Orthogonal Neighborhood Preserving Projections

E. Kokiopoulou and Y. Saad
Computer Science and Engineering Department
University of Minnesota
Minneapolis, MN 55455.
Email: {kokiopou,saad}@cs.umn.edu

***Abstract*— Orthogonal Neighborhood Preserving Projections (ONPP) is a linear dimensionality reduction technique which attempts to preserve both the intrinsic neighborhood geometry of the data samples and the global geometry. The proposed technique constructs a weighted data graph where the weights are constructed in a data-driven fashion, similarly to Locally Linear Embedding (LLE). A major difference with the standard LLE where the mapping between the input and the reduced spaces is implicit, is that ONPP employs an explicit linear mapping between the two. As a result, and in contrast with LLE, handling new data samples becomes straightforward, as this amounts to a simple linear transformation. ONPP shares some of the properties of Locality Preserving Projections (LPP). Both ONPP and LPP rely on a k-nearest neighbor graph in order to capture the data topology. However, our algorithm inherits the characteristics of LLE in preserving the structure of local neighborhoods, while LPP aims at preserving only locality without speci£cally aiming at preserving the geometric structure. This feature makes ONPP an effective method for data visualization. We provide ample experimental evidence to demonstrate the advantageous characteristics of ONPP, using well known synthetic test cases as well as real life data from computational biology and computer vision.**

Keywords: Linear dimensionality reduction, Locally Linear Embedding.

I. INTRODUCTION

The problem of dimensionality reduction appears in many £elds including data mining, machine learning and computer vision, to name just a few. The goal of dimensionality reduction is to map the high dimensional samples to a lower dimensional space such that certain properties are preserved. Usually, the property that is preserved is quanti£ed by an objective function and the dimensionality reduction problem is formulated as an optimization problem. For instance, Principal Components Analysis (PCA) is a traditional linear technique which aims at preserving the global variance and relies on the solution of an eigenvalue problem involving the sample covariance matrix. Locally Linear Embedding (LLE) [9], [11] is a nonlinear dimensionality reduction technique which aims at preserving the local geometries at each neighborhood.

While PCA is good at preserving the global structure, it does not preserve the locality of the data samples. In this paper, we propose a *linear* dimensionality reduction technique, named Orthogonal Neighborhood Preserving Projections (ONPP), which preserves the intrinsic geometry of the local neighborhoods. The high dimensional data samples are projected on a lower dimensional space by means of a linear transformation V. The dimensionality reduction matrix V is obtained by minimizing an objective function which captures the discrepancy of the intrinsic neighborhood geometries in the reduced space. Experimental evidence suggests that this feature is crucial in preserving the global geometry as well, via the interaction of overlapping neighborhoods. Thus, one could argue that ONPP can be effectively used for data visualization purposes and that it may be viewed as a synthesis of PCA and LLE.

ONPP constructs a weighted k-nearest neighbor (k-NN) graph which models explicitly the data topology. Similarly to LLE, the weights are built to capture the geometry of the neighborhood of each point. The linear projection step is determined by imposing the constraint that each data sample in the reduced space is reconstructed from its neighbors by the same weights used in the input space. However, in contrast to LLE, ONPP computes an explicit linear mapping from the input space to the reduced space. Note that in LLE the mapping is implicit and it is not clear how to embed new data samples (see e.g. research efforts by Bengio et al. [4]). In our case, the projection of a new data sample is straightforward and it simpli£es to a matrix vector product.

ONPP shares some properties with Locality Preserving Projections (LPP) [6]. Both are linear dimensionality reduction techniques which construct the k-NN graph in order to model the data topology. However, our algorithm uses the optimal data-driven weights of LLE which re¤ect the intrinsic geometry of the local neighborhoods, whereas the uniform weights (0/1) used in LPP aim at preserving locality without explicit consideration to the local geometric structure. Note that Gaussian weights can be used in LPP but these are somewhat arti£cial and require the selection of an appropriate value of the parameter σ, the width of the Gaussian envelope. Although this issue is often overlooked, it is crucial for the performance of the method and remains a serious handicap when using Gaussian weights. Experimental results suggest that ONPP is effective in conveying meaningful local and global geometric information from high dimensional samples to low dimensional ones.

Before starting, it is useful to de£ne in general terms the problem of dimensionality reduction. Given a dataset $X = [x_1, x_2, \ldots, x_n] \in R^{m\times n}$ and the dimension d of the reduced space, with $d \ll m$, the goal of dimensionality reduction is to determine a matrix $V \in R^{m\times d}$ such that the projected vectors $y_i = V^\top x_i$ in the reduced space satisfy a certain property. For instance, PCA computes V such that the variance of the projected vectors is maximized i.e,

$$\max_{\substack{V \in R^{m\times d}\\ V^\top V = I}} \left\| y_i - \frac{1}{n}\sum_{j=1}^{n} y_j \right\|_2^2, \quad y_i = V^\top x_i.$$

Our algorithm determines V such that the local geometry of the neighborhoods is preserved. This can be captured by a certain objective function and we show that optimizing this function results in an eigenvalue problem whose solution yields the matrix V. The next section describes the proposed method in more detail.

