

# 鲁棒主成分分析

陈煜<sup>1</sup>, 姜永鹏<sup>11</sup>

<sup>1</sup> 清华大学自动化系

2022 年 12 月 19 日

---

<sup>1</sup>两位小组成员的贡献比例为 1:1

## 1. 引言

本大作业考虑鲁棒主成分分析 (Robust Principal Analysis, RPCA) 任务，主要研究该问题的通用及特殊优化算法，在指定图片上测试，并分析超参数对算法性能的影响。本报告将按以下几部分展开：在第 2 章叙述一种基于梯度下降的求解方法，在第 3 章叙述一种基于增广拉格朗日函数与交替方向乘子法的求解方法，在第 4 章叙述一种基于矩阵低秩表示的替代模型及其优化方法，在第 5 章给出实验结果与分析，最后在第 6 章给出有关代码说明、小组分工等其他需要说明的内容。

## 2. 基于梯度下降的 RPCA 算法

根据作业文档中相关描述，鲁棒主成分分析是为了找到一个低维子空间来近似表示观测数据，该任务通常被简化成如下的优化问题：

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{E}} \quad & \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 \\ \text{s.t.} \quad & \mathbf{A} + \mathbf{E} = \mathbf{D} \end{aligned} \quad (1)$$

其中  $\mathbf{A}$  为低秩矩阵， $\mathbf{E}$  为高斯噪声矩阵， $\mathbf{D}$  为待处理数据

代入约束条件，可以得到无约束优化问题 (2)

$$\min_{\mathbf{A}} \|\mathbf{A}\|_* + \lambda \|\mathbf{D} - \mathbf{A}\|_1 \quad (2)$$

虽然这个问题是凸的，但是由于核范数在矩阵奇异值趋于 0 时以及 1-范数导数不连续会导致梯度不存在，只存在次梯度，会导致沿次梯度方向进行优化更新不能够保证函数值的下降。为此，采用梯度下降法来优化目标函数时需要针对次梯度做修正。

进行迭代更新时，若当前  $\mathbf{A}$  的奇异值不含 0，则其导数为

$$\nabla_{\mathbf{A}} \|\mathbf{A}\|_* = \mathbf{U}_r \mathbf{V}_r^\top \quad (3)$$

其中  $\mathbf{U}_r$  和  $\mathbf{V}_r$  分别为矩阵  $\mathbf{A}$  的前  $r$  个左奇异向量和右奇异向量组成的矩阵， $r = \text{Rank}(\mathbf{A})$

与此同时，通过数值微分计算当前迭代点的左右导数

$$\begin{aligned} \nabla_{\mathbf{A}_L} \|\mathbf{A}\|_{*ij} &= \lim_{h \rightarrow 0} \frac{\mathbf{A}(a_{ij} + h) - \mathbf{A}}{h} \\ h < 0 \end{aligned} \quad (4)$$

$$\begin{aligned} \nabla_{\mathbf{A}_R} \|\mathbf{A}\|_{*ij} &= \lim_{h \rightarrow 0} \frac{\mathbf{A}(a_{ij} + h) - \mathbf{A}}{h} \\ h > 0 \end{aligned} \quad (5)$$

其中  $\mathbf{A}(a_{ij} + h)$  表示矩阵  $\mathbf{A}$  的第  $(i, j)$  个元素  $a_{ij}$  在数值上增大  $h$

在使用梯度下降进行更新前，比较左右导数各个元素的符号，若符号不一致，代表  $\mathbf{0} \in \partial \|\mathbf{A}\|_*$ ，此时在该元素上不作梯度下降更新。

我们以同样的方法处理 1-范数的导数，采用定步长的一阶梯度下降方法，即可得到求解 RPCA 问题的伪代码 Algorithm 1

---

**Algorithm 1:** Gradient Decent for RPCA

---

**Input:** Raw Matrix  $\mathbf{D}$

**Output:** Result  $\mathbf{A}$   $\mathbf{E}$

```

1 Initialize  $\mathbf{A}$ 
2 Set hyper parameter    $h = 1e^{-8}$    stp =  $1e^{-4}$    iteration =  $1e^3$ 
3 for  $k^{th}$  iteration do
4    $\mathbf{g}^{1(k)} \leftarrow \nabla_{\mathbf{A}_L} \|\mathbf{A}^{(k)}\|_*$ 
5    $\mathbf{g}^{2(k)} \leftarrow \nabla_{\mathbf{A}_R} \|\mathbf{A}^{(k)}\|_*$ 
6    $\mathbf{g}^{(k)} \leftarrow \mathbf{U}_r^{(k)} \mathbf{V}_r^{(k)\top}$ 
7   for  $a_{ij}^{(k)} \in \mathbf{A}^{(k)}$  do
8     if  $\mathbf{g}_{ij}^{(k)} \times \mathbf{g}_{ij}^{(k)} < 0$  then
9        $\mathbf{g}_{ij}^{(k)} \leftarrow 0$ 
10      end
11    end
12     $\mathbf{g}_{ij}^{(k)} \leftarrow \mathbf{g}_{ij}^{(k)} - \nabla_{\mathbf{A}} \lambda \|\mathbf{D} - \mathbf{A}^{(k)}\|_1$ 
13     $\mathbf{A}^{(k+1)} \leftarrow \mathbf{A}^{(k)} - \mathbf{g}_{ij}^{(k)} \text{stp}$ 
14 end

```

---

为验证该梯度下降方法的有效性，我们在 Matlab 中随机生成  $3 \times 4$  的矩阵  $\mathbf{D}$ ，用梯度下降的方法计算矩阵  $\mathbf{A}$  和  $\mathbf{E}$ 。作为对照，我们采用 cvx 求解器来求解该凸优化问题，用于比较梯度下降方法所计算出的最优值和真实最优解的差距。共进行 3 次实验，实验结果如表 1 所示

实验表明，考虑次梯度进行梯度下降得到的优化结果与标准求解器计算出的最优解较为接近。二者是间隙是由于该算法在  $\mathbf{0} \in \partial(\|\mathbf{A}\|_* + \lambda \|\mathbf{D} - \mathbf{A}\|_1)$  时会停止优化。

表 1: Optimal Value of the Comparative Experiment

<i>Gradient Decent</i>	<i>CVX optimizer</i>
17.4902	17.2443
16.9765	16.6981
16.6340	16.5145

### 3. RPCA 特殊优化算法

#### 3.1 预备知识

**定义 1 (近端极小化问题)** 已知  $\mathbf{X}, \mathbf{F} \in \mathbb{R}^{m \times n}, \lambda \in \mathbb{R}$ , 定义近端极小化问题

$$\min_{\mathbf{X}} \lambda \|\mathbf{X}\|_* + \frac{1}{2} \|\mathbf{X} - \mathbf{F}\|_F^2 \quad (6a)$$

$$\min_{\mathbf{X}} \lambda \|\mathbf{X}\|_1 + \frac{1}{2} \|\mathbf{X} - \mathbf{F}\|_F^2 \quad (6b)$$

$$\min_{\mathbf{X}} \lambda \|\mathbf{X}\|_{2,1} + \frac{1}{2} \|\mathbf{X} - \mathbf{F}\|_F^2 \quad (6c)$$

其中  $\|\mathbf{X}\|_* = \text{tr}(\sqrt{\mathbf{X}^T \mathbf{X}})$ ,  $\|\mathbf{X}\|_1 = \sum_{i=1}^m \sum_{j=1}^n |\mathbf{X}_{ij}|$ ,  $\|\mathbf{X}\|_{2,1} = \sum_{j=1}^n \sqrt{\sum_{i=1}^m \mathbf{X}_{ij}^2}$  分别称为矩阵  $\mathbf{X}$  的核范数、 $1-$  范数与  $\{2, 1\}-$  范数。

