

$$S = \underline{\underline{\tau}}(M, \theta, \xi).$$

For instance, each weighted directed graph $G \in \text{Path}(\mathbf{a}, \mathbf{b})$ determines a rectifiable 1-current $G = \underline{\underline{\tau}}(\mathbf{G}, w, \xi)$ in \mathbb{R}^m , with $\partial G = \mathbf{b} - \mathbf{a}$ as currents.

3.2.2. Regularity of optimal transport paths

For any Radon measures $\mu^+, \mu^- \in \mathcal{M}_\Lambda(X)$ of equal total mass and $\alpha \in [0, 1)$, one of the main results in [25] says that if a transport path $T \in \text{Path}(\mu^+, \mu^-)$ has finite \mathbf{M}_α cost, then T determines a rectifiable 1-current $T = \underline{\underline{\tau}}(M, \theta, \xi)$ with boundary $\partial T = \mu^+ - \mu^-$. Moreover,

$$\mathbf{M}_\alpha(T) = \int_M \theta^\alpha \, d\mathcal{H}^k(x) < +\infty. \tag{3.1}$$

Under the assumption that one of μ^+ and μ^- is atomic, another theorem says: Suppose $T \in \text{Path}(\mu^+, \mu^-)$ is an optimal transport path with finite \mathbf{M}_α cost. For any point $\xi \in \text{spt}(T) \setminus \text{spt}(\mu^+ \cup \mu^-)$, there is an open neighborhood B_ξ of ξ , such that $T \llcorner B_\xi$ is a cone that consists of finite many line segments with suitable multiplicities.

3.2.3. Boundary regularity [29]

In general, the support of an optimal transport path T may not necessarily be 1 dimensional nearby its boundary, for the support of the measure $\mu^+ - \mu^-$ may even contain an open subset of \mathbb{R}^m . For instance, one may take μ^+ to be some Lebesgue measure on a domain Ω and μ^- to be some atomic measure on Ω . Then, the support of $\mu^+ - \mu^-$ has the same dimension of Ω , which is not necessarily 1 dimensional. So, the question is how to describe the behavior of T when a carrying set of T is possibly dense in the whole space X .

To study the boundary regularity of an optimal transport path, the main idea of our approach is to study its superlevel sets. This idea is motivated from observing vein structure of a tree leaf provided by the nature. Here, we show that each superlevel set of an optimal transport path is locally concentrated on the support of an 1-dimensional bi-Lipschitz chain, which is analogous to vein structures of a tree leaf.

We first clarify some terminology. For any $\lambda > 0$, the λ -superlevel set of a rectifiable current $T = \tau_{\underline{=}}(M, \theta, \xi)$ is the set

$$T_\lambda = \{x \in M : \theta(x) \geq \lambda\}. \tag{3.2}$$

Also, a *bi-Lipschitz chain* P is a finite sum of bi-Lipschitz curves in X with real coefficient multiplicities. That is,

$$P = \sum_{i=1}^K m_i \Gamma_i \tag{3.3}$$

for some real numbers $m_i > 0$, and some bi-Lipschitz curves Γ_i with $i = 1, 2, \dots, K$ in X . The support $spt(P)$ of the bi-Lipschitz chain P is the union of the image of every bi-Lipschitz curve Γ_i in X .

Now, we state our boundary regularity theorem as follows:

Theorem 3.4 ([29], Thm. 4.1). *For any $\mu^+, \mu^- \in \mathcal{M}_\Lambda(X)$, let $T = \tau_{\underline{=}}(M, \theta, \xi) \in \text{Path}(\mu^+, \mu^-)$ be any α -optimal transport path with finite \mathbf{M}_α cost for some $0 \leq \alpha < 1$. Then, for any $\epsilon > 0$ and any $p \in M$, there exists an open ball neighborhood $B_r(p)$ about p and a decomposition*

$$T \llcorner B_r(p) = P + R$$

as 1-dimensional rectifiable currents such that

- (a) P is a bi-Lipschitz chain in the form of (3.3).
- (b) $R \in \text{Path}(\mu_R^+, \mu_R^-)$ is an 1-dimensional rectifiable current for some Radon measures μ_R^+ and μ_R^- with mass $||\mu_R^+|| = ||\mu_R^-|| < \epsilon$.
- (c) Moreover, inside the ball $B_r(p)$, the ϵ -superlevel set T_ϵ of T as defined in (3.2) is a subset of the support of the bi-Lipschitz chain P .

4. NUMERICAL SIMULATIONS [30]

Let X be a convex compact subset in \mathbb{R}^d . Given two atomic measures in the form of

$$\mathbf{a} = m\delta_O \text{ and } \mathbf{b} = \sum_{i=1}^N m_i \delta_{y_i} \text{ with } m = \sum_{i=1}^N m_i \tag{4.1}$$

in X of equal total mass, we are interested in seeing what an optimal transport path G from the single source \mathbf{a} to \mathbf{b} look like numerically.

4.1. Optimal transport paths for small values of N

The case of $N = 1$ is trivial. When $N = 2$, there exists an explicit formula in [30] for constructing the optimal transport path. Things become complicated when $N \geq 3$, but still doable when N is small:

- **Step 1:** For each fixed topology t , find an \mathbf{M}_α minimizer G_t within the given topology using a modified version of the Melzak algorithm [16].
- **Step 2:** Set $G_{\text{small}}(\mathbf{a}, \mathbf{b}) \in \arg \min_t \{\mathbf{M}_\alpha(G_t)\}$ among all topology t .

This approach only work when N is small because the number of all possible topologies t becomes extremely large when N becomes larger. To visualize optimal transport paths (*i.e.* solutions of the NP-hard problem), we propose some heuristics approach. The idea is to construct an initial transport path $G \in \text{Path}(\mathbf{a}, \mathbf{b})$ first and then modify G as much as possible until we can not reduce the cost of G any further.

4.2. The construction of an initial transport path

Let

$$\lambda = \begin{cases} 3, & \text{if } d = 2 \\ 2, & \text{if } d \geq 3 \end{cases}$$

and $K = \lambda^d$ where d is the dimension of the ambient space \mathbb{R}^d .

Algorithm (Subdivision method):

Input: Two atomic measures \mathbf{a}, \mathbf{b} in the form of (4.1) and a parameter $\alpha \leq 1$;

Output: A transport path $G_{sd}(\mathbf{a}, \mathbf{b}) \in \text{Path}(\mathbf{a}, \mathbf{b})$ with $\text{degree}(v) \leq K$ for each $v \in V(G_{sd}(\mathbf{a}, \mathbf{b}))$.

Step 1: If $N \leq K$, then set $G_{sd}(\mathbf{a}, \mathbf{b}) = G_{\text{small}}(\mathbf{a}, \mathbf{b})$.

Step 2: If $N > K$, then let Q be a cube in \mathbb{R}^d that contains the supports of both \mathbf{a} and \mathbf{b} . We evenly split the cube Q into totally $K = \lambda^d$ smaller cubes $\{Q_i\}_{i=1}^K$ whose centers are denoted by $c(Q_i)$. Then, recursively we set

$$G_{sd}(\mathbf{a}, \mathbf{b}) = G_{\text{small}}(\mathbf{a}, \sum_{i=1}^K \mathbf{b}(Q_i) \delta_{c(Q_i)}) + \sum_{i=1}^K G_{sd}(\mathbf{b}(Q_i) \delta_{c(Q_i)}, \mathbf{b}|_{Q_i}). \tag{4.2}$$

4.3. Modification of an existing transport path

Now, suppose G is an existing transport path from \mathbf{a} to \mathbf{b} that contains no cycles. We want to modify G to reduce the transport cost as much as possible. Before describing algorithms, we first introduce some concepts about vertices of an transport path G .

For any two vertices $v, u \in V(G)$, we say that v is an *ancestor* of u and u is a *descendant* of v , if there exists a list of vertices $v_1 = v, v_2, \dots, v_{h-1}, v_h = u$ such that each $[v_i, v_{i+1}]$ is a directed edge in $E(G)$ for $i = 1, \dots, h-1$. Also, if $[v, u]$ is a directed edge in $E(G)$, then we say that v is a *parent* of u and u is a *child* of v .

Let $p(O) = O$. For each vertex $u \in V(G) \setminus \{O\}$, u has exactly one parent $p(u) \in V(G)$ because G contains no cycles and has a single source $\mathbf{a} = m\delta_O$. Let $m(u)$ be the associated weight on the directed edge $[p(u), u]$ in $E(G)$ for each $u \in V(G) \setminus \{O\}$, and also set $m(O) = m$. Note that $m(v) \geq m(u)$ whenever v is an ancestor of u . Moreover, the vertex O is always an ancestor of each u . That is, there exists a list of vertices v_1, v_2, \dots, v_k in $V(G)$ such that $[v_i, v_{i+1}] \in E(G)$ with $v_1 = O$ and $v_k = u$. Then, for each $t \in [-m(u), m(u)]$, we consider the path

$$R(G; t, u) := G - \sum_{i=1}^{k-1} t [v_i, v_{i+1}] \in \text{Path}(\mathbf{a} - t\delta_O + t\delta_u, \mathbf{b}).$$

When $t > 0$, we say that a mass of t is *removed* from the path G at vertex u . When $t < 0$, we say that a mass of t is *added* to the path G at vertex u . Moreover, the potential function of G at a vertex $u \in V(G)$ is defined by

$$P_G(u, t) = \begin{cases} P_G(p(u), t) + |p(u) - u| [m(u)^\alpha - (m(u) - t)^\alpha], & u \neq O \\ 0, & u = O \end{cases} \tag{4.3}$$

for $t \in [-m(u), m(u)]$. Note that $P_G(u, t)$ has the same sign as t .

4.3.1. Local minimization

Now, we describe a local minimization method to modify any existing transport path G containing no cycles.

Input: a transport path $G \in \text{Path}(\mathbf{a}, \mathbf{b})$ containing no cycles and $\alpha < 1$;

Output: a locally optimized path $\tilde{G} \in \text{Path}(\mathbf{a}, \mathbf{b})$ with $M_\alpha(\tilde{G}) \leq M_\alpha(G)$.

For each vertex u of G , let

$$\mathbf{a}_u = m(u) \delta_{p(u)} \text{ and } \mathbf{b}_u = \sum_{h \in V(G), p(h)=u} m(h) \delta_h$$

be two atomic measures corresponding to the parent and the children of u . Then,

$$G_{\text{old}}(u) = \sum_{h \in V(G), p(h)=u} m(h)[u, h] + m(u)[p(u), u] \in \text{Path}(\mathbf{a}_u, \mathbf{b}_u)$$

is the union of all weighted edges in G sharing u as their common endpoint. On the other hand, one may generate another path $G_{\text{small}}(\mathbf{a}_u, \mathbf{b}_u)$.

If $M_\alpha(G_{\text{old}}(u)) > M_\alpha(G_{\text{small}}(\mathbf{a}_u, \mathbf{b}_u))$, then replace G by a new path

$$\tilde{G} = G - G_{\text{old}}(u) + G_{\text{small}}(\mathbf{a}_u, \mathbf{b}_u) \in \text{Path}(\mathbf{a}, \mathbf{b}).$$

Continue this process for all vertices of G , and repeat until one can not reduce the cost any further. The final path is denoted by $G_{\text{locmin}}(G)$.

The main drawback of this algorithm is that the result is only local minimization rather than global minimization. For instance, edges may intersect with each other. Sometimes, using eyes of a human being, one can easily observe a better transport path. To overcome these drawbacks, we adopt the following algorithm.

4.3.2. Global minimization

Now, we introduce the following algorithm of global minimization:

Input: two probability measures \mathbf{a}, \mathbf{b} in the form of (4.1) and a parameter $\alpha \leq 1$;

Output: an approximately α -optimal transport path $G \in \text{Path}(\mathbf{a}, \mathbf{b})$.

Step 1: set the initial path $G = G_{sd}(\mathbf{a}, \mathbf{b})$ as in (4.2).

Step 2: modify the existing path G using the local minimization method.

Step 3: subdivide long edges of G into shorter edges.

Step 4: for each vertex u of G , remove a mass of $m(u)$ at vertex u from the path G ; change the parent $p(u)$ of u to a better one if possible and then add back a mass of $m(u)$ at vertex u . More precisely.

Substep 1: a list of potential parents of u is defined as

$$L(u) = \{v \in V(G) : |v - u| \leq \sigma, \text{ and } v \text{ is not a descendant of } u\},$$

where $\sigma = \frac{P_G(u, m(u))}{[m(u)]^\alpha}$ and P_G is defined in (4.3).

Substep 2: by removing a mass of $m(u)$ at vertex u from the path G , we get another path $\tilde{G} = R(G; m(u), u)$.

Substep 3: for each $v \in L(u) \setminus \{p(u)\}$, let $c(v) = -P_{\tilde{G}}(v, -m(u))$,

which measures the extra cost of transporting a mass of $m(u)$ on the system \tilde{G} from the source O to the vertex u via the vertex v .

Substep 4: find the maximum of $c(v)$ over all $v \in L(u) \setminus \{p(u)\}$. If $\max c(v) > \sigma [m(u)]^\alpha$, then we find a better parent for the vertex u . In this case, suppose the maximum of $c(v)$ is achieved at v^* . Then, let $G^* = R(\tilde{G}; -m(u), v^*) + m(u)[v^*, u]$. That is, we change the parent of u from $p(u)$ to v^* and then add a mass $m(u)$ at the vertex u to the modified path. For convenience, we still denote the final modified transport path G^* by G .

Step 5: Repeat steps 2–4 until one can not reduce the cost any further.

An example of our algorithms is illustrated in Figure 5.

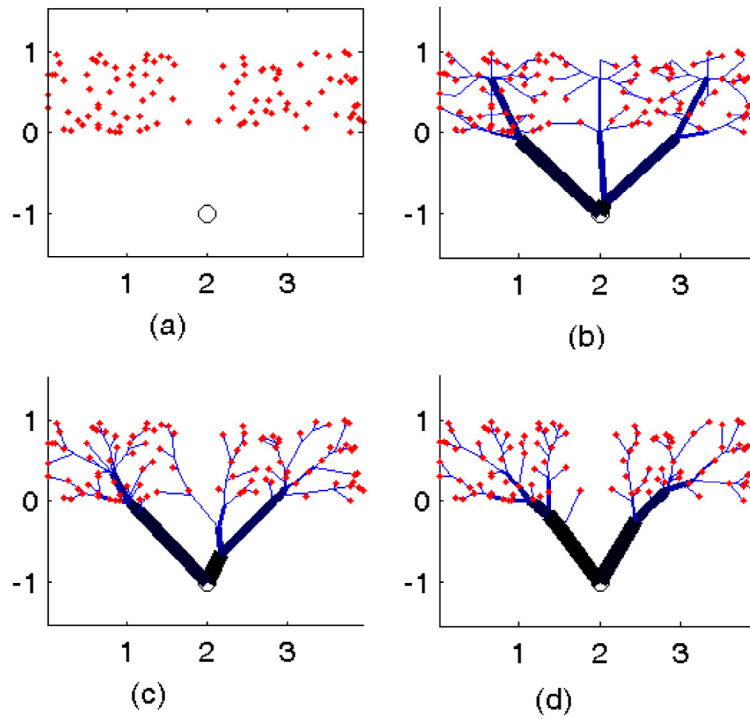


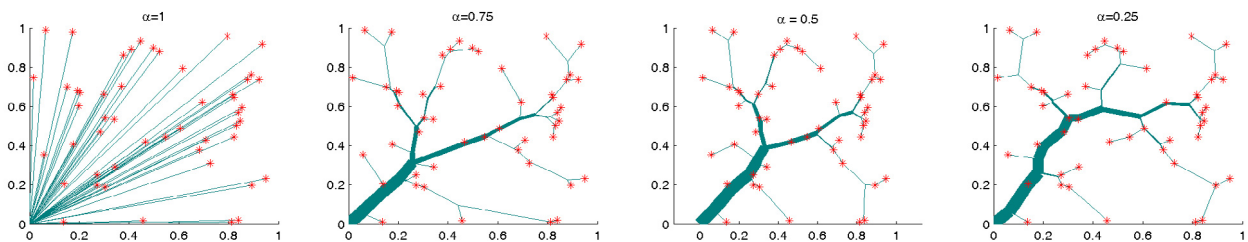
FIGURE 5. (a) A single source $\mathbf{a} = \delta_{(2,-1)}$ and the targeting measure \mathbf{b} represented by 100 random points of equal weight. (b) An initial transport path constructed by the subdivision method. (c) A modified transport path achieved by repeated modifying the initial path in (b) using the local minimization. (d) An approximately optimal transport path achieved by modifying the path in (c) using the global minimization method.

