Mathematisches Forschungsinstitut Oberwolfach

Report No. 20/2026

DOI: 10.4171/OWR/2026/20

S^7: Statistics and Stochastics for Sheaves, Shapes, Stratified and Sub-Riemannian Structures

Organized by
Karen Habermann, Coventry
Stephan Huckemann, Göttingen
Sayan Mukherjee, Leipzig
Stefan Sommer, Copenhagen

17 May – 22 May 2026

Abstract. The mathematical description of shapes as elements of non-linear geometric spaces has been central to the field of geometric statistics and has spawned a variety of new links to other fields of mathematics: Geometric data requires models for stochastic processes in geometric spaces, for example, for data in structural and evolutionary biology, in biomedical imaging, and in morphology such as plant sciences, and non-holonomic constraints lead to sub-Riemannian geometric structures. The prevalence of shape data has created a need for statistical inference tools to draw conclusions from observed shapes in complex, non-linear settings, and new numerical techniques now allow for studying geometries experimentally in novel ways, for example, using neural networks to allow simulation of diffusion bridges in sub-Riemannian spaces. The ability to model local information with sheaves gives the ability to analyze shapes with neural networks.

Mathematics Subject Classification (2020): 53-XX, 58-XX, 60-XX, 62-XX.

License: Unless otherwise noted, the content of this report is licensed under CC BY SA 4.0.

Introduction by the Organizers

The workshop S^7*: Statistics and Stochastics for Sheaves, Shapes, Stratified and Sub-Riemannian Structures* was dedicated to and held in memory of Sayan Mukherjee. Sayan Mukherjee (Leipzig) organized the workshop together with Karen Habermann (Coventry), Stephan Huckemann (Göttingen) and Stefan Sommer (Copenhagen), but Sayan sadly passed away before the workshop took

place. We are deeply saddened by the loss of Sayan, and we kept him in mind during the workshop, the workshop being just one part of Sayan's scientific legacy.

The workshop has brought together 24 scientists from the fields of geometry, probability, stochastic analysis, statistics and from major application areas. The focus was to strengthen interactions between otherwise mostly separate fields to accelerate progress and spawn new ideas, and to build on recent developments in geometric statistics, both for shapes and for broader classes of geometric data that cannot be modeled faithfully in vector spaces.

The scientific program included 21 invited talks of about 45 minutes each, which left plenty of room for informal discussions and collaboration throughout the week. Main themes of the talks at the workshop included

(1) metrics spaces and geometry of metric spaces in several specific cases,
(2) geometry of Wasserstein spaces,
(3) geometric statistics,
(4) machine learning on geometric objects, and
(5) shape space theory and methods.

On metric spaces and the geometry of metric spaces, Tom Nye and Fernando Galaz-García discussed the geometry of spaces of graphs and persistence diagrams, and Washington Mio gave a talk on statistics on metric spaces including estimation of barycenters and modes of probability distributions.

On the Wasserstein geometry theme, Tom Needham gave a talk on the geometry of Gromov-Wasserstein distances highlighting many links to the different fields this workshop brought together, Mao Nishino discussed diffusion processes on Wasserstein space and how to condition on observed data in the form of observed measures, Alice Le Brigant presented a Wasserstein geodesic principal component analysis for probability measures and raised an open question concerning the first principal geodesic component for a family of Gaussian distributions in higher dimensions, and François-Xavier Vialard introduced Gromov-Wasserstein solver that scale to hundreds of thousands of points in minutes.

Several aspects of geometric statistics were covered in the talks. Huiling Le discussed Fréchet p-functions and Fréchet p-means generalizing the Fréchet mean on manifolds. Susovan Pal presented asymptotic aspects of Fréchet mean estimation with detailed new conditions for when smeariness can occur and cannot occur. Yvo Pokern discussed estimation on complex projective spaces, and estimation of subspaces of manifolds was the topic of Ian Dryden's talk on principal nested cones and of Xavier Pennec's talk on flag manifolds and its role in statistics.

Machine learning and neural networks on complex geometric data was the topic of the talk by Christoph von Tycowicz about convolution on manifold valued graphs using diffusion, and of Erlend Grong's talk on PDE learning on manifolds and operator learning, including a discussion on operators equivariant under isometry groups of sub-Riemannian model spaces. Infinite dimensional diffusion models and training were the topic of Elizabeth Baker's talk.

Concerning shape space analysis, Martin Bauer presented a recent result on the completeness of reparametrization-invariant Sobolev metrics on the space of

surfaces, concluding a ten year quest. Erik Jansson discussed gradient flows for template-based reconstruction, Nicolas Charon talked about regression and classification of shapes in the space of varifolds, and Levin Maier spoke about periodic geodesics on manifolds of mappings. Peter Michor discussed shape spaces quotient the special Euclidean group (rigid motion, translations and rotations) with geometry inherited from metrics on the diffeomorphism group with derivatives of rigid motions in their null-space. Karen Habermann presented a complete characterization of stochastic completeness for landmark manifolds with inherited metrics. This problem has been open for several years, and it has been discussed intensively in Oberwolfach at previous workshops. The presentation at this workshop marks the solution of the problem.

The workshop concluded with a perspective on possible future research directions, both theoretical and applied in nature, for geometric statistics, for instance, linking Wasserstein approaches to diffusion models and generative AI, and various other newly evolving machine learning methods.

Acknowledgement: The MFO and the workshop organizers would like to thank the National Science Foundation for supporting the participation of junior researchers in the workshop by the grant DMS-2230648, "US Junior Oberwolfach Fellows".

Workshop: S^7: Statistics and Stochastics for Sheaves, Shapes, Stratified and Sub-Riemannian Structures

Table of Contents

Elizabeth Baker (joint with Alexander Denker, Jes Frellsen)
Infinite-dimensional diffusion models for Bayesian inverse problems 1263

Martin Bauer (joint with Cy Maor and Benedikt Wirth)
Completeness of reparametrization-invariant Sobolev metrics on the space of surfaces 1266

Nicolas Charon (joint with Emmanuel Hartman, Murad Hossen)
Regression and classification of shapes in the space of varifolds 1268

Ian L. Dryden (joint with Yanyan Zhan, Yuexuan Wu)
Principal Nested Cones 1270

Fernando Galaz-García (joint with Mauricio Che, Fernando Galaz-Fontes, Luis Guijarro, Ingrid Amaranta Membrillo Solis, and Motiejus Valiunas)
Metric Geometry of Spaces of Persistence Diagrams 1272

Erlend Grong (joint with F. Ballerin)
Equivariant nonlinear operators 1273

Karen Habermann (joint with Stefan Sommer)
Stochastic completeness for landmark space 1275

Erik Jansson (joint with Carola–Bibiane Schönlieb, Ozan Öktem, Jonathan Krook, Klas Modin)
Gradient flows for template-based reconstruction 1277

Huiling Le (joint with Karthik Bharath and Andrew T.A. Wood)
Fréchet p-functions and Fréchet p-means 1279

Alice Le Brigant (joint with Nina Vesseron, Elsa Cazelles and Thierry Klein)
Principal Component Analysis in the Wasserstein space 1280

Levin Maier (joint with Martin Bauer)
On Periodic Geodesics on Half–Lie Groups 1282

Peter W. Michor (joint with Sarang Joshi, Peter W. Michor and Stefan Sommer)
Not seeing Euclidean motions on landmark space: An LDDMM-metric on $\mathrm{Land}^n(\mathbb{R}^d)/\mathrm{SE}(d)$ 1284

Washington Mio (joint with Tom Needham)
On Stability of Barycenters and Modes of Probability Measures 1287

Tom Needham (joint with Martin Bauer, Facundo Mémoli, Mao Nishino)
Geometry of Gromov-Wasserstein Distances 1289

Mao Nisihino (joint with Martin Bauer, Tom Needham, Stefan Sommer)
Conditioning a measure-valued diffusion on the Wasserstein space 1290

Tom M. W. Nye (joint with William M. Woodman)
From limits of the matrix exponential to a geodesic metric space of graphs 1292

Susovan Pal
Support conditions for smeary probability measures on spheres 1294

Xavier Pennec
Flag for stats 1297

François-Xavier Vialard (joint with Guillaume Houry and Jean Feydy)
From limits of the matrix exponential to a geodesic metric space of graphs 1298

Christoph von Tycowicz (joint with Martin Hanik, Gabriele Steidl)
Manifold GCN: Diffusion-based Convolutional Neural Network for Manifold-valued Graphs 1300

Abstracts

Infinite-dimensional diffusion models for Bayesian inverse problems

Elizabeth Baker
(joint work with Alexander Denker, Jes Frellsen)

1. Introduction and Motivation

Bayesian inverse problems in function spaces arise throughout the sciences and engineering: one observes a noisy, indirect measurement of an unknown function and seeks to characterise the full posterior distribution over that function. Formally, given Hilbert spaces H and Y, a measurable forward operator $G : H \to Y$, and observations of the form

$$y = G(f) + \eta, \quad f \sim \pi, \quad \eta \sim \pi_0,$$

the Bayesian posterior π^y satisfies, by the infinite-dimensional Bayes theorem [4],

$$\frac{d\pi^y}{d\pi}(f) = \frac{1}{\xi} \exp(-\Phi(f, y)),$$

where $\Phi : H \times Y \to \mathbb{R}$ is the negative log-likelihood potential and ξ is a normalising constant.

Score-based diffusion models (SDMs) have recently been extended to infinite-dimensional function spaces [9, 7, 8], providing expressive data-driven priors on such spaces. However, the problem of conditioning an infinite-dimensional SDM to sample from a posterior π^y, rather than the prior π, has remained open. This work addresses that gap, building on the infinite-dimensional h-transform in [2].

2. Infinite-Dimensional Diffusion Models

We work within the framework of [9]. Let C be a trace-class covariance operator on H and let W^H be the associated cylindrical Wiener process. The SDE

$$dX_t = -\tfrac{1}{2} X_t \, dt + \sqrt{C} \, dW_t^H, \quad X_0 \sim \pi,$$

has marginal distributions converging to $\mathcal{N}(0, C)$ as $t \to \infty$. The time-reversed process $Z_t := X_{T-t}$ satisfies

$$dZ_t = \big(\tfrac{1}{2} Z_t + s(T - t, Z_t)\big) \, dt + \sqrt{C} \, dW_t^H,$$

where the *score* $s(t, x)$ is defined as

$$s(t, x) := -(1 - e^{-t})^{-1} \big(x - e^{-t/2} \, \mathbb{E}[X_0 \mid X_t = x]\big).$$

In finite dimensions this coincides with $\nabla_x \log p_t(x)$, but the formulation above is well-posed in infinite dimensions.

3. Conditioning via Doob's h-Transform

The central theoretical contribution [1] is to show that the reverse SDE can be conditioned on $Y = y$ by adding a *guidance term* to its drift, via an infinite-dimensional Doob's h-transform. We work under one of two general settings:

Setting 1.: The support of π lies in a ball of radius R in the Cameron-Martin space of C.

Setting 2.: The prior is absolutely continuous with respect to a Gaussian, $d\pi/d\mathcal{N}(0, C_\pi) \propto \exp(-\Psi)$, with Ψ having at most quadratic growth and Lipschitz gradient, and Φ Fréchet differentiable.

In both settings the time-reversed SDE has a strong solution [9, Theorems 12-13]. We define the h-transform

$$h^y(t,x) := \xi^{-1}\,\mathbb{E}_P[\exp(-\Phi(X_0,y)) \mid X_t = x]\,,$$

where P denotes the path measure of the unconditional process.

Theorem 1 (Conditional SDE)**.** *Under Setting 1 or Setting 2, the conditional path measure $P^y = P(\cdot \mid Y = y)$ satisfies $dP^y = h^y(T, Z_T)\,dP$. Moreover, the conditional process Z_t^y satisfies*

$$dZ_t^y = \Big(\tfrac{1}{2}Z_t^y + s(T-t, Z_t^y) + C\nabla \log h^y(T-t, Z_t^y)\Big)\,dt + \sqrt{C}\,dW_t^H,$$

where $\nabla \log h^y$ denotes the Riesz representative of the Fréchet derivative of $\log h^y$ with respect to x.

Differentiability of h^y is established by different arguments in each setting. In Setting 2 it follows from the differentiability of Φ and Ψ via a Gaussian change of variables. In Setting 1 it follows from the Cameron-Martin theorem applied to the transition kernels of the forward SDE.

Theorem 2 (Score Decomposition)**.** *Under Setting 1 or Setting 2, setting $\beta_t = e^{-t/2}(1-e^{-t})^{-1}$, the guidance term decomposes as*

$$\begin{aligned} C\nabla \log h^y(t,x) &= s^y(t,x) - s(t,x) \\ &= \beta_t\Big(\mathbb{E}[X_0 \mid X_t = x,\, Y = y] - \mathbb{E}[X_0 \mid X_t = x]\Big), \end{aligned}$$

where $s^y(t,x) := -(1-e^{-t})^{-1}\big(x - e^{-t/2}\mathbb{E}[X_0 \mid X_t = x, Y = y]\big)$ is the conditional score.

This is the infinite-dimensional analogue of the finite-dimensional identity $\nabla_x \log p(y \mid x_t) = \nabla_x \log p(x_t \mid y) - \nabla_x \log p_t(x_t)$ [5], and justifies the modular approach of learning only the guidance term on top of a pre-trained unconditional score.

4. Approximating the Guidance Term

Since h^y involves the full transition law of the infinite-dimensional process, the guidance term $C\nabla \log h^y$ is intractable. To learn this we introduce Supervised Guidance Training [1] which recasts the problem as a denoising score matching problem.

Proposition 1 (Supervised Guidance Training)**.** *Define*

$$B := -\frac{e^{-t}}{(1-e^{-t})^2}\,\mathbb{E}\Big[\big\|\mathbb{E}[X_0 \mid X_t, Y] - X_0\big\|_K^2\Big].$$

If B is bounded, then the minimiser of the score-matching objective

$$\mathbb{E}\Big[\big\|u(t, X_t, Y) - C\nabla \log h^Y(t, X_t)\big\|_K^2\Big]$$

coincides with the minimiser of the simulation-free denoising objective

$$\mathbb{E}\left[\left\|\frac{X_t - e^{-t/2}X_0}{1-e^{-t}} + s(t, X_t) + u(t, X_t, Y)\right\|_K^2\right],$$

where the expectation is over $(X_0, Y) \sim \mathrm{Law}(X_0, Y)$ and $X_t \sim \mathrm{Law}(X_t \mid X_0)$. The minimiser u^* is almost surely unique and satisfies $u^*(t, x, y) = C\nabla \log h^y(t, x)$.

The space K is chosen as the Cameron-Martin space of C under Setting 1, or H itself under Setting 2, and boundedness of B follows from [9, Lemma 8]. Training requires only the forward pass of the pre-trained score network s_θ and paired samples (X_0, Y) available by forward simulation of the PDE. For a new observation setting, only the lightweight guidance network u_φ need be retrained; the unconditional score is kept fixed.

5. Conclusion

In this work, based on [1], we developed a framework for using diffusion models to learn posterior distributions for Bayesian inverse problems. We show that this can be done via conditioning of infinite-dimensional SDEs. The resulting guidance term can be decomposed into the score and conditional score terms. Therefore, by learning the conditional score we can learn posterior distributions.

References

[1] E. L. Baker, A. Denker, and J. Frellsen. *Supervised guidance training for infinite-dimensional diffusion models.* arXiv preprint arXiv:2601.20756, 2026. `https://arxiv.org/abs/2601.20756`.

[2] E. L. Baker, G. Yang, M. L. Severinsen, C. A. Hipsley, and S. Sommer. *Conditioning non-linear and infinite-dimensional diffusion processes.* Advances in Neural Information Processing Systems, 37:10801–10826, 2024.

[3] H. Chung, J. Kim, M. T. McCann, M. L. Klasky, and J. C. Ye. *Diffusion posterior sampling for general noisy inverse problems.* In International Conference on Learning Representations, 2023.

[4] M. Dashti and A. M. Stuart. *The Bayesian approach to inverse problems.* In Handbook of Uncertainty Quantification, pp. 311–428. Springer, 2017.

[5] A. Denker, F. Vargas, S. Padhy, K. Didi, S. Mathis, R. Barbano, V. Dutordoir, E. Mathieu, U. J. Komorowska, and P. Lio. *DEFT: Efficient fine-tuning of diffusion models by learning the generalised h-transform.* Advances in Neural Information Processing Systems, 37:19636–19682, 2024.

