一种新的全局嵌入降维算法

刘胜蓝 [1] 闫德勤 [1]

摘 要 目前大多数流形学习算法都以距离来度量数据间的相似度, 并取得满意的效果, 但都难以处理噪音造成的子空间偏离. 针对此问题, 提出了一种基于角度优化的全局降维算法. 通过给出多样本增量的协方差阵更新方式, 从理论上证明了中心化样本长度与其偏离低维空间角度为子空间偏离的主要因素, 进而解决了噪音造成的子空间偏离问题. 同时, 与主成分分析相比, 能够更好地与其他算法融合解决小样本问题. 实验证实了该算法在手工和真实数据集上的有效性.

关键词 全局嵌入, 不规则 M 数据, 角度, 正交投影

DOI 10.3724/SP.J.1004.2011.00828

A New Global Embedding Algorithm

LIU Sheng-Lan[1] YAN De-Qin[1]

Abstract Recently, most manifold learning algorithms take advantage of distance to measure similarity of data, and obtain satisfactory results, but most of them can not handle subspace deviation caused by noise. To solve this problem, a global dimensionality reduction algorithm based on angle optimization is proposed in this paper. Theoretically it proves that the main factors of subspace deviation are the length of the center sample and the angle of deviation from the low-dimensional space by providing covariance matrix update mode of multi-sample incremental. Consequently, the algorithm solves the subspace deviation problem caused by noise. Compared with the principal component analysis, it can integrate better with other algorithms to solve small sample problems. Experiments carried out on handwork and real data sets show a clear improvement over the results of other linear algorithms.

Key words Global embedding, anomalistic M data, angle, orthogonal projection

流形学习的主要目的是从高维数据中恢复出低维流形结构, 从而达到降维的目的, 是机器学习有潜力的重要方法, 并在计算机的很多领域有着深刻的应用和研究, 成为近年来的研究重点[1−4]. 主成分分析 (Principal component analysis, PCA)[1]、多维尺度变换 (Multi-dimensional scaling, MDS)[2]、局部线性嵌入 (Locally linear embedding, LLE)[3] 等都是基于欧氏距离流形学习的著名方法. PCA 考虑整体流形是线性或近似线性的, 通过计算中心化样本的协方差阵最大化方差来寻找主成分进行降维, MDS 和 PCA 尽管目标函数的含义不同, 但在一定条件下它们可以相互转化. LLE 是一个经典的局部算法, 其主要思想是计算每个样本的 k 最近邻 (或 ε 近邻) 邻域, 寻求该局部的线性关系并由此重建权值, 进而构建出权值矩阵. 它希望这个权值在降维过程中保持不变, 从而在保持原来拓扑结构的基础上, 映射到低维流形. 而后出现了基于切空间的流形学习方法, 首先由 Donoho 等提出了 Hessian LLE (Hessian locally linear embedding, HLLE)[4], 利用 Hessian 矩阵能表达数据曲率的几何性质, 通过计算样本局部切空间坐标并将其进行 Gram-Schmidt 正交化来构建近似的 Hessian 阵, 得到最小化高维样本曲率的优化模型并嵌入到其潜在的低维流形. Min 等在几何上进一步证明局部 PCA 低维空间可以逼近以邻域均值为中心的切空间[5], 这为基于切空间的流形学习算法提供了可靠的理论依据.

Zhang 等提出了局部切空间排列算法 (Local tangent space alignment, LTSA)[6], 在局部计算出以样本均值为中心的高维样本数据的协方差阵, 进而利用 PCA 寻找局部的低维空间, 将所有局部切空间的坐标进行整合, 从而得到低维空间的整体坐标. 由于 LTSA 很难进行增量学习, 时间复杂度也很高, 杨剑等改进了 LTSA, 使之具有更强的学习能力[7]. LTSA 基于假设数据分布在近似一阶光滑可微的函数结构中, 这样的理想前提使得算法对数据的要求较高. 事实上, 由于 LTSA 建立在由数据形成的几何结构上, 所以对数据几何结构的变化非常敏感, 因此, 切空间运算并不总是适合的. 真实世界里高维数据往往是复杂的[6], 当存在某类噪音时, 切空间的偏差会增大, 影响全局的降维效果. 局部 MDS[8] 也

收稿日期 2010-07-28 录用日期 2010-11-03
Manuscript received July 28, 2010; accepted November 3, 2010
中国科学院自动化研究所复杂系统与智能科学重点实验室开放课题基金 (20070101), 辽宁省教育厅高等学校科学研究基金 (2008344)
Supported by the Open Fundation of State Key Laboratory of Complex Systems and Intelligent Sciences of Institute of Automation, Chinese Academy of Sciences (20070101) and Higher Education Department of Liaoning Province Research Foundation (2008344)

1. 辽宁师范大学计算机与信息技术学院 大连 116029
1. College of Computer and Information Technology, Liaoning Normal University, Dalian 116029

面临同样的问题. Adaptive LTSA[9] 将曲率引入到局部切空间, 克服了 LTSA 对线性化程度低的局部流形恢复效果不好的问题. 虽然可以自适应地选取邻域, 但是它没有充分利用近邻点的信息, 而是利用邻域形成的整体切空间的信息, 在估计某个近邻点测地线曲率时, 会因为其他近邻点的偏差而增大该近邻点曲率的误差, 而且曲率的存在要求假设数据分布在近似二阶光滑的函数结构中, 这对原始数据的几何结构提出了更高的要求. 由于欧氏距离在估计数据结构上的弱点, 所以在全局中, 等距映射 ISOMAP[10] 采用最短路径的方法估计测地线距离, 避免了上述的问题. 但是, 在处理大批量样本时, 最短路径的时间复杂度远大于欧氏距离.

为了更好地解决实际问题, 近几年涌现出局部保持投影 (Locality preserving projection, LPP)[11]、邻域保持嵌入 (Neighborhood preserving embedding, NPE)[12]、正交的邻域保持投影 (Orthogonal neighborhood preserving projection, ONPP)[13] 等能够提取局部信息的线性流形学习算法. 这些算法都对应于已有的非线性算法, 通过将高维数据映射到低维子空间的非线性隐式映射明确为一种线性映射而得到的. 近期, Zheng 等将这种非线性隐式映射明确为一种多项式非线性映射, 并将其应用于 LLE, 提出了邻域保持多项式嵌入 (Neighborhood preserving polynomial embedding, NPPE) 算法[14], 该算法很好地保持了高维数据分布的非线性特征. 尽管这些算法都能很好地把握局部信息[15], 但都无法进行含有噪音的学习, 而真实世界的数据不可避免地存在异常点或引入噪音, 这对流形学习算法带来了挑战.