Another example is given below to indicate changing behaviors of α -optimal transport paths as the parameter α changes:

Example 4.1. Let $\{y_i\}$ be 50 random points in the square $[0, 1] \times [0, 1]$. Then, $\{y_i\}$ determines an atomic probability measure

$$\mathbf{b} = \sum_{i=1}^{50} \frac{1}{50} \delta_{y_i}.$$

Let $\mathbf{a} = \delta_O$ where $O = (0, 0)$ is the origin. Then an optimal transport path from \mathbf{a} to \mathbf{b} looks like the following figures with $\alpha = 1, 0.75, 0.5$ and 0.25 respectively:



5. AN APPLICATION IN MEASURE THEORY: TRANSPORT DIMENSION OF MEASURES AND DIMENSIONAL DISTANCE

Given a probability measure μ on \mathbb{R}^m , when can it be transported to a Dirac mass δ_O with finite d_α cost? The answer to this question depends on the dimensional information of the measure. In [6], Devillanova and Solimini introduced the concept of irrigability dimension of measures. For any given probability measure μ on \mathbb{R}^m , the irrigability dimension of μ is defined by

$$\dim_I(\mu) := \inf_{0 \leq \alpha < 1} \left\{ \frac{1}{1 - \alpha} : \text{if } \mu \text{ is } \alpha\text{-irrigable, i.e. } d_\alpha(\mu, \delta_O) < +\infty \right\}.$$

The main theorem in [6] shows that

$$\max \{ \dim_H(\mu), 1 \} \leq \dim_I(\mu) \leq \max \{ 1, \dim_M(\mu) \} \tag{5.1}$$

where $\dim_H(\mu)$ (or $\dim_M(\mu)$) denotes the infimum of the Hausdorff dimension (or the Minkowski dimension, respectively) of sets that μ is concentrated on. By definition, the irrigability dimension $\dim_I(\mu)$ of a measure μ must be larger or equal to 1 as the parameter α is in the range of $[0, 1)$.

In [32], we removed the maximum constraint from (5.1) by using a different approach of ramified optimal transportation (i.e. using optimal transport paths), and also allowing the parameter α to be negative. This generalization allows us to consider measures which have fractal dimensions (e.g. the Cantor measure) less than 1. We introduced an analogous concept called “the transport dimension of μ ”, and showed that

$$\dim_H(\mu) \leq \dim_T(\mu) \leq \dim_M(\mu)$$

with a slight modification of the definition of $\dim_M(\mu)$. The major difference between $\dim_I(\mu)$ and $\dim_T(\mu)$ is that $\dim_T(\mu)$ is allowed to be less than 1. For instance, for the Cantor measure, we show that $\dim_T(\mu)$ is $\frac{\ln 2}{\ln 3}$ which is exactly the dimension of the Cantor set.

We now describe how we defined the transport dimension $\dim_T(\mu)$. Let $\mathcal{A}(\mathbb{R}^m)$ be the space of all atomic probability measures on \mathbb{R}^m . We first introduced the d_α metric on $\mathcal{A}(\mathbb{R}^m)$ for any $\alpha \in (-\infty, 1]$. Let

$$\mathcal{P}_\alpha(\mathbb{R}^m)$$

be the completion of the metric space $\mathcal{A}(\mathbb{R}^m)$ with respect to the metric d_α . One may check that if $\beta < \alpha$, then $\mathcal{P}_\beta(\mathbb{R}^m) \subseteq \mathcal{P}_\alpha(\mathbb{R}^m)$, and for all μ, ν in $\mathcal{P}_\beta(\mathbb{R}^m)$ we have $d_\beta(\mu, \nu) \geq d_\alpha(\mu, \nu)$. Note that when $\alpha = 1$, the metric d_1 is the usual Monge’s distance on $\mathcal{A}(\mathbb{R}^m)$ and $\mathcal{P}_1(\mathbb{R}^m)$ is just the space $\mathcal{P}(\mathbb{R}^m)$ of all probability measures on \mathbb{R}^m . Therefore, each element in $\mathcal{P}_\alpha(\mathbb{R}^m)$ can be viewed as a probability measure on \mathbb{R}^m when $\alpha < 1$.

Let $\{\mathbf{a}_k\}_{k=1}^\infty$ be a sequence of atomic measures on \mathbb{R}^m of equal total mass in the form of

$$\mathbf{a}_k = \sum_{i=1}^{N_k} m_i^{(k)} \delta_{x_i^{(k)}}$$

for each k , and $\alpha < 1$. We say that this sequence is a d_α -admissible Cauchy sequence if for any $\epsilon > 0$, there exists an N such that for all $n > k \geq N$ there exists a partition of

$$\mathbf{a}_n = \sum_{i=1}^{N_k} \mathbf{a}_{n,i}^{(k)}$$

with respect to \mathbf{a}_k as sums of disjoint atomic measures and a path

$$G_{n,i}^k \in \text{Path}(m_i^{(k)} \delta_{x_i^{(k)}}, \mathbf{a}_{n,i}^{(k)})$$

for each $i = 1, 2, \dots, N_k$ such that

$$\sum_{i=1}^{N_k} \mathbf{M}_\alpha(G_{n,i}^k) \leq \epsilon.$$

Clearly, each d_α -admissible Cauchy sequence of probability atomic measures corresponds to an element in $\mathcal{P}_\alpha(\mathbb{R}^m)$. Let

$$\mathcal{D}_\alpha(\mathbb{R}^m) \subseteq \mathcal{P}_\alpha(\mathbb{R}^m)$$

be the set of all probability measures μ which corresponds to a d_α -admissible Cauchy sequence of probability measures. Also, note that if $\mu, \nu \in \mathcal{D}_\alpha(X)$, one automatically has $d_\alpha(\mu, \nu) < \infty$.

We now introduce the following concept: for any probability measure μ on \mathbb{R}^m , we define the *transport dimension* of μ to be

$$\dim_T(\mu) := \inf_{\alpha < 1} \left\{ \frac{1}{1-\alpha} : \mu \in \mathcal{D}_\alpha(\mathbb{R}^m) \right\}.$$

Note that if $\frac{1}{1-\alpha} > \dim_T(\mu)$, then $\mu \in \mathcal{D}_\alpha(\mathbb{R}^m)$, and thus $d_\alpha(\mu, \delta_O) < +\infty$ for any fixed point $O \in \mathbb{R}^m$. In addition $\alpha \geq 0$, then there exists an α -optimal transport path from μ to δ_O .

For any $\alpha < 1$, let

$$\mathcal{S}_\alpha(\mathbb{R}^m) = \{ \Lambda(\mu - \nu) : \Lambda \geq 0, \mu, \nu \in \mathcal{D}_\alpha(\mathbb{R}^m) \}$$

be a collection of signed measures. Clearly, $\mathcal{S}_{\alpha_1}(\mathbb{R}^m) \subseteq \mathcal{S}_{\alpha_2}(\mathbb{R}^m)$ if $\alpha_1 \leq \alpha_2$.

Definition 5.1. For any two probability measures μ, ν on \mathbb{R}^m , we define

$$D(\mu, \nu) := \inf_{\alpha < 1} \left\{ \frac{1}{1-\alpha} : \mu - \nu \in \mathcal{S}_\alpha(\mathbb{R}^m) \right\}.$$

It is easy to see that $D(\mu, \nu)$ is a pseudometric² on the space of probability measures on \mathbb{R}^m . This pseudometric D is called the *dimensional distance* because it gives the following geometric meaning to the transport dimension of measures.

Theorem 5.2. For any positive probability measure μ , we have

$$\dim_T(\mu) = D(\mu, \mathbf{a})$$

where \mathbf{a} is any atomic probability measure.

This theorem says that the transport dimension of a probability measure μ is the distance from μ to any atomic measure with respect to the dimensional distance. In other words, the dimension information of a measure tells us quantitatively how far the measure is from being an atomic measure.

6. AN APPLICATION IN FRACTAL GEOMETRY: DIFFUSION LIMITED AGGREGATION DRIVEN BY OPTIMAL TRANSPORTATION

Diffusion-limited aggregation, or DLA, has been extensively employed (see [12]) since its proposition in 1981 by Witten and Sander [23] to model cluster growth controlled by the random process of diffusion. This leads to structures with very regular fractal properties: for instance, off-lattice DLA in the plane evolves a cluster with fractal dimension 1.71 [15]. While the use of different lattices has an effect on the resulting fractal dimension, they fall within a narrow range. This makes it difficult to model processes in which varying a certain parameter affects the shape of the cluster. In particular, electrodeposition experiments use a range of voltages and produce a range of cluster shapes only one of which can correspond to DLA.

²A pseudometric D means that it is nonnegative, symmetric, satisfies the triangle inequality, and $D(\mu, \mu) = 0$. But $D(\mu, \nu) = 0$ does not imply $\mu = \nu$.

In [33], we modified the standard DLA model with ideas from ramified optimal transportation. The key idea in [33] is that the probability of sticking is inversely proportional to the additional cost of transporting the new particle to the root *via* the existing transport system in the current aggregate.

Let us describe this idea more precisely as follows. The model of diffusion-limited aggregation begins with any number of seeds in a space. A particle is released at a radius slightly larger than the maximum radius of the current aggregate and undergoes a random walk (Brownian motion). Once it comes within some critical distance of the existing aggregate it sticks and the process starts over. We may represent the current aggregate by a weighted directed tree G . When a new particle arrives at a position x which is adjacent to a vertex v of G , then we get a new aggregate represented by another weighted directed tree \tilde{G} . Suppose Γ_v is the unique path on the weighted directed tree G from the vertex v to the root. Then the additional transport cost for transporting a mass ϵ at x through Γ_v to the root is

$$\begin{aligned} \mathbf{M}_\alpha(\tilde{G}) - \mathbf{M}_\alpha(G) &= \sum_{e \in \Gamma_v} ([w(e) + \epsilon]^\alpha - [w(e)]^\alpha) \text{length}(e) + \epsilon^\alpha L \\ &= \epsilon^\alpha L \left(\sum_{e \in \Gamma_v} \left(\left[\frac{w(e)}{\epsilon} + 1 \right]^\alpha - \left[\frac{w(e)}{\epsilon} \right]^\alpha \right) \frac{\text{length}(e)}{L} + 1 \right), \end{aligned}$$

where L is the distance from x to v . Now we take a unit mass $\epsilon = 1$, and also take the length of each edge of the path to be a constant (*e.g.* the diameter of the particle). DLA is often done on a square, triangular or hexagonal lattice in which case the edge length would be the lattice size. Then we may assume the additional cost is

$$c(v) = \sum_{e \in \Gamma_v} ([w(e) + 1]^\alpha - [w(e)]^\alpha) + 1,$$

and set the probability $p(v)$ of a new particle sticking at vertex v to be inversely proportional to $c(v)$, *e.g.* $p(v) = \frac{1}{c(v)}$.

Note that when $\alpha = 0$ we always have $c(v) = 1$ and $p(v) = \frac{1}{c(v)} = 1$. Thus we get the standard DLA structure. When $\alpha > 0$, $c(v) \geq 1$ and $p(v) = \frac{1}{c(v)} \leq 1$. However, when $\alpha < 0$ one can show that $0 < c(v) \leq 1$ and thus $\frac{1}{c(v)} \geq 1$. To get a probability we normalize $\frac{1}{c(v)}$ and set $p(v) = \frac{c_G}{c(v)}$, where c_G is the normalization factor. An ideal normalization factor c_G is the minimum of all values $c(v)$ over all vertices v of G . Nevertheless, instead of calculating this number directly we approximate it by taking c_G to be the minimum of all possible $c(v)$ that have been calculated before (including the current one). When $\alpha \geq 0$ we have $c_G = 1$, and when $\alpha < 0$ then c_G becomes a small positive number. So for each v , $p(v) \leq 1$ as $c_G \leq c(v)$.

We use the following process to aggregate a cluster:

- (1) We release a particle from infinity which in practice is a point on a circle whose radius is slightly larger than the maximum radius of the cluster.
- (2) The particle undergoes a random walk through unoccupied spaces until it moves adjacent to the cluster.
- (3) The additional cost $c(v)$ of attaching to one of the adjacent cluster points is calculated for each adjacent point v and this is translated into a probability $p(v)$.
- (4) If the particle sticks, it is added to the cluster and the process starts anew at Step 1. If the particle does not stick it continues its random walk from Step 2.

Using this algorithm we form in Figure 6 the clusters on a square lattice with different parameters α .

In this model, positive α promotes growth near the root whereas negative α promotes growth at the tips of the cluster. $\alpha = 0$ is a phase transition point and corresponds to standard DLA. What makes this model interesting is that when α is negative enough the final cluster is an one dimensional curve. On the other hand, when α is positive enough we get a nearly two dimensional disk. Thus our model encompasses *the full range of fractal dimension* from 1 to 2.

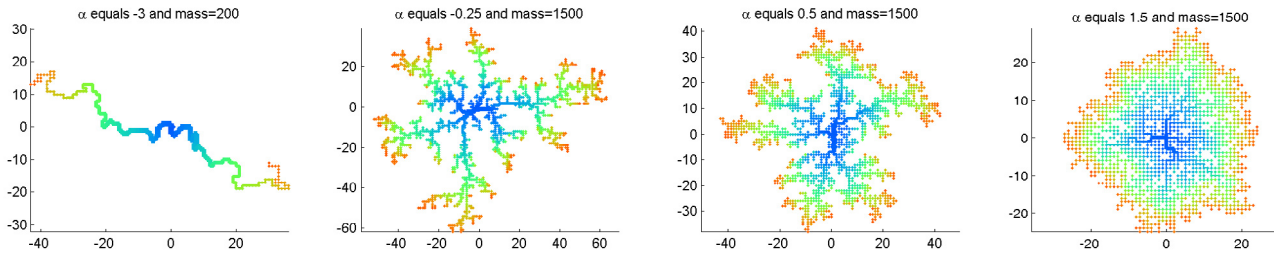


FIGURE 6. Modified DLAs with varying parameter α .

