
SCIENTIFIC REPORTS

IMS-NTU joint workshop on Applied Geometry for Data Sciences Part II

02 June 2025–06 June 2025

Organizing Committee

Han Fei

National University of Singapore

Wilderich Tuschmann

Karlsruhe Institute of Technology

Theodore Papamarkou

Zhejiang Normal University

Yuguang Wang

Shanghai Jiao Tong University

Kelin Xia

Nanyang Technological University

Rex Ying

Yale University

CONTENTS PAGE

	Page	
Frédéric Barbaresco Thales Group, France	Symplectic Foliation-informed Neural Network (SFINN) and Lie Groups Machine Learning based on Jean-marie Souriau Lie Groups Thermodynamics & Koszul Information Geometry	4
Tolga Birdal Imperial College London, UK	Topological Complexity Measures as Proxies for Generalization in Neural Networks	11
Cristian Bodnar Silurian AI, UK	Aurora: A Foundation Model for the Earth System	12
Baris Coskunuzer University of Texas Dallas, USA	Topological Compound Fingerprinting in Computer Aided Drug Discovery	13
Xiaowen Dong University of Oxford, UK	Bayesian Optimisation of Graph-based Functions	15
Mustafa Hajij University of San Francisco, USA	Frontiers and Opportunities in Topological Deep Learning	16
Niu Huang National Institute of Biological Sciences, China	Integrating HPC and AI: A New Paradigm for Predicting Protein-ligand Binding	17
Wei Huang RIKEN, Japan	Decoding Deep Graph Neural Networks: An Optimization and Generalization Perspective	19
Stephan Huckemann Georg-August-Universität Göttingen, Germany	Dirty Limit Theorems and Applications	21
Stephan Klaus Mathematisches Forschungsinstitut Oberwolfach, Germany	Nonlinear Regression with Real Algebraic Varieties and their Topology	26
Patrice Koehl University of California, Davis, USA	A Physicist's View on Partial 3D Shape Comparison	30
Ran Levi The University of Aberdeen, UK	Foundations of Differential Calculus for Modules over Small Categories	35
Zheng Ma Shanghai Jiao Tong University, China	Solving PDE Inverse Problems with Generative Models and Their Applications	37
Anthea Monod Imperial College London, UK	Algebraic Geometry Learns Machines and Machines Learn Algebraic Geometry	38
Frank Nielsen Sony Computer Science Laboratories, Japan	Computational Information Geometry on Bregman Manifolds and Submanifolds	39

Page

Hans Riess Georgia Institute of Technology, USA	Categories and Sheaves for Optimization: From Multi-Stage to Distributed	40
Roman Sauer Karlsruhe Institute for Technology, Germany	Expanders, Waists, and the Kazhdan Property	43
Jian Tang HEC Montréal, Canada	Geometric Deep Learning for Protein Design	45
Ben Yang University of Oxford, UK	Large Isometry Invariant Topological Transform Shape Descriptor	46
Yaoyu Zhang Shanghai Jiao Tong University, China	Towards Understanding the Condensation Phenomenon of Deep Neural Networks	48
Yipeng Zhang Nanyang Technological University, Singapore	Multi-Cover: A Mathematical Framework for Topological Data Analysis and Deep Learning	49
Difan Zou University of Hong Kong, Hong Kong SAR	Towards Understanding the Representation Learning of Diffusion Models	53

SYMPLECTIC FOLIATION-INFORMED NEURAL NETWORK (SFINN) AND LIE GROUPS MACHINE LEARNING BASED ON JEAN-MARIE SOURIAU LIE GROUPS THERMODYNAMICS & KOSZUL INFORMATION GEOMETRY

FRÉDÉRIC BARBARESCO

Keywords: Symplectic foliation, Lie groups, thermodynamics, metriplectic flow, information geometry, neural networks

INTRODUCTION

The symplectic model of statistical mechanics developed by Jean-Marie Souriau, termed the “Thermodynamics of Lie Groups”, extends the structures of information geometry to the realm of Lie groups. This framework enables the definition of Maximum Entropy Gibbs densities possessing covariance under the action of the group operating on the system. Moreover, it generalises the Fisher–Rao–Fréchet metric to Lie groups, rendering it invariant under the group’s action. Crucially, Shannon’s axiomatic definition of entropy is supplanted by a purely geometric construction, wherein entropy emerges as a Casimir invariant function defined on the leaves of the foliation induced by coadjoint orbits through the moment map associated with the group action (the moment map being the geometric counterpart of Noether’s theorem).

Souriau’s thermodynamics of Lie groups introduces a web-like geometric structure composed of two transverse foliations: a symplectic foliation generated by coadjoint orbits (corresponding to the level sets of entropy) and a transverse Riemannian foliation (corresponding to the level sets of energy). The dynamics on each foliation make it possible to distinguish between non-dissipative phenomena (with constant entropy) and dissipative phenomena (with constant energy and entropy production). This dynamic behaviour is governed by a metriplectic flow that encapsulates the first law of thermodynamics through Poisson bracket (quantitative conservation of energy) and the second law through metric flow bracket (qualitative degradation of energy and generation of entropy).

We shall explore the connections between TINNs (Thermodynamics-Informed Neural Networks), metriplectic flows, and the Lie groups thermodynamics. The overarching aim is for TINNs not merely to learn from data, but also to adhere to thermodynamic constraints, thereby enabling more accurate predictions and a deeper understanding of physical systems, particularly those characterised by dissipative phenomena.

Souriau Lie Groups Thermodynamics is studied in the framework of two european action, European CaLISTA COST action and European CaLIGOLA MSCA action.

WEBS AND FOLIATIONS MODEL OF DISSIPATIVE LIE GROUPS THERMODYNAMICS

Jean-Marie Souriau's model, known as "Lie Group Thermodynamics", is a symplectic formulation of statistical mechanics. This framework incorporates geometric methods into statistical mechanics, wherein the Gibbs states of a system are represented as points within a symplectic manifold, and Lie groups express the symmetries of the system. We offer an original interpretation of "Lie Group Thermodynamics" through a symplectic foliation generated by the coadjoint orbits of the Lie group acting upon the system, wherein the entropy is characterised as an invariant Casimir function. Transverse to this symplectic foliation, considered as the level sets of entropy, we associate a Riemannian foliation, defined as the level sets of energy. These transverse foliations together define a web structure. We describe the dynamics along both sets of transverse leaves by means of a metriplectic flow: this flow combines a Poisson bracket on the symplectic leaves (level sets of entropy), which accounts for non-dissipative phenomena by preserving entropy, with a metric bracket on the Riemannian leaves (level sets of energy), which characterises dissipative phenomena through the production of entropy. Within the framework of information geometry for statistical manifolds, we may associate the symplectic foliation with the Fisher metric, and the Riemannian foliation with the dual of the Fisher metric (the Hessian of entropy). In the context of Information geometry, the Riemannian metric of an exponential family is given by the Fisher metric:

$$g_{ij} = - \left[\frac{\partial^2 \Phi}{\partial \theta_i \partial \theta_j} \right]_{ij} \quad \text{with } \Phi(\theta) = -\log \int_R e^{-\langle \theta, y \rangle} d\lambda(y) = -\log \psi_R \quad (1)$$

The Shannon Entropy is given by the Legendre transform:

$$S(\eta) = \langle \theta, \eta \rangle - \Phi(\theta) \quad \text{with } \eta_i = \frac{\partial \Phi(\theta)}{\partial \theta_i} \quad \text{and} \quad \theta_i = \frac{\partial S(\eta)}{\partial \eta_i} \quad (2)$$

$\Phi(\theta) = -\log \int_R e^{-\langle \theta, y \rangle} d\lambda(y)$ the generating function of cumulants in statistics. In the Souriau model, the structure of the information geometry is preserved and extended for Lie Groups over the Symplectic manifold M associated with the coadjoint orbits:

$$I(\beta) = -\frac{\partial^2 \Phi}{\partial \beta^2} \quad \text{and} \quad \Phi(\beta) = -\log \int_M e^{-\langle \beta, U(y) \rangle} d\lambda \quad \text{with } U: M \rightarrow \mathfrak{g}^* \quad (3)$$

$$S(Q) = \langle \beta, Q \rangle - \Phi(\beta) \quad \text{with } Q = \frac{\partial \Phi(\beta)}{\partial \beta} \in \mathfrak{g}^* \quad \text{and} \quad \beta = \frac{\partial S(Q)}{\partial Q} \in \mathfrak{g} \quad (4)$$

From these equations, we can prove that we have a complete integrable system. We can observe first that the Entropy is a Casimir function on symplectic leaves generated by coadjoint orbits:

$$\forall H, \{S, H\}_{\tilde{\Theta}}(Q) = \left\langle ad_{\frac{\partial S}{\partial Q}}^* Q + \Theta\left(\frac{\partial S}{\partial Q}\right), \frac{\partial H}{\partial Q} \right\rangle = 0 \Rightarrow ad_{\frac{\partial S}{\partial Q}}^* Q + \Theta\left(\frac{\partial S}{\partial Q}\right) = 0 \quad (5)$$

This equation appears in the Souriau paper published in 1974, observing that geometric temperature β is a kernel of $\tilde{\Theta}_\beta$, which is written as follows:

$$\beta \in Ker \tilde{\Theta}_\beta \Rightarrow \langle Q, [\beta, Z] \rangle + \tilde{\Theta}(\beta, Z) = 0 \quad (6)$$

and can be developed to recover the Casimir equation:

$$\begin{aligned} \langle Q, ad_\beta Z \rangle + \tilde{\Theta}(\beta, Z) &= 0 \Rightarrow \langle ad_\beta^* Q, Z \rangle + \tilde{\Theta}(\beta, Z) = 0 \\ \xrightarrow{\beta = \frac{\partial S}{\partial Q}} \left\langle ad_{\frac{\partial S}{\partial Q}}^* Q, Z \right\rangle + \tilde{\Theta}\left(\frac{\partial S}{\partial Q}, Z\right) &= \left\langle ad_{\frac{\partial S}{\partial Q}}^* Q + \Theta\left(\frac{\partial S}{\partial Q}\right), Z \right\rangle = 0, \forall Z \Rightarrow ad_{\frac{\partial S}{\partial Q}}^* Q + \Theta\left(\frac{\partial S}{\partial Q}\right) = 0 \end{aligned} \quad (7)$$

Geometric heat equation is then given by:

$$\frac{dQ}{dt} = ad_{\frac{\partial S}{\partial Q}}^* Q + \Theta\left(\frac{\partial H}{\partial Q}\right) \quad (8)$$

We introduce a web structure model, grounded in information geometry and heat theory, based on Jean-Marie Souriau's thermodynamics of Lie groups, to describe the transverse Poisson structure of the metriplectic flow associated with dissipative phenomena. The metriplectic bracket was first introduced in 1983 by A. N. Kaufman and P. J. Morrison. This formalism ensures the conservation of energy and the non-decrease of entropy, and it reduces to the traditional Poisson bracket formalism in the limit where dissipation is absent. The axiomatic formulation of this model was developed concurrently by Grmela and Öttinger. There are three principal types of dissipation: thermal diffusion, which conserves energy while producing entropy via heat transfer; viscosity, which extracts energy from the system (as exemplified by the Navier–Stokes equations); and transport equations incorporating collision operators. These classes of dissipative systems, which conform to both the first and second laws of thermodynamics, are encompassed within metriplectic dynamics. A novel bracket $\{\{.,.\}\}$ within the metriplectic formalism yields the corresponding evolution equation:

$$\frac{df}{dt} = \{\{f, F\}\} = \{f, F\} + (f, F) \text{ with Hamiltonian } F = H + S \quad (9)$$

The 2^{nd} bracket is a metric bracket checking the 2 constraints $(f, F) = (F, f)$ and $(f, f) \geq 0$ with the entropy S selected in the set of Casimir invariants of the non-canonical Poisson bracket. The metriplectic flow thus conforms to the 1^{st} and 2^{nd} principles of thermodynamics:

1st principle of thermodynamics: conservation of energy

$$\frac{dH}{dt} = \{H, F\} + (H, F) = \{H, H\} + \{H, S\} + (H, H) + (H, S) = 0 \quad (10)$$

because $\{H, H\} = 0$ by symmetry, $\{f, S\} = 0, \forall f$ and $(H, f) = 0, \forall f$

2nd principle of thermodynamics: the production of entropy

$$\frac{dS}{dt} = \{S, F\} + (S, F) = 0 + (S, H) + (S, S) = (S, S) \geq 0 \quad (11)$$

with $\{S, f\} = 0 \forall f, (f, H) = 0 \forall f$ and (S, S) positive half-definite

Finally, two compatible brackets, a Poisson bracket and a symmetric bracket, determine the geometry in metriplectic systems, as illustrated in Figure 15:

$$\frac{df}{dt} = \{\{f, F\}\} = \{f, H\} + (f, S) \quad (12)$$

The energy H is a Casimir invariant of the dissipative bracket, and the entropy S is a Casimir invariant of the Poisson bracket:

$$\{S, H\} = 0 \forall H \text{ and } (H, S) = 0 \forall S \quad (13)$$

The symmetry requirement generalizes Onsager symmetry from irreversible linear The metriplectic equation linked to transverse symplectic foliations is illustrated in Figure 1.

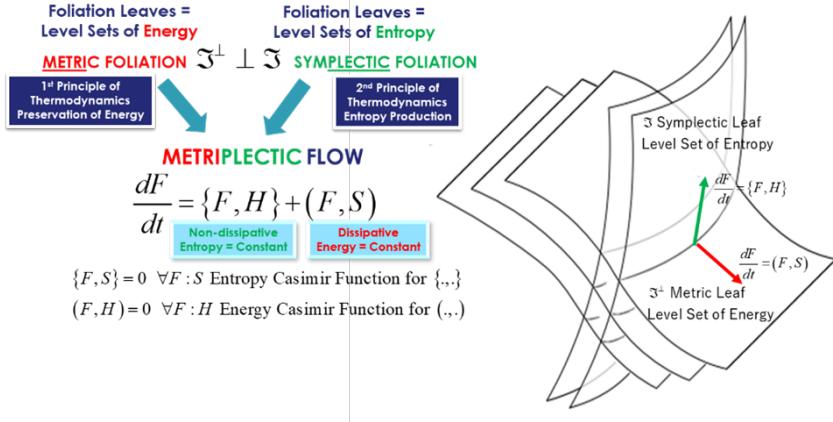


FIGURE 1. Metriplectic Flow on Webs structure: Symplectic Foliation (level set of Entropy) & Transverse Riemannian Foliation (level sets of Energy).

Carnot's cycle can be defined within Souriau's Lie Group Thermodynamics on the symplectic leaves (level sets of entropy) and the transverse Riemannian leaves (level sets of energy), as illustrated in Figure 2. The Carnot cycle is a theoretical thermodynamic cycle for a dithermal engine, comprising four reversible processes: reversible isothermal expansion, reversible adiabatic (isentropic) expansion, reversible isothermal compression, and reversible adiabatic (isentropic) compression. Alternatively, the Carnot cycle may be represented on a temperature–entropy diagram: AB denotes isothermal expansion; BC, adiabatic (isentropic) expansion; CD, isothermal compression; and DA, adiabatic (isentropic) compression.

ENTROPY DEFINITION: Entropy is Invariant Casimir Function on Leaves of Symplectic Foliation associated to Coadjoint Orbits generated via Moment Map of Symmetry Group acting on the System.

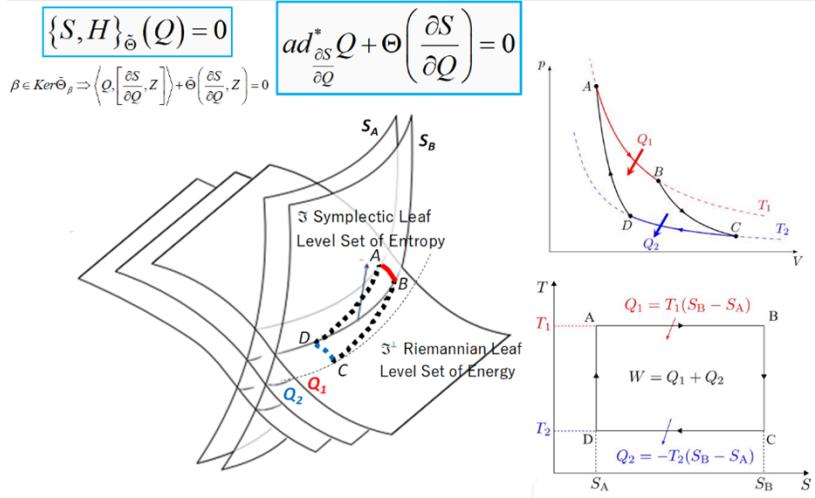


FIGURE 2. Carnot cycle in T–P and T–S representations (right) and Souriau cycle on symplectic and transverse Riemannian foliations (left).

