# THE GEODESIC PROBLEM IN NEARMETRIC SPACES

QINGLAN XIA

Abstract. In this article, we study the geodesic problem in a generalized metric space, in which the distance function satisfies a relaxed triangle inequality $d(x,y) \leq \sigma(d(x,z)+d(z,y))$ for some constant $\sigma \geq 1$, rather than the usual triangle inequality. Such a space is called a nearmetric space. We show that many well-known results in metric spaces (e.g. Ascoli-Arzelà theorem) still hold in nearmetric spaces. Moreover, we explore conditions under which a nearmetric will induce an intrinsic metric. As an example, we introduce a family of nearmetrics on the space of atomic probability measures. The associated intrinsic metrics induced by these nearmetrics coincide with the $d_\alpha$ metric studied early in [6]. Moreover, optimal transport paths between atomic probability measures turn out to be geodesics in these intrinsic metric spaces.

## 1. Introduction

This article aims at studying some classical analysis problems in semimetric spaces, in which the distance does not required to satisfy the triangle inequity. Researches on semimetric spaces are mainly carried out by topologist so far (see [2] and references there). Analysts have not shown enough interest in studying semimetric spaces, partially because of lacking some interesting modeling examples of semimetric spaces. Nevertheless, during the author's recent study of optimal transport path between probability measures, he observes that there exists a family of very interesting semimetrics on the space of atomic probability measures. These semimetrics satisfy a relaxed triangle inequality $d(x,y) \leq \sigma(d(x,z)+d(z,y))$ for some constant $\sigma \geq 1$, rather than the usual triangle inequality. Such semimetric spaces were called nearmetric spaces in [4]. Moreover, these family of nearmetrics induce a family of intrinsic metrics on the space of atomic probability measures. Furthermore, optimal transport paths studied in [6], [7],[8],[9] etc turn out to be geodesics in these induced metric spaces. This observation motivates us to study functions in nearmetric spaces in this article.

This article is organized as follows. In section 2, we first introduce the concept as well as some basic properties of nearmetric spaces, then we extend some well-known results (e.g. Ascoli-Arzelà theorem) about continuous functions in metric spaces to continuous functions in nearmetric spaces. After that, in section 3, we consider the geodesic problem in nearmetric spaces. We show that every continuous nearmetric will induce an intrinsic pseudometric on the space. In case that the nearmetric is nice enough (e.g. either "ideal" or "perfect" in the sense of Definition 2.5 or Definition 3.14), then the nearmetric will indeed induce an intrinsic metric. In

2000 *Mathematics Subject Classification.* Primary 54E25, 51F99; Secondary 49Q20, 90B18.
*Key words and phrases.* optimal transport path, nearmetric, semimetric, intrinsic metric.
This work is supported by an NSF grant DMS-0710714.

the end, we spend the last section in discussing our motivation example: optimal transport paths between atomic probability measures. We first introduce a family of nearmetrics on the space of atomic probability measures. Each of these nearmetric is both ideal and perfect, and thus it induces an intrinsic metric on the space of atomic probability measures. We showed that the $d_\alpha$-metrics introduced in [6] is simply the intrinsic metrics induced by these nearmetrics. Furthermore, each geodesic in these length spaces corresponds to an optimal transport path studied in [6].

## 2. Continuous maps in nearmetric spaces

### 2.1. Nearmetric Spaces.

**Definition 2.1.** *Let $X$ be any nonempty set. A function $J : X \times X \to \mathbb{R}$ is called a nearmetric if for any $x, y, z \in X$, we have*

1. *(non-negativity) $J(x,y) \geq 0$;*
2. *(identity of indiscernibles) $J(x,y) = 0$ if and only if $x = y$*
3. *(symmetry) $J(x,y) = J(y,x)$;*
4. *(relaxed triangle inequality) $J(x,y) \leq \sigma [J(x,z) + J(z,y)]$ for some constant $\sigma \geq 1$.*

*When $J$ is a nearmetric on $X$, the pair $(X, J)$ is called a nearmetric space. Let $\sigma(J)$ denote the smallest number $\sigma$ satisfying condition (4).*

Every metric space is clearly a nearmetric space with $\sigma = 1$.

**Example 2.2.** *Suppose $d$ is a metric on a nonempty set $X$. Then, for any $\beta > 1, \lambda \geq 0, \mu > 0$, $J(x,y) = \lambda d(x,y) + \mu d(x,y)^\beta$ is typically not a metric on $X$. However, $J$ defines a nearmetric on $X$ with $\sigma(J) \leq 2^{\beta-1}$. Indeed,*

$$\begin{aligned} J(x,y) &= \lambda d(x,y) + \mu d(x,y)^\beta \\ &\leq \lambda [d(x,z) + d(y,z)] + \mu [d(x,z) + d(y,z)]^\beta \\ &\leq \lambda [d(x,z) + d(y,z)] + 2^{\beta-1}\mu \left[d(x,z)^\beta + d(y,z)^\beta\right] \\ &\leq 2^{\beta-1} [J(x,z) + J(z,y)]. \end{aligned}$$

In section 4, we will provide a family of interesting nearmetrics on the space of atomic probability measures.

More generally, suppose $J$ is a distance function on $X$ satisfying conditions (1),(2),(3) in Definition 2.1. For each $n$, let $\sigma_n(J)$ be the smallest number $\sigma_n \geq 1$ satisfying

$$J(x_1, x_{n+1}) \leq \sigma_n \sum_{i=1}^{n} J(x_i, x_{i+1}), \tag{2.1}$$

for any $x_1, \cdots, x_{n+1} \in X$. In particular, $\sigma_1(J) = 1$ and $\sigma_2(J) = \sigma(J)$.

**Lemma 2.3.** *Suppose $(X, J)$ is a nearmetric space. Then, for each $n$,*

$$\sigma_n(J) \leq \sigma(J)^{n-1}.$$

*Proof.* We show this using the mathematical induction. It is trivial when $n = 1$ or 2. Then, from condition (4), we see that for any $n$ and any points $\{x_1, x_2, \cdots, x_n\}$

in $X$, we have

$$\begin{aligned} J(x_1,x_n) &\le \sigma(J)\left(J(x_1,x_{n-1})+J(x_{n-1},x_n)\right) \\ &\le \sigma(J)\left(\sigma(J)^{n-2}\sum_{i=1}^{n-2}J(x_i,x_{i+1})+J(x_{n-1},x_n)\right) \\ &\le \sigma(J)^{n-1}\sum_{i=1}^{n-1}J(x_i,x_{i+1}) \text{ since } \sigma(J)\ge 1. \end{aligned}$$

Therefore, $\sigma_n(J)\le\sigma(J)^{n-1}$ for all $n$. □

**Proposition 2.4.** *Suppose* $(X,J)$ *is a nearmetric space. Then, for each* $n$ *and* $m$ *in* $\mathbb{N}$*,*

$$\sigma_{nm}(J)\le\sigma_n(J)\,\sigma_m(J).$$

*Proof.* Note that, for any $\{x_1,x_2,\cdots,x_{mn+1}\}$ in $X$, from (2.1), we have

$$\begin{aligned} &J(x_1,x_{mn+1}) \\ \le\ & \sigma_n(J)\left(J(x_1,x_{m+1})+J(x_{m+1},x_{2m+1})+\cdots+J\left(x_{(n-1)m+1},x_{nm+1}\right)\right) \\ \le\ & \sigma_n(J)\left(\sigma_m(J)\sum_{i=1}^{m}J(x_i,x_{i+1})+\cdots+\sigma_m(J)\sum_{i=(n-1)m+1}^{nm}J(x_i,x_{i+1})\right) \\ =\ & \sigma_n(J)\,\sigma_m(J)\sum_{i=1}^{nm}J(x_i,x_{i+1}). \end{aligned}$$

Therefore,

$$\sigma_{nm}(J)\le\sigma_n(J)\,\sigma_m(J).$$

□

Clearly, $\sigma_n(J)$ is nondecreasing as $n$ increases. Thus, we define

$$\sigma_\infty(J):=\lim_n\sigma_n(J) \tag{2.2}$$

for any nearmetric $J$ on $X$.

**Definition 2.5.** *Suppose* $J$ *is a nearmetric on* $X$*. If* $\sigma_\infty(J)<\infty$*, then* $J$ *is called an ideal nearmetric on* $X$*.*

Note that $J$ is an ideal nearmetric if and only if for some $\sigma\ge 1$,

$$J(x,y)\le\sigma\sum_{i=1}^{n}J(x_i,x_{i+1}), \tag{2.3}$$

for any finitely many points $x_1,\cdots,x_{n+1}\in X$ with $x_1=x$, $x_{n+1}=y$. The smallest $\sigma$ satisfying (2.3) is just $\sigma_\infty(J)$.

A sequence $\{x_n\}$ is *convergent* to $x$ in a nearmetric space $(X,J)$ if $J(x_n,x)\to 0$, and we denote it by $x_n\xrightarrow{J}x$. A sequence $\{x_n\}$ is *Cauchy in* $(X,J)$ if for any $\epsilon>0$, there exists an $N\in\mathbb{N}$ such that $J(x_n,x_m)\le\epsilon$ for all $n,m\ge N$. Since $J(x_n,x_m)\le\sigma(J)\left(J(x_n,x)+J(x,x_m)\right)$, it follows that every convergent sequence in $(X,J)$ is a Cauchy sequence. If every Cauchy sequence in $(X,J)$ is convergent, then we say $J$ is a *complete* nearmetric on $X$. A nearmetric $J$ on $X$ always gives a topology on $X$ where a subset $A$ is closed if it contains every point $a \in X$ for which there is some sequence $a_i \in A$ with $\lim_{i\to\infty} J(a_i, a) = 0$.

