

# 一种改进的局部切空间排列算法\*

杨 剑<sup>1,2+</sup>, 李伏欣<sup>1,2</sup>, 王 珏<sup>1</sup>

<sup>1</sup>(中国科学院 自动化研究所,北京 100080)

<sup>2</sup>(中国科学院 研究生院,北京 100049)

## A Better Scaled Local Tangent Space Alignment Algorithm

YANG Jian<sup>1,2+</sup>, LI Fu-Xin<sup>1,2</sup>, WANG Jue<sup>1</sup>

<sup>1</sup>(Institute of Automation, The Chinese Academy of Sciences, Beijing 100080, China)

<sup>2</sup>(Graduate School, The Chinese Academy of Sciences, Beijing 100049, China)

+ Corresponding author: Phn: +86-10-82616599, E-mail: jian.yang@mail.ia.ac.cn

Received 2004-11-26; Accepted 2005-05-27

Yang J, Li FX, Wang J. A better scaled local tangent space alignment algorithm. *Journal of Software*, 2005,16(9):1584–1590. DOI: 10.1360/jos161584

**Abstract:** Recently, a new manifold learning algorithm, LTSA (local tangent space alignment), has been proposed. It is efficient for many nonlinear dimension reduction problems but unfit for large data sets and newcome data. In this paper, an improved algorithm called partitional local tangent space alignment (PLTSA) is presented, which is based on VQPCA (vector quantization principal component analysis) and LTSA. In the algorithm, the sample space is first divided into overlapping blocks using the X-Means algorithm. Then each block is projected to its local tangent space to get local low-dimensional coordinates of the points in it. At last, the global low-dimensional embedded manifold is obtained by local affine transformations. PLTSA is better than VQPCA in that it gives the global coordinates of the data. It works on a much smaller optimization matrix than that of LTSA and leads to a better-scaled algorithm. The algorithm also provides a set of transformations that allow to calculate the global embedded coordinates of the newcome data. Experiments illustrate the validity of this algorithm.

**Key words:** dimensionality reduction; manifold learning; principal component analysis; local principal component analysis; local tangent space alignment; X-means

**摘 要:** 局部切空间排列算法(local tangent space alignment,简称 LTSA)是一种新的流形学习算法,能有效地学习出高维采样数据的低维嵌入坐标,但也存在一些不足,如不能处理样本数较大的样本集和新来的样本点.针对这些缺点,提出了一种基于划分的局部切空间排列算法(partitional local tangent space alignment,简称 PLTSA).它建立在 VQPCA(vector quantization principal component analysis)算法和 LTSA 算法的基础上,利用 X-均值算法把样本空间划

---

\* Supported by the National Natural Science Foundation of China under Grant No.60175023 (国家自然科学基金); the National Grand Fundamental Research 973 Program of China under Grant No.2004CB318103 (国家重点基础研究发展规划(973))

作者简介: 杨剑(1976 - ),男,四川简阳人,博士生,主要研究领域为流形学习,维数约简;李伏欣(1984 - ),男,博士生,主要研究领域为流形学习,Boosting 算法,蛋白质组学;王珏(1948 - ),男,研究员,博士生导师,主要研究领域为粗糙集理论,数据挖掘,人工智能及其应用.

分成一些相互有重叠的块,通过把样本点投影到它所在块的局部切空间上得到其局部低维坐标,对局部低维坐标施加平移、旋转、伸缩变换,求出整体低维坐标. PLTSA 解决了 VQPCA 不能求出整体低维坐标和 LTSA 中大规模矩阵的特征值分解问题,且能够有效处理新来的样本点,这是很多流形学习算法所不能的.通过实验说明了 PLTSA 的有效性.

关键词: 维数约简;流形学习;主成分分析;局部主成分分析;局部切空间排列; $X$ -均值

中图法分类号: TP18 文献标识码: A

维数约简一直是模式识别、机器学习、多元数据分析等领域的重要研究课题之一.特别是随着信息时代的到来,人们获取大数据量、高维数、非结构化的数据变得越来越容易,这就使数据约简变得更为迫切.传统的维数约简方法之一是对样本特征进行线性组合,如主成分分析(principal component analysis,简称 PCA).PCA 是一种经典且广泛应用的线性维数约简方法,它通过把样本点投影到由样本散布矩阵最大特征值所对应的特征向量张成的低维线性子空间来实现维数约简.