本文以增量子空间研究为基础, 证明了中心化样本的长度与偏离子空间的角度是子空间降维产生误差的主要因素, 进而提出了一种基于角度优化的全局嵌入 (Angle optimized global embedding, AOGE) 算法. 该算法利用中心化样本偏离它在低维空间的正交投影的角度来刻画数据之间的关系, 基于角度优化实现全局降维. 该方法消除了基于距离度量对子空间带来的误差, 改用样本数据偏离低维空间的程度来度量样本之间的差异, 进而得到潜在的低维流形. AOGE 方法可以很好地应用于图像识别问题, 尤其可以应用于解决小样本问题.

为了在实验上更好地与其他全局算法进行对比, 我们将 Frey face[3] 进行个别样本删除并分类, 建立 FFC (Frey face classification, FFC) 数据库. 在识别实验中, AOGE 与 LPP, 2DPCA 等全局算法[16−17] 有很好的可比性. 由于 AOGE 方法用角度来度量样本, 故对距离数据中心较远的分散点没有其他全局算法敏感, 即具有较好的抗噪音能力. 此外, AOGE 方法是对单位化后的样本进行特征分解, 避免了应用 PCA 计算维数过高的样本时特征值过大溢出的可能. 在第 3 节, 将 AOGE 和部分全局算法分别在表情的分类、人脸识别等实验中进行对比. 实验表明, AOGE 算法在大多数情况下比 PCA, MDS 及 2DPCA 等全局算法有更好的嵌入效果, 并在真实数据中表现出更为优越的性能.

1 子空间分析

PCA 作为一种线性降维算法, 由于其算法结构简单、速度快、易应用而倍受关注. 我们把由数据所投影的低维空间称为子空间. PCA 子空间往往受到噪音及异常点的影响很大, 这将导致其学习能力大大降低. 为了使 PCA 有更好的鲁棒性, De la Torre 等提出 Robust PCA[18], 它力求通过迭代法对经典的 PCA 加权, 进而减小奇异点对算法的影响. Robust LLE[19] 也希望借助鲁棒 PCA, 在原 LLE 的基础上增加描述样本奇异程度的权值矩阵来进行鲁棒学习. 但迭代法时间复杂度大, 这将降低算法对样本数量较大的数据集的学习能力.

考虑到减小子空间的偏差成为鲁棒学习的根本, 以下通过数据在子空间上的正交投影来分析影响子空间偏差的因素.

1.1 数据的正交投影

设低维空间的一组基为: $Q \in \mathbf{R}^{D\times d}$, 流形上的样本空间 $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n] \in \mathbf{R}^{D\times n}$ 在这组基下的投影为低维流形 $Y = Q^{\mathrm{T}}X$,$Y = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n] \in \mathbf{R}^{d\times n}$. 将样本 X 进行中心化 $\hat{X} = X\left(I - \dfrac{1}{n}\boldsymbol{l}\boldsymbol{l}^{\mathrm{T}}\right)$, 其中, $\boldsymbol{l} = (1, \cdots, 1)^{\mathrm{T}}_{1\times k}$, I 为单位阵. 由样本空间 X 所形成的变换及正交投影矩阵的定义, 可以得到以下定理:

定理 1. 经过中心化的某样本 $\hat{\boldsymbol{x}}$ 的正交投影为 $QQ^{\mathrm{T}}\hat{x}$; 该点偏离切空间的夹角余弦 β 为

$$\cos\beta = \sqrt{\frac{\hat{\boldsymbol{x}}^{\mathrm{T}}QQ^{\mathrm{T}}\hat{\boldsymbol{x}}}{\|\hat{\boldsymbol{x}}\|_2^2}} \tag{1}$$

进一步地

$$\beta = \arccos\sqrt{R(\hat{\boldsymbol{x}})} \tag{2}$$

其中, $\beta \in \left[0, \dfrac{\pi}{2}\right]$, $R(\hat{\boldsymbol{x}})$ 为 QQ^{T} 的 Rayleigh 商.

证明. 在 $\mathbf{R}^D$ 空间中, 令 $U = [Q, P]$ 为 $\mathbf{R}^D$ 变换后的一组新基底, 其中 Q 为 U 的前 d 列, 为低维空间的一组基, P 为 U 的后 $n-d$ 列. 设 $\hat{X}$ 的正交投影阵为 T, 则由正交投影矩阵的定义, 有:

$T[Q,P]=[Q,O]$, 进而可知 $T=[Q,O][Q,P]^{-1}$, 其中 O 为零矩阵, 再由正交投影矩阵为幂等阵, 可得到:

$$TT^{\mathrm{T}}=[Q,O][Q,P]^{\mathrm{T}}[Q,P]^{-1}[Q,P]^{\mathrm{T}}=$$
$$[Q,O](U^{\mathrm{T}}U)^{-1}[Q,P]^{\mathrm{T}}$$

再由 $U^{\mathrm{T}}U=I$, 得 $T=QQ^{\mathrm{T}}$, 及 $Q^{\mathrm{T}}Q=I$.

由正交投影阵 $T=QQ^{\mathrm{T}}$, 欧氏空间内积的定义, 及 $Q^{\mathrm{T}}Q=I$, 可以得到 $\cos\beta=\sqrt{\dfrac{\hat{\boldsymbol{x}}^{\mathrm{T}}QQ^{\mathrm{T}}\hat{\boldsymbol{x}}}{\|\hat{\boldsymbol{x}}\|_2^2}}$. 进一步, 可以得到偏离角 $\beta=\arccos\sqrt{R(\hat{\boldsymbol{x}})}$. □

为了计算方便, 我们将其改写为 $(\cos\beta)^2=\hat{\boldsymbol{x}}_e^{\mathrm{T}}QQ^{\mathrm{T}}\hat{\boldsymbol{x}}_e$, $(\cdot)_e$ 为 "·" 的单位化向量或矩阵, 即 $\hat{\boldsymbol{x}}_e=\dfrac{\hat{\boldsymbol{x}}}{\|\hat{\boldsymbol{x}}\|}$, 满足 $\hat{\boldsymbol{x}}_e^{\mathrm{T}}\hat{\boldsymbol{x}}_e=1$.

1.2 协方差阵的更新

得到数据的正交投影后, 考虑在原数据集中增加一个或若干个噪音或奇异点来分析这些点对子空间的影响, 首先要得出增加这些点前后协方差阵的关系, 因此, 给出下面的引理.

