

分类号: TP391.4

UDC: 510

学 号: 17451082117

密级: 公开级

# 温州大学

## 硕士学位论文



题 目: 基于  $L_0$  范数非凸替代方法  
低秩恢复及应用

作者姓名: 郑晶晶

培养类型: 学术型

专业名称: 应用数学

研究方向: 智能系统与控制

指导教师: 张笑钦(教授)

完成日期: 2020年5月30日

温州大学学位委员会



## 温州大学学位论文独创性声明

本人郑重声明：所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得温州大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

论文作者签名：

日期： 年 月 日

## 温州大学学位论文使用授权声明

本人完全了解温州大学关于收集、保存、使用学位论文的规定，即：学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权温州大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。本人在导师指导下完成的论文成果，知识产权归属温州大学。

保密论文在解密后遵守此规定。

论文作者签名：

导师签名：

日期： 年 月 日

日期： 年 月 日



# 基于 $L_0$ 范数非凸替代方法的低秩恢复及其应用

## 摘 要

低秩矩阵恢复问题作为近几年来研究的热点，被广泛地应用于数据降维及数据分析等问题中。低秩矩阵恢复问题，即从带有噪声的观测矩阵中恢复低秩矩阵以找到数据所分布的低维子空间。由于原始低秩恢复模型涉及  $L_0$  范数和秩函数，这使得多数模型都是 NP 难的。一种常见的解决方法是将原问题近似成为凸优化问题进行求解。对于大部分的单子空间低秩模型，其精确恢复性已在理论上被证明：当某些条件被满足时，基于凸优化方法可精确恢复问题的真实解。然而，在现实生活中，这些严格的条件往往难以被满足。这意味着基于凸方法可能无法得到精确的低秩稀疏解。为了权衡模型的有效性和可解性，本论文主要研究两类非凸低秩矩阵恢复框架的通用算法，并讨论了基于非凸替代方法的低秩恢复模型在异常识别问题和图像去噪上的应用。现总结文章的主要贡献与创新如下：

(1) 文章首先给出了一个新的不动点算法 (GAI 算法)，并在理论上证明了 GAI 算法的超线性收敛于一个关键问题的全局最优解。结合广义近端梯度算法和 GAI 算法，给出了第一类低秩恢复框架的一个通用算法。文章从实验和理论上证明了：相比传统算法而言，本文所给算法更高效。

(2) 对于第二类低秩恢复框架，文章利用块坐标下降算法的一种变种算法将原始非凸框架转化为两个非凸子问题。文章还给出一个重要定理以用于非凸子问题的求解，并由此得到了第二类低秩恢复框架的一个通用算法。文章从实验和理论出发证明了该算法的有效性。另外由实验结果也能看出，相比较凸方法而言，基于非凸替代的低秩恢复能够在异常识别问题中得到更优的结果。

(3) 本文利用参数化方法给出一种新的非凸替代方法 (参数化范数和参数化核范数)，并利用所给的非凸替代得到一个新颖的非凸低秩恢复模型。实验证明了非凸低秩恢复模型在图像去噪问题中的有效性。

**关键词：**低秩矩阵恢复, 非局部图像去噪算法, 非凸替代, 不动点



# LOW RANK RECOVERY BASED ON $L_0$ NORM NON-CONVEX SURROGATE METHODS AND ITS APPLICATION

## ABSTRACT

In recent years, low rank recovery problems are getting more attention, and widely used for dimension reduction and data analysis. Low rank recovery is proposed to exploit low-dimensional structure in high-dimensional data. Because the original low rank recovery problems are involved with rank function and  $L_0$  norm. This causes that the most of models are NP-hard. A common solution is to use the corresponding convexly relaxed problems to approximate the original problems for effective solving. For most single subspace low rank models, its exact recovery has been proved theoretically: if some conditions are satisfied, the convex optimization methods can exactly recover the real solution of the problems. However, since these strictly conditions are hardly met in reality, which means that the solution by convex problem can be far away from the true value. Therefore, to balance the validity and solvability of the model, general solver for two types of low rank recovery frames are studied. And low rank recovery method based on non-convex surrogate and its related applications in outlier pursuit and image denoising problems are discussed in this work. The main contributions of this paper are listed in the following:

(1) A new fixed-point algorithm (GAI) is given for solving a key optimization problem to global optimality with superlinear convergence rate. Based on generalized proximal gradient and GAI, a general solver for first kind of low rank recovery frame is given in this paper. It is proven theoretically and experimentally that the proposed algorithm is more efficient compared with traditional method.