虽然 PCA 思想简单,易于理解,执行速度快,能够真实地反映数据的线性结构,但是它对于学习嵌套于一个低维流形的高度非线性、属性强相关数据的内在几何结构却是无能为力的.而实际中的很多高维数据都是嵌套于一个低维非线性流形的,且这种低维流形可能与人的认知流形是一致的<sup>[1]</sup>.例如,几百维的人脸图像实际上是由光线亮度、人离相机的距离、人的头部姿势、人的面部肌肉等几个因素决定的<sup>[2]</sup>,这还可能是人们能够在不同时间、不同环境下快速识别一个人的原因.近年来,很多学者都在致力于研究能够学习高维数据所在低维流形的流形学习算法,且一致认为它将是一种非常有效的非线性维数约简方法<sup>[3]</sup>.假设数据是均匀采样于一个高维欧氏空间中的低维流形,流形学习的目标是从高维采样数据中恢复低维流形结构,即找到高维空间中的低维流形,并求出相应的嵌入映射,以实现维数约简或者数据可视化.它是从观测到的现象中去寻找事物的本质,找到产生数据的内在规律.当然,要达到这一目标必须对低维流形或者嵌入映射做适当的约束或限制,而不同的约束和限制也就得到不同的流形学习算法.等距映射算法(Isomap)<sup>[4]</sup>是用测地距离代替欧氏距离来改进多维尺度变换(multidimensional scaling,简称 MDS)的一种算法.它假定嵌入映射是等距映射.局部线性嵌入算法(locally linear embedding,简称 LLE)<sup>[5]</sup>假设采样数据所在低维流形是局部线性的,利用近邻构建局部几何,然后寻找能够在平移、旋转、伸缩变换下保持局部几何结构的局部线性映射,从而达到数据约简的目的.拉普拉斯特征映射算法(Laplacian Eigenmap)<sup>[6]</sup>是以把高维空间中离得很近的点映射到低维空间中也离得很近为目的,利用图拉普拉斯算子的性质进行求解的一种流形学习算法.局部追踪嵌入算法(locality pursuit embedding,简称 LPE)<sup>[7]</sup>通过局部 PCA 投影来逼近局部切空间的投影,但它只是一种学习低维流形的线性方法.局部切空间排列算法(local tangent space alignment,简称 LTSA)<sup>[8]</sup>通过逼近每一样本点的切空间来构建低维流形的局部几何,然后利用局部切空间排列求出整体低维嵌入坐标.

LTSA 是一种很好的流形学习算法,能够有效地学习出体现数据集低维流形结构的整体嵌入坐标,但是它也存在一些不足:(1) 算法中用于特征值分解的矩阵的阶数等于样本数,对于较大的样本集将无法处理;(2) 算法对于新来的样本点不能进行有效的处理.另一方面,基于向量量化(vector quantization,简称 VQ)的局部主成分分析方法(vector quantization principal component analysis,简称 VQPCA)<sup>[9]</sup>只是求出了样本点的低维局部坐标,而不能求出整体坐标,所以不同的点可能有相同的局部坐标,这将给进一步在低维空间中的研究和处理带来很大困难.针对这两种算法的缺点,我们提出了一种新的局部主成分分析算法(partitional local tangent space alignment,简称 PLTSA),它建立在 LTSA 的基础上.算法用  $X$ -均值<sup>[10]</sup>把样本空间划分成相对于样本数很少且有重叠的块,将每一块中的点投影到它所在块的局部切空间中得到局部坐标,再把局部坐标经过平移、旋转、伸缩变换,排列成整体低维嵌入坐标. PLTSA 减小了 LTSA 中用于特征值分解的矩阵的规模,从而适用于更多的数据集,特别地,更适合于样本空间的维数不是很高但样本数较大的数据集. PLTSA 在存储少量数据的条件下能对新来的样本进行有效的处理,这是目前大部分流形学习算法所做不到的,且 PLTSA 算法更能体现出数学上流形概念的局部坐标化本质<sup>[11]</sup>.

本文第 1 节简单介绍 VQPCA,LTSA 和  $X$ -均值这 3 种算法.第 2 节详细推导 PLTSA.第 3 节给出 PLTSA 的

几个实验结果.第4节对全文进行总结.

## 1 基于向量量化的局部主成分分析,局部切空间排列算法及 $X$ -均值算法

由于 PLTSA 建立在 VQPCA 和 LTSA 基础上,且利用 LTSA 中通过对局部坐标进行平移、旋转、伸缩变换求取整体嵌入坐标的思想,我们在第 1.1 节和第 1.2 节中分别简单介绍 VQPCA 和 LTSA.为了减少参数,使 PLTSA 能自动求出空间划分的块数,我们将利用  $X$ -均值算法求出数据集的基于贝叶斯信息标准(Bayesian information criterion,简称 BIC)的最优聚类数.第 1.3 节对  $X$ -均值作了简单介绍,更详细的内容可以参见文献[10].

### 1.1 基于向量量化的局部主成分分析(VQPCA)

Kambhatla 等人在文献[9]中提出了 VQPCA.它是对经典 PCA 的一种基于局部线性化的非线性推广.在给定的参考中心个数的条件下,VQPCA 通过竞争学习算法找到参考向量 VQ(一般采用  $K$ -均值算法),用 VQ 把样本空间划分为不相交的块,对每一块进行 PCA,求出低维坐标.文献[9]还提出了用重构误差作为距离度量求取参考向量的改进 VQPCA 算法.