FROM THERMODYNAMICS-INFORMED NEURAL NETWORKS TO SFINNS

The integration of physical knowledge into learning models is a rapidly growing field aimed at enhancing the modelling of complex physical phenomena across a wide range of applications. Accurate modelling of physical systems plays a fundamental role in many areas of military products, both for engineering purposes and operational use. Traditional models, based on physical laws and expressed through partial differential equations (PDEs), are powerful but limited. These conventional techniques often exhibit slow inference, rely on overly simplistic assumptions, and tend to neglect or underutilise data collected from the modelled system. While data-driven approaches, such as neural networks, can overcome these limitations, the data available are typically too scarce to support purely data-driven methods. Research efforts address the challenge of modelling physical systems using data-driven techniques in contexts where system data are limited and sparse. These research projects advance modelling methodologies by integrating prior physical knowledge at various stages of the modelling process. This form of hybridisation between neural networks and analytical models is entitled PINN (Physics-Informed Neural Network), a neural network informed by physics, injecting the physical model into the neural network through both hard and soft constraints. PINN could integrate Thermodynamics constraints to model dissipative process, and are called TINN (Thermodynamics-Informed Neural Network), a neural network based on thermodynamics, hybriding data integration with principles of energy conservation and entropy production.

This type of formulations are also known as metriplectic formulations, since they combine metric and symplectic terms. However, in GENERIC equations, known as degeneracy conditions, play a fundamental role. They are key ingredients in the demonstration of the a priori satisfaction of the two laws of thermodynamics:

- **Conservation of energy** in closed systems Given the anti-symmetry of L : $\frac{dE}{dt}(z) = \frac{\partial E}{\partial z} \frac{\partial z}{\partial t} = 0$
- **Non-negative entropy production**, given the positive semi-definiteness of M : $\frac{dS}{dt}(z) = \frac{\partial S}{\partial z} \frac{\partial z}{\partial t} = \frac{\partial S}{\partial z} M(z) \frac{\partial S}{\partial z} \geq 0$

Therefore, the GENERIC structure consistently guarantees the **satisfaction of the laws of thermodynamics by construction**. This makes GENERIC a very appealing choice for the construction of inductive biases in the learning of physical phenomena.

For Thermodynamics of learning physical phenomena, we assume that data sets D_i contain labelled pairs of a single-step state vector z_t and its evolution in time z_{t+1} , $D = \{D_i\}_{i=1}^{N_{\text{sim}}}, D_i = \{(z_t, z_{t+1})\}_{t=0}^T$ so that a neural network can be constructed by means of two loss terms, a data loss term that takes into account the correct prediction of the state vector time evolution using the following integrator, called GENERIC/METRIPLECTIC, defined as $\Lambda_n^{\text{data}} = \left\| \frac{dz^{\text{GT}}}{dt} - \frac{dz^{\text{net}}}{dt} \right\|_2^2$ where $\|\cdot\|_2$ denotes the L2-norm.

The choice of the time derivative instead of the state vector itself is to regularize the global loss function to a uniform order of magnitude with respect to the degeneracy terms. A second loss term takes into account the fulfillment of the

degeneracy equations, $|\Lambda_n^{\text{deg}}| = \left\| L \frac{\partial S}{\partial z_n} \right\|_2^2 + \left\| M \frac{\partial E}{\partial z_n} \right\|_2^2$. This formulation gave rise to the so-called structure-preserving neural networks and thermodynamics-informed neural networks. These networks have been employed recently in the development of physics perception with the help of computer vision techniques. The global loss term is a weighted mean of the two terms over the shuffled Nbatch batched snapshots: $\Lambda = \frac{1}{N_{\text{batch}}} \sum_{n=0}^{N_{\text{batch}}} (\lambda \Lambda_n^{\text{data}} + \Lambda_n^{\text{deg}})$.

We call Symplectic Foliations-Informed Neural Network (SFINN), integrator using symplectic foliation and transverse Riemannian foliation described by the Souriau model in previous chapter.

REFERENCES

- [1] F. Barbaresco, F. Gay-Balmaz. Lie Group Cohomology and (Multi)Symplectic Integrators: New Geometric Tools for Lie Group Machine Learning Based on Souriau Geometric Statistical Mechanics. *Entropy*, 22:498, 2020.
- [2] F. Barbaresco. Koszul lecture related to geometric and analytic mechanics, Souriau's Lie group thermodynamics and information geometry. *Information Geometry*, 4:245–262, 2021.
- [3] F. Barbaresco. Jean-Marie Souriau's Symplectic Model of Statistical Physics: Seminal Papers on Lie Groups Thermodynamics. In: Nielsen F., Barbaresco F. (eds) *Geometric Structures of Statistical Physics, Information Geometry, and Learning*. Springer, 2021.
- [4] F. Barbaresco. Symplectic theory of heat and information geometry. *Handbook of Statistics*, 46(4):107–143, 2022.
- [5] F. Barbaresco. Symplectic Foliation Structures of Non-Equilibrium Thermodynamics as Dissipation Model: Application to Metriplectic Nonlinear Lindblad Quantum Master Equation. *Entropy*, 24:1626–1662, 2022.
- [6] F. Barbaresco. Symplectic foliation model of information geometry for statistics and learning on Lie groups. *Geometry, topology and statistics in data sciences*, Institut Henri Poincaré, Paris, 2022. Available at <https://www.carmin.tv/fr/video/symplectic-foliation-model-of-information-geometry-for-statistics-and-learning-on-lie-groups>
- [7] F. Barbaresco. Symplectic Foliation Transverse Structure and Libermann Foliation of Heat Theory and Information Geometry. In: Nielsen F., Barbaresco F. (eds) *Geometric Science of Information GSI'23*. Lecture Notes in Computer Science 14072, Springer, 2023, pp. 152–164.
- [8] F. Barbaresco. Transverse Symplectic Foliation Structure for Thermodynamics-Informed Neural Network and Lie-Groups Machine Learning. Erwin Schrödinger Institute Seminar, Infinite-dimensional Geometry: Theory and Applications trimester, 2025. Available at <https://www.esi.ac.at/events/t2213/>
- [9] F. Barbaresco. Jean-Marie Souriau's Symplectic Foliation Model of Sadi Carnot's Thermodynamics. *Entropy*, 27:509, 2025. <https://doi.org/10.3390/e27050509>
- [10] F. Barbaresco. Transverse Symplectic Foliation Structure for Dissipative Thermodynamics based on Souriau Model of Statistical Mechanics. 18th Joint European Thermodynamics Conference, Mini-Symposia “Foundations and philosophy of thermodynamics”, Belgrade, 2025.
- [11] J.-M. Souriau. *Structure des systèmes Dynamiques*. Dunod, 1969. English translation: *Structure of Dynamical Systems: A Symplectic View of Physics*, Birkhäuser, 1997. Available at <https://empg.maths.ed.ac.uk/Activities/Souriau/Souriau.pdf>
- [12] J.-M. Souriau. Mécanique statistique, groupes de Lie et cosmologie. In: *Colloque International du CNRS Géométrie symplectique et physique Mathématique*, CNRS, Marseille, 1974.
- [13] J.-L. Koszul, Y.-M. Zou. *Introduction to Symplectic Geometry*. Springer Nature Singapore, 2019. <https://link.springer.com/book/10.1007/978-981-13-3987-5>

- [14] F. Barbaresco, F. Nielsen (eds). *Geometric Structures of Statistical Physics, Information Geometry, and Learning: SPIGL'20, Les Houches, France, July 27–31*. Springer, 2021. <https://link.springer.com/book/10.1007/978-3-030-77957-3>
- [15] European COST network CaLISTA. <https://site.unibo.it/calista/en>
- [16] European HORIZON-MSCA CaLIGOLA. <https://site.unibo.it/caligola/en>
- [17] Special Issue, Dynamics Beyond the Hamiltonian: Dissipation in Classical Metriplectic Systems and Quantum Non-Unitary Systems. Guest Editors: M. Materassi & F. Barbaresco. *Entropy*, MDPI. <https://www.mdpi.com/si/229747>

THALES, 19/21 AVENUE MORANE SAULNIER, 78140 VÉLIZY-VILLACOUBLAY
Email address: frederic.barbaresco@thalesgroup.com

TOPOLOGICAL COMPLEXITY MEASURES AS PROXIES FOR GENERALIZATION IN NEURAL NETWORKS

TOLGA BIRDAL

Deep neural networks (DNNs) exhibit remarkable generalization abilities, yet the mechanisms behind these capabilities remain poorly understood, defying the established wisdom of statistical learning theory. Recent research has revealed a compelling link between the fractal structures formed during iterative training and the resulting generalization performance. In this talk, Dr. Birdal sheds new light on these connections by presenting a novel framework that ties complexity measures to the topological properties of the training process. The presentation begins by bounding the generalization error through the fractal dimension of training trajectories, practically computed using tools from persistent homology—introducing the ‘persistent homology dimension’ as a new, insightful proxy for generalization. Building on this, Dr. Birdal introduces more computationally efficient topological complexity measures that bypass the need for continuous training trajectories. These measures consistently show strong correlations with the generalization gap across diverse models, including transformers and graph networks. The findings hold transformative implications for both theory and practice, offering a new lens to study, understand and optimize the generalization power of modern AI systems.

Relevant Publications:

<https://arxiv.org/abs/2111.13171> [arxiv.org] (NeurIPS 2021)

<https://arxiv.org/abs/2407.08723> [arxiv.org] (NeurIPS 2024)

IMPERIAL COLLEGE LONDON, UK

Email address: tbirdal@imperial.ac.uk

AURORA: A FOUNDATION MODEL FOR THE EARTH SYSTEM

CRISTIAN BODNAR

Keywords: Atmospheric and Oceanic Physics; Machine Learning

Reliable forecasts of the Earth system are crucial for human progress and safety from natural disasters. Artificial intelligence offers substantial potential to improve prediction accuracy and computational efficiency in this field, however this remains underexplored in many domains. Here we introduce Aurora, a large-scale foundation model for the Earth system trained on over a million hours of diverse data. Aurora outperforms operational forecasts for air quality, ocean waves, tropical cyclone tracks, and high-resolution weather forecasting at orders of magnitude smaller computational expense than dedicated existing systems. With the ability to fine-tune Aurora to diverse application domains at only modest computational cost, Aurora represents significant progress in making actionable Earth system predictions accessible to anyone.

<https://arxiv.org/abs/2405.13063>

SILURIAN AI, UK

Email address: crisbodnar96@gmail.com

TOPOLOGICAL TRANSFORMERS FOR MOLECULAR PROPERTY PREDICTION

JOSHEM UDDIN, ASTRIT TOLA, CUNEYT AKCORA, AND BARIS COSKUNUZER

Classification AMS 2020: 62R40, 55N31

Keywords: Topological Data Analysis, Persistent Homology, Molecular Property Prediction, Transformers

ABSTRACT. We introduce *Topo-Transformers*, a novel framework that integrates topological data analysis with transformer architectures to advance graph representation learning. Our approach circumvents the computational bottlenecks of traditional Persistent Homology (PH) by employing *Topo-Scan*, an efficient, sequential encoding of topological signatures. This encoding enables scalable and parallelizable integration of topological features within attention-based models, improving both performance and generalization across graph domains. Topo-Transformers demonstrate state-of-the-art performance on multiple benchmarks for graph classification and molecular property prediction, offering a robust and extensible foundation for topology-aware deep learning on graphs.

1. INTRODUCTION

Graphs are powerful data structures for modeling complex relational systems in domains such as biology, social networks, and finance. Despite recent progress in Graph Neural Networks (GNNs), current models often overfit to specific datasets or tasks and struggle to generalize across heterogeneous domains. In contrast, foundation models in vision and language exhibit broad transferability, typically achieved through large-scale pretraining. Developing such general-purpose Graph Foundation Models (GFMs) remains an open challenge, largely due to the non-Euclidean, irregular structure of graphs that vary in size, sparsity, and topology [1, 2].

Topological Data Analysis (TDA) provides a principled approach to encoding multiscale structural information that is invariant to node permutations and stable under perturbations. However, the use of topological methods in deep graph learning is limited, primarily due to the computational expense of Persistent Homology (PH), which has cubic complexity in worst-case scenarios. Existing approximations often focus on point cloud data and fail to exploit graph-specific structures [3].

2. METHOD OVERVIEW

We propose *Topo-Transformers*, a scalable graph learning framework that unifies TDA and transformer architectures. At the core of our model is *Topo-Scan*, a novel topological encoder that generates fixed-length, sequential representations of graph topology. These signatures are optimized for transformer-based processing and can be integrated directly into attention mechanisms. Unlike prior approaches that rely on message passing or spectral analysis, Topo-Transformers capture both local and global graph patterns using a unified attention framework.

3. EXPERIMENTAL RESULTS

We evaluate Topo-Transformers on eight standard graph classification datasets and seven molecular property prediction tasks. Our model achieves state-of-the-art accuracy on 6 out of 8 classification benchmarks and ranks first on 4 of 7 molecular prediction tasks. These results demonstrate the effectiveness of incorporating topological structure into modern deep learning pipelines.

4. THEORETICAL ANALYSIS

We provide theoretical guarantees showing that our Topo-Scan descriptors are stable: small perturbations in the input graph lead to bounded changes in the encoded representation. This robustness is essential for real-world applications, where input noise or sampling variability is common.

5. CONTRIBUTIONS

- **Topo-Scan:** A novel encoding that transforms graph topology into sequential, fixed-size vectors suitable for transformers.
- **Topo-Transformers:** A scalable framework that fuses topological descriptors with attention mechanisms for end-to-end graph learning.
- **Empirical Success:** Extensive experiments demonstrate state-of-the-art performance on diverse benchmarks.
- **Theoretical Stability:** We establish robustness guarantees for Topo-Scan encodings under input perturbations.

REFERENCES

- [1] Mao, H., Chen, Z., Tang, W., Zhao, J., Ma, Y., Zhao, T., Shah, N., Galkin, M., and Tang, J. (2024). Position: Graph foundation models are already here. In *Proceedings of the International Conference on Machine Learning (ICML)*.
- [2] Wang, H., Jiang, Z., You, Y., Han, Y., Liu, G., Srinivasa, J., Kompella, R., and Wang, Z. (2023). Graph mixture of experts: Learning on large-scale graphs with explicit diversity modeling. In *Advances in Neural Information Processing Systems*, 36, 50825–50837.
- [3] Coskunuzer, B. and Akçora, C.G. (2024). Topological methods in machine learning: A tutorial for practitioners. *arXiv preprint arXiv:2409.02901*.

UNIVERSITY OF TEXAS AT DALLAS

Email address: joshem.uddin@utdallas.edu

UNIVERSITY OF TEXAS AT DALLAS

Email address: astrit.tola@utdallas.edu

UNIVERSITY OF CENTRAL FLORIDA

Email address: cuneyt.akcora@ucf.edu

UNIVERSITY OF TEXAS AT DALLAS

Email address: coskunuz@utdallas.edu

BAYESIAN OPTIMISATION OF GRAPH-BASED FUNCTIONS

XIAOWEN DONG

Classification AMS 2020: 90C27

Keywords: Graph signal processing, Gaussian processes, Bayesian optimisation

The increasing availability of graph-structured data motivates a new type of optimisation problems over graph-based functions, i.e., searching for the graph or node that maximises the value of an underlying function. Such optimisation problems are challenging due to the search space that is discrete and high-dimensional, as well as the underlying function that is often black-box and expensive to evaluate. In this talk, I will provide several examples on how Bayesian optimisation can be used to optimise graph-based functions defined on graphs [1, 2], node set of a graph [3], and node subsets of a graph [4]. These are enabled by generalising Gaussian processes to graph-structured data [5, 6], and they demonstrate the promise in combining probabilistic and geometric reasoning for analysing complex functions or solving machine learning tasks [7, 8]. Practical applications include automated machine learning, epidemiological source identification, and social influence maximisation.

