

# 基于迹范数的 L1-PCA 算法

刘丽敏<sup>1a</sup>, 樊晓平<sup>1a,2</sup>, 廖志芳<sup>1b</sup>

(1. 中南大学 a. 信息科学与工程学院; b. 软件学院, 长沙 410075; 2. 湖南财政经济学院信息管理系, 长沙 410205)

**摘要:** L1-PCA 相比传统的主成分分析(PCA)更具鲁棒性, 但是 L1-PCA 算法存在很多局部最优解且秩约束计算较为复杂。为此, 提出一种基于迹范数的 L1-PCA 算法。利用迹范数近似代替矩阵的秩, 以解决秩约束存在很多局部最优解的问题, 采用基于增强拉格朗日乘子的方法对算法求解, 并将其应用于图像的降噪处理。实验结果表明, 利用该算法降噪后的图像轮廓清晰、同类图像特征明显趋同。

**关键词:** 主成分分析; 迹范数; 增强拉格朗日乘子; 闭合形式解; 奇异值分解

## L1-PCA Algorithm Based on Trace Norm

LIU Li-min<sup>1a</sup>, FAN Xiao-ping<sup>1a,2</sup>, LIAO Zhi-fang<sup>1b</sup>

(1a. School of Information Science and Engineering; 1b. School of Software, Central South University, Changsha 410075, China;  
2. Department of Information Management, Hunan College of Finance and Economics, Changsha 410205, China)

**【Abstract】** Compared with Principal Component Analysis(PCA), L1-PCA has better robustness. But, there are some problems in the L1-PCA such as locally optimal solutions, computational complexity of rank. In order to solve the problems, the paper proposes a new algorithm of L1-PCA based on trace norm, it uses trace norm approximate to instead of matrix rank, and solves the problem that rank constraint has many local optimal solution, and the solution algorithm is based on Augmented Lagrange Multiplier(ALM) and applies it in noise reduction of images. Experimental results show that the image outline of this algorithm after doing noise reduction is clear and objects within the same class become more similar.

**【Key words】** Principal Component Analysis(PCA); trace norm; Augmented Lagrange Multiplier(ALM); closed form solution; Singular Value Decomposition(SVD)

DOI: 10.3969/j.issn.1000-3428.2013.04.046

### 1 概述

随着技术的进步, 许多应用领域的的数据呈现出高维的特点。维数的增长给数据分析带来了巨大的挑战。通过降噪提取数据中重要的特征来降低高维数据的维数引起人们的重视。近些年来, 国内外很多学者对此问题进行了大量的研究, 提出了很多算法<sup>[1-2]</sup>。其中, 主成分分析(Principal Component Analysis, PCA)是一种无监督的数据分析方法, 已经广泛应用于数据降维、特征提取、图像识别等领域。但是传统的 PCA 是基于 L2 范数(以下简称为 L2-PCA), 然而 L2 范数是基于通常意义下的欧拉距离, 它对异常值和特征噪声敏感。

为了解决这个问题并优化算法, 文献[3]提出基于 L1 范数协方差估计的方法; 文献[4]提出利用半正定规划问题来优化 L1 范数的方法; 文献[5]提出联合 L2 范数和 L1 范

数来提取健壮子空间特征的方法; 文献[6]提出基于核函数的 L1-PCA 算法; 文献[7]提出利用信息熵来调节重建误差权值的 PCA 算法。但是上述算法都存在 2 个问题:

(1)存在很多局部最优解。

(2)这些算法通常是通过降低矩阵的秩实现降维, 然而在数值计算中, 矩阵的秩计算复杂。文献[8-9]提出用迹范数代替矩阵的秩, 并从理论上证明了在满足一定条件下, 迹范数代替矩阵的秩是可行的。

受此启发, 本文提出一种基于迹范数的 L1-PCA 算法, 针对算法优化问题提出了基于增强拉格朗日乘子的方法, 并将算法应用于 Yale 人脸数据集和 bin-alpha 数据集进行降噪处理。

