arXiv:2507.20088v1 [cs.LG] 27 Jul 2025

Feed-anywhere ANN (I) Steady Discrete $\to$ Diffusing on Graph Hidden States

Dmitry Pasechnyuk-Vilensky[1*], **Daniil Doroshenko**[2*]
[1]Mohamed bin Zayed University of Artificial Intelligence
Abu Dhabi, UAE
[2]Moscow State University
Moscow, Russia
dmivilensky1@gmail.com
daniildorosenko@gmail.com

Abstract

We propose a novel framework for learning hidden graph structures from data using geometric analysis and nonlinear dynamics. Our approach: (1) Defines discrete Sobolev spaces on graphs for scalar/vector fields, establishing key functional properties; (2) Introduces gauge-equivalent nonlinear Schrödinger and Landau–Lifshitz dynamics with provable stable stationary solutions smoothly dependent on input data and graph weights; (3) Develops a stochastic gradient algorithm over graph moduli spaces with sparsity regularization. Theoretically, we guarantee: topological correctness (homology recovery), metric convergence (Gromov–Hausdorff), and efficient search space utilization. Our dynamics-based model achieves stronger generalization bounds than standard neural networks, with complexity dependent on the data manifold's topology.

Code — https://aaai.org/example/code

Introduction

Learning latent graph structures from the data is a fundamental challenge in artificial intelligence, with critical applications in physical and social modeling. Although graph neural networks have shown remarkable success, they inherently assume known graph topologies, which is a significant limitation when confronting real-world data with unobserved relational structures. Existing structure-learning methods lack rigorous theoretical guarantees for topological correctness and metric convergence, failing to provably recover a data manifold's homology groups or geodesic distances. This theoretical gap becomes particularly problematic in scientific applications where interpretable latent structures are essential.

The basis of our methodology was the following reasoning. Based on the hypothesis that the data represented by pairs (X, y) approximately form a manifold, the question of how to encode the structure of this manifold arises. If such a method of encoding were found, a suitable structure could be found by approximate numerical methods. The simplest structures are topology and metric, both of which are easily represented as a weighted graph. At the same time, weighted graphs form a topological space, which in some areas is a manifold, which means that gradient descent can be organized on it, and where the gradient is not defined, other, nonlocal, oracles can be used. We assume that the data structure is restored and that the objects in this structure are assigned initial values. It is necessary to obtain a description from these values that would help characterize the class to which the corresponding point belongs. Of course, y is the class, but we understand the class more broadly, so that when the target variable is replaced by another reasonable one, the description remains useful. We assume, in a simplified way, that the class is related to localization on a manifold. Thus, the problem of composing a description is the problem of finding a mapping from functions on a graph to functions on a graph whose values are concentrated in suitable regions. It is reasonable to define such mappings based on elementary operators, differential and integral, which for a graph reduces to the Laplace operator and convolution. It is reasonable to consider idempotent mappings. In this paper, we have decided to restrict ourselves to mappings defined via the Laplace operator, that is, ultimately, differential equations, if we understand them as mappings from initial values to a stationary stable solution such as $t \to +\infty$.

Frameworks

It is necessary to separately comment on the approaches we used to describe those very mappings defined by the differential equation, since they are connected with mathematical physics and are intended to provide a basis for finding connections between the proposed model and physical problems in further work.

We extend the standard Sobolev theory (Adams and Fournier 2003) to graphs via discrete Laplacian operators (Chung 1997). After describing $L^2(V), H^1(V)$, and $H^2(V)$, we construct their counterparts $H^1(V; \mathbb{R}^3)$ for spin fields, establishing norm equivalences $\|\cdot\|_{L^2} \sim \|\cdot\|_{H^1}$ and compact embeddings (Heinonen et al. 2015). This provides a functional-analytical foundation for considering the differential dynamics of scalar and vector fields.

Our stereographic projection $\psi_j \leftrightarrow \vec{S}_j$ extends $SU(2)$ formalism from the non-Abelian gauge theory framework (Yang and Mills 1954) to graph domains, where $\vec{S}_j$ trans-

*These authors contributed equally.

forms as an isospin doublet under $G_j \in SU(2)$. The physical role of the gauge equivalence between nonlinear Schrödinger and Landau–Lifshitz equations is described in (Gomes 2025) and is interpreted as "descriptive redundancy". In addition, it clarifies how $\mathcal{M}_{\text{LL}} \simeq \mathbb{S}^{2n-1}$ and $\mathcal{M}_{\text{Sch}}$ represent identical physical states despite mathematical differences.

Contribution

There exist approaches to physics-aware neural networks (Chuprov et al. 2023), networks that use not forwardly connected hidden spaces (Hopfield 1982), and networks that operate on graphs (Veličković 2023). One can notice the same physical motives related to diffusion or spin physical systems as motivations or as applications of those models. But none of these approaches allows one to recover the hidden topology of the data and model the evolution of the hidden state on the recovered topology simultaneously. Our framework unifies these strands through dynamics-driven geometry inference, delivering:

- Topological correctness. Guaranteed recovery of Betti numbers (β_0, β_1) for latent graphs, not achieved by GNNs.
- Geometric convergence. Bounded Gromov-Hausdorff distance $d_{\text{GH}} \leq \mathcal{O}(\delta + t^{-1})$ to ground truth manifolds.
- Efficient stratified optimization. $\mathcal{O}(2^{|E_{\text{true}}|})$ search complexity via gradient-based search of edge set.
- Physics-informed interpretability. Gauge equivalent NLSE and spin dynamics (Melcher and Ptashnyk 2013) provides mechanistic insight into the learned structures.

This establishes a new paradigm for data-driven geometric discovery with theoretical guarantees unmatched by existing methods.

Structure

The first two of the following sections develop the theoretical framework that is behind the FANN model.

In section "Evolution of hidden state", we study the nonlinear Schrödinger and Landau–Lifshitz equations on the weighted graphs, demonstrate their gauge-equivalence, prove that they admit stationary solution which is smoothly dependent on weights of graph and initial values, in the Sobolev spaces on graph which are formerly introduced. The non-complex counterpart of NSE and the gauge-equivalent LL equation are presented in the appendices.

Section "Moduli space of graph hidden spaces" studies the problem of optimizing some function w.r.t. the weighted graph, which is the hidden space. We introduce the moduli space of weighted graphs and suggest a gradient-based method which is identical to gradient descent on the interior of moduli space strata (when all weights are separated from zero) and turns into the addition / deletion procedure on the boundaries of strata. Guarantees of convergence in this space are provided.

Finally, Section "Statistical learning" introduces the FANN model, or rather, its special case, and studies its complexity and efficiency using statistical learning tools. The results are compared with those of a standard neural network.

Evolution of hidden state

Sobolev space

Consider an undirected graph $G = (V, E)$ with vertices $V = \{1, \dots, n\}$, edges $E \subseteq V \times V$ (such that $(u, v) \in E \Rightarrow (v, u) \in E$), vertex measure $\mu : V \to \mathbb{R}_+$ (by default, we assume $\mu(v) = 1$, another possible option is $\mu(v) = \deg(v)$), and edge measure $\rho : E \to \mathbb{R}_+$ (by default, we assume $\rho(e) = 1$).

$L^2(V)$ consists of functions $f : V \to \mathbb{C}$ with finite norm:

$$\|f\|^2_{L^2(V)} = \sum_{v \in V} |f(v)|^2 \mu(v).$$

$L^2(E)$ consists of edge functions: $\omega : E \to \mathbb{C}$ such that $\omega(u, v) = -\omega(v, u)$ with norm

$$\|\omega\|^2_{L^2(E)} = \frac{1}{2} \sum_{e \in E} |\omega(e)|^2 \rho(e)$$

(Bondy, Murty et al. 1976).

***Definition* 1.** *Discrete gradient $d : L^2(V) \to L^2(E)$:*

$$(df)(e) = f(u) - f(v) \quad \textit{for oriented edge } e = (v, u).$$

Laplacian $\Delta : L^2(V) \to L^2(V)$:

$$(\Delta f)(v) = \frac{1}{\mu(v)} \sum_{e=(v,u)} \big(f(u) - f(v)\big) \rho(e)$$

(Chung 1997). $\Delta = d^* d + dd^*$, where $d^* : L^2(E) \to L^2(V)$ is the adjoint of d (Diestel 2025).

***Definition* 2.** *(Shao, Yang, and Zhao 2025) Sobolev space*

$$H^1(V) = \{f \in L^2(V) : df \in L^2(E)\}$$

consists of all functions $f \in L^2(V)$ for which the discrete gradient df belongs to $L^2(E)$. $H^1(V)$ is inner-product space with

$$\langle f, g \rangle_{H^1(V)} = \sum_{v \in V} (f\overline{g})(v)\mu(v) + \sum_{e \in E} (df\overline{dg})(e)\rho(e)$$

and associated norm

$$\|f\|^2_{H^1(V)} = \|f\|^2_{L^2(V)} + \|\nabla f\|^2_{L^2(E)},$$

where

$$\begin{aligned} \|\nabla f\|^2_{L^2(E)} &= -\langle \Delta f, f \rangle_{L^2(V)} \\ &= \sum_{e=(v,u) \in E} |f(u) - f(v)|^2 \rho(e). \end{aligned}$$

In the norm $\| \cdot \|_{H^1(V)}$ defined above, first summand controls the "magnitude" of the function, and second summand controls the "smoothness" of the function — the sum of the squares of its changes along all edges. The smaller the changes, the "smoother" the function on the graph.

