Monge-Kantorovich Fitting Under a Sobolev Budget

Or: Sobolev-Constrained Principal Curves and Surfaces

Forest Kobayashi (UBC)

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

Friday, January 10th 2025

Introduction

- Goal: "Efficiently" "approximate" n-d measures via m-surfaces
  - "Approximate" in what sense?
  - "Efficient" in what sense?
- Outline:
  - Motivation
  - Defining terms
  - Problem Statement + Previous works
  - Local improvement
  - Numerics; ML Applications

Example: Extracting signal from noisy data

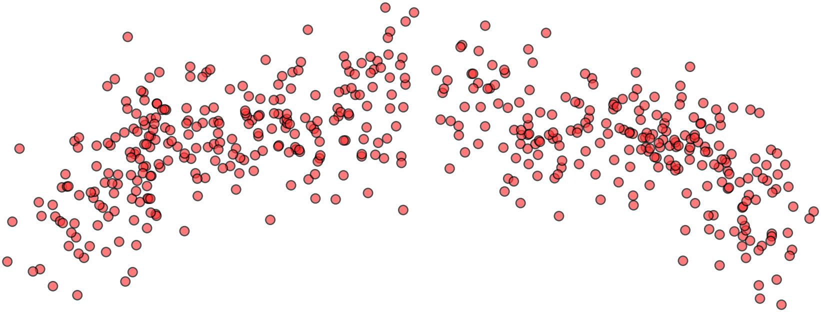


Figure: Noisy observations of a 1d process

Example: Extracting signal from noisy data

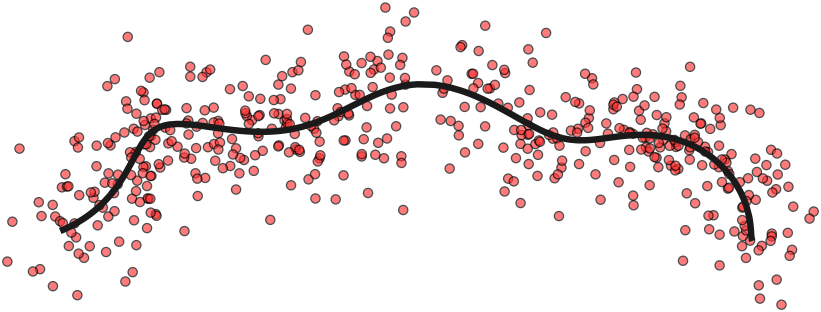


Figure: Guess for a generating curve

Example: Approximating $\rho \in \mathcal{P}(\mathbb{R}^2)$ by a curve

Fire danger rating

The fire danger rating (the risk of a wildfire starting) for the province is updated daily at approximately 2 pm.

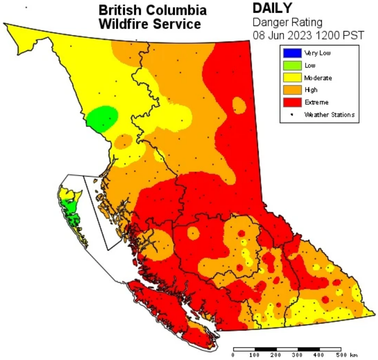
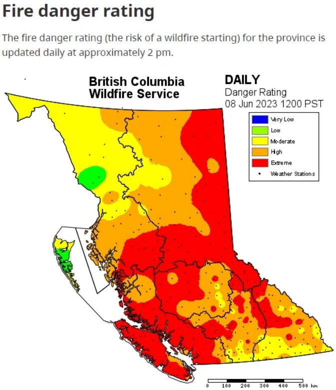