REFERENCES

- [1] B. Ru, X. Wan, X. Dong and M. A. Osborne. Interpretable neural architecture search via Bayesian optimisation with Weisfeiler-Lehman kernels. *International Conference on Learning Representations (ICLR)*, 2021.
- [2] X. Wan, H. Kenlay, B. Ru, A. Blaas, M. A. Osborne and X. Dong. Adversarial attacks on graph classifiers via Bayesian optimisation. *Conference on Neural Information Processing Systems (NeurIPS)*, 2021.
- [3] X. Wan, P. Osselin, H. Kenlay, B. Ru, M. A. Osborne and X. Dong. Bayesian optimisation of functions on graphs. *Conference on Neural Information Processing Systems (NeurIPS)*, 2023.
- [4] H. Liang, X. Wan and X. Dong. Bayesian optimization of functions over node subsets in graphs. *Conference on Neural Information Processing Systems (NeurIPS)*, 2024.
- [5] F. L. Opolka, Y.-C. Zhi, P. Liò and X. Dong. Adaptive Gaussian processes on graphs via spectral graph wavelets. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.
- [6] Y.-C. Zhi, Y. C. Ng and X. Dong. Gaussian processes on graphs via spectral kernel learning. *IEEE Transactions on Signal and Information Processing over Networks*, 2023.
- [7] F. L. Opolka, Y.-C. Zhi, P. Liò and X. Dong. Graph classification Gaussian processes via spectral features. *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2023.
- [8] H. Sáez de Ocáriz Borde, Á. Arroyo, I. Morales, I. Posner and X. Dong. Neural latent geometry search: Product manifold inference via Gromov-Hausdorff-informed Bayesian optimization. *Conference on Neural Information Processing Systems (NeurIPS)*, 2023.

EAGLE HOUSE, WALTON WELL ROAD, OXFORD OX2 6ED, UK
Email address: xiaowen.dong@eng.ox.ac.uk

FRONTIERS AND OPPORTUNITIES IN TOPOLOGICAL DEEP LEARNING

MUSTAFA HAJI

Classification AMS 2020:

Keywords:

In recent years, deep learning has achieved remarkable success across various domains, particularly in tasks involving data structured as regular grids or sequences, such as images and text. However, scientific and real-world data frequently exhibit complex, nonEuclidean structures—including point clouds, meshes, graphs, and higher-order topological spaces—that challenge the assumptions of traditional neural network architectures. Topological Deep Learning (TDL) is an emerging field that extends deep learning methods to handle these rich and intricate data types. By incorporating topological constructs such as simplicial complexes, cell complexes, and hypergraphs, TDL enables the modelling of higher-order relationships, global dependencies, and qualitative spatial properties that are otherwise inaccessible to standard approaches. This talk will explore the foundational ideas behind TDL, the computational and theoretical challenges it presents, and its broad potential to enhance learning in diverse areas—from physics.

UNIVERSITY OF SAN FRANCISCO, USA

Email address: mhajij@usfca.edu

INTEGRATING HPC AND AI: A NEW PARADIGM FOR PREDICTING PROTEIN-LIGAND BINDING INTERACTIONS

NIU HUANG

Classification AMS 2020: 81V55, 92C05, 82B05

Keywords: Protein-ligand Interaction, High Performance Computing, Artificial Intelligence, Data Limitations

In the process of small molecule drug discovery, the prediction of protein-ligand interactions urgently demands enhancements in computational accuracy and efficiency, given its crucial role in identifying novel lead compounds for new targets. However, current artificial intelligence (AI) models are constrained by the scarcity of large, high-quality protein-ligand complex structures and binding data, which consequently impairs their generalization ability, limiting their effectiveness in real-world applications [1, 2].

The integration of physics-based high-performance computing (HPC) has emerged as a transformative approach to address these computational challenges. The exceptional computational power of modern HPC systems enables the generation of vast, high-quality datasets that serve as invaluable resources for both training sophisticated AI models and conducting comprehensive validation studies. This computational revolution allows researchers to explore molecular interaction spaces at unprecedented scales, generating the massive datasets required for robust machine learning model development.

Our research team has successfully redesigned SWDOCK, a specialized port of the widely-used UCSF DOCK software optimized for the Sunway TaihuLight supercomputer. This enhanced implementation enables ensemble docking with optimized compound database management, while achieving superior performance portability through an innovative emulator of the Sunway architecture. The optimized version, SWDOCKP², demonstrates substantial speedup improvements over the original SWDOCK implementation on Sunway CPUs and significantly outperforms UCSF DOCK on traditional X86 CPU architectures. The computational capabilities of SWDOCKP² are particularly impressive, enabling the docking of 1 trillion ligands to 8 different protein receptors within a single day. This unprecedented throughput allows researchers to explore chemical space at scales previously considered computationally intractable. Such massive screening capabilities are essential for identifying rare but highly potent lead compounds that might be missed by smaller-scale computational approaches [3, 4].

We have developed BindingNet v2, an extensively expanded dataset comprising 689,796 modeled protein-ligand binding complexes spanning 1,794 distinct protein targets. This comprehensive dataset was constructed using an enhanced template-based modeling workflow evolved from BindingNet v1, incorporating advanced pharmacophore and molecular shape similarity algorithms to improve complex

prediction accuracy. The template-based approach leverages known structural information to predict binding modes for novel ligands, significantly expanding the available training data for machine learning models. The effectiveness of BindingNet v2 in binding pose generation has been rigorously evaluated, demonstrating significantly improved generalization ability of the Uni-Mol model for novel ligands not present in the training data. The success rate on the challenging PoseBusters dataset increased dramatically from 38.55% when using the PDBbind dataset alone to 64.25% when augmented with BindingNet v2. This substantial improvement highlights the critical importance of large, diverse training datasets for developing robust predictive models [2].

We're continuing to explore how HPC and AI work together to enable more accurate, efficient calculations of molecular interactions. This integrated approach points to ways to overcome current data limitations and steadily improve AI models' predictive power for drug discovery. Pairing massive computational resources with advanced machine learning algorithms is changing how we approach computational drug discovery. What's more, developing AI architectures tailored for HPC environments will make molecular interaction predictions even more efficient. These advances could speed up the drug discovery timeline, cut development costs, and ultimately get new therapies to patients faster and more effectively.

REFERENCES

- [1] Yang, J.; Shen, C.; Huang, N. Predicting or Pretending: Artificial Intelligence for Protein-Ligand Interactions Lack of Sufficiently Large and Unbiased Datasets. *Front. Pharmacol.*, 2020, 11.
- [2] Zhu, H.; Li, X.; Chen, B.; Huang, N. Augmented BindingNet Dataset for Enhanced Ligand Binding Pose Predictions Using Deep Learning. *npj Drug Discov.*, 2025, 2 (1), 1.
- [3] Xu, K.; Zhang, J.; Duan, X.; Wan, X.; Huang, N.; Schmidt, B.; Liu, W.; Yang, G. Redesigning and Optimizing UCSF DOCK3.7 on Sunway TaihuLight. *IEEE Transactions on Parallel and Distributed Systems*, 2022, 33 (12), 4458–4471.
- [4] Xiaohui Duan, Cheng Shen, Gaowei Chen, Shanshan Wu, Yizhen Wang, Qiancheng Xia, Yizhen Chen, Qixin Chang, Zekun Yin, Yibing Shan, Guangwen Yang, Weiguo Liu, Niu Huang. Trillion Ligands per Day: Performance-Portable Virtual Screening via Compound Database Optimization and Multi-Target Docking. *The International Conference for High Performance Computing, Networking, Storage, and Analysis.*, 2025.

NO. 7 SCIENCE PARK ROAD, ZHONGGUANCUN LIFE SCIENCE PARK, CHANGPING DISTRICT, BEIJING 102206, P. R. CHINA

Email address: huangniu@nibs.ac.cn

DECODING GRAPH NEURAL NETWORKS: AN OPTIMIZATION PERSPECTIVE

WEI HUANG

Classification AMS 2020: 68T05, 68R10, 90C26

Keywords: Graph Neural Network, Optimization, Over-smoothing, Graph Convolution, DropEdge, Neural Tangent Kernel

Graph Neural Networks (GNNs) have emerged as a powerful framework for modeling graph-structured data and have achieved remarkable success in various domains such as social network analysis, recommendation systems, and molecular property prediction. However, several fundamental challenges remain, including the *over-smoothing phenomenon*—where node representations become indistinguishable as layers deepen—trainability degradation in deep architectures, and a lack of quantitative understanding of how graph structural information benefits learning.

In this talk, we present two complementary research directions addressing these challenges from both optimization and generalization perspectives.

The first direction investigates the *optimization dynamics of deep GNNs* through the lens of the Graph Neural Tangent Kernel (GNTK) framework [1]. Our analysis reveals that as the network depth increases, the trainability of GNNs—defined as the ability to optimize via gradient descent—suffers an exponential decline, a limitation that traditional residual connections can only partially alleviate. To fundamentally counteract this decay, we propose the *Critical DropEdge* method, which is a connectivity-aware and graph-adaptive edge sampling strategy. This approach not only mitigates the trainability degradation but also leads to consistent performance improvements across both infinite-width and finite-width GNNs.

The second direction focuses on the *feature learning and generalization properties of GNNs* [2]. By adopting a signal-noise data model and comparing two-layer graph convolutional networks (GCNs) with multilayer perceptrons (MLPs), we quantitatively analyze how graph convolution leverages the inherent structural information in graphs. Our findings demonstrate that GNNs significantly prioritize learning meaningful signals while suppressing noise memorization, thereby enlarging the regime of low test error compared to MLPs. Specifically, we show that graph convolution achieves a D^{q-2} -fold advantage in generalization, where D denotes a node's expected degree and q is the degree of the ReLU activation function (with $q > 2$). This provides a theoretical and empirical foundation for the observed superior generalization of GNNs in practical applications.

Together, these studies provide a unified view on why and how deepening GNNs remains challenging, as well as when and why GNNs outperform traditional neural architectures in learning from graph-structured data. Our results not only deepen the theoretical understanding of GNN optimization and generalization but also offer practical guidelines for designing more effective graph learning algorithms.

REFERENCES

- [1] Wei Huang, Yayong Li, Weitao Du, Jie Yin, Richard Yi Da Xu, Ling Chen, and Miao Zhang. Towards Deepening Graph Neural Networks: A GNTK-based Optimization Perspective. In *International Conference on Learning Representations (ICLR)*, 2023.
- [2] Wei Huang, Yuan Cao, Haonan Wang, Xin Cao, and Taiji Suzuki. Quantifying the Optimization and Generalization Advantages of Graph Neural Networks Over Multilayer Perceptrons. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2025.

7-CHŌME-3-1 HONGŌ, BUNKYO CITY, TOKYO

Email address: wei.huang.vr@riken.jp

DIRTY LIMIT THEOREMS AND APPLICATIONS

STEPHAN F. HUCKEMANN

Classification AMS 2020: 60D05, 53C20, 62H11

Keywords: Stratified spaces, Fréchet means, smeariness, stickiness, bootstrap

1. INTRODUCTION

Fréchet means are generalizations of the expected value to manifolds and stratified spaces. Their asymptotic rates, however, may deviate from those of their Euclidean kin. For instance, faster rates let the sample mean "stick" to the population mean, whereas slower rates let it appear to be "smearily" spread out. We sketch some relationships between geometry and statistics via dirty (sticky, smeary) asymptotic rates of Fréchet means, and illustrate the relevance of these results for statistical testing for geometric shape.

Some typical data sets and research questions.

- (1) Weekly mean wind directions taken at Basel over the periods (1990-1994, 2000-2004, 2010-214): Did the mean direction change over time hinting toward climate change [11]?
- (2) Fragments of DNA regions from different species: Build, average and compare phylogenetic trees [14].

In quantile based asymptotic Euclidean statistics one has random variables $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} X$ on \mathbb{R}^D independent from random variables $Y_1, \dots, Y_m \stackrel{\text{i.i.d.}}{\sim} Y$ on \mathbb{R}^D , a hypothesis H_0 : e.g. $\mathbb{E}[X] = \mathbb{E}[Y]$, and a statistic $T_{n,m} = T(X_1, \dots, X_n, Y_1, \dots, Y_m) \in \mathbb{R}^k$ with known asymptotic distribution $T_{n,m} \xrightarrow{D} T$ ($n, m \rightarrow \infty$). Building a confidence region C with $\mathbb{P}\{T \in C\} \geq 1 - \alpha$ for a level $0 < \alpha < 1$, reject H_0 at level α if $T_{n,m} \notin C$.

In bootstrap based asymptotic Euclidean statistics one resamples suitably (!) from $X_1, \dots, X_n, Y_1, \dots, Y_m$ to obtain a confidence region $C_{n,m}$ (e.g. [7]).

2. PROBABILTY ON METRIC SPACES

For random variables $X_1, X_2, \dots \stackrel{\text{i.i.d.}}{\sim} X$ on a metric space (M, d) , define the population and sample *Fréchet functions*, respectively,

$$F(p) := \frac{1}{2} \mathbb{E}[d(X, p)^2], \quad F_n(p) := \frac{1}{2n} \sum_{j=1}^n d(X_j, p)^2,$$

and the sets of population and sample *Fréchet means*, respectively,

$$E(X) := \operatorname{argmin}_{p \in M} F(p), \quad E_n := \operatorname{argmin}_{p \in M} F_n(p).$$

Theorem 2.1 (Ziezold, Bhattacharya-Patrangenaru). If (M, d) is separable and $\mathbb{E}(d(X, p)) < \infty$ for some $p \in M$, then [20] (Ziezold strong consistency):

$$\bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} E_k} \subset E \text{ a.s. ,}$$

if additionally $E \neq \emptyset$ and (M, d) is Heine-Borel then [3] (Bhattacharya- Patrangenaru strong consistency): $\forall \epsilon > 0$ a.s. $\exists n = n(\epsilon) \in \mathbb{N}$ such that

$$\bigcup_{k=n}^{\infty} E_k \subset \{p \in M : d(E, p) \leq \epsilon\}.$$

Statistics on stratified spaces. Additionally to seperability of (M, d) assume

- (1) $E = \{\mu\}$ is unique,
- (2) $\exists C^2$ open finite-dimensional manifold U with $\mu \in U \subset M$,
- (3) \exists chart ϕ of U at μ such that $x \mapsto d(X, \phi^{-1}(x))^2$ is a.s. C^2 ,
- (4) $\exists \mathbb{E}(\text{grad}_2 d(X, \mu)^2), \text{cov}(\text{grad}_2 d(X, \mu)^2) =: C_\phi$,
- (5) $\text{Hess}F_n(\tilde{\mu}_n) \xrightarrow{\text{a.s.}} \text{Hess}F(\mu) =: H_\phi$ a.s. for all $\tilde{\mu}_n \xrightarrow{\text{a.s.}} \mu$, exists and is of full rank.

Theorem 2.2 ([4, 2]). Under these assumption let E_n be a Bhattacharya- Patrangenaru strongly consistent estimator of $\{\mu\}$. Then for every measurable choice $\mu_n \in E_n$,

$$\sqrt{n}(\phi(\mu_n) - \phi(\mu)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, H_\phi^{-1} C_\phi H_\phi^{-1}).$$

For limitations due to cut loci in general see [6]. Here is a special one:

Theorem 2.3 (First CLT on the Circle, [16, 8]). Consider $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} X$ on $S^1 = [-\pi, \pi]/\sim$ with local density f near $-\pi \sim \pi$ and Fréchet means $\mu = 0$ (population), μ_n (sample). Then, $0 < f(-\pi) < \frac{1}{2\pi}$ is possible with

$$\text{Hess}F_n(\mu_n) \xrightarrow{\text{a.s.}} 1 \neq 1 - 2\pi f(-\pi) = \text{Hess}F(0).$$

3. SMEARINESS

Theorem 3.1 (Smeary CLT on the Circle, [8, 12]). Under the assumptions of the CLT on the circle, even $f(-\pi) = \frac{1}{2\pi}$ is possible, i.e. $\text{Hess}F(0) = 0$ with

$$\sqrt{n} \cdot \text{sign}(x_n) 2\pi G(|x_n|) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \text{var}[X])$$

where for some $\delta > 0$ and a nonnegative, strictly convex, and \mathcal{C}^1 function $G: [0, \delta) \rightarrow \mathbb{R}_{\geq 0}$ that fulfills $G(0) = G'(0) = 0$,

$$\begin{aligned} f(-\pi + \epsilon) &= \frac{1}{2\pi} - G'(\epsilon) + o(G'(\epsilon)) \\ f(\pi - \epsilon) &= \frac{1}{2\pi} - G'(\epsilon) + o(G'(\epsilon)) \end{aligned}$$

forall $\epsilon \in [0, \delta]$.

