NONLINEAR DIMENSIONALITY RECUTCION USING ISOMAP

J. Suh.

Center of Advanced Vehicular System
Mississippi State University
suh@cavs.msstate.edu

ABSTRACT

Dimension reduction techniques are widely used for the analysis and visualization of complex sets of data. Contrarily to the traditional linear Principle Component Analysis (PCA), the non linear methods work like Multidimensional Scaling (MDS), by reproducing in the projection space the pairwise distances measure in the data space, which is called Isomap. Isomap differs from the classical linear MDS by metrics they use and by the way they build the mapping. I describe an approach to solving dimensionality reduction problems that uses easily measured local metric information to learn the underlying global geometry of a data set. The proof of the existence of Isomap is explained in Section 3. The Isomap algorithm is performed on human handwriting and face images, so the algorithm discovers the nonlinear degrees of freedom that underlying complex natural observations.

1. INTRODUCTION

While complex stimuli of complex data can be represented by points in a high dimensional vector space, they typically have a much more compact description. Coherent structure in the world leads to strong correlations between inputs, generating observations that lie on or close to a smooth low dimension manifold. Several dimension reduction techniques has been studied by the respect of linear method, like PCA and MDS, and these techniques allow the user to better analyze or visualize complex data sets in linear data space. A collected data points has an irregular time series, and then non linear tool needs to be employed to analysis the data.

The task of recovering meaningful low dimensional structures hidden in high dimensional data is main task for this paper. Classical techniques for manifold learning, such as PCA or MDS, are designed to operate when the submanifold is embedded linearly in the observation space. More generally there is a wider class of techniques, involving iterative optimization procedures, by which unsatisfactory linear representations obtained by PCA or MDS may be improved toward more successful non linear representation of the data.

The PCA or MDS fail when non linear structure cannot simply be regarded as a perturbation from a linear approximation. The iterative approach has a tendency to get stuck at locally optimal solutions that grossly misrepresent the true geometry of the situation.

The Section 2 describes basic concepts of MDS. The MDS enables that we can find the intrinsic manifolds dimensions. The Section 3 is main task for this paper to explain about Isomap and proof for existence of Isomap. The Section 4 gives a result for high dimensional face images and hand written data sets. One goal of this paper is to describe a class of practical techniques for dealing with nonlinear distributed signals.

2. MULTIDIMENSIONAL SCALING

MDS can explore the underlying structure of relations between entities by providing a geometrical representation of each data inputs. MDS algorithm has three main steps to gain insight of relations of data [1]. First, the algorithm computes a scale of comparative distances between all pairs of data points. The comparative distances, which are relative to one another from data, are lack of null point to measure absolute distances. The second step involves estimating an additive constant and using this constant estimate to convert the comparative distances into absolute distances. In the third step, the absolute distances between data points are projected onto the real Euclidean space. The third step makes the MDS technique unique compared to PCA analysis. The next paragraph provides more rigorous background to support the existence of projected dimension of Euclidean space.

Young and Householder proposed the method, which states the absolute distances in any space can be considered to be the distances lying in Euclidean space [2]. The distance in Euclidean distance determines the dimension of the space by projection of points on a set of orthogonal axes of the space. For example, the three points, which are i, j, and k, be alternate scripts for n points (i, j, k = 1, 2, …, n) and d_{ij}, d_{ik}, and d_{jk} be the absolute distance between the points, then B_i is an (n-1)x(n-1) symmetric matrix with elements.

$$b_{jk} = \frac{1}{2}(d_{ij}^2 + d_{ik}^2 - d_{jk}^2) \qquad (1)$$

The element b_{jk} can be considered to be the scalar product of vectors from point i to points j and k. Matrix B_i is a matrix of scalar products of vectors with origin at point i.

Young and Householder have shown that:

1. If any matrix B_i is positive semidefinite, the distanc es may be considered to be the distances between p oints lying in a real Euclidean space.
2. The rank of any positive semidefinite matrix B_i is equal to the dimensionality of the set of points.
3. Any positive semidefinite matrix B_i may be factor-ed to obtain a matrix A_i such that $B_i = A_i A_i^T$. If the rank of B_i is r, where $r \leq (n-1)$, then matrix Ai is a (n-1) x r matrix of projections of points on r orth-o gonal axes with origin at the i^{th} point of the r-dimen sional, real Euclidean space.

