Spectral Methods for Semi-supervised Manifold Learning

Zhenyue Zhang
Department of Mathematics, Zhejiang University
Hangzhou 310027, China
zyzhang@zju.edu.cn

Hongyuan Zha and Min Zhang
College of Computing, Georgia Tech
Atlanta, GA 30332
{zha, minzhang}@cc.gatech.edu

Abstract

Given a finite number of data points sampled from a low-dimensional manifold embedded in a high dimensional space together with the parameter vectors for a subset of the data points, we need to determine the parameter vectors for the rest of the data points. This problem is known as semi-supervised manifold learning, and in this paper we propose methods to handle this problem by solving certain eigenvalue-problems. Our proposed methods address two key issues in semi-supervised manifold learning: 1) fitting of the local affine geometric structures, and 2) preserving the global manifold structures embodied in the overlapping neighborhoods around each data points. We augment the alignment matrix of local tangent space alignment (LTSA) with the orthogonal projection based on the known parameter vectors, giving rise to the eigenvalue problem that characterizes the semi-supervised manifold learning problem. We also discuss the roles of different types of neighborhoods and their influence on the learning process. We illustrate the performance of the proposed methods using both synthetic data sets as well as data sets arising from applications in video annotations.

1. Introduction

Many problems in pattern recognition and computer vision generate high-dimensional data whose variations can be characterized by a small number of parameters. Manifold learning is a popular unsupervised learning approach aiming at recovering the low-dimensional parameterization from a finite set of high-dimensional data points sampled from the manifold [6, 7]. Several manifold learning algorithms have been proposed and applied to problems in pattern recognition and computer vision. One important issue, however, has largely been ignored, at least not explicitly addressed in the past: the fact that there exist infinitely many nonlinear parameterizations for the same manifold. If the parameterization is isometric, it makes sense to aim at recovering the isometric parameterization (up to a rigid motion). In the more general case, the problem seems to be ill-defined, which parameterization out of the infinitely many possible ones to aim at? We argue that in applications one is interested in the *semantically meaningful* parameterizations,[1] and extra information is needed to tell those parameterizations from the rest. One type of extra information is in the form of labeled data: a subset of the data points are labeled with the corresponding semantically meaningful parameter vectors, and we need to compute the semantically meaningful parameter vectors for the rest of the data points. This type of semi-supervised manifold learning problems are the focus of this paper. Some variations of the problems have been discussed before [3, 8], but we want to emphasize the crucial role it plays in recovering semantically meaningful and physically relevant parameterizations. We now proceed to a more formal discussion of the problems.

2. Semi-supervised manifold learning

Let $\mathcal{M}$ be a parameterized d-dimensional manifold embedded in $\mathcal{R}^D$ with $D \gg d$, i.e., there is a one-to-one function ϕ defined on a sub-domain Ω of $\mathcal{R}^d$ such that $\mathcal{M} = \phi(\Omega)$ [4]. We call ϕ a parameterization of $\mathcal{M}$. Obviously, the parameterization of $\mathcal{M}$ is not unique; for any one-to-one transformation ψ from a sub-domain Δ to Ω, $\phi \circ \psi$ is another parameterization of $\mathcal{M}$. Generally, we need to determine some parameterization ϕ which is semantically meaningful to the application we are considering.

Given data points $x_1, \ldots, x_N$ sampled from the manifold $\mathcal{M}$, we want to compute the corresponding semantically meaningful parameter vectors which we designate as $y_1, \ldots, y_N$. Generally, the y_i's may even live in $\mathcal{R}^n$ with $n \neq d$. To deal with this generality, we assume that $y_1, \ldots, y_N$ are sampled from another parameterized manifold $\mathcal{M}_s$ embedded in $\mathcal{R}^n$, and both $\mathcal{M}$ and $\mathcal{M}_s$ share the same d-dimensional parameter space Ω. So there are two functions ϕ and ψ defined on the same domain Ω such that

$$\mathcal{M} = \phi(\Omega), \quad \mathcal{M}_s = \psi(\Omega). \tag{1}$$

[1] The video annotation problem discussed in section is a very good illustrative example.

Furthermore, $x_i = \phi(\tau_i), y_i = \psi(\tau_i), i = 1, \ldots, N$. We see

$$y = \psi(\tau) = \psi(\phi^{-1}(x)) = \psi \circ \phi^{-1}(x) = g(x),$$

and $y_i = g(x_i)$. Now we can state the problem of semi-supervised manifold learning as follows.

