Received: 2 September 2024 | Accepted: 30 September 2024

DOI: 10.1002/aaai.12210

HIGHLIGHT

Geometric Machine Learning

Melanie Weber

School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA

Correspondence
Melanie Weber, School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA.
Email: mweber@seas.harvard.edu

Funding information
NSF, Grant/Award Number: CBET-2112085,DMS-2406905

Abstract

A cornerstone of machine learning is the identification and exploitation of structure in high-dimensional data. While classical approaches assume that data lies in a high-dimensional Euclidean space, *geometric machine learning* methods are designed for non-Euclidean data, including graphs, strings, and matrices, or data characterized by symmetries inherent in the underlying system. In this article, we review geometric approaches for uncovering and leveraging structure in data and how an understanding of data geometry can lead to the development of more effective machine learning algorithms with provable guarantees.

INTRODUCTION

Many classical machine learning methods assume that data lies in a high-dimensional Euclidean space. However, many applications involve structured data, such as sets, graphs, and matrices, data concentrated in low-dimensional subspaces, and data exhibiting symmetries derived from fundamental laws of physics. Recognizing and leveraging such structure can lead to more efficient and interpretable models.

One of the early examples of this idea are convolutional neural networks (CNN) (Krizhevsky, Sutskever, and Hinton 2012; LeCareun et al. 1998), which revolutionized image classification in the 2010s. In image classification, the goal is to assign labels based on the objects within an image, regardless of their position. CNNs achieve this by encoding translation invariance as an inductive bias, that is, by restricting the model to functions whose output is unchanged under translations of the input. Geometric machine learning extends this idea by encoding various types of geometric structures into model architectures (Bronstein et al. 2021; Cohen and Welling 2016). These approaches have been particularly impactful in scientific machine learning, with applications spanning material sciences (Batzner et al. 2022; Subramanian et al. 2024), biochemistry (Azizzadenesheli et al. 2024; Diepeveen et al. 2024; Ganea et al. 2021; Watson et al. 2023; Zhao and Singer 2014), climate science (Kashinath et al. 2021; Li et al. 2021), biomedical sciences (Bekkers et al. 2018; Bhasker et al. 2024; Klimovskaia et al. 2020; Weber et al. 2017c), and robotics (Wang et al. 2022), among others. The development of geometric machine learning approaches is guided by two central questions:

1. How can we characterize useful geometric structures in data?
2. How can we design (provably) more efficient algorithms and architectures that explicitly encode such structure?

The first question relates to *data geometry*, characterizing which involves understanding the local and global geometric properties of data on both discrete and continuous domains. The second question concerns the design of algorithms and architectures that leverage these geometric properties and the fundamental question of when and why geometric approaches may outperform classical methods.

This article discusses key geometric tools for characterizing data geometry and how such characterizations can be leveraged in learning tasks on graphs, matrix spaces,

and symmetric spaces. It accompanies a New Faculty Highlights talk given at AAAI 2024 (Weber 2024).

CHARACTERIZING DATA GEOMETRY

The first step in leveraging data geometry is to identify and parametrize such structure in a way that is amenable to downstream applications. Data and models can exhibit various geometric structures, such as symmetries arising from fundamental physical laws or low-dimensional structure that arises from a few latent factors.

Continuous setting

Symmetries

Symmetries, sometimes called "invariances," are transformations that leave crucial properties of the input data unchanged (or "invariant"). Mathematically, such structure can be characterized with algebraic and geometric tools through the language of group theory: The set of transformations $U : \mathcal{M} \to \mathcal{M}$, which, when applied to a data point $x \in \mathcal{M}$, do not change the value of the objective (i.e., $f(UxU^{\dagger}) = f(x)$ for all $x \in \mathcal{M}$) has a group structure G. This concept is crucial in understanding a variety of structures in machine learning and data science applications. For example, in relational data like sets or graphs, there is often no natural order relation among the elements, which is indicative of permutation symmetries. In image classification, the focus is usually on identifying the type of object in the image, regardless of its location, which corresponds to translation symmetry. When predicting the structural properties of molecules, the output should not be affected by global rotations of the input, that is, it should be rotation-invariant. In these cases, it is important that machine learning models respect these fundamental structural properties of the data. We will later discuss how to encode such structural information as inductive bias into machine learning architectures.

Smooth Data Manifolds

In learning tasks it is often assumed that high-dimensional data lies on or near a lower-dimensional manifold (the so-called "manifold hypothesis"). Leveraging such structure in downstream tasks requires characterizing the geometric properties of this manifold, such as its intrinsic dimension and curvature. It is common to assume certain smoothness conditions, for instance, by bounding the *reach*. The reach of a manifold corresponds to the largest distance at which points in the surrounding space still have a unique closest point on the manifold (Federer 1959). Restrictions on the reach can ensure that the manifold does not have sharp corners or other irregularities. Another concept that is useful in this context is *Ricci curvature*, a local, intrinsic notion of curvature, which describes how the volume within small regions of the manifold changes.

Discrete setting

As discussed above, graph- and set-structured data, which are prime examples of discrete data structures, are inherently permutation-invariant. As such, their structure can be characterized through the lens of permutation symmetries. Another approach to studying their structure involves discrete notions of *curvature*. This requires the extension of classical concepts of curvature, which are typically defined in continuous settings, to discrete spaces. We introduce such notions below and describe how they can be used to analyze both local and global geometric properties of data.

Discrete Notions of Curvature

Curvature is a fundamental concept in Differential Geometry, which characterizes the geometric properties of geodesic spaces. Here we consider discretizations of *Ricci curvature* that allow for characterizing the local and global geometric properties of data represented in discrete spaces, such as graphs, networks, hypergraphs, and simplicial complexes. One of the key challenges in defining curvature in discrete spaces is the lack of a (natural) differential structure. To address this, we define curvature using *curvature analogies*: In continuous spaces, the curvature is related to several classical tools, many of which have well-studied counterparts in discrete settings. By defining a curvature concept that maintains the relationship with one of these discrete characteristics, we can define a discrete curvature notion that retains some properties of its continuous counterpart.

Two curvature notions, introduced by Ollivier (Ollivier 2010) and Forman (Forman 2003), are particularly popular in machine learning on graph-structured data. *Ollivier's curvature* is based on a key relationship between Ricci curvature and the behavior of random walks. We will introduce this notion in detail below and discuss how it motivates the applications discussed in the next section. *Forman's curvature*, while closely related to Ollivier's, is often used as a more scalable alternative. Although Ollivier's curvature typically offers a richer geometric characterization, Forman's curvature is computationally more efficient, especially for large and dense graphs. Consequently, applications of both notions have been explored in the machine learning literature.

