Nonlinear Dimension Reduction via Local Tangent Space Alignment

Zhenyue Zhang[1]⋆ and Hongyuan Zha[2]⋆⋆

[1] Zhejiang University, Hangzhou 310027, P.R. China,
zyzhang@zju.edu.cn
[2] Penn State University, University Park, PA 16802, USA,
zha@cse.psu.edu

Abstract. In this paper we present a new algorithm for manifold learning and nonlinear dimension reduction. Based on a set of unorganized data points sampled with noise from the manifold, we represent the local geometry of the manifold using tangent spaces learned by fitting an affine subspace in a neighborhood of each data point. Those tangent spaces are aligned to give the internal global coordinates of the data points with respect to the underlying manifold by way of a partial eigendecomposition of the neighborhood connection matrix. We present a careful error analysis of our algorithm and show that the reconstruction errors are of second-order accuracy. Numerical experiments including 64-by-64 pixel face images are given to illustrate our algorithm.

1 Introduction

Many high-dimensional data in real-world applications can be modeled as data points lying close to a low-dimensional nonlinear manifold. Discovering the structure of the manifold from a set of data points sampled from the manifold possibly with noise represents a very challenging unsupervised learning problem [1,4,6,8,9]. The discovered low-dimensional structures can be further used for classification, clustering, outlier detection and data visualization. The key observation is that the dimensions of the embedding spaces can be very high, the intrinsic dimensionality of the data points, however, are rather limited due to factors such as physical constraints and linguistic correlations.

Recently, there have been much renewed interests in developing efficient algorithms for constructing nonlinear low-dimensional manifolds from sample data points in high-dimensional spaces, emphasizing simple algorithmic implementation and avoiding optimization problems prone to local minima [6,9]. Two lines of research of manifold learning and nonlinear dimension reduction have emerged: one is exemplified by [1,9] where pairwise *geodesic* manifold distances are estimated, and then project the data points into a low-dimensional space that

⋆ The work of this author was supported in part by the Special Funds for Major State Basic Research Projects (project G19990328) and NSF grants CCR-9901986.
⋆⋆ The work of this author was supported in part by NSF grants CCR-9901986.

J. Liu et al. (Eds.): IDEAL 2003, LNCS 2690, pp. 477–481, 2003.

best preserves the geodesic distances. Another line of research follows the long tradition starting with self-organizing maps[4], principal curves/surfaces[3], and topology-preserving networks [5]. The key idea is that the information about the global structure of a nonlinear manifold can be obtained from a careful analysis of the interactions of the *overlapping* local structures. In particular, the local linear embedding (LLE) method constructs a local geometric structure and seeks to project the data points into a low-dimensional space that best preserves those local geometries [6,7].

Our approach draws inspiration from and improves upon the work in [6,7]. In this paper, The basic idea of our approach is to use the tangent space in the neighborhood of a data point to represent the local geometry, and then align those local tangent spaces to construct the global coordinate system for the nonlinear manifold by minimizing the alignment error for the global coordinate learning. This minimization problem is equivalent to an eigenvalue problem that can be efficiently solved. We call the new algorithm *local tangent space alignment* (LTSA) algorithm.

2 The Local Tangent Space Alignment Algorithm

We assume that we are given a data set $X = [x_1, \ldots, x_N],\ x_i \in \mathcal{R}^m$ sampled with noise from an underlying d-dimensional nonlinear manifold $\mathcal{F}$ embeded into a m-dimensional space with *unknown* generating function $f(\tau),\ \tau \in R^d,\ d < m$,

$$x_i = f(\tau_i) + \epsilon, \quad i = 1, \ldots, N,$$

where $\tau_i \in \mathcal{R}^d$ are unknown. The objective for nonlinear dimension reduction is to reconstruct τ_i's from the data points x_i's without explicitly constructing f.

