Geometry and Machine Learning: A Survey for Data Scientists and Machine Learning Researchers

By Colleen M. Farrelly, Independent Researcher

[cfarrelly@med.miami.edu](mailto:cfarrelly@med.miami.edu)

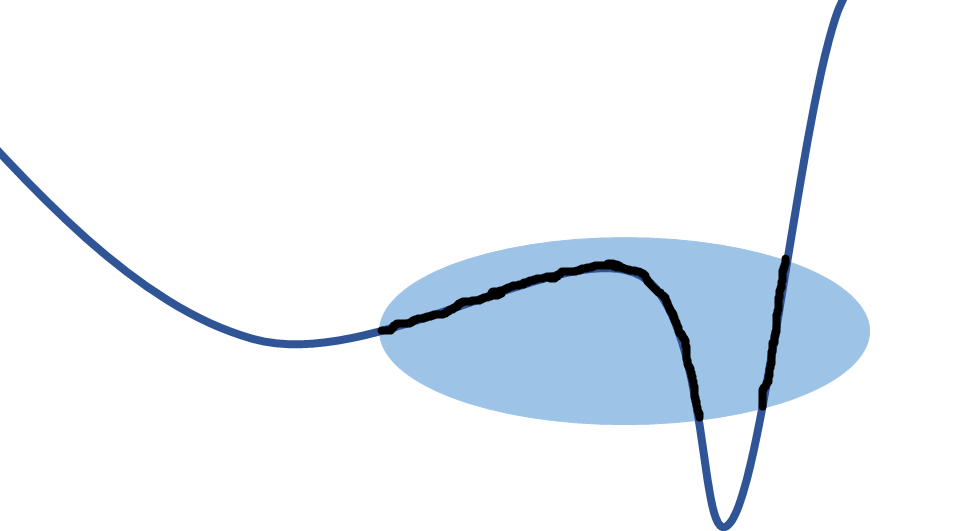
**Introduction**

Many machine learning algorithms and statistical models rely heavily upon matrices and linear algebra for computation. Linear algebra is well-suited to modern computing, as operations can be computed quickly, have decent enough accuracy, and only rely on a few assumptions about the data (usually relating to linear independence of columns/rows, sample sizes being larger than the number of predictors, and determinant values of the matrix). However, assumptions sometimes fail in the real world, and accuracy is not always as good as a machine learning practitioner might need. Data may not even lie within a linear space. Fortunately, a plethora of alternatives to linear-algebra-based algorithms are being actively developed, providing machine learning researchers and data scientists with many useful tools.

This new set of tools and algorithms is highlighting problem areas within many popular machine learning frameworks, creating tools that can work on extremely small datasets or datasets with many correlated predictors, and exploring the synergy between disparate fields of mathematics and computer science. Many of these algorithms rely on data and model geometry, and the methods detailed in Part 1 (algebraic geometry) and Part 2 (differential geometry) of this overview are only a small subset of those being actively explored and developed.

**Part 1: Algebraic Geometry**

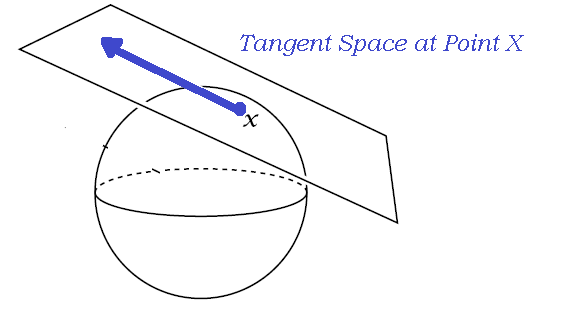
The subfield of algebraic geometry has yielded several recent advances in algorithm design, model selection, and robustness of algorithms; it is also finding a niche in the study of engineering problems and dynamic systems. Algebraic geometry is a field of mathematics focusing on the intersection of curves in real and complex spaces. For instance, in the example below, a 1-dimensional curve (dark blue) intersects a 2-dimensional ellipse (light blue), creating a 1-dimensional overlap curve within the ellipse (black line). These overlaps among curves/shapes of varying dimension are called algebraic varieties, and they represent solutions to nonlinear systems of equations.