(2) For second kind of low rank recovery frame, this paper presents a

variant algorithm of block coordinate descent algorithm. Based on that, the original frame is splited into two non-convex subproblems. Besides, an important theorem is given to solve the non-convex subproblems. Therefore, the general solver for the second kind of low rank recovery frame is obtained. The validity of the proposed solver is proved theoretically and experimentally. In addition, it can be seen from the experiment that low rank recovery methods based on non-convex surrogate achieve a better results compared with the convex method in the outlier persuit problems.

(3) This paper gives a novel non-convex surrogate method using parameterize method (include parameterized norm and parameterized nuclear norm), and a new non-convex model is obtained based on the two norms. Experiments illustrate the effectiveness of the proposed approach in the image denoising problems.

**KEY WORDS:** Low rank matrix recovery, Image denoising algorithm based on non-local method, Non-convex surrogate, Fixed-point algorithm



# 目录

摘要	I
ABSTRACT	III
第一章 绪 论	1
1.1 研究背景及意义 . . . . .	1
1.2 国内外研究现状 . . . . .	2
1.3 论文的选题和主要工作 . . . . .	4
1.4 本文的结构安排 . . . . .	5
第二章 相关工作	7
2.1 低秩矩阵恢复模型 . . . . .	7
2.2 基于非局部相似块的图像去噪算法介绍 . . . . .	9
2.2.1 非局部均值算法 . . . . .	9
2.2.2 基于块的 SVD 的图像去噪方法 . . . . .	10
2.2.3 基于同步稀疏编码方法 . . . . .	10
第三章 不动点算法	13
3.1 不动点算法及其理论分析 . . . . .	13
3.2 数值实验与分析 . . . . .	17
第四章 第二类低秩矩阵恢复框架的解法以及应用	19
4.1 第二类低秩恢复框架 . . . . .	19

4.2	块坐标下降算法的一种变种算法及其理论分析 .	19
4.3	实验结果与分析 . . . . .	23
4.3.1	异常识别 . . . . .	23
<b>第五章</b>	<b>基于非凸低秩恢复模型的非局部图像去噪方法</b>	<b>29</b>
5.1	基于 $L_0$ 范数的参数化非凸替代的低秩恢复模型	30
5.2	优化算法 . . . . .	32
5.3	基于非凸低秩恢复的非局部图像去噪算法 . .	35
5.4	实验结果 . . . . .	35
5.4.1	零均值高斯噪声 . . . . .	36
5.4.2	零均值高斯脉冲混合噪声 . . . . .	37
<b>第六章</b>	<b>论文小结与展望</b>	<b>41</b>
6.1	论文小结 . . . . .	41
6.2	展望 . . . . .	42
	<b>参考文献</b>	<b>43</b>
	<b>附录 A</b>	<b>47</b>
	<b>致谢</b>	<b>55</b>
	<b>攻读硕士期间的科研成果</b>	<b>57</b>

# 第一章 绪论

## 1.1 研究背景及意义

随着计算机科学技术与互联网的蓬勃发展，图像、视频等成为人们从外界获取信息的主要渠道以及日常生活与工作最常用的信息载体。我们每天都能获得前人无法想象的海量数据。这些数据中虽然含有丰富的信息可供利用，但数据量的庞大也加大了分析处理与传输的难度。幸运的是，高维数据之间通常存在较大的相关性。这种相关性可表现为两种不同类型的稀疏结构：(1) 一阶向量的稀疏性，即信号在某个变换域下的表示系数稀疏或是近似稀疏 [1–3] (这时系数向量元素大部分为“0”或是近似为“0”)；(2) 二阶矩阵的低秩性，即信号对应的矩阵为低秩或是近似低秩 [4–6] (这时矩阵的奇异值大部分为“0”或是近似为“0”)。另外，由于大规模数据在获取和传输过程中受到人为或是自然的干扰，常常出现缺失、损坏的情况。这使得很多传统的分析与处理手段失效。因此，研究如何从不完整或带有损毁的数据中精确且高效地恢复原有数据成为一个重要问题。E. Candès 等人于 2008 年提出的矩阵填充理论以用于矩阵填充问题 (Matrix completion, MC) 的解决：已知某矩阵在某些位置的值，如何恢复该矩阵？该理论被广泛应用于许多实际问题，其中包括 Netflix 挑战、量子态断层成像等。矩阵填充理论的广泛应用使得近年又重新涌现出大量低秩模型。这些低秩模型被广泛的应用于计算机视觉问题中 [5, 7–13]。

