A New Orthogonalization of Locality Preserving Projection and Applications

Gitam Shikkenawis[1,⋆], Suman K. Mitra, and Ajit Rajwade[2]

[1] Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, Gujarat, India
[2] Indian Institute of Technology Bombay, Mumbai, Maharashtra, India
{201221004,suman_mitra}@daiict.ac.in, ajitvr@cse.iitb.ac.in

Abstract. Locality Preserving Projection (LPP) is a linear projection method that preserves the local structure to find the underlying manifold of the data. Non-orthogonality of LPP basis makes its use difficult not only for reconstruction but also for other applications such as denoising. At present, orthogonal basis of LPP (OLPP) are obtained in an iterative manner which is computationally expensive. In this article, a new orthogonalization of LPP (NOLPP) basis is proposed by relaxing the constraint used to minimize the objective function giving rise to the basis. The reducibility capacity of NOLPP for data clustering is validated by performing experiments on several databases. Use of NOLPP for image denoising shows its efficiency in comparison to the state of the art research. Fine structures present in the images are preserved even at high noise levels.

Keywords: Locality Preserving Projection, Dimensionality Reduction, Image Denoising.

1 Introduction

Data analysis and representation in a transformed domain has fetched a lot of attention. Many linear and non-linear methods have been proposed that try to learn the manifold structure from the given set of observations. Principle components analysis (PCA) [11], Linear Discriminant Analysis (LDA) [3] and Independent Component Analysis (ICA) [10] are the most widely used linear methods that learn the basis preserving global structure of the data. Locality Preserving Projection (LPP)[9], added in last decade, tries to preserve the local structure using the neighbourhood information. LPP is a linear method, but due to the characteristics of preserving the local structure, it captures the non-linearity present in the data. Variants of LPP [12], [6], [7] have also been proposed in recent years. Though, conventional LPP and its variants well preserve the local manifold structure, the basis vectors learnt are not orthogonal. Having orthonormal basis is advantageous in many applications and it makes the data

⋆ The author is acknowledging Tata Consultancy Services for providing the financial support for carrying out this work.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 277–283, 2013.

reconstruction much simpler. Orthogonal LPP [4] orthogonalized the basis of LPP in an iterative manner by selecting a basis vector orthogonal to previously learnt basis vectors. In this article, we are proposing a new orthogonalization of LPP (NOLPP) which learns orthogonal basis solving the eigen value problem only once.

LPP has been applied in various applications of dimensionality reduction, data classification and clustering, recognition tasks as well as document representation and indexing. Experiments of data clustering using the basis of currently proposed NOLPP have been performed. A new application of NOLPP in the area of Image Denoising has also been explored. Patches from the noisy image are used to learn the NOLPP basis and the denoising is performed in the transformed domain. As mentioned earlier, inverse transformation of the denoised patches into the spatial domain becomes much simpler because of the orthogonal property.

A brief discussion about conventional LPP, OLPP and the proposed NOLPP is included in Section 2. Section 3 presents clustering results on various datasets in using much less dimensions in the NOLPP domain. A new application of LPP for image denoising is covered in Section 4 along with the concluding remarks in Section 5.

2 Locality Preserving Projection

The objective function of LPP ensures closeness of the neighbouring data points in the projection domain by assigning them higher weights in the weight matrix $\mathbf{W}$. $\mathbf{x_i}$'s represent the data points in vector format and $\mathbf{y_i} = \mathbf{a^T x_i}$ are their projections on the learned basis $\mathbf{a}$ as follows:
$min \sum_{ij} \|\mathbf{y_i} - \mathbf{y_j}\|^2 \mathbf{W_{ij}} = min \sum_{ij} \left(\mathbf{a^T x_i} - \mathbf{a^T x_j}\right)^2 \mathbf{W_{ij}} = min\ \mathbf{a^T X L X^T a}$
Here, $\mathbf{X}$ is the matrix containing the data points $(\mathbf{x_i}'s)$ in columns, $\mathbf{D} = \sum_i \mathbf{W_{ij}}$ is a diagonal matrix. $\mathbf{D} - \mathbf{W}$ is called the laplacian matrix i.e. $\mathbf{L}$. The weighing matrix $\mathbf{W}$ is computed as : $\mathbf{W_{ij}} \equiv e^{-\frac{\|\mathbf{x_i} - \mathbf{x_j}\|^2}{t}}$, if $\mathbf{x_i}$ belongs to k-NN of $\mathbf{x_j}$. Here, t controls the width of the Gaussian kernel. A normalization constraint $\mathbf{a^T X D X^T a} = \mathbf{1}$ is imposed on the given objective function which results in solving the generalized eigenvalue solution of the $\mathbf{XLX^T a} = \lambda \mathbf{XDX^T a}$. Eigenvectors corresponding to the minimum eigenvalues are the strongest basis vectors for the given dataset. Basis vectors learnt using LPP are not orthonormal. There are many applications of the projection based approaches where orthonormal basis vectors are more advantageous. Also, reconstruction of the data becomes difficult using non-orthonormal basis.

