Out of samples extensions for SC-LLE, new nonlinear dimensionality reduction algorithm.

Hala Hijazi, Oussama Bazzi
UL, Faculté des sciences 1
Beyrouth, Liban
Email: {halaht,obazzi}@ul.edu.lb

André Bigand
LISIC, ULCO
Calais Cedex, 62228, France
Email: bigand@lisic.univ-littoral.fr

***Abstract*—In a previous paper a new version of nonlinear dimensionality reduction algorithm was proposed, the SC-LLE approach. This approach combines a supervised method, linear discriminant analysis (LDA, a simple but widely used algorithm in pattern recognition) with an unsupervised method, local linear embedding (LLE, manifold learning). SC-LLE method can generalize any linear classifier (like LDA) to nonlinear by transforming data into some low-dimensional feature space. This new concept (SC-LLE) applied to nonlinear data projection seems to be promising, and we show in this new paper that semi-supervised learning (SSL) is another interesting property of SC-LLE. Applications on 3D data show the interest of this method.**

***Index Terms*—Semi-supervised learning; spectral clustering; dimensional reduction; Pattern Recognition.**

I. Introduction

In pattern recognition and machine learning fiels, many applications face high-dimensional problem (the dimensionality curse!). Finding a low-dimensional representation of the high-dimensional data is a basic task (reduced features). Using supervised classification problem, e.g. linear discriminant analysis (LDA) [1], the reduced low-dimensional features contain discriminative information based on the labeled data.

Recent developments in so-called manifold learning techniques has opened new perspectives in nonlinear data analysis, such as Isomap [2], LLE (linear locally embedding [3]), Laplacian eigenmaps [4], Hessian eigenmaps [5] and methods based on Riemannian normal coordinates [6]. Characteristic for this new generation of manifold learning techniques is efficiency and global convergence (many of these techniques are based on the solution of very large eigenvalue problems). Projection algorithms make it possible visualization and exploratory analysis of multidimensional data. Projective methods consist of detecting a linear (or nonlinear) structure (in a reduced dimensional space) from the input multidimensional data space. Most of dimension reduction techniques are unsupervised. Recently, some work was done to combine supervised and unsupervised dimensional reduction algorithms. Tang *et al.* [7] proposed a nonlinear discriminant mapping (NDM) using the Laplacian of a graph and combining the advantages of LDA and Isomap. They obtained interesting results. Nevertheless Isomap can be only applied to intrinsically flat manifolds. So in the same way we propose a new supervised spectral space classifier named SC-LLE (spectral clustering-LLE). This LLE variant of spectral clustering, in which such an embedding is an an intermediate step before obtaining a clustering of the data that can capture flat, elongated and curved clusters. The two tasks (manifold learning and clustering) are linked because clusters found by spectral clustering can be arbitrary curved manifolds. Moreover, this new technique combines the advantages of LDA and LLE in the general framework of multidimensional reduction methods with minimal shortcomings, and using the results of [8] for extension of spectral methods. We have shown that SC-LLE is a supervised dimensionality reduction method [9] wich overcomes vulnerability of LDA and LLE against within-class multimodality or outliers [10].

When only a small number of labeled samples are available, supervised dimensionality methods tend to find embedding spaces which are overfitted to the labeled samples. In such situations, using unlabeled samples is often effective. Chapelle *et al.* [11] showed that preserving the global structure of all samples in an unsupervised manner can be better than strongly relying on class information provided by a small number of labeled samples. So in this paper, we show that combining discriminative properties of LDA algorithm with manifold learning properties of LLE algorithm leads to an interesting semi-supervised learning (SSL) framework. Based on the relationship between regularized discriminant analysis and regularized least-squares (RLS), we add a regularization term to the original criteria of LAD. The regularization term is based on the prior knowledge proviede by both labeled and unlabeled data, and can be constructed using graph Laplacian [12]. This new algorithm is then applied to classic geometrical problems with good results.

The paper is organized as follows:

- In section 2 we briefly review LDA and LLE algorithms and we briefly describe our previous SC-LLE algorithm;
- In section 3 we present the SSL framework we propose;
- In section 4 we will present some results;
- Finally, the paper is summarized with some conclusions in section 5.

