



Generalized Spectral Dimensionality Reduction Based on Kernel Representations and Principal Component Analysis

MacArthur C. Ortega-Bustamante^{1,2}, Waldo Hasperué², Diego H. Peluffo-Ordóñez^{3,4},
Juan González-Vergara^{4(✉)}, Josué Marín-Gaviño^{4,5}, and Martín Velez-Falconi⁴

¹ Universidad Técnica Del Norte, Ibarra, Ecuador

² III-LIDI, Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina

³ Modeling, Simulation and Data Analysis (MSDA) Research Program,
Mohammed VI Polytechnic University, 47963 Ben Guerir, Morocco

⁴ SDAS Research Group, 47963 Ben Guerir, Morocco

juan.gonzalez@sdas-group.com

⁵ School of Mathematical and Computational Science, Yachay Tech, Urcuquí, Ecuador
<https://www.sdas-group.com>

Abstract. Very often, multivariate data analysis problems require dimensionality reduction (DR) stages to either improve analysis performance or represent the data in an intelligible fashion. Traditionally DR techniques are developed under different frameworks and settings what makes their comparison a non-trivial task. In this sense, generalized DR approaches are of great interest as they enable both to power and compare the DR techniques in a proper and fair manner. This work introduces a generalized spectral dimensionality reduction (GSDR) approach able to represent DR spectral techniques and enhance their representation ability. To do so, GSDR exploits the use of kernel-based representations as an initial nonlinear transformation to obtain a new space. Then, such a new space is used as an input for a feature extraction process based on principal component analysis. As remarkable experimental results, GSDR shows to be able to outperform the conventional implementation of well-known spectral DR techniques (namely, classical multidimensional scaling and Laplacian eigenmaps) in terms of the scaled version of the average agreement rate. Additionally, relevant insights and theoretical developments to understand the effect of data structure preservation at local and global levels are provided.

Keywords: Dimensionality reduction · Kernel representations · Principal component analysis · Spectral methods

1 Introduction

Dimensionality reduction (DR) aims to embed relevant information from high-dimensional data into a lower dimension representation, being of great use among data-related

M. Velez-Falconi—This work is supported by SDAS research group (www.sdas-group.com).

© Springer Nature Switzerland AG 2021

O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12952, pp. 512–523, 2021.

https://doi.org/10.1007/978-3-030-86973-1_36

areas such as big data, pattern recognition, clustering or data visualization. Many DR techniques have been extensively studied, ranging from distance-based preservation criteria, e.g. classical approaches such as classical multidimensional scaling (CMDS) [1], to graph-based approaches as Laplacian eigenmaps (LE) [2]. Given such a wide range of techniques developed under different frameworks and settings, generalized DR approaches are of great interest as it enables to contrast and enhance them in a fair, proper manner.

In this work, a generalized spectral dimensionality reduction (GSDR) approach capable of representing DR spectral techniques and exploit their representation ability is introduced. To do so, in the here-studied GSDR a new space representation is obtained through an initial nonlinear transformation employed by the use of kernel-based representations. Then, a feature extraction process based on principal component analysis is applied in such new space. Remarkable experimental results shows that in terms of the scaled version of the average agreement rate, GSDR is able to outperform the conventional implementation of well-known spectral DR techniques -namely, classical multidimensional scaling and Laplacian eigenmaps. Additionally, aimed at better understanding the effect of data structure preservation at local and global levels simultaneously, theoretical developments and relevant insights are provided.

The remaining of this manuscript is structured as follows: Sect. 2 states the notation used throughout this work and presents a brief overview on kernels. In Sect. 3, we introduce the GSDR method. Both the nonlinear mapping and the feature extraction are explained in theoretical and computational terms. Section 4 describes the setup and parameter settings for experiments. Section 5 gathers and discusses the experimental results. Finally, conclusions and final remarks are drawn in Sect. 6.

