STUDY OF MANIFOLD GEOMETRY USING MULTISCALE NON-NEGATIVE KERNEL GRAPHS

Carlos Hurtado[*], *Sarath Shekkizhar*[†], *Javier Ruiz-Hidalgo*[*], *Antonio Ortega*[†]

[*]Universitat Politècnica de Catalunya, Barcelona, Spain
[†]University of Southern California, Los Angeles, USA

ABSTRACT

Modern machine learning systems are increasingly trained on large amounts of data embedded in high-dimensional spaces. Often this is done without analyzing the structure of the dataset. In this work, we propose a framework to study the geometric structure of the data. We make use of our recently introduced non-negative kernel (NNK) regression graphs to estimate the point density, intrinsic dimension, and linearity of the data manifold (curvature). We further generalize the graph construction and geometric estimation to multiple scales by iteratively merging neighborhoods in the input data. Our experiments demonstrate the effectiveness of our proposed approach over other baselines in estimating the local geometry of the data manifolds on synthetic and real datasets.

Index Terms— Local neighborhoods, Manifold geometry, Multi-scale graphs, Intrinsic dimension, Curvature

1. INTRODUCTION

The geometry of a dataset can be summarized using properties such as point density, curvature, and intrinsic dimensionality (ID). The ID of a dataset refers to the minimum number of parameters required for its characterization while maintaining its structure [1]. Approaches for ID estimation [2] often rely on the construction of similarity-based graphs such as K-nearest neighbor (KNN) or ε-neighborhood graphs (ε-graphs). However, the choice of these "neighborhood parameters" (K/ε) is generally ad hoc, which can severely affect the estimation of ID and other geometric properties of the data [3, 4, 5]. Furthermore, these similarity-based graph methods define the scale at which the geometry is estimated through the choice of the same neighborhood parameters (K/ε). As a consequence, the only way to analyze the data at different scales is by increasing the number of neighbors connected to a given query point.

In this paper, we propose new local methods for studying the geometrical properties of manifolds, using novel metrics we have developed from local data neighborhoods defined with the non-negative kernel (NNK) regression graphs [6, 7]. An NNK graph is built by first selecting an initial neighborhood, e.g., a KNN graph, and then using optimization to eliminate connections to geometrically redundant neighbors. NNK graphs are more robust to the initial neighborhood definition (e.g., KNN graphs with different K choices can lead to the same NNK graph). More importantly, the number of NNK neighbors is explicitly dependent on the local geometry of the data and results in a local polytope around any query point.

We propose new metrics derived from NNK graphs to gain insights into three aspects of the geometry of a manifold. First, we study local manifold properties directly derived from NNK neighbors. The NNK optimization implies that the number of points selected in a neighborhood and the size of the local polytope depend on the local geometry of the data. This is in contrast to the measures obtained from a KNN or ε-graph where the number of neighbors or the diameter will depend only on the threshold parameters.

Second, we propose local linear subspace estimation through low-rank approximation of similarity-based graphs via principal component analysis (PCA) on the features of the points in each neighborhood. These lower-dimensional projections are associated with the tangent plane to the manifold's surface, and their dimension has been used [8, 9, 10, 11] to estimate ID. For nonlinear manifolds, the low-rank approximations of similarity-based graphs have been shown [3, 4, 5] to depend heavily on the definition of the local neighborhoods, and therefore, on the choice of the threshold parameter (K/ε). In the NNK neighborhood optimization, only one point in each direction will be selected, and, while only locally relevant directions will be chosen, stronger directions will not be reinforced. This way, while NNK is more robust to nonlinearities in the data representation space, KNN projections will be more robust when there is linearity. This way, the change in KNN projections as a function of scale can be useful in assessing linearity, while NNK low-rank approximations will provide more reliable estimates for the local tangent plane to the manifold.