注意到矩阵的低秩性可以看作是向量稀疏性概念的扩展。作为稀疏性最自然的刻画方式，由于  $L_0$  范数 (系数向量中所有非零元素的个数，即  $\|x\|_0 = \sum_i |x_i|^0$ ，其中当  $x_i = 0$  时  $|x_i|^0$  被定义为 0 否则  $|x_i|^0 = 1$ ) 的非凸非连续性常常导致其相关优化问题是 NP 难的。一种常用的解决方法是：利用  $L_0$  范数的凸包络  $L_1$  范数 (系数向量中所有非零元素的绝对之和，即  $\|x\|_1 = \sum_i |x_i|$ ) 来替换模型中的  $L_0$  范数，并通过求解相应的凸问题以得到想要的稀疏解。类似于一维向量的场景，核范数 (矩阵奇异值向量的  $L_1$  范数)，秩函数的凸包络 (矩阵奇异值向量的  $L_0$  范数)，常常被用来替换模型中的秩函数以解决原优化模型为 NP 难的问题。尽管现存部分基于凸方法的低秩恢复模型的精确可恢复性已经在理论上被证明，即当某些条件被满足时，通过求解凸优化问题能够得到真实的解。然而在实际应用中，这些严格的前提条件往往难以被满足。这意味着由凸优化得到的解可能会严重偏离问题的

表 1-1:  $L_0$  范数的非凸替代。Table 1-1: Some non-convex surrogate functions of  $L_0$  norm.

Name	$g(x), x \geq 0$
$L_p$	$x^p, 0 < p < 1$
LSP	$\log(1 + \frac{x}{\gamma})$
Laplace	$(1 - \exp(-\frac{x}{\gamma}))$
LOG	$\log(\gamma + x)$
Logarithm	$\frac{1}{\log(\gamma+1)} \log(\gamma x + 1)$
ETP	$\frac{1 - \exp(-\gamma x)}{1 - \exp(-\gamma)}$

真实解。

## 1.2 国内外研究现状

为了解决上述问题，研究者们给出了一些非凸替换函数  $g(x)$  (见表格1-1) 来取代  $\|x\|_0 = \sum_i |x_i|^0$  或是  $\text{rank}(X) = \sum_i |\sigma_i(X)|^0$  (其中  $\sigma_i(X)$  是矩阵  $X$  第  $i$  个奇异值) 中的  $|\cdot|^0$  以权衡问题的可解性和有效性。为了将  $L_0$  范数的非凸替换方法有效的应用于低秩恢复问题中，文章 [14] 考虑了如下低秩恢复框架，并给出了迭代重加权核范数方法 (Iteratively Reweighted Nuclear Norm Method, IRNN) 用于下面框架 (第一类低秩恢复框架) 的求解：

$$\min_X \mathbf{F}_0(X) = h(X) + \sum_{i=1}^m g(\sigma_i(X)), \quad (1-1)$$

其中  $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  在区间  $[0, +\infty)$  是一个非增连续的凹函数，而函数  $h : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^+$  的导数为 Lipschitz 连续函数 (其 Lipschitz 常数被记为  $L(h)$ )。正如 [15] 所述，IRNN 算法尽管保证了目标函数值能够随着迭代次数的增加而递减。但由于其上界替代 (Upper bound surrogate) 比较宽松，使得目标函数值的下降速度较为缓慢。为了解决该问题，Lu 等人给出了广义梯度算法 (Generalized proximal gradient, GPG) 以用于问题 (1-1) 的有效求解：

$$\begin{aligned} X^{(k+1)} &= \arg \min_X h(X^{(k)}) + \langle \nabla h(X^{(k)}), X - X^{(k)} \rangle \\ &\quad + \frac{\xi}{2} \|X - X^{(k)}\|_F^2 + \sum_{i=1}^m g(\sigma_i(X)) \\ &= \arg \min_X \frac{1}{2} \|X - O^{(k)}\|_F^2 + \sum_{i=1}^m g(\sigma_i(X)), \end{aligned} \quad (1-2)$$

