Research Statement

Sarath Shekkizhar*

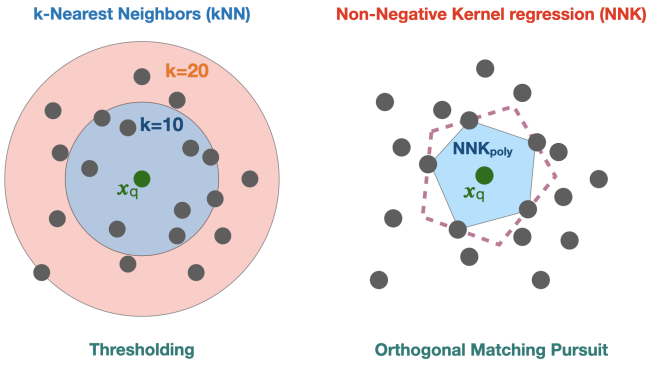
My research is on the algorithmic foundations and applications of data science. Over the last decade, machine learning has achieved remarkable performance in many application domains. However, the research on data itself, particularly the role of its structure in learning and generalization in the real world, is still in its early stages. To this end, my research focuses on fundamental problems in algorithms used for representing data and in the understanding of machine learning algorithms based on the geometrical aspects of the data.

My long-term research goal is to develop simple algorithms that can be applied across domains and can help illuminate the interactions between data, learning models, and humans. Achieving this goal is not an easy task, and practical issues can (and do) expose inadequacies in conventional frameworks. However, I believe that these shortcomings present new research opportunities, and solving these problems has the potential to have a significant impact. Over the past few years, I have organized and collaborated on various research efforts aimed at addressing such challenges, which are discussed below.

Dissertation work: Local neighborhoods and Graphs – a sparse signal approximation view

A critical step in the understanding and analysis of data is the construction of graphs to represent data. Graphs can generalize across different data types and capture complex relationships within a dataset that are non-obvious by only looking at the data. A popular approach to construct graphs, where one is not known *a priori*, is to consider local neighborhoods at each data such as k-nearest neighbor or ϵ-neighborhood to build the graph. However, the choice of k and ϵ in these methods is often ad-hoc and lacks a clear interpretation. Further, a shortcoming common to all existing approaches is their limited geometrical interpretation: they only consider the similarity (or distance) between data and its candidate neighbors, and ignore the relative position of the neighbors themselves.

A key contribution in my dissertation is a novel interpretation of neighborhood definition as a non-negative sparse approximation [1, 2]. This view of neighborhoods allows one to analyze the problem using tools from sparse signal processing and opens up several directions for further research. In particular, I show that kNN and ϵ-neighborhood are equivalent to a signal approximation based on thresholding, with their hyperparameters, k or ϵ, used to control the level of sparsity. This perspective allowed me to (i) establish a new notion of optimality for neighborhoods and (ii) an improved neighborhood and graph definition, non-negative kernel regression (**NNK**), motivated by the well-known limitations of thresholding. Moreover, I show that when applied to data problems the orthogonality in approximation corresponds to the selection of points that are not "geometrically redundant". The key differences between kNN and NNK neighborhoods are shown in figure.



More generally, this study has allowed me to revisit several graph signal processing and machine learning applications such as (1) images [3], where I showed that one can obtain better image-graph representations with up to 90% fewer edges. This work was recognized as *Best Student paper* at IEEE International Conference of Image Processing (ICIP), 2020; (2) non-parametric estimation where I show that NNK representation leads to robust and better interpolation in several learning tasks [4, 5, 6, 7]; (3) a k-means like summarization via geometric dictionary learning where each data is adaptively associated to nearest neighbors [8]; and (4) Multi-scale neighborhoods definitions for graph representations of data [9].