Algorithm: **Orthogonal Neighborhood Preserving Projections**
Input: Dataset $X \in R^{m\times n}$ and d: dimension of reduced space.
Output: Embedding vectors $Y \in R^{d\times n}$.
1. Compute the k nearest neighbors of data point $x_1, \cdots, x_n$.
2. Compute the weights W_{ij} which give the best linear reconstruction of each data point x_i by its neighbors (Equ. (3))
3. Compute the projected vectors $y_i = V^\top x_i,\ i = 1, \ldots, n$ where V is determined by computing the $d+1$ eigenvectors of $\tilde{M} = X(I - W^\top)(I - W)X^\top$ associated with smallest eigenvalues

TABLE I

THE ONPP ALGORITHM.

II. ORTHOGONAL NEIGHBORHOOD PRESERVING PROJECTIONS

Consider a dataset represented by the columns of a matrix $X = [x_1, x_2, \ldots, x_n] \in R^{m\times n}$. The £rst part of ONPP is identical with that of LLE [9], [11] and consists of computing some optimal weights in each neighborhood. The basic assumption is that each data sample along with its k nearest neighbors (approximately) lies on a locally linear manifold. Hence, each data sample x_i is reconstructed by a linear combination of its k nearest neighbors. The reconstruction errors are measured by minimizing the objective function

$$\mathcal{E}(W) = \sum_i \|x_i - \sum_j W_{ij} x_j\|_2^2. \tag{1}$$

The weight W_{ij} represent the linear coef£cient for reconstructing the sample x_i from its neighbors $\{x_j\}$. The following constraints are imposed on the weights:

1) $W_{ij} = 0$, if x_j is not one of the k nearest neighbors of x_i;
2) $\sum_j W_{ij} = 1$, that is x_i is approximated by a convex combination of its neighbors.

Note that the optimization problem (1) can be recast in matrix form as $\min_W \|X(I - W^\top)\|_F$, where W is an $n \times n$ sparse matrix which has a speci£c sparsity pattern (condition (1)) and satis£es the constraint that its row-sums be equal to one (condition (2)). The weights for a speci£c data point x_i are computed as follows. De£ne

$$C_{pl} = (x_i - x_p)^\top (x_i - x_l) \in R^{k\times k}, \tag{2}$$

the local Gram matrix containing the pairwise inner products among the neighbors of x_i, given that the neighbors are centered with respect to x_i. It can be shown that the weights of the above constrained least squares problem are given in closed form [9] using the inverse of C,

$$w_i = \frac{\sum_p C_{ip}^{-1}}{\sum_{pl} C_{pl}^{-1}}\ , \tag{3}$$

where w_i represents the i-th column of W. The weights W_{ij} satisfy certain optimality properties. They are invariant to rotations, scalings and translations. A direct consequence of these properties is that the weights preserve the intrinsic geometric characteristics of each neighborhood.

Consider now the second part of projecting the data samples X to the reduced space $Y = [y_1, y_2, \ldots, y_n] \in R^{d\times n}$. ONPP imposes an explicit linear mapping from $X \to Y$ such that $y_i = V^\top x_i,\ i = 1, \ldots, n$ for an appropriately determined matrix $V \in R^{m\times d}$. In order to determine the matrix V, ONPP imposes the constraint that each data sample y_i in the reduced space is reconstructed from its k neighbors by exactly the same weights as in the input space. This leads to the solution of the following optimization problem, where we set $M = (I - W^\top)(I - W)$ and $\tilde{M} = XMX^\top$

$$\begin{aligned} \min_Y \Phi(Y) &= \min_Y \sum_i \|y_i - \sum_j W_{ij} y_j\|_2^2 \\ &= \min_{V\in R^{m\times d}} \sum_i \|V^\top x_i - \sum_j W_{ij} V^\top x_j\|_2^2 \\ &= \min_{V\in R^{m\times d}} \|V^\top X(I - W^\top)\|_F^2 \\ &= \min_{V\in R^{m\times d}} \mathrm{tr}(V^\top XMX^\top V) \\ &= \min_{V\in R^{m\times d}} \mathrm{tr}(V^\top \tilde{M} V)\ . \end{aligned} \tag{4}$$

If we impose the additional constraint that the columns of V are orthonormal, i.e. $V^\top V = I$, then the solution V to the above optimization problem is the basis of the eigenvectors associated with the d smallest eigenvalues of $\tilde{M}$. We observed in practice that ignoring the smallest eigenvector of $\tilde{M}$ is helpful. This is an issue to be investigated in future work. Note that the embedding vectors of LLE are obtained by computing the eigenvectors of matrix M associated with its smallest eigenvalues.

New data points. Consider now a new test data sample x_t that needs to be projected. The test sample is projected onto the subspace $y_t = V^\top x_t$ using the dimensionality reduction matrix V. Therefore, the computation of the new projection simpli£es to a matrix vector product.

Computational cost. The £rst part of ONPP consists of forming the k-NN graph. This scales as O(n^2). Its second part requires the computation of a few of the smallest eigenvectors of $\tilde{M}$. Observe that in practice this matrix is not computed explicitly. Rather, iterative techniques are used to compute the corresponding smallest singular vectors of matrix $X(I-W)^\top$ [10]. The inner computational kernel of these techniques is the matrix-vector product which scales quadratically with the dimensions of the matrix at hand.