Matrices and matrix operations can be used to solve these systems analogously to linear algebra; numbers populating a matrix are simply replaced with polynomial equations. Various software platforms, including Bertini and Macauley, can then be used to solve the system of nonlinear equations using sets of algorithms. Several recent papers and workshops in numerical algebraic geometry suggest that nonlinear algebra is a viable alternative to linear algebra in machine learning algorithms, yielding algorithms with fewer mathematical assumptions about the matrix with increased accuracy.

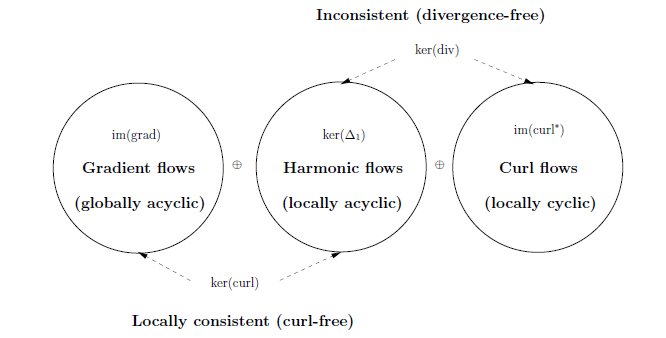
Nonlinear algebra handles the non-convexity of real-world optimization problems in ways that linear algebra can’t, and it has been used to explore machine learning model selection and parameter estimation methods. A recent paper by Evans (2018) found that convexity problems exist for many types of statistical models (including LASSO, Bayesian networks, and ARIMA models for time series data), as the local geometry of several possible solutions overlap. This means that algorithms can’t distinguish between parameter values or predictor sets within the model space at the sample sizes suggested through power analysis or simulation studies; as the distance between models decreases, the sample needed to distinguish between models increases polynomially.

Two solutions to this issue exist: using nonlinear algebra for optimization or fitting the model using another geometric tool called tangent space. Tangent spaces are a generalization of derivatives at a given point in space (see below for a simple example of tangent space for a point on a sphere). They are linear spaces, and maps between a point’s neighborhood and that point’s tangent space exist, allowing geometers to develop local metrics in the high-dimensional, curved space that are common in point clouds. Even if parameter values or predictor sets overlap within the model space geometry, they typically are distinguishable in the tangent space. One nice machine learning algorithm that fits models using tangent space calculations is dgLARS, which will be detailed in Part 2.



Another intriguing connection between machine learning and algebraic geometry involves matrix completion algorithms for missing data (Ongie et al., 2017), in which a kernel based on the nonlinear structures of the observed data is derived. Creating an algebraic variety through unions and intersections of extant nonlinear data structures provides a rigid framework in which to constrain and compute missing values. The method performs quite favorably relative to state-of-the-art matrix completion algorithms like low-rank matrix completion algorithms and non-convex Schatten minimization. Though the algorithm does not have an open-source package in R or Python yet, connections to the Bertini and Macauley software used in numerical algebraic geometry exist for both R and Python, allowing users to implement the algorithm and experiment with related methods.

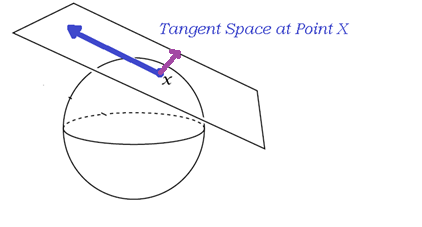
Another recent machine learning advance comes from a subfield of algebraic geometry called Hodge theory, which deals with geometric objects built from algebraic equations and ways in which to decompose them. Jiang et al. (2008) leveraged a common decomposition method, called the Hodge-Helmholtz equation (which partitions these types of objects into curl, gradient, and harmonic pieces—see below), to generalize the PageRank algorithm for paired-ranking problems; the math boils down to a simple least squares problem. This allows for an assessment of both global ranking and local ranking to check for consistency. Many real-world ranking problems involve ties or near-ties among items, which can be problematic in production, and HodgeRank can at least flag items that might not have reliable ranking. In addition, the algorithm can handle a lot of missing data intrinsically.