Figure: Wildfire danger map

Example: Approximating $\rho \in \mathcal{P}(\mathbb{R}^2)$ by a curve



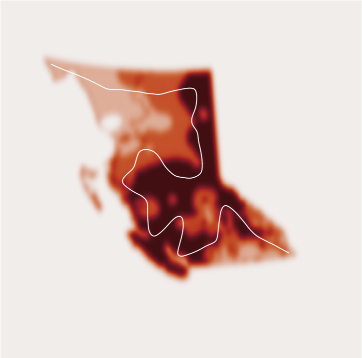
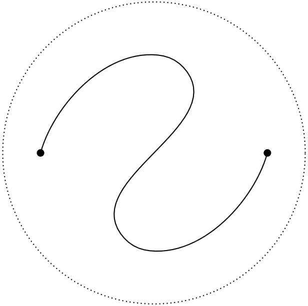


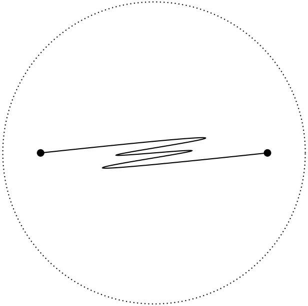
Figure: Example trajectory

How to quantify approximation?

Q: Which shape better “fills” the disk? Why?



(a) One candidate

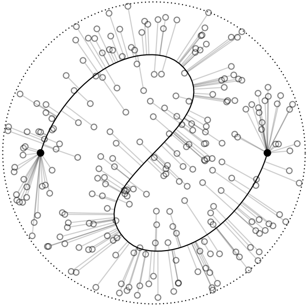


(b) ...And another

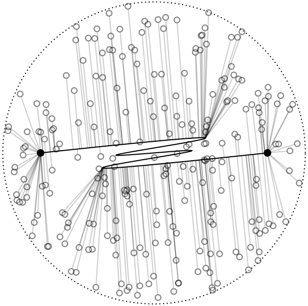
How to quantify approximation?

Q: Which shape better “fills” the disk? Why?

One idea...



(a) A “good” filling



(b) A “bad” filling

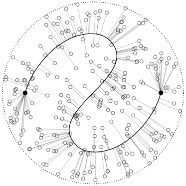
Quantifying "approximation," cont.

Given

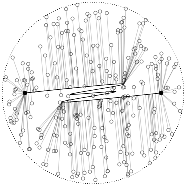
- $\Omega \subseteq \mathbb{R}^n$,
- $f : X \subseteq \mathbb{R}^m \to \Omega$

Let

$$\mathcal{J}(f) = \frac{1}{|\Omega|} \int_\Omega d(\omega, \operatorname{img}(f)) \; d\omega.$$



(a)



(b)

Quantifying “approximation,” cont.

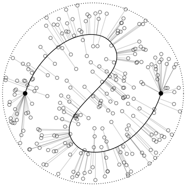
Given

- $\Omega \subseteq \mathbb{R}^n$,
- $f : X \subseteq \mathbb{R}^m \to \Omega$

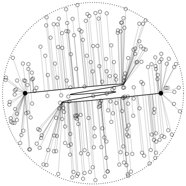
Let

$$\begin{aligned}\mathcal{J}(f) &= \frac{1}{|\Omega|}\int_\Omega d(\omega, \mathrm{img}(f))\ d\omega,\\ &= \mathbb{E}_{\mathrm{Unif}(\Omega)}[d(\omega, \mathrm{img}(f))].\end{aligned}$$

General $\rho \in \mathcal{P}_1(\Omega)$...?



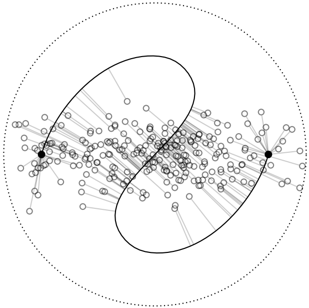
(a)



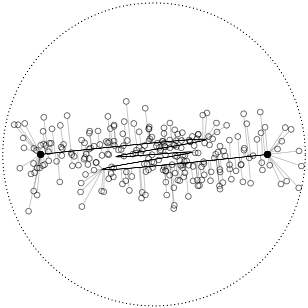
(b)

General $\rho \in \mathcal{P}(\Omega)$

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



(a) Now a “bad” approximation



(b) Now a “good” approximation

Summary of "Approximation"

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

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

- ...Connection to OT?

