

基于 TL_1 范数约束的子空间聚类方法

李海洋* 王恒远

(西安工程大学理学院 西安 710048)

摘要: 该文将 TL_1 范数应用于子空间聚类的研究中, 提出基于 TL_1 范数约束的子空间聚类优化模型。尽管该优化模型是非凸的, 在无噪音的情形下, 证明了它的最优解为具有块对角结构的系数矩阵, 这对随后进行的谱聚类提供了理论保证; 在有噪音的情形下, 它的约束条件等价于以干净数据为字典的优化模型, 因而求解出的系数矩阵提高了聚类的精确度。进一步, 利用增广拉格朗日-交替方向乘子方法给出该优化模型的求解方法。实验结果表明, 基于 TL_1 范数的子空间聚类方法不仅增强了系数矩阵的稀疏性, 而且在聚类精确度, 对噪音的鲁棒性方面要优于低秩子空间聚类方法和稀疏子空间聚类方法。

关键词: TL_1 范数; 子空间聚类方法; 稀疏; 低秩; 谱聚类

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2017)10-2428-09

DOI: 10.11999/JEIT170193

Subspace Clustering Method Based on TL_1 Norm Constraints

LI Haiyang WANG Hengyuan

(School of Science, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: The TL_1 norm is applied to propose a new optimization model for the study of subspace clustering. Although the optimization is nonconvex, in the case of non-noise, it proves that the optimal solution of the proposed model is the coefficient matrix with block-diagonal structure, which provides the theoretical guarantee for the latter spectral clustering. In the case of dealing with noise, the constraint condition of this model is presented to be equivalent with the optimal model using the corrected data as the dictionary, which contributes to improving the clustering accuracy. Then, the alternating direction method of Lagrangian multipliers is applied to solving the unknown matrices. Experimental results show that subspace clustering method based on TL_1 norm not only enhances the sparsity of coefficient matrix, but also is superior to low-rank subspace clustering and sparse subspace clustering method in terms of clustering accuracy and robustness to noise.

Key words: TL_1 norm; Subspace clustering method; Sparse; Low-rank; Spectral clustering

1 引言

高维数据在信息高速发展的今天越来越普遍, 如何有效地组织这些高维数据成为亟待解决的难题。高维数据聚类的目标是根据数据点的特征或属性将属于同一类的数据组织在一起^[1]。子空间聚类是实现高维数据集聚类的有效途径, 被广泛地应用在计算机视觉^[2,3]与模式识别^[4,5]等领域。常见的子空间聚类算法主要分为4类: 代数方法、统计方法、迭代方法和基于谱聚类的方法^[6]。

近年来, 诸多学者提出了许多有效的基于谱聚类的子空间聚类算法^[7-10]。这些算法都需要两个过程: (1)基于数据点的自表达性质(子空间中的数据点能够表示为同一子空间中其它数据点的线性组合), 找到属于同一子空间的数据点, 建立数据点之间的关系; (2)根据第(1)步得出数据点之间的关系, 构建相似度矩阵, 然后运用谱聚类^[11]的方法进行聚类。其中, 如何有效地构建数据点之间的关系将直接影响之后的聚类效果。

为了找到子空间中每个数据点的最优表示, 文献^[7-10]提出基于稀疏表示(Sparse Representation, SR)的子空间聚类方法(Sparse Subspace Clustering, SSC)^[7,8]、基于低秩表示(Low Rank Representation, LRR)的低秩子空间聚类方法^[9,10], 这些方法通过寻找数据在自身数据字典上的稀疏或者低秩表示来得到系数矩阵, 与其他子空间聚类算法相比, 基于稀

收稿日期: 2017-03-03; 改回日期: 2017-06-27; 网络出版: 2017-08-14

*通信作者: 李海洋 fplihaiyang@126.com

基金项目: 国家自然科学基金(11271297), 陕西省自然科学基金(2015JM1012)

Foundation Items: The National Natural Science Foundation of China (11271297), The Natural Science Foundation of Shaanxi Province (2015JM1012)

疏和低秩表示的方法可以有效地处理噪声和异常值。然而，稀疏子空间聚类方法是对每一个数据进行单独地分析，忽略了数据之间的关联性。基于矩阵核范数的 LRR 方法尽管考虑了数据点的全局结构，但是该方法所求出的系数矩阵不够稀疏，使得子空间表示能力受到很大限制。针对此，文献[12]利用矩阵的 Forbenius 范数对系数矩阵进行约束提高了子空间聚类的效率。文献[13,14]同时对系数矩阵进行稀疏和低秩约束。文献[15]通过联合矩阵的核范数和 Forbenius 范数作为正则项，提高了算法的稳定性。实验结果表明这些方法要优于单一约束的 SSC 方法和 LRR 方法。

遗憾的是，在数据点受噪声影响的情况下，上述方法均采用噪声数据本身作为表示字典，而这会对寻找数据间相关性的准确程度造成不利影响。为了同时追求表示系数矩阵之间的紧连接和模型对噪声的鲁棒性，本文做了以下几方面的工作：(1)受到向量的 TL_1 范数在压缩感知中良好表现的启发，在构建目标函数时，我们使用矩阵的 TL_1 范数来代替秩函数，提高了系数矩阵的稀疏性和对噪声的鲁棒性；(2)证明所提模型得到的系数矩阵具有块对角结构，这对后面的聚类在理论上具有至关重要的作用；(3)针对有噪声的情形，本文给出的约束条件等价于先利用列稀疏约束的鲁棒主成分分析去除噪声，再以干净数据为字典对数据点进行表达，能够更好地展现聚类模型对噪声的鲁棒性；(4)利用增广拉格朗日-交替方向乘子方法给出模型的有效求解方法，对求得的系数矩阵再利用经典的谱聚类方法^[11]进行聚类。

本文的内容安排如下：第2节主要介绍稀疏表示子空间聚类模型和低秩表示子空间聚类模型；第3节详细阐述基于矩阵 TL_1 范数的子空间聚类模型，并对该模型进行了理论分析；第4节利用交替方向乘子方法给出本文所提模型的最优解，并对收敛性进行简要分析；第5节是在合成数据集、Extend Yale B 数据库^[6]、Hopkins155 数据集、USPS 数据库上的仿真实验及结果；第6节对本文进行了总结。

