Dynamic optimization on graphons

Zaid Harchaoui[1,3], Sewoong Oh[1,4], Soumik Pal[2], Raghav Somani[1] and Raghav Tripathi[2]

[1]UW CSE, [2]UW Math, [3]UW Statistics & [4]Google

February 2, 2023





Plan

- Motivation(s)
- Detour: Interacting particle system
- Optimization on graphons: Setup and Results
- Proof Sketches
- Future directions

Motivation (Problem 1)

Mantel-Turán Problem

Among all graphs on n vertices **containing no triangles**, maximize the number of edges.

Since we are interested in large n, we normalize. Let's define

$$t(K_3, G) = \frac{\text{No. of triangles in } G}{n^3}, \qquad t(K_2, G) = \frac{\text{No. of edges in } G}{n^2}.$$

Problem

Maximize $t(K_2, G)$ subject to the constraint $t(K_3, G) = 0$.

Mantel-Turán Theorem

$t(K_2, G) > \frac{1}{2} \implies t(K_3, G) > 0.$

Motivation (Problem 2)

Erdös Theorem

$t(C_4, G) \geq t(K_2, G)^4$ where $t(C_4, G) = (\text{No. of 4-cycles in } G)/n^4$.

Problem

Find the minimum of $t(C_4, G)$ over all graphs with $t(K_2, G) \geq 1/2$.

- We know that the 4-cycle density must be $\geq 1/16$.
- $t(C_4, G) = 1/16$ is not achieved by any finite graph.
- $t(C_4, G)$ can be arbitrarily close to $1/16$ for appropriate families of graphs.

Motivation (Problem 3)

- Let G be a weighted graph on n vertices with weighted adjanceny matrix A.
- Let F be a finite simple graph on m vertices on m vertices.
- We define *homomorphism density* of F into K

$$t(F,G) = \frac{1}{n^m} \sum_{i_1,i_2,\ldots,i_m} \prod_{\{u,v\}\in E(F)} A(i_u,i_v)\ .$$

Ising model on graphs

- $F :=$ A graph on m vertices. Every vertex may have a state $1,2,\ldots,q$.
- Between two neighboring vertices with states i,j, there is an interaction energy J_{ij}.
- A configuration is a map $\sigma : V(F) \to [q]$.
- The partition function is given by

$$Z = \sum_{\sigma:V(F)\to[q]} \exp\left(-\sum_{uv\in E(F)} J_{\sigma(u),\sigma(v)}\right) = \sum_{\sigma:V(G)\to[q]} \prod_{uv\in E(F)} \beta_{\sigma(u),\sigma(v)},$$

  where $\beta_{ij} = \exp\left(-J_{ij}\right)$.
- Miniminizing Z is equivalent to minimizing$t(F, K_q^{\beta})$, where K_q^{β} is complete graph with edge weights β_{ij}.

Summary

- There are interesting optimization problems on graphs.
- Some of these optimization problem may not admit solutions in the space of finite graphs.

Summary

- There are interesting optimization problems on graphs.
- Some of these optimization problem may not admit solutions in the space of finite graphs.

Plan

- Fill in the holes in the space of graphs, that is, take a completion of the space of all finite graphs.
- Try solving optimization problem on the complete space.
- These optimization problems have rich symmetries (exchangeability). Can we exploit that?

Detour: Interacting Diffusion (McKean, Kac, Snitzman, JKO ...)

Consider the following example of interacting diffusions

$$dX_t^{i,N} = \frac{1}{N}\sum_{j=1}^{N} \nabla b(X_t^{i,N} - X_t^{j,N})\,\mathrm{d}t + dW_t^i, \quad i = 1, \dots, N$$
$$X_0^{i,N} = x_0^i\ .$$

Let $\mu_t^N \coloneqq N^{-1}\sum_{i=1}^{N} \delta_{X^{i,N}(t)}$.

Detour: Interacting Diffusion (McKean, Kac, Snitzman, JKO ...)