*Electrical and Computer Engineering, University of Southern California. Email: shekkizh@usc.edu. Web: shekkizh.github.io

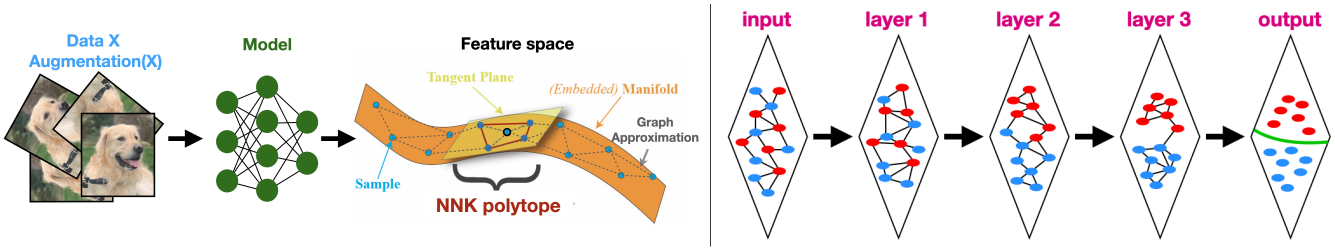


Figure 1: **Left:** Data-driven view of geometry of the embedding manifold using graph. Because neighborhoods and graphs are intrinsically independent of the exact data position, we can compare observations that are fundamentally heterogeneous (such as representations from different dimensional spaces or models). **Right:** Progressive transformation of input space over successive layers of a DNN. The samples in the dataset are the same, and thus their attributes (e.g., labels) are the same, but their position in feature space, and hence the graph constructed, changes as the model is optimized for a particular task.

Ongoing work: A graph-based empirical framework for understanding deep learning

The overarching goal of this research pursuit is to develop insights into the geometry of input-output mappings learned by deep neural networks (DNN). Currently, I focus on addressing three questions in this area:

1) Trust: Can we trust a DNN model whose decisions and parameters are not robust to perturbations? [4, 10]

2) Explainability: How can we explain a particular prediction by a model? Is it biased? [5, 11]

3) Transfer: Can we determine if a model trained with one dataset or task be used for another? [12, 13]

Although these questions are well-known, the lack of explicit model for the data, the common denominator used to build the DNNs, makes it difficult to compare and distinguish DNNs trained with the same input data. For example, recent architectures and optimizations make it possible to achieve zero error on the training data. However, not all of these overparameterized models generalize equally. These differences arise from complex design choices that are hard to detect with current performance metrics. This raises concerns about model *reliability* and *trustworthiness*, especially for deployment in critical applications.

To address this gap, I have been developing **geometric tools** for the understanding of DNNs by characterizing the local and global geometry of the data manifolds using neighborhoods and graphs (Figure 1). The motivation for this approach is the observation that while DNNs involve complex non-linear mappings, the induced transformations and the structure of the representation space can be inferred by observing the relative positions of data in that space. Importantly, this approach enables us to compare the space surrounding a given data point (regardless of its modality) in the embedded space corresponding to different layers of the DNN and at any stage in the system (training, inference, or transfer). Moreover, this approach can be extended to compare the representations of the same data point given by different models.

The concept of a manifold has been widely utilized in machine learning. However, the current state-of-the-art methods for DNNs, which depend on high-dimensional data representations, have made it increasingly difficult to determine whether the common assumption that data belongs to a manifold does, in fact, lead to valid insights. Three main challenges arise: (i) developing computationally efficient metrics to quantify local manifold structure, (ii) verifying the reliability of these metrics for very large datasets in high-dimensional space, and (iii) incorporating knowledge about the structure of complex feature extractors, such as layers, channels, or multi-attention structures in DNNs, into these metrics and their computation.

The development of NNK graph construction as part of my dissertation work has provided essential insights and tools for addressing these challenges, and it forms the cornerstone of this research endeavor. Compared to kNN, NNK graphs are also computed locally and have only a modest increase in complexity. However, unlike kNN, whose neighbors are determined by choice of k, NNK selects neighbors that span the local tangent space.