**定理 1** 近端极小化问题 (1) 均为无约束的凸规划问题, 其解分别为

$$\hat{\mathbf{X}} = \mathcal{D}_\lambda(\mathbf{F}) = \mathbf{U} \mathcal{S}_\lambda(\boldsymbol{\Sigma}) \mathbf{V}^T. \quad (7a)$$

$$\hat{\mathbf{X}} = \mathcal{S}_\lambda(\mathbf{F}), \hat{x}_{ij} = \text{sgn}(f_{ij}) \cdot \max(|f_{ij}| - \lambda, 0) = \begin{cases} f_{ij} - \lambda, & \lambda < f_{ij} \\ 0, & |f_{ij}| \leq \lambda \\ f_{ij} + \lambda, & f_{ij} < -\lambda \end{cases}. \quad (7b)$$

$$\hat{\mathbf{X}} = \mathcal{W}_\lambda(\mathbf{F}), \hat{\mathbf{x}}_j = \frac{\max(\|\mathbf{f}_j\| - \lambda, 0)}{\|\mathbf{f}_j\|} \cdot \mathbf{f}_j = \begin{cases} 1 - \frac{\lambda}{\|\mathbf{f}_j\|} \cdot \mathbf{f}_j, & \lambda < \|\mathbf{f}_j\| \\ 0, & \|\mathbf{f}_j\| \leq \lambda \end{cases}. \quad (7c)$$

其中  $\mathbf{F} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T$  为矩阵  $\mathbf{F}$  的奇异值分解,  $\mathbf{f}_j$  为矩阵  $\mathbf{F}$  的第  $j$  个列向量。

**证明 1** 首先证明 (1a) 的极小化条件。根据<sup>[1]</sup>, 核范数的次梯度满足

$$\partial \|\mathbf{X}\|_* \triangleq \{\mathbf{U}_r \mathbf{V}_r^T + \mathbf{W} | \mathbf{W}^T \mathbf{U}_r = \mathbf{0}, \mathbf{W} \mathbf{V}_r = \mathbf{0}, \|\mathbf{W}\| \leq 1\}, \quad (8)$$

其中  $r = \text{rank}(\mathbf{X})$ ,  $\mathbf{U}_r = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r]$ ,  $\mathbf{V}_r = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r]$ 。简记目标函数  $g(\mathbf{X}) = p(\mathbf{X}) + q(\mathbf{X})$ , 其中  $p(\mathbf{X}) = \lambda \|\mathbf{X}\|_*$ ,  $q(\mathbf{X}) = \frac{1}{2} \|\mathbf{X} - \mathbf{F}\|_F^2$ 。下证  $\hat{\mathbf{X}} = \mathbf{U} \mathcal{S}_\lambda(\boldsymbol{\Sigma}) \mathbf{V}^T$  是原问题最优解, 不妨设  $\mathbf{F}$  的奇异值  $\sigma_1 \geq \dots \geq \sigma_k > \lambda \geq \sigma_{k+1} \geq \dots \geq \sigma_{\min(m,n)} \geq 0$ , 则最优解可简化为  $\hat{\mathbf{X}} = \mathbf{U} \max(\boldsymbol{\Sigma} - \lambda \mathbf{I}, 0) \mathbf{V}^T = \sum_{i=1}^k (\sigma_i - \lambda) \mathbf{u}_i \mathbf{v}_i^T = \mathbf{U}_k (\boldsymbol{\Sigma}_k - \lambda \mathbf{I}) \mathbf{V}_k^T$ 。验证最优解等价于验证  $\hat{\mathbf{X}}$  满足

$$\mathbf{0} \in \partial g \Leftrightarrow \mathbf{F} - \hat{\mathbf{X}} \in \lambda \partial \|\hat{\mathbf{X}}\|_*. \quad (9)$$

上式利用了 Frobenius 范数的梯度  $\nabla_{\mathbf{X}} q = \mathbf{X} - \mathbf{F}$ ,  $\partial g$  为次梯度算子, 由以上分析, 矩阵  $\mathbf{F}$  的奇异值分解  $\mathbf{F} = \mathbf{U}_k \boldsymbol{\Sigma}_k \mathbf{V}_k^T + \sum_{i=k+1}^{\min(m,n)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ , 因此  $\mathbf{F} - \hat{\mathbf{X}} = \lambda (\mathbf{U}_k \mathbf{V}_k^T + \mathbf{W})$ , 其中  $\mathbf{W} = \frac{1}{\lambda} \sum_{i=k+1}^{\min(m,n)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$  显然  $k = \text{rank}(\mathbf{X})$ ,  $\|\mathbf{W}\| = \frac{1}{\lambda} \sigma_{k+1} \leq 1$ , 由奇异向量的正交性可知  $\mathbf{W}^T \mathbf{U}_r = 0, \mathbf{W} \mathbf{V}_r = 0$ 。综上,  $\mathbf{0} \in (\partial g) \hat{\mathbf{X}}$ 。

随后证明 (1b) 的极小化条件。简记目标函数  $g(\mathbf{X}) = p(\mathbf{X}) + q(\mathbf{X})$ , 其中  $p(\mathbf{X}) = \lambda \|\mathbf{X}\|_1$ ,  $q(\mathbf{X}) = \frac{1}{2} \|\mathbf{X} - \mathbf{F}\|_F^2$ 。  
 $\hat{\mathbf{X}}$  是原问题最优解, 当且仅当

$$\mathbf{0} \in \partial g = (\partial p) \hat{\mathbf{X}} + (\hat{\mathbf{X}} - \mathbf{F}) \Leftrightarrow \mathbf{F} \in (I + \partial p) \hat{\mathbf{X}} \Leftrightarrow \hat{\mathbf{X}} = (I + \partial p)^{-1} \mathbf{F},$$

$I$  为恒等映射, 逆映射  $(I + \partial p)^{-1}$  为单射。

$$(\partial p)\hat{\mathbf{X}} = \lambda \text{sgn}(\hat{\mathbf{X}}) \Rightarrow (I + \partial p)\hat{\mathbf{X}} = \lambda \text{sgn}(\hat{\mathbf{X}}) + \hat{\mathbf{X}} \Rightarrow (I + \partial p)^{-1}\mathbf{F} = \text{sgn}(\mathbf{F}) \cdot \max(|\mathbf{F}| - \lambda, 0) \quad (10)$$

注意上式中各算子直接作用于矩阵的每个元素。

最后证明 (1c) 的极小化条件。简记目标函数  $g(\mathbf{X}) = p(\mathbf{X}) + q(\mathbf{X})$ , 其中  $p(\mathbf{X}) = \lambda \|\mathbf{X}\|_{2,1}$ ,  $q(\mathbf{X}) = \frac{1}{2} \|\mathbf{X} - \mathbf{F}\|_F^2$ 。

当  $\|\mathbf{x}_j\| \neq 0$ , 对矩阵的 2,1-范数求梯度得

$$(\nabla_{\mathbf{X}} \|\mathbf{X}\|_{2,1})_{ij} = \frac{x_{ij}}{\|\mathbf{x}_j\|},$$

因此我们将次梯度的列作为整体考虑。若  $\hat{\mathbf{X}}$  是原问题最优解, 则次梯度条件  $\mathbf{0} \in \partial g$  等价于  $\mathbf{0} \in (\partial g)\hat{\mathbf{x}}_j$  对于次梯度矩阵的列  $\forall j \in \{1, 2, \dots, n\}$  均成立, 因此有

$$\mathbf{0} \in (\partial g)\hat{\mathbf{x}}_j \Leftrightarrow \mathbf{f}_j \in (I + \partial p)\hat{\mathbf{x}}_j.$$

