Constrained Wasserstein Fitting Problems

Or: "How Best to Fill a Region with a Curve?"

Forest Kobayashi

The University of British Columbia

Wednesday, November 15$^{\text{th}}$, 2023

The Problem

Given:

1. A region $\Omega \subseteq \mathbb{R}^2$,
2. A finite length ℓ of rope,

Find:

- The rope shape that best "fills" Ω.

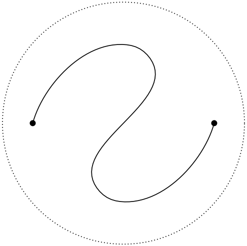
The Problem

Given:

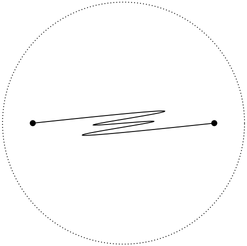
1. A region $\Omega \subseteq \mathbb{R}^2$,
2. A finite length ℓ of rope,

Find:

- The rope shape that best "fills" Ω.



(a) A "good" filling

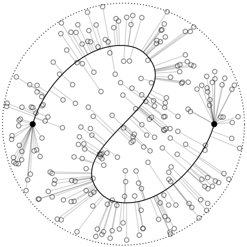


(b) A "bad" filling

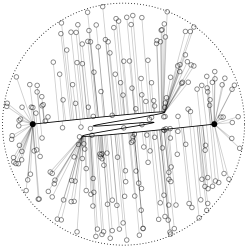
The Problem

Given:
1. A region $\Omega \subseteq \mathbb{R}^2$,
2. A finite length ℓ of rope,

Find:
- The rope shape that best "fills" Ω.



(a) A "good" filling



(b) A "bad" filling

Takeaway

Objective: Find a curve $f : [0,1] \to \Omega$ minimizing

$$\mathcal{J}(f) = \int_\Omega d(\omega, f([0,1]))\ d\omega$$

subject to length$(f) \leq \ell$.

Takeaway

Objective: Find a curve $f : [0,1] \to \Omega$ minimizing

$$\mathcal{J}(f) = \int_{\Omega} d(\omega, f([0,1]))\ d\omega$$

subject to length$(f) \leq \ell$.

Some questions:

- Can we use d^p instead of d?

Takeaway

Objective: Find a curve $f : [0,1] \to \Omega$ minimizing

$$\mathcal{J}(f) = \int_\Omega d(\omega, f([0,1]))\ d\omega$$

subject to $\text{length}(f) \leq \ell$.

Some questions:

- Can we use d^p instead of d? ✓ Sure.
- We're really "filling" the uniform measure $d\omega$. Can we try other probability measures?

Takeaway

Objective: Find a curve $f : [0,1] \to \Omega$ minimizing

$$\mathcal{J}(f) = \int_\Omega d(\omega, f([0,1]))\ d\omega$$

subject to $\text{length}(f) \leq \ell$.

Some questions:

- Can we use d^p instead of d? ✓ Sure.
- We're really "filling" the uniform measure $d\omega$. Can we try other probability measures? ✓ Sure, with finite p-th moment.
- Can we work with $f : X \to \Omega$ where $X \subseteq \mathbb{R}^m$, $\Omega \subseteq \mathbb{R}^n$, $m < n$?

Takeaway

Objective: Find a curve $f : [0,1] \to \Omega$ minimizing

$$\mathcal{J}(f) = \int_{\Omega} d(\omega, f([0,1]))\ d\omega$$

subject to $\text{length}(f) \leq \ell$.

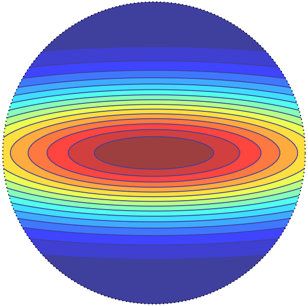
Some questions:

- Can we use d^p instead of d? ✓ Sure.
- We're really "filling" the uniform measure $d\omega$. Can we try other probability measures? ✓ Sure, with finite p-th moment.
- Can we work with $f : X \to \Omega$ where $X \subseteq \mathbb{R}^m$, $\Omega \subseteq \mathbb{R}^n$, $m < n$? ✓ Sure, but some care required.

