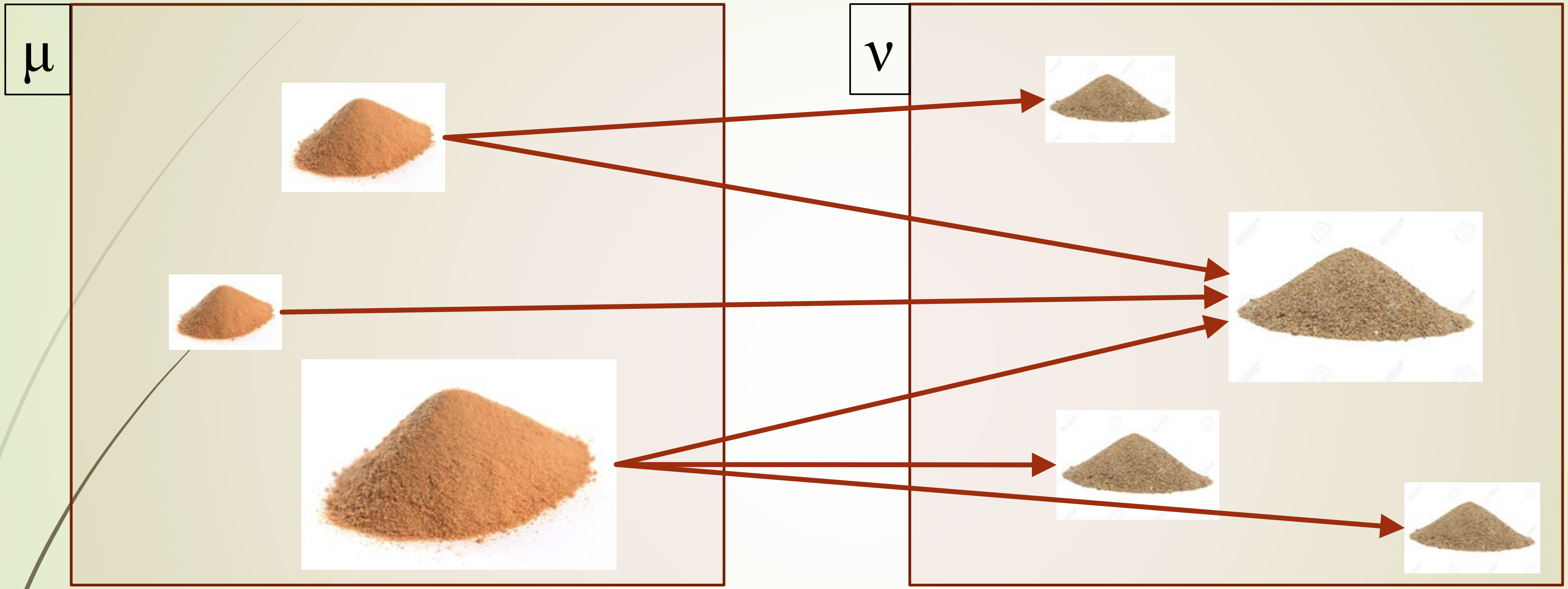
Impossibility of Sketching of the 3D Transportation Metric with Quadratic Cost

Alexandr Andoni, Assaf Naor, Ofer Neiman

Transportation Metrics

- A distance function between probability distributions
- Let (X,d) be some underlying metric space (e.g. $\mathbb{R}^3$)
- Let μ,ν be two probability distributions over X
- A coupling π is a distribution over $X \times X$ with marginals μ,ν
  - Intuitively, it is a matching between the mass of μ to that of ν

Moving Sand..