2 Background on Kernel Functions and Notation

2.1 Notation

Let us define the input data matrix as $\mathbf{X} \in \mathbb{R}^{N \times D}$ holding N samples represented by D variables, in the form: $\mathbf{X} = (\mathbf{x}_1^\top, \dots, \mathbf{x}_N^\top)^\top$, with $\mathbf{x}_i \in \mathbb{R}^D$ and $i \in \{1, \dots, N\}$. Likewise, let $\mathbf{Y} \in \mathbb{R}^{N \times d}$ be the output data matrix, such that $\mathbf{Y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_N^\top)^\top$, $\mathbf{y}_i \in \mathbb{R}^d$ and $d \leq D$. In terms of feature extraction, matrix \mathbf{Y} is the embedded (also extracted, projected, or mapped) space. In such vein, it is traditionally set $d < D$ for DR purposes. That said, the aim of DR is to embed the space \mathbf{X} into a lower-dimensional space \mathbf{Y} .

2.2 Concept of Kernel Function

Roughly speaking, the so-named kernel function can be understood as an approach that allows for estimating the similarity among input data samples [3]. In general, such similarity is calculated over samples from either independent or associated spaces [4,5]. In this work, the concept of kernel is referred to the pairwise similarity or affinity measures intended to represent the input data. Naturally, similarity measures must

be ruled by a positive semi-definite function. In mathematical terms, we can express a positive semi-definite kernel function $\mathcal{K}(\cdot, \cdot)$ as follows:

$$\begin{aligned} \mathcal{K}(\cdot, \cdot) : \mathbb{R}^D \times \mathbb{R}^D &\longrightarrow \mathbb{R} \\ \mathbf{x}_i, \mathbf{x}_j &\longmapsto \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j), \end{aligned} \tag{1}$$

satisfying

$$\sum_{i=1}^N \sum_{j=1}^N c_i \bar{c}_j \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) \geq 0, \tag{2}$$

for all $c_i \in \mathbb{C}$, being \bar{c}_i the complex conjugate of c_i .

3 Proposed Generalized Spectral Dimensionality Reduction (GSDR) Approach

The here-proposed Generalized Spectral Dimensionality Reduction, short termed as GSDR, is based on the premise that data can be mapped onto another space $\mathbf{Z} \in \mathbb{R}^{N \times M}$ before going through a feature extraction procedure itself. In this connection and inspired by works devoted to dissimilarity-based representations [6], we alternatively propose to explore the possibility of a nonlinear mapping $\mathcal{T}\{\cdot\}$ based on pairwise similarities, such that:

$$\mathbf{Z} = \mathcal{T}\{\mathbf{X}\}, \tag{3}$$

where $z_{ij} = \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j)$. Therefore, \mathbf{Z} is said to be a kernel matrix as well as $M = N$. Then, a linear projection is performed over the mapped space to obtain the embedded space \mathbf{Y} , such that $\mathbf{Y} = \mathbf{Z}\mathbf{R}$ where $\mathbf{R} \in \mathbb{R}^{N \times d}$ is a rotation matrix to be defined.

Figure 1 depicts a high-level outline of the proposed GSDR.

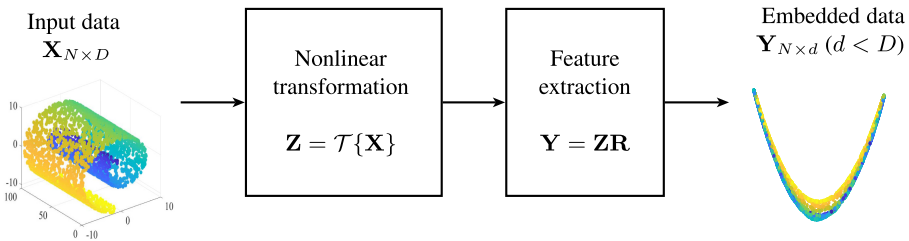


Fig. 1. Block diagram of the proposed GSDR. It mainly involves two steps: nonlinear transformation based on kernel functions (similarities) and linear feature extraction using PCA.

Notice that in this work, either an element of the space (matrix) or the space itself is indistinctly referred as space.