Orthonormal Locality Preserving Projection (OPLL) [4] imposed the orthonormality constraint on the basis vectors in the conventional LPP framework. The objective function now becomes, min $\mathbf{a^T XLX^T a}$ subject to the constraints $\mathbf{a^T XDX^T a} = \mathbf{1}$ and $\mathbf{a_m^T a_k} = 0, \forall m = \{1, 2, ..., k-1\}$.

The basis vectors $\mathbf{a_k}$s are found in the iterative manner assuring that the vector learnt is orthonormal to all the previously found basis vectors. $\mathbf{a_1}$ is the eigenvector associated with the smallest non-zero eigenvalue of $\mathbf{M^{(1)}} = \left(\mathbf{XDX^T}\right)^{-1} \mathbf{XLX^T}$. $\mathbf{a_k}, \forall k$ are computed as follows:

$$\mathbf{M}^{(k)} = \left\{ \mathbf{I} - \left(\mathbf{XDX^T}\right)^{-1} \mathbf{A}^{(k-1)} \cdot \left[\mathbf{B}^{(k-1)}\right]^{-1} \left[\mathbf{A}^{(k-1)}\right]^{T} \right\} \mathbf{M}^{(\mathbf{1})} \tag{1}$$

$\mathbf{a_k}$ is the eigenvector associated with the smallest eigenvalue of $\mathbf{M^{(k)}}$.

$$\mathbf{A}^{(k-1)} = [\mathbf{a_1}, \ldots, \mathbf{a_{k-1}}];\ \mathbf{B}^{(k-1)} = \left[\mathbf{A}^{(k-1)}\right]^{T} \left(\mathbf{XDX^T}\right)^{-1} \mathbf{A}^{(k-1)} \tag{2}$$

The basis matrix **a** thus found is orthonormal. Though, we get the orthonormal basis vectors, the method described above is computationally very expensive.

2.1 A New Orthogonalization of LPP

The fundamental need for the basis vectors to be orthogonal is $\mathbf{a^T a} = I$. Hence, in order to have orthogonal basis vectors, the constraint $\mathbf{a^T a} = I$ should be satisfied. We introduce this constraint while minimizing the objective function of conventional LPP as opposed to $\mathbf{a^T XDX^T a} = 1$. The minimization problem now reduces to the eigenvalue solution of $\mathbf{XLX^T a} = \lambda \mathbf{a}$. The solution is no more the generalized eigenvalue problem and produces orthonormal basis vectors directly in the first step, hence is computationally efficient. The computation of the weight matrix is carried out as per Extended LPP [12]. The weights are assigned using a z-shaped weighing function 3 that weighs the data points a monotonically decreasing fashion according to the Euclidean distance between them assuring their closeness in the projection domain as well as adding more discriminating power than the conventional LPP [12]. l and u are the lower and upper limits of the distance range within which the function changes its values.

The weighing function specified in Eqn. 3 is used throughout the article for experiments. l and u are the lower and upper limits of the distance range within which the function changes its values otherwise 1 and 0 values are directly assigned.

$$\mathbf{W}_{ij} \equiv \left\{ \begin{array}{ll} 1 - 2\left(\frac{\|\mathbf{x_i}-\mathbf{x_j}\|^2 - l}{u-l}\right)^2; & \text{if } l \leq \|\mathbf{x_i} - \mathbf{x_j}\|^2 \leq \frac{l+u}{2} \\ 2\left(\frac{\|\mathbf{x_i}-\mathbf{x_j}\|^2 - u}{l-u}\right)^2; & \text{if } \frac{l+u}{2} \leq \|\mathbf{x_i} - \mathbf{x_j}\|^2 \leq u \end{array} \right\} \tag{3}$$

3 Data Dimensionality Reduction and Clustering

Dimensionality Reduction is the key application of LPP and other such projection techniques. Several experiments have been performed on various datasets to test the data discrimination capacity of the proposed approach with very few dimensions. The results with varying no. dimensions (maximum 100) are included in order to show the capability of discrimination of the proposed approach with very less dimensions. Current approach is compared with the conventional LPP and one of its variants.