Wasserstein Distances

- The Wasserstein-p distance W_p is

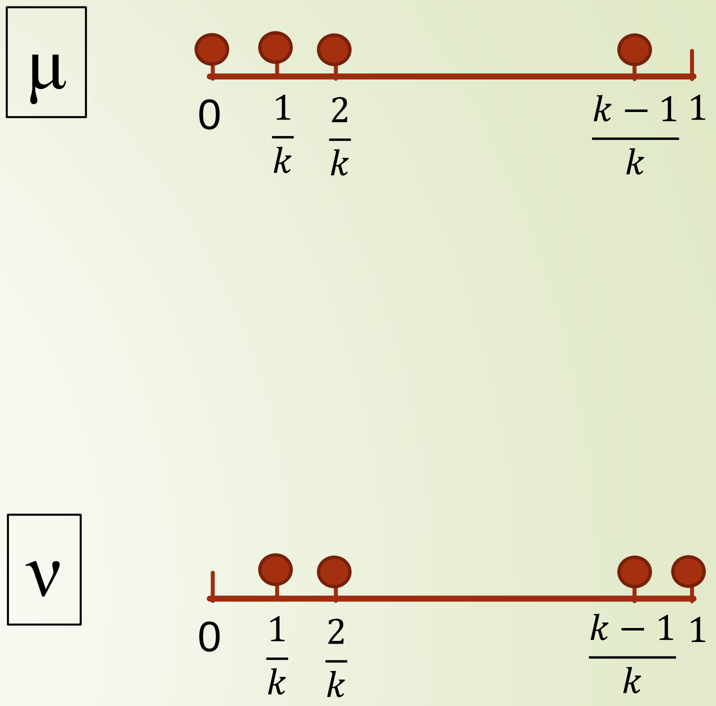
$$\inf_{\pi}\left(\iint_{X\times X} d(x,y)^p \, d\pi(x,y)\right)^{1/p}$$

- That is, for each grain of sand we take the distance it has to travel under π, and take the p-norm of these distances
- The W_1 distance is also known as EMD (Earth Mover Distance)
- The W_2 distance generalizes RMS (Root-Mean-Square) distance

Example

- 2 distributions over the real line
- The (un-normalized) W_1 distance between them is 1
- The (un-normalized) W_2 distance between them is

$$\sqrt{\sum_{i=1}^{k}\left(\frac{1}{k}\right)^2} = \frac{1}{\sqrt{k}} \longrightarrow 0$$

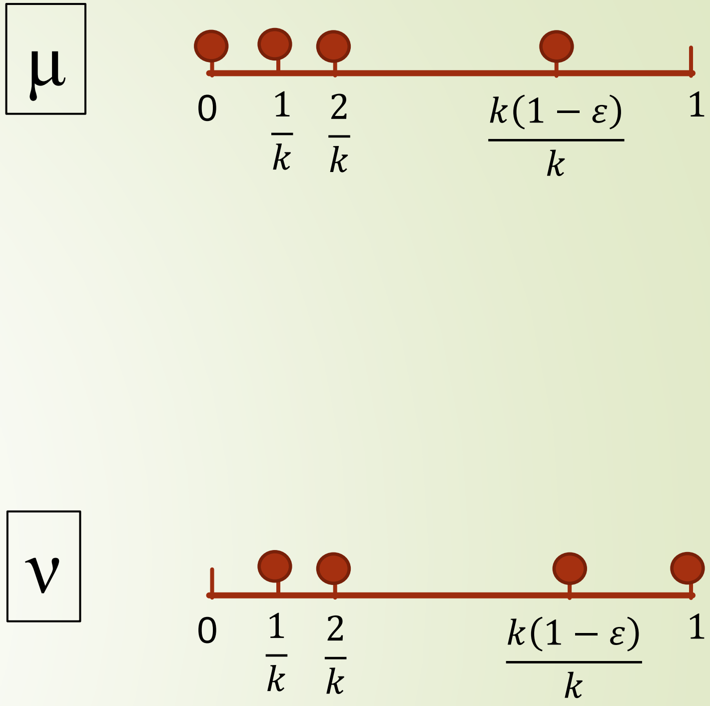


Example

- 2 distributions over the real line
- The W_1 distance between them is 1
- The W_2 distance between them is

$$\sqrt{\sum_{i=1}^{k}\left(\frac{1}{k}\right)^2} = \frac{1}{\sqrt{k}} \longrightarrow 0$$

- A “gap” of size εk would increase the W_2 distance to ≈ ε



Applications

- Transportation metrics play an important in various CS areas
  - Vision, learning, pattern recognition…
- Research on the computational properties of these metrics:
  - Efficiently computing the distance between distributions (with finite supports),
  - Nearset-Neighbor search, clustring
  - Embedding, sketching, etc.
- Almost all results were derived for W_1 (EMD)
- However, in some applications, W_2 is a more natural distance than W_1