7. AN APPLICATION IN METRIC GEOMETRY: GEODESIC PROBLEMS IN QUASIMETRIC SPACES

A quasimetric space is a generalized metric space, in which the distance satisfy a relaxed triangle inequality $d(x, y) \leq c(d(x, z) + d(z, y))$ for some $c \geq 1$, rather than the usual triangle inequality. Quasimetric concepts are critical for studying harmonic analysis on spaces of homogeneous type [5]. Further exploration of optimal transport paths provides us a motivational example to study the geodesic problem in quasimetric spaces. We first introduce a quasimetric on the space of atomic probability measures as follows:

Definition 7.1. For any two atomic probability measures

$$\mathbf{a} = \sum_{i=1}^m a_i \delta_{x_i} \text{ and } \mathbf{b} = \sum_{j=1}^n b_j \delta_{y_j}$$

on a metric space (Y, d) , and $\alpha < 1$, define

$$J_\alpha(\mathbf{a}, \mathbf{b}) := \min \left\{ \sum_{i=1}^m \sum_{j=1}^n (\gamma_{ij})^\alpha d(x_i, y_j) : \gamma = \sum_{i=1}^m \sum_{j=1}^n \gamma_{ij} \delta_{(x_i, y_j)} \in \text{Plan}(\mathbf{a}, \mathbf{b}) \right\}.$$

For any given natural number $N \in \mathbb{N}$, let $\mathcal{A}_N(Y)$ be the space of all atomic probability measures $\sum_{i=1}^m a_i \delta_{x_i}$ on Y with $m \leq N$. One can show that J_α defines a quasimetric on $\mathcal{A}_N(Y)$. Note that, in general, J_α may fail to be a metric on $\mathcal{A}_N(Y)$ as demonstrated in the following example.

Example 7.2. For any $\alpha < 1$, let y be a positive real number. Then, we consider three atomic measures in $Y = \mathbb{R}^2$:

$$\mathbf{a} = \frac{1}{2} \delta_{(-1, y+1)} + \frac{1}{2} \delta_{(1, y+1)}, \mathbf{b} = \delta_{(0, 0)} \text{ and } \mathbf{c} = \delta_{(0, y)}.$$

Then,

$$J_\alpha(\mathbf{a}, \mathbf{c}) + J_\alpha(\mathbf{c}, \mathbf{b}) - J_\alpha(\mathbf{a}, \mathbf{b}) = 2 \left(\frac{1}{2}\right)^\alpha \sqrt{2} + y - 2 \left(\frac{1}{2}\right)^\alpha \sqrt{1 + (y + 1)^2} < 0$$

whenever y is large enough. Thus, J_α does not satisfy the triangle inequality.

As we know, the intrinsic metric induced by a metric has played an important role in the study of metric geometry. A natural question is: will a quasimetric be able to induce an intrinsic *metric*?

To answer this question, we first studied properties of a quasimetric space (X, J) in [28]. Note that, as J only satisfies the relaxed triangle inequality, J is not necessarily continuous. One can easily show that J_α is not continuous by constructing a sequence of measures located on a “Y-shaped” optimal transport path. Nevertheless, in [28], we showed that many well-known results in metric spaces still hold in suitable quasimetric

spaces. For instance, the Ascoli-Arzelà theorem holds when (X, J) is a complete quasimetric space and J is lower semi-continuous.

Also, to study the geodesic problem in a quasimetric space (X, J) , one needs to define the length of a continuous curve $f : [a, b] \rightarrow (X, J)$. In a metric space (X, d) , we usually define

$$L(f) = \sup_P \sum_{i=1}^N d(f(t_{i-1}), f(t_i)),$$

where the supremum is over all partitions $P = \{a = t_0 < t_1 < \dots < t_N = b\}$ of $[a, b]$. However, on a quasimetric space, subdivision of a partition may *decrease* rather than increase the variation of f . One can not take the supremum of variations as before. Nevertheless, in the study of ramified optimal transportation, we observed in ([28], Rem. 4.16) that for each transport G between two atomic probability measures in $\mathcal{A}_N(Y)$, there exists a piecewise metric Lipschitz curve f in the quasimetric space $(\mathcal{A}_N(Y), J_\alpha)$ such that the transport cost

$$\mathbf{M}_\alpha(G) = \int_0^1 |\dot{f}(t)|_{J_\alpha} dt$$

where the quasimetric derivative

$$|\dot{f}(t)|_{J_\alpha} := \lim_{s \rightarrow t} \frac{J_\alpha(f(s), f(t))}{|s - t|}$$

provided the limit exists. This motivates us to define the length of a curve f in a general quasimetric space (X, J) to be

$$L_J(f) := \int_a^b |\dot{f}(t)|_J dt$$

whenever it is well defined. In particular, for any natural number M , every curve f in the family $\mathcal{P}_M([a, b], (X, J))$ of all M -piecewise metric Lipschitz curves in (X, J) has a well-defined length $L_J(f)$. For any $x, y \in X$, we consider the geodesic problem

$$\min\{L_J(f)\} \tag{7.1}$$

among all f in the family

$$\text{Path}_M(x, y) = \{f \in \mathcal{P}_M([0, 1], (X, J)) \text{ with } f(0) = x; f(1) = y\}.$$

Definition 7.3. Suppose J is a quasimetric on X . For any $x, y \in X$, and $M \in \mathbb{N}$, define

$$D_J^{(M)}(x, y) = \inf\{L_J(f) : f \in \text{Path}_M(x, y)\}$$

whenever $\text{Path}_M(x, y)$ is not empty, and set $D_J^{(N)}(x, y) = \infty$ when $\text{Path}_M(x, y)$ is empty. Since $D_J^{(M)}(x, y)$ is a decreasing function of M , we define

$$D_J(x, y) = \lim_{M \rightarrow \infty} D_J^{(M)}(x, y).$$

For each quasimetric J on X , D_J is automatically a pseudometric. When D_J is indeed a metric on X , it is called the *intrinsic metric*, or *geodesic distance*, on X induced by the quasimetric J .

We explored some sufficient conditions on X and J that ensure D_J to be a metric. One of them says that if $D_J^{(M)}(x, y)$ becomes a real valued constant $D_J(x, y)$ (i.e. independent of M) whenever M is large enough, then D_J is a metric on X . If in addition, the geodesic problem (7.1) has a solution for M large enough, then (X, D_J) is a length space. It turns out the quasimetric space $(\mathcal{A}_N(Y), J_\alpha)$ satisfies these conditions, and thus J_α induces a metric D_{J_α} . The metric d_α coincides with this induced metric D_{J_α} on $\mathcal{A}_N(Y)$. Consequently, an optimal transport path between two atomic probability measures is simply a geodesic in the length space $(\mathcal{A}_N(Y), d_\alpha)$ in the usual sense of metric geometry.

8. GENERALIZATION: RAMIFIED MASS TRANSPORTATION IN METRIC SPACES

In [31], we generalized the study of ramified optimal transportation from Euclidean spaces to geodesic metric spaces. Such generalization is not only mathematically natural but is also expected to be useful for considering specific examples of metric spaces later. By exploring various properties of the metric, we showed that many results about ramified optimal transportation is not limited to Euclidean spaces, but can be extended to suitable metric spaces.

When (X, d) is a geodesic metric space, we defined a family of metrics d_α on the space $\mathcal{A}(X)$ of atomic probability measures on X for a (possibly negative) parameter $\alpha < 1$. The space $(\mathcal{A}(X), d_\alpha)$ is still a geodesic metric space when $0 \leq \alpha < 1$. A geodesic, also called an optimal transport path, in this space is a weighted directed graph whose edges are geodesic segments. It is still true that an optimal transport path contains no cycles when $0 \leq \alpha < 1$.

In the settings of an Euclidean space \mathbb{R}^m , the degree (*i.e.* the total number of edges) at each vertex of an α -optimal transport path is uniformly bounded above by a constant $D(m, \alpha)$ as in (2.5). This is not necessarily true in the general metric space setting (as a counter example is given by [31], Example 2.0.4). We now explore sufficient conditions for the existence of such a uniform upper bound D .

Recall that, in the setting of a Euclidean space, there is a universal lower bound θ_α in (2.4) for the angle between any two neighboring edges on an optimal transport path. To extend this result in a more general setting, we consider geodesic metric spaces of bounded curvature as follows: We now recall the definition of a space of bounded curvature [3]. For a real number k , the model space M_k^2 is the simply connected surface with constant curvature k . That is, if $k = 0$, then M_k^2 is the Euclidean plane. If $k > 0$, then M_k^2 is obtained from the sphere \mathbb{S}^2 by multiplying the distance function by the constant $\frac{1}{\sqrt{k}}$. If $k < 0$, then M_k^2 is obtained from the hyperbolic space \mathbb{H}^2 by multiplying the distance function by the constant $\frac{1}{\sqrt{-k}}$. The diameter of M_k^2 is denoted by $D_k := \pi/\sqrt{k}$ for $k > 0$ and $D_k := \infty$ for $k \leq 0$.

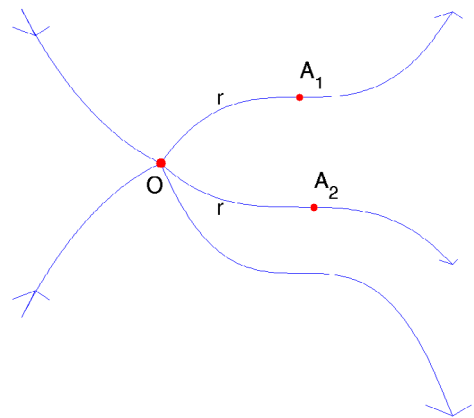
Let (X, d) be a geodesic metric space, and let ΔABC be a geodesic triangle in X with geodesic segments as its sides. A *comparison* triangle $\Delta \bar{A}\bar{B}\bar{C}$ is a triangle in the model space M_k^2 such that $d(A, B) = |\bar{A} - \bar{B}|_k$, $d(B, C) = |\bar{B} - \bar{C}|_k$ and $d(A, C) = |\bar{A} - \bar{C}|_k$, where $|\cdot|_k$ denotes the distance function in the model space M_k^2 . Such a triangle is unique up to isometry. Also, the interior angle of $\Delta \bar{A}\bar{B}\bar{C}$ at \bar{B} is called the *comparison angle* between A and C at B . A geodesic metric space (X, d) is a *space of curvature bounded above by a real number k* if for every geodesic triangle ΔABC in X and every point h in the geodesic segment γ_{AC} , one has $d(h, B) \leq |\bar{h} - \bar{B}|_k$ where \bar{h} is the point on the side $\gamma_{\bar{A}\bar{C}}$ of a comparison triangle $\Delta \bar{A}\bar{B}\bar{C}$ in M_k^2 such that $|\bar{h} - \bar{C}|_k = d(h, C)$.

Now, let X be a geodesic metric space with curvature bounded above by a real number k . Suppose $\alpha < 1$ and G is an α -optimal transport path between two atomic probability measures $\mathbf{a}, \mathbf{b} \in \mathcal{A}(X)$.

Let O be any vertex of G and e_i be any two distinct directed edges with $e_i^+ = O$ (or $e_i^- = O$ simultaneously) and weight $m_i > 0$ for $i = 1, 2$. Also, for $i = 1, 2$, let A_i be the point on the edge e_i with $d(O, A_i) = r$ for some r satisfying $0 < r \leq \frac{1}{2}D_k$ and $r \leq \text{length}(e_i)$.

By means of the spherical/ordinary/hyperbolic law of cosines in the model space M_k^2 , one can show the following estimates for the distance:

$$d(A_1, A_2) \geq \begin{cases} \frac{2}{\sqrt{k}} \sin^{-1} \left(\frac{R}{2} \sin \left(r\sqrt{k} \right) \right), & \text{if } k > 0 \\ Rr, & \text{if } k = 0 \\ \frac{2}{\sqrt{-k}} \sinh^{-1} \left(\frac{R}{2} \sinh \left(r\sqrt{-k} \right) \right), & \text{if } k < 0, \end{cases} \quad (8.1)$$



where R is the constant given by

$$R = \sqrt{\frac{(m_1^\alpha + m_2^\alpha)^2 - (m_1 + m_2)^{2\alpha}}{m_1^\alpha m_2^\alpha}}. \tag{8.2}$$

Let θ be the comparison angle between A_1 and A_2 at O in the model space M_k^2 . Using the above estimates on $d(A_1, A_2)$, one can show that

$$\theta \geq \arccos\left(1 - \frac{R^2}{2}\right) = \arccos\left(\frac{1 - k_1^{2\alpha} - k_2^{2\alpha}}{2k_1^\alpha k_2^\alpha}\right),$$

where

$$k_1 = \frac{m_1}{m_1 + m_2} \text{ and } k_2 = \frac{m_2}{m_1 + m_2}.$$

Moreover,

$$\theta \geq \theta_\alpha := \begin{cases} \frac{\pi}{2}, & \text{if } 0 < \alpha \leq \frac{1}{2} \\ \arccos(2^{2\alpha-1} - 1), & \text{if } \frac{1}{2} < \alpha < 1 \text{ or } \alpha \leq 0. \end{cases}$$

Note that when $k = 0$, this agrees with what we have found in (2.4). Also, when α approaches $-\infty$, θ_α approaches π ; and when α approaches 1, θ_α approaches 0.

A metric space X is called *doubling* if there is a doubling constant $C_d \geq 1$ so that every subset of diameter r in X can be covered by at most C_d subsets of diameter at most $\frac{r}{2}$.

When X is a geodesic doubling metric space of curvature bounded above, we showed that the degree of any vertex of an α -optimal transport path between two atomic measures on X is bounded above by a constant D depending only on α and the doubling constant of X . In the other direction, we also provided a lower bound of the curvature of X by a quantity related to the degree of vertices.

Furthermore, when X is a complete geodesic metric space, as in Section 5, we considered optimal transportation between any two probability measures on X by considering the *completion* of the metric space $(\mathcal{A}(X), d_\alpha)$. A geodesic, if it exists, in the completed metric space is viewed as an α -optimal transport path between measures. The existence of an optimal transport path is closely related to the dimensional information of the measures. Analogous to Section 5, we considered concepts such as transport dimension as well as dimensional distance of measures. It turns out that main results in Section 5 still hold when \mathbb{R}^m is replaced with a complete geodesic doubling metric space X with Assouad dimension m .

When X is a compact geodesic doubling metric space with Assouad dimension m and the parameter $\alpha > \max\{1 - \frac{1}{m}, 0\}$, then we showed that the space $\mathcal{P}(X)$ of probability measures on X with respect to d_α is a geodesic metric space. In other words, there exists an α -optimal transport path between any two probability measures on X .

9. APPLICATIONS IN MATHEMATICAL BIOLOGY

9.1. The formation of a plant leaf

In [27], we built a mathematical model to understand the dynamic formation of a plant leaf. Plant leaves have diverse and elaborate shapes and venation patterns. Why do tree leaves grow in such an amazing way? What is the mathematics behind it? To answer these questions, we first look at basic functions of leaves. A tree leaf will transport resources like water and solutes from its root to its tissues *via* xylem, absorb solar energy at their cells through photosynthesis (a function of leaf surface area), and then transport the chemical products (carbohydrates) synthesized in the leaf back to its root by phloem. Thus, a leaf tends to increase the surface area as large as possible to maximize metabolic capacity, because metabolism produces the energy and materials required to sustain and reproduce life. On the other hand, and more important, the leaf tends to maximize

internal efficiency by developing an efficient transporting system. One of the main tasks here is to describe this efficient biological transport system.

