

---

# Isometric Embedding and Continuum ISOMAP

---

Hongyuan Zha

Department of Computer Science and Engineering, The Pennsylvania State University, University Park, PA 16802, USA

ZHA@CSE.PSU.EDU

Zhenyue Zhang

Department of Mathematics, Zhejiang University, Yuquan Campus, Hangzhou, 310027, P. R. China

ZYZHANG@ZJU.EDU.CN

## Abstract

Recently, the Isomap algorithm has been proposed for learning a nonlinear manifold from a set of unorganized high-dimensional data points. It is based on extending the classical multidimensional scaling method for dimension reduction. In this paper, we present a continuous version of Isomap which we call continuum isomap and show that manifold learning in the continuous framework is reduced to an eigenvalue problem of an integral operator. We also show that the continuum isomap can perfectly recover the underlying natural parametrization if the nonlinear manifold can be isometrically embedded onto an Euclidean space. Several numerical examples are given to illustrate the algorithm.

Traditional dimension reduction techniques such as principal component analysis and factor analysis usually work well when the data points lie close to a *linear* (affine) subspace of the high-dimensional data space [5]. They can not, in general, discover nonlinear structures embedded in the set of data points. Recently, two novel methods for manifold learning, the locally linear embedding method (LLE) in [7] and the Isomap method in [11], have drawn great interests. Unlike other nonlinear dimension reduction methods, both LLE and Isomap methods emphasize simple algorithmic implementation and avoid nonlinear optimization formulations that are prone to local minima. The focus of this paper is on analyzing the Isomap method which extends the classical multidimensional scaling (MDS) method by exploiting the use of geodesic distances of the underlying nonlinear manifold (details of Isomap will be presented in section 2). Regarding the Isomap method, a fundamental question of great theoretical as well as practical interest is the following:

## 1. Introduction

The continuous increase in computing power and storage technology makes it possible for us to collect and analyze ever larger amount of data. In many real-world applications, the data are of high-dimension, those applications include computational genomics, image analysis and computer vision, document analysis in information retrieval and text mining. Fortunately, in many of those applications, all of the components of those high-dimensional data vectors are not independent of each other and in many cases the data points can be considered as lying on or close to a low-dimensional nonlinear manifold embedded in a high-dimensional space. Learning the nonlinear low-dimensional structures hidden in a set of unorganized high-dimensional data points known as *manifold learning* represents a very useful and challenging unsupervised learning problem.

What is the low-dimensional nonlinear structure that Isomap tries to discover and for what type of nonlinear manifolds, can Isomap perfectly recover the low-dimensional nonlinear structure?

Since a nonlinear manifold can be parametrized in infinitely many different ways, it is not apparent what is actually being discovered by those nonlinear dimension reduction methods, and relying on experimental verification will tend to be very unreliable as we will show later.

The general question of when Isomap performs well is first addressed in [1, 11] where asymptotic convergence results are derived for Isomap, highlighting the importance of geodesic convexity of the underlying manifold

and isometry for the success of Isomap at recovering the low-dimensional nonlinear structure. Extensions to conformal mappings is also discussed in [8]. Some of the aspects of the question has further been analyzed in [4] under the framework of continuum Isomap emphasizing nonlinear manifolds constructed from collections of images. In [4] it is *defined* that continuum Isomap obtains a perfect recovery of the natural parameter space of the nonlinear manifold in question if the geodesic distance on the nonlinear manifold is proportional to the Euclidean distance in the parameter space. Unfortunately, no continuous version of the Isomap is given and in all the work we have just mentioned the reason why (continuum) Isomap should work perfectly is explained using the discrete framework of the classical MDS. The purpose of this paper is to fill the gap by presenting a continuous version of Isomap using integral operators. In particular, we show that for a nonlinear manifold that can be isometrically embedded onto an open and convex subset of an Euclidean space, the continuum Isomap computes a set of eigenfunctions which forms the canonical coordinates of the Euclidean space up to a rigid motion. For non-flat manifolds, we argue that certain information will be lost if the Isomap only makes use of a finite number of the eigenfunctions. More importantly, we emphasize that isometry is a more fundamental property than geodesic distance with regard to manifold learning. A global method such as Isomap can not make full use of isometry which is basically a local property between two manifolds. Local manifold learning methods such as LLE [7] and LTSA [13] are called for when global methods such as Isomap fail. The rest of the paper is organized as follows: In section 2 we review both the classical MDS and its generalization Isomap proposed in [11]. In section 3, we recall several basic concepts from differential geometry such as isometric embedding and geodesic distances. In section 4, we derive the continuum Isomap and show that the continuum Isomap can perfect recover the natural parameter space of the nonlinear manifold in question if the geodesic distance on the nonlinear manifold is proportional to the Euclidean distance in the parameter space. We also illustrate the role played by an Isometry. Section 5 contains several concluding remarks.