[6] B. Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

[7] G. Franzese, G. Corallo, S. Rossi, M. Heinonen, M. Filippone, and P. Michiardi. Continuous-time functional diffusion processes. *Advances in Neural Information Processing Systems*, 36:37370–37400, 2023.
[8] J. H. Lim, N. B. Kovachki, R. Baptista, C. Beckham, K. Azizzadenesheli, J. Kossaifi, V. Voleti, J. Song, K. Kreis, J. Kautz, et al. Score-based diffusion models in function space. *Journal of Machine Learning Research*, 26(158):1–62, 2025.
[9] J. Pidstrigach, Y. Marzouk, S. Reich, and S. Wang. Infinite-dimensional diffusion models. *Journal of Machine Learning Research*, 25(414):1–52, 2024.
[10] J. Yao, A. Mammadov, J. Berner, G. Kerrigan, J. C. Ye, K. Azizzadenesheli, and A. Anandkumar. Guided diffusion sampling on function spaces with applications to PDEs. In *Advances in Neural Information Processing Systems*, 2025.

Completeness of reparametrization-invariant Sobolev metrics on the space of surfaces

Martin Bauer
(joint work with Cy Maor and Benedikt Wirth)

1. Completeness of reparametrization-invariant Sobolev metrics on the space of surfaces

The results presented in this talk are based on joint work with Cy Maor and Benedikt Wirth [3].

Motivation: Reparametrization-invariant Riemannian metrics on spaces of immersions have become a central tool in shape analysis and geometric data science [6, 1]. They provide a mathematically rigorous framework for comparing geometric objects and induce Riemannian metrics on shape spaces of unparametrized submanifolds. At the same time they form a natural generalization of right-invariant metrics on diffeomorphism groups and thus constitute an important class of examples in infinite-dimensional Riemannian geometry. A fundamental question concerns completeness properties of these metrics. Around 2013, David Mumford conjectured that Sobolev metrics of sufficiently high order on spaces of immersions should be complete, analogously to strong Sobolev metrics on diffeomorphism groups. This conjecture was verified for spaces of immersed curves by Bruveris, Michor and Mumford [4] and subsequently generalized in several directions, see e.g. [5, 2] and the references therein. The main result of the present work establishes conditions ensuring metric and geodesic completeness as well as the existence of minimizing geodesics on the space of immersed surfaces, thereby confirming the completeness conjecture of Mumford in this important special case. We note, that prior to the present work, no completeness result was known for reparametrization invariant metrics on this space.

Statement of Main Results: To formulate the main result, we first introduce some notation. Let $\mathrm{Imm}^l(M, \mathbb{R}^d)$ denote the Hilbert manifold of H^l-immersions

of a closed oriented surface M into $\mathbb{R}^d$, where $d, l \geq 3$. We then consider the curvature-weighted Sobolev metric

$$G_f^k(h_1, h_2) = \int_M \Big(h_1 \cdot h_2 + |H|^4 \left(g(\nabla h_1, \nabla h_2) + g(\nabla^2 h_1, \nabla^2 h_2) \right) + g(\nabla^k h_1, \nabla^k h_2) \Big) \operatorname{vol}, \tag{1}$$

of order $3 \leq k \leq l$, where g, H, ∇ and vol denote the metric, mean curvature, Levi–Civita connection and volume form induced by the immersion f. It can be easily verified, that the metric is invariant under the action of the group of Sobolev diffeomorphisms $\operatorname{Diff}^l(M)$ and therefore descends to the shape space

$$\mathcal{S}^l(M, \mathbb{R}^d) := \operatorname{Imm}^l(M, \mathbb{R}^d) / \operatorname{Diff}^l(M).$$

The main result of the paper is the following theorem:

Theorem 1 (Completeness Properties of the Space of Immersed Surfaces [3, Theorem 1.1])**.** *Let M be a two-dimensional, closed manifold, let $d, l \geq 3$ and let $\operatorname{Imm}^l(M, \mathbb{R}^d)$ be the space of all immersions of M into $\mathbb{R}^d$ of regularity H^l, equipped with the Riemannian metric G^k from (1) of order $3 \leq k \leq l$. We have:*

- *(a)* **Metric Completeness of $\operatorname{Imm}^k(M, \mathbb{R}^d)$:** *The space $(\operatorname{Imm}^k(M, \mathbb{R}^d), G^k)$ is metrically complete.*
- *(b)* **Geodesic Completeness of $\operatorname{Imm}^l(M, \mathbb{R}^d)$ for $l \geq k$:** *The space of immersions $(\operatorname{Imm}^l(M, \mathbb{R}^d), G^k)$ is geodesically complete. The result continues to hold for $l = \infty$, i.e., the space of smooth immersions $\operatorname{Imm}(M, \mathbb{R}^d)$ equipped with the metric G^k is geodesically complete.*
- *(c)* **Existence of Minimizing Geodesics on $\operatorname{Imm}^k(M, \mathbb{R}^d)$:** *For any two immersed surfaces f_0, f_1 in the same connected component of $\operatorname{Imm}^k(M, \mathbb{R}^d)$ there exists a minimizing G^k-geodesic in $\operatorname{Imm}^k(M, \mathbb{R}^d)$ connecting them.*
- *(d)* **Metric Completeness of $\mathcal{S}^k(M, \mathbb{R}^d)$:** *The space $\mathcal{S}^k(M, \mathbb{R}^d)$ is metrically complete with respect to the metric induced by G^k.*
- *(e)* **Existence of Optimal Reparametrizations:** *For any two elements $[f_1], [f_2] \in \mathcal{S}^k(M, \mathbb{R}^d)$ in the same connected component there exists an optimal reparametrization $\bar{\varphi} \in \operatorname{Diff}^k(M)$ attaining the infimum in the definition of the quotient distance,*

  $$\operatorname{dist}([f_1], [f_2]) = \inf_{\varphi \in \operatorname{Diff}^k(M)} \operatorname{dist}_{G^k}(f_1, f_2 \circ \varphi) = \operatorname{dist}_{G^k}(f_1, f_2 \circ \bar{\varphi}).$$

- *(f)* **Shape Space is a Geodesic Length Space:** *For any two elements $[f_1], [f_2] \in \mathcal{S}^k(M, \mathbb{R}^d)$ in the same connected component there exists a minimizing connecting geodesic in $\mathcal{S}^k(M, \mathbb{R}^d)$, in the sense of metric geometry, for the metric induced by G^k.*

Proof Strategy and Open Questions: The proof proceeds in two steps. First, we establish a general criterion for metric completeness, geodesic completeness and existence of minimizing geodesics for strong Riemannian metrics defined on open subsets of complete Hilbert manifolds. This result is formulated in terms of bounds on metric balls and may be of independent interest. Second, we verify the

assumptions of this abstract theorem for spaces of immersions. The main technical difficulty is obtaining control of the induced metric, mean curvature and their derivatives on G^k-metric balls. This is achieved by combining intrinsic Sobolev estimates with Michael–Simon–Sobolev inequalities. These estimates exclude degenerations such as curvature blow-up and loss of immersedness at finite geodesic distance, thereby yielding the desired completeness properties.

We believe that the methods developed in this work provide a direct route towards proving analogous completeness results for immersed submanifolds of arbitrary dimension and codimension. In particular, we expect that suitably weighted Sobolev metrics of order above the critical threshold $\dim(M)/2+1$ should satisfy metric completeness, geodesic completeness, and existence of minimizing geodesics on both the space of immersions and the corresponding shape space.

At the same time, several important questions remain open. Most notably, it is not known whether the curvature weights used in the present work are essential. In particular, it remains unclear whether unweighted Sobolev metrics of sufficiently high order are complete on spaces of immersed surfaces. Further interesting directions include extensions to metrics of fractional order and to immersions into general ambient manifolds, the construction of complete geometries on spaces of embeddings, and the development of convergent discretizations of the corresponding geodesic boundary value problems.

References

[1] M. Bauer, M. Bruveris, and P. W. Michor, *Overview of the geometries of shape spaces and diffeomorphism groups*, J. Math. Imaging Vision **50** (2014), no. 1–2, 60–97.

[2] M. Bauer, C. Maor, and P. W. Michor, *Sobolev metrics on spaces of manifold valued curves*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **24** (2023), no. 3, 1895–1948.

[3] M. Bauer, C. Maor, and B. Wirth, *Completeness of reparametrization-invariant Sobolev metrics on the space of surfaces*, preprint, arXiv:2512.01566, 2025.

[4] M. Bruveris, P. W. Michor, and D. Mumford, *Geodesic completeness for Sobolev metrics on the space of immersed plane curves*, Forum Math. Sigma **2** (2014), e19, 38 pp.

[5] M. Bruveris, *Completeness properties of Sobolev metrics on the space of curves*, J. Geom. Mech. **7** (2015), 125–150.

[6] L. Younes, *Shapes and Diffeomorphisms*, 2nd ed., Appl. Math. Sci., vol. 171, Springer, Cham, 2019.

Regression and classification of shapes in the space of varifolds

Nicolas Charon

(joint work with Emmanuel Hartman, Murad Hossen)

Despite important progress in the rapidly developing field of geometric deep learning [1], performing statistical analysis on geometric data—where each observation is a shape such as a curve, graph, or surface—remains challenging due to several factors, the most fundamental of which being the non-Euclidean nature of shape spaces. Indeed, shapes are generically defined as equivalence classes of parametrization functions, e.g. Lipschitz embeddings $q : M \to \mathbb{R}^n$ where M is a base parameter space such as S^1 or S^2, modulo the action of some invariance

groups. Among those is the invariance to reparametrization, namely to the right composition $q \circ \gamma$ by any bi-Lipschitz homeomorphism γ of M. Building machine learning frameworks that incorporate those fundamental invariances is often crucial to ensure generalizability of the trained models to new observations.

In this work, we proposed a new framework to tackle the tasks of classification and regression on datasets of curves, spatial networks and surfaces. At its core, our approach leverages the representation of such objects as *varifolds* [2], which in our context, are defined as positive measures on the product space $\mathbb{R}^n \times S^{n-1}$ and can be interpreted as distributions of points with tangent/normal directions attached. A key property is that any parametrized shape q as above can be naturally associated with a varifold μ_q which in addition is invariant to reparametrizations of q. This varifold $\mu_q \in \mathcal{M}^+(\mathbb{R}^n \times S^{n-1})$ can be viewed via its dual action on any test function $h \in C_c(\mathbb{R}^n \times S^{n-1})$ as

$$\langle \mu_q, h \rangle = \int_M h(q(m), \vec{t}_q(m))\, \mathrm{d}\,\mathrm{vol}_q(m).$$

Our proposed framework, that we coin *SVarM* [3], consists in building regressor or classifier functionals in the space of varifolds, which, thanks to the mapping $q \mapsto \mu_q$ from shape spaces to varifolds, will in turn lead to a corresponding invariant shape regression/classification model. Our primary focus has so far been on the simplest possible class of varifold functionals, which, similar to SVM in Euclidean spaces, consist of the affine maps of the form $\mu \mapsto \langle \mu, h \rangle + \beta$ parametrized by a "support" test function $h \in C_c(\mathbb{R}^n \times S^{n-1})$ and a bias $\beta \in \mathbb{R}$. By modelling h as a Multi Layer Perceptron neural network, one can then estimate h and β from a given training set of shapes (viewed as varifolds) with their corresponding labels. We showed, on several benchmark datasets and examples, that this leads to a competitive approach in terms of classification or regression performance when evaluated on test sets, while keeping the size and number of trainable parameters of the model noticeably smaller than typical alternative geometric deep learning architectures.

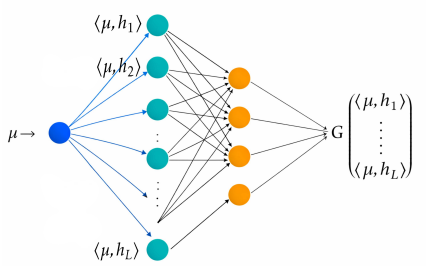

FIGURE 1. Towards parametrizing nonlinear functionals on varifold spaces.

Although those results were obtained based on simple affine functionals on varifold space, one could extend this setting in various ways. An approach in

particular would consist in combining multiple functionals $\mu \mapsto \langle \mu, h_\ell \rangle$ for different test functions $h_1, \ldots, h_L$ via some nonlinear function G as illustrated in Figure 1, which, in principle, could allow one to approximate (in the limit of $L \to +\infty$) any weak-* continuous functional on the space of varifolds. We plan in further exploring such a model in the future and evaluate its potential interest for problems beyond regression and classification.

References

[1] M. Bronstein, J. Bruna, Y. Le Cun, A. Szlam and P. Vandergheynst, *Geometric deep learning: going beyond euclidean data*, IEEE Signal Processing Magazine **34** (2017), 18–42.

[2] W. Allard, *On the first variation of a varifold*, Annals of mathematics **95** (1972), 417–491.

[3] E. Hartmann and N. Charon, *SVarM: Linear Support Varifold Machines for Classification and Regression on Geometric Data*, Preprint (2025).

Principal Nested Cones

Ian L. Dryden

(joint work with Yanyan Zhan, Yuexuan Wu)

In many statistical applications, it is common to encounter data that lie on a type of cone, where there is a distinction between an overall scale variable and the remaining scale-free structure. For example, in biology, it is of interest to examine the relationship between size (a scale variable) and shape (geometrical properties which are scale-free). Another example is a covariance matrix Σ which can be represented as the total variation $\mathrm{tr}(\Sigma)$ and a scale-free trace 1 positive definite matrix $\Sigma/\mathrm{tr}(\Sigma)$. The distinction between scale and scale-free structure leads to the representation of the data on a cone of a metric space, which is the product of a positive real line (the scale) and an underlying metric space M. The use of the cone representation can be helpful for examining the joint relationship between the scale variable and data on the underlying metric space M.

Among such settings, object and shape data arise in a wide range of applications, including biology, medical imaging, computer vision, and molecular analysis [2, 4, 5]. These data are commonly represented by landmark coordinates or derived measurements such as lengths, widths, and angles, and can therefore be high-dimensional. In addition, complex interactions between shape and size are common. Dimension reduction and joint size-and-shape analysis are therefore fundamental challenges in statistical modeling.

A common approach to dimension reduction is to carry out Principal Components Analysis (PCA), which is typically performed in the tangent space of shape or size-and-shape space [2, Section 7.7]. Principal Nested Spheres (PNS; [3]) provides an alternative by performing a sequence of projections onto nested spherical submanifolds rather than relying on tangent space approximations. An extension to Principal Nested Shape Spaces (PNSS) was given by [1]. However, PNS and PNSS remove scale by normalizing observations to unit size, thereby discarding information about size.

In order to retain scale information [6] propose Principal Nested Cones (PNC), a nonlinear dimension-reduction framework that jointly models size and shape. Through a sequence of nested cones, size is retained as an explicit geometric component throughout the analysis. It leads to a novel representation for size-and-shape data from a scale–structure decomposition, in which a radial component encodes size and a manifold-valued component encodes shape. Moreover, the approach generalizes naturally to broader Euclidean and non-Euclidean settings. Some details are as follows.

Let M be a compact Riemannian manifold with intrinsic geodesic distance $\rho(\cdot,\cdot)$, which ensures well-defined local projections onto submanifolds. A cone $\mathrm{Con}(M)$ of a metric space consists of the product of a metric space (M,ρ) with the non-negative real numbers $[0,\infty)$. Each point on the cone can be written as (x,t) where $x \in M$ and $t \geq 0$, and in addition all apex points $(x,0)$ are identified with each other. The cone is therefore a quotient space $\{M \times [0,\infty)\}/\sim$, with the apex equivalence relation $\sim$. For PNC we construct a nested sequence of submanifolds

$$M = M_d \supset M_{d-1} \supset \cdots \supset M_1,$$

which induces a nested sequence of cone submanifolds

$$\mathrm{Con}(M) = \mathrm{Con}(M_d) \supset \mathrm{Con}(M_{d-1}) \supset \cdots \supset \mathrm{Con}(M_1).$$

To estimate such a sequence of submanifolds, at each stage k, we can minimize the corresponding empirical objective function

$$Q_{n,k} = \sum_{i=1}^{n} r_i^2 \, \rho\left(x_i, \pi_k(x_i)\right)^2,$$

where $(x_i, r_i) \in \mathrm{Con}(M)$, $r_i \geq 0$ is the size of observation x_i, and π_k is the intrinsic projection onto the lower-dimensional submanifold M_{k-1}. Thus, in the general setting, we minimize a size-weighted sum of squared intrinsic distances on the manifold M, where size corresponds to the radial coordinate. This general formulation is first illustrated via the spherical case $M = S^d$, where the cone admits an explicit Euclidean realization via hypercones and the criterion reduces to the sum of squared hypercone residuals.

To enable scalable computation in ultra-high-dimensional settings, we develop a fast approximation combining PCA-based transformation with standard PNC. Simulation studies and real data applications demonstrate that PNC captures nonlinear size-and-shape variation, improves representation and reconstruction, and yields interpretable insights across morphometric, developmental, and molecular datasets.