Consider the following example of interacting diffusions

$$dX_t^{i,N} = \frac{1}{N}\sum_{j=1}^{N} \nabla b(X_t^{i,N} - X_t^{j,N})\,\mathrm{d}t + dW_t^i, \quad i = 1,\ldots,N$$
$$X_0^{i,N} = x_0^i\,.$$

Let $\mu_t^N \coloneqq N^{-1}\sum_{i=1}^{N}\delta_{X^{i,N}(t)}$.Then, $\mu_t^N \to \mu_t$ weakly where μ_t is a gradient flow with respect to 2-Wasserstein metric, given as

$$\partial_t \mu_t(x) = -\mathrm{div}_x\left[\mu_t(x)\cdot(\nabla b * \mu_t)(x)\right] + \frac{1}{2}\Delta\mu_t(x)\,. \tag{1.1}$$

Detour: Interacting Diffusion (McKean, Kac, Snitzman, JKO ...)

Consider the following example of interacting diffusions

$$dX_t^{i,N} = \frac{1}{N}\sum_{j=1}^{N} \nabla b(X_t^{i,N} - X_t^{j,N})\,\mathrm{d}t + dW_t^i, \quad i = 1, \dots, N$$
$$X_0^{i,N} = x_0^i\,.$$

Let $\mu_t^N \coloneqq N^{-1}\sum_{i=1}^{N} \delta_{X^{i,N}(t)}$.Then, $\mu_t^N \to \mu_t$ weakly where μ_t is a gradient flow with respect to 2-Wasserstein metric, given as

$$\partial_t \mu_t(x) = -\mathrm{div}_x\left[\mu_t(x)\cdot(\nabla b * \mu_t)(x)\right] + \frac{1}{2}\Delta\mu_t(x)\,. \tag{1.1}$$

Interacting particles system converges to McKean-Vlasov

Suppose $X_0^{i,N}$ are i.i.d. with distribution μ_0. As $N \to \infty$, each $X_\cdot^{i,N}$ has a natural limit $\bar{X}^i$. Each $\bar{X}^i$ is an independent copy of following McKean-Vlasov process

$$dX_t = (\nabla b * \mu_t)(X_t)\,\mathrm{d}t + dB_t\,, X_{t=0} = X_0 \sim \mu_0\,.$$

Detour continued...

Morality

- Think of each particle $X^{i,N}$ as doing a (noisy) gradient flow.
- Drift of the particle $X^{i,N}$ depends on 'itself' $X^{i,N}$ and 'on the ensemble' $N^{-1}\sum_{i=1}^{N}\delta_{X^{i,N}}$ in a symmetric way.
- Then, 'the ensemble limit' also performs a gradient flow in suitable sense.
- And, the evolution of a typical particle can be described by a McKean-Vlasov equation.

Introduction

Objective

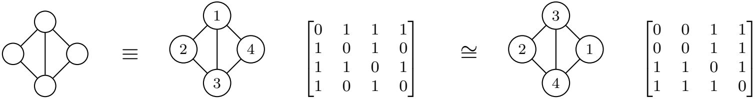
Study large scale optimization problems over dense weighted **unlabeled graphs**.

Introduction

Objective

Study large scale optimization problems over dense weighted **unlabeled graphs**.

Let $G = (V, E)$ be a graph and let A be an adjacency matrix of G.



$$\equiv \quad \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad \cong \quad \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

Figure: Symmetry in unlabeled graphs.

Examples of functions

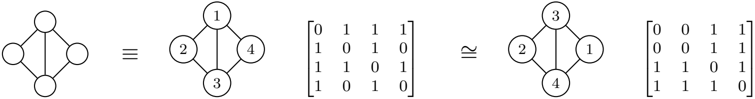
- Edge density: $h_{-}(G) = (\# \text{ of edges in } G)/n^2$.
- Triangle density: $h_{\triangle}(G) = (\# \text{ of } \triangle\text{s in } G)/n^3$.

Introduction

Objective

Study large scale optimization problems over dense weighted **unlabeled graphs**.

Let $G = (V, E)$ be a graph and let A be an adjacency matrix of G.



$$\equiv \quad \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad \cong \quad \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

Figure: Symmetry in unlabeled graphs.