### 1.2 局部切空间排列算法(LTSA)

LTSA 是一种基于切空间的流形学习算法.假定样本集是采样于某个参数流形且含有噪声的无序点集,LTSA 用近似的局部切空间来表示样本点所在流形的局部几何.而 Min 等人在文献[7]中证明了局部切空间可以用局部样本协方差矩阵的特征向量来表示,因此可以把求取表示局部信息的样本点在局部切空间中的投影坐标的问题转换为局部主成分分析问题.LTSA 用  $K$ -近邻求出每一个样本点的局部散布矩阵,对其进行特征值分解,求出局部主成分,通过向主成分投影得到样本点的低维局部坐标.算法进一步假定样本点的低维整体嵌入坐标可以由局部坐标经过一些变换得到,具体为平移、旋转、伸缩变换.最后,经过一系列数学推导,LTSA 把求解整体嵌入坐标问题转换为矩阵的特征值问题.

### 1.3 $X$ -均值算法

$X$ -均值是推广和完善经典  $K$ -均值的一种聚类算法.虽然  $K$ -均值是一种思想简单、具有很好的局部收敛性和广泛应用的聚类算法,但是它也存在计算复杂度高、运行时间长、算法中的  $K$  值必须事先给定及可能收敛到局部最小等缺点<sup>[12]</sup>.针对这些问题, $X$ -均值首先加快每一次  $K$ -均值,在  $K$ -均值的每一次迭代中,用每一个聚类中心的信息代表它所在的子集,并且利用  $kd$ -树的层次结构大幅度提高其求解速度.有了快速的  $K$ -均值算法就可以求解最优的  $K$  值, $X$ -均值在指定的上、下界范围内寻找最优的  $K$ ,其最优性的判断标准是基于所谓的贝叶斯信息标准(BIC)<sup>[13]</sup>:  $BIC(C|X) = L(X|C) - \frac{p}{2} \log n$ ,其中  $L(X|C)$  是数据集  $X$  关于模型  $C$  的对数似然, $p = k(d+1)$  是  $d$  维的,  $K$  个聚类中心的模型  $C$  中的参数个数,  $n$  是样本数. $X$ -均值主要由以下两步组成:(1) 改进聚类参数,即搜索最优的  $K$  个中心;(2) 改进聚类结构,即寻找最优的  $K$  值.

## 2 基于划分的局部切空间排列算法(PLTSA)

PLTSA 是将 VQPCA 和 LTSA 相结合而产生的.首先用  $X$ -均值找到样本集的最优划分的个数,把样本空间划分成有重叠的块,通过把每一样本点投影到它所在块的局部切空间中求出其低维局部坐标,然后利用重叠部分中点的局部坐标来构建整体坐标.构建的基本思想与 LTSA 类似,但是在数学推导上有些不同.下面我们给出其推导过程.

用  $I$  表示单位矩阵,  $e$  表示每个分量都是 1 的列向量,  $A^+$  表示矩阵  $A$  的 Moore-Penrose 广义逆,  $\|\cdot\|_2$  表示 2 范数,而  $\|\cdot\|_F$  表示 Frobenius 范数.

假设样本集  $X = [x_1, \dots, x_N]$  是从含有噪声的低维流形上均匀采样得到的,即  $x_j = f(\tau_j) + \varepsilon_j$ , 其中  $j = 1, \dots, N$ ,  $\tau_j \in R^d$ ,  $x_j \in R^D$  且  $d < D$ .进一步地,设由  $X$ -均值找到  $X$  的  $t$  个中心为  $\{m_1, \dots, m_t\}$ .接下来,把样本空间划分成  $t$  个彼此有重叠的块,对每一块分别进行 PCA 求得局部坐标,然后利用重叠部分中点的局部坐标来构建整体嵌入坐

标,构建的前提假设是整体嵌入坐标可以由重叠部分中的点局部坐标经平移、旋转、伸缩变换来求得.对样本空间进行划分,有基于  $\varepsilon$  带和基于  $l$  点两种方式.(1) 基于  $\varepsilon$  带的划分.对  $x_j, j = 1, \dots, N$ , 找出离它最近的两个中心

$m_1^{(j)}, m_2^{(j)}$ . 如果  $c_j = \frac{\langle 2x_j - (m_1^{(j)} + m_2^{(j)}), m_1^{(j)} - m_2^{(j)} \rangle}{\|m_1^{(j)} - m_2^{(j)}\|_2} \leq \varepsilon$  且  $m_i$  是离  $x_j$  最近的中心,则令  $x_j \in M_i$ , 如果  $c_j > \varepsilon$  且  $m_i$  是

离  $x_j$  最近的中心,则令  $x_j \in M'_i$ , 其中  $c_j$  是相邻两块间重叠部分的宽度占两个中心间距离的百分比,  $\varepsilon$  是参数.(2)