当  $\|\mathbf{x}_j\| \neq 0$  时,  $\mathbf{f}_j = \mathbf{x}_j + \lambda \frac{\mathbf{x}_j}{\|\mathbf{x}_j\|}$ , 即  $\|\mathbf{f}_j\| = \lambda + \|\mathbf{x}_j\| > \lambda$ ; 当  $\|\mathbf{x}_j\| = 0$  时, 有  $f_{ij} \in [-\lambda, \lambda]$ , 由此可得

$$\hat{\mathbf{x}}_j = (I + \partial p)^{-1}\mathbf{f}_j = \begin{cases} 1 - \frac{\lambda}{\|\mathbf{f}_j\|}\mathbf{f}_j, & \|\mathbf{f}_j\| > \lambda \\ \mathbf{0}, & \|\mathbf{f}_j\| \leq \lambda \end{cases} \quad (11)$$

## 3.2 鲁棒主成分分析的特殊优化方法

这里我们介绍增广拉格朗日法求解 RPCA 的方法:

**定理 2** 原 RPCA 问题 (1) 等价于迭代更新超参  $\mathbf{Y}$  和  $\rho$  后对问题 (12) 的求解

$$\min_{\mathbf{A}, \mathbf{E}} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 + \langle \hat{\mathbf{Y}}, \mathbf{D} - \mathbf{A} - \mathbf{E} \rangle + \frac{\rho}{2} \|\mathbf{D} - \mathbf{A} - \mathbf{E}\|_F^2 \quad (12)$$

Updated with  $\rho \leftarrow \mu\rho, \mu > 1$

$$\hat{\mathbf{Y}} \leftarrow \hat{\mathbf{Y}} + \rho(\mathbf{D} - \mathbf{A} - \mathbf{E})$$

其中  $\hat{\mathbf{Y}}$  是拉格朗日乘子矩阵的估计,  $\rho$  是权重因子

**证明 2** 根据拉格朗日乘子法, 原问题 (1) 等价于下述的 minmax 问题

$$\min_{\mathbf{A}, \mathbf{E}} \max_{\mathbf{Y}} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 + \langle \mathbf{Y}, \mathbf{D} - \mathbf{A} - \mathbf{E} \rangle \quad (13)$$

其中  $\mathbf{Y}$  是拉格朗日乘子矩阵。

由于求解关于拉格朗日乘子矩阵  $\mathbf{Y}$  的内层 max 优化问题会得到一个非光滑的结果, 从而导致外层的 min 问题在某些点只有次梯度, 无法保证梯度下降。因此采用增广拉格朗日法, 给出原问题的增广拉格朗日函数 (*Augmented Lagrangian*)

$$L(\mathbf{A}, \mathbf{E}, \mathbf{Y}) = \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 + \langle \mathbf{Y}, \mathbf{D} - \mathbf{A} - \mathbf{E} \rangle - \frac{1}{2\rho} \|\mathbf{Y} - \hat{\mathbf{Y}}\|_F^2 \quad (14)$$

其中  $\hat{\mathbf{Y}}$  是我们对最优拉格朗日乘子矩阵  $\mathbf{Y}$  的先验估计, 当  $\rho \rightarrow \infty$  时, 求解 (14) 的 minmax 问题与问题 (13) 等效

由于 (14) 引入了对  $\hat{\mathbf{Y}}$  的二次项  $\frac{1}{2\rho}\|\mathbf{Y} - \hat{\mathbf{Y}}\|_F^2$ , 消除了内层  $\max$  问题求解结果的非光滑性, 从而保证了外层  $\min$  问题的梯度下降。

内层的  $\max$  问题是一个二次优化问题, 其解为

$$\mathbf{Y}^*(\hat{\mathbf{Y}}) = \hat{\mathbf{Y}} + \rho(\mathbf{D} - \mathbf{A} - \mathbf{E}) \quad (15)$$

将 (15) 代入 (14) 的  $\min\max$  中即可得到等效优化问题

$$\min_{\mathbf{A}, \mathbf{E}} \|\mathbf{A}\|_* + \lambda\|\mathbf{E}\|_1 + \langle \hat{\mathbf{Y}}, \mathbf{D} - \mathbf{A} - \mathbf{E} \rangle + \frac{\rho}{2}\|\mathbf{D} - \mathbf{A} - \mathbf{E}\|_F^2 \quad (16)$$

因此, 每次求解优化问题 (16) 并根据求解结果, 用 (15) 更新对拉格朗日乘子的估计值  $\hat{\mathbf{Y}}$ , 逐步增大  $\rho$ , 直至满足约束  $\mathbf{D} - \mathbf{A} - \mathbf{E} = \mathbf{0}$ , 便能够得到原问题 (1) 的解

根据定理 2, 我们只需要求解无约束优化问题 (12), 采用坐标下降 (Coordinate Descent) 的方法, 分别求解关于  $\mathbf{A}$  和  $\mathbf{E}$  的优化子问题:

固定  $\mathbf{A}$  时的优化子问题:

$$\begin{aligned} & \min_{\mathbf{E}} \lambda\|\mathbf{E}\|_1 + \langle \hat{\mathbf{Y}}, \mathbf{D} - \mathbf{A} - \mathbf{E} \rangle + \frac{\rho}{2}\|\mathbf{D} - \mathbf{A} - \mathbf{E}\|_F^2 \\ &= \min_{\mathbf{E}} \lambda\|\mathbf{E}\|_1 + \text{tr}(\hat{\mathbf{Y}}^\top(\mathbf{D} - \mathbf{A} - \mathbf{E})) + \frac{\rho}{2}\text{tr}((\mathbf{D} - \mathbf{A} - \mathbf{E})^\top(\mathbf{D} - \mathbf{A} - \mathbf{E})) \\ &= \min_{\mathbf{E}} \lambda\|\mathbf{E}\|_1 + \frac{\rho}{2}\text{tr}((\mathbf{D} - \mathbf{A} - \mathbf{E})^\top(\mathbf{D} - \mathbf{A} - \mathbf{E}) + \frac{2\hat{\mathbf{Y}}^\top}{\rho}(\mathbf{D} - \mathbf{A} - \mathbf{E})) \\ &= \min_{\mathbf{E}} \lambda\|\mathbf{E}\|_1 + \frac{\rho}{2}\text{tr}((\mathbf{D} - \mathbf{A} - \mathbf{E} + \frac{\hat{\mathbf{Y}}}{\rho})^\top(\mathbf{D} - \mathbf{A} - \mathbf{E} + \frac{\hat{\mathbf{Y}}}{\rho})) \\ &= \min_{\mathbf{E}} \frac{\lambda}{\rho}\|\mathbf{E}\|_1 + \frac{1}{2}\|\mathbf{E} - (\mathbf{D} - \mathbf{A} + \frac{\hat{\mathbf{Y}}}{\rho})\|_F^2 \end{aligned} \quad (17)$$

固定  $\mathbf{E}$  时的优化子问题:

$$\begin{aligned} & \min_{\mathbf{A}} \|\mathbf{A}\|_* + \langle \hat{\mathbf{Y}}, \mathbf{D} - \mathbf{A} - \mathbf{E} \rangle + \frac{\rho}{2}\|\mathbf{D} - \mathbf{A} - \mathbf{E}\|_F^2 \\ &= \min_{\mathbf{A}} \|\mathbf{A}\|_* + \text{tr}(\hat{\mathbf{Y}}^\top(\mathbf{D} - \mathbf{A} - \mathbf{E})) + \frac{\rho}{2}\text{tr}((\mathbf{D} - \mathbf{A} - \mathbf{E})^\top(\mathbf{D} - \mathbf{A} - \mathbf{E})) \\ &= \min_{\mathbf{A}} \|\mathbf{A}\|_* + \frac{\rho}{2}\text{tr}((\mathbf{D} - \mathbf{A} - \mathbf{E})^\top(\mathbf{D} - \mathbf{A} - \mathbf{E}) + \frac{2\hat{\mathbf{Y}}^\top}{\rho}(\mathbf{D} - \mathbf{A} - \mathbf{E})) \\ &= \min_{\mathbf{A}} \|\mathbf{A}\|_* + \frac{\rho}{2}\text{tr}((\mathbf{D} - \mathbf{A} - \mathbf{E} + \frac{\hat{\mathbf{Y}}}{\rho})^\top(\mathbf{D} - \mathbf{A} - \mathbf{E} + \frac{\hat{\mathbf{Y}}}{\rho})) \\ &= \min_{\mathbf{A}} \frac{1}{\rho}\|\mathbf{A}\|_* + \frac{1}{2}\|\mathbf{A} - (\mathbf{D} - \mathbf{E} + \frac{\hat{\mathbf{Y}}}{\rho})\|_F^2 \end{aligned} \quad (18)$$