Finally, we propose a geometric analysis at multiple scales. Different approaches have been proposed to analyze manifolds at multiple scales. [10, 11] build on the technique of applying PCA locally by taking a multiscale approach in the KNN graph construction. This approach, however, relies on choosing an appropriate range of values for $K \geq$ ID, where the value is small enough that the manifold is linear and large enough to mitigate noise in the data. Thus, [10, 11] are highly sensitive to the density and distribution of points in the manifold. An alternative approach followed by TwoNN [12] is to work on smaller subsets of the initial dataset, generated by random sampling of data points. The estimates are then aggregated, under the assumption that the errors that arise from the sampling will average to zero for a large enough number of subsets. In practice, ID estimates based on random sampling have high variance on sparse manifolds and do not account for changes in the local manifold structure.

We propose an alternative approach to dataset sampling in which the points in the dataset are merged iteratively based on the neighborhood defined at the current scale. This step is repeated until a dataset of the desired size is obtained. Subsets with different geometrical properties can be achieved based on the choice of similarity metric and neighborhood definition. For example, when using a Euclidean distance-based KNN similarity graph, the closest points in space will be selected and denser areas will be merged faster. On the contrary, when using the distance-based similarity but with NNK similarity graphs we can preserve the geometry of the initial data and maintain

areas of different density in the resulting sampled datasets.

In summary, we propose a framework to study the local geometry of data using the properties of NNK graphs. We demonstrate via experiments: (i) ID estimation using NNK is in line with the state-of-the-art methods, (ii) linearity of data manifolds using KNN and NNK graphs, and (iii) the impact of neighborhood choice in merging examples for scale. Practical applications of some of the metrics (at one scale) presented here are studied in the context of transfer performance of pre-trained neural networks in [13].

2. NON-NEGATIVE KERNEL (NNK) REGRESSION GRAPHS

A positive definite kernel $k(\boldsymbol{x_i}, \boldsymbol{x_j})$ corresponds to a transformation of points in $\mathbb{R}^d$ to points in a Hilbert space $\mathcal{H}$, such that similarities can be interpreted as dot products in this transformed space (generally referred to as *Kernel Trick*). This way, $k(\boldsymbol{x_i}, \boldsymbol{x_j}) = \phi_i^T \phi_j$, where $\phi : \mathbb{R}^d \to \mathcal{H}$ and ϕ_i represents the transformed observation $\boldsymbol{x_i}$. A popular kernel based on the distance between points that has this property is the Gaussian kernel,

$$k(\boldsymbol{x_i}, \boldsymbol{x_j}) = \exp\left(-\frac{\|\boldsymbol{x_i} - \boldsymbol{x_j}\|^2}{2\sigma^2}\right), \qquad (1)$$

where σ corresponds to the bandwidth parameter of the kernel.

A KNN (or ε-graph) can be constructed by choosing the K largest inner products $\phi_i^T \phi_j$ (or those above a threshold ε). Therefore, these approaches are analogous to a sparse approximation of ϕ_i using a thresholding approach.

In contrast, an NNK [7] graph corresponds to an improved strategy for representation using basis pursuit. Starting from an initial KNN or ε-neighborhood $\mathcal{S}$, the NNK neighborhood at each node is obtained by solving

$$\boldsymbol{\theta}_\mathcal{S} = \min_{\boldsymbol{\theta}:\boldsymbol{\theta} \geq 0} \|\phi_i - \boldsymbol{\Phi}_\mathcal{S}\boldsymbol{\theta}\|_2^2, \qquad (2)$$

where $\boldsymbol{\theta}_\mathcal{S}$ corresponds to the weights of neighbors ($\boldsymbol{\Phi}_\mathcal{S}$) used to approximate ϕ_i. Using the *Kernel Trick*, the objective in (2) can be rewritten as:

$$\boldsymbol{\theta}_\mathcal{S} = \operatorname*{argmin}_{\boldsymbol{\theta}:\boldsymbol{\theta} \geq 0} \frac{1}{2}\boldsymbol{\theta}^T \boldsymbol{K}_{\mathcal{S},\mathcal{S}}\boldsymbol{\theta} - \boldsymbol{K}_{\mathcal{S},i}^T\boldsymbol{\theta}, \qquad (3)$$