基于  $l$  点的划分.对  $x_j$ , 找出离它最近的两个中心  $m_1^{(j)}, m_2^{(j)}$ , 且计算出它与两个中心的垂直平分超平面的距离  $d_j = \langle 2x_j - (m_1^{(j)} + m_2^{(j)}), m_1^{(j)} - m_2^{(j)} \rangle$ . 对每个  $m_i$ , 把它以它为两个最近中心之一的点按  $d_j$  由小到大排序.对  $x_j$ , 如果  $d_j$  排在前  $l$  个中, 则令  $x_j \in M_i$ , 如果  $d_j$  不排在前  $l$  个中且  $m_i$  是离  $x_j$  最近的中心, 则令  $x_j \in M'_i$ , 其中参数  $l$  是重叠部分中点的个数.把样本集划分成有交的  $t$  个子集  $X_i = M_i \cup M'_i, i = 1, \dots, t$ , 用重叠部分  $M_i, i = 1, \dots, t$  中点的低维局部坐标来引导构建整体坐标.

对每一块作局部主成分分析,只给出基于  $l$  点划分的推导,基于  $\varepsilon$  带的推导与其类似.设  $X_i$  由  $M_i$  中的  $l$  个点  $x_{i1}, \dots, x_{il}$  和  $M'_i$  中的  $q-l$  个点  $x_{i(l+1)}, \dots, x_{iq}$  组成.现在计算  $X_i$  的最优  $d$  维仿射空间逼近,即求解最优化问题

$$\min_{x, \theta} \sum_{j=1}^q \|x_{ij} - (x + Q\theta_j)\|_2^2 = \min_{x, \theta} \|X_i - (xe^T + Q\theta)\|_F^2$$
, 其中  $Q$  是  $D \times d$  的正交矩阵,  $\theta = [\theta_1, \dots, \theta_q]$ . 容易知道  $x$  的最优解是  $x_{ij}, j = 1, \dots, q$  的均值  $\bar{x}_i$ , 令  $X_i - \bar{x}_i e^T$  的奇异值分解为  $X_i - \bar{x}_i e^T = Q' \Sigma V^T$ ,  $Q' \in R^{D \times D}, \Sigma \in R^{D \times q}, V \in R^{q \times q}$ , 则  $Q\theta$  的最优解为  $Q^d \Sigma_d V_d^T$ , 其中  $\Sigma_d = \text{diag}(\sigma_1, \dots, \sigma_d)$  是由  $X_i - \bar{x}_i e^T$  的最大的  $d$  个奇异值按由大到小的顺序构成的对角矩阵,  $Q^d$  和  $V_d$  分别是由这  $d$  个奇异值的左、右奇异向量所组成的矩阵.所以,  $Q$  的最优解  $Q_i = Q^d$ .  $\theta$  的最优解为  $Q_i^T X_i (I - \frac{1}{q} ee^T) = [\theta_1^{(i)}, \dots, \theta_q^{(i)}]$ , 其中  $\theta_j^{(i)} = Q_i^T (x_{ij} - \bar{x}_i)$ . 从而  $x_{ij} = \bar{x}_i + Q_i \theta_j^{(i)} + \xi_j^{(i)}$ , 其中  $\xi_j^{(i)} = (I - Q_i Q_i^T)(x_{ij} - \bar{x}_i)$  是重构误差,  $\theta_j^{(i)}$  是  $x_{ij}$  的局部坐标, 记  $\theta_i = [\theta_1^{(i)}, \dots, \theta_q^{(i)}]$ .

现在考虑从表示局部几何的局部坐标  $\theta_j^{(i)}, j = 1, \dots, l, i = 1, \dots, t$  出发构建整体嵌入坐标,特别地,假定整体坐标可以由局部坐标经平移、旋转、伸缩变换得到.设  $x_{ij}$  的整体嵌入坐标是  $\tau_{ij}$ , 则  $\tau_{ij} = \bar{\tau}_i + L_i \theta_j^{(i)} + \varepsilon_j^{(i)}, j = 1, \dots, l$ , 其中  $\bar{\tau}_i$  是  $\tau_{ij}, j = 1, \dots, l$  的均值,  $L_i$  是第  $i$  块所对应的仿射变换矩阵, 是待求解的,  $\varepsilon_j^{(i)}$  是局部变换的重构误差.由于平移、旋转和伸缩变换都是线性的且局部坐标的均值是坐标原点, 一般情况下, 局部坐标系的坐标原点总是可以由  $\theta_j^{(i)}, j = 1, \dots, l$  的线性组合来表示, 所以, 在不考虑重构误差的情况下, 均值  $\bar{\tau}_i$  可以由第  $i$  块的局部坐标的均值经线性变换得到, 而局部坐标均值的表示是不唯一的, 我们选择一个解  $(I - \theta_i^+ \theta_i) e$ , 将其归一化, 令

$$v_i = \frac{(I - \theta_i^+ \theta_i) e}{e^T (I - \theta_i^+ \theta_i) e}$$
,  $\theta^{(i)} = [v_i, \dots, v_i]$  ( $l$  列), 则  $\bar{\tau}_i = [\tau_{i1}, \dots, \tau_{il}] v_i$ . 记  $T_i = [\tau_{i1}, \dots, \tau_{il}]$ ,  $E_i = [\varepsilon_i^{(1)}, \dots, \varepsilon_i^{(l)}]$ , 则