For simplicity, a leaf may be viewed as a finite union of square cells, centered on a given grid $\Gamma_h = \{(mh, nh) : m, n \in \mathbb{Z}\}$ in \mathbb{R}^2 of size h . A leaf will be a subset of Γ_h . However, not every subset of Γ_h gives a leaf. One aim is to understand the speciality of a reasonable tree leaf. Let

$$\Omega = \{x_1, x_2, \dots, x_N\} \subset \Gamma_h$$

be any finite subset representing a prospective leaf.

The amount of water needed at each cell is proportional to its area ($= h^2$). Without losing generality, we may assume it is h^2 . So, each $x_i \in \Omega$ corresponds to a particle of mass h^2 located at x_i . The aim is to transport these particles to the node (*i.e.* the intersection point of the leaf with the branch) O in some cost efficient way.

A transport system for the leaf is modeled by a transport path G from O to $\sum_{i=1}^N (h^2)\delta_{x_i}$. The cost function $\mathbf{M}_{\alpha,\beta}(G)$ on the transport system G is a modified version of $\mathbf{M}_\alpha(G)$. As before, the parameter α comes from the transport economy of scales. The other parameter β arises from the principle that when there exists a given transport direction, it is cheaper to transport in the given direction than transport in any other direction. The cost is an increasing function of the angle between the favored direction and other directions. More precisely, let $\alpha \in [0, 1)$ and $\beta > 0$ be two fixed real numbers. For any transport system $G = \{V(G), E(G), w\}$, the cost of G is defined by

$$\mathbf{M}_{\alpha,\beta}(G) := \sum_{e \in E(G)} m_\beta(e^+) (w(e))^\alpha \text{length}(e).$$

The function m_β on $V(G)$ is defined inductively as follows: we first set $m_\beta(O) = 1$. For any other $v \in V(G)$, let $p(v)$ be the “parent vertex” of v , then we set

$$m_\beta(v) = m_\beta(p(v)) H_\beta(\vec{e}_v, \vec{e}_{p(v)}),$$

where for any two unit vectors u, v with θ being the angle between them,

$$H_\beta(u, v) = \begin{cases} |u \cdot v|^{-\beta} = \frac{1}{\cos^\beta(\theta)}, & \text{if } u \cdot v > 0 \\ +\infty, & \text{otherwise.} \end{cases}$$

Note that the function m_β depends on G .

The growing of a plant leaf is a dynamic process of generating new cells. At every stage, the leaf will develop an optimal transport system to transport water between the root and the existing cells with respect to the given transport cost function. Let

$$\mathcal{A}_h := \left\{ (\Omega, G) : \Omega \subset \Gamma_h, G \text{ is an optimal transport system of } \Omega \text{ under the } \mathbf{M}_{\alpha,\beta} \text{ cost} \right\}.$$

Also, as the environment changes, the leaf may generate some new cells nearby its boundary. The selection of those new cells is not random. Under the same environmental conditions, each potential new cell outside the existing leaf produces about the same amount of revenue (such as the absorbed solar energy). The revenue that a new cell produces is proportional to its area. That is, for some constant $\epsilon > 0$,

$$\text{revenue} = \epsilon h^2.$$

However, the expense corresponding to each potential new cell varies with respect to the position of the cell. Here, the expense is mainly the transport cost of water and nutrients between the cell and the node O . Given $(\Omega, G) \in \mathcal{A}_h$, let B be the boundary set of Ω . For any $x \in \Gamma_h \setminus \Omega$, and any “boundary point” $b \in B$, we define “the transport cost of x via b ” to be

$$C_\Omega(x, b) := h^{2\alpha} |x - b| m_\beta(b) H_\beta\left(\frac{x - b}{|x - b|}, \vec{e}_b\right) + P_G(h^2, b),$$

where

$$P_G(y, v) = \sum_{u \in P_v \setminus \{O\}} m_\beta(u) [(w(e_u) + y)^\alpha - (w(e_u))^\alpha] \text{length}(e_u)$$

measures “the increment of the total cost” if one adds an extra mass of weight y to the point $v \in V(G)$. Then, the expense for generating a new cell at x is mainly given by the transport cost

$$C_\Omega(x) := \min_{b \in B} C_\Omega(x, b) = C_\Omega(x, b(x)).$$

The selection of new cells during the generation process is governed by a *selection principle* which says that a new cell is generated only if the expense is less than the revenue it produces. Based on the selection principle, for any given $\epsilon > 0$, we set

$$\begin{aligned} \tilde{\Omega} &= \{x \in \Gamma_h \setminus \Omega : C_\Omega(x) \leq \epsilon h^2\} \cup \Omega, \\ \tilde{V} &= V(G) \cup \tilde{\Omega}, \quad \tilde{E} = E(G) \cup \{[x, b(x)] : x \in \tilde{\Omega} \setminus \Omega\} \end{aligned}$$

and \tilde{G} be the optimal transport system of $\tilde{\Omega}$ achieved by modifying $\tilde{G} = \{\tilde{V}, \tilde{E}, \bar{w}\}$. In other words, we take \tilde{G} as the initial transport path, and use algorithms analogous to those stated in Section 4 to reduce the $\mathbf{M}_{\alpha, \beta}$ transport cost as much as possible, and called the final transport path \tilde{G} . Thus, we get a map $L_{\epsilon, h} : \mathcal{A}_h \rightarrow \mathcal{A}_h$ by letting

$$L_{\epsilon, h}(\Omega, G) = (\tilde{\Omega}, \tilde{G}).$$

Note that \tilde{G} might reduce the transporting costs for cells outside $\tilde{\Omega}$. It is possible that $C_\Omega(x) > \epsilon h^2$ but $C_{\tilde{\Omega}}(x) \leq \epsilon h^2$. By the selection principle, we should also select such cells as new cells. Thus, we need to consider further: $L_{\epsilon, h}^2(\Omega, G) = L_{\epsilon, h} \circ L_{\epsilon, h}(\Omega, G), L_{\epsilon, h}^3(\Omega, G), \dots$ and so on. Will $L_{\epsilon, h}^n(\Omega, G)$ stop growing and stay in a bounded domain when n is large enough? The answer is yes if $\alpha \in (1/2, 1)$.

Indeed, for $\alpha > 1/2$, let $\mathcal{A}_{\epsilon, h} = \{(\Omega, G) \in \mathcal{A}_h : \Omega \subset B_{R_\epsilon}(O)\}$, where R_ϵ is a constant depending only on ϵ and α . A proposition in [27] says that $L_{\epsilon, h}$ maps $\mathcal{A}_{\epsilon, h}$ into itself. That is, for any $(\Omega, G) \in \mathcal{A}_{\epsilon, h}$, we have $(\tilde{\Omega}, \tilde{G}) := L_{\epsilon, h}(\Omega, G) \in \mathcal{A}_{\epsilon, h}$. Now, for $\alpha \in (1/2, 1)$ and $(\Omega, G) \in \mathcal{A}_{\epsilon, h}$, we may define $(\Omega_n, G_n) = L_{\epsilon, h}(\Omega_{n-1}, G_{n-1}) \in \mathcal{A}_{\epsilon, h}$ with $(\Omega_0, G_0) = (\Omega, G)$. This leads to a sequence of nested sets $\Omega = \Omega_0 \subset \Omega_1 \subset \Omega_2 \subset \dots \subset B_{R_\epsilon}(O) \cap \Gamma_h$, which is a finite set. Thus, when N is large enough, $\Omega_N = \Omega_{N+1} = \dots$. As a result, one may define $g_{\epsilon, h} : \mathcal{A}_{\epsilon, h} \rightarrow \mathcal{A}_{\epsilon, h}$ by sending

$$g_{\epsilon, h}(\Omega, G) = (\Omega_N, G_N) = \lim_{n \rightarrow \infty} L_{\epsilon, h}^n(\Omega, G).$$

This map is called the generation map.

With a given initial growing direction, a leaf originates from a bud which is modeled by

$$\Omega_0 = \{O\} \quad \text{and} \quad G_0 = \{\{O\}, \emptyset, -\}. \tag{9.1}$$

So, Ω_0 consists only the node O , and G_0 contains no edges. This element (Ω_0, G_0) generates a subset of \mathcal{A}_h by the generation map. When the environmental conditions change, the corresponding revenue that a cell can produce also changes. When the corresponding revenue of each cell increased to a certain degree, it becomes beneficial to produce some new cells, and thus the leaf will grow. Due to limited resources, the revenue that a cell can possibly produce is bounded above. This fact forces the leaf to stop growing after some time. More precisely,

Definition 9.1. For any $\epsilon > 0$ and $h > 0$, a pair $(\Omega, G) \in \mathcal{A}_{\epsilon, h}$ is called an (ϵ, h) leaf if there exists a list $\{(\Omega_n, G_n)\}_{n=1}^k$ of elements in \mathcal{A}_h such that for each $n = 1, 2, \dots, k$,

$$(\Omega_n, G_n) = g_{\epsilon, h}(\Omega_{n-1}, G_{n-1})$$

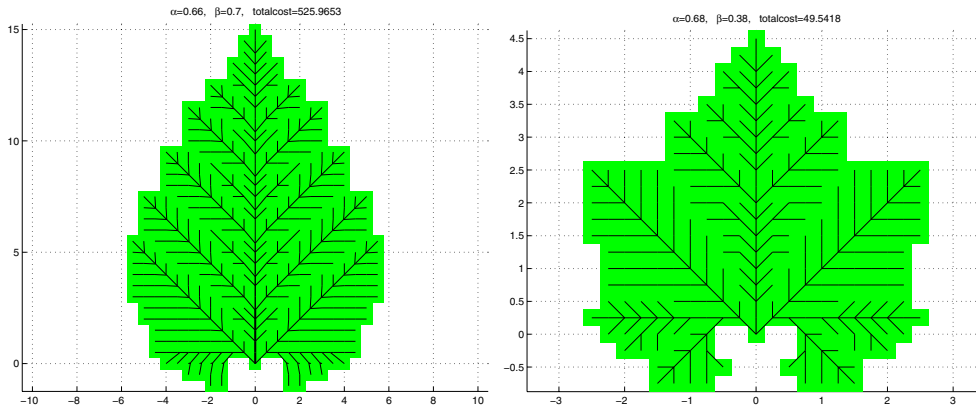


FIGURE 7. Some numerically generated leaves.

and

$$(\Omega, G) = (\Omega_k, G_k)$$

for some positive numbers ϵ_i 's satisfying

$$0 < \epsilon_1 < \epsilon_2 < \dots < \epsilon_k = \epsilon.$$

As a result, the final shape and venation pattern of a leaf are mainly determined by the cost function defined on the collection of all possible transport systems, as well as the actual environment. Based on this model, we also provided some computer visualization of tree leaves, which resemble many known leaves including the maple and mulberry leaf.

9.2. Transport efficiency of transport systems in human placentas

The human newborn is a reflection of the entirety of nutrients transferred from the maternal circulatory system to the fetal system across the placenta during gestation. By extension, birth weight and newborn health depend on placental function. In [37, 38], we quantify efficiency of the transport system in the human placenta and study its role played in the birth weight as well as the gestational age of the human newborn.

As stated in [40], 1110 placentas were collected by an academic health center in central North Carolina. For each placenta (see Fig. 8a), a trained technician captured series of x, y coordinates that marked the site of the umbilical cord insertion and the perimeter of the fetal surface (see Fig. 8b). Thus, each placenta P is represented by a pair (D, O) where D is a bounded planar domain and O is a fixed point inside D representing the site of the umbilical cord insertion.

We say two placentas $P_1 = (D_1, O_1)$ and $P_2 = (D_2, O_2)$ have the same shape if there exists a number $\lambda > 0$ such that the mapping $f_\lambda : D_1 \rightarrow D_2$ given by

$$f(x) = \lambda(x - O_1) + O_2, x \in D_1$$

is one-to-one and onto. To encode the shape information of the placenta for analysis and quantifying its impact on the birth weight of the newborn, we introduced the variable “shape factor” and removed the impact of size as follows: For each placenta $P = (D, O)$ and $\alpha < 1$, the α -shape factor of P is defined by

$$S_\alpha(P) := \frac{d_\alpha(\mu_D, |D| \delta_O)}{|D|^{\alpha+0.5}},$$

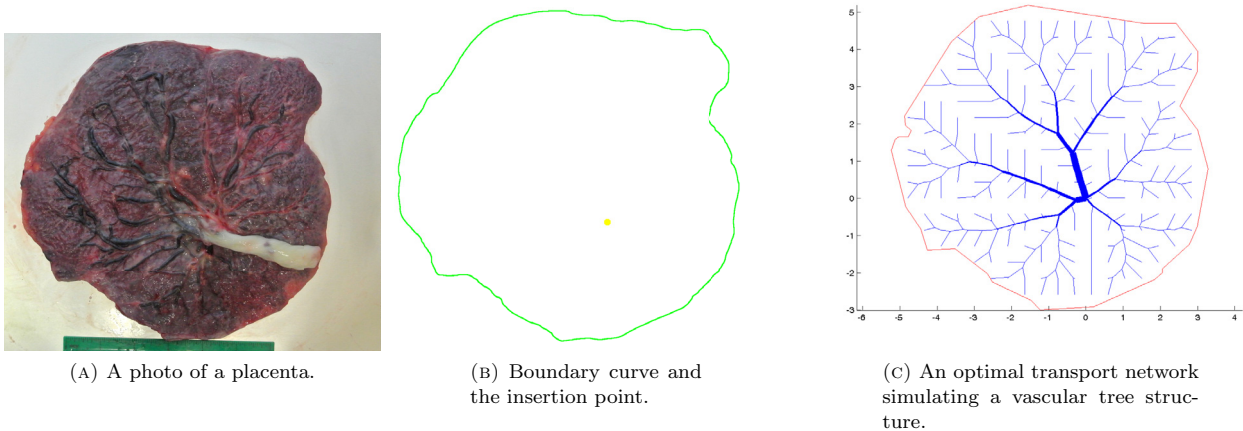


FIGURE 8.

where μ_D denotes the Lebesgue measure on \mathbb{R}^2 restricted to the set D , $|D|$ denotes the area of D , and d_α is the distance defined in (2.1). It turns out that two placentas of the same shape have the same shape factor independent of their sizes.

Since transportation cost $d_\alpha(\mu_D, |D|\delta_O)$ is proportional to the shape-factor S_α , among placentas of similar sizes, the smaller the shape-factor S_α the more efficient the transport system corresponding to the placenta. This motivates us to consider the following definition: for each placenta P and $\alpha < 1$, the transport α - efficiency of P is defined to be

$$E_\alpha(P) := \frac{S_\alpha(B(0,1))}{S_\alpha(P)},$$

where $S_\alpha(B(0,1))$ denotes the α -shape factor of the unit ball $B(0,1)$ in \mathbb{R}^2 with respect to the origin $(0,0)$. Clearly, two placentas of the same shape have the same transport efficiency $E_\alpha(P)$.