3.1 Nonlinear Transformation Using a Kernel Matrix

Since vectors \mathbf{x}_i are assumed to be real and D -dimensional, and a collection of N vectors is available (just as stated in notation given above), a matrix $\mathbf{Z} \in \mathbb{R}^{N \times N}$ with entries $z_{ij} = \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j)$ can be formed. Such a matrix is known as kernel matrix (Gram or generalized co-variance matrix as well). Therefore, a real symmetric $N \times N$ matrix \mathbf{Z} whose ij entries satisfy Eq. (2) for all $c_i \in \mathbb{R}$ is also a positive semi-definite matrix.

A remarkable benefit of this property is that all eigenvalues of \mathbf{Z} are ensured to be non-negative, which enables to readily carry out useful spectral developments for feature extraction purposes.

Figure 2 depicts the effect of the kernel-based data representation. By nature, a kernel entries can be understood as pairwise similarities, and therefore non-directed, weighted graph becomes a suitable geometric representation thereof.

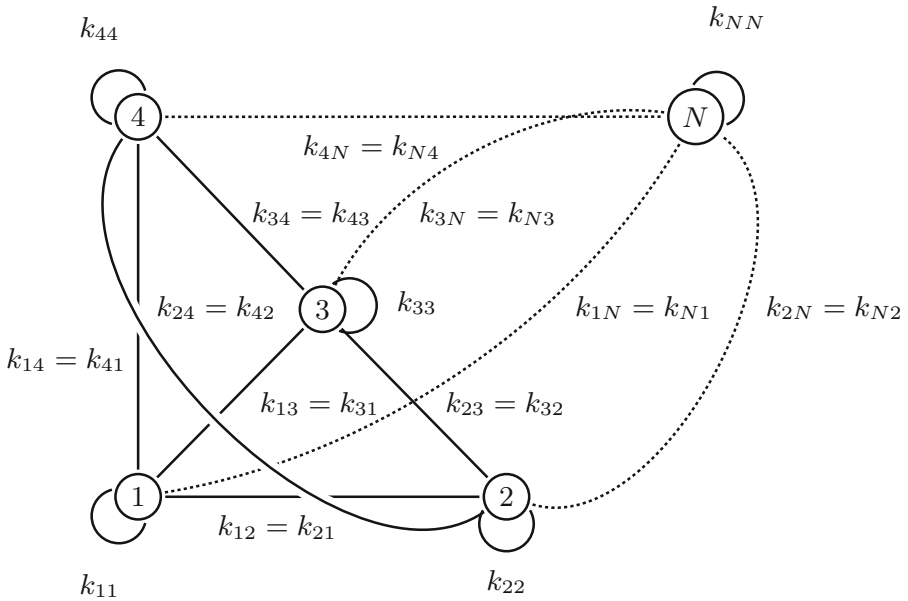


Fig. 2. An explanatory diagram depicting the relationship between the kernel entries and the weights of an N -node weighted, non-directed graph within a similarity-based representation framework.

As a matter of fact, kernel matrix entries may be related to the similarity among nodes (data points), which is in turn related to an opposite notion of distance, and therefore the concept of close neighborhood (local structure) takes place. Such a local structure of data can be preserved by a kernel function if its corresponding kernel matrix is properly tuned and selected, and subsequently used as an input to either a robust enough kernelized DR method [7] or similarity-driven generalized DR [8].

3.2 PCA-Based Feature Extraction

For the dimensionality reduction process to be carried out over the new space \mathbf{Z} , we use a PCA-based feature extraction approach. It works as follows: First, let us consider a linear projection in the form:

$$\mathbf{Y} = \mathbf{Z}\mathbf{R}, \quad (4)$$

where $\mathbf{R} \in \mathbb{R}^{N \times d}$ is a projection or rotation matrix. In this connection, the condition $d < D$ takes place to extract features in a lower dimensional space.

To ensure linear independence and prevent from length effects, an orthonormal rotation matrix is considered, i.e. $\mathbf{R}^\top \mathbf{R} = \mathbf{I}_d$, where \mathbf{I}_d is d -dimensional identity matrix.

The estimation of \mathbf{R} follows from the distance-based framework widely explained in [8].