2 预备知识

给定一组待聚类的数据 $D = [d_1, d_2, \dots, d_N] \in R^{M \times N}$ ， D 的 N 个数据点来自于 k 个未知维数的子空间 $\{S_i\}_{i=1}^k$ 的并，且子空间 $\{S_i\}_{i=1}^k$ 是相互独立的，即 $\dim(\bigoplus_{i=1}^k S_i) = \sum_{i=1}^k \dim(S_i)$ 。子空间聚类的目标是求解子空间的数目 k 和它们的维数，并将数据点分割到其相应的子空间中。

2.1 稀疏子空间聚类

稀疏子空间聚类^[7,8]包含两步：首先，针对每一个数据点 d_i ，通过求解以下稀疏表示问题获取稀疏系数向量 C_i ：

$$\min_{C_i} \|C_i\|_0, \quad \text{s.t. } d_i = DC_i$$

其中， $D = [d_1, d_2, \dots, d_{i-1}, d_{i+1}, \dots, d_n]$ ，由于 l_0 范数最小化的稀疏求解问题是 NP-hard 问题，一定条件下，上述 l_0 优化问题等价于式(1)所示的凸 l_1 优化问题：

$$\min_{C_i} \|C_i\|_1, \quad \text{s.t. } d_i = DC_i \quad (1)$$

其次，把求得的系数向量 C_i 按列排成系数矩阵 $C = [C_1, C_2, \dots, C_n]$ ，构建相似度矩阵 $W = |C| + |C^T|$ ，其中 $|x|$ 表示逐项取绝对值。最后再利用谱聚类^[11]的方法来实现最终的聚类。

2.2 低秩表示子空间聚类

与稀疏子空间聚类相比，低秩子空间聚类方法^[9,10]更加全局地考虑了数据点之间的相互关系，能够找到数据点之间的紧连接。具体地，低秩表示子空间聚类模型可以表示为

$$\min_C \text{rank}(C), \quad \text{s.t. } D = DC$$

事实上，该优化问题依然是 NP-hard，求解非常困难。相关研究表明，秩优化问题在一定条件下等价于式(2)所示的核范数凸优化问题^[8]：

$$\min_C \|C\|_*, \quad \text{s.t. } D = DC \quad (2)$$

其中， $\|x\|_*$ 表示矩阵的核范数即矩阵的所有奇异值之和。对于数据点含有噪声或野点的情形，可以通过加入误差项 E 来增加鲁棒性：

$$\min_{C, E} \|C\|_* + \lambda \|E\|_{2,1}, \quad \text{s.t. } D = DC + E \quad (3)$$

其中， $\|E\|_{2,1} = \sum_{j=1}^N \sqrt{\sum_{i=1}^M (E_{ij}^2)}$ 用来度量表示误差， λ 为平衡参数。

3 基于 TL_1 范数的子空间聚类模型

2000 年，Nikolova 在研究变量选择时首次提出了 TL_1 范数的概念^[16]。2009 年文献[17]研究了 TL_1 的无偏性、稀疏性和连续性。2014 年文献[18,19]将 TL_1 范数应用到压缩感知，提出了基于 TL_1 范数的阈值迭代算法等。本文将矩阵 TL_1 范数应用到子空间聚类的研究中，获得了很好的效果。下面我们先给出矩阵 TL_1 范数的定义及其有关性质^[18,19]：

矩阵的 TL_1 范数定义如下： $\|x\|_{TL_1} = \sum_{i=1}^{\text{rank}(*)} \rho_a(\sigma_i)$ ，其中 σ 代表矩阵 $*$ 的奇异值， $\rho_a(x) = \frac{(a+1)|x|}{a+|x|}$ ， $a \in (0, +\infty)$ 为参数。

上述函数 $\rho_a(*)$ 满足以下性质: (1) 正定性: $\rho_a(\mathbf{x}) \geq 0$ 且 $\rho_a(x) = 0 \Leftrightarrow x = 0$; (2) 三角不等式: $\rho_a(\mathbf{x}_i + \mathbf{x}_j) \leq \rho_a(\mathbf{x}_i) + \rho_a(\mathbf{x}_j)$; (3) $\rho_a(|c\mathbf{x}|) \begin{cases} \leq |c|\rho_a(|\mathbf{x}|), & |c| > 1 \\ \geq |c|\rho_a(|\mathbf{x}|), & |c| \leq 1 \end{cases}$, 但我们仍把它称作 TL_1 范数。另外, 它显然满足以下性质: (1) $\rho(t)$ 递增且在 $t \in [0, \infty)$ 是凹的; (2) $\rho'(t)$ 在 $t \in (0, \infty)$ 是连续的。

在无噪音的情况下, 本文提出了基于 TL_1 范数子空间聚类模型, 具体可以表示为

$$\min_C \|C\|_{TL_1}, \quad \text{s.t. } D = DC \quad (4)$$

下面我们首先证明式(4)得到的系数矩阵 C^* 具有块对角结构, 这对后面的聚类在理论上具有至关重要的作用。

定理 1 假设数据点样本充分采样于 k 个相互独立的子空间 $S_i, i = 1, 2, \dots, k$, 式(4)的解 C^* 具有块对角的结构:

$$C^* = \begin{bmatrix} C_1^* & 0 & 0 & 0 \\ 0 & C_2^* & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & C_k^* \end{bmatrix}$$

为了证明定理 1, 首先给出以下引理:

引理 1^[20] 针对于 $A \in M_{n,p}, B \in M_{p,m}, q = \min\{n, p, m\}$, A, B, AB 的奇异值按顺序排列分别为 $\sigma_1(A) \geq \sigma_2(A) \geq \dots \geq \sigma_{\min\{n,p\}}(A) \geq 0, \sigma_1(AB) \geq \sigma_2(AB) \geq \dots \geq \sigma_{\min\{n,m\}}(AB) \geq 0, \sigma_1(B) \geq \sigma_2(B) \geq \dots \geq \sigma_{\min\{n,m\}}(B) \geq 0$, 则有以下结论成立: $\sum_{i=1}^k \sigma_i(AB) \leq \sum_{i=1}^k \sigma_i(A)\sigma_i(B)$ 。