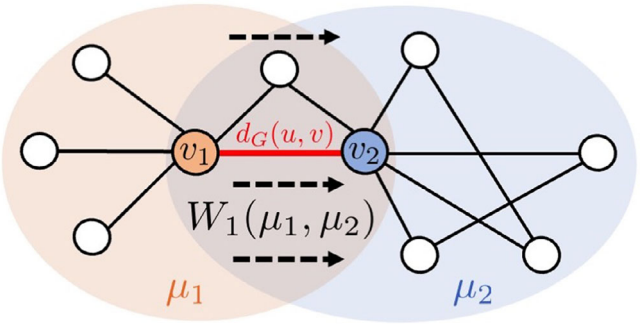

FIGURE 1 Computing Ollivier's Ricci curvature.

Ollivier's Ricci Curvature

Ollivier (Ollivier 2010) introduces a Ricci curvature, which relates the curvature along a geodesic between nearby points x, y on a manifold $\mathcal{M}$ with the optimal transportation distance between their neighborhoods.

This relies on a close relation between the behavior of random walks starting at nearby points, which can be characterized via optimal transport, and Ricci curvature. Specifically, random walks are likely to draw closer together if the Ricci curvature is positive and further apart, if the Ricci curvature is negative. Random walks are well-studied on graphs as well, and one can use this knowledge to define a meaningful curvature notion. In particular, if we start uniform random walks at neighboring nodes v_1, v_2, they are likely to stay close-by, if their neighborhoods overlap significantly, and likely to draw apart if there is no significant overlap. This can be quantified using the Wasserstein-1 distance $W_1(\mu_1, \mu_2)$ between two probability distributions μ_1, μ_2 induced by the random walks (see Figure 1). Then the following notion of curvature is positive, if the random walks stay close-by, and negative, if they draw apart ($d_G(x, y)$ denoting the shortest path distance):

$$\kappa(v_1, v_2) = 1 - \frac{W_1(m_1, m_2)}{d_G(v_1, v_2)}.$$

To compute Ollivier's curvature in practice, we need to solve an optimal transport problem for each pair of connected nodes in the graph. This can be done, e.g., using the Hungarian algorithm (Kuhn 1955). Unfortunately, this can be computationally expensive, especially for large and/or dense graphs. Faster approximations can be obtained by replacing the Wasserstein-1 distance with the Sinkhorn distance (Sinkhorn and Knopp 1967) or by using a combinatorial approximation of Ollivier's curvature (Tian, Lubberts, and Weber 2023b).

Continuum limits

Continuum limits of graphs and their geometric characteristics can provide a bridge between the discrete and continuous regimes. A common setting in which such limits are studied are geometric graphs that are constructed from ϵ-nets of a point cloud (i.e., nodes corresponding to pairs of points are connected, if they are within distance ϵ) and capture spatial information in the underlying data sets. Here, continuum limits characterize the geometry of the graph as the number of nodes grows larger and the resolution becomes finer. Limits of graph characteristics have found applications in the study of point clouds, in geometry processing tasks within computer vision, as well as in manifold learning, a classical task in geometric data analysis.

Continuum limits have been explored for a range of geometric characteristics. A key example are geodesic distances, where measuring distances in the (discrete) observation space can allow for inferring the "true" distances on the underlying data manifold (see, e.g., Davis and Sethuraman; Diaz et al. 2016). Another important example is the Graph Laplacian and its relation to the Laplace-Beltrami operator of the underlying data manifold (Calder and Trillos 2022; Hein, Audibert, and von Luxburg 2005). The Graph Laplacian is a tool widely used in graph machine learning applications, such as spectral clustering and semi-supervised learning (Von Luxburg 2007).

Recently, continuum limits of discrete Ricci curvature have also been studied. The first asymptotic convergence guarantee of Ollivier's curvature was given by (van der Hoorn et al. 2021). More comprehensive non-asymptotic convergence guarantees, which provide concrete error bounds with respect to the size of the data set, have been established in (Trillos and Weber 2023). Importantly, (Trillos and Weber 2023) also demonstrates that Ollivier's discrete Ricci curvature can be used to learn global bounds on the curvature of the underlying data manifold. This could have implications for manifold learning, where intrinsic curvature offers complementary insights to intrinsic dimension While intrinsic dimension is implicitly learned by classical algorithms such as MDS and Isomap, it does not allow for a characterization of the data's intrinsic curvature. Notably, these non-asymptotic results do not assume access to the true geodesic distances, a commonly used assumption that simplifies the analysis, but renders the results impractical as it requires knowledge of the true data manifold.

LEARNING ON GRAPHS

In this section, we review graph machine learning approaches that leverage the inherent geometric structure of graphs. This includes exploiting permutation symmetries in graph- and set-structured data, as well characterizations of a graph's geometry via discrete curvature.

We briefly comment on another geometric tool, the Graph Laplacian, which is frequently used in graph

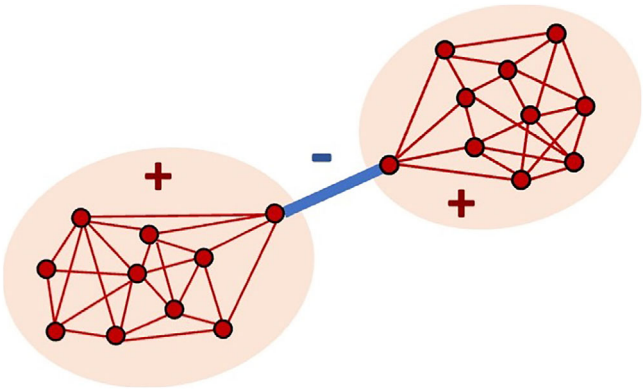

FIGURE 2 Curvature characterization of a graph.

machine learning, including for clustering and coarsening tasks. Its spectrum is particularly effective at capturing global and mesoscale structures in the graph. In contrast, discrete curvature characterizes the geometric properties of the graph more locally. This makes both approaches, to some extent, complementary. Since computing curvature only requires access to local information, it can be easily parallelized, which enhances scalability, specifically on large and dense graphs. On the other hand, computing the spectrum of the Graph Laplacian requires a spectral decomposition of the adjacency matrix, which can come at a higher computational cost. By understanding the strengths and limitations of both tools, practitioners can better choose the appropriate geometric framework depending on the specific learning task.