## 2. Classical Multidimensional Scaling and Isomap

Suppose for a set of  $N$  points  $\{x_i\}_{i=1}^N, x_i \in \mathcal{R}^m$  with  $N > m$ , we are given the set of pairwise Euclidean

distances

$$d(x_i, x_j) = \|x_i - x_j\|_2 \equiv \left( \sum_{k=1}^m (x_i^{(k)} - x_j^{(k)})^2 \right)^{1/2},$$

where  $x_i^{(k)}$  denotes the  $k$ -th component of  $x_i$ , and we are asked to reconstruct the  $\{x_i\}$ 's from the above set of pairwise distances. We can proceed as follows: without loss of generality, we can assume that the  $\{x_i\}$ 's are centered, i.e.,

$$\sum_{i=1}^N x_i = 0.$$

Notice that the squared pairwise distance

$$d^2(x_i, x_j) = \|x_i - x_j\|_2^2 = x_i^T x_i - 2x_i^T x_j + x_j^T x_j.$$

Let the  $N$ -dimensional vector  $\psi$  be

$$\psi = [x_1^T x_1, \dots, x_N^T x_N]^T.$$

Then the squared-distance matrix  $D = [d^2(x_i, x_j)]_{i,j=1}^N$  can be written as

$$D = \psi e^T - 2X^T X + e \psi^T,$$

where  $e$  is the  $N$ -dimensional vector of all ones, and  $X = [x_1, \dots, x_N]$ . Let  $J = I - ee^T/N$ . Then it follows that

$$H \equiv -JDJ/2 = X^T X. \quad (2.1)$$

To recover  $X$ , let the eigendecomposition of  $H$  be

$$H = U \text{diag}(\lambda_1, \dots, \lambda_m) U^T$$

with  $U \in \mathcal{R}^{N \times m}$ . Then we can set

$$X = \text{diag}(\lambda_1^{1/2}, \dots, \lambda_m^{1/2}) U^T.$$

It is easy to see that the set of  $x_i$ 's can be recovered up to a rigid motion, i.e., a composition of a translation and an orthogonal transformation. More details of multidimensional scaling (MDS) and connection with eigendecomposition can be found in [2, 12].

When the set of data points  $x_i$ 's lie on or close to a low-dimensional nonlinear manifold embedded in a high-dimensional space and the nonlinear structure can not be adequately represented by a linear approximation, classical MDS as discussed above usually fails to recover the low-dimensional structure of the nonlinear manifold. We illustrate this issue using a set of examples. On the top row of Figure 1, we plot two sets of sample points from 1D curves represented as  $x_i = f(\tau_i) + \epsilon_i, i = 1, \dots, N, N$  is the total number

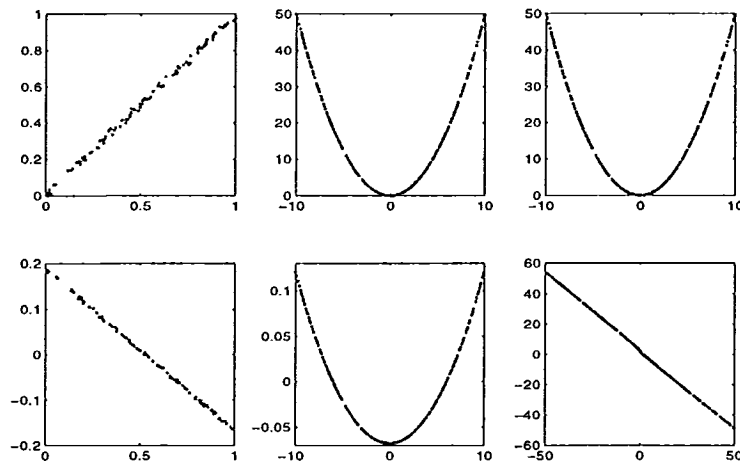


Figure 1. 2D point sets and their 1D projections: (left) points on a straight line, projections using MDS, (middle) points on a parabola, projections using MDS, (right) points on a parabola, projections using Isomap

of points plotted, the  $\tau_i$ 's are chosen uniformly from a finite interval and  $\epsilon_i$  represents noise. The first curve is a straight line and the second and the third curves are the same and both are parabola. Each figures in the second row plots the  $\tau_i$ 's against the 1D projections computed by the classical MDS discussed above and the Isomap method presented in the next paragraph. If the  $\tau_i$ 's are perfectly recovered, we should see a straight line in the figures of the second row with a slope either  $\pi/4$  or  $-\pi/4$ . We see that for the straight line example in the left panel of Figure 1, the classical MDS can recover the underlying 1D parametrization, but it fails for the nonlinear curve in the middle panel. However, Isomap can still recover the 1D parametrization for the same nonlinear curve as is shown in the right panel of Figure 1.