引理 2^[20] 对任意 $C \in M_{m,n}, V_k \in M_{m,k}, W_k \in M_{n,k}$, 满足 $k \leq \min\{m, n\}, V_k^H V_k = I, W_k^H W_k = I$, 有以下结论成立:

(1) $\sigma_i(V_k^H C W_k) \leq \sigma_i(C)$, (2) $|\det V_k^H C W_k| \leq \sigma_1(C) \cdot \sigma_2(C) \cdot \dots \cdot \sigma_k(C)$ 。

引理 3 令 A 和 D 是方形矩阵, 对于任意兼容维数的矩阵 B 与矩阵 C , 有

$$\left\| \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right\|_{TL_1} \geq \left\| \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix} \right\|_{TL_1}$$

证明 设矩阵 $R = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ 的奇异值按顺序排列有 $\sigma_1(R) \geq \sigma_2(R) \geq \dots \geq \sigma_k(R) \geq 0$, 矩阵 $r =$

$\begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}$ 的奇异值按顺序排列为 $\sigma_1(r) \geq \sigma_2(r) \geq \dots$

$\geq \sigma_k(r) \geq 0$ 。对任意 $C \in M_{m,n}, V_k \in M_{m,k}, W_k \in M_{n,k}$, 满足 $k \leq \min\{m, n\}, V_k^H V_k = I, W_k^H W_k = I$, 我们有

$$\begin{aligned} \|M\|_{TL_1} &= \sum_{i=1}^k \frac{(a+1)\sigma_i(M)}{a+\sigma_i(M)} = \sum_{i=1}^k \frac{(a+1)\sigma_i(M)\sigma_i(C)}{a+\sigma_i(M)\sigma_i(C)} \\ &\geq \sum_{i=1}^m \frac{(a+1)\sigma_i(MC)}{a+\sigma_i(MC)} \\ &\geq \sum_{i=1}^m \frac{(a+1)\sigma_i(V_k^H M W_k)}{a+\sigma_i(V_k^H M W_k)} \end{aligned}$$

其中, $C = W_k V_k^H$ 且 $C^H C = V_k W_k^H W_k V_k^H = V_k V_k^H$ 。第 2 个等式根据 $V_k V_k^H$ 的第 k 个最大奇异值是 1 可得。第 1 个不等式根据引理 1 可得。第 2 个不等式根据引理 2 可得。所以

$$\begin{aligned} \left\| \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right\|_{TL_1} &\geq \sum_{i=1}^k \frac{(a+1)\sigma_i \left[\begin{bmatrix} V_1^H & V_2^H \\ 0 & 0 \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} W_1 & 0 \\ W_2 & 0 \end{bmatrix} \right]}{a+\sigma_i \left[\begin{bmatrix} V_1^H & V_2^H \\ 0 & 0 \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} W_1 & 0 \\ W_2 & 0 \end{bmatrix} \right]} \\ &= \left\| \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix} \right\|_{TL_1} \end{aligned}$$

定理 1 的证明 令 C 是式(4)的任意一个最小解, 从 C 中, 我们可以构建一个块对角矩阵 C^* 具有以下形式:

$$C_{ij}^* = \begin{cases} C_{ij}, & d_i \text{ 与 } d_j \text{ 属于同一子空间} \\ 0, & \text{其它} \end{cases}$$

这里, d_i 与 d_j 分别代表数据点矩阵的第 i, j 列。

令 $Q = C - C^*$ 是 C 的非对角块的一部分。一般地, 假设 x_j 存在于第 l 个子空间 S_l , 那么有 $[DC]_{:,j} \in S_l$ 。从 C^* 与 Q 的定义可知, $[DC^*]_{:,j} \in S_l, [DQ]_{:,j} \in \cup_{m \neq l} S_m, [DQ]_{:,j} = [DC]_{:,j} - [DC^*]_{:,j} = [D]_{:,j} - [DC^*]_{:,j} \in S_l$ 。又由于假设子空间是相互独立的, 因此 $S_l \cap \oplus_{m \neq l} S_m = \{0\}$, 即 $[DQ]_{:,j} = 0$ 。故 C^* 也是式(4)的可行解。根据引理 3 知 $\|C\|_{TL_1} \geq \|C^*\|_{TL_1}$, 所以 C^* 是式(4)的最小解, 即式(4)的最优解有块对角的形式。证毕

在真实情况下, 数据点往往含有噪音, LRR 方法直接运用噪音数据 D 本身作为字典而不是干净

的数据点 D ，当噪音或异常值的比例相对较大时，这显然是不合理的。为了更好地处理噪音，我们考虑运用干净的数据作为字典，式(4)转变为

$$\begin{aligned} \min_{C,E,X} & \|C\|_{TL_1} + \lambda \|E\|_{2,1}, \\ \text{s.t.} & X = XC, D = X + E \end{aligned} \quad (5)$$

其中， $\|E\|_{2,1} = \sum_{j=1}^n \sqrt{\sum_{i=1}^n (E_{ij})^2}$ 叫做 $l_{2,1}$ 范数，参数 $\lambda > 0$ 是平衡参数。可以将式(5)等价地写成式(6)：

$$\min_{C,E} \|C\|_{TL_1} + \lambda \|E\|_{2,1}, \quad \text{s.t.} \quad D = (D - E)C + E \quad (6)$$

首先，相对于式(5)，式(6)的约束减少了未知量 X ，有利于模型的求解。更重要的是，式(6)利用干净的字典求解出的系数矩阵对之后的聚类有显著影响。

4 基于 TL_1 范数的子空间聚类模型求解

4.1 模型的求解

对模型式(6)，采用增广拉格朗日-交替方向乘子方法^[21,22]进行求解，引入辅助变量 J, S, F ，那么式(6)可以等价为式(7)的优化模型：

$$\begin{aligned} \min_{C,E,J} & \|J\|_{TL_1} + \lambda \|F\|_{2,1}, \quad \text{s.t.} \quad D = (D - E)S + E, \\ & S = C, F = E, J = C \end{aligned} \quad (7)$$

