Introduction to optimal transport paths
A Tutorial of IPAM 2008 Workshop on Optimal Transportation

Qinglan Xia
University of California at Davis

March 11, 2008
For a given cost function $c : X \times X \to [0, +\infty)$, we have considered

- **Monge problem:** Minimize
  \[
  I[\psi] := \int_X c(x, \psi(x)) \, d\mu^+(x)
  \]
  among all transport maps.

- **Monge-Kantorovich problem:** Minimize
  \[
  J(\gamma) := \int_{X \times X} c(x, y) \, d\gamma(x, y)
  \]
  among all transport plans.
For a given cost function $c : X \times X \to [0, +\infty)$, we have considered

- **Monge problem**: Minimize

  $$I[\psi] := \int_X c(x, \psi(x)) \, d\mu^+(x)$$

  among all transport maps.

- **Monge-Kantorovich problem**: Minimize

  $$J(\gamma) := \int_{X \times X} c(x, y) \, d\gamma(x, y)$$

  among all transport plans.

But, should we always define transportation cost as an integral of a cost function $c(x, y)$?
For a given cost function $c : X \times X \to [0, +\infty)$, we have considered

- **Monge problem:** Minimize
  
  $$ I [\psi] := \int_X c (x, \psi (x)) \, d\mu^+ (x) $$

  among all transport maps.

- **Monge-Kantorovich problem:** Minimize
  
  $$ J (\gamma) := \int_{X \times X} c (x, y) \, d\gamma (x, y) $$

  among all transport plans.

But, should we always define transportation cost as an integral of a cost function $c(x, y)$?

**Answer:** Not always.
A simple example
What is the best way to ship two items from nearby cities to the same destination far away.
A simple example

What is the best way to ship two items from nearby cities to the same destination far away.

First Attempt: Move them directly to their destination.
A simple example

What is the best way to ship two items from nearby cities to the same destination far away.

Another way: put them on the same truck and transport together!
A V-shaped path

A Y-shaped path

Answer: Transporting two items together might be cheaper than the total cost of transporting them separately. As a result,

- A “Y shaped” path is preferable to a “V shaped” path.
- Here, the cost is naturally given by the actual transport “path”, while the transport maps for both types are trivially same. Knowing only maps is not enough here.

In general, a ramified structure might be more efficient than a “linear” structure consisting of straight lines.
Examples of Ramified Structures

- Trees
- Circulatory systems
- Cardiovascular systems
- Railways, Airlines
- Electric power supply
- River channel networks
- Post office mailing system
- Urban transport network
- Marketing
- Ordinary life
- Communications
- Superconductor

Conclusion: Ramified structures are very common in living and non-living systems. It deserves a more general theoretic treatment.
Problem: Given two arbitrary probability measures $\mu^+$ and $\mu^- \in P(X)$ on a convex compact subset $X \subset \mathbb{R}^m$, find an optimal path transporting $\mu^+$ to $\mu^-$. 

Need:

- A class of “transport paths”.
  - Broad enough to ensure the existence of optimal transport paths;
- A reasonable cost functional on the category.
  - Optimal transport paths should allow some parts overlap in a cost efficient fashion. Should be “$Y$-shaped” rather than “$V$ shaped”.
  - Nice regularity of optimal transport paths.

Idea: figuring out simple cases first!
Atomic measures

An atomic measure is a (finite) sum of Dirac measures with positive multiplicities.

\[ a = \sum_{i} a_i \delta_{x_i} \]

for some \( x_i \in X \) and \( a_i > 0 \). Let \( A(X) \) be the space of all atomic measures on \( X \).

Question: What is a transport path between two atomic probability measures \( a \) and \( b \)?
Transport atomic measures

A transport path from $a$ to $b$ is a weighted directed graph

\[ G = \{ V(G), E(G), w : E(G) \to (0, +\infty) \} \]

satisfying Kirchhoff’s laws (for electrical circuits):

\[ \sum_{v=e^-} w(e) = \sum_{v=e^+} w(e) \]

for any interior vertex $v$.

Notation: For atomic measures $a, b \in P(X)$, let

\[ Path(a, b) \]

be the family of all transport paths from $a$ to $b$. 
Cost Functionals

Note that in general the space $\text{Path}(a, b)$ might be very large.
Cost Functionals

Note that in general the space $\text{Path}(a, b)$ might be very large.