References

[1] Dryden, I. L., Kim, K., Laughton, C. A., and Le, H. (2019). *Principal nested shape space analysis of molecular dynamics data.* Annals of Applied Statistics, 13:2213–2234.

[2] Dryden, I. L. and Mardia, K. V. (2016). *Statistical Shape Analysis: with applications in R. 2nd edition.* John Wiley & Sons.

[3] Jung, S., Dryden, I. L., and Marron, J. S. (2012). *Analysis of principal nested spheres.* Biometrika, 99(3):551–568.
[4] Marron, J. S. and Dryden, I. L. (2021). *Object Oriented Data Analysis.* CRC Press/Chapman and Hall, Boca Raton.
[5] Wu, Y., Huang, C., and Srivastava, A. (2024). *Shape-based functional data analysis. Test*, 33(1):1–47.
[6] Zhan, Y., Dryden, I. L. and Wu, Y. (2026). *Principal nested cones.* arXiv:2604.19972

Metric Geometry of Spaces of Persistence Diagrams

Fernando Galaz-García
(joint work with Mauricio Che, Fernando Galaz-Fontes, Luis Guijarro, Ingrid Amaranta Membrillo Solis, and Motiejus Valiunas)

Persistence diagrams are a standard tool in topological data analysis (TDA) for summarizing the shape of data across scales. They were introduced by Cohen-Steiner, Edelsbrunner, and Harer as equivalent representations for persistence modules indexed by the positive real numbers [5]. The set of persistence diagrams may be metrized with the *p-Wasserstein metrics*, parametrized by $1 \leq p \leq \infty$ (see [6]), with the metric corresponding to $p = \infty$ also known as the *bottleneck distance*. In this way, one obtains a family of metric spaces $\mathcal{D}_p(\mathbb{R}^2, \Delta)$, where Δ is the diagonal of $\mathbb{R}^2$, placing metric geometry and TDA in a common framework.

Mileyko, Mukherjee and Harer formalized these metric constructions in [7]. They also examined the completeness, separability, and compactness of subsets of the space $\mathcal{D}_p(\mathbb{R}^2, \Delta)$ with the p-Wasserstein metric, $1 \leq p < \infty$, and the existence of Fréchet means for probability distributions on $\mathcal{D}_p(\mathbb{R}^2, \Delta)$. Motivated by the computation of Fréchet means of finite subsets of $\mathcal{D}_p(\mathbb{R}^2, \Delta)$, Turner, Mileyko, Mukherjee, and Harer later showed in [8], among other results, that $\mathcal{D}_2(\mathbb{R}^2, \Delta)$ is an infinite-dimensional Alexandrov space with non-negative curvature. This result established a connection between the still poorly understood geometry of infinite-dimensional Alexandrov spaces and the metric geometry of spaces of persistence diagrams.

The construction of spaces of persistence diagrams $\mathcal{D}_p(\mathbb{R}^2, \Delta)$ may be generalized by considering a metric pair (X, A), where X is a metric space and A is a distinguished nonempty closed subset. In this context, (generalized) persistence diagrams may be viewed as countable multisets of points in X, considered relative to A ([2]; see also [1]).

In this talk, I will discuss the functorial construction $(X, A) \mapsto \mathcal{D}_p(X, A)$, which assigns to each metric pair (X, A) a pointed metric space $\mathcal{D}_p(X, A)$ of persistence diagrams equipped with the p-Wasserstein metric, $1 \leq p \leq \infty$. Several basic properties of the metric space X, such as completeness and geodesicity, are preserved by the functor $\mathcal{D}_p$, with X assumed proper in the geodesic case (see [2, 3]). In particular, if X is a proper Alexandrov space of non-negative curvature and $A \subset X$ is any non-empty closed subset, then $\mathcal{D}_2(X, A)$ is an infinite-dimensional Alexandrov space of non-negative curvature.

I will also discuss more recent work [4] in which the bottleneck and p-Wasserstein distances are recovered as special cases of a broader construction obtained by replacing the usual ℓ^p norm by the norm of a normalized permutation-invariant Banach sequence ideal E. Under natural assumptions on E, the resulting matching formula defines a metric and the associated spaces of persistence diagrams are complete, separable, and geodesic. This enlarges the class of metrics available on spaces of persistence diagrams while preserving the basic geometric properties of the classical cases.

References

[1] P. Bubenik and I. Hartsock, *Topological and metric properties of spaces of generalized persistence diagrams*, J. Appl. Comput. Topol. **8** (2024), no. 2, 347–399.

[2] M. Che, F. Galaz-García, L. Guijarro, and I.A. Membrillo Solis, *Metric geometry of spaces of persistence diagrams*, J. Appl. Comput. Topol. **8** (2024), no. 8, 2197–2246.

[3] M. Che, F. Galaz-García, L. Guijarro, I. Membrillo Solis, and M. Valiunas, *Basic metric geometry of the bottleneck distance*, Proc. Amer. Math. Soc. **152** (2024), no. 8, 3575–3591.

[4] M. Che, F. Galaz-Fontes, and F. Galaz-García, *Banach-ideal metrics on spaces of persistence diagrams*, in preparation, 2026.

[5] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, *Stability of persistence diagrams*, Discrete Comput. Geom. **37**(1)(2007), 103–120.

[6] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and Y. Mileyko, *Lipschitz functions have L_p-stable persistence*, Found. Comput. Math. **10**(2), (2010) 127–139.

[7] Y Mileyko, S Mukherjee, and J Harer, *Probability measures on the space of persistence diagrams*, Inverse Problems **27** (2011), no. 12, 124007, 22.

[8] K. Turner, Y Mileyko, S Mukherjee, and J Harer, *Fréchet means for distributions of persistence diagrams*, Discrete Comput. Geom. **52** (2014), no. 1, 44–70,

Equivariant nonlinear operators

Erlend Grong

(joint work with F. Ballerin)

We are motivated by the following question. Let P be a nonlinear partial differential operator on a Riemannian manifold (M, g). Let G be its group of isometries. We say that P is G-*equivariant* if

$$P(f \circ \varphi) = P(f) \circ \varphi, \qquad \text{for any } f \in C^\infty(M), \varphi \in G.$$

We consider this question for the case where (M, g) is a simply connected Riemannian model space, i.e., Riemannian manifolds with constant sectional curvature that, up to scaling, are isometric to a sphere, the hyperbolic space or the euclidean space. We will also consider nonlinear differential operators that are *polynomial*, meaning that they are linear combinations of operators on the form

$$f \mapsto (L_1 f)(L_2 f) \cdots (L_n f), \qquad \text{with each } L_j \text{ a linear differential operator.} \tag{1}$$

For an operator as in (1), if each L_j is linear order r_j, then we call $k = \max\{r_1, \dots, r_n\}$ *the order*, $p = r_1 + \cdots + r_n$ *the total order* and n *the polynomial degree* of the operator respectively. and extend these definitions by linearity to all polynomial differential operators. We write $\mathrm{NPDO}(M)$ for all nonlinear polynomial differential

operators, with $\mathrm{NPDO}^p(M)$ being the subspace of operators of total order $\leq p$. We write $\mathrm{NPDO}(M)^G$ for the subspace of equivariant operators under the isometry group G.

Our result is that all such operators can be categorized by multigraphs. *A multigraph* γ consists of vertices $V[\gamma] = \{v_1, \ldots, v_n\}$, and edges $E[\gamma]$ that are unordered pairs of vertices, allowing both multiple edges between vertices, as well as loops. We define the degree $\deg(v)$ of a vertex v as the number of edges adjacent to it, with loops counting twice, and define $\nabla^v = \nabla^{\deg(v),\mathrm{Sym}}$ for the corresponding symmetric covariant derivative. We then define a nonlinear differential operator $N([\gamma])$ from an isomorphism class $[\gamma]$ of a multigraph γ by

$$\nabla^{V[\gamma]} f = \nabla^{v_1} f \otimes \cdots \otimes \nabla^{v_n} f, \qquad \text{and} \qquad N([\gamma]) = \mathrm{tr}_{E[\gamma]} \nabla^{V[\gamma]} f,$$

where $\mathrm{tr}_{E[\gamma]}$ takes a trace for each edge $e = \{v_i, v_j\} \in E[\gamma]$ between $\nabla^{v_i} f$ and $\nabla^{v_j} f$. If γ has n vertices with maximal degree k and p edges, then $N([\gamma])$ has order k, total order $2p$ and polynomial degree n. As an example, if $\gamma =$, then

$$N([\gamma]) = \mathrm{tr}(\nabla^{3,\mathrm{Sym}}_{*_1,*_1,*_2} f)(\nabla_{*_2} f) = \langle \nabla \Delta f, \nabla f \rangle - \frac{2}{3} \mathrm{Ric}(\nabla f, \nabla f)$$

Our main result is as follows.

Theorem 1 ([1])**.** *Let* MG *be the collection of all equivalence classes of multigraphs, with subset* MG^p *being those that have at most* p *edges. If* (M, g) *is a simply connected* d*-dimensional complete manifold of constant sectional curvature and isometry group* G*, then*

$$N : \mathrm{span}_{\mathbb{R}} \mathrm{MG} \to \mathrm{NPDO}(M)^G \text{ is surjective.}$$

Furthermore for any

$$p \leq d, \qquad N|_{\mathrm{span}\,\mathrm{MG}^p} : \mathrm{span}_{\mathbb{R}} \mathrm{MG}^p \to \mathrm{NPDO}^{2p}(M)^G \text{ is bijective.}$$

This theorem generalizes a result by Helgason [2] stating that linear equivariant differential operators on the same spaces are polynomials in the Laplacian. The result is motivated by learning PDEs, see, e.g., [4, 3, 5]. The result shows that the number of candidate terms that need to be considered when learning an equivariant partial differential operator is asymptotically constant in the dimension d if we limit ourselves to a maximal total order and polynomial degree.

Studying the kernel of the operator N also gives us nontrivial linear dependence relations between differential operators which exist for lower dimensional manifolds, but in in higher dimensions. The smallest nontrivial case is below.

Theorem 2 ([1])**.** *On any Riemannian manifold of dimension* $d \geq 2$*, the operators*

$$Pf = \|\nabla f\|^2 (\Delta f)^2 - \|\nabla f\|^2 \|\nabla^2 f\|^2 + 2\|\nabla^2 f(\nabla f, \cdot)\|^2 - 2\Delta f \cdot \nabla^2 f(\nabla f, \nabla f),$$
$$Qf = (\Delta f)^3 - 3\Delta f \|\nabla^2 f\|^2 + 2\,\mathrm{tr}\, \nabla^2_{*_1,*_2} f \nabla^2_{*_1,*_3} f \nabla^2_{*_2,*_3} f,$$

are equivariant with respect to isometries, and they are identically zero if and only if $d = 2$*.*

The discussion described the mathematical foundations for these results. We also gave some comments on operators that were equivariant under the identity component of the isometry group as well as under the isometry group of a sub-Riemannian model space.

References

[1] Ballerin, Francesco, and Erlend Grong. *Equivariant nonlinear partial differential operators on constant curvature spaces.* arXiv preprint arXiv:2605.16847 (2026).

[2] Helgason, Sigurður *Differential operators on homogeneous spaces.* Acta Math. **102** (1959), 239–299.

[3] Rudy, Samuel H., et al. *Data-driven discovery of partial differential equations.* Science advances **3**, no. 4 (2017): e1602614.

[4] Brunton, Steven L., Joshua L. Proctor, and J. Nathan Kutz. *Discovering governing equations from data by sparse identification of nonlinear dynamical systems.* Proceedings of the national academy of sciences **113** no. 15 (2016): 3932–3937.

[5] Both, Gert-Jan, et al. *DeepMoD: Deep learning for model discovery in noisy data.* Journal of Computational Physics **428** (2021): 109985.

Stochastic completeness for landmark space

Karen Habermann

(joint work with Stefan Sommer)

In shape analysis, in computational anatomy and for stochastic models of shape evolution the Large Deformation Diffeomorphic Metric Mapping framework models shape variations as diffeomorphic deformations. Important shape spaces within this framework are spaces of landmarks with Riemannian metrics induced from right-invariant metrics on compactly supported diffeomorphisms, see [4]. Shapes are then characterized by $n \geq 2$ ordered distinct landmarks in $\mathbb{R}^d$ for $d \geq 1$.

In diffeomorphic landmark matching, two landmark configurations consisting of the same number of landmarks are compared by solving an optimization problem which minimizes a suitable energy functional associated with flows of compactly supported diffeomorphisms that transform one landmark configuration into the other one. The landmark manifold Land^n of n distinct landmarks in $\mathbb{R}^d$ can be endowed with a Riemannian metric g such that this optimization problem is equivalent to the geodesic boundary value problem for g on Land^n. For modeling stochastic shape evolutions and for performing data statistics on these spaces, it is of fundamental importance to study long-time existence of Brownian motion on the Riemannian landmark manifold (Land^n, g), that is, to analyze stochastic completeness properties of (Land^n, g).

The first progress [1] made in this direction provides a full characterization for long-time existence of Brownian motion for landmark configurations of exactly two landmarks and governed by a radial kernel. For Sobolev metrics one finds the threshold phenomenon that low-order Sobolev metrics allow landmark collision in finite time with positive probability, whereas sufficiently regular Sobolev metrics and Gaussian kernels give long-time existence of Brownian motion. The

results in [2] then establish a full characterization of geodesic completeness for spaces of landmarks with smooth translation- and rotation-invariant metrics for arbitrary numbers of landmarks, and give examples of geodesically complete but stochastically incomplete landmark manifolds.

In the recent work [3], we finally extend the stochastic completeness theory for landmark spaces from two landmarks to arbitrary numbers of landmarks and, for translationally and rotationally invariant kernels, we link stochastic completeness for any number of landmarks to specific properties of the Fourier transform of the kernels. We combine Grigor'yan's volume growth criterion with quantitative controls in terms of the Euclidean distance for geodesic balls in the landmark manifold. We show that in geodesic balls, one can control both the Euclidean size of individual landmarks and the rate at which pairwise landmark distances may approach zero. We then use a lower bound on the minimal eigenvalue of the landmark cometric to upper bound the Riemannian volume growth of geodesic balls.

An important class of translationally and rotationally invariant kernels used in applications arises from the Sobolev differential operator $L_{H^s} = (\mathrm{Id} - \sigma^2\Delta)^s$ for $s > \frac{d}{2}$ and $\sigma > 0$. A consequence of our analysis is the following characterization, which yields the same threshold previously observed in [1] for the two-landmark case.

Theorem. *Suppose the Riemannian metric g on the landmark configuration space* Land^n *arises from the Sobolev operator L_{H^s}. If we have $s \geq 1 + \frac{d}{2}$ then the landmark manifold* (Land^n, g) *is stochastically complete for all $n \geq 2$.*

This threshold observed in the stochastic completeness for cometric kernels arising from Sobolev operators is exactly the same one as for the geodesic completeness, where it is a consequence of [2] that, for any number of landmarks, the manifold (Land^n, g) is geodesically complete if and only if $s \geq 1 + \frac{d}{2}$.

References

[1] Karen Habermann, Philipp Harms, and Stefan Sommer, *Long-time existence of Brownian motion on configurations of two landmarks*, Bulletin of the London Mathematical Society **56** (2024), 1658–1679.

[2] Karen Habermann, Stephen C. Preston, and Stefan Sommer, *Characterization of geodesic completeness for landmark space*, Discrete and Continuous Dynamical Systems, **50** (2026), 224–240.

[3] Karen Habermann and Stefan Sommer, *Stochastic completeness for landmark space*, arXiv:2606.02570 (2026), 1–12.

[4] Sarang C. Joshi and Michael I. Miller, *Landmark matching via large deformation diffeomorphisms*, IEEE Transactions on Image Processing, **9** (2000), 1357–1370.

Gradient flows for template-based reconstruction

Erik Jansson
(joint work with Carola–Bibiane Schönlieb, Ozan Öktem, Jonathan Krook, Klas Modin)

Shape matching refers to the optimisation problem of aligning an initial template shape A to a target shape B by means of an optimal Lie group action. A shape is in turn just an element of a shape space V, which is a space acted upon by the Lie group G. The goal is then, abstractly, to find the group element minimising

$$g \mapsto \mathcal{L}(g.A, B) + \mathcal{R}(g),$$

where $\mathcal{L}\colon V \times V \to \mathbb{R}$ is a suitable fidelity term and $\mathcal{R}$ is a regularisation that penalises unwanted behaviours. In many applications, the group is the infinite-dimensional group of diffeomorphisms of some underlying manifold M, denoted by $\mathrm{Diff}(M)$. One possible choice of regularization, used in *large deformation diffeomorphic metric mapping*, see for instance Bruveris and Holm [2] or Younes [7], is to use

$$\mathcal{R}(g) = d^2\big(g, e\big). \tag{1}$$

Here, $e \in G$ is the identity element and $d\colon G \times G \to \mathbb{R}$ is a distance on G.