$T_i = T_i \theta^{(i)} + L_i \theta_i + E_i$ , 从而局部重构误差矩阵为  $E_i = T_i (I - \theta^{(i)}) - L_i \theta_i$ . 为了在整体坐标系中尽量保持每一块的局部信息, 应该寻找  $\tau_{ij}, L_i$  使得重构误差  $\varepsilon_j^{(i)}$  最小, 即最小化  $\|E_i\|_F^2 = \|T_i (I - \theta^{(i)}) - L_i \theta_i\|_F^2$ . 对于固定的  $T_i$ , 由最小二乘可知  $L_i = T_i (I - \theta^{(i)}) \theta_i^+$  为最小化  $\|E_i\|_F$  的解. 因此,  $E_i = T_i (I - \theta^{(i)}) (I - \theta_i^+ \theta_i)$ . 设  $\bigcup_{i=1}^t M_i$  中共有  $p$  个点  $\{x'_1, \dots, x'_p\}$  且  $x'_1, \dots, x'_p$  的排列顺序与它们在  $X$  中的顺序一致. 进一步设  $T = [\tau_1, \dots, \tau_p]$ ,  $S_i$  是满足  $TS_i = T_i$  的 0-1 选择矩阵, 其中  $\tau_r$  是  $x'_r$  的整体低维嵌入坐标  $r = 1, \dots, p$ . 则我们要寻求  $T$  使得最小化总的重构误差

$$\sum_{i=1}^t \|E_i\|_F^2 = \|TSW\|_F^2$$
, 其中  $S = [S_1, \dots, S_t]$ ,  $W = \text{diag}(W_1, \dots, W_t)$ ,  $W_i = (I - \theta^{(i)}) (I - \theta_i^+ \theta_i)$ . 为了唯一地确定  $T$ , 我们加上限制条件  $TT^T = I$ . 可以证明向量  $e$  是属于矩阵  $B = SWW^T S^T$  特征值 0 的特征向量. 因此, 根据 Rayleigh-Ritz 定理可知,  $T$  的最优解是由属于矩阵  $B$  的第 2 个 ~ 第  $d+1$  个最小特征值的特征向量组成的, 即  $T = [\tau_1, \dots, \tau_p] = [u_2^T, \dots, u_{d+1}^T]^T$ , 其中  $\tau_r$  是  $d$  维列向量,  $u_j$  是属于矩阵  $B$  的第  $j$  个特征值的特征向量. 这样,  $x'_r$  的低维嵌入坐标为  $\tau_r, r = 1, \dots, p$ , 我们将在注记中对此作进一步的解释.

对任意  $x' \in M'_i, i = 1, \dots, t$ , 它的低维嵌入坐标为  $\tau' = \bar{\tau}_i + L_i \theta' = T_i v_i + L_i \theta' = TS_i v_i + T_i (I - \theta^{(i)}) \theta_i^+ Q_i^T (x' - \bar{x}_i)$ , 其

中  $\theta'$  是  $x'$  的局部坐标.对于新来的样本点,找到离它最近的中心,把它划分给这一块并求出局部坐标,最后的处理就与  $x'$  一样了.

注:在此要证明  $x'_r$  对应的低维嵌入坐标是  $T$  的最优解的第  $r$  列  $\tau_r, r=1, \dots, p$ . 设  $P$  是一个  $p$  阶列变换矩阵,  $T' = TP$  且设  $T$  的最优解是  $Y$ , 则首先说明  $T'$  的最优解是  $Y' = YP$ . 由  $S$  的定义可知,对  $T$  作列变换  $P$  对应着对  $S$  作行变换  $P^T$ , 记  $S' = P^T S$ . 设  $T = [t_1^T, \dots, t_d^T]^T, T' = [t'_1{}^T, \dots, t'_d{}^T]^T$ , 其中  $t_j, t'_j$  都是  $p$  维行向量, 则

$$\min \sum_{j=1}^d t_j^T P^T S W W^T S^T P t_j = \min \|T' S' W\|_F^2 = \min \|T P P^T S' W\|_F^2 = \min \|T S W\|_F^2 = \min \sum_{j=1}^d t_j^T S W W^T S^T t_j.$$

令  $B = S W W^T S^T, B' = P^T S W W^T S^T P$ , 则  $s$  维列向量  $v$  是属于  $B$  的特征值  $\lambda$  的特征向量当且仅当  $P^T v$  是属于  $B'$  的特征值  $\lambda$  的特征向量. 所以,由 Rayleigh-Ritz 定理可知,  $Y' = YP$ . 假设  $P$  是一个只交换  $\tau_1, \tau_2$  的列变换, 则以上讨论说明只需交换  $Y$  的前两列就可以得到  $Y'$ . 再由  $\tau_r$  与  $x'_r$  的对应关系可知,  $x'_1$  和  $x'_2$  的整体嵌入坐标只与  $Y$  的前两列有关. 同理,  $x'_1$  和  $x'_3$  的低维嵌入坐标只与  $Y$  的第 1 列和第 3 列有关, 这样就可以得出  $x'_1$  的整体嵌入坐标是  $Y$  的第 1 列. 类似地, 可以得到  $x'_r$  的整体嵌入坐标是  $Y$  的第  $r$  列,  $r=1, \dots, p$ , 从而完成了证明.