Want: Find an optimal “Y shaped” or “ramified” transport path in $\text{Path}(a, b)$.
Cost Functionals

Note that in general the space $\text{Path}(a, b)$ might be very large.

**Want:** Find an optimal “Y shaped” or “ramified” transport path in $\text{Path}(a,b)$. Thus, we need a suitable cost functional on transport paths.
Cost Functionals

Note that in general the space $\text{Path}(a, b)$ might be very large.

Want: Find an optimal “Y shaped” or “ramified” transport path in $\text{Path}(a, b)$. Thus, we need a suitable cost functional on transport paths.

Answer: For each $G = \{V(G), E(G), w : E(G) \to (0, +\infty)\}$, define the $M_\alpha$ mass of $G$ by

$$M_\alpha (G) := \sum_{e} w(e)^\alpha \text{length}(e)$$

for some $\alpha \in [0, 1)$. 

Cost Functionals

Note that in general the space Path(a, b) might be very large.

Want: Find an optimal “Y shaped” or “ramified” transport path in Path(a,b). Thus, we need a suitable cost functional on transport paths.

Answer: For each \( G = \{V(G), E(G), w : E(G) \rightarrow (0, +\infty)\} \), define the \( M_\alpha \) mass of \( G \) by

\[
M_\alpha (G) := \sum_{e} w(e)^\alpha \text{length}(e)
\]

for some \( \alpha \in [0, 1) \).

Result: an \( M_\alpha \) mass minimizer is indeed “Y-shaped” or “ramified”.
Example 1: Two points to one point

It satisfies a balance equation:

\[ \sum_{i=1}^{3} m_i^\alpha \vec{n}_i = \vec{0}. \]

Using this equation, we have a formula to calculate the angles. In particular, if \( \alpha = 0 \), then the angles are 120°. Also, if \( \alpha = 1/2 \), then the top angle must be 90°.
Two points to two points

Three types from two points to two points

\[ \begin{align*}
&\quad \quad m_1 = m_3 \\
&\quad x_1 \quad \quad m_1 \\
&\quad m_2 = m_4 \\
&\quad x_2 \quad \quad m_2
\end{align*} \]

\[ \begin{align*}
&\quad \quad m_1 + m_2 \\
&\quad x_1 \quad \quad m_1 \\
&\quad x_2 \quad \quad m_2
\end{align*} \]

\[ \begin{align*}
&\quad \quad |m_2 - m_4| \\
&\quad x_1 \quad \quad m_1 \\
&\quad x_2 \quad \quad m_2
\end{align*} \]
Some lemmas (Xia, 2001)

**Lemma.** For any $G \in \text{Path}(a, b)$, there exists a $\tilde{G} \in \text{Path}(a, b)$ such that $\tilde{G}$ contains no cycles and

$$M_\alpha (\tilde{G}) \leq M_\alpha (G).$$
Some lemmas (Xia, 2001)

**Lemma.** For any $G \in \text{Path}(a, b)$, there exists a $\tilde{G} \in \text{Path}(a, b)$ such that $\tilde{G}$ contains no cycles and

$$M_\alpha \left( \tilde{G} \right) \leq M_\alpha \left( G \right).$$

Thus, we may consider only transport paths containing no cycles.
Some lemmas (Xia, 2001)

**Lemma.** For any \( G \in \text{Path}(a, b) \), there exists a \( \tilde{G} \in \text{Path}(a, b) \) such that \( \tilde{G} \) contains no cycles and

\[
\min(\alpha) \left( \tilde{G} \right) \leq \min(\alpha)(G).
\]

Thus, we may consider only transport paths containing no cycles.

**Lemma.** If \( G \) contains no cycles, then \( 0 < w(e) \leq 1 \) for any \( e \in E(G) \). Thus

\[
\max(G) \leq \min(\alpha)(G).
\]
Some lemmas (Xia, 2001)

Lemma. For any $G \in \text{Path}(a, b)$, there exists a $\tilde{G} \in \text{Path}(a, b)$ such that $\tilde{G}$ contains no cycles and

$$M_\alpha \left( \tilde{G} \right) \leq M_\alpha \left( G \right).$$

Thus, we may consider only transport paths containing no cycles.

Lemma. If $G$ contains no cycles, then $0 < w(e) \leq 1$ for any $e \in E(G)$.