Other Measures

Example: Choose $\rho \in \mathcal{P}(\Omega)$ more concentrated on the horizontal:

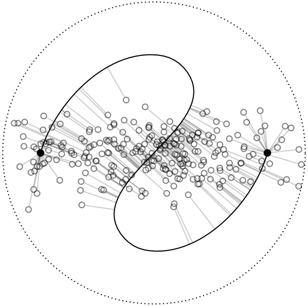
$$\rho \sim \mathcal{N}(0;\ [\sigma_X^2\ 0;\ 0\ \sigma_Y^2]) \qquad\qquad \sigma_Y \ll \sigma_X$$

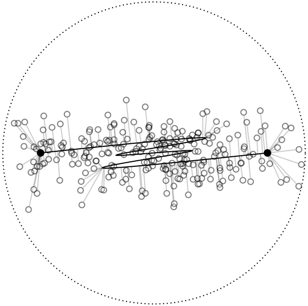


Other Measures

Example: Choose $\rho \in \mathcal{P}(\Omega)$ more concentrated on the horizontal:

$$\rho \sim \mathcal{N}(0;\ [\sigma_X^2\ 0;\ 0\ \sigma_Y^2]) \qquad\qquad \sigma_Y \ll \sigma_X$$





Other Dimensions

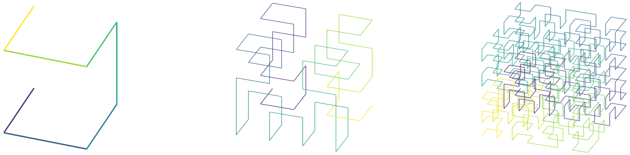
Example: If $f : [0,1]^2 \to [0,1]^3$, what should we pick to be the constraint?

- Maybe surface area?

Other Dimensions

Example: If $f : [0,1]^2 \to [0,1]^3$, what should we pick to be the constraint?

- Maybe surface area? ✗ Unfortunately, no.
- Let $g : [0,1] \to [0,1]^3$ and take $f(x_1, x_2) = g(x_1)$
- g can still "fill" space while $\text{Area}(f) = 0$:



Choosing a Constraint $\mathcal{C}(f)$

Want $\mathcal{C}(f)$ to measure "complexity" of f:

- If f nonconstant, $\mathcal{C}(f) > 0$
  - prev slide: m-dimensional Hausdorff measure fails

Choosing a Constraint $\mathcal{C}(f)$

Want $\mathcal{C}(f)$ to measure "complexity" of f:

- If f nonconstant, $\mathcal{C}(f) > 0$
  - prev slide: m-dimensional Hausdorff measure fails
- (Optional): If f highly oscillatory, $\mathcal{C}(f) \gg 0$
  - penalizes effectively-1d solutions
  - relevance in applications

Choosing a Constraint $\mathcal{C}(f)$

Want $\mathcal{C}(f)$ to measure "complexity" of f:

- If f nonconstant, $\mathcal{C}(f) > 0$
  - prev slide: m-dimensional Hausdorff measure fails
- (Optional): If f highly oscillatory, $\mathcal{C}(f) \gg 0$
  - penalizes effectively-1d solutions
  - relevance in applications

Choice:

$$\mathcal{C}(f) = \|f\|_{W^{k,q}(X;\Omega)}$$

(assume $kq > m$ for Sobolev inequality)

Revised Problem Statement

Given:

- $X \subseteq \mathbb{R}^m$, $\Omega \subseteq \mathbb{R}^n$
- $\rho \in \mathcal{P}(\Omega)$ (measure to approximate)
- $\ell \geq 0$ (budget)
- $p \geq 1$ (weight of far-away points)
- Some technical hypotheses

Minimize:

$$\mathcal{J}_p(f) = \int_\Omega d^p(\omega, f)\ d\rho(\omega)$$

Subject to:

$$\|f\|_{W^{k,q}(X;\Omega)} \leq \ell$$

Related Work

- Quantization problem for measures (e.g. [Iac17])
  - Optimal approximation of ρ by Dirac masses
  - $m = 0$, $\mathcal{C} = \#$ of points
- Principal Curves ([KS17], [LS21], [DF20])
  - Like our problem for $m = 1$, $\mathcal{C}(f) = \text{length}(f)$
  - Common to use soft penalty:

$$\mathcal{J}_p^\lambda(f) = \mathcal{J}_p(f) + \lambda \mathcal{C}(f)$$

- Unequal-dimensional OT ([MP19])