Examples of functions

- Edge density: $h_{-}(G) = (\#\text{ of edges in } G)/n^2$.
- Triangle density: $h_{\triangle}(G) = (\#\text{ of }\triangle\text{s in } G)/n^3$.

Invariant functions

A function $F\colon \mathcal{M}_n \to \mathbb{R}$ is said to be *invariant function/graph function* if $F(A) = F(A^\sigma)$ for all permutations $\sigma \in S_n$ and $A \in \mathcal{M}_n$, where $A^\sigma(i,j) = A(\sigma(i), \sigma(j))$.

Plan and analogies with interacting diffusion

Objective

Let F be graph function. Our goal is to minimize F over large graphs.

Can perform gradient descent on finite graphs/symmetric matrices.

Exploiting the symmetry

- Think of the problem as an optimization problem on the space of 'graphons'.
- Hope-Pray-Prove! The gradient descent process on finite graphs/symmetric matrices converge to a limit as $n \to \infty$.

Plan and analogies with interacting diffusion

Objective

Let F be graph function. Our goal is to minimize F over large graphs.

Can perform gradient descent on finite graphs/symmetric matrices.

Exploiting the symmetry

- Think of the problem as an optimization problem on the space of ‘graphons’.
- Hope-Pray-Prove! The gradient descent process on finite graphs/symmetric matrices converge to a limit as $n \to \infty$.
- Can we show that the limit of GD is a gradient flow on graphons?
- Can one construct natural Markov processes on graphs that converge to the gradient flow?

Plan and analogies with interacting diffusion

Objective

Let F be graph function. Our goal is to minimize F over large graphs.

Can perform gradient descent on finite graphs/symmetric matrices.

Exploiting the symmetry

- Think of the problem as an optimization problem on the space of 'graphons'.
- Hope-Pray-Prove! The gradient descent process on finite graphs/symmetric matrices converge to a limit as $n \to \infty$.
- Can we show that the limit of GD is a gradient flow on graphons?
- Can one construct natural Markov processes on graphs that converge to the gradient flow?

Graphons vs Wasserstein space

- Given a graph on n vertices is akin to particle ensemble
- Think of every edge as a *particle* and edge-weights are evolving

Setup and Results

Graphons

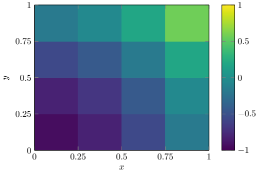
Kernels $\mathcal{W}$

A kernel is a measurable function $W\colon [0,1]^2 \to [-1,1]$ such that $W(x,y) = W(y,x)$.

- Adjacency matrix $\equiv$ *kernel*.

$$\frac{1}{16}\begin{bmatrix} -16 & -15 & -12 & -7 \\ -15 & -14 & -11 & 1 \\ -12 & -11 & -6 & 4 \\ -7 & 1 & 4 & 9 \end{bmatrix}$$

Symmetric matrix A



Kernel representation of A

Graphons

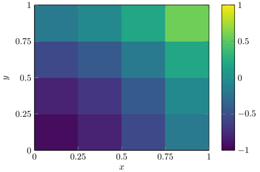
Kernels $\mathcal{W}$

A kernel is a measurable function $W\colon [0,1]^2 \to [-1,1]$ such that $W(x,y) = W(y,x)$.

- Adjacency matrix $\equiv$ *kernel.*

$$\frac{1}{16}\begin{bmatrix} -16 & -15 & -12 & -7 \\ -15 & -14 & -11 & 1 \\ -12 & -11 & -6 & 4 \\ -7 & 1 & 4 & 9 \end{bmatrix}$$

Symmetric matrix A



Kernel representation of A

- Identify adjacency matrix/kernel up to ‘permutations’.
- Identify $W_1 \cong W_2$ if one can be obtained by ‘relabeling’ the vertices of the other, i.e.,

$$W_1(\varphi(x), \varphi(y)) = W_2(x,y), \qquad \text{where} \phi : [0,1] \to [0,1] \text{is a measure preserving map.}$$

Graphons

Graphons $\widehat{\mathcal{W}}$ (Lovász & Szegedy, 2006): $\widehat{\mathcal{W}} := \mathcal{W}/\cong$

Cut metric :: Weak convergence

- Cut metric, $\delta_\square$, metrizes graph convergence.
- $(\widehat{\mathcal{W}}, \delta_\square)$ is **compact**.

