On the Whitney extension problem for near distortions, near alignment, near manifold learning, shortest paths.

Steven Damelin, University of Michigan

Ariel University, 2023,

.

- **Main reference**: S. B. Damelin, "On the Whitney near extension problem, BMO, alignment of data, best approximation in algebraic geometry, manifold learning and their beautiful connections: A modern treatment", arXiv:2103.09748. To appear: John Wiley & Sons.
- Collaborators:
- Charles Fefferman, Princeton: (Part 1)
- Alfred Hero, University of Michigan: (Part 2)
- Sung Jin Wang, Google: (Part 2)
- Daniel McKenzie, Colorodo School of Mines: (Part 1)

- We will work in $\mathbb{R}^d$, $d \geq 2$. Unless stated otherwise, $|\cdot|$ will denote the Euclidean norm in $\mathbb{R}^d$.
- Moving forward all constants depend on d and the same symbol may denote a different constant at a different time.

- Part 1: The near Whitney extension problem.
- A map $A : \mathbb{R}^d \to \mathbb{R}^d$ is an improper Euclidean motion (transformation) if there exist $M \in O(d)$ and $x_0 \in \mathbb{R}^d$ so that for every $x \in \mathbb{R}^d$, $A(x) = Mx + x_0$.
- If $M \in SO(d)$, then A is a proper (orientation preserving) Euclidean motion. Here, $O(d)$ and $SO(d)$ are respectively the orthogonal and special orthogonal groups. A Euclidean motion can either be proper or improper.

- The Whitney extension problem:
- Let $\phi : E \to \mathbb{R}$ be a map defined on an arbitrary set $E \subset \mathbb{R}^D$. How can one decide whether ϕ extends to a map $\Phi : \mathbb{R}^D \to \mathbb{R}$ which agrees with ϕ on E and is in $C^m(\mathbb{R}^D)$, $m > 1$, the space of functions from $\mathbb{R}^D$ to $\mathbb{R}$ whose derivatives of order m are continuous and bounded.

- The near distorted Whitney extension problem:
- Let $E \subset \mathbb{R}^D$ be compact with some geometry.
- Suppose that we are given a map $\phi : E \to \mathbb{R}^D$ with ϕ a small distortion on E. What this means is that there exists $\delta > 0$ small enough (depending on d) so that $(1-\delta)|x-y| \leq |\phi(x)-\phi(y)| \leq |x-y|(1+\delta)$ for every $x, y \in E$.
- Note that ϕ is distance preserving (rigid), if $|\phi(x)-\phi(y)| = |x-y|$ for every $x, y \in E$. Otherwise it is non rigid.

- How can one decide whether ϕ extends to a smooth small distortion $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ which agrees with ϕ on E.
- Note that the distortion constant for Φ say ε will in general be different to the distortion constant of ϕ namely δ.
- The maps ϕ and Φ are examples of bi-Lipchitz maps.
- A bi-lipchitz map will not extend to another bi-lipchitz map unless the distortion constants are small.

- We also ask if how to decide if and where in addition the smooth small distortion Φ can be approximated well by by proper and improper Euclidean motions in small balls in $\mathbb{R}^d$.

- If we take E to be a finite set, we obtain the following problem. How to understand the following.
- Given $\varepsilon > 0$ small enough. There exists $\delta > 0$ depending on ε such that the following holds: Let $y_1, ...y_k$ and $z_1, ...z_k$ be two $k \geq 1$ sets of distinct points in $\mathbb{R}^d$. Suppose that

$$(1+\delta)^{-1} \leq \frac{|z_i - z_j|}{|y_i - y_j|} \leq (1+\delta),\ 1 \leq i,j \leq k,\ i \neq j.$$