Shallow graph machine learning

Curvature-Based Clustering

Community detection, or unsupervised node clustering, is a fundamental task in graph learning, which seeks to identify densely connected substructures within a graph. Many, by now classical, algorithms have been proposed for this task, including spectral clustering (Von Luxburg 2007) and the Louvain algorithm (Blondel et al. 2008). More recently, approaches that leverage discrete Ricci curvatures and associated geometric flows have been proposed, including curvature-based thresholding (Fesser et al. 2023; Gosztolai and Arnaudon 2021; Sia, Jonckheere, and Bogdan 2019), as well as reweighting and thresholding guided by an associated discrete Ricci flow (Ni et al. 2019; Tian, Lubberts, and Weber 2023b). Both types of methods rely on the observation that the curvature of edges between communities is low, whereas edges within communities have high curvature (Figure 2). Curvature-based thresholding removes edges with low curvature, upon which the resulting connected components reveal the communities. The associated Ricci flow can be leveraged to refine this detection procedure. While these methods have primarily focused on single-membership community detection, where each node belongs to only one community, recent work has generalized curvature-based methods to the more difficult task of mixed-membership community detection (Tian, Lubberts, and Weber 2023a, 2023b). In these cases, curvature-based techniques are applied to the line graph, the dual representation of the original graph, which allows for better characterizing the geometric structure induced by nodes that belong to multiple communities.

Curvature-Based Coarsening

Graph coarsening allows for reducing the size of an input graph while preserving its key properties. Typically, this involves merging nodes through a process known as edge contraction, resulting in a smaller, simplified graph. Several geometric approaches have been leveraged to identify candidate edges for contraction, including spectral methods (Fiedler 1973; Spielman and Teng 1996), the Louvain algorithm (Blondel et al. 2008), and Graclus (Dhillon, Guan, and Kulis 2007). From the perspective of discrete curvature, edges with high curvature are good candidates for contraction, as they represent less critical connections in the graph. As such, curvature-based pooling operators, which implement similar algorithmic ideas as presented above in the context of community detection may be leveraged (Feng and Weber 2024; Weber, Saucan, and Jost 2017a, 2017b). Graph Coarsening has applications in exploratory data analysis via visualization of the coarsened graph, as well as the reduction and simplification of large input graphs to enable more computationally expensive downstream analysis (Weber et al. 2017c).

Deep graph machine learning

Message-passing Graph Neural Networks

Graph Neural Networks (GNNs) have emerged as a popular architecture for deep learning on graphs. Many GNNs implement the *message-passing* paradigm (Gori, Monfardini, and Scarselli 2005; Hamilton, Ying, and Leskovec 2017), which iteratively learns node embeddings h_v^l that jointly represent information encoded in the adjacency and the attributes of the input graph (Figure 3). During each iteration ("layer") l, nodes aggregate information from their neighbors (the "message" m_v^l) and then update their own representation based on this new information. This allows for capturing increasingly complex structure as the depth of the network increases. The aggregation f_{Agg} and update functions f_{Up} that specify each layer are typically implemented using neural networks with trainable parameters. Examples of message-passing GNNs

include Graph Convolutional Networks (Kipf and Welling 2017), Graph Isomorphism Networks (Xu et al. 2018), GraphSAGE (Hamilton, Ying, and Leskovec 2017), and Graph Attention Networks (Veličković et al. 2017).

Challenges during training

While Graph Neural Networks have achieved significant success across various learning tasks and domains, they are not without challenges. Two primary issues are the inability to effectively utilize information encoded in long-range connections (an effect known as *over-squashing* (Alon and Yahav 2021)), and the difficulty in maintaining distinct node representations as the network depth increases (known as *over-smoothing* (Li, Han, and Wu 2018)). Over-squashing occurs when information encoded in long-range connections is compressed through "bottlenecks" in message-passing, which can make it difficult to capture this information in the learned representations as the number of layers in the network grows. From a geometric perspective, it can be observed that edges that cause such "bottlenecks" have low curvature (Topping et al. 2022). On the other hand, over-smoothing happens when, with increasing network depth, the representations of dissimilar nodes become increasingly indistinguishable. This arises in particular in dense regions of the graph, which are often characterized by high curvature (see Figure 2).

As such, discrete curvature offers a useful framework for characterizing both effects. Curvature can also be used to mitigate both effects during training, specifically through rewiring, a preprocessing technique that perturbs the edges of the input graph. Several rewiring approaches have been proposed, many guided by graph characteristics like discrete curvature (Topping et al. 2022; Fesser and Weber 2023; Nguyen et al. 2023) or the spectrum of the Graph Laplacian (Karhadkar, Banerjee, and Montufar 2023). Curvature-based rewiring performs the following edge perturbations: In regions with high curvature, edges are removed to prevent over-smoothing, while in areas with low curvature, synthetic edges are added to counteract over-squashing. A key question is to determine a suitable number of edges to add and remove. Here, again, the geometry of the graph can provide guidance: The *curvature gap* (Gosztolai and Arnaudon 2021) is a global graph characteristic derived from the distribution of curvature values across the graph, from which a threshold for identifying regions with "too high" or "too low" curvature can be deduced. This, in turn, informs the number of edges that need to be perturbed for effective rewiring.

Improving the representational power of GNNs via geometric augmentation

Another key question regarding the effectiveness of GNNs is what types of functions they can and cannot learn. This question can be analyzed through the following lens: *Does a GNN map two graphs to the same representation if and only if they are topologically identical?* Unfortunately, standard message-passing GNNs are unable to learn such a mapping for notable classes of graphs, including simple examples such as regular graphs (Morris et al. 2019; Xu et al. 2018). As a consequence, GNNs are unable to learn some basic functions, such as calculating the diameter of a graph or determining the size of its largest cycle. What can be done to address this? More powerful GNN architectures, such as Higher-order GNNs (Morris et al. 2019) and topology-aware message passing that encodes local substructures (Bouritsas et al. 2022; Zhao et al. 2022) are able to distinguish a larger set of non-topologically identical graphs. However, often these architectures are computationally less efficient. A simpler alternative is to augment the graph with geometric information that allows for learning richer node representations, which enhances their utility in downstream tasks. Various types of such augmentations, also known as "encodings", have been introduced, leveraging tools like the spectrum of the Graph Laplacian (Dwivedi et al. 2022) and discrete curvature (Fesser and Weber 2024).