[1] Gradient flows on graphons - Oh, Pal, Somani, Tripathi, 2021

[2] Gradient Flows: In Metric Spaces and in the Space of Probability Measures - Ambrosio, Gigli, Savaré, 2008

Graphons

Graphons $\widehat{\mathcal{W}}$ (Lovász & Szegedy, 2006): $\widehat{\mathcal{W}} := \mathcal{W}/\cong$

Cut metric :: Weak convergence

- Cut metric, $\delta_\square$, metrizes graph convergence.
- $(\widehat{\mathcal{W}}, \delta_\square)$ is **compact**.

Invariant L^2 metric δ_2 :: 2-Wasserstein metric $\mathbb{W}_2$

- Stronger than the cut metric (i.e., $\delta_\square$ convergence $\not\Rightarrow$ δ_2 convergence).
- **Gromov-Wasserstein distance** between $([0,1], \mathrm{Leb}, W_1)$ and $([0,1], \mathrm{Leb}, W_2)$.

[1] Gradient flows on graphons - Oh, Pal, Somani, Tripathi, 2021

[2] Gradient Flows: In Metric Spaces and in the Space of Probability Measures - Ambrosio, Gigli, Savaré, 2008

Graphons

Graphons $\widehat{\mathcal{W}}$ (Lovász & Szegedy, 2006): $\widehat{\mathcal{W}} := \mathcal{W}/\cong$

Cut metric :: Weak convergence

- Cut metric, $\delta_\square$, metrizes graph convergence.
- $(\widehat{\mathcal{W}}, \delta_\square)$ is **compact**.

Invariant L^2 metric δ_2 :: 2-Wasserstein metric $\mathbb{W}_2$

- Stronger than the cut metric (i.e., $\delta_\square$ convergence $\not\Rightarrow$ δ_2 convergence).
- **Gromov-Wasserstein distance** between $([0,1], \mathrm{Leb}, W_1)$ and $([0,1], \mathrm{Leb}, W_2)$.

We show[1]

- The metric δ_2 is **geodesic** (just like $\mathbb{W}_2$). Geodesic convexity on $(\widehat{\mathcal{W}}, \delta_2)$.
- Notion of 'gradient' on $(\widehat{\mathcal{W}}, \delta_2)$ called 'Frechét-like derivative'!
- Construction of 'gradient flows' on $(\widehat{\mathcal{W}}, \delta_2)$[2].

[1] Gradient flows on graphons - Oh, Pal, Somani, Tripathi, 2021

[2] Gradient Flows: In Metric Spaces and in the Space of Probability Measures - Ambrosio, Gigli, Savaré, 2008

Existence of gradient flow on Graphons

Theorem [OPST '21]

If $R\colon \widehat{\mathcal{W}} \to \mathbb{R}$

- has a Fréchet-like derivative,
- is geodesically semiconvex in δ_2,

then starting from any $W_0 \in \widehat{\mathcal{W}}$, $\exists!$ gradient flow curve $(W_t)_{t\in\mathbb{R}_+}$ for R

Existence of gradient flow on Graphons

Theorem [OPST '21]

If $R\colon \widehat{\mathcal{W}} \to \mathbb{R}$

- has a Fréchet-like derivative,
- is geodesically semiconvex in δ_2,

then starting from any $W_0 \in \widehat{\mathcal{W}}$, $\exists!$ gradient flow curve $(W_t)_{t\in\mathbb{R}_+}$ for R satisfying

$$W_t \coloneqq W_0 - \int_0^t DR(W_s)\,\mathrm{d}s, \qquad t \in \mathbb{R}_+,$$

inside $\widehat{\mathcal{W}}$. At the boundary $\{-1, 1\}$ of $\widehat{\mathcal{W}}$, add constraints to contain it.