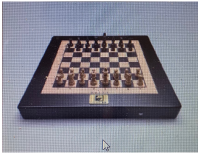
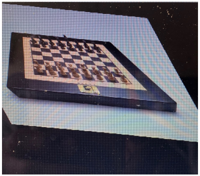
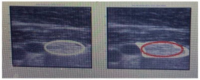
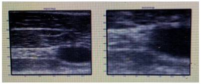
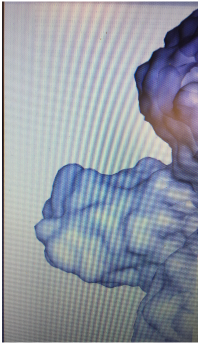
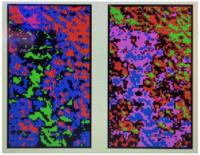
Then there exists a diffeomorphism (differentiable, invertible map (hence 1-1 and onto)) $\Phi : \mathbb{R}^d \to \mathbb{R}^D$ with

$$(1+\varepsilon)^{-1}|x-y| \leq |\Phi(x) - \Phi(y)| \leq |x-y|(1+\varepsilon),\ x, y \in \mathbb{R}^d$$

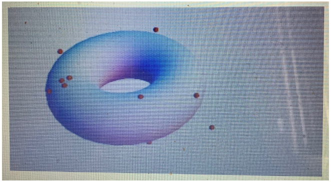
satisfying

$$\Phi(y_i) = z_i,\ 1 \leq i \leq k.$$

- Is it possible that Φ at the same time agrees with a Euclidean motion "globally away from" the points $\{y_1, ..., y_k\}$ and sometimes also with Euclidean motions "locally close" to each point in $\{y_1, ..., y_k\}$.
- Can one say how ε, δ are related.
- Near alignment problem.


- Near manifold learning.



- **Theorem: [DF]**
- Let $\varepsilon > 0$ small enough depending on d. Let $\{y_1, ...y_k\}$ and $\{z_1, ...z_k\}$ be two $1 \leq k \leq d$ sets of distinct points in $\mathbb{R}^d$. There exists $\delta > 0$ small enough depending on ε with the following property: Suppose that

$$(1+\delta)^{-1} \leq \frac{|z_i - z_j|}{|y_i - y_j|} \leq (1+\delta),\ 1 \leq i,j \leq k,\ i \neq j.$$

Then there exists a ε-distorted diffeomorphism $f : \mathbb{R}^d \to \mathbb{R}^d$ satisfying

$$f(y_i) = z_i,\ 1 \leq i \leq k.$$

- Building f:
- Slow twists, Smooth distorted rotations which rotate slowly as a function of the distance from the center of mass. Non Rigidity.
- Slides, Smooth distorted translations which translate slowly as a function of the distance from the origin. Non Rigidity
- Partition of unity.
- Clusters (shortest paths, similarity kernels, herarchical clustering in $\mathbb{R}^d$.
- Gluing.

- **Slow twists and slides.**
- Slow Twist: Let $x \in \mathbb{R}^d$ and $\varepsilon >$ small enough depending on d. Let $S(x)$ be the $d \times d$ block-diagonal matrix

$$\begin{pmatrix} D_1(x) & 0 & 0 & 0 & 0 & 0 \\ 0 & D_2(x) & 0 & 0 & 0 & 0 \\ 0 & 0 & . & 0 & 0 & 0 \\ 0 & 0 & 0 & . & 0 & 0 \\ 0 & 0 & 0 & 0 & . & 0 \\ 0 & 0 & 0 & 0 & 0 & D_r(x) \end{pmatrix}$$

where for each i either $D_i(x)$ is the 1×1 identity matrix or else

$$D_i(x) = \begin{pmatrix} \cos f_i(|x|) & \sin f_i(|x|) \\ -\sin f_i(|x|) & \cos f_i(|x|) \end{pmatrix}$$

where $f_i : \mathbb{R} \to \mathbb{R}$ are functions satisfying the condition: $t|f_i'(t)| < c\varepsilon$ some $c > 0$ small enough, uniformly for $t \geq 0$.

- Let $\Phi(x) = \Theta^T S(\Theta x)$ where Θ is any fixed matrix in $SO(d)$.

- Slide: Let $\varepsilon > 0$ be small enough depending on d. Let $g : \mathbb{R}^d \to \mathbb{R}^d$ be smooth and have the following property. There exists $c > 0$ small enough depending on d such that $|g'(x)| < c\varepsilon$ uniformly for $x \in \mathbb{R}^d$. Consider the map $\Phi(x) = x + g(x)$, $x \in \mathbb{R}^d$.