To this end, let us assume that the function f is smooth enough. Using first-order Taylor expansion at a fixed τ, we have

$$f(\bar{\tau}) = f(\tau) + J_f(\tau) \cdot (\bar{\tau} - \tau) + O(\|\bar{\tau} - \tau\|^2), \tag{1}$$

where $J_f(\tau) \in \mathcal{R}^{m \times d}$ is the Jacobi matrix of f at τ. Ignoring the second order term, the shifted coordinate $\bar{\tau} - \tau$ is a local coordinate of $f(\bar{\tau}) - f(\tau)$ with respect to the basis of the d column vectors of $J_f(\tau)$. Without knowing the function f, we can not explicitly compute the Jacobi matrix $J_f(\tau)$. However, if we know an orthonormal basis, say Q_τ in matrix form, of the tangent space $\mathcal{T}_\tau$ spanned by the columns of $J_f(\tau)$ and the coordinate θ_τ of $f(\bar{\tau}) - f(\tau)$ corresponding to Q_τ, we have

$$\bar{\tau} - \tau \approx L_\tau \theta_\tau$$

with (unknown) matrix $L_\tau = (Q_\tau^T J_f(\tau))^{-1}$ provided $J_f(\tau)$ is of full rank, i.e., $\mathcal{F}$ is *regular*. It is clear that the global coordinate τ can be extracted by minimizing

$$\int d\tau \int_{\Omega(\tau)} \|\bar{\tau} - \tau - L_\tau \theta_\tau\| d\bar{\tau}. \tag{2}$$

Here $\Omega(\tau)$ is a local neighborhood of τ and $\bar{\tau}$ can be taken as its mean.

The above approach can be applied to the data set X for approximately extracting the underline coordinates τ_i because the orthogonal basis Q_i of the tangent space with respect to τ_i and the local corresponding coordinate θ_i can be approximately determined by the neighborhood $x_{i_1}, \cdots, x_{i_k}$ of x_i. This can be done by finding the best d-dimensional affine subspace approximation for the data points $x_{i_1}, \cdots, x_{i_k}$

$$\min_{c_i, \theta_j^{(i)}, Q_i} \sum_{j=1}^{k} \left\| x_{i_j} - (c_i + Q_i \theta_j^{(i)}) \right\|_2^2 .$$

The optimal solutions are given by $c_i = \bar{x}_i$, the mean of all the x_{i_j}'s, Q_i the d left singular vectors of $[x_{i_1} - \bar{x}_i, \cdots, x_{i_k} - \bar{x}_i]$ corresponding to its d largest singular values, and the orthogonal project $\theta_j^{(i)} = Q_i^T (x_{i_j} - \bar{x}_i)$.

To retrieval the global coordinates $\tau_i, i = 1, \ldots, N$, in the low-dimensional feature space based on the local coordinates $\theta_j^{(i)}$, the global coordinates should respect the local geometry determined by the $\theta_j^{(i)}$, $\tau_{i_j} = \bar{\tau}_i + L_i \theta_j^{(i)} + \epsilon_j^{(i)}$ with $\bar{\tau}_i$ the mean of $\tau_{i_j}, j = 1, \ldots, k$, corresponding to the neighborhood index set of x_i. In matrix form, $T_i = \frac{1}{k} T_i e e^T + L_i \Theta_i + E_i$ with $T_i = [\tau_{i_1}, \ldots, \tau_{i_k}]$, $\Theta_i = [\theta_1^{(i)}, \ldots, \theta_k^{(i)}]$, and $E_i = [\epsilon_1^{(i)}, \cdots, \epsilon_k^{(i)}]$ is the local reconstruction error matrix. To preserve as much of the *local* geometry in the low-dimensional feature space, we seek to find τ_i and the local affine transformations L_i to minimize the reconstruction errors $\epsilon_j^{(i)}$, i.e.,

$$\sum_i \|E_i\|^2 \equiv \sum_i \|T_i (I - \frac{1}{k} e e^T) - L_i \Theta_i\|^2 = \min . \tag{3}$$

