Metric Distortion of Random Sets

Uri Grupel
Joint work with Mark Braverman and Itai Benjamini

Online Asymptotic Geometric Analysis Seminar, April 2020
Random Measures

- Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be independent uniform random points on the circle $S^1$.

- In Bobkov’s talk (about his work with Ledoux) we defined random measures

$$
\mu_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k}, \quad \nu_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{Y_k},
$$

and looked at the Wasserstein distance $W_p(\mu, \nu)$.

- The optimal solution is of the same order as passing through the evenly spaced (average) case

$$
\sigma_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{e^{2\pi i k/n}}.
$$
The Setup

- The metric distortion of $A, B \subseteq S^1$ with cardinality $n$ is

$$\text{dist}(A, B) = \inf \{ \| f \|_{Lip} \| f^{-1} \|_{Lip} ; \ f : A \rightarrow B \ \text{bijection} \}.$$ 

- Assume that $A, B$ is defined by $n$ i.i.d uniform points.
  - What is the "typical" $\text{dist}(A, B)$?
  - Let $C \subseteq S^1$ be $n$ evenly spaced points. How does $\text{dist}(A, C)$ compare to $\text{dist}(A, B)$?
Remarks on the Question

- Similarly, one may ask if the following algorithm is good:
  - Choose starting points in both $A$ and $B$.
  - Move the $i$–th point in $A$ to the $i$–th point in $B$.
  - Optimize over all choices of starting points.

- This question does not make sense in a non probabilistic setting (worst case scenario).
Structure of Random Sets

Let \( \{X_1, \ldots, X_n\} \subseteq S^1 \) be a random set, ordered (by choosing a starting point).

- There are no unusual accumulations or empty stretches. For any \( s > 10 \), with probability at least \( 1 - Cn^{1-s/8} \) we have
  \[
  \frac{j - i}{4n} \leq |X_j - X_i| \leq \frac{4(j - i)}{n},
  \]
  for any \( i, j \) such that \( \max\{|j - i|, |n + 1 - j - i|\} > s \log n \)

- Let \( Y_i = X_{i+1} - X_i \) then \((Y_1, \ldots, Y_{n-1})\) is a uniform random vector on the boundary of the \( n - 1 \) dimensional simplex.

- By standard arguments the distribution of \( Y_i \) is the same as \( Z_i/(Z_1 + \cdots + Z_{n-1}) \) where \( Z_1, \ldots, Z_{n-1} \) are i.i.d exponential random variables. The sum \( Z_1 + \cdots + Z_{n-1} \) is concentrated around \( n \).
Distortion to Evenly Spaced Set

- Since the distances in a random sets are comparable to exponent random variables:
  
  \[ \mathbb{P} \left( \frac{1}{n^2 \log n} \leq \min |X_{i+1} - X_i| \leq \frac{\log n}{n^2} \right) \to 1. \]

- \[ \mathbb{P} \left( \frac{1}{n} \leq \max |X_{i+1} - X_i| \leq \frac{10 \log n}{n} \right) \to 1. \]

- By these bounds, with high probability the metric, the metric distortion is \( \Theta^*(n) \).
Let $\beta(A) = \min |X_{i+1} - X_i|$, then $\text{dist}(A, B) \geq \beta(A)/\beta(B)$.

Let $Y_1, \ldots, Y_n$ be the points in $B$. Then,

$$\mathbb{P}(\beta(B) \leq t) \geq \mathbb{P}(|Y_2 - Y_1| \leq t) \geq Ct.$$ 

$$\mathbb{P} \left( \beta(A) \geq \frac{1}{2n^2} \right) \geq \frac{1}{2}.$$ 

$$\mathbb{P} \left( \text{dist}(A, B) \geq t \right) \geq \mathbb{P} \left( \beta(A) \geq \frac{1}{2n^2} \text{ and } \beta(B) \leq \frac{1}{2tn^2} \right) \geq \frac{C}{tn^2}.$$ 

$$\mathbb{E}\text{dist}(A, B) \geq \frac{C}{n^2} \int_1^\infty \frac{1}{t} dt = \infty.$$
Main Results

Theorem
For $n$ large enough, we have

$$\mathbb{P}\left( \exists f : A \to B, \ dist(f) \leq n^{2/3-\varepsilon} \right) \leq \frac{3}{8}.$$

Theorem
For $n$ large enough, we have

$$\mathbb{P}\left( \exists f : A \to B, \ dist(f) \leq n^{2/3+\varepsilon} \right) \geq \frac{5}{8}.$$
By choosing a starting point, we associate a function $f : A \to B$ with a permutation $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$.

We define (reverse) permutation distortion by

\[
\begin{align*}
  s(f) &:= \max_{i=1,\ldots,n} |i - \pi(i)| \\
  s_r(f) &:= \max_{i=1,\ldots,n} |i - (n + 1 - \pi(i))|.
\end{align*}
\]

Let $0 < \alpha < 1/2$. With high probability, no map $f : A \to B$ with $\min\{s(f), s_r(f)\} \leq n^\alpha$ has $\|f\|_{Lip} \leq n^{1-\alpha}/\log^2 n$.

Sketch of the proof:

- Consider the shortest segment. It is of length smaller than $\log n/n^2$.
- It is mapped to one of $2n^\alpha$ consecutive segment.
- With high probability, the shortest segment there is bigger than $n^{1-\alpha}/\log n$. 