where $\boldsymbol{K}_{i,j} = k(\boldsymbol{x_i}, \boldsymbol{x_j})$. Thus, the i-th row of the graph adjacency matrix $\boldsymbol{W}$ is given by $\boldsymbol{W}_{i,\mathcal{S}} = \boldsymbol{\theta}_\mathcal{S}$ and $\boldsymbol{W}_{i,\mathcal{S}^c} = 0$.

NNK performs a selection similar to the orthogonal step in orthogonal matching pursuits [14] which makes NNK robust to the choice of sparsity parameters in the initialization (i.e., K in KNN). Additionally, the resulting graph has a geometric interpretation where each edge in an NNK graph corresponds to a hyperplane with normal in the edge direction, points beyond which are not connected (edge weight zero) [6, 7].

NNK has been shown to perform well in several machine learning tasks [15], image representation [16], and generalization estimation in neural networks [17]. Furthermore, NNK has also been used to understand convolutional neural networks (CNN) channel redundancy [18] and to propose an early stopping criterion for them [19]. Graph properties (not necessarily based on NNK graphs) have been also proposed for the understanding and interpretation of deep neural network performance [20], latent space geometry [21, 22] and to improve model robustness [23]. The specific contribution of this work is to explore the effectiveness of NNK graphs in providing insights into the local geometry of the data, which can be useful in

understanding the properties and structure of the whole dataset. The metrics we propose are not limited to analyzing features in deep neural networks and can be on any dataset embedded in some space.
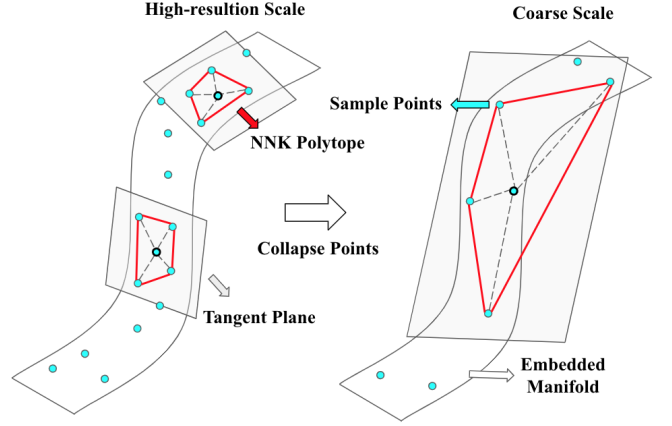
Fig. 1: Our proposed approach to the geometric analysis of data using multi-scale NNK graphs. We assume observed data belongs to a manifold embedded in a higher dimensional space. Properties derived from NNK graphs allow us to capture the local geometry of the data. Changes in the properties of NNK graphs at multiple scales reflect changes in the manifold geometry.

3. MULTI-SCALE ANALYSIS OF NNK GRAPHS

3.1. Local NNK neighborhood data

For the Gaussian kernel, the local geometry of the NNK graph for a given node is a convex polytope around the node. Given a sufficiently large number of initial neighbors, the local connectivity of an NNK graph will be a function of the local dimension of the manifold, as depicted in Fig. 1. We can derive a set of geometrical properties from an NNK graph, and by comparing these properties at different points in a manifold we can gain insight into its geometry.

The number of neighbors in an NNK graph can be insightful, but it can vary locally based on (i) the distribution of the points sampled from a manifold and (ii) the location of the points relative to the geometry of the manifold (e.g., on the edges vs. the middle of the manifold). We can obtain information on the local geometry of the manifold by analyzing other properties of an NNK graph.

