Orthogonal Neighborhood Preserving Projection using L1-norm Minimization

Purvi A. Koringa and Suman K. Mitra
Dhirubhai Ambami Institute of Information and Communication Technology, Gandhinagar, Gujarat, India
{*201321010, suman_mitra*}*@daiict.ac.in*

Keywords: L1-norm, L2-norm, Outliers, Dimensionality Reduction.

Abstract: Subspace analysis or dimensionality reduction techniques are becoming very popular for many computer vision tasks including face recognition or in general image recognition. Most of such techniques deal with optimizing a cost function using L2-norm. However, recently, due to capability of handling outliers, optimizing such cost function using L1-norm is drawing the attention of researchers. Present work is the first attempt towards the same goal where Orthogonal Neighbourhood Preserving Projection (ONPP) technique is optimized using L1-norm. In particular the relation of ONPP and PCA is established in the light of L2-norm and then ONPP is optimized using an already proposed mechanism of L1-PCA. Extensive experiments are performed on synthetic as well as real data. It has been observed that L1-ONPP outperforms its counterpart L2-ONPP.

1 INTRODUCTION

Images are very high dimensional data which poses many challenges while handling them in fields like computer vision, machine learning etc. Though, image seems to be high dimensional data, it is observed that it lies in comparatively very low linear or non-linear manifold (He et al., 2005), (Kokiopoulou and Saad, 2007). This leads to the development of data dimensionality reduction techniques. The fundamental idea is to seek a linear or non-linear transformation to map the high dimensional data to a lower dimensional subspace which makes the same class of data more compact. This leads to favorable outcomes for classification tasks or reduces computational burden. Such manifold learning based methods have drawn considerable interests in recent years. Some of the examples are Principal Component Analysis (PCA) (Turk and Pentland, 1991), Linear Discriminant Analysis (LDA) (Lu et al., 2003), Locality Preserving Projection (LPP) (He and Niyogi, 2004), (Shikkenawis and Mitra, 2012) and Neighborhood Preserving Embedding (NPE) (He et al., 2005), (Koringa et al., 2015). Techniques such as PCA and LDA preserve global geometry of data. On the other hand, techniques such as LPP and NPE tend to preserve local geometry by a graph structure, based on local neighborhood information.

The linear dimensionality reduction method Orthogonal Neighborhood Preserving Projection (ONPP) proposed in (Kokiopoulou and Saad, 2007) preserves global geometry of data as well as captures innate relationship of local neighborhood. An extended version of the same is presented in (Koringa et al., 2015). ONPP is linear extension of Locally Linear Embedding (LLE) presented in (Roweis and Saul, 2000) which assumes that the data point lying on a small patch have linear relationship with its neighbours. LLE uses a weighted nearest neighborhood graph to represent local geometry by representing each data point as linear combination of its neighbors and it embeds sample points into lower dimensional space such that the linear relationship is also preserved in lower dimensional space. Being a non-linear dimensionality reduction technique, LLE does not have any mechanism of accommodating out-of-sample data. ONPP uses the same philosophy as that of LLE and projects the sample data onto linear subspace and thus allows an out-of-sample data point to be projected in low dimensional space.

All these dimensionality reduction techniques mainly use cost function in the form of optimizing error in L2-norm, which are not robust to out-liers (Chang and Yeung, 2006). L1-norm on other hand, is known for its robustness to out-liers. In recent times, L1-norm optimization is employed to dimensionality reduction techniques. This paper uses one such algorithm used in L1-PCA (Kwak, 2008) to achieve

Koringa, P. and Mitra, S.
Orthogonal Neighborhood Preserving Projection using L1-norm Minimization.
DOI: 10.5220/0006196101650172
In *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017)*, pages 165-172
ISBN: 978-989-758-222-6

L1-ONPP. This article contains the experiments on synthetic data showing susceptibility of L2-ONPP towards outliers and comparison of the performance of L2-ONPP and L1-ONPP on data having outliers. The relationship between PCA and ONPP is established and proved theoretically, the experiment performed on synthetic as well as real data supports the claim that ONPP basis can be obtained using PCA. Experimental results suggest that L1-ONPP outperforms L2-ONPP while dealing with the outliers.

In the next section, L1-norm based PCA is explained in detail, following section III establishes relation between ONPP and PCA. Section IV consists of experimental results followed by conclusion in Section V.

2 L1-NORM FOR DIMENSIONALITY REDUCTION

All conventional Dimensionality reduction techniques employs optimization of cost function in terms of L2-norms. Conventional ONPP is also based on L2-norm optimization (Kokiopoulou and Saad, 2007). Although it has been successful in many problems, it is prone to the presence of outliers because the effect of the outliers with a large norm is exaggerated by the use of the L2-norm. In order to alleviate this problem and achieve robustness, research has been performed on L1-norm based dimensionality reduction techniques. Many works have been done in PCA based on L1-norm (Ding et al., 2006), (Baccini et al., 1996), (Ke and Kanade, 2005), (Kwak, 2008). Not much works has been carried to propose L1-norm based methods of recently proposed dimensionality reduction techniques such as LPP and ONPP.