Thus

$$M \left( G \right) \leq M_\alpha \left( G \right).$$

Now, given any two probability measures $\mu^+$ and $\mu^-$, what is a transport path from $\mu^+$ to $\mu^-$?

$$\mu^+ \rightarrow \mu^-$$
Transport general probability measures

Idea:
• Approximate $\mu^+$, $\mu^-$ by atomic measures $a_i, b_i$;
• Transport $a_i$ to $b_i$ by a graph $G_i$;
• The limit $T$ of $G_i$ (in a suitable sense) is a transportation of $\mu^+$ to $\mu^-$.  

The sequence of triples $\{a_i, b_i, G_i\}$ is called an approximating graph sequence of $T$. 
Dyadic approximation of Radon measures

Assume $X \subset Q$, a cube in $\mathbb{R}^m$ of the edge length $d$, with center $c$. Let

$$Q_i = \{ Q^h_i : h \in \mathbb{Z}^m \cap [0, 2^i)^m \}$$

be a partition of $Q$ into smaller cubes of edge length $\frac{d}{2^i}$.

For any Radon measure $\mu$ on $X$, let

$$A_i(\mu) = \sum_{Q^h_i} \mu(Q^h_i) \delta_{c^h_i}$$

where $c^h_i$ is the center of $Q^h_i$. Then, $A_i(h)$ converges to $\mu$ weakly as measures. This is called “Dyadic approximation of $\mu$”.
How to take limits of $G_i$’s? —— Duality!!

Answer: View each $G_i$ as a 1 dimensional normal current with $\partial G_i = b_i - a_i$.

Let $U \subset \mathbb{R}^m$ be any open set.

- $\mathcal{D}^n(U)$: $C^\infty$ differential $n$—forms in $U$ with compact support.
- An $n$—current is an element of the dual space $\mathcal{D}_n(U)$ of $\mathcal{D}^n(U)$. i.e. an $n$—current is a continuous linear functional on $\mathcal{D}^n(U)$. Thus, $0$—currents are just distributions.
- For any $T \in \mathcal{D}_n(U)$, its boundary $\partial T \in \mathcal{D}_{n-1}(U)$ is given by
  \[
  \partial T(\psi) = T(d\psi), \forall \psi \in \mathcal{D}^{n-1}(U).
  \]
- The mass of $T \in \mathcal{D}_n(U)$ is given by
  \[
  M(T) = \sup\{T(\omega) : |\omega| \leq 1, \omega \in \mathcal{D}^n(U)\}
  \]
- $T \in \mathcal{D}_n(U)$ is normal if $M(T) + M(\partial T) < +\infty$. 
Examples of n-current

- Oriented $n$-dimensional submanifold $M$ of $U$ with $\mathcal{H}^n(M) < +\infty$.

  $$[M](\omega) = \int_M \omega = \int_M <\omega(x), \xi(x)> \ d\mathcal{H}^n(x)$$

  for any $\omega \in \mathcal{D}^n(U)$. Note that $\partial[M] = [\partial M]$ and $M([M]) = \mathcal{H}^n(M)$.

- Differential $m - n$ forms $\phi \in \mathcal{D}^{m-n}(U)$;

  $$\phi(\omega) = \int_U \phi \wedge \omega.$$ 

- Rectifiable currents $\tau(M, \theta, \xi)$

  $$\tau(M, \theta, \xi)(\omega) = \int_M <\omega(x), \xi(x)> \theta(x) d\mathcal{H}^n(x)$$

  Here: $M$ is a rectifiable n-set, $\theta$ is a locally $\mathcal{H}^n$ integrable function and $\xi(x)$ is the orientation of $T_xM$. 
Transport paths between Radon measures

Definition. Given \( \mu^+, \mu^- \in P(X) \), a normal 1-current \( T \) is called a transport path from \( \mu^+ \) to \( \mu^- \) if there exists a sequence of approximating graphs \( \{a_i, b_i, G_i\} \) such that

\[
a_i \rightharpoonup \mu^+, \quad b_i \rightharpoonup \mu^-, \quad G_i \rightharpoonup T
\]

in the sense of distributions.

Note that we automatically have \( \partial T = \mu^+ - \mu^- \) as distributions.

For each transport path \( T \), we define

\[
M_\alpha(T) := \inf \{a_i, b_i, G_i\} \lim_{i \to \infty} \inf M_\alpha(G_i).
\]

Let \( \text{Path}(\mu^+, \mu^-) \) be the family of all transport paths from \( \mu^+ \) to \( \mu^- \).
Example: How to transport a Lebesgue measure to a Dirac measure?