Without loss of generalities, we assume $x_1, \cdots, x_\ell$ are labelled with parameter vectors $y_1, \cdots, y_\ell$, we want to determine the parameter vectors y_i for $x_i, i = \ell + 1, \ldots, N$, or more importantly, we also want to estimate the function $g : \mathcal{M} \to \mathcal{M}_s$ as accurately as possible from the data.

3. Local tangent space alignment (LTSA)

Our proposed semi-supervised manifold learning methods can be considered as an adaptation of LTSA to the semi-supervised setting. For this reason, we first give a brief review of LTSA [9]. Along the way we also introduce the concepts of *tangent coordinates* and the *alignment matrix* which are key ingredients of our proposed algorithms.

For a given set of sample points $x_1, \ldots, x_N$, we begin with building a connectivity graph on top of them which specifies, for each sample point, which subset of the sample points constitutes its neighborhood [6]. Let the set of neighbors for the sample point x_i be $X_i = [x_{i_1}, \ldots, x_{i_{k_i}}]$, including x_i itself. We approximate those neighbors using a d-dimensional (affine) linear space,

$$x_{i_j} \approx \bar{x}_i + Q_i \theta_j^{(i)}, \; Q_i = [q_1^{(i)}, \ldots, q_d^{(i)}], \quad j = 1, \ldots, k_i,$$

where $\bar{x}_i$ is the mean of the x_{i_j}'s, d is the dimension of the manifold, Q_i is an orthonormal basis matrix. The $\theta_j^{(i)}$ are the local *tangent coordinates* associated with the basis matrix, i.e., $\theta_j^{(i)} = Q_i^T(x_j - \bar{x}_i)$. Using the singular value decomposition of the centered matrix $X_i - \bar{x}_i e^T$, Q_i can be computed as the left singular vectors corresponding to the d largest singular values of $X_i - \bar{x}_i e^T$ [2].

We postulate that in each neighborhood, the corresponding global parameter vectors $T_i = [\tau_{i_1}, \ldots, \tau_{i_{k_i}}]$ differ from the local ones $\Theta_i = [\theta_1^{(i)}, \ldots, \theta_{k_i}^{(i)}]$ by a local affine transformation. The errors of the optimal affine transformation is then given by

$$\begin{aligned} & \min_{c_i, L_i} \sum_{j=1}^{k_i} \|\tau_{i_j} - (c_i + L_i \theta_j^{(i)})\|_2^2 \\ = \; & \min_{c_i, L_i} \|T_i - (c_i e^T + L_i \Theta_i)\|_F^2 = \|T_i \Phi_i\|_F^2, \end{aligned} \tag{2}$$

where Φ_i is the orthogonal projection whose null space is spanned by the columns of $[e, \Theta_i^T]$. We seek to compute the parameter vectors $\tau_1, \ldots, \tau_N$ by minimizing the following objective function,

$$\sum_i \|T_i \Phi_i\|_F^2 = \mathrm{trace}(T \Phi T^T) \tag{3}$$

over $T = [\tau_1, \ldots, \tau_N]$. Here

$$\Phi = \sum_{i=1}^N S_i \Phi_i S_i^T \tag{4}$$

is the *alignment matrix* with $S_i \in R^{N \times k_i}$, the 0-1 *selection matrix*, such that $T_i = T S_i$. Imposing certain normalization conditions on T such as $T^T T = I_d$ and $Te = 0$,[2] the corresponding optimization problem

$$\min\{\mathrm{trace}(T \Phi T^T) \mid T T^T = I_d, Te = 0, \}$$

is solved by computing the eigenvectors corresponding to the 2nd to $d+1$st smallest eigenvalues $\lambda_2, \cdots, \lambda_{d+1}$ of Φ, here the eigenvalues of Φ are arranged in nondecreasing order, i.e., $\lambda_1 = 0 \le \lambda_2 \le \cdots \le \lambda_{d+1} \le \cdots \le \lambda_N$. The first $d+1$ smallest eigenvalues are quite small generally, compared with others. We mention that if the parameterization ϕ is isometric, then under certain conditions, LTSA can recover the isometric parameterization.

4. Local structures of parameterizations

A manifold can admit many different parameterizations. However, each parameterization locally can be approximately characterized by the tangent coordinates discussed above. This property motivates us to consider a spectral solution for the semi-supervised problem: find a global parameter system that minimizes this kind of local approximation errors (on all labelled and unlabelled points) as well as the least squares approximation errors on labels. In this section, we first give a detailed analysis for the local structures of parameterizations.