As we discussed above, curvature captures local structural information within a node's two-hop neighborhood in contrast to message-passing, which only captures information in the one-hop neighborhood. As a

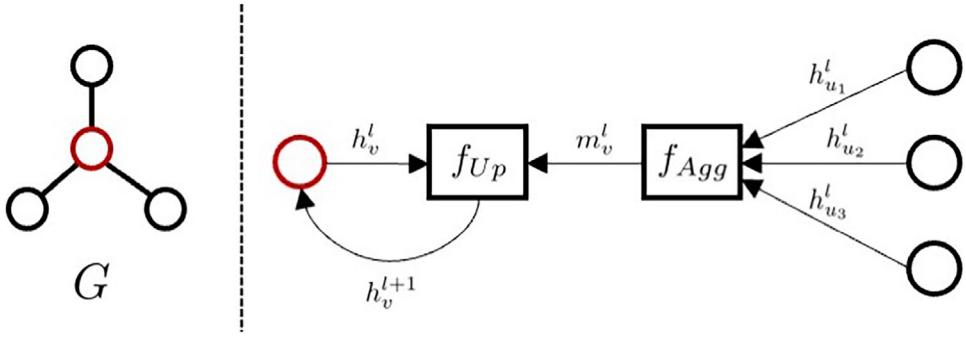

FIGURE 3 Message-passing paradigm.

consequence, simple curvature augmentations are surprisingly effective encodings and can lead to significant performance improvements, often outperforming encodings based on other graph characteristics (Fesser and Weber 2024). Importantly, curvature encodings are relatively scalable compared to other methods. Since encodings represent a preprocessing routine, scalability is important to balance the benefits of enhanced representational power with the computational resources required for computing the encodings.

LEARNING ON MATRIX SPACES

In many machine learning and data science applications we encounter optimization tasks that involve matrices, both in the form of standalone algorithms or as subroutines of more complex algorithms. Traditionally, these tasks are solved with tools from Euclidean optimization, which treat the underlying matrix space as an Euclidean space. Here, the structure of the matrices, such as orthogonality or positive definiteness, is enforced through constraints, necessitating the use of constrained optimization techniques. However, matrix spaces can also be viewed as Riemannian manifolds, where the structural properties are inherently encoded in the domain's parametrization. This perspective can offer two significant advantages. First, because the structure is implicitly encoded in the problem's formulation, there is no need to impose explicit constraints. This transforms the constrained problem into an *unconstrained* one, which is often much easier to solve. Second, the difficulty of solving an optimization problem depends crucially on its convexity. Interestingly, many matrix-valued optimization problems that are nonconvex in the Euclidean setting admit a *geodesically convex* formulation in the Riemannian setting. This allows for the application of (geodesically) convex optimization techniques with global optimality guarantees.

Identifying and certifying geodesic convexity

Geodesic convexity (short *g-convexity*) is a simple extension of the classical Euclidean convexity concept to the geometric setting. Formally, we say that a function f is g-convex, if for all $x, y \in \mathcal{M}$, the inequality $f(\gamma(t)) \leq (1-t)f(x) + tf(y)$ holds for all $t \in [0,1]$. This corresponds to evaluating the usual convexity condition along a geodesic $\gamma : [0,1] \to \mathcal{M}$, which represents the shortest path between the points $x = \gamma(0)$ and $y = \gamma(1)$ on the manifold $\mathcal{M}$; its shape is determined by the geometry of the manifold. Importantly, g-convexity guarantees the existence of a global optimum, which can be found with first-order methods from any initialization. In order to leverage this useful property, identifying g-convexity in optimization tasks is of great interest. In the Euclidean setting, *Disciplined Convex Programming* (DCP) (Grant, Boyd, and Ye 2006) is a well-established framework that automates the verification of convexity for a wide range of functions. DCP achieves this by decomposing functions into basic convex components (atoms) and applying convexity-preserving operations (rules). This idea can be extended to the geometric setting by leveraging the algebraic properties of g-convex functions. *Disciplined Geodesically Convex Programming* (Cheng, Dixit, and Weber 2024) generalizes this framework to Cartan-Hadamard manifolds, specifically for optimization on the manifold of symmetric, positive definite matrices.

Algorithms for Riemannian optimization

Motivated by the advantages discussed above, Riemannian optimization has gained significant traction in machine learning and data science. Many classical algorithms have been extended to the geometric setting, resulting in generalized approaches for convex (Bacák 2014; Udriste 1994; Zhang and Sra 2016), nonconvex (Boumal, Absil, and Cartis 2019; Weber and Sra 2021), stochastic (Bonnabel 2013; Weber and Sra 2021; Zhang, Reddi, and Sra 2016), constrained (Bergmann et al. 2022; Weber and Sra 2021, 2022), and min-max optimization problems (Jordan, Lin, and Vlatakis-Gkaragkounis 2022; Martínez-Rubio et al. 2023). Moreover, several numerical software packages for solving geometric optimization problems have been developed across different programming languages (Bergmann 2022; Boumal, Mishra, Absil, and Sepulchre 2014; Townsend, Koep, and Weichwald 2016), allowing for ease of application in practice. However, the computational overhead associated with the use of Riemannian tools in algorithms can be non-negligible. This has motivated the investigation of "mixed approaches," where Euclidean algorithms are used to solve g-convex problems, while geometric tools are employed to certify global optimality. For instance, in problems with a difference of convex structure, a purely Euclidean CCCP algorithm can be used as a solver, and a Riemannian analysis can then be leveraged to certify its global optimality due to g-convexity (Weber and Sra 2023).

LEARNING UNDER SYMMETRY

Lastly, we consider a setting where geometric structure in data and models arises from symmetries. As discussed above, symmetries can be understood through transformations of the input space that leave the labels unchanged, that is, $f(UxU^T) = f(x)$ for all $x \in \mathcal{X}$. Geometric Deep Learning (Bronstein et al. 2021) provides a framework for explicitly encoding such structure into models, by constraining it to learn only functions with the desired

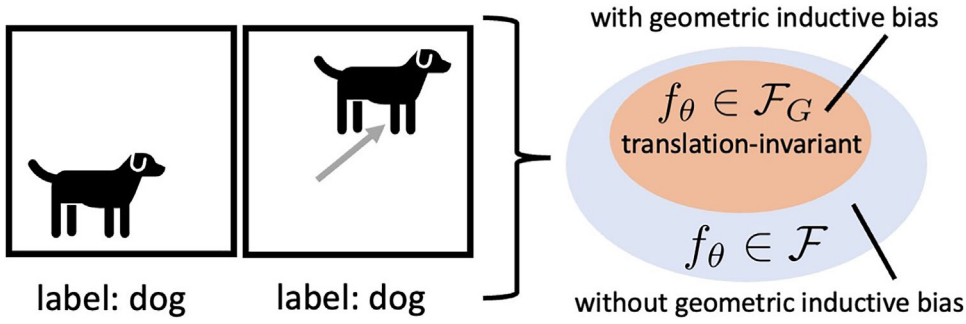

FIGURE 4 Learning with and without geometric inductive bias.

symmetries. Such *geometric inductive bias* can improve the model's efficiency, including by improving the model's ability to generalize well.