The **diameter of an NNK polytope** is defined as the maximum distance between points in the NNK neighborhood around datapoint k:

$$d_k = \max_{i,j \in \mathcal{S}_k} \|\boldsymbol{x_i} - \boldsymbol{x_j}\|, \qquad (4)$$

where $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ are the features of nodes i and j in the NNK neighborhood of k, $\mathcal{S}_k$. Given that NNK will select the nearest point along each direction in space, we can assess the local point density of a manifold using the diameter d_k of polytopes around different points.

3.2. Linear subspace estimation from an NNK graph

An existing approach to ID estimation consists of performing a local parametrization by finding the local tangent plane in the NNK/ε-graph neighborhood of a point and aggregating the estimated ID for each data neighborhood analyzed [8, 9, 10, 11]. Given an appropriate neighborhood, PCA returns the local linear tangent space to
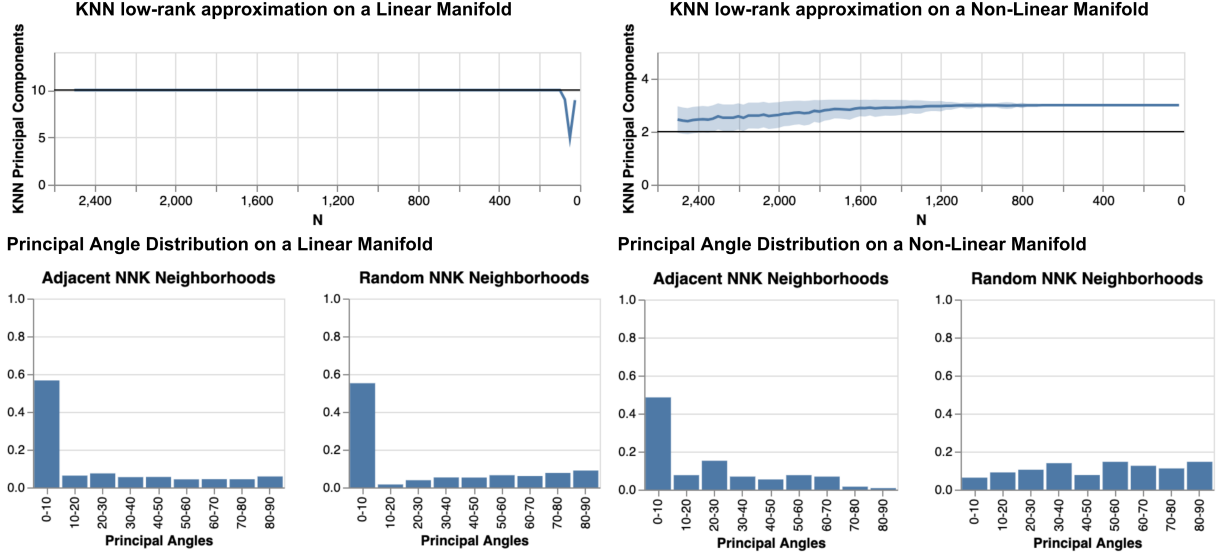
Fig. 2: **Average number of principal components of the low-rank approximation of KNN graphs** as a function of point merging, together with a **black line indicating the datasets' ID**, and distribution of the **principal angles between NNK low-rank approximation subspaces** on a linear manifold and a non-linear manifold. We merge points using Euclidean distance, which, given that the manifolds are uniformly sampled, will result in points being merged at the same rate throughout the manifolds. We show the angles between pairs of adjacent NNK neighborhoods (i.e., one of the center nodes is a neighbor of the other) and the angles for random pairs of NNK neighborhoods. On a linear manifold, the chosen principal components of a KNN graph will be equal to the manifold's ID, and any pair of neighborhoods will have angles close to 0. On a nonlinear manifold, the principal components of a KNN graph will over-estimate the ID, and close neighborhoods have angles close to zero, while random neighborhoods are more dissimilar