This optimization problem is equivalent to an eigenvalue problem (4) if we impose an normalization constraint on τ_i's. Obviously, for a fixed T_i, the optimal alignment matrix L_i that minimizes the local reconstruction error $\|E_i\|_F$ is given by $L_i = T_i (I - \frac{1}{k} e e^T) \Theta_i^+ = T_i \Theta_i^+$ and therefore $E_i = T_i W_i$ with $W_i = (I - \frac{1}{k} e e^T)(I - \Theta_i^+ \Theta_i)$, where Θ_i^+ is the Moor-Penrose generalized inverse of Θ_i. Let $T = [\tau_1, \ldots, \tau_N]$ and S_i be the 0-1 selection matrix such that $T S_i = T_i$. Then $E_i = T S_i W_i$. To uniquely determine T, we impose the constraints $T T^T = I_d$. We then need to find T to minimize the overall reconstruction error,

$$\min \sum_i \|E_i\|_F^2 = \min_{T T^T = I_d} \operatorname{trace}(T B T^T), \tag{4}$$

where

$$B = S_1 W_1 W_1^T S_1^T + \cdots + S_N W_N W_N^T S_N^T . \tag{5}$$

Note that the vector e of all ones is an eigenvector of B corresponding to a zero eigenvalue, therefore, the optimal T is given by the d eigenvectors of the matrix B, corresponding to the 2nd to d+1st smallest eigenvalues of B. Because of the sparse structure (5), it is not difficult to construct B.

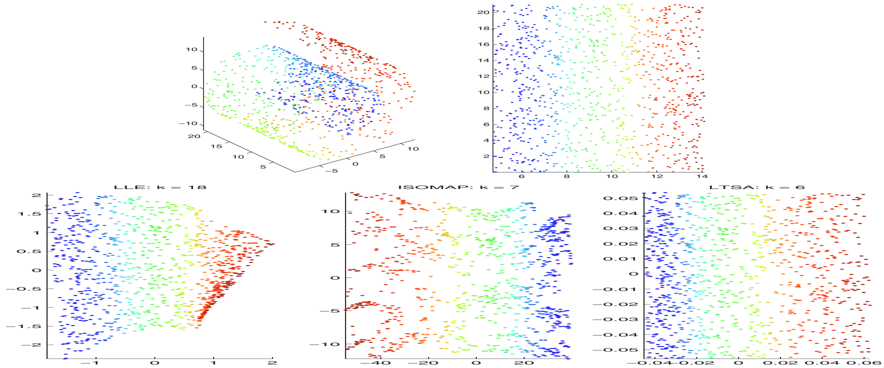

Fig. 1. Top line: Swiss roll data points and the coordinates. Bottom line: 2D coordinates computed by LLE (left, $k = 18$), ISOMAP (middle, $k = 7$), and LTSA (right, $k = 6$).

3 Experimental Results

We have applied our LTSA algorithm to many data sets. The test data sets include curves in 2D/3D Euclidean spaces and surfaces in 3D Euclidean spaces. To show that our algorithm can also handle data points in high-dimensional spaces, we also consider curves and surfaces in Euclidean spaces with dimension equal to 100 and an image data set with dimension 4096. Here we report two applications of the LTSA algorithm.

First we apply LTSA, LLE [6], and ISOMAP [9] to the swissroll data set constructed as $x_i = [t_i \cos(t_i),\ s_i,\ t_i \sin(t_i)]^T$ with t_i and s_i are randomly chosen in the interval $(3\pi/2, 3\pi)$ and $(0,\ 21)$, respectively. We set $n = 1000$. LTSA always produces coordinates T that has similar geometric structure as the generating coordinates. There are little geometric deformations in the coordinates generated by LTSA, see the bottom in Figure 1 for the swissroll data set. The surface has zero Gaussian curvature, and therefore it can be flattened without any geometric deformation, i.e., the surface is *isometric* to a 2D plane. In the left and middle in the bottom line of Figure 1, we also plot the results for LLE and ISOMAP, the deformations (stretching and compression) in the generated coordinates are quite prominent.