通过以上的推导, PLTSA 可以概括如下:

输入: 样本集  $X$ 、低维嵌入空间的维数  $d$ 、重叠参数  $\varepsilon$  或者  $l$ , 划分块数的上下界;

输出: 低维嵌入坐标  $T$ 、投影方向  $Q_i$ 、选择矩阵  $S$ 、权值向量  $v_i$ 、仿射变换矩阵  $L_i$ 、重叠部分中点的局部坐标  $\theta_i$ .

1. 用  $X$ -均值找到  $X$  的  $t$  个中心;
2. 把  $X$  划分成有重叠的  $t$  块  $X_i = M_i \cup M'_i$ ;
3. 对每一个  $X_i$  进行局部主成分分析, 求得局部坐标  $\theta_i^{(i)}$ ;
4. 对于重叠部分  $M_i$  中的点, 其整体坐标  $\tau$  由属于  $B$  的第 2 个~第  $d+1$  个最小特征值的特征向量给出;
5. 对于  $M'_i$  中的点, 其整体坐标为  $\tau = T S_i v_i + T_i (I - \theta_i^{(i)}) \theta_i^+ Q_i^T (x - \bar{x}_i)$ ;
6. 对于新来的样本点  $x$ , 找到离它最近的中心  $m_i$ , 其整体坐标为  $\tau = T S_i v_i + T_i (I - \theta_i^{(i)}) \theta_i^+ Q_i^T (x - \bar{x}_i)$ .

PLTSA 中用于特征值分解的矩阵的大小为  $p \leq l \times t$  阶, LTSA 中的矩阵为  $N$  阶, 而一般地,  $p \ll N$ , 所以相对于 LTSA, PLTSA 可以处理更大的样本集. 此外, PLTSA 在处理新来的样本点时需要的存储空间远小于 LTSA. 根据算法中的第 6 步, 对新来的点, PLTSA 需要计算它到所有聚类中心的距离, 然后把该点映射到低维空间中. 所以, 需要存储  $t$  个聚类中心的坐标,  $t$  个  $D \times d$  维的矩阵  $Q_i$  和  $t$  个  $d \times d$  维的矩阵  $T_i (I - \theta_i^{(i)}) \theta_i^+$ , 总共需要的空间是  $t \times (D + d) + t \times D \times d + t \times d \times d = \Theta(tDd)$ . 而 LTSA 需要的空间为  $\Theta(NDd)$ . 由于通常情况下  $N \gg t$ , 当样本个数很大时, LTSA 所需要的存储空间容易失控, 而在 PLTSA 中可以通过控制  $t$  来控制所需要的存储空间.

### 3 实验结果

本节给出了 PLTSA 在  $S$ -curve 数据集、Swiss-roll 数据集和 MNIST 数据集上的一些实验结果.

#### 3.1 模拟数据

$S$ -Curve 数据集和 Swiss-roll 数据集分别是 3 维欧氏空间中无噪声的  $S$ -curve 曲面和 Swiss-roll 曲面上均匀采样得到的<sup>[7]</sup>. 为了说明 PLTSA 在样本数较大的样本集上比 LTSA 更有空间上的优势, 在实验中我们取  $N=20000, d=2$ , 而  $X$ -均值中分块数的上、下界分别取 300 和 650, 其他参数取值为 Pelleg 等人在实验中的推荐值<sup>[10]</sup>.  $X$ -均值把  $S$ -curve 和 Swiss-roll 数据集划分的块数在 450 左右, 这与数据的采样有关, 且有一定的随机性. 图 1 和图 2 分别给出了  $S$ -curve 数据集, Swiss-roll 数据集在不同  $l$  值情况下的实验结果以及相应的排列矩阵  $B$  的阶数  $p$ , 其中  $S$ -curve 数据集的划分块数为 458, Swiss-roll 数据集的划分块数为 468. 文献[7]中给出了 LTSA 在  $N=2000, d=2$  和不同近邻值的结果. 虽然在 PLTSA 中用于重构低维整体坐标的样本的个数 ( $p$ ) 相对于 LTSA 算法中的所有样本 ( $N$ ) 要少很多, 但实验效果与 LTSA 相当. 图 3 和图 4 分别给出了 PLTSA 在  $S$ -curve 数据集和 Swiss-roll 数据集上对新来样本的处理结果. 在实验中,  $N=2000, d=2, l=5$ ,  $X$ -均值中划分块数的上、下界分别取 20 和 100, 其他参数取值为 Pelleg 等人在实验中的推荐值. 新来的样本点 (圆圈内的点) 是  $S$ -curve 曲面或 Swiss-roll

曲面上的一个随机采样.由图中可以看出,PLTSA 能有效地求出新来样本点的整体低维嵌入坐标.