Equivariant Neural Networks

Equivariant neural networks are a deep learning architecture, which encode symmetries as inductive biases by adapting the layers to preserve symmetries via equivariances, that is, ensuring that $f(UxU^T) = Uf(x)U^T$ (Figure 4). For example, in a classical CNN, which encodes translation invariance, the convolutional layer is translation-equivariant. Nonlinear layers in these networks use activation functions that preserve equivariance. Various architectures for equivariant networks have been proposed to handle specific groups. These include networks designed for rotation-invariant domains (Cohen, Geiger, Köhler, and Welling 2018), permutation-invariant domains like graphs (Kipf and Welling 2017) and sets (Zaheer et al. 2017), as well as more general groups (Cohen and Welling 2016; Dym and Maron 2020; Finzi, Welling, and Wilson 2021; Kiani, Fesser, and Weber 2024a; Kondor and Trivedi 2018; Villar et al. 2021).

Representation trade-offs

A key question in the context of equivariant neural networks and, in fact, more general architectures, is that of representation trade-offs: *What do we gain from geometric inductive biases, such as symmetries?* Empirically, equivariant neural networks have been shown to offer computational benefits in various applications (Wang, Walters, and Yu 2021), but can this be formalized in terms of computational complexity?

Most research on this question has focused on the impact of symmetry and the corresponding inductive biases on generalization. Notably, (Bietti, Venturi, and Bruna 2021; Mei, Misiakiewicz, and Montanari 2021) established theoretical guarantees for improved generalization in shallow linear group-convolutional neural networks when compared to their fully-connected counterparts. Such trade-offs have also been explored for other geometric methods, such as kernel regression (Tahmasebi and Jegelka 2023) and robust classification in hyperbolic space (Weber et al. 2020).

Beyond generalization, one can also analyze the complexity of learning algorithms designed for symmetric function classes by assessing how effectively neural networks can be trained. Formally, such an analysis can be performed in the statistical query (SQ) framework (Kearns (1998)), where hardness is quantified by the number of queries (akin to steps in a learning algorithm) required to achieve a certain level of accuracy. Gradient descent, the standard method for training deep learning models, is a prime example of this approach. What can we say about the complexity of learning neural networks, geometric or not? Studies on the hardness of learning fully-connected neural networks (i.e., without considering symmetries) have shown that learning even shallow networks on simple data distributions is computationally hard (Chen et al. 2022; Diakonikolas et al. 2020). These findings suggest that, in general, it is not feasible to efficiently learn every function that a small neural network can represent under typical data distributions. This raises an important question: Can incorporating symmetry as inductive bias help overcome these computational barriers? Unfortunately, no. Imposing symmetry on a function class does indeed reduce the computational complexity of learning. For example, for finite groups, the reduction is proportional to the size of the group (Kiani et al. 2024b). However, simply encoding symmetries does not eliminate the inherent difficulty of training neural networks. Overcoming these challenges requires additional geometric assumptions. For instance, recent work (Kiani, Wang, and Weber 2024c) has shown that while learning remains hard under input manifolds with bounded curvature, it becomes feasible when there are additional constraints on the volume of the data manifold, such as those arising from bounded Ricci curvature.

CONCLUSIONS

This article surveyed work at the intersection of geometry and machine learning, focusing on characterizing geometric structure in data and the design of algorithms and architectures that leverage such structure to learn more efficiently. The field of *Geometric Machine Learning* has gained significant momentum, not least driven by the successful application of geometric architectures in analyzing scientific data and developing physics-informed models that can aid scientific discovery. There are still many challenges and opportunities in this area, such as encoding more diverse geometric structures into models; developing computational techniques that enhance the scalability of geometric models to ensure broader applicability; and gaining a deeper understanding of when and why geometric architectures provide algorithmic advantages.

ACKNOWLEDGMENTS

This work was supported by NSF awards CBET-2112085 and DMS-2406905.

CONFLICT OF INTEREST STATEMENT

The author declares that there is no conflict.

ORCID

Melanie Weber https://orcid.org/0000-0003-1104-7181

REFERENCES

Alon, Uri and Eran Yahav. 2021, "On the Bottleneck of Graph Neural Networks and Its Practical Implications." In *International Conference on Learning Representations*.

Azizzadenesheli, Kamyar, Nikola Kovachki, Zongyi Li, Miguel Liu-Schiaffini, Jean Kossaifi, and Anima Anandkumar. 2024. "Neural Operators for Accelerating Scientific Simulations and Design." *Nature Reviews Physics* 6: 320–28.

Bacák, Miroslav. 2014. "Convex Analysis and Optimization in Hadamard Spaces." In *Convex Analysis and Optimization in Hadamard Spaces*. de Gruyter.

Batzner, Simon, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. 2022. "E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials." *Nature Communications* 13(1): 1–11.

Bekkers, Erik J., Maxime W. Lafarge, Mitko Veta, Koen A. J. Eppenhof, Josien P. W. Pluim, and Remco Duits. 2018. "Roto-Translation Covariant Convolutional Networks for Medical Image Analysis." In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Proceedings, Part I*, 440–48. Granada, Spain: Springer.

Bergmann, Ronny. 2022. "Manopt.jl: Optimization on Manifolds in Julia." *Journal of Open Source Software* 7(70): 3866. https://doi.org/10.21105/joss.03866.

Bergmann, Ronny, Roland Herzog, Julián Ortiz López, and Anton Schiela. 2022. "First-and Second-Order Analysis for Optimization Problems With Manifold-Valued Constraints." *Journal of Optimization Theory and Applications* 195(2): 596–623.

Bhasker, Nithya, Hattie Chung, Louis Boucherie, Vladislav Kim, Stefanie Speidel, and Melanie Weber. 2024. "Contrastive Poincaré Maps for Single-Cell Data Analysis." In *ICLR 2024 Workshop on Machine Learning for Genomics Explorations*. ICLR.

