# Nonnegative Embeddings and Projections for Dimensionality Reduction and Information Visualization

Stefanos Zafeiriou and Nikolaos Laskaris
*Artificial Intelligence and Information Analysis Lab*
*Department of Informatics*
*Aristotle University of Thessaloniki,Greece*
*drzafeiriou@gmail.com, laskaris@aiia.csd.auth.gr*

## Abstract

*In this paper, we propose novel algorithms for low dimensionality nonnegative embedding of vectorial and/or relational data, as well as nonnegative projections for dimensionality reduction. We start by introducing a novel algorithm for Metric Multidimensional Scaling (MMS). We propose algorithms for Nonnegative Locally Linear Embedding (NLLE) and Nonnegative Laplacian Eigenmaps (NLE). By reformulating the problem of MMS, NLLE and NLE for finding projections we propose algorithms for Nonnegative Principal Component Analysis (NPCA), for Nonnegative Orthogonal Neighbourhood Preserving Projections (NONPP) and Nonnegative Orthogonal Locality Preserving Projections (NOLPP). We demonstrate some first preliminary results of the proposed methods in data visualization.*

## 1 Introduction

Machine learning, pattern recognition and data mining constitute areas of great development during the past decade. One of the primary goals of many data mining and machine learning systems is dimensionality reduction. The goal of dimensionality reduction is to reduce the number of features of data in order to perform tasks like clustering and/or training a classifier. The motivation behind data dimensionality reduction is to reduce the cost and the number of parameters of the classifier. In some cases of learning algorithms (like Linear Discriminant Analysis (LDA) [2]) and for Small Sample Size (SSS) problems (where $N \ll F$, $N$ is the number of samples and $F$ is the feature dimensionality) it is crucial to perform dimensionality reduction in order perform invertion of matrices and eigenanalysis. To that end, a lot of methods were proposed [2, 7, 1, 11, 3, 4, 6]. Metric Multidimension Scaling [5] (MMS) and Princi-

pal Component Analysis (PCA) [2] were among the first methods proposed for dimensionality reduction. LDA was one of the most popular method for feature extraction especially for face recognition [2]. Nonlinear and linear dimensionality reduction methods based on preserving local structure of data were also proposed [7, 1, 11, 3, 4, 6]. Another group of methods created having the same aim was the one based on Nonnegative Matrix Factorization [9, 6, 8].

In this paper we enrich the arsenal of dimensionality reduction algorithms by applying methods proposed in [6] in order to perform Nonnegative MMS (NMMS), Nonnegative Locally Linear Embedding (NLLE) as well as to find Nonnegative Laplacian Eigenmaps (NLE). Subsequently, we reformulate the problem and instead of finding directly the data embedding we approximate it trying to find projections based on Nonnegative Principal Component Analysis (NPCA), Nonnegative Orthogonal Neighbourhood Preserving Projections (NONPP) and Nonnegative Orthogonal Locallity Preserving Projections (NOLPP). We present some initial results of the proposed methods using artificial data.

## 2 Nonnegative Embedding Algorithms

Let a set of original samples $\mathbf{x}_i \in \Re^F$ be represented by a matrix $\mathbf{X} = [\mathbf{x}_1|\cdots|\mathbf{x}_N]$ and $\mathbf{m} = \sum_{i=1}^{N} \mathbf{x}_i$ be the center of samples. Let also $\bar{\mathbf{X}} = [\mathbf{x}_1 - \mathbf{m}|\cdots|\mathbf{x}_N - \mathbf{m}]$ be the centered data matrix. All the algorithms that will be described below aim at finding a low dimensional embedding $\mathbf{Y} = [\mathbf{y}_1|\cdots|\mathbf{y}_N]$ with $\mathbf{y}_i \in \Re^P$ with $P < F$ such that a particular criterion is satisfied.