***Definition* 3.** *The space*

$$H^2(V) = \{f \in H^1(V) : \Delta f \in L^2(V)\}$$

consists of all functions $f \in H^1(V)$, the Laplacian Δf of which lies in $L^2(V)$. $H^2(V)$ is normed space with

$$\|f\|^2_{H^2(V)} = \|f\|^2_{L^2(V)} + \|\Delta f\|^2_{L^2(V)}.$$

We can define the operator d^2 that acts on the edge functions and require $d^2(df) \in L^2(F)$

Let us now consider the vector field of spins $\vec{S} = (\vec{S}_1, \dots, \vec{S}_n)$, where $\vec{S}_j \in \mathbb{R}^3$ and $\|\vec{S}_j\| = 1$. We are goind to define the Sobolev spaces for this vector field on the graph.

$L^2(V;\mathbb{R}^3)$ consists of functions $\vec{S} : V \to \mathbb{R}^3$ with finite norm:

$$\|\vec{S}\|^2_{L^2(V;\mathbb{R}^3)} = \sum_{v\in V} \|\vec{S}(v)\|^2_{\mathbb{R}^3}\mu(v),$$

where

$$\|\vec{S}(v)\|^2_{\mathbb{R}^3} = (S^x(v))^2 + (S^y(v))^2 + (S^z(v))^2.$$

***Definition* 4.** *Sobolev space*

$$H^1(V;\mathbb{R}^3) = \left\{ \vec{S} \in L^2(V;\mathbb{R}^3) \mid d\vec{S} \in L^2(E;\mathbb{R}^3) \right\}$$

consists of all functions $f \in L^2(V;\mathbb{R}^3)$ for which the discrete gradient $d\vec{S}$ belongs to $L^2(E;\mathbb{R}^3)$. The norm is

$$\begin{aligned}\|\vec{S}\|^2_{H^1(V;\mathbb{R}^3)} = &\sum_{v\in V} \|\vec{S}(v)\|^2\mu(v) + \\ &+ \frac{1}{2}\sum_{e=(u,v)\in E} \|\vec{S}(u) - \vec{S}(v)\|^2\rho(e).\end{aligned}$$

In the norm defined above, the $\sum \|\vec{S}(v)\|^2\mu(v)$ term controls the amplitude of the field, while $\sum \|\vec{S}(u)-\vec{S}(v)\|^2\rho(e)$ term controls the smoothness along the edges.

***Definition* 5.** *The space*

$$H^2(V;\mathbb{R}^3) = \left\{ \vec{S} \in H^1(V;\mathbb{R}^3) \mid \Delta\vec{S} \in L^2(V;\mathbb{R}^3) \right\}$$

consists of all functions $\vec{S} \in H^2(V;\mathbb{R}^3)$, the Laplacian $\Delta\vec{S}$ of which lies in $L^2(V;\mathbb{R}^3)$. $H^2(V;\mathbb{R}^3)$ is normed space with

$$\|\vec{S}\|^2_{H^2(V;\mathbb{R}^3)} = \|\vec{S}\|^2_{H^1} + \sum_{v\in V} \|\Delta\vec{S}(v)\|^2\mu(v).$$

***Theorem* 1** (Equivalence of norms and compact inclusions)**.** *(Heinonen et al. 2015)*

1. $\|\cdot\|_{L^2(V,\mathbb{R}^3)} \sim \|\cdot\|_{H^1(V,\mathbb{R}^3)} \sim \|\cdot\|_{H^2(V,\mathbb{R}^3)}$*, that is*

$$\begin{gathered}\exists\, C_1, C_2, C_3, C_4 > 0 \quad \forall \vec{S} \in \mathbb{R}^{3|V|} : \\ C_1\|\vec{S}\|_{H^2} \le \|\vec{S}\|_{H^1} \le C_2\|\vec{S}\|_{L^2}, \\ C_3\|\vec{S}\|_{H^1} \le \|\vec{S}\|_{H^2} \le C_4\|\vec{S}\|_{H^1}.\end{gathered}$$

2. $H^2(V;\mathbb{R}^3) \overset{c}{\hookrightarrow} H^1(V;\mathbb{R}^3) \overset{c}{\hookrightarrow} L^2(V;\mathbb{R}^3)$*, where* $\overset{c}{\hookrightarrow}$ *denotes compact inclusion.*

Nonlinear Schrödinger equation

Consider the following differential equation on graph $G = (V,E)$, $|V| = N$:

$$\frac{d\psi}{dt} = F(\psi, w), \quad \psi(0) = \psi^0,$$

where $\psi^0, \psi(t) \in H^1(V)$ are wave functions, $w : E \to \mathbb{R}_+$ is an edge weight function.

$$\begin{aligned}F(\psi, w) = &- i(\Delta(w) + \operatorname{diag}(|\psi(0)|^2))\psi \\ &- \gamma P^{\perp}_{\psi} D(\psi, w),\end{aligned} \tag{1}$$

where

$$P^{\perp}_{\psi} = I - \frac{\psi\psi^{\dagger}}{\|\psi\|^2_{\mathbb{C}^V}},\ D(\psi,\theta) = \Delta\psi + (|\psi|^2 - |\psi(0)|^2)\psi,$$

and $\Delta(w)$ is the discrete Laplace operator (Definition 1), which is a linear function of w, in particular $(\Delta\psi)_i = \sum_{e=(i,j)\in E} w(e)(\psi_i - \psi_j)$.

The term $\operatorname{diag}(|\psi|^2)\psi$ in F introduces the nonlinear potential, which is equal to $|\psi_i|^2\psi_i$ in $i \in V$. In the dissipation term, the projector $P^{\perp}_{\psi}$ eliminates the component parallel to ψ, preserving the norm of the solution, i.e. $\|\psi\|^2_{\mathbb{C}^V} = const$.

***Theorem* 2** (Existence of $\psi(+\infty)$ and C^∞-smoothness as a function of w)**.** *Assume that for the weight function $w_0 \in \mathbb{R}^E_+$ there exists an isolated asymptotically stable stationary solution $\psi(+\infty, w_0) \in H^2(V)$ of the non-linear Schrödinger equation* (1)*. Then:*

1. $\lim_{t\to+\infty} \psi(t) = \psi(+\infty, w)$ *exists for w from the vicinity of w_0,*

2. $(w \mapsto \psi(+\infty, w)) \in C^\infty(\mathbb{R}^E_+, H^2(V))$.

***Theorem* 3** (C^∞-smoothness of $\psi(+\infty)$ as a function of $\psi(0)$)**.** *Given the $\psi(0) = \psi^0 \in H^1(V)$, let $\psi(+\infty, \psi^0) \in H^2(V)$ be the isolated asymptotically stable stationary solution of the non-linear Schrödinger equation* (1)*. Then $\exists\delta > 0$:*

1. $\forall\psi' \in B_\delta(\psi(+\infty, \psi^0)) \subset H^1(V)$, $\lim_{t\to+\infty}\psi(t) = \psi(+\infty, \psi')$,

2. $(\psi^0 \mapsto \psi(+\infty, \psi^0)) \in C^\infty(H^1(V), H^2(V))$.

The phase space of the nonlinear Schrödinger equation under consideration is a unit sphere in $\mathbb{C}^n$:

$$\mathcal{M}_{\text{Sch}} = \mathbb{S}^{2n-1} = \{\psi \in \mathbb{C}^n : \|\psi\| = 1\}.$$

Landau–Lifshitz equation

Consider the stereographic projection, which maps complex plane $\mathbb{C}$ which is the range of ψ onto the unit Bloch sphere:

$$\vec{S}_j = \begin{pmatrix} \dfrac{\psi_j + \bar{\psi}_j}{1 + |\psi_j|^2} & \dfrac{i(\bar{\psi}_j - \psi_j)}{1 + |\psi_j|^2} & \dfrac{1 - |\psi_j|^2}{1 + |\psi_j|^2}\end{pmatrix}^{\top},$$

$$\psi_j = \frac{S^x_j + iS^y_j}{1 + S^z_j}.$$

This mapping is isometric, that is, preserves angles and spheres, and conformal, i.e. preserves infinitesimal forms, and has singularities at $S^z_j \to -1$ ($|\psi_j| \to \infty$). In addition, it satisfies $\|\vec{S}_j\| = 1$ if $\|\psi\| < \infty$.

There is a natural connection between this gauge transformation and the integrable structure of the system, described by the Lax pair:

$$\begin{cases} \dfrac{dL}{dt} = [B, L], \\ L = \begin{pmatrix} -i\lambda & \psi \\ -\bar{\psi} & i\lambda \end{pmatrix},\ B = \begin{pmatrix} -i|\psi|^2 & -i\frac{d\psi}{dt} \\ i\frac{d\bar{\psi}}{dt} & i|\psi|^2 \end{pmatrix}. \end{cases}$$

After the transformation, the Lax matrix takes the spin form:

$$L_{\mathrm{LL}} = \lambda \begin{pmatrix} S_j^z & S_j^x - iS_j^y \\ S_j^x + iS_j^y & -S_j^z \end{pmatrix} = G_j L_{\mathrm{Sch}} G_j^{-1},$$

where

$$G_j = \frac{1}{\sqrt{1+|\psi_j|^2}} \begin{pmatrix} 1 & \bar{\psi}_j \\ -\psi_j & 1 \end{pmatrix}, \quad G_j \in SU(2),$$

so that the commutation relation is preserved:

$$\frac{dL_{\mathrm{LL}}}{dt} = [B_{\mathrm{LL}}, L_{\mathrm{LL}}],$$

where B_{LL} corresponds to the precession in the effective field.