III. SUPERVISED ONPP

It is possible to implement ONPP in either an unsupervised or a supervised setting. Note that in the later case where the class labels are available, ONPP can be modi£ed appropriately and yield a projection which carries not only geometric information but discriminating information as well. In a supervised setting we £rst build the data graph $G = (N, E)$, where the nodes N correspond to data samples and an edge $e_{ij} = (x_i, x_j)$ exists if and only if x_i and x_j belong to the same class. In other words, we make adjacent those nodes (data samples) which belong to the same class. Notice that in this case one does not need to set the parameter k, the number of nearest neighbors, and the method becomes fully automatic.

Denote by c the number of classes and n_i the number of data samples which belong to the i-th class. The data graph G consists of c cliques, since the adjacency relationship between two nodes re¤ects their class relationship. This implies that with an appropriate reordering of the columns and rows, the weight matrix W will have a block diagonal form where the size of the i-th block is equal to the size n_i of the i-th class. In this case W will be of the following form,

$$W = \operatorname{diag}(W_1, W_2, \ldots, W_c).$$

The weights W_i within each class are computed in the usual way as described by equation (3). The rank of W de£ned above, is restricted as is explained by the following proposition.

Proposition 3.1: The rank of W is at most $n - c$.

Proof: Recall that the row sum of the weight matrix W_i is equal to 1, because of the constraint (2). This implies that $W_i e_i = 0, e_i = [1, \ldots, 1]^\top \in R^{n_i}$. Thus, the following c vectors

$$\begin{bmatrix} e_1 & 0 & \cdots & 0 \\ 0 & e_2 & \cdots & 0 \\ 0 & 0 & \cdots & e_c \end{bmatrix},$$

are linearly independent and belong to the null space of W. Therefore, the rank of W is at most $n - c$. ■

Consider now the case $m > n$ where the number of samples (n) is less than their dimension (m). This case is known as the *undersampled size* problem. A direct consequence of the above proposition is that in this case, the matrix $\tilde{M} \in R^{m \times m}$ will have rank at most $n - c$. In order to ensure that the resulting matrix $\tilde{M}$ will be nonsingular, we may employ an initial PCA projection that reduces the dimensionality of the data vectors to $n - c$. Call V_{PCA} the dimensionality reduction matrix of PCA. Then the ONPP algorithm is performed and the total dimensionality reduction matrix is given by

$$V = V_{\text{PCA}} V_{\text{ONPP}},$$

where V_{ONPP} is the dimensionality reduction matrix of ONPP.

IV. Kernel ONPP

It is possible to formulate a kernelized version of ONPP. Kernels have been extensively used in the context of Support Vector machines (SVMs) [12]. Essentially, we employ a nonlinear mapping $\Phi : R^m \to \mathcal{H}$. Denote by $\Phi(X) = [\Phi(x_1), \Phi(x_2), \ldots, \Phi(x_n)]$ the transformed dataset in the feature space $\mathcal{H}$. Denote also by $K_{ij} = k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$ the Gram matrix induced by the kernel $k(x, y)$ associated with the feature space. Consider the case $d = 1$

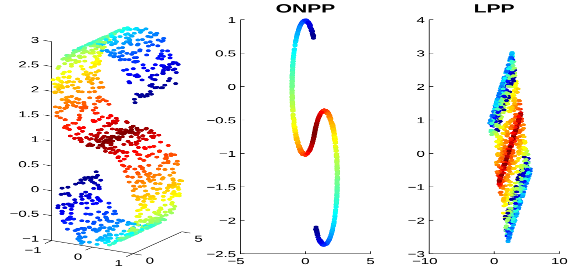

Fig. 1. Results of applying ONPP and LPP on the `s-curve`.

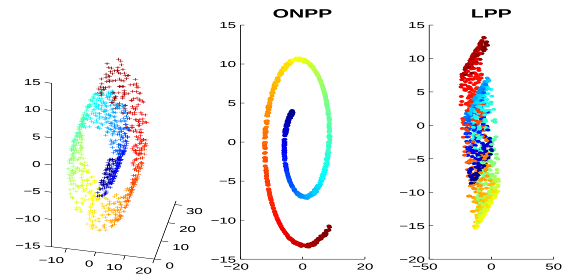

Fig. 2. Results of applying ONPP and LPP on the `swissroll`.

i.e., the data samples are projected on a line $y_i = v^\top x_i$. Then, observe that the optimization problem (4) in the feature space is formulated as follows

$$\min_v v^\top \Phi(X) M \Phi(X)^\top v,$$

where $v = \sum_{i=1}^n \alpha_i \Phi(x_i) = \Phi(X)\alpha$. Then, the above optimization problem is rewritten as follows

$$\min_\alpha \alpha^\top \Phi(X)^\top \Phi(X) M \Phi(X)^\top \Phi(X) \alpha = \min_\alpha \alpha^\top K M K \alpha. \tag{5}$$

Therefore, α should be the eigenvector of KMK associated with its smallest eigenvalue. For the general case of an arbitrary dimension d of the reduced space, we compute the eigenvectors $\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(d)}$ of KMK associated with its smallest eigenvalues. Then, the projection of a new point x_t is computed by computing its components along the $v^{(i)}$'s. In particular, the dot product of x_t with $v^{(i)}$ is computed as

$$\langle v^{(i)}, \Phi(x_t) \rangle = \langle \sum_{j=1}^n \alpha_j^{(i)} \Phi(x_j), \Phi(x_t) \rangle = \sum_{j=1}^n \alpha_j^{(i)} k(x_j, x_t).$$

Now consider again for simplicity the case $d = 1$. Concerning the training points, observe that the projections along the eigenvector v are given by $y = Ka$. Then, notice that the optimization problem (5) is rewritten as $\min_y y^\top M y$ which is exactly the eigenvalue problem solved by LLE. Therefore, LLE (nonlinear) could be viewed as performing ONPP (linear) implicitly in the feature space $\mathcal{H}$.