Young and Householder's Euclidean model enables the absolute distance to project the mapped data points onto true dimension of the underlying manifold.

3. ISOMAP

To explore a broad class of nonlinear manifolds, nonlinear technique is employed with MDS. Nonlinear Dimensionali-ty reduction problem is known as "manifold learning" [6]. This approach seeks to preserve the intrinsic geometry of the data, as captured in the geodesic, which is shortest path, manifold distances between all pairs of data points. The algorithm, Isomap, is estimating the geodesic distance between faraway points, given only input-space distances. Global approaches may similarly seek to map nearby points on the manifold to nearby points in low-dimensional space, but at the same time faraway points on the manifold must be mapped to faraway points in low dimensional space.

The basic idea behind Isomap consists in overcoming the limitations of the traditional metric MDS, which is linear, by replacing the Euclidean distance by another metrics [5]. Indeed, the MDS encounters difficulties when projecting nonlinear structures like the spiral illustrated in Figure 1 (a). Actually, the spiral is embedded in a two-dimensional space, but clearly its intrinsic dimension does not exceed one: only one parameter suffices to describe the spiral. Unfortunately, the projection from two dimensions to only one dimension is not easy because the spiral needs to be unrolled onto a straight line. This unfolding is difficult for MDS because the pairwise Euclidean distances after projection are much larger than in the embedding space: they cannot go through shortcuts like in Figure 1 (b). They have to be measured like in Figure 1 (c) along the spiral.

From a technical point of view, Isomap processes a d-dimensional set of n data points as follows [4]:

1. Construct neighborhood graph: Define the graph G over all data points by connecting point i and j if they are closer than ε (ε-Isomap), or if i is one of the K nearest neighborhoods of j (K-Isomap). Set edge lengths equal to $d_x(i,j)$. Choosing the value of ε and K will be discussed in next section.
2. Compute shortest paths: Initialize $d_G(i,j) = d_x(i,j)$ if i, j are linked by an edge; $d_G(i,j) = \infty$ otherwise. Then for each value of $k = 1,2,\ldots,n$ in turn, replace all entries $d_G(i,j)$ by $\min\{ d_G(i,j), d_G(i,k)+d_G(k,j)\}$. The matrix of final value of $D_G = \{d_G(i,j)\}$ will constant the shortest path distance between all pairs of points in G.

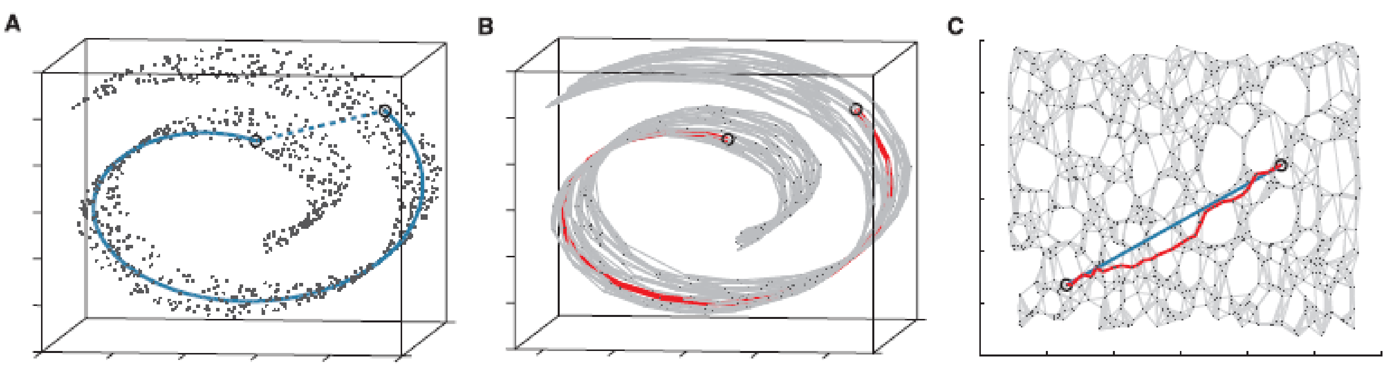

Figure 1 (a): Two empty circle points are measure by Euclidean distance as dotted line and by geodesic distance along the low dimensional manifold as solid line

Figure 1 (b): The neighborhood graph G constructed by first step of Isomap algorithm as K=7, and the red line indicate graph G.