Rates $\psi(n)$ such that $\psi(n)(\phi(\mu_n) - \phi(\mu))$ converges to a nondegenerate distribution are called *smeary* if $\psi(n)/\sqrt{n} \rightarrow 0$. There are examples of

- *power smeariness* $\mu_n \sim n^{-\frac{1}{2(k+1)}}$ $\forall k \in \mathbb{N}$ on the circle [8],
- *2nd (k = 2) order smeariness* on arbitrary spheres [7],

- arbitrary power smeariness $k \in \mathbb{R}$ on the circle [12],
- log smeariness on the circle [12]), e.g. for arbitrary $r > 0$:
 - $\mu_n \xrightarrow{a.s.} 0$,
 - $(\log \sqrt{n})^{1/r} \mu_n \xrightarrow{\mathcal{D}} \frac{1}{2} \cdot d\delta_{-1} + \frac{1}{2} \cdot d\delta_1$,
 - $r (\log \sqrt{n})^{(1+r)/r} \left(\mu_n - \frac{\text{sign}(\mu_n)}{(\log \sqrt{n})^{1/r}} \right) \xrightarrow{\mathcal{D}} \text{sign}(Z) \cdot \log |Z|$ with $Z \sim \mathcal{N}(0, \sigma^2)$,
- log smeariness on Wasserstein–2 space [13].

For the CLT by [7] on arbitrary manifolds featuring arbitrary power smeariness of order $2 < k \in \mathbb{N}$, only require uniqueness: $\mu_n \xrightarrow{\mathbb{P}} \mu$, existence of second moments and first moments of k -th derivatives, and Donsker conditions, e.g. from [19, Secns 5 & 19.].

[5] discriminates:

- topological = cut locus smeariness (if cut locus has codimension 1, e.g. tori, real projective spaces),
- geometric smeariness e.g. on \mathbb{S}^m without a polar cap for $m \geq 5$, cap \rightarrow hemisphere as $m \rightarrow \infty$.

[18] conjectured that every compact manifold features smeariness and showed this for directional smeariness.

4. STICKINESS

Definition 4.1. A random variable X on a metric space (M, d) with sample means μ_n , $n \in \mathbb{N}$ is sample sticky to a subset $S \subset M$ if

$$\exists \text{ random } Z \text{ in } \mathbb{N} \text{ such that all } \mu_n \in S \text{ a.s. for all } n \geq Z.$$

Theres are sample sticky distributions on

- the spider for $S = \{0\}$ and the open book for $S = \text{spine}$ [9],
- the kale for $S = \{0\}$ [10],
- BHV_N space for phylogenetic trees, for any lower dimensional stratum $S \subset \text{BHV}_N$, of dimension $0 \leq m \leq N - 3$, e.g. [1].

This is a dead end for (some) nonparametric asymptotic statistics.

Definition 4.2. For a CAT(κ) space (M, d) , $\kappa \in \mathbb{R}$, $\sigma = [\gamma] \in \Sigma_x M$ (space of directions, γ is a unit speed geodesic starting at x), $x \in M$, let

$$\nabla_\sigma f(x) := \frac{d}{dt} \Big|_{t=0} f \circ \gamma(t)$$

if it exists and is well-defined. Then call X directionally sticky at its mean $\mu \in M$, if

$$\nabla_\sigma F(\mu) > 0 \quad \forall \sigma \in \Sigma_x M.$$

Theorem 4.3 (CLT for directional stickiness, [15]). $\nabla_\sigma F(x) = \mathbb{E}[\phi_{x,\sigma} \circ X]$ for all $x \in M, \sigma \in \Sigma_x M$, with $\phi_{x,\sigma}(z) := -d(x, z) \cdot \cos \angle_x(\sigma, z)$, the pull of z in direction σ at x .

Moreover, on a finite dimensional CAT(κ) space, with a measurable selection μ_n of sample means, the empirical process

$$\sqrt{n} (\nabla F_n(\mu) - \nabla F(\mu))$$

converges weakly to a Gaussian process with zero mean and covariance

$$(\mathbb{E}[\phi_{\mu,\sigma}(X) \phi_{\mu,\tau}(X)])_{\sigma,\tau \in \Sigma_\mu}.$$

This allows testing [14] for:

- presence of stickiness,
- discriminating two sticky distributions by their directional sticky covariance.

5. UNIFIED FRAMEWORK

Definition 5.1 ([17]). For $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} X$ with second moments, population mean μ and sample mean μ_n , the variance modulation is

$$\mathfrak{m}_n^2 := \frac{n \mathbb{E}[d(\mu, \mu_n)^2]}{\mathbb{E}[d(\mu, X)^2]}.$$

Remark 5.2. X is \mathfrak{m}^2 sticky at $S = \{\mu\}$ if $\mathfrak{m}_n^2 \rightarrow 0$.

- (1) $M = \text{Euclidean} \Rightarrow \mathfrak{m}_n^2 \equiv 1$ (hence quantile based asymptotic statistics possible),
- (2) $\text{smeary} \Rightarrow \mathfrak{m}_n^2 \rightarrow \infty$,
- (3) \exists also
 - finite sample smeary: $\mathfrak{m}_n^2 > 1 \forall 2 \leq n \in \mathbb{N}$ (e.g. most of the wind data),
 - finite sample sticky: $\mathfrak{m}_n^2 < 1 \forall 2 \leq n \in \mathbb{N}$ (typical for phylogenetic tree data),
then, only bootstrap based asymptotic statistics are meaningful.

REFERENCES

- [1] Barden, D., H. Le, and M. Owen (2018). Limiting behaviour of Fréchet means in the space of phylogenetic trees. *Annals of the Institute of Statistical Mathematics* 70(1), 99–129.
- [2] Bhattacharya, R. and L. Lin (2017). Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proceedings of the American Mathematical Society* 145(1), 413–428.
- [3] Bhattacharya, R. N. and V. Patrangenaru (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds I. *The Annals of Statistics* 31(1), 1–29.
- [4] Bhattacharya, R. N. and V. Patrangenaru (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds II. *The Annals of Statistics* 33(3), 1225–1259.
- [5] Eltzner, B. (2022). Geometrical smeariness – a new phenomenon of Fréchet means. *Bernoulli* 28(1), 239–254.
- [6] Eltzner, B., F. Galaz-García, S. F. Huckemann, and W. Tuschmann (2021). Stability of the cut locus and a central limit theorem for Fréchet means of Riemannian manifolds. *Proceedings of the American Mathematical Society* 149(9), 3947–3963.
- [7] Eltzner, B. and S. F. Huckemann (2019). A smeary central limit theorem for manifolds with application to high-dimensional spheres. *Ann. Statist.* 47(6), 3360–3381.
- [8] Hotz, T. and S. Huckemann (2015). Intrinsic means on the circle: Uniqueness, locus and asymptotics. *Annals of the Institute of Statistical Mathematics* 67(1), 177–193.
- [9] Hotz, T., S. Huckemann, H. Le, J. S. Marron, J. Mattingly, E. Miller, J. Nolen, M. Owen, V. Patrangenaru, and S. Skwerer (2013). Sticky central limit theorems on open books. *Annals of Applied Probability* 23(6), 2238–2258.
- [10] Huckemann, S., J. C. Mattingly, E. Miller, and J. Nolen (2015). Sticky central limit theorems at isolated hyperbolic planar singularities. *Electronic Journal of Probability* 20(78), 1–34.
- [11] Hundrieser, S., B. Eltzner, and S. Huckemann (2024a). Finite sample smeariness of Fréchet means with application to climate. *Electron. J. Statist.* 18(2), 3274–3309.
- [12] Hundrieser, S., B. Eltzner, and S. F. Huckemann (2024b). A lower bound for estimating Fréchet means. arXiv 2402.12290.
- [13] Kroshnin, A. (2021). *Inside and around Wasserstein barycenters*. Ph. D. thesis, Université de Lyon; Kharkevich Institute.
- [14] Lammers, L., T. M. Nye, and S. F. Huckemann (2024). Statistics for phylogenetic trees in the presence of stickiness. arXiv 2407.03977.
- [15] Lammers, L., D. T. Van, and S. F. Huckemann (2023). Sticky flavors. *arXiv preprint arXiv:2311.08846*.

- [16] McKilliam, R. G., B. G. Quinn, and I. V. L. Clarkson (2012). Direction estimation by minimum squared arc length. *IEEE Transactions on Signal Processing* 60(5), 2115–2124.
- [17] Pennec, X. (2019). Curvature effects on the empirical mean in Riemannian and affine manifolds: a non-asymptotic high concentration expansion in the small-sample regime. *arXiv preprint arXiv:1906.07418*.
- [18] Tran Van, D., B. Eltzner, and S. F. Huckemann (2021). Smeariness begets finite sample smeariness. In *Geometric Science of Information 2021 proceedings*, pp. 29–36. Springer.
- [19] van der Vaart, A. (2000). *Asymptotic statistics*. Cambridge Univ. Press.
- [20] Ziezold, H. (1977). Expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Transaction of the 7th Prague Conference on Information Theory, Statistical Decision Function and Random Processes A*, 591–602.

FELIX-BERNSTEIN-INSTITUTE FOR MATHEMATICAL STATISTICS IN THE BIOSCIENCES, GEORGIA AUGUSTA
UNIVERSITY GOETTINGEN, GOLDSCHMIDTSTRASSE 7, D-37077 GOETTINGEN, GERMANY

Email address: huckeman@math.uni-goettingen.de

NONLINEAR REGRESSION WITH REAL ALGEBRAIC VARIETIES AND THEIR TOPOLOGY

STEPHAN KLAUS

Keywords: topological data analysis, algebraic variety, nonlinear regression, Betti numbers, lemniscate

MSC: 14P05, 14P25, 55N31, 57R12, 57R19, 62R40, 68T09

1. TOPOLOGICAL DATA ANALYSIS

A central problem in *topological data analysis* is the approximation of a *point data cloud* $X := \{x^1, x^2, \dots, x^N\}$, $x^\alpha \in \mathbb{R}^n$, (where the x^α are distinct, thus N is the size of X and n is its *ambient dimension*) by a subset $M \subset \mathbb{R}^n$ of a specific topological type and to compute topological invariants of M out of X .

The standard method is to construct M as the *Vietoris-Rips complex* of the point cloud X , $K_d(X) := \{\sigma \subset \{1, 2, \dots, N\} | \text{diam}\{x^\alpha | \alpha \in \sigma\} < d\}$ for a given *distance parameter* $d \in \mathbb{R}^+$. Here the choice of d is essential as $K_d(X) = X$ is 0-dimensional for d too small (smaller than the minimum of all distances in X) and $K_d(X)$ is the N -dimensional standard simplex if d is too large (larger than the diameter of X). Enlarging the parameter $d < d'$ induces an inclusion of simplicial complexes $K_d(X) \subset K_{d'}(X)$. As a topological invariant one usually takes the simplicial homology of $K_d(X)$. Analysing the dependence on d leads to the *persistant homology* of X . As the number of simplices in $K_d(X)$ can be very large the computational cost is in general high.

2. APPROXIMATION BY MANIFOLDS OR VARIETIES

Here we aim to approximate X by a nice space, e.g. an embedded manifold M . In order to avoid overfitting, we have to formulate this task as an optimization problem: 1) minimize the distance of X to M , and at the same time, 2) bound the "complexity" of M from above. Here the "complexity" can be defined in different ways, e.g. as the number of parameters to define M , or curvature conditions for M with its induced Riemannian metric.

More generally we can use for M the zero set of a function f in a *finite dimensional function space* F with a given basis of functions $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, 2, \dots, m$ (the f_i do not need to be linear). Thus we consider the zero set $V_f := f^{-1}(0)$ of the function $f := \sum_{i=1}^m a_i f_i$ for given parameters $a_i \in \mathbb{R}$.

As a standard example, we can consider the space F_r of polynomials of degree at most r (with $r \in \mathbb{N}$ fixed). A basis of F_r is given by the monomial functions $f_{i_1, \dots, i_n} := x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$ with $i_1 + i_2 + \dots + i_n \leq r$. In particular, F_r has dimension $m = \binom{n+r}{r}$. By definition V_f is then an affine real algebraic variety (where we do not assume irreducibility in the definition of a variety).

As V_f does not change if we replace f by λf with $\lambda \in \mathbb{R} - \{0\}$, we can introduce the *normalization condition* $\sum_{i=1}^m a_i^2 = 1$.

3. REGRESSION WITH NON-LINEAR FUNCTIONS

The distance of X to V_f is given by the *quadratic approximation error*

$$\delta := \sum_{\alpha=1}^m f(x^\alpha)^2 = \sum_{\alpha=1}^N \left(\sum_{i=1}^m a_i f_i(x^\alpha) \right)^2 = \sum_{i,j=1}^m \delta_{i,j} a_i a_j = a^t \Delta a$$

with $\delta_{i,j} := \sum_{\alpha=1}^N f_i(x^\alpha) f_j(x^\alpha)$. Here $a := (a_1, a_2, \dots, a_m)$ denotes the *parameter vector* and $\Delta := (\delta_{i,j})$ is the *error matrix* associated to the basis f_i and X . The optimization problem amounts then to minimize δ within F . In particular, the fixed function space F represents the upper bound on the complexity of V_f . As δ is a homogeneous quadratic form in a , this is equivalent to find a *minimal eigen-value with its eigen-vector* of the symmetric $m \times m$ matrix Δ .

4. MANIFOLDS AND ALGEBRAIC VARIETIES

There are strong relations between smooth manifolds and real algebraic varieties. First of all, the *Regular Value Theorem* [6] implies that for f smooth and 0 a regular value of f , the variety $V_f \subset \mathbb{R}^n$ is a smooth submanifold of codimension 1. By the *Theorem of Sard*, regular values are dense (singular values have Lebesgue measure 0), thus for generic smooth f we get a smooth manifold V_f . In particular, we can smooth out singularities in a variety by changing f to $f + \epsilon$ ("singularities have probability 0"). Moreover, over \mathbb{R} one polynomial is enough to get a variety because the condition $f_{(k)} = 0$ for $k = 1, \dots, s$ is equivalent to $f := f_{(1)}^2 + f_{(2)}^2 + \dots + f_{(s)}^2 = 0$ (of course, this makes 0 a singular value of f).

In general singularities allow more flexibility concerning the topological type of an approximation for X . As an example, imagine a data cloud in \mathbb{R}^2 with the rough shape of the infinity symbol. There is a singular algebraic curve of degree 4, the lemniscate, which is defined by the function $f(x, y) := ((x-a)^2 + y^2)((x+a)^2 + y^2) - a^2$ and has focal points $(\pm a, 0)$. As it contains a singularity in $(0, 0)$, for a given data cloud it appears as a minimum of δ with probability zero only. If the singularity is smoothed out (by some ϵ as above), the lemniscate splits into either one or two closed lines (deformed circles) which have a different topological type than that one would expect from the shape "infinity" of the data cloud.

Here is a solution of this problem: Instead of strict optimization we could use the generalization "regression with tolerance" by considering V_g for a neighborhood $g \in U$ of the minimal f . The union of the V_g gives then an n -dimension thickening \hat{V} of V_f in \mathbb{R}^n which can contain also a singular V_g . If there is only one singular g in U , one can show that the thickening \hat{V} has the same homotopy type as V_g using Morse-Theory. Even more general, δ gives a Gaussian probability distribution on the unit sphere in the parameter space \mathbb{R}^m defined by the normalization condition which is proportional to $\exp(-\delta)$. This can be interpreted as a probability distribution on the space of all algebraic varieties with degree bounded by r .

Last but not least, there holds the deep *Theorem of Nash-Tognoli* [5]: Every compact smooth manifold is diffeomorphic to a smooth real algebraic variety.

5. ON THE TOPOLOGY OF ALGEBRAIC VARIETIES

It is an obvious question how we can compute topological invariants of a real algebraic variety from its representation as the zero set of a polynomial. However this is a very difficult problem in general.