Connection to OT

- Wasserstein Cost: Given $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ $(X, Y \subseteq \mathbb{R}^n)$

$$\mathbb{W}_p(\mu,\nu) = \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{X \times Y} |x-y|^p \, d\gamma(x,y) \right)^{1/p}$$

- $\mathbb{W}_p(\mu,\nu)$ gives metric that respects "structure" of $\mathbb{R}^n$

Connection to OT

- Wasserstein Cost: Given $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ $(X, Y \subseteq \mathbb{R}^n)$

$$\mathbb{W}_p(\mu, \nu) = \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{X \times Y} |x - y|^p \; d\gamma(x, y) \right)^{1/p}$$

- $\mathbb{W}_p(\mu, \nu)$ gives metric that respects "structure" of $\mathbb{R}^n$
- $\mathcal{J}_p \approx$ measures closeness of f and ρ:

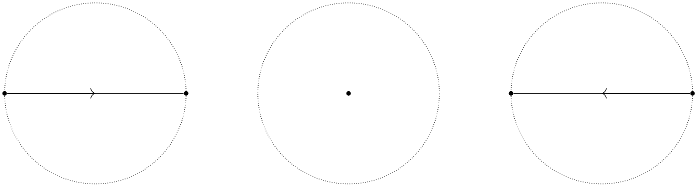
Proposition: For all f,

$$\mathcal{J}_p(f) = \inf_{\nu \in \mathcal{P}(f(X))} \mathbb{W}_p^p(\rho, \nu).$$

- Analogous to how $d(p, A) = \inf_{a \in A} d(p, a)$.

Flavor of Problem: $\mathscr{J}_p$

1. $\mathscr{J}_p$ nonconvex



(a) f_0 (b) $f_{.5}$ (c) f_1

2. $\mathscr{J}_p$ nonconcave as well
3. But, jointly weakly continuous in f and ρ

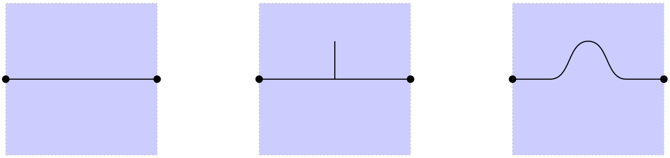
Flavor of Problem: $\mathcal{C}(f)$

Given ε extra budget, can we improve J?

Flavor of Problem: $\mathcal{C}(f)$

Given ε extra budget, can we improve J?

- 2^{nd}-order term makes local modifications very hard.

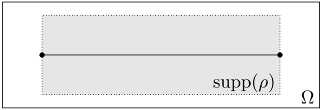


(a) f (b) Desired perturbation (c) Best we can do

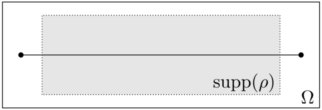
Flavor of Problem: $\mathcal{C}(f)$

Given ε extra budget, can we improve J?

- 2^{nd}-order term makes local modifications very hard.
- Can't just "extend" f past $\partial f(X)$...



(a) $f(X)$



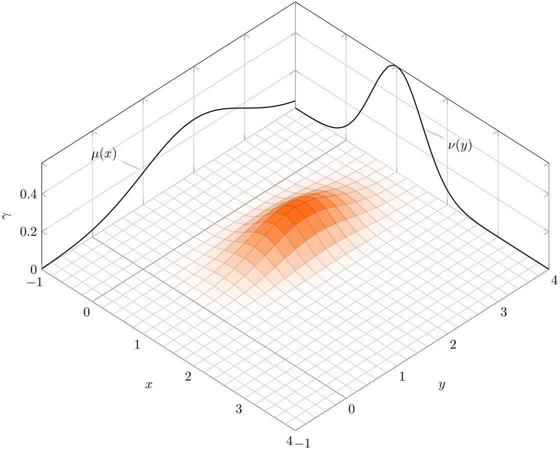
(b) Extended $f(X)$

Summary of Main Results

1. Optimizers?
   - Exist under mild hypotheses
   - Generally nonunique (e.g. if X, Ω have symmetries)
2. Relationship between J and ℓ?
   - J continuous in ℓ
   - J (trivially) nonincreasing...more than this, hard to say.
   - Coarse asymptotic estimates from covering numbers
   - Important tool: Can find directional derivative of $\mathcal{J}_p$ in $C(f(X); \mathbb{R}^n)$.
3. Discretization?
   - Yes; nice consistency results

