

指导教师： 杨涛

提交时间： 2015.3.29

The task of
Digital Image Processing

数字图像处理

School of Computer Science

No: 1

姓名： 杨翰林

学号： 2012302465

班号： 10011204



具有鲁棒性的分块对角先验信息的子空间分割

Jiashi Feng¹, Zhouchen Lin², Huan Xu³, Shuicheng Yan¹

¹Department of ECE, National University of Singapore, Singapore

²Key Lab. of Machine Perception, School of EECS, Peking University, China

³Department of ME, National University of Singapore, Singapore

¹{a0066331,eleyans}@nus.edu.sg, ²zlin@pku.edu.cn, ³mpexuh@nus.edu.sg

摘要

本文通过有效构建了一个完全的分块对角关联矩阵样例涉及了子空间分割问题。精确的样本聚类非常需要分块对角结构来解决，但获得这样的结构是困难的。当前技术水平下的子空间分割方法（例如 SSC_[4] 和 LRR_[12]）需要非一般性结构的先验信息（例如稀疏或低秩）来构造关联矩阵。在本文的工作中，我们直接实行基于拉普拉斯约束方程构建分块对角结构，并且开发了一个有效的随机次梯度算法来对结果进行最优化。另外，本文给出了两种新的子空间分割方法：分块对角的 SSC 和 LRR。据我们所知，这是第一个有关直接构建分块对角结构的研究。经过在面部识别、动作识别和图构建的半监督学习实验，可以证实我们用于子空间分割问题的新方法具有明显的优越性。

1. 介绍

高维度的图像数据，比如人脸面部图像和刚体运动轨迹等，通常可以用由多个低维的子空间组成的并集描述_[8,21,25]。要找到这样的低维结构，我们经常要将数据聚类为多组，同时为每组对应一个子空间。

由此引出对子空间分割问题的重要定义：

定义 1（子空间分割_[22]）给定一个由 k 个子空间组成的并集 $\{S_i\}_{i=1}^k$ 产生的样本向量 $X = [X_1, \dots, X_k] = [x_1, \dots, x_n] \in R^{d \times n}$ ，令 X_i 为 n_i 个由子空间 S_i 产生的样本集， $n = \sum_{i=1}^k n_i$ 。子空间分割的作用是将样本根据其底层的子空间进行划分。



图 1. 一个用作说明的比较：用不同的先验信息在综合噪声环境中构造 4 子空间关联矩阵：

(a) 稀疏 (b) 低秩 (c) 分块对角。分块对角关联矩阵对样本的聚类 and 子空间分割更为准确。

最近，基于谱聚类的子空间分割方法被广泛提出^[4,12,13]。这些方法使用围绕每个样本的全局或局部的信息来构建关联矩阵。之后，样本的分割即可使用谱聚类^[15]构建出的关联矩阵得到。

特别地，稀疏子空间聚类 (SSC)^[4] 和低秩表达 (LRR)^[12] 两种当前技术水平下的子空间分割方法，是通过寻找各个样本关于全部样本集的稀疏或低秩的线性表示构建关联矩阵的。所获得的表达系数被直接用于构建关联矩阵。在限定的条件下，这些方法可以生成一个分块对角关联矩阵。然而，当信噪比低时，得到的分块对角矩阵很脆弱，会出现不同的分块之间过于接近或是分块不独立的情况。因此，子空间分割的效果被严重地影响^[19]。

在这项的工作中，我们提出明确寻求分块对角结构的关联矩阵用于子空间分割。我们在图的拉普拉斯变换中规定一个给定的秩约束，这样等价于约束了关联矩阵中的连通分支的数目。于是，一个分块对角关联矩阵便可有效的得出，即使是在不利的条件下，比如低信噪比或是不同子空间的定位不恰当的情况等。图 1 给出了一个用作说明的例子。它显示了与稀疏和低秩的先验信息相比，我们的方法获得了准确度更高的关联矩阵。要高效地解决所产生的最优化问题，我们使用了随机次梯度算法与投影操作来确保得到的关联矩阵在循环内是分块对角的。

我们用当前的子空间分割技术，即 SSC 与 LRR，作为与我们提出方法的对比。我们将明确地展示该方法如何能够生成准确的分块对角关联矩阵，并改善它在子空间分割中的表现。值得注意的是，我们的方法并不局限于这两种情况。实际上，这种方法具有普适性，可被直接应用于构造其他的关联矩阵。多方面的子空间分割实验，包括综合数据聚类，人脸图像聚类和运动轨迹分割等，都可证明该方法的有效性。我们也进行了数字和人脸识别的半监督学习实验来进一步证实该方法相比于其他通用的图构建方法更具优越性。

2. 相关工作

在过去的十年内，许多子空间分割或归类方法被开发出来^[17,4,12,23,13,5]。由于在该工作中我们主要关注基于谱聚类的方法，以下我们将详细回顾这方面的相关工作。对于其他的方法，一个不错的总结可以在这里得到^[22]。

稀疏子空间聚类 (SSC)^[4]表达的是将集合中的每个样本看做是所有其他样本的线性组合，并且组合系数需要是稀疏的。之后，Liu 等人^[12]提出低秩表达 (LRR) 用于子空间分割。LRR 构建的是低秩关联矩阵，这样捕捉到的隐含子空间的全局先验信息仍是低维的。这种先验信息下的 LRR 对严重损坏的数据样本具有很强的鲁棒性。Wang 等人^[23]介绍了另一种诱导分块对角的先验信息，但是，他们也证明了只有在子空间两两正交的情况下，样本的关联矩阵才是完全分块对角的。这个条件过于苛刻，并不适用于真实数据。最近，Lu 等人^[13]提出了一个基于最小二乘回归的方法，用于构建关联矩阵。它声称，基于该方法的对同一子空间的分类效果能够改善子空间分割的结果。