Now, let us express the derivative of $\vec{S}$ through ψ:

$$\frac{d\vec{S}_j}{dt} = \frac{\partial \vec{S}_j}{\partial \psi_j}\frac{d\psi_j}{dt} + \frac{\partial \vec{S}_j}{\partial \bar{\psi}_j}\frac{d\bar{\psi}_j}{dt}$$

and substitute in to the former equation on ψ_j:

$$\frac{d\psi_j}{dt} = -i\sum_{k\in V} L_{jk}\psi_k - i|\psi_j(0)|^2\psi_j - \gamma P_j^{\perp} D_j.$$

After some algebra, using the properties of projector, one has:

$$\begin{aligned} \frac{d\vec{S}_j}{dt} =& \vec{S}_j \times \left(-2\sum_k w_{jk}\vec{S}_k + 2|\psi_j(0)|^2 e_3\right) - \\ & - \gamma \vec{S}_j \times \left(\vec{S}_j \times \vec{\mathcal{D}}_j\right),\ \vec{S}_j(0) = \vec{S}_j^0 \end{aligned} \tag{2}$$

where

$$\begin{aligned} \vec{\mathcal{D}}_j =& -2\sum_k w_{jk}(\vec{S}_k - \vec{S}_j) + \\ & + 2\left(|\psi_j(0)|^2 - \frac{1}{|V|}\sum_i |\psi_i(0)|^2\right) e_3 \end{aligned}$$

is the dissipation field, and $e_3 = (0\ 0\ 1)^{\top}$.

The phase space of the Landau–Lifshitz equation is a submanifold of $(\mathbb{S}^2)^n$:

$$\begin{aligned} \mathcal{M}_{\mathrm{LL}} = \big\{ & (\vec{S}_1, \dots, \vec{S}_n) \in (\mathbb{S}^2)^n : \\ & \sum_{j=1}^{n} \frac{1 - S_j^z}{1 + S_j^z} = 1,\ S_j^z > -1 \big\}, \end{aligned} \tag{3}$$

which is diffeomorphic to $\mathbb{S}^{2n-1}$ (diffeomorphism is given by the gauge transformation considered in the previous subsection).

Moduli space of graph hidden spaces

Given the number of vertices N and $V = \{1, ..., N\}$, the universal moduli space $\mathcal{M} = \{(E, (w(e))_{e\in E}) : E \subset V \times V, w : E \to \mathbb{R}\}$ of the hidden spaces contains all configurations of graphs $G = (V, E, w)$ with weighted edges and vertices V. $\mathcal{M}$ is stratified space: each set of edges $E \subset V \times V$ defines the strata $\mathcal{M}(E) \simeq \mathbb{R}_{>0}^{|E|}$.

Algorithm 1: Gradient descent on the moduli space of graph hidden spaces

Parameter: $T \in \mathbb{N}, p \in (0, 1], 0 < \theta, \Theta \ll 1, (\eta_t > 0)_{t=0}^T$
E_0 is such that $e \in E$ with probability p, $w_0(E_0) = \{1\}$, batch size $B \in \mathbb{N}$
Output: E_T, w_T

Let $t = 0$.
while $t < T$ **do**
  Sample $D \sim \mathcal{D}$, $|D| = B$
  for all $e \in E_t$ **do**
    $w_{t+1}(e) \leftarrow w_t(e) - \eta_t g(E_t, w_t; D)$
  end for
  $E' \leftarrow E$
  for all $e \in N \times N \setminus E_t$ **do**
    if $g(E_t \cup \{e\}, w_t + (e, \theta); D) < -\Theta$ **then**
      $E' \leftarrow E' \cup \{e\}$
      $w_{t+1}(e) \leftarrow 2\theta$
    end if
  end for
  for all $e \in E_t : w(e) < \theta$ **do**
    $E' \leftarrow E' \setminus \{e\}$
  end for
  $E_{t+1} \leftarrow E'$
  $t \leftarrow t + 1.$
end while
return E_T, w_T

Consider the function $\mathcal{L}_{\text{out}} \in \mathcal{C}^{1,1}(H^2(V), \mathbb{R})$. Let $\psi_\infty \in H^2(V)$ be a stable solution of Schrödinger equation on $G = (V, E, w)$ with initial condition $\psi_0 \in H^1(V)$. ψ_∞ continuously depends on $\psi_0 \in H^1(V)$ and $(w(e))_{e\in E}$, therefore $\mathcal{L}$, that is, for fixed ψ_0, $\psi_\infty \in \mathcal{C}^{1,1}(\mathcal{M}(E), H^2(V))$. Therefore, $\mathcal{L} = \mathcal{L}_{\text{out}} \circ \psi_\infty \in \mathcal{C}^{1,1}(\mathcal{M}(E), \mathbb{R})$. $\frac{\partial \mathcal{L}}{\partial w(e)}$ is well-defined at $(\psi_0, (E, w))$ if $e \in E$ and $w(e) > 0$. Assume additionally that $\mathcal{L}$ is strongly convex w.r.t. $w(e)$ if $e \in E$.

The algorithm 1 suggests a gradient method to minimize $\mathcal{L}$ w.r.t. (E, w).

Given the data distribution $(X, y) \sim \mathcal{D}$, define

$$\begin{aligned} \mathcal{L}(E, w) = & \mathbb{E}_{(X,y)\sim\mathcal{D}}[\mathcal{L}_{\text{sample}}(X, y, (E, w))] + \\ & + \frac{\mu_2}{2}\|w(E)\|_2^2 + \mu_1\|w(E)\|_1, \end{aligned} \tag{4}$$

where $\mathcal{L}_{\text{sample}}(\psi_0, y, (E, w)) = (k(\psi_\infty(\psi_0, (E, w))) - y)^2$, $\|\cdot\|_2^2$ regularization guarantees strong convexity w.r.t. $w(e)$ if $e \in E$, and $\|\cdot\|_1$ enforces the sparsity of the graph, and $k \in \mathcal{C}^{1,1}(H^2(V), \mathbb{R})$. The stochastic gradient of $\mathcal{L}$ w.r.t. w is defined by $g_e(E, w; D) = \frac{1}{|D|}\sum_{(X,y)\in D} \frac{\partial \mathcal{L}_{\text{sample}}}{\partial w(e)}(X, y, (E, w))$ for a given D. $\mathbb{E}_{D\sim\mathcal{D}}[g_e(E, w; D)] = \frac{\partial \mathcal{L}}{\partial w(e)}(E, w)$. Assume that $\mathrm{Var}_{D\sim\mathcal{D}}[g_e(w; D)] \le \frac{\sigma^2}{|D|}$.

Assume that $\mathcal{G}$ is a compact connected smooth Riemannian manifold with injectivity radius $\rho > 0$, sectional curvature $\le \kappa$, and $\operatorname{diam}\mathcal{G} < +\infty$. Let $0 < \delta \ll 1$, and for every $(X, y) \sim \mathcal{D}$ there exists $\|\delta_{X,y}\| \le \delta$, such that $(X, y) + \delta_{X,y} \in \mathcal{G}$.

The geodesic triangulation edge set:

$$E_{\text{true}} = \{(u,v) \in V \times V : 0 < d_{\mathcal{G}}(u,v) < \rho\}.$$

Let us assume that $\delta < \frac{\mu_1+\mu_2\theta-\Theta}{C}$ for $C > 0$, and that batch size B satisfies $\text{Var}(g_e) \leq \sigma^2/B$. Then, one has that with probability $\geq 1-\epsilon$, the edge set becomes $E_t = E_{\text{true}}$ for all $t \geq T_0$, where $T_0 = O\left(\frac{\log(|E_{\text{true}}|/\epsilon)}{\log\log(N^2/\epsilon)}\right)$, and the topological and metric structures of the graph hidden space converge to those of the data manifold in the following sense:

***Theorem* 4** (Homology convergence)**.** *(Hatcher 2002) For $t \geq T_0$, $\beta_k(G_t) = \beta_k(\mathcal{G})$ for $k = 0, 1$.*

***Theorem* 5** (Gromov-Hausdorff convergence)**.** *For $t \geq T_0$:*

$$d_{GH}\left((V, d_{G_t}), \mathcal{G}\right) \leq C_1\delta + C_2 t^{-1},$$

(Gromov 1981) where $d_{G_t}(u,v) = \inf_{p:u\to v} \sum_{e\in p} \frac{1}{w_t(e)}$.

Statistical learning

The effective moduli space $\mathcal{M}_{\text{eff}}$ consists of all strata $\mathcal{M}(E) \simeq \mathbb{R}_{>0}^{|E|}$ that are visited with positive probability during the execution of the algorithm. The number of strata depends on the probability of visitation, which differs significantly between the true edges ($e \in E_{\text{true}}$) and the spurious edges ($e \notin E_{\text{true}}$). The asymptotic estimate for the expected number of strata is provided below.