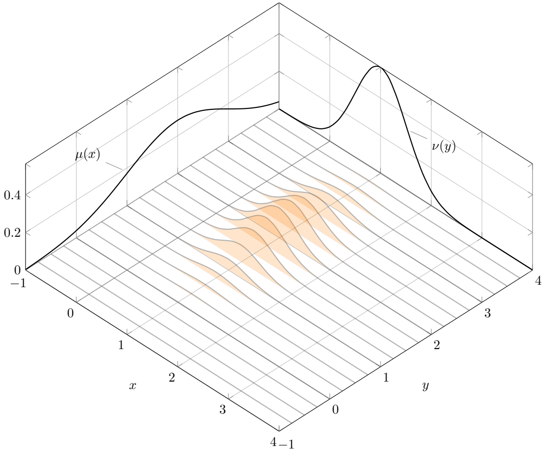
Definining a "Gradient:" Disintegration of Measures

- Given $\gamma \in \mathcal{P}(X \times Y)$ with marginals μ, ν:



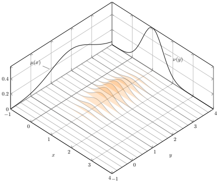
Definining a "Gradient:" Disintegration of Measures

- "Disintegrate" γ via projection onto x-axis:



Definining a “Gradient:” Disintegration of Measures

- “Disintegrate” γ via projection onto x-axis:



- Write μ as $(\pi_X)_{\#\gamma}$ (“projection” of γ onto x axis)
- For each x, let γ_x be the “slice” of γ living on $\pi_X^{-1}(x)$. Then

$$\int_{X\times Y} f(x,y)\ d\gamma(x,y) = \int_X \left(\int_{\pi_X^{-1}(x)} f(x,y)\ d\gamma_x(y) \right) d(\pi_X)_{\#\gamma}(x)$$

Defining a "Gradient"

Define a Gradient: Let $\pi_f : \Omega \to f(X)$ be closest point projection. Disintegrate ρ under π_f into $(\{\rho_y\}_{y \in f(X)}, \nu_\rho)$ and let

$$F_p(y) = p \int_{\pi_f^{-1}(\{y\})} (\omega - y)|\omega - y|^{p-2} \, d\rho_y(\omega).$$

Defining a "Gradient"

Define a Gradient: Let $\pi_f : \Omega \to f(X)$ be closest point projection. Disintegrate ρ under π_f into $(\{\rho_y\}_{y\in f(X)}, \nu_\rho)$ and let

$$F_p(y) = p\int_{\pi_f^{-1}(\{y\})} (\omega - y)|\omega - y|^{p-2}\ d\rho_y(\omega).$$

- Trivially also a function of x by taking $F_p(x) = F_p(f(y))$.

Defining a "Gradient"

Define a Gradient: Let $\pi_f : \Omega \to f(X)$ be closest point projection. Disintegrate ρ under π_f into $(\{\rho_y\}_{y\in f(X)}, \nu_\rho)$ and let

$$F_p(y) = p \int_{\pi_f^{-1}(\{y\})} (\omega - y)|\omega - y|^{p-2}\ d\rho_y(\omega).$$

- Trivially also a function of x by taking $F_p(x) = F_p(f(y))$.
- $\langle F_p, \xi \rangle_{L^2}$ gives first variation in direction ξ.