**Definition 2.6.** *A nearmetric $J$ on $X$ is continuous if for any convergent sequences $x_n \xrightarrow{J} x$, $y_n \xrightarrow{J} y$, we have*

$$J(x_n, y_n) \to J(x, y), \text{ as } n \to \infty. \tag{2.4}$$

*If for any convergent sequences $x_n \xrightarrow{J} x$, $y_n \xrightarrow{J} y$, we have*

$$J(x, y) \le \liminf_n J(x_n, y_n), \tag{2.5}$$

*then we say $J$ is lower semicontinuous.*

For instance, suppose $J$ satisfies conditions (1),(2),(3) in Definition 2.1, and also the following condition

$$|J(x, y) - J(z, w)| \le \sigma (J(x, z) + J(w, y)) \tag{2.6}$$

for any $x, y, z, w \in X$ and some $\sigma \ge 1$. By setting $z = w$, we get $J(x, y) \le \sigma [J(x, z) + J(z, y)]$, and hence $J$ is a nearmetric on $X$. Also, since for each $n$,

$$|J(x_n, y_n) - J(x, y)| \le \sigma (J(x, x_n) + J(y, y_n)),$$

$J$ is automatically satisfying the continuous condition (2.4) in this case. When $J$ is indeed a metric on $X$, then (2.6) trivially holds.

2.2. **Continuous maps in nearmetric spaces.** In this section, we extend some well-known results (see for instance in [5] or [1]) about continuous maps in metric spaces to continuous maps in nearmetric spaces.

Suppose $(X, J)$ is a nearmetric space, and $K$ is a compact metric space with a metric $d_K$. A map $f : K \to (X, J)$ is *continuous* if $J(f(x_n), f(x)) \to 0$ in $X$ whenever $d_K(x_n, x) \to 0$ in $K$ as $n \to \infty$. A map $f : K \to (X, J)$ is *uniformly continuous* if for every $\epsilon > 0$, there exists a $\delta > 0$ such that $J(f(x), f(y)) \le \epsilon$ whenever $x, y \in K$ with $d_K(x, y) \le \delta$. A map $f : K \to (X, J)$ is *Lipschitz* if there exists a constant $C \ge 0$ such that

$$J(f(x), f(y)) \le C d_K(x, y)$$

for any $x, y \in K$. Let $C(K, (X, J))$ be the family of all continuous maps from $K$ to $(X, J)$, and $Lip(K, (X, J))$ be the family of all Lipschitz maps from $K$ to $(X, J)$.

**Proposition 2.7.** *Suppose $J$ is a continuous nearmetric on $X$. Then, every continuous map $f : K \to (X, J)$ is uniformly continuous.*

*Proof.* Suppose $f : K \to (X, J)$ is continuous. If $f$ is not uniformly continuous, then there exists an $\epsilon > 0$, and two sequences $\{x_n\}, \{y_n\}$ in $K$ such that $d(x_n, y_n) \le \frac{1}{n}$, but $J(f(x_n), f(y_n)) \ge \epsilon$. By the compactness of $K$ and taking subsequence if necessary, we may assume that both $\{x_n\}$ and $\{y_n\}$ converge to the same point $x^* \in K$. So, by the continuity of $J$ in (2.4) and the continuity of $f$ at $x^*$, we have

$$0 = J(f(x^*), f(x^*)) = \lim_{n\to\infty} J(f(x_n), f(y_n)) \ge \epsilon.$$

A contradiction. Thus, $f$ must be uniformly continuous. □

For any maps $f, h : K \to (X, J)$, let

$$J_\infty(f, h) := \sup_{x\in K} J(f(x), h(x)). \tag{2.7}$$

If $J_\infty(f_n, f) \to 0$, then we say that $f_n$ is *uniformly convergent* to $f$.

**Proposition 2.8.** *Suppose $J$ is a nearmetric on $X$. Then, $J_\infty$ is a nearmetric on $C(K,(X,J))$.*

*Proof.* For any $f,h \in C(K,(X,J))$, by definition (2.7), we have $J_\infty(f,h) \geq 0$ and $J_\infty(f,h) = J_\infty(h,f)$. Also, $J_\infty(f,h) = 0$ if and only if $f(x) = h(x)$ for all $x \in K$. Moreover, for any $g \in C(K,(X,J))$,

$$\begin{aligned} J_\infty(f,h) &= \sup_{x\in K} J(f(x),h(x)) \\ &\leq \sup_{x\in K} \sigma(J)\left[J(f(x),g(x)) + J(g(x),h(x))\right] \\ &\leq \sigma(J)\left[\sup_{x\in K} J(f(x),g(x)) + \sup_{x\in K} J(g(x),h(x))\right] \\ &= \sigma(J)\left[J_\infty(f,g) + J_\infty(g,h)\right]. \end{aligned}$$

Therefore, $(C(K,(X,J)),J_\infty)$ is also a nearmetric space. □

**Proposition 2.9.** *Suppose $\{f_n : K \to (X,J)\}$ is a sequence of continuous maps. If $J_\infty(f_n,f) \to 0$, then $f$ is also continuous.*

*Proof.* Since $J_\infty(f_n,f) \to 0$, for any $\epsilon > 0$, there exists an $n$ such that

$$\sup_{x\in K} J(f_n(x),f(x)) \leq \epsilon/3 \tag{2.8}$$

For any $x \in K$, since $f_n$ is continuous at $x$, there exists a $\delta = \delta(x) > 0$ such that $J(f_n(x),f_n(y)) \leq \epsilon/3$ whenever $y \in K$ with $d_K(x,y) \leq \delta$. Therefore, by lemma 2.3 and (2.8), we have

$$\begin{aligned} J(f(x),f(y)) &\leq \sigma(J)^2\left[J(f(x),f_n(x)) + J(f_n(x),f_n(y)) + J(f_n(y),f(y))\right] \\ &\leq \epsilon\sigma(J)^2 \end{aligned}$$

and thus $f$ is continuous at every $x \in K$. □

**Theorem 2.10.** *Suppose $(X,J)$ is a complete nearmetric space and $J$ is lower semicontinuous. Then, the space $(C(K,(X,J)),J_\infty)$ is also a complete nearmetric space.*

*Proof.* Let $\{f_n\}$ be any Cauchy sequence in $C(K,(X,J))$ with respect to $J_\infty$. That is, for any $\epsilon > 0$, there exists an $N$ such that whenever $m,n \geq N$, we have $J_\infty(f_n,f_m) \leq \epsilon$. So, for each $x \in K$, $\{f_n(x)\}$ is Cauchy in $X$. Since $X$ is complete, $\{f_n(x)\}$ converges to some $f(x) \in X$ with respect to $J$. Now,

$$\begin{aligned} J_\infty(f_n,f) &= \sup_{x\in K} J(f_n(x),f(x)) \\ &\leq \sup_{x\in K} \lim_{m\to\infty} J(f_n(x),f_m(x)), \text{ because } J \text{ is lower semicontinuous} \\ &\leq \limsup_{m\to\infty}\left[\sup_{x\in K} J(f_n(x),f_m(x))\right] \leq \epsilon \end{aligned}$$

So, $J_\infty(f_n,f) \to 0$. By proposition 2.9, $f$ is continuous. Hence, by proposition 2.8, $J_\infty$ is a complete nearmetric on $C(K,(X,J))$. □

**Definition 2.11.** *A subset $\mathcal{F}$ of $C(K,(X,J))$ is equicontinuous if for every $x \in K$ and $\epsilon > 0$, there is a $\delta = \delta(x,\epsilon) > 0$, such that whenever $y \in K$ with $d_K(x,y) \leq \delta$, we have $J(f(x),f(y)) \leq \epsilon$ for all $f \in \mathcal{F}$.*

Now, we have the following Ascoli-Arzelà theorem in nearmetric spaces:

**Theorem 2.12.** *Suppose $(X, J)$ is a complete nearmetric space and $J$ is lower semicontinuous. A subset $\mathcal{F}$ of $(C(K,(X,J)), J_\infty)$ is precompact if and only if it is bounded and equicontinuous.*

*Proof.* Suppose $\mathcal{F}$ is a precompact (i.e. every sequence has a convergent subsequence) subset of $C(K,(X,J))$. Then, for each fixed $\epsilon > 0$ , there exists a finite subset $\{f_1, \cdots, f_k\}$ of $\mathcal{F}$ such that

$$\mathcal{F} \subset \bigcup_{i=1}^{k} B_{\epsilon/3}(f_i), \tag{2.9}$$

where the notation $B_\epsilon(g) = \{h \in C(K,(X,J)) \,|\, J_\infty(g,h) < \epsilon\}$. Otherwise, for any finite subset $\{f_1, \cdots, f_k\}$, there exists an $f_{k+1} \notin \bigcup_{i=1}^{k} B_{\epsilon/3}(f_i)$, and thus we get a sequence $\{f_k\}$ in $\mathcal{F}$. Since $J_\infty(f_m, f_n) \geq \epsilon/3$ for any $m \neq n$, we know $\{f_n\}$ does not contain any Cauchy subsequence, which contradicts to $\mathcal{F}$ being precompact. Therefore, (2.9) must be true, which also implies that $\mathcal{F}$ is bounded.