设有 k 个样本点构成的样本矩阵为 $\boldsymbol{x}_k$, 通过中心化矩阵 $H_k=I-\dfrac{1}{k}\boldsymbol{l}\boldsymbol{l}^{\mathrm{T}}$(其中, $\boldsymbol{l}=(1,\cdots,1)_{1\times k}^{\mathrm{T}}$)把其中心化为 $\hat{X}_k=\boldsymbol{x}_kH_k$. 则相应的中心化的协方差阵为: $G_k=\hat{X}_k\hat{X}_k^{\mathrm{T}}$. 设 k 个样本的样本协方差矩阵为 C_k, 则 $C_k=\dfrac{1}{k-1}\hat{X}_k\hat{X}_k^{\mathrm{T}}=\dfrac{1}{k-1}G_k$. 当计算 C_k 后, 若增加一个新样本 $\boldsymbol{x}_{k+1}$, 按协方差定义重新计算新的样本协方差阵 C_{k+1}, 就会浪费存储空间同时增加时间复杂度. 考虑到能否由新增的 $\boldsymbol{x}_{k+1}$ 及原有的协方差阵的信息, 得到其秩一更新. 加以推广, 给出增加 r 个样本后更新样本协方差矩阵的方法.

引理 1. 设 C_k 为 k 个样本的样本协方差矩阵, 则增加 r 个样本点 $\boldsymbol{x}_{k+1},\cdots,\boldsymbol{x}_{k+r}$ 后, 更新样本协方差阵 C_{k+r} 可由下列递推式得到: $C_{k+r}=\dfrac{k-1}{k+r-1}C_k+\dfrac{1}{k+r-1}\sum_{i=1}^{r}\dfrac{k+i}{k+i-1}\Delta_{k+i}$, 其中, $\Delta_{k+i}=\boldsymbol{\delta}_{k+i}\boldsymbol{\delta}_{k+i}^{\mathrm{T}}$, $\boldsymbol{\delta}_{k+i}=\boldsymbol{x}_{k+i}-\bar{\boldsymbol{x}}_{k+i}$, $\bar{\boldsymbol{x}}_{k+i}$ 为 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{k+i}$ 的均值, $i=1,\cdots,r$(证明见附录 A1).

1.3 子空间分析

在分析采样于一个高维欧氏空间中低维流形的数据时, 存在这样一些数据: 它们存在于真实世界中, 彼此相互孤立, 且与其他数据也彼此孤立, 即不能和其他数据建立密切联系; 同时它们也不能满足或近似满足流形学习中数学模型合理的假设条件, 我们把这样的数据称为不规则 M 数据. 比如对于拍照时的光线, 相机的颤抖形成了图像数据的噪音. 需要注意的是, 不规则 M 数据并不能完全归结为噪音; 同样对于拍照, 给出一个运动员获奖的表情采样集合: {表情采样 |{高兴}, {悲伤}, {恐惧}, {惊讶}, {惊讶地喜极而泣}}, {惊讶地喜极而泣}只存在于一个很短的时间, 不能称之为噪音. 下面给出不规则 M 数据的数学定义.

定义 1. 不规则 M 数据: 设 M 为一个近似光滑的流形, 若 M 中, $\exists a_1,\cdots,a_r\in M$, 对 $\forall i\in\{0,\cdots,r\}$, $r\in\mathbf{N}^+$, $\exists\nu_i>0$, 使得 $\mathrm{U}(a_i,\nu_i)\cap M=\{a_i\}$, 则 a_i 为不规则 M 数据.

比如在降维过程中某些数据 "远离"[1]低维空间, 记为 $data_d$; 某些数据 "偏离"[2]低维空间, 记为 $data_\theta$. 这些数据都是不规则 M 数据.

考察每个不规则数据对降维算法的影响, 可动态地把每个不规则数据看作新增样本或噪音. 为了减少不规则数据给低维子空间带来的误差, 用主成分来分析影响子空间偏离度的因素, 进而在下一节中考虑如何给出一种新算法, 消除或削弱这些因素带来的误差. 设高维数据空间的维数为 D, 降维后的低维空间的维数为 d. 给出如下定理:

定理 2. 设 C_k 特征分解得到前 d 个最大特征值 $\mu_d\le\mu_{d-1}\le\cdots\le\mu_1$(假定 μ_i 各异, $i=1,\cdots,d$), 对应的特征向量为低维子空间的一组基 $Q=[\boldsymbol{q}_1,\cdots\boldsymbol{q}_d]$, C_{k+1} 特征分解得到前 d 个最大特征值 $\lambda_d\le\lambda_{d-1}\le\cdots\le\lambda_1$, 对应的特征向量为低维子空间的一组基 $\hat{Q}=[\hat{q}_1,\cdots\hat{q}_d]$, 则:

1) $\forall\boldsymbol{x}\in\mathbf{R}^D$ 有 $d(x)f(\lambda)=1$, λ 与 $d(\boldsymbol{x})$ 同增减. 对于单位化的样本 $\forall(\boldsymbol{x}-\bar{\boldsymbol{x}}_k)_e\in\mathbf{R}^D$, 则 $f(\lambda)=1$. 其中, $f(\lambda)=\sum_{j=1}^{d}\dfrac{\cos^2\alpha_j}{\lambda-\mu_j}$, $d(\boldsymbol{x})=\|\boldsymbol{x}-\bar{\boldsymbol{x}}_k\|_2$, $\alpha_{k+1,j}\in\left[0,\dfrac{\pi}{2}\right]$ 为 $(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k)$ 偏离切空间第 j 个主成分 $\boldsymbol{q}_j$ 的角度, $j=1,\cdots,d$, $k=1,2\cdots,n-1$.

2) 前 d 个主成分夹角之和为: $\sum_{j=1}^{d}\cos\hat{\theta}_j=\sum_{j=1}^{d}\hat{\boldsymbol{q}}_j^{\mathrm{T}}\boldsymbol{q}_j=\boldsymbol{l}^{\mathrm{T}}\left(\hat{\boldsymbol{\xi}}_{k+1}\right)_e$. 其中, $\hat{\theta}_j=\langle\hat{\boldsymbol{q}}_j,\boldsymbol{q}_j\rangle$, $\hat{\boldsymbol{\xi}}_{k+1}=(\dfrac{\cos\alpha_{k+1,1}}{\lambda_1-\mu_{k+1,1}},\dfrac{\cos\alpha_{k+1,2}}{\lambda_2-\mu_{k+1,2}},\cdots,\dfrac{\cos\alpha_{k+1,d}}{\lambda_d-\mu_{k+1,d}})^{\mathrm{T}}$, $j=1,\cdots,d$, $k=1,2,\cdots,n-1$ (证明见附录 A2).

2 AOGE 算法

两个子空间 space(Q) 和 space$(\hat{Q})$ 的距离为 $\mathrm{dist}(\hat{Q},Q)=\sqrt{1-\sigma_d^2}$[20], σ_d 为 $\hat{Q}^{\mathrm{T}}Q$ SVD 分解的最小奇异值; 它反映了子空间基底的偏差程度,

[1] "远离" 是指离中心较远的数据, 即满足 $\|\boldsymbol{\delta}_{k+i}\|_2>\gamma$, 其中 γ 根据 $\|\boldsymbol{\delta}_j\|_2$ $(j=1,2,\cdots,k)$ 设定.