978-1-4673-5307-6/13/$31.00 ©2013 IEEE

II. SC-LLE: NEW NONLINEAR DIMENSIONALITY REDUCTION ALGORITHM

A. Existing methods: LLE and LDA algorithms

Let a data set $X = \{x_1, ..., x_i, ...x_m, x_{m+1}, ..., x_n\}$, including m training samples ($X_S = \{x_1, ..., x_m\}$, with a known label set $Y_S = \{y_1, ...y_i, ..., y_m\}$) and $(n-m)$ test data ($X_T = \{x_{m+1}, ..., x_n\}$, with a label set $Y_T = \{y_{m+1}, ..., y_n\}$ to be predicted). Assume that there exists p classes in samples ($y_i \in C = \{c_1, ..., c_p\}$). The goal is to compute outputs y_i that provide a faithfull embedding in $d < m$ dimensions to solve the problem of manifold learning. LLE is one possible solution.

B. Linear locally embedding (LLE)

LLE algorithm tends to solve the same problem as Isomap (graph-based method) but using linear locally approach ([3]). This method preserves the local topology present in input data space in three steps. In first step, a directed graph G is constructed whose edges indicate nearest neighbors relations. Each point and its nearer neighbors are approximatively located on a locally linear surface. In second step, weights w_{ij} (assigned to the edges of the graph) are computed by reconstructing each input x_i from its *k*-nearest neighbors. There are two constraints for w_{ij}: $w_{ij} = 0$ if x_j is not among the *k*-nearest neighbors of x_i and $\sum_{j=1}^{m} w_{ij} = 1$, thus the weights constitute a sparse matrix $W = (w_{ij})_{m \times m}$.
In third step the *d-dimensional* embedding (with $d < m$) that minimizes objective matrix $\Phi(Y)$ is obtained by computing the bottom *d+1* eigenvectors of the matrix $\Phi = (I-W)^T(I-W)$. These eigenvectors often produce reasonable embedding but do not map exactly input submanifolds. These shortcomings were presented in a previous communication [9]. Then the eigenvector solutions are usally collected in an indicating matrix for post-processing, such as k-means clustering. At this step, LDA may be another classifier possibility as we now recall.

C. Linear discriminant analysis (LDA)

LDA is based on the Fisher linear discriminant function which is given by:

$$J(A) = \frac{A^T S_b A}{A^T S_w A} \tag{1}$$

where A is the objective projecting matrix, S_w and S_b are within and between class scatter matrices defined as:

$$S_w = \frac{1}{m} \sum_{i=1}^{p} \sum_{j=1}^{l_i} \left[\left(x_i^j - m_i \right) . \left(x_i^j - m_i \right)^T \right] \tag{2}$$

$$S_b = \frac{1}{m} \sum_{i=1}^{p} \left[l_i . (m_i - m_o) . (m_i - m_o)^T \right] \tag{3}$$

where x_i^j is the jth sample of the ith class, l_i is the number of samples of the ith class and m_i and m_o are the mean vectors of the ith class samples and all samples respectively.

Let $S = S_b - S_w$. A novel discriminant criterion function is given by:

$$\widetilde{J(A)} = \frac{A^T S A}{A^T A} \tag{4}$$

It is easy to prove that $J(A)$ and $\widetilde{J(A)}$ will reach the maximum at the same points ([13]). This other discriminant function is used in the following.

D. The design of new SC-LLE algorithm

1) Proposed non linear discriminant analysis: To outperform mentioned shortcomings, a new framework is built from extension of spectral methods proposed by [8]. The three parts of this framework are the following ones:

- Supervised manifold mapping
- Classifier construction;
- Test of the method.