Now, for any $x \in K$ and each $f_i$ in (2.9), there exists a $\delta_i > 0$ such that whenever $y \in K$ with $d_K(x,y) < \delta_i$, we have $J(f_i(x), f_i(y)) \leq \frac{\epsilon}{3}$. For every $f \in \mathcal{F}$, by (2.9), there is an $1 \leq i \leq k$ such that $J_\infty(f, f_i) \leq \frac{\epsilon}{3}$. We conclude that for any $y \in K$ with $d_K(x,y) < \delta = \min\{\delta_1, \cdots, \delta_k\}$, we have

$$\begin{aligned} J(f(x), f(y)) &\leq \sigma(J)^2 \left[J(f(x), f_i(x)) + J(f_i(x), f_i(y)) + J(f_i(y), f(y))\right] \\ &\leq \epsilon\sigma(J)^2 . \end{aligned}$$

Therefore, $\mathcal{F}$ is equicontinuous at every $x \in K$.

On the other hand, suppose $\mathcal{F}$ is equicontinuous and bounded. Then, for any sequence $\{f_n\}$ in $\mathcal{F}$, by using the diagonal process and taking subsequence if necessary, we may assume $\{f_n\}$ is convergent to $f$ on a countable dense subset $S$ in $K$. We now prove that $\{f_n\}$ is Cauchy in $C(K,(X,J))$ with respect to $J_\infty$. Indeed, for any $\epsilon > 0$, since $\mathcal{F}$ is equicontinuous and $K$ is compact, there exists a finite many points $\{r_1, \cdots, r_k\}$ in $S$ such that for any $x \in K$, there is a $r_i$, such that

$$J(f_n(x), f_n(r_i)) \leq \frac{\epsilon}{3}$$

for all $n$. Now, whenever $m, n$ are large enough, for all $x \in K$,

$$\begin{aligned} & J(f_n(x), f_m(x)) \\ \leq\; & \sigma(J)^2 \left[J(f_n(x), f_n(r_i)) + J(f_n(r_i), f_m(r_i)) + J(f_m(r_i), f_m(x))\right] \\ \leq\; & \sigma(J)^2 \epsilon. \end{aligned}$$

Therefore, $\{f_n\}$ is a Cauchy sequence in $C(K,(X,J))$. By the completeness of $C(K,(X,J))$ stated in theorem 2.10, the sequence $\{f_n\}$ is convergent with respect to $J_\infty$. Thus, $\mathcal{F}$ is precompact. □

**Corollary 2.13.** *Suppose $(X, J)$ is a complete nearmetric space and $J$ is lower semicontinuous. A subset $\mathcal{F}$ of $C(K,(X,J))$ is sequentially compact with respect to $J_\infty$ if and only if it is closed, bounded and equicontinuous.*

## 3. Intrinsic Metrics induced by nearmetrics

This section is devoted to study the geodesic problem in a nearmetric space $(X,J)$. Let $[a,b]$ be a bounded closed interval.

**Definition 3.1.** *Let $N$ be a natural number. A curve $f\in C([a,b],(X,J))$ is called an $N$-piecewise Lipschitz curve in $(X,J)$ if there exists a partition*

$$P_f=\{a=a_0<a_1<\cdots<a_N=b\}$$

*of $[a,b]$ such that for each $i=0,1,\cdots,N-1$,*

1. *$J$ is a metric on the subset $f([a_i,a_{i+1}])$ of $X$ and*
2. *the restriction of $f$ on $[a_i,a_{i+1}]$ is Lipschitz.*

Here, requiring $J$ to be a metric on $f([a_i,a_{i+1}])$ is the same as asking it to satisfy the triangle inequality: $J(f(t_1),f(t_2))\le J(f(t_1),f(t_2))+J(f(t_2),f(t_3))$ for any $t_1,t_2,t_3\in[a_i,a_{i+1}]$. Let

$$\mathcal{P}_N([a,b],(X,J))$$

be the family of all $N-$piecewise Lipschitz curves in $(X,J)$, and $\mathcal{P}([a,b],(X,J))$ be the union of $\mathcal{P}_N([a,b],(X,J))$ over all $N$'s.

### 3.1. Length of rectifiable curves.

Recall that when $(X,d)$ is a metric space, and $f:[a,b]\to(X,d)$ is a (continuous) curve. Then, one may define its length as

$$L(f)=\sup_P V_P(f)\in[0,+\infty],$$

where the supremum is over all partitions $P$ of $[a,b]$, and $V_P(f)$ is the variation of $f$ over the partition $P=\{a=t_0<t_1<\cdots<t_N=b\}$ given by

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

In case $f$ is Lipschitz, an equivalent formula for the length of $f$ is

$$L(f)=\int_a^b\left|\dot f(t)\right|_d dt,$$

where $\left|\dot f(t)\right|_d$ is the metric derivative of $f$ at $f(t)$ defined by

$$\left|\dot f(t)\right|_d:=\lim_{s\to t}\frac{d(f(s),f(t))}{|s-t|},$$

provided the limit exists. When $f$ is Lipschitz, $\left|\dot f(t)\right|_d$ exists almost everywhere, and is bounded and measurable in $t$.

Now, suppose $(X,J)$ is a nearmetric space, and $f\in\mathcal{P}_N([a,b],(X,J))$. Then on each interval $[a_i,a_{i+1}]$, $f:[a_i,a_{i+1}]\to(X,J)$ is a Lipschitz curve in the metric space $(f([a_i,a_{i+1}]),J)$, and thus the length of the restriction of $f$ on $[a_i,a_{i+1}]$ is well defined. As a result, we may define the length of $f$ to be

$$L(f):=\sum_{i=0}^{N-1}L\left(f\lfloor_{[a_i,a_{i+1}]}\right).$$

In other words, we have

**Definition 3.2.** *For any $f \in \mathcal{P}_N([a,b],(X,J))$, the length of $f$ is defined as*

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

*where the metric derivative*

$$\left|\dot{f}(t)\right|_J := \lim_{s\to t} \frac{J(f(s),f(t))}{|s-t|}$$

*provided the limit exists. We may simply write $L_J(f)$ as $L(f)$ if $J$ is obvious.*

**Lemma 3.3.** *Suppose $J$ is a continuous nearmetric on $X$, $C>0$ is a constant, and $P=\{a=a_0<a_1<\cdots<a_N=b\}$ is a partition of the interval $[a,b]$. Then, for any $x,y\in X$, the family*

$$\mathcal{F}=\left\{\begin{array}{l} f\in C([a,b],(X,J)):\ f(a)=x, f(b)=y,\ \text{and } J \text{ is a metric on} \\ f([a_i,a_{i+1}])\ \text{ and } Lip\left(f\lfloor_{[a_i,a_{i+1}]}\right)\le C,\ \text{for each } i=0,\cdots,N-1 \end{array}\right\}$$

*is a bounded, closed and equicontinuous subset of $C([a,b],(X,J))$. Moreover, if $f_n$ is uniformly convergent to $f$ in $J_\infty$, then,*

$$L(f)\le \liminf_n L(f_n).$$

*Proof.* For any $g\in\mathcal{F}$ and any $t\in[a,b]$, we have $t\in[a_j,a_{j+1}]$ for some $j\le N-1$ and

$$\begin{aligned} J(g(t),x) &= J(g(t),g(a)) \\ &= \sigma(J)^j\left(\sum_{i=0}^{j-1} J(g(a_i),g(a_{i+1}))+J(g(a_j),g(t))\right) \\ &\le \sigma(J)^j C|t-a| \le C\sigma(J)^{N-1}|b-a| \end{aligned}$$

Therefore, $\mathcal{F}$ is bounded.

Suppose $\{f_n\}$ is any convergent sequence in $\mathcal{F}$ with respect to $J_\infty$ with $f\in C([a,b],(X,J))$ being the limit. Then, for each fixed $i$, and any $t_1,t_2,t_3\in[a_i,a_{i+1}]$, we have

$$J(f_n(t_1),f_n(t_2))\le J(f_n(t_1),f_n(t_3))+J(f_n(t_3),f_n(t_2))$$

and

$$J(f_n(t_1),f_n(t_2))\le C|t_1-t_2|.$$

Let $n\to\infty$, we have $J$ is a metric on $f([a_i,a_{i+1}])$ and $Lip(f\lfloor[a_i,a_{i+1}])\le C$. Therefore, $f\in\mathcal{F}$. This shows that $\mathcal{F}$ is closed and also equicontinuous. Moreover, for any partition $Q$ of $[a_i,a_{i+1}]$, the variation

$$V_Q(f\lfloor[a_i,a_{i+1}])=\lim_n V_Q((f_n)\lfloor[a_i,a_{i+1}])\le\liminf_n L((f_n)\lfloor[a_i,a_{i+1}]).$$

So,

$$L(f\lfloor[a_i,a_{i+1}])=\sup_Q V_Q(f\lfloor[a_i,a_{i+1}])\le\liminf_n L(f_n\lfloor[a_i,a_{i+1}]).$$

Hence,

$$\begin{aligned} L(f) &= \sum_{i=0}^{N-1} L\left(f\lfloor_{[a_i,a_{i+1}]}\right)\le\sum_{i=0}^{N-1}\liminf_n L\left(f_n\lfloor_{[a_i,a_{i+1}]}\right) \\ &= \liminf_n L(f_n). \end{aligned}$$

□

**Proposition 3.4.** *Suppose* $(X, J)$ *is a nearmetric space, and* $f \in \mathcal{P}_N([a,b], (X,J))$. *If* $L(f) = 0$, *then* $f$ *is a constant map.*

*Proof.* $L(f) = 0$ implies that $L\left(f\lfloor_{[a_i, a_{i+1}]}\right) = 0$ for each $i$. Thus, $f$ is a constant on $[a_i, a_{i+1}]$ for each $i$. Since $f$ is continuous, $f$ is a constant on $[a,b]$. □

Since any Lipschitz curve in a metric space has an arc parametrization, by applying arc parametrizations piecewisely, we also have

**Proposition 3.5.** *(Reparametrization) For any* $f \in \mathcal{P}_N([a,b], (X,J))$ *and* $L = L(f)$, *there exists a homeomorphism* $\phi : [0,L] \to [a,b]$ *so that* $\gamma = f \circ \phi \in \mathcal{P}_N([0,L], (X,J))$ *has* $|\dot{\gamma}(t)|_J = 1$ *almost everywhere in* $[0,L]$.