First attempt:

\[
\sum_{i=1}^{n} \left( \frac{1}{n} \right)^{\alpha} l_i 
\approx C \sum_{i=1}^{n} \left( \frac{1}{n} \right)^{\alpha} = C n^{1-\alpha} \rightarrow +\infty.
\]
Example: How to transport a Lebesgue measure to a Dirac measure?

First attempt:
\[
\sum_{i=1}^{n} \left( \frac{1}{n} \right)^{\alpha} l_i 
\approx C \sum_{i=1}^{n} \left( \frac{1}{n} \right)^{\alpha} = C n^{1-\alpha} \to +\infty.
\]

Second attempt:
\[
\sum_{n=1}^{\infty} \sum_{i=1}^{2^n} \left( \frac{1}{2^n} \right)^{\alpha} l_i 
\approx C \sum_{n=1}^{\infty} \sum_{i=1}^{2^n} \left( \frac{1}{2^n} \right)^{\alpha} \frac{1}{2^n}
\]
\[
= C \sum_{n=1}^{\infty} \left( \frac{1}{2^n} \right)^{\alpha} = \frac{C}{1 - \frac{1}{2^\alpha}}
\]
In higher dimension case, if $\alpha > 1 - \frac{1}{m}$, then

$$
\sum_{n=1}^{\infty} \sum_{i=1}^{(2n)^m} \left( \frac{1}{(2n)^m} \right)^\alpha l_i
\approx C \sum_{n=1}^{\infty} \sum_{i=1}^{(2n)^m} \left( \frac{1}{(2n)^m} \right)^\alpha \frac{1}{2n}
= C \sum_{n=1}^{\infty} \left( \frac{1}{(2n)^m} \right)^\alpha 2n(m-1)
= C \sum_{n=1}^{\infty} \left( 2^{m(1-\alpha)-1} \right)^n < +\infty
$$

**Proposition. [Finite Cost] (Xia, 2001)** Suppose $\alpha > 1 - \frac{1}{m}$. For any $\mu \in P(X)$, there exists a $T \in \text{Path}(\mu, \delta_c)$ from $\mu$ to a Dirac measure $\delta_c$ with $M_\alpha(T) < +\infty$. 
Existence theorem (Xia, 2001)

**Theorem.** Given $\mu^+$ and $\mu^- \in \mathcal{M}_\Lambda(X)$, $\alpha \in (1 - \frac{1}{m}, 1]$, there exists an $M_\alpha$ mass minimizer $S$ in the family $\text{Path}(\mu^+, \mu^-)$. Moreover, $M_\alpha(S) < \frac{\Lambda^\alpha}{2^{1-m(1-\alpha)-1}} \frac{\sqrt{md}}{2}$.

**Sketch of the proof:**

- Pick $\{a_i, b_i, G_i\}$ with $M_\alpha(G_i) \downarrow \inf \{M_\alpha(T) : T \in \text{Path}(\mu^+, \mu^-)\}$
- We may assume $\{G_i\}$ has no cycleless $M(G_i) \leq M_\alpha(G_i) < C$ bounded.
- By the compactness of normal currents, $G_{i_k} \rightarrow T \in \text{Path}(\mu^+, \mu^-)$
- lower semicontinuity of $M_\alpha$. 
A new distance on $P(X)$

**Definition.** Given $\mu^+$ and $\mu^- \in P(X)$, define

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

**Theorem.** (Xia, 2001) $d_\alpha$ is a distance on $P(X)$.

Remark: $d_\alpha$ is different from any of the Wasserstein distances.

**Theorem.** (Xia, 2001) $d_\alpha$ metrizes the weak * topology of $P(X)$. 
Optimal transport paths

**Lemma.** If $G_i \in \text{Path}(a_i, b_i)$ is an $M_\alpha$ minimizer, then $T \in \text{Path}(\mu^+, \mu^-)$ is also an $M_\alpha$ minimizer in $\text{Path}(\mu^+, \mu^-)$.

**Definition.** A transport path $T \in \text{Path}(\mu^+, \mu^-)$ is called an optimal transport path if there exists a sequence of approximating graphs $\{a_i, b_i, G_i\}$ such that each $G_i \in \text{Path}(a_i, b_i)$ is an $M_\alpha$ minimizer.
Error estimate

By the lemma, we can pick our favorite approximating atomic measures \( \{a_i\}, \{b_i\} \).
We choose “dyadic approximation” \( \{A_n(\mu)\} \).