虽然现有的方法已经取得了相当的成功，但对于子空间分割问题，并没有人能够为真实样本产生一个完全的分块对角矩阵。就现有的结果来看，这些方法对隐含子空间的噪声、数据错误和定位不当等干扰非常敏感。

3. 问题公式化

在这项工作中，我们专注于将谱聚类应用于基于全局信息的子空间分割关联矩阵的构建。对于关联矩阵的构建，每一个样本 $x_i \in R^d$ 都可视为标准样本 Xz_i 的线性组合的近似。这里 $X = [x_1, \dots, x_n] \in R^{d \times n}$ 是标准样本的集合。系数向量 z_i 中元素的绝对值 $|z_i(j)|$ 可直接用作 x_i 与第 j 个标准样本的相关值。指定 $S(z_i)$ 为 z_i 的相关样本的索引集： $S(z_i) = \{j \mid z_i(j) \neq 0, \forall j = 1, \dots, n\}$ 。要将样本准确的分类到相对应的子空间，就需要当第 i 个和第 j 个样本来自不同子空间时，有 $S(z_i) \cap S(z_k) \equiv \emptyset$ 。也就是说，当两个样本来自于不同的子空间时，它们的相关值应为零。因此，对于由元素 $Z = [z_1, \dots, z_n] \in R^{n \times n}$ 并取绝对值 $|z|$ 所构成的

对称关联矩阵 $W = (|z| + |z^T|)/2$ ，经过适当的行或列变换后，我们获得了如下的分块对角矩阵：

$$\tilde{W} = \text{blockdiag}(W_1, W_2, \dots, W_k).$$

在分块对角矩阵中的每一块 W_i 相当于一个子空间 S_i 。这样样本的分割可以快速的获得。尽管现有的几种子空间分割方法也能够获得这样的分块对角结构，但是，就如之前提到过的，这些方法的条件要求非常苛刻，并不适用于处理真实数据。比如，当数据信噪比较小时，样本会被噪音干扰，导致分割结果偏离其所对应的子空间。某一个样本可能与另一个不同子空间的样本非常接近，而它们的关联值是非零的。这种错误的关联关系可能会影响之后的谱聚类过程，并降低子空间分割的效果。作为对比，如果分块对角的先验信息是给定的，这样的错误就可以被构建绝对分块对角矩阵的限定所纠正。这样谱聚类的子空间分割的效果便得以提高。

3.1 分块对角化的拉普拉斯约束

在介绍应用于分块对角矩阵的拉普拉斯约束之前，我们首先来回顾一下拉普拉斯矩阵的定义以及拉普拉斯矩阵谱不变量与关联矩阵的结构之间的关系：

定义 2（拉普拉斯矩阵）对一个 n 个加权 $W(j, j')$ 样本的关联矩阵 $W \in R^{n \times n}$ ，

拉普拉斯矩阵 $L_w \in R^{n \times n}$ 定义如下：

$$L_w(j, j') = -W(j, j'), \text{ if } j \neq j'; \sum_{\ell \neq j} W(j, \ell) \text{ otherwise.}$$

下面著名的定理^[14]叙述了拉普拉斯矩阵的秩与其对应的关联矩阵中块的数量间的关系：

定理 1^[14] 令 W 为一关联矩阵。这样，其对应的拉普拉斯矩阵 L_w 特征值为 0 的重数 k 与 W 中连通分支的数量（块数）相等。

基于以上定理，我们可以将一个一般的方阵经由以下的拉普拉斯约束(LC)转变为一个 k 个分块的分块对角矩阵： $\text{rank}(L_w) = n - k$ 。之后，我们可以定义一个 k 个分块的分块对角矩阵集合：

$$\kappa = \{Z \mid \text{rank}(L_w) = n - k, W = \frac{1}{2}(|Z| + |Z^T|)\}$$

在上述约束集合中，参数 k 是子空间分割问题中需要进行聚类的个数（参见定义 1）。在构建了 k 个分块的分块对角关联矩阵之后，样本就被很快地分成 k 个聚类。 k 值常由用户来指定。

3.2 对于分块对角 SSC 的拉普拉斯约束

稀疏子空间聚类(SSC)通过构建稀疏的稀疏矩阵^[4]来解决子空间分割问题。它的目标函数为：

$$\min_Z \|Z\|_1 + \frac{\lambda}{2} \|X - XZ\|_F^2, \text{ s.t. } \text{diag}(Z) = 0.$$

在获得系数矩阵 Z 之后，可以构建相似矩阵： $W = (\|Z\|_1 + \|Z^T\|_1) / 2$ 。在确定的条件下，比如无噪声的数据，SSC 只会选择那些与目标样本的线性表示在同一子空间中的样本。然而，对于真实的数据，这样的条件很难满足。为了获取分块对角的关联矩阵，我们利用对 SSC 的拉普拉斯约束，提出如下的分块对角 SSC 目标函数：

$$\begin{aligned} \min_Z f_1(Z) &= \|Z\|_1 + \frac{\lambda}{2} \|X - XZ\|_F^2, \\ \text{s.t. } \text{diag}(Z) &= 0, Z \in \kappa. \end{aligned} \quad (1)$$