In (Baccini et al., 1996), (Ke and Kanade, 2005), each component of the error between the original data point and its projection was assumed to follow a Laplacian distribution instead of Gaussian and maximum likelihood estimation was used to formulate L1-norm PCA (L1- PCA) to the given data. In (Baccini et al., 1996), a heuristic estimation for general L1 problem was used to obtain a solution of L1-PCA. While, in (Ke and Kanade, 2005), the weighted median method and convex programming methods were proposed for L1-norm PCA. Despite the robustness of L1-PCA, it has several drawbacks and it is computationally expensive because it is based on linear or quadratic programming. In (Ding et al., 2006), R1-PCA was proposed, which combines the merits of L2-PCA and those of L1-PCA. R1-PCA is rotational invariant like L2-PCA and it successfully suppresses the effect of outliers as L1-PCA does. However, these methods are highly dependent on the dimension d of a subspace to be found. For example, the projection vector obtained when $d = 1$ may not be in a subspace obtained when $d = 2$. Moreover, as it is an iterative algorithm so for a large dimensional input space, it takes a lot of time to achieve convergence. Let us now discuss the work on L1-norm based PCA.

2.1 L1-norm PCA

Let $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}] \in \mathbf{R}^{m\times n}$ be the given data where m and n denotes dimensions of the original input space and number of data samples, respectively. Without losing generality, data is assumed to have zero mean i.e. $\bar{\mathbf{x}} = \mathbf{0}$. L2-PCA tries to find a $d(<m)$ dimensional linear subspace such that the basis vectors capture the direction of maximum variance by minimizing the error function:

$$\arg\max \mathcal{E}(\mathbf{y}) = \arg\max \sum_{i=1}^{n} \parallel \mathbf{y_i} - \bar{\mathbf{y}} \parallel_2$$

$$\mathbf{y_i} = \mathbf{V^T x_i}$$

$$\arg\max \mathcal{E}(\mathbf{V}) = \arg\max \sum_{i=1}^{n} \parallel \mathbf{V^T x_i} - \mathbf{V^T}\bar{\mathbf{x}} \parallel_2$$

$$\arg\max \mathcal{E}(\mathbf{V}) = \arg\max \sum_{i=1}^{n} \parallel \mathbf{V^T x_i} \parallel_2$$

$$\arg\max \mathcal{E}(\mathbf{V}) = \arg\max \parallel \mathbf{V^T X} \parallel_2 \tag{1}$$

subject to $\mathbf{V^T V} = \mathbf{I_d}$

where, $\mathbf{V} \in \mathcal{R}^{\mathbf{m\times d}}$ is the projection matrix whose columns constitute the bases of the d dimensional linear subspace.

In (Kwak, 2008), instead of maximizing variance in original space which is based on the L2-norm, a method that maximizes the dispersion in L1-norm in the feature space is presented to achieve robust and rotational invariant PCA. The approach presented in (Kwak, 2008) for L1-norm optimization is simple, iterative and easy to implement. It is also proven to find a locally maximal solution. Maximizing dispersion using L1-norm in the feature space can be presented as

$$\arg\max \mathcal{E}(\mathbf{V}) = \arg\max \parallel \mathbf{V^T X} \parallel_1 \tag{2}$$

The closed form solution is not possible in L1-norm maximization problem, thus the basis are sought

iteratively as follows:
For $d = 1$

$$v_1 = \arg\max \parallel \mathbf{v^T X} \parallel_1 = \arg\max \sum_{i=1}^{n} |\mathbf{v^T x_i}| \qquad (3)$$

subject to $\parallel \mathbf{v} \parallel_\mathbf{2} = 1$
For $d>1$
Once the basis in the direction of maximum variance $\mathbf{v_i}$ (here, $\mathbf{v_1}$) is found as explained above, the data is projected on this basis vector. For rest of the basis vectors the same maximization problem given in 2 is solved for projected samples $(\mathbf{X_d} = \mathbf{X_{d-1}} - \mathbf{v_{d-1}}(\mathbf{v_{d-1}^T X_{d-1}}))$ iteratively, which essentially means in every iteration, direction of maximum variance is sought, the data is projected on this basis and from this projected data the direction of maximum variance is sought until desirable dimensional space is achieved.
Algorithm to Compute L1-PCA Basis:(Kwak, 2008)