Consider a parameterization $\phi : \Omega \to \mathcal{M}$ of the manifold $\mathcal{M}$ with $\Omega \subset \mathcal{R}^d$. Let $\mathcal{T}_{\hat{x}}$ be the tangent space of $\mathcal{M}$ at $\hat{x} = \phi(\hat{\tau})$. Consider a small neighborhood $\Omega(\hat{\tau})$ of $\hat{\tau}$ such that the orthogonal projection of $\phi(\tau)$ onto $\mathcal{T}_{\hat{x}}$ is one-to-one. Let θ be the tangent coordinate of $\phi(\tau)$ corresponding to an orthogonal basis $Q_{\hat{x}}$ of $\mathcal{T}_{\hat{x}}$, i.e.,

$$\theta = Q_{\hat{x}}^T(\phi(\tau) - \phi(\hat{\tau})), \quad \tau \in \Omega(\hat{\tau}). \tag{5}$$

Consider the Taylor expansion of first order of ϕ at point $\hat{\tau}$,

$$\phi(\tau) \;\; = \;\; \phi(\hat{\tau}) + J_{\hat{\tau}}(\tau - \hat{\tau}) + \epsilon(\tau, \hat{\tau}). \tag{6}$$

Here $J_{\hat{\tau}}$ is the gradient matrix of ϕ at $\hat{\tau}$ and $\epsilon(\tau, \hat{\tau})$ the second order term of $\|\tau - \hat{\tau}\|$,

$$\epsilon(\tau, \hat{\tau}) = \frac{1}{2} J_{\hat{\tau}}^T H_{\hat{\tau}}(\tau - \hat{\tau}, \tau - \hat{\tau}) + o(\|\tau - \hat{\tau}\|^2)$$

with the Hessian tensor $H_{\hat{\tau}}$ of ϕ at $\hat{\tau}$, respectively. It follows from (6) that

$$\begin{aligned} \tau - \hat{\tau} \;\; = \;\; & (J_{\hat{\tau}}^T J_{\hat{\tau}})^{-1} J_{\hat{\tau}}^T \big(\phi(\tau) - \phi(\hat{\tau})\big) \\ & + (J_{\hat{\tau}}^T J_{\hat{\tau}})^{-1} \epsilon(\tau, \hat{\tau}). \end{aligned} \tag{7}$$

[2] e is a vector of all 1's with suitable dimension.

Note that $J_{\hat\tau} = Q_{\hat\tau}P_{\hat\tau}$ with nonsingular matrix $P_{\hat\tau} = Q_{\hat\tau}^T J_{\hat\tau}$. Substituting it into (7) and using (5), we obtain that

$$\tau - \hat\tau = P_{\hat\tau}^{-1}\theta + (P_{\hat\tau}^T P_{\hat\tau})^{-1}\epsilon(\tau,\hat\tau).$$

Equivalently,

$$\tau = \hat\tau + P_{\hat\tau}^{-1}\theta + (P_{\hat\tau}^T P_{\hat\tau})^{-1}\epsilon(\tau,\hat\tau). \tag{8}$$

The equality (8) shows the affine relation between the local tangent coordinates θ of x, which can be estimated from a neighbor set of x, and the global coordinates τ. Note that the affine transformation $\hat\tau + P_{\hat\tau}^{-1}\theta$ is also local. The "affine" difference between the two coordinates is of second order of $\|\tau - \hat\tau\|$,

$$\begin{aligned}&\|\tau - (\hat\tau + P_{\hat\tau}^{-1}\theta)\| \le \sigma_{\min}(J_{\hat\tau})^{-2}\|\epsilon(\tau,\hat\tau)\| \\ \le\ & \frac{\|J_{\hat\tau}^T H_{\hat\tau}\|}{2\sigma_{\min}^2(J_{\hat\tau})}\|\tau-\hat\tau\|^2 + o(\|\tau-\hat\tau\|^2).\end{aligned} \tag{9}$$

Here $\sigma_{\min}(\cdot)$ denotes the smallest singular value of a matrix and we have used $P_{\hat\tau}^T P_{\hat\tau} = J_{\hat\tau}^T J_{\hat\tau}$. Furthermore, the error bound above is $o(\|\tau-\hat\tau\|^2)$ for isometric ϕ since $J_\tau^T H_\tau(u,v) = 0$ for any τ, u, and v.

