

# 若干低秩子空间恢复模型的闭解及其应用

林宙辰

北京大学, 北京 100871

## 1 引言

我们正处于大数据时代。其所面对的数据除了常提及的 4 个 V (Volume (海量的数据规模)、Velocity (快速的数据更新)、Variety (多样的数据类型)、Value (巨大的数据价值)) [1] 特性外, 往往还伴随高维性, 即数据表达成了维数极高的向量, 例如采集一张千万像素的照片即是在千万维空间中采集了一个样本点, 视频、网页、基因阵列、用户行为日志等等也都是高维数据。数据的高维性给高效鲁棒的数据处理提出了巨大挑战。

幸运的是, 高维数据并非杂乱无章。图 1 给出了某个人的人脸图像分布的 2 维可视化。从中可见, 人脸图像其实并未充满整个空间, 而是集中在若干低维流形附近。当低维流形为最简单的子空间时, 找出样本点所在不同维数子空间的问题就称为子空间聚类问题[2], 它是子空间恢复问题的一种。

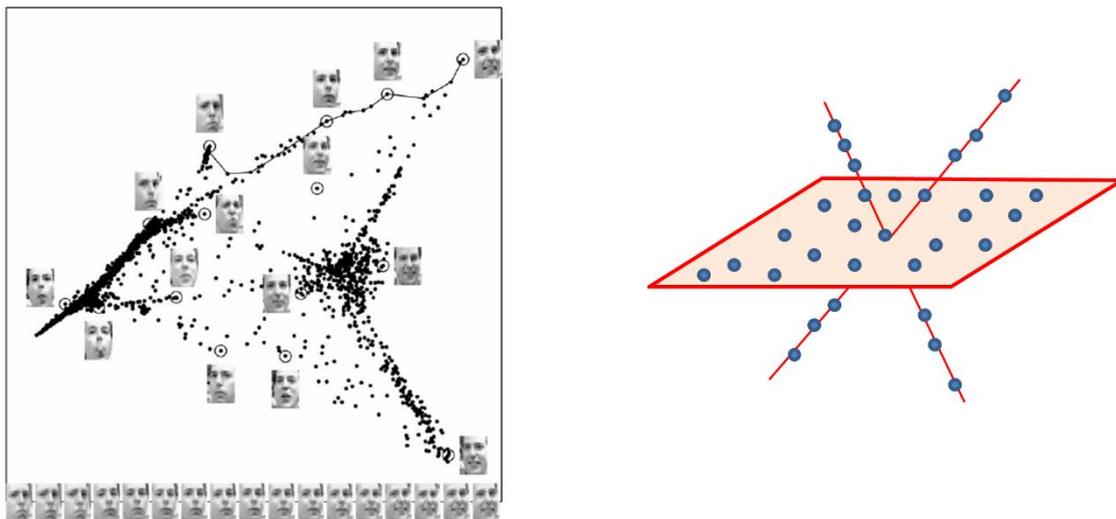


图 1. 人脸图像分布的 2 维可视化 (左, 取自[3]) 及子空间聚类问题 (右)。

由于子空间的维数等于子空间上样本点所构成的数据矩阵的秩, 因此子空间恢复问题往往可建模成低秩问题, 即求解秩尽可能小的某矩阵的问题。目前有代表性的低秩模型有[4]: 矩阵填充 (Matrix Completion, MC) 模型[5]:

$$\min_{\mathbf{A}} \text{rank}(\mathbf{A}), \quad s.t. \quad \mathbf{d} = \pi_{\Omega}(\mathbf{A}); \quad (1)$$

鲁棒主成分分析 (Robust Principal Component Analysis, RPCA) 模型[6]:

$$\min_{\mathbf{A}, \mathbf{E}} \text{rank}(\mathbf{A}) + \lambda \|\mathbf{E}\|_0, \quad s.t. \quad \mathbf{D} = \mathbf{A} + \mathbf{E}; \quad (2)$$

低秩表示 (Low-Rank Representation, LRR) 模型[7]:

$$\min_{\mathbf{Z}, \mathbf{E}} \text{rank}(\mathbf{Z}) + \lambda \|\mathbf{E}\|_{2,0}, \quad s.t. \quad \mathbf{D} = \mathbf{DZ} + \mathbf{E}; \quad (3)$$

潜在 (或隐性) 低秩表示 (Latent Low-Rank Representation, LatLRR) 模型[8]:

$$\min_{\mathbf{Z}, \mathbf{L}, \mathbf{E}} \text{rank}(\mathbf{Z}) + \text{rank}(\mathbf{L}) + \lambda \|\mathbf{E}\|_0, \quad s.t. \quad \mathbf{D} = \mathbf{DZ} + \mathbf{LD} + \mathbf{E}, \quad (4)$$

其中  $\mathbf{d}$  和  $\mathbf{D}$  为已知的数据向量或数据矩阵（每个样本为  $\mathbf{D}$  的一个列向量）， $\pi_{\Omega}(\mathbf{A})$  表示提取  $\mathbf{A}$  中下标在  $\Omega$  中的元素的算子， $\|\mathbf{E}\|_0$  为  $\mathbf{E}$  的 0-范数<sup>1</sup>（即  $\mathbf{E}$  中非零元的个数）， $\|\mathbf{E}\|_{2,0}$  为  $\mathbf{E}$  的(2,0)-范数（即  $\mathbf{E}$  中非零列的个数）。MC 模型是通过已知矩阵的某些位置的值推断该矩阵的缺失值。RPCA 模型是从已知样本点中提取一个低维子空间，使得样本点偏离该子空间的误差稀疏。LRR 模型是从已知样本点中提取多个低维子空间，使得样本点偏离所属子空间的误差稀疏。LatLRR 模型则是为了解决 LRR 模型在样本不足时面对的问题而提出（其推导过程见第 2 节）。

由于模型（1）-（4）都是离散优化问题，一般情况下是 NP-难问题，因此常见的处理方案是把它们近似成凸优化问题。具体说来，是把 rank 函数换成核范数（Nuclear Norm），即矩阵奇异值之和；把 0-范数换成 1-范数，即矩阵元素绝对值之和；把(2,0)-范数换成(2,1)-范数，即各列 2-范数（元素平方和开根号）之和。之所以这样替换，是因为可以粗略地说核范数、1-范数、(2,1)-范数分别是离 rank 函数、0-范数和(2,0)-范数“最近”的凸函数<sup>2</sup>[9]。由此得到如下相应的松弛后的凸优化问题：

松弛的矩阵填充（Relaxed Matrix Completion）模型[5]：

$$\min_{\mathbf{A}} \|\mathbf{A}\|_*, \quad s.t. \quad \mathbf{d} = \pi_{\Omega}(\mathbf{A}); \quad (5)$$

松弛的鲁棒主成分分析（Relaxed Robust Principal Component Analysis）模型<sup>3</sup>[6]：

$$\min_{\mathbf{A}, \mathbf{E}} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1, \quad s.t. \quad \mathbf{D} = \mathbf{A} + \mathbf{E}; \quad (6)$$

松弛的低秩表示（Relaxed Low-Rank Representation）模型[7]：

$$\min_{\mathbf{Z}, \mathbf{E}} \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1}, \quad s.t. \quad \mathbf{D} = \mathbf{DZ} + \mathbf{E}; \quad (7)$$