Figure 1 (c): The two dimensional embedding recovered by Isomap in neighborhood graph. Straight line represents cleaner approximation to true geodesic distance.

3. Construct d-dimensional embedding: Apply the traditional metric MDS on matrix D_G, i.e. comput-e the eigenvectors of D_G and keep the ones associ-ated with the p largest eigenvalues, giving the coor dinates of the landmarks points in the p-dimension al projection space.

The final step applies classical MDS to the matrix of graph distance D_G construing an embedding of the data in a d-dimensional Euclidean space S that best preserves the manifold's estimated intrinsic geometry. The coordinate vectors s_i for points in S are chosen to minimize the cost function

$$E = \| \tau(D_G) - \tau(D_S) \| \quad (2)$$

where D_S denotes the matrix of Euclidean distances. The τ operator coverts distances to inner products, which uniquely characterize the geometry of the data in a form that supports efficient optimization. The global minimum of Eq (2) is achieved by setting the coordinates si to the top d eigenvectors of the matrix $\tau(D_G)$. Actually, if Isomap had used Euclidean distances, step 3 computed with D_G would have given the same results as PCA directly applied to the d-dimensional coordinates. However, the use of geodesic distances introduces an implicit nonlinear transform of the coordinates and forbids the use of PCA.

3.1. Proof of ISOMAP

Isomap deals with finite data sets of points in R^n which are assumed to lie on a smooth submanifold M^d of low dimension $d < n$. The algorithm attempts to recover M given only the data points. A crucial stage in the algorithm involves estimating the unknown geodesic distance in M between data points in terms of the graph distance with respect to some graph G constructed on the data points.

Let $\{x_i\} \subset M$ be a finite set, whose elements we will refer to as data points. These points may be chosen randomly, or obtained in some other manner. The Isomap algorithm attempts to recover the manifold distances $d_M(i,j)$, given only the data points $\{x_i\} \subset R^n$. Of course, this can be done approximately. Suppose that the data points are chosen randomly, with a certain density function α..

For example, the sample set $\{x_i\}$ is chosen according to a Poisson process with density function α, meaning that for any reason measurable subset $A \subseteq M$,

$$P(\text{A contains exactly k points in } \{x_i\}) = e^{-a} a^k / k! \quad \text{where } a = \int_A \alpha \quad (3)$$

The expected number of points in A is just a. The Poisson process is constructed so that disjoint regions behave independently of each other. Here we start a simple condition on the data set $\{x_i\}$ and the graph G which guarantee that d_S is a good approximation to d_M.

Asymptotic Convergence Theorem

Given $\lambda_1, \lambda_2, \mu > 0$, then for α sufficiently large the inequalities

$$1 - \lambda_1 \le \frac{\text{graph distance}}{\text{geodesic distance}} \le 1 + \lambda_2 \quad (4)$$

hold with probability at least 1-μ. Since the Poisson distribution is constructed in disjoint regions, the probability at least hold 1-μ.

Let's prove that the λ_1 and λ_2 can be defined by the inequality properties. Let $M = M^d$ is a compact d-dimensional smooth submanifold of the Euclidean space R^n. The natural Riemannian structure on M (induced from the Euclidean metric on R^n) gives rise to a manifold metric d_M defined by:

$$d_M(x,y) = \inf_{\gamma} \{length(\gamma)\} \quad (5)$$

where *r* varies over the set of smooth arcs connecting x to y in M. Notes that $d_M(x,y)$ is generally different from the Euclidean distance $\|x-y\|$.

The construction makes use of a graph G on the data points. Given such a graph we can define two further metrics, just on the set of data points. Let x, y belong to the set {xi}. We define:

$$\begin{aligned} d_G(x,y) &= \min_p(\| x_0 - x_1 \| + \ldots + \| x_{p-1} - x_p \|) \\ d_S(x,y) &= \min_p(d_M(x_0 - x_1) + \ldots + d_M(x_{p-1} - x_p)) \end{aligned} \quad (6)$$

where $P=(x_0, \ldots, x_p)$ varies over all paths along the edges of G connecting x to y. Given the data points and graph G, one can compute d_G without knowledge of the manifold M. This is the key stage in the Isomap algorithm.