其中  $O^{(k)} = X^{(k)} - \frac{\nabla h(X^{(k)})}{\xi}$  且  $\xi > L(h)$ 。因此问题 (1-2) 被转为下面形式的优化问题

$$\arg \min_X \frac{1}{2} \|Y - X\|_F^2 + \lambda \sum_{i=1}^m g(\sigma_i(X)). \quad (1-3)$$

对于 (1-3), Lu 等人 [15] 已经证明, 当函数  $g(x)$  有下界时有

$$\begin{aligned} \arg \min_X \frac{1}{2} \|Y - X\|_F^2 + \lambda \sum_{i=1}^m g(\sigma_i(X)) \\ = U \text{Diag}(\text{Prox}_g(\sigma(Y), \lambda)) V^T, \end{aligned} \quad (1-4)$$

其中  $U$ 、 $\sigma(Y)$  和  $V$  由  $Y$  的 SVD 分解  $Y = U \text{Diag}(\sigma(Y)) V^T$  得到。因此, 如下关键问题被广泛讨论:

$$\text{Prox}_g(y, \lambda) = \arg \min_{x \geq 0} \frac{1}{2} (y - x)^2 + \lambda g(x), \quad (1-5)$$

其中对于某些特定  $g$  情况, 问题 (1-5) 的闭式解已经被找到 [16–19]。为了求解一类广泛  $g$  下的问题 (1-5), Lu 等人 [15] 给出了一个迭代算法 (具体细节见算法1, 其中  $\varphi(x) = y - \lambda g'(x)$ )。论文 [15] 证明当  $g$  满足某些条件时该算法收敛到问题 (1-5) 的全局最优解。因此, 对于一类广泛的  $g$ , 我们能够结合算法1, 结论 (1-4) 以及 GPG 算法, 得到 (1-1) 的一个通用算法。尽管框架 (1-1) 已经包括了较为广泛的一类低秩恢复问题, 但依然存在一类常见的低秩恢复问题 (1-6) (在本文中被称为第二类低秩恢复框架) 是不能简单被归纳到框架 (1-1) 里的。

$$\begin{aligned} \min_{L, E} \sum_{i=1}^{\max(m, n)} g(\sigma_i(A)) + \alpha \|E\|_{\mathcal{G}, g}, \\ \text{s.t. } \|P - A - E\|_F^2 \leq \epsilon, \end{aligned} \quad (1-6)$$

其中  $\|E\|_{\mathcal{G}, g} = \sum_{i=1}^h g(\sqrt{\sum_{(i,j) \in \mathcal{G}_i} E_{ij}^2})$ , 其中  $E_{ij}$  是矩阵  $E \in \mathbb{R}^{m \times n}$  (假设  $m < n$ ) 中位于第  $i$  行第  $j$  列的元素, 且簇族  $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \dots, \mathcal{G}_h\}$  满足

$$\begin{cases} \cup_{i=1}^h \mathcal{G}_i = \{(i, j) | 1 \leq i \leq m, 1 \leq j \leq n\}; \\ \cap_{i=1}^h \mathcal{G}_i = \emptyset. \end{cases}$$

正如我们后面所将要讨论的那样, 第二类低秩恢复框架是能够利用分离变量迭代更新的策略 (如块坐标下降算法) 被简化为 (1-3) 和 (1-7):

$$\arg \min_X \frac{1}{2} \|Y - X\|_F^2 + \lambda \|X\|_{\mathcal{G}, g}. \quad (1-7)$$

由  $\|X\|_{\mathcal{G}, g}$  的定义可知, 问题 (1-8) 能够看作是问题 (1-7) 的一个特殊形式:

$$\min_X \frac{1}{2} \|Y - X\|_F^2 + \lambda \sum_i g(\|X_i\|_2), \quad (1-8)$$

其中  $X_i$  是矩阵  $X$  的第  $i$  个列向量。当  $g(x) = |x|$  时, 凸问题 (1-8) 有如下闭式解 [20,21]

$$\hat{X}_j = \begin{cases} \frac{\|Y_j\|_2 - \lambda}{\|Y_j\|_2} Y_j & \text{若 } \|Y_j\|_2 > \lambda \\ 0 & \text{if } \|Y_j\|_2 \leq \lambda \end{cases} \quad \text{对于 } j = 1, \dots, m. \quad (1-9)$$