Isomap was proposed as a general technique for nonlinear dimension reduction (i.e., uncovering the natural parameter space of a nonlinear manifold): the pairwise Euclidean distance  $d(x_i, x_j)$  in classical MDS is replaced by the geodesic distance between  $x_i$  and  $x_j$  on the manifold (defined as the length of the shortest path between two points) [11]. In particular, 1) a so-called neighborhood graph  $G$  of the data points  $x_i$ 's is constructed with an edge connecting points  $x_i$  and  $x_j$  if  $x_i$  is one of the  $k$  nearest neighbors of  $x_j$ , for example.<sup>1</sup> The edge  $(x_i, x_j)$  is then assigned a weight  $\|x_i - x_j\|_2$  which equals the Euclidean distance between  $x_i$  and  $x_j$ ; 2) the geodesic distance between  $x_i$  and  $x_j$  is then approximated by the shortest path within

<sup>1</sup>The number of nearest neighbors  $k$  is a parameter of the algorithm that needs to be pre-set. One can also choose an  $\epsilon$ -neighborhood, i.e., considering  $x_i$  and  $x_j$  as connected if  $\|x_i - x_j\|_2 \leq \epsilon$  (see [11] for details).

the weighted graph  $G$ , call it  $\hat{d}(x_i, x_j)$ ; 3) the classical MDS is then applied to the squared-geodesic-distance matrix  $\hat{D} = [\hat{d}^2(x_i, x_j)]_{i,j=1}^N$ . As we can see that the key difference between the classical MDS and Isomap is that in the classical MDS pairwise Euclidean distance is used while in Isomap it is the pairwise geodesic distance. Empirical success of Isomap for discovering nonlinear structures of high-dimensional data sets has been demonstrated in [4, 11].

From both a practical as well as a theoretical viewpoint, one is naturally led to the following questions: What is the low-dimensional nonlinear structure that Isomap tries to discover? And for what type of nonlinear manifolds, can Isomap perfectly recover the low-dimensional nonlinear structure? With issues such as discretization errors, sampling density of the data points and errors due to noise, it is not easy to answer those questions in a clear way using the current framework of Isomap based on a discrete set of data points. It is generally agreed, however, that reasoning in a continuous framework can sometimes crystallize the real issue of the problem and provide intuition for further development. This is the viewpoint taken in [4] and the same we will follow in this paper as well. Before we discuss Isomap in a continuous framework, we need to first introduce some basic notions from differential geometry.

### 3. Isometric Embedding of Manifolds

In this section, we recall several basic facts of differential geometry [3, 6, 9, 10]. The general theory of manifold learning can be cast in the framework of Riemannian geometry, but to avoid unnecessary abstraction,

we consider manifolds represented as hypersurfaces in Euclidean spaces. First, we introduce the concept of a *tangent map*.

**DEFINITION.** Let  $f : \Omega \rightarrow \mathcal{M}$  be a mapping between two manifolds. The tangent map  $f_*$  of  $f$  assigns to each tangent vector  $v$  to  $\Omega$  the tangent vector  $f_*(v)$  to  $\mathcal{M}$  such that if  $v$  is the initial velocity of a curve  $\alpha$  in  $\Omega$ , then  $f_*(v)$  is the initial velocity of the image curve  $f(\alpha)$  in  $\mathcal{M}$ .

When  $\Omega$  is an open set of the  $d$ -dimensional Euclidean space  $\mathcal{R}^d$  and  $\mathcal{M}$  is embedded in  $\mathcal{R}^m$ , and assume that we can write  $f$  as

$$f(\tau) = \begin{bmatrix} f_1(\tau) \\ \vdots \\ f_m(\tau) \end{bmatrix},$$

then

$$J_f(\tau) = \begin{bmatrix} \partial f_1/\partial \tau_1 & \cdots & \partial f_1/\partial \tau_d \\ \vdots & \vdots & \vdots \\ \partial f_m/\partial \tau_1 & \cdots & \partial f_m/\partial \tau_d \end{bmatrix}$$

gives the Jacobian matrix of  $f$ , and the tangent map  $f_*(v)$  is simply  $f_*(v) = J_f v$ , here  $\tau = [\tau_1, \dots, \tau_d]^T$ .

**DEFINITION.** The mapping  $f : \Omega \rightarrow \mathcal{M}$  is an *isometry* if  $f$  is one-to-one and onto and  $f$  preserves inner products in the tangent spaces, i.e., for the tangent map  $f_*$ ,

$$f_*(v)^T f_*(w) = v^T w$$

for any two vectors  $v$  and  $w$  that are tangent to  $\Omega$ .

In the case that  $\Omega$  is an open set of  $\mathcal{R}^d$ , it is easy to see that  $f$  is an isometry if and only if  $f$  is one-to-one and onto, and the Jacobian matrix  $J_f$  is orthonormal, i.e.,  $J_f^T J_f = I_d$ .