V. Discussion

Principal Components Analysis (PCA) computes the eigenvectors of the sample covariance matrix

$$C = (X - \mu e^\top)(X - \mu e^\top)^\top,$$

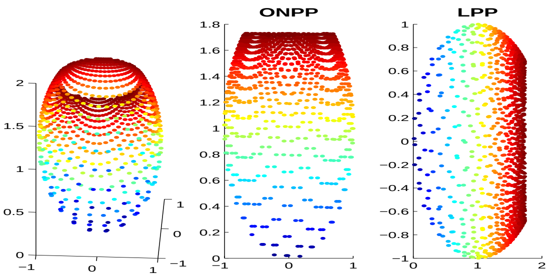

Fig. 3. Results of applying ONPP and LPP on the `punctured sphere`.

associated with its largest eigenvalues. In the above formula we used $\mu = \sum_{i=1}^{n} X_i$ which is the mean vector of the data. The data points are projected using the principal eigenvectors of C. PCA projects along the axes of maximum variance [13]. Therefore, PCA tends to preserve the global characteristics of the dataset. Recall that ONPP preserves the local geometries in each neighborhood. Notice that due to the overlapping neighborhoods, the combination of the local geometric information is expressed as preserving the global geometry as well. Moreover, the columns of V are orthonormal and this causes ONPP to tend to preserve the angles as much as possible. Therefore, one could argue that ONPP can be effectively used for data visualization purposes.

ONPP shares some of the properties of Locality Preserving Projections (LPP) [6], a recently proposed linear dimensionality reduction technique which aims at preserving the locality of the data samples. The locality information is quanti£ed by the following objective function

$$\sum_{ij} (y_i - y_j)^2 W_{ij}.$$

In particular, LPP employs the same objective function with Laplacian Eigenmaps [3] which is a nonlinear technique for dimensionality reduction. However, LPP is linear and employs an explicit linear mapping from $X \rightarrow Y$.

Both ONPP and LPP tend to preserve the locality of the data samples. However, ONPP is designed to preserve not only the locality (neighborhood graph) but also the local geometry. This is because the local geometric characteristics are captured by the optimal weights W_{ij} (see 3). In contrast, LPP does not consider the local geometries explicitly. Furthermore, the Gaussian weights used in LPP are somewhat arti£cial and do not necessarily re¤ect the underlying geometry. Recall also that a suitable value for the parameter σ, the width of the Gaussian envelope, must be determined. This parameter tuning is crucial for the performance of the algorithm.

Note also that ONPP is semi-automated since it only requires one parameter k, the number of nearest neighbors. Finally, note that the dimensionality reduction matrix V has orthonormal columns, and so ONPP amounts to an orthogonal projection of the data. In contrast, the dimensionality reduction matrix of LPP does not have orthonormal columns since these are the solutions of a generalized eigenproblem.

VI. Experimental Results

In this section we evaluate our algorithm in an unsupervised setting and compare it with PCA and LPP, which are linear dimensionality reduction methods. We use an implementation of LPP which is publicly available[1].

A. Synthetic data

We use three well known synthetic datasets from [11]: the `s-curve`, the `swissroll` and the `punctured sphere`. Figures 1 and 2 illustrate the two dimensional projections obtained by the ONPP and LPP methods in the `scurve` and `swissroll` datasets. We use $k = 12$ in both algorithms. Observe that both methods preserve locality and this is indicated by the color shading. However, ONPP preserves local and global geometric characteristics as well, since it gives a faithful projection which conveys information about how the manifold is folded in the high dimensional space.

Figure 3 illustrates the outcome of both algorithms on the `punctured sphere`. This dataset veri£es that although LPP preserves locality (illustrated by color shading) it seems to perturb the local geometries of the neighborhoods. On the other hand, ONPP respects not only the locality but the local geometries as well. This feature of ONPP is crucial for data visualization purposes.

B. Digit visualization

The next experiment involves digit visualization. We use 20×16 images of handwritten digits which are publically available from S. Roweis' web page[2]. The dataset contains 39 samples from each class (digits from '0'-'9'). Each digit image sample is represented lexicographically as a high dimensional vector of length 320. We project the dataset in the two dimensional space and the results are depicted in Figure 4. We use $k = 6$ for both ONPP and LPP. The left panels corresponds to digits '0'-'4' and the right panels corresponds to digits '5'-'9'.

Observe that the projections of PCA are spread out since PCA aims at maximizing the variance. However, the classes of different digits seem to heavily overlap. This means that PCA does not do well at discriminating between data. On the other hand, observe that ONPP and LPP yield more meaningful projections since samples of the same class are mapped close to each other. This is because these methods aim at preserving the locality. Finally, ONPP seems to provide slightly better projections than LPP since the clusters of the former appear more cohesive.