Briefly put, this framework minimizes the distance between of \mathbf{Z} and a low-rank representation thereof $\widehat{\mathbf{Z}} \in \mathbb{R}^{N \times N}$, as follows:

$$\begin{aligned} \min_{\mathbf{R}} \|\mathbf{Z} - \widehat{\mathbf{Z}}\|_2^2 \\ \mathbf{R}^\top \mathbf{R} = \mathbf{I}_d, \end{aligned} \quad (5)$$

where $\|\cdot\|_2$ stands for the Euclidean (L_2) norm. As demonstrated in [8], previous formulation is equivalent to the following dual problem:

$$\begin{aligned} \max_{\mathbf{R}} \text{tr}(\mathbf{R}^\top \mathbf{\Sigma} \mathbf{R}) \\ \mathbf{R}^\top \mathbf{R} = \mathbf{I}_d, \end{aligned} \quad (6)$$

where

$$\mathbf{\Sigma} = \mathbf{Z}^\top \mathbf{Z} \quad (7)$$

and $\text{tr}(\cdot)$ denotes the conventional matrix trace operator.

As the functional of the dual optimization problem presented in (6) is quadratic and \mathbf{R} is an orthonormal matrix, it is easy to demonstrate that a feasible solution is selecting \mathbf{R} are the eigenvectors corresponding to the d largest eigenvalues of $\mathbf{\Sigma}$. It is worth noticing that, once centered the matrix \mathbf{Z} with

$$\mathbf{Z} \leftarrow \left(\mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\top \right) \mathbf{Z}, \quad (8)$$

being $\mathbf{1}_N$ an N -dimensional all ones vector, $\mathbf{\Sigma}$ becomes an estimation of the covariance matrix of \mathbf{Z} .

3.3 GSDR Algorithm

The Algorithm 1 is a pseudocode gathering the steps of the proposed GSDR.

Algorithm 1. Generalized spectral dimensionality reduction (GSDR)

 $\mathbf{Y}_{N \times d} = \text{GSDR}(\mathbf{X}_{N \times D}, \mathcal{K}(\cdot, \cdot), d)$

Input: Data matrix $\mathbf{X}_{N \times D} = (\mathbf{x}_1^\top, \dots, \mathbf{x}_N^\top)^\top$, given kernel function $\mathcal{K}(\cdot, \cdot)$, desired dimension d ($d < D$)

- 1: Calculate the transformation $\mathbf{Z}_{N \times N} = \mathcal{T}\{\mathbf{X}_{N \times D}\}$ as $z_{ij} = \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j)$
- 2: Center matrix $\mathbf{Z}_{N \times N}$ with $\mathbf{Z}_{N \times N} \leftarrow (\mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\top) \mathbf{Z}_{N \times N}$
- 3: Calculate $\Sigma_{N \times N}$ as $\Sigma_{N \times N} = \mathbf{Z}_{N \times N}^\top \mathbf{Z}_{N \times N}$
- 4: Calculate the eigenvalue and eigenvector decomposition of $\Sigma_{N \times N}$ as $[\Lambda_N, \mathbf{V}_{N \times N}] = \text{eig}(\Sigma_{N \times N})$, with $\Lambda_N = (\lambda_1, \dots, \lambda_N)$, $\mathbf{V}_N = (\mathbf{v}_1, \dots, \mathbf{v}_N)$ and $\lambda_1 > \dots > \lambda_N$.
- 5: Form $\mathbf{R}_{N \times d}$ as the eigenvectors corresponding to the first d largest eigenvalues of $\Sigma_{N \times N}$ organized in decreasing order as: $\mathbf{R}_{N \times d} = (\mathbf{v}_1, \dots, \mathbf{v}_d)$
- 6: Calculate $\mathbf{Y}_{N \times d} = \mathbf{Z}_{N \times N} \mathbf{R}_{N \times d}$

Output: Embedded data matrix $\mathbf{Y}_{N \times d}$

4 Experimental Setup

Kernels for DR: Two kernel approximations for spectral DR methods [9] are considered, namely CMDS and LE, as detailed in Table 1.

All previously mentioned kernels are widely described in [9].

Table 1. Brief description of the here-used kernel matrices representing dimensionality reduction.