### 2.1 Nonnegative Metric Multidimensional Scaling

The aim of MMS [5] is to find an embedding $\mathbf{Y}$ such that the distances of the low dimensional embedded data

$||\mathbf{y}_i - \mathbf{y}_j||^2$ approximatelly preserve the distances of the original data $||\mathbf{x}_i - \mathbf{x}_j||^2$. In MMS the problem is reformulated into an equivalent problem of preserving the inner products in the embedded space. Let that we define the centered inner product matrix in the original space $\mathbf{K} = [\bar{\mathbf{x}}_i^T \bar{\mathbf{x}}_j]$, where $\bar{\mathbf{x}}_i = \mathbf{x}_i - \mathbf{m}$ are the centered data. Let that $\mathbf{D} = \left[||\mathbf{x}_i - \mathbf{x}_j||^2\right]$ be the distance matrix. Then, the centered dot product matrix is defined as $\mathbf{K} = -(\mathbf{I} - \frac{1}{N}\mathbf{E})\mathbf{D}(\mathbf{I} - \frac{1}{N}\mathbf{E})$, where $\mathbf{E}$ is a matrix of ones. The optimization problem of MMS is:

$$\mathbf{Y}_o = \arg\min_{\mathbf{Y}} ||\mathbf{K} - \mathbf{Y}^T\mathbf{Y}||_F^2. \qquad (1)$$

The solution is provided by the eigendecomposition of $\mathbf{K} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ and for a $P$ dimensional reduction we select $\mathbf{Y}_o = \mathbf{\Lambda}_P^{\frac{1}{2}}\mathbf{U}_P^T$. A natural generalization of the above procedure is the ISOMAP [1]. In this case we use the geodesic distance instead of the typical $L_2$ norm.

By observing that $\mathbf{U}_P^T\mathbf{U}_P = \mathbf{I}_P$ we reformulate the above optimization problem so as to enforce nonnegativity constrains for the resulted embedding $\mathbf{Y} = \mathbf{S}^{\frac{1}{2}}\mathbf{H}^T$:

$$\mathbf{H}_o = \arg\min_{\mathbf{H}} ||\mathbf{K} - \mathbf{HSH}^T||_F^2 \\ \text{s.t. } \mathbf{H} \geq 0, \ \mathbf{S} \geq 0, \ \mathbf{HH}^T = \mathbf{I} \qquad (2)$$

which is the proposed nonnegative MMS (NMMS).

Using similar reasoning as the one developed in [6] in order to find nonnegative decompositions under orthogonality constraints, we can prove that the following update rules converge to a local minimum of optimization problem (2):

$$\begin{aligned} \mathbf{H} &\leftarrow \mathbf{H} * \sqrt{\frac{\mathbf{K}^+\mathbf{HS} + \mathbf{HH}^T\mathbf{K}^-\mathbf{HS}}{\mathbf{K}^-\mathbf{HS} + \mathbf{HH}^T\mathbf{K}^+\mathbf{HS}}} \\ \mathbf{s} &\leftarrow \mathbf{s} * \sqrt{\frac{(\mathbf{H}\odot\mathbf{H})\text{vec}(\mathbf{K}^+)}{(\mathbf{H}\odot\mathbf{H})(\mathbf{H}\odot\mathbf{H})'\mathbf{s} + (\mathbf{H}\odot\mathbf{H})\text{vec}(\mathbf{K}^-)}} \end{aligned} \qquad (3)$$

where $\mathbf{S} = \text{diag}(\mathbf{s})$, $*$ is the Hadamard product, $\odot$ is the Khatri-Rao product, $\text{vec}(\mathbf{A}) \in \Re^{NM}$ is the vectorized matrix $\mathbf{A} \in \Re^{N \times M}$ and for a matrix $\mathbf{G}$ we define the matrices $\mathbf{G}^+$ and $\mathbf{G}^-$ as $G_{ij}^+ \triangleq \begin{cases} G_{ij} & \text{if } G_{ij} \geq 0 \\ 0 & \text{otherwise.} \end{cases}$ and $G_{ij}^- \triangleq \begin{cases} |G_{ij}| & \text{if } G_{ij} \leq 0 \\ 0 & \text{otherwise.} \end{cases}$