The semantically meaningful parameter vector y of a manifold $\mathcal{M}$ generally will have dimension different from the intrinsic dimension of $\mathcal{M}$. Referring to (1), by the Taylor expansion of ψ,

$$\psi(\tau) = \psi(\hat\tau) + J_{\hat\tau}^{\psi}(\tau-\hat\tau) + \epsilon^{\psi}(\tau,\hat\tau),$$

and using (8), we see that

$$y = \hat y + J_{\hat\tau}^{\psi}P_{\hat\tau}^{-1}\theta + J_{\hat\tau}^{\psi}(P_{\hat\tau}^T P_{\hat\tau})^{-1}\epsilon(\tau,\hat\tau) + \epsilon^{\psi}(\tau,\hat\tau).$$

Therefore, we also have the following estimation for a general parameterization,

$$\|y - (\hat y + L_{\hat\tau}\theta)\| \lessapprox c_{\phi,\psi}(\hat\tau)\|\tau-\hat\tau\|^2 \tag{10}$$

with

$$c_{\phi,\psi}(\hat\tau) = \frac{1}{2}\left\{\frac{\|J_{\hat\tau}^{\psi}\|\|J_{\hat\tau}^T H_{\hat\tau}\|}{\sigma_{\min}^2(J_{\hat\tau})} + \|(J_{\hat\tau}^{\psi})^T H_{\hat\tau}^{\psi}\|\right\}. \tag{11}$$

We summarize the above discussion in the following theorem.

Theorem 1 *Assume that $\phi:\Omega\to\mathcal{M}$ is a smooth function. $\Omega(\hat\tau)$ is a neighborhood of $\hat\tau\in\Omega$ and $\rho(\Omega(\hat\tau))$ denotes its diameter. Let $y=\psi(\tau)\in\mathcal{R}^m$ is a user interested parameter variable of the manifold. If ϕ is not rank-deficient at $\hat\tau$, i.e. $\sigma_{\min}(J_{\hat\tau})\neq 0$, then the coordinate y is affine equal to its tangent coordinate θ with error $O(\rho^2(\Omega(\hat\tau)))$ in $\Omega(\hat\tau)$. Furthermore, if both ϕ and ψ are isometric, the error is $o(\rho^2(\Omega(\hat\tau)))$.*

5. Spectral methods

Theorem 1 shows that the local structures of the desired parameterization $y=g(x)$ with $g=\psi\circ\phi^{-1}$ can be recovered by an affine transformation of the local tangent coordinates. Denote by z_i *estimates* of $y_i = g(x_i)$, $i=1,\cdots,N$. Let $x_{i_1},\cdots,x_{i_{k_i}}$ be a set of neighbors of x_i with the corresponding $y_{i_1},\ldots,y_{i_{k_i}}\in\Omega(y_i)$, a neighborhood of y_i. The affine transformation $\hat y_i + L_{\hat\tau_i}^{-1}\theta$ in (10) can be estimated as

$$(c+L\theta) = \arg\min_{c,L}\sum_{j=1}^{k_i}\|z_{i_j} - (c+L\theta_j^{(i)})\|^2,$$

here $\theta_j^{(i)}$ are the local tangent coordinates. Then by (10)

$$\begin{aligned}&\frac{1}{k_i}\mathrm{trace}(Z_i\Phi_i Z_i^T) = \frac{1}{k_i}\sum_{j=1}^{k_i}\|z_{i_j} - (c+L\theta_j^{(i)})\|^2 \\ \approx\ & c_{\phi,\psi}^2(\tau_i)\rho^4(\Omega(\tau_i)) + o(\rho^4(\Omega(\tau_i))).\end{aligned}$$

Here $Z_i = [z_{i_1},\cdots,z_{i_{k_i}}]$, and Φ_i is the orthogonal projection whose null space is $\mathrm{span}([e,\Theta_i^T])$ with $\Theta_i = [\theta_1^{(i)},\cdots,\theta_{k_i}^{(i)}]$. Summing over all the data points, and let Φ be the alignment matrix in (4),

$$\begin{aligned}&\frac{1}{N}\mathrm{trace}(Z\Phi Z^T) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{k_i}\sum_{j=1}^{k_i}\|z_{i_j} - (c+L\theta_j^{(i)})\|^2 \\ \le\ & \max_i\left\{c_{\phi,\psi}^2(\tau_i)\rho^4(\Omega(\tau_i)) + o(\max_i\rho^4(\Omega(\tau_i)))\right\}.\end{aligned}$$

The above analysis shows that if g has bounded gradient and Hessian, the parameter vectors y_i give rise to small $\mathrm{trace}(Z\Phi Z^T)$. We emphasize that Φ embodies two requirements: 1) local affine equivalence to the tangent coordinates as illustrated above; and 2) smooth global manifold structure through the overlaps among the neighborhoods.