Defining a "Gradient"

Define a Gradient: Let $\pi_f : \Omega \to f(X)$ be closest point projection. Disintegrate ρ under π_f into $(\{\rho_y\}_{y\in f(X)}, \nu_\rho)$ and let

$$F_p(y) = p \int_{\pi_f^{-1}(\{y\})} (\omega - y)|\omega - y|^{p-2} \, d\rho_y(\omega).$$

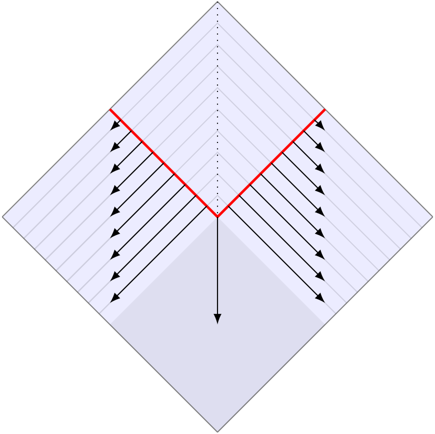
- Trivially also a function of x by taking $F_p(x) = F_p(f(y))$.
- $\langle F_p, \xi \rangle_{L^2}$ gives first variation in direction ξ.
- F_p can be discontinuous even when
  - $f \in C^\infty(X; \Omega)$
  - $f(X)$ is a C^1 manifold

Example of F_p discontinuous with $f \in C^\infty(X;\Omega)$

Let f smoothly parametrize the graph of $|x|$ with Ω a diamond.

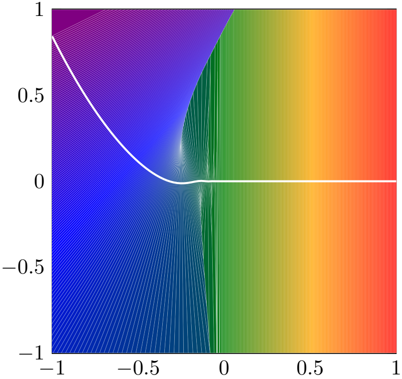
Example of F_p discontinuous with $f \in C^\infty(X;\Omega)$

Let f smoothly parametrize the graph of $|x|$ with Ω a diamond.



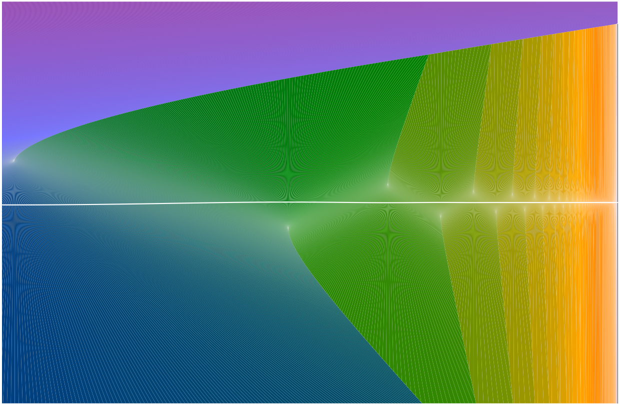
Example of F_p discontinuous with $f(X)$ a C^1-manifold

Let $f(x) = (x, x^3 \sin(1/x))$ with Ω a square.



Example of F_p discontinuous with $f(X)$ a C^1-manifold

Let $f(x) = (x, x^3 \sin(1/x))$ with Ω a square(not to scale!).



Local Modification Theorems

Theorem (A): Suppose $f \in C(X;\Omega)$ is ν_ρ-a.e. injective and $\xi : f(X) \to \mathbb{R}^n$ is cont. except perhaps on a ν_ρ-null closed set. Then

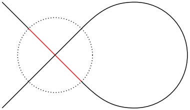
$$\lim_{\varepsilon\to 0} \frac{\mathcal{J}_p(f+\varepsilon\xi) - \mathcal{J}_p(f)}{\varepsilon} = -\int_X \langle F_p(x), \xi(x)\rangle \; d\nu_\rho(f(x))$$

Theorem (B): Suppose $f \in C(X;\Omega)$ and $\xi \in C(f(X);\Omega)$. Then

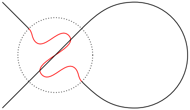
$$\lim_{\varepsilon\to 0} \frac{\mathcal{J}_p(f+\varepsilon\xi) - \mathcal{J}_p(f)}{\varepsilon} = -\int_Y \langle F_p(y), \xi(y)\rangle \; d\nu_\rho(y)$$