Scaling limits of GD [OPST '21 + HOPST '22]

Euclidean GD/SGD of R_n over $n \times n$ symmetric matrices, converges to the 'gradient flow' of R on the metric space of graphons.

Example

For $p \in [0,1]$, define the entropy function $I(p) = p\log(p) + (1-p)\log(p)$. In the following we assume $W : [0,1]^2 \to [0,1]$.
For a kernel W, define

$$I(W) := \iint I(W(x,y))\,\mathrm{d}x\,\mathrm{d}y.$$

Gradient flow of $F(W) = t(K_3, W) + \beta I(W)$

$$W_t(x,y) = W_0(x,y) - 3\int_0^t \int W_s(x,z)W_s(z,y)\,\mathrm{d}z\,\mathrm{d}s - \beta\int_0^t \log\left(\frac{W_s(x,y)}{1-W_s(x,y)}\,\mathrm{d}s\right).$$

Example

For $p \in [0,1]$, define the entropy function $I(p) = p\log(p) + (1-p)\log(p)$. In the following we assume $W : [0,1]^2 \to [0,1]$.
For a kernel W, define

$$I(W) := \iint I(W(x,y))\,\mathrm{d}x\,\mathrm{d}y.$$

Gradient flow of $F(W) = t(K_3, W) + \beta I(W)$

$$W_t(x,y) = W_0(x,y) - 3\int_0^t \int W_s(x,z)W_s(z,y)\,\mathrm{d}z\,\mathrm{d}s - \beta\int_0^t \log\left(\frac{W_s(x,y)}{1-W_s(x,y)}\,\mathrm{d}s\right) .$$

Finite dimensional gradient descent

$$W_t^{(n)}(i,j) = W_0^{(n)} - 3n^2\int_0^t \frac{1}{n^3}\left(W_s^{(n)}\right)^2(i,j)\,\mathrm{d}s - \beta\int_0^t \log\left(\frac{W_s^{(n)}(i,j)}{1-W_s^{(n)}(i,j)}\right) .$$

Markov Chain converging to gradient flow

Suppose we want to construct a Markov process on graphs that converges to the gradient flow of triangle density $t(K_3, \cdot)$.

Markov Chain converging to gradient flow

Suppose we want to construct a Markov process on graphs that converges to the gradient flow of triangle density $t(K_3, \cdot)$.

- Start with $G_{n,0}$.
- At each time step τ_n, all the edges in $G_{n,k}$ flip (or don't flip according to following rule).
  - If $\{i, j\}$ is not an edge in $G_{n,k}$ then $\{i, j\}$ remains a non-edge in $G_{n,k+1}$.
  - If $\{i, j\}$ is an edge in $G_{n,k}$ then drop it with probability

$$p_{ij} = \tau_n \frac{\Delta_{ij}}{n} ,$$

where Δ_{ij} = Number of triangles with containing $\{i, j\}$.

Scaling limit of Noisy SGD

For $n \in \mathbb{N}$, let $\nabla R_n(A) = \mathbb{E}_\xi[\nabla \ell_n(A;\xi)]$ for $A \in \mathcal{M}_n$.

SGD

Given the k-th iterate $W_k^{(n)} \in \mathcal{M}_n$, sample ξ,

$$W_{k+1}^{(n)} = W_k^{(n)} - \tau_n \cdot n^2 \underbrace{\nabla \ell_n(W_k^{(n)};\xi)}_{\text{stochastic Euclidean gradient}}$$

Scaling limit of Noisy SGD

For $n \in \mathbb{N}$, let $\nabla R_n(A) = \mathbb{E}_\xi[\nabla \ell_n(A;\xi)]$ for $A \in \mathcal{M}_n$.