**Proposition.** For any \( \mu \in P(X) \),

\[
d_\alpha(\mu, A_n(\mu)) \leq C\lambda^n
\]

with some constant \( C > 0 \) and \( \lambda = 2^{m(1-\alpha) - 1} \in (0, 1) \).

**Corollary.** If each \( G_n \) is optimal, then

\[
M_\alpha(T) \leq M_\alpha(G_n) + 2C\lambda^n
\]
**Length Space Property**

**Theorem.** (Xia, 2002) $(P(X), d_\alpha)$ is a length space.

That is, for any $\mu^+, \mu^- \in P(X)$, there exists a continuous map

$$\psi : [0, t] \rightarrow (P(M), d_\alpha)$$

with $t = d_\alpha(\mu^+, \mu^-)$ such that

$$\psi(0) = \mu^+, \psi(T) = \mu^-$$

and for any $0 \leq s_1 < s_2 \leq t$,

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

In other words, an optimal transport path between Radon measures plays the role of a **geodesic** between two points.

Later, we will see that in fact each $\psi(s)$ is purely atomic for any $0 < s < t$. 

![Diagram of optimal transport path between Radon measures](image-url)
Atomic approximation ($\alpha = 0.1$)

![Graphs showing atomic approximation with different total values.](image)
Atomic approximation ($\alpha = 0.5$)
Atomic approximation \((\alpha = 0.95)\)
From Lebesgue to Dirac
Transporting general measures
optimal transportation of 50 random points with alpha=0.95
Transport Path & Transport Plan
Let $a$ and $b$ be any two atomic measures. For example,
Transport Path & Transport Plan

Let $a$ and $b$ be any two atomic measures. For example,

- Each transport plan $\gamma \in \text{Plan}(a, b)$ is given by a real valued matrix
  
  \[
  U = (u_{ij}).
  \]

  e.g.

  \[
  U_1 = \begin{pmatrix}
    1/4 & 1/2 & 0 \\
    0 & 1/5 & 1/12 \\
    1/4 & 0 & 1/12
  \end{pmatrix}
  \quad \text{or} \quad
  U_2 = \begin{pmatrix}
    0 & 1/2 & 1/4 \\
    1/4 & 1/5 & 0 \\
    1/4 & 1/12 & 0
  \end{pmatrix}
  \]

- Each transport path $G \in \text{Path}(a, b)$ gives a 1-current valued matrix
  
  \[
  g(G) = (g_{ij}). \ (\text{no cycles!})
  \]

\[
X_1, X_2, X_3, Y_1, Y_2
\]
Compatible Pair of Transport Path & Plan

A transport path $G$ and a transport plan $\gamma$ are said to be compatible if

$$G = \sum u_{ij} \cdot g_{ij}.$$ 

A compatible pair gives a decomposition of $G$. 

Compatible Pair of Transport Path & Plan

A transport path $G$ and a transport plan $\gamma$ are said to be compatible if

$$G = \sum u_{ij} \cdot g_{ij}.$$ 

A compatible pair gives a decomposition of $G$.

For instance, $U_1$ is compatible with $G$ while $U_2$ is not.
Compatible Pair of Transport Path & Plan

A transport path $G$ and a transport plan $\gamma$ are said to be compatible if

$$G = \sum u_{ij} \cdot g_{ij}.$$  

A compatible pair gives a decomposition of $G$. For instance, $U_1$ is compatible with $G$ while $U_2$ is not.

A compatible pair of transport path and transport plan provides the necessary transporting information by its unique matrix representation $((u_{ij}), (g_{ij}))$.

$u_{ij} =$ amount of mass from $x_i$ to $y_j$, while $g_{ij} =$ actual transport path.
Some Results (Xia, 2001)

- There exists \( G \in Path(a, b) \) compatible with all \( \gamma \in Plan(a, b) \).
- For any \( G \in Path(a, b) \), there exists a \( \gamma \in Plan(a, b) \) compatible with \( G \).
- Given a transport plan \( \gamma \in Plan(\mu^+, \mu^-) \), there exists an optimal transport path \( T \in Path(\mu^+, \mu^-) \) with least finite \( M_\alpha \) cost among all compatible pairs \((T, \gamma)\). (mailing problem)
- Given a transport path \( T \in Path(\mu^+, \mu^-) \), there exists an optimal transport plan \( \gamma \in Plan(\mu^+, \mu^-) \) with least \( I(\gamma) \) cost among all compatible pairs \((T, \gamma)\).
How nice is an optimal transport path?