如同之前的解释， $Z \in \kappa$ 的条件使获得的矩阵 Z 为分块对角的关联矩阵。

3.3 对于分块对角 LRR 的拉普拉斯约束

低秩表达 (LRR) 对一组给定的样本，可由样本本身与其基元组成的相关性寻找它的一个低秩表达的系数矩阵^[12]。根据相关的理论，低秩正规化利用了下层低秩的子空间的集合仍为低秩的这一全局先验信息。通常地，LRR 解决了如下的最优化问题：

$$\min_Z \|Z\|_* + \|X - XZ\|_F^2.$$

在数据量充足、无噪声和子空间独立的条件下，所获得的相似矩阵会是严格分块对角的。然而，对于真实数据，比如人脸面部图像或其他物品的图像，这样的要求过于严格。这样，所获得的矩阵 W 常不是分块对角的，这样会影响子空间分割的结果。要使获得的矩阵 W 是一个 k 个分块的分块对角矩阵，我们使用上述的对 LRR 的拉普拉斯约束，提出如下的分块对角低秩表达：

$$\min_{Z,E} f_*(Z) = \|Z\|_* + \frac{\lambda}{2} \|X - XZ\|_F^2, s.t. Z \in \kappa. \quad (2)$$

式(1)与式(2)各自表明了 SSC 和 LRR 的最优化问题是不同的，并且需要各自独立的方法来解决。在这项工作中，我们的目标便是开发一个普适的最优化方法来将这两个问题在一个统一的框架之内解决。这个最优化方法的细节将在以下的章节给出。

4. 拉普拉斯约束的最优化方法

我们利用了一个有效的随机次梯度算法 (SSGD) 来解决式(1)与式(2)中的最优化问题，他们都包含了一个高度非凸的 k 分块对角矩阵集 (k -DBMS) 约束。在每次循环内，关联矩阵 Z 向负的次梯度方向移动以减少结构化带来的误差。之后 Z 立刻由 $\prod_{\kappa}(\cdot)$ 投影回约束集合 κ ，以确保它的分块对角结构。

次梯度的计算方法如下：对式(1)中的目标函数 $f_1(Z)$ ，其次梯度为

$$\zeta(f_1(Z)) = \lambda X^T X(Z - I) + \partial \|Z\|_1 / \partial Z. \text{ 这里}$$

$$\partial \|Z\|_1 / \partial Z_{ij} = \text{sign}(Z_{ij}) \text{ if } Z_{ij} \neq 0 \text{ and } \theta \in [-1, 1] \text{ otherwise.}$$

由于关联矩阵

Z 的对角元素并不必须更新，我们手动的将 $\zeta(f_1(Z))$ 的对角元素置零。对式(2)

的目标函数 $f_*(Z)$ ，其次梯度为 $\zeta(f_*(Z)) = \lambda X^T X(Z - I) + \partial \|Z\|_* / \partial Z$ 。令

$$U\Sigma V^T \text{ 为 } Z \text{ 的致密奇异值分解 (SVD), 这样 } \partial \|Z\|_* / \partial Z = UV^T + H^2_{[24]},$$

H 为零矩阵的集合。算法 1 中给出了 SSGD 算法的细节，而 $\prod_{\kappa}(\cdot)$ 的射影的计算将在算法 2 中给出。

算法 1 基本 SSGD

输入：数据矩阵 X ，目标函数 f ，分块数目 k ，最大循环次数 T ，交换参数 λ 。

1. 初始化 $Z^{(0)} = 0, t = 0, p = \text{rank}(X)$.
2. 设置步长 $\eta = 1.5\sqrt{np} / (1.5\lambda n \|XX^T\|_2 + \sqrt{p})\sqrt{T}$
3. 当 $t \leq T$ 时:

4. 从 $\{\sqrt{ne_1}, \dots, \sqrt{ne_n}\}$ 中随机选取 p 列，组成一个 $n \times p$ 的探索矩阵 Y
5. 计算次梯度: $g^{(t)} \leftarrow \zeta(f(Z^{(t)}))YY^T$
6. 预测: 通过算法 2 计算 $Z^{(t+1)} \leftarrow \Pi_{\kappa}(Z^{(t)} - \eta g^{(t)})$
7. $t \leftarrow t + 1$
8. 结束

输出: 系数矩阵 Z 。

4.1 次梯度无偏估计

令 SSGD 收敛和有效的一个最重要的因素是目标函数的次梯度为无偏估计量。如果 $E(\tilde{g}^{(t)}) = g^{(t)}$ ，则 $\tilde{g}^{(t)}$ 被称为次梯度 $g^{(t)}$ 的无偏估计量。在这项工作中，我们使用如下的随机次梯度方法来获得次梯度的无偏估计，同时减少次梯度的计算量_[7]。特别地，我们用一个探索矩阵来试出次梯度的估计值。探索矩阵的定义如下：

定义 3 (探索矩阵_[7]) 一个随机的 $n \times p$ 矩阵 Y ， $p < n$ 。当 $E[YY^T] = I_{n \times n}$ ， $I_{n \times n}$ 为 $n \times n$ 的单位矩阵，且期望值基于 Y 中值的选择时，称 Y 为探索矩阵。