Let $\varepsilon > 0$ and let $0 < \alpha < 1$. Let $f : A \to B$ be a bi-Lipschitz function such that $\|f\|_{\text{Lip}}\|f^{-1}\|_{\text{Lip}} \leq n^{\alpha - \varepsilon}$ and $\min\{s(f), s_r(f)\} \geq n^\alpha$. Then,

$$\mathbb{P} \left( \exists i; \min\{|i - \pi(i)|, |n + 1 - i - \pi(i)|\} > n^{\alpha - \varepsilon/2} \right) \to 1.$$ 

- We partition the indices to left $\{i \leq n/2 - n^{\alpha - \varepsilon/2}\}$, center and right $\{i \geq n/2 + n^{\alpha - \varepsilon/2}\}$.
- An index is projected if it is mapped $n^{\alpha - \varepsilon/4}$ close to itself and reflected if to $n + 1 - i$.
- Either all the left (right) is projected or reflected. Otherwise there is a pair that drift too far apart. A contradiction to $\min\{|i - \pi(i)|, |n + 1 - i - \pi(i)|\}$.
Theorem (Benjamini and Shamov)
Let $\pi : \mathbb{Z} \to \mathbb{Z}$ be a bi-Lipschitz bijection. Then,

$$\pi(x) = \pm x + \text{const} + r(x),$$

where $|r(x)| \leq \|\pi\|_{\text{Lip}} \|\pi^{-1}\|_{\text{Lip}}$.

Lemma
Let $0 < \alpha < 1$ and let $\epsilon > 0$. Then, with high probability there is no $f : A \to B$ bi-Lipschitz such that

$$\|f\|_{\text{Lip}} \|f^{-1}\|_{\text{Lip}} < n^{\alpha - \epsilon} \text{ and } \min\{s(f), s_r(f)\} \geq n^\alpha.$$
Lower Bound

Define the events

\[ E_1 = \left\{ \exists f : A \to B \mid \|f\|_{\text{Lip}} < n^{1/3 - \varepsilon/2} \quad \& \quad \min\{s(f), s_r(f)\} < n^{2/3} \right\} \]

\[ E_2 = \left\{ \exists g : B \to A \mid \|g\|_{\text{Lip}} < n^{1/3 - \varepsilon/2} \quad \& \quad \min\{s(g), s_r(g)\} < n^{2/3} \right\} \]

\[ E_3 = \left\{ \exists f : A \to B \mid \|f\|_{\text{Lip}} \|f^{-1}\|_{\text{Lip}} < n^{2/3 - \varepsilon/2} \quad \& \quad \min\{s(g), s_r(g)\} \geq n^{2/3} \right\} \]

Then,

\[ \left\{ \text{dist}(A, B) \leq n^{2/3 - \varepsilon} \right\} \subseteq E_1 \cup E_2 \cup E_3, \]

and each of them occur with probability less than 1/8 (when \( n \) is big enough).
Upper Bound - The Setup

- We partition our sets to segments of $n^{1/3}$ points.
- We group together consecutive $n^{1/3}$ segments.
- We condition of the high probability events:
  - Up to a constant, each segment is of length $n^{-2/3}$.
  - The shortest distance between any two points is bigger than $1/n^2$.
  - The longest distance between any two points is smaller than $1/n^{1+\varepsilon}$.
- Different segments are “almost” independent.
- In each segment we are interested in two scales:
  - Short intervals, of length less than $n^{-4/3 - \varepsilon}$.
  - Long intervals, of length bigger than $n^{-4/3 - \varepsilon}$.
Upper Bound - The Mapping

We base the map between $A$ and $B$ on the following observations:

- If two segments have the same sequences of short intervals they are mapped to each other with distortion less than $n^{2/3+2\varepsilon}$.

- If we increase the short scale to $n^{-4/3+\varepsilon}$, then the distortion is still bounded by $n^{2/3+2\varepsilon}$.

- If each segments ends and start with a long interval, and two grouping have the same sequences of segments of the previous observation, then they are mapped to each other with the same distortion bound.

- If we end and start each grouping with a segment without short intervals, and all previous conditions are fulfilled then $A$ is mapped to $B$ with distortion less than $n^{2/3+2\varepsilon}$.
Upper Bound - Conditioning

In order to have the conditions to be able to map $A$ to $B$, it is enough to use the previous slide.

- The probability for a segment to have short intervals is bounded by $C/n^{-\varepsilon}$.
- By rotations, we can ensure that we begin and end with long intervals, and that each grouping begin and end with a segment free of short intervals.
- We can count the sequences of short intervals as a *bins and balls* problem. Note that each of them is a concentrated binomial random variable.
- In order to cover all sequences, we allow scaling up by $n^{-1/\varepsilon}$. By the concentration inequalities, it is enough (we just add the missing pieces).
- We use a similar argument for the groupings.
Further Questions

- How does the metric distortion change with dimension? I find three cases that interest me the most:
  - The $d$-dimensional torus $\mathbb{R}^d/\mathbb{Z}^d$.
  - The $d$-dimensional sphere $S^d$.
  - Fractals, such as Cantor sets.
- How the result would change if we are allowed to throw away $n^\beta$ points?
Thank You