Besides a small $\mathrm{trace}(Z\Phi Z^T)$, the estimate Z should also satisfy the constraints that it should be close to the known parameter vectors. Without loss of generality, we assume that $Y_L = [y_1,\cdots,y_\ell], \ell \ll N$ are known. Then $Z_L = [z_1,\cdots,z_\ell]$ should be close to Y_L. This can be achieved by minimizing the difference between Z_l and an affine transformation of Y_L,

$$\begin{aligned}&\min_{c,L}\frac{1}{\ell}\sum_{i=1}^{\ell}\|z_i - (c-Ly_i)\|^2 \\ =\ & \frac{1}{\ell}\mathrm{trace}(Z_L P_{Y_L} Z_L^T) = \frac{1}{\ell}\mathrm{trace}(ZS_L P_{Y_L} S_L^T Z^T),\end{aligned}$$

where P_{Y_L} is the orthogonal projection whose null space is $\mathrm{span}([e,Y_L^T])$.

We now combine the two types of constraints to arrive at following optimization problem,

$$\min_{\substack{ZZ^T=I\\ Ze=0}}\frac{1}{N}\mathrm{trace}(Z\Phi Z^T) + \frac{\lambda}{\ell}\mathrm{trace}(ZS_L P_{Y_L} S_L^T Z^T), \tag{12}$$

which amounts to computing the 2nd to (d+1)-st smallest eigenvectors $u_2, \cdots, u_{\ell+1}$ of the symmetric and positive semi-definite matrix

$$\Psi = \Phi + \beta S_L P_{Y_L} S_L^T \tag{13}$$

with $\beta = \frac{\lambda N}{\ell}$, since e is a trivial eigenvector of Ψ corresponding to the zero eigenvalue.

Affine Transformation. The constraint $ZZ^T = I_d$ is a convenient one to use in (12), but the resulting Z needs an affine transformation $z \to c + Wz$ in order to match the known parameter vectors. There are several possible ways to do this. We present a simple one as follows. We compute

$$(c + Wz) = \arg\min_{c,W} \sum_{j=1}^{p} \|y_j - (c + Wz_j)\|_2^2. \tag{14}$$

Its solution is $[c, W]Q_{11}^T = Y_1$, where Q_{11} is the submatrix of rows in $U_1 = [u_1, \cdots, u_{\ell+1}]$ corresponding to the labeled set. In terms of numerical stability, the labeled set should be selected such that Q_{11} has as small a condition number as possible. This is the main idea of AE selection given in [8]. Note that here Q_{11} also depends on the selection of the labelled set. So it is hard to select the prior points in the "optimal" way. In general, if Q_{11} is close to singular, one may regularize the linear system using a small $\eta > 0$ and obtain the regularized solution,

$$[c, W] = Y_1 Q_{11}(Q_{11}^T Q_{11} + \eta \|Q_{11}\|^2 I)^{-1}.$$

Once c and W are computed, the $\{z_i\}$ are then affinely transformed to $\tilde{z}_i = c + Wz_i, i = \ell + 1, \cdots, N$ with $Z_2 = [z_{\ell+1}, \cdots, z_N]$.

6. Some improvements

We now present some improvements over the approach outlined in the above section by taking a closer look at the different roles played by the various data points in their contribution to the alignment matrix Φ. To this end, we split Φ into three parts: 1) neighborhoods of labeled points, 2) neighborhoods containing at least a labeled points, and 3) other neighborhoods,

$$\Phi = \sum_{j=1}^{\ell} S_j \Phi_j S_j^T + \sum_{i=\ell+1}^{p} S_i \Phi_i S_i^T + \sum_{i=p+1}^{N} S_i \Phi_i S_i^T,$$

where we assume that the neighborhood of each x_i, $i = \ell + 1, \cdots, p$, contains at least one labeled point and other neighbor sets of x_i, $i > p$, have no labeled points in their neighbors. We will modify Φ by introducing two tuning parameters $\alpha = [\alpha_1, \alpha_2]$ with $1 > \alpha_1, \alpha_2 > 0$,

$$\Phi(\alpha) = \alpha_1 \sum_{j=1}^{\ell} S_j \Phi_j S_j^T + \sum_{i=\ell+1}^{p} S_i \Phi_i S_i^T + \alpha_2 \sum_{i=p+1}^{N} S_i \Phi_i S_i^T.$$

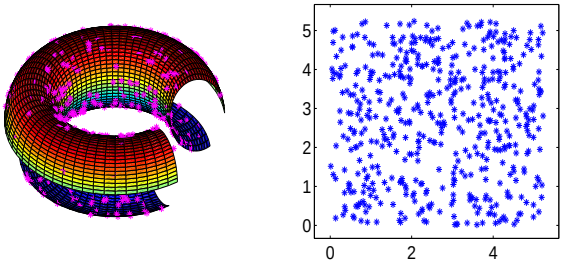

Figure 1. The incomplete tire manifold with samples (left) and generating coordinates of the samples (right).