利用增广拉格朗日-交替方向乘子方法，有

$$\begin{aligned} \min_{Y,W,P,Q,J,F,Z,E,S} & \|J\|_{TL_1} + \text{tr}(Q^T(E - F) + P^T(C - J)) \\ & + \frac{\mu}{2} (\|J - C\|_F^2) + \frac{\mu}{2} (\|D - (D - E)S - E\|_F^2) \\ & + \frac{\mu}{2} (\|S - C\|_F^2) + \text{tr}(\|D - (D - E)S - E\|_F^2) \\ & + \text{tr}(W^T(C - S)) + \frac{\mu}{2} (\|F - E\|_F^2) + \lambda \|F\|_{2,1} \end{aligned} \quad (8)$$

其中， $\text{tr}(\ast)$ 代表给定矩阵的迹算子。 Y, W, P, Q 为拉格朗日乘子矩阵， μ 为增广项参数。

利用拉格朗日交替方向乘子方法求解式(8)如算法1。

输入：原始矩阵 D ，参数 λ 。

初始化： $J = C = 0$ ， $S = 0$ ， $F = E = 0$ ， $Y = 0$ ， $W = 0$ ， $P = 0$ ， $Q = 0$ ， $k = 0$ ， $\mu_0 = 10^{-6}$ ， $\rho = 1.1$ ， $\mu_{\max} = 10^{10}$ ， $\varepsilon = 10^{-8}$ 。

(1)固定其他，对 J 求极小，通过

$$J_{k+1} = \arg \min_J \|J\|_{TL_1} + \frac{1}{2} \left\| J - \left(C_k + \frac{P_k}{\mu_k} \right) \right\|_F^2 \quad (9)$$

(2)固定其他，对 F 求极小，通过

$$F_{k+1} = \arg \min_F \frac{\lambda}{\mu_k} \|F\|_{2,1} + \frac{1}{2} \left\| F - \left(E_k + \frac{Q_k}{\mu_k} \right) \right\|_F^2 \quad (10)$$

(3)固定其他，对 S 求极小，通过

$$S_{k+1} = \left[(E_k - D)^T (E_k - D) + I \right]^{-1} \cdot \left[C_k + (D - E_k)^T (D - E_k) + \frac{W_k + (D - E_k)^T Y_k}{\mu_k} \right] \quad (11)$$

(4)固定其他，对 C 求极小，通过

$$C_{k+1} = \frac{1}{2} \left\| C - \left(\frac{J_{k+1} + S_{k+1}}{2} - \frac{W_k + P_k}{2\mu_k} \right) \right\|_F^2 \quad (12)$$

(5)固定其他，对 E 求极小，通过

$$E_{k+1} = \left[D(S_{k+1} - I)(S_{k+1} - I)^T + F_{k+1} - \frac{Y_k(S_{k+1} - I)^T + Q_k}{\mu_k} \right] \cdot \left[(S_{k+1} - I)(S_{k+1} - I)^T + I \right]^{-1} \quad (13)$$

(6)更新乘子，通过

$$\left. \begin{aligned} Q_{k+1} &= Q_k + \mu_k (E_{k+1} - F_{k+1}) \\ P_{k+1} &= P_k + \mu_k (C_{k+1} - J_{k+1}) \\ W_{k+1} &= W_k + \mu_k (C_{k+1} - S_{k+1}) \\ Y_{k+1} &= Y_k + \mu (D - (D - E_{k+1})S_{k+1} - E_{k+1}) \end{aligned} \right\} \quad (14)$$

(7)更新参数，通过

$$\mu_{k+1} = \min \{ \rho \mu_k, \mu_{\max} \} \quad (15)$$

(8)终止条件：

$$\begin{aligned} \|Z - J\|_{\infty} < \varepsilon, \quad \|Z - S\|_{\infty} < \varepsilon, \quad \|E - F\|_{\infty} < \varepsilon, \\ \|D - (D - E)S - E\|_{\infty} < \varepsilon \end{aligned} \quad (16)$$

其中，第(1)步对 J 求极小由定理2确定。

定理 2^[18,19] 给定矩阵 $Y \in R^{M \times N}$ 有奇异值分解： $Y = U \text{Diag}(\sigma) V^T$ ， $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_m)$ ，如果 $\min_X \frac{1}{2} \|X - Y\|_F^2 + \lambda \|X\|_{TL_1}$ 的最优解是 X^* ，那么 X^* 有以下形式：

$$X^* = U \text{Diag}(g_{\lambda,a}(\sigma)) V^T$$

其中，

$$g_{\lambda,a}(\omega) = \begin{cases} 0, & |\omega| \leq t \\ h_{\lambda}(\omega), & |\omega| > t \end{cases}$$

$$h_{\lambda}(\omega) = \text{sgn}(\omega) \left\{ \frac{2}{3} (a + |\omega|) \cos \left(\frac{\phi(\omega)}{3} \right) - \frac{2a}{3} + \frac{|\omega|}{3} \right\}$$

$$\phi(\omega) = \arccos \left(1 - \frac{27\lambda a(a+1)}{2(a+|\omega|)^3} \right)$$

第(2)步对 F 求极小化由定理3确定：

定理 3^[23] 令 $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_i, \dots]$ 是给定的矩阵, $\|\cdot\|_F$ 代表矩阵的 Forbenius 范数, 如果 $\min_{\mathbf{W}} \lambda \|\mathbf{W}\|_{2,1} + \frac{1}{2} \|\mathbf{W} - \mathbf{Q}\|_F^2$ 的最优解是 \mathbf{W}^* , 那么 \mathbf{W}^* 的第 i 列有如下形式:

$$\mathbf{W}^*(:, i) = \begin{cases} \frac{\|\mathbf{q}_i\| - \lambda}{\|\mathbf{q}_i\|} \mathbf{q}_i, & \lambda < \|\mathbf{q}_i\| \\ 0, & \text{其他} \end{cases}$$