[2] "偏离" 是指数据偏离其正交投影的角度大, 即满足 $\beta>\zeta$, 其中 ζ 根据 β_j $(j=1,2,\cdots,k)$ 设定.

故减少不规则数据带来误差的本质在于两个子空间 space(Q) 和 space($\hat{Q}$) 的距离最小化问题. 希望最大化 $\cos\theta_d$ 以减小子空间的误差, 它等价于 $\max\sum_{j=1}^{d}\cos\theta_j$[3], 由于 $\cos\theta$ 在实际计算中很难找到, 故利用 $\cos\hat{\theta}_j$ 近似代替, 原最优化问题变为 $\max E_{AOGE}$, 其中, $E_{AOGE}=\sum_{j=1}^{d}\cos\hat{\theta}_j$.

假设中心化的样本已经单位化, 由定理 2 的结论 1), $\hat{\boldsymbol{q}}(\lambda)$ (当 $\lambda\in(\mu_j,\mu_{j-1})$, $\hat{\boldsymbol{q}}(\lambda)\triangleq\hat{\boldsymbol{q}}_j$, $j=1,\cdots,d$) 由 $\cos^2\alpha(\boldsymbol{x})_i$ 决定, 在 $f'(\lambda)=-\sum_{j=1}^{d}\dfrac{\cos^2\alpha_j}{(\lambda-\mu_j)^2}$ 中, $\max\cos^2\alpha$ 即最大化 λ 的减小速度, 这将导致定理 2 的结论 2) 最大化, 即 $\max E_{AOGE}$ 等价于 $\max\sum_{k=1}^{n}\sum_{j=1}^{d}\cos^2\alpha_{kj}$.

当增加 $\boldsymbol{x}_{k+1}$ 时, 令 $\boldsymbol{p}_{k+1,d}=(\cos\alpha_{k+1,1},\cos\alpha_{k+1,2},\cdots,\cos\alpha_{k+1,d})^{\mathrm{T}}$, 由定理 1, $\boldsymbol{p}_d=Q^{\mathrm{T}}(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k)/\|\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k\|_2$, 再由定理 2, 有 $\cos^2\beta_{k+1}=\left\|\boldsymbol{p}_{k+1,d}\right\|_2^2=\sum_{j=1}^{d}\cos^2\alpha_{k+1,j}$. 因此, 考虑 $\max_k\cos^2\beta_k$, 并适当地选取 d, 同时使中心化的样本单位化, 具体算法过程描述如下:

将 $\hat{X}$ 单位化 $\hat{X}_e=\hat{X}D_0$, 其中, $D_0=\mathrm{diag}\left\{\dfrac{1}{\|\hat{\boldsymbol{x}}_1\|_2},\cdots,\dfrac{1}{\|\hat{\boldsymbol{x}}_1\|_n}\right\}$, 可得: $\hat{X}_e\hat{X}_e^{\mathrm{T}}=\hat{X}\tilde{D}\hat{X}^{\mathrm{T}}$, 其中 $\tilde{D}=D_0^2$. 把每一个样本点 $\boldsymbol{x}_i$ 看作新增点, 经中心化后, 构造优化函数为

$$E_{\mathrm{AOGE}}=\sum_{i=1}^{n}\cos^2\beta_i$$

$$E_{\mathrm{AOGE}}=\sum_{i=1}^{n}\hat{\boldsymbol{x}}_{e_i}^{\mathrm{T}}QQ^{\mathrm{T}}\hat{\boldsymbol{x}}_{e_i}=\mathrm{Tr}\left(Q^{\mathrm{T}}\hat{X}_{e_i}\hat{X}_{e_i}^{\mathrm{T}}Q\right)$$

规定低维空间的一组基 $Q\in\mathbf{R}^{D\times d}$ 为标准正交基, 令 $M=\hat{X}_e\hat{X}_e^{\mathrm{T}}=\hat{X}\tilde{D}\hat{X}^{\mathrm{T}}$, 得到有约束的最优化模型为

$$\begin{cases}\max\mathrm{Tr}\,(Q^{\mathrm{T}}MQ)\\ \text{s.t. } Q^{\mathrm{T}}Q=I\end{cases}\tag{3}$$

Q 为 M 前 d 个最大特征值所对应的特征向量. 最终得到 $Y=Q^{\mathrm{T}}X$.

由以上分析, 得到 AOGE 算法的一般步骤如下:

算法 1. AOGE 算法

步骤 1. 初始化样本并将样本中心化 $\hat{X}=X\left(I-\dfrac{1}{n}\boldsymbol{l}\boldsymbol{l}^{\mathrm{T}}\right)$;

步骤 2. 单位化协方差阵: $\hat{\boldsymbol{x}}_i\leftarrow\dfrac{\hat{\boldsymbol{x}}_i}{norm(\hat{\boldsymbol{x}}_i)}$, $i=1,2,\cdots,n$;

步骤 3. 计算 M, 并将其进行特征分解, 得到特征值 $\lambda_1\geq\cdots\geq\lambda_D$ 及对应的特征向量阵 U;

步骤 4. 取 U 的前 d 个最大特征值所对应的特征向量 $Q=[\boldsymbol{u}_1,\cdots,\boldsymbol{u}_d]$, 得到 $Y=Q^{\mathrm{T}}X$.

由定理 2 的结论 1), 式 (3) 对 $\forall\boldsymbol{x}_e\in\mathbf{R}^D$, λ 只受 $\cos\alpha(x)_i$, $i=1,\cdots,d$ 的影响, 且影响最小化. 若每个样本点的模都相同, 即 $\|\hat{\boldsymbol{x}}_1\|_2^2=\|\hat{\boldsymbol{x}}_2\|_2^2=\cdots=\|\hat{\boldsymbol{x}}_n\|_2^2$, 则有 $\tilde{D}=1/\|\hat{\boldsymbol{x}}_1\|_2^2\mathrm{diag}\{1,\cdots,1\}=1/\|\hat{\boldsymbol{x}}_1\|_2^2\times\mathrm{I}$. 因此, $\hat{X}_e\hat{X}_e^{\mathrm{T}}=\hat{X}\tilde{D}\hat{X}^{\mathrm{T}}=k^2\hat{X}\tilde{D}\hat{X}^{\mathrm{T}}$, 其中, $k=1/\|\hat{\boldsymbol{x}}_1\|_2$ 为常数; 此时, 模型 (3) 就是著名的 PCA 模型. 可见基于定理 2 得到的优化模型是对 PCA 模型的推广. 得到的模型 (3) 更充分地提取了样本点的信息. 图 1 给出了二维椭圆 (长轴 5, 短轴 2) 随机采样得到的 AOGE 及 PCA 主成分.

