Monge-Kantorovich Fitting Under a Sobolev Budget

Or: Sobolev-Constrained Principal Curves and Surfaces

Forest Kobayashi

Joint work with Young-Heon Kim (UBC) and Jonathan Hayase (UW)

Thursday, August 22^{nd} 2024

Overview

- Summary: Approximating n-dimensional measure with m-dimensional measure
- Talk is front-loaded
- Outline
  - Our *initial* motivation
  - Related work
  - A few results
  - Numerics
  - Follow-up work

Our Initial Motivation

Given:

1. A compact $\Omega \subseteq \mathbb{R}^n$
2. A rope constrained by $\mathcal{C}(\text{rope}) \leq \ell$,

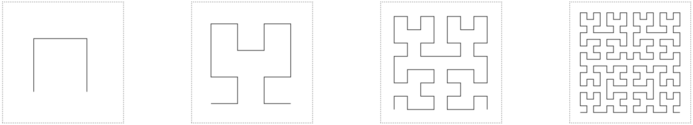
Find:

- The rope shape that best "fills" Ω.

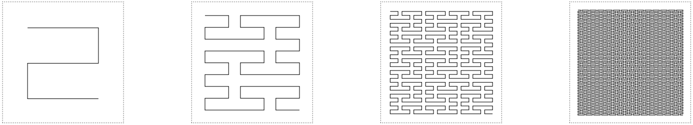
Goal: Quantify "efficiency" of space-filling curves

Our Initial Motivation

Hilbert Curve:



Peano Curve:



(not exactly apples-to-apples; length scales differently in order)

Our Initial Motivation

Given:

1. A compact $\Omega \subseteq \mathbb{R}^n$
2. A rope constrained by $\mathcal{C}(\text{rope}) \leq \ell$,

Find:

- The rope shape that best "fills" Ω.

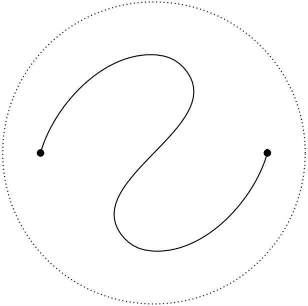
Goal: Quantify "efficiency" of space-filling curves

To discuss:

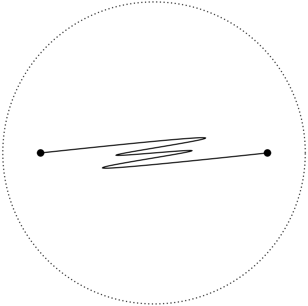
- Meaning of "filling?"
- What should $\mathcal{C}$ be?

Notion of "Filling"

Q: Which is "better?" Why?



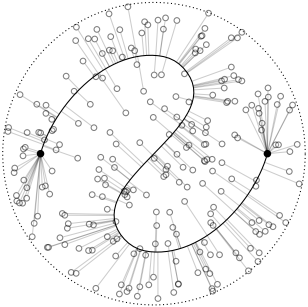
(a) One candidate



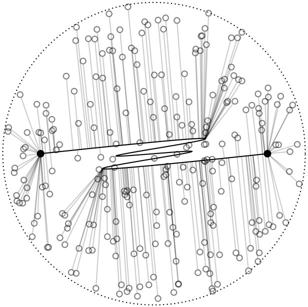
(b) ...And another

Notion of "Filling"

Q: Which is "better?" Why? *One idea...*



(a) A "good" filling



(b) A "bad" filling

Notion of "Filling," cont.

Proposal: Average distance

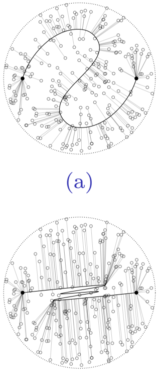
$$\mathcal{J}(f) = \frac{1}{m(\Omega)} \int_{\Omega} d(\omega, \mathrm{img}(f)) \; d\omega$$

Notion of “Filling,” cont.

Proposal: Average distance

$$\mathcal{J}(f) = \frac{1}{m(\Omega)} \int_\Omega d(\omega, \mathrm{img}(f)) \; d\omega$$
$$= \mathbb{E}_{\mathrm{Unif}(\Omega)}[d(\omega, \mathrm{img}(f))]$$

...What's the fundamental object?

