ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

Geometrical homotopy for data visualization

Diego H. Peluffo-Ordóñez [1], Juan C. Alvarado-Pérez [2], John A. Lee [3,4], and Michel Verleysen [3] *

1- Eslinga research group, Faculty of Engineering
Universidad Cooperativa de Colombia

2- Universidad de Salamanca

3-Machine Learning Group - ICTEAM,
Université Catholique de Louvain

4-Molecular Imaging Radiotherapy and Oncology - IREC,
Université Catholique de Louvain

Abstract. This work presents an approach allowing for an interactive visualization of dimensionality reduction outcomes, which is based on an extended view of conventional homotopy. The pairwise functional followed from a simple homotopic function can be incorporated within a geometrical framework in order to yield a bi-parametric approach able to combine several kernel matrices. Therefore, the users can establish the mixture of kernels in an intuitive fashion by only varying two parameters. Our approach is tested by using kernel alternatives for conventional methods of spectral dimensionality reduction such as multidimensional scalling, locally linear embedding and laplacian eigenmaps. The proposed mixture represents every single dimensionality reduction approach as well as helps users to find a suitable representation of embedded data.

1 Introduction

Dimensionality reduction (DR) methods are often developed under determined design parameters and pre-established optimization criterion, and therefore they still lack properties such as user interaction and controllability. These properties are characteristic of information visualization procedures. The field of information visualization (IV) is aimed at developing graphical ways of representing data so that information can be more usable and intelligible for the user [1, 2]. Then, one can intuit that DR can be improved by importing some properties of the IV methods.

In this work, with the aim to enable users to provide parameters for data visualization tuning, we propose to combine the effects of different DR approaches through a homotopy approach for kernels. From the pairwise functional of a conventional homotopy function, an extension able to combine more than two kernels is introduced. Such an extension consists of a linear combination where the coefficients are related to the points inside the surface of a polygon. This geometrical approach enables users to visualize every single method as well as *deformations* or combinations of the methods. To facilitate the localization of the polygonal surface, the entire surface can be spanned by varying only two parameters. In other words, kernels are combined through a weighted sum where coefficients are interactively provided by users within a simple bi-parametric framework. Our approach is tested by using kernel alternatives for

*J.A. Lee is a Research Associate with the FRS-FNRS (Belgian National Scientific Research Fund). This work is funded by FRS-FNRS (Belgian National Scientific Research Fund) project 7.0175.13 DRedVis.

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

conventional methods of spectral dimensionality reduction such as multidimensional scaling, locally linear embedding and laplacian eigenmaps [3]. The quality of obtained embedded data is quantified by a scaled version of the average agreement rate between K-ary neighborhoods as described in [4]. Provided mixture represents every single dimensionality reduction approach as well as helps users to find a suitable representation of embedded data.

The outline of this paper is as follows: Proposed geometrical homotopy and its application for kernel-based DR methods is presented in section 2. Section 3 states the experimental setup for this work. Results and discussion are shown in section 4. Finally, section 5 draws the conclusions and final remarks.

2 Geometrical homotopy

In topology, the general concept of homotopy refers to the mapping process of a continuous function onto another one. Such a mapping is done through a continuous deformation of one function into the other [5]. Mathematically, homotopy function for two topological spaces can be defined as follows: Let f_1 and f_2 be two continuous functions associated to the topological spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. A homotopy function can be written as follows:

$$\begin{aligned} h: \mathcal{X} \times [0,1] &\rightarrow \mathcal{Y} \\ f_1, f_2 \quad \lambda \quad &\mapsto h(f_1, f_2, \lambda), \end{aligned} \tag{1}$$

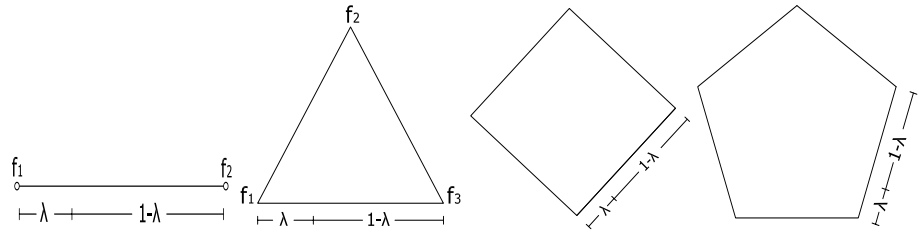
such that $h(f_1, f_2, 0) = f_1$ and $h(f_1, f_2, 1) = f_2$. A simple function fulfilling the homotopy conditions can be expressed in the form $h(f_1, f_2, \lambda) = \lambda f_1 + (1 - \lambda) f_2$. Such form is also used for regularization purposes. Graphically, this homotopy approach can be represented as a line of length 1 drawn between two points so that points represent the two homotopic functions (See Figure 1(a)). Then, within a visualization framework, the homotopy parameter λ can be seen as a slider bar.