A challenge with using the regularisation (1) is to evaluate d, as this involves solving an optimisation problem, so the shape matching problem is a coupled optimisation problem on a Lie group. To solve this, the distance is taken as the distance functional of a right-invariant Riemannian metric. In this case, the minimisation problem can be replaced by finding a minimising curve $\gamma\colon [0,1] \to G$ in the group

$$\gamma \mapsto \mathcal{L}(\gamma(1).A, B) + \lambda \int_0^1 \langle \dot\gamma(t), \dot\gamma(t)\rangle_{\gamma(t)} \mathrm{d}t \quad \text{such that } \gamma(0) = e, \tag{2}$$

with $\langle \cdot, \cdot \rangle$ denoting a right-invariant metric on G. This, however, requires working with curves in the group, or generating them from a curve of vector fields, and therefore incurs additional computational cost. It turns out that the minimisation problem is equivalent to the Euler–Poincaré equation on the Lie algebra coupled to a reconstruction equation for the group curve. The actual optimisation is performed by shooting: given a first guess of the initial velocity of the curve γ, integrate the governing Euler–Poincaré equation and flow equation to obtain the corresponding curve γ, calculate the gradient of the functional with respect to the initial velocity, and take a gradient descent update step. This is iterated until convergence, and each gradient descent step requires solving two differential equations.

To avoid this, Balehowsky et al. [1] suggested to solve the matching problem via a gradient flow directly on the group. This offers some benefits, mainly that we now solve one differential equation on the group, rather than having to solve two equations each gradient descent update. A central result in Balehowsky et al. [1] is that if the group is equipped with a right-invariant metric and the group

action of G on V has a momentum map (see for instance Marsden and Ratiu [6]), the gradient flow is straightforward to write down.

In this talk, we are interested in *indirect matching*, namely, the case when the target shape B is replaced by observed data y in some data space Y. The geometric setup changes very little: indeed, all we require is a *forward model* $\mathcal{F}$ that takes us from V to Y, and models how data in Y is generated from data in V. The optimisation problem then becomes

$$g \mapsto \mathcal{L}(\mathcal{F}(g.A), y) + \mathcal{R}(g).$$

The found g thus produces a deformed template $g.A$, which should be close to whatever element of V that generates the data. Thus, we obtain a reconstruction that incorporates the bias that whatever is reconstructed is a deformation of A. If the template is chosen appropriately, this allows for incorporation of an informative anatomical prior which may aid reconstruction performance (see for instance Lang et al. [5] for an example of this in the context of tomographic reconstruction). Also in the indirect matching setting, two routes can be taken, either the LDDMM approach, which was done in for instance Chen and Öktem [3] and Lang et al. [5], or the gradient flow route, which is the topic of this talk. We present an application of the gradient flow, namely reconstruction from cryo-EM data, i.e., the reconstruction of protein conformations from noisy images. This is a challenging task, but one well-suited for template-based reconstruction as we can incorporate prior knowledge of which protein is being imaged, and use as a template a reference conformation. In this setting, we use a coarse-grained protein model, meaning that we can work with a *finite-dimensional group*. The LDDMM approach was explored in Jansson et al. [4], and while it produces good reconstructions, the repeated solving of the governing differential equation slows the approach down, rendering it less useful for heterogeneous cryo-EM problems. In this talk, we showcase the applicability of the gradient flow approach, and show that it means faster convergence than the LDDMM approach and opens up for reconstruction in more complicated settings, for instance for reconstruction of multi-chain proteins or simple problems with conformational heterogeneity.

References

[1] T. Balehowsky, C.-J. Karlsson, and K. Modin, "Shape analysis via gradient flows on diffeomorphism groups," *Nonlinearity*, vol. 36, p. 862, 2022. doi: 10.1088/1361-6544/aca73c.

[2] M. Bruveris and D. D. Holm, "Geometry of Image Registration: The Diffeomorphism Group and Momentum Maps," in *Geometry, Mechanics, and Dynamics*, Springer, New York, 2015, pp. 19–56. doi: 10.1007/978-1-4939-2441-7_2.

[3] C. Chen and O. Öktem, "Indirect Image Registration with Large Diffeomorphic Deformations," *SIAM Journal on Imaging Sciences*, vol. 11, no. 1, pp. 575–617, 2018. doi: 10.1137/17M1134627.

[4] E. Jansson, J. Krook, K. Modin, and O. Öktem, "Geometric shape matching for recovering protein conformations from single-particle cryo-EM data," *SIAM Journal on Imaging Sciences*, vol. 18, no. 4, pp. 2509–2546, 2025. doi: 10.1137/24M1704336.

[5] L. F. Lang, S. Neumayer, O. Öktem, and C.-B. Schönlieb, "Template-Based Image Reconstruction from Sparse Tomographic Data," *Applied Mathematics & Optimization*, vol. 82, no. 3, pp. 1081–1109, 2019. doi: 10.1007/s00245-019-09573-2.

[6] J. E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, Texts in Applied Mathematics, Springer, New York, 1999. doi: 10.1007/978-0-387-21792-5.

[7] L. Younes, *Shapes and Diffeomorphisms*, Applied Mathematical Sciences, Springer, Berlin Heidelberg, 2010. doi: 10.1007/978-3-642-12055-8.

Fréchet p-functions and Fréchet p-means

Huiling Le

(joint work with Karthik Bharath and Andrew T.A. Wood)

Let $\boldsymbol{M}$ be a complete and simply connected Riemannian manifold (without boundary) of dimension d, and ρ be the distance function on $\boldsymbol{M}$ induced from the Riemannian metric. Also, let μ be a probability measure on $\boldsymbol{M}$ and $p \geq 1$ be a constant.

The Fréchet p-function $F_{\mu,p}$ on $\boldsymbol{M}$, associated with μ and p, is defined as

$$F_{\mu,p} : x \mapsto \frac{1}{p} \int_{\boldsymbol{M}} \rho(x, x')^p d\mu(x').$$

A point $\hat{x}_p \in \boldsymbol{M}$ is a Fréchet p-mean of μ if it minimises $F_{\mu,p}$. In particular, a Fréchet 1-mean is a Fréchet median and a Fréchet 2-mean is commonly known simply as a Fréchet mean.

In the first part of the talk, we discuss the conditions required on μ to ensure the existence of the gradient and Hessian of $F_{\mu,p}$ at any given point x, and also find their expressions. These conditions on μ, as well as the term that is additional to the expected Hessian of $\rho(x, x')^p/p$ in the expression for the Hessian of $F_{\mu,p}$, are primarily related to the cut locus of x. It turns out that, for the case $d = 1$, the Hessian of $F_{\mu,1}$ requires a separate investigation, due to the different behaviour of $F_{\mu,1}$ on a 1-dimensional manifold.

In the second part, we discuss some probabilistic and statistical properties of Fréchet p-means. In particular, the results in the first part of the talk enable us to apply the so-called δ-method to obtain the Central Limit Theorem for Fréchet p-means under some mild conditions, and to investigate the influence function of Fréchet p-means. The latter will play an important role in our future study of the robustness of Fréchet p-means.

This work builds upon much previous work of various authors, some of which are listed below.

References

[1] T. Hotz and S.F. Huckemann, *Intrinsic means on the circle: uniqueness, locus and asymptotics*, The Annals of the Institute of Statistical Mathematics **67** (2015), 177–193.

[2] T. Hotz, H. Le and A.T.A. Wood, *Central limit theorem for intrinsic Fréchet means in smooth compact Riemannian manifolds*, Probability Theory and Related Fields **189** (2024), 1219–1246.

[3] H. Le and D. Barden, *On the measure of the cut locus of a Fréchet mean*, Bulletin of London Mathematical Society **46** (2014), 698–708.

[4] A. Lytchak and S.F. Huckemann, *Zero probability of the cut locus of a Fréchet mean on a Riemannian manifold*, arXiv:2508.00747v1 (2025).
[5] X. Pennec, *Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements*, Journal of Mathematical Imaging and Vision **25** (2006), 127–154.
[6] L. Yang, *Some properties of Fréchet medians in Riemannian manifolds*, arXiv:1110.3899v2 (2011).

Principal Component Analysis in the Wasserstein space

Alice Le Brigant
(joint work with Nina Vesseron, Elsa Cazelles and Thierry Klein)

In this talk we discuss how to perform Principal Component Analysis (PCA) of a dataset whose elements are probability distributions. The goal is to identify (geodesic) curves in the space of probability measures that best capture the modes of variation of the underlying dataset.

A natural strategy is to identify the probability measures with their probability density functions and to perform PCA on these using the L^2 Hilbert metric. Unfortunately, as highlighted in [2], the components computed in this manner fail to capture the intrinsic structure of the dataset: the projections onto the components most likely result in non-positive and un-normalized functions.

Using the 2-Wasserstein metric instead has proven to overcome these limitations by taking into account the geometry of the space of distributions. The 2-Wasserstein metric endows the space of probability distributions with a Riemannian-like structure, framing the problem as PCA on a (positively) curved Riemannian manifold.

A first approach to solve this task, known as Tangent PCA (TPCA), consists in embedding the data into the tangent space at a reference point, and applying classical PCA in this flat space, as in [3]. TPCA is computationally advantageous but can generically induce distortion in the embedded data, depending on the curvature of the manifold at the reference point and the dispersion of the data.

A more geometrically coherent approach is Geodesic PCA (GPCA) proposed for Riemannian manifolds in [5, 4], where principal modes of variations are geodesics that minimize the variance of the projection residuals. Following this approach, the first geodesic component of a set of probability measures $\nu_1, \dots, \nu_n$ in the Wasserstein space solves

$$\inf_{t \mapsto \mu(t) \text{ geodesic}} \sum_{i=1}^{n} \inf_{t_i} W_2^2(\mu(t_i), \nu_i). \tag{1}$$

The second principal component is defined to be the geodesic that minimizes the same cost function, with the constraint of intersecting the previous component orthogonally. The subsequent principal components have the additional constraint of going through the intersection of the first two principal geodesics.

In this talk, we introduce two algorithms to solve the exact GPCA problem in the Wasserstein space of (i) centered Gaussian distributions and (ii) a.c. probability measures on $\mathbb{R}^d$. Our methods are exact in the sense that they do not

rely on a linearization of the Wasserstein space, and the components are true geodesics that minimize the cost in (1). In the Gaussian case, we leverage the Bures-Wasserstein geometry [1] to lift the computations to the flat space of invertible matrices. We show examples where GPCA and TPCA differ significantly, and relate this effect to curvature. In the general case of a.c. probability distributions, we lift the probability distributions to the space of (non necessarily optimal) maps that pushforward a given reference measure, as described by Otto in [6]. This approach is independent of the chosen reference measure and yields a convenient way to parametrize geodesic components and define orthogonality with respect to the Wasserstein metric. In practice, we parametrize geodesic components using multilayer perceptrons (MLPs), trained to minimize the cost in (1). We show illustrations on images and 3D point clouds. Along the way, we prove that for univariate Gaussian distributions, GPCA yields the same results whether it is performed in the space of a.c. distributions or restricted to the Gaussian submanifold.

The work described in this talk corresponds to [7]. In future work, it would be interesting to modify the GPCA cost so that the residuals are computed from the projection onto the subspace generated by all the previous components, and not just the projections on the current geodesic component, so as to allow for better reconstruction.

References

[1] R. Bhatia, T. Jain, and Y. Lim, *On the Bures–Wasserstein distance between positive definite matrices*, Expositiones mathematicae **37.2** (2019): 165–191.

[2] E. Cazelles, Elsa, V. Seguy, J. Bigot, M. Cuturi and N. Papadakis. *Geodesic PCA versus log-PCA of histograms in the Wasserstein space*, SIAM Journal on Scientific Computing **40.2** (2018): B429-B456.

[3] T. Fletcher, C. Lu, and S. Joshi, *Statistics of shape via principal geodesic analysis on Lie groups*, 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. Vol. 1. IEEE, 2003.

[4] S. Huckemann, T. Hotz, and A. Munk, *Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions*, Statistica Sinica (2010): 1–58.

[5] S. Huckemann and H. Ziezold, *Principal component analysis for Riemannian manifolds, with an application to triangular shape spaces*, Advances in Applied Probability **38.2** (2006): 299–319.

[6] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Communications in Partial Differential Equations **26** (2001), 101–174.

[7] N. Vesseron, E. Cazelles, A. Le Brigant and T. Klein, *On the Wasserstein geodesic principal component analysis of probability measures*, arXiv preprint arXiv:2506.04480, 2025.

On Periodic Geodesics on Half–Lie Groups

Levin Maier
(joint work with Martin Bauer)

Infinite-dimensional Riemannian geometry differs from the finite-dimensional theory in several basic ways: Hopf–Rinow fails in general [4], weak metrics may have vanishing geodesic distance [5], and a Levi–Civita connection need not exist. Nevertheless, certain mapping groups retain enough structure for global Riemannian arguments to survive. The results reported here, obtained in joint work with Martin Bauer [1], show that topological mechanisms forcing closed geodesics in finite dimensions still produce periodic geodesics on a broad class of infinite-dimensional groups.

A Hilbert half–Lie group is a topological group carrying a Hilbert manifold structure such that all right translations are smooth. A Riemannian metric $\mathcal{G}$ is called strong if the musical map

$$\mathcal{G}^\flat : TG \longrightarrow T^*G$$

is a bundle isomorphism. We focus on strong right-invariant metrics, for which recent completeness results for half–Lie groups provide an infinite-dimensional substitute for the compactness and completeness input used in the classical finite-dimensional theory.

Periodic geodesics from the fundamental group. The first main result is the analogue of minimizing the energy in a nontrivial free homotopy class, in the spirit of classical closed-geodesic theory [6, 7, 8].

Theorem 1 (Closed geodesics in nontrivial homotopy classes, [1, Theorem 3.1])**.** *Let $(G, \mathcal{G})$ be a Hilbert half–Lie group with a strong right-invariant metric. Under mild regularity and closedness assumptions, if $\pi_1(G) \neq 0$, then*

$$0 \neq [\alpha] \in \pi_1(G) \quad \Longrightarrow \quad [\alpha] \text{ contains a nonconstant closed geodesic.}$$

If $[\alpha]$ is primitive, this geodesic can be chosen prime. If $\dim G = \infty$, then right translates give infinitely many geometrically distinct periodic geodesics.

The proof has two complementary forms: a direct minimization argument in the prescribed homotopy class after normalization by right translation, and a universal-cover argument in which a minimizing geodesic from the identity to the relevant deck transformation projects to the required closed geodesic.

A Lyusternik–Fet theorem. The second result treats contractible periodic geodesics, where the relevant topology is higher homotopy rather than π_1. Let $\Lambda_0 G$ denote the component of contractible H^1-loops in G, and let E be the Riemannian energy. A nonzero class $[\Gamma] \in \pi_{q-1}(\Lambda_0 G)$ determines the minimax value

$$c(\Gamma) = \inf_{\gamma \in [\Gamma]} \sup_{\xi \in S^{q-1}} E(\gamma(\xi)).$$

Because the metric is right-invariant, compactness has to be imposed modulo right translations. This is essential rather than cosmetic: for noncompact G, the

ordinary Palais–Smale condition fails, since right translates of any critical loop have the same energy and need not contain a convergent subsequence.

Theorem 2 (Lyusternik–Fet theorem for half–Lie groups, [1, Theorem 3.9])**.** *Let $(G, \mathcal{G})$ be a half–Lie group equipped with a strong right-invariant Riemannian metric. Assume that $\pi_q(G) \neq 0$ for some $q \geq 2$, and let $0 \neq [\Gamma] \in \pi_{q-1}(\Lambda_0 G)$ be the induced class. If the energy functional satisfies the Palais–Smale condition at the level $c(\Gamma)$ modulo right translations, then $(G, \mathcal{G})$ admits a nonconstant contractible periodic geodesic of energy $c(\Gamma) > 0$. If $\dim G = \infty$, then right translates give infinitely many geometrically distinct contractible periodic geodesics.*

The proof is the classical minimax scheme with quotient-type compactness: one first proves $c(\Gamma) > 0$, obtains a Palais–Smale sequence at this level, and then uses compactness modulo translations to extract a critical point of E.