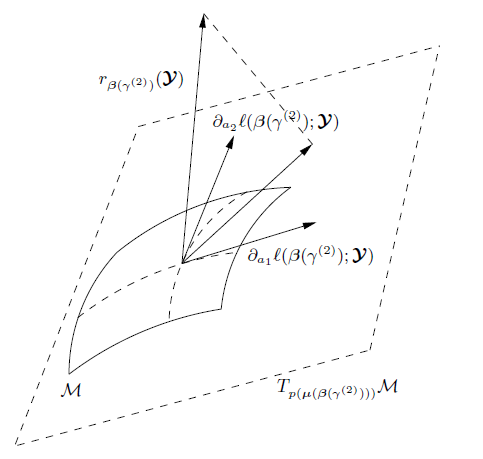
**Part 2: Differential Geometry**

Many types of data are sampled from curved spaces (images, 3-dimensional object models, high-dimensional data that may live a lower-dimensional subspace that is nonlinear…), and researchers focusing on these nonlinear data spaces have made many algorithmic and theoretical contributions to machine learning. In fact, algorithms like linear regression, boosted regression, and various clustering methods have been extended to certain types of data spaces (Riemannian manifolds) common in image/video analytics, which are included in the reference section.

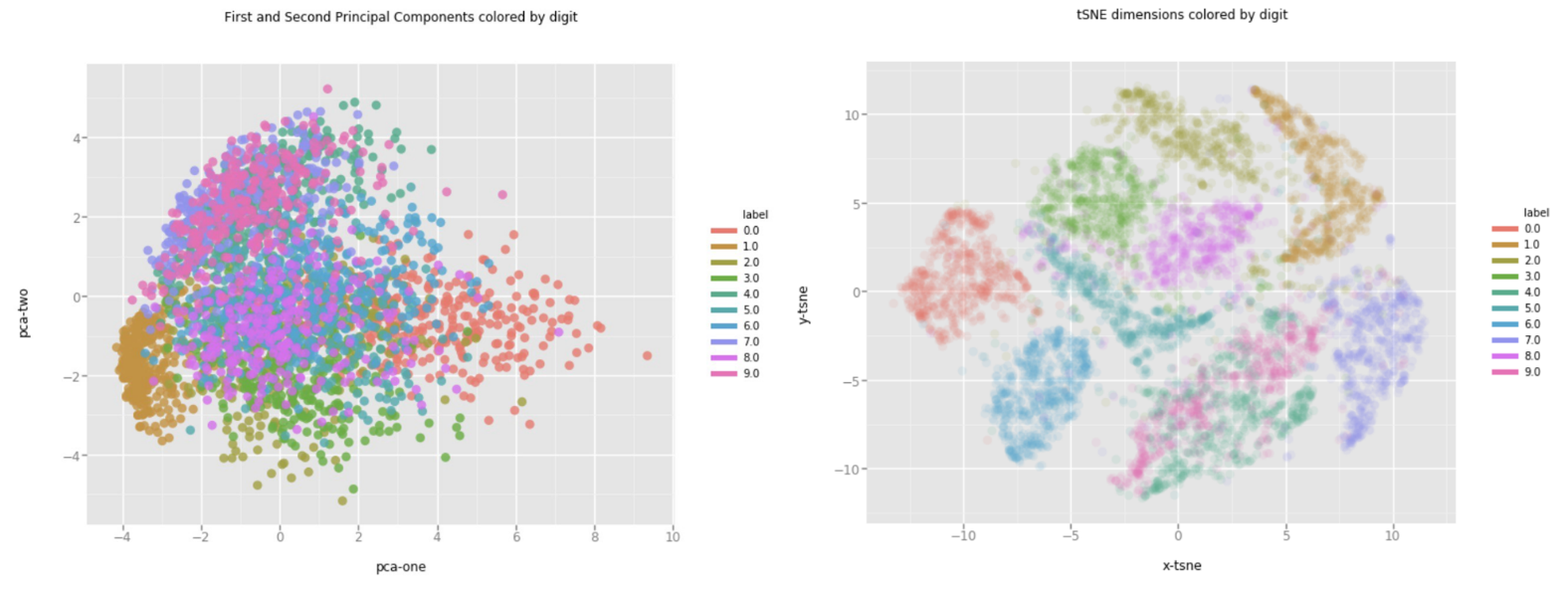
One key idea in differential geometry is the generalization of derivatives from basic calculus to tangent spaces, which were overviewed in Part 1. Briefly, tangent spaces are linear and tangent to a given point in all directions (see below), and maps between a point’s neighborhood and its tangent space exist, allowing geometers to develop local metrics (called geodesics) in high-dimensional, curved spaces.