The geodesic distance between two points on a manifold is defined as the length of the shortest path between the two points in question. For an isometry  $f$  defined on an open convex set of  $\mathcal{R}^d$ , it is easy to show that the geodesic distance between two points  $f(\tau_1)$  and  $f(\tau_2)$  on  $\mathcal{M}$  is given by

$$d(f(\tau_1), f(\tau_2)) = \|\tau_1 - \tau_2\|_2. \tag{3.2}$$

In fact, let  $\alpha : [0, 1] \rightarrow \mathcal{M}$  be a curve on  $\mathcal{M}$  such that  $\alpha(0) = \tau_1$  and  $\alpha(1) = \tau_2$ , i.e.,  $f(\alpha(0)) = f(\tau_1)$  and  $f(\alpha(1)) = f(\tau_2)$ . The length of the curve  $\alpha$  is then

$$\begin{aligned} L(\alpha) &= \int_0^1 \left\| \frac{df(\alpha(t))}{dt} \right\|_2 dt \\ &= \int_0^1 \|J_f(\alpha(t))\alpha'(t)\|_2 dt = \int_0^1 \|\alpha'(t)\|_2 dt. \end{aligned}$$

Geodesic curves on  $\mathcal{R}^d$  are straight lines, and we can choose a minimizer of  $L(\alpha)$  as  $\alpha(t) = \tau_1 + t(\tau_2 - \tau_1)$ . The corresponding minimum value of  $L(\alpha)$  is  $\|\tau_1 - \tau_2\|_2$ , and a geodesic curve on  $\mathcal{M}$  is given by  $f(\tau_1 + t(\tau_2 - \tau_1))$ . We next give two examples to illustrate the concepts we have introduced.

**EXAMPLE.** First we consider the set of scaled 2-by-2 rotations of the form

$$R(\theta) = \frac{1}{\sqrt{2}} \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}, \quad \theta \in (-\infty, \infty).$$

We embed  $\{R(\theta)\}$  into  $\mathcal{R}^4$  by

$$R(\theta) \rightarrow f(\theta) = \frac{1}{\sqrt{2}} [\cos \theta, \sin \theta, -\sin \theta, \cos \theta]^T.$$

It is easy to check that  $\|J_f(\theta)\|_2 = 1$ , and the geodesic distance between  $R(\theta_1)$  and  $R(\theta_2)$  is  $|\theta_1 - \theta_2|$ .

**REMARK.** If  $d = 1$ , and  $\mathcal{M}$  represents a *regular* curve, i.e.,  $f'(\tau) \neq 0$  for all  $\tau \in \Omega$ , we can always reparametrize  $\mathcal{M}$  by its arc-length  $s$  to obtain  $g : s \rightarrow \mathcal{M}$  and  $\|g'(s)\|_2 = 1$ , i.e.,  $g$  is an isometry.

**EXAMPLE.** Next, we consider the 2-dimensional swiss-roll surface in 3-dimensional Euclidean space parametrized as for  $u > 0^2$

$$f(u, v) = \left[ \frac{1}{\sqrt{2}} u \cos(\log u), v, \frac{1}{\sqrt{2}} u \sin(\log u) \right]^T.$$

It can be verified that  $J_f(u, v)$  is orthonormal, i.e.,  $(J_f(u, v))^T J_f(u, v) = I_2$ , and the swiss-roll surface is isometric to  $\{(u, v) \mid u > 0\}$ .

### 4. Continuum Isomap

With the above preparation, we are now ready to present a continuous version of Isomap. Let  $d(x, y)$  define the geodesic distance between two points  $x$  and  $y$  on the manifold  $\mathcal{M}$ . We define a continuous version of the matrix  $H$  defined in (2.1) in the form of a continuous kernel  $K(x, y)$  as follows (this can be considered as the case when the sample points are uniformly concentrated on  $\mathcal{M}$ ),

$$\begin{aligned} K(x, y) &= \frac{1}{2} \int_{\mathcal{M}} (d^2(x, t) + d^2(t, y) - d^2(x, y)) dt / \int_{\mathcal{M}} dt \\ &\quad - \frac{1}{2} \int_{\mathcal{M}} ds \int_{\mathcal{M}} d^2(t, s) dt / \left( \int_{\mathcal{M}} dt \right)^2. \end{aligned}$$

We will restrict ourselves to the case:  $f : \Omega \rightarrow \mathcal{M}$ , and  $\Omega \subset \mathcal{R}^d$  is an open convex subset. Consequences

<sup>2</sup>A different swiss-roll surface was used in [7, 11] and will be discussed in the next section.