Classical multidimensional scaling (CMDS)	Laplacian eigenmaps (LE)
<p>CMDS kernel is the double centered distance matrix $\mathbf{D} \in \mathbb{R}^{N \times N}$, such that $\mathbf{K}_{\text{CMDS}} = -\frac{1}{2}(\mathbf{I}_N - \mathbf{1}_N \mathbf{1}_N^\top) \mathbf{D} (\mathbf{I}_N - \mathbf{1}_N \mathbf{1}_N^\top)$</p> <p>where the ij entry of \mathbf{D} is given by $d_{ij} = \ \mathbf{y}_i - \mathbf{y}_j\ _2^2$ and $\ \cdot\ _2$ stands for Euclidean norm.</p>	<p>The LE kernel can be expressed as $\mathbf{K}_{\text{LE}} = \mathbf{L}^\dagger$,</p> <p>where $\mathbf{L} = \mathcal{D} - \mathbf{S}$ is the graph Laplacian, \mathbf{S} is a similarity matrix, $\mathcal{D} = \text{Diag}(\mathbf{S} \mathbf{1}_N)$ is the degree matrix, and \mathbf{L}^\dagger stands for the pseudo-inverse of \mathbf{L}</p> <p>Following the study done in [10], \mathbf{S} is calculated by keeping the relative bandwidth parameter as $\log K$ to keep the entropy across the neighbors, being K a given number of neighbors. In this work, K is set to be 30</p>

Performance Measure: The performance of the considered methods is quantified by the scaled version, ranged within the interval $[0, 1]$, of the average agreement rate $R_{NX}(K)$ presented in [11]. Given that $R_{NX}(K)$ is obtained at each perplexity value from 2 to $N - 1$, a numerical indicator of the overall performance is acquired through calculating its area under the curve (AUC). Therefore, this AUC is an overall indicator of quality preservation of a DR approach, as it evaluates the most appropriate weights at all scales.

Both the dimensionality reduction techniques (GSDR, LE, and CMDS) and the performance measure ($R_{NX}(K)$ curve) are implemented on MATLAB Version: 9.10 (R2021a).

Databases: Experiments are carried out over fourth conventional data sets. Figure 3 depicts examples/views of the considered data sets.