3.2. **The geodesic problem.** Let $N$ be a fixed natural number. For any $x, y \in X$, we consider the geodesic problem

$$\min\{L(f)\} \tag{3.1}$$

among all $f$ in the family

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

Note that, by a linear change of variable, one may replace $[0,1]$ in $Path_N(x,y)$ by any closed interval $[a,b]$ without changing the infimum value in the geodesic problem (3.1).

**Definition 3.6.** *Suppose* $J$ *is a nearmetric on* $X$. *For any* $x, y \in X$, *and* $N \in \mathbb{N}$, *define*

$$D_N(x,y) = \inf\{L(f) : f \in Path_N(x,y)\}$$

*whenever* $Path_N(x,y)$ *is not empty, and set* $D_N(x,y) = \infty$ *when* $Path_N(x,y)$ *is empty. Since* $D_N(x,y)$ *is a decreasing function of* $N$, *we define*

$$D_J(x,y) = \lim_{N\to\infty} D_N(x,y).$$

**Theorem 3.7.** *Suppose* $J$ *is a continuous complete nearmetric on a nonempty set* $X$. *For any* $N \in \mathbb{N}$, *and* $x, y \in X$, *the geodesic problem (3.1) admits a solution* $f \in Path_N(x,y)$ *provided that* $Path_N(x,y)$ *is not empty. So,* $L(f) = D_N(x,y)$.

*Proof.* Suppose $Path_N(x,y)$ is not empty. Let $L = \inf\{L(f) : f \in Path_N(x,y)\}$. Note that for each $f \in Path_N(x,y)$, we have

$$\begin{aligned} J(x,y) &\leq \sigma(J)^{N-1} \sum_{i=0}^{N-1} J(f(a_i), f(a_{i+1})) \\ &\leq \sigma(J)^{N-1} \sum_{i=0}^{N-1} L\left(f\lfloor_{[a_i,a_{i+1}]}\right) = \sigma(J)^{N-1} L(f). \end{aligned}$$

This implies that if $L = 0$, then we have $J(x,y) = 0$. Therefore, $x = y$ and the constant $f(t) \equiv x$ is the desired solution.

So, without losing generality, we may assume that $L > 0$. Let $\{f_n\}$ be a length minimizing sequence in $Path_N(x,y)$ with $L(f_n) \to L$. Let

$$P_{f_n} = \left\{0 = a_0^{(n)} < a_1^{(n)} < \cdots < a_N^{(n)} = 1\right\}$$

be the partition of $[0,1]$, associated with $f_n$. By reparametrization if necessary, we may assume that each $f_n$ is Lipschitz with $Lip\,(f_n) \le 1.5L$ on $\left[a_i^{(n)}, a_{i+1}^{(n)}\right]$ for each $i = 0, \cdots, N-1$. Then, by choosing a subsequence if necessary, we may assume that each sequence $\left\{a_i^{(n)}\right\}$ is convergent to some point $a_i$ as $n \to \infty$ for each $i = 0, 1, \cdots, N$. Using a linear change of variable, we may assume that for each $i$, $a_i^{(n)} = a_i$ and $Lip\,(f_n) \le 2L$ on $[a_i, a_{i+1}]$. Now, $\{f_n\}$ is a sequence in the family

$$\mathcal{F} = \left\{ \begin{array}{c} f \in C\left([0,1], (X,J)\right) : \ f(0) = x, f(1) = y, \text{ and } J \text{ is a metric on} \\ f\left([a_i, a_{i+1}]\right) \text{ and } Lip\left(f\lfloor_{[a_i,a_{i+1}]}\right) \le 2L, \text{ for each } i = 0, \cdots, N-1 \end{array} \right\}.$$

By lemma 3.3, $\mathcal{F}$ is a bounded, closed and equicontinuous subset of $C\left([0,1], (X,J)\right)$. By the Ascoli-Arzelà theorem shown in corollary 2.13, a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ in $\mathcal{F}$ is uniformly convergent to some $f \in \mathcal{F}$ with respect to $J_\infty$. By the lower semicontinuity of $L$ in the family $\mathcal{F}$, we have $L(f) \le \liminf_k L(f_{n_k}) = L$. Therefore, $f$ is a length minimizer in $Path_N(x,y)$. □

Recall that a function $d : X \times X \to [0, +\infty)$ is a *pseudometric* on $X$ if $d$ satisfies conditions (1),(3) in Definition 2.1, and the triangle inequality $d(x,y) \le d(x,z) + d(z,y)$ for any $x, y, z \in X$. But $d(x,y) = 0$ does not necessarily imply $x = y$. A function $d : X \times X \to [0, +\infty)$ is a *semimetric* on $X$ if $d$ satisfies conditions (1),(2),(3) in Definition 2.1. So, a semimetric $d$ is not required to satisfy the triangle inequality.

Note that each $D_N$ is a semimetric on $X$ in the sense that $D_N(x,y) \ge 0$, $D_N(x,y) = 0$ if and only if $x = y$, and $D_N(x,y) = D_N(y,x)$. In general, $D_N$ may fail to satisfy the triangle inequality. Nevertheless, we have

$$D_{n+m}(x,y) \le D_n(x,z) + D_m(z,y)$$

for any $m, n$ and $x, y, z \in X$. As a result, by letting $N \to \infty$, we have

**Proposition 3.8.** *Suppose $J$ is a nearmetric on $X$, then $D_J$ is a pseudometric on $X$.*

Since $D_J$ is a pseudometric, $D_J$ is a metric on $X$ if and only if

$$D_J(x,y) > 0 \text{ whenever } x \neq y.$$

When $D_J$ becomes a metric on $X$. This metric is called the *intrinsic metric* on $X$ induced by the nearmetric $J$.

### 3.3. **Examples of metrics induced by nearmetrics.**

Now, we are interested in cases that $D_J$ is indeed a metric on $X$.

3.3.1. *Ideal nearmetrics.* Let $J$ be any semimetric on $X$. For any $x, y \in X$, we set

$$d_J(x,y)$$

to be the infimum of

$$\sum_{i=1}^{n-1} J(x_i, x_{i+1})$$

over all finitely many points $x_1, \cdots, x_n \in X$ with $x_1 = x$ and $x_n = y$.

This $d_J$ defines a pseudometric on $X$, but not necessarily a metric on $X$.

**Example 3.9.** *For instance, let* $X = [0,1]$ *and* $J(x,y) = |x-y|^p$ *for some* $p > 1$ *defines a nearmetric on* $X$*. Then, for each* $n$*,*

$$\begin{aligned} d_J(0,1) &\leq \sum_{i=0}^{n-1} J\left(\frac{i}{n}, \frac{i+1}{n}\right) \\ &= \sum_{i=0}^{n-1} \left(\frac{1}{n}\right)^p = \frac{1}{n^{p-1}} \to 0 \text{ as } n \to \infty. \end{aligned}$$

*Thus,* $d_J(0,1) = 0$*, but* $0 \neq 1$*. Hence* $d_J$ *is not a metric on* $X$*. Also, note that in this example,* $Path_N(x,y)$ *is empty whenever* $x \neq y$*. Thus,* $D_J(x,y) = \infty$ *whenever* $x \neq y$*.*

As in the case of $D_J$, $d_J$ is a metric on $X$ if and only if

$$d_J(x,y) > 0 \text{ whenever } x \neq y.$$

Note also that

$$d_J(x,y) \leq D_N(x,y)$$

for each $N$, and thus,

$$d_J(x,y) \leq D_J(x,y).$$

Therefore, $d_J(x,y) > 0$ will automatically imply $D_J(x,y) > 0$. As a result, we have

**Proposition 3.10.** *Suppose* $J$ *is a nearmetric on* $X$*. If* $d_J$ *is a metric on* $X$ *and* $D_J(x,y) < \infty$ *for every* $x, y \in X$*, then* $D_J$ *also defines a metric on* $X$ *.*

**Remark 3.11.** *When* $J$ *is indeed a metric on* $X$*, then both* $d_J$ *and* $D_J$ *are metrics. In this case,* $d_J$ *is just the metric* $J$ *itself, while* $D_J$ *is the intrinsic metric induced by* $J$*.*

In general, by means of definition, we have

$$d_J(x,y) \leq J(x,y) \leq \sigma_\infty(J)\, d_J(x,y),$$

where $\sigma_\infty(J)$ is defined as in (2.2).