The basic idea is that $x_i, i = \ell + 1, \cdots, p$ are the neighborhoods that connect the labeled data points with the unlabeled data points, and should be assigned larger weights. Then the eigen-problem (12) becomes

$$\min_{ZZ^T = I, Ze = 0} Z\Psi(\alpha, \beta)Z^T, \tag{15}$$

where $\Psi(\alpha, \beta) = \Phi(\alpha) + \beta S_L P S_L^T$ with $\beta = \frac{\lambda \ell}{N}$.

7. Numerical experiments

We present several numerical expamples to illustrate our proposed algorithms and compare them with the least squares (LS) method [8] and the Laplacian regression method (LapRLS) [1]. The data selection method is the one used in [8] which is briefly discussed at the end of section 5. Two data sets are considered: one is a toy data set from a non-isometric 2D manifold embedded in 3D space. The other is a video sequence and we are interested in annotations of the frames in the video sequence.

Example 1. Consider a section of a torus with a middle slice removed and its parameterization given by

$$g(s,t) = \begin{bmatrix} (3 + \cos s)\cos t \\ (3 + \cos s)\sin t \\ \sin s \end{bmatrix}, \quad (s,t) \in [0, \frac{5\pi}{3}]^2.$$

It is a 2D manifold embedded in $\mathcal{R}^3$ and is not isometric ($d = 2,\ D = 3$). We sample $N = 500$ points $x_j = g(s_j, t_j)$ from the manifold with the generating parameter vectors (s_j, t_j) uniformly distributed in the domain $[0, \frac{5\pi}{3}]^2$. Figure 6 shows the manifold, the sample points and their generating parameters.

We labeled $\ell = 50$ of the data points with their corresponding 2D generating parameter coordinates, and applied the proposed spectral method with 8 neighbors of each x_i (including x_i itself) to recover the parameters (s,t), and $\alpha_1 = 2\alpha, \alpha_2 = \alpha$ with a range of values of α and λ. The spectral approach works well and is not sensitive to the choice of the turning parameters α and β. In Tables 1 we list the value of $10^2\epsilon(\alpha, \beta)$, where $\epsilon(\alpha, \beta)$ is the relative Frobenius norm of the total errors of unlabeled points Z_{UL},

$$\epsilon(\alpha, \beta) = \|Z_{UL} - Y_{UL}\|_F / \|Y_{UL}\|_F$$

with respect to different α's and β's. We set $d = 2$ for construct the alignment matrix $\Phi(\alpha)$. We point out that the optimal parameters α and β depend on the selection of the prior points when the number of labelled points is fixed.

Table 1. $10^2\epsilon(\alpha, \beta)$ of the eigen-system approach with LM selection.

$\alpha \backslash \beta$	10^{-1}	10^0	10^1	10^2	10^3	10^4
0.01	2.12	1.56	1.47	1.46	1.46	1.46
0.02	2.17	1.49	1.39	1.38	1.38	1.38
0.03	2.32	1.49	1.38	1.37	1.37	1.37
0.04	2.52	1.53	1.40	1.39	1.38	1.38
0.05	2.72	1.59	1.44	1.42	1.42	1.42

We also compared our spectral method with the LS method in [8] and LapRLS in [1]. A range of parameters were used with 'rbf' kernel and 'heat' weights for Laplacian graph, including the neighborhood size k of each point x_i (not including x_i itself) listed as follows,

$$\begin{aligned} KernelParam &\in [10^{-3},\ 10^3], \\ GraphWeightParam &\in [10^{-4},\ 10^3], \\ \gamma_A &\in [10^{-4},\ 10^3], \\ \gamma_I &\in [10^{-3},\ 10^3]. \end{aligned}$$

The smallest result is achieved with error 0.0715 at $k = 10$, kernel parameter 1.5, graph weight parameter 1000, $\gamma_A = 0.0005$, and $\gamma_I = 0.004$. Generally, LapRLS is quite sensitive to the choice of the parameters. The smallest error is much larger than the error 0.01365 by our spectral method using same labelling. The error of the LS method is 0.03363. Figure 2 plots the computed generating coordinates by the three methods: the spectral method, the LS method and LapRLS.