松弛的潜在低秩表示（Relaxed Latent Low-Rank Representation）模型[8]：

$$\min_{\mathbf{Z}, \mathbf{L}, \mathbf{E}} \|\mathbf{Z}\|_* + \|\mathbf{L}\|_* + \lambda \|\mathbf{E}\|_1, \quad s.t. \quad \mathbf{D} = \mathbf{DZ} + \mathbf{LD} + \mathbf{E}. \quad (8)$$

目前低秩模型已经获得了广泛的应用，代表性的有背景建模[6]、图像对齐[11]、图像校正[12]、运动分割[7]、图像分割[13]、显著性检测[14]、图像标签优化[15]等，具体介绍可见本人写的综述文章[4]。

## 2 无噪低秩表示模型的闭解

我们首先发现松弛的无噪低秩模型：

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_*, \quad s.t. \quad \mathbf{D} = \mathbf{DZ} \quad (9)$$

具有唯一的闭解[16]。

定理 1. 松弛的无噪低秩模型（9）具有唯一的闭解：假设  $\mathbf{D}$  的秩为  $r$ ， $\mathbf{D} = \mathbf{U}_r \Sigma_r \mathbf{V}_r^T$  是  $\mathbf{D}$  的

<sup>1</sup>严格地说，0-范数和下文的(2,0)-范数并不是数学意义上的范数，这里称它们为范数只是为了方便。

<sup>2</sup>严格地说，是在某凸集上的凸包络（Convex Envelope）[10]。

<sup>3</sup>该模型正式名称为主成分追踪（Principal Component Pursuit）模型[6]，但这里为了统一指称的方便，称其为松弛的鲁棒主成分分析模型。

瘦型奇异值分解，则 (9) 的唯一解为  $\mathbf{Z}^* = \mathbf{V}_r \mathbf{V}_r^T$ 。

定理 1 有点让人意外之处在于核范数并不是严格凸<sup>4</sup>的，因此理论上在线性约束下可能会有无穷多个解，但是特定的线性约束  $\mathbf{D} = \mathbf{DZ}$  刚好排除了这种可能性。 $\mathbf{V}_r \mathbf{V}_r^T$  在计算机视觉三维重建理论中称为形状交互矩阵 (Shape Interaction Matrix) [16]。把  $\mathbf{Z}^* = \mathbf{V}_r \mathbf{V}_r^T$  代入 (9) 的目标函数，得到

$$\min_{\{\mathbf{Z} | \mathbf{D} = \mathbf{DZ}\}} \|\mathbf{Z}\|_* = \text{rank}(\mathbf{D}), \quad (10)$$

即 (9) 的最小值为  $\mathbf{D}$  的秩。更深入的分析可以得到：当样本从独立的线性子空间里采样得到时， $\mathbf{Z}^*$  为块对角矩阵，而且每一对角块对应于一个线性子空间、每一对角块的秩等于相应线性子空间的维数[7]。

定理 1 可以有推广形式[7]：

定理 2. 松弛的广义无噪低秩模型

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_*, \quad s.t. \quad \mathbf{D} = \mathbf{BZ} \quad (11)$$

具有唯一的闭解：

$$\mathbf{Z}^* = \mathbf{B}^+ \mathbf{D}, \quad (12)$$

其中  $\mathbf{B}^+$  为  $\mathbf{B}$  的 Moore-Penrose 伪逆。

Yu 和 Schuurmans 进一步把定理 2 推广为[17]：

定理 3. 对任意酉不变范数  $\|\cdot\|_{\text{UI}}$  (即对任意矩阵  $\mathbf{X}$  和尺寸匹配的列单位正交矩阵  $\mathbf{U}$  和  $\mathbf{V}$  都有

$\|\mathbf{UXV}^T\|_{\text{UI}} = \|\mathbf{X}\|_{\text{UI}}$ ), 如果约束  $\{\mathbf{Z} | \mathbf{D} = \mathbf{BZ}\}$  不为空集，则问题

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_{\text{UI}}, \quad s.t. \quad \mathbf{D} = \mathbf{BZ} \quad (13)$$

都具有唯一的闭解： $\mathbf{Z}^* = \mathbf{B}^+ \mathbf{D}$ 。

对于原始的广义无噪低秩模型：

$$\min_{\mathbf{Z}} \text{rank}(\mathbf{Z}), \quad s.t. \quad \mathbf{D} = \mathbf{BZ}, \quad (14)$$

我们也找到了它的闭解，但和问题 (9) 不同，其解不唯一[18]：

定理 4. 如果约束  $\{\mathbf{Z} | \mathbf{D} = \mathbf{BZ}\}$  不为空集， $\mathbf{D} = \mathbf{U}_D \mathbf{\Sigma}_D \mathbf{V}_D^T$  和  $\mathbf{B} = \mathbf{U}_B \mathbf{\Sigma}_B \mathbf{V}_B^T$  分别是  $\mathbf{D}$  和  $\mathbf{B}$  的瘦型奇异值分解，则原始的广义无噪低秩模型 (12) 的所有解为：

$$\mathbf{Z}^* = \mathbf{B}^+ \mathbf{D} + \mathbf{S} \mathbf{V}_D^T, \quad (15)$$

<sup>4</sup>严格凸的直观涵义为函数的图像和任一超平面相切时只有一个切点。

其中  $\mathbf{S}$  是任一满足  $\mathbf{V}_B^T \mathbf{S} = \mathbf{0}$  的矩阵。

定理 4 也稍有点出乎意料，因为离散优化问题经常是 NP-难的。(15) 和 (12) 的差别在于多了一个尾项  $\mathbf{S}\mathbf{V}_D^T$ 。定理 2 曾被用于 Structure from Motion 矩阵分解中[19] (该文是 CVPR2012 最佳论文)。

## 2.1 应用：潜在低秩表示模型的推导

基于定理 2，可以给出潜在低秩表示模型的推导过程。低秩表示模型在样本不足时，已有样本  $\mathbf{D}_O$  可能不足以自我表达，Liu 等人[8]于是引入未观测的隐藏样本  $\mathbf{D}_H$ ，使得它和  $\mathbf{D}_O$  一起可以表达  $\mathbf{D}_O$ ：

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_*, \quad s.t. \quad \mathbf{D}_O = [\mathbf{D}_O \ \mathbf{D}_H] \mathbf{Z}. \quad (16)$$

但是由于  $\mathbf{D}_H$  未知，相应的低秩表示模型不能求解，所以要想办法消去  $\mathbf{D}_H$ 。有如下引理[8]：

引理 1. 设  $[\mathbf{D}_O \ \mathbf{D}_H]$  的瘦型奇异值分解为  $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ ，并把  $\mathbf{V}$  分块为  $\mathbf{V} = \begin{bmatrix} \mathbf{V}_O \\ \mathbf{V}_H \end{bmatrix}$ 。则 (16) 的解

可表达为：

$$\mathbf{Z}^* = \mathbf{V}\mathbf{V}_O^T. \quad (17)$$

证明：由  $[\mathbf{D}_O \ \mathbf{D}_H] = \mathbf{U}\mathbf{\Sigma} \begin{bmatrix} \mathbf{V}_O \\ \mathbf{V}_H \end{bmatrix}^T$  可得：

$$\mathbf{D}_O = \mathbf{U}\mathbf{\Sigma}\mathbf{V}_O^T, \quad \mathbf{D}_H = \mathbf{U}\mathbf{\Sigma}\mathbf{V}_H^T. \quad (18)$$