### 2 基于迹范数的 L1-PCA 算法

传统的 PCA 通常是利用矩阵的秩来实现数据降维。假

**基金项目:** 国家“863”计划基金资助项目(2007AA022008); 国家科技支撑计划基金资助项目(2012BAH08B00)

**作者简介:** 刘丽敏(1976—), 女, 讲师、博士研究生, 主研方向: 数据挖掘, 机器视觉, 智能信息处理; 樊晓平, 教授、博士、博士生导师; 廖志芳, 副教授、博士

**收稿日期:** 2012-05-07 **修回日期:** 2012-07-11 **E-mail:** liulimin@csu.edu.cn

设给定数据矩阵  $X = \{x_1, x_2, \dots, x_n\} \in R^{p \times n}$  :

$$X \approx FG \quad (1)$$

其中, 矩阵  $F$  和矩阵  $G$  的优化解为:

$$\min_{F,G} \|X - FG\|_2^2, \text{ 且 } \text{rank}(FG) = k \quad (2)$$

其中,  $\|A\|_2$  指 L2 范数, 定义为  $\|A\|_2 = \left(\sum_{j=1}^n \sum_{i=1}^p |a_{ij}|^2\right)^{1/2}$ ,  $\text{rank}(A)$  指矩阵的秩。L2 范数对异常值和特征噪声敏感, 更具鲁棒性的 PCA 是基于 L1 范数(即 L1-PCA)。L1-PCA 的数学描述如下:

$$\min_{F,G} \|X - FG\|_1, \text{ 且 } \text{rank}(FG) = k \quad (3)$$

式(3)的  $\|A\|_1$  是指 L1 范数, 定义为:

$$\|A\|_1 = \sum_{j=1}^n \sum_{i=1}^p |a_{ij}|$$

然而基于 L1 范数的 PCA 存在很多局部最优解的问题。为此, 令式(3)中  $FG$  近似等于  $XZ$ , 即  $FG \equiv XZ$ , 式(3)近似等价于:

$$\min_Z \|X - XZ\|_1 \quad (4)$$

其中,  $k \ll p, k \ll n$  且  $\text{rank}(FG) = \text{rank}(XZ)$ 。从式(4)可以看出  $\text{rank}(XZ) = \min(\text{rank}(X), \text{rank}(Z))$ , 因此, 可以推出  $\text{rank}(Z) = k$ 。

然而, 秩约束存在很多局部最优解的问题。文献[10]提出利用迹范数近似代替秩, 而迹范数是凸问题, 并且可以求得唯一最优解。遵循这种思想, 本文利用迹范数  $\|Z\|_{\text{tr}}$  代替其秩。式(4)可以近似等价于式(5):

$$\min_Z \|X - XZ\|_1, \text{ 且 } \|Z\|_{\text{tr}} \leq k \quad (5)$$

式(5)可近似等价于:

$$\min_Z \alpha \|X - XZ\|_1 + \|Z\|_{\text{tr}} \quad (6)$$

迹范数  $\|Z\|_{\text{tr}}$  定义为:

$$\|Z\|_{\text{tr}} = \text{Tr}(ZZ^T)^{1/2} = \sum_{k=1}^{\text{rank}(Z)} \sigma_k$$

其中,  $\sigma_k$  为  $Z$  的特征值;  $\alpha$  是一个正常数。这就是本文提出的基于迹范数的 L1-PCA 算法。算法利用迹范数近似代替矩阵的秩, 从而解决了秩约束存在很多局部最优解的问题。算法的隐含意义是在一定的条件下, 利用图像数据中部分元素就能以很高的概率近似恢复原始图像。

### 3 基于增强拉格朗日乘子的求解方法

对于式(6)的求解问题, 本文采用基于增强拉格朗日乘子的优化方法。在式(6)中, 令  $U = X - XZ$ ,  $V = Z$ , 式(6)可以等价于:

$$\min_{U,V,Z} \alpha \|U\|_1 + \|V\|_{\text{tr}} \quad (7)$$

针对式(7)的求解问题, 本文采用增强拉格朗日乘子(Augmented Lagrange Multiplier, ALM)<sup>[11]</sup>的方法解决。根据