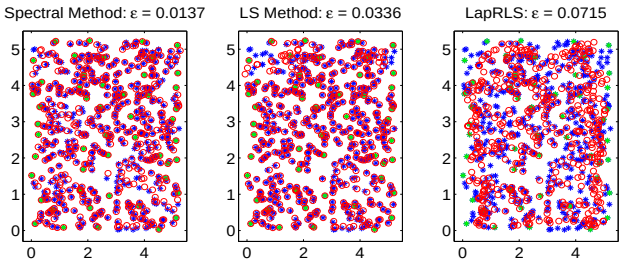

Figure 2. The computed coordinates (red circles) by our spectral method (left), the LS method (middle), and LapRLS (right), together with the generating coordinates (blue dots) of the incomplete tire.

Example 2. This is video sequence consisting of 393 frames of a subject sitting in a sofa and waving his arms.[3]

[3]From http://www.csail.mit.edu/ rahimi/manif. The original data set has 984 frames. We deleted identical frames from the data set.

Table 2. Numbers of correct position vectors of *unlabeled* points computed by spectral method (top rows), LS method (middle row), and LapRLS (bottom rows) within given accuracy.

$\ell \backslash \eta$	0.03	0.05	0.07	0.09	0.11	0.13	0.15
20	23	84	142	218	287	341	360
	5	29	60	116	194	279	328
	0	12	64	151	242	306	337
30	39	116	218	265	303	324	338
	22	68	115	193	264	311	336
	2	36	119	203	286	326	344
40	35	133	225	274	291	305	323
	30	91	176	267	302	330	343
	8	75	175	268	308	326	340
50	50	192	280	314	330	336	342
	35	139	240	309	326	332	336
	7	98	210	277	312	330	337

We are interested in annotating the frames with the positions of the subject's arms that are determined by the four points of the elbows and wrists in the frames. So y_i are now 8-dimensional vectors containing the four 2D points. The relative error function of a computed vector z_i is defined as

$$\epsilon_\ell(z_i) = ||z_i - y_i||/a, \quad i = \ell + 1, ..., N, \tag{16}$$

where a is the diameter of image rectangle.

In table 2, we list the number of the computed position vectors of the unlabeled points satisfying a given accuracy threshold η, i.e., $\epsilon_\ell(z_i) < \eta$. The turning parameters of the spectral method are simply set as $\alpha_1 = \alpha_2 = 1$ and $\beta = 200$ for all ℓ. We search the parameter space for the 'rnf' kernel parameter, γ_A and γ_I in the interval $[10^{-3}, 10^3]$ and choose the following: 'rbf' kernel parameter 100, heat weight parameter 10^{-3}, and $\gamma_A = 10^{-3}$, $\gamma_I = 10^{-3}$, which give the best results. Figure 3 plots the computed arm positions (blue bars) by the spectral method ($\ell = 50$) against the manually labeled positions (red dots) on several illustrative examples with increasing amount of error.

Our spectral method performs better than the LS method and LapRLS. First, it has higher accuracy on the unlabeled images as shown in Table 2. Second, our method is not as sensitive to the turning parameters. For example, the spectral method gives almost the same results for $(\alpha_1, \alpha_2, \beta) \in [1,\ 100] \times [1,\ 100] \times [50,\ 500]$ if $\ell = 50$. We remark that the factor $\tilde{\gamma}_I = \frac{\gamma_I}{N^2}$ of the term used in the Laplacian matrix in (35) of [1] is very small, $\tilde{\gamma}_I = 6.4746 \times 10^{-9}$ for the optimal γ_I. It seems that the Laplacian does not play a positive role for this real data set. If we increase $\tilde{\gamma}_I$ to 10^{-4} with the others unchanged, the smallest position error $\min \epsilon_{30}(z_i)$ of unlabeld points computed by LapRLS is 0.085. Figure 4 plots the error cures for the spectral method, LS method, and LapRLS method ($\ell = 30$) with $\frac{\gamma_I}{N^2} = 10^{-6}, 10^{-5}, \frac{1}{2}10^{-4}, 10^{-4}$ while other tuning param-