***Theorem* 6** (Expected size of effective moduli space)**.**

$$\mathbb{E}_{D\sim\mathcal{D}}\left[|\mathcal{M}_{\textit{eff}}|\right] = \mathcal{O}\left(2^{|E_{\textit{true}}|}\left(1 + \frac{N^2|E_{\textit{true}}|\sigma^2}{B(\Delta')^2}\right)\right).$$

The dominant term in the estimate above is $2^{|E_{\text{true}}|}$, which is the cardinality of the power set of true edges: it reflects that the deterministic algorithm (where $g(E, w; D)$ is replaced with $\frac{\partial\mathcal{L}}{\partial w(e)}(E, w)$) visits states corresponding to all subsets of E_{true}. The spurious edge term $\frac{2^{|E_{\text{true}}|}N^2|E_{\text{true}}|\sigma^2}{B(\Delta')^2}$ is proportional to the number of possible edges N^2 and variance σ^2/B, so it decreases with larger batch size B or stronger curvature (via Δ').

Let us consider the particular machine learning model built on the frameworks of the evolution of hidden state and of the moduli space of graph hidden spaces. Given the $(X, y) \sim \mathcal{D}$ ($X \in \mathbb{C}^n$ or, which is a particular case, $\mathbb{R}^n$, $y \in \mathbb{C}$ or $\mathbb{R}$ correspondingly), define the first fully connected layer as follows: $L^1_{A_1,b_1} : X \mapsto \psi^0 = \sigma_1(A_1X + b_1)$, where $A_1 \in \mathbb{C}^{n\times|V|}$, $b_1 = \mathbb{C}^{|V|}$, $\sigma_1 : \mathbb{C} \to \mathbb{C}$ is holomorphic (except for enumerable singularities, e.g. $\sigma_1 = \tanh$) and is applied to vectors in a component-wise manner. The second layer is defined as $L^2_{E,w} : \psi^0 \mapsto \psi(+\infty; w, E, \psi^0)$, where L^2 is C^∞-smooth as a function of ψ^0 and w. In addition, $(E, w) \in \mathcal{M}_{\text{eff}} \subset \mathcal{M}$, and $G = (V, E, w)$ converge to the Delaunay triangulation of the true data manifold w.r.t. Gromov-Hausdoff metric and the β_0, β_1 Betti numbers after the finite number of iterations. The third layer is $L^3_{a_3,b_3} : \psi(+\infty) \mapsto \sigma_3(\langle a_3, \psi(+\infty)\rangle + b_3)$, where $a_3 \in \mathbb{C}^{|V|}, b_3 \in \mathbb{C}$, σ_3 satisfies the same requirements as σ_1. The model as a whole is

$$f : X \mapsto y = (L_3 \circ L_2 \circ L_1)(X). \tag{5}$$

Assume that σ_1, σ_3 are $L_{\sigma_1}, L_{\sigma_3}$-Lipschitz correspondingly, and that $\|A_1\|_F \leq R_A$, $\|b_1\| \leq R_b$, $\|a_3\| \leq R_a$, $|b_3| \leq R_{b_3}$, and the weights are bounded $w \in [0, R_w]^{|E_{\text{true}}|}$.

***Theorem* 7** (Generalization Bound)**.** *(Shalev-Shwartz and Ben-David 2014) With probability $\geq 1 - \delta - \epsilon$ over S and algorithm randomness:*

$$\mathbb{E}_{(X,y)\sim\mathcal{D}}[\ell(f(X), y)] \leq \frac{1}{|D|}\sum_{i=1}^{|D|}\ell(f(X_i), y_i) + 3\sqrt{\frac{\log(2/\delta)}{2|D|}} + 2L_\ell K_1\sqrt{\frac{\beta_1(\mathcal{G}) + n|V|}{|D|}} + 2L_\ell K_2\frac{\beta_1(\mathcal{G})\log|D|}{|D|},$$

where $L_\ell = 2(C_f + C_y)$, and ϵ is the failure probability of graph convergence.

Let us now compare our model with its simpler counterpart, in particular, the fully connected 3-layered neural network. The first and third layers are identical to $L^1_{A_1,b_1}$ and $L^3_{a_3,b_3}$, while the second layer is $L^2_{W_2,b_2} : \psi_0 \mapsto \sigma_2(W_2\psi_0 + b_2)$, where $W_2 \in \mathbb{C}^{H\times H}, b_2 \in \mathbb{C}^H$, $\|W_2\|_F \leq R_{W_2}, \|b_2\|_2 \leq R_{b_2}$. Standard computations show that the generalization bound for this model is

$$\mathbb{E}_{(X,y)\sim\mathcal{D}}[\ell(f(X), y)] \leq \frac{1}{|D|}\sum_{i=1}^{|D|}\ell(f(X_i), y_i) + 3\sqrt{\frac{\log(1/\delta)}{2|D|}} + 2L'_\ell K\sqrt{\frac{H(n + H + 1)}{|D|}}$$

with probability $\geq 1 - \delta$, where L'_ℓ depends on the model constants.

Conclusion

We presented a rigorous framework for learning latent graph structures with certified geometric and topological guarantees. Our contributions are formally established as follows:

1. Topological correctness. For data on Riemannian manifolds, Algorithm 1 recovers graphs matching the ground-truth Betti numbers (β_0, β_1) with logarithmic iteration complexity.
2. Metric convergence. Reconstructed graphs converge to the data manifold in Gromov–Hausdorff distance, with error bounded by data noise and iteration count.
3. Efficient Optimization. Our stratified approach achieves search complexity scaling with the true edge set size, outperforming full edge-space exploration.
4. Generalization guarantees. The generalization bound depends on the manifold topology rather than the ambient dimension.

Another contribution is the novel physical interpretation of the hidden state. Gauge-equivalent Schrödinger / Landau–Lipfshitz equations on the graph describe the evolution of the scalar / vector fields, which allows one to interpret activation of neurons as the concentration of probability / spin alignment due to the external field.

Thus, our work establishes the first graph learning framework with simultaneous topological and geometric certification, as well as the fruitful framework to describe the hidden state transformation using the equations of mathematical physics.

References

Absil, P.-A.; Mahony, R.; and Sepulchre, R. 2010. Optimization on manifolds: Methods and applications. In *Recent Advances in Optimization and its Applications in Engineering: The 14th Belgian-French-German Conference on Optimization*, 125–144. Springer.

Adams, R. A.; and Fournier, J. J. 2003. *Sobolev spaces*, volume 140. Elsevier.

Bartlett, P. L.; and Mendelson, S. 2002. Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3(Nov): 463–482.

Boissonnat, J.-D.; Dyer, R.; and Ghosh, A. 2018. Delaunay triangulation of manifolds. *Foundations of Computational Mathematics*, 18(2): 399–431.

Bondy, J. A.; Murty, U. S. R.; et al. 1976. *Graph theory with applications*, volume 290. Macmillan London.

Bottou, L.; Curtis, F. E.; and Nocedal, J. 2018. Optimization methods for large-scale machine learning. *SIAM review*, 60(2): 223–311.

Boucheron, S.; Lugosi, G.; and Massart, P. 2003. Concentration inequalities using the entropy method. *The Annals of Probability*, 31(3): 1583–1614.

Chung, F. R. 1997. *Spectral graph theory*, volume 92. American Mathematical Soc.

Chuprov, I. A.; Gao, J.; Efremenko, D. S.; Kazakov, E. A.; Buzaev, F. A.; and Zemlyakov, V. 2023. Optimization of physics-informed neural networks for nonlinear Schrödinger equation. *Doklady Rossijskoj Akademii Nauk. Mathematika, Informatika, Processy Upravlenia*, 514(2): 28–38.

Combes, J.-M.; and Thomas, L. 1973. Asymptotic behaviour of eigenfunctions for multiparticle Schrödinger operators. *Communications in Mathematical Physics*.

Deimling, K. 2013. *Nonlinear functional analysis*. Springer Science & Business Media.

Diestel, R. 2025. *Graph theory*, volume 173. Springer Nature.

Edelsbrunner, H.; and Harer, J. 2010. *Computational topology: an introduction*. American Mathematical Soc.

Gomes, H. D. A. 2025. *Gauge Theory and the Geometrisation of Physics*. Cambridge University Press.

Gromov, M. 1981. Structures métriques pour les variétés riemanniennes. *Textes Math.*, 1.

Hatcher, A. 2002. *Algebraic Topology*. Cambridge University Press.

Heinonen, J.; Koskela, P.; Shanmugalingam, N.; and Tyson, J. T. 2015. *Sobolev spaces on metric measure spaces*. 27. cambridge university press.

Hopfield, J. J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8): 2554–2558.

Khalil, H. K.; and Grizzle, J. W. 2002. *Nonlinear systems*, volume 3. Prentice hall Upper Saddle River, NJ.

Maz'ya, V. 2013. *Sobolev spaces*. Springer.

Melcher, C.; and Ptashnyk, M. 2013. Landau-Lifshitz-Slonczewski equations: global weak and classical solutions. *SIAM Journal on Mathematical Analysis*, 45(1): 407–429. Copyright 2013 Elsevier B.V., All rights reserved.