Now we look at the results of applying LTSA algorithm to the face image data set [9]. The data set consists of a sequence 698 64-by-64 pixel images of a face rendered under various pose and lighting conditions. Each image is converted to an $m = 4096$ dimensional image vector. We apply LTSA with $k = 12$ neighbors and $d = 2$. The constructed coordinates are plotted in the middle of Figure 2. We also extracted four paths along the boundaries of of the set of the 2D coordinates, and display the corresponding images along each path. It can be seen that the computed 2D coordinates do capture very well the pose and lighting variations in a continuous way.

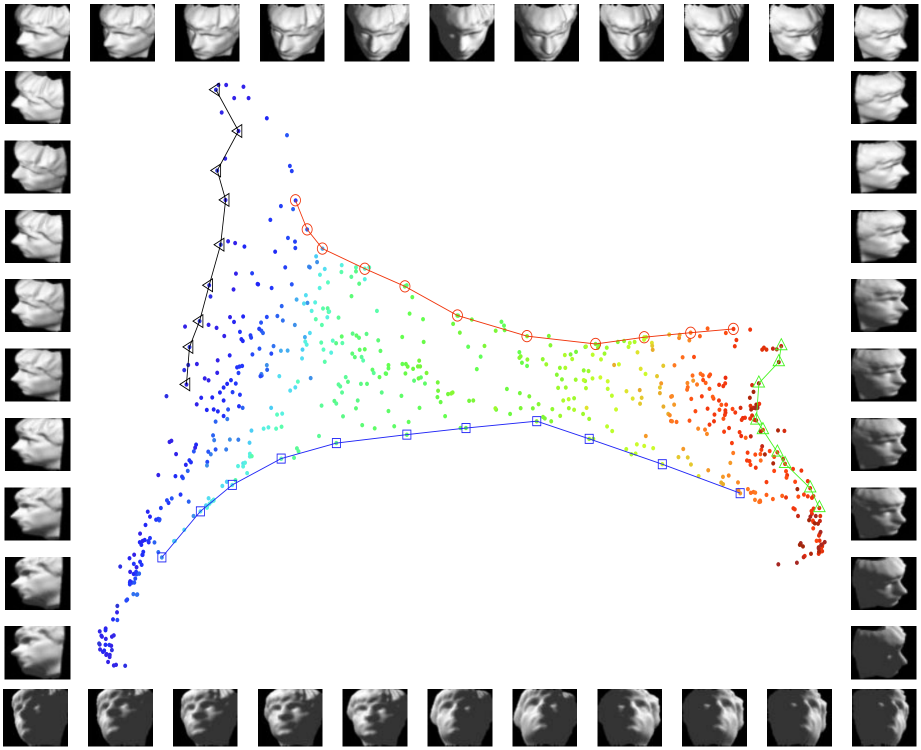

Fig. 2. Coordinates computed by Algorithm LTSA with $k = 12$ (middle) and images corresponding to the points on the bound lines (top, bottom, left, and right).

References

1. D. Donoho and C. Grimes. Local ISOMAP perfectly recovers the underlying parameterization for families of occluded/lacunary images. To appear in IEEE Computer Vision & Pattern Recognition, 2003.
2. G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, Maryland, 3nd edition, 1996.
3. T. Hastie and W. Stuetzle. Principal curves. *J. Am. Statistical Assoc.*, 84: 502–516, 1988.
4. T. Kohonen. *Self-organizing Maps*. Springer-Verlag, 3rd Edition, 2000.
5. T. Martinetz and K. Schulten. Topology representing networks. *Neural Networks*, 7: 507–523, 1994.
6. S. Roweis and L. Saul. Nonlinear dimension reduction by locally linear embedding. *Science*, 290: 2323–2326, 2000.
7. L. Saul and S. Roweis. Think globally, fit locally: unsupervised learning of nonlinear manifolds. Technical Reports, MS CIS-02-18, Univ. Pennsylvania, 2002.
8. Y. Teh and S. Roweis. Automatic Alignment of Hidden Representations. To appear in *Advances in Neural Information Processing Systems*, 15, MIT Press (2003).
9. J. Tenenbaum, V. De Silva and J. Langford. A global geometric framework for nonlinear dimension reduction. *Science*, 290:2319–2323, 2000.