Let $T \in \text{Path}(\mu^+, \mu^-)$ be any transport path with $M_\alpha(T) < +\infty$, not necessarily optimal.

**Theorem.** *(rectifiability) (Xia, 2001)* $T$ is a real multiplicity 1-rectifiable current $T = \tau(M, \theta, \xi)$ with $\partial T = \mu^+ - \mu^-$. Moreover,

$$M_\alpha(T) = \int_M \theta(x)^\alpha d\mathcal{H}^1(x)$$

Idea of proof: Follows from the rectifiable slicing theorem.

Now, assume that $T$ is optimal. Let us see how nice $T$ is.
Interior regularity: a local finiteness property (Xia, 2002)

Suppose one of $\mu^+$ or $\mu^-$ is atomic. For any $p \in spt(T) \setminus spt(\partial T)$, there exists an open ball neighborhood $B_p$ of $p$ such that

$$T|_{B_p}$$

is a cone at $p$ consisting of finite union of segments with suitable multiplicities. These segments are balanced by a simple balance equation.
How about the boundary?

Observation: The support of $T$ may not necessarily be 1-dimensional nearby its boundary, which is the difference of the given two measures. This is because the boundary itself may even be dense in the space, as demonstrated by letting the initial measure to be the Lebesgue measure.
How about the boundary?

Observation: The support of $T$ may not necessarily be 1-dimensional nearby its boundary, which is the difference of the given two measures. This is because the boundary itself may even be dense in the space, as demonstrated by letting the initial measure to be the Lebesgue measure.

Solution: Relax yourself and enjoy the nature.

The nature has provided a wonderful solution for us: the leaf vein.
How about the boundary?

Observation: The support of $T$ may not necessarily be 1-dimensional nearby its boundary, which is the difference of the given two measures. This is because the boundary itself may even be dense in the space, as demonstrated by letting the initial measure to be the Lebesgue measure.

Solution: Relax yourself and enjoy the nature.

The nature has provided a wonderful solution for us: the leaf vein.

But, how to read this information?
Boundary Regularity

To understand the boundary behavior, a suitable approach is to study the “level sets” of the rectifiable current $T = \tau(M, \theta, \xi)$ instead. For each $\lambda > 0$, let

$$M_\lambda = \{ x \in M : \theta(x) \geq \lambda \}.$$
Boundary Regularity

To understand the boundary behavior, a suitable approach is to study the “level sets” of the rectifiable current $T = \tau(M, \theta, \xi)$ instead. For each $\lambda > 0$, let

$$M_\lambda = \{x \in M : \theta(x) \geq \lambda\}.$$ 

Theorem (Xia, 2003): Each level set of an optimal transport path is locally concentrated on a finite union of bilipschitz curves. These curves enjoy some nice properties similar to those satisfied by segments near an interior point.
Key Idea of Proof: Decomposition!

- For any optimal weighted directed graph $G \in \text{Path}(a, b)$, if $M^\alpha(a) + M^\alpha(b)$ is bounded above, then we can decompose $a, b, G$

  $$a = a_P + a_R, \quad b = b_P + b_R, \quad G = P + R$$

  so that $P \in \text{Path}(a_P, b_P), R \in \text{Path}(a_R, b_R)$, the total number of vertices and edges of $P$ are uniformly bounded. The level set $G_\lambda$ is contained in $P$. Edges of $P$ are “nice”.

- Taking the limits to get the decomposition of optimal transport paths.

  Advantage: Graphs are much easier to deal with. Just using combinatory.
Key Idea of Proof: Decomposition!

- For any optimal weighted directed graph $G \in \text{Path}(a, b)$, if $M^\alpha(a) + M^\alpha(b)$ is bounded above, then we can decompose $a, b, G$

$$a = a_P + a_R, \quad b = b_P + b_R, \quad G = P + R$$

so that $P \in \text{Path}(a_P, b_P), R \in \text{Path}(a_R, b_R)$, the total number of vertices and edges of $P$ are uniformly bounded. The level set $G_\lambda$ is contained in $P$. Edges of $P$ are “nice”.

- Taking the limits to get the decomposition of optimal transport paths.

  Advantage: Graphs are much easier to deal with. Just using combinatorial.