For $d = 1$**:**

1. Initialization: Pick any $\mathbf{v(0)}$. Set $\mathbf{v(0)} \leftarrow \mathbf{v(0)}/\parallel \mathbf{v(0)} \parallel_\mathbf{2}$ and $t = 0$.
2. Polarity Check: For all $i \in 1,...,n$, if $\mathbf{v^T(t)x_i} < 0$, $p_i(t) = -1$, otherwise $p_i(t) = 1$
3. Flipping and maximization: Set $t \leftarrow t+1$, and $\mathbf{v(t)} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{p_i(t)x_i}$. Set $\mathbf{v(t)} \leftarrow \mathbf{v(t)}/\parallel \mathbf{v(t)} \parallel_\mathbf{2}$
4. Convergence Check:
   a. if $v(t) \neq v(t-1)$, go to Step 2.
   b. Else if there exists i such that $\mathbf{v^T(t)x_i = 0}$, set $\mathbf{v(t)} \leftarrow (\mathbf{v(t)}+\Delta\mathbf{v})/\parallel \mathbf{v(t)}+\Delta\mathbf{v} \parallel_\mathbf{2}$ and go to step 2, Here, $\Delta\mathbf{v}$ is a small nonzero random vector.
   c. Otherwise, set $\mathbf{v}^* = \mathbf{v(t)}$ and stop.

For $d>1$:
For $j = 2$ to d,

1. Projecting Data: $\mathbf{X_j} = \mathbf{X_{j-1}} - \mathbf{v_{j-1}}(\mathbf{v_{j-1}^T X_{j-1}})$
2. Finding L1-PCA basis: in order to find $\mathbf{v_j}$, apply L1-PCA procedure to $\mathbf{X_j}$

end

3 ONPP IN TERMS OF PCA OF RECONSTRUCTION ERRORS

Now coming to ONPP(Kokiopoulou and Saad, 2007), which is a linear extension of LLE and thus inherits the sensitivity towards outliers. The degradation in manifold learning in the presence of outliers inspired the use of L1-norm minimization of ONPP to tackle the outliers. In the Section 3.1 traditional L2-ONPP is explained and then the relationship between L2-PCA and L2-ONPP established in Section 3.2. Section 3.3 explains how L1-PCA can be used to compute L1-ONPP bases.

3.1 L2-ONPP and L2-MONPP

LLE is a nonlinear dimensionality reduction technique that embeds high dimension data samples on lower dimensional subspace. The drawback of this embedding is the non-explicit mapping, in the sense that embedding is data dependent. In LLE, learned data manifold will change with the inclusion or exclusion of data point. Hence, problem such as recognition or classification of out-of-sample, LLE fails. ONPP is linear extension of LLE which resolves this problem and finds the explicit mapping of the data in lower dimensional subspace through a linear orthogonal projection matrix. In presence of this orthogonal projection matrix, new data point can be embedded into lower dimensional subspace making classification or recognition task of out-of-sample data possible. However, like LLE, ONPP is also susceptible to presence of outliers. Another variant of ONPP is Modified ONPP (Koringa et al., 2015), which considered local non-linearity in neighbourhood patch and uses non-linear weight to reconstruct the data point. ONPP and MONPP both uses L2-norm optimization, the difference between both dimensionality reduction algorithm is the mechanism to assign weights to neighbours of a data points as explained below.

Let $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}] \in \mathbf{R}^{m\times n}$ be the data matrix such that $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}$ are data points from m-dimensional space. The key task of the subspace based dimensionality reduction techniques is to find an orthogonal or non-orthogonal projection matrix $\mathbf{V}^{m\times d}$ such that $\mathbf{Y} = \mathbf{V}^T\mathbf{X}$, where $\mathbf{Y} \in \mathbf{R}^{d\times n}$ is the embedding of $\mathbf{X}$ in lower dimension as d is assumed to be less than m.

ONPP achieves the projection matrix in two step algorithm, the first step considers local patches, where each data point is expressed as a linear combination of its neighbors. In the second step, ONPP tries to preserve this linear relationship in neighbourhood and achieves data compactness through a minimization problem.

Let $\mathcal{N}_{x_i}$ be the set of k neighbors $\mathbf{x_j}$s of data point $\mathbf{x_i}$. First, data point $\mathbf{x_i}$ is expressed as linear combination of its neighbors as $\sum_{j=1}^{k} w_{ij}\mathbf{x_j}$ where, $\mathbf{x_j} \in \mathcal{N}_{x_i}$. The weight w_{ij} are computed by minimizing the reconstruction errors i.e. error between $\mathbf{x_i}$ and linear combination of its neighbours $\mathbf{x_j} \in \mathcal{N}_{x_i}$. The minimization problem can be posed as:

$$\arg\min \mathcal{E}(\mathbf{W}) = \arg\max \frac{1}{2}\sum_{i=1}^{n} \| \mathbf{x_i} - \sum_{j=1}^{k} w_{ij}\mathbf{x_j} \|_2 \quad (4)$$

subject to $\sum_{j=1}^{k} w_{ij} = 1$.