One algorithm based on model tangent spaces is dgLARS, an adaptation of the least-angle regression (LARS) framework. Originally, dgLARS was designed to handle the problem of small sample sizes and correlated predictors often found in microarray and genome-wide association studies (Augugliaro, Mineo, & Wit, 2009). This algorithm searches tangent spaces of predictors in different point neighborhoods and compares their local tangent space to the geometry of the data manifold, obtaining angles between the two (see below, where variable a1 is selected for the model over a2 based on relative angles). As dgLARS iterates over local point sets and sets of predictors, it carves out a solution curve based on the Fisher information of the generalized linear model’s geometry relative to the tangent space, adding predictors according to the ranking of their relative angles. This allows the algorithm to group predictors with similar geometry (collinear predictors) and select the best predictor in a given set, as well as produce model fit statistics familiar to those who use generalized linear models: deviance, Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC)… Several studies have shown favorable results compared to GLMs, LARS, and other penalized regression models for genetic datasets, other medical datasets, and educational datasets. The R package dglars offers an easy-to-use set of functions to fit generalized linear models from the exponential family of distributions (ex. logistic regression, Poisson regression, gamma regression…) and derive important model fit statistics.



Another useful class of differential-geometry-based methods includes manifold learning for dimensionality reduction. Principle component analysis (PCA) is a dimensionality reduction methods that assumes high-dimensional data lives on lower-dimensional linear spaces, to which the high-dimensional data can be mapped. Manifold learning algorithms address the question of what to do if data naturally lie in a curved space rather than the linear one assumed by PCA. Sets of tools like Laplacian eigenmaps (LE), t-Distributed Stochastic Neighbor Embedding (t-SNE), and multidimensional scaling (MDS) aim to map high-dimensional data to lower-dimensional spaces without making the assumption of subspace linearity. This allows the algorithms to explore nonlinear spaces, as well as linear spaces. One such dataset that seems to lie on a nonlinear space rather than a linear one is the NMIST dataset, as demonstrated in Luuk Derksen’s Python tutorial (results shown below; note markedly better separation of digits, which are color-coded, in the t-SNE plot compared to the PCA plot). Some algorithms prioritize the local geometry of the data when fitting the map (such as LE), while others consider the global geometry/topology of the data space (such as MDS). One recent study suggests that combining multiple dimensionality reduction strategies (particularly local and global methods) can improve predictive modeling results while reducing dimensionality (Farrelly, 2017), and manifold learning is an active area of machine learning research.



Network analytics has also benefited from differential geometry. One recently-developed tool is Forman-Ricci curvature (Weber, Jost, & Saucan, 2016), which measures the amount of “stuff” weighing down and spreading out each edge in the network, signaling network growth/shrinkage and importance of that edge to the overall shape of the network (with highly-curved edges serving as part of the network’s underlying skeleton). Understanding key pieces underlying a network’s structure allows researchers to identify high-value targets to disrupt a terrorist communication networks or pinpoint sets of possible drug targets within a gene expression network. In addition, this new tool can simplify visualizations of large networks by identifying and plotting the network’s underlying skeleton (“backbone”), capturing the essential pieces of the network without cluttering the plot with less important vertices or edges. Many other tools related to geometric curvature (Ricci flow, Laplacians…) can be defined and used through Forman-Ricci curvature on graphs to understand dynamic network behavior, remove noise on a given network through a method akin to regularization, partition a given network into major connected components, and data mine for interesting patterns. It’s likely that new tools will grow from this initial set.