I am currently exploring the use of **Manifold Graph Metrics (MGMs)** to characterize the geometry of deep learning models. These metrics are derived from the set of NNK neighbors and provide information about the structure of the embedding space, such as the intrinsic dimension and stability of representations for different augmentations. Other MGMs are derived from local tangent spaces estimated from vectors connecting data points and their NNK neighbors, such as the orientation and curvature. By comparing these metrics across points in the dataset, we can assess the homogeneity of the data space. These MGMs enable us to understand the structure of the data and its variation across layers, training epochs, and channels in a quantitative manner.

Research motivation: The quest for principles of representation learning

I tend to look for *similarities* among different types of data and explore ways to connect them. I often wonder if a learning principle that works well for analyzing one type of signal, such as images, can be applied to other modalities. This curiosity is a key motivation for me to pursue research in neighborhood and graph methods.

Recently, there have been significant advancements in representation learning through supervised and self-supervised approaches. However, the nature of a *learned* representation and the process behind its formation remains unclear. Several studies have enhanced our understanding of learning systems and have made important contributions, often building upon one another and adopting diverse perspectives. These theories help make progress in constructing simple, interpretable, and intelligent systems from data. While we do not yet have a complete view of the principles underlying learning, significant components are beginning to take shape.

In my research, I bring together three such components for representation learning: non-negative sparse coding, dictionary learning, and local neighborhoods. These pieces form the foundational pillars of my dissertation work. In neighborhood definitions, a fixed dictionary (the dataset) is used to perform non-negative sparse approximation. However, by learning a dictionary (which is a summary of the training dataset), we can approximate the input data space using polytope patches whose vertices are the dictionary atoms. This view of representation learning aligns with observations in deep learning models, where each ReLU activated layer encodes a patch in the input, and drives my current research views for building better learning models.

References

[1] S. Shekkizhar and A. Ortega, "Graph construction from data by non-negative kernel regression," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2020.

[2] S. Shekkizhar and A. Ortega, "Neighborhood and Graph constructions using non-negative kernel regression (NNK)," *arXiv:1910.09383*, 2019.

[3] S. Shekkizhar and A. Ortega, "Efficient graph construction for image representation," in *IEEE International Conference on Image Processing (ICIP)*, IEEE, 2020.

[4] S. Shekkizhar and A. Ortega, "DeepNNK: Explaining deep models and their generalization using polytope interpolation," *arXiv:2007.10505*, 2020.

[5] S. Shekkizhar and A. Ortega, "Model selection and explainability in neural networks using a polytope interpolation framework," in *55th Asilomar Conference on Signals, Systems, and Computers*, IEEE, 2021.

[6] S. Shekkizhar and A. Ortega, "Revisiting local neighborhood methods in machine learning," in *Data Science and Learning Workshop (DSLW)*, IEEE, 2021.

[7] K. Nonaka, S. Shekkizhar, and A. Ortega, "Graph-based deep learning analysis and instance selection," in *IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*, IEEE, 2020.

[8] S. Shekkizhar and A. Ortega, "NNK-Means: Data summarization using dictionary learning with non-negative kernel regression," in *30th European Signal Processing Conference (EUSIPCO)*, IEEE, 2022.

[9] C. Hurtado, S. Shekkizhar, J. Ruiz-Hidalgo, and A. Ortega, "Study of manifold geometry using multiscale non-negative kernel graphs," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[10] D. Bonet, A. Ortega, J. Ruiz-Hidalgo, and S. Shekkizhar, "Channel-wise early stopping without a validation set via nnk polytope interpolation," 2021.

[11] D. Bonet, A. Ortega, J. Ruiz-Hidalgo, and S. Shekkizhar, "Channel redundancy and overlap in convolutional neural networks with channel-wise nnk graphs," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2022.

[12] R. Cosentino, S. Shekkizhar, M. Soltanolkotabi, S. Avestimehr, and A. Ortega, "The geometry of self-supervised learning models and its impact on transfer learning," *arXiv:2209.08622*, 2022.

[13] P. Das, S. Shekkizhar, and A. Ortega, "Towards a geometric understanding of spatio temporal graph convolution networks," 2022.