Finite-dimensional reduction. A third contribution is a transfer principle from compact finite-dimensional totally geodesic submanifolds. If a half–Lie group G carries a right-invariant metric whose Levi–Civita connection exists, and if G contains a compact finite-dimensional totally geodesic submanifold $\bar{M}$, then closed geodesics of $\bar{M}$ are closed geodesics of G. Hence nontrivial $\pi_1(\bar{M})$ yields closed geodesics in the corresponding nontrivial classes, while nontrivial higher homotopy of $\bar{M}$ yields contractible periodic geodesics. This reduction is useful also for weak right-invariant metrics, where the strong-metric variational hypotheses above may not be available.

Applications. First, strong Sobolev metrics on $\mathrm{Diff}^s(M)$ provide examples for Theorem 1. Let (M, g) be closed, let $s > n/2 + 1$, and let L be a positive, elliptic, self-adjoint pseudo-differential operator of order $2s$. Then the associated right-invariant H^s-metric is strong, and if $\pi_1(\mathrm{Diff}^s(M)) \neq 0$, every nontrivial homotopy class contains infinitely many geometrically distinct nonconstant periodic geodesics. This applies, for example, to $\mathbb{S}^1$, $\mathbb{S}^2$, tori $\mathbb{T}^n$, and several principal $\mathbb{S}^1$-bundles such as lens spaces whose fibre class is nontrivial.

Second, the finite-dimensional reduction applies to Sobolev diffeomorphism groups through isometry groups. If $K = \mathrm{Isom}_0(M, g)$ and the inertia operator L is K-equivariant and preserves Killing fields, then K is totally geodesic in $(\mathrm{Diff}^s(M), \mathcal{G}^L)$. Thus nontrivial $\pi_1(K)$ gives periodic geodesics in the corresponding nontrivial classes, while nontrivial $\pi_q(K)$ for some $q \geq 2$ gives contractible periodic geodesics. The same statement holds for the volume-preserving Sobolev diffeomorphism group. This covers standard spheres, flat tori, complex projective spaces with the Fubini–Study metric, and many compact homogeneous or symmetric spaces.

For strong Sobolev metrics on diffeomorphism groups, the corresponding Palais–Smale condition remains open. Finally, this hypothesis can be verified in a genuinely infinite-dimensional model: for $G = H \times K$, with H an infinite-dimensional separable Hilbert space and K a compact non-aspherical Lie group, a suitable strong right-invariant metric satisfies the Palais–Smale condition modulo right

translations. Hence Theorem 2 yields a nonconstant contractible periodic geodesic. For generic coupling this geodesic is not obtained from a finite-dimensional totally geodesic subgroup, showing that the minimax theorem is not merely formal.

References

[1] M. Bauer and L. Maier, *On periodic geodesics on half–Lie groups*, preprint, 2026.
[2] M. Bauer, P. Harms, and P. W. Michor, *Regularity and completeness of half–Lie groups*, J. Eur. Math. Soc., 2025.
[3] M. Bruveris and F.-X. Vialard, *On completeness of groups of diffeomorphisms*, J. Eur. Math. Soc. **19** (2017), 1507–1544.
[4] C. J. Atkin, *The Hopf–Rinow theorem is false in infinite dimensions*, Bull. London Math. Soc. **7** (1975), 261–266.
[5] P. W. Michor and D. Mumford, *Vanishing geodesic distance on spaces of submanifolds and diffeomorphisms*, Doc. Math. **10** (2005), 217–245.
[6] L. A. Lyusternik and A. I. Fet, *Variational problems on closed manifolds*, Dokl. Akad. Nauk SSSR (N.S.) **81** (1951), 17–18.
[7] D. Gromoll and W. Meyer, *Periodic geodesics on compact Riemannian manifolds*, J. Differential Geom. **3** (1969), 493–510.
[8] W. Klingenberg, *Lectures on Closed Geodesics*, Springer, 1978.
[9] T. Marquis and K.-H. Neeb, *Half–Lie groups*, Transform. Groups **23** (2018), 801–840.

Not seeing Euclidean motions on landmark space: An LDDMM-metric on $\mathrm{Land}^n(\mathbb{R}^d)/\,\mathrm{SE}(d)$

Peter W. Michor

(joint work with Sarang Joshi, Peter W. Michor and Stefan Sommer)

Abstract: Motivated by Kendall's landmark shape spaces, which remove translation, rotation, and scale from landmark configurations equipped with Euclidean geometry, we present an LDDMM-type Riemannian metric on the diffeomorphism group $\mathrm{Diff}_c(\mathbb{R}^d)$ which induces Riemannian metric on $\mathrm{Land}^n(\mathbb{R}^d)/SE(d)$ whose geodesics are easily computable as a Hamiltonian system on its cotangent space without the need to invert the kernel matrix. The main ingredient is a screened (regularized) elasticity operator on the space of vector fields with compact support whose nullspace (when extended) consists exactly of the infinitesimal rigid motions. The construction also works on other shape spaces modulo rigid motions like $\mathrm{Emb}(M,\mathbb{R}^d)/\,\mathrm{Diff}(M)$ for compact M or spaces of varifolds, or direction fields. This talk is based on [2].

Infinitesimal rigid motions and the elasticity operator. Let $\mathrm{SE}(d) = \{A(x) = R(x) + b\} = \mathrm{SO}(d) \ltimes \mathbb{R}^d$ be the special Euclidean group of rigid motions of $\mathbb{R}^d$ with Lie algebra $\mathfrak{se}(d) = \mathfrak{so}(d) \ltimes \mathbb{R}^d = \{A.x + a\}$.

The classical elasticity operator, also known as the Killing operator, $\operatorname{sym}\nabla : C^\infty(\mathbb{R}^d,\mathbb{R}^d) \to C^\infty(\mathbb{R}^d, S^2\mathbb{R}^d)$ is given by $\operatorname{sym}\nabla V = (\partial_i V^j + \partial_j V^i)_{i,j=1}^d$. It has the infinitesimal rigid motions as its null-space.

We can identify the corresponding inertia operator by partial integration

$$\int_{\mathbb{R}^d} \operatorname{Tr}\big(\operatorname{sym}\nabla V.\operatorname{sym}\nabla w\big)\,dx = -\int_{\mathbb{R}^n} \big\langle \operatorname{div}\operatorname{sym}\nabla V.w\big\rangle\,dx$$

where in the last line div acts on each column (equiv. each line) of $\operatorname{sym}\nabla V$. We see that (a) if $\xi \neq 0$ then the quadratic form in $\hat{V}(\xi)$ is strictly positive definite, so the symbol is elliptic for $\xi \neq 0$; (b) if the integral vanishes then $\hat{V}$ has support at $\xi = 0$ and then V is an infinitesimal rigid motion, a Killing field.

From this we define the operator $L_{\operatorname{sym}} = -\operatorname{div}\circ\operatorname{sym}\nabla$ which is elliptic on the orthogonal complement of its null-space which coincides exactly with the Lie algebra $\mathfrak{se}(d)$ of infinitesimal rigid motions. Up to a factor this is the Helmholtz operator of linear elasticity, i.e.,

$$L_{\operatorname{sym}}V = -2(\Delta V + \nabla\operatorname{div}V) \quad\text{or}\quad L_{\text{Lamé}}V = -\mu\Delta V - (\mu+\lambda)\nabla\operatorname{div}V)$$

By [3, 5, 4] $L_{\text{Lamé}}$ has Green's function on the orthogonal complement of infinitesimal rigid motions in the form of the Kelvin matrix Φ given by

$$\text{In } \mathbb{R}^2: \qquad \Phi(x) = \frac{1}{4\pi\mu(2\mu+\lambda)}\left((3\mu+\lambda)\log\frac{1}{|x|}\operatorname{Id}_2 + (\mu+\lambda)\frac{xx^T}{|x|^2}\right)$$

$$\text{In } \mathbb{R}^3: \qquad \Phi(x) = \frac{1}{8\pi\mu(2\mu+\lambda)}\left((3\mu+\lambda)\frac{1}{|x|}\operatorname{Id}_3 + (\mu+\lambda)\frac{xx^T}{|x|^3}\right)$$

Landmark space modulo rigid motions. $\operatorname{SE}(d)$ acts diagonally on landmark space $\operatorname{Land}^n(\mathbb{R}^d)$ as a proper action. For $q \in \operatorname{Land}^n(\mathbb{R}^d)$, the infinitesimal action is $\zeta_q : \mathfrak{se}(d) \to T_q\operatorname{Land}^n(\mathbb{R}^d)$, given by $\zeta_q(X,a) = (Xq_1 + a, \dots, Xq_n + a)$ for $(X,a) \in \mathfrak{se}(d) = \mathfrak{so}(d) \ltimes \mathbb{R}^d$. The tangent space to the$\operatorname{SE}(d)$-orbit through q is

$$\mathfrak{se}(d)(q) := \zeta_q\big(\mathfrak{se}(d)\big) = \big\{(Xq_i + a)_{i=1}^n : (X,a) \in \mathfrak{se}(d)\big\} \subset T_q\operatorname{Land}^n(\mathbb{R}^d).$$

Its annihilator $\mathfrak{se}(d)(q)^{\text{ann}}$ consists of all $\alpha \in T_q^*\operatorname{Land}^n(\mathbb{R}^d)$ with $\sum_{i=1}^n \alpha_i(Xq_i + a) = 0$ for all $(X,a) \in \mathfrak{se}(d)$, equivalently $\sum_{i=1}^n \alpha_i = 0$ and $\sum_{i=1}^n (q_i\alpha_i^T - \alpha_i q_i^T) = 0$.

The action of the translation subgroup is free, and the quotient by all translations is the space of centered landmarks $\operatorname{Land}_0^n(\mathbb{R}^d) := \{q = (q_i) : C(q) := \frac{1}{n}\sum_i q_i = 0\}$ where $\operatorname{SO}(d)$ is still acting; but it is not invariant under the action of $\operatorname{Diff}_{\mathcal{A}}(\mathbb{R}^d)$. We have the centroid mapping, the span mapping, and the rank of a landmark

$$C(q) = \bar{q} = (q_i - C(q))_{i=1}^n = t_{C(q)}^{-1}(q), \quad \text{where } t_b \text{ is translation by } b,$$

$$\operatorname{Span} : \operatorname{Land}^n(\mathbb{R}^d) \to \bigcup_{k=1}^d \operatorname{Gr}_n(k), \quad \operatorname{rk}(q) := \dim(\operatorname{Span}(\bar{q})),$$

which maps a landmark q to the linear subspace of $\mathbb{R}^d$ spanned by its centered translate $\bar{q} := (q_i - C(q))_{i=1}^n = t_{C(q)}^{-1}(q)$.

For a landmark q, since $\bar{q}$ contains a basis of $\operatorname{Span}(\bar{q})$, the stabilizer or isotropy group $\operatorname{SE}(d)_q \subset \operatorname{SE}(d)$ is $\operatorname{SE}(d)_q = t_{C(q)}.(\operatorname{SO}(\operatorname{Span}(\bar{q})^\perp) \times \{\operatorname{Id}_{\operatorname{Span}(\bar{q})}\}.t_{C(q)}^{-1}$, conjugate to $\operatorname{SO}(d-k) \times \{\operatorname{Id}_k\}$ where $k = \operatorname{rk}(q)$. The conjugacy class of the isotropy

subgroup in $\mathrm{SE}(d)$ depends only on $k = \dim(\mathrm{Span}(\bar{q}))$ and the orbit space is the union of strata $\overline{\mathrm{Land}}^n(\mathbb{R}^d) := \mathrm{Land}^n(\mathbb{R}^d)/\,\mathrm{SE}(d) = \mathrm{Land}^n_0(\mathbb{R}^d)/\,\mathrm{SO}(d) =: \bigcup_k \mathcal{S}_k$ with $\dim(S_k) = nk - \frac{1}{2}k(k+1)$. By the general theory the regular stratum is connected, open and dense. Moreover, $\mathcal{S}_{k'}$ is contained in the boundary of $\mathcal{S}_k$ whenever $k' < k$: move some landmark components so that the rank drops.

The screened elasticity operator $L_s = -(\mathrm{Id} - \sigma^2\Delta)^{s-2} \circ \mathrm{div} \circ \mathrm{sym}\,\nabla$ inherits the null-space $\mathfrak{se}(d)$ of the elasticity operator. Its zero order term does not see infinitesimal rigid motions. The Green's kernel of L_s is obtained by convolving the Kelvin matrix Φ with the scalar Sobolev kernel $k_{H^{s-2}}$, i.e. $k_s = k_{H^{s-2}} * \Phi$. Closed form solutions for operators of the form $a(\Delta)L_{\mathrm{sym}}$ depend on the roots of the polynomial $a(\Delta)$. In the case $s = 3$, the Green's kernel in dim 2 is given by

$$\begin{aligned}(k_2)_{ij}(x) &= -\frac{1}{2\pi\mu}\,\delta_{ij}\,(\log r + k_0(r/\sigma)) \\ &+ \left(\frac{1}{2\pi\mu} - \frac{1}{2\pi(\lambda+2\mu)}\right)\partial_i\partial_j\left(\frac{r^2}{4}\log r - \frac{r^2}{4} + \sigma^2\,(\log r + k_0(r/\sigma))\right)\end{aligned}$$

where k_0 denotes the modified Hankel function and $r = |x|$. In dimension 3, the fundamental solution is (σ is the root of $a(\Delta)$)

$$\begin{aligned}(k_2)_{ij}(x) = \frac{1}{4\mu}\delta_{ij}\left(\frac{1}{r} - \frac{e^{-r/\sigma}}{r}\right) + \\ + \left(\frac{1}{4\pi}\frac{1}{\lambda+2\mu} - \frac{1}{4\mu}\frac{1}{\mu}\right)\partial_i\partial_j\left(\frac{r}{2} + \sigma^2\left(\frac{1}{r} - \frac{e^{-r/\sigma}}{r}\right)\right)\end{aligned}$$

For $s > 3$, the polynomial $a(\Delta)$ has multiple roots. See [6], e.g.

The kernel k_s is conditionally positive definite with respect to $\mathfrak{se}(d)$: for $q = (q_i)$, the quadratic form $(\alpha, \beta) \mapsto \sum_{i,j=1}^n \langle \alpha_i, k_s(q_i - q_j)\beta_j\rangle$ is positive definite on $\alpha, \beta \in T_q^*\,\mathrm{Land}^n$ satisfying $\sum_{i=1}^n \langle \alpha_i, X(q_i)\rangle = 0$ and $\sum_{i=1}^n \langle \beta_i, X(q_i)\rangle = 0$. for all $X \in \mathfrak{se}(d)$; i.e., pullbacks of covectors from $\mathrm{Land}^n(\mathbb{R}^d)/\,\mathrm{SE}(d)$.

These are exactly the rigid-motion constraints that appear when passing from $\mathrm{Land}^n(\mathbb{R}^d)$ to the quotient $\mathrm{Land}^n(\mathbb{R}^d)/\,\mathrm{SE}(d)$.

Metrics and geodesics on $\mathrm{Land}^n(\mathbb{R}^d)/\,\mathrm{SE}(d)$**.** Let $\bar{\alpha} \in T^*_{\pi(q)}\big(\mathrm{Land}^n(\mathbb{R}^d)/\,\mathrm{SE}(d)\big)$ be a covector. Its pullback $\tilde{\alpha} = \pi^*_{\mathrm{Land}}\bar{\alpha} \in T_q^*\,\mathrm{Land}^n(\mathbb{R}^d)$ by the projection π_{Land} satisfies $\langle \tilde{\alpha}, v\rangle = \tilde{\alpha}^T v = \bar{\alpha}(T_q\pi_{\mathrm{Land}}v)$ for all $v \in T_q\,\mathrm{Land}^n(\mathbb{R}^d)$. Moreover $\sum_{i=1}^n \langle \alpha_i, X(q_i)\rangle = 0$ for all $X \in \mathfrak{se}(d)$. The kernel is strictly positive definite on $\mathfrak{se}(d)(q)^{\mathrm{ann}}$ and therefore defines a cometric on $\mathrm{Land}^n(\mathbb{R}^d)/\,\mathrm{SE}(d)$. By inversion, this cometric induces a positive definite inner product on $T_{\pi(q)}\big(\mathrm{Land}^n(\mathbb{R}^d)/\,\mathrm{SE}(d)\big)$.