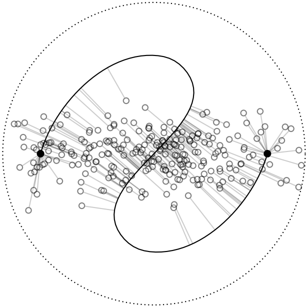


(a)

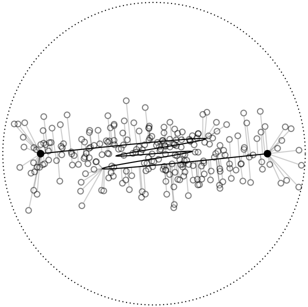
(b)

“Filling” Other Measures

Proposal: Minimize $\int_\Omega d(\omega, \mathrm{img}(f))\ d\rho(\omega) = \mathbb{E}_\rho[d(\omega, \mathrm{img}(f))]$



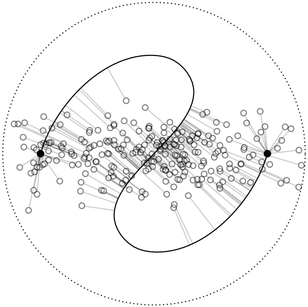
(a) A “bad” filling



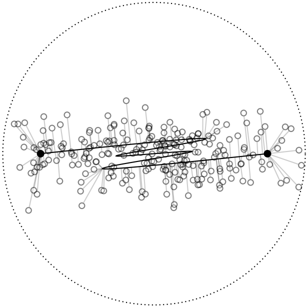
(b) A “good” filling

“Filling” Other Measures

Proposal: Minimize $\int_\Omega d^{p}(\omega, \mathrm{img}(f))\ d\rho(\omega) = \mathbb{E}_\rho[d^{p}(\omega, \mathrm{img}(f))]$



(a) A “bad” filling



(b) A “good” filling

Summary of “Filling”

For $p \geq 1$, $\rho \in \mathcal{P}(\Omega)$, let

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

...Connection to OT?

Summary of "Filling"

For $p \geq 1$, $\rho \in \mathcal{P}(\Omega)$, let

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

...Connection to OT?

- Denote $\pi(\omega) = \operatorname{proj}_{\operatorname{img}(f)}(\omega)$ and $\nu = \pi_{\#\rho}$:

$$\mathcal{J}_p(f;\rho) = \mathbb{W}_p^p(\rho, \nu)$$

Summary of "Filling"

For $p \geq 1$, $\rho \in \mathcal{P}(\Omega)$, let

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

...Connection to OT?

- Denote $\pi(\omega) = \mathrm{proj}_{\mathrm{img}(f)}(\omega)$ and $\nu = \pi_{\#\rho}$:

$$\begin{aligned}\mathcal{J}_p(f;\rho) &= \mathbb{W}_p^p(\rho,\nu) \\ &= \inf_{\nu' \in \mathcal{P}(\mathrm{img}(f))} \mathbb{W}_p^p(\rho,\nu')\end{aligned}$$

- Next: m-surfaces; higher-order constraints

Quantifying "Complexity"

- No meaningful constraint $\implies$ "efficiency" meaningless
- Our choice: $W^{k,q}(X;\Omega)$ Sobolev norm (denoted $\mathcal{C}(f)$)

Quantifying "Complexity"

- No meaningful constraint $\implies$ "efficiency" meaningless
- Our choice: $W^{k,q}(X;\Omega)$ Sobolev norm (denoted $\mathcal{C}(f)$)
- Problem statement:

Given:	**Goal:**
1. A measure $\rho \in \mathcal{P}(\Omega)$, 2. A budget $\ell \geq 0$, 3. A space $W^{k,q}(X;\Omega)$	■ Minimize: $\mathcal{J}_p(f;\rho)$ ■ Subject to: $\mathcal{C}(f) \leq \ell$

  - Technical hypotheses: X, Ω are "nice," $1 < q < \infty$, and $kq > m$
  - Example applications (Routing, Catalyst Design, ML)

Related work

- Average-Distance Problem: $m = 1$, $\mathcal{C}(f) = \mathcal{H}^1(\text{img}(f))$
  - Buttazzo, Oudet, and Stepanov, 2002; many more (see review in Lemenant, 2012)
  - Doesn't generalize obviously to $m > 1$
- Principal Curves and Surfaces: $\mathcal{C}(f) = \text{length}(f)$
  - Hastie and Stuetzle, 1989; Kégl et al., 2000; Kirov and Slepčev, 2017; Lu and Slepčev, 2013, 2016, 2020; Delattre and Fischer, 2018
- Smoothing splines: ρ essentially m-dimensional; $p = 2$, $\mathcal{C}(f) = \sum_{|\alpha|=k} \|D^\alpha f_i\|_{L^2}$
  - Tons of references
  - Some formulations more similar to ours (Wang, Pottman, and Liu, 2006)

Comparison to Principal Curves

- Principal curves: $\mathcal{C}(f) = \text{length}(f)$
- Key distinction:
  - $\text{length}(f)$ parametrization-independent.
  - For $k > 1$, $\|f\|_{W^{k,q}(X;\Omega)}$ **severely** parametrization-dependent.