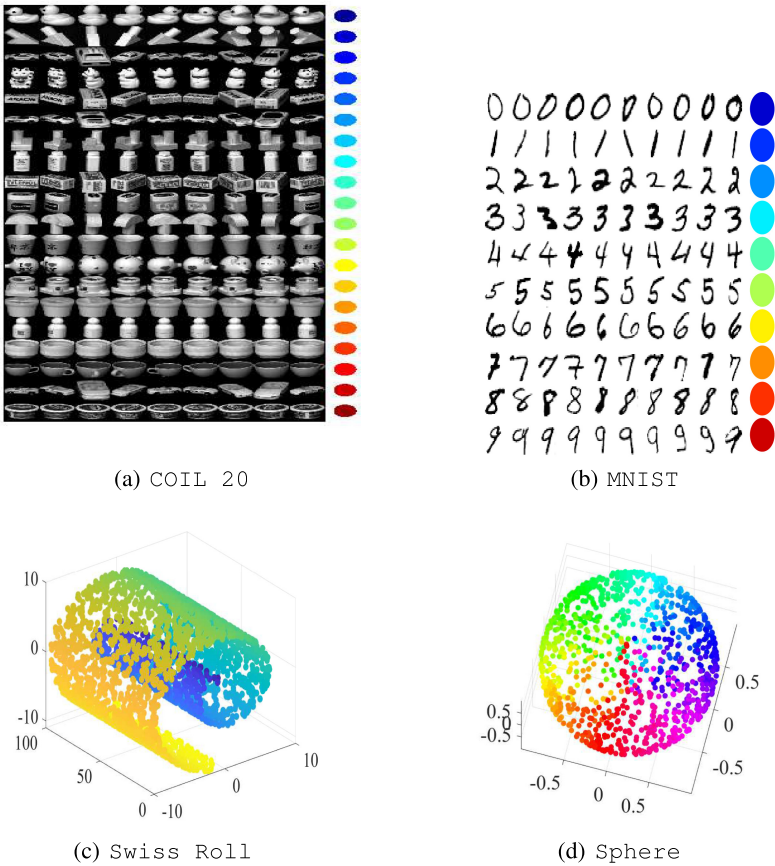


Fig. 3. The four considered data sets: COIL-20, MNIST, Swiss roll and Sphere

The first data set is a randomly selected subset of the MNIST image bank [12], which is formed by 6000 gray-level images of each of the 10 digits ($N = 1500$ data points –150 instances for all 10 digits– and $D = 24^2$), a sample is presented in Fig. 3(b). The second one is the COIL-20 image bank [13], which contains 72 gray-level images representing 20 different objects ($N = 1440$ data points –20 objects in 72 poses/angles– with $D = 128^2$), as seen in Fig. 3(a). The third data set (Sphere) is an artificial spherical shell ($N = 1500$ data points and $D = 3$), and the fourth data set is a toy set here called `Swiss roll` ($N = 3000$ data points and $D = 3$), depicted in Fig. 3(d) and 3(c), respectively.

5 Results and Discussion

The plot of the R_{NX} curve and its AUC obtained from reducing the data sets by the conventional implementation (no kernelized) of CMDS [1] and LE [2] are compared with the ones reached by GSDR ($\mathbf{X}, \mathcal{K}(\cdot, \cdot), 2$) (according to Algorithm 1) when selecting $\mathcal{K}(\cdot, \cdot)$ as kernel functions producing respectively the kernel matrices \mathbf{K}_{CMDS} and \mathbf{K}_{LE} . Results are shown in Fig. 4, 5, 6 and 7.

As can be observed, GSDR slightly outperforms conventional CMDS and LE in all cases. A remarkable advantage of the proposed GSDR is its ability to both unfold manifolds (as seen in Fig. 4 and 5) and reach separable-classes visualization from complex, high dimensional real data (as seen in Fig. 6 and 7).

It is also worth noticing that some rotation occurs over the embedded spaces as can be appreciated in Fig. 6(b), 6(d), 7(b) and 7(d). This fact is due to the orthogonal rotation done at the second step of GSDR procedure, which adds an effect of global structure preservation.

That said, as it performs a kernel-based representation and linearly projects the data with a PCA-based rotation, GSDR is able to preserve both the global and local structure of the input data.

Even though these preliminary results exhibit no great improvement regarding conventional DR methods, its mathematical development and versatility is highly promising as it opens possibilities to exploit simultaneously a kernel-matrix-based representation together with simple PCA.

As demonstrated in previous works [14, 15], a joint formulation involving linear projections (just as PCA) and either similarity-based or kernelized representations is a suitable framework to design DR alternatives able to preserve local and global structure of the data.

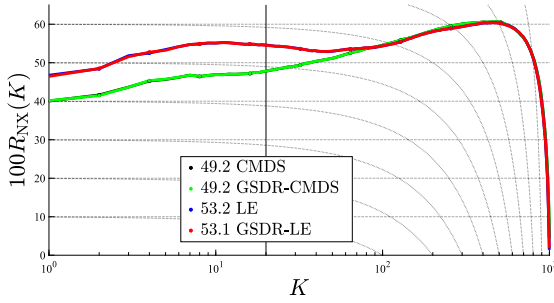
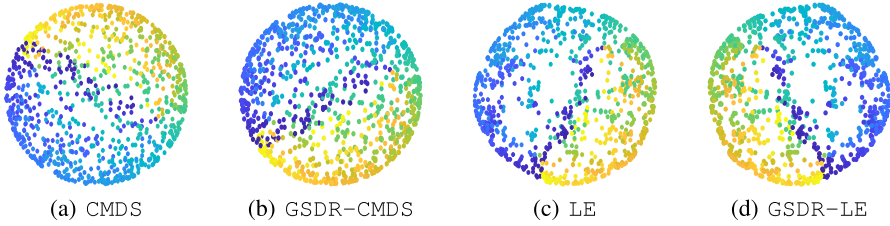


Fig. 4. Results for Sphere. Both the embedded spaces of GSDR and conventional CMDS and LE are shown. Comparison is made in terms of the R_{NX} curve