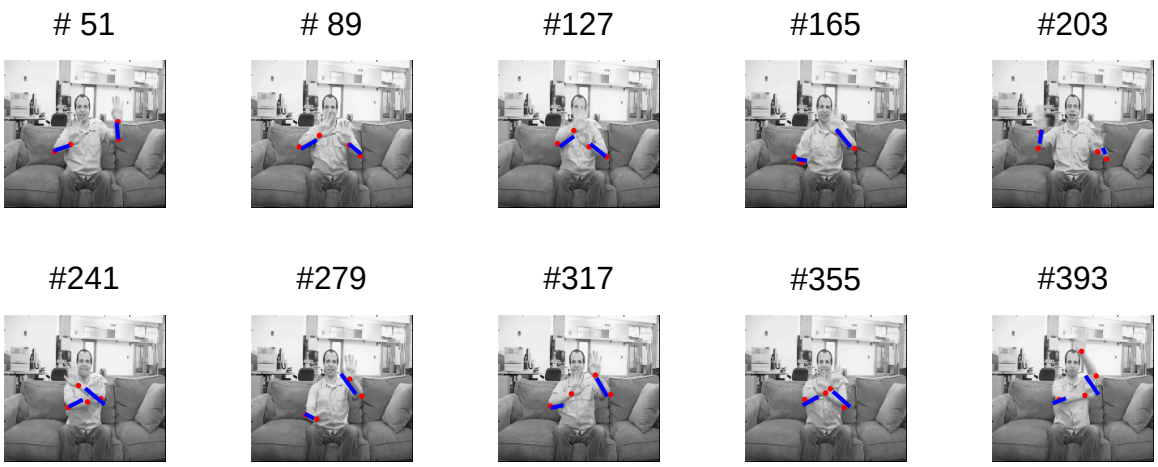

Figure 3. Computed arm-positions (blue bars) by the spectral method ($\ell = 50$) and labelled points (red dots) in increasing order of errors.

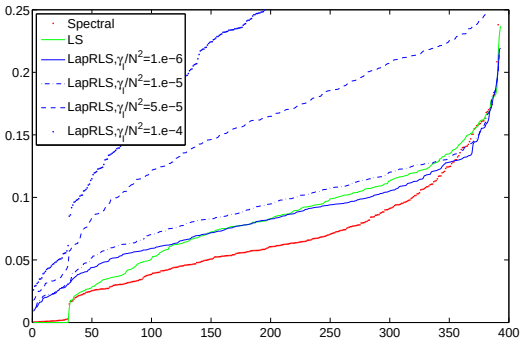

Figure 4. The error curves of the spectral method, LS method, and LapRLS method with $\frac{\gamma_I}{N^2} = 10^{-6}, 10^{-5}, \frac{1}{2}10^{-4}, 10^{-4}, \ell = 30$.

eters remain unchanged.

8. Conclusion

We proposed methods for semi-supervised manifold learning by solving eigenvalue problems. We emphasize the important role played by semi-supervised manifold learning in computing semantically meaningful parameterizations of manifolds embedded in high-dimensional spaces. Several topics need further investigation: 1) developing active learning methods to select the data points to label for the proposed spectral methods; and 2) explore other problem structures such as the dynamics of the scene in video annotations for semi-supervised manifold learning.

Acknowledgments. The work was supported in part by NSFC grant 10771194 and NSF grant DMS-0736328.

References

[1] M. Belkin, P. Niyogi and V. Sindhwani. Manifold regularization: a geometric framework for learning from Labeled and Unlabeled Examples. *Journal of Machine Learning Research*, 6: 2399–2424, 2006. 4, 5

[2] G. H. Golub and C. F Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, Maryland, 3nd edition, 1996. 2

[3] J. Ham, I. Ahn and D. Lee. Learning a manifold-constrained map between image sets: applications to matching and pose estimation. *CVPR*, 2006. 1

[4] J. Munkres. *Analysis on Manifold* Addison Wesley, Redwood City, CA, 1990. 1

[5] A. Rahimi, B. Recht and T. Darrell. Learning appearance manifolds from video. CVPR, 2005.

[6] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290: 2323–2326, 2000. 1, 2

[7] J. Tenenbaum, V. De Silva and J. Langford. A global geometric framework for nonlinear dimension reduction. *Science*, 290:2319–2323, 2000. 1

[8] X. Yang, H. Fu, H. Zha and J. Barlow. Semi-supervised nonlinear dimensionality reduction. *ICML*, 2006. 1, 4, 5

[9] Z. Zhang and H. Zha. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM J. Scientific Computing*. 26:313-338, 2005. 2