在这项工作中，我们随机地从度量标准基 $\{\sqrt{ne_1}, \dots, \sqrt{ne_n}\}$ 中选出 p 个向量来组成探索矩阵: $Y = [\sqrt{ne_{(1)}}, \dots, \sqrt{ne_{(p)}}] / \sqrt{p}$ 。容易证得 $E[YY^T] = I$ 。这样在每步次梯度下降中次梯度可以如下计算: $g^{(t)} = \zeta(f(Z^{(t)}))YY^T$ ，其中 $\zeta(\cdot)$ 为之前定义过的次梯度。

要看出这项技术如何减轻了计算量，我们用 $BD-LRR$ 作为例子 (同样适用于 $BD-SSC$): 在算法 1 的第 5 步，梯度为 $\zeta(f(Z))YY^T$ ，计算 $\zeta(f(Z))$ 包括了计算 $XX^T Z$ 加上对 Z 的奇异值分解。对于 $XX^T Z$ ，在乘 YY^T 之后，它变为了 $XX^T (ZY)Y^T$ 。 ZY 实际上是从 Z 中的 n 列抽取的 p 个样本。这样 Z 的大小便从 n^2 变为 np ，计算量由此得到了降低。关于 Y^T 的乘法，由于左边矩阵的零值而变得非常快速。对 Z 的奇异值分解，我们使用了_[7] 中引理 2.4 的方法。

这样计算量从 $O(n^2r)$ 变为 $O(npr)$, $r = \text{rank}(Z)$ 。该技术将次梯度计算的运算复杂度从 $O(n^3)$ 变为 $O(pn^2)$ 。如果 $p \ll n$, 改善的幅度会很明显。

4.2 k -BDMS 投影的解决

在一步次梯度下降之后, 变量矩阵 Z 可能会偏离约束集合, 并不再保持分块对角结构。因此, 我们要将 Z 投影进 k -BDMS 中。对一个矩阵 Z_0 的投影算子定义如下:

$$\Pi_{\kappa}(Z_0) = \arg \min_{Z \in \kappa} \|Z - Z_0\|_F^2 \quad (3)$$

这个投影运算找到了一个与 Z_0 的欧氏距离最短的集合 κ 中的矩阵。与之相关的最优化问题可以写作:

$$\min_Z \frac{1}{2} \|Z - Z_0\|_F^2, s.t. Z \in \kappa.$$

上述的最优化问题被两个强加的限制严格约束, 且都包含在变量 Z 的结构变换之中。然而, 在经过对问题的检查之后, 我们发现约束条件是解耦的, 这意味着我们不需要同时处理两个条件, 这样问题的简化便很有意义。因此, 我们提出了辅助变量 \tilde{Z} 来代替原先的拉普拉斯矩阵 L_W 。这样目标函数便等价:

$$\begin{aligned} \min_{Z, \tilde{Z}} \frac{1}{2} \|Z - Z_0\|_F^2, \\ s.t. \text{rank}(\tilde{Z}) = n - k, W = \frac{1}{2}(|Z| + |Z^T|), \tilde{Z} = L_W. \end{aligned}$$

我们进一步重述了约束条件 $\tilde{Z} = L_W$ 作为扩展的拉格朗日乘子_[11]的代价:

$$\begin{aligned} \min_{Z, \tilde{Z}} \frac{1}{2} \|Z - Z_0\|_F^2 + \langle J, \tilde{Z} - L_W \rangle + \frac{\beta}{2} \|\tilde{Z} - L_W\|_F^2, \\ s.t. \text{rank}(\tilde{Z}) = n - k, W = \frac{1}{2}(|Z| + |Z^T|). \end{aligned}$$

这里 J 为拉格朗日乘子, β 为 $\tilde{Z} = L_W$ 约束条件下的增量重量参数。由于两个约束条件解耦, 我们可以交替地对 Z 和 \tilde{Z} 进行最优化。特别地, 每个最优化问题只包含一个约束条件。

算法 2 k-DBMS 的投影

输入：目标矩阵 Z_0 ，分块个数 k

输出：一个分块对角矩阵 Z

1. 初始化： $Z^{(0)} = Z_0; t = 0; \rho = 1.1; \beta^{(0)} = 1 \times 10^{-4}$
2. 当 非收敛 时：
 3. 用(4)式更新 $Z^{(t+1)}$
 4. 用(5)式更新 $\tilde{Z}^{(t+1)}$
 5. 更新 $Y^{(t+1)} = Y^{(t)} + \beta^{(t)}(\tilde{Z} - L_W)$
 6. 更新 $\beta^{(t+1)} = \rho\beta^{(t)}$
 7. $t \leftarrow t + 1$
8. 结束

迭代计算变量 \tilde{Z} 和乘数 J 以及排除不包含 Z 的条件，我们用如下的方法：

$$\min_Z \frac{1}{2} \|Z - Z_0\|_F^2 - \langle J, L_W \rangle + \frac{\beta}{2} \|\tilde{Z} - L_W\|_F^2, \quad (4)$$

$$s.t. W = \frac{1}{2} (|Z| + |Z^T|).$$

注意到除去 $\|Z - Z_0\|_F^2$ 的条件，所有其他在上述问题中的情况只包括 Z 的绝对值。因此 Z 的解中的元素与对应 Z_0 中的元素有相同的符号。否则，我们就能够经常变换 Z 中元素的符号，在不改变其他值的情况下减少 $\|Z - Z_0\|_F^2$ 的值。因此，基于这种现象，解就能写成 $Z = \hat{Z} \otimes \text{sign}(Z_0)$ ， \otimes 代表叉积，而 Z 便是下述问题的解：

$$\min_{\hat{Z}} \frac{1}{2} \|\hat{Z} - |Z_0|\|_F^2 - \langle J, L_W \rangle + \frac{\beta}{2} \|\tilde{Z} - L_W\|_F^2,$$

$$s.t. W = \frac{1}{2} (\hat{Z} + \hat{Z}^T), \hat{Z} \geq 0.$$

上述目标函数可以被任何现有的二次规划求解程序高效地解决。在解出 \hat{Z} 之后, $Z = \hat{Z} \otimes \text{sign}(Z_0)$ 也就被解出。