Now, suppose $J$ is an ideal nearmetric, then $\sigma_\infty(J) < \infty$ and $J$ satisfies the condition

$$J(x_1,x_n) \leq \sigma_\infty(J) \sum_{i=1}^{n-1} J(x_i, x_{i+1})$$

for any finitely many points $\{x_1, x_2, \cdots, x_n\} \subset X$. Clearly, we have the following proposition:

**Proposition 3.12.** *Suppose* $(X,J)$ *is an ideal nearmetric space. Then for any* $N$ *and any* $f \in \mathcal{P}_N([a,b], (X,J))$*, we have*

$$J(f(a), f(b)) \leq \sigma_\infty(J)\, L(f).$$

**Lemma 3.13.** *Suppose* $J$ *is an ideal nearmetric on* $X$ *. Then,* $d_J$ *is a metric on* $X$*. Moreover, if* $D_J(x,y) < \infty$ *for every* $x, y \in X$*, then* $D_J$ *also defines a metric on* $X$*.*

*Proof.* This is simply because when $x \neq y$, $d_J(x,y) \geq \frac{1}{\sigma_\infty(J)} J(x,y) > 0$. □

3.3.2. *Perfect nearmetrics.* Here is another kind of nearmetric $J$ which also induces a metric $D_J$.

**Definition 3.14.** *A nearmetric $J$ on $X$ is a perfect near metric if for any $x, y \in X$, the value $D_N(x,y)$ becomes a real valued constant $D_J(x,y)$ when $N$ is large enough.*

Since for each $N$, $D_N(x,y) = 0$ if and only if $x = y$, we have the following theorem.

**Proposition 3.15.** *On a perfect nearmetric space $(X, J)$, $D_J$ defines a metric on $X$.*

When $J$ is indeed a metric on $X$, then for each $N$, the metric $D_N$ agrees with the intrinsic metric induced by $J$. Thus, every metric space is automatically a perfect nearmetric space. In section 4, we will discuss a family of very important perfect nearmetric spaces, which are not metric spaces.

**Theorem 3.16.** *Suppose $(X, J)$ is a perfect nearmetric space, and the geodesic problem 3.1 has solution for $N$ large enough. Then, $(X, D_J)$ is a length space in the sense that for every $x, y \in X$, there exists a curve $f : [0, L] \to (X, D_J)$ such that $f(0) = x$, $f(L) = y$ and*

$$D_J(f(t), f(s)) = |t - s|$$

*for every $t, s \in [0, L]$ where $L = D_J(x, y)$.*

*Proof.* For every $x, y \in X$, since $(X, J)$ is a perfect nearmetric space, we have $D_N(x,y) = D_J(x,y) < \infty$ whenever $N$ is large enough. Now, for each large enough $N$, there exists a curve $f : [0, L] \to (X, J)$ such that $f$ is the length minimizer in $Path_N(x,y)$ with $L(f) = D_N(x,y) = D_J(x,y)$. Without losing generality, we may assume $f$ has its arc parametrization. Now for any $0 \le s < t \le L$, we have

$$D_J(f(s), f(t)) \le L\left(f\lfloor_{[s,t]}\right) = \int_s^t \left|\dot{f}\right|_J dt = t - s.$$

Similarly, $D_J(f(0), f(s)) \le s$ and $D_J(f(t), f(L)) \le L - t$. Thus, we have

$$\begin{aligned} L &= D_J(x,y) \le D_J(f(0), f(s)) + D_J(f(s), f(t)) + D_J(f(t), f(L)) \\ &\le s + (t-s) + (L-t) = L. \end{aligned}$$

Therefore, all inequalities becomes equalities at every step and for any $t, s \in [0, L]$, we have $D_J(f(t), f(s)) = |t - s|$ . □

**Corollary 3.17.** *Suppose $J$ is a complete, continuous, perfect nearmetric on $X$. Then, $(X, D_J)$ is a length space.*

The curve $f$ in the theorem 3.16 is called a *geodesic* from $x$ to $y$ in the perfect nearmetric space $(X, J)$.

## 4. Optimal transport paths as geodesics

We now begin to introduce a family of both ideal and perfect nearmetrics on the space of atomic probability measures.

4.1. **A family of nearmetrics on the space of atomic probability measures.**

Let $(Y, d)$ be any metric space. For any $y \in Y$, let $\delta_y$ be the Dirac measure centered at $y$. An atomic probability measure in $Y$ is in the form of

$$\sum_{i=1}^{m} a_i \delta_{y_i}$$

with distinct points $y_i \in Y$, and $a_i > 0$ with $\sum_{i=1}^{m} a_i = 1$.

Given two atomic probability measures

$$\mathbf{a} = \sum_{i=1}^{m} a_i \delta_{x_i} \text{ and } \mathbf{b} = \sum_{j=1}^{n} b_j \delta_{y_j} \tag{4.1}$$

in $Y$, a *transport plan* from $\mathbf{a}$ to $\mathbf{b}$ is an atomic probability measure

$$\gamma = \sum_{i=1}^{m} \sum_{j=1}^{n} \gamma_{ij} \delta_{(x_i, y_j)} \tag{4.2}$$

in the product space $Y \times Y$ such that

$$\sum_{i=1}^{m} \gamma_{ij} = b_j \text{ and } \sum_{j=1}^{n} \gamma_{ij} = a_i \tag{4.3}$$

for each $i$ and $j$. Let $Plan\,(\mathbf{a}, \mathbf{b})$ be the space of all transport plans from $\mathbf{a}$ to $\mathbf{b}$.

For any $\alpha < 1$, we now introduce the functional $H_\alpha$ on transport plans. For any atomic probability measure $\gamma$ in $Y \times Y$ of the form (4.2), we define

$$H_\alpha(\gamma) := \sum_{i=1}^{m} \sum_{j=1}^{n} (\gamma_{ij})^\alpha \, d(x_i, y_j),$$

where $d$ is the given metric on $Y$.

Using $H_\alpha$, we may define

**Definition 4.1.** *For any two atomic probability measures* $\mathbf{a}, \mathbf{b}$ *on* $Y$*, and* $\alpha < 1$*, define*

$$J_\alpha(\mathbf{a}, \mathbf{b}) := \min \{H_\alpha(\gamma) : \gamma \in Plan\,(\mathbf{a}, \mathbf{b})\}.$$

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 $m \leq N$, and $\mathcal{A}(Y) = \bigcup_N \mathcal{A}_N(Y)$ be the space of all atomic probability measures on $Y$.

**Proposition 4.2.** *$J_\alpha$ defines a nearmetric on $\mathcal{A}_N(Y)$ with $\sigma(J_\alpha) \leq N$.*

*Proof.* For any $\mathbf{a}, \mathbf{b} \in \mathcal{A}_N(Y)$ in the form of (4.1), clearly $J_\alpha(\mathbf{a}, \mathbf{b}) \geq 0$ and $J_\alpha(\mathbf{a}, \mathbf{b}) = J_\alpha(\mathbf{b}, \mathbf{a})$.

If $J_\alpha(\mathbf{a}, \mathbf{b}) = 0$, then there exists a $\gamma \in Plan\,(\mathbf{a}, \mathbf{b})$ such that $H_\alpha(\gamma) = 0$. Thus, $d(x_i, y_j) = 0$ whenever $\gamma_{ij} \neq 0$. Since $\{y_j\}$'s are distinct, at most one of $\gamma_{ij}$ can be nonzero for each $i$. On the other hand, by (4.3), at least one of $\gamma_{ij}$ must be nonzero for each $i$. Therefore, for each $i$, there is a unique $j = \sigma(i)$ such that $x_i = y_j$ and $\gamma_{ij} = a_i = b_j$. This shows that $\mathbf{a} = \mathbf{b}$.

Now, we prove that $J$ satisfies the relaxed triangle inequality as in condition 4 in Definition 2.1. Indeed, for any

$$\mathbf{a}=\sum_{i=1}^{m}a_i\delta_{x_i},\ \mathbf{b}=\sum_{j=1}^{n}b_j\delta_{y_j}\text{ and }\mathbf{c}=\sum_{k=1}^{h}c_k\delta_{z_k}$$

in $\mathcal{A}_N(Y)$, and any

$$u_{\mathbf{a}}^{\mathbf{c}}=\sum_{i=1}^{m}\sum_{k=1}^{h}u_{ik}\delta_{(x_i,z_k)}\in Path(\mathbf{a},\mathbf{c})\text{ and }\tau_{\mathbf{c}}^{\mathbf{b}}=\sum_{j=1}^{n}\sum_{k=1}^{h}\tau_{kj}\delta_{(z_k,y_j)}\in Path(\mathbf{c},\mathbf{b}),$$

we denote

$$\gamma_{ij}=\sum_{k=1}^{h}\frac{u_{ik}\tau_{kj}}{c_k}$$

for each $i,j$. Note that

$$\sum_{i=1}^{m}\gamma_{ij}=\sum_{i=1}^{m}\left(\sum_{k=1}^{h}\frac{u_{ik}\tau_{kj}}{c_k}\right)=\sum_{k=1}^{h}\left(\sum_{i=1}^{m}\frac{u_{ik}\tau_{kj}}{c_k}\right)=\sum_{k=1}^{h}\tau_{kj}=b_j$$

and similarly $\sum_j\gamma_{ij}=a_i$.Therefore, we find a transport plan

$$\gamma=\sum_{i=1}^{m}\sum_{j=1}^{n}\gamma_{ij}\delta_{(x_i,y_j)}\in Plan(\mathbf{a},\mathbf{b}).$$