of nonconvexity of  $\Omega$  has been discussed in [1, 4] and will also be mentioned at the end of this section. By the formula for change of variables from multi-variate calculus, the integral of a function  $F$  defined on  $\mathcal{M}$  can be expressed as

$$\int_{\mathcal{M}} F(x)dx = \int_{\Omega} F(f(\tau))h(\tau)d\tau,$$

where  $h(\tau) = \sqrt{\det(J_f^T J_f)}$ . Define for short and with an abuse of notation,

$$d_f(\tau_1, \tau_2) \equiv d(f(\tau_1), f(\tau_2)),$$

and

$$K_f(\tau_1, \tau_2) \equiv K(f(\tau_1), f(\tau_2)),$$

then the kernel can be represented as

$$\begin{aligned} & K_f(\tau_1, \tau_2) \\ = & \frac{\frac{1}{2} \int_{\Omega} \left( d_f^2(\tau_1, \tau) + d_f^2(\tau, \tau_2) - d_f^2(\tau_1, \tau_2) \right) h(\tau) d\tau}{\int_{\Omega} h(\tau) d\tau} \\ & - \frac{\frac{1}{2} \int_{\Omega} d\tau \int_{\Omega} d_f^2(\tau, \hat{\tau}) h(\tau) h(\hat{\tau}) d\hat{\tau}}{\left( \int_{\Omega} h(\tau) d\tau \right)^2}. \end{aligned}$$

More generally, we can also consider data points sampled from an arbitrary density function  $\rho$  concentrated on  $\Omega$  to obtain

$$\begin{aligned} & K_f(\tau_1, \tau_2) \\ = & \frac{\frac{1}{2} \int_{\Omega} \left( d_f^2(\tau_1, \tau) + d_f^2(\tau, \tau_2) - d_f^2(\tau_1, \tau_2) \right) H(\tau) d\tau}{\int_{\Omega} H(\tau) d\tau} \\ & - \frac{\frac{1}{2} \int_{\Omega} d\tau \int_{\Omega} d_f^2(\tau, \hat{\tau}) H(\tau) H(\hat{\tau}) d\hat{\tau}}{\left( \int_{\Omega} H(\tau) d\tau \right)^2}, \end{aligned}$$

where  $H(\tau) = \rho(\tau)h(\tau)$ .

Parallel to the development in the classical MDS, we consider the eigenvalue problem of the integral operator with kernel  $K_f(\tau_1, \tau_2)$ . Let  $\phi(x)$  be an eigenfunction of the kernel  $K(x, y)$ , i.e.,

$$\int_{\mathcal{M}} K(x, y)\rho(y)\phi(y)dy = \lambda\phi(x), \quad x \in \mathcal{M}, \quad (4.3)$$

or equivalently on  $\Omega$ ,

$$\int_{\Omega} K_f(\tau_1, \tau_2)H(\tau_2)\phi(f(\tau_2))d\tau_2 = \lambda\phi(f(\tau_1)), \quad \tau_1 \in \Omega.$$

It is not difficult to verify that  $\phi(x)$  has zero mean, i.e.,

$$\int_{\mathcal{M}} \rho(x)\phi(x)dx / \int_{\mathcal{M}} \rho(x)dx =$$

$$\int_{\Omega} H(\tau)\phi(f(\tau))d\tau / \int_{\Omega} H(\tau)d\tau = 0.$$

We now show that if  $f$  is an isometry, then the first  $d$  largest eigenfunctions form the canonical coordinates of  $\Omega$  up to a rigid motion.

**Theorem 4.1** *Let  $f : \Omega \subset \mathcal{R}^d \rightarrow \mathcal{M}$  be an isometry, i.e.,*

$$d_f(\tau_1, \tau_2) = \|\tau_1 - \tau_2\|_2$$

*with  $\Omega$  open and convex. Let the mean vector  $c$  be*

$$c = \int_{\Omega} \tau H(\tau) d\tau / \int_{\Omega} H(\tau) dx.$$

*Assume that  $\phi_1(x), \dots, \phi_d(x)$  are the  $d$  eigenfunctions of the kernel  $K(x, y)$  corresponding to the  $d$  largest eigenvalues  $\lambda_j, j = 1, \dots, d$ , and*

$$\int_{\mathcal{M}} \rho(x)\phi_j^2(x)dx = \lambda_j. \quad (4.4)$$

*Then the vector function*

$$\theta \equiv [\phi_1, \dots, \phi_d]^T = P(\tau - c),$$

*where  $P \in \mathcal{R}^{d \times d}$  is a constant orthogonal matrix. Furthermore,  $\lambda_j, j = 1, \dots, d$  are the eigenvalues and  $P$  is the eigenvector matrix of the  $d$ -by- $d$  symmetric positive definite matrix,*