## 2.2 Nonnegative Locally Linear Embedding

In LLE [10, 11], instead of preserving the global distances $||\mathbf{x}_i - \mathbf{x}_j||^2$, as in MMS, an embedding is found that preserves the neighbourhoods of the original samples $\mathbf{x}_i$. LLE works under the assumption that the original space lie on some high dimensional manifold, thus every sample can be faithfully reconstructed using only a few neighbors. The reconstruction error is defined as:

$$E(\mathbf{W} = [w_{ij}]) = \sum_{i=1}^{N} \left|\left|\mathbf{x}_i - \sum_{j=1}^{N} w_{ij}\mathbf{x}_j\right|\right|^2. \qquad (4)$$

The weights $w_{ij}$ for a sample $\mathbf{x}_i$ are equal to zero if $\mathbf{x}_j$ does not belong in the neighbourhood of this sample. Moreover, $\sum_{j=1}^{N} w_{ij} = 1$ and $w_{ii} = 0$. A lot of methods and optimization problems were proposed for finding the weights matrix $\mathbf{W}$ [11]. Afterwards, having found the optimal weights matrix $\mathbf{W}$ we want to find a low dimensional embedding $\mathbf{Y}$ by optimizing:

$$\begin{aligned} \mathbf{Y}_o &= \arg\min_{\mathbf{Y}} \sum_{i=1}^{N} \left|\left|\mathbf{y}_i - \sum_{j=1}^{N} w_{ij}\mathbf{y}_j\right|\right|^2 \\ &= \arg\min_{\mathbf{Y}} \text{tr}\left[\mathbf{Y}(\mathbf{I}-\mathbf{W})(\mathbf{I}-\mathbf{W})^T\mathbf{Y}^T\right] \\ \text{s.t.} \quad & \mathbf{YY}^T = \mathbf{I} \end{aligned} \qquad (5)$$

As in the NMMS we enforce nonnegativity constraints in the embedding $\mathbf{Y}$ and propose the following optimization problem:

$$\mathbf{Y}_o = \arg\min_{\mathbf{Y}} \text{tr}\left[\mathbf{YMY}^T\right] \\ \text{s.t.} \mathbf{Y} \geq 0, \ \mathbf{YY}^T = \mathbf{I}. \qquad (6)$$

where $\mathbf{M} = (\mathbf{I} - \mathbf{W})(\mathbf{I} - \mathbf{W})^T$. In a similar manner as above we can prove that the following update rules converge to a local minimum:

$$\mathbf{Y} \leftarrow \mathbf{Y} * \sqrt{\frac{\mathbf{YM}^- + \mathbf{YM}^+\mathbf{Y}^T\mathbf{Y}}{\mathbf{YM}^+ + \mathbf{YM}^-\mathbf{Y}^T\mathbf{Y}}} \qquad (7)$$

## 2.3 Nonnegative Laplacian EigenMaps

LE [3] follows similar reasoning as LLE. The main difference is in the way the weights are chosen to represent locality. The heat kernel $w_{ij} = e^{-\frac{||\mathbf{x}_i - \mathbf{x}_j||_2}{\gamma}}$ or constant weights ($w_{ij} = 1$ if the $i$-th and the $j$-th vectors are adjacent and $w_{ij} = 0$ otherwise) constitute common choices for the weights. They also have a slightly different objective function [3]. LE finds a low dimensional embedding $\mathbf{Y}$ by minimizing the following:

$$\begin{aligned} \mathbf{Y}_o &= \arg\min_{\mathbf{Y}} \frac{1}{2} \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} ||\mathbf{y}_i - \mathbf{y}_j||_2^2 \\ &= \arg\min_{\mathbf{Y}} \text{tr}\left(\mathbf{YLY}^T\right) \\ \text{s.t.} \quad & \mathbf{YDY}^T = \mathbf{I} \end{aligned} \qquad (8)$$

where $\mathbf{L} = \mathbf{D} - \mathbf{W}$ where $\mathbf{D} = \text{diag}(\mathbf{W1})$ (where $\text{diag}(\mathbf{a})$ is the diagonal matrix having as main diagonal vector $\mathbf{a}$). The objective function with the chosen weights $w_{ij}$ results in a heavy penalty if the neighboring points $\mathbf{x}_i$ and $\mathbf{x}_j$ are mapped far apart. Therefore, its minimization ensures that if $\mathbf{x}_i$ and $\mathbf{x}_j$ are "near", then $\mathbf{y}_i$ and $\mathbf{y}_j$ are "near", as well.