We can use general results (see [1]) on the induced metric on orbit spaces to get information about $\mathrm{Land}^n(\mathbb{R}^d)/SE(d) \cong \mathrm{Land}^n_0(\mathbb{R}^d)/\,\mathrm{SO}(d)$

Geodesics. The screened elasticity kernel k_s gives a Hamiltonian

$$H(q, \alpha) = \frac{1}{2}\sum_{i,j=1}^n \langle \alpha_i, k_s(q_i - q_j)p_j\rangle$$

on the cotangent bundle $T^*\,\mathrm{Land}^n(\mathbb{R}^d)$, similarly to the Hamiltonian for the positive definite metrics. If the initial momentum is a pullback of a covector in

$T^*_{\pi(q)}\big(\operatorname{Land}^n(\mathbb{R}^d)/\operatorname{SE}(d)\big)$, then it is in $\mathfrak{se}(d)(q)^{\mathrm{ann}}$, and, as shown in the proposition below, it remains in $\mathfrak{se}(d)(q)^{\mathrm{ann}}$ along the Hamiltonian flow.

The flow lines on $T^*\operatorname{Land}^n(\mathbb{R}^d)$ with initial momentum in $\mathfrak{se}(d)(q(0))^{\mathrm{ann}}$ project to horizontal geodesics on $\operatorname{Land}^n(\mathbb{R}^d)$ and hence to geodesics on $\operatorname{Land}^n(\mathbb{R}^d)/\operatorname{SE}(d)$ as long as they do not cross more singular strata.

Proposition. *For $(X,b) \in \mathfrak{se}(d)$, define the momentum map $J_{(X,b)}(q,\alpha) := \sum_{i=1}^n \alpha_i(Xq_i + b) \in \mathbb{R}$. Then the function $J_{(X,v)}$ is preserved along the Hamiltonian flow of H. In particular, if $\alpha(0) \in \mathfrak{se}(d)(q(0))^{\mathrm{ann}}$, then $\alpha(t) \in \mathfrak{se}(d)(q(t))^{\mathrm{ann}}$ for all t for which the flow exists.*

This follows from Emmy Noether's theorem.

References

[1] D. Alekseevsky, A. Kriegl, M. Losik, and P. W. Michor. *The Riemannian geometry of orbit spaces—the metric, geodesics, and integrable systems Publ. Math. Debrecen*, 62(3-4):247–276, 2003. Dedicated to Professor Lajos Tamássy on the occasion of his 80th birthday.

[2] S. Joshi, P. W. Michor, and S. Sommer. *Quotients of landmark spaces with induced invariant metrics.* in preparation, 2026.

[3] J. E. Marsden and T. J. R. Hughes. *Mathematical Foundations of Elasticity.* Dover Publications, New York, 1994.

[4] W. C. H. McLean. *Strongly Elliptic Systems and Boundary Integral Equations.* Cambridge University Press, Jan. 2000.

[5] J. Nečas and I. Hlaváček, editors. *Mathematical Theory of Elastic and Elasto-Plastic Bodies.* Feb. 2017.

[6] D. Rogula. *Some basic solutions in strain gradient elasticity theory of an arbitrary order.* Archiwum Mechaniki Stosowanej, 25:43–68, 1973.

On Stability of Barycenters and Modes of Probability Measures

Washington Mio

(joint work with Tom Needham)

We address the problem of stable representation of the barycenters and modes of probability distributions on a Polish metric space. The challenge comes, in part, from the fact that even a small perturbation to a distribution can lead to a large change in Fréchet mean sets on spaces as simple as a circle. We address this issue by introducing a merge tree representation of barycenters, which takes the form of a merge tree equipped with a height function and a metric-measure structure. Modes are treated as special cases of barycenters through the use of generalized forms of the Fréchet variance function. We prove that barycentric merge trees are stable with respect to optimal transport metrics and describe an algorithmic approach for approximating these merge trees.

Let (X,d) be a connected and locally path connected Polish metric space and $\theta\colon X \times X \to \mathbb{R}$ a pseudo-metric. We say that θ is *L-admissible*, $L > 0$, if $\theta(x,y) \leq Ld(x,y)$, $\forall x,y \in X$. (In applications to mode analysis, θ is typically a diffusion distance derived from a kernel function [1], such as the heat kernel of a Riemannian

manifold, thus providing a multiscale view of modes.) The *p-deviation function* of a Borel probability measure μ on X with finite p-moments, $p \geq 1$, is defined as

$$\sigma_p(x) = \Big(\int_X \theta^p(x,y)\, d\mu(y) \Big)^{1/p},$$

the pth root of the Fréchet p-variance function of μ with respect to θ. The merge tree T_p is the quotient space of X under the equivalence relation $x \sim y$ if there is $t \in \mathbb{R}$ such that $\sigma_p(x) = \sigma_p(y) = t$ and x and y lie in the same connected component of the sublevel set $\sigma_p^{-1}(-\infty, t]$. Clearly, σ_p induces a height function $\hat{\sigma}_p \colon T_p \to \mathbb{R}$ satisfying $\sigma_p = \hat{\sigma}_p \circ \pi$, where $\pi \colon X \to T_p$ is the quotient map. The leaves of T_p represent the connected components of the local minimum sets (the barycenters) of σ_p and the tree structure captures interrelationships between these local minima. We metrize T_p with a *cophenetic distance* denoted d_p and let $\mu_p = \pi_\sharp \mu$. The (functional) *barycentric merge tree* representation of μ is the quadruple $\mathcal{T}_p = (T_p, d_p, \mu_p, \hat{\sigma}_p)$.

Stability of barycentric merge trees is formulated with respect to (a variant of) an optimal transport distance introduced by Sturm [3]. To describe this distance, given $\mathcal{T}_p = (T_p, d_p, \mu_p, \hat{\sigma}_p)$ and $\mathcal{T}_p' = (T_p', d_p', \mu_p', \hat{\sigma}_p')$, consider: (i) all metric couplings δ between (T_p, d_p, μ_p) and (T_p', d_p', μ_p'); that is, metrics δ on the disjoint union $X \sqcup X'$ that restrict to d_p and d_p' on the supports of μ_p and μ_p', respectively; (ii) all probabilistic couplings h between (T_p, μ_p) and (T_p', μ_p'). If μ, μ' have finite p-moments, the *Kantorovich-Sturm distance* $d_{KS,p}(\mathcal{T}_p, \mathcal{T}_p')$ is defined as

$$d_{KS,p}(\mathcal{T}_p, \mathcal{T}_p') = \inf_{\delta, h} \max\{D(\delta, h), F(h)\},$$

where

$$D(\delta, h) = \Big(\int_{T_p \times T_p'} \delta^p(x,y) dh(x,y) \Big)^{1/p}$$

is the structural offset of (δ, h) and

$$F(h) = \Big(\int_{T_p \times T_p'} |\sigma_p(x) - \sigma_p'(y)|^p dh(x,y) \Big)^{1/p}.$$

is the functional offset of h.

We state our main stability result for the special case of a geodesic space (X, d). The full statement can be found in [2], as well as discussion of matters related to computation.

Theorem. Suppose (X, d) is a geodesic space and θ is an L-admissible pseudometric. If μ, μ' have finite p-moments, then

$$d_{KS,p}(\mathcal{T}_p, \mathcal{T}_p') \leq 2L w_p(\mu, \mu'),$$

where w_p denotes the Wasserstein p-distance.

Corollary. Empirical barycentric merge trees converge to the population barycentric merge tree almost surely; that is, $\lim_{n\to\infty} d_{KS,p}(\mathcal{T}_p, \mathcal{T}_{n,p}) = 0$, almost surely.

References

[1] R.R. Coifman and S. Lafon, *Diffusion maps*, Appl. Comput. Harmon. Anal. **21** (2006), 5–30.
[2] W. Mio and T. Needham, *Robust Representation and Estimation of Barycenters and Modes of Probability Measures on Metric Spaces*, arXiv:2505.09609.
[3] K.-T. Sturm, *On the geometry of metric measure spaces*, Acta Mathematica **196** (2006), 65–131.

Geometry of Gromov-Wasserstein Distances

Tom Needham

(joint work with Martin Bauer, Facundo Mémoli, Mao Nishino)

The *Gromov-Wasserstein (GW) distances* [1] comprise a family of metrics on the space of metric measure spaces—that is, triples of the form (X, d_X, μ_X), where (X, d_X) is a compact metric space and μ_X is a fully supported Borel probability measure—defined as follows. For metric measure spaces (X, d_X, μ_X) and (Y, d_Y, μ_Y) and $p \in [1, \infty)$, the associated GW distance is

$$\begin{aligned}\mathrm{GW}_p(X,Y) &= \frac{1}{2}\inf_{\pi\in\mathcal{C}(\mu_X,\mu_Y)} \|d_X - d_Y\|_{L^p(\pi\otimes\pi)} \\ &= \frac{1}{2}\inf_{\pi\in\mathcal{C}(\mu_X,\mu_Y)} \left(\iint_{(X\times Y)^2} |d_X(x,x') - d_Y(y,y')|^p \mathrm{d}\pi(x,y)\mathrm{d}\pi(x',y')\right)^{1/p},\end{aligned}$$

where the infimum is over the space of *measure couplings* $\mathcal{C}(\mu_X, \mu_Y)$, or Borel probability measures π on $X \times Y$ with marginals μ_X and μ_Y.

Over the past several years, there has been an influx of variants of GW distances which are designed to compare complex data types, such as node- and edge-attributed graphs. A general framework that recovers many of these variants as special cases was introduced in our recent paper [2]. The construction depends on a fixed separable metric space (Z, d_Z). For $p \in [1, \infty)$, we define a (Z, p)-*network* to be a triple (X, ω_X, μ_X) such that (X, μ_X) is a Polish probability space and $\omega_X : X \times X \to Z$ is a Z-valued kernel, which is p-integrable in the sense that

$$\iint_{X\times X} d_Z(\omega_X(x,x'), z_0)^p \mathrm{d}\mu_X(x)\mathrm{d}\mu_X(x') < \infty \tag{1}$$

for some (and hence, for any) $z_0 \in Z$. Given two (Z, p)-networks (X, ω_X, μ_X) and (Y, ω_Y, μ_Y), we define the Z-*Gromov-Wasserstein distance* between them to be

$$\mathrm{GW}_p^Z(X,Y) = \frac{1}{2}\inf_{\pi\in\mathcal{C}(\mu_X,\mu_Y)} \|d_Z(\omega_X,\omega_Y)\|_{L^p(\pi\otimes\pi)}.$$

This naturally generalizes the classical GW distance, which is recovered in the case that $Z = \mathbb{R}$, with its standard metric. Moreover, it recovers the p-Wasserstein distance over the ground metric space Z, as well as many other metrics which have been introduced in the literature. These special cases demonstrate the applicability of the Z-GW framework to shape analysis and object-oriented data science.

The generality of the framework suggests additional applications which have not previously appeared.

The main goal of our work is to describe properties of the Z-GW distance and, in particular, how they depend on properties of Z. Results appearing in [2] concern its metric and topological properties: it is complete if Z is complete, geodesic if Z is geodesic, and the resulting metric space is always contractible. Ongoing work focuses on the Alexandrov geometry of the Z-GW metric space, such as curvature bounds, which depend strongly on the geometry of the underlying space Z. These geometric results are largely based on the following main result from our forthcoming paper, concerning submetries between some metric spaces associated to Z. Recall that a submetry is a map between metric spaces $\phi : M \to N$ which takes the open ball $B_M(m, r)$ to the open ball $B_N(\phi(m), r)$.

Theorem. *There exists a chain of submetries*

$$\mathcal{L}_p([0,1]^2, Z) \to \mathcal{M}_p^Z \to \mathcal{P}_p^Z,$$

where

- $\mathcal{L}_p([0,1]^2, Z)$ *is the space of (isomorphism classes of) measurable maps* $[0,1]^2 \to Z$ *satisfying an integrability condition analogous to* (1)*, with respect to Lebesgue measure, endowed with a natural* L^p*-type distance;*
- $\mathcal{M}_p^Z$ *is the space of (isomorphism classes of)* (Z, p)*-networks, endowed with* Z*-GW distance;*
- $\mathcal{P}_p^Z$ *is the space of Borel probability measures on* Z *with finite* p*th moments, endowed with the classical* p*-Wasserstein distance.*

This theorem, and its various geometric consequences, will appear in a forthcoming paper [3].

References

[1] F. Mémoli, *Gromov-Wasserstein distances and the metric approach to object matching*, Foundations of Computational Mathematics, **11** (2011), 417–487.

[2] M. Bauer, F. Mémoli, T. Needham, M. Nishino, *The Z-Gromov-Wasserstein distance*, Journal of Machine Learning Research **26** (2025), 1–57.

[3] M. Bauer, F. Mémoli, T. Needham, M. Nishino, *Geometry of Gromov-Wasserstein distances*, To appear (2026).

Conditioning a measure-valued diffusion on the Wasserstein space

Mao Nisihino

(joint work with Martin Bauer, Tom Needham, Stefan Sommer)

In this work in progress, we study a conditioning of a diffusion on the space of probability measures equipped with the Wasserstein metric. We use the Dirichlet-Ferguson diffusion [1] for an unconditioned process, and we will control the diffusion so that the terminal measure follows the preference defined by a given function

$g : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ where $\mathcal{P}(\mathcal{X})$ is the space of probability measures on a closed Riemannian manifold $\mathcal{X}$. The unconditioned Dirichlet-Ferguson diffusion $\{\mu_t\}_{t\in[0,1]}$ can be written as

$$\mu_t = \sum_{i=1}^{\infty} s_i \delta_{X_i(t)}, \quad dX_i(t) = \sqrt{\frac{2}{s_i}} dW_i(t),$$

where $\{s_i\}_{i=1}^{\infty}$ is a sequence of strictly decreasing positive numbers such that $\sum_{i=1}^{\infty} s_i = 1$ and $\{W_i(t)\}_{i=1}^{\infty}$ is a sequence of independent Brownian motions on $\mathcal{X}$. Therefore, the unconditioned process can be seen as a superposition of infinitely many independent Brownian motions on $\mathcal{X}$, where particles of smaller mass s_i move faster. To condition the process, we need to solve the backward heat equation

$$\begin{cases} \partial_t u_t + L u_t = 0, & \text{for a.e. } t \in (0,T), \\ u_T = g. \end{cases} \tag{1}$$

on the L^2 space on the Wasserstein space with the Dirichlet-Ferguson measure, and L is the generator for the Dirichlet-Ferguson diffusion, which is informally a sum of Laplacian for each particle coordinate weighted by the inverse of the mass. Once we identify the solution, the following Girsanov argument will produce the desired conditioning. By Girsanov's theorem in Hilbert spaces [2, Theorem I.0.6], the processes

$$\widehat{\beta}_i(t) := \beta_i(t) - \int_0^t \sqrt{2s_i}\, D \log u_r(\mu_r, X_i(r))\, dr \tag{2}$$

are independent Brownian motions under $\widetilde{\mathbb{P}}$, where $\widetilde{\mathbb{P}}$ is a measure on the path space satisfying

$$\left.\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}\right|_{\mathcal{F}_T} \propto \frac{g(\mu_T)}{u_0(\mu_0)}.$$

Here, $\mathbb{P}$ is the original measure on the path space under which $\{\mu_t\}_{t\in[0,1]}$ is a Dirichlet-Ferguson diffusion, and $\mathcal{F}_T$ is the sigma-algebra generated by the process up to time T.

Consequently, under $\widetilde{\mathbb{P}}$ each particle satisfies the SDE

$$dX_i(t) = \sqrt{\frac{2}{s_i}}\, d\widehat{\beta}_i(t) + 2\, D \log u_t(\mu_t, X_i(t))\, dt, \tag{3}$$

and the induced measure-valued process $\mu_t = \sum_{i\geq 1} s_i \delta_{X_i(t)}$ solves a drifted version of the Dirichlet–Ferguson diffusion. Now that we have the desired conditioning, we would like to approximate the infinite dimensional diffusion by a finite dimensional one. The naive way is to truncate the infinite sum $\sum_{i=1}^{\infty} s_i \delta_{X_i(t)}$ at a finite level N. Our result is that the semigroup of the truncated process converges to the semigroup of the original process as $N \to \infty$:

Theorem 1. *The semigroup of the truncated conditioned process converges to the semigroup of the original conditioned process as $N \to \infty$ in the sense of generalized Hilbert convergence* [1, Section A.3].

References

[1] L. Dello Schiavo. *The Dirichlet–Ferguson diffusion on the space of probability measures over a closed Riemannian manifold* The Annals of Probability, 50(2):591–648, 2022. doi:10.1214/21-AOP1541.

[2] W. Liu and M. R"ockner. *Stochastic Partial Differential Equations: An Introduction.* Universitext. Springer, Cham, 2015. ISBN 978-3-319-22353-7. doi: 10.1007/978-3-319-22354-4.