To calculate the transport cost $d_\alpha(\mu_D, |D|\delta_O)$, we used algorithms described in [30] to generate an approximated optimal transport path $G_P \in \text{Path}(\mu_D, |D|\delta_O)$. This leads to Figure 8. We used this idealized transport system G_P to model the vein (or the arterial) structure inside the placenta. The value $d_\alpha(\mu_D, |D|\delta_O)$ is approximated by $\mathbf{M}_\alpha(G)$. Thus,

$$S_\alpha(P) \approx \frac{\mathbf{M}_\alpha(G_P)}{|D|^{\alpha+0.5}}. \tag{9.2}$$

Now, for each placenta P in the available data and $\alpha = 0.85$, we calculate the transport cost $\mathbf{M}_\alpha(G_P)$, shape-factor $S_\alpha(P)$ and the transport efficiency $E_\alpha(P)$.

Statistically, the transport cost $\mathbf{M}_\alpha(G)$ is highly correlated with the birth weight. To see it geometrically, we plot the average transport cost as a function of birth weight in Figure 9. Here, for each value B in $\{500, 1000, \dots, 5000\}$, we calculate the average transport cost of all placentas whose birth weights are among $(B - 500, B + 500)$ grams. Figure 9 indicates that the average transport cost is nearly a linear function of birth weight. This positive correlation between transport cost and birth weight is expected given that transport cost primarily reflects placental size which will, on average, vary with larger and smaller fetal weights.

We also study transport efficiency of each placenta and investigate its correlations with birth weight and gestational age at birth. We show that averaged birth weight and averaged gestational age are both roughly increasing functions of the calculated placental transport efficiency (see for instance Fig. 10 for the birth weight case). Significantly, both preterm and low birth weight are associated with lower placental transport efficiency.

Furthermore, we also show that the relationship of transport efficiency to these outcomes is nonlinear, reaching a plateau at 38–39 weeks gestational age and 3200g birth weight. As illustrated in Figure 11, the contribution of

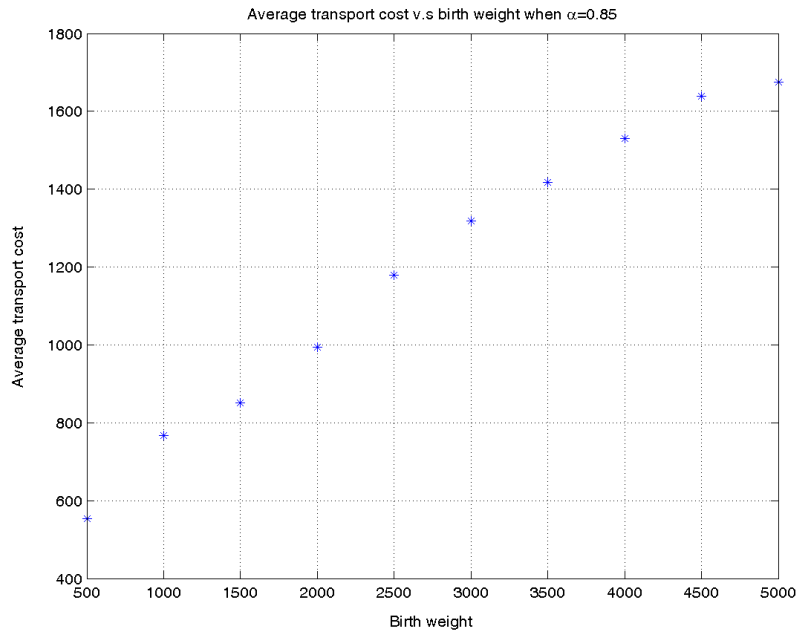


FIGURE 9. Average transport cost M_α is nearly a linear function of birth weights.

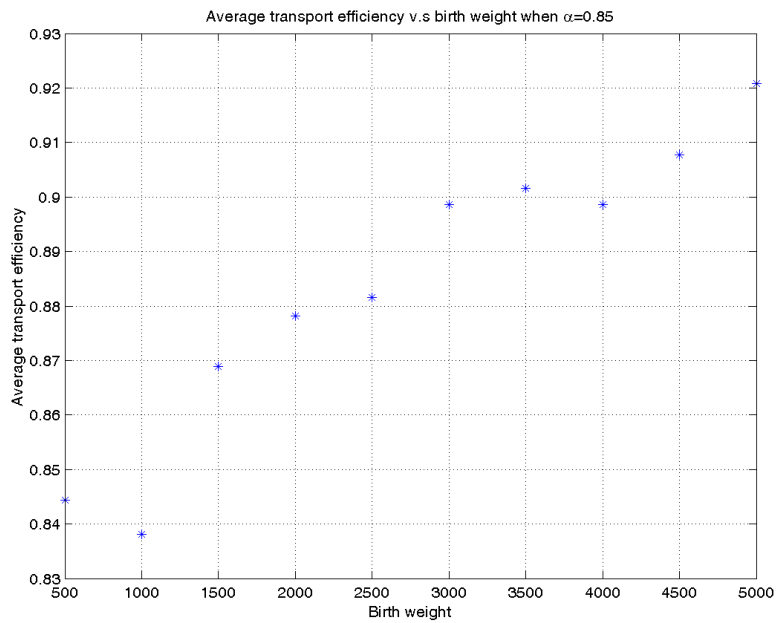


FIGURE 10. Average transport efficiency is roughly an increasing function of birth weights.

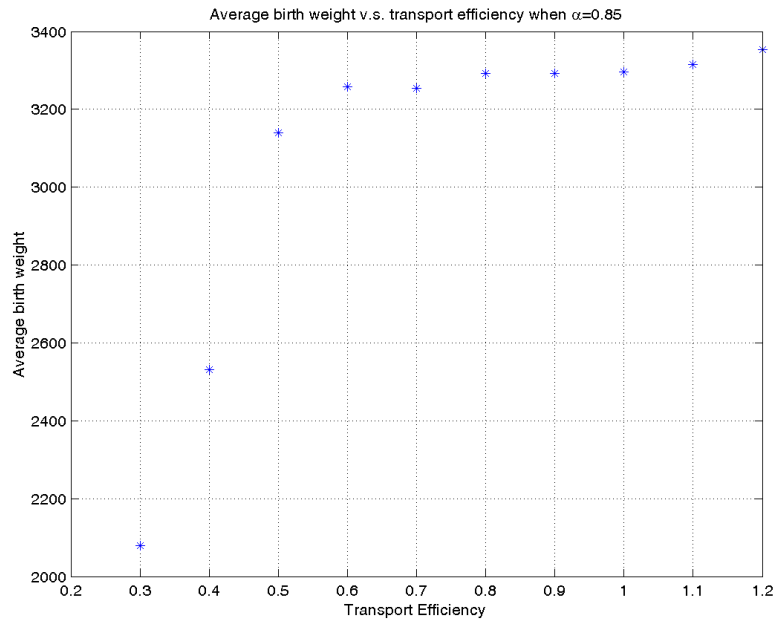


FIGURE 11. Transport efficiency still has a positive relationship with averaged birth weights.

placental transport efficiency to the eventual birth weight plateaus above a certain level of transport efficiency. Average birth weight still increases as transport efficiency increases, but with a decreased slope, indicating a smaller effect.

9.3. Human placentas, optimal transportation and Autism

Autism spectrum disorder (ASD) is one of the fastest-growing developmental disabilities. An effective way to reduce the symptoms of ASD is early intervention. However, obvious signs and symptoms of ASD tend to emerge between 2 and 3 years of age.

A natural question arises: *Is there any measurable anatomical indicator of ASD in infants at birth?*

Our studies of vascular structures in placentas provide a positive answer. The placenta is a temporary organ that serves in the exchange of nutrients and wastes between the mother and the fetus. ASD appears to have its roots in fetal life, which is highly dependent on the dynamic growth function of the placenta. Information about the timing and (possibly) the nature of early brain maldevelopment may be paralleled by the geometric structure of the placenta as measured by its size, shape, and depth as well as the thickness, length, and geometric locations of veins/arteries in the placenta.

Recently, we have traced arterial (Fig. 12a) and venous (Fig. 12b) vascular networks from the following two groups of placentas:

- 201 placentas collected by NCS. NCS is a population cohort study of children in the United States and as such would be expected to experience the baseline ASD risk seen in the general U.S. population.
- 75 placentas collected by the Early Autism Risk Longitudinal Investigation (EARLI) Study. The EARLI cohort is a network of research sites that is enrolling and following a group of mothers of children with autism at the start of another pregnancy and documenting the development of the newborn child (cases of siblings with autism) through three years of age.

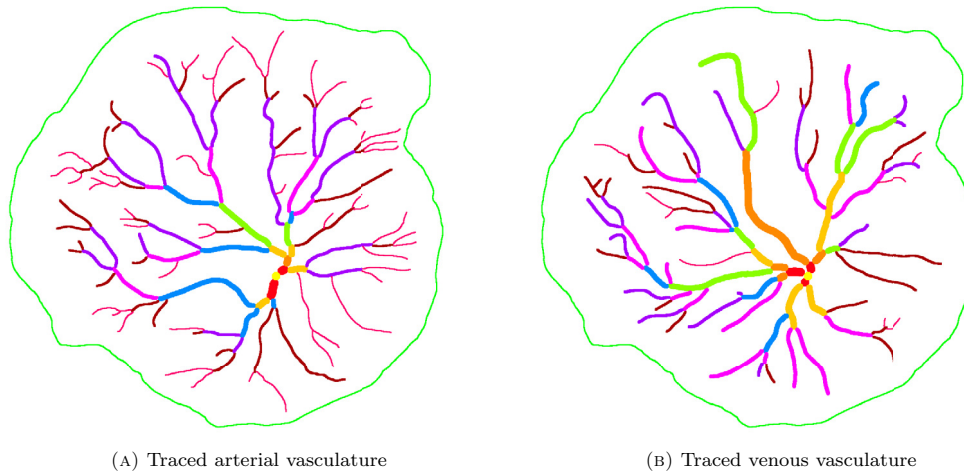


FIGURE 12. Examples of arterial and venous vasculature manually traced with electronic device.

Newborns whose placentas are in the NCS group tend to have lower ASD risk when compared with those from the EARLI group.

Since the conclusions in [39] for veins and arteries are very similar, we only discuss the situation of arteries here. Note that each traced network of arteries gives a directed weighted graph G . For each blood vessel segment e of length $l(e)$, let $f(w(e), l(e))$ be the cost of transporting a volumetric flow rate $w(e)$ along e . In particular, we are interested in the cost function $f(x, y) = x^\alpha y^\beta$ for real valued parameters α and β . The corresponding transportation cost G is

$$M_{\alpha, \beta}(G) = \sum_{e \in E(G)} w(e)^\alpha l(e)^\beta.$$

A benchmark for vascular systems was proposed by Murray in 1926 as a compromise between the frictional and metabolic costs, with the latter expressed as a cost function. The formulation of a minimum energy hypothesis led to a scaling law, $w(e) \propto d(e)^3$, between the volumetric flow rate $w(e)$ and the diameter $d(e)$ of a blood vessel segment e . By Murray's law, it holds that

$$M_{\alpha, \beta}(G) \propto \sum_{e \in E(G)} d(e)^{3\alpha} l(e)^\beta. \quad (9.3)$$

Motivated by (9.2), we use a normalization process to remove the effect of the size of the placenta on $M_{\alpha, \beta}(G)$. For each placenta $P = (D, O)$ with a traced vascular structure G , we define the shape factor of G to be

$$S_{\alpha, \beta}(G) := \frac{\sum_{e \in E(G)} d(e)^{3\alpha} l(e)^\beta}{|D|^{\alpha+\beta/2}} \propto \frac{M_{\alpha, \beta}(G)}{|D|^{\alpha+\beta/2}}.$$

Let \mathcal{P}_L^a (and \mathcal{P}_H^a) denote the set of all traced arterial vascular networks G from placentas in the NCS group (or EARLI group, respectively). For each value α and β , we consider the *relative transport efficiency* function of the two groups:

$$R^a(\alpha, \beta) := \frac{\text{mean}(\{S_{\alpha, \beta}(G) : G \in \mathcal{P}_L^a\})}{\text{mean}(\{S_{\alpha, \beta}(G) : G \in \mathcal{P}_H^a\})}.$$

For different values of α , Figure 13 plots the graphs of $R^a(\alpha, \beta)$ as functions of β .

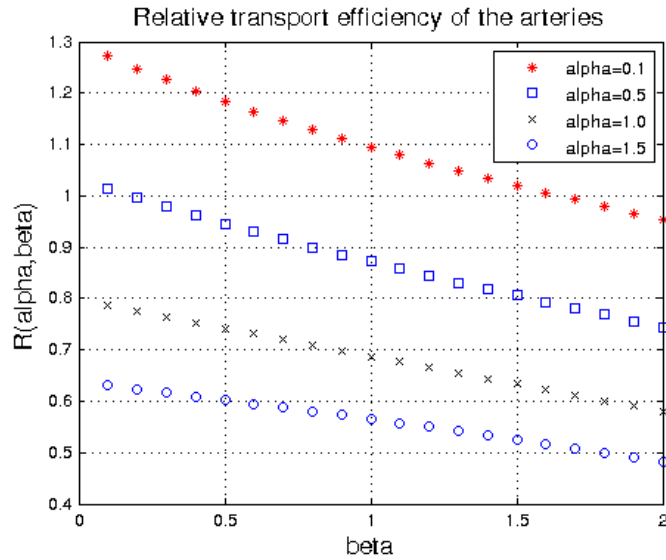


FIGURE 13. Relative transport efficiency $R^a(\alpha, \beta)$ of arteries of the two groups.

Based on Figure 13, we have the following observations:

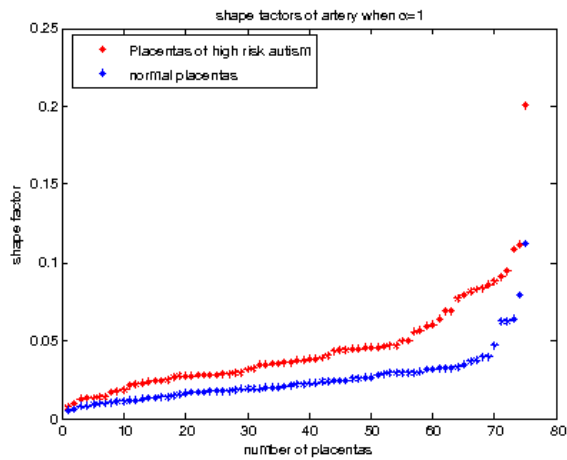
- \mathcal{P}_H^a is not a randomly selected subset of \mathcal{P}_L^a , otherwise the values $R^a(\alpha, \beta)$ would be around 1. The behaviors of traced vascular graphs of NCS placentas are different from those in the EARLI group. These differences are quantified by the relative transport efficiencies $R^a(\alpha, \beta)$.
- The values of the relative transport efficiency $R^a(\alpha, \beta)$ vary according to the transport cost function $x^\alpha y^\beta$. Under some transport costs, EARLI placentas are relatively more transport efficient than NCS placentas, while for other transport costs, the opposite situation appears.
- $R^a(\alpha, \beta)$ is a decreasing function in both variables α and β .
- For each fixed α , $R^a(\alpha, \beta)$ is nearly a linear function of β with a negative slope.