根据定理 1, 优化问题 (17) 和 (18) 作为近端极小值问题, 它的解为

$$\begin{aligned} \hat{\mathbf{A}} &= \mathcal{D}_{\frac{1}{\rho}}(\mathbf{D} - \mathbf{E} + \frac{\hat{\mathbf{Y}}}{\rho}) \\ \hat{\mathbf{E}} &= \mathcal{S}_{\frac{\lambda}{\rho}}(\mathbf{D} - \mathbf{A} + \frac{\hat{\mathbf{Y}}}{\rho}) \end{aligned} \quad (19)$$

根据上述结论，采用增广拉格朗日法求解 RPCA 问题的伪代码如下：

---

**Algorithm 2:** Augmented Lagrange multiplier for RPCA

---

**Input:** Raw Matrix  $\mathbf{D}$

**Output:** Result  $\mathbf{A} \mathbf{E}$

```

1 Initialize  $\mathbf{A} \mathbf{E} \hat{\mathbf{Y}} \rho$ 
2 Set hyper parameter  $\mu = 1.5 \delta = 1e^{-8}$ 
3 Criterion  $\leftarrow \|\mathbf{D} - \mathbf{A} - \mathbf{E}\|_F^2$ 
4 while Criterion  $> \delta$  do
5    $\mathbf{A} \leftarrow \mathcal{D}_{\frac{1}{\rho}}(\mathbf{D} - \mathbf{E} + \frac{\hat{\mathbf{Y}}}{\rho})$ 
6    $\mathbf{E} \leftarrow \mathcal{S}_{\frac{\lambda}{\rho}}(\mathbf{D} - \mathbf{A} + \frac{\hat{\mathbf{Y}}}{\rho})$ 
7    $\hat{\mathbf{Y}} \leftarrow \hat{\mathbf{Y}} + \rho(\mathbf{D} - \mathbf{A} - \mathbf{E})$ 
8    $\rho \leftarrow \mu\rho$ 
9   Criterion  $\leftarrow \|\mathbf{D} - \mathbf{A} - \mathbf{E}\|_F^2$ 
10 end

```

---

经过数值实验，以上算法求解速度快于梯度下降算法。

## 4. 矩阵的低秩表示与优化方法

由 RPCA 问题的约束条件  $\mathbf{D} = \mathbf{A} + \mathbf{E}$ ，该模型假设数据来自单一的低秩子空间。当数据来自多个低秩子空间时，RPCA 模型假设生成数据的低秩子空间是若干独立低秩子空间的并。(1) 本身很难保证恢复上述独立低秩子空间，因此由 RPCA 模型得到的低秩矩阵  $\mathbf{A}$  存在近似误差<sup>[2]</sup>。

矩阵的低秩表示 (Low Rank Representation, LRR) 能够解决上述问题。其约束条件  $\mathbf{D} = \mathbf{Z}\mathbf{A} + \mathbf{E}$  中包含字典矩阵  $\mathbf{Z}$ ， $R(\mathbf{A})$  张成产生原始数据的空间，由于  $\text{rank}(\mathbf{Z}\mathbf{A}) \leq \text{rank}(\mathbf{A})$ ，故  $\mathbf{Z}\mathbf{A}$  可看作对原始数据  $\mathbf{D}$  的低秩恢复，即字典矩阵各列重新线性组合。注意 RPCA 可看作 LRR 在字典矩阵  $\mathbf{Z} = \mathbf{I}$  时的特例。

与 RPCA 问题类似，A 通过在目标函数中引入  $\text{rank}(\mathbf{A})$  惩罚低秩矩阵的秩，并松弛为  $\|\mathbf{A}\|_*$ 。LRR 通过  $2,1-$  范数  $\|\mathbf{E}\|_{2,1}$  惩罚噪声矩阵  $\mathbf{E}$ 。若假设噪声向量为稠密形式，而非某些维度接近 0 的稀疏解，对列向量  $\mathbf{e}_i, i = 1, \dots, n$  取  $2-$  范数，保证算法对噪声的鲁棒性。此外，LRR 假设原始数据的误差与样本相关，即仅部分数据向量受噪声污染，因此对列向量  $2-$  范数结果再取  $1-$  范数，最终得到  $2,1-$  范数。若假设噪声向量为稀疏形式，部分维度接近 0，则  $\|\mathbf{E}\|_{2,1}$  可通过  $\|\mathbf{E}\|_1$  代替<sup>[2]</sup>。

**定义 2 (矩阵的低秩表示)** 通过求解优化问题

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{E}} \quad & \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_{2,1} \\ \text{s.t.} \quad & \mathbf{D} = \mathbf{Z}\mathbf{A} + \mathbf{E} \end{aligned} \tag{20}$$

的最优解  $\{\mathbf{A}^*, \mathbf{E}^*\}$ ，我们可以得到对原始数据  $\mathbf{D}$  的低秩表示  $\mathbf{Z}\mathbf{A}^* = \mathbf{D} - \mathbf{E}^*$ 。

问题 (20) 是凸的, 为推导其优化算法, 首先引入辅助变量  $\mathbf{B} \in \mathbb{R}^{n \times n}$ , 将其化为等价形式

$$\begin{aligned} & \min_{\mathbf{A}, \mathbf{B}, \mathbf{E}} \|\mathbf{B}\|_* + \lambda \|\mathbf{E}\|_{2,1} \\ \text{s.t. } & \mathbf{D} = \mathbf{Z}\mathbf{A} + \mathbf{E} \\ & \mathbf{A} = \mathbf{B} \end{aligned} \quad (21)$$

字典矩阵  $\mathbf{Z}$  的选取不唯一, 且影响低秩恢复效果, 一种可行的方案是  $\mathbf{Z} = \mathbf{D}$ , 即考虑  $R(\mathbf{D})$  作为产生数据的空间, 注意其隐含  $\mathbf{D}$  未受噪声污染。为便于求解, 引入辅助变量  $\mathbf{X} \in \mathbb{R}^{m \times n}$ , 将此条件考虑在约束中, (21) 可进一步化为

$$\begin{aligned} & \min_{\mathbf{A}, \mathbf{B}, \mathbf{X}, \mathbf{E}} \|\mathbf{B}\|_* + \lambda \|\mathbf{E}\|_{2,1} \\ \text{s.t. } & \mathbf{D} = \mathbf{X}\mathbf{A} + \mathbf{E} \\ & \mathbf{A} = \mathbf{B}, \mathbf{X} = \mathbf{D} \end{aligned} \quad (22)$$

与 (14) 类似, 通过增广拉格朗日方法优化原问题 (21)。相应无约束优化问题的目标函数为

$$\begin{aligned} L(\mathbf{A}, \mathbf{B}, \mathbf{E}, \mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3) = & \|\mathbf{B}\|_* + \lambda \|\mathbf{E}\|_{2,1} \\ & + \langle \mathbf{Y}_1, \mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E} \rangle + \langle \mathbf{Y}_2, \mathbf{A} - \mathbf{B} \rangle + \langle \mathbf{Y}_3, \mathbf{X} - \mathbf{D} \rangle \\ & + \frac{\rho}{2} (\|\mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E}\|_F^2 + \|\mathbf{A} - \mathbf{B}\|_F^2 + \|\mathbf{X} - \mathbf{D}\|_F^2) \end{aligned} \quad (23)$$