Following Sampling Theorem exemplifies that there is an upper limit, λ_2.

Sampling Theorem

Let δ and ε be positive, with $4\delta < \varepsilon$. Suppose:

1. The graph G contains all edges x and y for which $d_M(x,y) \le \varepsilon$.
2. For every point m in M there is a data point x_i for which $d_M(m,x_i) \le \delta$.

Then for all pairs of data points x, y we have

$$d_M(x,y) \le d_S(x,y) \le (1+4\delta/\varepsilon)\, d_M(x,y) \quad (7)$$

We refer to the second condition in the theorem as the "δ-sampling condition".

Using the geometry property of M in addition with distribution property, we defined the some parameter. The minimum radius of curvature is $r_0=r_0(M)$. Any Euclidean sphere of radius r_0 has minimum radius of curvature equal to r_0; in particular this is true of circles of radius r_0 contained in some 2-dimensional plane. Intuitively, geodesics in M curl around "less tightly" than circles of radius less than $r_0(M)$. The minimum branch separation $s_0= s_0$ (M) is defined to be the largest positive number for which $||x-y||< s_0$ implies $d_M(x,y)\leq\pi r_0$, for $x,y\in$ M.

To prove the lower limit, λ_1, we use following Lemma and Corollary.

Lemma

If r is a geodesic in M connecting points x an y, and if l = length(r) $\leq \pi$ then:

$$2r_0 \sin(l/2r_0) \leq \| x-y \| \leq l \quad (8)$$

Using the fact that $\sin(t) \geq t - t^3/6$ for $t > 0$, we can write down a weakened form of Lemma:

$$(1-l^2/24r_0^2)l \leq \| x-y \| \leq l \quad (9)$$

Finally, using the first-order weakening:

$$(2/\pi)l \leq \| x-y \| \leq l, \quad (10)$$

which is valid in the range $l \leq \pi r_0$.

Corollary

Let $\lambda > 0$ be given, Suppose the points x_i, x_{i+1} in M satisfy the conditions:

$$\| x_i - x_{i+1} \| < s_0$$

$$\| x_i - x_{i+1} \| \leq (2/\pi)r_0\sqrt{24\lambda}$$

Suppose also that there is a geodesic arc of length $d_M(x_i, x_{i+1})$ connecting x_i to x_{i+1}.
Then:

$$(1-\lambda)d_M(x_i,x_{i+1}) \leq \| x_i - x_{i+1} \| \leq d_M(x_i,x_{i+1}) \quad (11)$$

The corollary condition results and equation (9) gives the more clear condition for λ, which is $1 - \lambda \leq 1 - l^2/24r_0^2$. Finally we can decide the lower limit of λ.

$$\sqrt{24\lambda}r_0 \geq l \text{ where } \lambda \geq \frac{l^2}{24r_0^2} \quad (12)$$

Using sampling theorem, lemma, and corollary can provide how we can set the λ_1 and λ_2.

Main Theorem for ε

Let M be a compact submanifold of R^n isometrically equivalent to a convex domain in R^d. Let λ_1, λ_2 and μ be given, and $\varepsilon > 0$ be chosen so that $\varepsilon < s_0$ and $\varepsilon \leq (2/\pi)r_0\sqrt{24\lambda_1}$. A sample data set $\{x_i\}$ is chosen randomly from a Poisson distribution with density function α, and the ε–rule is used to construct a graph G on $\{x_i\}$.

By viewing geometry point of view, we can set the volume formula $V\min(r) = \eta_d r_d$, where η_d is the volume of the unit ball in R^d.

l_{max} Theorem

Let l_{max} be chosen to satisfy:

$$\alpha_{\min}V_{\min}(l_{\max}/2) > 2(K+1)$$

Then, with probability at least $1 - \mu$, no ball $B_x(l_{max})$ of radius l_{max} contains fewer than K + 1 data points.