We now want to show

$$H_\alpha(\gamma)\le N\left(H_\alpha(u_{\mathbf{a}}^{\mathbf{c}})+H_\alpha\left(\tau_{\mathbf{c}}^{\mathbf{b}}\right)\right).$$

Indeed,

$$\begin{aligned}
H_\alpha(\gamma)&=\sum_{i=1}^{m}\sum_{j=1}^{n}(\gamma_{ij})^\alpha d(x_i,y_j)=\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\sum_{k=1}^{h}\frac{u_{ik}\tau_{kj}}{c_k}\right)^\alpha d(x_i,y_j)\\
&\le\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{h}\left(\frac{u_{ik}\tau_{kj}}{c_k}\right)^\alpha(d(x_i,z_k)+d(z_k,y_j)),\ \text{because }\alpha<1\\
&=\sum_{i=1}^{m}\sum_{k=1}^{h}\left(\sum_{j=1}^{n}\left(\frac{u_{ik}\tau_{kj}}{c_k}\right)^\alpha\right)d(x_i,z_k)+\sum_{j=1}^{n}\sum_{k=1}^{h}\left(\sum_{i=1}^{m}\left(\frac{u_{ik}\tau_{kj}}{c_k}\right)^\alpha\right)d(z_k,y_j)\\
&\le N\left(\sum_{i=1}^{m}\sum_{k=1}^{h}(u_{ik})^\alpha d(x_i,z_k)+\sum_{j=1}^{n}\sum_{k=1}^{h}(\tau_{kj})^\alpha d(z_k,y_j)\right),\ \text{since }\tau_{kj}\le c_k\text{ and }u_{ik}\le c_k\\
&=N\left(H_\alpha(u_{\mathbf{a}}^{\mathbf{c}})+H_\alpha\left(\tau_{\mathbf{c}}^{\mathbf{b}}\right)\right)
\end{aligned}$$

Therefore, by taking infimum, we have

$$J_\alpha(\mathbf{a},\mathbf{b})\le N(J_\alpha(\mathbf{a},\mathbf{c})+J_\alpha(\mathbf{c},\mathbf{b})).$$

□

Note that, in general, $J_\alpha$ may fail to be a metric on $\mathcal{A}_N(Y)$ as demonstrated in the following example.

**Example 4.3.** *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,*

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

*whenever $y$ is large enough. Thus, $J_\alpha$ does not satisfy the triangle inequality.*

4.2. **Optimal transport paths between atomic probability measures.** Now, we want to show that the nearmetric $J_\alpha$ is both ideal and perfect. To achieve these results, we first recall some concepts about optimal transport paths between probability measures as studied in [6].

Let $\mathbf{a}$ and $\mathbf{b}$ be two fixed atomic probability measures in the form of (4.1).

**Definition 4.4.** *A transport path from $\mathbf{a}$ to $\mathbf{b}$ is a weighted directed graph $G$ consists of a vertex set $V(G)$, a directed edge set $E(G)$ and a weight function*

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

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

$$\sum_{\substack{e \in E(G) \\ e^- = v}} w(e) = \sum_{\substack{e \in E(G) \\ e^+ = v}} w(e) + \begin{cases} a_i, & \text{if } v = x_i \text{ for some } i = 1, \cdots, k \\ -b_j, & \text{if } v = y_j \text{ for some } j = 1, \cdots, l \\ 0, & \text{otherwise} \end{cases} \tag{4.4}$$

*where $e^-$ and $e^+$ denotes the starting and ending endpoints of each edge $e \in E(G)$.*

**Remark 4.5.** *The balance equation (4.4) simply means that the total mass flows into $v$ equals to the total mass flows out of $v$. When $G$ is viewed as a polyhedral chain or current, (4.4) can be simply expressed as*

$$\partial G = \mathbf{b} - \mathbf{a}.$$

*Also, when $G$ is viewed as a vector valued measure, the balance equation is simply*

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

*in the sense of distributions.*

Let $Path(\mathbf{a}, \mathbf{b})$ be the space of all transport paths from $\mathbf{a}$ to $\mathbf{b}$.

**Definition 4.6.** *For any $\alpha \leq 1$, and any $G \in Path(\mathbf{a}, \mathbf{b})$, define*

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

**Remark 4.7.** *In [6], the parameter $\alpha$ was restricted in $[0,1]$. Later, the author observed that $\alpha < 0$ is also very interesting, and related to studying the dimension of fractals. So, negative $\alpha$ is also allowed here.*

We first recite two lemmas that were proved in [6, Proposition 2.1] and [6, Definition 7.1 and Lemma 7.1] respectively.

**Lemma 4.8.** *For any transport path $G \in Path(\mathbf{a}, \mathbf{b})$, there exists another transport path $\tilde{G} \in Path(\mathbf{a}, \mathbf{b})$ such that*

$$\mathbf{M}_\alpha\left(\tilde{G}\right) \leq \mathbf{M}_\alpha(G),$$

*vertices $V\left(\tilde{G}\right) \subset V(G)$ and $\tilde{G}$ contains no cycles.*

Here, a weighted directed graph $G = \{V(G), E(G), W : E(G) \to (0,1]\}$ contains a *cycle* if for some $k \geq 3$, there exists a list of distinct vertices $\{v_1, v_2, \cdots, v_k\}$ in $V(G)$ such that for each $i = 1, \cdots, k$, 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 $v_{k+1} = v_1$. When a directed graph $G$ contains no cycles, it becomes a directed tree.

**Lemma 4.9.** *For any transport path $G \in Path(\mathbf{a}, \mathbf{b})$ containing no cycles, there exists*

(1) *an $m \times n$ real matrix*

$$u = (u_{ij}) \text{ with}$$

$$u_{ij} \geq 0, \sum_{i=1}^{m} u_{ij} = a_i, \sum_{j=1}^{n} u_{ij} = b_j \text{ for each } i, j \text{ and} \sum_{i=1}^{m}\sum_{j=1}^{n} u_{ij} = 1,$$

(2) *and an $m \times n$ matrix*

$$g = (g_{ij})$$

*with each $g_{ij}$ is either 0 or an oriented polyhedral curve $g_{ij}$ from $x_i$ to $y_j$,*

*such that*

$$G = \sum_{i,j} u_{ij} g_{ij}$$

*as real coefficients polyhedral chains.*

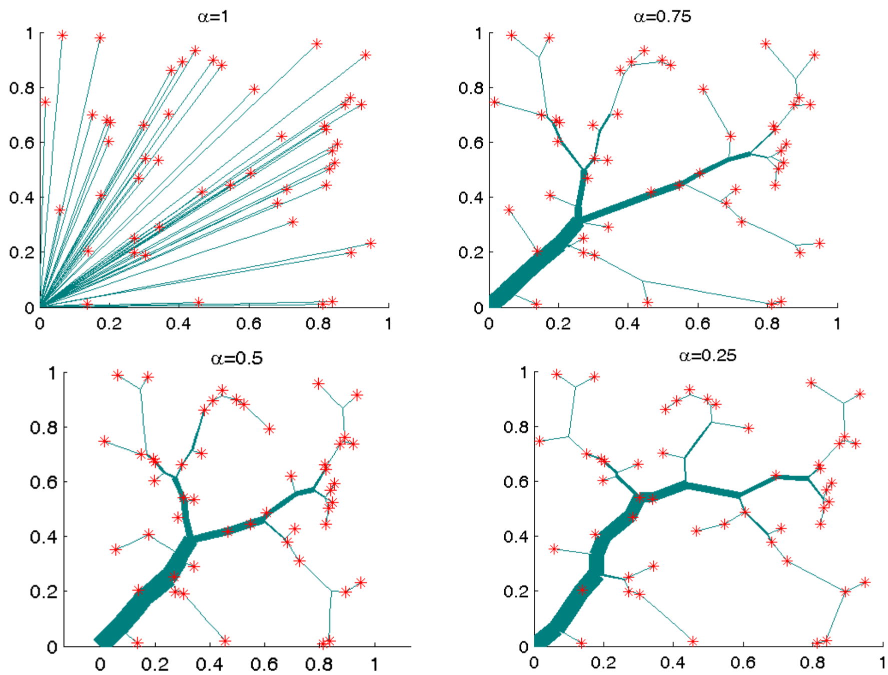
By means of lemma 4.8, it is easy to see that for each $\alpha \leq 1$, there exists an optimal transport path in $Path(\mathbf{a}, \mathbf{b})$ which minimizes the cost functional $\mathbf{M}_\alpha$. To help readers have a better understanding of optimal transport paths, we provided some numerical simulation of optimal transport paths in the following examples, but leaving details of generating algorithms in a forthcoming article.