In case of data discrimination and clustering, ambiguities arise in the structurally similar data points. Sets of structurally similar data points from various

Table 1. Results (Errors) of clustering on various databases using LPP, ELPP and proposed NOLPP

Data	# Dimensions Used											
	2			10			50			100		
	LPP	ELPP	NO-LPP	LPP	ELPP	NO-LPP	LPP	ELPP	NO-LPP	LPP	ELPP	NO-LPP
Digits 3-8	42.2	**17.0**	18.75	52.0	**14.5**	**14.5**	45.7	10.5	**4**	51.2	9.75	**3.25**
Digits 1-7	61.2	10.2	**3.25**	13.5	9.25	**3.25**	5.5	3.0	**1**	4.25	2.75	**1**
YALE Per 1,5	8.33	11.11	**6.94**	6.94	5.55	**2.78**	**0**	**0**	**0**	**0**	**0**	**0**
YALE Per 3,4	26.38	11.11	**8.33**	16.66	5.55	**2.77**	**1.38**	2.77	2.77	**1.38**	2.77	2.77
Person 1	72.47	3.89	**0.61**	71.98	0.85	**0**	70.52	**0**	**0**	67.11	**0**	**0**
Person 2	71.80	0.73	**0**	73.92	**0**	**0**	69.24	**0**	**0**	70.06	**0**	**0**

databases are selected for the experiments. The data points are projected on the learnt basis using LPP, ELPP and the current proposal and clustered using nearest neighbour approach. Digit pairs $3-8$ and $1-7$ from the MNIST database of Handwritten Digits [1] are chosen as they share similar structures. Two pairs of persons from the YALE Face DatabaseB [8] having resemblance are selected. Also, it is difficult to discriminate various expressions of a person, hence results of clustering four different facial expressions of two persons from the Video database [12] are shown in Table 1.

4 Image Denoising : A New Application

A new application of LPP has been explored for Image Denoising tasks. The basic noisy image formation model is $\mathbf{I}_\eta = \mathbf{I} + \eta$ where $\mathbf{I}$ is the original image, η is i.i.d. Gaussian noise with standard deviation σ and $\mathbf{I}_\eta$ is the noisy image. The procedure of removing noise is carried out in the projection domain of new Orthogonal LPP. Image is processed in the moving window fashion, considering a patch of size $n \times n$ at a time.

The noisy image formation model in the transformed domain using basis $\mathbf{a}$ can be given as, $\mathbf{I}_{\eta\mathbf{a}} = \mathbf{I}_\mathbf{a} + \eta_\mathbf{a}$. In order to remove noise, the coefficients of the noisy image in transformed domain $\mathbf{I}_{\eta\mathbf{a}}$ are modified using shrinkage rule such as Soft thresholding, Hard thresholding and Weiner filter update. Weiner filter update, $\frac{\sigma^2_{I_{\eta a}} - \sigma^2}{\sigma^2_{I_{\eta a}}}\mathbf{I}_{\eta\mathbf{a}}$ is used for all the experiments in this article as it produced the best results. Here, σ is the standard deviation of noise and $\sigma_{I_{\eta a}}$ is computed from the degraded image.

4.1 Experimental Results

The experiments are performed on greyscale texture and natural image databases. For all the experiments, the images are divided into overlapping patches of size 6×6. Each patch is considered as a data point for learning the NOLPP basis. For computation of weight matrix as per [12], the distance

Table 2. Denoising results - PSNR (SSIM): UIUC Texture Database ($\sigma \in \{20, 30, 40\}$)