Following the same reasoning as above we propose a nonnegative embedding $\tilde{\mathbf{Y}} = \mathbf{YD}^{\frac{1}{2}}$ and solve the following optimization problem as:

$$\arg\min_{\mathbf{Y}} \text{tr}\left(\tilde{\mathbf{Y}}\mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}}\tilde{\mathbf{Y}}^T\right) \\ \text{s.t.} \ \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T = \mathbf{I}, \ \tilde{\mathbf{Y}} \geq 0. \qquad (9)$$

The update rules for $\tilde{\mathbf{Y}}$ are the same as the ones given in (7) but instead of $\mathbf{M}$ we use the matrix $\mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}}$.

# 3 From Nonnegative Embeddings to Nonnegative Projections

In this section we confine ourselves to finding a projection matrix $\mathbf{U}$, in order to find the desired embedding. That is, the embedding $\mathbf{Y}$ is given by $\mathbf{Y} = \mathbf{U}^T\mathbf{X}$.

## 3.1 Nonnegative Principal Component Analysis

We can prove that by substituting $\mathbf{Y} = \mathbf{U}^T\bar{\mathbf{X}}$ into the MMS problem we can derive the well-known PCA optimization problem:

$$\begin{aligned} \mathbf{U}_o &= \arg\max_{\mathbf{U}} \operatorname{tr}\left[\mathbf{U}^T\bar{\mathbf{X}}\bar{\mathbf{X}}^T\mathbf{U}\right] \\ \text{s.t.} \quad &\mathbf{U}^T\mathbf{U} = \mathbf{I} \end{aligned} \tag{10}$$

The corresponding NPCA problem is given by:

$$\begin{aligned} \mathbf{U}_o &= \arg\max_{\mathbf{U}} \operatorname{tr}\left[\mathbf{U}^T\bar{\mathbf{X}}\bar{\mathbf{X}}^T\mathbf{U}\right] \\ \text{s.t.} \quad &\mathbf{U} \geq 0, \ \mathbf{U}^T\mathbf{U} = \mathbf{I}. \end{aligned} \tag{11}$$

By letting $\mathbf{C} = \bar{\mathbf{X}}\bar{\mathbf{X}}^T$ we can guarantee that the optimization problem will reach a local maximum, given by:

$$\mathbf{U} \leftarrow \mathbf{U} * \sqrt{\frac{\mathbf{C}^+\mathbf{U} + \mathbf{U}\mathbf{U}^T\mathbf{C}^-\mathbf{U}}{\mathbf{C}^-\mathbf{U} + \mathbf{U}\mathbf{U}^T\mathbf{C}^+\mathbf{U}}}. \tag{12}$$

## 3.2 Orthogonal Nonnegative Neighbourhood Preserving Projections and Orthogonal Nonnegative Locallity Preserving Projections

The Neighbourhood preserving projections (NPP) [7] method is similar to the LLE method but instead of seeking for a general mapping $\mathbf{Y}$ it seeks for particular projection matrix $\mathbf{U}$. That is, $\mathbf{Y}$ is now $\mathbf{Y} = \mathbf{U}^T\mathbf{X}$. Then, optimization problem (5) is reformulated as:

$$\begin{aligned} \mathbf{U}_o &= \arg\min_{\mathbf{U}} \operatorname{tr}\left[\mathbf{U}^T\mathbf{X}\mathbf{M}\mathbf{X}^T\mathbf{V}\right] \\ \text{s.t.} \quad &\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U} = \mathbf{I} \end{aligned} \tag{13}$$