Here we illustrate the concept of a slow twist on $\mathbb{R}^2$. Given a fixed $\varepsilon > 0$, and a function $f : \mathbb{R} \to \mathbb{R}$ satisfying $t|f'(t)| < c\varepsilon$ for every $t \geq 0$ and for suitably small $c > 0$, define the **slow-twist matrix** $S(x)$ for any $x \in \mathbb{R}^2$ via

$$S(x) := \begin{bmatrix} \cos f(|x|) & \sin f(|x|) \\ -\sin f(|x|) & \cos f(|x|) \end{bmatrix}.$$

Then given any pure rotation $\Theta \in \mathrm{SO}(2)$, the following map $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ is a **slow twist**:

$$\Phi(x) := \Theta^T S(\Theta x)x.$$

For illustration purposes at first, we fix Θ to be the identity matrix. For a first set of illustrations, we will look only at one application of a slow twist with f being an exponential function with differing scaling parameter.

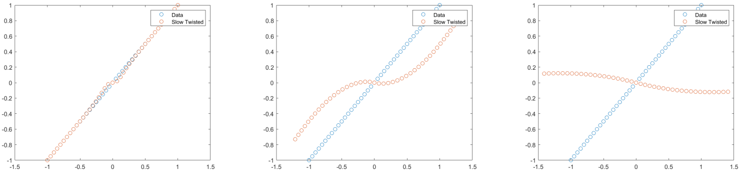


Figure: Initial data lying on the line $y = x$, and the application of a slow twist with $f(x) = \exp(-\mu|x|)$, with $\mu = 10$ (top left), $\mu = 1$ (top right), and $\mu = 0.1$ (bottom).

For large values of μ, it can be seen that the twist is quite rigid, and even outside a small cube centered at the origin, the data is left fixed. On the other hand, as μ tends to 0, the twist becomes closer to a pure rotation near the origin. Nevertheless, at a far enough distance, the slow-twist Φ will leave the data unchanged. Indeed, the next figures illustrate this:

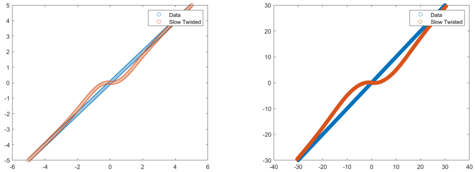


Figure: Large scale for slow twists with $f(x) = \exp(-\mu|x|)$. Left: $\mu = 1$, the twist leaves the data static outside $[-5, 5]^2$; Right: $\mu = 0.1$, the twist only starts to leave the data static outside about $[-30, 30]^2$.

Let us pause to consider what happens when the decay condition on the twist function f is not satisfied; in this case we will dub the function Φ a **fast twist** for reasons that will become apparent presently.

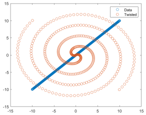


Figure: Fast twist with function $f(x) = |x|$.

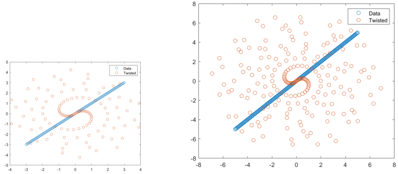


Figure: Fast twist with $f(x) = |x|^2$ for a small interval $[-3, 3]$ (left) and large interval $[-10, 10]$.

From Figure 3, one can see that when f is the identity function, the rate of twisting is proportional to the distance away from the origin, and hence there is no way that Φ will leave points fixed outside of any ball centered at the origin. Likewise, one sees from Figure 4 that the fast twist with function $f(x) = |x|^2$ rapidly degenerates points into a jumbled mess.

Here we illustrate what happens when one iteratively applies a slow twist to a fixed initial data.