the manifold. PCA estimates are robust on linear manifolds, but on nonlinear manifolds, the task of finding an appropriate neighborhood makes these estimators unstable [3, 4, 5]. We can also obtain a low-rank approximation from the NNK neighborhood vector subspace, such that the **number of relevant principal components** would be a robust estimate of the ID of the manifold. The sparsity enforced by the NNK selection makes NNK graphs less sensitive to the instability of the neighborhood in the initial KNN graph. To find the number of significant eigenvalues, we introduce a threshold on the size of the eigenvalues as proposed in [8], namely

$$\lambda_j \geq \frac{1}{10} \cdot \lambda_{max}. \tag{5}$$

In addition, we use the change with scale in the size of the low-rank approximations of KNN graphs to validate our observations on the curvature of a manifold. While the KNN graph's projection will not change on flat manifolds, it will be unstable and overestimate the flatness on highly curved manifolds.

By comparing the **principal angles between the low-rank approximation of NNK neighborhoods** we can better understand the geometry of a manifold. Principal angles [24] refer to the generalization of the concept of angles between lines in the plane to any arbitrary dimension. This way, on a flat manifold the distribution of the angles will be similar in neighborhoods in different positions of the manifold, and many of these principal angles will be close to zero. In contrast, on a highly curved manifold, the distribution of the angles between NNK subspaces will change at different regions in the manifold, and the angles will be larger. Moreover, on locally smooth manifolds, we would expect the majority of the angles between the low-rank approximation of adjacent NNK neighborhoods (i.e., neighborhoods of points where one is in the NNK neighborhood of the other) to be close to zero.

3.3. NNK graph construction at multiple scales

To study the manifold at multiple scales, we must increase the size of the NNK neighborhood. Intuitively, we would do so by adjusting the hyperparameters of the NNK algorithm to observe points that are further away. However, by definition, NNK will choose a single neighbor in each relevant direction, therefore increasing the sparsity parameter K in the initial KNN graph is unlikely to affect the size of the resulting NNK neighborhood after optimization. On the other hand, increasing the bandwidth σ of the Gaussian kernel will make points that are very close to each other collapse (similarity value of 1), in which case the NNK optimization will only select one point for the neighborhood. To overcome this issue, we propose an approach that simultaneously increases σ while merging the closest data points in the manifold. Thus, we change the scale of the analysis by subsampling the manifold in a controlled way, increasing the distances between points in the process, while also increasing σ to allow farther away points to be connected.

We would expect a linear manifold to have similar properties at different scales. However, merging on a highly curved manifold would result in connecting points that were initially in different local neighborhoods, thus changing the shape of the manifold and leading to changes on the NNK graphs at coarser scales (see Fig. 1). Thus, studying our proposed properties at different scales can lead to a better understanding of the shape of the manifold.

To achieve a sparser representation we iteratively merge the two closest points according to some selected metric. For example, we can merge the two nodes with the shortest pairwise distance (largest KNN graph weight), and the nodes with the largest NNK pairwise weight, or use alternative metrics. At each step, we combine the points that are closest so that we can increase the window of observation. After each merging iteration, we can recalculate the NNK

graph. Since the decay parameter σ has been defined based on the distances in the dataset, we will in turn increase the size of the NNK graph as we merge points. This way, we can construct larger NNK graphs and thus be able to analyze the manifold at different scales. This merged dataset can be achieved as described in Algorithm 1.

Algorithm 1 Two-closest Merging

Input: X features, I merging steps

1: **for** iter in I **do**
2: **for** each node $i = 1, 2, ..., N$ **do**
3:    $S_i = \{\text{neighborhood of node } i\}$
4:    $K_{i,S} = \{\text{similarity to each neighbor}\}$
5: **end for**
6: $i, j = \{i, j : \max_{i,j} K_{i,j}\}$
7: $X = X \cup \frac{X_i + X_j}{2} \setminus X_i, X_j$
8: **end for**