于是约束  $\mathbf{D}_O = [\mathbf{D}_O \ \mathbf{D}_H] \mathbf{Z}$  就退化为：

$$\mathbf{V}_O^T = \mathbf{V}^T \mathbf{Z}.$$

根据定理 2，其核范数最小的解为：

$$\mathbf{Z}^* = \mathbf{V}\mathbf{V}_O^T = \begin{bmatrix} \mathbf{V}_O \mathbf{V}_O^T \\ \mathbf{V}_H \mathbf{V}_O^T \end{bmatrix}. \quad (19)$$

证毕。

由 (18) - (19)，约束  $\mathbf{D}_O = [\mathbf{D}_O \ \mathbf{D}_H] \mathbf{Z}$  可作如下改写：

$$\begin{aligned} \mathbf{D}_O &= [\mathbf{D}_O \ \mathbf{D}_H] \mathbf{Z}^* \\ &= \mathbf{D}_O \mathbf{V}_O \mathbf{V}_O^T + \mathbf{D}_H \mathbf{V}_H \mathbf{V}_O^T \\ &= \mathbf{D}_O (\mathbf{V}_O \mathbf{V}_O^T) + (\mathbf{D}_H \mathbf{V}_H \mathbf{\Sigma}^{-1} \mathbf{U}^T) \mathbf{D}_O. \end{aligned}$$

容易看到括号里的矩阵都是低秩的。因此我们可以试图求解两个低秩矩阵  $\mathbf{Z}$  和  $\mathbf{L}$  (其中  $\mathbf{D}_o$  改写为  $\mathbf{D}$ ):

$$\min_{\mathbf{Z}, \mathbf{L}} \text{rank}(\mathbf{Z}) + \text{rank}(\mathbf{L}), \quad s.t. \quad \mathbf{D} = \mathbf{DZ} + \mathbf{LD}. \quad (20)$$

其凸松弛形式为:

$$\min_{\mathbf{Z}, \mathbf{L}} \|\mathbf{Z}\|_* + \|\mathbf{L}\|_*, \quad s.t. \quad \mathbf{D} = \mathbf{DZ} + \mathbf{LD}. \quad (21)$$

有噪声时, 就是模型 (4) 和 (8)。

### 3 无噪潜在低秩模型的闭解

我们发现无噪潜在低秩表示模型 (20) 也有闭解, 但是解不唯一[18]:

定理 5. 原始无噪潜在低秩表示模型 (20) 的全部解为:

$$\mathbf{Z}^* = \mathbf{V}_D \tilde{\mathbf{W}} \mathbf{V}_D^T + \mathbf{S}_1 \tilde{\mathbf{W}} \mathbf{V}_D^T, \quad \mathbf{L}^* = \mathbf{U}_D \Sigma_D (\mathbf{I} - \tilde{\mathbf{W}}) \Sigma_D^{-1} \mathbf{U}_D^T + \mathbf{U}_D \Sigma_D (\mathbf{I} - \tilde{\mathbf{W}}) \mathbf{S}_2, \quad (22)$$

其中  $\tilde{\mathbf{W}}$  是任意幂等矩阵 (即  $\tilde{\mathbf{W}}^2 = \tilde{\mathbf{W}}$ ),  $\mathbf{S}_1$  和  $\mathbf{S}_2$  是满足下列条件的任意矩阵:

1.  $\mathbf{V}_D^T \mathbf{S}_1 = \mathbf{0}$ ,  $\mathbf{S}_2 \mathbf{U}_D = \mathbf{0}$ ;
2.  $\text{rank}(\mathbf{S}_1) \leq \text{rank}(\tilde{\mathbf{W}})$ ,  $\text{rank}(\mathbf{S}_2) \leq \text{rank}(\mathbf{I} - \tilde{\mathbf{W}})$ .

得到定理 5 的原始想法是通过观察 (20) 的形式, 得知  $\mathbf{Z}$  与  $\mathbf{L}$  的地位是对称的, 因此猜测 (20) 可以拆成两个子问题:

$$\begin{aligned} \min_{\mathbf{Z}} \text{rank}(\mathbf{Z}), \quad s.t. \quad \frac{1}{2} \mathbf{D} &= \mathbf{DZ}, \\ \min_{\mathbf{L}} \text{rank}(\mathbf{L}), \quad s.t. \quad \frac{1}{2} \mathbf{D} &= \mathbf{LD}. \end{aligned}$$

它们的闭解由定理 4 给出。进一步发现两个系数  $\frac{1}{2}$  可以分别改为  $\alpha$  和  $1 - \alpha$ , 其中  $\alpha$  为任意

实数, 由此再深入挖掘, 才得到定理 5。

类似地, 松弛的无噪潜在低秩表示模型 (21) 也有闭解, 但是解也不唯一[18]:

定理 6. 松弛的无噪潜在低秩表示模型 (21) 的全部解为:

$$\mathbf{Z}^* = \mathbf{V}_D \hat{\mathbf{W}} \mathbf{V}_D^T, \quad \mathbf{L}^* = \mathbf{U}_D (\mathbf{I} - \hat{\mathbf{W}}) \mathbf{U}_D^T, \quad (23)$$

其中  $\hat{\mathbf{W}}$  是任意满足下列条件的块对角矩阵:

1.  $\hat{\mathbf{W}}$  的对角块与  $\Sigma_D$  相容, 即如果  $[\Sigma_D]_{ii} \neq [\Sigma_D]_{jj}$  则  $[\hat{\mathbf{W}}]_{ij} = 0$  ( $[\mathbf{X}]_{ij}$  表示  $\mathbf{X}$  的第  $i$  行第  $j$  列元素);
2.  $\hat{\mathbf{W}}$  和  $\mathbf{I} - \hat{\mathbf{W}}$  均为半正定矩阵。

注意： $\hat{\mathbf{W}}$  不必为对称矩阵。图 2 是一个  $\hat{\mathbf{W}}$  与  $\Sigma_D$  相容的例子。

$$\Sigma_D = \begin{bmatrix} 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix} \Rightarrow \hat{\mathbf{W}} = \begin{bmatrix} 0.8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.1 & 0.3 & 0 & 0 \\ 0 & 0.1 & 0.6 & 0.2 & 0 & 0 \\ 0 & 0.3 & 0.2 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

图 2. 一个  $\hat{\mathbf{W}}$  与  $\Sigma_D$  相容的例子。当  $\Sigma_D$  的第  $i$  个对角元和第  $j$  个对角元不相等时， $\hat{\mathbf{W}}$  的第  $i$  行第  $j$  列元素就为 0。

### 3.1 应用：提升潜在低秩表示模型的聚类性能[20]

我们首先把数据用 RPCA 去噪，得到的数据  $\mathbf{D}$  可以认为是无噪的（或噪声可以忽略），然后在  $\mathbf{D}$  上应用松弛的无噪潜在低秩表示模型，但是由定理 6，其解不唯一，我们可以试图从中选取一个最适合于聚类的解。受稀疏子空间聚类模型[21]的启发，我们要求表达系数矩阵  $\mathbf{Z}^*$  稀疏（因为  $\mathbf{Z}^*$  本身已经低秩，再对它进行低秩约束没有意义）：

$$\min_{\mathbf{Z}, \mathbf{W}} \|\mathbf{Z}\|_1, \quad s.t. \quad \mathbf{Z} = \mathbf{V}_D \mathbf{W} \mathbf{V}_D^T, \mathbf{W} \text{ is diagonal}, 0 \leq \text{diag}(\mathbf{W}) \leq 1, \quad (24)$$

$$\text{tr}(\mathbf{W}) = 1,$$

其中  $\mathbf{W}$  限定为对角阵是因为对于随机抽取的数据，其奇异值一般来说是各不相同的，由于  $\mathbf{W}$  要与奇异值矩阵相容，一般来说它就必须是对角阵。于是  $\mathbf{W}$  和  $\mathbf{I} - \mathbf{W}$  均为半正定矩阵的条件就退化为  $0 \leq \text{diag}(\mathbf{W}) \leq 1$ 。额外的约束  $\text{tr}(\mathbf{W}) = 1$  只是为了避免平凡解  $\mathbf{Z} = \mathbf{0}$ 。(24)