Shalev-Shwartz, S.; and Ben-David, S. 2014. *Understanding machine learning: From theory to algorithms*. Cambridge university press.

Shao, M.; Yang, Y.; and Zhao, L. 2025. Sobolev spaces on locally finite graphs. *Proceedings of the American Mathematical Society*, 153(02): 693–708.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1): 267–288.

Vaart, A. v. d.; and Wellner, J. A. 1997. Weak convergence and empirical processes with applications to statistics. *Journal of the Royal Statistical Society-Series A Statistics in Society*, 160(3): 596–608.

Veličković, P. 2023. Everything is connected: Graph neural networks. *Current Opinion in Structural Biology*, 79: 102538.

Wainwright, M. J. 2019. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press.

Yang, C.-N.; and Mills, R. L. 1954. Conservation of isotopic spin and isotopic gauge invariance. *Physical review*, 96(1): 191.

Proofs for evolution of hidden state

Proof of the Theorem 1. Let us prove the first point about the equivalence of the norms $\|\cdot\|_{L^2}, \|\cdot\|_{H^1}, \|\cdot\|_{H^2}$. Consider the space $\mathcal{V} = \mathbb{R}^{3|V|}$. Firstly, prove the lower bound on

$$\|\vec{S}\|_{H^1}^2 = \underbrace{\sum_{v\in V} \|\vec{S}(v)\|^2 \mu(v)}_{A} + \underbrace{\frac{1}{2}\sum_{e=(u,v)} \|\vec{S}(u) - \vec{S}(v)\|^2 \rho(e)}_{B}.$$

One can see that $A \le \|\vec{S}\|_{H^1}^2$ and $B \le \|\vec{S}\|_{H^1}^2$. Since $A \ge \mu_{\min}\|\vec{S}\|^2_{\mathbb{R}^{3|V|}}$, one has

$$\|\vec{S}\|_{L^2} \le \|\vec{S}\|_{H^1}.$$

Second, we prove the upper bound in $\|\cdot\|_{H^1}$. Denote

$$\deg_{\max} = \max_{v\in V} \deg(v), \rho_{\max} = \max_{e\in E} \rho(e).$$

Then,

$$B \le \frac{1}{2}\rho_{\max} \sum_{e=(u,v)} \|\vec{S}(u) - \vec{S}(v)\|^2.$$

Using $\|\vec{S}(u) - \vec{S}(v)\|^2 \le 2(\|\vec{S}(u)\|^2 + \|\vec{S}(v)\|^2)$:

$$\sum_{e=(u,v)} \|\vec{S}(u) - \vec{S}(v)\|^2 \le 2\deg_{\max} \sum_{v\in V} \|\vec{S}(v)\|^2.$$

Then:

$$\|\vec{S}\|_{H^1}^2 \le \sum_v \|\vec{S}(v)\|^2\mu(v) + \rho_{\max}\deg_{\max}\sum_v \|\vec{S}(v)\|^2 \le \left(1 + \frac{\rho_{\max}\deg_{\max}}{\mu_{\min}}\right)\|\vec{S}\|_{L^2}^2.$$

Thus:

$$\|\vec{S}\|_{L^2} \le \|\vec{S}\|_{H^1} \le C_2\|\vec{S}\|_{L^2},\ C_2 = \sqrt{1 + \frac{\rho_{\max}\deg_{\max}}{\mu_{\min}}},$$

and $\|\cdot\|_{L^2} \sim \|\cdot\|_{H^1}$.

Let us move on to $\|\cdot\|_{H^2} \sim \|\cdot\|_{H^1}$. One can see that in finite dimensional space, the norm of Laplace operator is bounded:

$$\|\Delta\|_{\text{op}} = \sup_{\vec{S}\neq 0} \frac{\|\Delta\vec{S}\|_{L^2}}{\|\vec{S}\|_{L^2}} \le C_\Delta.$$

Therefore,

$$\|\vec{S}\|_{H^2}^2 \le \|\vec{S}\|_{H^1}^2 + C_\Delta^2\|\vec{S}\|_{L^2}^2 \le (1 + C_\Delta^2)\|\vec{S}\|_{H^1}^2.$$

Conversely,

$$\|\vec{S}\|_{H^1}^2 \le \|\vec{S}\|_{H^2}^2.$$

Thus:

$$\|\vec{S}\|_{H^1} \le \|\vec{S}\|_{H^2} \le C_4\|\vec{S}\|_{H^1},\ C_4 = \sqrt{1 + C_\Delta^2},$$

and $\|\cdot\|_{H^1} \sim \|\cdot\|_{H^2}$.

Let us now consider the second point on inclusions. In the finite-dimensional space $\mathcal{V}$ it is true that 1) the bounded set in H^2-norm is bounded in H^1-norm, due to the equivalence; 2) closed ball in $\mathcal{V}$ is compact; 3) the identity correspondence $I : H^2 \to H^1$ is continuous, due to the equivalence, and the image of the bounded set is relatively compact. Therefore,

$$H^2(V;\mathbb{R}^3) \overset{c}{\hookrightarrow} H^1(V;\mathbb{R}^3),$$

and similarly for $H^1 \hookrightarrow L^2$. □

Proof of the Theorem 2. The equation on the stationary state is

$$G(\psi, w) = -iH(w)\psi - \gamma P_\psi^\perp D(\psi, w) = 0,$$

where $H(w) = \Delta(w) + \operatorname{diag}(|\psi^0|^2)$, $D(\psi, w) = \Delta(w)\psi + \operatorname{diag}(|\psi|^2 - |\psi^0|^2)\psi$. If $w = w_0$, $G(\psi(+\infty), w_0) = 0$. G is smooth, since $\Delta(w)$ is linear w.r.t. $w \implies L \in C^\infty$, projector $P_\psi^\perp$ is analytical if $\psi \neq 0$, the nonlinear term $|\psi|^2\psi$ is polynomial. Thus,

$$G \in C^\infty(H^2(V) \times \mathbb{R}_+^E, L^2(V)).$$

Consider the Fréchet derivative

$$J_\psi = D_\psi G(\psi(+\infty), w_0) : H^2(V) \to L^2(V).$$

Having the stability condition $\operatorname{Re}(\sigma(J_\psi)) < -\alpha < 0$ one has that J_ψ is invertible.

By the theorem of implicit function in Banach spaces (Deimling 2013), $\exists\delta > 0, \epsilon > 0$,

$$\psi(+\infty) \in C^\infty(B_\delta(w_0); B_\epsilon(\psi(+\infty)) \subset H^2(V)) :$$
$$G(\psi(+\infty, w), w) = 0 \quad \forall w \in B_\delta(w_0).$$

The spectrum of $J_\psi(w)$ is continuous w.r.t. $w \Rightarrow \exists\delta' > 0$:

$$\operatorname{Re}(\sigma(J_\psi(w))) < -\alpha/2 < 0 \quad \forall w \in B_{\delta'}(w_0),$$

and therefore preserves stability. □

Proof of the Theorem 3. According to the definition of asymptotic stability (Khalil and Grizzle 2002), $\exists\delta > 0$ and the Lyapunov function $V(\psi) \ge 0$ such that:

$$\frac{dV}{dt} \le -\beta\|\psi - \psi_s\|_{H^1}^2, \quad \beta > 0$$

for $\|\psi^0 - \psi(+\infty, \psi^0)\|_{H^1} < \delta$.

Consider the extended system

$$\frac{d}{dt}\Phi(t, \psi^0) = F(\Phi(t, \psi^0)), \quad \Phi(0, \psi^0) = \psi^0,$$

where the right-hand side $F \in C^\infty$. By the theorem on the smooth dependence of solutions on a parameter,

$$(\Phi : [0, T] \times \mathbb{S}^{2n-1} \to \mathbb{S}^{2n-1}) \in C^\infty.$$

For $\psi_0 \in B_\delta(\psi_s)$ one has the exponential convergence to the stationary state:

$$\|\Phi(t, \psi^0) - \psi(+\infty)\|_{H^1} \le Ce^{-\alpha t}\|\psi^0 - \psi(+\infty)\|_{H^1},\ \alpha > 0,$$

and $\psi(+\infty, \psi^0) = \lim_{t\to\infty}\Phi(t, \psi^0)$ is well defined.

Let $T > 0$. The mapping $\psi^0 \mapsto \Phi(T, \psi^0)$ is smooth. For large enough T, one has

$$\|\Phi(T, \psi^0) - \psi(+\infty, \psi^0)\|_{H^1} < \varepsilon.$$

Since $\psi(+\infty, \psi^0)$ is the uniform limit of smooth functions $\Phi(T_n, \psi_0)$ for $T_n \to \infty$, and the convergence is exponential, we have $\psi(+\infty, \psi^0) \in C^\infty$.

For ψ' in the vicinity $\psi(+\infty)$:

$$\psi(+\infty, \psi^0) = \psi(+\infty) + \int_0^\infty \frac{\partial \Phi}{\partial t}(t, \psi^0) dt.$$

Exponential convergence guarantees the convergence of the integral and smoothness.