增强拉格朗日乘子的相关定义构造拉格朗日函数, 式(7)可以等价于:

$$\begin{aligned} \min_{U,V,Z} J(U,V,Z / A,B,\lambda) = & \alpha \|U\|_1 + \langle B, X - XZ - U \rangle + \\ & \lambda/2 \|X - XZ - U\|_2^2 + \|V\|_{\text{tr}} + \\ & \langle A, Z - V \rangle + \lambda/2 \|Z - V\|_2^2 \end{aligned} \quad (8)$$

其中,  $A$ 、 $B$  是拉格朗日乘子;  $\lambda$  是惩罚因子;  $\langle M, N \rangle$  定义为  $\langle M, N \rangle = \sum_{ij} M_{ij} N_{ij} = \text{tr} M^T N$ 。

ALM 是一种迭代更新算法, 在迭代更新过程中需要给出问题求解过程和相关参数的迭代方法。

#### 3.1 问题求解过程

对于输入  $A$ 、 $B$  和  $\lambda$ , 算法的关键就是解决式(8)的最优解问题。这个问题可以通过  $U$ 、 $V$  和  $Z$  闭合形式解决:

(1) 算法固定  $(V, Z)$  来求解  $U$ , 式(8)可以等价于:

$$\min_U \alpha \|U\|_1 + \langle B, X - XZ - U \rangle + \lambda/2 \|X - XZ - U\|_2^2 \quad (9)$$

式(9)可以等价于:

$$\min_U \alpha \|U\|_1 + \lambda/2 \|U - (X - XZ + B/\lambda)\|_2^2 \quad (10)$$

式(10)闭合形式解的求解过程如下:

令  $Q = X - XZ + B/\lambda$ , 式(10)等价于:

$$\min_U \alpha \|U\|_1 + \lambda/2 \|U - Q\|_2^2 \quad (11)$$

式(11)的问题可以分解为  $p$  个子问题:

$$\min_U \sum_{i=1,2,\dots,p} \alpha \|u_i\|_1 + \lambda/2 \|u_i - q_i\|_2^2 \quad (12)$$

根据文献[12], 式(12)的解为:

$$u_i^* = \text{sign}(q_i) (|q_i| - \alpha/\lambda)_+$$

其中,  $u_i$  是矩阵  $U$  的第  $i$  个分量;  $q_i$  是矩阵  $Q$  的第  $i$  个分量。利用这个结果, 式(9)的闭合形式解为:

$$U_y^* = \text{sign}(Q_{ij}) (|Q_{ij}| - \alpha/\lambda)_+$$

其中,  $Q = X - XZ + B/\lambda$ 。

(2) 算法固定  $(U, Z)$  来求解  $V$ 。式(8)可写成:

$$\min_V \|V\|_{\text{tr}} + \langle A, Z - V \rangle + \lambda/2 \|Z - V\|_2^2 \quad (13)$$

式(13)可以等价于:

$$\min_V \|V\|_{\text{tr}} + \lambda/2 \|V - (Z + A/\lambda)\|_2^2 \quad (14)$$

式(14)可通过奇异值分解(Singular Value Decomposition, SVD)获得, 即:

$$SVD(R) = P \Sigma Q^T$$

其中,  $P = (p_1, p_2, \dots, p_r)$  是矩阵向量  $R$  的左奇异向量;  $Q = (q_1, q_2, \dots, q_r)$  是矩阵向量  $R$  的右奇异向量;  $\Sigma =$