One intriguing application of flows in data analysis involves mapping images to a common space for more accurate comparison of images. In medical imaging, variance of landmark position across subjects often presents issues for statistical modeling and inference across subjects or comparison of diseased/non-diseased samples, and many image standardization/pre-processing methods distort important structures in the image, adding substantial noise. Wang et al. (2012) explored the use of Ricci flow to map brain images to a disc of predefined curvature, such that image landmarks (which can vary in location by subject and image position) could be mapped to a specific point and angles between landmarks could be preserved, minimizing noise introduction from pre-processing steps. Results from this method and subsequent methods with another type of flow (Yamabe flow) are robust to noise and variance among participant images. The image below shows a 3-D brain image with 4 landmarks and the disc created by the Ricci-flow-based map between spaces.



**Conclusion**

It is an exciting time for geometry and its machine learning applications. As computing power increases and quantum computing becomes a reality, it is likely that geometry will play a larger role in machine learning and data science. Familiar algorithms may switch to nonlinear-algebra-based solvers, offering greater precision and flexibility for use on real-world data. The language of algebraic varieties might even become commonplace in model selection, multi-objective optimization problems, and missing data imputation. Differential geometry offers a wide range of tools for exploring nonlinear spaces like networks, imaging data, and video data, and as more data is collected in this form, it is likely new tools will emerge to wrangle and model that data. Both are useful toolsets in data science today.

**References:**

Abbena, E., Salamon, S., & Gray, A. (2017). *Modern differential geometry of curves and surfaces with Mathematica*. Chapman and Hall/CRC. ISBN: 9781420010312

Augugliaro, L., Mineo, A. M., & Wit, E. C. (2009). A differential geometric approach to identify important variables in GLMs when p>> n. *Statistics and the Life Sciences: High-dimensional inference and complex data, University of Groningen, Groningen, September*, 9-11. DOI: 10.1111/rssb.12000

Desbrun, M., Hirani, A. N., Leok, M., & Marsden, J. E. (2005). Discrete exterior calculus. *arXiv preprint math/0508341*.

Dufresne, E., Edwards, P. B., Harrington, H. A., & Hauenstein, J. D. (2018). Sampling real algebraic varieties for topological data analysis. *arXiv preprint arXiv:1802.07716*.

Evans, R. J. (2018). Model selection and local geometry. *arXiv preprint arXiv:1801.08364*.

Farrelly, C. M. (2017). Dimensionality Reduction Ensembles. *arXiv preprint arXiv:1710.04484*.

Farrelly, C. M. (2017). Topology and Geometry in Machine Learning for Logistic Regression. DOI: 10.17605/OSF.IO/V8JGK

<https://medium.com/@luckylwk/visualising-high-dimensional-datasets-using-pca-and-t-sne-in-python-8ef87e7915b>

Jiang, X., Lim, L. H., Yao, Y., & Ye, Y. (2008). Statistical ranking and combinatorial Hodge theory. *arXiv preprint arXiv:0811.1067*.

Ongie, G., Willett, R., Nowak, R. D., & Balzano, L. (2017). Algebraic variety models for high-rank matrix completion. *arXiv preprint arXiv:1703.09631*.

Sommese, A. J., Verschelde, J., & Wampler, C. W. (2005). Introduction to numerical algebraic geometry. In *Solving polynomial equations* (pp. 301-337). Springer, Berlin, Heidelberg. DOI: 10.1007/3-540-27357-3\_8

Tosato, D., Farenzena, M., Spera, M., Murino, V., & Cristani, M. (2010, September). Multi-class classification on riemannian manifolds for video surveillance. In *European Conference on Computer Vision* (pp. 378-391). Springer, Berlin, Heidelberg. ISBN:3-642-15551-0 978-3-642-15551-2

Wang, Y., Shi, J., Yin, X., Gu, X., Chan, T. F., Yau, S. T., ... & Thompson, P. M. (2012). Brain surface conformal parameterization with the Ricci flow. *IEEE transactions on medical imaging*, *31*(2), 251-264. DOI: 10.1109/TMI.2011.2168233

Weber, Melanie, Jürgen Jost, and Emil Saucan. "Forman-Ricci flow for change detection in large dynamic data sets." *Axioms* 5, no. 4 (2016): 26. DOI: 10.3390/axioms5040026