Main Theorem for K

Let M be a compact submanifold of Rn isometrically equivalent to a convex domain in Rd. Let λ_1, λ_2 and μ be given, and $\varepsilon > 0$ be chosen so that $\varepsilon < s_0$ and $\varepsilon \leq (2/\pi)r_0\sqrt{24\lambda_1}$. A sample data set {xi} is chosen randomly from a Poisson distribution with density function alpha, which has bounded variation $A=\alpha_{min}/\alpha_{max}$. Fix the ratio

$$\frac{K+1}{a_{\min}} = \frac{\eta_d(\varepsilon/2)^d}{2}$$

And use the K-rule to construct a graph G on {xi}.

The proof of Theorem for ε and K are explicitly provided in paper [3].

How quickly $d_G(x,y)$ converges to $d_M(x,y)$ depends on certain parameters of the manifold as it lies within the high-dimensional space (radius of curvature and branch separation) and on the density of data points. To the extent that a data set presents extreme values of these parameters or deviates from a uniform density, asymptotic convergence still holds in general, but the sample size required estimating geodesic distance accurately may be impractically large.

4. EXPERIMENTAL RESULTS

The Isomap performs its dimensionality reduction analysis on high dimensional person's face images and handwriting images. The face input consists of many images of a person's face observed under different pose and lighting conditions in no particular order. These images can be thought of as points in a high-dimension corresponding to the brightness of on a pixel in the image. The handwriting inputs consists of bottom loop and top arc of 2's in any order. Although the input dimensionality may be quite high (i.e., 4096 for these 64 pixel by 64 pixel images), the perceptually meaningful structure of these image has many fewer independent degrees of freedom. Applying the Isomap technique, underlying manifolds or intrinsic dimensions are appreciated.

To analysis the face images, Isomap takes 698 raw images, and neighborhood value is set to 6 to learn the embedding dimension of data's intrinsic data structure. Figure 2 (a) shows that person's face images within the 4096-dimensional input space lie on an intrinsically three dimensional manifold that can be parameterized by two pose variables plus an azimuthal lighting angle. For handwritten data, Isomap apply for total number of images as 1000. We use ε Isomap, because we do not expect a constant dimensionality to hold over the

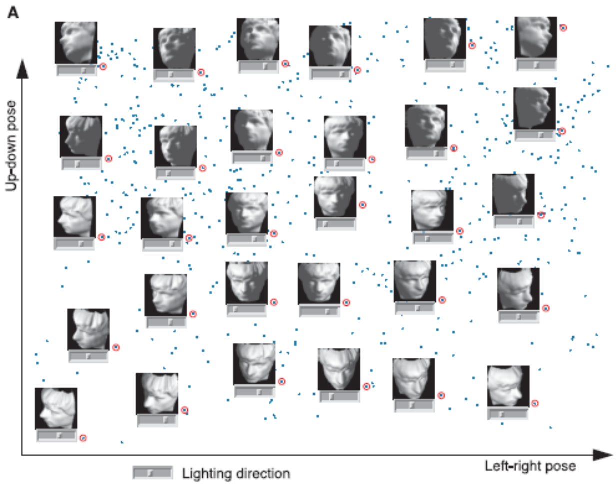

Figure 2 (a): N = 698 raw images, K = 6 Isomap to represent 3 dimensions, which are left right pose, up down pose and lighting directions.

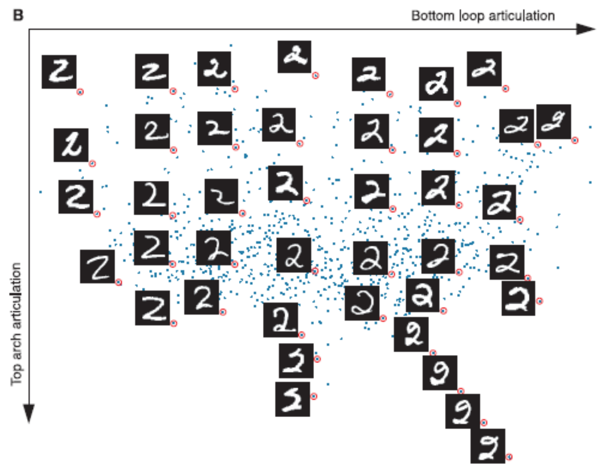

Figure 2 (b): N = 1000 handwritten images, ε = 4.2 Isomap to represent 2 dimensions, which are bottom loop articulation and top arch articulation.

whole data set. Figure 2 (b) similarly shows handwritten "2"s images place on two dimensional space which can be parameterized by bottom loop articulation and top arch articulation. Our goal is to discover, given only the unordered high-dimensional inputs, low-dimensional representations such as Figure 2 with coordinates that capture the intrinsic degrees of freedom of a data set. The Figures 2 clearly indicate the Isomap can extract the intrinsic lower dimensional features from high dimensional data inputs.