For example, for real algebraic curves in \mathbb{R}^2 there is *Harnack's Theorem*: The number b_0 of path components is bounded by the degree r as $b_0 \leq 1 + \binom{r-1}{2}$. The closer investigation of the topology of real algebraic curves and surfaces (e.g. relative positions of the branches) is just the first part of *Hilbert's 16th Problem*. For curves there has been a lot of progress in the second half of the past century by Thom, Arnold, Gudkov and others, but for surfaces only very few results are known. In fact, the huge number of examples for algebraic surfaces with very interesting shapes is astonishing, see the picture galleries in [3] and [4] and on the website of Imaginary [1].

Concerning topological properties of algebraic varieties, we report the following results from [5]. Here we also include the *complexification* and/or *projectivization* and use the following notation, where $p_i \in \mathbb{R}[x_1, x_2, \dots, x_n]$, $i = 1, \dots, s$ (of course, the polynomials p_i have to be homogeneous in the case of projective varieties):

$$\begin{aligned} V &:= \{x \in \mathbb{R}^n | p_i(x) = 0 \forall i\}, & \mathbb{C}V &:= \{x \in \mathbb{C}^n | p_i(x) = 0 \forall i\}, \\ PV &:= \{x \in \mathbb{R}P^n | p_i(x) = 0 \forall i\}, & \mathbb{C}PV &:= \{x \in \mathbb{C}P^n | p_i(x) = 0 \forall i\}. \end{aligned}$$

The *Theorem of Milnor* on the sum of Betti numbers of V , where r denotes the maximal degree of the polynomials p_i : $\sum_{i=0}^n b_i(V) \leq r(2r-1)^{n-1}$

The *Theorem of Floyd-Thom* compares the sum of $\mathbb{Z}/2$ -Betti numbers of the real and complex projective varieties:

$$\sum_{i=0}^n b_i(PV; \mathbb{Z}/2) \leq \sum_{i=0}^n b_i(\mathbb{C}PV; \mathbb{Z}/2)$$

The *Theorem of Sullivan* states that the Euler-characteristics of the real and complex projective varieties have the same parity: $(PV) \equiv \chi(\mathbb{C}PV) \pmod{2}$

These three theorems can be considered as a generalization of the fact that a real polynomial $p(z)$ has less real roots than its degree and that the number of non-real roots is even (as they appear in complex conjugate pairs).

The best understood class of varieties (from the topological point of view) is formed by *complex complete intersections*: Here the projective hypersurfaces of the polynomials p_i , $i = 1, \dots, s$ (with degrees r_i) have to be smooth and have to intersect transversally. In this case, the following strong results hold true [2]:

The *Theorem of Thom* that the real diffeomorphism type of the $2(n-s)$ -dimensional manifold $\mathbb{C}PV$ depends only on the integers n and r_1, r_2, \dots, r_s .

The *Lefschetz Hyperplane Theorem* which implies that the integral cohomology rings of $\mathbb{C}PV$ and $\mathbb{C}P^{n-s}$ are isomorphic with the exception of the middle dimension.

The *Theorem of Hirzebruch* which computes the total Chern class:

$$c(\mathbb{C}PV) = (1+z)^{n+1} / \prod_{i=1}^s (1+r_i z)$$

with $z \in H^2(\mathbb{C}P^n)$ the generator which satisfies $i^* z^n [\mathbb{C}PV] = r_1 r_2 \cdots r_s$.

Acknowledgements The author would like to thank Stephan Huckemann and Wilderich Tuschmann for helpful discussions.

REFERENCES

- [1] Website of Imaginary open mathematics, a non-profit organization for open and interactive mathematics: www.imaginary.org
- [2] Fang Fuquan, Stephan Klaus: *Topological classification of 4-dimensional complete intersections*, Manuscripta Math. 90 (1996), no. 2, p. 139–147.
- [3] Stephan Klaus: *Möbius strips, knots, pentagons, polyhedra, and the SURFER software*, In: Decker, Wolfram (ed.) et al., Singularities and computer algebra. Festschrift for Gert-Martin Greuel on the occasion of his 70th birthday. Based on the conference, Lambrecht (Pfalz), Germany, June 2015. Springer (2017), p. 161–172.
- [4] Stephan Klaus: *Cutting, gluing, squeezing, and twisting: visual design of real algebraic surfaces*, In: Sriraman, B. (eds) Handbook of the Mathematics of the Arts and Sciences, Springer (2021), p. 863–877.
- [5] Janos Kollar, *The topology of real algebraic varieties*, Current Developments in Mathematics (2000), International Press, p. 197-231.
- [6] John Milnor, *Topology from the Differentiable Viewpoint*, Princeton Landmarks (1993), Princeton University Press, 78 pages

MATHEMATISCHES FORSCHUNGSIINSTITUT OBERWOLFACH, GERMANY

Email address: klaus@mfo.de

A PHYSICIST'S VIEW ON PARTIAL 2D SHAPE COMPARISON

PATRICE KOEHL

Classification AMS 2020: 49Q22, 82M30

Optimal unbalanced transport; Mean field approximation; 3D shape comparison

1. DISCRETE OPTIMAL TRANSPORT

The optimal transport (OT) problem has a long history in mathematics and statistics. Originally introduced by Monge to formalize the problem of moving earth from one area to another as a linear optimization problem [1], it has been expanded as a means to handle allocation resources and therefore has become integral to operation research. OT, however, is not specific to applied problems. At its core, it can be seen as an option for computing the distance and a correspondence between two probability distributions (for review, see [2]). Finding such distance and mapping between probability measures is of relevance to most, if not all data science disciplines, and as such applications of OT have exploded in the recent years, in domains such as signal and image processing, machine learning, computer vision and image analysis , linguistics, differential geometry, and geometric shape matching. We propose a novel approach to solving a generalized version of the OT problem, which we refer to as the discrete variable mass optimal transport (VMOT) problem, using techniques adapted from statistical physics.

In the discrete variable mass optimal transport setting we work with two sets of points S_1 and S_2 , of size N_1 and N_2 . In 2D shape comparison, these are vertices in two shapes to be compared. Each point k in S_1 (respectively S_2) is assigned a “mass” $\rho_1(k)$ (respectively $\rho_2(k)$). The cost of transport between S_1 and S_2 is encoded by a positive matrix $C(k, l)$ with $k \in \{1, \dots, N_1\}$ and $l \in \{1, \dots, N_2\}$. Our goal is to find a transportation plan G , with $G(k, l)$, the amount of mass transported between $k \in S_1$ and $l \in S_2$, to minimize the total transportation cost U :

$$(1.1) \quad U = \sum_{kl} C(k, l)G(k, l) + \tau_1 \sum_k \left(\frac{m_1(k)}{\rho_1(k)} \right)^2 + \tau_2 \sum_l \left(\frac{m_2(l)}{\rho_2(l)} \right)^2,$$

where τ_1 and τ_2 are parameters. We require that $G(k, l)$ be positive for each k and l . In this setting, the mass assignments ρ_1 and ρ_2 are seen as perturbations of yet to be determined balanced mass measures m_1 and m_2 :

$$m_1(k) = \sum_l G(k, l); \quad m_2(l) = \sum_k G(k, l).$$

Here $m_1(k)$ is the actual amount of mass transported from point k in S_1 by the transport plan and $m_2(l)$ is the corresponding actual amount of mass received by point l in S_2 . These masses are variables of the system, thereby allowing for mass creation or deletion. Finally, we ensure that there is some mass transferred between S_1 and S_2 by imposing the constraint $\sum_{kl} G(k, l) = 1$.

From a statistical physics perspective, each state S of the variable mass OT system is identified with a transport plan G , vectors of masses \mathbf{m}_1 and \mathbf{m}_2 for the points in S_1 and S_2 , and its energy $U(S)$ is defined in equation 1.1. The collection of all feasible states S is a set that we note $\mathcal{S}(S_1, S_2)$. Note that this set is convex.

The probability distribution function for this system, $P(S)$, also referred to as the Gibbs distribution, is defined as:

$$P(S) = \frac{1}{Z_\beta(S_1, S_2)} e^{-\beta U(S)}.$$

In this equation, $\beta = 1/(k_B T)$ where k_B is the Boltzmann constant and T the temperature, and $Z_\beta(S_1, S_2)$ is the partition function computed over all states of the system. The partition function is given by

$$Z_\beta(S_1, S_2) = \int_{S \in \mathcal{S}(S_1, S_2)} e^{-\beta U(S)} d\mu_{12},$$

where $d\mu_{12}$ can be seen as the Lebesgue measure on $\mathcal{S}(S_1, S_2)$. Taking into account the constraints that define $\mathcal{S}(S_1, S_2)$,

$$\begin{aligned} Z_\beta(S_1, S_2) &= \int_0^1 \prod_{kl} dG(k, l) \int_{-\infty}^{+\infty} \prod_k dm_1(k) \int_{-\infty}^{+\infty} \prod_l dm_2(l) \\ &\quad e^{-\beta(\sum_{kl} C(k, l)G(k, l) + \sum_k \alpha_1(k)m_1^2(k) + \sum_l \alpha_2(l)m_2^2(l))} \\ &\quad \prod_k \delta\left(\sum_l G(k, l) - m_1(k)\right) \prod_l \delta\left(\sum_k G(k, l) - m_2(l)\right) \delta\left(\sum_{kl} G(k, l) - 1\right), \end{aligned}$$

where we have set $\alpha_1(k) = -\tau_1/\rho_1(k)^2$ and $\alpha_2(l) = -\tau_2/\rho_2(l)^2$. (both are constants). Note the limits of integrations. For the $G(k, l)$ values, they are set to be $\{0, 1\}$. For the masses $m_1(k)$ and $m_2(l)$, we set them unconstrained in \mathbb{R} , although we know that they are positive (as sums of elements of the transport plan G that are all positive), and smaller than 1 (from equation ??). Those constraints are enforced by the delta functions.

Using Fourier analysis, we can represent a delta function as an integral of an exponential, where the integration is usually performed along the real axis. Introducing new auxiliary variables $\lambda(k)$ and $\mu(l)$, with $(k, l) \in [1, N_1] \times [1, N_2]$, and x , integrating explicitly the variable $G(k, l)$, $m_1(k)$ and $m_2(l)$, and omitting the unessential normalization factors, the partition function can be written as,

$$Z_\beta(S_1, S_2) = \int_{-\infty}^{+\infty} \prod_k d\lambda(k) \int_{-\infty}^{+\infty} \prod_l d\mu(l) \int_{-\infty}^{+\infty} dx e^{-\beta F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu}, x)},$$

where we have defined F_β as

$$F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu}, x) = -x - \frac{1}{4} \sum_k \frac{\lambda^2(k)}{\alpha_1(k)} - \frac{1}{4} \sum_l \frac{\mu^2(l)}{\alpha_2(l)} - \frac{1}{\beta} \sum_{kl} \ln \left[\frac{1 - e^{-\beta(C(k, l) + \lambda(k) + \mu(l) + x)}}{\beta(C(k, l) + \lambda(k) + \mu(l) + x)} \right].$$

F_β is the functional free energy of the system. We note that it is a function of $N_1 + N_2 + 1$ variables, i.e. we have reduced the dimensionality of the problem from quadratic to linear in the number of points considered. The expected solution of the VMOT problem is then obtained by applying a saddle point approximation (SPA) to F_β . The SPA is

computed by looking for extrema of the effective free energy with respect to the variables $\lambda(k)$, $\mu(l)$, and x :

$$\frac{\partial F_\beta}{\partial \lambda(k)} = 0, \quad \frac{\partial F_\beta}{\partial \mu(l)} = 0, \quad \text{and} \quad \frac{\partial F_\beta}{\partial x} = 0.$$

After some rearrangements, those equations lead to the following system of equations:

$$(1.2) \quad \begin{aligned} \forall k, \quad \sum_l H(k, l) &= \frac{\lambda(k)}{2\alpha_1(k)}, \\ \forall l, \quad \sum_k H(k, l) &= \frac{\mu(l)}{2\alpha_2(l)}, \\ \sum_{kl} H(k, l) &= 1. \end{aligned}$$

where,

$$(1.3) \quad H(k, l) = \phi(\beta(C(k, l) + \lambda(k) + \mu(l) + x)), \quad \phi(t) = \frac{e^{-t}}{e^{-t} - 1} + \frac{1}{t}.$$

The function $\phi(t)$ is defined and continuous for all real values t (with the extension that $\phi(0) = 0.5$), monotonically decreasing over \mathbb{R} , with asymptotes $y = 1$ and $y = 0$ at $-\infty$ and $+\infty$, respectively.

Equations 1.2 define a non linear system of $N_1 + N_2 + 1$ equations, with the same number of variables. The following theorem shows that this system has a unique solution:

Theorem 1.1. *The Hessian of the effective free energy $F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu}, x)$ is negative definite. It is therefore strictly concave, and the system of equations 1.2 has a unique solution.*

Proof. See [3]. □

For a given value of the parameter β , the expected values $\bar{G}(k, l)$ that are solutions to the system of equations (1.2) form a transport plan G_β^{MF} between S_1 and S_2 that is optimal with respect to the free energy defined in (1.2). Similarly, $\bar{\mathbf{m}}_1 = \mathbf{m}_1^{MF}$ and $\bar{\mathbf{m}}_2 = \mathbf{m}_2^{MF}$ are the optimal masses transferred from, and received by S_1 and S_2 , respectively. We can associate to these values an optimal free energy F_β^{MF} and an optimum energy $U_\beta^{MF} = \sum_{k,l} G_\beta^{MF}(k, l)C(k, l) + \sum_k \alpha_1(k)(m_1^{MF}(k))^2 + \sum_l \alpha_2(l)(m_2^{MF}(l))^2$. Note that those two values are the mean field approximations of the exact free energy and internal energy defined in Eqn. ?? and ??, respectively. We now list important properties of U_β^{MF} and F_β^{MF} .

Proposition 1.2. *F_β^{MF} and U_β^{MF} are monotonic decreasing function of the parameter β . In addition, both converge to the optimal variable mass transport energy $d_u(S_1, S_2)$.*

Proof. See [3]. □

Theorem 1.1 and the proposition 1.2 above highlight a number of advantages of the proposed framework that rephrases the variable mass transport problem as a temperature dependent process. First, at each temperature the variable mass transport problem is turned into a strongly concave problem with a unique solution. This problem has a linear complexity in the number of variables, compared to the quadratic

complexity of the original problem. The concavity allows for the use of simple algorithms for finding a minimum of the effective free energy functional. We note also that equations 1.2 provide good numerical stability for computing the transport plan, because of the ratio of exponentials. Second, the convergence as a function of temperature is monotonic.

2. A SIMPLE APPLICATION TO 2D SHAPE COMPARISON

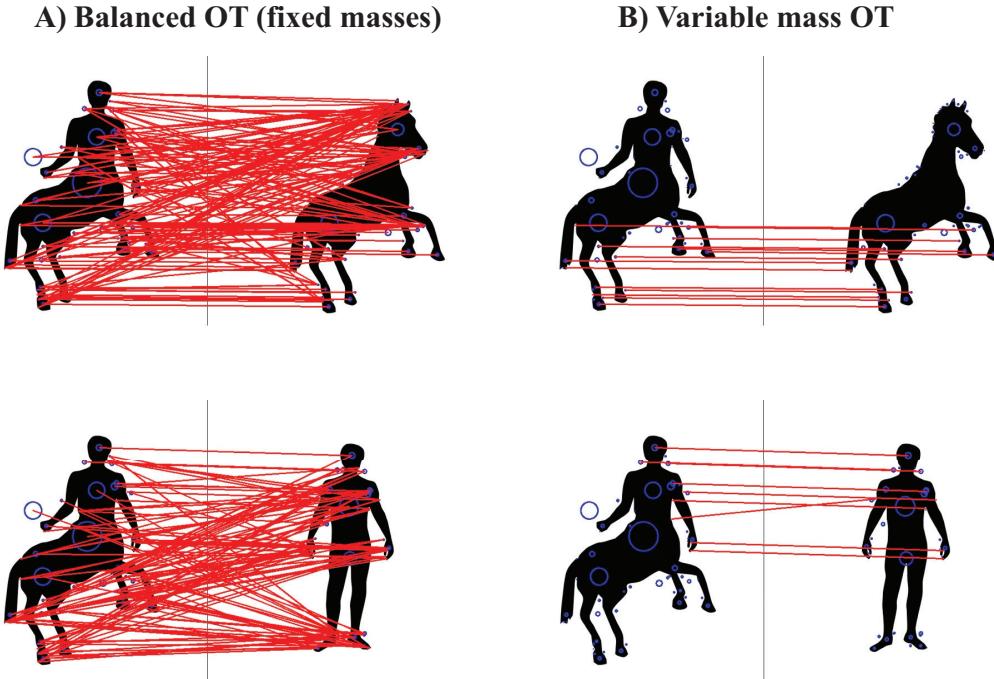


Figure 1: Visualization of global similarity (balanced OT distance, panel A) and partial similarity (variable mass OT distance, panel B) between of image of a horse and the image of a centaur (top) and between the image of the same centaur and the image of a man (bottom). See text for details.