通过交替方向乘子法 (Alternative Direction Multiplier Method, ADMM) 求解上述可分离的凸优化问题 (22)。

当  $\mathbf{B}$  为未知量, 其他矩阵为已知量时的优化子问题:

$$\begin{aligned} & \arg \min_{\mathbf{B}} \|\mathbf{B}\|_* + \langle \mathbf{Y}_2, \mathbf{A} - \mathbf{B} \rangle + \frac{\rho}{2} \|\mathbf{A} - \mathbf{B}\|_F^2 \\ & = \arg \min_{\mathbf{B}} \|\mathbf{B}\|_* + \langle \mathbf{Y}_2, \mathbf{A} - \mathbf{B} \rangle + \frac{\rho}{2} \langle \mathbf{A} - \mathbf{B}, \mathbf{A} - \mathbf{B} \rangle + \frac{1}{2\rho} \langle \mathbf{Y}_2, \mathbf{Y}_2 \rangle \\ & = \arg \min_{\mathbf{B}} \|\mathbf{B}\|_* + \frac{\rho}{2} \langle \mathbf{A} - \mathbf{B} + \frac{1}{\rho} \mathbf{Y}_2, \mathbf{A} - \mathbf{B} + \frac{1}{\rho} \mathbf{Y}_2 \rangle \\ & = \arg \min_{\mathbf{B}} \frac{1}{\rho} \|\mathbf{B}\|_* + \frac{1}{2} \|\mathbf{B} - \left( \mathbf{A} + \frac{1}{\rho} \mathbf{Y}_2 \right)\|_F^2 \end{aligned} \quad (24)$$

根据 (7a) 可知以上优化问题的解

$$\hat{\mathbf{B}} = \mathcal{D}_{\frac{1}{\rho}} \left( \mathbf{A} + \frac{1}{\rho} \mathbf{Y}_2 \right) \quad (25)$$

当  $\mathbf{A}$  为未知量, 其他矩阵为已知量时的优化子问题:

$$\begin{aligned} & \arg \min_{\mathbf{A}} \langle \mathbf{Y}_1, \mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E} \rangle + \langle \mathbf{Y}_2, \mathbf{A} - \mathbf{B} \rangle + \frac{\rho}{2} (\|\mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E}\|_F^2 + \|\mathbf{A} - \mathbf{B}\|_F^2) \\ & = \arg \min_{\mathbf{A}} \langle \mathbf{Y}_1, \mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E} \rangle + \frac{\rho}{2} \|\mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E}\|_F^2 + \frac{1}{2\rho} \langle \mathbf{Y}_1, \mathbf{Y}_1 \rangle \\ & \quad + \langle \mathbf{Y}_2, \mathbf{A} - \mathbf{B} \rangle + \frac{\rho}{2} \|\mathbf{A} - \mathbf{B}\|_F^2 + \frac{1}{2\rho} \langle \mathbf{Y}_2, \mathbf{Y}_2 \rangle \\ & = \arg \min_{\mathbf{A}} \frac{1}{\rho} \|\mathbf{Y}_1 + \mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E}\|_F^2 + \frac{1}{\rho} \|\mathbf{Y}_2 + \mathbf{A} - \mathbf{B}\|_F^2 \end{aligned} \quad (26)$$

令上式对  $\mathbf{A}$  的微分

$$\begin{aligned} & -2\mathbf{X}^T \left( \frac{1}{\rho} \mathbf{Y}_1 + \mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E} \right) + 2 \left( \frac{1}{\rho} \mathbf{Y}_2 + \mathbf{A} - \mathbf{B} \right) \\ & = 2(\mathbf{I} + \mathbf{X}^T \mathbf{X}) \mathbf{A} - 2\mathbf{X}^T \left( \frac{1}{\rho} \mathbf{Y}_1 + \mathbf{D} - \mathbf{E} \right) + 2 \left( \frac{1}{\rho} \mathbf{Y}_2 - \mathbf{B} \right) = 0 \end{aligned}$$

可得优化子问题 (30) 的解

$$\hat{\mathbf{A}} = (\mathbf{I} + \mathbf{X}^T \mathbf{X})^{-1} \left( \mathbf{X}^T \left( \frac{1}{\rho} \mathbf{Y}_1 + \mathbf{D} - \mathbf{E} \right) + \left( \mathbf{B} - \frac{1}{\rho} \mathbf{Y}_2 \right) \right) \quad (27)$$

当  $\mathbf{E}$  为未知量, 其他矩阵为已知量时的优化子问题:

$$\begin{aligned} & \arg \min_{\mathbf{E}} \lambda \|\mathbf{E}\|_{2,1} + \langle \mathbf{Y}_1, \mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E} \rangle + \frac{\rho}{2} \|\mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E}\|_F^2 \\ & = \arg \min_{\mathbf{E}} \lambda \|\mathbf{E}\|_{2,1} + \langle \mathbf{Y}_1, \mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E} \rangle + \frac{\rho}{2} \|\mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E}\|_F^2 + \frac{1}{2\rho} \langle \mathbf{Y}_1, \mathbf{Y}_1 \rangle \\ & = \arg \min_{\mathbf{E}} \frac{\lambda}{\rho} \|\mathbf{E}\|_{2,1} + \frac{1}{2} \|\mathbf{E} - \left( \mathbf{D} - \mathbf{X}\mathbf{A} + \frac{1}{\rho} \mathbf{Y}_1 \right)\|_F^2 \end{aligned} \quad (28)$$

根据 (7a) 可知以上优化问题的解

$$\hat{\mathbf{E}} = \mathcal{W}_{\frac{\lambda}{\rho}} \left( \mathbf{D} - \mathbf{X}\mathbf{A} + \frac{1}{\rho} \mathbf{Y}_1 \right) \quad (29)$$

当  $\mathbf{X}$  为未知量, 其他矩阵为已知量时的优化子问题:

$$\begin{aligned} & \arg \min_{\mathbf{D}} \langle \mathbf{Y}_1, \mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E} \rangle + \langle \mathbf{Y}_3, \mathbf{X} - \mathbf{D} \rangle + \frac{\rho}{2} (\|\mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E}\|_F^2 + \|\mathbf{X} - \mathbf{D}\|_F^2) \\ & = \arg \min_{\mathbf{D}} \langle \mathbf{Y}_1, \mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E} \rangle + \frac{\rho}{2} \|\mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E}\|_F^2 + \frac{1}{2\rho} \langle \mathbf{Y}_1, \mathbf{Y}_1 \rangle \\ & \quad + \langle \mathbf{Y}_3, \mathbf{X} - \mathbf{D} \rangle + \frac{\rho}{2} \|\mathbf{X} - \mathbf{D}\|_F^2 + \frac{1}{2\rho} \langle \mathbf{Y}_3, \mathbf{Y}_3 \rangle \\ & = \arg \min_{\mathbf{D}} \frac{1}{\rho} \|\mathbf{Y}_1 + \mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E}\|_F^2 + \frac{1}{\rho} \|\mathbf{Y}_3 + \mathbf{X} - \mathbf{D}\|_F^2 \end{aligned} \quad (30)$$