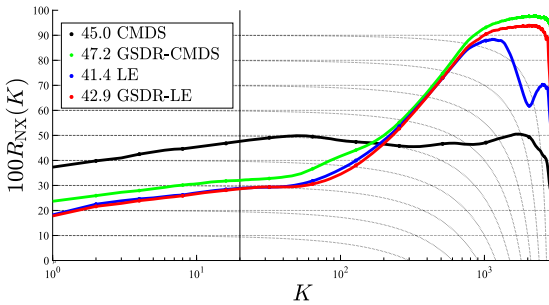
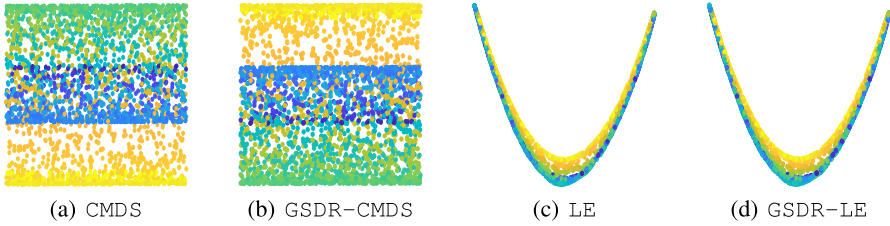
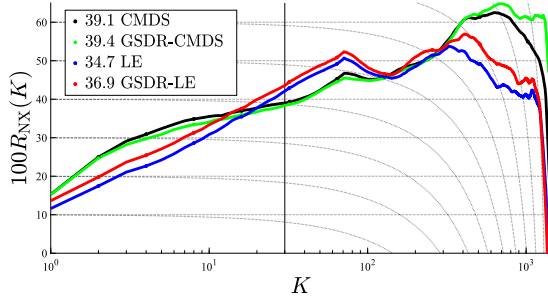
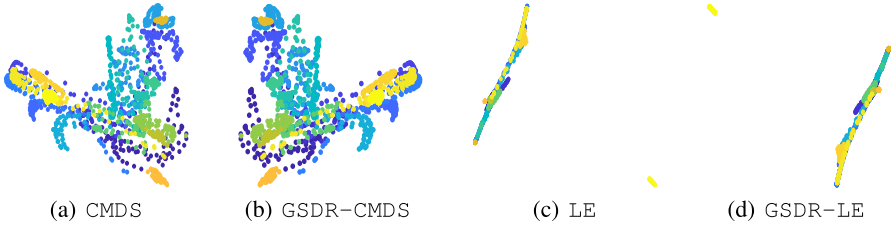
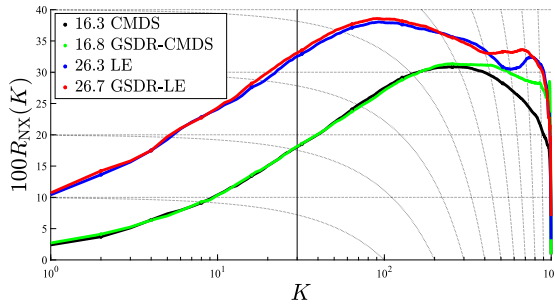
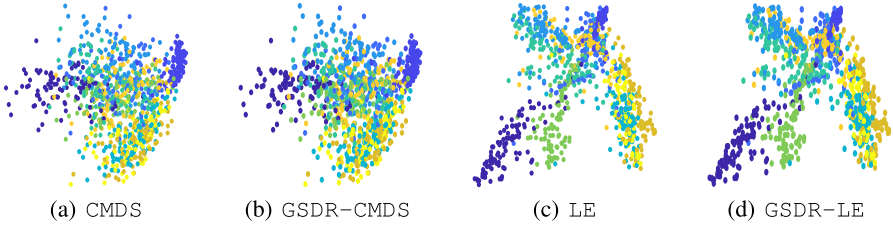


Fig. 5. Results for Swiss roll.



(e) R_{NX} curve

Fig. 6. Results for COIL 20.



(e) R_{NX} curve

Fig. 7. Results for MNIST.