现在我们转向求解 \tilde{Z} 。类似地, 设置了变量 Z 和乘子 J 之后, 我们通过计算下式来更新 \tilde{Z} 的值:

$$\min_{\tilde{Z}} \langle J, \tilde{Z} \rangle + \frac{\beta}{2} \|\tilde{Z} - L_W\|_F^2, \text{ s.t. } \text{rank}(\tilde{Z}) = n - k. \quad (5)$$

上式也等价于:

$$\min_{\tilde{Z}} \|\tilde{Z} - (L_W - 1/\beta J)\|_F^2, \text{ s.t. } \text{rank}(\tilde{Z}) = n - k.$$

由 Eckart-Young 定理^[3] 可得, 该问题允许一个闭合式的解。更特殊的是, 这个闭合式的解可由对 $L_W - \frac{1}{\beta} J = U\Sigma V^T$ 进行奇异值分解, 并选择首 $n - k$ 个奇异向量获得: $\tilde{Z} = U_{1:(n-k)} [\Sigma]_{1:(n-k)} V_{1:(n-k)}^T$ 。最优化相关的算法细节在算法 2 中表述。在运行中, 当 $\beta^{(t)} \max(\|\tilde{Z}^{(t+1)} - \tilde{Z}^{(t)}\|, \|Z^{(t+1)} - Z^{(t)}\|) < 1 \times 10^{-4}$ 或 $\|\tilde{Z}^{(t)} - L_{W^{(t)}}\| < 1 \times 10^{-6}$ 时, 最优化终止。

4.3 收敛的相关事项

式(1)与式(2)的最优化问题, 由于 $k - \text{BDMS}$ 约束的存在, 是强烈不收敛的。幸运的是, 我们可以证明算法 1 的解收敛于全局最优值。这里我们简单地从理论上进行解释, 而相关的实验验证将在后续章节给出。算法 1 的收敛与否取决于^[1] 中的结果, 特别地, 对于非凸约束的仿射问题, 如果满足可扩展的有限等距性 (SRIP), 带有投影的梯度下降将会收敛于全局最优值。就式(3)而言, 尽管它是非凸的, 我们同样可以获得最优值。在实际操作中, 当 $Z_0 = Z^{(t)} - \eta g^{(t)}$ 时, 我们只需令 $Z^{(t+1)}$ 为最优值即可。我们控制 η 的值很小直到结束。这样, $Z^{(t)}$ 便与 Z_0 在 κ 中的投影距离非常近。我们从 $Z^{(t)}$ 开始进行 ALM 循环, ALM 将收敛于其 KKT 点 $Z^{(t+1)}$ ^[11], 由于 $Z^{(t)}$ 与 Z_0 距离相近, 得到的将是最优值。

5. 相关实验

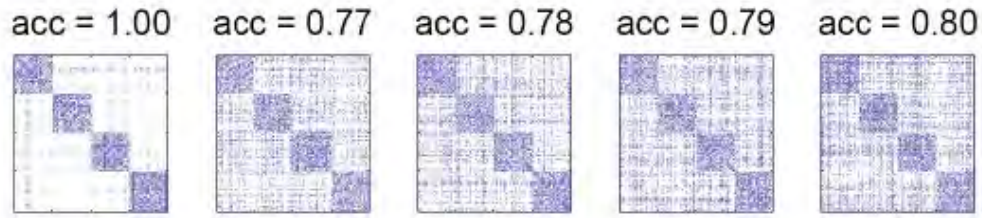
5.1 合成数据

数据生成 我们在不同的噪声条件下合成了 5 组数据样本。样本从^[12]设计的方案中生成。我们构造了 $k = 4$ 的三维独立子空间，周围的维度为 30，*i.e.*, $\{S_i\}_{i=1}^4 \subset R^{30}$ 。每个子空间 S_i 包含了 50 个样本。有 30% 的样本是由零均值、方差 $\sigma \|x\|_2$ 的高斯噪声生成并掺杂入数据当中。这里 $\|x\|_2$ 表示相对应样本的 ℓ_2 规范， σ 可看作信噪比。我们分别用信噪比为 $\{0, 0.1, 0.2, 0.3, 0.4\}$ 的样本数据测试 SSC, LRR, BD-SSC 和 BD-LRR 的性能。 $\sigma = 0$ 代表无噪样本。对称关联矩阵由 $W = (|Z| + |Z^T|)/2$ 获得，矩阵 Z 即为上述方法的输出。