Bietti, Alberto, Luca Venturi, and Joan Bruna. 2021. "On the Sample Complexity of Learning Under Geometric Stability." In *Advances in Neural Information Processing Systems*, edited by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, and J. Wortman Vaughan, Vol 34. 18673–84. Curran Associates, Inc.

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics: Theory and Experiment* 2008(10): P10008.

Bonnabel, Silvere. 2013. "Stochastic Gradient Descent on Riemannian Manifolds." *IEEE Transactions on Automatic Control* 58(9): 2217–29.

Boumal, Nicolas, Bamdev Mishra, P-A Absil, and Rodolphe Sepulchre. 2014. "Manopt, a Matlab Toolbox for Optimization on Manifolds." *The Journal of Machine Learning Research* 15(1): 1455–59.

Boumal, Nicolas, Pierre-Antoine Absil, and Coralia Cartis. 2019. "Global Rates of Convergence for Nonconvex Optimization on Manifolds." *IMA Journal of Numerical Analysis* 39(1): 1–33.

Bouritsas, Giorgos, Fabrizio Frasca, Stefanos Zafeiriou, and Michael M. Bronstein. 2022. "Improving Graph Neural Network Expressivity Via Subgraph Isomorphism Counting." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45(1): 657–68.

Bronstein, Michael M., Joan Bruna, Taco Cohen, and Petar Veličković. 2021. "Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges." *arxiv.2104.13478*.

Calder, Jeff and Nicolás García Trillos. 2022. "Improved Spectral Convergence Rates for Graph Laplacians on ε-Graphs and k-NN Graphs." *Applied and Computational Harmonic Analysis* 60: 123–75.

Chen, Sitan, Aravind Gollakota, Adam Klivans, and Raghu Meka. 2022. "Hardness of Noise-Free Learning for Two-Hidden-Layer Neural Networks." *Advances in Neural Information Processing Systems* 35: 10709–24.

Cheng, Andrew, Vaibhav Dixit, and Melanie Weber. 2024. "Disciplined Geodesically Convex Programming." *arXiv preprint arXiv:2407.05261*.

Cohen, Taco, and Max Welling. 2016. "Group Equivariant Convolutional Networks." In *International Conference on Machine Learning*, 2990–99. PMLR.

Cohen, Taco S., Mario Geiger, Jonas Köhler, and Max Welling. "Spherical CNNs." *arXiv preprint arXiv:1801.10130*.

Davis, Erik and Sunder Sethuraman. "Approximating Geodesics Via Random Points." *The Annals of Applied Probability* 29(3): 1446–86.

Dhillon, Inderjit S., Yuqiang Guan, and Brian Kulis. 2007. "Weighted Graph Cuts Without Eigenvectors a Multilevel Approach." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(11): 1944–57.

Diakonikolas, Ilias, Daniel M. Kane, Vasilis Kontonis, and Nikos Zarifis. 2020. "Algorithms and sq Lower Bounds for Pac Learning One-Hidden-Layer Relu Networks." In *Conference on Learning Theory*, 1514–39. PMLR.

Diaz, Josep, Dieter Mitsche, Guillem Perarnau, and Xavier Pérez-Giménez. 2016. "On the Relation Between Graph Distance and Euclidean Distance in Random Geometric Graphs." *Advances in Applied Probability* 48(3): 848–64.

Diepeveen, Willem, Carlos Esteve-Yagüe, Jan Lellmann, Ozan Öktem, and Carola-Bibiane Schönlieb. 2024. "Riemannian Geometry for Efficient Analysis of Protein Dynamics Data." *Proceedings of the National Academy of Sciences* 121(33): e2318951121.

Prakash Dwivedi, Vijay, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. 2022. "Graph Neural Networks With Learnable Structural and Positional Representations." In *International Conference on Learning Representations*.

Dym, Nadav and Haggai Maron. 2020. "On the Universality of Rotation Equivariant Point Cloud Networks." *arXiv preprint arXiv:2010.02449*

Federer, Herbert. 1959. "Curvature Measures." *Transactions of the American Mathematical Society* 93(3): 418–91.

Feng, Amy and Melanie Weber. 2024. "Graph Pooling Via Ricci Flow." *Transactions on Machine Learning Research*

Fesser, Lukas, and Melanie Weber. 2023. "Mitigating Over-Smoothing and Over-Squashing Using Augmentations of Forman-Ricci Curvature." In *Learning on Graphs Conference*.

Fesser, Lukas, and Melanie Weber. 2024. "Effective Structural Encodings Via Local Curvature Profiles." In *International Conference on Learning Representations*.

Fesser, Lukas, Sergio Serrano de Haro Iváñez, Karel Devriendt, Melanie Weber, and Renaud Lambiotte. 2023. "Augmentations of Forman's Ricci Curvature and Their Applications in Community Detection." Journal of Physics: Complexity 5(3)

Fiedler, Miroslav. 1973. "Algebraic Connectivity of Graphs." *Czechoslovak Mathematical Journal* 23(2): 298–305.

Finzi, Marc, Max Welling, and Andrew Gordon Gordon Wilson. 2021. "A Practical Method for Constructing Equivariant Multilayer Perceptrons for Arbitrary Matrix Groups." In *Proceedings of the 38th International Conference on Machine Learning*, 3318-28. PMLR.

Forman, Robin. 2003. "Bochner's Method for Cell Complexes and Combinatorial Ricci Curvature." *Discrete & Computational Geometry* 29(3): 323–74.

Ganea, Octavian, Lagnajit Pattanaik, Connor Coley, Regina Barzilay, Klavs Jensen, William Green, and Tommi Jaakkola. 2021. "Geomol: Torsional Geometric Generation of Molecular 3d Conformer Ensembles." *Advances in Neural Information Processing Systems* 34: 13757–69.

Gori, Monfardini, and Scarselli. "A New Model for Learning in Graph Domains." In *Proceedings. 2005 IEEE international joint conference on neural networks*, Vol 2. 729–34. IEEE.

Gosztolai, Adam, and Alexis Arnaudon. 2021. "Unfolding the Multiscale Structure of Networks With Dynamical Ollivier-Ricci Curvature." *Nature Communications* 12(1): 4561.

Grant, Michael, Stephen Boyd, and Yinyu Ye. 2006. *Disciplined Convex Programming*. Springer.

Hamilton, Will, Zhitao Ying, and Jure Leskovec. 2017. "Inductive Representation Learning on Large Graphs." *Advances in neural information processing systems* 30: 1025–35.