The isolation and stability guarantee the invertibility of J_ψ. Then, the theorem of implicit function is applicable and

$$\|\Phi(t, \psi^0) - \psi(+\infty)\|_{H^1} \le Ce^{-\alpha t}$$

allows one to differentiate under the integral.

Using the fact that $\Delta : H^2(V) \to L^2(V)$ is bounded for graphs:

$$\|\Delta\psi\|_{L^2} \le C_\Delta \|\psi\|_{H^2}$$

and $H^2(V) \hookrightarrow H^1(V)$ is compact, one obtains the result. $\square$

Analogous theorems are applicable to the Landau–Lifshitz equation (Maz'ya 2013).

Proofs for moduli space of graph hidden space

***Lemma* 1** (Gradient of $\psi(+\infty)$)**.** *For $e = (i, j) \in V \times V$, the derivative $\frac{\partial \psi(+\infty)}{\partial w(e)}$ satisfies:*

$$\left\| \frac{\partial \psi(+\infty)}{\partial w(e)} \right\|_{H^2} \le \begin{cases} C_1 & \text{if } e \in E_{true} \\ C_2 \exp(-c \cdot d_{\mathcal{G}}(i, j)) & \text{otherwise} \end{cases}$$

Proof. The steady-state solution satisfies:

$$(H + D)\psi(+\infty) = F(\psi^0),\ F(\psi^0) = \int_0^\infty e^{-Dt} H\psi^0 dt.$$

Differentiating w.r.t. $w(e)$:

$$\frac{\partial (H + D)}{\partial w(e)} \psi(+\infty) + (H + D) \frac{\partial \psi(+\infty)}{\partial w(e)} = 0.$$

Thus:

$$\frac{\partial \psi(+\infty)}{\partial w(e)} = -(H + D)^{-1} \frac{\partial H}{\partial w(e)} \psi(+\infty).$$

Case 1 ($e \in E_{\text{true}}$): Since $d_{\mathcal{G}}(i, j) < \rho$, the operator $(H + D)^{-1}$ is bounded:

$$\|(H + D)^{-1}\| \le \lambda_{\min}^{-1} \le C_3,$$

where $\lambda_{\min} > 0$ by the compactness of $\mathcal{G}$ and the dissipativity. The term $\frac{\partial H}{\partial w(e)}$ is the edge derivative of the Laplacian:

$$\left[\frac{\partial H}{\partial w(e)} \psi(+\infty)\right]_k = \begin{cases} \psi_i(+\infty) - \psi_j(+\infty) & k = i \\ \psi_j(+\infty) - \psi_i(+\infty) & k = j \\ 0 & \text{otherwise} \end{cases}.$$

By Lipschitz continuity of $\psi(+\infty)$ on $\mathcal{G}$:

$$|\psi_i(+\infty) - \psi_j(+\infty)| \le L \cdot d_{\mathcal{G}}(i, j) \le L\rho.$$

Thus:

$$\left\| \frac{\partial H}{\partial w(e)} \psi(+\infty) \right\|_{H^2} \le C_4,\ \left\| \frac{\partial \psi(+\infty)}{\partial w(e)} \right\|_{H^2} \le C_3 C_4 = C_1.$$

Case 2 ($e \notin E_{\text{true}}$): For $d_{\mathcal{G}}(i, j) \ge \rho$, the Green's function $G = (H + D)^{-1}$ decays exponentially:

$$|G(x, y)| \le C_5 \exp\left(-c \cdot d_{\mathcal{G}}(x, y)\right),\ c = c(\kappa, \rho) > 0.$$

This follows from the Combes-Thomas estimate for elliptic operators on manifolds (Combes and Thomas 1973). Thus:

$$\begin{aligned} \left\| \frac{\partial \psi(+\infty)}{\partial w(e)} \right\|_{H^2} &\le C_5 \exp(-c \cdot d_{\mathcal{G}}(i, j)) \left\| \frac{\partial H}{\partial w(e)} \psi(+\infty) \right\|_{H^2} \\ &\le C_2 \exp(-c\rho). \end{aligned}$$

$\square$

***Lemma* 2** (Residual Bound))**.** *The residual $r = k(\psi(+\infty)) - y$ satisfies:*

$$\mathbb{E}_{(X,y)\sim\mathcal{D}}[|r|] \ge \begin{cases} \gamma_1 d_{\mathcal{G}}(i, j) - C_6\delta & \text{if } e \in E_{true} \\ C_7\delta & \text{otherwise} \end{cases}.$$

Proof. For $e \in E_{\text{true}}$: By finite propagation speed of the Schrödinger equation on $\mathcal{G}$, if e is absent, the solution error is bounded below:

$$\inf_{\text{no } e} |k(\psi(+\infty)) - y| \ge \gamma_1 d_{\mathcal{G}}(i, j).$$

With δ-perturbations:

$$\mathbb{E}[|r|] \ge \gamma_1 d_{\mathcal{G}}(i, j) - C_6\delta.$$

For $e \notin E_{\text{true}}$: Since $d_{\mathcal{G}}(i, j) \ge \rho$, r depends only on δ-close data:

$$|r| \le C_7\delta \implies \mathbb{E}[|r|] \le C_7\delta.$$

$\square$

***Lemma* 3** (Expected Gradient)**.**

$$\mathbb{E}\left[\frac{\partial \mathcal{L}_{sample}}{\partial w(e)}\right] \le \begin{cases} -\gamma_0 & \text{if } e \in E_{true} \\ C_8\delta & \text{otherwise} \end{cases},$$

where $\gamma_0 = \gamma_1\rho_{\min} - C_9\delta > 0$ and $\rho_{\min} = \min_{(u,v)\in E_{true}} d_{\mathcal{G}}(u, v)$.

Proof. From $\mathcal{L}_{\text{sample}} = r^2$:

$$\frac{\partial \mathcal{L}_{\text{sample}}}{\partial w(e)} = 2r \cdot \nabla k(\psi(+\infty)) \cdot \frac{\partial \psi(+\infty)}{\partial w(e)}.$$

Case 1 ($e \in E_{\text{true}}$):

$$\mathbb{E}\left[\frac{\partial \mathcal{L}_{\text{sample}}}{\partial w(e)}\right] \le -2\gamma_1 d_{\mathcal{G}}(i, j) \cdot \|\nabla k\| \cdot C_1 + 2C_6\delta \cdot \|\nabla k\| \cdot C_1.$$

Since $d_{\mathcal{G}}(i, j) \ge \rho_{\min} > 0$:

$$\mathbb{E}\left[\frac{\partial \mathcal{L}_{\text{sample}}}{\partial w(e)}\right] \le -2C_1\|\nabla k\|(\gamma_1\rho_{\min} - C_6\delta) =: -\gamma_0.$$

Case 2 ($e \notin E_{\text{true}}$):

$$\left|\mathbb{E}\left[\frac{\partial \mathcal{L}_{\text{sample}}}{\partial w(e)}\right]\right| \le 2C_7\delta \cdot \|\nabla k\| \cdot C_2 \exp(-c\rho) \le C_8\delta.$$

$\square$

Lemma 4 (Necessary Edges are Added). *For $e \in E_{true}$ and $e \notin E_t$, the full gradient at test weight θ satisfies:*

$$\frac{\partial \mathcal{L}}{\partial w(e)}(E_t \cup \{e\}, w_t + (e, \theta)) \leq -\gamma_0 + \mu_2 \theta + \mu_1 < -\Theta.$$

Proof. From the definition of $\mathcal{L}$:

$$\frac{\partial \mathcal{L}}{\partial w(e)} = \mathbb{E}\left[\frac{\partial \mathcal{L}_{\text{sample}}}{\partial w(e)}\right] + \mu_2 w(e) + \mu_1 \text{sign}(w(e)).$$

At test weight $w(e) = \theta$:

$$\frac{\partial \mathcal{L}}{\partial w(e)} \leq -\gamma_0 + \mu_2 \theta + \mu_1.$$

Choose $\theta < \frac{\gamma_0 - \mu_1}{2\mu_2}$ and $\Theta < \gamma_0 - \mu_1 - \mu_2 \theta$, so:

$$-\gamma_0 + \mu_2 \theta + \mu_1 < -\Theta.$$

Thus, the condition for addition is met. □

Lemma 5 (Spurious Edges are Not Added). *For $e \notin E_{true}$, the full gradient satisfies:*

$$\frac{\partial \mathcal{L}}{\partial w(e)}(E_t \cup \{e\}, w_t + (e, \theta)) \geq -C_8 \delta + \mu_2 \theta + \mu_1 > -\Theta.$$

Proof.

$$\frac{\partial \mathcal{L}}{\partial w(e)} \geq -C_8 \delta + \mu_2 \theta + \mu_1.$$

Choose $\delta < \frac{\mu_1 + \mu_2 \theta - \Theta}{C_8}$, so:

$$-C_8 \delta + \mu_2 \theta + \mu_1 > -\Theta.$$

Thus, the condition for addition is never met. □

Theorem 8. $\exists T_0 < \infty$ *such that* $\forall t \geq T_0$, $E_t = E_{true}$ *with probability* $\geq 1 - \epsilon$.