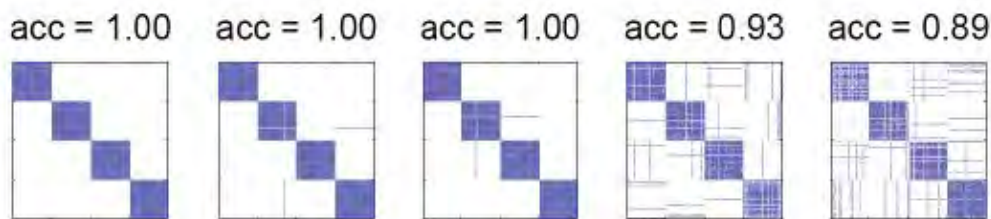
初始化 对 BD-SSC 和 BD-LRR 而言，交换参数 λ 置为 10，最大循环次数 T 为 600。算法 1 中的变量 Z 初始值为全零矩阵。对于奇异值分解，我们实行在式 (5) 中 $L_w - Y/\beta$ 上的部分奇异值分解，直到 k 的最小奇异值 $U_k \Sigma_k V_k$ ，令 $\tilde{Z} = L_w - Y/\beta - U_k \Sigma_k V_k$ 。因为常有 $k \ll n - k$ ，该方法极大地降低了部分奇异值分解的计算量。BD-SSC 在四核 CPU，2.83GHz 和 8GB 内存的 PC 机上，使用 matlab 平台对 2000 个样本的关联矩阵的构建花费大约 20 秒，而 BD-LRR 由于计算了奇异值分解，耗费了大约 70 秒的时间。作为比较，SSC 与 LRR 各自花费了 5.5 秒和 2.3 秒。对大数据处理而言，计算量是该方法的一个瓶颈。我们将进一步研究在大数据量下降低该方法的运算量的方法。由于有过在 LRR 方法中成功的例子^[20]，一个有希望的方向是应用分治法解决最优化问题。我们通过实验发现，最优化问题的收敛是高度非凸的。特别地，在补充材料中我们画出了 BD-SSC 与 BD-LRR 的客观价值收敛曲线。

结果 图 2 与图 3 展示了所获得的、行列的序列由样本标签的情况做过改变的关联矩阵。与之对应的分割准确率也给予标记。我们可以看出，对于 SSC 与 LRR 方法，当信噪比大于 0.1 时，有些样本便被选入了其他的子空间，矩阵的分块对角结构便被破坏，分割准确率也随之降低。对于无噪的情况，所有的方法都达到了完全准确的分割。然而，信噪比为 0.2 时，SSC 与 LRR 的分割准确率分别降到了 0.79 和 0.83。而 BD-SSC 与 BD-LRR 方法则能够保持矩阵的分块对角结构，并有着非常高的分割准确率。这个结果表明，与现有的 SSC 和 LRR 方法

相比, BD-SSC 与 BD-LRR 在分割中有着更好地表现, 对噪声也有更好的鲁棒性。注意到 BD-SSC 与 BD-LRR 生成的关联矩阵中对样本也有错误的分割, 但经过适当的行列变换后, 仍能得到严格的分块对角矩阵。



(a) SSC 的结果

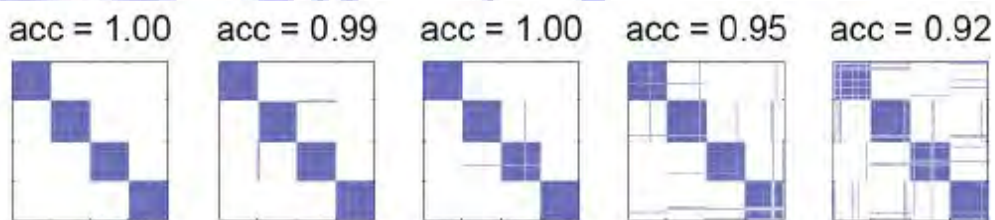


(b) BD-SSC 的结果

图 2 SSC 与 BD-SSC 所得到的关联矩阵。从左向右依次是信噪比为 0、0.1、0.2、0.3 和 0.4 的情况, 子空间块数为 4, 每块包括 50 个样本数据。分割的准确率在每个矩阵图示上方标记。



(a) LRR 的结果



(b) BD-LRR 的结果

图 3 LRR 与 BD-LRR 所得到的关联矩阵。从左向右依次是信噪比为 0、0.1、0.2、0.3 和 0.4 的情况, 子空间块数为 4, 每块包括 50 个样本数据。分割的准确率在每个矩阵图示上方标记。

5.2 面部聚类

我们使用扩展的耶鲁大学数据库 $B_{[6]}$ ，来评价上述方法和现有的方法的性能。数据集包括 38 种不同的人脸图像，每种图像包括 64 张不同光照条件下的人脸正面照片。在这项实验中，我们选取了前 $c=\{2,3,5,8,10\}$ 种图像进行聚类。为减少计算量和内存开销，我们将每张图像的大小从 192×168 像素压缩至 48×42 像素。由此形成的 2016 维的向量，被使用标准主成分分析而进一步投影至 $9 \times c$ 维的子空间内。这里每种图像的 9 个维度是从面部图像矩阵的奇异值曲线中观测得到的。对于 SSC 和 LRR，我们使用了作者的默认参数设置。也就是说，SSC 的 $\lambda = 20$ ，LRR 的 $\lambda = 0.18$ 。对于 BD-SSC 和 BD-LRR，我们各自设置其交换参数与 SSC 和 LRR 的 λ 相同。循环次数 T 设为 1000，参数 β 为学习速率，被设置为 1。因此，BD-SSC 的学习速率为 $\eta = 7.6 \times 10^{-4}$ ，BD-LRR 的学习速率为 $\eta = 3.4 \times 10^{-6}$ 。