将目标函数化简为迹的形式, 其中

$$\begin{aligned} & \frac{1}{\rho} \|\mathbf{Y}_1 + \mathbf{D} - \mathbf{X}\mathbf{A} - \mathbf{E}\|_F^2 \\ & = \text{tr} \left( (\mathbf{X}\mathbf{A})^T (\mathbf{X}\mathbf{A}) \right) + \text{tr} \left( - \left( \frac{1}{\rho} \mathbf{Y}_1 + \mathbf{D} - \mathbf{E} \right)^T \mathbf{X}\mathbf{A} \right) + \text{tr} \left( \left( \frac{1}{\rho} \mathbf{Y}_1 + \mathbf{D} - \mathbf{E} \right)^T \left( \frac{1}{\rho} \mathbf{Y}_1 + \mathbf{D} - \mathbf{E} \right) \right), \\ & \frac{1}{\rho} \|\mathbf{Y}_3 + \mathbf{X} - \mathbf{D}\|_F^2 \\ & = \text{tr} ((\mathbf{X}^T \mathbf{X})^T) + \text{tr} \left( \left( \frac{1}{\rho} \mathbf{Y}_3 - \mathbf{D} \right)^T \mathbf{X} \right) + \text{tr} \left( \left( \frac{1}{\rho} \mathbf{Y}_3 - \mathbf{D} \right)^T \left( \frac{1}{\rho} \mathbf{Y}_3 - \mathbf{D} \right) \right). \end{aligned} \quad (31)$$

令上式对  $\mathbf{X}$  的微分

$$2\mathbf{X}\mathbf{A}\mathbf{A}^T - \left( \frac{1}{\rho} \mathbf{Y}_1 + \mathbf{D} - \mathbf{E} \right) \mathbf{A}^T + 2\mathbf{X} + \left( \frac{1}{\rho} \mathbf{Y}_3 - \mathbf{D} \right) = 0$$

可得优化子问题 (30) 的解

$$\hat{\mathbf{X}} = \left( \left( \frac{1}{\rho} \mathbf{Y}_1 + \mathbf{D} - \mathbf{E} \right) \mathbf{A}^T + \left( \mathbf{D} - \frac{1}{\rho} \mathbf{Y}_3 \right) \right) (\mathbf{A}\mathbf{A}^T + \mathbf{I})^{-1} \quad (32)$$

根据上述结论，采用增广拉格朗日法求解 LRR 问题的伪代码如下：

---

**Algorithm 3:** Augmented Lagrange multiplier for LRR

---

**Input:** Raw Matrix  $\mathbf{D}$

**Output:** Result  $\mathbf{A}, \mathbf{B}, \mathbf{X}, \mathbf{E}$

```

1 Initialize  $\mathbf{E} = \mathbf{Y}_1 = \mathbf{Y}_3 = \mathbf{O}_{m \times n}$ ,  $\mathbf{A} = \mathbf{B} = \mathbf{Y}_2 = \mathbf{O}_{n \times n}$ ,  $\mathbf{X} = \mathbf{D}$ ,  $\rho$ 
2 Set hyper parameter  $\mu = 1.5$ ,  $\delta = 1e^{-8}$ 
3 Criterion  $\leftarrow \max(\|\mathbf{D} - \mathbf{XA} - \mathbf{E}\|_F^2, \|\mathbf{A} - \mathbf{B}\|_F^2, \|\mathbf{X} - \mathbf{D}\|_F^2)$ 
4 while Criterion  $> \delta$  do
5    $\hat{\mathbf{B}} \leftarrow \mathcal{D}_{\frac{1}{\rho}} \left( \mathbf{A} + \frac{1}{\rho} \mathbf{Y}_2 \right)$ 
6    $\hat{\mathbf{A}} = (\mathbf{I} + \mathbf{X}^T \mathbf{X})^{-1} \left( \mathbf{X}^T \left( \frac{1}{\rho} \mathbf{Y}_1 + \mathbf{D} - \mathbf{E} \right) + \left( \mathbf{B} - \frac{1}{\rho} \mathbf{Y}_2 \right) \right)$ 
7    $\hat{\mathbf{E}} = \mathcal{W}_{\frac{\lambda}{\rho}} \left( \mathbf{D} - \mathbf{XA} + \frac{1}{\rho} \mathbf{Y}_1 \right)$ 
8    $\hat{\mathbf{X}} = \left( \left( \frac{1}{\rho} \mathbf{Y}_1 + \mathbf{D} - \mathbf{E} \right) \mathbf{A}^T + \left( \mathbf{D} - \frac{1}{\rho} \mathbf{Y}_3 \right) \right) (\mathbf{AA}^T + \mathbf{I})^{-1}$ 
9    $\mathbf{Y}_1 = \mathbf{Y}_1 + \rho (\mathbf{D} - \mathbf{XA} - \mathbf{E})$ 
10   $\mathbf{Y}_1 = \mathbf{Y}_1 + \rho (\mathbf{A} - \mathbf{B})$ 
11   $\mathbf{Y}_1 = \mathbf{Y}_1 + \rho (\mathbf{X} - \mathbf{D})$ 
12   $\rho \leftarrow \mu\rho$ 
13  Criterion  $\leftarrow \max(\|\mathbf{D} - \mathbf{XA} - \mathbf{E}\|_F^2, \|\mathbf{A} - \mathbf{B}\|_F^2, \|\mathbf{X} - \mathbf{D}\|_F^2)$ 
14 end

```

---

## 5. 实验与分析

使用给定的三幅图片（如图 1 所示）测试算法效果。考虑到图片的空间关联性和各通道的相对独立性，对测试图片单独提取 R, G, B 三通道，分别作为数据矩阵  $\mathbf{D}$  输入算法，将输出的低秩矩阵和噪声矩阵分别重组得到低秩成分图与噪声图。



图 1：给定测试图片

### 5.1 不同算法性能横向比较

对三张图片分别采用 PCA, RPCA 及 LRR 处理。其中 PCA 限制低秩成分的秩  $r(\mathbf{A}) \leq 0.025 \cdot \min(m, n)$ ，其中  $m, n$  分别为矩阵的行数与列数；RPCA 与 LRR 的超参数选择  $\lambda = \frac{1}{\max(m, n)}$ ,  $\mu = 1.1$ , 系数  $\rho$  的初值  $\rho_0 = 10^{-6}$ , 最大值  $\rho_{\max} = 10^8$ , 迭代停止阈值  $\delta = 10^{-14} \cdot \|\mathbf{D}\|_F^2$ 。对三幅测试图片的实验效果展示如下。