$\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ , 且  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ ,  $r = \text{rank}(\mathbf{R})$ 。  
则  $V$  的解为:

$$V_{ij}^* = P(\Sigma - \frac{1}{\lambda} I)_+ Q^T$$

其中, 矩阵  $S_+$  定义为  $S_+ = 0.5(\|S_{ij}\| + S_{ij})$ 。

(3) 算法固定  $(U, V)$  来求解  $Z$ , 式(8)可以等价:

$$\begin{aligned} \min_Z J = & \langle B, X - XZ - U \rangle + \\ & \frac{\lambda}{2} \|X - XZ - U\|_2^2 + \\ & \langle A, Z - V \rangle + \frac{\lambda}{2} \|Z - V\|_2^2 \end{aligned} \quad (15)$$

令  $\frac{\partial J}{\partial Z} = 0$ , 可以求得  $Z$  的解。即为:

$$Z = (I + X^T X)^{-1} (V + X^T (X - U + \frac{B}{\lambda}) - \frac{A}{\lambda})$$

### 3.2 参数迭代方法

在增强拉格朗日乘子迭代过程中, 每次迭代得到  $U$ 、 $V$  和  $Z$ , 其中, 参数  $A$ 、 $B$  和  $\lambda$  的迭代方法如下:

$$\begin{aligned} A & \leftarrow A + \lambda(Z - V) \\ B & \leftarrow B + \lambda(Z - XZ - U) \\ \lambda & \leftarrow \lambda\rho \end{aligned} \quad (16)$$

其中,  $\rho$  是一个常量, 并且  $\rho > 1$ 。从式(8)可以看出, 公式不仅使用拉格朗日乘子  $A$  和  $B$  来约束  $U$  和  $V$ , 而且还增加了 2 项惩罚项  $\frac{\lambda}{2} \|X - XZ - U\|_2^2$  和  $\frac{\lambda}{2} \|Z - V\|_2^2$ , 这就从两方面约束了  $U$  和  $V$ 。

### 3.3 基于迹范数的 L1-PCA 算法框架

本文算法框架如下:

**输入**  $X$

**输出**  $U$ 、 $V$ 、 $Z$

初始化:  $A = 0, B = 0, \rho = 1.2$

计算:  $\lambda = \frac{1}{\|X\|_1}$

Do  $\lambda$  to  $\min(\lambda\rho, 10^{10})$

$U = \arg \min_U \|U\|_1 + \langle B, X - XZ - U \rangle +$

$$\frac{\lambda}{2} \|X - XZ - U\|_2^2$$

$V = \arg \min_V \|V\|_{tr} + \langle A, Z - V \rangle + \frac{\lambda}{2} \|V - (Z + \frac{A}{\lambda})\|_2^2$

$Z = \arg \min_Z \langle A, Z - V \rangle + \langle B, X - XZ - U \rangle +$

$$\frac{\lambda}{2} \|X - XZ - U\|_2^2 + \frac{\lambda}{2} \|V - (Z + \frac{A}{\lambda})\|_2^2$$

Update  $A$ ,  $A \leftarrow A + \lambda(Z - V)$ ;

Update  $B$ ,  $B \leftarrow B + \lambda(Z - XZ - U)$ ;

Update  $\lambda$ ,  $\lambda \leftarrow \lambda\rho$ ;

undo;

return  $U, V, Z$

从算法框架和式(16)可看出, 在优化迭代过程中,  $\lambda$  的值越来越大, 但是  $\lambda$  不能无限大, 因此, 在所提出的算

法框架中  $\lambda$  的取值为  $\min(\lambda\rho, 10^{10})$ 。按照拉格朗日乘子的定义, 算法中的  $A$ 、 $B$  并不是标准意义下的拉格朗日乘子。实际上是拉格朗日乘子的近似估计。从算法中还可以看出, 在每次迭代后,  $U$ 、 $V$  和  $Z$  值的变化越来越小, 因此,  $U$ 、 $V$  和  $Z$  可解, 算法可以收敛。