注意到对于 SSC 的子空间分割，作者给出了一个额外的启发式阈值运算作为对得到的关联矩阵的后处理以得到更好的结果_[4]。对于 LRR，也有一个启发式搜索过程来刻画获得的关联矩阵_[12]。在该实验中，我们并未采用这样额外的后处理过程以确保性能评测的公正性。表格 1 列出了每种方法聚类结果的错误比率。根据表格我们可以看出，BD-SSC 与 BD-LRR 均比其对应的现有方法的识别率有 3 到 6 个百分点的提升。这能够说明分块对角结构能够提高识别的准确率，所构建的关联矩阵也有较强的鲁棒性。

# Classes	SSC	BD-SSC	LRR	BD-LRR
2	9.37	3.90	8.59	3.91
3	20.13	17.70	13.54	10.02
5	30.00	27.50	15.00	12.97
8	36.33	33.20	31.64	27.70
10	43.59	39.53	35.16	30.84

表格 1 在扩展的耶鲁大学数据库 B 中获得的聚类结果错误率。

5.3 运动分割

运动分割是指将多种物体的运动轨迹在时空集合中聚类，从中寻找出轨迹与物体的对应关系_[21]。一个物体形成的运动轨迹的集合构成了一个低维子空间。

因此运动分割问题可以通过对运动轨迹的空间坐标进行子空间分割来解决。我们

使用基准霍普金斯 155 数据集^[21]进行试验。数据集包含 156 个两或三个物体的运动视频，并被分为了三类：棋盘，交通信息，和铰接序列。运动轨迹被一个追踪器自动的标出，并且异常值也被手动去除。

表格 2 显示了四种方法的分割的错误率统计，其中包括了最大误差、平均误差、中值误差和标准误差的值。对所有的方法，都没有做任何的数据的预处理或后处理。通过结果来看，BD-SSC 和 BD-LRR 相对于 SSC 和 LRR 分别减少了 0.21% 和 1.48% 的错误率。值得注意的是，通过 SSC 和 LRR 得到的错误率已经相当的低，所以获得这样的改进结果是有意义的。如同^{[4][12]}中所提到的，对 SSC 和 LRR 加以启发式搜索的后处理也能够降低错误率。在表格 3 中，我们给出了对 SSC 和 BD-SSC 加以阈值后处理、对 LRR 和 BD-LRR 加以比例缩放后处理之后的结果。从结果能够看出，分别做这样的处理使错误率进一步下降了 0.31% 和 0.77%。就我们所知，我们的 BD-LRR 方法取得了对该数据集中 156 个数据的最佳错误率 0.97%^[4]。Pham 等人^[16]在该数据集上曾取得了错误率 0.16% 的结果，然而他们的方法依赖于额外的空间信息^[16]。相比之下，我们的方法并不需额外信息。

	SSC	BD-SSC	LRR	BD-LRR
Max	42.34	42.34	41.18	38.97
Mean	3.11	2.90	4.83	3.35
Med	0	0	0.52	0.37
Std	7.78	7.48	9.35	8.84

表格 2 霍普金斯 155 数据集的错误率统计。对 156 个数据分析的最大误差(Max)、平均误差(Mean)、中位误差(Med)和标准误差(Std)在表中依次给出。

	SSC	BD-SSC	LRR	BD-LRR
Max	47.20	43.07	43.38	14.93
Mean	1.99	1.68	1.74	0.97
Med	0	0	0	0
Std	6.97	5.97	5.51	2.46

表格 3 加以后处理的霍普金斯 155 数据集错误率统计。对 156 个数据分析的最大误差(Max)、平均误差(Mean)、中位误差(Med)和标准误差(Std)在表中依次给出。

5.4 半监督学习中的应用

我们所提出的 BD-SSC 和 BD-LRR 方法，通过对带标签或无标签的训练数据构建关联图，可直接应用于半监督学习任务中。注意到基于图的半监督学习方法^[26]通常包括两步：图构建和对图的标签扩展。我们的工作主要集中在第一步。我们主要在图构建时与其他方法作比较，而采用相同的标签扩展方法。监督学习的方法并不适用于无标签的训练数据，因此在此我们没有给出它们的结果。

在该实验中，我们使用了两个数据集。一个是 USPS 手写数字数据集^[9]，其中包括了 10 种数字(0 到 9)共 11000 个样本。根据^[2]中的实验设置，我们在每种数字样本里随机选取了 200 个，共 2000 个样本用于实验。另一个数据集是之前已经介绍过的扩展的耶鲁大学数据集 B。这里用到了所有的 38 种人脸图像样本。在构建关联矩阵之后，我们使用了随机游走算法^[10]来将有标签的数据扩展至无标签的数据。在每个数据集中，构建了四种流行的图、矩阵作为基准，包括应用 SSC^[2]构建 LLE 图^[18]、kNN 图、 ℓ_1 图，以及 LRR 图^[12]。对于 LLE 图和 kNN 图的构建，考虑到稀疏图通常会获得更好的效果^[2]，邻接的大小被设置为总数据量的 1%。对于 ℓ_1 图的构建，正规化参数 λ 分别被设为 5×10^{-3} (耶鲁大学数据集 B)和 3×10^{-2} (USPS 数据集)。对于 LRR 图， λ 都被设为 0.2。结果分别在表格 4 和表格 5 中给出。

# Labeled	LLE	kNN	ℓ_1	BD-SSC	LRR	BD-LRR
10	35.8	21.4	16.1	12.4	13.4	11.2
20	24.8	17.0	11.4	8.9	11.8	9.4
30	18.0	15.2	11.5	10.1	11.2	10.7
40	15.7	11.5	8.2	6.4	10.9	6.9