Noisy SGD

Given the k-th iterate $W_k^{(n)} \in \mathcal{M}_n$, sample ξ,

$$W_{k+1}^{(n)} = W_k^{(n)} - \tau_n \cdot n^2 \underbrace{\nabla \ell_n(W_k^{(n)};\xi)}_{\text{stochastic Euclidean gradient}} + \tau_n^{1/2} \cdot \underbrace{\xi_k \sim N(0,I)}_{\text{independent added noise}}$$

Scaling limit of Noisy SGD

For $n \in \mathbb{N}$, let $\nabla R_n(A) = \mathbb{E}_\xi[\nabla \ell_n(A;\xi)]$ for $A \in \mathcal{M}_n$.

Noisy SGD

Given the k-th iterate $W_k^{(n)} \in \mathcal{M}_n$, sample ξ,

$$W_{k+1}^{(n)} = P\left(W_k^{(n)} - \tau_n \cdot n^2 \underbrace{\nabla \ell_n(W_k^{(n)};\xi)}_{\text{stochastic Euclidean gradient}} + \tau_n^{1/2} \cdot \underbrace{\xi_k \sim N(0,I)}_{\text{independent added noise}} \right)$$

Scaling limit of Noisy SGD

For $n \in \mathbb{N}$, let $\qquad \nabla R_n(A) = \mathbb{E}_\xi[\nabla \ell_n(A;\xi)] \qquad$ for $A \in \mathcal{M}_n$.

Noisy SGD

Given the k-th iterate $W_k^{(n)} \in \mathcal{M}_n$, sample ξ,

$$W_{k+1}^{(n)} = P\left(W_k^{(n)} \quad - \quad \tau_n \cdot n^2 \underbrace{\nabla \ell_n(W_k^{(n)};\xi)}_{\text{stochastic Euclidean gradient}} \quad + \quad \tau_n^{1/2} \cdot \underbrace{\xi_k \sim N(0,I)}_{\text{independent added noise}} \right)$$

If $W_0^{(n)} \xrightarrow{\delta_2} W_0$, and $\tau_n \to 0$, as $n \to \infty$, then a.s.

$$W^{(n)} \overset{\delta_\square}{\rightrightarrows} \Gamma, \qquad \text{as } n \to \infty,$$

where $\Gamma \colon t \mapsto \Gamma(t)$ is the curve described by the McKean-Vlasov equation.

McKean-Vlasov equation

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a Brownian Motion $B(t)$, and $(U, V) \overset{\text{i.i.d.}}{\sim} \text{Uni}[0,1]$.
- Consider the process $(X(t), \Gamma(t))$ such that

Existence + uniqueness when DR is L^2 Lipschitz - [HOPST '22]

McKean-Vlasov equation

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a Brownian Motion $B(t)$, and $(U, V) \overset{\text{i.i.d.}}{\sim} \text{Uni}[0,1]$.
- Consider the process $(X(t), \Gamma(t))$ such that on $\{U = u, V = v\}$,

$$\mathrm{d}X(t) = -(DR)(\Gamma(t))(u,v)\,\mathrm{d}t + \mathrm{d}B(t) \quad \underbrace{+\,\mathrm{d}L^{-}(t) - \mathrm{d}L^{+}(t)}_{\text{constrain in } [-1,1]}, \qquad \text{(McKean-Vlasov)}$$

$$\Gamma(t)(x,y) = \mathbb{E}[X(t) \mid (U,V) = (x,y)], \quad \forall\, (x,y) \in [0,1]^2.$$

Existence + uniqueness when DR is L^2 Lipschitz - [HOPST '22]

McKean-Vlasov equation

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a Brownian Motion $B(t)$, and $(U, V) \overset{\text{i.i.d.}}{\sim} \text{Uni}[0,1]$.
- Consider the process $(X(t), \Gamma(t))$ such that on $\{U = u, V = v\}$,

$$\mathrm{d}X(t) = -(DR)(\Gamma(t))(u,v)\,\mathrm{d}t + \mathrm{d}B(t) \quad \underbrace{+\,\mathrm{d}L^{-}(t) - \mathrm{d}L^{+}(t)}_{\text{constrain in } [-1,1]}, \qquad \text{(McKean-Vlasov)}$$

$$\Gamma(t)(x,y) = \mathbb{E}[X(t) \mid (U,V) = (x,y)], \quad \forall\, (x,y) \in [0,1]^2.$$

Expected to arise as limit of large number of graph dynamics:

- "Mean-field interaction": For any edge-weight, the effect of all others edge-weights on its evolution is invariant under vertex relabeling.
- "Propagation of chaos": Every edge-weight between a set of m randomly chosen vertices evolves independently in the limit.