**Example 4.10.** *Let $\{x_i\}$ be 50 random points in the square $[0,1] \times [0,1]$. Then,*

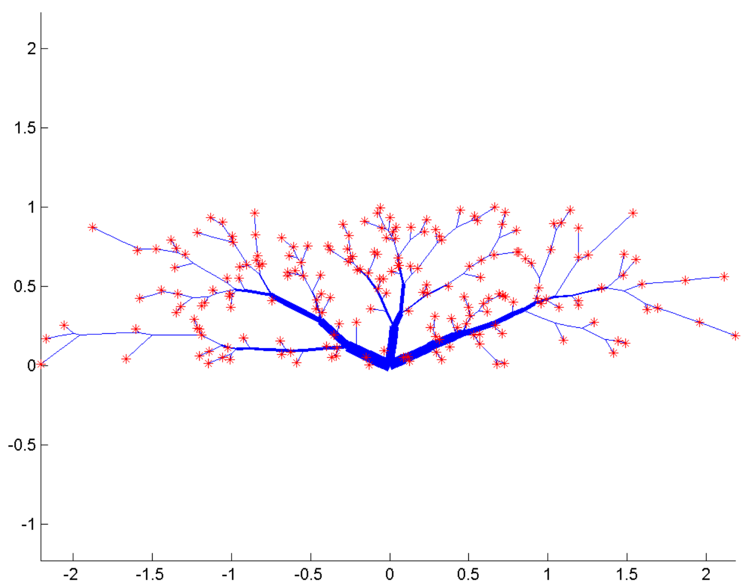
*$\{x_i\}$ determines an atomic probability measure*

$$\mathbf{a} = \sum_{i=1}^{50} \frac{1}{50} \delta_{x_i}.$$

*Let $\mathbf{b} = \delta_O$ 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:*




**Example 4.11.** *Let $\{x_i\}$ be 100 random points in the rectangle $[-2.5, 2.5] \times [0, 1]$. Then, $\{x_i\}$ determines an atomic probability measure $\mathbf{a} = \sum_{i=1}^{100} \frac{1}{100}\delta_{x_i}$. Let $\mathbf{b} = \delta_O$ where $O = (0, 0)$ is the origin, and let $\alpha = 0.85$. Then an optimal transport path from $\mathbf{a}$ to $\mathbf{b}$ looks like the following figure.*



4.3. **Relation between optimal transport paths and nearmetrics $J_\alpha$.** We now start to investigate the relationship between optimal transport path and the nearmetric $J_\alpha$ on $\mathcal{A}_N(Y)$. We first observe that any transport plan $\gamma \in Plan(\mathbf{a}, \mathbf{b})$ in the form of (4.2) determines a transport path $G_\gamma \in Path(\mathbf{a}, \mathbf{b})$. Indeed, we consider the weighted directed graph $G_\gamma$ with

$$\begin{aligned} V(G_\gamma) &= \{x_1, \cdots, x_m, y_1, \cdots, y_n\}, \\ E(G_\gamma) &= \{\text{a pair } [x_i, y_j] \text{ if } \gamma_{ij} \neq 0\}, \end{aligned}$$

and setting the weight $W([x_i, y_j]) = \gamma_{ij}$ for each $i, j$ with $\gamma_{ij} \neq 0$. Moreover,

$$\mathbf{M}_\alpha(G_\gamma) = \sum_{e \in E(G_\gamma)} w(e)^\alpha \, length(e) = \sum_{i,j} (\gamma_{ij})^\alpha \, d(x_i, y_j) = H_\alpha(\gamma).$$

**Proposition 4.12.** *For any $\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \cdots, \mathbf{a}^{(k)} \in \mathcal{A}(Y)$, there exists a transport path $G \in Path\left(\mathbf{a}^{(1)}, \mathbf{a}^{(k)}\right)$ such that*

$$\mathbf{M}_\alpha(G) \leq \sum_{i=1}^{k-1} J_\alpha\left(\mathbf{a}^{(i)}, \mathbf{a}^{(i+1)}\right)$$

*and $G$ contains no cycles.*

*Proof.* Let $\gamma_i$ be an optimal transport path from $\mathbf{a}^{(i)}$ to $\mathbf{a}^{(i+1)}$, for each $i = 1, 2, \cdots, k-1$. Each $\gamma_i$ determines a transport path $G_{\gamma_i} \in Path\left(\mathbf{a}^{(i)}, \mathbf{a}^{(i+1)}\right)$ as above. Then, viewed as real coefficients polyhedral chains,

$$G = \sum_{i=1}^{k-1} G_{\gamma_i}$$

is a transport path from $\mathbf{a}^{(1)}$ to $\mathbf{a}^{(k)}$. Moreover, we have

$$\mathbf{M}_\alpha(G) \leq \sum_{i=1}^{k-1} \mathbf{M}_\alpha(G_{\gamma_i}) = \sum_{i=1}^{k-1} H_\alpha(\gamma_i) = \sum_{i=1}^{k-1} J_\alpha\left(\mathbf{a}^{(i)}, \mathbf{a}^{(i+1)}\right).$$

By lemma 4.8, there exists a transport path $\tilde{G}$ from $\mathbf{a}^{(1)}$ to $\mathbf{a}^{(k)}$ such that $\tilde{G}$ contains no cycles, $V\left(\tilde{G}\right) \subset V(G)$, and

$$\mathbf{M}_\alpha\left(\tilde{G}\right) \leq \mathbf{M}_\alpha(G) \leq \sum_{i=1}^{k-1} J_\alpha\left(\mathbf{a}^{(i)}, \mathbf{a}^{(i+1)}\right).$$

□

**Theorem 4.13.** *$J_\alpha$ is an ideal nearmetric on $\mathcal{A}_N(Y)$ with $\sigma_\infty(J_\alpha) \leq N^2$.*

*Proof.* For any $k \in \mathbb{N}$ and any points $\left\{\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \cdots, \mathbf{a}^{(k)}\right\} \subset \mathcal{A}_N(Y)$, by proposition 4.12, there exists a transport path $G \in Path\left(\mathbf{a}^{(1)}, \mathbf{a}^{(k)}\right)$ such that

$$\mathbf{M}_\alpha(G) \leq \sum_{i=1}^{k-1} J_\alpha\left(\mathbf{a}^{(i)}, \mathbf{a}^{(i+1)}\right)$$

and $G$ contains no cycles. Moreover, by lemma 4.9, there exists a matrix $(u_{ij})$ of real numbers and a matric $(g_{ij})$ of polyhedral curves such that

$$G = \sum_{ij} u_{ij} g_{ij}$$

as real coefficients polyhedral chains. Let

$$\gamma = \sum_{ij} u_{ij} \delta_{(x_i, y_j)}$$

be any transport plan in $Plan\left(\mathbf{a}^{(1)}, \mathbf{a}^{(k)}\right)$. Then,

$$\begin{aligned}
H_\alpha(\gamma) &= \sum_{ij} (u_{ij})^\alpha d(x_i, y_j) \leq \sum_{ij} (u_{ij})^\alpha \, length(g_{ij}) \\
&= \sum_{e \in E(G)} \left( \sum_{g_{ij} \text{ contains } e} (u_{ij})^\alpha \right) length(e) \\
&\leq \sum_{e} \left( N^2 \left( \sum_{g_{ij} \text{ contains } e} u_{ij} \right)^\alpha \right) length(e) \\
&= N^2 \sum_{e \in E(G)} (w(e))^\alpha \, length(e) \\
&= N^2 \mathbf{M}_\alpha(G) \leq N^2 \sum_{i=1}^{k-1} J_\alpha\left(\mathbf{a}^{(i)}, \mathbf{a}^{(i+1)}\right).
\end{aligned}$$

Therefore,

$$J_\alpha\left(\mathbf{a}^{(1)}, \mathbf{a}^{(k)}\right) \leq N^2 \sum_{i=1}^{k-1} J_\alpha\left(\mathbf{a}^{(i)}, \mathbf{a}^{(i+1)}\right)$$

and thus $J_\alpha$ is an ideal nearmetric on $\mathcal{A}_N(Y)$ with $\sigma_\infty(J_\alpha) \leq N^2$. □

Suppose $(Y, d)$ is a geodesic metric space. That is, for any $x, y \in Y$, there exists a Lipschitz curve $\Gamma_{x,y} : [0,1] \to (Y, d)$ with $\Gamma_{x,y}(0) = x$, $\Gamma_{x,y}(1) = y$ and length $L(\Gamma_{x,y}) = d(x, y)$.