C. Face recognition

In this experiment we employ ONPP and LPP in a supervised setting (see Section III for more details) for face recognition. Note that in this case we employ LPP with Gaussian weights. We determined the value of the width σ of the Gaussian envelope as follows. First, we sample 1000 points

[1] `http://people.cs.uchicago.edu/~xiaofei/LPP.m`

[2] `http://www.cs.toronto.edu/~roweis/data.html`

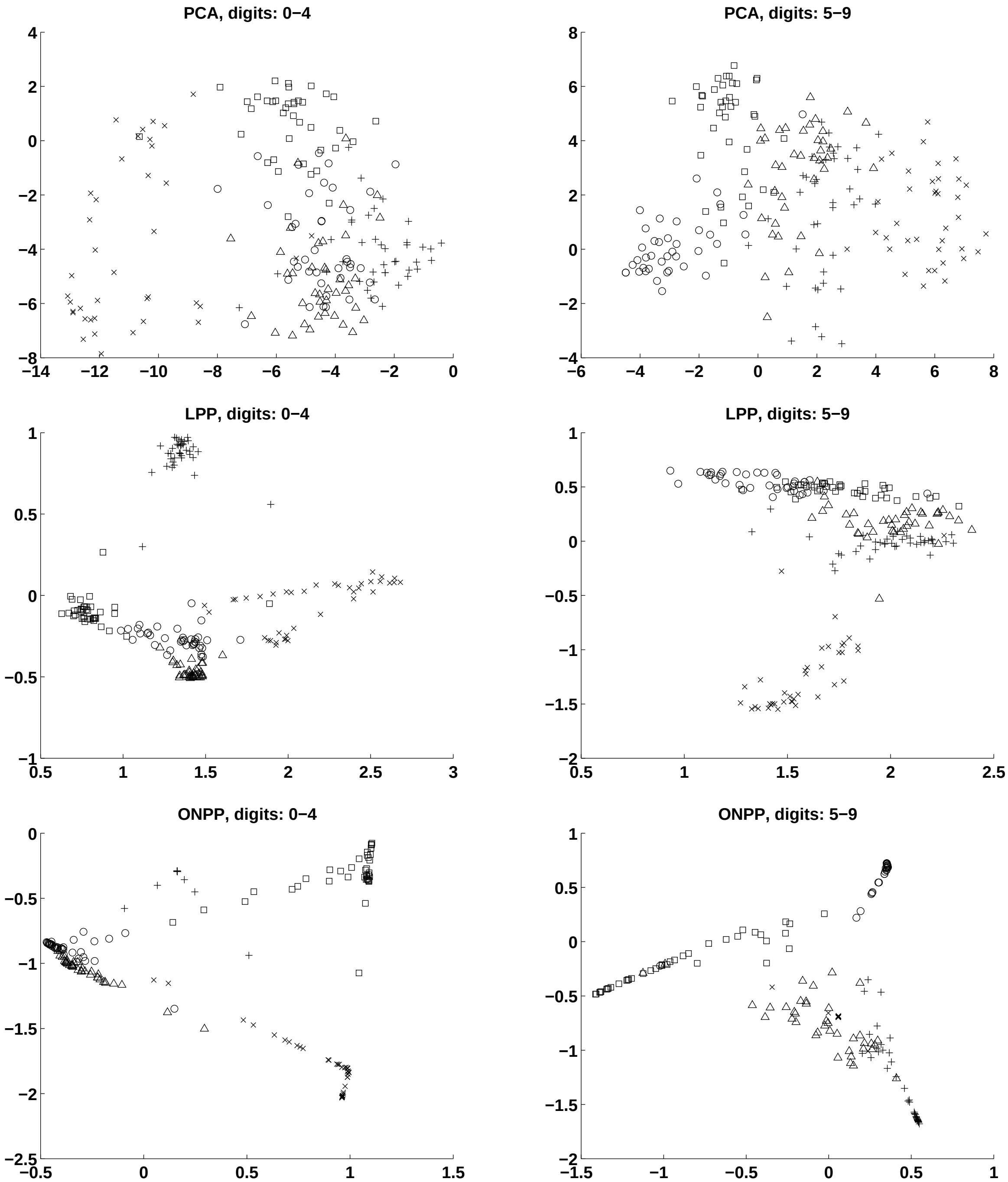

Fig. 4. Two dimensional projections of digits. Left panel: '+' denotes 0, 'x' denotes 1, 'o' denotes 2, '△' denotes 3 and '□' denotes 4. Right panel: '+' denotes 5, 'x' denotes 6, 'o' denotes 7, '△' denotes 8 and '□' denotes 9. Top panels: PCA, middle: LPP and bottom: ONPP.

randomly and then compute the pairwise distance among them. Then σ is set equal to half the median of those pairwise distances. This gives a good and reasonable estimate for the value of σ.

We use a subset of the AR face database [8] which contains 126 subjects under 8 different facial expressions and variable lighting conditions for each individual. We form the training set by a random subset of 4 different facial expressions/poses per subject and use the remaining 4 as a test set. We plot the average error rate (%) across 20 random realizations of the training/test set, for $d = [30 : 10 : 100]$ (in MATLAB notation).

The results are illustrated in Figure 5. Observe that ONPP outperforms the other methods. Furthermore, Table II reports the best achieved error rate by each method and the corresponding dimension d of the reduced space. It appears that PCA (unsupervised) has inferior performance versus the other methods which are supervised. Notice that ONPP is the best of all the methods tested, with a clear margin of superiority.