Starting from the set of m data points X_S, and according to a specific similarity measure we have the affinity matrix W $\in R^{m \times m}$. From W we can construct a weighted graph $G = (V, E)$ with each vertex $v_i \in V$ corresponding to the data point x_i and each edge $e(i, j) \in E$ carries a weight w_{ij} which represents the similarity between point x_i and x_j. The clustering problem is equivalent to choosing a partition $c_1, c_2, ..., c_p$ of G which minimizes a specific objective function. Basic idea of spectral method is to transform the graph segmentation problem to the eigenvalue problem of Laplacian matrix L, without estimating any explicit model of the data distribution. According to Rayleigh-Ritz theory [14], the approximate solution could be derived from the leading eigenvectors of L. The use of Laplacian matrix eigenvector for approximating the graph minimum cut is called spectral clustering. So let classically be $L = D - W$ the Laplacian L of G where:

$$L = \begin{cases} -w_{ij} & \text{if i} \neq \text{j} \\ \sum_{j=1}^{m} w_{ij} & \text{if i} \in \text{[1, m].} \end{cases} \tag{5}$$

and where D is a diagonal matrix with $D_{ii} = \sum_{j=1}^{m} w_{ij}$.
Supervised information integration present in S (equation 4) is made using the global and local Laplacians, as proposed by [7] for NDM. So we use a modified local-scaled graph construction method based on the global and local Laplacians of a graph G, L_g and L_l^L respectively. From equations 2 and 3 , S in equation 4 can be rewritten as:

$$S = \frac{1}{M}(XL_gX^T - 2XL_l^LX^T). \tag{6}$$

From equation 6, it is obvious that LDA is translated into spectral decomposition of the two Laplacians, L_g and L_l^L. So the proposed method is a combination of LDA and LLE that makes it possible to combine the discriminant property of LDA and the nonlinear projective property of LLE. These two properties will be illustrated in the following experimental results.

2) Supervised spectral space classifier: It is easy to show that the matrix W may be transformed such as:

$$L_g W L_g^T = -Y^T Y \tag{7}$$

Let $S_g = Y^T Y$, equation 7 can be rewritten:

$$S_g = Y^T Y = -L_g W L_g^T \tag{8}$$

According to Rayleigh-Ritz theory, the objective function of MDS (multidimensional scaling method) in LLE is to maximize:

$$J_g(Y) = \frac{Y^T S_g Y}{Y^T Y} \tag{9}$$

Thus optimal scaling Y is equal to some eigenvectors of S_g corresponding to several eigenvalues and this eigenvector expansion makes it possible to outperform the classification performances of LLE using SC-LLE. The SC-LLE algorithm is similar to LLE, but the addition of both global and local Laplacians in a spectral framework, i.e., L_g and L_l^L which can be easily derived from class label of the training samples of LDA, and contributes to discriminant property of LDA.

Inspired from the conclusions established in [7], and the similarity matrix W being block diagonal, the largest p eigenvectors are defined using the objective matrix of SC-LLE as:

$$\widetilde{S_g} = \frac{-1}{m}(L_g W L_g^T - 2 L_l W L_l^T). \tag{10}$$

where L_g and L_l^L are the same as those defined in equation 6. The criterion function of SC-LLE is then:

$$\widetilde{J_g}(Y) = \frac{Y^T \widetilde{S_g} Y}{Y^T Y} \tag{11}$$

So the criterion function of SC-LLE $\widetilde{J_g}(Y)$ can be considered as a nonlinear version of the criterion function $\widetilde{J}(A)$. The nonlinear attribute is captured in the Euclidean distance between all pairs of data points.

3) Algorithm: The implementation of this new spectral analysis method is given by the following algorithm:

Algorithm 1

Require: a data set X

1: Construct near-neighbor graph G
2: Calculate similarity matrix W
3: Calculate global and local Laplacian matrices, i.e., L_g and L_l^L
4: Calculate the eigenvalues and eigenvectors of objective matrix $\widetilde{S_g}$
5: Mapping of the training data into the spectral space
6: Give clusters from this projected representation using LDA classifier
7: Extension of the manifold mapping to the test data

III. Semi-supervised dimensionality reduction

In the semi-supervised problem, we only have partially labeled observed examples $D = \{X, Y\}$. We suppose there are l labeled points and u unlabeled points (with $l + u = m$, each point x is a d-dimensional vector). Then the observed input points can be written as $X = (X_L, X_U)$, where $X_L = \{x_1, ..., x_i, ...x_l\}$, and $X_U = \{x_{l+1}, ..., x_{l+u}\}$, with a label set $Y_L = \{y_1, ..., y_l\}$. We propose now a semi-supervised version of SC-LLE algorithm, starting from regularized LDA.