A variation of this method is the orthogonal NPP (ONNP) where the projection $\mathbf{U}$ is considered orthogonal [7]

$$\begin{aligned} \mathbf{U}_o &= \arg\min_{\mathbf{U}} \operatorname{tr}\left[\mathbf{U}^T\mathbf{X}\mathbf{M}\mathbf{X}^T\mathbf{U}\right] \\ \text{s.t.} \quad &\mathbf{U}^T\mathbf{U} = \mathbf{I}. \end{aligned} \tag{14}$$

The proposed optimization problem for finding nonnegative ONNP is given by:

$$\begin{aligned} \mathbf{U}_o &= \arg\min_{\mathbf{U}} \operatorname{tr}\left[\mathbf{U}^T\mathbf{X}\mathbf{M}\mathbf{X}^T\mathbf{U}\right] \\ \text{s.t.} \quad &\mathbf{U} \geq 0, \ \mathbf{U}^T\mathbf{U} = \mathbf{I}. \end{aligned} \tag{15}$$

Let that $\mathbf{P} = \mathbf{X}\mathbf{M}\mathbf{X}^T$, then the update rules that guarantee that the above optimization problem will lead to a local minimum are given by:

$$\mathbf{U} \leftarrow \mathbf{U} * \sqrt{\frac{\mathbf{P}^-\mathbf{U} + \mathbf{U}\mathbf{U}^T\mathbf{P}^+\mathbf{U}}{\mathbf{P}^+\mathbf{U} + \mathbf{U}\mathbf{U}^T\mathbf{P}^-\mathbf{U}}}. \tag{16}$$
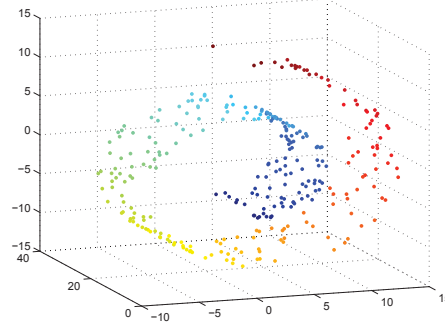


**Figure 1. Random Sampled Swiss Roll.**

The merit of the using orthogonal projections were also discussed in [4] where orthogonal LLP (OLLP) were proposed (the counterpart of Laplacian eigenmaps in case that we seek for projections). In OLLP the optimization problem (8) was reformulated in:

$$\begin{aligned} \mathbf{U}_o &= \arg\min_{\mathbf{U}} \operatorname{tr}\left[\mathbf{U}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{U}\right] \\ \text{s.t.} \quad &\mathbf{U}^T\mathbf{U} = \mathbf{I}. \end{aligned} \tag{17}$$

in order to find a set of orthogonal projections.

As in NONPP we can define the nonnegative OLPP (NOLPP) by solving the following optimization problem:

$$\begin{aligned} \mathbf{U}_o &= \arg\min_{\mathbf{U}} \operatorname{tr}\left[\mathbf{U}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{U}\right] \\ \text{s.t.} \quad &\mathbf{U} \geq 0, \ \mathbf{U}^T\mathbf{U} = \mathbf{I}. \end{aligned} \tag{18}$$

The update rules are the same as (16) but instead of using matrix $\mathbf{P}$ we use matrix $\mathbf{X}\mathbf{L}\mathbf{X}^T$.