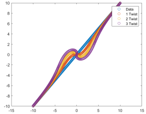
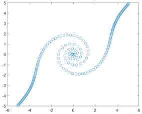


Figure: Iterated slow twist with $f(x) = \exp(-0.5|x|)$. Shown is the initial data along the line $y = x$, $\Phi(x)$, $\Phi \circ \Phi(x)$, and $\Phi \circ \Phi \circ \Phi(x)$.

In Figure 5, we see an illustration of the fact that the composition of slow twists remains a slow twist, but the distortion changes slightly; indeed notice that as we take more iterations of the exponential slow twist, we have to go farther away from the origin before the new twist leave the data unchanged.



Recall that given an $\varepsilon > 0$, a slide is defined to be

$$\Phi(x) = t + g(t),$$

where g satisfies $|g'(t)| < c\varepsilon$ for sufficiently small $c > 0$. Here we illustrate some simple examples of slides on $\mathbb{R}^2$.
First consider equally spaced points on the line $y = -x$, and the slide given by the function

$$g(t) := \begin{bmatrix} \frac{1}{1+|t_1|^2} \\ \\ \frac{1}{2}e^{-|t_2|} \end{bmatrix}.$$

This is illustrated in Figure 6 shown below.

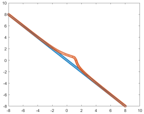


Figure: Slide with the function g given above.

To give some more sophisticated examples, we consider first the slide function

$$g(t) := \begin{bmatrix} e^{-|t_1|} \\ \\ e^{-0.1|t_2|} \end{bmatrix},$$

acting iteratively on uniform points along both the lines $y = x$ and $y = -x$.

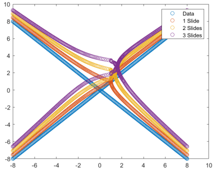


Figure: Lines $y = x$ and $y = -x$ along with $\Phi(x)$, $\Phi \circ \Phi(x)$ and $\Phi \circ \Phi \circ \Phi(x)$for g.

Similarly, the following figure shows the slide function

$$g_2(t) := \begin{bmatrix} 1 - e^{-|t_1|} \\ \\ 1 - e^{-0.1|t_2|} \end{bmatrix}$$

acting iteratively on uniform points along the lines $y = x$ and $y = -x$.

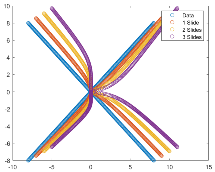


Figure: Lines $y = x$ and $y = -x$ along with $\Phi(x)$, $\Phi \circ \Phi(x)$ and $\Phi \circ \Phi \circ \Phi(x)$ for g_2.

To illustrate the effect of the distance of data from the origin, we illustrate here how slides affect uniform points on circles of different radii.

We use again the asymmetric sliding function

$$g(t) = \begin{bmatrix} \frac{1}{1+|t_1|^2} \\ \\ \frac{1}{2} e^{-|t_2|} \end{bmatrix}.$$

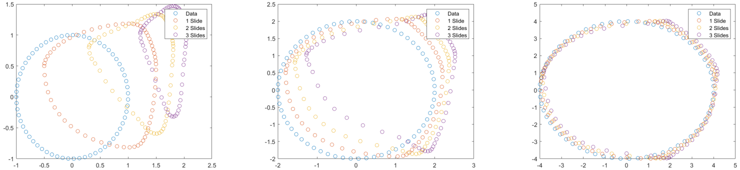


Figure: Circles under 3 iterated slides with the function g above, beginning with a circle of radius 1 (top left), 2 (top right) and 4 (bottom).

We see from Figure 9 that the farther out the data is (i.e. the larger the radius of the initial circle, the less the effect of the slide, which makes sense given the definition and the fact that the slides must be ε–distortions of $\mathbb{R}^2$.

Here we illustrate some of the motions above in $\mathbb{R}^3$.
The Matlab code here allows the user to generate a generic rotation matrix in $SO(3)$ by specifying parameters a, b, c, d satisfying $a^2 + b^2 + c^2 + d^2 = 1$, and the rotation matrix Θ is defined by

$$\Theta = \begin{bmatrix} a^2 + b^2 - c^2 - d^2 & 2(bc - ad) & 2(bd + ac) \\ 2(bc + ad) & a^2 - b^2 + c^2 - d^2 & 2(cd - ab) \\ 2(bd - ac) & 2(cd + ab) & a^2 - b^2 - c^2 + d^2 \end{bmatrix}.$$

As a reminder, our slow twist on $\mathbb{R}^3$ is thus $\Theta^T S(x) \Theta x$.