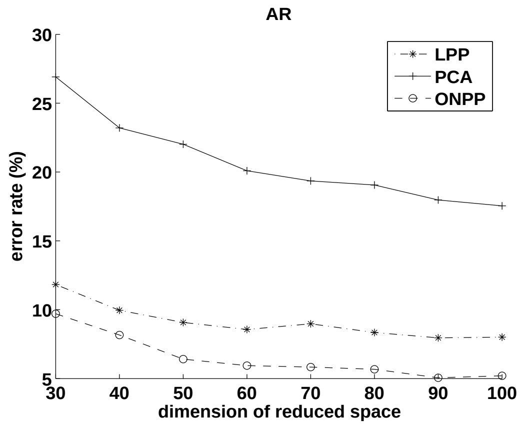

Fig. 5. Error rate with respect to the reduced dimension d on the AR database.

	dim	error (%)
PCA	100	17.53
LPP	90	7.96
ONPP	90	**5.05**

TABLE II

THE BEST ERROR RATE ACHIEVED BY ALL METHODS ON THE AR DATABASE.

Additionally, recall that the supervised version of ONPP is fully automated, since the number of nearest neighbors k is determined by the class membership and the user need not set any parameters. In contrast, LPP with Gaussian weights requires to determine the parameter σ.

D. Tissue visualization using gene expression data

The evolution of DNA microarray technology has resulted in an abundance of gene expression data. Those data sets need to be analyzed in a meaningful way to improve our understanding and knowledge of the functional role of genes in the development of diseases and malignancies. Each entry of the DNA microarray hybridization experiment is the ratio of the expression level of a particular gene under two different conditions. Usually, the numerator measures the expression level of a certain gene at the condition of interest and the denominator corresponds to the reference condition. The results of n DNA microarray experiments across m genes, is a $m \times n$ data matrix whose rows correspond to the expression levels of a particular gene and each column corresponds to a speci£c tissue sample experiment.

The goal of the experiments in this section is to use gene expression data for tissue visualization using unsupervised ONPP and LPP. We use a subset of the datasets in [14] for tissue classi£cation. Table III summarizes the main characteristics of the datasets. Details on each of these are provided next.

Dataset	No of tissues	No of genes
Leukemia	72	7129
Lymphoma	62	4026
Colon	62	2000

TABLE III

MAIN CHARACTERISTICS OF THE GENE EXPRESSION DATASETS.

Leukemia: The **Leukemia** dataset [5] consists of two different types of acute leukemias: acute lymphoblastic leukemia (ALL) and acute myeloblastic leukemia (AML). It is available at `http://www.genome.wi.mit.edu/MPR`. The dataset contains 72 leukemia tissue samples of which 47 are ALL samples and 25 are AML samples. The number of genes included at the microchip is 7129. We employ both ONPP and LPP with $k = 6$ and illustrate the two dimensional projections of all tissue samples in Figure 6. Observe that in ONPP the tissues of the same type of leukemia cluster together and the clusters are cohesive. On the other hand, in the case of LPP the ALL cluster seems to be separated in two parts.

Lymphoma: The **Lymphoma** dataset [1] comes from a study of discrimination among different types of non-Hodgkins lymphoma using gene expression data. It is available at `http://genome-www.stanford.edu/lymphoma`. There are three classes of lymphoma: B-cell chronic lymphocytic leukemia (CLL), follicular lymphoma (FL) and diffuse large B-cell lymphoma (DLCL). In particular, there 11 samples of CLL-type, 9 samples of FL-type and 42 samples of DLCL-type lymphoma. The number of genes used in the 'Lymphochip' is 4026. We employ ONPP and LPP with $k = 8$ and project the tissue samples on a two dimensional space. The resulting projections are shown at Figure 7. Observe that although both methods are used in unsupervised mode, the tissues of the same type cluster together in the reduced space. However, in ONPP the FL-type lymphoma samples seem to be spread out and this may be due to the geometry of those samples in the high dimensional space. Additionally, the observed elongated cluster of the CLL-type lymphoma samples may convey information about how those samples are arranged in the high dimensional space.

Colon: The **Colon** dataset [2] consists of colon tissue samples. In particular, it contains 40 tumor and 22 normal colon tissues. It is available at `http://microarray.princeton.edu/oncology`. The total number of genes included at the microchip is 2000. We employ ONPP and LPP with $k = 6$ and illustrate the two dimensional projections of all tissue samples in Figure 8. Observe that ONPP respects the intrinsic geometries in the local neighborhoods in the higher dimensional space. On the other hand, LPP appears to disturb the local geometries. Furthermore, the projections of this dataset suggest that ONPP respects the class information of the tissue samples as well.