A. Regularized LDA

LDA is based on the Fisher linear discriminant function which is given by equation 1 (where A is the objective projecting matrix, S_w and S_b are within and between class scatter matrices defined previously). When data size is small, LDA algorithm encounters the ill-posed problem. In this case, the within-class scatter matrix S_w is singular for LDA, and the mixture scatter matrix S_m is poorly conditioned (S_m= S_b+ S_w). This can be solved by adding a regularization term to the original objective function. For regularized LDA, Song *et al.* [15] have shown that the Fisher linear discriminant function is given by:

$$J(A) = \frac{A^T S_b A}{A^T (S_w + \lambda I) A} \tag{12}$$

In the same way, it is obvious [15] that the scatter matrix $S = S_b - S_w$ can be transformed into $S = S_b - \lambda S_w$ used in $\widetilde{J(A)}$ (equation 4).

B. Semi-supervised regularization framework

Let introduce the classic regularization framework for LS in the binary classification case:

$$min_{f \in \mathcal{F}} \int\limits_{X \times Y} (y - f(x))^2 \, dp(x, y) + \lambda_1 \|f\|_T^2 + \lambda_2 \|f\|_M^2 \tag{13}$$

where $\lambda_1 \|f\|_T^2$ is the Tikhonov regularization term in function space, and $\lambda_2 \|f\|_M^2$ is the regularization term based on manifold analysis [12]. λ_1 and λ_2 are the parameters which control the tradeoffs of these two terms. To use the unlabeled information in the regularization term, graph Laplacian can be used to approximate the Laplace-Beltrami operator on data manifold [12], as we now recall the main result. Starting from defined typical adjacency matrix W of neighborhood graph and the graph Laplacian L, the following regularization term (in the projected space) using both labeled and unlabeled data can be formulated as:

$$tr(A^T X L X^T A) = \sum_{i,j}^{l+u} W_{i,j} \left\| \frac{1}{\sqrt{d_i}} y_i - \frac{1}{\sqrt{d_j}} y_j \right\|_2^2 \tag{14}$$

Minimizing this term means that we desire y_i and y_j to be close if they are close in the input space. This term is also the prior smoothness of labeled and unlabeled data. We now use it to regularize SC-LLE algorithm, to preserve the local information of the manifold structure. In our case, considering the local and global graph Laplacians (L_l and L_g), the previous equation is easily modified in that way.

C. Semi-supervised SC-LLE

Based on the relationship between equations 12 and 13, by adding the new regularization term 14 to LDA, the objective function of semi-supervised SC-LLE is to maximize:

$$\widetilde{S_g} = \frac{-1}{m}(L_g W L_g^T + \lambda_1 I - 2\lambda_2 L_l W L_l^T). \quad (15)$$

where L_g and L_l^L are the same as those defined in equation 6.