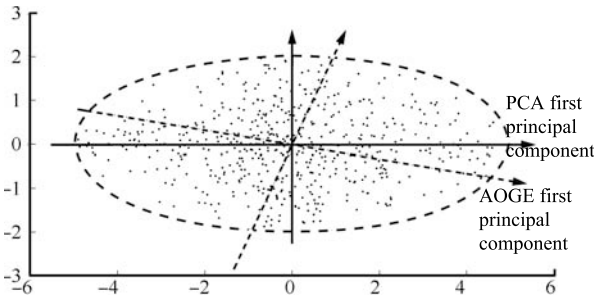

图 1　椭圆采样中 PCA 及 AOGE 主成分对比

Fig. 1　The principal component comparision of PCA and AOGE in the elliptical sampling

图 1 更加明确了 AOGE 和 PCA 在寻找主成分时的不同: PCA 考虑了 $data_\theta$ 和 $data_d$ 的中心化投影方差, AOGE 考虑了中心化样本的投影角度.

3　实验结果与分析

在本节中, 第 3.1 节为真实世界图片采样 Frey 人脸表情的无监督分类, 第 3.2 节为 AR 人脸表情数据库的识别实验, 第 3.3 节为手工流形数据实验.

3.1　Frey face 表情分类

为检验算法对实际数据的降维能力, 我们利用 Frey face 表情图片库, 每个样本为 20 像素 × 28 像素的灰度图片, 该表情库是由视频设备连续拍摄得到的. 进一步地, 将 Frey face 人脸的表情分成六

[3] σ_l 为半正交阵 $\hat{Q}^{\mathrm{T}}Q$ 的 SVD 分解的第 l 个奇异值, $l=1,2,\cdots,d$, $\sigma_d\leq\cdots\leq\sigma_1$, 其中, $\cos\theta_l=\sigma_l$, 由于 $\cos\theta_d\leq\cdots\leq\cos\theta_1$, 故最大化 $\cos\theta_d$ 等价于 $\max\sum_{j=1}^{d}\cos\theta_j$

类, 并除去 10 幅不确定表情, 中性 (自然) 571 幅、吐舌 79 幅、高兴 560 幅、微笑 59 幅、不高兴 (悲伤或生气) 623 幅、撅嘴 63 幅, 共 1 955 幅图像, 简称 FFC (Frey face classification). 选择前五类表情进行分类实验, 与 Frey face 库相比, FFC 可以更好地分析算法的性能.

图 2 表明, 高兴、中性、不高兴这三大类利用 AOGE 算法可以比较清晰地进行分类, 并在三维欧氏空间呈三角形分布. 微笑低维散点基本分布在高兴散点区域的边缘, 并向不高兴和中性表情靠近; 而吐舌散点则分布在不高兴和中性表情中间的区域. 将 FFC 中不同表情随机选取了 400 个样本作为训练集, 剩下的样本作为测试集, 共做了 10 次实验, 表 1 给出了这几种算法在 Frey 人脸数据集中不同的平均分类率及标准差.

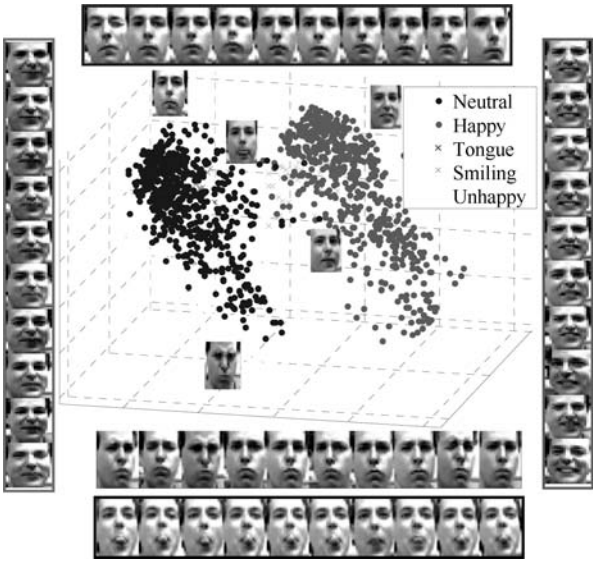

图 2 FFC 利用 AOGE 算法降至三维散点图

Fig. 2 Scatter diagram of dimensionality reduction to three-dimension by FFC using AOGE algorithm

表 1 Frey 人脸分类率对比

Table 1 Comparison of Frey face classification rates

约简维数	平均分类率 ± 标准差 (%)		
	AOGE	PCA	MDS
2	67.14 ± 2.79	54.84 ± 1.27	54.61 ± 1.74
3	74.38 ± 1.34	71.29 ± 2.88	70.80 ± 3.87
7	86.46 ± 0.88	86.29 ± 1.16	86.21 ± 0.99

从表 1 的数据可以看出, 在维数约简到比较低时, AOGE 可以表现出比 PCA 及 MDS 更优越的性能.

3.2 AR 人脸识别[21]

AR 人脸库共 126 人, 每人由 26 幅图像构成, 并有比较明显的光照和表情变化. 我们从中选取 120 人去掉遮挡和墨镜图像, 根据眼睛和嘴的定位处理并裁剪为 50 像素 × 40 像素的灰度图像, 组成新图像库: 每人 14 幅, 共 1 680 幅. 从中随机选取 20 人作为实验样本, 其中每人选取前 5 幅作为训练集, 后 9 幅样本作为测试集. 为了考察 AOGE 算法对不规则数据的效果, 首先对图像加以均值为 0、方差为 0.1 的高斯白噪声, 图 3 及表 2 给出了 AOGE 算法同其他线性算法约简到不同维数的识别率比较.

再从剩余的 100 人中随机选取 20 人, 替换原训练集每类的一个训练样本, 作为噪音样本. 在此情况下, AOGE 算法同其他线性算法约简到不同维数的识别率比较如表 3 所示.