From limits of the matrix exponential to a geodesic metric space of graphs

Tom M. W. Nye

(joint work with William M. Woodman)

Learning about populations of graphs via statistical analysis of samples from those populations has become an important task in statistics and machine learning, with applications in diverse areas. Here we focus on graphs which are vertex-labelled by some set $[n] = \{1, \ldots, n\}$ and which are edge-weighted, with weights in $\mathbb{R} > 0$. If we consider a sequence of graphs in which some given edge weight tends to zero, while the graphs otherwise remain fixed there are two possible limits: either the edge is removed in the limiting graph, or the vertices at either end of the edge can be merged (see Figure 1) below. If the edge weights represent a measure of direct association between the objects represented by the vertices, e.g. co-expression in a gene network or degree of sharing information in a social network, then the former limit is appropriate. In contrast, if the weights represent physical distance or a cost of interaction, for example, the latter limit is appropriate.

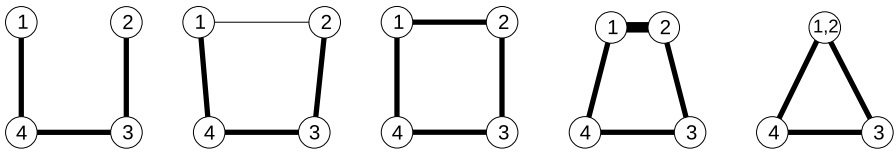

Figure 1. Limiting graphs as weight $w_{1,2}$ varies. Left: in the limit $w_{1,2} \to 0$ vertices 1 & 2 have no direct association and the edge is removed. Right: as edge weight $w_{1,2} \to \infty$, the vertices are infinitely strongly associated, so merge.

However, for some applications these limits could be viewed as being reciprocal: if edge weights represent association, then in the limit of infinite edge weight the vertices should merge. For example, infinite association in a gene network would occur between two genes which are only ever co-expressed, in which case the genes would be represented by a single vertex. These limits are illustrated in Figure 1. How can we construct a metric geometry for graphs which respects both these limits? Can we do this in such a way that the space of graphs is a geodesic metric space, since that will enable development of richer statistical methodology?

The diffusion metric is defined as follows. Given a labelled, edge-weighted graph g, let Q_g be the rate matrix of the continuous-time random walk on the vertices of g, where the instantaneous rate of moving between vertices is given by the edge weights. The diffusion metric [3, 4] is $d_t(g_1, g_2) = \|\mathrm{Exp}(tQ_{g_1}) - \mathrm{Exp}(tQ_{g_2})\|$ where $t > 0$ is a fixed constant, the norm is the Frobenius norm and Exp is the matrix exponential function. The diffusion metric compares the transition probability matrices of the random walk on the graphs g_1, g_2 at time t, thereby accounting for indirect connections between vertices. The time t must be fixed for any analysis, but for simplicity we fix $t = 1$ here. If all the edge weights in a connected graph on $[n]$ are scaled by a parameter α, then the limit as $\alpha \to \infty$ can be interpreted as a graph consisting of a single collapsed vertex. Conversely, when $\alpha \to 0$, the limit is n disconnected vertices. With this interpretation, the distance between two arbitrary connected graphs should tend to zero in both limits as $\alpha \to 0$ or ∞. This is indeed observed for the diffusion metric: see Figure 2.

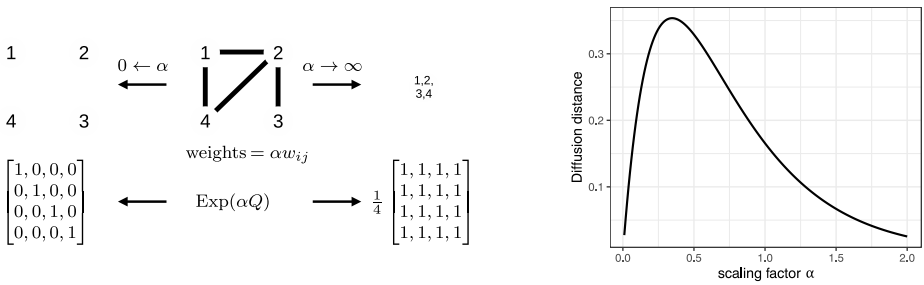

FIGURE 2. Limits as all edges weights are scaled by a factor α. Left: limits of the transition matrix of the random walk, and proposed behavior of the graph. Right: diffusion metric between two connected graphs under the same scaling.

This shows that the diffusion metric is consistent with the idea that vertices merge under the limit of infinite edge-weight, at least under this global scaling limit. In contrast, other existing metrics do not show this behaviour. A related collection of metrics [2] are defined by $\|L_{g_1}^p - L_{g_2}^p\|$ for some power $p \in \mathbb{R}$, where $L_g = -Q_g$ is the graph Laplacian. These metrics cannot reproduce this behaviour under both limits $\alpha \to 0$ and $\to \infty$, and they blow up as $\alpha \to \infty$ for $p > 0$.

Beyond the global scaling limit, behaviour of the diffusion metric under more general limits has not been studied. An existing result [1] can be used to calculate certain limits of the matrix exponential: if Q and $\hat{Q}$ are the rate matrices of two labelled graphs $g, \hat{g}$ on $[n]$, then

$$\lim_{\rho \to \infty} \mathrm{Exp}(Q + \rho\hat{Q}) = P\,\mathrm{Exp}(PQP)P$$

where P is orthogonal projection onto $\ker \hat{Q}$. By extending this result, the closure of the image of the matrix exponential on the set of all rate matrices Q_g can be

identified. The boundary points are rank deficient and uniquely correspond to graphs containing merged vertices (multi-labelled graphs). By applying the Hopf-Rinow theorem, this is a geodesic metric space. The details of this construction and the generalisation of the Campbell-Rose [1] result will be published elsewhere.

References

[1] S. L. Campbell, and N. J. Rose, *Singular perturbation of autonomous linear systems*, SIAM Journal of Mathematical Analysis **10(3)** (1978), 542–551.

[2] K. Severn, I. Dryden, and S. Preston, *Manifold valued data analysis of samples of networks*, Annals of Applied Statistics **16(1)** (2022), 368–390.

[3] D. Hammond, Y. Gur, and C. Johnson, *Graph diffusion distance: A difference measure for weighted graphs based on the graph Laplacian exponential kernel*, In IEEE Global Conference on Signal and Information Processing (2013), 419–422.

[4] S. Lunagómez, S. Olhede, and P. Wolf, *Modeling network populations via graph distances*, Journal of the American Statistical Association **116(536)** (2020), 2023–2040.

Support conditions for smeary probability measures on spheres

Susovan Pal

Let X be a random variable on the unit sphere $\mathbb{S}^m$, with Riemannian distance d, and let

$$F_\mu(p) = \mathbb{E}[d^2(p, X)]$$

be its Fréchet function. Assume that $N \in \mathbb{S}^m$ is the prescribed Fréchet mean and write the lifted Fréchet function in normal coordinates as

$$\widetilde{F}_\mu(u) = F_\mu(\exp_N u), \qquad u \in T_N\mathbb{S}^m.$$

The talk concerned the following support question: how far from N must the support of an absolutely continuous probability measure extend before the Hessian of $\widetilde{F}_\mu$ can degenerate in a way which produces full or directional smeariness?

The starting point is an explicit fourth-order Taylor expansion, at $u = 0$, of

$$g(u; v) = d^2(\exp_N u, \exp_N v), \qquad v = R\Theta, \quad R \in (0, \pi), \quad \Theta \in \mathbb{S}^{m-1}.$$

Assume first that μ has a density which is rotationally symmetric about N and vanishes in a neighbourhood of the antipode $-N$. Then the Hessian and fourth derivative of $\widetilde{F}_\mu$ reduce to one-dimensional radial integrals. More precisely, if

$$b_m(R) = 1 + (m-1)R\cot R,$$

then

$$D^2\widetilde{F}_\mu(0) = \frac{2}{m}\left(\int_0^\pi \rho(R)(\sin R)^{m-1} b_m(R)\, dR\right) I_m.$$

Similarly, the fourth derivative has the form

$$D^4\widetilde{F}_\mu(0)[w, w, w, w] = 24|\mathbb{S}^{m-1}|\left(\int_0^\pi \rho(R)(\sin R)^{m-1} h_m(R)\, dR\right) \cdot \|w\|^4,$$

where h_m is an explicit scalar function coming from the fourth-order term in the expansion of $g(u;v)$. Thus, in the rotationally symmetric case, the support problem is reduced to the sign and zero structure of the two geometry-only functions b_m and h_m.

The function b_m is strictly decreasing on $(0,\pi)$ and has a unique zero $R_m \in (\pi/2,\pi)$. For $m \geq 4$, the function h_m has a first zero $S_m \in (\pi/2,\pi)$, and the two thresholds satisfy

$$R_m < S_m.$$

Moreover, both thresholds are super-hemispherical and approach the equator as $m \to \infty$; for example $R_m = \pi/2 + O(1/m)$ and $S_m = \pi/2 + O(1/m)$.

The resulting rotationally symmetric support picture is as follows. If

$$\operatorname{supp}\mu \subset \overline{B}(N;R_m),$$

then $D^2\widetilde{F}_\mu(0)$ is positive definite, hence no smeariness occurs. In dimensions $m = 2,3$, rotationally symmetric 2-smeariness does not occur: if the Hessian is forced to vanish, then the fourth derivative is negative in every non-zero direction, so 0 is a strict local maximum rather than a smeary local minimum. For $m \geq 4$, the quartic obstruction persists up to the larger threshold S_m: if

$$\operatorname{supp}\mu \subset \overline{B}(N;S_m) \quad \text{and} \quad D^2\widetilde{F}_\mu(0) = 0,$$

then the fourth derivative is still negative in every non-zero direction. Thus a rotationally symmetric 2-smeary density can only occur after the support crosses the radius S_m.

Conversely, in the high-dimensional regime covered by the construction, this obstruction is sharp up to an arbitrarily small excess beyond S_m. For every $\varepsilon > 0$, and for all sufficiently large m, one can construct a rotationally symmetric probability measure with continuous density such that

$$\operatorname{supp}\mu \subset B(N;S_m+\varepsilon), \qquad \mu\{x : S_m < d(N,x) < S_m+\varepsilon\} > 0,$$

with

$$D^2\widetilde{F}_\mu(0) = 0, \qquad D^4\widetilde{F}_\mu(0)[w,w,w,w] > 0 \quad (w \neq 0).$$

Hence N is a local 2-smeary Fréchet mean. Since rotational symmetry makes the Hessian a scalar multiple of the identity, full and directional smeariness coincide in this part of the theory.

I then discussed the non-rotationally symmetric case. For a general absolutely continuous density $\rho(R,\Theta)$, still supported away from the antipode, the Hessian is

$$D^2\widetilde{F}_\mu(0) = 2\int_0^\pi \int_{\mathbb{S}^{m-1}} \left(\frac{R}{\tan R} I_m + \left(1 - \frac{R}{\tan R}\right)\Theta\Theta^T\right) \cdot \rho(R,\Theta)(\sin R)^{m-1}\, d\Theta\, dR.$$

This formula gives a closed-hemisphere obstruction: if

$$\operatorname{supp}\mu \subset \overline{B}(N;\pi/2),$$

then the Hessian is positive definite, so neither full nor directional smeariness occurs. More generally, support in $\overline{B}(N; R_m)$ prevents the Hessian from vanishing identically, and hence rules out full smeariness.

The full and directional problems differ sharply beyond the hemisphere. For every $m \geq 2$ and every $\varepsilon > 0$, there exists an absolutely continuous probability measure supported in

$$B(N; \pi/2 + \varepsilon)$$

for which the Hessian at N is positive semidefinite of rank one, the third derivative vanishes, and the fourth derivative is strictly positive on the kernel of the Hessian. This gives a directionally 2-smeary local Fréchet mean with support arbitrarily close to a closed hemisphere. Thus the hemispherical support bound is rigid for directional smeariness.

Finally, I mentioned one finite-sample consequence in the non-smeary regime. If $H = D^2\widetilde{F}_\mu(0)$ is positive definite, $\Sigma = \mathrm{Cov}(\log_N X)$, and the usual non-smeary central limit theorem together with a uniform $(2+\delta)$ moment bound gives convergence of second moments, then the limiting variance modulation is

$$m_\infty = \frac{\mathrm{trace}(4H^{-1}\Sigma H^{-1})}{\mathrm{trace}(\Sigma)}.$$

For absolutely continuous measures on $\mathbb{S}^m$, $m \geq 2$, the spectral bound

$$\lambda_{\max}(H/2) < 1$$

implies $m_\infty > 1$ whenever the above limit is valid. Thus, under hemispherical support and the stated moment condition, one obtains Type-I finite sample smeariness, although ordinary smeariness is excluded by the positive-definite Hessian.

References

[1] B. Afsari, *Riemannian L^p center of mass: existence, uniqueness, and convexity*, Proc. Amer. Math. Soc. **139** (2011), 655–673.

[2] R. N. Bhattacharya and V. Patrangenaru, *Large sample theory of intrinsic and extrinsic sample means on manifolds. II*, Ann. Statist. **33** (2005), 1225–1259.

[3] B. Eltzner, *Geometrical smeariness – a new phenomenon of Fréchet means*, Bernoulli **28** (2022), 239–254.

[4] S. Huckemann and B. Eltzner, *A smeary central limit theorem for manifolds with application to high-dimensional spheres*, Ann. Statist. **47** (2019), 3360–3381.

[5] B. Eltzner, S. Hundrieser and S. F. Huckemann, *Finite sample smeariness on spheres*, in *Geometric Science of Information*, Lecture Notes in Comput. Sci. **12829**, Springer, 2021, 12–19.

[6] D. Tran, *Behavior of the Fréchet mean and central limit theorems on spheres*, Braz. J. Probab. Stat. **35** (2021), 590–608.

Flag for stats

Xavier Pennec

Dimension reduction is a key issue in numerous statistical and machine learning problems. When data belong to a manifold, we first need to define families of parametric subspace in manifolds which could play the role of principal subspaces. Geodesic shooting along the main eigenvectors of the covariance matrix at the Fréchet mean point generates a Geodesic Subspace (GS) which is used in tangent PCA. The point and tangent vectors defining that GS that can also be optimized to best fit the data in more principled methods such as Principal Geodesic Analysis (PGA) or Geodesic PCA (GPCA). To restore the full symmetry between the parameters, we proposed in [1] to use the **Exponential Barycentric subspace** (EBS) defined as the locus of weighted means of $k+1$ reference points (with positive or negative weights summing up to 1). The EBS is locally a stratified spaces of maximal dimension k comprised of critical points of the weighted variance satisfying the barycentric equation $\sum_i \lambda_i log_x(x_i) = 0$. Its metric completion is called the **affine span** $\mathrm{Aff}(x_0, \dots x_k)$. Such spaces generalise the notion of Geodesic Subspaces which can be shown to be the limit when reference points coalesce towards a local 1-jet. If the exponential barycentric subspace is locally a regular manifold when the reference points are affinely independent, the affine span is naturally a stratified space as it includes the metric completion. Less natural stratifications may also appear when the reference points are affinely dependent.

Barycentric subspaces and affine spans can naturally be nested by defining an ordering of the reference points. This allows the construction of forward or backward nested sequence of subspaces. However, these methods optimized for one subspace at a time and cannot optimize the explained (or unexplained) variance simultaneously for all the subspaces of the flag. In order to obtain a global criterion, PCA in Euclidean spaces was rephrased in [1] as an optimization on the flags of linear subspaces of the accumulated unexplained variance criterion. It is well known that the best k-dimensional subspace in the least square sense is given by the k first eigenvectors of the PCA, but their order cannot be decided without adding an extra iterative step over the dimensions. With the optimization of our accumulated unexplained variance criterion on the flags of subspaces, the order of these eigenvectors is automatically decided. Thus we obtain a variational formulation of the PCA algorithm but on the space of flags rather than on the space of Grassmanians. This feature generalizes nicely to flags of affine spans in Riemannian manifolds and gives a particularly appealing generalization of PCA on manifolds, called **Barycentric Subspaces Analysis** (BSA).

Considering PCA as an optimization of flags spaces gives an interesting geometric point of view, even in Euclidean spaces. Indeed, one usually consider the succession of unidimensional eigenmodes for the interpretation of the data in PCA, but more general flags with higher dimensional subspaces naturally arise with the geometric point of view. They correspond to covariance matrix with repeated eigenvalues, in which case eigenspaces are stable but eigenvectors may be freely

rotated within each eigenspace. This raises an an important issue about the interpretation of PCA modes, called **the curse of isotropy** [2]: principal components associated with equal eigenvalues show large intersample variability and are arbitrary combinations of potentially more interpretable components.