另外当 (1-8) 中的  $g(x)$  为  $L_p$  范数时, Wang 等人 [22] 给出了相应非凸问题的全局最优解。

尽管对于某些特殊的  $g$ , (1-8) 的闭式解已经被找到。但是, 对于一类广泛的  $g$ , 问题 (1-8) 仍缺乏一种统一的解法。

---

#### 算法 1 Lu 的工作 [15]

---

输入:  $y > 0$ ,  $\tau > 0$  以及最大迭代次数  $\kappa > 0$ 。

- 1:  $f'_y(y) = 0$ , 返回  $\bar{x}_y = y$  且执行第 6 步。否则执行第 2 步。
- 2: 初始化  $x_L^{(0)} = y$ ,  $x_L^{(1)} = \varphi(x_L^{(0)})$  和  $k = 1$ 。继续执行第 3 步。
- 3: 若  $|x_L^{(k)} - x_L^{(k-1)}| > \tau$  &  $k < \kappa$ , 返回  $x_L^{(k+1)} = \varphi(x_L^{(k)})$  且执行第 4 步。否则执行第 5 步。
- 4: 若  $x_L^{(k+1)} < 0$ , 返回  $\bar{x}_y = 0$  且执行第 6 步。否则  $k = k + 1$  且返回第 3 步。
- 5:  $\bar{x}_y = x_L^{(k+1)}$ 。
- 6: 通过比较  $f_y(0)$  和  $f_y(\bar{x}_y)$  得到最优解  $x_L^*$ 。

输出: 问题 (4-21) 的最优解  $x_L^*$ 。

---

### 1.3 论文的选题和主要工作

如上讨论可知, Lu 等人给出了 GPG 算法, 并基于该算法将第一类低秩恢复框架 (1-1) 转为 (1-5) 进行求解。尽管 Lu 等人已经证明: 对于广泛的一类  $g$ , (1-5) 的全局最优解都能够利用算法 1 找到。然而正如我们之后所证明的那样, 算法 1 至多是线性收敛的。另外, 对于第二类低秩恢复框架, 目前还依然缺少一种通用算法。为了解决这些问题, 本文主要给出如下工作: (1) 文章首先给出了一个新的不动点算法 (GAI 算法) 用于问题 (1-5) 的求解。论文从理论上证明了, 当问题 (1-5) 中的非凸替代  $g(\cdot): \mathbb{R}^+ \rightarrow \mathbb{R}^+$  满足 **A1** – **A3** 时, 该算法超线性收敛于问题 (1-5) 的全局最优解, 其中

**A1**  $g(x)$  严格凹且严格增, 且  $g(0) = 0$ ;

**A2**  $g'(x)$  严格凸;

**A3**  $g''(x)$  在区间  $(0, +\infty)$  上连续。

另外, 实验也直观证明了 GAI 算法的有效性。(2) 为了给出第二类低秩恢复框架的一个通用算法, 本文利用块坐标下降算法的一种变种算法, 将框架转化为形式分别为 (1-3) 和 (1-7) 的两个非凸子问题进行解决, 并对该算法进行了理论分析。为了给出问题 (1-7) 在某一类广泛  $g$  下的一个一般算法。文章给出了一个重要定理 (该定理证明了两类问题的等价性, 基于此, 我们能够将原本的高维的优化问题被转化为一维的优化问题进行解决。)。因此结合 GAI 算法, 我们便得到了广义非凸低秩恢复框架的一个一般解法。(3) 文章最后利用参数化方法给出了一种新的  $L_0$  范数的非凸替代方法, 并将由此得到的非凸低秩恢复模型应用于图像去噪问题之中。基于该模型, 本文所给算法能够同时处理高斯噪声与混合噪声两种不同类型的噪声。实验结果证明了非凸替代方法的有效性。

## 1.4 本文的结构安排

第一章是绪论, 阐述了论文的研究背景和意义, 并介绍了一些常见非凸替代方法, 以及与其相关的一些重要的优化问题。最后指出了现有方法所存在的一些局限性, 并且介绍了本文的主要研究内容与成果。