The problem corresponding to point $\mathbf{x_i}$ can be solved as a least square problem. let $\mathbf{X_{N_i}}$ be a neighbourhood matrix having $\mathbf{x_j}$ as its columns, where $\mathbf{x_j} \in N_{x_i}$. Note that $\mathbf{X_{N_i}}$ includes $\mathbf{x_i}$ as its own neighbor. Hence, dimension of $\mathbf{X_{N_i}}$ is $m \times k+1$. Now equation (4) can be written as a least square problem $(\mathbf{X_{N_i}} - \mathbf{x_i}\mathbf{e^T})\mathbf{w_i} = \mathbf{0}$ with a constraint $\mathbf{e^T}\mathbf{w_i} = \mathbf{1}$. Here, $\mathbf{w_i}$ is a weight vector of dimension $k \times 1$ and e is a vector of ones. A closed form solution, as shown in equation (5) is derived for $\mathbf{w_i}$. Here, $\mathbf{e}$ is a vector of ones having dimension $k \times 1$ same as $\mathbf{w_i}$.

$$\mathbf{w_i} = \frac{\mathbf{G}^{-1}\mathbf{e}}{\mathbf{e}^T\mathbf{G}^{-1}\mathbf{e}} \quad (5)$$

where, $\mathbf{G}$ is Gramiam matrix of dimension $k \times k$. Each element of $\mathbf{G}$ is calculated as $\mathbf{g_{pl}} = (\mathbf{x_i} - \mathbf{x_p})^T(\mathbf{x_i} - \mathbf{x_l}),\ for \forall \mathbf{x_p}, \mathbf{x_l} \in \mathcal{N}_{x_i}$

On the other hand, MONPP stresses on the fact that the local neighbourhood patch assumed to be linear may have some non-linearity and thus uses non-linear weights incorporating Z-shaped function (Koringa et al., 2015) to reconstruct a data point using its neighbours. Equation(6) is used to assign weight to each neighbor $\mathbf{x_j}$ corresponding to $\mathbf{x_i}$ using Z-shaped function based on the distance d between them. Note that this equation is same as equation(5), where $\mathbf{G}^{-1}$ is replaced by $\mathbf{Z}$. The new weights are

$$\mathbf{w_i} = \frac{\mathbf{Ze}}{\mathbf{e}^T\mathbf{Ze}} \quad (6)$$

Next step is dimensionality reduction or finding the projection matrix V such that the data point $\mathbf{x_i} \in \mathbf{R}^m$ is projected on lower dimensional space as $\mathbf{y_i} \in \mathbf{R}^d$ $(d << m)$ with the assumption that the linear combination of neighbors $\mathbf{x_j}$s which reconstruct the data point $\mathbf{x_i}$ in higher dimensional space would also reconstruct $\mathbf{y_i}$ in lower dimensional space with corresponding neighbors $\mathbf{y_j}$s along with same weights w_{ij} as in higher dimensional space. Such embedding can be obtained by solving a minimization problem of reconstruction errors in the lower dimensional space. Hence, the objective function is given by

$$\arg\min \mathcal{E}(\mathbf{Y}) = \arg\max \sum_{i=1}^{n} \| \mathbf{y_i} - \sum_{j=1}^{n} w_{ij}\mathbf{y_j} \|_2 \quad (7)$$

subject to, $\mathbf{V}^T\mathbf{V} = \mathbf{I_d}$ (orthogonality constraint).

This optimization problem results in a eigen-value problem and the closed form solution is eigen-vectors corresponding to the smallest d eigen values of matrix $\mathbf{X}(\mathbf{I}-\mathbf{W})(\mathbf{I}-\mathbf{W}^T)\mathbf{X}^T$. ONPP explicitly maps $\mathbf{X}$ to $\mathbf{Y}$, which is of the form $\mathbf{Y} = \mathbf{V}^T\mathbf{X}$, where, each column of $\mathbf{V}$ is an eigen-vector.

3.2 L2-norm ONPP using PCA

Consider the philosophy of ONPP, where each data point $\mathbf{x_i}$ is reconstructed with its neighbours $\mathbf{x_j} \in \mathcal{N}_{x_i}$. Let $\mathbf{X}' = [\mathbf{x}_1', \mathbf{x}_2', ..., \mathbf{x}_n'] \in \mathbf{R}^{m \times n}$ be the reconstructed data matrix. The same can be written as a product of data matrix $\mathbf{X}$ and weight matrix $\mathbf{W} = [\mathbf{w_1}, \mathbf{w_2}, ...\mathbf{w_n}]$ i.e. $\mathbf{XW}$. Thus reconstruction error for each data point can be denoted as $\mathbf{er_i} = \mathbf{x_i} - \mathbf{x_i}'$ as shown in Figure(1) and the reconstruction error matrix is $\mathbf{Er} = \mathbf{X} - \mathbf{XW}$ such that each column of matrix $\mathbf{Er}$ represents error vector $\mathbf{er_i}$ between $\mathbf{x_i}$ and its reconstruction $\mathbf{x_i}'$.