Local Modification Theorems

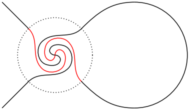
- Version A: can change topology, injectivity hypotheses on f
- Version B: can't change topology, no injectivity hypotheses on f



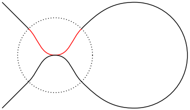
(a) Initial f



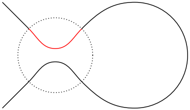
(b) An f_ε for Version A



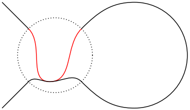
(c) An f_ε for Version B



(d) Another initial f



(e) An f_ε for Version A



(f) An f_ε for Version B

Local Modifications, Continued

- Challenge to gradient flows: Since F_p not even guaranteed to be continuous, $f + \varepsilon F_p$ might not be $W^{k,q}(X;\Omega)$
- However...

Theorem: Suppose $\nu_\rho(\{F_p \neq 0\}) > 0$. Then there exists $\xi \in C^\infty(f(X);\Omega)$ and $\varepsilon > 0$ such that $f + \varepsilon\xi \in W^{k,q}(X;\Omega)$ and

$$\lim_{\varepsilon \to 0} \frac{\mathcal{J}_p(f + \varepsilon\xi) - \mathcal{J}_p(f)}{\varepsilon} > 0.$$

Special Case: Strict Monotonicity!

Proposition: With very mild assumptions, when $m = 1$, $n = 2$, and $k = 1$, then if f is an optimizer for budget ℓ, then

$$\nu_\rho(\{F_p \neq 0\}) > 0.$$

Sketch (Contradiction):

Special Case: Strict Monotonicity!

Proposition: With very mild assumptions, when $m = 1$, $n = 2$, and $k = 1$, then if f is an optimizer for budget ℓ, then

$$\nu_\rho(\{F_p \neq 0\}) > 0.$$

Sketch (Contradiction):

- Consider perturbation $-f$.
- $\mathcal{C}((1-\varepsilon)f) = (1-\varepsilon)\mathcal{C}(f)$. So we recover $O(\varepsilon)$ budget.

Special Case: Strict Monotonicity!

Proposition: With very mild assumptions, when $m = 1$, $n = 2$, and $k = 1$, then if f is an optimizer for budget ℓ, then

$$\nu_\rho(\{F_p \neq 0\}) > 0.$$

Sketch (Contradiction):

- Consider perturbation $-f$.
- $\mathcal{C}((1-\varepsilon)f) = (1-\varepsilon)\mathcal{C}(f)$. So we recover $O(\varepsilon)$ budget.
- Perturbation is continuous on $f(X)$, so local modification theorem gives $\mathcal{J}_p(f - \varepsilon f) - \mathcal{J}_p(f) = O(\varepsilon^2)$
- Adding a "spike" in the right place changes objective by $O(\varepsilon^{3/2})$

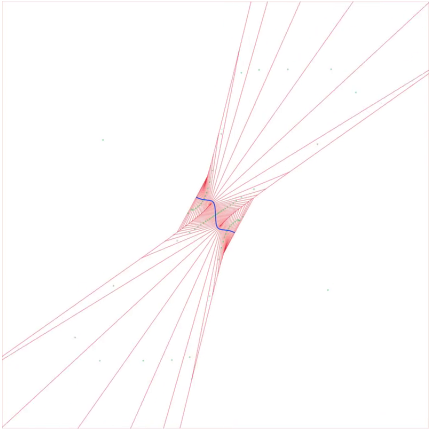
Consistency Results

- Discretizing ρ:

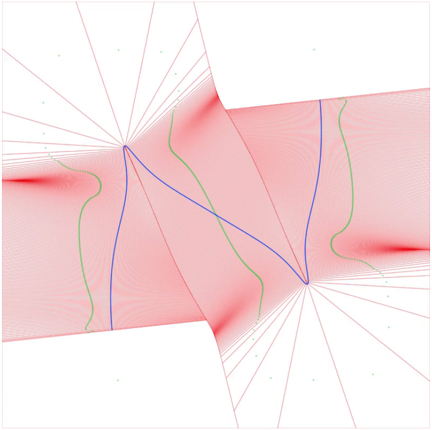
> **Proposition:** Let $\hat{\rho}_N$ be an empirical measure for ρ, and let f_N be associated ℓ-optimizers. Then almost surely, every limit point of $\{f_N\}$ is an ℓ-optimizer for ρ.