Output: Dataset X after I merging steps

---

4. EXPERIMENTS

4.1. ID estimation benchmark

| d	MLE	kNNG	BPCA	Hein	CD	DANCo	MLSVD	NNK
1	**1.00**	1.07	5.70	**1.00**	1.14	**1.00**	**1.00**	**1.00**
2	2.21	2.03	1.55	**2.00**	2.19	**2.00**	1.00	1.00
2	1.97	2.06	**2.00**	**2.00**	1.98	**2.00**	**2.00**	**2.00**
2	1.96	2.09	**2.00**	**2.00**	1.93	**2.00**	2.35	**2.00**
3	2.88	3.03	**3.00**	**3.00**	2.88	**3.00**	**3.00**	**3.00**
4	3.83	3.82	**4.00**	**4.00**	3.23	**4.00**	2.08	**4.00**
4	3.95	4.76	4.25	**4.00**	3.88	**4.00**	8.00	**4.00**
6	6.39	11.24	12.00	**5.95**	5.91	7.00	12.00	8.00
10	8.26	10.21	5.20	8.90	8.09	9.86	**10.00**	**10.00**
10	9.10	9.98	5.45	9.45	9.12	10.09	**10.00**	**10.00**
3	4.05	4.32	4.00	**3.00**	3.37	4.00	1.00	**3.00**
8-11	**10.29**	9.58	11.00	8.00	6.96	9.98	1.00	10.00

Table 1: ID estimation on synthetic datasets [2] (first 10 rows), Isomap [25], and MNIST [26] (last 2 rows), using 7 state-of-the-art methods and our proposed NNK approach.

We estimate ID from the number of principal components selected from the NNK graph following equation (5). We compare our proposed method for ID estimation (NNK), with state-of-the-art methods for each of the categories defined in an ID estimator literature review and benchmark proposal [2]. We include BPCA [27] and MLSVD [28], which are projective estimators; the kNNG [29] graph-based estimator; CD [30] and Hein [31] as examples of topological fractal estimators; and MLE [5] and DANCo [32], topological nearest neighbor-based estimators. We show the results on a series of synthetic datasets [31] generated by uniformly drawing samples from manifolds of known ID that are embedded linearly or nonlinearly on higher dimensional spaces. We also show results for the Isomap [25] faces and MNIST [26] datasets.

Table 1 shows that the estimate of ID derived from NNK graphs achieves performance in line with that of other estimators in the literature. While a more complete benchmark would help assess the capabilities and limitations of our NNK estimates, our results show that meaningful information about the local geometry of a data manifold can be derived from NNK neighborhoods.

4.2. Synthetic manifold analysis

We use two of the metrics we previously described to assess the linearity of two synthetic data manifolds. In the first row, we show the size of the low-rank approximations of KNN neighborhoods as a function of the points in the dataset. Below, we plot the distribution of the principal angles between the low-rank approximation of NNK neighborhoods. Fig. 2 shows the distributions obtained by comparing adjacent and random NNK neighborhoods on a linear (left) and nonlinear (right) manifold. For the linear manifold, the distribution of both adjacent and random pairs of neighborhoods are almost the same, since the geometry of a neighborhood in a linear manifold will be similar throughout the manifold. In contrast, on the nonlinear manifold, there is a difference in the distribution of the angles.
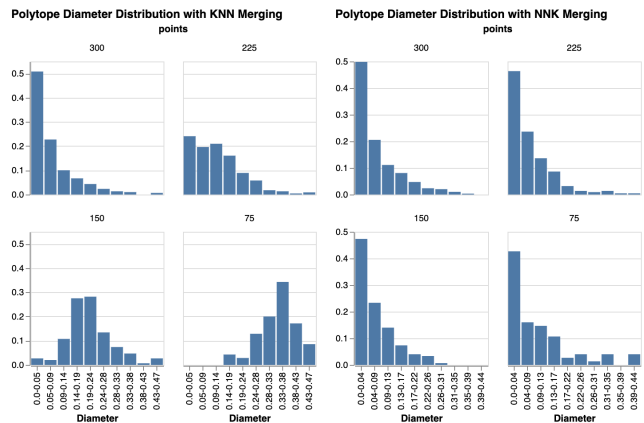
Fig. 3: Distribution of the diameter of NNK graph polytopes constructed on the *control* dataset while merging with KNN or NNK similarity. When merging based on KNN similarity, we observe the polytope diameter distribution shift to larger diameters. When using NNK similarity in the selection of the points to merge, the distribution of the polytope sizes remains the same after merging.