Hein, Matthias, Jean-Yves Audibert, and Ulrike von Luxburg. 2005. "From Graphs to Manifolds—Weak and Strong Pointwise Consistency of Graph Laplacians." In *Learning Theory, Lecture Notes in Computer Science*, edited by Peter Auer and Ron Meir, 470–85, Berlin, Heidelberg, Springer.

Jordan, Michael, Tianyi Lin, and Emmanouil-Vasileios Vlatakis-Gkaragkounis. 2022. "First-Order Algorithms for Min-Max Optimization in Geodesic Metric Spaces." *Advances in Neural Information Processing Systems* 35: 6557–74.

Karhadkar, Kedar, Pradeep Kr. Banerjee, and Guido Montufar. 2023. "FoSR: First-Order Spectral Rewiring for Addressing Oversquashing in GNNs." In *The Eleventh International Conference on Learning Representations*.

Kashinath, K., M. Mustafa, A. Albert, J. L. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, R. Wang, A. Chattopadhyay, A. Singh, et al. 2021. "Physics-Informed Machine Learning: Case Studies for Weather and Climate Modelling." *Philosophical transactions. Series A, Mathematical, Physical, and Engineering Sciences* 379(2194): 20200093–20200093.

Kearns, Michael. 1998. "Efficient Noise-Tolerant Learning From Statistical Queries." *Journal of the ACM (JACM)* 45(6): 983–1006.

Kiani, Bobak, Lukas Fesser, and Melanie Weber. 2024a. *Unitary Convolutions for Learning on Graphs and Groups*. Advances in Neural Information Processing Systems 37 (NeurIPS)

Kiani, Bobak T., Thien Le, Hannah Lawrence, Stefanie Jegelka, and Melanie Weber. 2024b. "On the Hardness of Learning Under Symmetries." In *International Conference on Learning Representations*.

Kiani, Bobak T., Jason Wang, and Melanie Weber. 2024c. "Hardness of Learning Neural Networks Under the Manifold Hypothesis." *Advances in Neural Information Processing Systems 37 (NeurIPS)*.

Kipf, Thomas N. and Max Welling. 2017. "Semi-Supervised Classification With Graph Convolutional Networks." In *Proceedings of the 5th International Conference on Learning Representations*, ICLR '17. ICLR.

Klimovskaia, Anna, David Lopez-Paz, Léon Bottou, and Maximilian Nickel. June 2020. "Poincaré Maps for Analyzing Complex Hierarchies in Single-Cell Data." *Nature Communications* 11(1): 2966. ISSN 2041-1723.

Kondor, Risi, and Shubhendu Trivedi. 2018. "On the Generalization of Equivariance and Convolution in Neural Networks to the Action of Compact Groups." In *International Conference on Machine Learning*, 2747–55. PMLR.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet Classification With Deep Convolutional Neural Networks." In *Advances in Neural Information Processing Systems*, Vol. 25. MIT Press.

Kuhn, H. 1955. "The Hungarian Method for the Assignment Problem." *Naval Research Logistics Quarterly* 2(1-2): 83–97.

Lecun, Y., L. Bottou, Y. Bengio, and P. Haffner. 1998. "Gradient-Based Learning Applied to Document Recognition." *Proceedings of the IEEE* 86(11): 2278–324.

Li, Kovachki, Azizzadenesheli, Liu, Bhattacharya, Stuart, and Anandkumar. 2021. "Fourier Neural Operator for Parametric Partial Differential Equations." In *International Conference on Learning Representations*.

Li, Qimai, Zhichao Han, and Xiao-Ming Wu. "Deeper Insights Into Graph Convolutional Networks for Semi-Supervised Learning." In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32. 3538–45. AAAI Press.

Martínez-Rubio, David, Christophe Roux, Christopher Criscitiello, and Sebastian Pokutta. 2023. "Accelerated Methods for Riemannian Min-Max Optimization Ensuring Bounded Geometric Penalties." *arXiv preprint arXiv:2305.16186*.

Mei, Song, Theodor Misiakiewicz, and Andrea Montanari. 15–19 Aug 2021. "Learning With Invariances in Random Features and Kernel

Models." In *Proceedings of Thirty Fourth Conference on Learning Theory*, edited by Mikhail Belkin and Samory Kpotufe, *Proceedings of Machine Learning Research*, Vol. 134. 3351–418. PMLR.

Morris, Christopher, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. 2019. "Weisfeiler and Leman Go Neural: Higher-Order Graph Neural Networks." In *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33, 4602–9. AAAI Press.

Nguyen, Khang, Nong Minh Hieu, Vinh Duc Nguyen, Nhat Ho, Stanley Osher, and Tan Minh Nguyen. July 23–29 2023. "Revisiting Over-Smoothing and Over-Squashing Using Ollivier-Ricci Curvature." In *Proceedings of the 40th International Conference on Machine Learning, Proceedings of Machine Learning Research*, Vol. 202, 25956–79. PMLR.

Ni, Chien-Chun, Yu-Yao Lin, Feng Luo, and Jie Gao. 2019. "Community Detection on Networks With Ricci Flow." *Scientific Reports* 9(1): 9984.

Ollivier, Yann. 2010. "A Survey of Ricci Curvature for Metric Spaces and Markov Chains." In *Probabilistic Approach to Geometry*, 343–81. Mathematical Society of Japan.

Sia, Jayson, Edmond Jonckheere, and Paul Bogdan. 2019. "Ollivier-Ricci Curvature-Based Method to Community Detection in Complex Networks." *Scientific Reports* 9(1): 9800.

Sinkhorn, Richard, and Paul Knopp. 1967. "Concerning Nonnegative Matrices and Doubly Stochastic Matrices." *Pacific Journal of Mathematics* 21(2): 343–48.

Spielman, Daniel A., and Shang-Hua Teng. 1996. "Spectral Partitioning Works: Planar Graphs and Finite Element Meshes." In *Proceedings of 37th Conference on Foundations of Computer Science*, 96–105. IEEE.

Subramanian, Akshay, Wenhao Gao, Regina Barzilay, Jeffrey C. Grossman, Tommi Jaakkola, Stefanie Jegelka, Mingda Li, Ju Li, Wojciech Matusik, Elsa Olivetti, Connor W. Coley, et al. March 27 2024. "Closing the Execution Gap in Generative AI for Chemicals and Materials: Freeways or Safeguards." *An MIT Exploration of Generative AI*, https://mit-genai.pubpub.org/pub/681kpeoa.