		LPG-PCA1	LPG-PCA2	BM3D1	BM3D2	NOLPP
Bark	$\sigma = 20$	25.66 (0.79)	25.53 (0.78)	25.55 (0.78)	25.79 (0.79)	**26.04** (**0.82**)
	$\sigma = 30$	23.71 (0.68)	23.58 (0.66)	23.70 (0.67)	23.88 (0.68)	**24.26** (**0.74**)
	$\sigma = 40$	22.56 (0.61)	22.46 (0.58)	22.25 (0.58)	22.73 (0.61)	**23.08** (**0.67**)
Water	$\sigma = 20$	28.23 (0.76)	28.20 (0.76)	28.24 (0.76)	28.56 (0.78)	**28.96** (**0.80**)
	$\sigma = 30$	26.24 (0.65)	26.32 (0.66)	26.46 (0.66)	26.87 (0.69)	**27.17** (**0.72**)
	$\sigma = 40$	24.91 (0.56)	25.18 (0.58)	25.19 (0.58)	25.76 (0.63)	**25.91** (**0.65**)
Pebbles	$\sigma = 20$	26.81 (0.79)	26.75 (0.79)	26.70 (0.78)	27.05 (0.80)	**28.14** (**0.85**)
	$\sigma = 30$	24.74 (0.69)	24.70 (0.68)	24.82 (0.68)	25.15 (0.71)	**26.16** (**0.78**)
	$\sigma = 40$	23.44 (0.60)	23.46 (0.60)	23.35 (0.59)	23.92 (0.63)	**24.87** (**0.71**)
Brick	$\sigma = 20$	26.16 (0.80)	26.09 (0.79)	26.04 (0.79)	**26.28** (0.80)	26.02 (**0.82**)
	$\sigma = 30$	24.14 (0.70)	24.08 (0.69)	24.21 (0.70)	**24.43** (0.71)	24.17 (**0.72**)
	$\sigma = 40$	22.88 (0.63)	22.85 (0.61)	22.69 (0.61)	**23.23** (0.63)	23.00 (**0.65**)
Fur	$\sigma = 20$	28.36 (0.80)	28.41 (0.80)	30.63 (0.82)	**31.35** (**0.86**)	31.24 (0.85)
	$\sigma = 30$	26.18 (0.68)	26.31 (0.69)	28.45 (0.73)	**29.31** (**0.79**)	29.22 (0.78)
	$\sigma = 40$	24.75 (0.59)	25.04 (0.60)	26.97 (0.65)	**27.98** (**0.73**)	27.72 (0.72)

Table 3. Denoising Results - PSNR (SSIM): Natural Images ($\sigma \in \{20, 30, 40\}$)

		LPG-PCA1	LPG-PCA2	BM3D1	BM3D2	NOLPP
Barbara	$\sigma = 20$	30.85 (0.86)	31.42 (0.89)	31.12 (0.88)	**31.74** (**0.90**)	30.46 (0.87)
	$\sigma = 30$	28.43 (0.77)	29.22 (0.85)	28.95 (0.82)	**29.81** (**0.86**)	28.32 (0.81)
	$\sigma = 40$	26.57 (0.68)	27.55 (0.81)	26.74 (0.75)	**27.98** (**0.82**)	26.81 (0.75)
Goldhill	$\sigma = 20$	30.06 (0.77)	30.22 (0.78)	30.40 (0.79)	**30.70** (**0.80**)	30.30 (0.79)
	$\sigma = 30$	28.02 (0.67)	28.51 (0.72)	28.67 (0.72)	**29.13** (**0.75**)	28.50 (0.72)
	$\sigma = 40$	26.38 (0.58)	27.23 (0.67)	27.21 (0.65)	**27.93** (**0.70**)	27.80 (0.66)
Lena	$\sigma = 20$	31.84 (0.81)	32.61 (0.87)	32.45 (0.85)	**32.98** (**0.87**)	31.91 (0.83)
	$\sigma = 30$	29.46 (0.71)	30.71 (0.83)	30.49 (0.80)	**31.25** (**0.84**)	29.91 (0.78)
	$\sigma = 40$	27.51 (0.61)	29.14 (0.80)	28.64 (0.74)	**29.79** (**0.81**)	28.43 (0.72)
Mandrill	$\sigma = 20$	26.52 (0.78)	26.42 (0.77)	26.28 (0.77)	26.60 (0.79)	**26.61** (**0.81**)
	$\sigma = 30$	24.26 (0.67)	24.20 (0.67)	24.22 (0.67)	24.56 (0.70)	**24.61** (**0.71**)
	$\sigma = 40$	22.93 (0.58)	22.92 (0.59)	22.44 (0.57)	23.11 (0.61)	**23.31** (**0.64**)
Stream	$\sigma = 20$	27.05 (0.77)	26.96 (0.77)	27.01 (0.77)	27.27 (0.78)	**27.30** (**0.79**)
	$\sigma = 30$	25.08 (0.67)	25.04 (0.67)	25.23 (0.68)	25.47 (0.69)	**25.50** (**0.71**)
	$\sigma = 40$	23.78 (0.58)	23.83 (0.59)	23.82 (0.59)	24.29 (0.62)	**24.34** (**0.64**)

range of similar patches is set to be $\sqrt{3\sigma^2 n^2}$ which is the widely used similarity measure for patches. Up to the specified distance, the patches are considered to be similar and are weighed accordingly, all other patches are weighed 0. Instead of learning different NOLPP basis for each overlapping patch, only one set of basis are learnt for windows of size 120×120 from the input noisy images of size 512×512. Each noisy patch is then projected on the corresponding NOLPP basis, and the coefficients are modified using the Weiner filter update.