表格 4 不同的半监督学习图构建下的 USPS 数据集的识别错误率(%)。使用了不同的每种数据下带标签的数据量(Labeled)。

# Labeled	LLE	kNN	ℓ_1	BD-SSC	LRR	BD-LRR
5	28.1	76.5	26.8	17.1	28.2	15.1
10	15.0	66.3	14.7	7.6	24.6	8.2
20	11.0	60.2	12.6	4.2	19.4	5.1
30	10.5	59.4	10.8	2.6	16.1	4.7

表格 5 不同的半监督学习图构建下的耶鲁-B 数据集的人脸识别错误率(%)。使用了不同的每种数据下带标签的数据量(Labeled)。

从结果中可以看出，因为得益于稀疏的条件， ℓ_1 图较LLE图和kNN图在USPS数据集上的表现更佳。在耶鲁-B数据集中，由于LLE图利用了人脸图像的相关性， ℓ_1 图和LLE图的表现都很好。在两个数据集中，BD-SSC构建的 ℓ_1 图大约减少了3%到8%的错误率。BD-LRR相对于LRR的错误率减少的更为明显。这样的改进主要是由于加上了对样本关联矩阵的分块对角约束，使得样本聚类的结果更为准确了。由于人脸图像相比于手写数字的噪声干扰更为强烈，BD-SSC和BD-LRR在耶鲁-B数据集上获得的性能改进较USPS数据集的改进更为明显。这也证实了基于分块对角先验信息的图构建对于噪声有较强的鲁棒性。

6. 结论

在这项工作中，我们提出了基于拉普拉斯图约束方程的绝对分块对角关联矩阵的构建。两种新的子空间分割方法，BD-SSC与BD-LRR，基于上述构想被设计出来。我们做了大量的实验，对合成数据、真实的人脸数据和运动轨迹进行了分割。我们所设计的方法相比于现有方法有了明显的性能提升。此外，BD-SSC和BD-LRR也能够应用于半监督学习中分块对角的 ℓ_1 图和LRR图的构建，在人脸识别和手写数字识别中也取得了不错的表现。

鸣谢 新加坡国家研究基金会国际研究中心通过IDM项目办公室给予了J.Feng和S.Yan资助和支持。中国国家自然科学基金会给予Z.Lin以支持。新加坡教育部AcRF Tier 2项目给予了H.Xu部分支持。

参考文献

- [1] A. Beck and M. Teboulle. A linearly convergent algorithm for solving a class of nonconvex/affine feasibility problems. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer, 2011. 5
- [2] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. Huang. Learning with ℓ_1 -graph for image analysis. *TIP*, 2010. 7
- [3] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1936. 5
- [4] E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *TPAMI*, 2013. 1, 2, 3, 6, 7
- [5] Y. Fang, R. Wang, and B. Dai. Graph-oriented learning via automatic group sparsity for data analysis. In *ICDM*, 2012. 2

- [6] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. TPAMI, 2001. 6
- [7] A. Haim, K. Satyen, K. Shiva, and S. Vikas. Efficient and practical stochastic subgradient descent for nuclear norm regularization. In ICML, 2012. 4
- [8] J. Ho, M. Yang, J. Lim, K. Lee, and D. Kriegman. Clustering appearances of objects under varying illumination conditions. In CVPR, 2003. 1
- [9] J. Hull. A database for handwritten text recognition research. TPAMI, 1994. 7
- [10] M. S. T. Jaakkola and M. Szummer. Partially labeled classification with markov random walks. NIPS, 2002. 7
- [11] Z. Lin, M. Chen, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. Arxiv preprint arXiv:1009.5055, 2010. 4, 5
- [12] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. TPAMI, 2013. 1, 2, 3, 5, 7
- [13] C. Lu, H. Min, Z. Zhao, L. Zhu, D. Huang, and S. Yan. Robust and efficient subspace segmentation via least squares regression. In ECCV, 2012. 1, 2
- [14] U. Luxburg. A tutorial on spectral clustering. Statistics and Computing, 2007.3
- [15] A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. NIPS, 2002. 1
- [16] D. Pham, S. Budhaditya, D. Phung, and S. Venkatesh. Improved subspace clustering via exploitation of spatial constraints. In CVPR, 2012. 7
- [17] S. Rao, R. Tron, R. Vidal, and Y. Ma. Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories. TPAMI, 2010. 2
- [18] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. Science, 2000. 7
- [19] M. Soltanolkotabi and E. Candes. A geometric analysis of subspace clustering with outliers. The Annals of Statistics, 2012. 1
- [20] A. Talwalkar, L. Mackey, Y. Mu, S.-F. Chang, and M. I. Jordan. Distributed low-rank subspace segmentation. In ICCV. IEEE, 2013. 6
- [21] R. Tron and R. Vidal. A benchmark for the comparison of 3-d motion Segmentation algorithms. In CVPR, 2007. 1, 7
- [22] R. Vidal. Subspace clustering. IEEE Signal Processing Magazine, 2010. 1, 2
- [23] S. Wang, X. Yuan, T. Yao, S. Yan, and J. Shen. Efficient subspace segmentation via quadratic programming. AAI, 2011. 2
- [24] G. Watson. Characterization of the subdifferential of some matrix norms. Linear

Algebra and its Applications, 1992. 4

[25] A. Yang, J. Wright, Y. Ma, and S. Sastry. Unsupervised segmentation of natural images via lossy data compression. CVIU, 2008. 1

[26] X. Zhu. Semi-supervised learning literature survey. Computer Science, University of Wisconsin-Madison, 2006. 7