IV. Experimental results

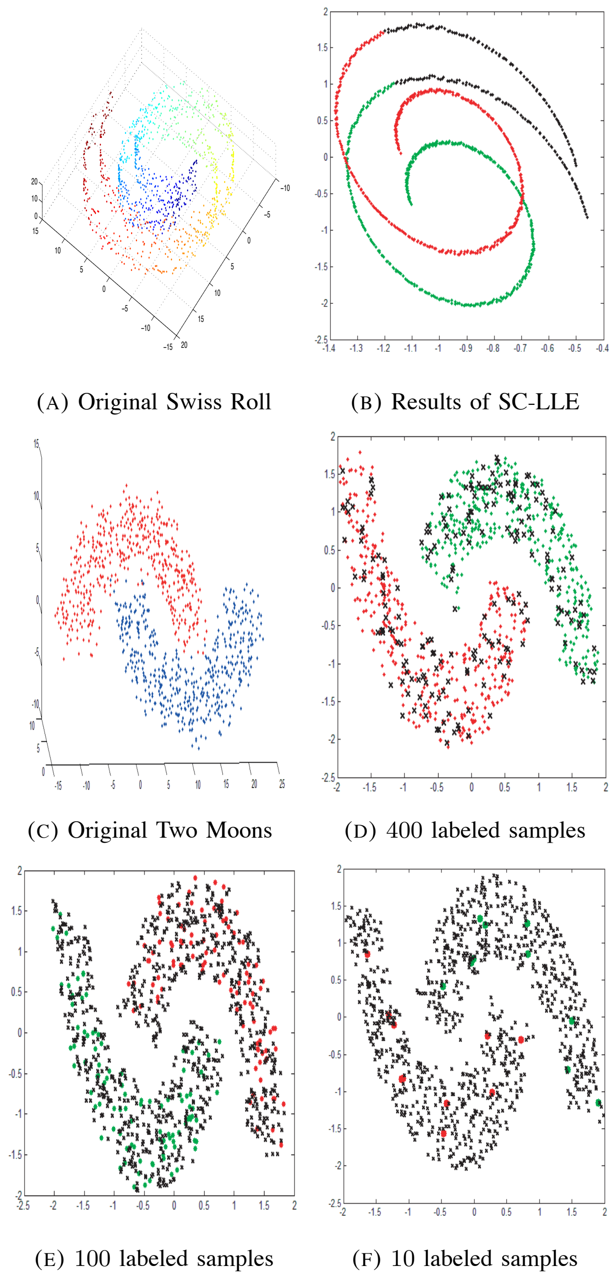
In this section, we will evaluate the performance of SC-LLE on three synthetic data sets (a 3D data set named "two-layer Swiss roll" and presented figure 1, where each roll of the original data set has 500 samples, and the classical "‘two-moons"’ data set) and classic Iris data problem. The semi-supervised manifold learning is illustrated in the low-dimensional spectral space (2D space).

A. SC-LLE approach, semi-supervised learning

In this section, we will evaluate the performance of SC-LLE by comparing its implementation with different values of l labeled data. In figure 1, the first image (top left) is the original (3D) "‘Swiss Roll "‘data set, the second image is the result obtained with SC-LLE, the third image (middle left) is the original "‘Two-moons"’ data set, the fourth image (middle right) is the result obtained 400 initial labeled samples (so 100 unlabeled samples represented using black points) and the two following images are relative to "‘Two-moons"’ data set, using 100 and 10 labeled samples.
It is interesting to analyze these results. First, we can easily notice that “two-moons” discrimination is realized whatever the value “l”. Then we can notice that good nonlinear projection is observed with this semi-supervised SC-LLE algorithm, and submanifold geometry is preserved.
By following a semi-supervised learning paradigm of LLE, the structure of the neighborhood graph is preserved in the linear embedding space and can be visualized. Running time remains reasonable (algorithms were implemented on the well-known software Matlab® on PC and it is important to note that the code has not been optimized).

B. Iris data, Out-of-sample extension

In order to evaluate the built classifier performances, we attempt to classify the out-of-samples using well-known Iris data. Iris (overlapping) data are shown in figure 2. Each (Iris) species are represented by 50 points, where only 30 are labeled samples. It turns out from figure 2 that the (unknown) unlabeled data still holds in the same classes as (supervised) training data, according to their similarity with the training data. Indeed the discriminative information is integrated into the similarity computation. More experiments on real test data and test ratios are being conducted and will be presented in the future.



(A) Original Swiss Roll (B) Results of SC-LLE

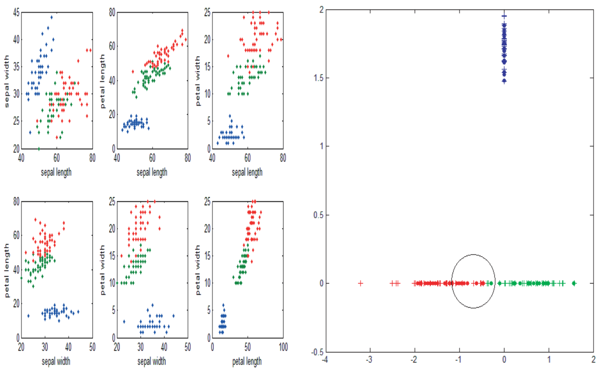
(C) Original Two Moons (D) 400 labeled samples

(E) 100 labeled samples (F) 10 labeled samples

Fig. 1. Clustering results on 3D data "‘Swiss Roll" and "‘Two Moons"’.