Existence + uniqueness when DR is L^2 Lipschitz - [HOPST '22]

Proof Sketch:Scaling limits of gradient flow

- We show that the cut topology is *consistent* with the invariant L^2 metric δ_2[3].
- At every $n \in \mathbb{N}$, consider *implicit Euler update* rule with positive a step size τ_n.
- The limit is obtained by showing Γ-convergence.

[3] Gradient Flows: In Metric Spaces and in the Space of Probability Measures - Ambrosio, Gigli, Savaré '08

Proof Sketch: Scaling limits of noisy SGD

- The existence of the deterministic limit Γ is obtained as a limit of a Picard iterations.
- Independently sample a sequence of vertices.
  - From SGD iterations $W^{(n)}(t)$, sample a random $m \times m$ submatrix process $W^{(n)}(t)[m]$.
  - Couple and get matrix processes $X(t)[m]$ & $\Gamma(t)[m]$ from McKean-Vlasov type SDEs.
- Use concentration estimates to show that as curves,

$$W^{(n)}[m] \overset{\delta_\square}{\rightrightarrows} \Gamma, \qquad \text{as} \quad n \to \infty, \text{ and } m \to \infty, \qquad \text{a.s.}$$

We recover the scaling limit of SGD (without added noise) as a corollary.

Future directions-I

Cut convergence gives limited information

- What can we infer if $W_n \to W$ in cut topology?
- We can infer the convergence of $t(F, W_n) \to t(F, W)$ for any finite graphs.
- Unfortunately, we can't say $\iint W_n(x,y)^2 \,\mathrm{d}x\,\mathrm{d}y \to \iint W(x,y)^2 \,\mathrm{d}x\,\mathrm{d}y$.

Future directions-I

Cut convergence gives limited information

- What can we infer if $W_n \to W$ in cut topology?
- We can infer the convergence of $t(F, W_n) \to t(F, W)$ for any finite graphs.
- Unfortunately, we can't say $\iint W_n(x,y)^2 \,\mathrm{d}x\,\mathrm{d}y \to \iint W(x,y)^2 \,\mathrm{d}x\,\mathrm{d}y$.

Cut topology is not good for weighted graphs

- Let $G(n,p)$ be the Erdös-Rènyi graph.
- $G(n,p) \to W_p \equiv p$.
- Let $K(n,p)$ be the complete weighted graph withe edge weights p.
- $K(n,p) \to W_p$.
- We would want to say $G(n,p)$ converges to an infinite exchangeable array $G(\infty,p)$ with i.i.d. Bernoulli random variables.
- And, $K(n,p)$ converges to an infinite (deterministic) array $K(\infty,p)$.

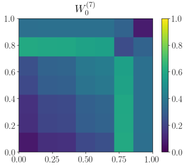
- Stronger but natural topology? Measure-valued graphons? In progress.

Simulations

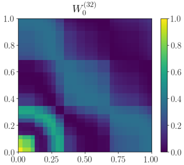
- Turán's theorem: The n-vertex triangle-free graph with the maximum number of edges is a complete bipartite graph.

Q. Can we recover this theorem through an optimization problem on graphons?

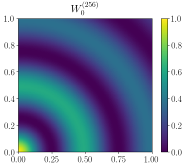
$$F(W) = t(K_3, W) - \frac{1}{10} t(K_2, W) \, .$$



(a) GD ($n = 7$)



(b) GD ($n = 32$)



(c) GD ($n = 256$)

Thank you!

Thank you!

ArXiv version[4]: https://arxiv.org/abs/2210.00422

[4] Stochastic optimization on matrices and a graphon McKean-Vlasov limit - Harchaoui, Oh, Pal, Somani, Tripathi, 2022