E. Face manifold visualization

In the last experiment we use a collection of face images from the ISOMAP [7] web-page[3]. The face images contain a virtual face under different degrees of freedom including rotation/pose and lighting/shading. There are 698 facial images of size 64×64. In the pre-processing step each face image is downsampled to 32×32 to reduce computational complexity. We then use ONPP with $k = 12$ to project the high

[3] `http://isomap.stanford.edu/datasets.html`

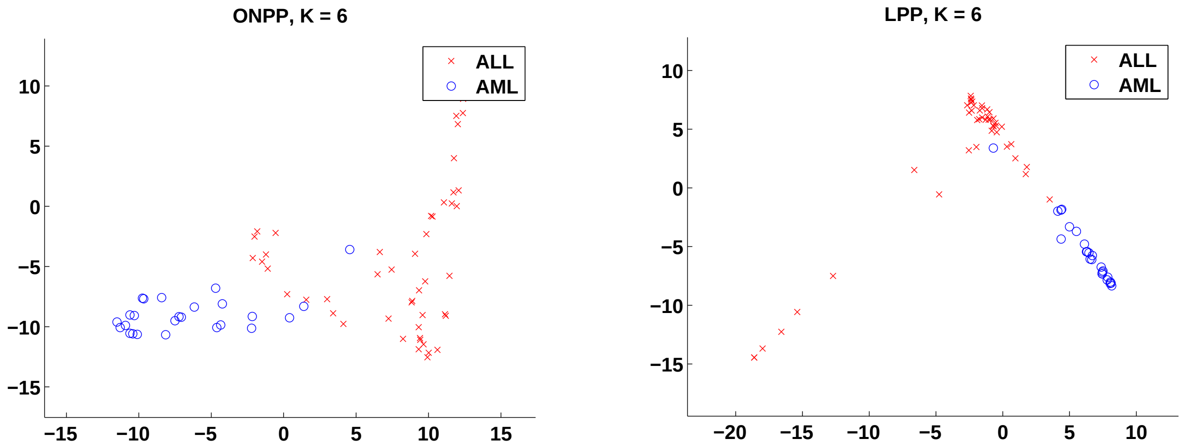

Fig. 6. Two dimensional projections of the **Leukemia** tissue samples using ONPP (left panel) and LPP (right panel).

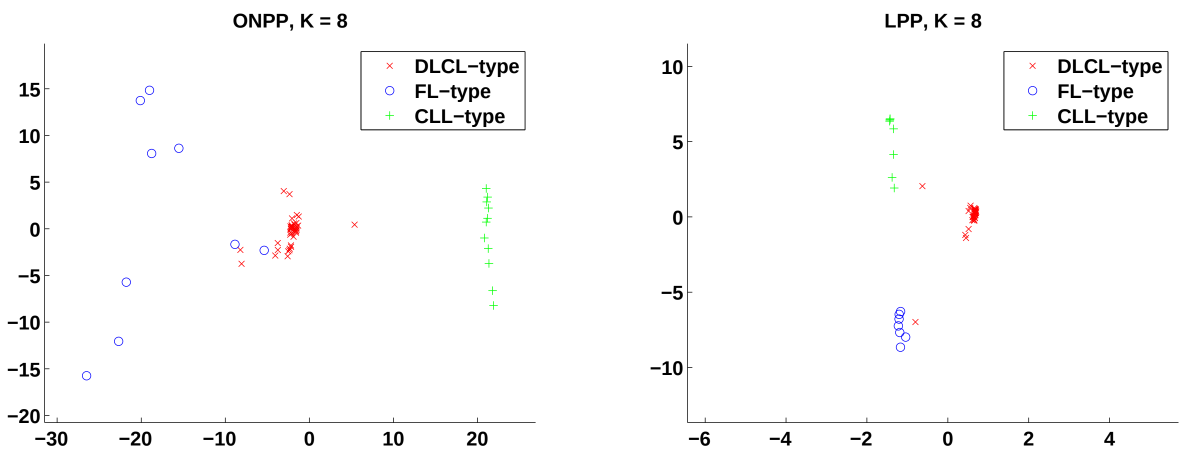

Fig. 7. Two dimensional projections of the **Lymphoma** tissue samples using ONPP (left panel) and LPP (right panel).

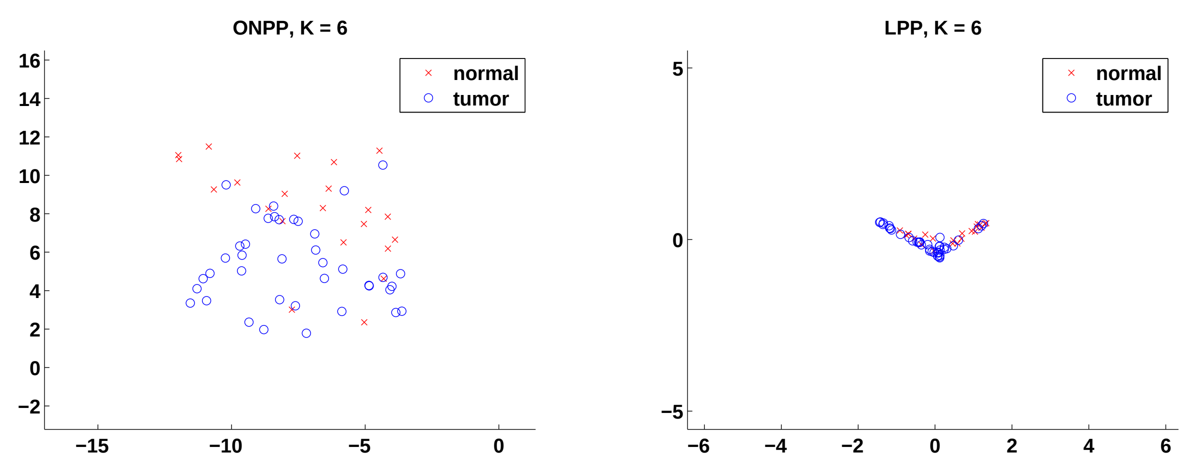

Fig. 8. Two dimensional projections of the **Colon** tissue samples using ONPP (left panel) and LPP (right panel).