是凸优化问题，可以很方便地应用交错方向法求解 [22][23]。求得最优解  $\mathbf{Z}^*$  后，以  $\mathbf{S} = |\mathbf{Z}^*| + |(\mathbf{Z}^*)^T|$  为相似度矩阵，可以对样本进行聚类，如谱聚类[24]。图 3 给出了一个仿真的例子，可以看出，新的算法（Robust LatLRR）比直接在带噪声数据上应用带噪声的潜在低秩表示模型（8）精度和鲁棒性都好。在真实数据库 Hopkins 155 和 Extended Yale B 上，新的算法的精度也明显领先（表 1、2）。

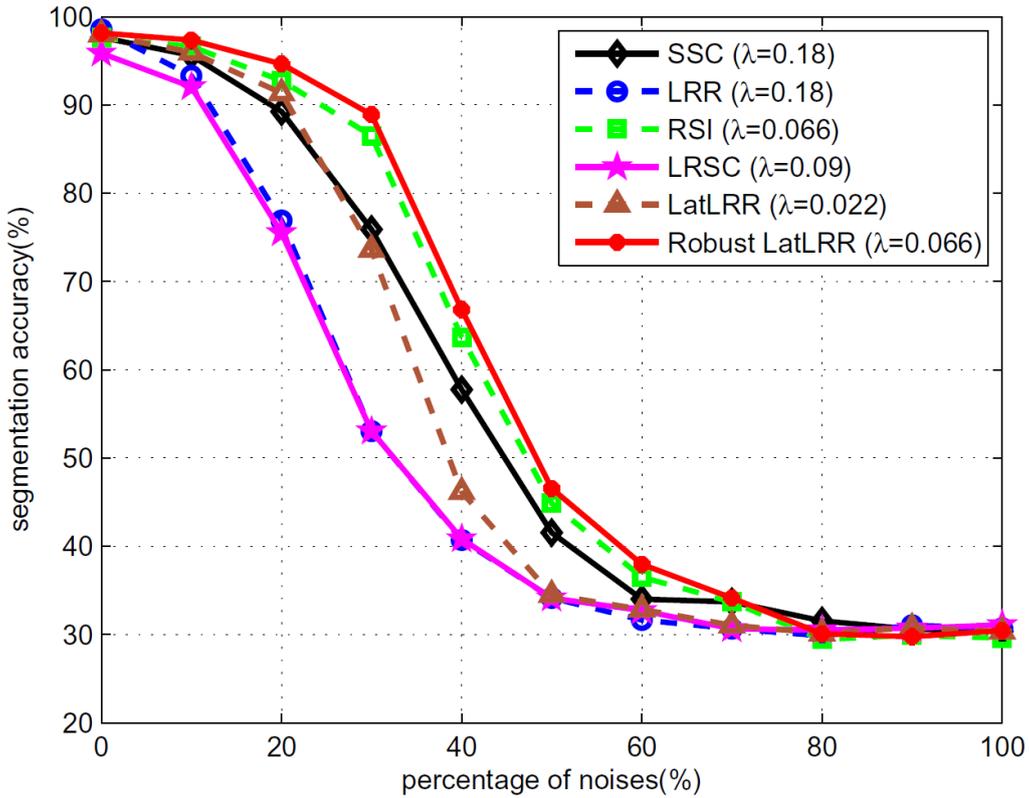


图 3. 各子空间聚类模型在仿真数据上的性能比较, 各模型的主要参数  $\lambda$  都在 20% 数据损毁时各自调到最优。

表 1: 在 Hopkins 155 数据库上的聚类错误率 (%)。对于鲁棒的潜在低秩表示 (RobustLatLRR), 参数  $\lambda$  设为  $0.806/\sqrt{n}$ 。其他方法的参数也各自调到最优。

	SSC	LRR	RSI	LRSC	LatLRR	Robust LatLRR
MAX	46.75	49.88	47.06	40.55	42.03	<b>35.06</b>
MEAN	<b>2.72</b>	5.64	6.54	4.28	4.17	3.74
STD	8.20	10.35	9.84	8.55	9.14	<b>7.02</b>

表 2: 在 Extended Yale B 数据库上不同人数情况下的聚类准确率 (%)。对于鲁棒的潜在低秩表示 (RobustLatLRR), 参数  $\lambda$  依次设为 0.014, 0.013, 0.0135。其他方法的参数也各自调到最优。

Persons	SSC	LRR	RSI	LRSC	LatLRR	Robust LatLRR
5	90.31	91.25	88.44	75.94	69.06	<b>95.00</b>
7	87.05	72.77	89.73	65.63	42.86	<b>92.86</b>
9	71.35	60.42	87.67	51.04	36.11	<b>91.49</b>

#### 4 若干主要低秩模型的解之间的关系[25]

本节我们对噪声项所使用的范数不做限制, 因为本节的结论不依赖于  $\mathbf{E}$  的范数。

前面讨论的都是无噪声低秩模型, 但是现实中不可能数据没有噪声。这个理论与实践不一致的问题可以通过如下方式解决。我们首先看到, 在低秩表示模型和潜在低秩表示模型中, 当数据中噪声较大时, 用带噪声的数据本身作为字典来表达数据是不太合理的。因此我们考虑先把数据去噪, 再运用低秩表示模型和潜在低秩表示模型, 得到相应的鲁棒的低秩表示模

型和鲁棒的潜在低秩表示模型，其中  $f(\mathbf{E})$  未明确指定。

鲁棒的低秩表示 (Robust LRR) 模型[16]:

$$\min_{\mathbf{Z}, \mathbf{E}} \text{rank}(\mathbf{Z}) + \lambda f(\mathbf{E}), \quad s.t. \quad \mathbf{A} = \mathbf{AZ}, \mathbf{D} = \mathbf{A} + \mathbf{E}, \quad (25)$$

鲁棒的潜在低秩表示 (Robust LatLRR) 模型[20]:

$$\min_{\mathbf{Z}, \mathbf{L}, \mathbf{E}} \text{rank}(\mathbf{Z}) + \text{rank}(\mathbf{L}) + \lambda f(\mathbf{E}), \quad s.t. \quad \mathbf{A} = \mathbf{AZ} + \mathbf{LA}, \mathbf{D} = \mathbf{A} + \mathbf{E}, \quad (26)$$

当把  $\text{rank}$  函数换成核范数时，得到相应的凸松弛模型。

松弛的鲁棒低秩表示 (Relaxed Robust LRR) 模型[16]:

$$\min_{\mathbf{Z}, \mathbf{E}} \|\mathbf{Z}\|_* + \lambda f(\mathbf{E}), \quad s.t. \quad \mathbf{A} = \mathbf{AZ}, \mathbf{D} = \mathbf{A} + \mathbf{E}, \quad (27)$$