Our Results

- W_2 over $\mathbb{R}^3$ cannot be represented faithfully in a rich class of normed spaces
  - In particular, for measures supported on n points:
    - Distortion $\Omega(\sqrt{\log n})$ is required for any L_1 embedding.
    - Distortion $\Omega(\sqrt{\log n})$ is required for any constant size sketching
  - For W_p over $\mathbb{R}^3$, $1<p<\infty$, the lower bound is $\Omega(\sqrt[p]{\log n})$
- More generally, W_2 over $\mathbb{R}^3$ does not admit a coarse embedding into any Banach space of non-trivial type

Snowflake Universality

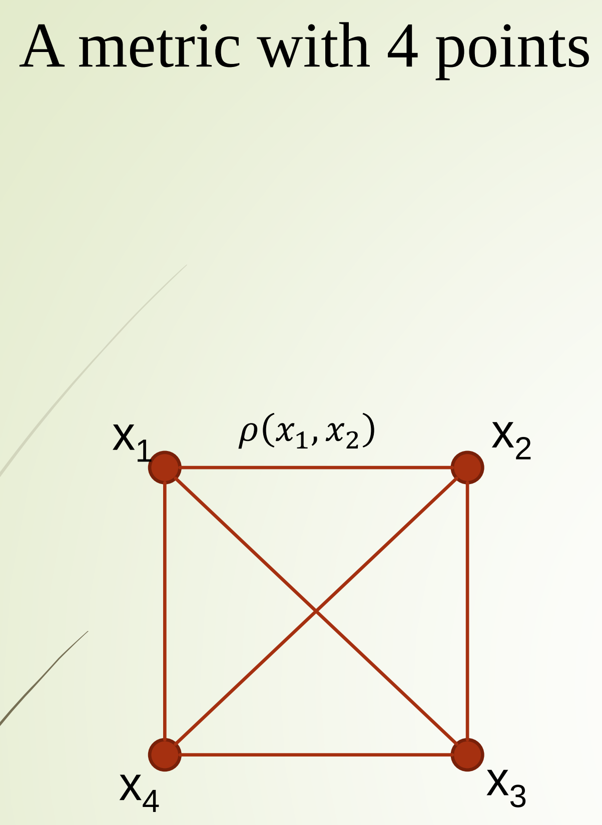
- The ½-snowflake of a metric (Y,ρ) is the metric $(Y,\rho^{1/2})$
- Thm: W_2 over $\mathbb{R}^3$ is ½-snowflake universal.
  - It contains all finite ½-snowflakes with distortion arbitrarily close to 1
- More generally W_p over $\mathbb{R}^3$ is (1/p)-snowflake universal for $1<p<\infty$
- Known lower bounds for embedding and sketching translate to the snowflake version (loosing a √ factor)

Tightness of Snowflake Universality

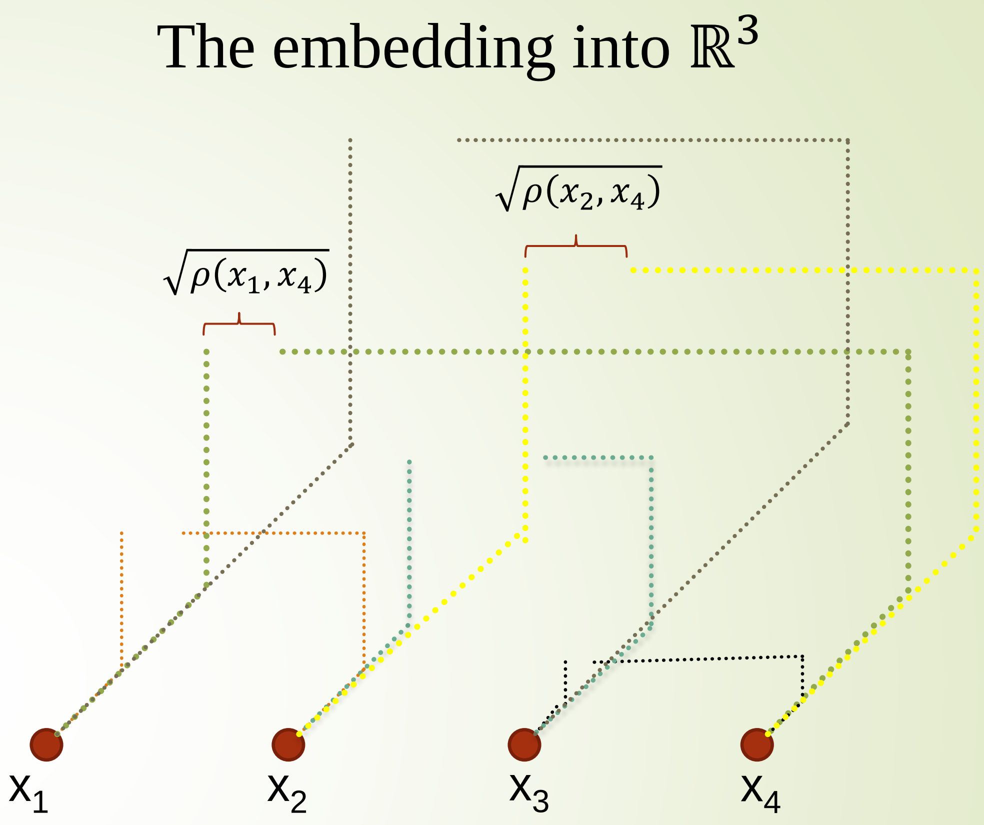
- Thm: There exists an n-point metric space that for any ½<α<1, its α-snowflake requires distortion at least $\Omega(\log n)^{\alpha-1/2}$ when embedded into W_2 over $\mathbb{R}^3$
- Similar lower bound: $\Omega(\log n)^{\alpha-1/p}$ for embedding into W_p over $\mathbb{R}^3$ for 1<p<2