The results of the proposed method are compared with some of the state of the art methods for denoising such as LPG-PCA [13] and BM3D [5]. Both the methods work in 2 stages, hence denoising performances in terms of PSNR and SSIM of both the stages are included in Table 2 and 3 along with the results of the proposed approach for some of the images from UCIC Texture database [2] and Lansel database of natural images respectively for 3 different noise levels i.e. $\sigma = 20,\ 30\ \&\ 40$. Proposed approach out performs both LPG-PCA and BM3D in most of the cases for texture images as these approaches tend to smooth out the fine textural details , whereas in case of natural images, proposed approach is comparable i.e. the finer details are still preserved but it under smoothes some homogeneous parts of the image (Figure 1).

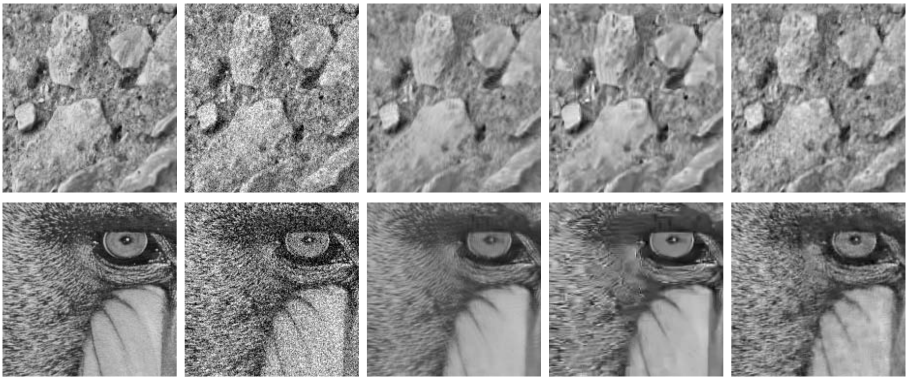

Fig. 1. From left to right, top to bottom: Original texture image, Noisy image with i.i.d. Gaussian noise ($\sigma = 30$), Results of denoising using LPG-PCA, BM3D, NOLPP. The same order for a natural image with i.i.d Gaussian noise ($\sigma = 40$) in the next row.

5 Conclusion

Locality Preserving Projection (LPP) though preserves the local structure, which could be essential for many applications, yet yields non-orthogonal basis that may not be acceptable for applications such as image denoising. This paper presented a method of obtaining orthogonal LPP basis. The application of such basis for pattern clustering and image denoising has been presented. It shows a strong reducibility capacity of data dimension which is effectively utilized in clustering. The denoising mechanism presented is almost outperforming most of the state of the art methods of denoising for textured images while comparable in case of the natural images. Denoising of natural images having some texture is appeared to be the best. This research will extend the applicability of New Orthogonal LPP towards other problems such as image deblurring and inpainting.

References

1. Mnist database of handwritten digits, `http://yann.lecun.com/exdb/mnist/`
2. Ucic texture database,
   `http://wwwcvr.ai.uiuc.edu/ponce_grp/data/texture_database/samples/`
3. Belhumeur, P., Hespanha, J., Kriengman, D.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. Pattern Analysis Machine Intelligence 19(13), 711–720 (1997)
4. Cai, D., He, X., Han, J., Zhang, H.: Orthogonal laplacianfaces for face recognition. IEEE Transactions on Image Processing 15(11), 3608–3614 (2006)
5. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3d transform domain collaborative filtering. IEEE Trans. on Image Processing 16(8), 2080–2095 (2007)
6. Dornaika, F., Assoum, A.: Enhanced and parameterless locality preserving projections for face recognition. Neurocomputing 99, 448–457 (2013)
7. Yu, J.W.G., Peng, H., Ma, Q.: Enhanced locality preserving projections using robust path based similarity. Neurocomputing 71, 598–605 (2011)
8. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intelligence 23(6), 643–660 (2001)
9. He, X., Niyogi, P.: Locality preserving projections. In: NIPS (2003)
10. Hyvarinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley, New York (2001)
11. Jolliffe, I.: Principal Component Analysis. Springer, Berlin (2002)
12. Shikkenawis, G., Mitra, S.K.: Improving the locality preserving projection for dimensionality reduction. In: International Conference on Emerging Applications of Information Technology, pp. 161–164. IEEE Computer Society (2012)
13. Zhang, L., Dong, W., Zhang, D., Shi, G.: Two-stage image denoising by principal component analysis with local pixel grouping. Pattern Recognition 43(4), 1531–1549 (2010)