Summary of "Approximation"

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

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

- ...Connection to OT?

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

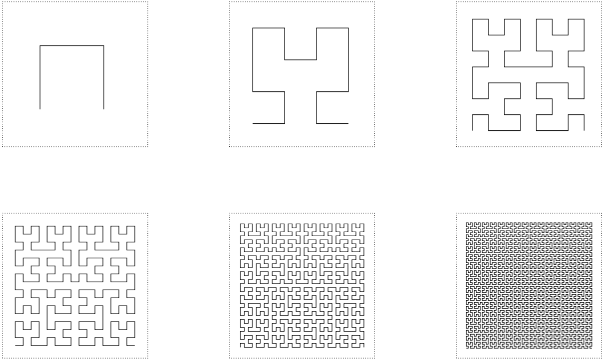
- Next: Defining "Efficiency"

Rationale for "efficiency"

- Need to bound complexity of f, or else problem degenerates

Rationale for “efficiency”

- Need to bound complexity of f, or else problem degenerates



Rationale for “efficiency”

- Need to bound complexity of f, or else problem degenerates

Controlling "complexity" of approximation

- Restrict to some function class $\mathcal{F}$
- Choose some $\mathcal{C} : \mathcal{F} \to [0, \infty)$ ("complexity functional")
- Enforce "efficiency" via:
  - "Soft penalty:" Given $\lambda > 0$,

$$\min_{f \in \mathcal{F}} \left[\mathcal{J}_p(f; \rho) + \lambda \mathcal{C}(f) \right]$$

  - "Hard constraint:" Given $\ell \geq 0$,

$$\min_{\mathcal{C}(f) \leq \ell} \mathcal{J}_p(f; \rho)$$

Some previously-studied $\mathcal{C}$'s

- 1d trees: $\mathcal{C}(f) = \mathcal{H}^1(\mathrm{img}(f))$
  - Buttazzo, Oudet, and Stepanov, 2002; many more (see review in Lemenant, 2012)
  - Doesn't generalize obviously to $m > 1$
- "Classical" principal curves: $\mathcal{C}(f) = \mathrm{length}(f)$
  - Hastie and Stuetzle, 1989; Kégl et al., 2000; Kirov and Slepčev, 2017; Lu and Slepčev, 2013, 2016, 2020; Delattre and Fischer, 2018
  - (Some works add curvature penalty; see Lu and Slepčev 2012)

Our “smoothing” $\mathcal{C}$

- Use $W^{k,q}(X;\mathbb{R}^n)$ Sobolev norm

$$\boxed{\mathcal{C}(f) := \|f\|_{W^{k,q}(X,\mathbb{R}^n)} = \left(\sum_{j=1}^{n} \sum_{|\alpha|\le k} \|D^\alpha f_j\|_{L^q}^q\right)^{1/q}.}$$

- (Technical hypotheses: X, Ω “nice”; $1 < q < \infty$; $kq > m$)
- Similar to smoothing splines: $\mathcal{C}(f) = \sum_{j=1}^{n} \sum_{|\alpha|=k} \|D^\alpha f_j\|_{L^2}$
  - Typically assume labeled data
  - Some approaches similar to ours (Wang, Pottman, and Liu, 2006)

Comparison to $\mathcal{C}(f) = \text{length}(f)$

- Role of parametrization?
  - $\text{length}(f) \to$ invariant
  - For $k > 1$, $\|f\|_{W^{k,q}(X;\Omega)}$ **severely** parametrization-dependent.

Comparison to $\mathcal{C}(f) = \text{length}(f)$

- Role of parametrization?
  - length$(f) \to$ invariant
  - For $k > 1$, $\|f\|_{W^{k,q}(X;\Omega)}$ **severely** parametrization-dependent.
- Tractability of Analysis?
  - length(f): Analysis can be subtle
  - Sobolev norm: Analysis can be *brutal*.