综上所述, 本文提出的基于 TL_1 范数约束的子空间聚类算法如算法 2。

输入: 数据点矩阵 $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_N] \in R^{M \times N}$, 子空间个数 k 。

步骤 1 通过求解式(6)获得最优的系数矩阵 \mathbf{C} 。

步骤 2 由系数矩阵 \mathbf{C} 构建相似度矩阵 $\mathbf{W} = \frac{1}{2}(|\mathbf{C}| + |\mathbf{C}^T|)$ 。

步骤 3 再利用谱聚类算法^[11]对相似度矩阵 \mathbf{W} 进行划分, 从而得到最终的聚类结果。

4.2 收敛性分析

当目标函数平滑时, 文献[24]给出了拉格朗日乘子法收敛性的证明。在有两个变量矩阵交替迭代的情况下, 拉格朗日乘子方法可以收敛到最优解也已经得到证明^[25]。但是, 求解 LRR 算法时, 通常包含 3 个变量矩阵(\mathbf{J} , \mathbf{Z} 和 \mathbf{E})交替迭代, 其收敛性很难在理论上证明^[10]。本文运用 ADMM 算法求解非凸 5 个变量矩阵(\mathbf{J} , \mathbf{F} , \mathbf{S} , \mathbf{C} , \mathbf{E})的优化问题, 证明其收敛性更加困难。但是, 实验中, 我们设定当迭代的次数超过 10000 时即停止迭代或者当问题式(16)中的 ε 足够小时即停止迭代。通常在实验中 $\varepsilon = 10^{-4}$ 。大量的实验证明, 拉格朗日乘子法在最优优化问题中通常能够很好地收敛。

5 数值实验

5.1 数据聚类及分析

为了验证本文算法的性能, 首先, 对本文所提出的 TL_1 算法与目前较好的子空间聚类算法 LRR, SSC 进行对比实验。在构造的合成数据集中, 我们采用分割的精确度作为衡量各个算法性能的标准, 而在 Extended Yale B 数据库中, 本文将子空间聚类错误率作为评价各个算法性能的标准。其中, 错误率越小, 子空间聚类效果越好; 反之, 则越差。其定义为

$$\text{子空间聚类错误率} = \frac{\text{错误分类的样本数}}{\text{样本总数}}$$

算法采用编程软件 Matlab R2011b 进行仿真实验,

实验采用的条件为: Microsoft Windows7, 处理器: Inter(R) Corei7, CPU@2.50 GHz, 内存: 4 GB。在参数设置方面, 文中的比较算法 LRR, LSR 的正则参数 λ 取值为 $\{0.0001, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 10\}$, SSC, LSA, LRSC 的 λ 取值为 $\{0.0001, 0.001, 0.005, 0.01, 0.05, 0.1\}$ 。本文参数的选择情况如表 1 所示。其中 Tolerance 为终止参数, μ 为拉格朗日系数, λ 为平衡参数, a 为 TL_1 范数中的参数。

表 1 本文方法的参数选择情况

| 参数 | Hopkins155 数据库 | USPS 数据库 | Extended Yale B |
|-----------|-------------------|-----------|--------------------|
| Tolerance | 10^{-4} | 10^{-4} | 10^{-4} |
| λ | 1 | 10^{-1} | 1 |
| μ | 1 | 10^3 | 1.6 |
| a | 1 | 0.3 | 0.3 |

5.1.1 人工数据集 首先按照文献[10]的方法, 构建 5 个相互独立的子空间 $\{S_i\}_{i=1}^5$, 子空间的维数为 4 维, 它的基 $\{U_i\}_{i=1}^5$ 能够通过 $U_{i+1} = \mathbf{T}U_i$, $1 \leq i \leq 4$, 计算出来, 这里 \mathbf{T} 是一个随机的矩阵, 取 U_1 维度为 100×4 的随机正交矩阵。首先, 本文通过 $\mathbf{D}_i = U_i Q_i$, 从子空间中取样 20 个 100 维的数据点向量, Q_i 是一个 4×20 的随机矩阵。由此, 我们得到了 5 个子空间样本为 20×100 的数据点矩阵。选择一部分数据点随机加入服从均值为 0, 标准差为 $0.3\|\cdot\|$ 的高斯分布噪音 ($\|\cdot\|$ 表示选取向量的 l_2 范数), 以用来检验算法对噪音的鲁棒性。在得到系数矩阵之后, 运用谱算法^[11]将数据点分成 5 类并观察每类方法分割的准确率。其中, 对于稀疏表示的子空间聚类方法, 我们采用两种形式。第 1 种形式是: $\text{SR}_{2,1} : \min_{\mathbf{Z}, \mathbf{E}} \|\mathbf{Z}\|_1 + \lambda_2 \|\mathbf{E}\|_{2,1}$, s.t. $\mathbf{D} = \mathbf{DZ} + \mathbf{E}$, $\text{diag}(\mathbf{Z}) = 0$ 。第 2 种是标准形式: $\text{SR}_1 : \min_{\mathbf{Z}, \mathbf{E}} \|\mathbf{Z}\|_1 + \lambda_1 \|\mathbf{E}\|_1$, s.t. $\mathbf{D} = \mathbf{DZ} + \mathbf{E}$, $\text{diag}(\mathbf{Z}) = 0$; 这里, $\|\mathbf{Z}\|_1 = \sum_{i,j=1}^n \|\mathbf{Z}_{ij}\|$ 。其中参数的选择均为文献所给的最优参数 ($\lambda_1 = 0.81$, $\lambda_2 = 4.24$), 实验结果如图 1。

由图 1 可以看出, 本文提出的 TL_1 方法在合成数据集中处理噪音方面比 LRR, SR_1 , $\text{SR}_{2,1}$ 方法获得了更好的聚类效果。具体地, SSC 方法分别采用 L_2 范数与 $L_{2,1}$ 范数来约束相似度矩阵 \mathbf{C} , 虽然能获得系数矩阵 \mathbf{C} 的块对角结构, 但是 SSC 是单独地一一列对数据进行处理, 受噪音影响较大。而 LRR 方法使用低秩来约束相似度矩阵 \mathbf{C} , 能够很好地在