2.1 Polygonal approach

In this work, from the conventional homotopy concept relying on a pairwise function, we introduce a simple approach able to combine more than two functions. Since two homotopic functions can be represented as a line, we can intuitively extend the mixture given by a pairwise function to polygons for representing more than two functions. By doing so, each vertex represent a function and the homotopic parameter can span every pair of functions working as a slider on all the edges singly. Similarly as conventional homotopy, the length of edges for polygonal approach is 1. Figures 1(b) to 1(d) depict graphically this idea.

Following the idea of combining functions by varying a parameter, the edges of the polygons can be stretched to result in a 1D representation that allows to represent different functions in a pairwise fashion. In other words, the whole set of functions is spanned by varying λ around the edges of the polygon -rounding the polygon. Nonetheless, since the homotopic function considered here is a binary operator no more than two functions are simultaneously taken into account within the mixture of functions.

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.



(a) $M = 2$ (b) $M = 3$ (c) $M = 4$ (d) $M = 5$

Figure 1: Polygonal approach to apply a homotopy functional on a set of functions. When considering more than two functions, the mixture of functions can be done in a pairwise fashion by varying ("sliding") the homotopy parameter λ.

Let us consider a mixture of M functions given by: $f = \sum_{m=1}^{M} \alpha_m(\lambda) f_m$. So far, by using the first geometrical approach, two coefficients $\alpha_m(\lambda)$ take values λ and $1 - \lambda$, meanwhile the remaining ones are zero. In order to involve the effect of the whole set of functions (at least more than the two ones located at the vertexes of a specific edge), we introduce a geometrical homotopy approach using the entire surface rather than the perimeter only. To this end, every point inside the surface of the polygon represents a set of coefficients. Therefore, a grid or graph representing the inner points is required. We propose the use of roll-like grid as shown in Figure 2. Assuming that λ takes discrete values, proposed grid can be seen as a dashed concentric line in which the resolution is given by a constant term ε. Such resolution establishes the distance between two contiguous parallel lines. According to ε, the resolution level μ_ℓ with $\ell \in 0, \ldots, n$ is set in such a way that μ_1 and μ_n denotes the outer and inner line, respectively. Thus, $\varepsilon = 1/n$, being n the number of levels. Now, the coefficients are bi-parametric $\alpha_m(\lambda, \mu)$ and yield a mixture in the form $f = \sum_{m=1}^{M} \alpha_m(\lambda, \mu) f_m$. Since maximum length of the edges is 1, a necessary condition to satisfy homotopy is that the maximum value of functions is 1, in case of real functions.

To explain our approach, consider a set of M kernel matrices $\{\boldsymbol{K}^{(1)}, \ldots, \boldsymbol{K}^{(M)}\}$ representing different DR methods. The geometrical homotopy is then aimed at accomplishing a mixture in the form $\boldsymbol{K} = \sum_{m=1}^{N} \alpha_m(\lambda, \mu) \boldsymbol{K}^{(m)}$. As explained above, kernel matrices should also be normalized so that the maximum entry is 1.

3 Experimental setup

The kernel resulting from the mixture provided here is tested on kernel PCA as explained in [6]. Kernel PCA, as any dimensional reduction approach, is aimed to embed a high dimensional data matrix $\boldsymbol{Y} \in \mathbb{R}^{D \times N}$ into a low-dimensional, latent data matrix $\boldsymbol{X} \in \mathbb{R}^{d \times N}$, being $d < D$. Then, observed data and latent data matrices are formed by N observations, denoted respectively by $\boldsymbol{y}_i \in \mathbb{R}^D$ and $\boldsymbol{x}_i \in \mathbb{R}^d$, with $i \in \{1, \ldots, N\}$.