**Lemma 4.14.** *Suppose $(Y, d)$ is a geodesic metric space. Let $G \in Path(\mathbf{a}, \mathbf{b})$ for some $\mathbf{a}, \mathbf{b} \in \mathcal{A}_N(Y)$. If each edge of $G$ is a geodesic curve between its endpoints in the metric space $Y$, then there exists a piecewise Lipschitz curve $g \in \mathcal{P}_{N_G}([0,1], (\mathcal{A}_N(Y), J_\alpha))$ such that*

$$L_{J_\alpha}(g) = \mathbf{M}_\alpha(G),$$

*where $N_G$ is total number of edges in the graph $G$.*

*Proof.* We may prove it using the mathematical induction on $N_G$. When $N_G = 1$, $G$ itself is a geodesic in $Y$. Then, it is clearly true in this case. Now, assume $N_G > 1$. Pick an edge $e$ of $G$ with its starting endpoint $e^-$ being a vertex in $\mathbf{a}$. Let

$$\tilde{\mathbf{a}} = \mathbf{a} + w(e)\left(\delta_{e^+} - \delta_{e^-}\right),$$

where $e^+$ is the targeting endpoint of the directed edge $e$, and $w(e)$ is the associated weight on $e$. Removing edge $e$ from $G$, we get another transport path $\tilde{G} \in Path(\tilde{\mathbf{a}}, \mathbf{b})$. Then, $N_{\tilde{G}} = N_G - 1 \geq 1$. By the principle of the mathematical induction, we may assume that $\tilde{G}$ corresponds to a piecewise Lipschitz curve $\tilde{g} \in \mathcal{P}_{N_{\tilde{G}}}([0,1], (\mathcal{A}_N(Y), J_\alpha))$ such that

$$L_{J_\alpha}(\tilde{g}) = \mathbf{M}_\alpha\left(\tilde{G}\right).$$

Now, let

$$g(t)=\begin{cases}\tilde{g}\left(\frac{t}{\lambda}\right), & 0\le t\le\lambda\\ \Gamma_e\left(\frac{t-\lambda}{1-\lambda}\right), & \lambda\le t\le 1\end{cases},$$

where $\lambda=\frac{N_G-1}{N_G}$, and $\Gamma_e$ is the associated geodesic in $Y$ from $e^-$ to $e^+$. Then, $g\in\mathcal{P}_{N_G}([0,1],(\mathcal{A}_N(Y),J_\alpha))$ and

$$L_{J_\alpha}(g)=L_{J_\alpha}(\tilde{g})+L_{J_\alpha}(\Gamma_e)=\mathbf{M}_\alpha\left(\tilde{G}\right)+w(e)^\alpha\, length(e)=\mathbf{M}_\alpha(G).$$

□

**Remark 4.15.** *From this lemma, we see that for any transport path $G\in Path(\mathbf{a},\mathbf{b})$ in a geodesic metric space $(Y,d)$, we have a simple formula for the transport cost:*

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

*On the other hand, in [3], the authors studied another kind of ramified transportation in which the cost of a path is given by*

$$\int_0^1|\dot{g}(t)|_W\, J(g(t))\,dt$$

*where $W$ is the Wasserstein distance on probability measures, and $J$ is some function on the space of atomic probability measures. It is interesting to see this difference between these two different approaches.*

**Theorem 4.16.** *Suppose $(Y,d)$ is a geodesic metric space. Then, $J_\alpha$ is a perfect nearmetric on $\mathcal{A}_N(Y)$, and thus it induces a metric $D_{J_\alpha}$ on $\mathcal{A}_N(Y)$.*

*Proof.* Suppose $\mathbf{a},\mathbf{b}$ are two points in $\mathcal{A}_N(Y)$. For any $f\in\mathcal{P}_k([0,1],(\mathcal{A}_N(Y),J_\alpha))$ with $f(0)=\mathbf{a}$ and $f(1)=\mathbf{b}$, there exists a partition $P=\{0=a_0<\cdots<a_k=1\}$ of $[0,1]$ such that $J_\alpha$ is a metric on $f([a_i,a_{i+1}])$ and $f\lfloor_{[a_i,a_{i+1}]}$ is Lipschitz for each $i=0,1,\cdots,k-1$. Let $x_i=f(a_i)$ for each $i$, by proposition 4.12, there exists a transport path $G$ from $f(0)=\mathbf{a}$ to $f(1)=\mathbf{b}$ such that

$$\mathbf{M}_\alpha(G)\le\sum J_\alpha(x_i,x_{i+1})\le\sum_i L\left(f\lfloor_{[a_i,a_{i+1}]}\right)=L(f)$$

and $G$ contains no cycles.When $(Y,d)$ is a geodesic metric space, each edge of $G$ is realized by a geodesic curve between its endpoints. By lemma 4.14, $G$ determines a curve $g\in\mathcal{P}_{N_G}([0,1],(\mathcal{A}_N(Y),J_\alpha))$ with $L(g)=\mathbf{M}_\alpha(G)\le L(f)$. Since $\mathbf{a},\mathbf{b}\in\mathcal{A}_N(Y)$ and $G\in Path(\mathbf{a},\mathbf{b})$, the total number of vertices of $G$ with degree one is no more than $2N$. Since $G$ contains no cycles, the total number $N_G$ of edges of $G$ is no more than $4N-3$. Thus, $g\in\mathcal{P}_{4N-3}([0,1],(\mathcal{A}_N(Y),J_\alpha))$. Hence, for any $\mathbf{a},\mathbf{b}\in\mathcal{A}_N(Y)$,

$$D_k(\mathbf{a},\mathbf{b})=D_{4N-3}(\mathbf{a},\mathbf{b})$$

for any $k\ge 4N-3$. This shows that $J_\alpha$ is a perfect nearmetric on $\mathcal{A}_N(Y)$. □

**Corollary 4.17.** *Suppose $(Y,d)$ is a geodesic metric space. Then, for any $\mathbf{a},\mathbf{b}\in\mathcal{A}_N(Y)$ and $\alpha\le 1$, we have*

$$D_{J_\alpha}(\mathbf{a},\mathbf{b})=\min\{\mathbf{M}_\alpha(G): G\in Path(\mathbf{a},\mathbf{b})\}.$$

*Proof.* Let $G$ be any optimal transport path from $\mathbf{a}$ to $\mathbf{b}$. From the proof of the above theorem, we see $D_{J_\alpha}(\mathbf{a},\mathbf{b}) \leq \mathbf{M}_\alpha(G) \leq L(f)$ for any $f \in \mathcal{P}_k([0,1],(\mathcal{A}_N(Y),J_\alpha))$ with $k \geq 4N-3$. Hence, $D_{J_\alpha}(\mathbf{a},\mathbf{b}) = \mathbf{M}_\alpha(G)$. □

**Corollary 4.18.** *Suppose $(Y,d)$ is a geodesic metric space. Then, $(\mathcal{A}_N(Y),D_{J_\alpha})$ is a length space.*

*Proof.* By corollary 4.17, each optimal transport path $G$ determines a solution $g$ to the geodesic problem (3.1). Then, by theorem 3.16, $(\mathcal{A}_N(Y),D_{J_\alpha})$ becomes a length space. □

Since $\mathcal{A}_1(Y) \subset \mathcal{A}_2(Y) \subset \cdots \subset \mathcal{A}_N(Y) \subset \cdots$, and $(\mathcal{A}_N(Y),D_{J_\alpha})$ is a length space for each $N$, we have

**Proposition 4.19.** *Suppose $(Y,d)$ is a geodesic metric space. Then, $D_{J_\alpha}$ is a metric on the space $\mathcal{A}(Y)$ of all atomic probability measures on $Y$. Moreover, $(\mathcal{A}(Y),D_{J_\alpha})$ is a length space.*

We now give some conclusive remarks:

**Remark 4.20.** *In* [6], *we defined $d_\alpha(\mathbf{a},\mathbf{b}) := \min\{\mathbf{M}_\alpha(G) : G \in Path(\mathbf{a},\mathbf{b})\}$ for $0 \leq \alpha < 1$ and showed that $d_\alpha$ defines a metric on the space of (atomic) probability measures. Moreover, we showed $(\mathcal{A}(Y),d_\alpha)$ is a length space. Now, from corollary 4.17, we see that $d_\alpha = D_{J_\alpha}$. That is, the metric $d_\alpha$ is just the intrinsic metric on $\mathcal{A}(Y)$ induced by the nearmetric $J_\alpha$. Proposition 4.19 simply gives another proof of $(\mathcal{A}(Y),d_\alpha)$ being a length space. Furthermore, an optimal transport path studied in* [6] *is simply a geodesic in the length space $(\mathcal{A}(Y),D_{J_\alpha})$.*

**Remark 4.21.** *Suppose $(Y,d)$ is a geodesic metric space, and $\mathcal{P}_\alpha(Y)$ is the completion of the metric space $(\mathcal{A}(Y),D_{J_\alpha})$. Then, $(\mathcal{P}_\alpha(Y),D_{J_\alpha})$ is also a length space. A geodesic in the length space $(\mathcal{P}_\alpha(Y),D_{J_\alpha})$ is also called an optimal transport path between its endpoints.*

## References


[1] Luigi Ambrosio, Tilli Paolo; Topics on Analysis in Metric Spaces, Oxford University Press, 2004.

[2] Arkhangel'skii, A.V.; Pontryagin, L.S. (1990). General Topology I: Basic Concepts and Constructions Dimension Theory, Springer.

[3] A. Brancolini, G. Buttazzo, F. Santambrogio, Path functions over Wasserstein spaces. J. Eur. Math. Soc. Vol. 8, No.3 (2006),415–434.

[4] R. Fagin, R. Kumar and D. Sivakumar, Comparing Top k Lists, SIAM Journal on Discrete Mathematics, Volume 17, Issue 1 (2003),134-160.

[5] John Hunter and B. Nachtergaele, Applied Analysis, World Scientific Publishing Company.(2001).

[6] Qinglan Xia, Optimal paths related to transport problems. Communications in Contemporary Mathematics. Vol. 5, No. 2 (2003) 251-279.

[7] Qinglan Xia. Interior regularity of optimal transport paths. Calculus of Variations and Partial Differential Equations. 20 (2004), no. 3, 283–299.

[8] Qinglan Xia. Boundary regularity of optimal transport paths. Preprint.

[9] Qinglan Xia. An application of optimal transport paths to urban transport networks. Discrete and Continuous Dynamical Systems, supp. Volume 2005, pp.904-910.



University of California at Davis, Department of Mathematics, Davis,CA,95616
*E-mail address*: `qlxia@math.ucdavis.edu`
*URL*: `http://math.ucdavis.edu/~qlxia`