Comparing the linear tool and nonlinear tool for estimating the true underlying factors, the residual variance with dimensionality experiment results are shown in Figure 3 and 4. The residual variance is decreased as the dimensionality increasing. Linear tool, PCA and MDS, and nonlinear tool, Isomap, is applied to face images and hand written images. In Figure 3, the Isomap clearly points out that the 3 dimension can reduce the variance error close to zero. In the mean time, the PCA and MDS do not give same variance until at dimension as 10. The handwritten image has more dimension requirement, since the handwritten data has successive exaggerations of an extra stroke or ornament in the digit. In Figure 4, the empty circle and triangle represent the MDS and PCA, and the linear tool takes much higher residual variance than Isomap.

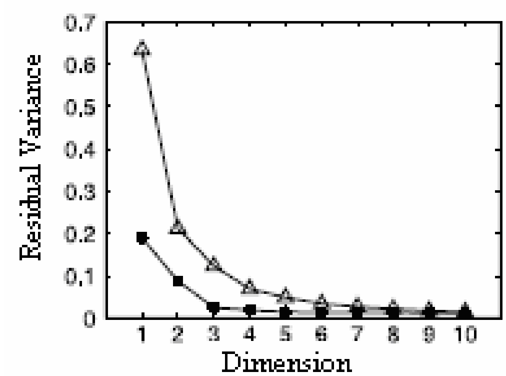

Figure 3: Face images varying in pose and illumination. (empty triangle: PCA and MDS, and dark circle: Isomap)

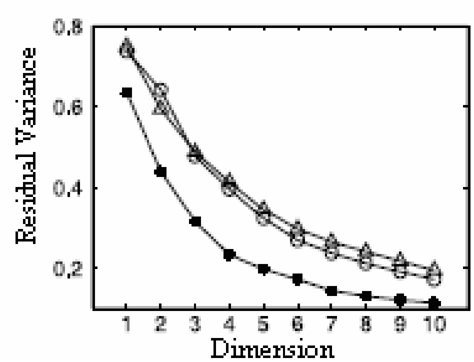

Figure 4: Handwritten "2" images. (empty triangle: PCA empty circle: MDS, and dark circle: Isomap)

5. CONCLUSIONS AND FUTURE WORKS

In this paper we have presented how the higher dimensional data can be analyzed to find the lower dimensional embedding of the data's intrinsic geometric structure. Isomap shows the ability to estimate the most dominant parameters to describe the nonlinear data points. Using a proper value of K or ε , Isomap is able to construct the underlying manifold. The phone or power spectral of speech signal can be distinguished from well trained Isomap in further study. The way we approach by topological method is one of interesting area in non linear system.

6. REFERENCES

[1] Torgerson, W.S., "Multidimensional Scaling, I: Theory and Method," *Psychometrika*, vol. 17, pp. 401-419, 1952.

[2] Young, G. and Householder, A. S., "Discussion of a set of points in terms of their mutual distances," *Psychometrika*, vol. 2, p p. 19-22, 1938.

[3] Bernstein, M., Silva, V. de., Langford, J. C., and Tenenbaum, J. B., "Graph approximations to geodesics on embedded manifold, " *Technical report*, Stanford University, Stanford, 2000.

[4] Tenenbaum, J.B., Silva, V. de., Langford, J.C., "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319-2323, 2000.

[5] Lee, J. A., Lendasse, A., Verleysen, M., "Cuvlinear Distance Analysis versus Isomap," *ESANN proceedings*, vol. 24, pp. 185-192, 2002.

[6] V. de., Langford, J. C., and Tenenbaum J. B., "Global Versus local methods in nonlinear dimensionality reduction," *Advances in Neural Information Processing Systems*, vol. 15, 2003.