Given the placenta P of a newborn, how can one quantify the risk of the newborn having a future ASD diagnosis? For simplicity, we considered here a fixed transportation cost function, the quadratic transport cost $f(x, y) = xy^2$, with *e.g.* $\alpha = 1$ and $\beta = 2$. In Figure 13, the value $R^a(1, 2) = 0.58$, which indicates that average arteries of placentas in the NCS (*i.e.* low risk of ASD) group are more transport efficient than those of placentas in the EARLI (*i.e.* high risk of ASD) group under the quadratic transportation cost.

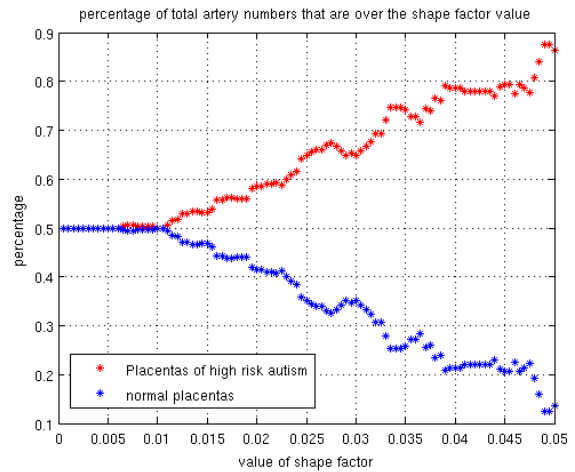
To better illustrate the difference of NCS and EARLI placentas, we order the values of both $\{S_{1,2}(G) : G \in \mathcal{P}_L^a\}$ and $\{S_{1,2}(G) : G \in \mathcal{P}_H^a\}$ increasingly, and plot them in Figure 14a: \mathcal{P}_L^a in blue, \mathcal{P}_H^a in red. Again, we find a significant difference between the two sets. It indicates that the shape factor $S_{1,2}(G)$ can be viewed as an indicator of risk of ASD. The higher the shape factor value is, the higher the risk of an ASD diagnosis. To quantify this correlation, we introduce the following definition: for any $s \geq 0$, define

$$r_H^a(s) := \frac{\#\{G \in \mathcal{P}_H^a : S_{1,2}(G) \geq s\}}{\#\{G \in \mathcal{P}_H^a \cup \mathcal{P}_L^a : S_{1,2}(G) \geq s\}}$$

and $r_L^a(s) := 1 - r_H^a(s)$. The graphs of the functions r_L^a and r_H^a are plotted in Figure 14b. For instance, for the shape factor value 0.04, the corresponding percentage value for the EARLI group shown in Figure 14b is 0.78. That is, among all placentas whose shape factors are over 0.04, 78% of them are from the EARLI group while 22% of them are from the NCS group. This big difference indicates that the shape factor of the placenta may be used as a neonatal indicator of the risk of ASD.



(A) Values of $S_{1,2}(G)$ of arteries of placentas in two groups.



(B) The graphs of the functions r_L^a and r_H^a .

FIGURE 14.

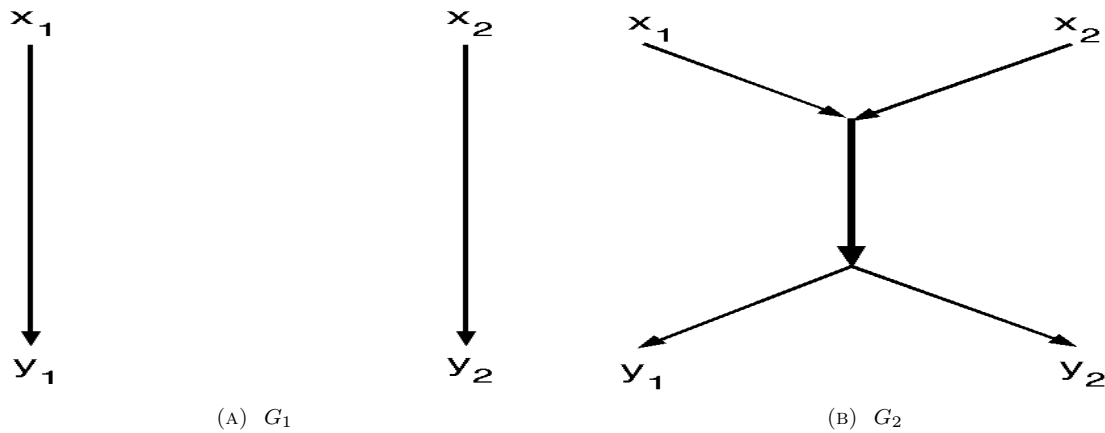


FIGURE 15. Unlike a traditional transport system G_1 , a ramified transport system G_2 provides an exchange value.

10. APPLICATIONS IN MATHEMATICAL ECONOMICS

10.1. The exchange value embedded in a transport system [34]

In the study of ramified optimal transportation, we have focused on the cost value of a branching transport system in terms of its effectiveness in reducing transportation cost. In [34], we showed that there is another value, named as *exchange value*, embedded in some ramified transport systems. As an illustration, we consider a spacial economy with two goods located at two distinct points $\{x_1, x_2\}$ and two consumers living at two different locations $\{y_1, y_2\}$. The spacial distribution is shown in Figure 15. Suppose consumer 1 favors good 2 more than good 1. However, good 2 may be more expensive than good 1 for some reason such as a higher transportation fee. As a result, she buys good 1 despite the fact that it is not her favorite. On the contrary, consumer 2 favors good 1 but ends up buying good 2, as good 1 is more expensive than good 2 for him. Given this purchase plan,

a traditional transporter will ship the ordered items in a transport system like G_1 (see Fig. 15b). However, as shown in [24] *etc.*, a transport system like G_2 (see Fig. 15b) with some branching structure might be more cost efficient than G_1 , so one may save some transportation cost by using a transport system like G_2 instead of G_1 . Now, we observe another very interesting phenomenon about G_2 . When using this transport system, one can simply switch the items which leads to consumer 1 getting good 2 and consumer 2 receiving good 1. This exchange of items makes both consumers better off since they both get what they prefer. More important, no extra transportation cost is incurred during this exchange process. In other words, a ramified transport system like G_2 may possess an exchange value, which cannot be found in other transport systems like G_1 .

In [34], we proposed an explicit valuation formula to measure this exchange value for a given compatible transport system. Suppose there are k sources of different goods which could be purchased by ℓ consumers distributed on a compact convex subset X of \mathbb{R}^m . Each source $x_i \in X$ supplies only one type of goods, $i = 1, \dots, k$. Each consumer j located at $y_j \in X$ derives utility from consuming k goods according to a utility function $u_j : \mathbb{R}_+^k \rightarrow \mathbb{R} : (q_{1j}, \dots, q_{kj}) \mapsto u_j$, $j = 1, \dots, \ell$, where $u_j : \mathbb{R}_+^k \rightarrow \mathbb{R}$ is *continuous, concave and increasing*, $j = 1, \dots, \ell$. Each consumer j has an initial wealth $w_j > 0$ and faces a price vector $p_j = (p_{1j}, \dots, p_{kj}) \in \mathbb{R}_{++}^k$, $j = 1, \dots, \ell$. We denote this economy as

$$\mathcal{E} = (U, P, W; x, y). \quad (10.1)$$

In the classical consumer's decision problem (see *e.g.* [14]), each consumer j will choose a utility maximizing consumption plan given the price p_j and wealth w_j . More precisely, the consumption plan \bar{q}_j is derived from the following utility maximizing problem:

$$\bar{q}_j \in \arg \max \{u_j(q_j) \mid q_j \in \mathbb{R}_+^k, p_j \cdot q_j \leq w_j\}. \quad (10.2)$$

Given the continuity and concavity of u_j , we know this problem has a solution.

Let $\bar{q} \in \text{Plan}(\mathbf{a}, \mathbf{b})$ be the initial plan given by (10.2). Denote

$$\Omega(\bar{q}) = \{G \in \text{Path}(\mathbf{a}, \mathbf{b}) \mid (G, \bar{q}) \text{ is compatible}\}. \quad (10.3)$$

For each probability measure $q = (q_{ij}) \in \mathcal{P}(X \times X)$, we define

$$S(q) = \sum_{j=1}^{\ell} e_j(p_j, u_j(q_j)) = \sum_{j=1}^{\ell} \min \{p_j \cdot t_j \mid t_j \in \mathbb{R}_+^k, u_j(t_j) \geq u_j(q_j)\}, \quad (10.4)$$

where $q_j = (q_{1j}, q_{2j}, \dots, q_{kj})$ for each $j = 1, \dots, \ell$. Here, $S(q)$ represents the least total expenditure for each individual j to reach utility level $u_j(q_j)$.

Each transport plan in the set

$$\mathcal{F}_G = \left\{ q \in \mathcal{P}(X \times X) \mid \begin{array}{l} q \text{ is compatible with } G \\ u_j(q_j) \geq u_j(\bar{q}_j), j = 1, \dots, \ell. \end{array} \right\} \quad (10.5)$$

is called a feasible plan for G , and the set \mathcal{F}_G is called the feasible set of G .

Let \mathcal{E} be an economy as in (10.1). For each transport path $G \in \Omega(\bar{q})$, we define the exchange value of G by

$$\mathcal{V}(G; \mathcal{E}) = \max_{q \in \mathcal{F}_G} S(q) - S(\bar{q}), \quad (10.6)$$

where S is given by (10.4).

Example 10.1. Let's return to the example discussed in introduction. More precisely, suppose $u_1(q_{11}, q_{21}) = q_{11} + 3q_{21}$, $w_1 = 1/2$, $p_1 = (1, 6)$ and $u_2(q_{12}, q_{22}) = 3q_{12} + q_{22}$, $w_2 = 1/2$, $p_2 = (6, 1)$. By solving (10.2), we find $\bar{q}_1 = (1/2, 0)$ and $\bar{q}_2 = (0, 1/2)$. This gives the initial plan

$$\bar{q} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

Now, solving expenditure minimization problems in (10.4) yields $e_1(p_1, \tilde{u}_1) = \tilde{u}_1$, $e_2(p_2, \tilde{u}_2) = \tilde{u}_2$, and thus

$$S(q) = e_1(p_1, u_1(q_1)) + e_2(p_2, u_2(q_2)) = u_1(q_1) + u_2(q_2)$$

for each probability measure $q \in \mathcal{P}(X \times X)$. Now, we find the exchange value embedded in the transport systems G_1 and G_2 as given in Figure 15.

- G_1 : the associated feasible set is $\mathcal{F}_{G_1} = \{\bar{q}\}$, and thus the exchange value of G_1 is

$$\mathcal{V}(G_1) = \max_{q \in \mathcal{F}_{G_1}} S(q) - S(\bar{q}) = S(\bar{q}) - S(\bar{q}) = 0.$$

- G_2 : the associated feasible set is

$$\begin{aligned} \mathcal{F}_{G_2} &= \left\{ q = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} \in \mathcal{P}(X \times X) \left| \begin{array}{l} q_{11} + q_{12} = 1/2, q_{21} + q_{22} = 1/2, \\ q_{11} + q_{21} = 1/2, \\ q_{11} + 3q_{21} \geq u_1(\bar{q}_1) = 1/2, \\ 3q_{12} + q_{22} \geq u_2(\bar{q}_2) = 1/2. \end{array} \right. \right\} \\ &= \left\{ q = \begin{pmatrix} q_{11} & 1/2 - q_{11} \\ 1/2 - q_{11} & q_{11} \end{pmatrix} \left| \begin{array}{l} q_{11} \leq 1/2 \\ q_{11} \geq 0 \end{array} \right. \right\} \end{aligned}$$

Thus, we have the following exchange value

$$\begin{aligned} \mathcal{V}(G_2) &= \max_{q \in \mathcal{F}_{G_2}} S(q) - S(\bar{q}) \\ &= \max_{q \in \mathcal{F}_{G_2}} \{(q_{11} + 3q_{21}) + (3q_{12} + q_{22})\} - 1 \\ &= \max_{0 \leq q_{11} \leq \frac{1}{2}} \{(q_{11} + 3(1/2 - q_{11})) + (3(1/2 - q_{11}) + q_{11})\} - 1 \\ &= \max_{0 \leq q_{11} \leq \frac{1}{2}} \{3 - 4q_{11}\} - 1 = 2. \end{aligned}$$

The exchange value is always nonnegative and bounded from above. It is jointly determined by transport structures, preferences and prices. Each of these factors may lead to a zero exchange value under very rare situations. More precisely, when the structure of the transport system yields a singleton feasible set \mathcal{F}_G or the utility functions are merely quantity dependent, or price vectors are collinear across consumers, the exchange value is zero. However, under many common situations, there exists a positive exchange value for a ramified transport system. For instance, there exists a positive exchange value in any of the following two cases:

- The feasible set \mathcal{F}_G is non-singleton, and each utility function $u_j : \mathbb{R}_+^k \rightarrow \mathbb{R}$ is homogeneous of degree $\beta_j > 0$, $(u_j(q_j))^{\frac{1}{\beta_j}}$ is concave in q_j satisfying the condition

$$(u_j((1 - \lambda_j)\tilde{q}_j + \lambda_j\hat{q}_j))^{\frac{1}{\beta_j}} > (1 - \lambda_j)(u_j(\tilde{q}_j))^{\frac{1}{\beta_j}} + \lambda_j(u_j(\hat{q}_j))^{\frac{1}{\beta_j}} \tag{10.7}$$

for each $\lambda_j \in (0, 1)$, and any non-collinear $\tilde{q}_j, \hat{q}_j \in \mathbb{R}_+^k$.

- The transport systems are of ramified structures with some order conditions on prices and non-degeneracy conditions on the utility.

When designing a transport system G , it is reasonable to consider both the minimization of the transportation cost $\mathbf{M}_\alpha(G)$ and the maximization of the embedded exchange value $\mathcal{V}(G)$. So, we propose the following minimization problem.

Problem 10.2. Given two atomic probability measures \mathbf{a} and \mathbf{b} on X in an economy \mathcal{E} given by (10.1), find a minimizer of

$$\mathbf{H}_{\alpha, \sigma}(G) := \mathbf{M}_\alpha(G) - \sigma \mathcal{V}(G) \tag{10.8}$$

among all $G \in \Omega(\bar{q})$, where $\Omega(\bar{q})$ is given by (10.3), and $\alpha \in [0, 1)$ and $\sigma \geq 0$ are fixed constants.