The Embedding

- Given a finite metric (Y,ρ), let G=(Y,E) be a complete graph with edge weights given by ρ
- Any graph can be represented in $\mathbb{R}^3$ without edge crossings
  - Use such a representation where vertices and non-neighboring edges are far
- Replace each edge by a set of (Steiner) points of distance $\frac{1}{k}$ between consecutive points (k is very large)
  - For the edge (u,v) introduce a gap of length proportional to $\sqrt{\rho(u,v)}$
- Embed $u \in Y$ into $f(u)$: a uniform distribution over all the (Steiner) points, and the point representing u


A metric with 4 points
x1
x2
x3
x4
$\rho(x_1, x_2)$

f


The embedding into $\mathbb{R}^3$
$\sqrt{\rho(x_1, x_4)}$
$\sqrt{\rho(x_2, x_4)}$
x1
x2
x3
x4

Proof of Snowflake Universality

- Lemma: For any $\varepsilon > 0$, if k is sufficiently large, then for all $u, v \in Y$:
  $$\sqrt{\rho(u,v)} \le C \cdot W_2\big(f(u), f(v)\big) \le (1+\varepsilon) \cdot \sqrt{\rho(u,v)}$$

- Right inequality: there exists a transport plan of cost at most $(1+\varepsilon) \cdot \sqrt{\rho(u,v)}$ that moves the mass along the u-v edge
- Left inequality: need to show that any transport has cost $\ge \sqrt{\rho(u,v)}$
  - Transport plan that moves mass between non-neighboring edges pay a lot
  - If a transport plan goes along the path $u = u_0, u_1, \dots, u_t = v$, then it will pay at least the sum of gaps to power 2
  - The triangle inequality implies this cost will be $\ge \sqrt{\rho(u,v)}$

Open Problems

- Is W_p over $\mathbb{R}^3$ ½-snowflake universal for p>2?
  - Our lower bounds are sharp only for $1<p\leq 2$
- Does W_2 over $\mathbb{R}^3$ embed into W_p over $\mathbb{R}^3$?
  - A natural counterpart to the fact: L_2 embeds into any L_p
- Is W_2 over $\mathbb{R}^2$ snowflake universal?
- Is W_1 over $\mathbb{R}^k$ universal? (i.e. contains all finite metrics)
  - Bourgain showed W_1 over ℓ_1 is indeed universal

Open Problems

- Does any n point metric embed to W_2 over $\mathbb{R}^3$ with distortion $\sqrt{\log n}$?
  - Our result imply an embedding with distortion $\sqrt{\Delta}$, where Δ is the aspect ratio of the metric
    - In particular, expanders embed with distortion $\sqrt{\log n}$
  - A positive resolution will solve the metric cotype dichotomy
  - (Just the fact that W_2 over $\mathbb{R}^3$ is ½-snowflake universal is not enough)
- Embedding W_2 over $\mathbb{R}^3$ into L_1?
  - Lower bound $\Omega\left(\sqrt{\log n}\right)$, nothing better than $O(\sqrt{n})$ is known..