dimensional facial images in two dimensional space. Figure 9 illustrates the two dimensional projection obtained. Next to some of the projected samples we place the corresponding facial image. Observe that the images are organized in two dimensional space according to the intrinsic degrees of freedom (underlying pose and lighting condition). For instance, facial

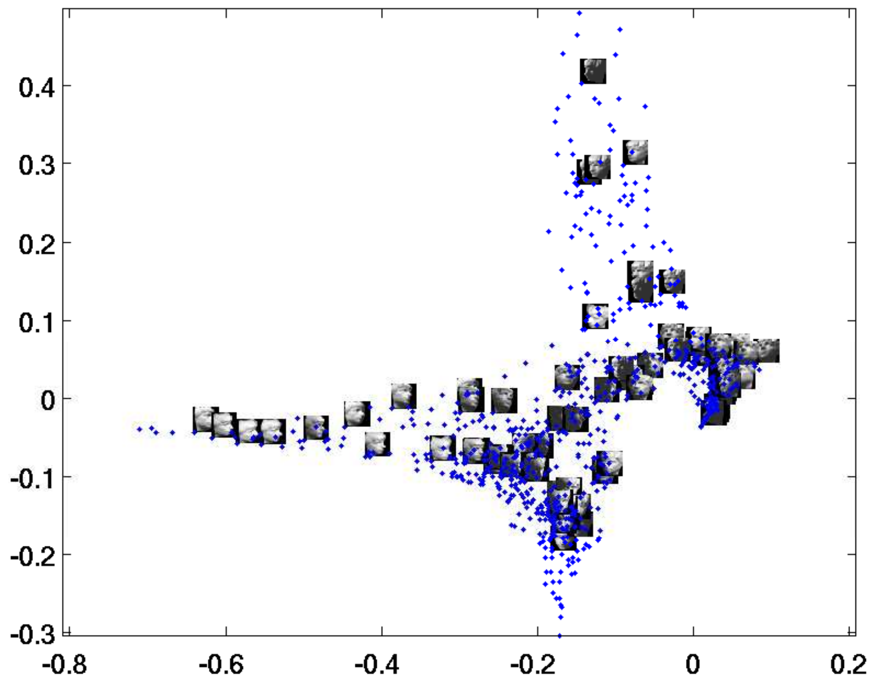

Fig. 9. Two dimensional projections of the face manifold using ONPP with $k = 12$ nearest neighbors.

images with similar head orientation are projected close by.

VII. Conclusion

The Orthogonal Neighborhood Preserving Projections (ONPP) introduced in this paper is a linear dimensionality reduction technique, which will tend to preserve not only the locality but also the local and global geometry of the high dimensional data samples. It can be extended to a supervised method and it can also be combined with kernel techniques. We showed that ONPP can be very effective for data visualization, and that it can be implemented in a supervised setting to yield a robust recognition technique. Experimental results spanning a few application areas from pattern recognition, computational biology and computer vision demonstrated the versatility and effectiveness of the proposed scheme.

Acknowledgements: We are grateful to Prof. D. Boley for his valuable help and insightful discussions on various aspects of the paper. We would also like to thank Dr. J. Ye for providing us with the gene expression data. This work was supported by NSF grant DMS 0510131 and by the Minnesota Supercomputer Institute.

References

[1] A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. H. Jr, L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, and L. M. Staudt. Distinct types of diffuse large b-cell lymphoma identi£ed by gene expression pro£ling. *Nature*, 403:503–511, 2000.

[2] U. Alon, N. Barkai, D. Notterman, K. Gish, S. Ybarra, D. Mack, and A. Levine. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Science*, 96:6745–6750, 1999.

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

[4] Y. Bengio, J-F Paiement, P. Vincent, O. Delalleau, N. Le Roux, and M. Ouimet. Out-of-Sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering. In Sebastian Thrun, Lawrence Saul, and Bernhard Schölkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

[5] S. Dudoit, J. Fridlyand, and T. Speed. Comparison of discrimination methods for the classi£cation of tumors using gene expression data. *Journal of the American Statistical Association*, 97:77–87, 2002.

[6] X. He and P. Niyogi. Locality preserving projections. *Advances in Neural Information Processing Systems 16 (NIPS 2003)*, 2003. Vancouver, Canada.

[7] V. de Silva J. B. Tenenbaum and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[8] A.M. Martinez and R. Benavente. The AR Face Database. Technical report, CVC no. 24, 1998.

[9] S. Roweis and L. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290:2323–2326, 2000.

[10] Y. Saad. *Numerical Methods for Large Eigenvalue Problems*. Halstead Press, New York, 1992.

[11] L. Saul and S. Roweis. Think Globally, Fit Locally: Unsupervised Learning of Nonlinear Manifolds. *Journal of Machine Learning Research*, 4:119–155, 2003.

[12] V. Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.

[13] A. Webb. *Statistical Pattern Recognition*. Wiley, 2nd edition, 2002.

[14] J. Ye, T. Li, T. Xiong, and R. Janardan. Using uncorrelated discriminant analysis for tissue classi£cation with gene expression data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1(4):181–190, 2004.