- Discretizing f:
  - "Discrete" version of F_p given by Voronoi cells
  - Somewhat intricate simulation algorithms (not included in paper)

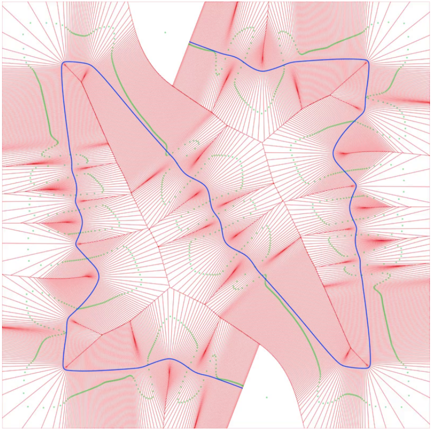
Some Simulations



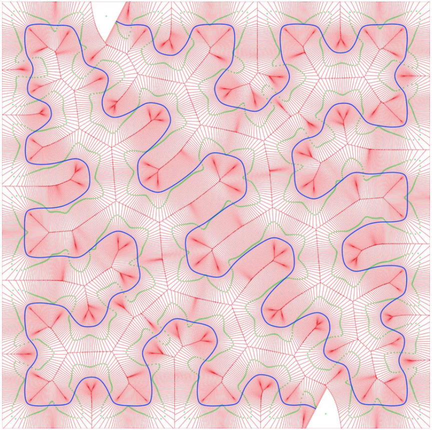
Some Simulations



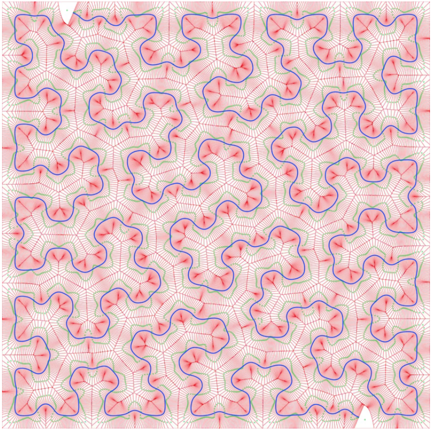
Some Simulations



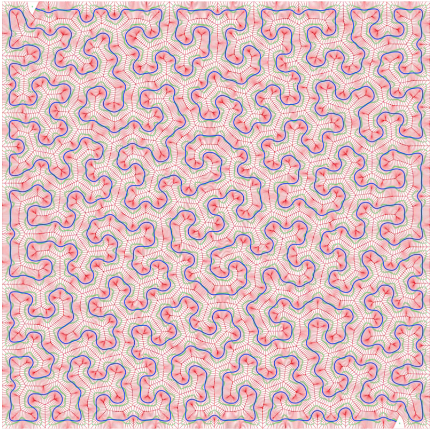
Some Simulations



Some Simulations



Some Simulations



Applications

- Routing problems
- Catalyst design
- Nonlinear Dimensional Reduction
- Generative Learning

Thank you!

References I

Sylvain Delattre and Aurélie Fischer, *On principal curves with a length constraint*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques **56** (2020), no. 3, 2108–2140, Publisher: Institut Henri Poincaré.

Mikaela Iacobelli, *A gradient flow perspective on the quantization problem*, November 2017, arXiv:1711.01652 [math].

Slav Kirov and Dejan Slepčev, *Multiple Penalized Principal Curves: Analysis and Computation*, Journal of Mathematical Imaging and Vision **59** (2017), no. 2, 234–256 (en).

References II

Xin Yang Lu and Dejan Slepčev, *Average-distance problem with curvature penalization for data parameterization: regularity of minimizers*, ESAIM: Control, Optimisation and Calculus of Variations **27** (2021), 8 (en), Publisher: EDP Sciences.

Robert J. McCann and Brendan Pass, *Optimal transportation between unequal dimensions*, May 2019, arXiv:1805.11187 [math].