Kernels for DR: Three kernel approximations for spectral DR methods [3] are considered. Namely, classical multidimensional scalling (CMDS), locally linear embed-

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

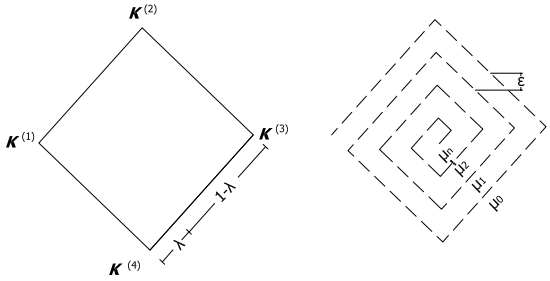


Figure 2: Geometrical homotopy for four kernel functions. This approach enables users to combine kernels by selecting a point inside the surface. Every point is localized by its corresponding pair (λ, μ), which is in turn associated with a set of M coefficients $\{\alpha_m(\lambda,\mu)\}_{m=1}^{M}$.

ding (LLE), and graph Laplacian eigenmaps (LE). CMDS kernel is the double centered distance matrix $\boldsymbol{D} \in \mathbb{R}^{N\times N}$ so

$$\boldsymbol{K}^{(1)} = \boldsymbol{K}_{CMDS} = -\frac{1}{2}(\boldsymbol{I}_N - \mathbf{1}_N\mathbf{1}_N^\top)\boldsymbol{D}(\boldsymbol{I}_N - \mathbf{1}_N\mathbf{1}_N^\top), \tag{2}$$

where the ij entry of $\boldsymbol{D}$ is given by $d_{ij} = \|\boldsymbol{y}_i - \boldsymbol{y}_j\|_2^2$, $\boldsymbol{I}_N$ denotes a N-dimensional identity matrix, and $\mathbf{1}_N$ is a N-dimensional all ones vector.

A kernel for LLE can be approximated from a quadratic form in terms of the matrix $\mathcal{W}$ holding linear coefficients that sum to 1 and optimally reconstruct observed data. Define a matrix $\boldsymbol{M} \in \mathbb{R}^{N\times N}$ as $\boldsymbol{M} = (\boldsymbol{I}_N - \mathcal{W})(\boldsymbol{I}_N - \mathcal{W}^\top)$ and λ_{max} as the largest eigenvalue of $\boldsymbol{M}$. Kernel matrix for LLE is in the form

$$\boldsymbol{K}^{(2)} = \boldsymbol{K}_{LLE} = \lambda_{max}\boldsymbol{I}_N - \boldsymbol{M}. \tag{3}$$

Since kernel PCA is a maximization of the high-dimensional covariance represented by a kernel, LE can be represented as the pseudo-inverse of the graph Laplacian $\boldsymbol{L}$:

$$\boldsymbol{K}^{(3)} = \boldsymbol{K}_{LE} = \boldsymbol{L}^\dagger, \tag{4}$$

where $\boldsymbol{L} = \mathcal{D} - \boldsymbol{S}$, $\boldsymbol{S}$ is a similarity matrix and $\mathcal{D} = \text{Diag}(\boldsymbol{S}\mathbf{1}_N)$ is the degree matrix. All previously mentioned kernels are widely described in [3]. The similarity matrix $\boldsymbol{S}$ is formed in such a way that the relative bandwidth parameter is estimated keeping the entropy over neighbor distribution as roughly $\log K$ where K is the given number of neighbors as explained in [7]. The number of neighbors is established as $K = 30$.

As well, a RBF kernel is also considered: $\boldsymbol{K}^{(4)} = \boldsymbol{K}_{RBF}$ whose ij entry are given by $\exp(-0.5\|\boldsymbol{y}_i - \boldsymbol{y}_j\|/\sigma^2)$ with $\sigma = 0.1$. For all methods, input data is embedded into a 2-dimensional space, then $d = 2$. Then, the homotopy approach is performed considering $M = 4$ kernels.

Database: Experiments are carried out over an artificial spherical shell with $N = 1500$ data points and $D = 3$ dimensions.

Performance measure: To quantify the performance of studied methods, the scaled version of the average agreement rate $R_{NX}(K)$ introduced in [4] is used, which is ranged within the interval $[0, 1]$. Since $R_{NX}(K)$ is calculated at each perplexity value from 2 to

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

$N-1$, a numerical indicator of the overall performance can be obtained by calculating its area under the curve (AUC). The AUC assesses the dimension reduction quality at all scales, with the most appropriate weights.

4 Results and discussion

Given that the mixture presented here is a linear combination, only two kernels (in a pairwise fashion) are evaluated in case of selecting coefficients from the outer perimeter. Doing so user can appreciate the deformation of the resulting embedding from a method onto that from another method by just varying parameter λ. Indeed, when selecting coefficients associated with the vertexes, DR process is perfomed under the effect of a single method. To involve the effect of more than two methods, different resolution levels μ_m should be explored through the polygon surface. So, the proposed homotopy approach enable users (even those no expert) to interact with the DR outcomes by intuitively selecting points inside or just in the boundary of the polygonal grid.