Example

Our first example is generated by the rotation matrix Θ as above with parameters $a = b = \frac{1}{\sqrt{3}}$ and $c = d = \frac{1}{\sqrt{6}}$, and the slow twist matrix S as

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(f(|x|)) & \sin(f(|x|)) \\ 0 & -\sin(f(|x|)) & \cos(f(|x|)) \end{bmatrix},$$

where $f(t) = e^{-\frac{t}{2}}$. Figures 10 and 11 show two views of the twisted motions generated by these parameters

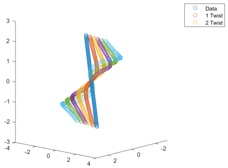


Figure: Slow Twist in $\mathbb{R}^3$ from Example 1

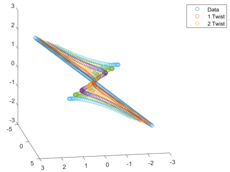
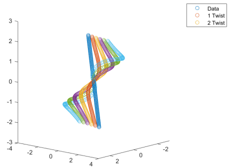


Figure: Alternate View of Slow Twist in $\mathbb{R}^3$ from Example 1



Here we generate 1000 random points on the unit sphere in $\mathbb{R}^3$ and allow them to move under a slide formed by

$$g(x) = x + \begin{bmatrix} e^{-0.5|x_1|} \\ e^{-|x_2|} \\ e^{-\frac{3}{2}|x_3|} \end{bmatrix}.$$

The results are shown in Figure 12

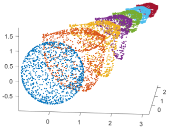


Figure: Anisotropic slide on the 2–sphere.

- **Theorem: [DF]**
- There exists $\eta > 0$ depending on ε for which the following holds. Let $\Theta \in SO(d)$, $r_1, r_2 > 0$ and let $0 < r_1 \leq \eta r_2$. Then, there exists an ε-distorted diffeomorphism $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ such that
$$\begin{cases} \Phi(x) = \Theta x, & |x| \leq r_1 \\ \Phi(x) = x, & |x| \geq r_2 \end{cases}$$
- There exists $\eta_1 > 0$ depending on ε such that the following holds. Let A be a proper Euclidean motion. Let $r_3, r_4 > 0$. Suppose $0 < r_3 \leq \eta_1 r_4$ and $|x_0| \leq c\varepsilon r_3$. Then there exists an ε-distorted diffeomorphism $\Phi_1 : \mathbb{R}^d \to \mathbb{R}^d$ such that
$$\begin{cases} \Phi_1(x) = A(x), & |x| \leq r_3 \\ \Phi_1(x) = x, & |x| \geq r_4 \end{cases}$$

- **Theorem: [DF]**
- $E \subset \mathbb{R}^d$ be a finite set with E of small diameter. There exists ε-distorted diffeomorphism $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ and C, C' depending on ε with C small enough and C' large enough so that:
- Φ coincides with an improper Euclidean motion on $\left\{x \in \mathbb{R}^d : \mathrm{dist}(x, E) \geq C\right\}$.
- Φ coincides with an improper Euclidean motion A_z on $B(z, C')$ for each $z \in E$.
- $\Phi(z) = z$ for each $z \in E$.

- The Orthogonal Procrustes Matrix-Tensor-Optimization Problem.
- For a $d \times d$ matrix, $M = (M_{ij})$, we will mean by $|M|$ the Hilbert-Schmidt norm (sometimes called the Frobenius norm or the Schatten-2-norm) on M which is: $|M| := \left(\sum_{ij} |M_{ij}|^2\right)^{1/2}$.
- Let now $\{y_1, \ldots, y_k\}$ and $\{z_1, ..., z_k\}$ be two sets of distinct sets of points in $\mathbb{R}^d$ with certain geometry.
- Concerning the form of $M \in O(d)$ for the Euclidean motion A in the Isometry extension alignment problem, the following optimization matrix problem below produces M when it is known that a solution $M \in O(d)$ exists.

$$\inf_{M \in O(d)} \sum_{i=1}^{k} |(z_i - M(y_i))|.$$

- How to formulate a matrix Procrustes problem for Slow twists and Slides and indeed near isometries where one should be able to "track" alignment?