Proof. **Stochastic Gradient Concentration** (Boucheron, Lugosi, and Massart 2003) For batch size B, by Chebyshev's inequality:

$$\mathbb{P}\left(\left|g_e - \frac{\partial \mathcal{L}}{\partial w(e)}\right| \geq \Delta\right) \leq \frac{\sigma^2}{B\Delta^2}.$$

Necessary Edge Addition For $e \in E_{\text{true}}$, $e \notin E_t$:

$$\mathbb{P}\left(g_e \geq -\Theta\right) \leq \mathbb{P}\left(g_e - \frac{\partial \mathcal{L}}{\partial w(e)} \geq \Delta_e\right) \leq \frac{\sigma^2}{B\Delta_e^2},$$

where $\Delta_e = -\frac{\partial \mathcal{L}}{\partial w(e)} - \Theta > 0$. Thus:

$$\mathbb{P}(\text{add } e) \geq 1 - \frac{\sigma^2}{B\Delta_e^2}.$$

Spurious Edge Addition For $e \notin E_{\text{true}}$:

$$\mathbb{P}\left(g_e < -\Theta\right) \leq \frac{\sigma^2}{B(\Delta'_e)^2}, \ \Delta'_e = \frac{\partial \mathcal{L}}{\partial w(e)} + \Theta > 0.$$

Union Bound Let $k = |E_{\text{true}}|$, $m = |V \times V|$, $\Delta = \min_e \min(\Delta_e, \Delta'_e)$. After T iterations:

$$\mathbb{P}(\text{error}) \leq k\left(\frac{\sigma^2}{B\Delta^2}\right)^T + mT\frac{\sigma^2}{B\Delta^2}.$$

Set $B = \frac{2mT\sigma^2}{\epsilon\Delta^2}$ and $T = \frac{\log(2k/\epsilon)}{\log(B\Delta^2/\sigma^2)}$. Then:

$$\mathbb{P}(E_t \neq E_{\text{true}} \text{ for some } t \geq T) \leq \epsilon.$$

Thus, $T_0 = O\left(\frac{\log(k/\epsilon)}{\log\log(m/\epsilon)}\right)$. □

Lemma 6 (Weights Remain Above Threshold). *For $e \in E_{true}$, after addition, $w_t(e) > \theta$ for all t with probability 1.*

Proof. At optimum $w^*(e)$:

$$0 = \frac{\partial \mathcal{L}}{\partial w(e)} \leq -\gamma_0 + \mu_2 w^*(e) + \mu_1$$

$$\implies w^*(e) \geq \frac{\gamma_0 - \mu_1}{\mu_2} > 2\theta.$$

By strong convexity, SGD converges to $w^*(e) > \theta$, and $w_t(e) > \theta$ for $t \geq T_e$. □

Proof of the Theorem 4. The graph $G^* = (V, E_{\mathcal{G}}, w^*)$ with $w^*(e) = d_{\mathcal{G}}(u,v)^{-1}$ is the 1-skeleton of the geodesic Delaunay complex (Boissonnat, Dyer, and Ghosh 2018) for the δ-net $V \subset \mathcal{G}$ ($\delta < \rho/4$). By the nerve lemma (Edelsbrunner and Harer 2010): 1. The Čech complex $C_\delta(\mathcal{G})$ (with radius δ) is homotopy equivalent to $\mathcal{G}$. 2. The Delaunay complex $D_{\mathcal{G}}$ (edges $E_{\mathcal{G}}$) is a subcomplex of $C_{\rho/2}(\mathcal{G})$ (since $d_{\mathcal{G}}(u,v) < \rho/2$ for edges). 3. For $\delta < \rho/4$, the inclusion $D_{\mathcal{G}} \hookrightarrow C_\delta(\mathcal{G})$ induces isomorphisms:

$$H_k(D_{\mathcal{G}}) \simeq H_k(C_\delta(\mathcal{G})) \simeq H_k(\mathcal{G}) \quad \text{for } k = 0, 1.$$

Thus, G_t (with $E_t = E_{\mathcal{G}}$ for $t \geq T_0$) satisfies $\beta_k(G_t) = \beta_k(\mathcal{G})$. □

Proof of the Theorem 5. Define the metric $d_{G^*}(u,v) = \inf_{p:u \to v} \sum_{e \in p} d_{\mathcal{G}}(u,v)$ (since $w^*(e) = d_{\mathcal{G}}(u,v)^{-2} \implies \frac{1}{w^*(e)} = d_{\mathcal{G}}(u,v)^2$, but we redefine $d(e) = \frac{1}{w(e)}$ here).

Step 1 $d_{\text{GH}}\left((V, d_{G^*}), \mathcal{G}\right) \leq C_1 \delta$: Since V is a δ-net in $\mathcal{G}$ and G^* uses edges $E_{\mathcal{G}} = \{(u,v) : d_{\mathcal{G}}(u,v) < \rho/2\}$: - For any $u, v \in V$, $d_{G^*}(u,v) \leq d_{\mathcal{G}}(u,v) + O(\delta)$ (by triangle inequality). - For any $x \in \mathcal{G}$, $\exists v \in V$ with $d_{\mathcal{G}}(x,v) \leq \delta$. Thus, $d_{\text{GH}} \leq C_1 \delta$.

Step 2 $d_{\text{GH}}\left((V, d_{G_t}), (V, d_{G^*})\right) \leq C_2 \|w_t - w^*\|_2$: For any $u, v \in V$, let p be the shortest path in G^*. Then:

$$|d_{G_t}(u,v) - d_{G^*}(u,v)| \leq \sum_{e \in p} \left|\frac{1}{w_t(e)} - \frac{1}{w^*(e)}\right|.$$

The function $x \mapsto 1/x$ is θ^{-2}-Lipschitz for $x \geq \theta$, and $|p| \leq \text{diam}(\mathcal{G}) \cdot \max_e \frac{1}{w^*(e)} \leq D$ (since $w^*(e) \geq (\rho/2)^{-2}$). Thus:

$$\begin{aligned} |d_{G_t}(u,v) - d_{G^*}(u,v)| &\leq D\theta^{-2} \|w_t - w^*\|_\infty \\ &\leq D\theta^{-2} \|w_t - w^*\|_2. \end{aligned}$$

Taking supremum over u, v and using $\mathbb{E}[\|w_t - w^*\|_2] \leq \sqrt{\frac{\sigma^2}{\mu_2 t}}$ (Bottou, Curtis, and Nocedal 2018), (Absil, Mahony, and Sepulchre 2010):

$$\mathbb{E}\left[d_{\text{GH}}\left((V, d_{G_t}), (V, d_{G^*})\right)\right] \leq D\theta^{-2}\sqrt{\frac{\sigma^2}{\mu_2 t}}.$$

$$d_{\text{GH}}\left((V, d_{G_t}), \mathcal{G}\right) \leq C_1\delta + C_2 t^{-1/2}.$$

□

Proofs for statistical learning

Proof of the Theorem 6. (Boucheron, Lugosi, and Massart 2003). On the one hand, addition probability (for an edge $e \notin E_t$) of the true edge $e \in E_{true}$ is $P^{\text{true}}_{\text{add}}(e) \geq 1 - \frac{\sigma^2}{B\Delta_e^2}$, while for a spurious edge it is $P^{\text{spur}}_{\text{add}}(e) \leq \frac{\sigma^2}{B(\Delta'_e)^2}$. On the other hand, removal probability for the true edge, due to ℓ_1-regularization, is $P^{\text{true}}_{\text{remove}}(e) \approx 0$, since weights stabilize above θ. For the spurious edges, $P^{\text{spur}}_{\text{remove}}(e) \approx 1$, due to ℓ_1 regularization (Tibshirani 1996).

Let us clarify the structure of the state space. The subsets $E \subseteq V \times V$ can be partitioned into $E = T \sqcup S$, where $T \subseteq E_{\text{true}}$ and $S \subseteq (V \times V) \setminus E_{\text{true}}$. The Algorithm 1 converges to an absorbing state $E_{\text{true}} = T_{\text{full}} \cup \emptyset$: once added, true edges persist. Thus, the algorithm progresses from smaller T to larger T to $T = E_{\text{true}}$.

The expected number of visited strata is

$$\mathbb{E}[|\mathcal{M}_{\text{eff}}|] = \sum_{T \subseteq E_{\text{true}}} \mathbb{E}[V_T],$$

where V_T is the number of distinct spurious edge sets S visited while the true edge set is T. This decomposes into the contribution of the initial state $E_0 = T_0 \sqcup S_0$, which is always visited: since T_0 and S_0 are chosen independently, this contributes $2^{|E_{\text{true}}|}$ possible T and $2^{N^2 - |E_{\text{true}}|}$ possible S, and the visitation at fixed T: the spurious edge set S changes via additions/removals, and the expected number of distinct S visited is bounded by:

$$\mathbb{E}[V_T] = O\left(1 + \mathbb{E}[A_T]\right),$$

where A_T is the number of spurious edge addition events while at T. It holds that

$$\mathbb{E}[A_T] \leq (N^2 - |E_{\text{true}}|) \cdot \mathbb{E}[\text{time at } T] \cdot \max_{e \notin E_{\text{true}}} P^{\text{spur}}_{\text{add}}(e),$$

where time to add next true edge $\mathbb{E}[\text{time at } T] = O(|E_{\text{true}}| - |T|)$, $\max_e P^{\text{spur}}_{\text{add}}(e) \leq \frac{\sigma^2}{B(\Delta')^2}$, where $\Delta' = \min_e \Delta'_e$.