Comparison to $\mathcal{C}(f) = \text{length}(f)$

- Role of parametrization?
  - $\text{length}(f) \to$ invariant
  - For $k > 1$, $\|f\|_{W^{k,q}(X;\Omega)}$ **severely** parametrization-dependent.
- Tractability of Analysis?
  - $\text{length}(f)$: Analysis can be subtle
  - Sobolev norm: Analysis can be *brutal.*
- Sobolev norm in applications?
  - Stronger "smoothing" of noise
  - Routing problems: no instantaneous turns; fuel efficiency
  - Numerics: very hard, but doable

Informal summary of problem

Given:

1. A measure $\rho \in \mathcal{P}(\Omega)$,
2. A budget $\ell \geq 0$,
3. A space $W^{k,q}(X;\Omega)$

Goal:

- Minimize: $\mathcal{J}_p(f;\rho)$
- Subject to: $\mathcal{C}(f) \leq \ell$ where $\mathcal{C}(f) = \|f\|_{W^{k,q}(X;\Omega)}$

- Givens: ρ, ℓ, X, Ω
- Decision variable: f
- Tweakable parameters:
  - $\mathcal{J}_p$: $p \geq 1$
  - $\mathcal{C}$: k, q (subject to $1 < q < \infty$ and $kq > m$)

Flavor of problem

- $\mathscr{J}_p$:
  - Continuity: ✓ Joint in f, ρ
  - Convexity/concavity: ✗ Neither
- Optimizers:
  - Exist ✓
  - ...but not necessarily unique
  - Characterization: harder than in length case
- Local improvements?

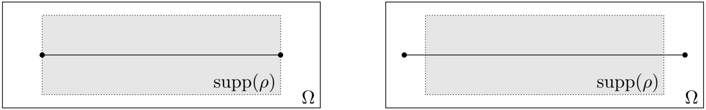
Case Study: Local Improvements

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 $\partial f(X)$...

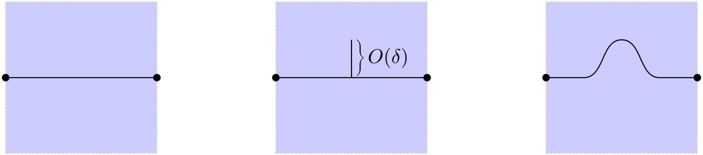


(a) $f(X)$ (b) Extended $f(X)$

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 (Informal)

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.

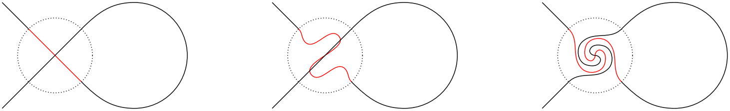
$$\lim_{\varepsilon \to 0} \frac{\mathcal{J}_p(f + \varepsilon\xi; \rho) - \mathcal{J}_p(f)}{\varepsilon} = \int_X -\langle F_\xi, \xi \rangle \, d\mu_\xi(x).$$

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

- F_ξ moves f toward "barycenters" of fibers
- But, can badly lack regularity

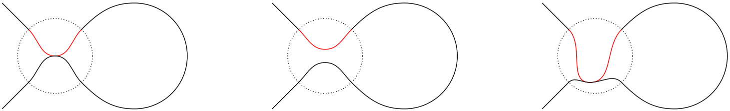
Can handle complicated modifications

Example 1:



(a) Initial f (b) Ex. perturbation 1 (c) Ex. perturbation 2

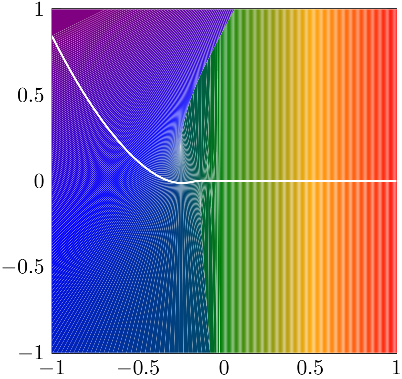
Example 2:



(a) Another initial f (b) Ex. perturbation 1 (c) Ex. perturbation 2

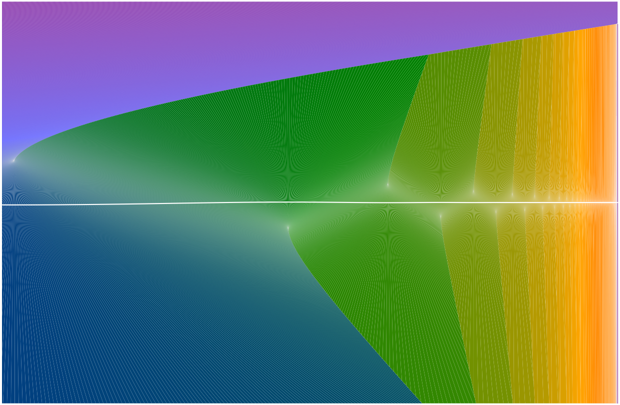
Problem: F_ξ inherits little regularity

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 (axes not to scale!).



Good News

- Improvements exist:

Theorem (Informal)

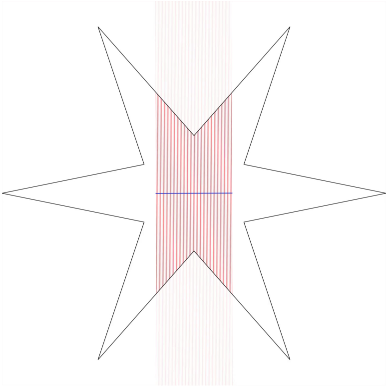
If there exists any ξ s.t. F_ξ is nontrivial, then there exists a local $\widetilde{\xi} \in C^\infty$ decreasing $\mathcal{J}_p$.

- F_ξ works for discretized f, ρ
- Framework of F_ξ useful in other problems
  - Forthcoming work (O'Brien, **K.**, Kim): Resolved a notable open problem for tree case

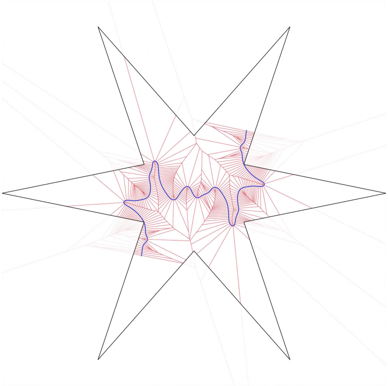
Numerics

- Special case: ρ uniform on 2d, PL set
  - Precise closed-form computation for F_ξ (no monte-carlo)!
- Case of discrete ρ, general dimension:
  - Paper to come...
  - Blog post explaining part:

Example numerics

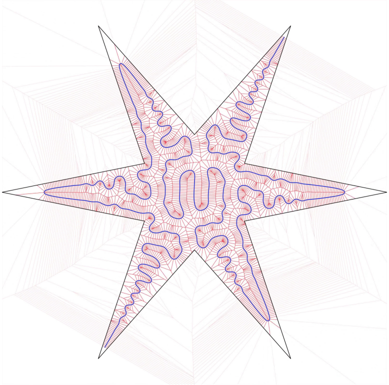


(a) $i = 1$

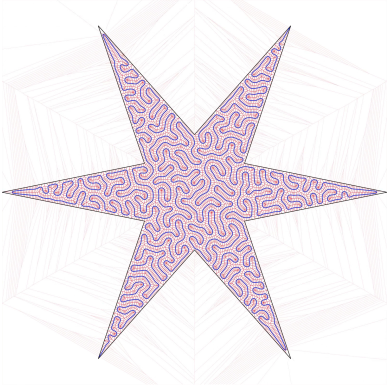


(b) $i = 150$

Example numerics



(a) $i = 300$

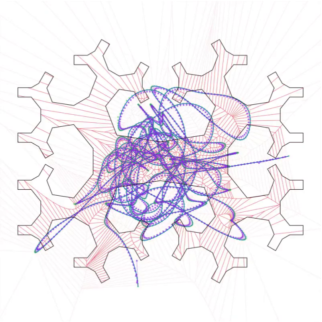


(b) $i = 1000$

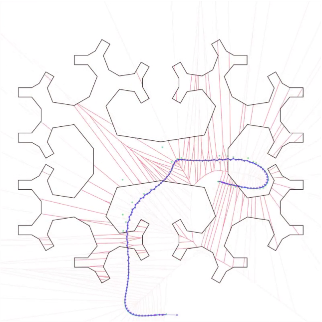
Generative Learning

- Soft Proposal: Monge-Kantorovich Fitting gives loosely-analogous toy problem for some black-box models
- Like (regularized) WGAN, but
  - Learns only support of approximating measure (not parametrization)
  - Regularization on *generator*, not critic
- Gives a possible interpretation for what regularization “does”

A regularized "generator"

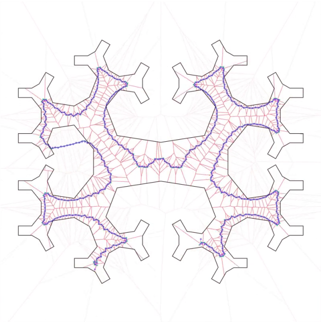


(a) $i = 1$

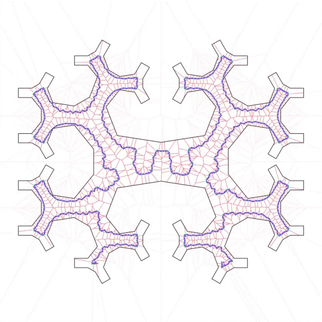


(b) $i = 110$

A regularized "generator"

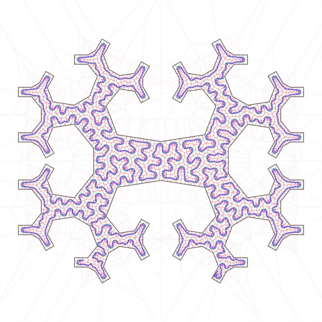


(a) $i = 800$



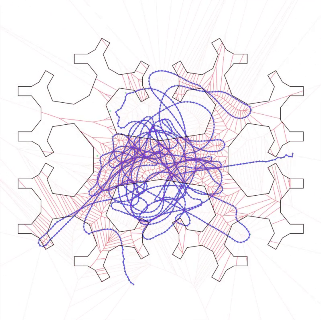
(b) $i = 1120$

A regularized “generator”