每组图片的展示顺序为：第一行给出低秩成分图，第二行给出噪声图。从左至右三列依次为 PCA、RPCA、LRR 算法的输出。

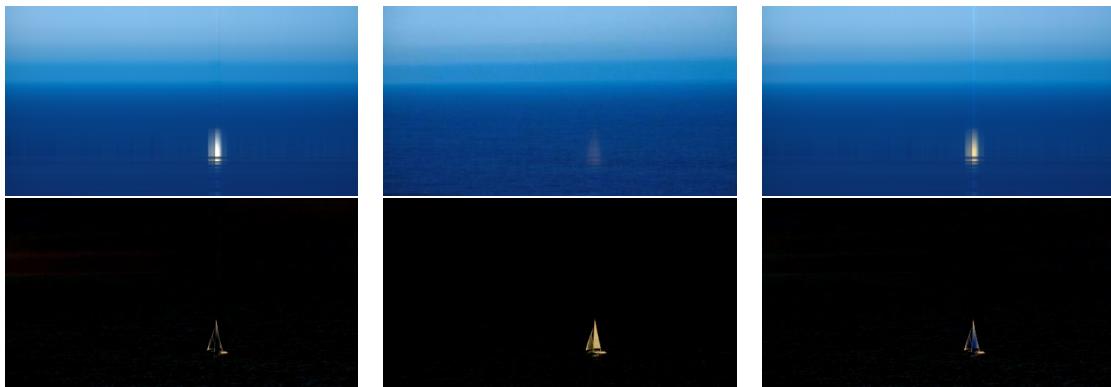


图 2: 对“大海航行”测试图片的处理效果

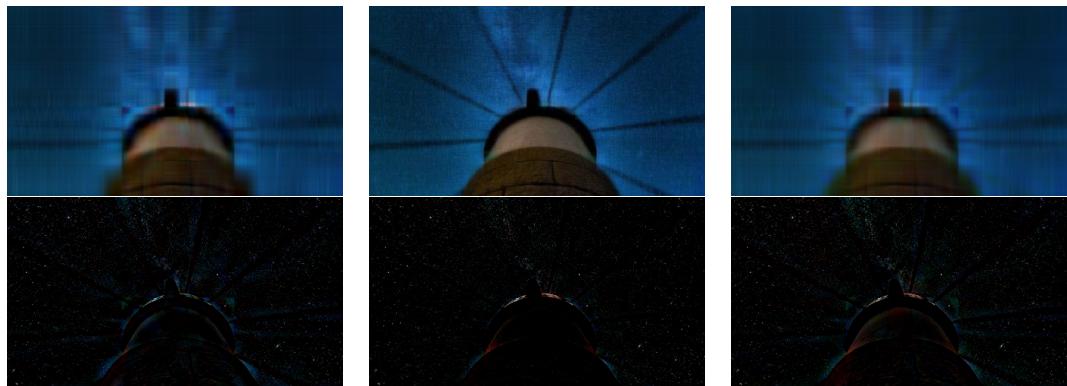


图 3: 对“星空灯塔”测试图片的处理效果

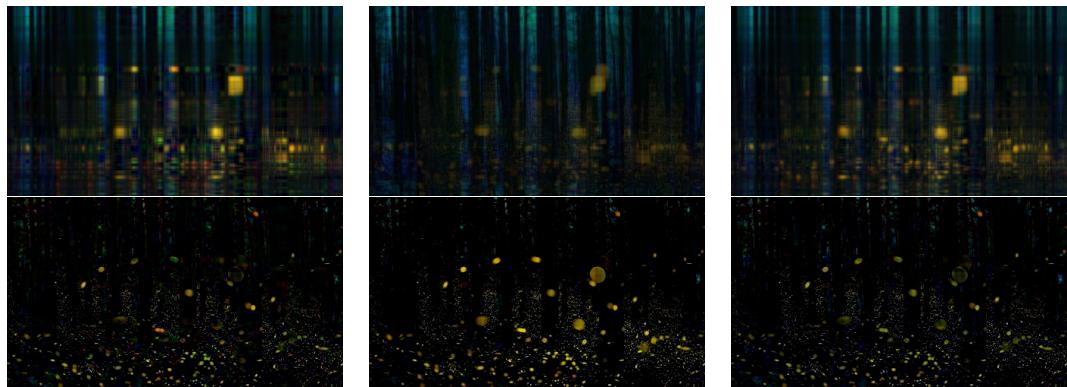


图 4: 对“幽林萤火”测试图片的处理效果

对比可见三种算法均能提取输入图像的低秩成分，表现为图像中的背景、连续的颜色或纹理，如测试图像中的海洋、天空与森林等。三种算法也能给出图像中的高频噪声，表现为图像中的亮点、亮斑与面积较小的前景，如测试图像中的繁星、萤火与帆船等。

其中 RPCA 算法效果最佳，低秩成分图清晰、细腻，对于原始图像的景物轮廓、大致纹理都能够保留，

对原始图像中的高频噪声、稀疏景物去除最为彻底，从其噪声图也可看出，噪点保留完整、轮廓清晰、具有较高的对比度。

相比之下，PCA 与 LRR 的处理效果类似，低秩成分图的轮廓与纹理被部分破坏，难以从中辨认原始图像。两种算法在去噪时对于尖锐噪声或离群点的鲁棒性不足，对高、低频成分的区分不够明显。为得到较好的去噪效果，通常低秩成分也遭到一定程度的破坏，图像呈现分块特征。如“大海航行”与“星空灯塔”两幅测试图片都含有明显的横纵向纹理，天空与海洋已模糊不清，帆船轮廓仍然可见，去噪不够彻底。“幽林萤火”测试图片仍有部分亮斑未去除，噪声成分轮廓模糊，出现色差。

## 5.2 超参数对算法性能的影响

对于 PCA 算法，超参数只有低秩成分的秩  $r$ ，对于 RPCA 与 LRR，由于均采用 ADMM 算法求解，超参数相同，其中对算法性能有明显影响的包括  $\lambda$ （对低秩矩阵范数的惩罚力度）、 $\rho_0$ （对增广拉格朗日项惩罚系数的初值）与  $\mu$ （ $\rho$  每步扩大的倍数）。纵向比较超参数对各类算法的影响时，以图 (2)~(4) 中的效果和参数为参考。本部分实验采取“幽林萤火”作为测试图片，因为其低秩成分图和噪声图在三幅测试图片中最为清晰可辨，容易对比。

首先展示超参数  $r = k \cdot \min(m, n)$  中的  $k$  对 PCA 的影响，实验结果如下：

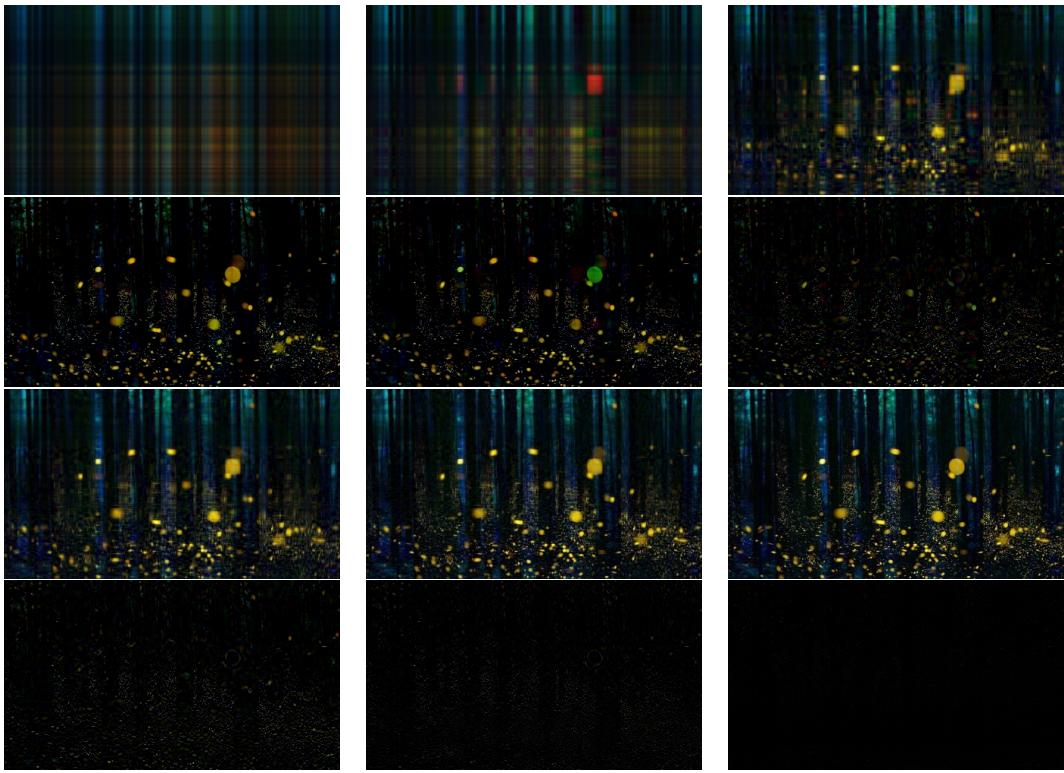


图 5：PCA 调整  $k$  的影响

前两行对应  $k = 0.005, 0.01, 0.05$ ，后两行对应  $k = 0.1, 0.2, 0.5$ ，第 1、3 行为低秩成分图，第 2、4 行为噪声图。可见调小  $r$  能够增加低秩成分的稀疏程度，更多地保留北京，过滤掉前景，当  $k \leq 0.01$  时，低秩成