松弛的鲁棒潜在低秩表示 (Relaxed Robust LatLRR) 模型[20]:

$$\min_{\mathbf{Z}, \mathbf{L}, \mathbf{E}} \|\mathbf{Z}\|_* + \|\mathbf{L}\|_* + \lambda f(\mathbf{E}), \quad s.t. \quad \mathbf{A} = \mathbf{AZ} + \mathbf{LA}, \mathbf{D} = \mathbf{A} + \mathbf{E}. \quad (28)$$

我们以 (25) 为例做如下推导。注意到给定  $\mathbf{A}$ ， $\min_{\mathbf{A}=\mathbf{AZ}} \text{rank}(\mathbf{Z}) = \text{rank}(\mathbf{A})$ 。因此 (25)

可以拆成两个相对简单的子问题:

$$\min_{\mathbf{A}, \mathbf{E}} \text{rank}(\mathbf{A}) + \lambda f(\mathbf{E}), \quad s.t. \quad \mathbf{D} = \mathbf{A} + \mathbf{E}, \quad (29)$$

$$\min_{\mathbf{Z}} \text{rank}(\mathbf{Z}), \quad s.t. \quad \mathbf{A} = \mathbf{AZ}. \quad (30)$$

(29) 可以称为广义的鲁棒主成分分析问题，而 (30) 的解由定理 4 给出。对于另外三个模型 (26) - (28)，也可以拆成两个子问题 (对松弛的模型，推导时需要用到 (10))，其中一个为广义的鲁棒主成分分析问题 (29)，另一个子问题的解由闭解给出。由于广义的鲁棒主成分分析问题 (29) 的求解相对来说要简单些，如果得到 (29) 的解，就马上能得到 (25) - (28) 的解。反之也对。具体来说，我们有如下的定理。

定理 7. 对于 (29) 的任一个解  $(\mathbf{A}^*, \mathbf{E}^*)$ ，假定  $\mathbf{A}^*$  的瘦型奇异值分解为  $\mathbf{U}_{A^*} \mathbf{S}_{A^*} \mathbf{V}_{A^*}^T$ ，则

$((\mathbf{A}^*)^+ \mathbf{A}^* + \mathbf{S} \mathbf{V}_{A^*}^T, \mathbf{E}^*)$  为 (25) 的解，其中  $\mathbf{S}$  是任一满足  $\mathbf{V}_{A^*}^T \mathbf{S} = \mathbf{0}$  的矩阵。反过来，如果

$(\mathbf{Z}^*, \mathbf{E}^*)$  是 (25) 的解，则  $(\mathbf{D} - \mathbf{E}^*, \mathbf{E}^*)$  是 (29) 的解。

定理 8. 对于 (29) 的任一个解  $(\mathbf{A}^*, \mathbf{E}^*)$ ， $((\mathbf{A}^*)^+ \mathbf{A}^*, \mathbf{E}^*)$  为 (27) 的解。反过来，如果  $(\mathbf{Z}^*, \mathbf{E}^*)$

是 (27) 的解，则  $(\mathbf{D} - \mathbf{E}^*, \mathbf{E}^*)$  是 (29) 的解。

定理 9. 对于 (29) 的任一个解  $(\mathbf{A}^*, \mathbf{E}^*)$ ，假定  $\mathbf{A}^*$  的瘦型奇异值分解为  $\mathbf{U}_{A^*} \mathbf{S}_{A^*} \mathbf{V}_{A^*}^T$ ，则

$(\mathbf{Z}^*, \mathbf{L}^*, \mathbf{E}^*)$  为 (26) 的解，其中

$$\mathbf{Z}^* = \mathbf{V}_{A^*} \tilde{\mathbf{W}} \mathbf{V}_{A^*}^T + \mathbf{S}_1 \tilde{\mathbf{W}} \mathbf{V}_{A^*}^T, \quad \mathbf{L}^* = \mathbf{U}_{A^*} \mathbf{\Sigma}_{A^*} (\mathbf{I} - \tilde{\mathbf{W}}) \mathbf{\Sigma}_{A^*}^{-1} \mathbf{U}_{A^*}^T + \mathbf{U}_{A^*} \mathbf{\Sigma}_{A^*} (\mathbf{I} - \tilde{\mathbf{W}}) \mathbf{S}_2,$$

$\tilde{\mathbf{W}}$  是任一幂等矩阵,  $\mathbf{S}_1$  和  $\mathbf{S}_2$  是任一满足下列条件的矩阵:

1.  $\mathbf{V}_{A^*}^T \mathbf{S}_1 = \mathbf{0}$ ,  $\mathbf{S}_2 \mathbf{U}_{A^*} = \mathbf{0}$ ;
2.  $\text{rank}(\mathbf{S}_1) \leq \text{rank}(\tilde{\mathbf{W}})$ ,  $\text{rank}(\mathbf{S}_2) \leq \text{rank}(\mathbf{I} - \tilde{\mathbf{W}})$ .

反过来, 如果  $(\mathbf{Z}^*, \mathbf{L}^*, \mathbf{E}^*)$  是 (26) 的解, 则  $(\mathbf{D} - \mathbf{E}^*, \mathbf{E}^*)$  是 (29) 的解。

定理 10. 对于 (29) 的任一个解  $(\mathbf{A}^*, \mathbf{E}^*)$ , 假定  $\mathbf{A}^*$  的瘦型奇异值分解为  $\mathbf{U}_{A^*} \mathbf{S}_{A^*} \mathbf{V}_{A^*}^T$ , 则

$(\mathbf{Z}^*, \mathbf{L}^*, \mathbf{E}^*)$  为 (28) 的解, 其中

$$\mathbf{Z}^* = \mathbf{V}_{A^*} \hat{\mathbf{W}} \mathbf{V}_{A^*}^T, \quad \mathbf{L}^* = \mathbf{U}_{A^*} (\mathbf{I} - \hat{\mathbf{W}}) \mathbf{U}_{A^*}^T,$$

$\hat{\mathbf{W}}$  是任一满足下列条件的块对角矩阵:

1.  $\hat{\mathbf{W}}$  的对角块与  $\Sigma_{A^*}$  相容;
2.  $\hat{\mathbf{W}}$  和  $\mathbf{I} - \hat{\mathbf{W}}$  均为半正定矩阵。

反过来, 如果  $(\mathbf{Z}^*, \mathbf{L}^*, \mathbf{E}^*)$  是 (28) 的解, 则  $(\mathbf{D} - \mathbf{E}^*, \mathbf{E}^*)$  是 (29) 的解。

以上四个定理所揭示的 (25) - (29) 解之间的互为表出关系可以用图 4 来总结。我们把广义鲁棒主成分分析模型 (29) 放在中心是因为它是这几个低秩模型里最简单的, 所以应当优先解它, 由它来获得其他低秩模型的解。

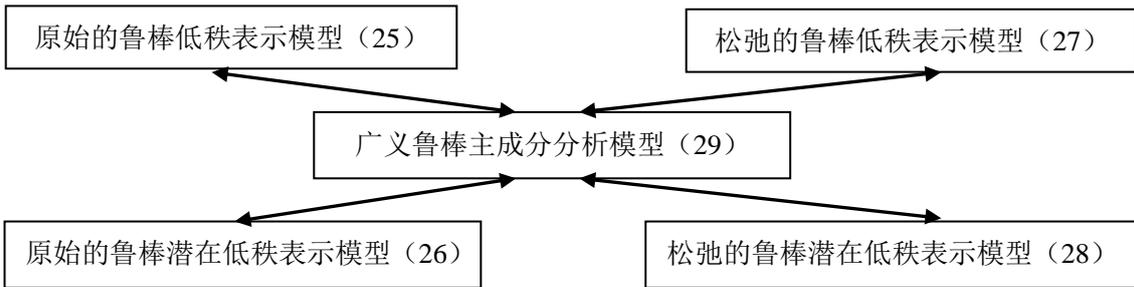


图 4. 五个低秩模型解之间的互为表出关系。如果 A 指向 B, 表示从 A 的解可以立即得到 B 的解。

以上解之间关系的意义在于:

1. 我们可能会得到非凸问题 (25) - (28) 的全局最优解。这是因为 (29) 在适当的条件下是可以求得全局最优解的[6][26]。
2. 我们可以得到求解 (25) - (28) 的快速算法。这是因为我们可以先解相对简单的 (29), 从而得到 (25) - (28) 的解。更进一步, 求解 (29) 有超快的随机算法, 如  $L_1$ -filtering[27]

或  $L_{2,1}$ -filtering 算法[20] (其基本原理是先恢复出矩阵的部分, 然后利用低秩性和噪声的稀

疏性对矩阵的其余部分进行并行去噪)。

#### 4.1 解的质量的比较

我们把先归结求解 (29) 再用闭解表达的算法称为 REDU-EXPR (为 Reduce-Express 的缩写)。图 5 是和直接使用部分交错方向法 (partial ADM) [30] 求解 (27) 的聚类准确率比较。可以看到, REDU-EXPR 算法得到的解对噪声抵抗能力明显高得多。

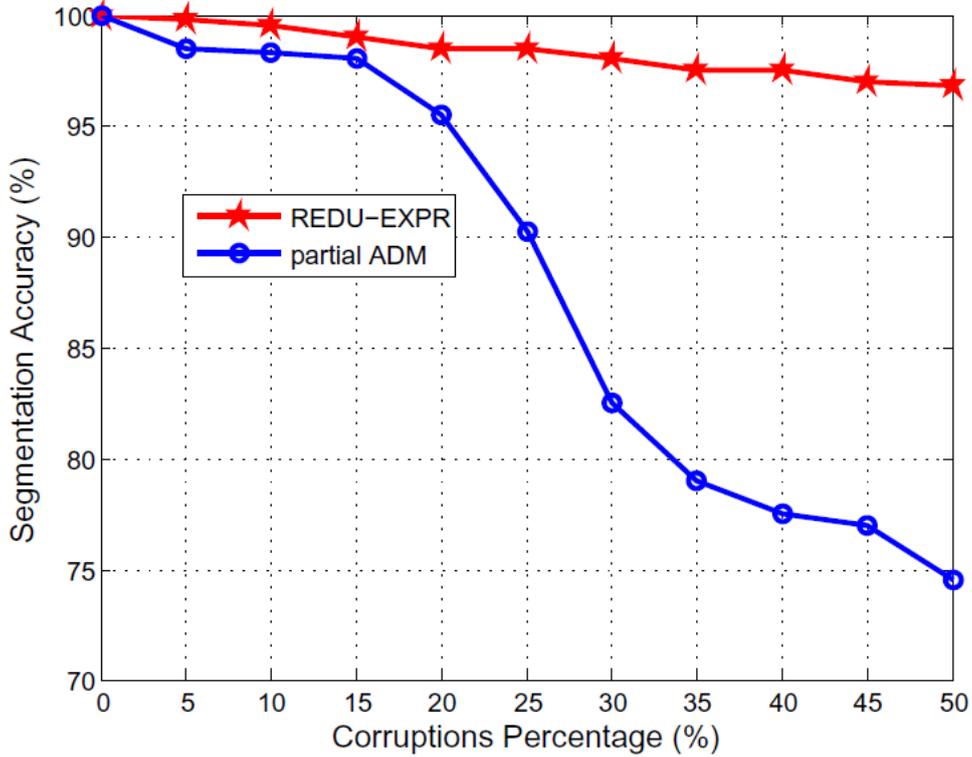


图 5. REDU-EXPR 算法和交错方向法求解(27)所得解的质量的比较,其中(27)中  $\lambda = \frac{1}{\sqrt{\ln n}}$ ,

$$f(\mathbf{E}) = \|\mathbf{E}\|_{2,1}, \quad n \text{ 为样本个数。聚类准确率为 10 次平均值。}$$

表 3 比较了两种算法得到的解的秩、稀疏度、目标函数值, 可以看到 REDU-EXPR 得到的解的数值质量明显优于交错方向法得到的。在运算速度方面, REDU-EXPR 也有明显优势。

表 3: 使用 partial ADM 与 REDU-EXPR 求解鲁棒低秩表示问题时随噪声占比的增长在鲁棒性与速度上的对比。所有的实验均执行 10 次取平均, 参数  $\lambda$  均设为:  $\lambda = 1/\sqrt{\ln n}$ , 其中  $n$  为数据规模。

Noise Percentage (%)	0	10	20	30	40	50
Rank(Z) (partial ADM)	20	30	30	30	30	30
Rank(Z) (REDU-EXPR)	20	20	20	20	20	20
$\ E\ _{2,0}$ (partial ADM)	0	99	200	300	400	500
$\ E\ _{2,0}$ (REDU-EXPR)	0	100	200	300	400	500

Objective (partial ADM)	<b>20.00</b>	67.67	106.10	144.14	182.19	220.24
Objective (REDU-EXPR)	<b>20.00</b>	<b>58.05</b>	<b>96.10</b>	<b>134.14</b>	<b>172.19</b>	<b>210.24</b>
Time (s, partial ADM)	<b>4.89</b>	124.33	126.34	119.12	115.20	113.94
Time(s, REDU-EXPR)	10.67	<b>9.60</b>	<b>8.34</b>	<b>8.60</b>	<b>9.00</b>	<b>12.86</b>

#### 4.2 求解速度的比较

表 4 给出了在真实数据库 Extended YaleB 上求解松弛的鲁棒低秩表示模型的计算时间的比较, REDU-EXPR 优势明显。表 5 给出了在仿真数据上的计算时间比较, 采用了随机优化算法  $\ell_{2,1}$ -filtering 算法[20]求解广义鲁棒主成分分析模型 (29) 后, REDU-EXPR 进一步获得了明显的加速。

表 4:在 Extended Yale B 数据库上非监督人脸图像的聚类结果。R-LRR 为鲁棒低秩表示模型。

Model	Method	Accuracy	CPUTime (h)
LRR	ADMM	-	>10
R-LRR	ADMM	-	Did not converge
R-LRR	Partial ADMM	-	>10
R-LRR	REDU-EXPR	61.6365%	0.4603

表 5:随着数据规模增长, 各方法在 CPU 耗时(秒)上的对比(方法包括: 用 ADMM 求解 LRR、用 partial ADMM 求解鲁棒低秩表示模型 (R-LRR)、用 REDU-EXPR 而不采用  $\ell_{2,1}$ -filtering 算法来求解 R-LRR, 以及用 REDU-EXPR 并采用  $\ell_{2,1}$ -filtering 算法来求解 R-LRR)。

Data Size	LRR (ADMM)	R-LRR (partial ADMM)	R-LRR (REDU-EXPR)	R-LRR (filtering REDU-EXPR)
250×250	33.0879	4.9581	1.4315	<b>0.6843</b>
500×500	58.9177	7.2029	1.8383	<b>1.0917</b>
1000×1000	370.1058	24.5236	6.1054	<b>1.5429</b>
2000×2000	>3600	124.3417	28.3048	<b>2.4426</b>
4000×4000	>3600	411.8664	115.7095	<b>3.4253</b>