6 Conclusions and Future Work

In this work, we present a generalized spectral dimensionality reduction (GSDR) approach, which exploits simultaneously the use of kernel-based representations and a feature extraction stage. The former is an initial nonlinear transformation aimed to generate a new space wherein the local-structure attributes are captured. The latter uses that new space as an input for a principal-component-analysis-driven projection, which enables the preservation of global-structure attributes. Experimentally, we prove that proposed GSDR reaches competitive performance in contrast to the conventional implementation of classical multidimensional scaling and Laplacian eigenmaps in terms of structure preservation criteria.

As a future work, more kernel representations are to be explored, which can be plugged to spectral dimensionality reduction approaches aiming at reaching a suitable trade-off between the preservation of local and global structure of data.

Acknowledgments. This work is supported by the research project “Proyecto PN223LH010-005 Desarrollo de nuevos modelos y métodos matemáticos para la toma de decisiones”. Authors thank the valuable support given by the SDAS Research Group (www.sdas-group.com).

References

1. Borg, I.: *Modern Multidimensional Scaling: Theory and Applications*. Springer, New York (2005)
2. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15**(6), 1373–1396 (2003)
3. Belanche Muñoz, L.A.: Developments in kernel design. In: *ESANN 2013 proceedings: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning: Bruges (Belgium), 24–26 April 2013*, pp. 369–378 (2013)
4. Bagchi, A.: *Lecture notes: Efficient approximation of kernel functions* (2020)
5. Ramon, E., Belanche-Muñoz, L., Molist, F., Quintanilla, R., Perez-Enciso, M., Ramayo-Caldas, Y.: kernint: a kernel framework for integrating supervised and unsupervised analyses in spatio-temporal metagenomic datasets. *Front. Microbiol.* **12**, 60 (2021)
6. Porro-Muñoz, D., Duin, R.P., Talavera, I., Orozco-Alzate, M.: Classification of three-way data by the dissimilarity representation. *Sig. Proc.* **91**(11), 2520–2529 (2011)
7. Peluffo-Ordóñez, D.H., Aldo Lee, J., Verleysen, M.: Generalized kernel framework for unsupervised spectral methods of dimensionality reduction. In: *Computational Intelligence and Data Mining (CIDM), 2014 IEEE Symposium on*, pp. 171–177. IEEE (2014)
8. Peluffo, D., Lee, J., Verleysen, M., Rodríguez, J., Castellanos-Domínguez, G.: Unsupervised relevance analysis for feature extraction and selection: a distance-based approach for feature relevance. In: *ICPRAM 2014 - Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods* (2014)
9. Ham, J., Lee, D.D., Mika, S., Schölkopf, B.: A kernel view of the dimensionality reduction of manifolds. In: *Proceedings of the Twenty-First International Conference on Machine Learning*, vol. 47 ACM (2004)
10. Cook, J., Sutskever, I., Mnih, A., Hinton, G.E.: Visualizing similarity data with a mixture of maps. In: *International Conference on Artificial Intelligence and Statistics*, pp. 67–74 (2007)
11. Lee, J.A., Renard, E., Bernard, G., Dupont, P., Verleysen, M.: Type 1 and 2 mixtures of kullback-leibler divergences as cost functions in dimensionality reduction based on similarity preservation. *Neurocomputing* **112**, 92–108 (2013)

12. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
13. Nene, S.A., Nayar, S.K., Murase, H.: Columbia object image library (coil-20). Dept. Comput. Sci. Columbia Univ. New York. **62** (1996). <http://www.cs.columbia.edu/CAVE/coil-20.html>
14. Rodríguez-Sotelo, J.L., Peluffo-Ordóñez, D., Cuesta-Frau, D., Castellanos-Domínguez, G.: Unsupervised feature relevance analysis applied to improve ECG heartbeat clustering. *Comput. Methods Programs Biomed.* **108**(1), 250–261 (2012)
15. Blanco Valencia, X.P., Becerra, M., Castro Ospina, A., Ortega Adarme, M., Viveros Melo, D., Peluffo Ordóñez, D.H., et al.: Kernel-based framework for spectral dimensionality reduction and clustering formulation: a theoretical study. *ADCAIJ: Adv. Distrib. Comput. Artif. Intell. J.* **6**(1) (2017)