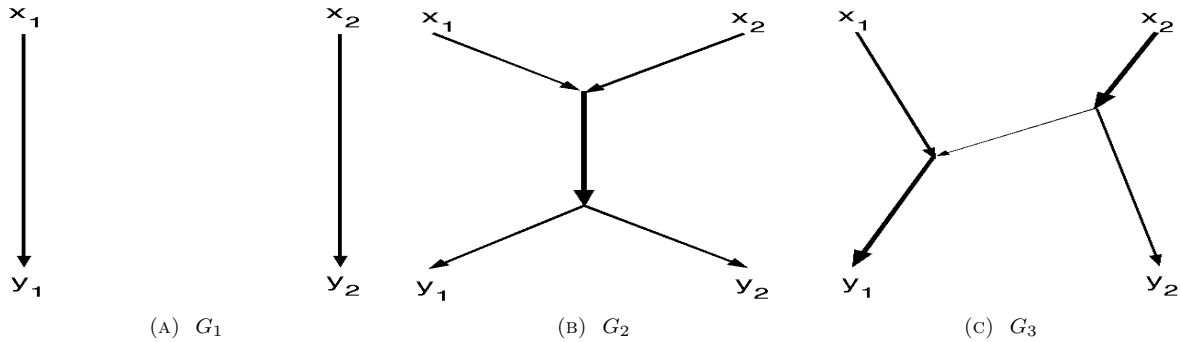


FIGURE 16. Three topologically different transport systems.

When the utility functions are merely quantity dependent or when price vectors are collinear across consumers, the exchange value of any $G \in \Omega(\bar{q})$ is always zero. In these cases, $\mathbf{H}_{\alpha,\sigma}(G) = \mathbf{M}_\alpha(G)$ for any σ . Thus, the study of $\mathbf{H}_{\alpha,\sigma}$ coincides with that of \mathbf{M}_α . However, as seen in the previous section, it is quite possible that $\mathbf{H}_{\alpha,\sigma}$ does not agree with \mathbf{M}_α on $\Omega(\bar{q})$ for $\sigma > 0$ in a general economy \mathcal{E} .

As \mathcal{V} is topologically invariant, many results that can be found in literature about \mathbf{M}_α still hold for $\mathbf{H}_{\alpha,\sigma}$. For instance, the Melzak algorithm for finding an \mathbf{M}_α minimizer ([16]) in a fixed topological class still applies to $\mathbf{H}_{\alpha,\sigma}$ because $\mathcal{V}(G)$ is simply a constant within each topological class. Also, as the balance equation (1.3) still holds, one can still calculate angles between edges at each vertex using existing formulas ([24]), and then get a universal upper bound on the degree of vertices on an optimal $\mathbf{H}_{\alpha,\sigma}$ path.

However, due to the existence of exchange value, one may possibly favor an optimal $\mathbf{H}_{\alpha,\sigma}$ path instead of the usual optimal \mathbf{M}_α path when designing a transport system. The topological type of the optimal $\mathbf{H}_{\alpha,\sigma}$ path may differ from that of the optimal \mathbf{M}_α path. This observation is illustrated by the following example.

Example 10.3. Let us consider the transportation from two sources to two consumers. If we only consider minimizing \mathbf{M}_α transportation cost, each of the three topologically different types shown in Figure 16 may occur. However, when σ is sufficiently large, only G_2 in Figure 16b may be selected under suitable conditions of u and p . This is because G_2 has a positive exchange value which does not exist in either G_1 or G_3 .

10.2. The ramified optimal allocation problem [35]

One of the lasting interests in economics is to study the optimal resource allocation in a spatial economy. For instance, the well known Monge–Kantorovich transport problem aims at finding an efficient allocation plan or map for transporting some commodity from factories to households. In the work [35], we proposed an optimal resource allocation problem where a social planner aims at finding an optimal allocation plan as well as an associated optimal allocation path to minimize the overall cost of transporting commodity from factories to households.

In the standard transport problems, *e.g.* Monge–Kantorovich or ramified optimal transportation, one typically assumes an exogenous fixed distribution on both sources and targets.

However, in many resource allocation practices, the distribution on either sources or targets is not pre-determined but rather determined endogenously. For instance, in a *production allocation* problem, suppose there are k factories and ℓ households located in different places in some area. The demand for some commodity from each household is fixed. Nevertheless, the allocation of production among factories is not pre-determined but rather depends on the distribution of demands among households as well as their relative locations to factories. A planner needs to make an efficient allocation plan of production over these k factories to meet given demands from these ℓ households. With ramified transportation, the transportation cost of each production plan is determined by an associated optimal transport path from factories to households. Consequently, the planner needs to find an optimal production plan as well as an associated optimal transport path to minimize

the total cost of distributing commodity from factories to households. We named this type of allocation problem the *ramified optimal allocation* problem. In this problem, the distribution of production over factories is not pre-determined as in either Monge–Kantorovich or ramified transport problem, but endogenously determined by the distribution of demands from households as well as their relative locations to factories. In other words, the atomic measure \mathbf{a} which represents the k factories can have varying production level m_i at each factory i .

We now describe this problem in more details. In a spatial economy, there are k factories and ℓ households located at $\mathbf{x} = \{x_1, x_2, \dots, x_k\}$ and $\{y_1, y_2, \dots, y_\ell\}$ in some area X , where X is a compact convex subset of a Euclidean space \mathbb{R}^m . In this model economy, there is only one commodity, and each household $j = 1, \dots, \ell$ has a fixed demand $n_j > 0$ for the commodity. We can thus represent the ℓ households as an atomic measure on X by

$$\mathbf{b} = \sum_{j=1}^{\ell} n_j \delta_{y_j}. \tag{10.9}$$

Without loss of generality, we may assume that \mathbf{b} is a probability measure.

Given \mathbf{x} and \mathbf{b} as above, an allocation plan from \mathbf{x} to \mathbf{b} is a probability measure q on $X \times X$ such that

$$q = \sum_{i=1}^k \sum_{j=1}^{\ell} q_{ij} \delta_{(x_i, y_j)} \text{ with } q_{ij} \geq 0 \text{ and } \sum_{i=1}^k q_{ij} = n_j \text{ for each } j.$$

Denote $\text{Plan}[\mathbf{x}, \mathbf{b}]$ as the set of all allocation plans from \mathbf{x} to \mathbf{b} . Note that any allocation plan $q \in \text{Plan}[\mathbf{x}, \mathbf{b}]$ corresponds to a transport plan q from $\mathbf{a}(q)$ to \mathbf{b} , where

$$\mathbf{a}(q) := \sum_{i=1}^k m_i(q) \delta_{x_i}, \text{ with } m_i(q) = \sum_{j=1}^{\ell} q_{ij}, \quad i = 1, \dots, k, \tag{10.10}$$

is the probability measure supported on \mathbf{x} representing k factories.

For any allocation plan $q \in \text{Plan}[\mathbf{x}, \mathbf{b}]$ and $\alpha \in [0, 1)$, the ramified transportation cost of q is defined by

$$\mathbf{T}_\alpha(q) := \min \{ \mathbf{M}_\alpha(G) : G \in \text{Path}(\mathbf{a}(q), \mathbf{b}), (G, q) \text{ compatible} \}. \tag{10.11}$$

Now, the ramified optimal allocation problem is: *Given \mathbf{x} and \mathbf{b} as above, minimize $\mathbf{T}_\alpha(q)$ among all allocation plans $q \in \text{Plan}[\mathbf{x}, \mathbf{b}]$.*

It is easy to prove that there exists a solution to the ramified optimal allocation problem. Moreover,

$$\min_{q \in \text{Plan}[\mathbf{x}, \mathbf{b}]} \mathbf{T}_\alpha(q) = \min \left\{ d_\alpha \left(\sum_{i=1}^k m_i \delta_{x_i}, \mathbf{b} \right) : m_i \geq 0, \sum_{i=1}^k m_i = 1 \right\}. \tag{10.12}$$

Similarly, given \mathbf{x} and \mathbf{b} as above, an allocation path from \mathbf{x} to \mathbf{b} is a transport path $G \in \text{Path}(\mathbf{a}, \mathbf{b})$ for some atomic probability measure \mathbf{a} supported on \mathbf{x} . Denote $\text{Path}[\mathbf{x}, \mathbf{b}]$ as the set of all allocation paths from \mathbf{x} to \mathbf{b} . By the definition of d_α , equation (10.12) can be alternatively written as

$$\min \{ \mathbf{T}_\alpha(q) : q \in \text{Plan}[\mathbf{x}, \mathbf{b}] \} = \min \{ \mathbf{M}_\alpha(G) : G \in \text{Path}[\mathbf{x}, \mathbf{b}] \}, \tag{10.13}$$

which shows that the ramified optimal allocation problem corresponds to a problem of finding an optimal allocation path.

The next lemma presents a key property of an optimal allocation path, as shown in Figure 17.

Lemma 10.4. *Let $G \in \text{Path}[\mathbf{x}, \mathbf{b}]$ be an optimal allocation path. Then, for any $i \neq s \in \{1, \dots, k\}$, x_i and x_s do not belong to the same connected component of G .*

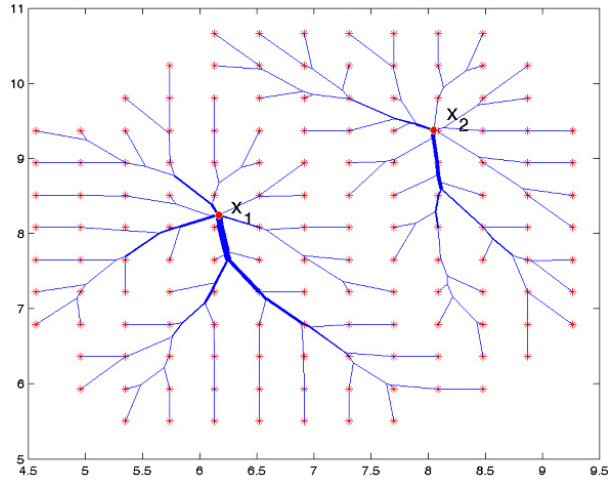


FIGURE 17. A partition in households and allocation path.

The above lemma says that no two factories will be connected on any optimal allocation path. Alternatively speaking, on an optimal allocation path, each single household will receive her commodity from only one factory, *i.e.*, each household is assigned to one factory. This result is attributed to the transport economy of scale underlying ramified transportation theory: As seen in Section 2, an $\alpha \in [0, 1)$ implies the existence of transport economy of scale with transporting in groups being more cost efficient than transporting separately. Any allocation path on which some single household receives commodity from two factories can not be optimal because the planner would be able to reduce transportation cost by transferring production of one factory to the other. This transfer makes the benefit of transport economy of scale more likely to be realized as commodity for this household is transported in a larger scale on the path. The result that each household is assigned to one factory on an optimal allocation path motivates the following notion of assignment map.

An *assignment map* is a function $S : \{1, \dots, \ell\} \rightarrow \{1, \dots, k\}$. Let $\text{Map}[\ell, k]$ be the set of all *assignment maps*. For any assignment map $S \in \text{Map}[\ell, k]$ and $\alpha \in [0, 1)$, define

$$\mathbf{E}_\alpha(S; \mathbf{x}, \mathbf{b}) := \sum_{i=1}^k d_\alpha(\mathbf{a}_i, \mathbf{b}_i),$$

where $\mathbf{a}_i = \left(\sum_{j \in S^{-1}(i)} n_j\right) \delta_{x_i}$ and $\mathbf{b}_i = \sum_{j \in S^{-1}(i)} n_j \delta_{y_j}$.

A main result in [35] is given by the following theorem:

Theorem 10.5. *Given a subset $\mathbf{x} = \{x_1, x_2, \dots, x_k\}$ in X , an atomic probability measure \mathbf{b} as in (10.9), and a parameter $\alpha \in [0, 1)$.*

- (1) *An allocation plan $q \in \text{Plan}[\mathbf{x}, \mathbf{b}]$ is optimal if and only if there exists an optimal assignment map $S \in \text{Map}[\ell, k]$ such that $q = q_S$, where $q_S = \sum_{j=1}^\ell n_j \delta_{(x_{S(j)}, y_j)}$.*
- (2) *An allocation path $G \in \text{Path}[\mathbf{x}, \mathbf{b}]$ is optimal if and only if there exists an optimal assignment map $S \in \text{Map}[\ell, k]$ such that $G = G_S$, where $G_S = \sum_{i=1}^k G_i \in \text{Path}[\mathbf{x}, \mathbf{b}]$ with each $G_i \in \text{Path}(\mathbf{a}_i, \mathbf{b}_i)$ being an optimal transport path.*
- (3) *Moreover,*

$$\min_{q \in \text{Plan}[\mathbf{x}, \mathbf{b}]} \mathbf{T}_\alpha(q) = \min_{S \in \text{Map}[\ell, k]} \mathbf{E}_\alpha(S; \mathbf{x}, \mathbf{b}) = \min_{G \in \text{Path}[\mathbf{x}, \mathbf{b}]} \mathbf{M}_\alpha(G). \tag{10.14}$$

The equivalence relation between an optimal allocation plan and an optimal assignment map found here has an analogous counterpart in Monge–Kantorovich problems and plays a crucial role there, but has never been observed in the current literature on ramified transport problems. Because of the equivalence, we can instead focus attention on studying the properties of optimal assignment maps in the ramified optimal allocation problem. We develop methods of marginal transportation analysis and projectional analysis to study geometric properties of optimal assignment maps. An important application of the properties of optimal assignment maps is that they can shed light on the search for those maps. We develop a search method utilizing these properties in the notion of a state matrix. In some non-trivial cases, it is shown that this method can exactly find an optimal assignment map.

11. LANDSCAPE FUNCTIONS ASSOCIATED WITH TRANSPORT PATHS FROM MULTIPLE SOURCES [36]

11.1. Background

In the study of river basins, geophysics yields problems very similar to ramified optimal transportation. While studying the configuration of a river basin, the main objects are the landscape function giving the altitude of any point of the region and a river channel network which is the datum of all the streams that concur to bring water to lakes. In a seminal paper [17] of Santambrogio, he introduced the concept “landscape function” associated with an optimal transport Path from a Dirac mass (representing a single lake) to another probability measure (representing the region). In the context of ramified optimal transportation, the landscape function at a point can be viewed as the marginal transportation cost from the point to the common single source. Independently, the author has also used analogous ideas of the landscape function in applications such as [33] in modeling diffusion-limited aggregation driven by optimal transportation and [35] about the marginal analysis of optimal assignment maps.

In the work of Santambrogio, it was assumed that the initial measure is a Dirac mass (*i.e.* a single common source). In the article [36], we generalized the concept of landscape function by allowing multiple sources rather than a single common source. For simplicity, we only considered the discrete version, *i.e.* transport paths (which are weighted directed graphs) between atomic measures. The continuous version will be left for future exploration.

11.2. Landscape functions

Let $\mathbf{a} = \delta_S$ be the initial source and the target \mathbf{b} be an atomic measure on X . Let $P = \{V(P), E(P), \omega\}$ be a transport path from \mathbf{a} to \mathbf{b} . Then for any x in the support of P , there is a unique polyhedral curve γ_x on P from the initial source S to x . In this case, Santambrogio’s landscape function $z(x)$ associated with P can be simply written as

$$z(x) = \int_{\gamma_x} \theta(s)^{\alpha-1} d\mathcal{H}^1(s)$$

where \mathcal{H}^1 represents the 1-dimensional Hausdorff measure, and $\theta(s)$ represents the mass flowing through the point $\gamma_x(s)$.

In general, an atomic measure \mathbf{a} is not necessarily a single source but of the form in (1.2) that contains multiple-sources. Here, we study a multiple-sources version of landscape functions associated with a transport Path.