We present some computational examples that illustrate the use of our framework for 2D image comparison. We used the publicly available Mythological Creatures 2D database [4]. We characterized each image by selecting a set of “keypoints” using the ORB procedure. ORB is an image feature detector and descriptor used for object recognition and image registration [5]. Within ORB, a keypoint is a pixel within the image that is expected to be significant, i.e. a signature feature of the image. The significance is defined from a local neighborhood of the pixel of interest, characterized by a vector of 64 features. The distance between two keypoints is then defined as the Euclidean distance between their feature vectors. We use this representation to compare images in our database. A pair of images is represented with their sets of keypoints, S_1 and S_2 , the cost matrix C between those keypoints, such that $C(k, l)$ between a keypoint k on image 1 and a keypoint l on image 2 is equal to the Euclidean distance between their feature vectors. The given masses $\rho_1(k)$ and $\rho_2(k)$ of the keypoints are set to 1.

In figure 1, we show the result of the comparisons of a man and a centaur, and of the same centaur and a horse. ORB keypoints for those images are shown as circles whose radii relate to the size of the corresponding meaningful neighborhoods. A line between two keypoints k and l from two different images indicates that the value of the optimized transport between k and l , $G^{MF}(k, l)$, is greater than a threshold, set to 0.01. On the left (panel A), we show the assignments based on regular optimal transport (see [6]). With the constraint that the transport need to be balanced, all keypoints on one image need to have correspondence with keypoints on the second image, leading to spurious assignments for example between the horse part of the centaur and the image of the human. In contrast, on the right of the figure (panel B) we illustrate the effectiveness of our variable mass OT framework in identifying only partial match. As expected, connections between the man and the centaur are found at the level of their heads and shoulder, while connections between the centaur and the horse are found at the level of their legs and tails.

As a final note, the variable mass OT problem considered in this paper assumes that the two sets of points considered are embedded in the same metric space, namely that we can build the cost matrix C that connect them. Situations in which the points lie in different, unregistered metric spaces have led to an extension to the optimal transport problem with the notion of Gromov-Wasserstein distances between metric measured spaces [7]. We have extended that the concept of finite temperature variable mass optimal transport into a finite temperature variable mass Gromov-Wasserstein transport formalism [8].

REFERENCES

- [1] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences et Mémoires de Mathématique et de Physique tirés des registres de cette Académie*, 1784, 666–704, 1781.
- [2] C. Villani. Topics in optimal transportation. *Graduate studies in mathematics*, 58, 2003.
- [3] P. Koehl, M. Delarue and H. Orland Physics approach to the variable-mass optimal-transport problem. *Phys. Rev. E*, 103, 012113, 2021.
- [4] A. Bronstein, M. Bronstein, A. Bruckstein, and R. Kimmel Analysis of two-dimensional non-rigid shapes. *Int. J. Comput. Vis.*, 78, 67–88, 2008.
- [5] E. Rublee, V. Rabaud, K. Konolige, and G.R. Bradski ORB: An efficient alternative to SIFT or SURF *Proceedings of the 2011 IEEE International Conference of Computer Vision (ICCV)*, 2564–2571, 2011.
- [6] P. Koehl, M. Delarue, and H. Orland A statistical physics formulation of the optimal transport problem. *Phys. Rev. E*, 100, 013310, 2019.
- [7] F. Mémoli Gromov-Wasserstein distances and the metric approach to object matching. *Found. Comput. Math.*, 11, 417–487, 2011.
- [8] P. Koehl, M. Delarue, and H. Orland Computing the Gromov–Wasserstein distance between two surface meshes using optimal transport. *Algorithms*, 16, 131, 2023.

DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF CALIFORNIA, DAVIS, CA 95616, USA
Email address: pakoehl@ucdavis.edu

FOUNDATIONS OF DIFFERENTIAL CALCULUS FOR MODULES OVER SMALL CATEGORIES

RAN LEVI

Classification AMS 2020: Primary: 16G20, 18G99, 18F30; Secondary: 05C90, 55N31

Keywords: Representations of categories, Representations of posets, Calculus on graphs, Categorification

Let k be a field and let \mathcal{C} be a small category. A k -linear representation of \mathcal{C} , or a $k\mathcal{C}$ -module, is a functor from \mathcal{C} to the category of finite dimensional vector spaces over k . Unsurprisingly, when the category \mathcal{C} is more general than a finite linear order, then its representation type is generally infinite, in some cases tame, and in most cases wild. This means that the classification of indecomposable $k\mathcal{C}$ -modules is hard in the tame case and impossible in the wild case. Hence the task of understanding such representations in terms of their indecomposable factors is generally intractable.

This talk is based the joint work with Jacek Brodzki and Henri Riihimäki [1] and further recent developments in work with Riihimäki, Markus Klemetti and Daniel Sölch. We offer a new set of ideas designed to enable studying modules *locally*. Specifically, inspired by work in discrete calculus on graphs, we set the foundations for a calculus type analysis of $k\mathcal{C}$ -modules, under some restrictions on the category \mathcal{C} . As a starting point, for a $k\mathcal{C}$ -module M we define its gradient $\text{gradient } \nabla[M]$ as a homomorphism from the Grothendieck group of isomorphism classes of $k\mathcal{C}$ -modules to the Grothendieck group of isomorphism classes of $k\widehat{\mathcal{C}}$ -modules, where $\widehat{\mathcal{C}}$ is a new category determined by \mathcal{C} and a generating quiver \mathcal{G} , which we call the *line category*. We show that the gradient thus defined, when restricted to the appropriate setup, reproduced and integral analog of the well known and used gradient on weighted graphs [2].

Pushing the analogy with ordinary differential calculus and discrete calculus on graphs [2], we define left and right divergence via the appropriate left and right Kan extensions and two bilinear pairings on modules and study their properties, specifically with respect to adjointness relations between the gradient and the left and right divergence. The left and right divergence are shown to be rather easily computable in favourable cases. Here too we are able to show that in the appropriate context our left and right divergence coincide and reproduce the an integral analog of the divergence on weighted graphs. Having left and right divergence in place, composition with the gradient allows us to introduce a left and a right Laplacians in our context.

Having set the scene, we concentrate specifically on the case where the category \mathcal{C} is a finite poset. Our main result is a necessary and sufficient condition for the gradient of a module M to vanish under certain hypotheses on the poset. The result can be summarise as saying that, under the right hypotheses, the gradient of a module M , restricted to a subcategory \mathcal{C}_0 of \mathcal{C} that is generated by a tree, vanishes if and only if $M(x) \cong M(y)$ for all objects in \mathcal{C}_0 and the homomorphisms induced by M on all indecomposable morphisms

are, in the appropriate sense, naturally conjugate. We next investigate implications for two modules whose gradients are equal. Finally we consider the resulting left and right Laplacians, namely the compositions of the divergence with the gradient, and study an example of the relationship between the vanishing of the Laplacians and the gradient.

Moving on, we investigate the kernel of the gradient when the poset in question is generated by a rooted tree. There we have an explicit description of the kernel, and find a representation any module M modulo the kernel in terms of kernels of iterated gradients. In another direction we study generalisations of the theory to arbitrary small categories and with varying generating quivers. We also define a new cohomology theory of module categories that in the appropriate sense is a categorification of the discrete Hodge cohomology for directed graphs [2].

REFERENCES

- [1] J. Brodzki, R. Levi and H. Riihimäki, Foundations for Differential Calculus of Modules over posets arXiv:2307.02444.
- [2] L.-H. Lim, Hodge laplacians on graphs, SIAM Review, 62:685–715, 2020.

UNIVERSITY OF ABERDEEN
Email address: r.levi@abdn.ac.uk

SOLVING PDE INVERSE PROBLEMS WITH GENERATIVE MODELS AND THEIR APPLICATIONS

ZHENG MA

Classification AMS 2020:

Keywords:

While deep learning has advanced PDE inverse problem solutions, current methods often depend on paired data or retraining when conditions change—limiting efficiency and flexibility. To overcome these challenges, we present an unsupervised inversion framework leveraging score-based generative diffusion models within a Bayesian inversion paradigm. Our approach recasts posterior estimation as a conditional generative process using reverse-time stochastic differential equations (SDEs). Furthermore, we introduce a diffusion posterior sampling algorithm based on ordinary differential equations (ODEs), ensuring accuracy through marginal probability consistency across forward Fokker-Planck dynamics. Experiments validate robust performance across diverse PDEs

SHANGHAI JIAO TONG UNIVERSITY, CHINA

Email address: zhengma@sjtu.edu.cn

ALGEBRAIC GEOMETRY LEARNS MACHINES AND MACHINES LEARN ALGEBRAIC GEOMETRY

ANTHEA MONOD

Classification AMS 2020:

Keywords:

In this talk I will overview some existing results and ongoing work at the intersection of algebraic geometry and machine learning. I will present how a piecewise linear and combinatorial variant of algebraic geometry—known as tropical geometry—has been shown to be relevant in defining neural networks and talk about some recent and current work that our group is doing that adapts tropical geometry theory in numerical studies towards a better understanding of neural network behaviour during training. While algebraic geometry holds much potential for better understanding machine learning, it turns out that machine learning is also a powerful tool that can help develop algebraic geometry theory. I will also overview some recent and ongoing work by researchers in my group where we use neural networks for theorem discovery in algebraic geometry.

IMPERIAL COLLEGE LONDON, UK

Email address: a.monod@imperial.ac.uk

COMPUTATIONAL INFORMATION GEOMETRY ON BREGMAN MANIFOLDS AND SUBMANIFOLDS

FRANK NIELSEN

In the first part, we review the construction of a Bregman manifold (M, g, ∇, ∇^*) from a Legendre-type convex function [2] $F : \Theta \rightarrow \mathbb{R}$. The Riemannian metric tensor can be expressed as $g(\theta) = \nabla^2 F(\theta)$ or equivalently in the dual coordinate system $\eta = \nabla F(\theta)$ induced by the Legendre transform $F^*(\eta) = \langle \theta, \eta \rangle - F(\theta)$ as $g(\eta) = \nabla^2 F^*(\eta)$. The torsion-free flat affine connections ∇ and ∇^* are dual with respect to the metric tensor g since $\frac{\nabla + \nabla^*}{2}$ coincide with the Levi-Civita connection induced by g . By further using a representation function $r(\theta)$, we show that α -divergences are representational Bregman divergences on the positive orthant cone and curved representational Bregman divergences on the probability simplex [4]. In the second part, we show that symmetrized Bregman divergences are curved Bregman divergences, and present a Bregman projection theorem for calculating the curved Bregman centroid [3]. In the third and last part, we describe clustering [5], nearest-neighbor query data structures [6], and Voronoi diagrams [1] on Bregman manifolds and sub-manifolds with several applications in statistics and data science. Finally, we present work in progress pyBregMan: A Python library for algorithms and data-structures on BREGman MANifolds[7].

REFERENCES

- [1] Jean-Daniel Boissonnat, Frank Nielsen, and Richard Nock. Bregman Voronoi diagrams. *Discrete & Computational Geometry*, 44:281–307, 2010.
- [2] Frank Nielsen. On geodesic triangles with right angles in a dually flat space. In *Progress in Information Geometry: Theory and Applications*, pages 153–190. Springer, 2021.
- [3] Frank Nielsen. Curved representational Bregman divergences and their applications. *arXiv preprint arXiv:2504.05654*, 2025.
- [4] Frank Nielsen and Richard Nock. The dual Voronoi diagrams with respect to representational Bregman divergences. In *2009 Sixth International Symposium on Voronoi Diagrams*, pages 71–78. IEEE, 2009.
- [5] Frank Nielsen, Richard Nock, and Shun-ichi Amari. On clustering histograms with k -means by using mixed α -divergences. *Entropy*, 16(6):3273–3301, 2014.
- [6] Frank Nielsen, Paolo Piro, and Michel Barlaud. Bregman vantage point trees for efficient nearest neighbor queries. In *2009 IEEE International Conference on Multimedia and Expo*, pages 878–881. IEEE, 2009.
- [7] Frank Nielsen and Alexander Soen. pyBregMan: A Python library for Bregman Manifolds. *arXiv preprint arXiv:2408.04175*, 2024.

SONY COMPUTER SCIENCE LABORATORIES INC, TOKYO, JAPAN

CATEGORIES AND SHEAVES FOR OPTIMIZATION: FROM MULTI-STAGE TO DISTRIBUTED

HANS RIESS

Classification AMS 2020: 90C25, 18F20, 18D20

Keywords: optimal control, category theory, sheaf theory, distributed optimization

1. INTRODUCTION

Optimization is a foundational pillar of modern data science and control theory, with applications spanning from machine learning to robotics. We argue that two abstract mathematical frameworks — category theory and sheaf theory — provide a powerful lens through which to structure, analyze, and solve complex optimization tasks. We demonstrate that compositional structures inherent in many optimization problems can be elegantly captured by categorical constructions. Specifically, multi-stage optimization problems find a natural home in the language of enriched category theory, while distributed optimization and coordination problems can be effectively modeled using cellular sheaves and their associated Laplacians. The results presented is joint work with Tyler Hanks, Baike She, Trevor Gross, Samuel Cohen, Matthew Hale, James Faribanks, Evan Patterson, and Matthew Klawonn.

2. MULTI-STAGE OPTIMIZATION VIA ENRICHED CATEGORIES

Many problems in control and robotics can be formulated as multi-stage optimization problems, where decisions are made sequentially over a time horizon. A canonical example is Model Predictive Control (MPC), where a system’s behavior is optimized over a finite future, subject to dynamics and constraints.

2.1. Modeling Convex MPC. We begin with convex optimization problems. A key insight is that a convex optimization problem can be represented by a mathematical object known as a convex bifunction, introduced by Rockafellar [1]. A convex bifunction $F : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{\pm\infty\}$ can be seen as a family of optimization problems parameterized by the second argument.

We demonstrate that these bifunctions form the morphisms of a symmetric monoidal category, which we denote Conv . The objects of this category are Euclidean spaces. Composition of morphisms corresponds to solving one optimization problem and using its output as a parameter for the next. The monoidal product allows for the parallel composition of independent problems.

To properly model MPC, we must distinguish between state variables, which couple stages sequentially, and control inputs, which are independent at each stage. This is achieved by moving to a parameterized category, denoted $\text{Para}(\text{Conv})$, whose morphisms are pairs (U, f) where $f : U \otimes X \rightarrow Y$ is a morphism in Conv . This construction formally separates the control parameters U from the state variables X, Y .

The central result for this section of the talk, detailed in [6], is that a standard multi-stage MPC problem can be constructed by the repeated composition of a single one-step morphism in the category $\text{Para}(\text{Conv})$.

Theorem 2.1 (Hanks et al. 2024). *A multi-stage convex MPC problem over a time horizon T is equivalent to the T -fold composition of a one-step MPC morphism in the category $\text{Para}(\text{Conv})$.*

This categorical framework not only provides a formal language for constructing MPC problems but also comes with a graphical calculus (string diagrams) that allows for intuitive visualization and manipulation of complex control architectures. We have implemented these ideas in the Julia library `AlgebraicControl.jl`, which allows for the automated and compositional construction of optimal control problems.

2.2. A Categorical View on Nonlinear MPC. Real-world systems are often nonlinear, posing significant challenges to traditional optimization methods. We extend our categorical framework to nonlinear MPC by assuming that the objective and constraint functions are polynomials, rendering the feasible sets as semi-algebraic. The naive composition of nonconvex bifunctions fails. To circumvent this, we redefine composition using a local minimization operator that depends on the basin of attraction of local minima, drawing on concepts from stratified Morse theory.

3. DISTRIBUTED OPTIMIZATION VIA CELLULAR SHEAVES

We now shift our focus from single-agent sequential problems to multi-agent coordination problems over networks. These distributed optimization problems are ubiquitous in modern data science, underpinning technologies like federated learning and recommendation systems.