#### 5 闭解应用的另一个例子: 仿射变换下特征点误匹配检测[28]

在三维重建或部分重复图像检测中, 为了加速计算往往需要用特征点寻找内容相同的区域。在特征点匹配好后, 需要进行误匹配操作, 去除误匹配, 才能计算出正确的两幅图匹配区域之间的几何变换 (图 6)。

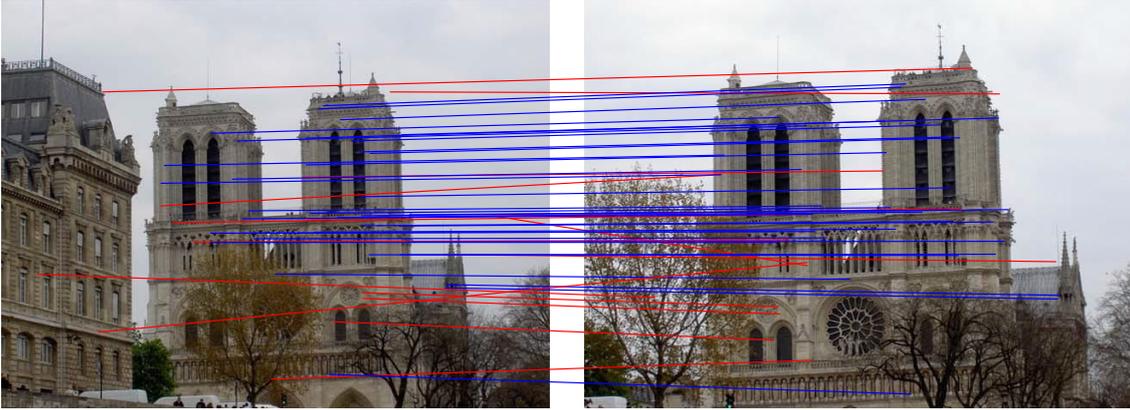


图 6. 图像误匹配检测问题示意图。

目前误匹配检测算法大部分适用于几何变换是欧氏（旋转、平移、放缩）的情形，能适用于仿射变换的算法非常稀少。一个有代表性的算法是 Guo 和 Cao 等人提出来的[29]，它需要假定一些锚点是正确匹配的，然后利用平面上的点关于锚点的面积坐标（表达成包含该点的锚点三角形顶点的线性组合）是仿射不变量的原理来检测误匹配点。我们提出了一个非常简洁的方法。不同于 Guo 和 Cao 等人的思路，我们不限定用包含该点的锚点三角形顶点来表达平面上的点，而是用所有的特征点来表达，但是要求表达系数矩阵低秩，这样就避免了假设锚点必须正确匹配、计算关于锚点的 Delauney 三角化和判断给定点在哪个 Delauney 三角形里。因此，根据定理 1，该表达系数矩阵为特征点坐标所构成的矩阵的形状交互矩阵。很容易证明，特征点坐标所构成的矩阵的形状交互矩阵是仿射不变量，通过比较两幅图各自特征点坐标所构成的矩阵的形状交互矩阵，可以检测出误匹配点（图 7）。在标准数据库上的部分重复图像检索任务的测试表明检索精度高于现有算法（图 8 (a) - (c)），在有明显仿射或透射变换时优势更为明显（图 8 (d)）。其运算速度也相当有竞争力（表 6）。