$$A \equiv \int_{\Omega} (\tau - c)H(\tau)(\tau - c)^T d\tau,$$

*respectively.*

**PROOF.** With the assumption  $d_f(\tau_1, \tau_2) = \|\tau_1 - \tau_2\|_2$ , we have

$$\begin{aligned} & K_f(\tau_1, \tau_2) \\ = & \frac{\frac{1}{2} \int_{\Omega} \left( \|\tau_1 - \tau\|^2 + \|\tau - \tau_2\|^2 - \|\tau_1 - \tau_2\|^2 \right) H(\tau) d\tau}{\int_{\Omega} H(\tau) d\tau} \\ & - \frac{\frac{1}{2} \int_{\Omega \times \Omega} \|\tau - \hat{\tau}\|^2 d\tau d\hat{\tau}}{\left( \int_{\Omega} H(\tau) d\tau \right)^2} \\ = & \frac{(\tau_1 - c)^T(\tau_2 - c) - (\tau_1 + \tau_2 - 2c)^T \int_{\Omega} (\tau - c)H(\tau) d\tau}{\int_{\Omega} H(\tau) d\tau} \\ & + \left\| \int_{\Omega} (\tau - c)H(\tau) d\tau / \int_{\Omega} H(\tau) d\tau \right\|^2. \end{aligned}$$

By the definition of  $c$ , we have  $\int_{\Omega} (\tau - c)H(\tau) d\tau = 0$ . Therefore after some algebraic manipulations,

$$K_f(\tau_1, \tau_2) = (\tau_1 - c)^T(\tau_2 - c).$$

Let  $\phi_j(x)$ ,  $j = 1, \dots, d$ , be the  $d$  eigenfunctions corresponding to the largest  $d$  eigenvalues  $\lambda_j$  of the integral operator with kernel  $K_f(\tau_1, \tau_2)$ ,

$$\int_{\Omega} (\tau_1 - c)^T (\tau_2 - c) H(\tau_2) \phi_j(f(\tau_2)) d\tau_2 = \lambda_j \phi_j(f(\tau_1)).$$

Defining

$$p_j = \frac{1}{\lambda_j} \int_{\Omega} (\tau - c) H(\tau) \phi_j(f(\tau)) d\tau, \quad (4.5)$$

we have

$$\phi_j(x) = \phi_j(f(\tau)) = (\tau - c)^T p_j. \quad (4.6)$$

Therefore

$$\theta \equiv [\phi_1(x), \dots, \phi_d(x)]^T = [p_1, \dots, p_d]^T (\tau - c) \equiv P(\tau - c).$$

To prove the orthogonality of  $P$ , let us substitute (4.6) into (4.5) and obtain

$$\lambda_j p_j = \int_{\Omega} (\tau - c) H(\tau) (\tau - c)^T d\tau \cdot p_j \equiv A p_j,$$

where

$$A = \int_{\Omega} (\tau - c) H(\tau) (\tau - c)^T d\tau$$

is a positive definite matrix. It clearly shows that  $p_j$  is the eigenvector of  $A$  corresponding to the eigenvalue  $\lambda_j$ . Therefore  $p_1, \dots, p_d$  can be chosen to be orthogonal to each other. Furthermore, by the normalization conditions (4.4) and (4.6),

$$\lambda_j = \int_{\Omega} H(\tau) \phi_j^2(f(\tau)) d\tau = p_j^T A p_j = \lambda_j p_j^T p_j.$$

Hence  $\|p_j\|_2 = 1$  and  $P$  is an orthogonal matrix. ■

Now if  $f(\tau)$  and  $f(\hat{\tau})$  are two different parametrizations of the same manifold  $\mathcal{M}$  and both  $f(\tau)$  and  $f(\hat{\tau})$  are isometries, then clearly  $\tau$  and  $\hat{\tau}$  only differ by a rigid motion. Now suppose  $\eta$  is a different parametrization of  $\mathcal{M}$  and is related to  $\tau$  by  $\tau = \tau(\eta)$ . What is the  $\theta$  function computed by the continuum Isomap in terms of  $\eta$ ? The following corollary answers this question.

**Corollary 4.2** *Let  $f : \tau \in \Omega \rightarrow \mathcal{M}$  be an isometry, and  $\eta$  is a parametrization of  $\mathcal{M}$  such that  $\tau = \tau(\eta)$ . Then the vector function  $\theta$  computed by the continuum Isomap considered as a function of  $\eta$  is given by*

$$\theta \equiv [\phi_1, \dots, \phi_d]^T = P(\tau(\eta) - c).$$

We now look at an application of this Corollary. In [7, 11], a 2-dimensional swiss-roll surface embedded in 3-dimensional space is parametrized as

$$f(u, v) = [u \cos u, v, u \sin u]^T.$$

It is easy to see that

$$J_f(u, v)^T J_f(u, v) = \text{diag}(\sqrt{1+u^2}, 1),$$

hence  $f$  is not an isometry for the given parametrization. However, the experimental data given in [7, 11] seem to indicate that data points sampled uniformly from a square in the parameter space are mapped back to a set of points uniformly distributed in a square by Isomap. How to explain this phenomenon? The answer lies in Corollary 4.2. Now assume that  $f$  has the following property,