4.3. Choice of neighborhoods for merging

From the distribution of the diameter of the NNK polytopes for the same dataset, in Fig. 3, we observe that when using KNN similarity for merging, the size of the polytopes increases, such that the size distribution shifts to larger polytopes while also growing in size. In contrast, when NNK similarity is used for merging, the distribution of the polytope diameters is preserved as their sizes grow overall. This shows that NNK merging is better if the goal is to preserve the differences in point densities in the multiscale analysis.

5. CONCLUSION

We present a framework based on the geometrical properties of NNK graphs to gain insight into the shape of data manifolds in terms of their intrinsic dimension, curvature, and point density. The proposed metrics are the dimension of the low-rank approximation of the KNN and NNK graphs, the diameter of NNK graphs, and the principal angles between the low-rank approximations of NNK graphs. Moreover, we compare these metrics at multiple scales, which we can do by using our proposed point merging algorithm. Experiments show that we can accurately estimate ID on popular benchmark manifolds. Furthermore, we have shown the effectiveness of NNK properties in characterizing data manifolds.

6. REFERENCES

[1] Robert Bennett, "The intrinsic dimensionality of signal collections," *IEEE Transactions on Information Theory*, vol. 15, no. 5, pp. 517–525, 1969.

[2] Paola Campadelli, Elena Casiraghi, Claudio Ceruti, and Alessandro Rozza, "Intrinsic dimension estimation: Relevant techniques and a benchmark framework," *Mathematical Problems in Engineering*, 2015.

[3] Peter J. Verveer and Robert P. W. Duin, "An evaluation of intrinsic dimensionality estimators," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 17, no. 1, pp. 81–86, 1995.

[4] Balázs Kégl, "Intrinsic dimension estimation using packing numbers," *Advances in neural information processing systems*, vol. 15, 2002.

[5] Elizaveta Levina and Peter Bickel, "Maximum likelihood estimation of intrinsic dimension," *Advances in neural information processing systems*, vol. 17, 2004.

[6] Sarath Shekkizhar and Antonio Ortega, "Graph construction from data using non negative kernel regression (nnk graphs)," *arXiv:1910.09383*, 2019.

[7] Sarath Shekkizhar and Antonio Ortega, "Graph construction from data by non-negative kernel regression," in *Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 3892–3896.

[8] Keinosuke Fukunaga and David R Olsen, "An algorithm for finding intrinsic dimensionality of data," *IEEE Transactions on Computers*, vol. 100, no. 2, pp. 176–183, 1971.

[9] Jörg Bruske and Gerald Sommer, "Intrinsic dimensionality estimation with optimally topology preserving maps," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 5, pp. 572–575, 1998.

[10] Anna V Little, Mauro Maggioni, and Lorenzo Rosasco, "Multiscale geometric methods for estimating intrinsic dimension," *Proc. SampTA*, vol. 4, no. 2, 2011.

[11] Daniel N Kaslovsky and François G Meyer, "Optimal tangent plane recovery from noisy manifold samples," *ArXiv eprints*, 2011.

[12] Elena Facco, Maria d'Errico, Alex Rodriguez, and Alessandro Laio, "Estimating the intrinsic dimension of datasets by a minimal neighborhood information," *Scientific reports*, vol. 7, no. 1, pp. 1–8, 2017.

[13] Romain Cosentino, Sarath Shekkizhar, Mahdi Soltanolkotabi, Salman Avestimehr, and Antonio Ortega, "The geometry of self-supervised learning models and its impact on transfer learning," *arXiv:2209.08622*, 2022.