3.1. Coordination Sheaves and Homological Programming. We model a network of interacting agents as a graph $G = (V, E)$. A cellular sheaf \mathcal{F} on this graph assigns a vector space (stalk) $\mathcal{F}(v)$ to each agent $v \in V$, representing its state space. To each communication link $e \in E$, it assigns a space $\mathcal{F}(e)$ and to each incidence of a vertex on an edge, a restriction map. These maps encode the constraints or desired relationships between the states of connected agents.

A global section of the sheaf is an assignment of states to agents that is consistent with all constraints. The problem of finding such a consistent assignment can be challenging. Sheaf Laplacians provide a dynamical way to approach this problem. For a cellular sheaf \mathcal{F} , the sheaf Laplacian $L_{\mathcal{F}}$ is an operator whose kernel corresponds to the space of global sections.

Building on this, we introduce a nonlinear homological program, which is a formulation for a broad class of distributed coordination tasks.

Definition 3.1. *A nonlinear homological program consists of an undirected graph G , a cellular sheaf \mathcal{F} on G , a set of local objective functions $\{f_v : \mathcal{F}(v) \rightarrow \mathbb{R}\}_{v \in V}$, and a set of potential functions $\{U_e : \mathcal{F}(e) \rightarrow \mathbb{R}\}_{e \in E}$. The optimization problem is to*

$$(3.1) \quad \begin{aligned} & \underset{x \in C^0(G; \mathcal{F})}{\text{minimize}} \quad \sum_{v \in V} f_v(x_v) \\ & \text{subject to} \quad L_{\mathcal{F}}^{\nabla U} x = 0 \end{aligned}$$

where $L_{\mathcal{F}}^{\nabla U}$ is the nonlinear sheaf Laplacian derived from the potentials U_e .

This framework, developed in [2], is highly expressive. By choosing different potential functions U_e , we can model various coordination goals such as consensus, formation control, flocking, and dissensus. When the objective and potential functions are convex, the homological program is a convex optimization problem. We propose a distributed solution method based on the Alternating Direction Method of Multipliers (ADMM).

3.2. Future Directions: Sheaves of Convex Bifunctions. A promising direction for future work is to unify the two perspectives presented here. We propose constructing cellular sheaves that are valued not in a simple category like vector spaces, but in the category Conv of convex bifunctions itself. The sheaf Laplacian in this context is outlined in [3].

4. RELATED WORK

The application of category theory to engineering and systems theory is a growing field. Our work on multi-stage optimization builds on foundational work in applied category theory and its application to areas like dynamical systems and control. The use of sheaf theory in network analysis and distributed systems has been pioneered by Jakob Hansen, Robert Ghrist and Michael Robinson. Our contribution lies in synthesizing these tools to create a unified framework specifically tailored for a broad class of optimization problems encountered in modern data science and engineering.

REFERENCES

- [1] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1997.
- [2] T. Hanks, H. Riess, S. Cohen, T. Gross, M. Hale, and J. Fairbanks. Distributed Multi-agent Coordination over Cellular Sheaves. *arXiv:2504.02049*, 2025.
- [3] R. Ghrist, M. Lopez, P. R. North, and H. Riess. Categorical Diffusion of Weighted Lattices. *arXiv:2501.03890*, 2025.
- [4] J. Hansen and R. Ghrist. Toward a spectral theory of cellular sheaves. *J Appl. and Comput. Topology*, vol. 3, no. 4, pp. 315358, 2019.
- [5] J. Hansen and R. Ghrist. Distributed Optimization with Sheaf Homological Constraints. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 565-571, 2019.
- [6] T. Hanks, B. She, E. Patterson, M. Hale, M. Klawonn, and J. Fairbanks. Modeling Model Predictive Control: A Category Theoretic Framework for Multistage Control Problems. In *2024 American Control Conference (ACC)*, pp. 48504857, 2024.

85 5TH ST NW, ATLANTA, GEORGIA 30332, UNITED STATES OF AMERICA
Email address: riess@gatech.edu

EXPANDERS, WAISTS, AND THE KAZHDAN PROPERTY

ROMAN SAUER

Classification AMS 2020: Primary 22D55; Secondary 58E99, 53C23, 05C48.

Keywords: Kazhdan property (T), expander graphs, Riemannian manifolds, uniform waist inequalities, higher-dimensional expanders.

The Kazhdan property (also known as property (T)) is a rigidity property for unitary representations of a group. It was introduced by Kazhdan in the 1960s and later exploited by Margulis to give the first explicit constructions of expander graphs, thereby inaugurating a line of research which has had far-reaching consequences in combinatorics and computer science. Expander graphs can be viewed as discrete analogues of geometric objects with strong connectivity properties. In this talk we explained how ideas originating from expander graphs lead to uniform geometric inequalities for Riemannian manifolds.

A closed Riemannian manifold M with Kazhdan fundamental group is known to satisfy a *uniform waist inequality* in codimension one. More precisely, Cheeger's inequality implies that there exists a positive constant $c > 0$ such that for every finite cover $\bar{M} \rightarrow M$ and every real analytic map $f: \bar{M} \rightarrow \mathbb{R}$ there is a regular value t for which the volume of the fiber $f^{-1}(t)$ is at least c times the total volume of \bar{M} . This result is a Riemannian analogue of the fact that Cayley graphs of finite quotients of a Kazhdan group form a one-dimensional expander family.

Our main result establishes a surprising improvement:

Theorem 0.1 (Main Theorem). *Let M be a closed Riemannian manifold whose fundamental group has property (T). Then there exists a constant $c > 0$ such that for every finite cover $\bar{M} \rightarrow M$ and every real analytic map $f: \bar{M} \rightarrow \mathbb{R}^2$ there is a regular value t for which the volume of the fiber $f^{-1}(t)$ is at least c times the volume of \bar{M} . In other words, the family of finite covers of M satisfies a uniform waist inequality in codimension two.*

This theorem, proved jointly with Uri Bader, shows that Kazhdan groups give rise to two-dimensional Riemannian expanders for free. Uniform waist inequalities in codimension two can be viewed as Riemannian analogues of higher-dimensional expander complexes. The key idea is to convert analytic information encoded in property (T) into combinatorial expansion properties of covering spaces and then translate these into geometric inequalities.

Our argument follows the strategy of Gromov's filling and topological overlap methods, combined with functional analytic techniques. The proof proceeds in several steps:

- (1) Using Kazhdan's property, we establish a uniform 1-coboundary expansion for the simplicial complexes arising from finite covers. This step relies on work of Bader–Gelander–Monod on cohomology with real coefficients.

- (2) We upgrade this to a uniform 2-coboundary expansion by bridging from cohomology with real coefficients to cohomology with integer coefficients using an integral linear programming argument.
- (3) The 2-coboundary expansion yields uniform isoperimetric inequalities via Poincaré duality and a deformation argument, showing that every cycle in low codimension bounds with controlled volume.
- (4) Finally, we deduce the uniform waist inequality from the isoperimetric inequality by a contradiction argument: assuming that all fibers of an analytic map are small produces a non-trivial top-dimensional cycle with arbitrarily small volume.

Examples of manifolds satisfying the hypotheses of Theorem 0.1 include compact locally symmetric spaces of higher rank. We expect that an analogous statement holds for smooth maps (without assuming real analyticity) and speculate that further codimension improvements might be possible for manifolds with stronger rigidity properties.

This work is part of a growing programme connecting geometric group theory, functional analysis and Riemannian geometry. By exploiting the algebraic rigidity of property (T), we obtain new geometric expansion phenomena for manifolds, complementing and extending classical results of Cheeger, Gromov and Margulis. In future work we hope to study whether property (T) yields uniform waist inequalities in even higher codimensions and to explore potential applications in data sciences and geometric deep learning.

REFERENCES

- [1] J. Cheeger, *A lower bound for the smallest eigenvalue of the Laplacian*, in *Problems in Analysis* (R. Gunning, ed.), Princeton Univ. Press, 1970, pp. 195–199.
- [2] M. Gromov, *Isoperimetry of waists and concentration of maps*, Geom. Funct. Anal. **13** (2003), pp. 178–215.
- [3] G. A. Margulis, *Explicit constructions of expanders*, in *Problems of information transmission*, 1968, pp. 325–332.
- [4] U. Bader and R. Sauer, *Uniform waist inequalities in codimension two for manifolds with Kazhdan fundamental group*, preprint (2024).
- [5] L. Guth *The waist inequality in Gromov's work.*, In: *The Abel Prize 2008–2012*, Springer, 2014, pp. 181–195.

INSTITUTE OF ALGEBRA, KARLSRUHE INSTITUTE OF TECHNOLOGY, KARLSRUHE, GERMANY
Email address: roman.sauer@kit.edu

GEOMETRIC DEEP LEARNING FOR PROTEIN DESIGN

JIAN TANG

Classification AMS 2020:

Keywords:

Proteins are workhorses of living cells. Understanding the functions of proteins is critical to many applications such as biomedicine and synthetic biology. Thanks to recent biotechnology breakthroughs such as gene sequencing and Cro-EM, a large amount of protein data (such as protein sequences and structures) are generated, providing a huge opportunity for AI. As the functions of proteins are determined by their structures, in this talk, I will introduce our recent work on protein understanding based on protein 3D structures with geometric deep learning. I will introduce three different topics including protein representation learning, generative models for protein structure prediction, and generative models for protein design, and also how these techniques are used for realworld applications in protein design.

HEC MONTRÉAL, CANADA
Email address: jian.tang@hec.ca

TOPOLOGICAL INTEGRAL TRANSFORMS WITH ISOMETRY INVARIANCE

HAOCHEN YANG

Classification AMS 2020: 62R40, 55-XX, 68U35

Keywords: Euler Characteristic, Euler Calculus, Topological Integral Transform, Shape Descriptor, Morphology, Organoids, Thymus

The Euler Characteristic Transform (ECT) [1], introduced in topological data analysis, is a powerful shape descriptor in both theory and practice. A key property of ECT is its invertibility: it encodes the full geometric and topological information of the input shape. In practice, however, shapes are represented as embedded simplicial complexes, and two complexes are considered equivalent if one can be obtained from the other by a rigid motion (translation, rotation, or reflection). Because ECT is invertible, it will still distinguish between complexes within the same equivalence class.

Based on Detecting Temporal shape changes with the Euler Characteristic Transform (DETECT) algorithm [2], we developed the DETECT metric to measure shape differences between pairs of input shapes. We applied DETECT and the metric to an assembloid dataset and successfully answered various biological questions [3].

We found that DETECT suffers from information loss for complex shapes. To address this challenge, we introduce **SampEuler**, a new shape descriptor based on the ECT pushforward measure of Curry–Mukherjee–Turner [4]. We establish convergence and stability results of SampEuler under Wasserstein distances showing it captures sufficient geometric information about the shape class represented by the input complex, while remaining robust under isometries of input complexes. I will also discuss the possibility of statistical tests based on SampEuler. In addition, we present the **EulerImage**, an isometry-invariant representation that both vectorizes the SampEuler (and ECT pushforward measure) and serves as a visualization tool.

We demonstrate these descriptors on a synthetic dataset and a real-world thymus image dataset, showing that they capture significant shape information with less noise than the standard ECT. Finally, we discuss how the EulerImage aids interpretation by linking analysis results back to the original geometric structures.

REFERENCES

- [1] Katharine Turner, Sayan Mukherjee, and Doug M. Boyer. Persistent Homology Transform for Modeling Shapes and Surfaces. *arXiv preprint arXiv:1310.1030*, 2013. :contentReference[oaicite:0]index=0
- [2] Lewis Marsh, Felix Y. Zhou, Xiao Qin, Xin Lu, Helen M. Byrne, and Heather A. Harrington. Detecting Temporal Shape Changes with the Euler Characteristic Transform. *arXiv preprint arXiv:2212.10883*, 2022. :contentReference[oaicite:0]index=0
- [3] Anna M. Dowbaj, Aleksandra Sljukic, Armin Niksic, Cedric Landerer, Julien Delpierre, Haochen Yang, Aparajita Lahree, Ariane C. Kühn, David Beers, Helen M. Byrne, Sarah Seifert, Heather A. Harrington, Marino Zerial, and Meritxell Huch. Mouse liver assembloids model periportal architecture and biliary fibrosis. In *Nature*, 2025. :contentReference[oaicite:0]index=0

- [4] Justin Curry, Sayan Mukherjee, and Katharine Turner. How Many Directions Determine a Shape and other Sufficiency Results for Two Topological Transforms. *arXiv preprint arXiv:1805.09782*, 2018. :contentReference[oaicite:0]index=0

UNIVERSITY OF OXFORD, RADCLIFFE OBSERVATORY, ANDREW WILES BUILDING, WOODSTOCK RD,
OXFORD OX2 6GG

Email address: haochen.yang@maths.ox.ac.uk

TOWARDS UNDERSTANDING THE CONDENSATION PHENOMENON OF DEEP NEURAL NETWORKS

YAOYU ZHANG

Classification AMS 2020:

Keywords: Machine Learning; Artificial Intelligence

Condensation (also known as quantization, weight clustering, or alignment) is a widely observed phenomenon where neurons in the same layer tend to align with one another during the nonlinear training of deep neural networks (DNNs). It is a key characteristic of the feature learning process of neural networks. However, due to the strong nonlinear nature of this phenomenon, establishing its theoretical understanding remains challenging. In this talk, I will present our systematic efforts to tackle this challenge in recent years. First, I will present results regarding the dynamical regime identification of condensation at the infinite width limit, where small initialization is crucial. Then, I will discuss the mechanism of condensation at the initial training stage and the global loss landscape structure underlying condensation in later training stages, highlighting the prevalence of condensed critical points and global minimizers. Finally, I will present results on the quantification of condensation and its generalization advantage, which includes a novel estimate of sample complexity in the best-possible scenario. These results underscore the effectiveness of the phenomenological approach to understanding DNNs, paving the way for a deeper understanding of deep learning in the near future.

<https://arxiv.org/abs/2105.11686>

SHANGHAI JIAO TONG UNIVERSITY, CHINA
Email address: zhyy.sjtu.edu.cn

MULTI-COVER: A MATHEMATICAL FRAMEWORK FOR TOPOLOGICAL DATA ANALYSIS AND DEEP LEARNING

YIPENG ZHANG

Classification AMS 2020:

Keywords: Computational Geometry, Topological Data Analysis, Topological deep learning

ABSTRACT. Topological Data Analysis (TDA) has made significant contributions to molecular and materials science. Multi-Cover Persistence (MCP) and its associated Rhomboid Tiling (RT) structure, as a generalization of the alpha shape in TDA, provide a powerful framework for capturing the shape and higher-order geometric features of objects. Leveraging these concepts, we developed three innovative computational models: First, a featurization-based machine learning model where the MCP framework extracts topological features from the persistent homology of molecular multi-covers to predict polymer properties with high accuracy. Second, the RT structure forms the basis for a hierarchical graph pooling model for molecular graph classification tasks. Third, building upon RT's multi-scale hierarchical structure, we developed a topological deep learning model that utilizes RT's higher-order geometric relationships to design an advanced message-passing mechanism . This framework demonstrates particular effectiveness in protein complex quality assessment. All three models show excellent performance, highlighting the versatility of MCP and RT in advancing polymer informatics, geometric graph learning, and topological deep learning applications.

1. BACKGROUND AND FRAMEWORK

Topological Data Analysis (TDA) provides a rigorous way to describe the shape of data through topological invariants such as homology and persistence. In molecular and materials data, atoms are naturally represented as points in \mathbb{R}^3 , and the overall structure of a molecule can be captured by placing a ball at each atomic position and examining how these balls overlap. This classical idea—representing data by a union of Euclidean balls—has been the foundation of many shape descriptors such as the *Alpha complex*. Multi-cover and Rhomboid tiling. Edelsbrunner and Osang extended this classical construction by not only considering the union of the balls themselves, but also the regions where multiple balls intersect [1]. Instead of recording only whether a point is covered by at least one ball, they studied the regions covered by at least k balls, called *k-fold covers*. This additional parameter k introduces a new geometric layer that reflects not only the global shape of the data but also the *density* and multi-scale organization of the point cloud. From the viewpoint of TDA, the multi-cover filtration generalizes the standard radius filtration: for each radius r and cover order k , one obtains the corresponding k -fold cover, leading to a natural *bi-filtration* structure. This bi-filtration simultaneously tracks how topological features evolve as the scale r grows and as the overlap order k increases, providing a richer description of geometric-topological behavior.