One may easily overlook the problem because empirical eigenvalues are almost surely distinct in practice due to sampling errors with a finite number of samples, but we showed in [2] that many datasets are likely to suffer from the curse of isotropy: in a dataset with 1000 samples, all the eigenvalue pairs with a relative eigengap lower than 21% should be assumed to actually come from one repeated eigenvalue. This demonstrates that the curse of isotropy cannot be overlooked. In this context, we propose to transition from fuzzy principal components to more interpretable principal subspaces, which can be seen as an optimization of the space of flags. The final methodology, coined **principal subspace analysis** is extremely simple and shows promising results on a variety of datasets from different fields.

References

[1] X. Pennec, *Barycentric subspace analysis in manifolds*, Annals of Statistics 46(6A) (2018), p.2711–2746.

[2] T. Szwagier, X. Pennec, *The curse of isotropy: from principal components to principal subspaces*, arXiv:2307.15348, (2023).

From limits of the matrix exponential to a geodesic metric space of graphs

François-Xavier Vialard

(joint work with Guillaume Houry and Jean Feydy)

Given two metric-measure spaces (X, c_X, α) and (Y, c_Y, β), the Gromov–Wasserstein distance is defined by

$$GW(\alpha, \beta) = \min_{\pi \in \Pi(\alpha,\beta)} \iint \big(c_X(x, x') - c_Y(y, y')\big)^2 \, d\pi(x, y) \, d\pi(x', y').$$

Gromov–Wasserstein distances compare metric-measure spaces by looking for a coupling that preserves pairwise distances, rather than minimizing pointwise displacements in a common ambient space. This makes them robust to rotations and relabelings, at the price of a nonconvex quadratic matching problem.

The main message of this talk is that a large class of Gromov–Wasserstein (GW) problems are implicitly an entropic optimal transport (EOT) problem after an (up to an unknown/optimizable) feature lift, which has important consequences for its computational complexity. This reduction was already known for the case of a squared Euclidean cost, which was very limiting in practice, since the optimal matchings were driven by large distance of pairwise points.

A cost c is conditionally of negative type (CNT) precisely when Schoenberg's theorem gives a Hilbert embedding

$$c(x, x') = \|\phi(x) - \phi(x')\|^2_{\mathcal{H}}.$$

Equivalently, after fixing a base point x_0, the centered kernel $k(x,x') = \frac{1}{2}(c(x,x_0)+c(x_0,x') - c(x,x'))$ is positive definite. Thus CNT costs play for squared distances the role that positive kernels play for inner products. With the GW lift

$$\Phi(x) = \big(\phi(x), \tfrac{1}{2}\|\phi(x)\|^2\big),$$

and similarly $\Psi(y)$, the entropy-regularized GW objective can be written, up to an additive constant, as

$$\begin{aligned} &GW_\varepsilon(\alpha,\beta) \\ =&C(\alpha,\beta) + 8\min_{\Gamma}\min_{\pi\in\Pi(\alpha,\beta)}\left\{\|\Gamma\|_{\mathrm{HS}}^2 + \frac{\varepsilon}{8}\mathrm{KL}(\pi \mid \alpha\otimes\beta) - 2\int\langle\Gamma, \Phi(x)\Psi(y)^\top\rangle\, d\pi\right\}. \end{aligned}$$

For fixed Γ, the minimization over π is exactly an EOT problem: one transports $\Phi_\#\alpha$ to $\Gamma_\#\Psi_\#\beta$ with the squared Euclidean cost and temperature $\varepsilon/8$. The non-convex GW search is therefore reduced to alternating between a Sinkhorn solve and a linear alignment update for Γ.

This reduction also explains the debiasing result. Entropic regularization breaks the desired property $GW_\varepsilon(\alpha,\alpha) = 0 (\geq 0)$ and makes the gradient flow converge to a biased steady state. The Sinkhorn GW divergence removes the self-interaction bias:

$$SGW_\varepsilon(\alpha,\beta) = GW_\varepsilon(\alpha,\beta) - \frac{1}{2}\big(GW_\varepsilon(\alpha,\alpha) + GW_\varepsilon(\beta,\beta)\big).$$

In the CNT setting, this symmetrization is nonnegative and vanishes whenever the two metric-measure spaces are isometric. For too large a temperatures, positivity is not guaranteed, while for the Euclidean squared cost, positivity is restored for small ε with respect to the data.

Finally, we present an algorithm that exposes a clear computational tradeoff. For N and M particles and feature dimensions D and E, the CNT implementation stores $O(ND + ME)$ numbers and costs

$$\begin{aligned} &O(N^2D + M^2E) && \text{for the kernel-PCA embeddings,} \\ &O(NM\, n_{\mathrm{inner}} n_{\mathrm{outer}}) && \text{for the Sinkhorn/GW loops.} \end{aligned}$$

This shows in particular that Gromov-Wasserstein estimations can be made quadratic for a regime of a large number of particles with a choice of embedding dimension usually of the order 20 to 200. Combining this with multiscale methods and with efficient implementation of kernel matrix products such as Keops, our method broadens the use of Gromov-Wasserstein distances by allowing more distance functions choices.

References

[1] G. Houry, J. Feydy and F.X. Vialard, *Gromov-Wasserstein at Scale, Beyond Squared Norms*, arXiv:2602.06658.

Manifold GCN: Diffusion-based Convolutional Neural Network for Manifold-valued Graphs

Christoph von Tycowicz
(joint work with Martin Hanik, Gabriele Steidl)

There is an increasing interest in applying the predictive power of deep learning to non-standard data types such as graphs, shapes, or special matrices. Often, incorporating the underlying regularities of such data into the learning task (or algorithm) leads to sizable empirical improvements. In this context, we investigate the case where the learning task is formulated for graphs that are embedded in a Riemannian manifold M, i.e. graphs $G = (V, E, w, f)$ with vertex set $V = \{v_1, \dots, v_n\}$, directed edge set $E \subset V \times V$, positive edge weights $w : E \to \mathbb{R}_+$, and manifold-valued feature map $f : V \to M$.

We present two novel network units that can be used with this kind of data to build deep architectures that adhere to the symmetries inherent to both the base and feature/signal domains. Both units are equivariant under node permutations and the feature manifold's isometries. These properties have led to a beneficial inductive bias in many deep-learning tasks. Furthermore, they enable novel, more flexible feature designs. The presented results are part of the work published in [1].

1. Diffusion Layer

We introduce a novel convolution-type layer that discretizes *graph diffusion in manifolds*. We thereby utilize a link between convolution and diffusion processes that graph neural networks have successfully used for Euclidean data [2]. Combined with an explicit solution scheme for the diffusion equation, the new diffusion layer exhibits the same local aggregation of features that makes convolution layers so successful.

Let $G = (V, E, w, f)$ be a graph with feature map f. We augment f with a time parameter t and consider, for $a > 0$, the M-valued graph diffusion equation

$$\begin{cases} \frac{\partial}{\partial t}\widetilde{f}(v,t) = -\Delta\widetilde{f}(v,t), & v \in V, \quad t \in (0,a), \\ \widetilde{f}(v,0) = f(v), & v \in V, \end{cases} \tag{1}$$

where Δ denotes the manifold-valued graph Laplacian derived in [3].

By discretizing this graph diffusion process, we can construct a convolution-type neural network layer with time t as learnable parameter. Furthermore, we need a nonlinear activation function that prevents depth collapse. We propose the following: With the sigmoid function $\mathrm{sig}(x) := e^x/(1+e^x)$ and $\theta = (\vartheta_1, \vartheta_2) \in \mathbb{R}^2_{\geq 0}$, we define, for any $p \in M$, the nonlinear activation function $\sigma_p^\theta : T_pM \to T_pM$ by

$$\sigma_p^\theta(X) := \mathrm{sig}\big(\vartheta_1 \|X\|_p - \vartheta_2\big)X.$$

The activation function has two learnable parameters ϑ_1 and ϑ_2, which increase the network capacity and realize an isometry-equivariant, nonlinear re-scaling of the input vector.

With this activation function, we define the 1-*step map* $\mathrm{step}_{t,\theta}$ by

$$\mathrm{step}_{t,\theta}(f)(v) := \exp_{f(v)}\Big(-t\sigma^{\theta}_{f(v)}(\Delta f(v))\Big), \quad v \in V.$$

The map $\mathrm{step}_{t,\theta}$ realizes a "nonlinearly activated" explicit diffusion step of length t of Equation (1). Each vector $\Delta f(v)$ may be viewed as a "force" acting on the feature $f(v)$, steering it toward the weighted Fréchet mean of its neighboring nodes in the graph; this displacement is realized through the exponential map. Since the graph Laplacian depends solely on the immediate neighbors of a node—its *1-hop neighborhood*—the resulting 1-step map aggregates information only from these local neighborhoods. Like numerical integration schemes, we extend the 1-step map for $\ell > 1$ to perform several steps by

$$\mathrm{step}^{\ell}_{t,\theta}(f) := \underbrace{\mathrm{step}_{t,\theta} \circ \cdots \circ \mathrm{step}_{t,\theta}}_{\ell \text{ times}}(f). \tag{2}$$

Again, $\mathrm{step}^{\ell}_{t,\theta}$ realizes ℓ "nonlinearly activated" explicit Euler steps of length $t > 0$ for Equation (1).

Now, we can define our diffusion layer. Let Ω be the set of graphs with M-valued vertex features.

Definition 1 (diffusion layer)**.** *For $c \in \mathbb{N}$, let $\boldsymbol{t} = (t_1, \ldots, t_c) \in \mathbb{R}^c_{\geq 0}$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_c) \in (\mathbb{R}^2_{\geq 0})^c$. A diffusion layer with $c \geq 1$ channels is a map $\mathrm{Diff}_{\boldsymbol{t},\boldsymbol{\theta}} : \Omega^c \to \Omega^c$ with*

$$\mathrm{Diff}_{\boldsymbol{t},\boldsymbol{\theta}}\big((G_i)^c_{i=1}\big) := \Big(\big(V_i, E_i, w_i, \mathrm{step}^{\ell}_{t_i,\theta_i}(f_i)\big)\Big)^c_{i=1}.$$

The diffusion layer takes c graphs as input and diffuses them for certain amounts of time. Thus, it consists of c diffusion "channels". It can be applied to graphs with varying numbers of nodes and connectivity patterns. Remarkably, the diffusion layer is equivariant under node permutations and isometric transformations of the feature manifold.

2. Tangent Multilayer Perceptron

Beyond the diffusion layer, we introduce a "1x1 convolution" layer that enables feature transformations at each node and allows the network width to vary. To construct this layer, we adapt concepts from the vector neuron framework [4] to the manifold-valued setting. Specifically, we define a linear transformation on tangent vectors followed by a nonlinear activation. Owing to this nonlinearity, multiple such layers can be composed to build deeper architectures, analogous to a standard multilayer perceptron.

The layer maps c_{in} input features $f_1, \ldots, f_{c_{\text{in}}}$ to c_{out} output features $g_1, \ldots, g_{c_{\text{out}}}$. To achieve this, the manifold-valued node features are first represented in a common reference tangent space. Within this space, the layer learns two sets of linear combinations. The first set defines vectors that determine positive and negative half-spaces, while the second set produces output vectors that are subsequently modified in a nonlinear manner depending on the half-space in which they lie.

The resulting tangent vectors are then mapped back onto the manifold via the exponential map.

3. Experiments

We evaluated networks built from the proposed layers on two graph classification benchmarks and obtained strong results. While they apply to a much broader class of problems, our models surpassed existing state-of-the-art approaches that were tailored specifically to the respective tasks.

The symmetry-equivariant nature of the feature representations further allowed us to develop a new graph embedding approach that leverages the connection between graph node permutations and isometries in hyperbolic and SPD manifolds. Compared with conventional degree-based embeddings, this method yielded substantial performance gains, highlighting the potential for geometric representation learning.

Experiments in low-data regimes indicate that the proposed architectures can achieve competitive performance with fewer training samples. A likely explanation is the additional inductive bias introduced by their equivariance properties. Such behavior is particularly attractive in domains where annotated data is limited. One prominent example is medical shape analysis, where we demonstrated that sphere-valued features can serve as compact shape descriptors and facilitate learning in small-sample-size settings.

References

[1] M. Hanik, G. Steidl, and C. von Tycowicz, *Manifold GCN: Diffusion-based Convolutional Neural Network for Manifold-valued Graphs*. International Journal of Computer Vision, vol. 134, p. 21, 2026. doi: 10.1007/s11263-026-02899-9.

[2] J. Atwood and D. Towsley, *Diffusion-convolutional neural networks*. Advances in Neural Information Processing Systems, vol. 29, pp. 2001–2009, 2016.

[3] R. Bergmann and D. Tenbrinck, *A graph framework for manifold-valued data*, SIAM Journal on Imaging Sciences, vol. 11, no. 1, pp. 325–360, 2018. doi: 10.1137/17M1118567.

[4] C. Deng, O. Litany, Y. Duan, A. Poulenard, A. Tagliasacchi, and L. J. Guibas, *Vector neurons: A general framework for SO(3)-equivariant networks*. Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), pp. 12200–12209, 2021.

Reporter: Johann Köhne

Participants

Dr. Elizabeth Baker
Technical University of Denmark
Richard Petersens Plads, Building 321
Copenhagen 2800
DENMARK

Prof. Dr. Martin Bauer
Department of Mathematics
Florida State University
208 Love Building
1017 Academic Way
Tallahassee, FL 32306-4510
UNITED STATES

Dr. Nicolas Charon
Department of Mathematics
University of Houston
Philip Guthrie Hoffman Hall
3551 Cullen Blvd, Room 669
Houston TX 77204-3476
UNITED STATES

Dr. Ian Dryden
School of Mathematical Sciences
The University of Nottingham
University Park
Nottingham NG7 2RD
UNITED KINGDOM

Dr. Fernando Galaz-García
Dept. of Mathematical Sciences
Durham University
Upper Mountjoy Campus, Stockton Rd,
Durham DH1 3LE
UNITED KINGDOM

Prof. Dr. Erlend Grong
Department of Mathematics
University of Bergen
P. O. Box 7803
5020 Bergen
NORWAY

Assoc. Prof. Dr. Karen Habermann
Department of Statistics
University of Warwick
Coventry CV4 7AL
UNITED KINGDOM

Prof. Dr. Stephan Huckemann
Institut für Mathematische Stochastik
Georg-August-Universität Göttingen
Goldschmidtstraße 7
37077 Göttingen
GERMANY

Dr. Erik Jansson
Department of Applied Mathematics &
Theoretical Physics (DAMTP)
Centre for Mathematical Sciences
Wilberforce Road
Cambridge CB3 OWA
UNITED KINGDOM

Johann Köhne
Institut für Mathematische Stochastik
Georg-August-Universität Göttingen
Goldschmidtstraße 7
37077 Göttingen
GERMANY

Prof. Dr. Huiling Le
School of Mathematical Sciences
The University of Nottingham
University Park
Nottingham NG7 2RD
UNITED KINGDOM

Dr. Alice Le Brigant
Dept. of Mathematics
University Paris I
90, rue de Tolbiac
75634 Paris Cedex 13
FRANCE

Levin Maier
Mathematisches Institut
Universität Heidelberg
Im Neuenheimer Feld 205
69120 Heidelberg
GERMANY

Prof. Dr. Peter W. Michor
Fakultät für Mathematik
Universität Wien
Oskar-Morgenstern-Platz 1
1090 Wien
AUSTRIA

Prof. Dr. Washington Mio
Department of Mathematics
Florida State University
Tallahassee, FL 32306-4510
UNITED STATES

Dr. Tom Needham
Department of Mathematics
Florida State University
1017 Academic Way
Tallahassee, FL 32306-4510
UNITED STATES

Mao Nishino
Department of Mathematics
Florida State University
Tallahassee, FL 32306-4510
UNITED STATES

Dr. Tom Nye
School of Mathematics, Statistics
and Physics
Newcastle University
Newcastle upon Tyne NE1 7RU
UNITED KINGDOM

Susovan Pal
Department of Mathematics
Vrije Universiteit
Pleinlaan 2
1050 Bruxelles Ixelles
BELGIUM

Dr. Xavier Pennec
Inria Sophia-Antipolis, Epione team
Côte d'Azur University and INRIA
2004 Route des Lucioles
P.O. Box BP93
06902 Sophia-Antipolis Cedex
FRANCE

Dr. Yvo Pokern
Department of Statistical Science
University College London
Gower Street
London WC1E 6BT
UNITED KINGDOM

Prof. Dr. Stefan H. Sommer
Department of Computer Science
University of Copenhagen
Universitetsparken 1
2100 København
DENMARK

Prof. Dr. François-Xavier Vialard
Laboratoire d'Informatique
Gaspard Monge
Université Gustave Eiffel
UMR 8049
77420 Champs-sur-Marne
FRANCE

Dr. Christoph von Tycowicz
Zuse Institute Berlin (ZIB)
Takustr. 7
14195 Berlin
GERMANY