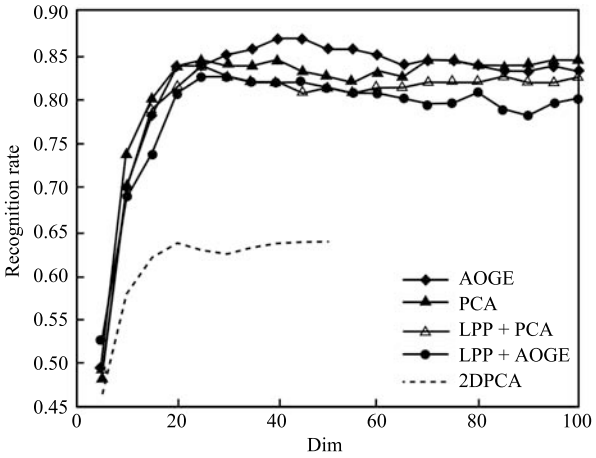

图 3 AR 人脸识别率随维数变化曲线

Fig. 3 The curves of AR face recogniton rate with the dimension change

表 2 AR 人脸加高斯白噪声的识别率对比 (%)

Table 2 Recognition rate comparison of AR face with white Gaussian noise (%)

方法/项目	Mean-RecR	Max-RecR	Max-Dim
AOGE	81.75	86.88	40
PCA	81.19	84.38	25
LPP + PCA	79.56	83.75	25
LPP + AOGE	78.22	82.50	25
2DPCA	60.87	63.75	20

从表 2 及图 3 中可以看出, 含噪音数据降低了提取局部信息的 LPP 等算法的学习效果, 而 AOGE 却表现出优越的不规则数据学习能力. 此外, 由表

3 给出的信息, 对于含有噪音样本的训练集, 由于 LPP + AOGE 经过 AOGE 处理信息保留的正确程度比 PCA 高, 所以该算法在本实验的各种线性算法中识别率最高; 而对于 2DPCA, 由于其对图像信息的敏感性及本文对 PCA 的分析, 一旦引入不规则数据, 低维嵌入空间将受到较大影响.

表 3　AR 人脸加噪音样本识别率对比 (%)

Table 3　Recognition rate comparison of AR face with noise samples (%)

方法/项目	Mean-RecR	Max-RecR	Max-Dim
AOGE	60.06	63.33	50
PCA	59.19	62.22	40
LPP + PCA	61.53	65.00	40
LPP + AOGE	63.25	66.11	45
2DPCA	35.44	42.22	5

3.3　手工流形数据

我们采用 Swiss-Roll 和 Twinpeaks 三维流形数据, 采样 1 000 个样本点, 分别利用 AOGE, PCA, LPP 降至二维, 降维效果如图 4 和图 5 所示.

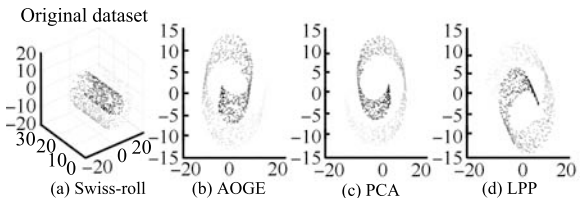

图 4　Swiss-Roll 降至二维

Fig. 4　Scatter diagram of dimensionality reduction to two-dimension of Swiss-Roll

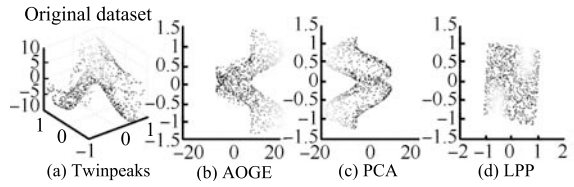

图 5　Twinpeaks 降至二维

Fig. 5　Scatter diagram of dimensionality reduction to two-dimension of Twinpeaks

从图 4 和图 5 中可以看出, AOGE 与 PCA 基本一致且效果比较理想, 而 LPP 因为提取局部信息时的短路现象, 降维效果并不理想, 在图 5 中 LPP 可以很好地提取局部信息, 表现出较好的流形学习能力, AOGE 在流形学习过程中也体现了原始数据双峰的特征, 而 PCA 的学习能力则差一些, 这是因为在双峰点处的数据的分布不能线性近似, 导致 PCA 在峰点处有偏差.

3.4　实验结果分析

第 3.1 节中 FFC 表情实验说明 AOGE 作为一种线性流形学习算法的有效性; 第 3.2 节的 AR 人脸识别实验主要考察各种线性流形学习算法对不规则数据的学习能力, 从中可以看到 AOGE 不仅本身表现出良好的学习能力, 而且在处理识别问题的小样本问题时, 表现出比 PCA 更强的抗噪音能力; 以上说明 AOGE 算法有较好的抗噪音识别及分类能力. 第 3.3 节的手工流形数据更加肯定了 AOGE 作为一种线性流形学习算法的有效性.

4　结论

本文给出了影响子空间误差的主要因素为中心化样本的长度及其偏离主成分的角度的结论, 同时给出多个样本点增量的协方差矩阵的更新方式, 分析了不规则 M 数据对低维子空间的影响, 并提出一种基于角度的全局降维算法 AOGE. 由于该算法克服了以往许多距离度量算法受不规则 M 数据影响较大的缺点, 因此对许多实际问题中的不规则 M 数据具有很强的学习能力, 这是其他同类全局算法所不具备的. 实验证明 AOGE 算法与 PCA 等算法比较有更好的抗噪音能力和降维效果. 由于高维数据往往是非线性的, 所以 AOGE 算法有待进一步研究, 将其局部化, 使其对复杂的数据有更强的学习能力.

附录 A

A1. 引理 1 证明

证明. 由 $\bar{\boldsymbol{x}}_{k+1}=\frac{1}{k+1}(k\bar{\boldsymbol{x}}_k+\boldsymbol{x}_{k+1})$ 得到 $\bar{\boldsymbol{x}}_{k+1}=\bar{\boldsymbol{x}}_k+\frac{1}{k+1}(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k)$,

$$\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_{k+1}=\frac{k}{k+1}(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k) \tag{A1}$$

对于样本矩阵增加一个样本后的中心化形式为

$$\hat{X}_{k+1}=(\boldsymbol{x}_1-\bar{\boldsymbol{x}}_{k+1},\cdots,\boldsymbol{x}_k-\bar{\boldsymbol{x}}_{k+1},\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_{k+1})$$

由式 (A1) 及 $H_k\mathbf{1}=0$, 增加一个样本后中心化的协方差矩阵为: $G_{k+1}=\hat{X}_k\hat{X}_k^{\mathrm{T}}+\frac{k}{(k+1)^2}(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k)(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k)^{\mathrm{T}}+\left(\frac{k}{k+1}\right)^2(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k)(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k)^{\mathrm{T}}=G_k+\frac{k}{k+1}(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k)(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k)^{\mathrm{T}}$

由上式得 $G_{k+1}=G_k+(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_{k+1})(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_{k+1})^{\mathrm{T}}$. 其中, $\Delta_{k+r}=\boldsymbol{\delta}_{k+r}\boldsymbol{\delta}_{k+r}^{\mathrm{T}}$, $\boldsymbol{\delta}_{k+r}=\boldsymbol{x}_{k+r}-\bar{\boldsymbol{x}}_{k+r}$, $r=1,2,\cdots$.

由递推关系: $G_{k+1}=G_k+\frac{k+1}{k}\Delta_{k+1}, G_{k+2}=G_{k+1}+\frac{k+2}{k+1}\Delta_{k+2},\cdots,G_{k+r}=G_{k+r-1}+\frac{k+r}{k+r-1}\Delta_{k+r}$.