  Feedback? A natural question: Can we use this idea to understand the dynamic formation of a tree leaf?
Key Idea of Proof: Decomposition!

For any optimal weighted directed graph $G \in Path(a, b)$, if $M^\alpha(a) + M^\alpha(b)$ is bounded above, then we can decompose $a, b, G$

$$a = a_P + a_R, b = b_P + b_R, G = P + R$$

so that $P \in Path(a_P, b_P), R \in Path(a_R, b_R)$, the total number of vertices and edges of $P$ are uniformly bounded. The level set $G_\lambda$ is contained in $P$. Edges of $P$ are “nice”.

Taking the limits to get the decomposition of optimal transport paths.

Advantage: Graphs are much easier to deal with. Just using combinatorial.

Feedback? A natural question: Can we use this idea to understand the dynamic formation of a tree leaf?

YES!! (Xia, 2004)
\[ \alpha = 0.6, \quad \beta = 0.5, \quad \varepsilon = 2 \]

\[ \alpha = 0.6, \quad \beta = 0.5, \quad \varepsilon = 3 \]

\[ \alpha = 0.6, \quad \beta = 0.5, \quad \varepsilon = 4 \]

\[ \alpha = 0.6, \quad \beta = 0.5, \quad \varepsilon = 5 \]
\( \alpha = 0.66, \beta = 0.7, \text{ totalcost} = 525.9653 \)
$\alpha = 0.68, \, \beta = 0.38, \, \text{total cost} = 49.5418$
Question: Given a measure $\mu, \nu$, for which $\alpha$, will we have $d_\alpha(\mu, \nu) < +\infty$?

For simplicity, we choose $\nu = \text{Dirac mass}$.

Recall that if $\mu = \text{Lebesgue measure}$ and $\alpha > 1 - \frac{1}{m}$, then

$$\sum_{n=1}^{\infty} \sum_{i=1}^{(2^n)^m} \left( \frac{1}{(2n)^m} \right)^\alpha l_i$$

$$\approx C \sum_{n=1}^{\infty} \sum_{i=1}^{(2^n)^m} \left( \frac{1}{(2n)^m} \right)^\alpha \frac{1}{2n}$$

$$= C \sum_{n=1}^{\infty} \left( \frac{1}{(2n)^m} \right)^\alpha 2^{n(m-1)}$$

$$= C \sum_{n=1}^{\infty} \left( 2^{m(1-\alpha)-1} \right)^n < +\infty$$

Here, dimension $m = \inf_{\alpha < 1} \left\{ \frac{1}{1-\alpha} : d_\alpha(\mu, \delta_0) < +\infty \right\}$
**Dimensional distance**

For any $\mu, \nu \in P(X)$, let

$$D(\mu, \nu) = \inf_{\alpha < 1} \left\{ \frac{1}{1 - \alpha} : d_\alpha(\mu, \nu) < +\infty \right\}$$

**Proposition.** $(P(X), D)$ is a pseudometric space.

That is, $D$ is a metric except that $D(\mu, \nu) = 0$ does not imply $\mu = \nu$.

*E.g.* $D(\delta_x, \delta_y) = 0$ for any $x, y \in X$ because $d_\alpha(\delta_x, \delta_y) = |x - y| < +\infty$, $\forall \alpha$.

**Definition.** For any $\mu$ and $\nu$, we say $\mu \simeq \nu$ if $D(\mu, \nu) = 0$. That is, $\mu$ and $\nu$ are equivalent if and only if $d_\beta(\mu, \nu) < +\infty$ for any $\beta$. The equivalent class of $\mu$ is denoted by $[\mu]$.

**Lemma.** If $\mu_1 \simeq \mu_2$, then for any $\nu$, $D(\mu_1, \nu) = D(\mu_2, \nu)$.

Thus, we may define

$$D([\mu], [\nu]) := D(\mu, \nu)$$
**Dimensional Distance**

**Theorem.** *(Xia, 2007)* $D$ defines a metric on the equivalent classes of probability measures.

In general, we have

$$d_{Haus}(spt(\mu)) \leq D(\mu, \delta_0) \leq d_{box}(spt(\mu)).$$

Thus, when support of $\mu$ is nice enough, we get

$$\text{dimension of } spt(\mu) = \text{the distance } D(\mu, \delta_0).$$

As a result, I call $D$ dimensional distance.