全局上除去噪音的影响。但是这两种算法仍然用含有噪音的数据作为字典，这显然是不合理的。本文提出的 TL_1 算法，首先，采用非凸约束，能够通过调节参数 a ，使系数矩阵 C 各块对角元素之间连接得更加紧密。其次，在处理噪音方面，本文提出约束条件(5)可以视为先利用鲁棒性PCA对噪音进行处理，最后再使用干净的数据点进行自表达。从图中可知，随着噪音的增大，本文算法仍表现出较好的性能。

5.1.2 Extended Yale B 数据库 为了进一步验证本文提出的 TL_1 算法的有效性，将本文的算法应用到Extend Yale B 数据库^[26]。该数据集包含38类灰度大小为 192×168 的人脸，每类包含64张在不同光照下的人脸图像。与文献[8]相同，我们把38个目标分为4组。1到4组分别有类别1~10, 10~20, 21~30和31~38，对于前3组中的每一组，我们的实验采用所有可以选择的2,3,5,8,10个类别，对于第4组，我们使用可选的2,3,5,8个类别进行重复实验。为比较不同算法的聚类误差，我们列出了局部子空间亲和(Local Subspace Affinity, LSA)^[27]，低

秩子空间聚类算法(LRSC)^[28,29]，最小二乘回归(LSR), LRR, SSC方法的结果。为了减小算法的计算时间，首先将图像灰度转化为 48×42 并将图片转化为2016维列向量。4种方法运用的参数均为各文献所述的最优参数。表2分别给出了不同算法的聚类误差。从表中可以看出，在2, 3, 5类的情况下，由于类别较少，即子空间的个数较少，SSC能够得到较高的准确率。本文提出的 TL_1 方法与SSC方法的聚类错误率相差无几，均要优于其他各方法。但是当在8类、10类的情况下，本文提出的算法在聚类错误率方面均要优于所列出的4种方法。图2给出了不同算法在取不同类Extend Yale B数据库中的平均计算时间，结果清楚地显示出本文 TL_1 算法的有效性。综上所述，本文提出的算法在聚类错误率和平均计算时间方面均要优于其它各类算法，也佐证了本文之前的分析。

5.1.3 Hopkins155 数据集 本文在Hopkins155运动分割数据集上进行对比实验。运动分割是指从视频序列中截取一系列2D的点的运动轨迹，这些运动轨迹是由不同的刚性运动所形成的。Hopkins155数

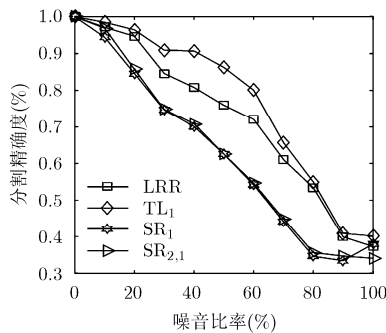


图1 TL_1 LRR SR_1 $SR_{2.1}$ 4种方法的分割精确率

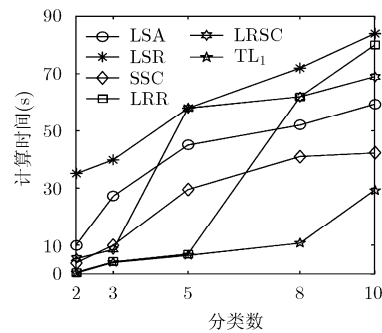


图2 不同算法在取不同类Extend Yale B数据库中的平均计算时间

表2 在Extended Yale B上的聚类错误率(%)

| 算法 | | LSA | SSC | LSR | LRR | LRSC | TL_1 |
|-----|-----|-------|-------|-------|-------|-------|--------|
| 2类 | 均值 | 32.80 | 1.86 | 3.86 | 2.54 | 5.32 | 2.01 |
| | 中位数 | 47.60 | 0 | 0.97 | 0.78 | 4.96 | 0.57 |
| 3类 | 均值 | 52.20 | 3.20 | 10.52 | 4.21 | 8.47 | 3.24 |
| | 中位数 | 50.00 | 1.04 | 9.32 | 2.60 | 7.86 | 1.52 |
| 5类 | 均值 | 58.02 | 4.31 | 29.49 | 6.90 | 12.24 | 4.52 |
| | 中位数 | 56.87 | 2.50 | 27.05 | 5.63 | 11.25 | 2.87 |
| 8类 | 均值 | 59.19 | 5.85 | 41.01 | 14.3 | 23.70 | 5.62 |
| | 中位数 | 58.59 | 4.49 | 40.35 | 14.2 | 28.03 | 4.19 |
| 10类 | 均值 | 60.42 | 10.94 | 42.25 | 22.92 | 30.36 | 7.14 |
| | 中位数 | 57.50 | 5.63 | 41.02 | 23.59 | 28.75 | 5.48 |

数据集包含 156 个不同序列，每一个序列都是从运动的物体中被提取出来的，每个序列是 2 个或 3 个类别。图 3 展示了一些 Hopkins155 数据集中带特征点的样本图片。

根据文献[27]，本文将 156 个数据集中的子空间聚类错误率的最大值，均值，中值，以及标准差作为评价标准，分别对 LSR, LRR, SSC 方法以及本文 TL_1 方法在该数据集上进行对比，实验结果如表 3 所示。

通过表 3 可以看出，相比于 LSR, LRR, SSC 方法，本文在 Hopkins155 数据集中的错误率均值方面比上述 3 种方法分别提高了 2.7%, 3.6%, 1.5%。由于该数据集大部分数据是比较干净的，只有小部分数据点受到污染，这样 SSC 方法充分利用样本间的相关性，取得了不错的结果。本文所提出的 TL_1 方法与 SSC 方法的实验错误率较为接近。究其原因是因为 Hopkins155 数据集中类别较少(2 或 3)，即