Nonetheless, the mixture of kernels not only allows for representing DR methods but also may improve the quality of resultant embedded data. In fact, the process of combining kernels naturally obtained from sliding the homotopy and resolution parameters may yield better embedding representations when the effects of DR methods are adequately blended. Figure 3 shows an instance where the mixture of kernels with $\lambda = 0.5$, $\epsilon = 0.1$, and μ_3 reaches better performance in terms of the considered quality measure. Such performance is associated with the ability to unfold the spherical shell as seen in Figure 4. Certainly, the mixture of kernels accomplishes a better representation of embedded data, which resembles the 2D sphere development.

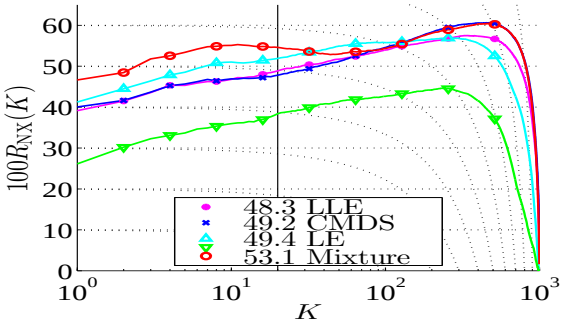
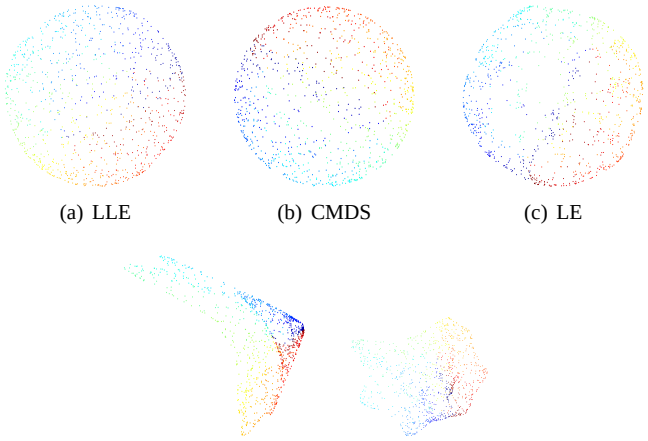


Figure 3: Results are shown regarding the quality measure $R_{NX}(K)$. Next to legend of each curve, its corresponding AUC is shown.

5 Conclusion

This work presents an interactive approach to visualize the embedded data resulting from dimensionality reduction methods. Such approach is based on a so-called geomet-

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.



(a) LLE (b) CMDS (c) LE

(d) RBF kernel (e) Mixture of kernels

Figure 4: Embedded data for each considered method. All the methods are performed by using the kernel representations on kernel PCA. The mixture of kernels is done with the coefficients corresponding to $\lambda = 0.5$, $\epsilon = 0.1$, and μ_3.

rical homotopy, which is aimed to facilitate the selection of a DR method that fulfills the user's needs. Even non-expert users might easily select a method or combination of methods by picking up points from a polygonal surface.

As a future work, more kernel representations as well as different mixtures for homotopy will be explored.

References

[1] Usama M Fayyad, Andreas Wierse, and Georges G Grinstein. *Information visualization in data mining and knowledge discovery*. Morgan Kaufmann, 2002.

[2] Matthew Ward, Georges Grinstein, and Daniel Keim. *Interactive data visualization: foundations, techniques, and applications*. AK Peters, Ltd., 2010.

[3] Jihun Ham, Daniel D Lee, Sebastian Mika, and Bernhard Schölkopf. A kernel view of the dimensionality reduction of manifolds. In *Proceedings of the twenty-first international conference on Machine learning*, page 47. ACM, 2004.

[4] John A Lee, Emilie Renard, Guillaume Bernard, Pierre Dupont, and Michel Verleysen. Type 1 and 2 mixtures of kullback-leibler divergences as cost functions in dimensionality reduction based on similarity preservation. *Neurocomputing*, 2013.

[5] Charles Nash and Siddhartha Sen. *Topology and geometry for physicists*. Courier Dover Publications, 2013.

[6] Diego Peluffo-Ordóñez, John Lee, and Michel Verleysen. Generalized kernel framework for unsupervised spectral methods of dimensionality reduction. In *IEEE Symposium Series on Computational Intelligence*, 2014.

[7] James Cook, Ilya Sutskever, Andriy Mnih, and Geoffrey E Hinton. Visualizing similarity data with a mixture of maps. In *International Conference on Artificial Intelligence and Statistics*, pages 67–74, 2007.