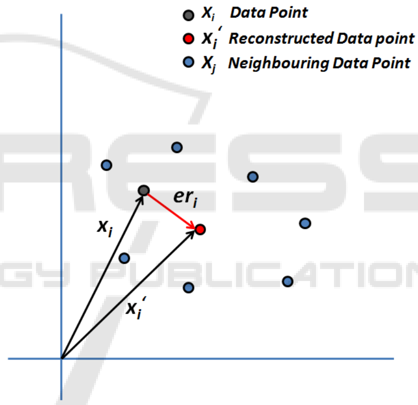


Figure 1: Illustration of data point $\mathbf{x_i}$, its reconstruction $\mathbf{x_i}'$ using neighbors $\mathcal{N}_{\mathbf{x_i}}$ and error vector $\mathbf{er_i}$.

To establish relationship between PCA and ONPP, rewrite the equation (7) in a matrix form:

$$\arg\min \mathcal{E}(\mathbf{Y}) = \arg\max \| \mathbf{Y} - \mathbf{YW} \|_2$$

$$\arg\min \mathcal{E}(\mathbf{V}) = \arg\max \| \mathbf{V^T X} - \mathbf{V^T XW} \|_2$$

$$\arg\min \mathcal{E}(\mathbf{V}) = \arg\max \| \mathbf{V^T}(\mathbf{X} - \mathbf{XW}) \|_2$$

$$\arg\min \mathcal{E}(\mathbf{V}) = \arg\max \| \mathbf{V^T Er} \|_2 \quad (8)$$

Now comparing the optimization problems for PCA in Equation(1) and the optimization problem for ONPP in Equation(8), both are eigen-value problems and have closed form solution in term of eigen-vectors. Equation(1) is maximization problem thus the desired bases are eigen-vectors corresponding to largest d eigen-values, where as Equation(8) is minimization problem thus the bases vectors are eigen-vectors corresponding to smallest d eigen-values.

In other word, ONPP is essentially finding bases vectors $\mathbf{V}$ such that it captures the direction in which the variance of reconstruction error is minimum. Thus the weakest basis of PCA when performed on the reconstruction errors $\mathbf{Er}$ is the strongest ONPP basis. This result was verified in experiments performed on the synthetic data and it is observed that the ONPP bases obtained in using conventional algorithm and ONPP bases obtained using PCA on reconstruction error are same. The experiments are documented in Section 4.

3.3 L1-norm ONPP using PCA on Reconstruction Error

The relationship established between L2-PCA and L2-ONPP in Section 3.2 led to the use of L1-PCA algorithm to solve L1-ONPP optimization problem. Rewriting ONPP optimization problem in Equation (7) as a L1-norm minimization problem, we have

$$\arg\min \mathcal{F}(\mathbf{Y}) = \arg\max \sum_{i=1}^{n} \| \mathbf{y_i} - \sum_{j=1}^{n} w_{ij}\mathbf{y_j} \|_1 \quad (9)$$

subject to, $\mathbf{V}^T\mathbf{V} = \arg\max \mathbf{I}$ (orthogonality constraint).

In matrix form,

$$\arg\min \mathcal{F}(\mathbf{Y}) = \arg\max \| \mathbf{Y} - \mathbf{YW} \|_1$$

$$\arg\min \mathcal{F}(\mathbf{V}) = \arg\max \| \mathbf{V^T X} - \mathbf{V^T XW} \|_1$$

$$\arg\min \mathcal{F}(\mathbf{V}) = \arg\max \| \mathbf{V^T}(\mathbf{X} - \mathbf{XW}) \|_1$$

$$\arg\min \mathcal{F}(\mathbf{V}) = \arg\max \| \mathbf{V^T Er} \|_1 \quad (10)$$

As established in Section 3.2, the problem stated in Equation(10) is similar to problem stated in L1-PCA (Equation(2)), and can be solved using L1-PCA algorithm performed on reconstruction error matrix $\mathbf{Er}$.

Comparing equation (10) with equation (2) of PCA, we can intuitively state that the component in the direction of minimum variance gives strong ONPP basis i.e. considering each reconstruction error vector $\mathbf{er_i}$ as one data point in m-dimensional space, the d-dimensional space can be sought such that the bases vectors are in the direction of minimum variances of the reconstruction error. Such bases can be computed using L1-PCA algorithm given in Section 2.1 where, each reconstruction error $\mathbf{er_i}$ is treated as a data point.