After summation over T,

$$\mathbb{E}[|\mathcal{M}_{\text{eff}}|] = \sum_{T \subseteq E_{\text{true}}} O\left(1 + \frac{(N^2 - |E_{\text{true}}|) \cdot (|E_{\text{true}}| - |T|)\sigma^2}{B(\Delta')^2}\right).$$

Taking into account that $\sum_T 1 = 2^{|E_{\text{true}}|}$ and

$$\sum_T (|E_{\text{true}}| - |T|) = |E_{\text{true}}| \cdot 2^{|E_{\text{true}}| - 1},$$

one obtains the result. □

Considering the model (5), due to Theorems 2, 3, one has

$$\|\psi_\infty(\psi', w') - \psi_\infty(\psi'', w'')\|_2 \leq$$
$$\leq L_\psi(\|\psi' - \psi''\|_2 + \|w' - w''\|_2).$$

This implies:

$$\|f(X)\| \leq C_f = L_{\sigma_3}\left(R_a\left(L_\psi(L_{\sigma_1}(R_A R_X + R_b) + \right.\right.$$
$$\left.\left. + \|\psi_\infty(0,0)\|_2 + L_\psi R_w \sqrt{|E_{\text{true}}|}\right) + R_{b_3}\right)$$

***Lemma* 7** (Covering Number for Graph Weights)**.** *Let $\mathcal{W} \subset [\theta, R_w]^{|E_{true}|}$ be the space of admissible edge weights for the graph $G = (V, E_{true}, w)$ converging to manifold $\mathcal{G}$. Then:*

$$\mathcal{N}(\mathcal{W}, \epsilon, \|\cdot\|_2) \leq \left(\frac{CR_w\sqrt{\beta_1(\mathcal{G})}}{\epsilon}\right)^{\beta_1(\mathcal{G})}$$

where $\beta_1(\mathcal{G})$ is the first Betti number of $\mathcal{G}$.

Proof. The weights w lie on a moduli space $\mathcal{M}_{\text{geo}}$ embedded in $\mathbb{R}^{|E_{\text{true}}|}$ with intrinsic dimension $d = \Theta(\beta_1(\mathcal{G}))$. Diameter

$$\text{diam}\, \mathcal{W} \leq R_w\sqrt{|E_{\text{true}}|} \leq R_w\sqrt{K\beta_1(\mathcal{G})}$$

since $|E_{\text{true}}| = N + \beta_1(\mathcal{G}) - 1$ by Euler's formula.

For a d-dimensional compact manifold:

$$\mathcal{N}(\mathcal{W}, \epsilon) \leq \left(\frac{C'\text{diam}(\mathcal{W})\sqrt{d}}{\epsilon}\right)^d$$

Substituting $d = c\beta_1(\mathcal{G})$ yields the result. □

***Lemma* 8** (Covering Number under Topological Constraints)**.** *The covering number of hypothesis class $\mathcal{F}$ of the model* (5) *satisfies:*

$$\log \mathcal{N}(\mathcal{F}, \epsilon, \|\cdot\|_\infty) \leq C_1\beta_1(\mathcal{G})\log\left(\frac{C_2}{\epsilon}\right)$$
$$+ C_3|E_{true}|\log\left(\frac{C_4}{\epsilon}\right),$$

where C_i depend on model constants and E_{true} is the converged edge set.

Proof. The parameters A_1, b_1, a_3, b_3 are in Euclidean balls, $\mathcal{O}(nN + N)$ parameters in total. The weights w are restricted by topology to a $\beta_1(\mathcal{G})$-dimensional submanifold of $[\theta, R_w]^{|E_{\text{true}}|}$. Taking into account Lemma 7, we get

$$\mathcal{N} \leq \underbrace{\left(\frac{C}{\epsilon}\right)^{2nN+4N+2}}_{\text{FC layers}} \times \underbrace{\left(\frac{C'\sqrt{\beta_1(\mathcal{G})}}{\epsilon}\right)^{\beta_1(\mathcal{G})}}_{\text{Weights}}$$

Taking log gives the result. □

***Lemma* 9** (Rademacher Complexity Bound)**.** *(Wainwright 2019)*

$$\widehat{\mathfrak{R}}_S(\mathcal{F}) \leq \inf_{\alpha > 0}\left(4\alpha + \frac{12}{\sqrt{m}}\int_\alpha^{C_f}\sqrt{\log \mathcal{N}(\mathcal{F}, \epsilon, \|\cdot\|_\infty)}d\epsilon\right)$$

Substituting Lemma 8:

$$\widehat{\mathfrak{R}}_S(\mathcal{F}_{graph}) \leq K_1\sqrt{\frac{\beta_1(\mathcal{G})+nN}{m}} + K_2\frac{\beta_1(\mathcal{G})\log m}{m}$$

where K_1, K_2 depend on model constants.

Proof. By applying Dudley's entropy integral (Vaart and Wellner 1997) and splitting the integral at $\alpha = \frac{1}{\sqrt{m}}$:

$$\int_{1/\sqrt{m}}^{C_f} \sqrt{C_1\beta_1(\mathcal{G})\log(1/\epsilon) + C_3(nN)\log(1/\epsilon)}d\epsilon \\ \leq \sqrt{C(\beta_1(\mathcal{G})+nN)}\int_0^\infty \sqrt{u}e^{-u}du$$

and using that $\int \sqrt{\log(1/\epsilon)}d\epsilon < \infty$. □

Proof of the Theorem 7. Apply the standard Rademacher generalization bound (Bartlett and Mendelson 2002), taking into account both the convergence event (with probability $\geq 1-\epsilon$) and the Rademacher bound (with probability $\geq 1-\delta$). □

Real counterparts of NSE and LLE

Diffusion System

The complex NSE is replaced by a real reaction-diffusion system with similar potential and dissipation terms. The state is a real-valued function $\phi : V \to \mathbb{R}$ (where $\phi(t) \in \mathbb{R}^N$), and the phase space is the unit sphere in $\mathbb{R}^n$.

$$\frac{d\phi}{dt} = F_{\text{real}}(\phi, w), \quad \phi(0) = \phi^0,$$

where $\phi^0, \phi(t) \in H^1(V)$, and $w : E \to \mathbb{R}_+$ is the edge weight function. The operator F_{real} is defined as:

$$F_{\text{real}}(\phi, w) = -\left(\Delta(w) + \operatorname{diag}((\phi(0))^2)\right)\phi \\ - \gamma P_\phi^\perp D_{\text{real}}(\phi, w)$$

with

$$P_\phi^\perp = I - \frac{\phi\phi^\top}{\|\phi\|^2_{\mathbb{R}^V}}, \; D_{\text{real}}(\phi, w) = \Delta\phi + (\phi^2 - \phi(0)^2)\phi.$$

The phase space is

$$\mathcal{M}_{\text{Heat}} = \mathbb{S}^{n-1} = \{\phi \in \mathbb{R}^n : \|\phi\| = 1\}.$$

2D Spin System

The complex LL equation is replaced by a real spin system where each spin $\vec{T}_j \in \mathbb{S}^1$ (unit circle). The stereographic projection maps $\mathbb{R}$ to $\mathbb{S}^1$ (omitting the point $(-1, 0)$).

For $\phi_j \in \mathbb{R}$ (real-valued), define $\vec{T}_j \in \mathbb{S}^1$ using stereographic projection:

$$\vec{T}_j = \left(\frac{2\phi_j}{1+\phi_j^2}, \frac{1-\phi_j^2}{1+\phi_j^2}\right)^\top, \quad \phi_j = \frac{T_j^x}{1+T_j^y}.$$

This map is conformal and singular at $\phi_j \to \infty$ (where $\vec{T}_j \to (-1, 0)$).

After substituting the real NSE into the stereographic projection, we obtain the following result.

$$\frac{d\vec{T}_j}{dt} = \vec{T}_j \times \left(-2\sum_k w_{jk}\vec{T}_k + 2(\phi_j(0))^2\vec{e}_2\right) \\ - \gamma\vec{T}_j \times \left(\vec{T}_j \times \vec{\mathcal{D}}_j\right),$$

where $\vec{e}_2 = (0, 1)^\top$, and the dissipation field is:

$$\vec{\mathcal{D}}_j = -2\sum_k w_{jk}(\vec{T}_k - \vec{T}_j) + 2\left((\phi_j(0))^2 - \frac{1}{|V|}\sum_i (\phi_i(0))^2\right)\vec{e}_2.$$

Here, $\times$ is the 2D cross product (scalar): $\vec{a} \times \vec{b} = a_x b_y - a_y b_x$. The term $\vec{T}_j \times (\vec{T}_j \times \vec{\mathcal{E}}_j)$ simplifies to $-\vec{\mathcal{E}}_j$ since $\|\vec{T}_j\| = 1$.

The phase space is

$$\mathcal{M}_{\text{LL-real}} = \Big\{(\vec{T}_1, \ldots, \vec{T}_n) \in (\mathbb{S}^1)^n : \\ \sum_{j=1}^n \frac{1-T_j^y}{1+T_j^y} = 1, \; T_j \neq (-1, 0)\Big\},$$

which is diffeomorphic to $\mathbb{S}^{n-1}$ via the stereographic projection.