[14] Joel A Tropp and Anna C Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[15] Sarath Shekkizhar and Antonio Ortega, "Revisiting local neighborhood methods in machine learning," in *Data Science and Learning Workshop (DSLW)*. IEEE, 2021, pp. 1–6.

[16] Sarath Shekkizhar and Antonio Ortega, "Efficient graph construction for image representation," in *Intl. Conf. on Image Processing (ICIP)*. IEEE, 2020, pp. 1956–1960.

[17] Sarath Shekkizhar and Antonio Ortega, "Model selection and explainability in neural networks using a polytope interpolation framework," in *55th Asilomar Conf. on Signals, Systems, and Computers*. IEEE, 2021, pp. 177–181.

[18] David Bonet, Antonio Ortega, Javier Ruiz-Hidalgo, and Sarath Shekkizhar, "Channel redundancy and overlap in convolutional neural networks with channel-wise nnk graphs," in *Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 4328–4332.

[19] David Bonet, Antonio Ortega, Javier Ruiz-Hidalgo, and Sarath Shekkizhar, "Channel-wise early stopping without a validation set via nnk polytope interpolation," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conf. (APSIPA ASC)*. IEEE, 2021, pp. 351–358.

[20] Vincent Gripon, Antonio Ortega, and Benjamin Girault, "An inside look at deep neural networks using graph signal processing," in *Information Theory and Applications Workshop (ITA)*. IEEE, 2018, pp. 1–9.

[21] Carlos Lassance, Myriam Bontonou, Ghouthi Boukli Hacene, Vincent Gripon, Jian Tang, and Antonio Ortega, "Deep geometric knowledge distillation with graphs," in *Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 8484–8488.

[22] Carlos Lassance, Vincent Gripon, and Antonio Ortega, "Laplacian networks: Bounding indicator function smoothness for neural networks robustness," *APSIPA Transactions on Signal and Information Processing*, vol. 10, 2021.

[23] Carlos Lassance, Vincent Gripon, and Antonio Ortega, "Representing deep neural networks latent space geometries with graphs," *Algorithms*, vol. 14, no. 2, pp. 39, 2021.

[24] Camille Jordan, "Essai sur la géométrie à n dimensions," *Bulletin de la Société mathématique de France*, vol. 3, pp. 103–174, 1875.

[25] Joshua B Tenenbaum, Vin de Silva, and John C Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[26] Yann LeCun, "The mnist database of handwritten digits," *http://yann. lecun. com/exdb/mnist/*, 1998.

[27] Christopher Bishop, "Bayesian pca," *Advances in neural information processing systems*, vol. 11, 1998.

[28] Anna V Little, Mauro Maggioni, and Lorenzo Rosasco, "Multiscale geometric methods for data sets i: Multiscale svd, noise and curvature," *Applied and Computational Harmonic Analysis*, vol. 43, no. 3, pp. 504–567, 2017.

[29] Jose A Costa, Abhishek Girotra, and AO Hero, "Estimating local intrinsic dimension with k-nearest neighbor graphs," in *IEEE/SP 13th Workshop on Statistical Signal Processing, 2005*. IEEE, 2005, pp. 417–422.

[30] Peter Grassberger and Itamar Procaccia, "Measuring the strangeness of strange attractors," in *The theory of chaotic attractors*, pp. 170–189. Springer, 2004.

[31] Matthias Hein and Jean-Yves Audibert, "Intrinsic dimensionality estimation of submanifolds in rd," in *Proceedings of the 22nd Intl. Conf. on Machine learning*, 2005, pp. 289–296.

[32] Claudio Ceruti, Simone Bassis, Alessandro Rozza, Gabriele Lombardi, Elena Casiraghi, and Paola Campadelli, "Danco: dimensionality from angle and norm concentration," *arXiv preprint arXiv:1206.3881*, 2012.