4 EXPERIMENTS

To validate the theoretical conclusion on the bases L2-ONPP and L2-PCA, experiments were performed on the synthetic as well as real data as documented in this Section.

4.1 A Toy Problem with Swiss Role Data

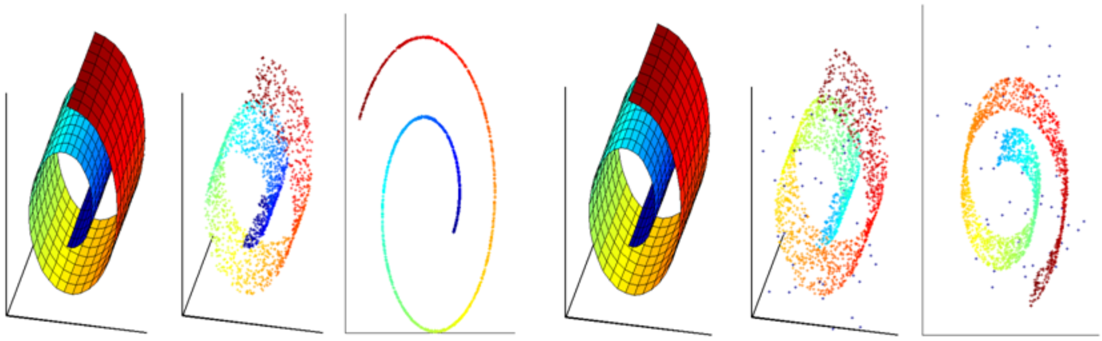
An experiment was performed on Swiss-role data to observe the effect of outliers on L2-ONPP embeddings. Over 2000 3D data points were randomly sampled from a continuous Swiss-role manifold, random uniform noise was added to nearly 2.5% of these data points i.e. 50 data points were chosen randomly and corrupted with uniform noise. As it can be seen from Figure 2(a), in clean data local as well as global geometry of data is well preserved. Whereas in case of noisy data (Figure 2(b)), all neighbours of the clean data point may not lie on locally linear patch of a manifold, which leads to the biased reconstruction. On the other hand, the neighbourhood of the outlier comparatively very larger and thus does not represent local geometry very well, as the effect of outliers is exaggerated by the use of L2-norm.

4.2 Comparing L2-ONPP Basis and L1-ONPP Basis on Synthatic Data

This experiment was performed on toy data to validate the relationship between PCA and ONPP as described in Section 3.2. 2D data was randomly generated from 7 clusters which are closely placed and slightly overlapping, 2 out of 7 were slightly separated. 100 samples from each cluster were taken resulting in 700 data points. L2-ONPP bases were found using conventional method and another set of bases vectors were computed by performing PCA on reconstruction error. The bases are same.

L2-norm ONPP basis [Figure 3(a)]
1^{st} basis : $[0.6361, 0.7716]^T$
2^{nd} basis: $[-0.7716, 0.6361]^T$

PCA basis on Reconstruction errors [Figure 3(b)]



(a) (b)

Figure 2: L2-ONPP performed on Swiss role data (a) Continuous manifold (left), sampled 3D data (middle) and its 2D representation using strongest 2 basis of ONPP(right) (b) Continuous manifold (left), sampled 3D data corrupted with uniform noise (middle) and its 2D representation using strongest 2 basis of ONPP (right).

1^{st} basis : $[0.6360, 0.7717]^T$
2^{nd} basis: $[-0.7717, 0.6360]^T$

Figure 3(c) shows that ONPP bases are basically searching the direction in which the variance of reconstruction error is minimum. For this data, L1-ONPP bases were computed using L1-PCA algorithm as can be seen from Figure 3(d) the projection basis are tilted towards the outlier data.

L1-norm ONPP basis [Figure 3(b)]
1^{st} basis : $[0.4741, 0.8805]^T$
2^{nd} basis: $[-0.8805, 0.4741]^T$

In this experiment, the residual error was observed for both, L2-ONPP and L1-ONPP, when data is projected on lower dimension space. In this case, the data was projected using only 1 dimension using the strongest basis vector. The average residual error was calculated using

$$\mathbf{e_{avg}} = \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{x_i} - \mathbf{v_1}(\mathbf{v_1^T x_i}) \qquad (11)$$

The average residual errors of L2-ONPP and L1-ONPP are 2.3221 and 0.7894, respectively. Thus. it can be concluded that L1-ONPP is less susceptible to outliers compared to L2-ONPP.