表 3 LSR, LRR, SSC 以及本文 TL_1 方法在 Hopkins155 数据集上的聚类错误率(%)

| 错误率 | LSR | LRR | SSC | TL_1 |
|-----|-------|-------|-------|--------|
| 最大值 | 0.397 | 0.476 | 0.414 | 0.402 |
| 平均值 | 0.042 | 0.051 | 0.030 | 0.018 |
| 中位数 | 0.005 | 0.005 | 0 | 0 |
| 标准差 | 0.086 | 0.100 | 0.076 | 0.048 |



图 3 Hopkins155 中的带特征点的运动物体的特征轨迹

子空间个数比较少，稀疏性在一定程度上会增加数据的局部性。本文在保证邻接矩阵块对角的基础上，利用矩阵的 TL_1 范数对邻接矩阵进行约束，增强了邻接矩阵的稀疏性，提升了数据的鉴别能力，从而提升了聚类准确率。因此，可以认为本文所提出的基于 TL_1 范数的子空间聚类算法是一种高效的子空间聚类方法。

5.1.4 USPS 手写字体数据库 USPS 数据库是一个有代表性的手写字体数据库，被广泛地应用于聚类实验。其中包含 9298 张 0~9 的数字图像。在本节实验中，本文考虑该数据库下不同类别的聚类误差，对于 $k = 2, 3, 4, \dots, 9$ ，我们随机选取 k 类样本，每一类随机选取 100 张图像重复进行实验，计算其误差率，并取 10 次误差率的均值作为最终的结果。

实验结果如表 4 所示。从表中可以看出，本文算法的性能均优于其他算法，平均聚类错误比 LRR 方法低了 7.53%。由于数据库中手写字体部分被随机遮挡(噪音)，在比较算法中各类算法稳定性不一。LRSC 方法没有对求得的系数矩阵做进一步处理，它的聚类错误较高。LSA 没有明确地处理受干扰的数据，在复杂噪音的数据库上表现出了较差的性能。可以看出 LRR 和 LSR 方法准确率下降明显，SSC 尝试探索数据的局部结构。但是在复杂噪音的数据库中，性能一般。而本文方法获得了不错的效果，说明基于 TL_1 方法对噪音的鲁棒性更强，聚类性能较为稳定。

表 4 USPS 数据库中的错误率(%)

| 类别数 | LRR | LSA | LSR | SSC | LRSC | TL_1 |
|------|-------|-------|-------|-------|-------|--------|
| 2 聚类 | 3.50 | 10.70 | 24.50 | 0.50 | 9.20 | 0.50 |
| 3 聚类 | 4.59 | 24.10 | 18.34 | 7.44 | 10.71 | 3.20 |
| 4 聚类 | 28.70 | 37.62 | 33.75 | 20.30 | 30.50 | 13.40 |
| 5 聚类 | 21.40 | 13.72 | 27.60 | 18.42 | 30.20 | 16.70 |
| 6 聚类 | 34.60 | 38.90 | 35.73 | 28.61 | 32.26 | 19.23 |
| 7 聚类 | 12.40 | 27.83 | 40.83 | 30.48 | 37.60 | 21.60 |
| 8 聚类 | 26.85 | 28.42 | 42.10 | 27.00 | 41.20 | 17.00 |
| 9 聚类 | 25.73 | 31.45 | 44.82 | 32.54 | 42.70 | 20.93 |
| 均值 | 19.72 | 26.59 | 33.46 | 20.67 | 28.17 | 14.07 |

6 结束语

本文提出了基于 TL_1 范数约束的子空间聚类模型, 尽管该优化模型是非凸的, 但它的最优解为具有块对角结构的表示系数矩阵, 这使得样本间相似的数据点达到紧连接的目的。针对含有噪音的情形, 所提出的约束条件可以看作是运用列稀疏约束的鲁棒主成分分析来有效地去除噪音, 这样, 文中优化模型的约束条件等价于以干净数据为字典的优化模型, 可以有效提高对噪音的鲁棒性。进一步运用增广拉格朗日-交替方向乘子方法对模型进行求解, 由求解出的系数矩阵得到相似度矩阵, 再采用经典的谱聚类方法进行聚类。实验表明, 本文提出的基于 TL_1 范数约束的子空间聚类算法在聚类准确率、系数矩阵的稀疏性、对噪音的鲁棒性等方面有较好的性能。尽管如此, 由于该优化模型是非凸的, 因此对所提出的基于 TL_1 范数模型的算法的收敛性等理论还需要做进一步的研究。