算法的输入参数为原始矩阵  $X$ , 需要初始化的参数包括增强拉格朗日乘子的迭代参数  $A$ 、 $B$ 、 $\lambda$  以及限制迭代次数的参数  $\rho$  和  $Z$  的初始估计值(后续的实验将其估计为  $0.9X$ )。算法首先初始化相应的参数值, 然后开始迭代计算  $U$ 、 $V$  和  $Z$  的值, 并更新参数  $A$ 、 $B$ 、 $\lambda$  的值, 当满足收敛条件  $\min(\lambda\rho, 10^{10})$  时, 算法停止迭代计算给出结果并返回。

算法的时间复杂度主要由求解  $U$ 、 $V$ 、 $Z$  和迭代系数决定。其中, 求解  $U$  利用了闭合形式解, 而求解闭合形式解的时间复杂度为  $O(pn)$ ; 求解  $V$  利用了奇异值分解, 而奇异值分解的时间复杂度为  $O(\min(p^2n, pn^2))$ ;  $Z$  的求解是线性时间。因此, 如果迭代次数为  $m$ , 算法的时间复杂度为  $O(m \times \min(p^2n, pn^2))$ ; 算法的空间复杂度可以表示为  $O(\max(p^2n, pn^2))$ 。

## 4 实验结果与分析

为了评价所提出的基于迹范数的 L1-PCA 及其求解算法的有效性, 将本文算法应用于 Yale 人脸数据集和 bin-alpha 数据集进行图像降噪处理。

首先从 Yale 人脸数据选出部分图像, 并将图像大小调整为  $32 \times 32$  像素。为了评估所提算法的有效性, 对每个图像增加随机噪声, 随机噪声的增加, 对图像设置方形遮挡。假如方形遮挡的尺寸设为  $d$ , 遮挡数量设为  $m$ , 对每个图像随机选取  $m$  个位置, 将其  $d \times d$  像素部分的值设为 0。图 1 显示了设置 4 种不同遮挡增加随机噪声的例子, 即增加随机噪声的图像。

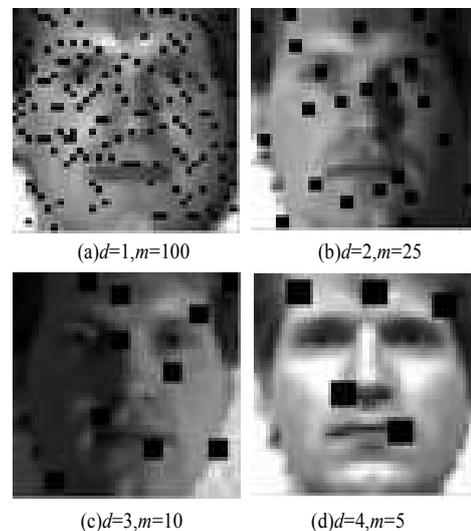
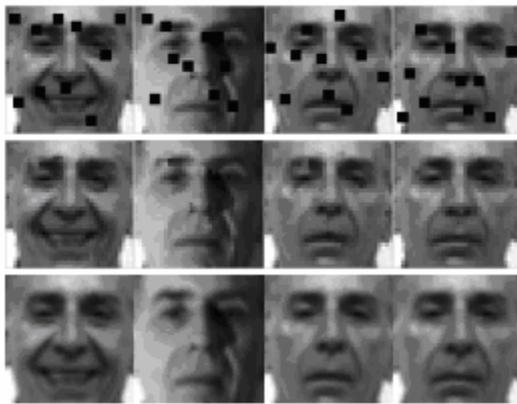
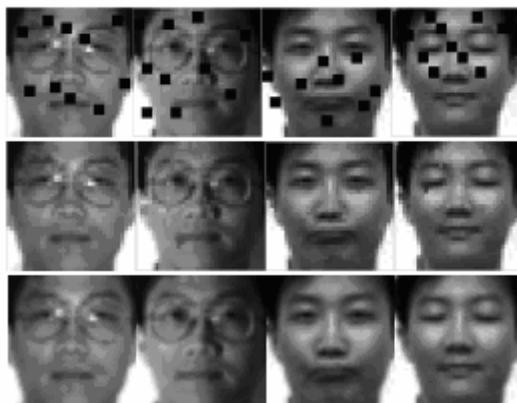