**Conclusion:** Dimension of a set/measure is just the distance from it to a Dirac mass.
Example: \( \mu = \text{Lebesgue measure}, \ \nu = \text{Dirac mass} \)

If \( \alpha > 1 - \frac{1}{m} \), i.e., \( m > \frac{1}{1 - \alpha} \) then

\[
\sum_{n=1}^{\infty} \sum_{i=1}^{(2^n)^m} \left( \frac{1}{(2n)^m} \right)^\alpha l_i
\]

\[
\approx C \sum_{n=1}^{\infty} \sum_{i=1}^{(2^n)^m} \left( \frac{1}{(2n)^m} \right)^\alpha \frac{1}{2^n}
\]

\[
= C \sum_{n=1}^{\infty} \left( \frac{1}{(2n)^m} \right)^\alpha 2^n(m-1)
\]

\[
= C \sum_{n=1}^{\infty} \left( 2^{m(1-\alpha)-1} \right)^n < +\infty
\]

So, the dimension of \( \mu = m = \inf_{\alpha < 1} \left\{ \frac{1}{1 - \alpha} : d_\alpha(\mu, \delta_0) < +\infty \right\} \)

\[
= D(\mu, \nu), \ \text{the distance from } \mu \text{ to } \delta_0
\]
Example: $\mu = \text{Cantor set}, \ \nu = \text{Dirac mass}$

$$\sum_{n=1}^{\infty} 2^n \left( \frac{1}{2n} \right)^{\alpha} \left( \frac{1}{3} \right)^n = \sum_{n=1}^{\infty} \left( \frac{2^{1-\alpha}}{3} \right)^n < \infty$$

$$\iff \frac{2^{1-\alpha}}{3} < 1$$

$$\iff 2^{1-\alpha} < 3$$

$$\iff 1 > \frac{\ln 2}{1 - \alpha} = \frac{\ln 3}{\ln 3}$$

Here again,

the dimension of $\mu = \frac{\ln 2}{\ln 3} = \inf_{\alpha < 1} \left\{ \frac{1}{1 - \alpha} : d_{\alpha}(\mu, \delta_0) < +\infty \right\}$

$= D(\mu, \nu)$, the distance from $\mu$ to $\delta_0$

Note, here $\alpha$ is allowed to be negative.
Example: \( \mu = \text{Fat Cantor set}, \quad \nu = \text{Dirac mass} \)

Examples: \( \mu = \text{Fat } \lambda \text{ Cantor set (i.e. remove an interval of length } \lambda \text{ from the middle of } [0, 1]). \)

\[
\sum_{n=1}^{\infty} 2^n \left( \frac{1}{2^n} \right)^\alpha \frac{1 + \lambda}{4} \left( \frac{1 - \lambda}{2} \right)^{n-1} = \frac{1 + \lambda}{2(1 - \lambda)} \sum_{n=1}^{\infty} \left( 2^{1-\alpha} p \right)^n < \infty
\]

\[
\iff 2^{1-\alpha} p < 1
\]

\[
\iff 2^{1-\alpha} < \frac{1}{p}
\]

\[
\iff \frac{1}{1 - \alpha} > -\frac{\ln 2}{\ln p} = \frac{\ln 2}{\ln 2 - \ln (1 - \lambda)}
\]

where \( p = \frac{1 - \lambda}{2} \).

Again, we have dimension of \( \mu = \inf_{\alpha < 1} \left\{ \frac{1}{1 - \alpha} : d_\alpha(\mu, \delta_0) < +\infty \right\} \)
Example: $\mu =$self-similar set, $\nu =$Dirac mass

Example: $A =$finite union of $A_i$ for $i = 1, \cdots k$. Each $A_i$ is a $\sigma -$rescale of $A$.

$$\sum_{n=1}^{\infty} k^n \left( \frac{1}{k^n} \right)^\alpha \sigma^{n-1} L = \frac{L}{\sigma} \sum_{n=1}^{\infty} \left( k^{1-\alpha} \sigma \right)^n < +\infty$$

$$\iff k^{1-\alpha} \sigma < 1$$

$$\iff \frac{1}{1-\alpha} > -\frac{\ln k}{\ln \sigma}$$

Therefore, $D(\mu) = -\frac{\ln k}{\ln \sigma}$.

Here again, self-similar dimension of $\mu = \inf_{\alpha<1} \{ \frac{1}{1-\alpha} : d_\alpha(\mu, \delta_0) < +\infty \}$
Thank You and Enjoy the Nature