- Removing the restriction of k. Removal of Degenerate Cases.
- $\mathrm{card}(\mathsf{E})$ cannot be too large.
- The diameter of the set E is not too large.
- The points of the set E cannot be too close to each other.
- The points of the set E are close to a hyperplane in $\mathbb{R}^D$.
- Roughly: the number of points still has to be finite but no longer bounded by d. Instead, roughly speaking, what is required isthat on any $d+1$ of the k points which form vertices of a relatively voluminous simplex, the mapping ϕ is orientation preserving. $\delta = \exp\left(-\frac{C}{\varepsilon}\right)$, $C = C(d)$. This makes a difference.

- Roughly: the number of points still has to be finite but no longer bounded by d. Instead, roughly speaking, what is required is that on any $d+1$ of the k points which form vertices of a relatively voluminous simplex, the mapping $\phi : E \to \phi(E)$ is orientation preserving. Here, $E := \{y_1, ..., y_k\}$ and $\phi(E) := \{z_1, ..., z_k\}$, $\delta = c \exp\left(-\frac{c_1}{\varepsilon}\right)$ and E cannot be too close to a hyperplane.
- Many theorems with different geometries on the points.

- Patches analogues and optimial: $\delta = c\varepsilon$.
- **Theorem [DF]**.
- Let $U \subset \mathbb{R}^d$ be an open set. Let $E' \subset U \subset \mathbb{R}^d$ be not too thin. [See book for details] . Let ε be small enough.
- Let $\phi : U \to \mathbb{R}^d$ be a $C^1(U)$ map satisfying for all $x, y \in E'$,

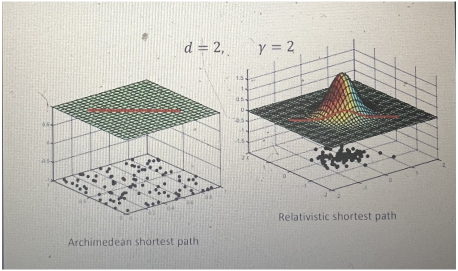
$$(1-\epsilon) \leq |x-y| \leq |\phi(x) - \phi(y)| \leq (1+\epsilon) \leq |x-y|.$$

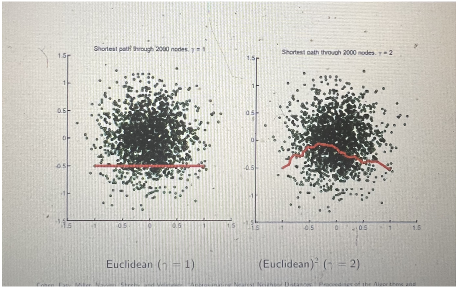
- There exists a C^1 map $f : \mathbb{R}^d \to \mathbb{R}^d$ and a Euclidean motion A with the following properties:

- $(1 - c'\varepsilon)|x - y| \leq |f(x) - f(y)| \leq (1 + c'\varepsilon)|x - y|$ for all $x, y \in \mathbb{R}^d$.
- $f = \phi$ in a neighborhood of E'.
- $f = A$ not to "far" out.
- f is one-to-one and onto.
- If $\phi \in C^m(U)$ for some given $m > 1$, then $f \in C^m(\mathbb{R}^d)$.

- The results in this first part given are part of a much larger collection of results which can be found in our main reference: S. B. Damelin, "On the Whitney near extension problem, BMO, alignment of data, best approximation in algebraic geometry, manifold learning and their beautiful connections: A modern treatment", arXiv:2103.09748. To appear: John Wiley & Sons.
- Unlabeled problems, Optimal transport, BMO (maps of Bounded Mean Oscillation), Algorthims, Other "Vision transformations" (for example, Projections, Affine maps, Near reflections.)
- Johnson-Lindenstrauss, noise, near manifold learning, uniquness, clustering, artificial intelligence.