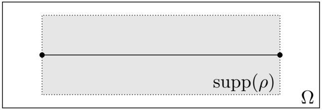
Case Study: Local Improvements

Question: Given δ extra budget, can we improve $\mathcal{J}_p(f)$?

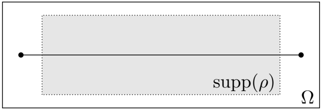
Case Study: Local Improvements

Question: Given δ extra budget, can we improve $\mathcal{J}_p(f)$?

- Can't just "extend" f past $\partial f(X)$...



(a) $f(X)$

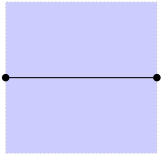


(b) Extended $f(X)$

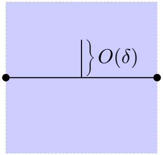
Case Study: Local Improvements

Question: Given δ extra budget, can we improve $\mathcal{J}_p(f)$?

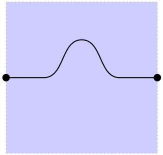
- Can't just "extend" f past $\partial f(X)$...
- Interior modifications? Challenging when $k > 1$



(a) Initial f



(b) length(f) case



(c) Sobolev case

Trying "Gradient Flow" of some kind?

Theorem

Let $f \in C(X;\Omega)$ and $\xi \in C(X;\mathbb{R}^n)$. Then if

1. $p > 1$, *or*
2. $p = 1$ *with a mild hypothesis,*

there exists a well-defined vector field F_ξ and measure μ_ξ on X s.t.

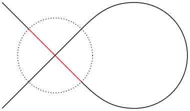
$$\delta_\xi \mathcal{J}_p(f;\rho) = \int_X -\langle F_\xi, \xi \rangle \, d\mu_\xi(x)$$

With some extra hypotheses, F_ξ, μ_ξ can be made independent of ξ.

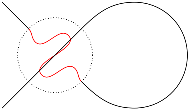
- Constraint-agnostic!
- ...But this causes problems of its own

Can handle complicated modifications

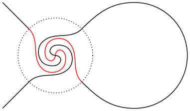
Example 1:



(a) Initial f

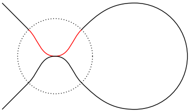


(b) Ex. perturbation 1

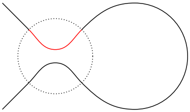


(c) Ex. perturbation 2

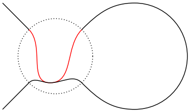
Example 2:



(a) Another initial f



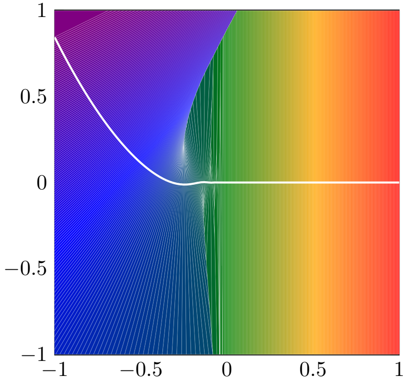
(b) Ex. perturbation 1



(c) Ex. perturbation 2

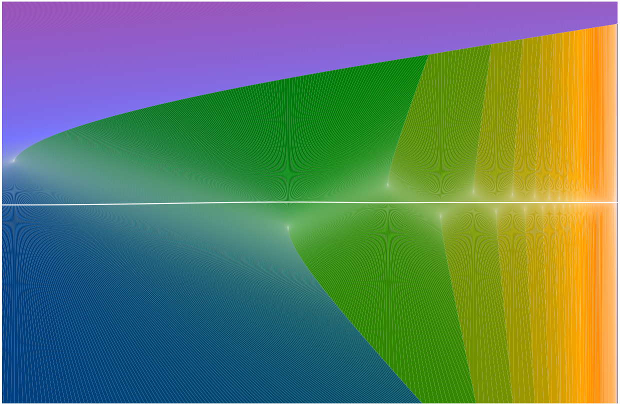
Problem: F_ξ inherits little regularity

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



Problem: F_ξ inherits little regularity

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



Good News

- Improvements exist:

Theorem
If there exists any ξ s.t. F_ξ is nontrivial, then there exists a local $\widetilde{\xi} \in C^\infty$ with $\delta_\xi \mathcal{J}_p(f;\rho) < 0$