将这 r 个等式相加可得: $G_{k+r}=G_k+\sum_{i=1}^{r}\frac{k+i}{k+i-1}\Delta_{k+i}$.

由于样本协方差阵 $C_k=\frac{1}{(k-1)}\hat{X}_k\hat{X}_k^{\mathrm{T}}=\frac{1}{(k-1)}G_k$, 从而得到增加 $\boldsymbol{x}_{k+1},\cdots,\boldsymbol{x}_{k+r}$ 的 r 个样本更新的协方差阵:

$$C_{k+r}=\frac{k-1}{k+r-1}C_k+\frac{1}{k+r-1}\sum_{i=1}^{r}\frac{k+i}{k+i-1}\Delta_{k+i} \qquad \square$$

Δ_{k+i} 的更新可由一个简单的求和算法实现. 令 $s_i=\sum_{j=1}^{i}(\boldsymbol{x}_{k+j}-\bar{\boldsymbol{x}}_k)$, 可求得 $\boldsymbol{\delta}_{k+i}=(\boldsymbol{x}_{k+i}-\bar{\boldsymbol{x}}_k)-\frac{1}{k+i}s_i$. 则: $\Delta_{k+i}=\boldsymbol{\delta}_{k+i}\boldsymbol{\delta}_{k+i}^{\mathrm{T}}$.

A2. 定理 2 证明

证明. 1) 在引理 1.1 的基础上, 首先考虑如何求得秩一更新 (即增加 $\boldsymbol{x}_{k+1}$ 后) C_{k+1} 的特征值和特征向量. $\mu(\cdot)$, $\lambda(\cdot)$ 为矩阵 "$\cdot$" 对应的特征值 (简写为 μ 和 λ, 分别表示增加样本前后, 样本协方差矩阵对应的特征值), $\mathrm{rank}(\cdot)$ 表示矩阵 "$\cdot$" 的秩. 根据文献 [20] 中的推论 8.1.6 我们可以得到: $\left|\mu_1(\frac{k-1}{k}C_k+\frac{k+1}{k^2}\Delta_{k+1})-\mu_1(\frac{k-1}{k}C_k)\right|\leq\frac{k+1}{k^2}\max_i\{\mu_i(\Delta_{k+1})\}$. 由引理 1, 该不等式也可写为: $\left|\lambda_1(C_{k+1})-\mu_1(\frac{k-1}{k}C_k)\right|\leq\frac{k+1}{k^2}\max_i\{\mu_i(\Delta_{k+1})\}$. 由式 (A1) 知 $\Delta_{k+1}=(\frac{k}{k+1})^2(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k)(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k)^{\mathrm{T}}$.

由于 $\mathrm{rank}(\Delta_{k+1})=1$, Δ_{k+1} 只有一个非零特征值, 因此, $\max_i\{\mu_i(\Delta_{k+1})\}=\mathrm{Tr}(\Delta_{k+1})=\|\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k\|_2^2$ ($\mathrm{Tr}(\cdot)$ 表示矩阵 "$\cdot$" 的迹). 我们得到

$$\frac{k-1}{k}\mu_1(C_k)\leq\lambda_1(C_{k+1})\leq \frac{k-1}{k}\mu_1(C_k)+\frac{k+1}{k^2}\|\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k\|_2^2 \tag{A2}$$

令 $\mu_0=\mu_1(C_k)+\frac{k+1}{k^2}\|\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k\|_2^2$, 由于 $\frac{k+1}{k^2}>0$, 可直接利用文献 [20] 中定理 8.18 的结论, 得到: $\lambda_i(C_{k+1})\in(\frac{k-1}{k}\mu_i(C_k),\frac{k-1}{k}\mu_{i-1}(C_k))$, 其中, $\mu_d\leq\mu_{d-1}\cdots\leq\mu_0$. 因此就确定了每个 λ 的区间. 在确定了每个 λ 的区间后, 就可以引用文献 [22] 给出更新特征值 λ 的方法: 引入 Weinstein-Aronszajn 阵 $W(\lambda)$. 设 $\omega(\lambda)=\det[W(\lambda)]$ 为 $W(\lambda)$ 的行列式. 由于 $\mathrm{rank}(\delta_{k+1})=1$, 得到:

$$\omega(\lambda)=1+\sum_{i=1}^{d}\frac{|\boldsymbol{q}_i^{\mathrm{T}}(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k)|^2}{\mu_i-\lambda} \tag{A3}$$

令 $\omega(\lambda)=0$, 可得到满足精度的特征值 $\lambda_j(j=1,\cdots,d)$. 则对 $\forall\boldsymbol{x}\in\mathbf{R}^D$, 有: $d(\boldsymbol{x})f(\lambda)=1$, 其中, $f(\lambda)=\sum_{j=1}^{d}\frac{\cos^2\alpha_j}{\lambda-\mu_j}$ 为 λ 的严格减函数, $\lambda\in\cup_{i=1}^{d}(\mu_i,\mu_{i-1})$, $d(\boldsymbol{x})=\|\boldsymbol{x}-\bar{\boldsymbol{x}}_k\|_2$. 为了说明 $\|\boldsymbol{x}-\bar{\boldsymbol{x}}_k\|_2$ 及 $\cos\alpha(\boldsymbol{x})_i$ ($\alpha_i\triangleq\alpha(\boldsymbol{x})_i$ 表示 α_i 与 $\boldsymbol{x}$ 相关) 对 λ 的影响, 考虑在高维空间中数据分布是稀疏的, 则 $d(\boldsymbol{x})\gg\cos^2\alpha(\boldsymbol{x})_i$ 与 $\lambda-\mu_i\gg\cos^2\alpha(\boldsymbol{x})_i$, 所以此时可以忽略 $\cos^2\alpha(\boldsymbol{x})_i$ 对 λ 的影响, 为了保证 $d(\boldsymbol{x})f(\lambda)=1$ 成立, 则 λ 与 $d(\boldsymbol{x})$ 同增减.

若 $\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k$ 单位化, 则由上述过程知, 对 $\forall\boldsymbol{x}\in\mathbf{R}^D$, 有: $f(\lambda)=1$.

2) 按照文献 [22] 中第 9.6.1 节和第 9.6.2 节的方法, 我们就可以计算出 C_{k+1} 特征分解对应的特征值 λ_j, 特征矩阵 $\hat{Q}$ 及特征向量 $\hat{\boldsymbol{q}}_j=\frac{\boldsymbol{y}_j}{\|\boldsymbol{y}_j\|_2}$, 其中 $\boldsymbol{y}_j=-\sum_{i=1}^{d}\frac{\boldsymbol{q}_i^{\mathrm{T}}(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k)}{\mu_i-\lambda_j}\boldsymbol{q}_i$.