Definition 11.1. Given two measures \mathbf{a}, \mathbf{b} as in (1.2), let $P = \{V(P), E(P), \omega\}$ be an acyclic transport path from \mathbf{a} to \mathbf{b} with $V(P) = \{v_1, v_2, \dots, v_J\}$, $E(P) = \{e_1, e_2, \dots, e_K\}$, and $0 \leq \alpha < 1$. A function $Z : V(P) \rightarrow \mathbb{R}$ is called an α -landscape function associated with P if for each edge $e \in E(P)$, it holds that

$$Z(e^+) - Z(e^-) = \omega(e)^{\alpha-1} \text{length}(e). \tag{11.1}$$

Using matrix notations the system (11.1) of linear equations on Z becomes

$$ZQ(G_P) = D \tag{11.2}$$

where $Z = [Z(v_1), Z(v_2), \dots, Z(v_J)]$, $Q(G_P)$ is the incidence matrix of the underlying graph G_P of P and $D = [\omega(e_1)^{\alpha-1} \text{length}(e_1), \dots, \omega(e_K)^{\alpha-1} \text{length}(e_K)]$.

One can show that the dimension of the solution space of the system (11.2) is equal to the connected components of the underlying graph G_P . In particular, for a connected transport path P , its α -landscape function is unique up to a constant. This agrees with the motivation that the landscape function represents the elevation of the landscape.

Landscape functions give another formula for the transport cost $M_\alpha(P)$:

Proposition 11.2. *For any α -landscape function Z associated with a transport path $P \in \text{path}(\mathbf{a}, \mathbf{b})$, it holds that*

$$\int_X Z d(\mathbf{b} - \mathbf{a}) = M_\alpha(P).$$

In particular, if P is an α -optimal transport path from \mathbf{a} to \mathbf{b} , then

$$\int_X Z d(\mathbf{b} - \mathbf{a}) = d_\alpha(\mathbf{a}, \mathbf{b}).$$

11.3. p -harmonic functions on directed graphs

It turns out that the study of α -landscape functions is closely related to the study of p -harmonic functions on directed graphs where $p = \frac{\alpha}{\alpha-1}$ is non-positive. More precisely, let G be an acyclic directed graph with a vertex set $V(G)$ and an edge set $E(G)$ of directed edges.

Definition 11.3. For any two vertices $v, \tilde{v} \in V(G)$, we say $v \prec \tilde{v}$ if there exists a list of edges $\{e_{i_1}, e_{i_2}, \dots, e_{i_k}\}$ in $E(G)$ with $e_{i_h}^+ = e_{i_{h+1}}^-$ for $h = 1, 2, \dots, k-1$, and $e_{i_1}^- = v, e_{i_k}^+ = \tilde{v}$.

Note that since G is acyclic, the partial order \prec is well-defined on $V(G)$. In particular, $e^- \prec e^+$ for any edge $e \in E(G)$.

Definition 11.4. Let \tilde{V} be a subset of $V(G)$. A function $u : \tilde{V} \rightarrow \mathbb{R}$ is monotone increasing with respect to G if for any $x, y \in \tilde{V}$ with $x \prec y$, it holds that $u(x) < u(y)$.

Let \mathcal{F}_G be the family of all monotone increasing functions $u : V(G) \rightarrow \mathbb{R}$ with respect to G . For instance, any landscape function Z associated with an acyclic transport path P is monotone increasing with respect to the underlying graph G_P .

For each $u \in \mathcal{F}_G$ and $p \leq 0$, the p -energy of u is given by

$$\mathbf{E}_p(u) = \sum_{e \in E(G)} |\nabla u(e)|^p \text{length}(e) \tag{11.3}$$

where for each $e \in E(G)$ with unit directional vector \vec{e} ,

$$\nabla u(e) = \frac{u(e^+) - u(e^-)}{\text{length}(e)} \vec{e}.$$

We are interested in minimizing $\mathbf{E}_p(u)$ among $u \in \mathcal{F}_G$ with a Dirichlet boundary condition. Let ∂G be a subset of $V(G)$ such that

$$\left\{ v \in V(G) : \begin{array}{l} \text{either there is no edge } e \in E(G) \text{ with } e^+ = v \\ \text{or there is no edge } e \in E(G) \text{ with } e^- = v \end{array} \right\} \subseteq \partial G \subseteq V(G).$$

In other words, ∂G contains all source and sink vertices in $V(G)$, and may contain some other vertices. We view ∂G as the boundary set of $V(G)$.

Problem 11.5. Suppose $u_0 : \partial G \rightarrow \mathbb{R}$ is monotone increasing with respect to G , and $p < 0$. Minimize

$$\mathbf{E}_p(u) = \sum_{e \in E(G)} |\nabla u(e)|^p \text{length}(e)$$

among all $u \in U = \{u : V(G) \rightarrow \mathbb{R} \mid u \in \mathcal{F}_G \text{ and } u = u_0 \text{ on } \partial G\}$.

By identifying $u \in \mathcal{F}_G$ with the value $(u(v_1), u(v_2), \dots, u(v_I))$, it turns out that this minimization problem is equivalent to a minimization of some strictly convex function f on a nonempty convex domain Ω in \mathbb{R}^I . The “nonempty convex” property of Ω is guaranteed by the monotonicity of u_0 . The associated Euler–Lagrange equation is

$$\sum_{\substack{e \in E(G) \\ e^+ = v_i}} |\nabla u(e)|^{p-1} = \sum_{\substack{e \in E(G) \\ e^- = v_i}} |\nabla u(e)|^{p-1} \tag{11.4}$$

at each $v_i \in V(G) \setminus \partial G$. Using the discrete version of the divergence notation:

$$\text{div}(\vec{V})(v) = \sum_{\substack{\text{either } e^+ = v \\ \text{or } e^- = v}} \vec{V}(e) \cdot \vec{e}$$

for any vector field $\vec{V} : E(G) \rightarrow \mathbb{R}^m$, equation (11.4) can be expressed as the p -Laplace equation on the graph G :

$$\text{div}(|\nabla u|^{p-2} \nabla u)(v) = 0 \tag{11.5}$$

for any $v \in V(G) \setminus \partial G$.

A solution to the p -Laplace equation (11.5) in U is called a p -harmonic function on the graph G . By the strict convexity of f , any p -harmonic function u is an \mathbf{E}_p -minimizer in \mathcal{F}_G with respect to its boundary datum.

The following theorem describes an equivalent relationship between landscape functions and p -harmonic maps:

Theorem 11.6. Let $\alpha \in [0, 1)$ and $p = \frac{\alpha}{\alpha-1} \leq 0$ be the conjugate of α .

1. If Z is an α -landscape function associated with an acyclic transport path P , then Z is a p -harmonic function on the underlying graph G_P . Moreover, $\mathbf{M}_\alpha(P) = \mathbf{E}_p(Z)$.
2. Conversely, if u is a p -harmonic function on an acyclic graph G , then u is an α -landscape function associated with an acyclic transport path P with G being its underlying graph. Moreover, $\mathbf{E}_p(u) = \mathbf{M}_\alpha(P)$.

11.4. Landscape functions associated with optimal transport paths

Proposition 11.7. Suppose $P \in \text{Path}(\mathbf{a}, \mathbf{b})$ is an α -optimal transport path. For any α -landscape function Z associated with P , it holds that at each vertex $v \in V(P) \setminus \{x_1, \dots, x_k, y_1, \dots, y_\ell\}$,

$$\sum_{e^+ = v} \omega(e) \nabla Z(e) = \sum_{e^- = v} \omega(e) \nabla Z(e) \tag{11.6}$$

and

$$\sum_{e^+ = v} |\nabla Z(e)|^{p-1} \nabla Z(e) = \sum_{e^- = v} |\nabla Z(e)|^{p-1} \nabla Z(e) \tag{11.7}$$

for $p = \frac{\alpha}{\alpha-1}$.

Remark 11.8. For any transport path P , by (1.3), mass is conserved at each interior vertex of P . When the transport path P is optimal, by (11.6), momentum is also conserved at each interior vertex when viewing $\nabla Z(e) = w(e)^{\alpha-1}\vec{e}$ as the velocity of a moving particle of mass $w(e)$ on each e .

The following proposition says that any α -landscape function associated with an α -optimal transport path is Lipschitz. Before stating the proposition, we first extend the domain of a landscape function $Z : V(P) \rightarrow \mathbb{R}$ to the support of P by linear extensions of Z on each edge of P .

Proposition 11.9. *Suppose $P \in \text{Path}(\mathbf{a}, \mathbf{b})$ is an α -optimal transport path for some $\alpha \in (0, 1)$. Let Z be an α -landscape function associated with P . Then, for any x, y on the same connected components of the support of G_P , it holds that*

$$|Z(x) - Z(y)| \leq \frac{1}{\alpha} \sigma^{\alpha-1} \|x - y\|$$

where $\sigma = \min_{e \subseteq \gamma} \omega(e)$, and γ is the unique curve on P from x to y . In particular, when P is connected, let $\sigma_P = \min_{e \in E(P)} \omega(e)$, then

$$|Z(x) - Z(y)| \leq \frac{1}{\alpha} (\sigma_P)^{\alpha-1} \|x - y\|$$

for any x, y on the support of G_P .

REFERENCES

- [1] L. Ambrosio, Lecture notes on Optimal Transport Problems. Mathematical Aspects of Evolving Interfaces (Funchal, 2000). In vol. 1812 of *Lect. Notes Math.* Springer, Berlin (2003) 1–52.
- [2] Y. Brenier, Décomposition polaire et ré arrangement monotone des champs de vecteurs [Polar decomposition and increasing rearrangement of vector fields]. *C. R. Acad. Sci. Paris Sér. I Math.* **305** (1987) 805–808.
- [3] D. Burago, Y. Burago and S. Ivanov, A Course in Metric Geometry. American Mathematical Society (2001).
- [4] L.A. Caffarelli, M. Feldman and R.J. McCann, Constructing optimal maps for Monge’s transport problem as a limit of strictly convex costs. *J. Amer. Math. Soc.* **15** (2002) 1–26.
- [5] R. Coifman and G. Weiss, Analyse Harmonique Non-Commutative sur Certains Espaces Homogenes. Vol. 242 of *Lect. Notes Math.* Springer-Verlag (1971).
- [6] G. Devillanova and S. Solimini, On the dimension of an irrigable measure. *Rend. Semin. Mat. Univ. Padova* **117** (2007) 1–49.
- [7] L.C. Evans and W. Gangbo, Differential equations methods for the Monge–Kantorovich mass transfer problem. *Mem. Amer. Math. Soc.* **137** (1999) 653.
- [8] H. Federer, Geometric measure theory. In vol. 153 of *Die Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, New York Inc. (1969).
- [9] M. Feldman and R.J. McCann, Uniqueness and transport density in Monge’s mass transportation problem. *Calc. Var. Partial Differ. Equ.* **15** (2002) 81–113.
- [10] W. Fleming, Flat chains over a finite coefficient group. *Trans. Amer. Math. Soc.* **121** (1966) 160–186.
- [11] W. Gangbo and R.J. McCann, The geometry of optimal transportation. *Acta Math.* **177** (1996) 113–161.
- [12] T.C. Halsey, Diffusion-Limited Aggregation: A Model for Pattern Formation. *Phys. Today* **53** (2000) 36–41.
- [13] F. Maddalena, S. Solimini and J.M. Morel, A variational model of irrigation patterns, *Interfaces and Free Boundaries* **5** (2003) 391–416.
- [14] A. Mas-Colell, M. Whinston and J. Green, Microeconomic Theory. Oxford University Press, New York (1995).
- [15] P. Meakin, Progress in DLA Research. *Physica D* **86** (1995) 104–112.
- [16] Z.A. Melzak, On the problem of Steiner. *Canad. Math. Bull.* **4** (1961) 143–148.
- [17] F. Santambrogio, Optimal Channel Networks, Landscape Function and Branched Transport. *Interfaces Free Bound.* **9** (2007) 149–169.
- [18] L. Simon, Lectures on geometric measure theory. In vol. 3 of *Proc. Centre Math. Anal.* Australian National University (1983).
- [19] D.P. Thierry and R. Hardt, Size minimization and approximating problems. *Calc. Var. Partial Differ. Equ.* **17** (2003) 405–442.
- [20] C. Villani, Topics in Mass Transportation. Vol. 58 of *AMS Grad. Stud. Math.* **58** (2003).
- [21] C. Villani, Optimal Transport: Old and New. *Grundlehren der mathematischen Wissenschaften* (2009).
- [22] B. White, Rectifiability of flat chains. *Ann. Math.* **150** (1999) 165–184.
- [23] T.A. Witten and L.M. Sander, Diffusion-Limited Aggregation, A Kinetic Critical Phenomenon. *Phys. Rev. Lett.* **47** (1981) 1400–1403.
- [24] Q. Xia, Optimal paths related to transport problems. *Commun. Contemp. Math.* **5** (2003) 251–279.
- [25] Q. Xia, Interior regularity of optimal transport paths. *Calc. Var. Partial Differential Equ.* **20** (2004) 283–299.

- [26] Q. Xia, An application of optimal transport paths to urban transport networks. *Discr. Contin. Dyn. Syst., Supp.* (2005) 904–910.
- [27] Q. Xia, The formation of tree leaf. *ESAIM: COCV* **13** (2007) 359–377.
- [28] Q. Xia, The geodesic problem in quasimetric spaces. *J. Geom. Anal.* **19** (2009) 452–479.
- [29] Q. Xia, Boundary regularity of optimal transport paths. *Adv. Calc. Var.* **4** (2011) 153–174.
- [30] Q. Xia, Numerical simulation of optimal transport paths. In vol. 1, *Proc. of the Second International Conference on Computer Modeling and Simulation ICCMS 2010* (2010) 521–525. DOI: [10.1109/ICCMS.2010.30](https://doi.org/10.1109/ICCMS.2010.30).
- [31] Q. Xia, Ramified optimal transportation in geodesic metric spaces. *Adv. Calc. Var.* **4** (2011) 277–307.
- [32] Q. Xia and A. Vershynina, On the transport dimension of measures. *SIAM J. Math. Anal.* **41** (2010) 2407–2430.
- [33] Q. Xia and D. Unger, Diffusion-limited aggregation driven by optimal transportation. *Fractals* **18** (2010) 1–7.
- [34] Q. Xia and S. Xu, The exchange value embedded in a transport system. *Appl. Math. Optim.* **62** (2010) 229–252.
- [35] Q. Xia and S. Xu, On the ramified optimal allocation problem. *Netw. heterog. Media* **8** (2013) 591–624.
- [36] Q. Xia, On landscape functions associated with transport paths. *Discr. Contin. Dyn. Syst. A* **34** (2014).
- [37] Q. Xia and C. Salafia, Transport efficiency of the human placenta. *J. Coupled Syst. Multiple Dyn.* **2** (2014).
- [38] Q. Xia, C. Salafia and M. Simon, Optimal transport and placental function. Vol. 17 of *Interdisciplinary Topics Appl. Math., Modeling and Computational Science. Springer Proc. Math. Stat.* Springer (2015). DOI: [10.1007/978-3-319-12307-3-73](https://doi.org/10.1007/978-3-319-12307-3-73).
- [39] Q. Xia and C. Salafia, Human placentas, Optimal transportation and Autism (submitted).
- [40] M. Yampolsky, C.M. Salafia and O. Shlakhter, Probability distributions of placental morphological measurements and origins of variability of placental shapes. *Placenta* **34** (2013) 493–6.