- Framework of F_ξ informs approach to other problems (project with Lucas)
- F_ξ trivially well-behaved for discretized f, ρ
  - Theorem: $\mathcal{J}_p(f,\rho)$ jointly continuous. Corollary: Consistency

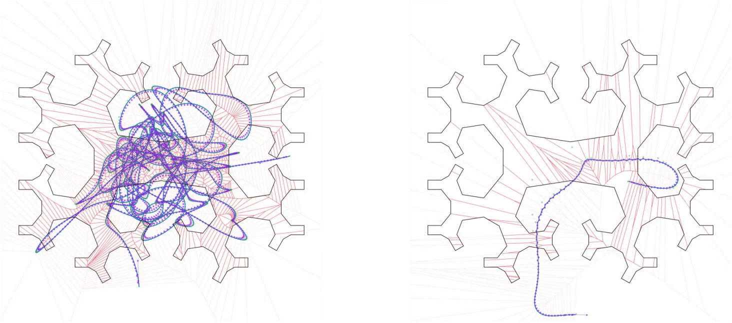
Numerics

- Assume $m = 1$, $n = 2$, ρ uniform, $\mathrm{supp}(\rho)$ convex
- Sample N points from f (assumed to be a spline)
- Fast routine for $G_\xi = \delta_\xi \mathcal{C}(f)$: $O(kN)$
  - $k =$ weak derivative order, so essentially $O(N)$
- Compute Voronoi cells: $O(N \log N)$
- Fast computation of F_ξ for *general* p: $O(NM)$, where M is the average number of boundary points per Voronoi cell
- Perturb by $\eta F_\xi + \lambda G_\xi$ and fit new spline
- Reparametrize by arc length via reduction to elliptic integral

Generative Learning

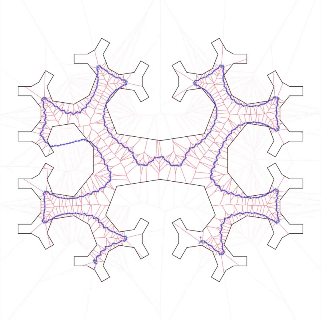
- Soft Proposal: Monge-Kantorovich Fitting gives loosely-analogous toy problem for some black-box models
- Like WGAN, but learns only geometry of support of approximating measure
- Regularizing *generator*, not critic