参 考 文 献

- [1] 张涛, 唐振民, 吕建勇. 一种基于低秩表示的子空间聚类改进算法[J]. 电子与信息学报, 2016, 38(11): 2811-2818. doi: 10.11999/JEIT160009.
ZHANG Tao, TANG Zhenmin, and LÜ Jianyong. Improved algorithm based on low rank representation for subspace clustering[J]. *Journal of Electronics & Information Technology*, 2016, 38(11): 2811-2818. doi: 10.11999/JEIT160009.
- [2] 王卫卫, 李小平, 冯象初, 等. 稀疏子空间聚类综述[J]. 自动化学报, 2015, 41(8): 1373-1384. doi: 10.16383/j.aas.2015.c140891.
WANG Weiwei, LI Xiaoping, FENG Xiangchu, et al. A survey on sparse subspace clustering[J]. *Acta Automatica Sinica*, 2015, 41(8): 1373-1384. doi: 10.16383/j.aas.2015.c140891.
- [3] YANG A, WRIGHT J, MA Y, et al. Unsupervised segmentation of natural images via lossy data compression[J]. *Computer Vision and Image Understanding*, 2008, 110(2): 212-225. doi: 10.1016/j.cviu.2007.07.005.
- [4] WRIGHT J, MAIRAL J, MA Y, et al. Sparse representation for computer vision and pattern recognition[J]. *Proceedings of the IEEE*, 2010, 98(6): 1031-1044. doi: 10.1109/JPROC.2010.2044470.
- [5] LI C G, YOU C, and VIDAL R. Structured sparse subspace clustering: A joint affinity learning and subspace clustering framework[J]. *IEEE Transactions on Image Processing*, 2017, 26(6): 2988-3001. doi: 10.1109/TIP.2017.2691557.
- [6] VIDAL R. Subspace clustering[J]. *IEEE Signal Processing Magazine*, 2011, 28(2): 52-68. doi: 10.1109/MSP.2010.939739.
- [7] ELHAMIFAR E and VIDAL R. Sparse subspace clustering [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009: 2790-2797. doi: 10.1109/CVPR.2009.5206547.
- [8] ELHAMIFAR E and VIDAL R. Sparse subspace clustering: Algorithm, theory, and applications[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(11): 2765-2781. doi: 10.1109/TPAMI.2013.57.
- [9] LIU G C, LIN Z C, YAN S C, et al. Robust recovery of subspace structures by low-rank representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1): 171-184. doi: 10.1109/TPAMI.2012.88.
- [10] LIU G C, LIN Z C, and YU Y. Robust subspace segmentation by low-rank representation[C]. *Proceedings of the International Conference on Machine Learning*, Haifa, Israel, 2010: 663-670.
- [11] NG A Y, JORDAN M, and WEISS Y. On spectral clustering: Analysis and an algorithm[C]. *Proceedings of the Advances in Neural Information Processing Systems*, Vancouver, Canada, 2001: 849-856.
- [12] PAN J, MATHIEU S, and HONG L. Efficient dense subspace clustering[C]. *IEEE Winter Conference on Applications of Computer Vision*, USA, 2014: 461-468. doi: 10.1109/WACV.2014.6836065.
- [13] ZHUANG L S, MA Y, LIN Z C, et al. Non-negative low-rank and sparse graph for semi-supervised learning[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, Rhode Island, 2012: 2328-2335. doi: 10.1109/CVPR.2012.6247944.
- [14] PATEL V M, NGUYEN H V, and VIDAL R. Latent space sparse and low-rank subspace clustering[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2015, 9(4): 691-701. doi: 10.1109/JSTSP.2015.2402643.
- [15] 李波, 卢春园, 冷成财, 等. 基于局部图拉普拉斯约束的鲁棒低秩表示聚类方法[J]. 自动化学报, 2015, 41(11): 1971-1980. doi: 10.16383/j.aas.2015.c150031.
LI Bo, LU Chunyuan, LENG Chengcai, et al. Robust low rank subspace clustering based on local graph laplace constraint[J]. *Acta Automatica Sinica*, 2015, 41(11): 1971-1980. doi: 10.16383/j.aas.2015.c150031.
- [16] NIKOLOVA M. Local strong homogeneity of a regularized estimator[J]. *SIAM Journal on Applied Mathematics*, 2000, 61(2): 633-658. doi: 10.1137/S0036139997327794.
- [17] LÜ J, and FAN Y. A unified approach to model selection and sparse recovery using regularized least squares[J]. *The Annals of Statistics*, 2009, 37(6A): 3498-3528. doi: 10.1214/09-AOS683.
- [18] ZHANG S, YIN P H, and JACK X. Transformed schatten

- iterative thresholding algorithms for matrix rank minimization and applications[J]. *Arxiv Preprint*, 2015, 1506.04444. <https://www.researchgate.net/publication/278413840>.
- [19] ZHANG S and JACK X. Minimization of transformed L1 penalty: Theory, difference of convex function algorithm, and robust application in compressed sensing[J]. *Arxiv Preprint*, 2016, 1411.5735v3. <https://arxiv.org/abs/1411.5735>.
- [20] HORN R and JOHNSON C. *Topics in Matrix Analysis*[M]. Cambridge University Press, 1991: 144–163.
- [21] LIN Z C, CHEN M, and MA Y. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices[R]. UIUC Technical Report UILU-ENG-09-2215, 2009.
- [22] 吴杰祺, 李晓宇, 袁晓彤, 等. 利用坐标下降实行并行稀疏子空间聚类[J]. *计算机应用*, 2016, 36(2): 372–376. doi: 10.11772/j.issn.1001-9081.2016.02.0372.
WU Jieqi, LI Xiaoyu, YUAN Xiaotong, *et al.* Parallel sparse subspace clustering via coordinate descent minimization[J]. *Journal of Computer Applications*, 2016, 36(2): 372–376. doi: 10.11772/j.issn.1001-9081.2016.02.0372.
- [23] VIDAL R and FAVARO P. A closed form solution to robust subspace estimation and clustering[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, USA, 2011, 1801–1807. doi: 10.11091/CVPR.2011.5995365.
- [24] BERTSEKAS D. *Constrained Optimization and Lagrange Multiplier Methods*[M]. Belmont, MA, USA: Athena Scientific, 1996: 326–340.
- [25] CANDÈS E J, LI X D, MA Y, *et al.* Robust principal component analysis[J]. *Journal of the ACM*, 2010, 58(3): 11. doi: 10.1145/1970392.1970395.
- [26] GEORGHIADES A, BELHUMEUR P, and KRIEGMAN D. From few to many: Illumination cone models for face recognition under variable lighting and pose[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 23(6): 643–660. doi: 10.1109/34.927464.
- [27] YAN J Y and POLLEYFEYS M. A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate[C]. *Proceedings of the European Conference on Computer Vision*, Graz, Austria, 2006: 94–106. doi: 10.1007/11744085_8.
- [28] FENG J S, LIN Z C, XU H, *et al.* Robust subspace segmentation with block-diagonal prior[C]. *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, Columbus, USA, 2014: 3818–3825. doi: 10.1109/CVPR.2014.482.
- [29] VIDAL R and FAVARO P. Low-rank subspace clustering (LRSC)[J]. *Pattern Recognition Letters*, 2014, 43: 47–61. doi: 10.1016/j.patrec.2013.08.006.
- 李海洋: 男, 1975年生, 教授, 硕士生导师, 主要研究方向为机器学习、稀疏信息处理、量子逻辑及格上拓扑学。
- 王恒远: 男, 1991年生, 硕士生, 研究方向为机器学习、稀疏信息处理。