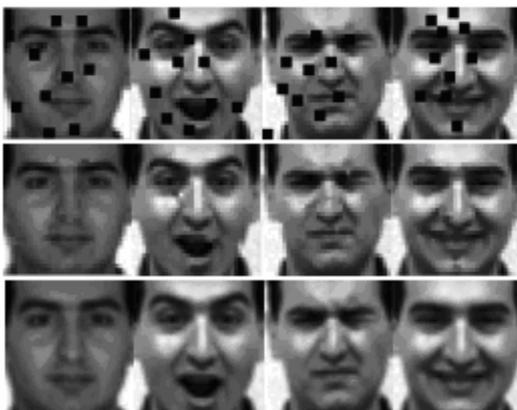
图 1 增加随机噪声的图像



(a)第1组图像对比分析



(b)第2组图像对比分析



(c)第3组图像对比分析

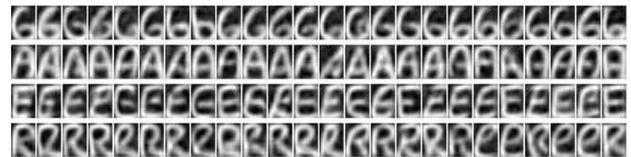
图2 Yale人脸数据集降噪效果比较

图2(a)~图2(c)显示了3组Yale人脸数据集的图像在随机增加了10块 $3 \times 3$ 遮挡图像的情况下,应用PCA算法和本文算法降噪效果对比分析。图2(a)~图2(c)的第1行是增加了随机噪声的图像;第2行是利用PCA算法的降噪效果;第3行是利用本文提出的算法降噪的效果。从图2(a)~图2(c)可以看出,本文算法降噪效果明显、图像特征更加清晰。

将本文算法应用于bin-alpha数据集进行降噪处理,每张图像大小调整为 $20 \times 16$ 像素,部分结果如图3所示。从图3(b)中可以看出,字母和数字的轮廓清晰、同类图像特征也明显趋同。



(a)bin-alpha数据集原始图像



(b)bin-alpha数据集降噪后图像

图3 bin-alpha数据集降噪效果比较

## 5 结束语

为了解决传统PCA和L1-PCA存在的问题,本文提出一种基于迹范数的L1-PCA算法。实验结果表明,该算法可以得到近似的唯一最优解。今后将针对算法的参数优化、收敛性、精度以及准确性做进一步研究,并通过理论证明和实验分析继续优化算法。

## 参考文献

- [1] Jolliffe I T. Principal Component Analysis[M]. 2nd ed. New York, USA: Springer, 2002.
- [2] Tenenbaum J B, Silva V D, Langford J C. A Global Geometric Framework for Nonlinear Dimensionality Reduction[J]. Science, 2000, 290(5500): 2319-2323.
- [3] Galpin J, Hawkins D. Methods of L1 Estimation of a Covariance Matrix[J]. Computational Statistics and Data Analysis, 1987, 5(4): 305-319.
- [4] Ke Qifa, Kanade T. Robust L1 Norm Factorization in the Presence of Outliers and Missing Data by Alternative Convex Programming[C]//Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE Press, 2005.
- [5] Chris D, Zhou Ding, He Xiaofang. R1-PCA: Rotational Invariant L1-norm Principal Component Analysis for Robust Subspace Factorization[C]//Proc. of the 23rd International Conference on Machine Learning. New York, USA: ACM Press, 2006.
- [6] 李勇,梁志贞,夏士雄.基于核函数的PCA-L1算法[J].计算机工程,2011,37(22):174-178.
- [7] 杨开睿,孟凡荣,梁志贞.一种自适应权值的PCA算法[J].计算机工程与应用,2012,48(3):189-191.
- [8] Candes E J, Recht B. Exact Low-rank Matrix Completion via Convex Optimization[C]//Proc. of the 46th Annual Allerton Conference on Communication, Control, and Computing. Monticello, USA: IEEE Press, 2008.

(下转第209页)