4.3 Experiment with IRIS Dataset

Iris data form UCI Machine Learning Reposatory (Fisher, 1999) is used to compare the classification performance of L1-ONPP and L2-ONPP. The Iris data set contains 4D data from 150 instances belonging to 3 different classes. Fig 4 shows the residual error obtained while reconstructing the data using varying number of dimensions. As shown in Table 1 the residual error is less in L1-ONPP as compared to L2-ONPP which significantly improves classification accuracy at lower dimensions. When dimension is 4, the projection spans entire original space, thus the reconstruction error drops nearly zero for both, L2-ONPP and L1-ONPP. As can be seen from Table 1 the classification error at 4 dimension yields greater than the lower dimension representation as it includes the redundant details present in higher dimension. The same results can be observed in all dimensionality reduction techniques at higher dimensions. Here, Nearest Neighbour (NN) classifier is used.

5 CONCLUSION

Liner dimensionality reduction techniques such as PCA, LPP and ONPP try to solve an optimization problem. Usually, the optimization is performed using L2-norm. However, these techniques based on L2-norm are susceptible to outliers present. The present work is first attempt to compute basis vector for ONPP using L1-norm. In particular, a relation is established to show that ONPP bases are same as that of PCA of reconstruction error. These phenomenon is established both theoretically and experimentally. An existing technique of finding PCA basis using L1-norm is utilized to compute the L1-norm ONPP basis. It has also been shown experimentally that the residual error after reconstructing data with less number of dimension is comparatively low in case of L1-norm ONPP than that of L2-norm ONPP. Experiments are performed for synthetic as well as real data, and the same conclusion as mentioned above is observed. As a future work one can employ L1-norm based ONPP on face reconstruction problem to observe its suitability to handle outliers such as a faces with occluded areas.

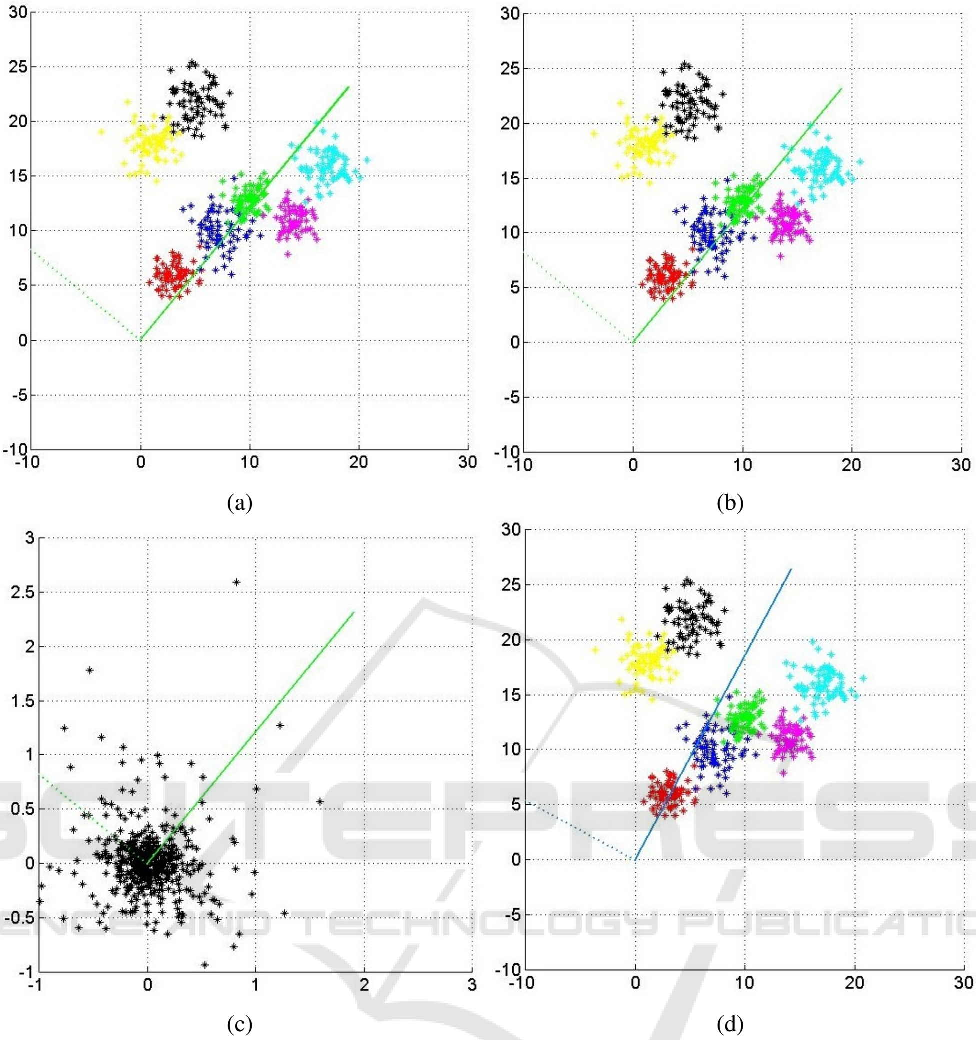


Figure 3: A toy example with 700 data samples from 7 clusters. Solid line : first projection basis, Dotted line : second projection basis (a) Projection basis using ONPP (b) Projection basis using PCA on reconstruction basis (c) Projection basis overlapped on reconstruction errors (d) Projection basis using L1-norm ONPP.