The *Rhomboid Tiling* (RT) structure, also introduced by Edelsbrunner and Osang [1], serves as a discrete topological model of the multi-cover. Formally, it is a polyhedral complex embedded in \mathbb{R}^{d+1} whose intersection with the hyperplane $x_{d+1} = -k$ recovers the order- k Delaunay complex. Intuitively, RT encodes the homotopy-equivalent relations among different k -levels of the multi-cover and gives an efficient combinatorial representation of its bi-filtration topology. Rather than describing all the geometric details, one may regard RT as the most compact discrete complex that links together the Delaunay slices of different orders and preserves their topological connectivity across scales. This construction provides the theoretical foundation upon which the subsequent computational frameworks—MCP-based feature extraction, RT-guided graph pooling, and RT-based topological deep learning—are built.

2. THREE MODELS BUILT ON MCP/RT

(A) MCP descriptors for polymer informatics. The first application of the MCP framework focuses on polymer property prediction. [2] By encoding molecular structures as Delaunay-slice persistence diagrams across multiple cover orders, MCP captures both shape and higher-order geometric information of polymer monomers. Combined with standard tree-based regressors, these descriptors achieve predictive accuracy superior to classical fingerprint representations such as the Extended Connectivity Fingerprints (ECFP) [3] and Polymer Genome descriptors [4]. Compared with these hand-crafted fingerprints that mainly encode local atomic connectivity, MCP features additionally embed global geometric topology and density variations of molecular structures, which proves particularly beneficial for larger and more complex monomers. Empirical studies demonstrate that MCP-based models yield state-of-the-art performance across benchmark polymer datasets, validating MCP as a scalable topological representation for polymer informatics.

(B) RTPool: RT-guided hierarchical geometric pooling. Building upon the discrete geometry of Rhomboid Tiling, we further developed a hierarchical pooling framework for geometric graphs, termed *RTPool*. [5] Instead of relying solely on graph connectivity, RTPool leverages the geometric relationships implied by Rhomboid Tiling to guide the coarse-graining of graph structures. This approach outperforms representative graph pooling methods such as DiffPool [6], MinCutPool [7], and the recently proposed TopoPool based on witness complexes [8]. The improvement arises from the fact that RTPool integrates explicit higher-order geometric cues into the pooling hierarchy, leading to more expressive and physically meaningful graph embeddings for molecular and bioinformatics datasets.

(C) RT-QA: Rhomboid-tiling-based topological deep learning for protein complexes. Finally, we extend the MCP/RT framework to topological deep learning by developing *RT-QA*, a Rhomboid-tiling-based model for protein complex quality assessment. RT-QA represents protein interfaces using higher-order Voronoi diagrams (orders $k = 1, 2, 3$), where each order corresponds to residue groups of increasing size. A cross-order message-passing mechanism, guided by the topology of Rhomboid Tiling, allows the network to learn interactions between different structural scales. Evaluated on the CASP15 EMA and antibody–antigen benchmark datasets, RT-QA outperforms existing learning-based quality assessment models such as TopoQA [9] and DProQA [10], and achieves accuracy approaching the self-assessment capability of AlphaFold3 [11]. These

results demonstrate that integrating Rhomboid Tiling topology enables a unified and geometry-consistent representation of multi-scale protein interactions.

3. DISCUSSION

The multi-cover framework and its discrete counterpart, the rhomboid tiling, reveal that data geometry is inherently multi-layered and density-sensitive. They encode not only the connectivity pattern of samples but also the overlapping intensity among local neighborhoods, which bridges geometric continuity and statistical concentration. This perspective transforms data from a mere collection of points into a structured hierarchy of coverings, where higher-order intersections convey the coherence and redundancy of spatial organization. Such representation captures how local structures reinforce global shape, providing an intrinsic regularization of geometric information.

These properties explain why multi-cover and rhomboid-tiling concepts serve as effective foundations for learning representations. When used to design feature descriptors, the induced topological invariants summarize data by stable, scale-aware statistics that are insensitive to perturbations yet sensitive to meaningful geometric variations. They naturally encode correlations between spatial arrangement and physical attributes—an ability that conventional fingerprints, largely based on local connectivity patterns, cannot achieve. When embedded into end-to-end neural architectures, the rhomboid-tiling topology defines coherent information pathways across different geometric scales, enabling networks to propagate messages along homotopy-consistent structures rather than arbitrary graph connections. In this sense, the success of MCP-based fingerprints and RT-based deep models stems from the same principle: both exploit the multi-scale, overlapping organization of data as a structural prior that aligns topology, geometry, and learning in a unified mathematical framework.

Acknowledgements. We thank collaborators and the IMS-NTU workshop community for insightful discussions.

REFERENCES

- [1] Herbert Edelsbrunner and Georg Osang. The multi-cover persistence of Euclidean balls. *Discrete & Computational Geometry*, 65(4):1296–1313, 2021.
- [2] Yipeng Zhang, Cong Shen, and Kelin Xia. Multi-Cover Persistence (MCP)-based machine learning for polymer property prediction. *Briefings in Bioinformatics*, 25(6):bbae465, 2024. Oxford University Press.
- [3] Adrià Cereto-Massagué, María José Ojeda, Cristina Valls, Miquel Mulero, Santiago Garcia-Vallvé, and Gerard Pujadas. Molecular fingerprint similarity search in virtual screening. *Methods*, 71:58–63, 2015.
- [4] Chiho Kim, Anand Chandrasekaran, Tran Doan Huan, Deya Das, and Rampi Ramprasad. Polymer genome: a data-powered polymer informatics platform for property predictions. *The Journal of Physical Chemistry C*, 122(31):17575–17585, 2018.
- [5] Yipeng Zhang, Longlong Li, and Kelin Xia. Rhomboid Tiling for Geometric Graph Deep Learning. In *Proceedings of the Forty-second International Conference on Machine Learning (ICML)*, 2025.
- [6] Zhitao Ying, Jiaxuan You, Christopher Morris, Xiang Ren, William Hamilton, and Jure Leskovec. Hierarchical graph representation learning with differentiable pooling. *Advances in Neural Information Processing Systems*, 31, 2018.
- [7] Filippo Maria Bianchi, Daniele Grattarola, and Cesare Alippi. Spectral clustering with graph neural networks for graph pooling. In *Proceedings of the International Conference on Machine Learning*, pages 874–883. PMLR, 2020.

- [8] Yuzhou Chen and Yulia R. Gel. Topological pooling on graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(6):7096–7103, 2023.
- [9] Xinyi Hu, et al. TopoQA: Topology-enhanced quality assessment of protein complexes. *Bioinformatics*, 40(2):btad712, 2024.
- [10] Jing Zhang, et al. DProQA: Deep learning-based protein complex quality assessment. *Bioinformatics*, 39(1):btac765, 2023.
- [11] DeepMind. AlphaFold3: A unified deep learning model for predicting atomic-resolution structures of biomolecules. *Nature*, in press, 2025.
- [12] First name Family name. Title of the paper. *journal name*, Volume number, pages, year.

SCHOOL OF PHYSICAL & MATHEMATICAL SCIENCES, NANYANG TECHNOLOGICAL UNIVERSITY, SPMS-04-05, 21 NANYANG LINK, SINGAPORE 637371

Email address: YIPENG001@e.ntu.edu.sg

ON THE REPRESENTATION LEARNING IN DIFFUSION MODEL

DIFAN ZOU

Classification AMS 2020:

Keywords: Diffusion Model, Representation Learning, Gradient Descent, Rule Learning, Generative Modeling

Background. Diffusion models [1, 2] have achieved remarkable success in generative tasks, such as image and audio synthesis, demonstrating state-of-the-art performance across diverse domains. However, their theoretical foundations, particularly regarding feature learning dynamics and the ability to capture hidden inter-feature rules, remain underexplored. While diffusion models have shown potential in downstream tasks like classification, their mechanisms for learning balanced representations and adhering to fine-grained feature dependencies contrast sharply with traditional discriminative models. Additionally, real-world applications reveal critical limitations: mainstream diffusion models often fail to capture subtle rules between features (e.g., lighting-shadow relationships or object reflections), highlighting a gap between their generative capabilities and the need for precise inter-feature consistency.

Related Works. Prior research on diffusion models has primarily focused on theoretical guarantees for distribution estimation [3, 4] and sampling error [12], and convergence [4, 10], showing minimax optimality under specific assumptions. Studies on denoising autoencoders, a related framework, have analyzed linear and non-linear variants but rarely addressed feature learning dynamics. For classification models, existing work emphasizes their tendency to prioritize easy-to-learn patterns, leading to shortcut learning [7, 8, 9].

In the context of rule learning, previous studies explored diffusion models' compositional abilities with independent features but reported contradictory results for dependent features, some showing success in shape reasoning, others failing at numerical constraints. These inconsistencies underscore the need for systematic evaluation of rule complexity and theoretical analysis of why diffusion models struggle with fine-grained dependencies.

Theoretical Results. We first performs the study on the general representation learning in diffusion model [5]. In particular, inspired by the image data structure, we employ a multi-patch data distribution $\mathbf{x} = [\boldsymbol{\mu}_y, \boldsymbol{\xi}]$ for both classification and diffusion model training. We consider a binary-class data setup with $y = \pm 1$ as the data label and $\boldsymbol{\mu}_1, \boldsymbol{\mu}_{-1} \in \mathbb{R}^d$ are two fixed orthogonal vectors, i.e., $\boldsymbol{\mu}_1 \perp \boldsymbol{\mu}_{-1}$, representing the signal. On the other hand, $\boldsymbol{\xi}$ is the label-independent noise, which is randomly sampled from a Gaussian distribution with standard deviation σ_ξ . Then, we develop the following theoretical results:

Theorem 0.1 (Informal). *Let SNR := $\|\boldsymbol{\mu}\| / (\sigma_\xi \sqrt{d})$ be the signal-to-noise ratio, then*

- For **diffusion model**, $|\langle \mathbf{w}, \mu_y \rangle|, |\langle \mathbf{w}, \xi \rangle|$ exhibit linear growth initially and there exists a stationary point along the path of the training dynamics that satisfies $|\langle \mathbf{w}, \mu_y \rangle|/|\langle \mathbf{w}, \xi \rangle| = \Theta(n \cdot \text{SNR}^2)$.
- For **classification**, $|\langle \mathbf{w}, \mu_y \rangle|, |\langle \mathbf{w}, \xi \rangle|$ exhibit exponential growth initially and when $n \cdot \text{SNR}^2 \geq \beta$ for some constant $\beta > 1$, $|\langle \mathbf{w}, \mu_y \rangle|/|\langle \mathbf{w}, \xi \rangle| = \omega(1)$, and when $n \cdot \text{SNR}^2 < 1/\beta$, $|\langle \mathbf{w}, \mu_y \rangle|/|\langle \mathbf{w}, \xi \rangle| = o(1)$.

Theorem 0.1 highlights differences in the feature learning process between diffusion models and classification. Especially in the regime where $n \cdot \text{SNR}^2 = \Theta(1)$, classification is sensitive to changes in SNR and tends to learn either the signal μ_y or the noise ξ . In contrast, diffusion model learns both signal and noise to the same order.

The theorem above demonstrates that **diffusion models can extract features more effectively than classification models in low-SNR settings**. However, it is still unclear whether diffusion models can capture more precise aspects of the data distribution, such as correlations between multiple features. To investigate this, we conduct an empirical study using four distinct datasets [6]. Each dataset contains images with multiple objects at different positions. The objects' geometry follows both: (1) a coarse rule (defined by inequality constraints), and (2) a fine-grained rule (defined by equality constraints). By our empirical investigation, we find our that (1) diffusion model can learn the coarse rule pretty well; but (2) cannot precisely capture the fine-grained rule between features.

To further demonstrate the capability, we also develop a theoretical framework based on a synthetically designed data model: Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ be two orthogonal feature vectors with unit norm, i.e., $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$ and $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. Let ζ be a random variable with its distribution \mathcal{D}_ζ supporting on a bounded domain $[\underline{c}_\zeta, \bar{c}_\zeta]$ for some constants $0 < \underline{c}_\zeta < \bar{c}_\zeta < \infty$. Each image data consists of multiple patches

$$\mathbf{x} = [\mathbf{x}^{(1)\top}, \mathbf{x}^{(2)\top}, \dots, \mathbf{x}^{(P)\top}]^\top,$$

where $\mathbf{x}^{(1)} = \zeta \mathbf{u}$, $\mathbf{x}^{(2)} = (1 - \zeta) \mathbf{v}$,

and $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}$ are *independent* with the remaining patches. Then, we establish the following theorem on the rule-conforming error achieved by the learned score function via a two-layer CNN model trained with gradient descent:

Theorem 0.2. *Let \mathbf{w}^* be a stationary point of the DDPM loss, then we can lower bound*

$$\text{Err}(\mathbf{w}^*) > \mathbb{E}_{\zeta, \epsilon_1} [\text{Var}_{|\zeta, \epsilon_1} [\text{poly}(\langle \mathbf{u}, \epsilon_2 \rangle, \langle \mathbf{v}, \epsilon_2 \rangle)]]],$$

where ϵ_1 and ϵ_2 are two independent gaussian random vectors and $\text{poly}(a, b)$ denotes a certain polynomial function of a and b .

Clearly, Theorem 0.2 shows that in many cases, as long as the polynomial function is not a constant, the rule-conforming error of the learned score network will always be positive, **showing that the inter-feature rule can not be precisely learned**. We also demonstrate this in many experiments on the synthetic data.

Future Direction. For feature learning dynamics, we will extend the framework to multi-feature distributions (e.g., features with varying frequencies or norms) and analyze conditional/latent diffusion models. We will explore how training objectives and optimizers affect feature balance in these new settings. For inter-feature rule learning: we will develop methods to enhance fine-grained rule capture in real-world scenarios, where rules are poorly defined. We can explore the improved guidance

strategies by addressing the fragility of classifier training for subtle rule differences and investigate whether architectural advancements (e.g., larger models or better priors) can reduce rule-learning errors.

These studies collectively advance understanding of diffusion models' strengths and limitations, paving the way for more robust and rule-aware generative systems.

REFERENCES

- [1] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840-6851.
- [2] Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2020). Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456.
- [3] Oko, K., Akiyama, S., & Suzuki, T. (2023, July). Diffusion models are minimax optimal distribution estimators. In International Conference on Machine Learning (pp. 26517-26582). PMLR.
- [4] Chen, M., Huang, K., Zhao, T., & Wang, M. (2023, July). Score approximation, estimation and distribution recovery of diffusion models on low-dimensional data. In International Conference on Machine Learning (pp. 4672-4712). PMLR.
- [5] Han, A., Huang, W., Cao, Y., & Zou, D. (2025, April) On the Feature Learning in Diffusion Models. In The Thirteenth International Conference on Learning Representations.
- [6] Han, Y., Han, A., Huang, W., Lu, C., & Zou, D. (2025). Can Diffusion Models Learn Hidden Inter-Feature Rules Behind Images?. arXiv preprint arXiv:2502.04725.
- [7] Cao, Y., Chen, Z., Belkin, M., & Gu, Q. (2022). Benign overfitting in two-layer convolutional neural networks. Advances in neural information processing systems, 35, 25237-25250.
- [8] Allen-Zhu, Z., & Li, Y. (2020). Towards understanding ensemble, knowledge distillation and self-distillation in deep learning. arXiv preprint arXiv:2012.09816.
- [9] Zou, D., Cao, Y., Li, Y., & Gu, Q. (2021). Understanding the generalization of adam in learning neural networks with proper regularization. arXiv preprint arXiv:2108.11371.
- [10] Li, P., Li, Z., Zhang, H., & Bian, J. (2023). On the generalization properties of diffusion models. Advances in Neural Information Processing Systems, 36, 2097-2127.
- [11] Chen, S., Kontonis, V., & Shah, K. (2024). Learning general gaussian mixtures with efficient score matching. arXiv preprint arXiv:2404.18893.
- [12] Chen, S., Chewi, S., Li, J., Li, Y., Salim, A., & Zhang, A. R. (2022). Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. arXiv preprint arXiv:2209.11215.

THE UNIVERSITY OF HONG KONG, HONG KONG

Email address: dzou@hku.hk