$$J_f(\tau)^T J_f(\tau) = \text{diag}(\rho_1(\tau_1), \dots, \rho_d(\tau_k)).$$

Assuming that all the  $\rho_i$ 's are positive, and define the following  $d$  one-variable functions

$$\eta_i(\tau_i) = \int^{\tau_i} \rho_i^{1/2}(\tau_i) d\tau_i, \quad i = 1, \dots, d,$$

based on indefinite integrals. Since the  $\rho_i$ 's are positive, the  $\eta_i$  are strictly monotonically increasing, and therefore the mapping  $\eta \rightarrow \tau$  is one-to-one. It is easy to verify that

$$\delta_{ij} = \sum_{k=1}^d \frac{\partial \eta_i}{\partial \tau_k} \frac{\partial \tau_k}{\partial \eta_j}, \quad i, j = 1, \dots, d,$$

where  $\delta_{ij}$  is the Kronecker  $\delta$ . Define

$$J_{\tau}(\eta) = [\partial \tau_i / \partial \eta_j]_{i,j=1}^d, \quad J_{\eta}(\tau) = [\partial \eta_i / \partial \tau_j]_{i,j=1}^d,$$

it follows that

$$J_{\tau}(\eta) J_{\eta}(\tau) = I_d.$$

Since  $J_{\eta}(\tau) = \text{diag}(\rho_1^{1/2}(\tau_1), \dots, \rho_d^{1/2}(\tau_d))$ , hence we have

$$J_{\tau}(\eta) = \text{diag}(\rho_1^{-1/2}(\tau_1), \dots, \rho_d^{-1/2}(\tau_d)),$$

and furthermore

$$J_f(\eta)^T J_f(\eta) = I_d,$$

i.e., the manifold parametrized by  $\eta$  is an isometry. It follows from Corollary 4.2 that, up to a rigid motion, the  $\theta$  function computed by the continuum Isomap has the form

$$\left[ \int^{\tau_1} \rho_1^{1/2}(\tau_1) d\tau_1, \dots, \int^{\tau_d} \rho_d^{1/2}(\tau_d) d\tau_d \right]^T.$$

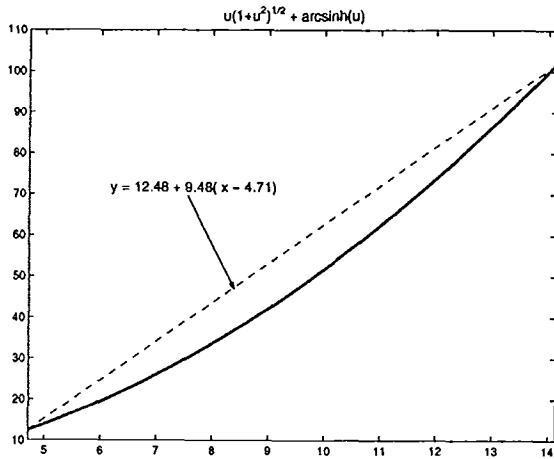


Figure 2. Deformation function for the swiss-roll surface

Notice that the geodesic distances are not preserved in the  $\tau$  space, but the deformation only occurs along the individual  $\tau_i$  directions.

Now going back to the swiss-roll surface

$$f(u, v) = [u \cos u, v, u \sin u]^T,$$

we see that up to a rigid motion, the  $\theta$  function computed by the continuum Isomap has the form

$$[(u\sqrt{1+u^2} + \operatorname{arcsinh}(u))/2, v]^T.$$

Hence no deformation (stretching and compressing) occurs in the  $v$  direction, but there is certain deformation in the  $u$  direction. However, in [7, 11], the data points are sampled uniformly from the interval  $[3\pi/2, 9\pi/2]$  along the  $u$ -direction. Within this interval, the function  $u\sqrt{1+u^2} + \operatorname{arcsinh}(u)/2$  is very close to a straight line as is illustrated in Figure 2. This explains why the points computed by Isomap seem to be uniformly distributed in a square.

Recall that we have assumed that  $\Omega$  is a *convex* open set. The convexity is crucial for the continuum Isomap to work correctly, this was clearly pointed out in [1, 4]. The reason is also quite simple, if there is a hole in the manifold, the geodesic curve needs to move *around* the hole and the relationship  $d_f(\tau_1, \tau_2) = \|\tau_1 - \tau_2\|_2$  will no longer hold even if  $J_f(\tau)^T J_f(\tau) = I_d$  still holds true. This is actually a drawback of methods such as Isomap that depend on *global* pairwise distances. As we have mentioned before, geodesic distance is a global property of a manifold while isometry is defined locally, i.e., property of the tangent spaces at each point. Proportionality of geodesic distances to Euclidean distances in the parameter space is a consequence of isometry. In the nonconvex case, however,

isometry can still hold but proportionality of geodesic distances to Euclidean distances will fail to be true. Global method such as Isomap can no longer handle this case and local method is called for. In fact if you roll a piece of paper into a swiss-roll shape, you can flatten it back without regard to whether the shape of the piece of paper is convex or not. Local methods such as the local tangent space alignment (LTSA) method proposed in [13] can still perfectly recover the low-dimensional structure as is illustrated in Figure 3, where the original data form a broken ring which is clearly nonconvex, Isomap fails to recover the original coordinates while LTSA does very well.