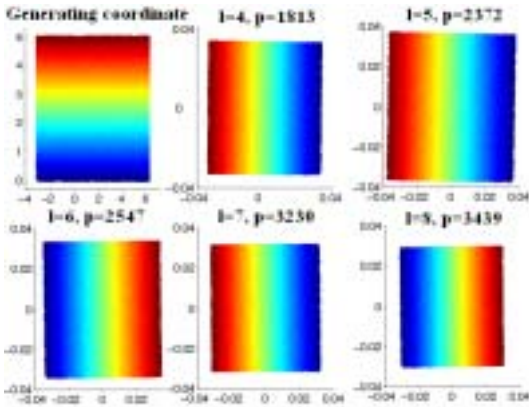


Fig.1 Generating coordinates of the S-curve and computed coordinates by PLTSA with various overlapping parameter  $l$  and corresponding aligning matrix size  $p$

图 1 S-curve 的生成坐标、PLTSA 算法在不同  $l$  值下计算出的 2 维坐标及相应排列矩阵的阶数  $p$

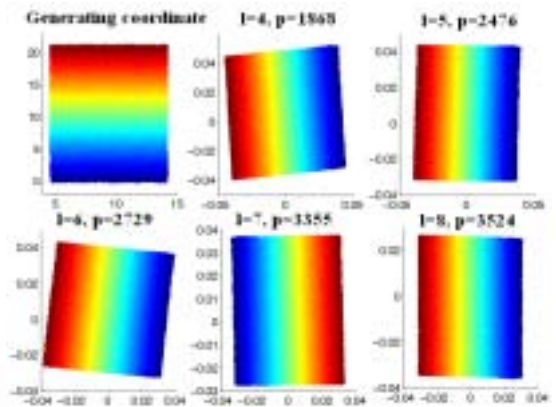


Fig.2 Generating coordinates of the Swiss-roll and computed coordinates by PLTSA with various overlapping parameter  $l$  and corresponding aligning matrix size  $p$

图 2 Swiss-roll 的生成坐标、PLTSA 算法在不同  $l$  值下计算出的 2 维坐标及相应排列矩阵的阶数  $p$

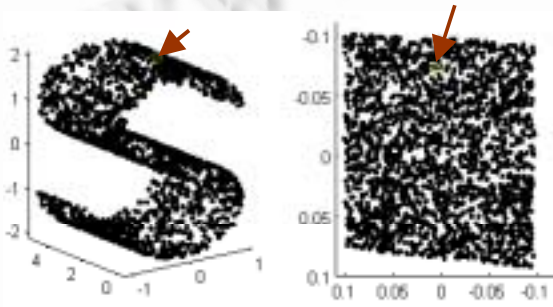


Fig.3 A newcomer point (the point in the circle) on S-curve and its computed 2D coordinate by PLTSA

图 3 S-Curve 上新来的样本点(圆圈内的点)及其在 PLTSA 算法下的低维坐标

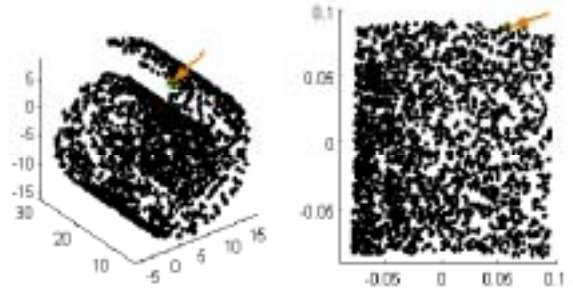


Fig.4 A newcomer point (the point in the circle) on Swiss-roll and its computed 2D coordinate by PLTSA

图 4 Swiss-Roll 上新来的样本点(圆圈内的点)及其在 PLTSA 算法下的低维坐标

### 3.2 MNIST数据集

MNIST 数据集上由 10 个手写数字的 60 000 个训练样本和 10 000 个测试样本组成,每个样本都是一帧  $28 \times 28$  的灰度图像<sup>[14]</sup>.我们从 60 000 个训练样本中随机选出数字 6,9 的图像各 1 500 幅,对它们分别用 PLTSA 进行处理.实验中, $d=2, l=5, X$ -均值中选取划分块数的上、下界为 20 和 100 ,而其他参数用 Pelleg 等人的推荐值. $X$ -均值对这两个数据集的划分的块数大概在 35 左右.图 5 和图 6 分别给出了它们的实验结果.从图中可以看出,数字 6 的 2 维投影随着横坐标的增加,其下面的圈的拐幅逐渐变小,随着纵坐标的增加,其向右倾斜的幅度不断减小;数字 9 的 2 维投影随着横坐标的增加,其向右的倾斜幅度加大,随着纵坐标的增加,其上面的圈逐渐变大.从而,PLTSA 的确能学习出 MNIST 手写数字的潜在低维变化规律.