分已基本模糊难辨,  $k \geq 0.2$  时, 噪声图几乎为全黑,  $0.01 < k < 0.2$  时低秩提取效果较好。

随后展示超参数对基于 ADMM 优化方法的影响, 鉴于 RPCA 与 LRR 具有相同的超参数, 只是取值范围有差别, 故本部分分析以 RPCA 为例, 最后集中给出 LRR 的实验效果。本部分实验只给出低秩成分, 噪声图所呈现的规律可以推理得出。

调整  $\lambda$  的实验结果如下:



图 6: RPCA 调整  $\lambda$  的影响

从左到右分别对应  $\lambda = \frac{i}{\max(m,n)}$ ,  $i = 1.0, 2.0, 4.0$ , 由于  $\lambda$  是目标函数中, 惩罚噪声矩阵范数的因子, 因此调大  $\lambda$ , 噪声矩阵越稀疏, 低秩成分对原始图像保留越完整, 图像越清晰。

调整  $\mu$  的实验结果如下:

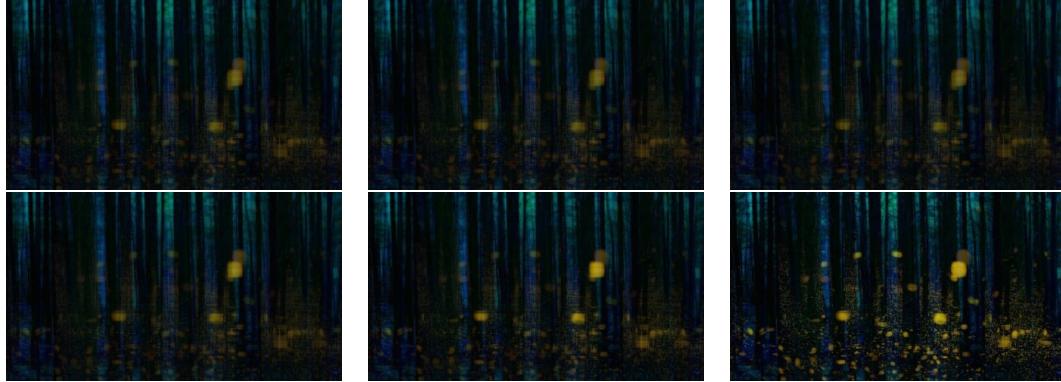
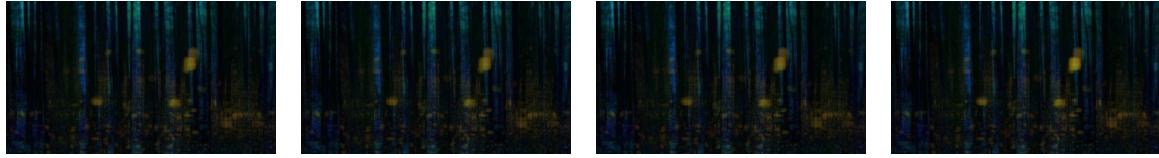


图 7: RPCA 调整  $\mu$  的影响

从左到右, 从上到下对应  $\mu = 1.01, 1.05, 1.1, 1.5, 2.0, 4.0$ 。增大  $\mu$  会导致惩罚项因子  $\rho$  上升较快, 算法收敛速度提高, 但目标函数迭代下降不够充分, 所得解远离目标值, 效果上表现为低秩矩阵稀疏度不足, 保留过多原始图像中的高频噪声。

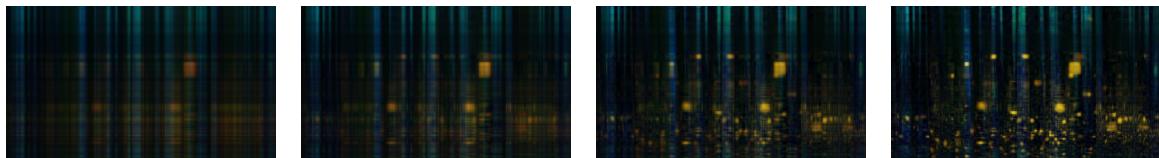
调整  $\rho_0$  的实验结果如下:

图 8: RPCA 调整  $\rho_0$  的影响

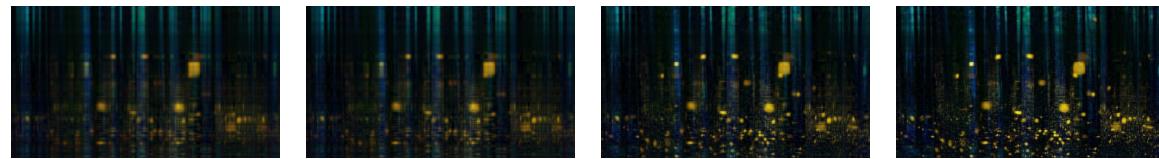
从左到右对应  $\rho_0 = 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}$ 。只要初值  $\rho_0$  不过大，调整  $\rho_0$  不会对算法收敛速度和目标值造成明显影响。

下面给出同样的超参数对 LRR 的影响。

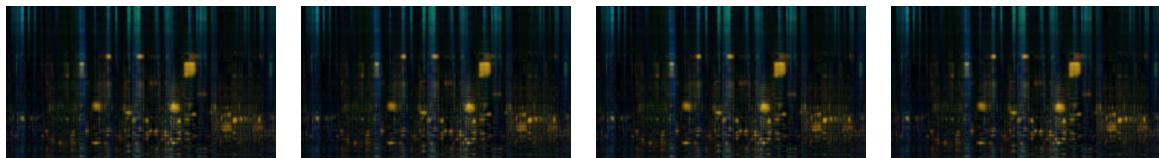
调整  $\lambda$  的实验结果如下，从左到右分别对应  $\lambda = \frac{i}{\max(m,n)}, i = 0.25, 0.5, 1.0, 2.0, :$

图 9: LRR 调整  $\lambda$  的影响

调整  $\mu$  的实验结果如下，从左到右对应  $\mu = 1.05, 1.1, 1.5, 2.0, :$

图 10: LRR 调整  $\mu$  的影响

调整  $\rho_0$  的实验结果如下，从左到右对应  $\rho_0 = 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, :$

图 11: LRR 调整  $\rho_0$  的影响

总体而言，LRR 的收敛速度要慢于 RPCA，这与其约束条件更复杂，交替下降方向更多，迭代终止条件需要检查的矩阵范数更多有关。从恢复效果而言，RPCA 要优于 LRR，推测是由于图像矩阵具有较强空间关联性。RPCA 约束条件  $D = A + E$  是矩阵的直接分解，而 LRR 约束条件  $D = XA + E$  是对列空

间的重新组合，RPCA 相比 LRR 能够在行空间也进行低秩运算，能够较好地保留图像矩阵的空间结构。此外，LRR 对于噪声矩阵以  $2, 1$ -范数呈现在目标函数中，本意是将各列  $e_j$  当作样本矩阵  $\mathbf{D}$  的噪声向量，但对于图片而言，噪声是以和原始图像同样大小的矩阵叠加的，因此 LRR 的模型在图像处理上不如 RPCA 合理。

## 6. 附加说明

- (a) 两位小组成员对大作业的贡献比例为 1:1。其中，陈煜负责 RPCA 通用优化算法与特殊优化算法的设计、推导，以及梯度下降优化的实验。姜永鹏负责 LRR 的设计、推导，以及 RPCA 和 LRR 的实验，以及算法的对比分析。
- (b) 代码运行说明：请将清华云盘提供的三张图片 1.jpg, 2.jpg, 3.jpg 与代码文件 PCA.py, RPCA\_lm.py, RPCA\_lrr.py，在主函数最前部能够修改读入图片的文件名，修改后运行 python 脚本即可。

**REFERENCES**

- [1] Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9:717–772, 2009.
- [2] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2013.