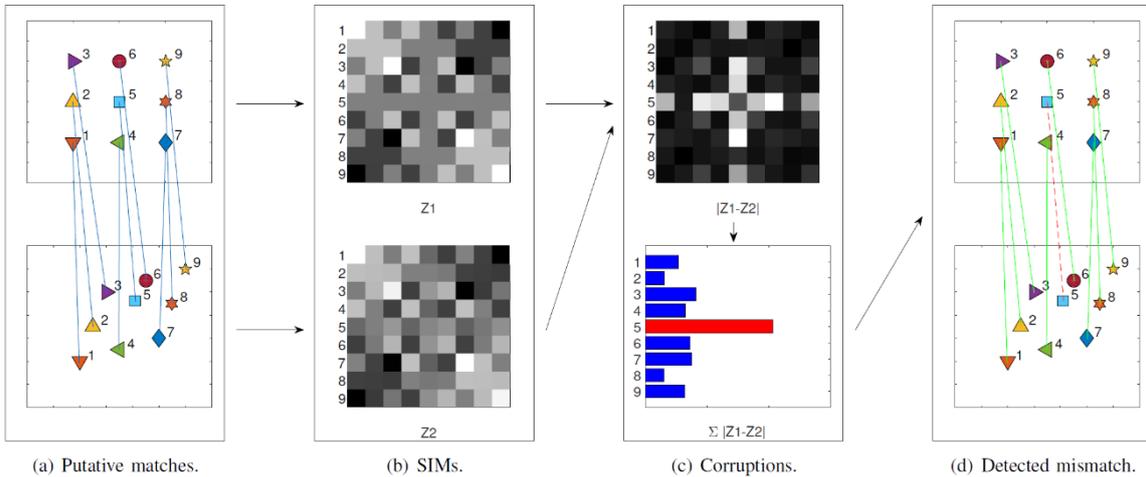


图 7. 基于形状交互矩阵的误匹配点检测算法示意图。

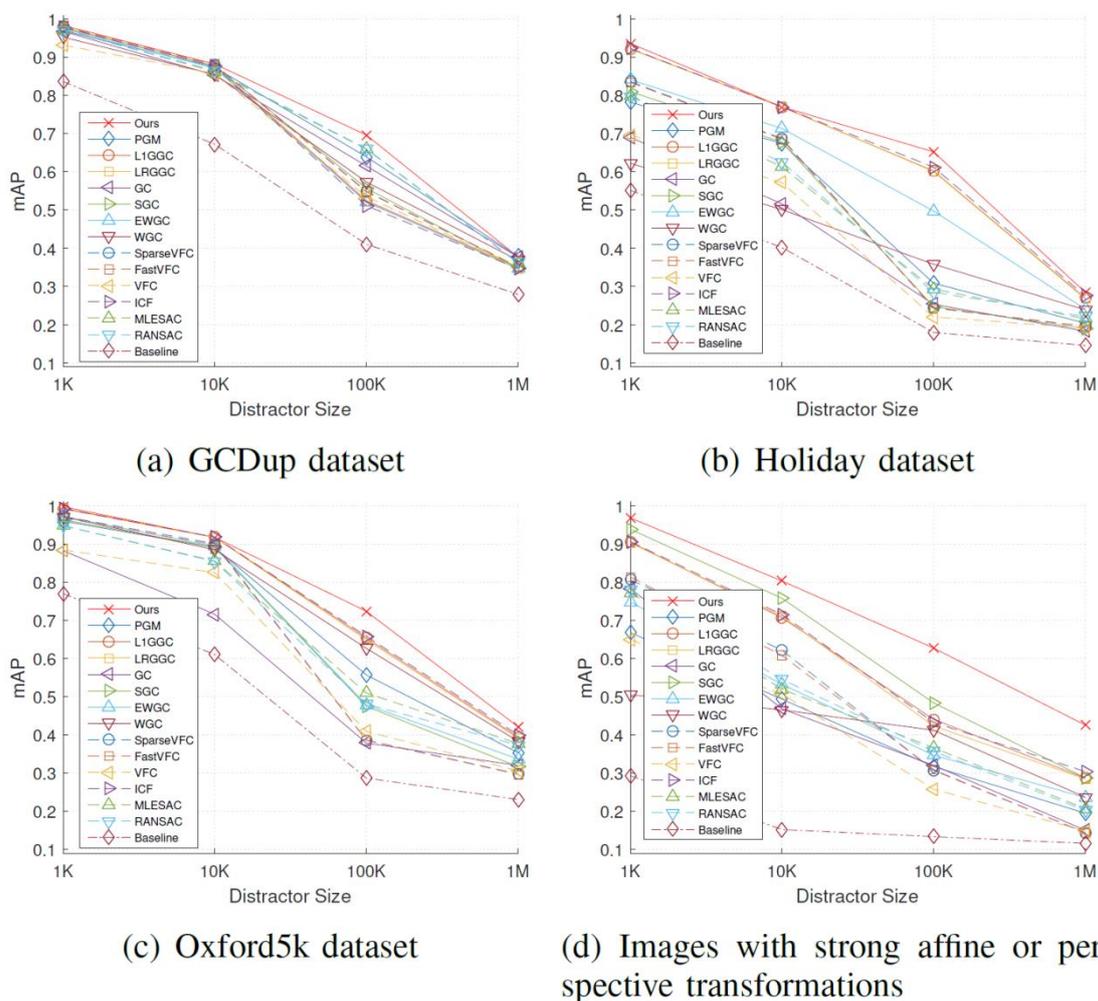


图 8. 在标准数据库上的部分重复图像检索任务的测试结果。

表 6: 所有方法在三个数据集上的平均计算耗时

Category	Method	Dataset		
		GCDup	Holiday	Oxford5k
Non-iterative	Ours	0.494s	0.175s	0.464s
	PGM	2.095s	0.684s	1.915s
	L1GGC	2.391s	0.819s	2.224s
	LRGGC	6.412s	2.243s	6.059s
	GC	7.532s	2.465s	6.957s
	SGC	4.742s	1.242s	4.355s
	EWGC	0.381s	0.111s	0.353s
	WGC	0.105s	0.056s	0.099s
Iterative	Sparse VFC	1.345s	0.710s	1.499s
	Fast VFC	5.396s	2.725s	6.000s
	VFC	7.301s	2.702s	6.658s
	ICF	13.35s	7.377s	13.89s
	MLESAC	2.553s	1.488s	2.593s
	RANSAC	2.479s	1.351s	2.476s

## 6 结语

由前面的介绍，低秩模型要比稀疏模型有更丰富的数学性质，其无噪声情形下的闭解在理论与应用上都有着重要的意义。事实上，有更多带闭解的低秩模型这里没有介绍，如[30]里所列出的。寻找更多的带闭解的低秩模型并应用于更多的实际问题将是有趣的探索。

### 参考文献

- [1] Mayer-Schönberger M, Cukier K. *Big Data: A Revolution That Will Transform How We Live*. Eamon Dolan/Houghton Mifflin Harcourt, 1994.
- [2] Vidal R. Subspace clustering, *IEEE Signal Processing Magazine*.2011, 28:52–68
- [3] Roweis S, Saul L. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, 290:2323–2326
- [4] Lin Z. A Review on Low-Rank Models in Signal and Data Analysis. *Big Data and Information Analytics*.2016, 1(2/3):139-161
- [5] Candès E, Recht B. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 2009, 9:717–772
- [6] Candès E, Li X, Ma Y, Wright J. Robust principal component analysis? *Journal of the ACM*, 2011, 58:1–37
- [7] Liu G, Lin Z, Yan S, Sun J, Yu Y, Ma Y. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1):171–184
- [8] Liu G, Yan S. Latent low-rank representation for subspace segmentation and feature extraction, In: *Proceedings of IEEE Conference on Computer Vision*, 2011, 1615–1622
- [9] Fazel M. Matrix Rank Minimization with Applications. Ph.D. Thesis, 2002.
- [10] Rockafellar R. *Convex Analysis*. Princeton University Press, 1970
- [11] Peng Y, Ganesh A, Wright J, Xu W, Ma Y. RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34 (11):2233–2246
- [12] Zhang Z, Ganesh A, Liang X, Ma Y. TILT: Transform invariant low-rank textures. *International Journal of Computer Vision*, 2012, 99 (1):1–24
- [13] Cheng B, Liu G, Wang J, Huang Z, Yan S. Multi-task low-rank affinity pursuit for image segmentation. In: *Proceedings of International Conference on Computer Vision*, 2011, 2439–2446
- [14] Lang C, Liu G, Yu J, Yan S. Saliency detection by multitask sparsity pursuit. *IEEE Transactions on Image Processing*, 2012, 21 (3):1327–1338
- [15] Zhu G, Yan S, Ma Y. Image Tag Refinement Towards Low-Rank Content-Tag Prior and Error Sparsity. In: *Proceedings of ACM International Conference on Multimedia*, 2010, 461-470
- [16] Wei S, Lin Z. Analysis and Improvement of Low Rank Representation for Subspace segmentation. arXiv: 1107.1561
- [17] Yu Y, Schuurmans D. Rank/norm regularization with closed-form solutions: Application to subspace clustering. In: *Proceedings of Uncertainty in Artificial Intelligence*, 2011, 778–785

- 
- [18] Zhang H, Lin Z, Zhang C. A counterexample for the validity of using nuclear norm as a convex surrogate of rank. In: *Proceedings of European Conference on Machine Learning*, 2013, 226–241
- [19] Dai Y, Li H, He M. A simple prior-free method for non-rigid structure-from-motion factorization. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2012
- [20] Zhang H, Lin Z, Zhang C, Gao J. Robust latent low rank representation for subspace clustering. *Neurocomputing*, 2014, 145:369–373.
- [21] Elhamifar E, Vidal R. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(11):2765–2781.
- [22] Lin Z, Liu R, Su Z. Linearized Alternating Direction Method with Adaptive Penalty for Low Rank Representation. In: *Proceedings of Advances in Neural Information Processing Systems*, 2011
- [23] Lin Z, Liu R, Li H. Linearized Alternating Direction Method with Parallel Splitting and Adaptive Penalty for Separable Convex Programs in Machine Learning. *Machine Learning*, 2015, 99(2):287-325
- [24] Shi J, Malik J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8):888–905
- [25] Zhang H, Lin Z, Zhang C, Gao J. Relation among some low rank subspace recovery models. *Neural Computation*, 2015, 27:1915–1950.
- [26] Zhang H, Lin Z, Zhang C, Chang E. Exact recoverability of robust PCA via outlier pursuit with tight recovery bounds. In: *Proceedings of AAAI Conference on Artificial Intelligence*, 2015, 3143–3149
- [27] Liu R, Lin Z, Su Z, Gao J. Linear time principal component pursuit and its extensions using  $\ell_1$  filtering. *Neurocomputing*, 2014, 142:529–541
- [28] Lin Y, Lin Z, Zha H. The shape interaction matrix based affine invariant mismatch removal for partial duplicated image search. *IEEE Transactions on Image Processing*, 2017, 26(2): 561-573.
- [29] Guo X, Cao X. Triangle-constraint for finding more good features. In: *Proceedings of International Conference on Pattern Recognition*, 2010, 1393–1396
- [30] Favaro P, Vidal R, Ravichandran A. A closed form solution to robust subspace estimation and clustering. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2011, 1801–1807