- Part 2: Shortest limits and continum limits on complete graphs, See [SDH]: (Annals of Applied Probability).
- Let $X := \{X_1, X_2, ..., X_n \in \mathbb{R}^d\}$ be i.i.d random vectors in $\mathbb{R}^d$ with marginal pdf f having compact support S with metric tensor g. Fix two points x_i and x_f in $\mathbb{R}^d$. Let $p > 0$ and define G as the complete graph spanning X with weights $\{|X_j - X_i|^p, 1 \leq i, j \leq n, i \neq j\}$ (power distance weight). Let $L_p(x_i, x_f)$ be the shortest path between x_i and x_f.


$d = 2,$ $\gamma = 2$
Archimedean shortest path
Relativistic shortest path


Shortest path through 2000 nodes, γ = 1
Shortest path through 2000 nodes, γ = 2
Euclidean (γ = 1)
(Euclidean)² (γ = 2)

- When $p = 1$, this is the Euclidean distance and the shortest path is a straight line. When $p < 1$, the shortest path is still a straight line. Looking at the transfer from uniform distribution to non uniform distribution, when $p > 1$ the shortest path is no longer a straight line and tends to move to the centre of the distribution favoring denser points.
- This is analogous to Newtonian verus special relativity where a photon bends at a region of high mass.
- Thus there is a "lensing effect". How can we explain this as $n \to \infty$?

- A continuum limit of shortest path.
- Let (M, g_1) be a smooth compact d-dimensional Riemannian manifold without boundary. Assume $d > 1$. Consider a probability distribution Pr over Borel subsets of M. Assume that the distribution has a smooth Lebesgue probability density function f with respect to g_1. Let $X_1, X_2, \ldots$ denote an i.i.d sequence drawn from this density.
- For $p > 1$, called the power parameter, define a new Riemannian metric $g_p = f^{2(1-p)/d} g_1$
- That is if Z_x and W_x are two tangent vectors at a point $x \in M$, then $g_p(Z_x, W_x) = f(x)^{2(1-p)/d} g_1(Z_x, W_x)$

- The main result here establshes an asymptotic limit of the lengths of the shortest path through finite subsets of points $\mathcal{X}_\backslash := \{X_1....X_n\}$ as $n \to \infty$.
- If $x, y \in M$, then let $L_n(x, y)$ denote the shortest path length from x to y through $\mathcal{X}_\backslash \cup \{x, y\}$.
- Here the edge weight between two points u and v is $\mathrm{dist}_1(u, v)^p$. where dist_1 denotes the Riemannian distance under g_1.
- The power weighted graph is defined as the complete graph over $\mathcal{X}_\backslash \cup \{x, y\}$.

- **Theorem: [SDH]**
- Suppose that $\inf_M f > 0$. Then

$$\lim_{n\to\infty} n^{(1-p)/d} L_n(x,y) = C(d,p)\mathrm{dist}_p(x,y), c.c$$

  with $C(d,\gamma)$ a positive constant independent of f.
- More general statements of this result are given in [SDH] with rate of convergence.
- Note: The scaling of the Riemannian metric by $f^{2(1-p))/d}$ is inversely proportional to the probability density function f. The theorem says that the density f shortens and lenghens paths which respectively pass through regions of high and low density.

- Metric spaces, brain and biological networks, Lensing for examples metric spaces of functions, distances, information theory, entropy, divergence dissimilarity, manifold learning, clustering, quantum clustering.
- Minimial paths and lensing on sparse graphs and hypergraphs. Currently we have: KNN, MST and Complete graphs.
- Complexity.
- **Main reference**: S. B. Damelin, "On the Whitney near extension problem, BMO, alignment of data, best approximation in algebraic geometry, manifold learning and their beautiful connections: A modern treatment", arXiv:2103.09748. To appear John Wiley & Sons.

Thank you very much!