# 4 Experimental Results

The proposed emdedding and projection methods are tested using simulated data (Randomly Sampled Swiss Roll Figure 1 ). The low dimension visualization using LE and the proposed NLE are depicted in Figures 2 (a) and (b), respectively. Visualization of the projections of the first two principal components is depicted in Figure 2 (c) and the corresponding visualization using two nonnegative principal components from the proposed method are depicted in Figure 2 (d). As can be seen the proposed methods resulted in very rich low-dimensional embedding. Finally, the convergence of two of the proposed methods (NLME and NPCA) are depicted in Figures 2 (e) and (f).
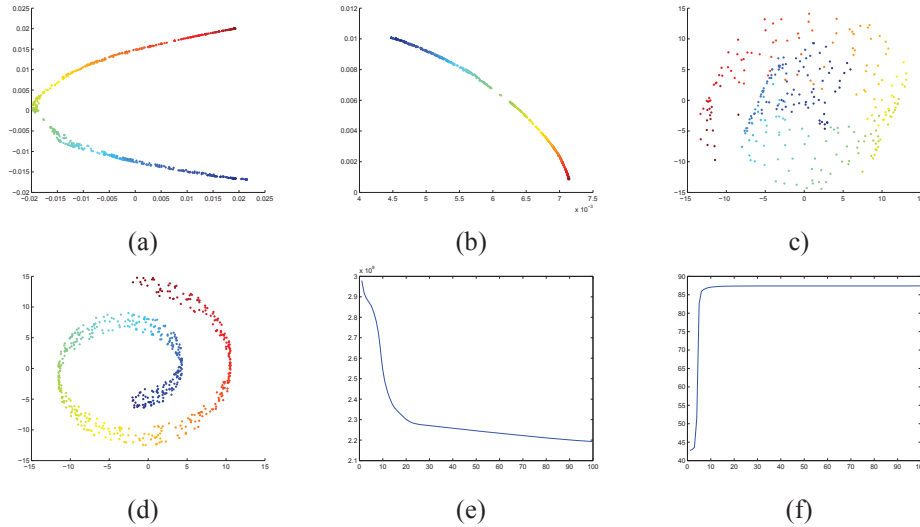
**Figure 2. (a) The embedding or the random sampled Swiss Roll using LE; (b) the embedding using NLE; (c) the embedding using the first two principal components; (d) the embedding using the first two nonnegative principal components; (e) objective function value versus the number of iterations for the NLE method; (f) objective function value versus the number of iterations for the NPCA method**

## 5  Conclusions

Novel algorithms for nonnegative embedding of high dimensional data were proposed. We have reformulated algorithms like MMS, LLE and LE by incorporating nonnegativity constraints. Moreover, by requiring the embedding to be a result of a projection of the original we proposed algorithms for performing NPCA, NONPP and NOLPP. We demonstrated the usefullness of the proposed algorithms in low dimensional visualization of prototypical manifolds. We believe that the proposed methods would be useful in a variety of dimensionality reduction problems (like face recognition).

## 6  Acknowlegdment

This work was partial funded by a postdoctoral fellowship of Stefanos Zafeiriou by the Greek State Scholarships Foundation.

## References

[1] M. Balasubramanian, E. Schwartz, J. Tenenbaum, V. de Silva, and J. Langford. The isomap algorithm and topological stability. *Science*, 295(5552):7, 2002.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligences*, 19(7):711–720, 1997.

[3] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

[4] D. Cai, X. He, J. Han, and H. Zhang. Orthogonal laplacianfaces for face recognition. *IEEE transactions on image processing*, 15(11):3608, 2006.

[5] M. Cox. *Multidimensional scaling*. CRC Press, 2000.

[6] C. Ding, T. Li, W. Peng, and H. Park. Orthogonal nonnegative matrix t-factorizations for clustering. In *Proceedings of the 12th ACM SIGKDD*, page 135, 2006.

[7] E. Kokiopoulou and Y. Saad. Orthogonal neighborhood preserving projections: A projection-based dimensionality reduction technique. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2143, 2007.

[8] I. Kotsia, S. Zafeiriou, and I. Pitas. A novel discriminant nonnegative matrix factorization method with application to facial image characterization problems. *IEEE Trans. Information Forensics and Security*, 2(3):588–595, 2007.

[9] D. Lee and H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[10] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323, 2000.

[11] L. Saul and S. Roweis. Think globally, fit locally: unsupervised learning of low dimensional manifolds. *The Journal of Machine Learning Research*, 4:119–155, 2003.