V. Conclusion

In this paper, we present a new semi-supervised dimensionality reduction framework, based on SC-LLE we presented previously. SC-LLE is based on a supervised algorithm (LDA) and an unsupervised algorithm (LLE) that can be combined in semi-supervised framework. By efficiently using the partially labeled data, interesting clustering properties of SC-LLE are preserved. Moreover the embedding results also show good discriminative information and manifold structure. The proposed algorithm preserves input space topology of data and makes it possible to find a linearly separable projection for nonlinear data in effectively combining dimensionality



(A) Iris test data (B) Results of SC-LLE

Fig. 2. Clustering results on test data Iris.

reduction (LDA) with discovering intrinsic structure of data (LLE). Nevertheless one problem is that the number of labeled samples is typically small in SSL and thus cross-validation is not easily reliable. So in the future we have to confirm the good results we obtain using huge data bases (like face data sets). We will also have to investigate the sensitivity to regularization parameters. Once this work made, this new method should be interesting for large data sets where dimensionality reduction is required. The algorithm should find a wide variety of potential applications (data analysis, visualization and classification). Particularly we want to extend the good performances (to noise or outliers) of the presented algorithm to real data sets (noisy color images) in the future.

References

[1] P.N.Belhumeur, J. Hespanda, and D.Kiregeman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. on PAMI*, vol. 19(7), pp. 711–720, 1997.

[2] J. Tenenbaum, V. Silvam, and J. Langford, "A global geometric framework for non linear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 2000.

[3] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[4] M. Belkin and P. Nyogi, *Laplacian eigenmaps and spectral techniques for embedding and clustering:*, M. MIT Press, Cambridge, Ed. T.G. Diettrich and S. Becker and Z. Gharamani, 2002.

[5] D. Donoho and C. G. C., "Hessian eigenmaps: locally linear embedding techniques for high-dimensional data," *PNAS*, vol. 100, pp. 5591–5596, 2003.

[6] A.Brun, C.F.Westin, M.Herbethson, and H.Knutsson, "Fast manifold learning based on riemannian normal coordinates," in *Proc. 14th Scandinavian Conf. on Image Analysis*, 2005.

[7] H. Tang, T. Fang, and P.-F. Shi, "Nonlinear discriminant mapping using the Laplacian of a graph," *Pattern Recognition*, vol. 39, pp. 156–159, 2006.

[8] F. Sha and L. Saul, "Analysis and extension of spectral methods for nonlinear dimensionality reduction," in *Proc. of Int. Conf. Machine Learning, Germany*, 2005.

[9] H. Hijazi, O.Bazzi, and A.Bigand, "A new nonlinear discriminant analysis algorithm using a combined version of LDA and LLE," in *Proc. of IPCV'11, Las Vegas, USA*, 2011.

[10] H.Hijazi, O.Bazzi, and A.Bigand, "A new manifold learning for nonlinear dimensionality reduction," in *Proc. of ISIVC'12 Congress, Valenciennes, France*, 2012.

[11] O.Chapelle, B.Scholkopf, and A.Zien, *Semi-supervised learning*, Cambridge, Ed. MIT Press, 2006.

[12] M.Belkin, P.Niyogi, and V.Sindhwani, "Manifold regularization: a geometric framework for learning from labeled and unlabeled examples," *Journal of Machine Learning Research*, vol. 7, pp. 2399–2434, 2006.

[13] H. Tang, T. Fang, and P.-F. Shi, "Laplacian linear discriminant analysis," *Pattern Recognition*, vol. 39, pp. 136–139, 2006.

[14] H. Lutkepohl, *Handbook of Matrices*, Chichester, Ed. Wiley, 1997.

[15] Y.Song, F.Nie, C.Zhang, and S.Siang, "A unified framework for semi-supervised dimensionality reduction," *Pattern Recognition*, vol. 41, pp. 2789–2799, 2008.