REMARK. The final remark we want to make is that if  $\mathcal{M}$  is not isometric to a flat space, then the number of nonzero eigenvalues of the integral operator with kernel  $K(x, y)$  defined at the beginning of the section will be *infinite*. If we select a finite number of the eigenfunctions, we can not expect them to fully represent the low-dimensional structure of the given nonlinear manifold, certain information has been lost going from infinite to finite.

## 5. Conclusions

Isomap is a generalization of the classical multidimension scaling method for nonlinear dimension reduction. We proposed a continuous version of the Isomap method and showed that for a nonlinear manifold that can be isometrically embedded onto an Euclidean space, the continuum Isomap computes a set of eigenfunctions that form the canonical coordinates of the Euclidean space up to a rigid motion. This answers the questions of what the low-dimensional nonlinear structure is that Isomap tries to discover and when it can perfectly discover it.

## Acknowledgements

The authors thank Prof. J. Tenenbaum for some helpful discussions and bringing our attention to the references [1, 8]. The work of first author was supported in part by NSF grants CCR-9901986. The work of the second author was done while visiting Penn State University and was supported in part by the Special Funds for Major State Basic Research Projects (project G19990328), Foundation for University Key Teacher by the Ministry of Education, China, and NSF grants CCR-9901986.

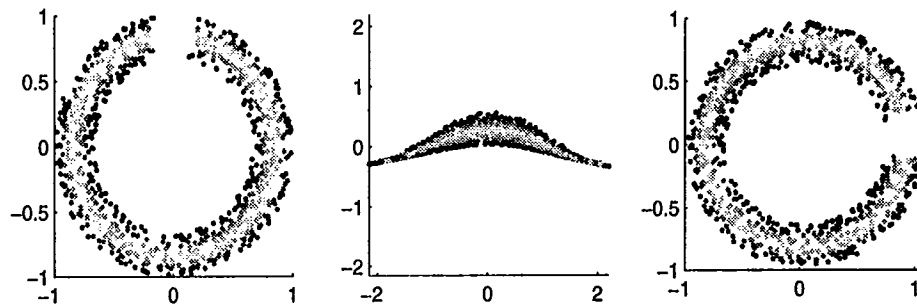


Figure 3. Broken ring data set: (left) the original data set, (middle) reconstruction using Isomap, (right) reconstruction using orthogonal LTSA

## References

- [1] M. Bernstien, V. de Silva, J. Langford, and J. Tenenbaum. Graph approximations to geodesics on embedded manifolds. Technical Report, Department of Psychology, Stanford University, 2000.
- [2] T. Cox and M. Cox. *Multidimensional Scaling*. Chapman & Hall, London, 2001.
- [3] M. do Carmo. *Differential Geometry of Curves and Surfaces*. Prentice-Hall, Englewood Cliffs, 1976.
- [4] D. Donoho and C. Grimes. When does ISOMAP recover the natural parametrization of families of articulated images? Technical Report 2002-27, Department of Statistics, Stanford University, 2002.
- [5] T. Hastie, R. Tibshirani and J. Friedman. *The Elements of Statistical Learning*. Springer, New York, 2001.
- [6] B. O'Neill. *Elementary Differential Geometry*. Academic Press, San Diego, 2nd Edition, 1997.
- [7] S. Roweis and L. Saul. Nonlinear dimension reduction by locally linear embedding. *Science*, 290: 2323–2326, 2000.
- [8] V. de Silva, J. B. Tenenbaum. Unsupervised learning of curved manifolds. In D. D. Denison, M. H. Hansen, C. C. Holmes, B. Mallick and B. Yu (eds.), *Nonlinear Estimation and Classification*, Springer-Verlag, New York, 2002.
- [9] M. Spivak. *Calculus on Manifolds*. Addison-Wesley, Redwood City, 1965.
- [10] M. Spivak. *A Comprehensive Introduction to Differential Geometry*. Publish or Perish, Boston, 2nd Edition, 1979.
- [11] J. Tenenbaum, V. De Silva and J. Langford. A global geometric framework for nonlinear dimension reduction. *Science*, 290:2319–2323, 2000.
- [12] C. Williams. On a connection between kernel PCA and metric multidimensional scaling. *Machine Learning*, 46: 11-19, 2002.
- [13] Z. Zhang and H. Zha. Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment. Technical Report, CSE-02-019, Department of Computer Science & Engineering, Pennsylvania State University, 2002.