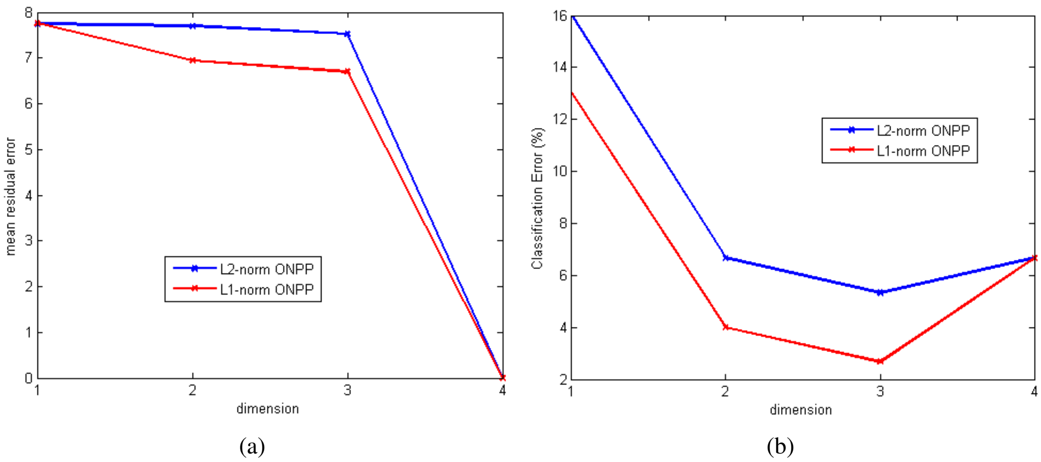


Figure 4: Performance comparision of L2-ONPP and L1-ONPP with respect to varying number of dimensions used to reconstruct the IRIS data in terms of (a) Residual Error (b) Classification Error.

Table 1: Comparison of performance in terms of residual error and classification error (in %) of L2-ONPP and L1-ONPP on IRIS data.

	Residual Error		Classification Error(%)	
Dim	L2-ONPP	L1-ONPP	L2-ONPP	L1-ONPP
1	7.8614	7.8546	16.00	13.00
2	7.7450	7.6730	6.67	4.00
3	7.0055	5.6545	5.33	2.67
4	1.44e-15	1.45e-15	6.67	6.67

REFERENCES

Baccini, A., Besse, P., and De Falguerolles, A. (1996). A l1-norm pca and a heuristic approach.

Chang, H. and Yeung, D.-Y. (2006). Robust locally linear embedding. *Pattern recognition*, 39(6):1053–1065.

Ding, C., Zhou, D., He, X., and Zha, H. (2006). R 1-pca: rotational invariant l 1-norm principal component analysis for robust subspace factorization. In *Proceedings of the 23rd international conference on Machine learning*, pages 281–288. ACM.

Fisher, R. A. (1999). UCI repository of machine learning databases – iris data set.

He, X., Cai, D., Yan, S., and Zhang, H. J. (2005). Neighborhood preserving embedding. In *Tenth IEEE International Conference on Computer Vision, ICCV 2005.*, volume 2, pages 1208–1213. IEEE.

He, X. and Niyogi, P. (2004). Locality preserving projections. In *Advances in Neural Information Processing Systems*, pages 153–160.

Ke, Q. and Kanade, T. (2005). Robust l 1 norm factorization in the presence of outliers and missing data by alternative convex programming. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 739–746. IEEE.

Kokiopoulou, E. and Saad, Y. (2007). Orthogonal neighborhood preserving projections: A projection-based dimensionality reduction technique. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2143–2156.

Koringa, P., Shikkenawis, G., Mitra, S. K., and Parulkar, S. (2015). Modified orthogonal neighborhood preserving projection for face recognition. In *Pattern Recognition and Machine Intelligence*, pages 225–235. Springer.

Kwak, N. (2008). Principal component analysis based on l1-norm maximization. *IEEE transactions on pattern analysis and machine intelligence*, 30(9):1672–1680.

Lu, J., Plataniotis, K. N., and Venetsanopoulos, A. N. (2003). Face recognition using lda-based algorithms. *Neural Networks, IEEE Transactions on*, 14(1):195–200.

Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326.

Shikkenawis, G. and Mitra, S. K. (2012). Improving the locality preserving projection for dimensionality reduction. In *Third International Conference on Emerging Applications of Information Technology (EAIT), 2012*, pages 161–164. IEEE.

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86.