### 4 小结及讨论

本文提出了一种以 VQPCA 和 LTSA 为基础的 PLTSA 算法.它吸取了前两种算法的优点,同时也弥补了它们的一些不足.与 VQPCA 相比,PLTSA 能够求出整体低维嵌入坐标,对进一步处理数据非常有利;与 LTSA 相

比,PLTSA 中用于特征值分解的矩阵  $B$  阶数  $p$  要比 LTSA 中的  $N$  小很多,它可以处理样本数更大的样本集.并且 PLTSA 在保存少量数据的情况下,对新来的样本点有很强的处理能力,这是大多数流形学习算法所不具备的.实验结果表明,PLTSA 算法的确是一种比较有效的流形学习算法.

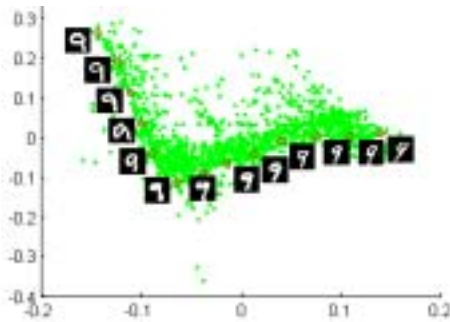


Fig.5 Computed 2D coordinates of handwritten digit 6 by PLTSA

图5 MNIST 手写数字 6 经 PLTSA 计算的 2 维坐标

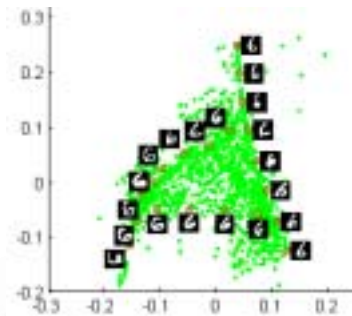


Fig.6 Computed 2D coordinates of handwritten digit 9 by PLTSA

图6 MNIST 手写数字 9 经 PLTSA 计算的 2 维坐标

在 PLTSA 中也存在一些需要改进和进一步研究的问题. PLTSA 中划分样本空间所用的  $X$ -均值的计算复杂度较高,在高维样本空间中的运行速度比较慢.在实验过程中,我们也考虑过其他空间划分算法,如 VQ 算法<sup>[9]</sup>、 $G$ -均值算法<sup>[15]</sup>,但它们的实验效果都不够理想.如何改进这些算法以适合我们的问题还有待研究,寻找新的样本空间划分算法也是努力的方向.另外,如果样本空间的维数很高,则 PLTSA 的运行速度较慢,寻找更快速的简化算法是我们面临的问题之一.

## References:

- [1] Seung HS, Lee DD. The manifold ways of perception. *Science*, 2000,290(5500):2268–2269.
- [2] Donoho DL, Grimes C. Hessian Eigenmaps: New locally linear embedding techniques for high-dimensional data. *Proc. of the National Academy of Sciences of the United States of American*, 2003,100(10):5591–5596.
- [3] <http://www.cse.msu.edu/~lawhiu/manifold/>
- [4] Tenenbaum J, Silva VD, Langford J. A global geometric framework for nonlinear dimensionality reduction. *Science*, 2000, 290(5500):2319–2323
- [5] Roweis S, Saul L. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000,290(5500):2323–2326.
- [6] Belkin M, Niyogi P. Laplacian Eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 2003,15(6): 1373–1396.
- [7] Min WL, Lu L, He XF. Locality pursuit embedding. *Pattern Recognition*, 2004,37(4):781–788.
- [8] Zhang ZY, Zha HY. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM Journal of Scientific Computing*, 2004,26(1):313–338.
- [9] Kambhatla N, Leen TK. Dimension reduction by local principal component analysis. *Neural Computation*, 1997,9(7):1493–1516.
- [10] Pelleg D, Moore A.  $X$ -means: Extending  $K$ -means with efficient estimation of the number of clusters. In: Langley P, ed. *Proc. of the 17th Int'l Conf. on Machine Learning*. San Francisco: Morgan Kaufmann Publishers, 2000. 727–734.
- [11] Chen WH. *An Introduction to Differentiable Manifold*. 2nd ed., Beijing: Higher Education Press, 2001 (in Chinese).
- [12] Duda RO, Hart PE, Stork DG. *Pattern Classification*. 2nd ed., New York: Wiley-Interscience, 2000.
- [13] Kass R, Wasserman L. Bayes factors. *Journal of the American Statistical Association*, 1995,90:773–795.
- [14] <http://yann.lecun.com/exdb/mnist/>
- [15] Hamerly G, Elkan C. Learning the  $k$  in  $k$ -means. In: Thrun S, ed. *Advances of the Neural Information Processing Systems 16*. Cambridge: MIT Press, 2004. 281–289.

## 附中文参考文献:

- [11] 陈维桓. *微分流形初步*. 第 2 版,北京:高等教育出版社,2001.