Tahmasebi, Behrooz, and Stefanie Jegelka. 2023. "The Exact Sample Complexity Gain From Invariances for Kernel Regression." *Advances in Neural Information Processing Systems* 36: 55616–46.

Tian, Yu, Zachary Lubberts, and Melanie Weber. 03 Dec 2023a. "Mixed-Membership Community Detection Via Line Graph Curvature." In *Proceedings of the 1st NeurIPS Workshop on Symmetry and Geometry in Neural Representations, Proceedings of Machine Learning Research*, Vol. 197, 219–33. PMLR.

Tian, Yu, Zachary Lubberts, and Melanie Weber. 2023b. "Curvature-Based Clustering on Graphs." *arXiv:2307.10155*.

Topping, Jake, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M. Bronstein. 2022. "Understanding Over-Squashing and Bottlenecks on Graphs Via Curvature." In *International Conference on Learning Representations*.

Townsend, James, Niklas Koep, and Sebastian Weichwald. 2016. "Pymanopt: A Python Toolbox for Optimization on Manifolds Using Automatic Differentiation." *Journal of Machine Learning Research* 17(137): 1–5.

Trillos, Nicolas Garcia and Melanie Weber. "Continuum Limits of Ollivier's Ricci Curvature on Data Clouds: Pointwise Consistency and Global Lower Bounds." *arXiv:2307.02378*.

Udriste, Constantin. 1994. *Convex Functions and Optimization Methods on Riemannian Manifolds*, Vol. 297, Springer Science & Business Media.

van der Hoorn, Pim, William J. Cunningham, Gabor Lippner, Carlo Trugenberger, and Dmitri Krioukov. 2021. "Ollivier-Ricci Curvature Convergence in Random Geometric Graphs." *Physical Review Research* 3(1): 013211.

Veličković, Petar, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. "Graph Attention Networks." *arXiv preprint arXiv:1710.10903*.

Villar, Soledad, David W. Hogg, Kate Storey-Fisher, Weichi Yao, and Ben Blum-Smith. 2021. "Scalars are Universal: Equivariant Machine Learning, Structured Like Classical Physics." *Advances in Neural Information Processing Systems* 34: 28848–63.

Von Luxburg, Ulrike. 2007. "A Tutorial on Spectral Clustering." *Statistics and Computing* 17: 395–416.

Wang, Dian, Robin Walters, Xupeng Zhu, and Robert Platt. 2022. "Equivariant q Learning in Spatial Action Spaces." In *Conference on Robot Learning*, 1713–23. PMLR.

Wang, Rui, Robin Walters, and Rose Yu. 2021. "Incorporating Symmetry Into Deep Dynamics Models for Improved Generalization." In *International Conference on Learning Representations*.

Watson, Joseph L., David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, et al. 2023. "De Novo Design of Protein Structure and Function With RFdiffusion." *Nature* 620(7976): 1089–100.

Weber, Melanie. 2024. "Exploiting Data Geometry in Machine Learning." *Proceedings of the AAAI Conference on Artificial Intelligence* 38(20): 22681–22681.

Weber, Melanie, and Suvrit Sra. 2021. "Projection-Free Nonconvex Stochastic Optimization on Riemannian Manifolds." *IMA Journal of Numerical Analysis* 42(4): 3241–71.

Weber, Melanie and Suvrit Sra. 2022. "Riemannian Optimization via Frank–Wolfe Methods." *Mathematical Programming* 199: 525–56.

Weber, Melanie and Suvrit Sra. 2023. "Global Optimality for Euclidean CCCP Under Riemannian Convexity." In *International Conference on Machine Learning*.

Weber, Melanie, Emil Saucan, and Jürgen Jost. 2017a. "Characterizing Complex Networks With Forman-Ricci Curvature and Associated Geometric Flows." *Journal of Complex Networks* 5(4): 527–50.

Weber, Melanie, Emil Saucan, and Jürgen Jost. "Coarse Geometry of Evolving Networks." *Journal of Complex Networks* 6(5): 706–32.

Weber, Melanie, Johannes Stelzer, Emil Saucan, Alexander Naitsat, Gabriele Lohmann, and Jürgen Jost. 2017c. "Curvature-Based Methods for Brain Network Analysis." *Technical Report (arXiv:1707.00180)*.

Weber, Melanie, Manzil Zaheer, Ankit Singh Rawat, Aditya Menon, and Sanjiv Kumar. 2020. "Robust Large-Margin Learning in Hyperbolic Space." *Advances in Neural Information Processing Systems 33 (NeurIPS)*. Cambridge, MA, USA: MIT Press.

Xu, Keyulu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2018. "How Powerful are Graph Neural Networks?" *arXiv preprint arXiv:1810.00826*.

Zaheer, Manzil, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R. Salakhutdinov, and Alexander J. Smola. 2017. "Deep Sets." *Advances in Neural Information Processing Systems* 30: 3394–404.

Zhang, Hongyi, and Suvrit Sra. 2016. "First-Order Methods for Geodesically Convex Optimization." *Conference on Learning Theory* 49: 1617–38.

Zhang, Hongyi, Sashank J. Reddi, and Suvrit Sra. 2016. "Riemannian SVRG: Fast Stochastic Optimization On Riemannian Manifolds." *Advances in Neural Information Processing Systems* 29: 2893–901.

Zhao, Lingxiao, Wei Jin, Leman Akoglu, and Neil Shah. "From Stars to Subgraphs: Uplifting any GNN With Local Structure Awareness." In *International Conference on Learning Representations*. ICLR.

Zhao, Zhizhen and Amit Singer. 2014. "Rotationally Invariant Image Representation for Viewing Direction Classification in Cryo-EM." *Journal of Structural Biology* 186(1): 153–66.

How to cite this article: Weber, M. 2025. "Geometric Machine Learning." *AI Magazine* 46: e12210. https://doi.org/10.1002/aaai.12210

AUTHOR BIOGRAPHY

Melanie Weber is an Assistant Professor of Applied Mathematics and of Computer Science at Harvard University, where she leads the Geometric Machine Learning Group. Her research studies geometric structure in data and how to leverage such information for the design of new, efficient Machine Learning algorithms with provable guarantees. Previously, she was a Hooke Research Fellow at the Mathematical Institute in Oxford and received her PhD from Princeton University. She is a recipient of the IMA Leslie Fox Prize in Numerical Analysis and a Sloan Research Fellowship in Mathematics.