Proposed mechanism: Regularizing Generator

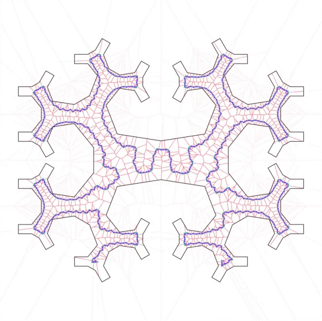


(a) $i = 1$ (b) $i = 110$

Proposed mechanism: Regularizing Generator

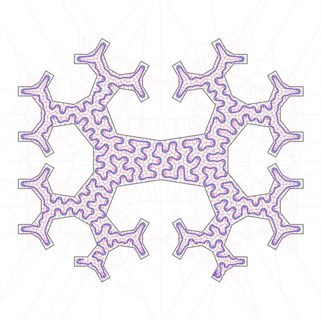


(a) $i = 800$



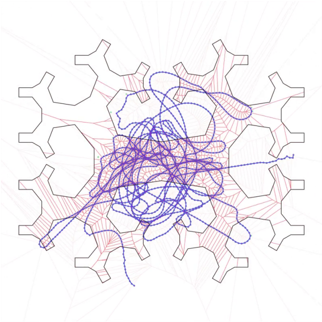
(b) $i = 1120$

Proposed mechanism: Regularizing Generator

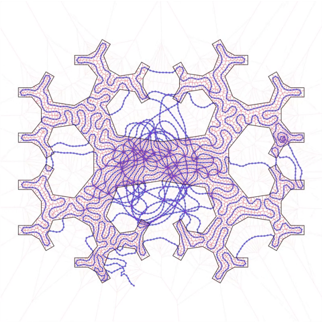


(a) $i = 1850$

(cont.): Unregularized Generator

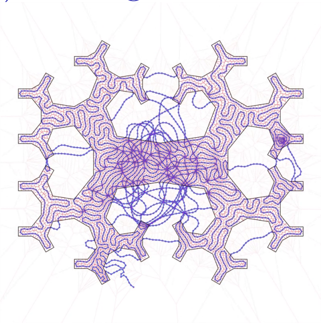


(a) $i = 110$

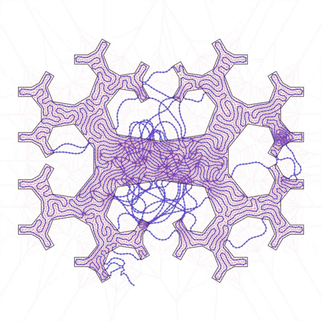


(b) $i = 800$

(cont.): Unregularized Generator

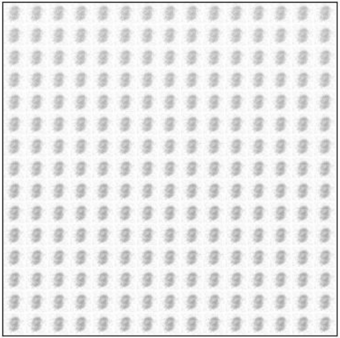


(a) $i = 1120$

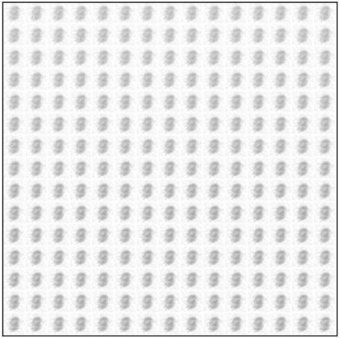


(b) $i = 1850$

Comparison on MNIST



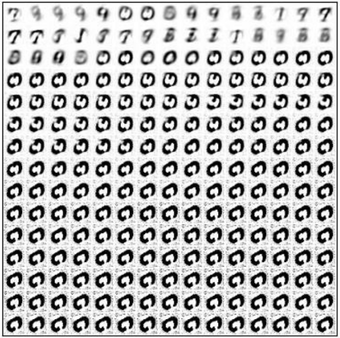
(a) 1 epoch (wd)



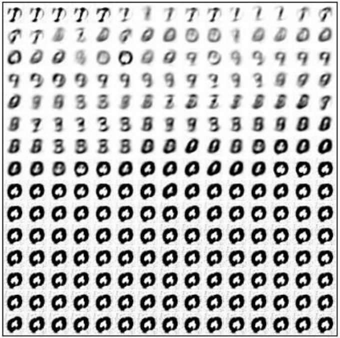
(b) 1 epoch (reg)

Figure: Visualization of f sampled at 15^2 uniformly-spaced points on $[0, 1]$

Comparison on MNIST



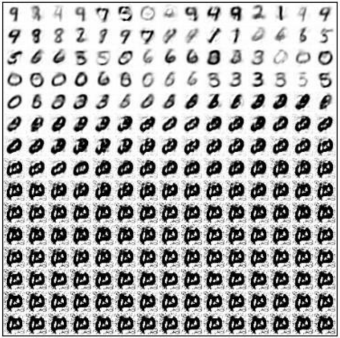
(a) 30 epochs (wd)



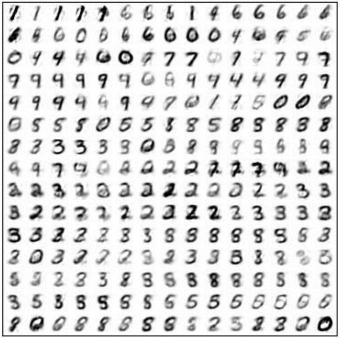
(b) 30 epochs (reg)

Figure: Visualization of f sampled at 15^2 uniformly-spaced points on $[0, 1]$

Comparison on MNIST



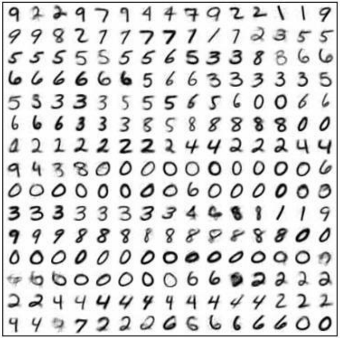
(a) 100 epoch (wd)



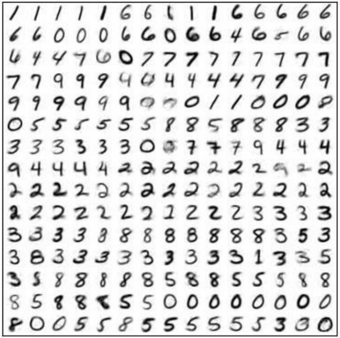
(b) 100 epoch (reg)

Figure: Visualization of f sampled at 15^2 uniformly-spaced points on $[0, 1]$

Comparison on MNIST



(a) 1000 epochs (wd)



(b) 1000 epochs (reg)

Figure: Visualization of f sampled at 15^2 uniformly-spaced points on $[0, 1]$

Thank you!