由内积定义知, $\boldsymbol{q}_i^{\mathrm{T}}(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k)=\cos\alpha_i\|\boldsymbol{q}_i\|_2\|\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k\|_2=\cos\alpha_i\|\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k\|_2$, 代入 $\boldsymbol{y}_j$ 的表示式得到:

$$\boldsymbol{y}_j=\sum_{i=1}^{d}\frac{\cos\alpha_i\|\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k\|}{\lambda_j-\mu_i}\boldsymbol{q}_i=\|\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_k\|Q\boldsymbol{\xi}_{k+1,j} \tag{A4}$$

其中, $\boldsymbol{\xi}_{k+1,j}=(\frac{\cos\alpha_{k+1,1}}{\lambda_j-\mu_{k+1,1}},\frac{\cos\alpha_{k+1,2}}{\lambda_j-\mu_{k+1,2}},\cdots,\frac{\cos\alpha_{k+1,d}}{\lambda_j-\mu_{k+1,d}})^{\mathrm{T}}$, 而 $\|Q\boldsymbol{\xi}_{k+1,j}\|_2^2=\boldsymbol{\xi}_{k+1,j}^{\mathrm{T}}Q^{\mathrm{T}}Q\boldsymbol{\xi}_{k+1,j}=\|\boldsymbol{\xi}_{k+1,j}\|_2^2$, 可计算 $\hat{\boldsymbol{q}}_j=\frac{Q\boldsymbol{\xi}_{k+1,j}}{\|\boldsymbol{\xi}_{k+1,j}\|_2}$. 再由式 (A4), 有 $\sum_{j=1}^{d}\cos\hat{\theta}_j=\sum_{j=1}^{d}\hat{\boldsymbol{q}}_j\boldsymbol{q}_j=\boldsymbol{l}^{\mathrm{T}}\left(\hat{\boldsymbol{\xi}}_{k+1}\right)_e$, 其中, $\hat{\boldsymbol{\xi}}_{k+1}=(\frac{\cos\alpha_{k+1,1}}{\lambda_1-\mu_{k+1,1}},\frac{\cos\alpha_{k+1,2}}{\lambda_2-\mu_{k+1,2}},\cdots,\frac{\cos\alpha_{k+1,d}}{\lambda_d-\mu_{k+1,d}})^{\mathrm{T}}$. $\square$

References

1 Jolliffe I T. *Principal Component Analysis (Second Edition)*. New York: Springer-Verlag, 2002

2 Cox T F, Cox M A A. *Multidimensional Scaling*. London: Chapman and Hall, 1994

3 Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, **290**(5500): 2323−2326

4 Donoho D L, Grimes C. Hessian eigenmaps: locally linear embedding, techniques for high-dimensional data. *Proceedings of the National Academy of Sciences*, 2003, **100**(10): 5591−5596

5 Min W L, Lu K, He X F. Locality pursuit embedding. *Pattern Recognition*, 2004, **37**(4): 781−788

6 Zhang Z Y, Zha H Y. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM Journal of Scientific Computing*, 2004, **26**(1): 313−338

7 Yang Jian, Li Fu-Xin, Wang Jue. A better scaled local tangent space alignment algorithm. *Journal of Software*, 2005, **16**(9): 1584−1590
(杨剑, 李伏欣, 王珏. 一种改进的局部切空间排列算法. 软件学报, 2005, **16**(9): 1584−1590)

8 Yang L. Alignment of overlapping locally scaled patches for multidimensional scaling and dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(3): 438−450

9 Wang J, Zhang Z Y, Zha H Y. Adaptive manifold learning. In: Proceedings of the Neural Information Processing Systems. Vancouver, Canada: The MIT Press, 2004. 1473−1480

10 Tenenbaum J B, De S V, Langford J C. A global geometric framework for nonlinear dimension reduction. *Science*, 2000, **290**(5500): 2319−2323

11 He X F, Niyogi P. Locality preserving projections. In: Proceedings of the Neural Information Processing Systems. Vancouver, Canada: The MIT Press, 2003. 153−160

12 He X F, Cai D, Yan S C, Zhang H J. Neighborhood preserving embedding. In: Proceedings of the 10th IEEE International Conference on Computer Vision. Beijing, China: IEEE, 2005. 1208−1213

13 Kokiopoulou E, Saad Y. Orthogonal neighborhood preserving projections. In: Proceedings of the 5th IEEE International Conference on Data Mining. Washington D. C., USA: IEEE, 2005. 1−7

14 Zheng S W, Qiao H, Zhang B, Zhang P. The application of intrinsic variable preserving manifold learning method to tracking multiple people with occlusion reasoning. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. St. Louis, USA: IEEE, 2009. 2993−2998

15 Li Le, Zhang Yu-Jin. Linear projection-based non-negative matrix factorization. *Acta Autmatica Sinica*, 2010, **36**(1): 23−39
(李乐, 章毓晋. 基于线性投影结构的非负矩阵分解. 自动化学报, 2010, **36**(1): 23−39)

16 Yang J, Zhang D, Frangi A F, Yang J Y. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(1): 131−137

17 Wen Ying, Shi Peng-Fei. An approach to face recognition based on common vector and 2DPCA. *Acta Autmatica Sinica*, 2009, **35**(2): 202−205
(文颖, 施鹏飞. 一种基于共同向量结合 2DPCA 的人脸识别方法. 自动化学报, 2009, **35**(2): 202−205)

18 De la Torre F, Black M J. Robust principal component analysis for computer vision. In: Proceedings of the 8th IEEE International Conference on Computer Vision. Vancouver, Canada: IEEE, 2001. 362−369

19 Chang H, Yeung D Y. Robust locally linear embedding. *Pattern Recognition*, 2006, **39**(6): 1053−1065

20 Golub G H, Van L C F. *Matrix Computations (Third Edition)*. London: The Johns Hopkins University Press, 1996. 396−397

21 Martinez A M. AR face database [Online], available: http://www2.ece.ohio-state.edu/∼aleix/ARdatabase.html, February 16, 2011

22 Zhang Xian-Da. *Matrix Analysis and Applications*. Beijing: Tsinghua University Press, 2004. 630−637
(张贤达. 矩阵分析与应用. 北京: 清华大学出版社, 2004. 630−637)

刘胜蓝 辽宁师范大学计算机与信息技术学院硕士研究生. 主要研究方向为模式识别.
E-mail: liushenglan-0787@163.com
(**LIU Sheng-Lan** Master student at the College of Computer and Information Technology, Liaoning Normal University. His main research interest is pattern recognition.)

闫德勤 辽宁师范大学计算机与信息技术学院教授. 1999 年获南开大学博士学位. 主要研究方向为模式识别. 本文通信作者. E-mail: yandeqin@163.com
(**YAN De-Qin** Professor at the College of Computer and Information Technology, Liaoning Normal University. He received his Ph. D. degree at Nankai University in 1999. His main research interest is pattern recognition. Corresponding author of this paper.)