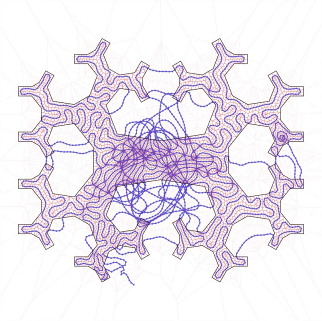


(a) $i = 1850$

(cont.): Unregularized Generator

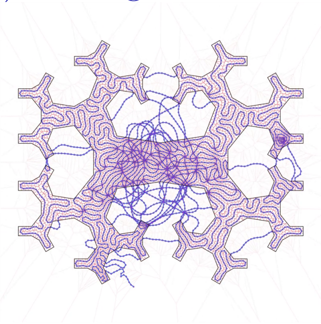


(a) $i = 110$

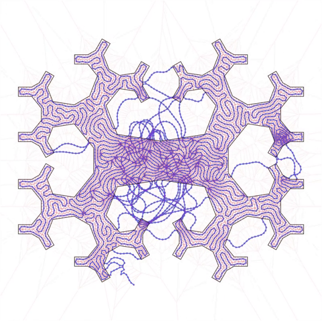


(b) $i = 800$

(cont.): Unregularized Generator

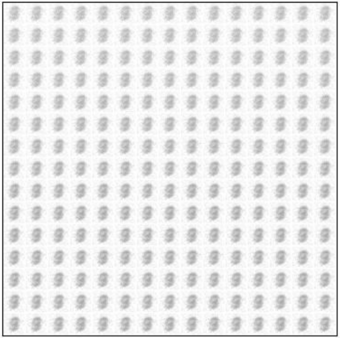


(a) $i = 1120$

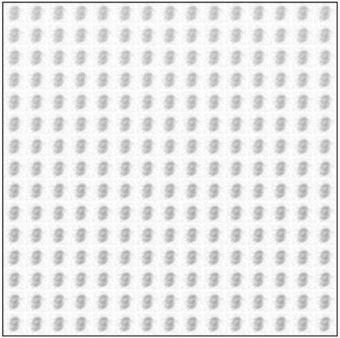


(b) $i = 1850$

Comparison on MNIST



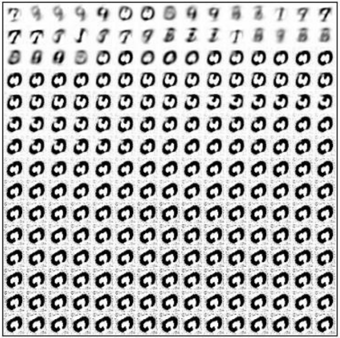
(a) 1 epoch (wd)



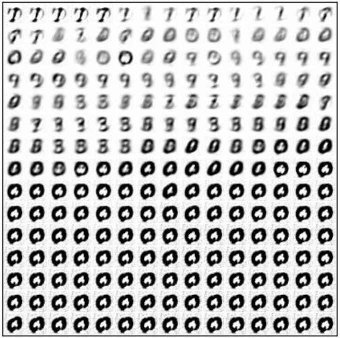
(b) 1 epoch (reg)

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

Comparison on MNIST



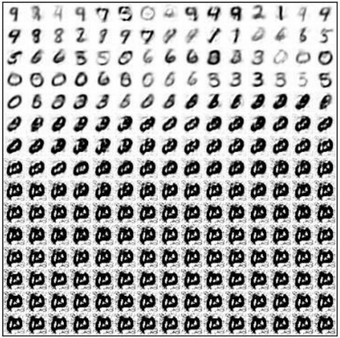
(a) 30 epochs (wd)



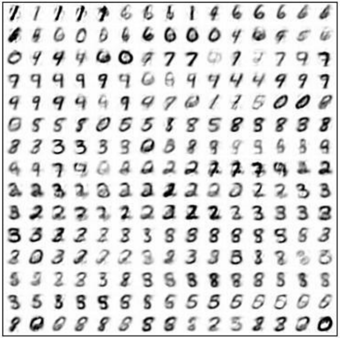
(b) 30 epochs (reg)

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

Comparison on MNIST



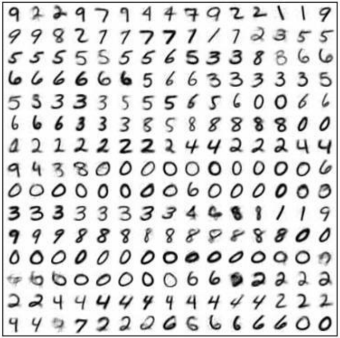
(a) 100 epoch (wd)



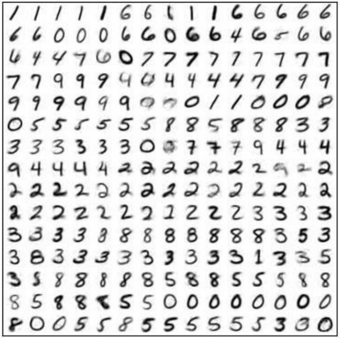
(b) 100 epoch (reg)

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!