第二章主要详细介绍了本文的相关工作, 主要被分为两个部分。第一部分介绍了主成分分析, 矩阵填充与鲁棒主成分分析等传统的低秩恢复方法。另外, 由于图像去噪是本文所给非凸替代方法的一个重要应用。在第二部分, 我们将会详细介绍了几个经典非局部图像去噪算法。

第三章主要给出了一个新颖的不动点算法 (GAI 算法), 并分别从理论分析与实验两个方面验明了该算法的有效性。

第四章首先对第二类低秩恢复框架进行了详细介绍, 并提出了块坐标下降算法的一种变种算法以用于该框架的求解, 然后给出了该算法的一个理论分析。一个重要定理也将在本章给出。结合该定理, 变种算法以及 GAI 算法, 最后得到第二类低秩恢复框架的一个一般解法。实验证明了文章所给解法的有效性。

第五章利用参数化的方法给出了一种新的  $L_0$  范数的非凸替代方法 (参数化范数和参数化核范数), 并基于所给的非凸替代给出了一个新的基于非凸低秩恢复的非局部图像去噪方法。最后的实验分别从高斯噪声和混合噪声两种情况出发, 将本文去噪方法与几种传统的图像去噪方法进行比较。

第六章对全文的工作进行了总结, 并提出了进一步研究的方向。

另外对于部分冗长的证明过程, 我们将在附录 A 给出。





## 第二章 相关工作

### 2.1 低秩矩阵恢复模型

近年低秩模型的兴起大致开始于矩阵填充问题的提出。但事实上，在 MC 被提出之前，主成分分析 (Principal component analysis, PCA) [9, 10] 就已经被给出并被广泛应用于数据降维中。PCA 假定数据近似分布于某个低维线性子空间，即由所有样本数据列向量化后所构成矩阵  $Y$  近似低秩。从数学上，矩阵  $Y$  可以表示为  $Y = X_0 + E_0$ ，其中  $X_0$  为低秩矩阵， $E_0$  为对  $X_0$  的小噪声扰动 (如高斯噪声)。因此，PCA 又能被看作是 Frobenius 范数约束下的矩阵秩最小化估计：

$$\min_X \text{rank}(X), \quad \text{s.t.} \quad \|Y - X\|_F^2 \leq \epsilon,$$

其中  $\|\cdot\|_F$  和  $\text{rank}(\cdot)$  分别用来表示 Frobenius 范数和秩函数 (矩阵奇异值向量上的  $L_0$  范数，即为矩阵非零奇异值的个数)， $\epsilon$  为重构误差的阈值。

PCA 的另一种常见表示形式为：

$$\min_X \|Y - X\|_F^2, \quad \text{s.t.} \quad \text{rank}(X) \leq l, \quad (2-1)$$

其中  $l$  被用来约束矩阵秩。利用拉格朗日乘子算法，上述两个问题能够被转化下面优化问题来进行求解：

$$\min_X \frac{1}{2} \|Y - X\|_F^2 + \alpha \text{rank}(X), \quad (2-2)$$

其中  $\frac{1}{2}$  的引入是为了方便模型的求解，而  $\alpha \geq 0$  为权衡系数。上述问题的闭式解能够通过下面硬阈值算法得到：

$$X = U\mathcal{H}_\alpha(\Sigma)V^T, \quad \mathcal{H}_\alpha(\Sigma) = \text{diag}(\text{sign}(\max(\sigma_i(Y) - \alpha, 0))\sigma_i(Y)),$$

其中  $U\Sigma V^T$  是  $Y$  的 SVD 分解:  $Y = U\Sigma V^T$ ,  $\sigma_i(Y)$  为  $Y$  的第  $i$  个奇异值向量，

$$\text{sign}(x) = \begin{cases} 1, & \text{if } x > 0; \\ 0, & \text{if } x = 0; \\ -1, & \text{if } x < 0. \end{cases} \quad (2-3)$$

当  $Y$  中仅有某些值被观测到时, 上述主成分分析问题被转换为矩阵填充问题, 即为数据缺失情况下的低秩结构恢复问题, 进行解决 [7]:

$$\min_X \text{rank}(Y), \quad \text{s.t.} \quad \|\pi_\Omega(Y) - \pi_\Omega(X)\|_F^2 \leq \epsilon,$$

其中  $\Omega$  为矩阵中观测元素位置的集合, 而  $\pi_\Omega$  为在集合  $\Omega$  上的投影算子。由于秩函数  $\text{rank}(\cdot)$  非凸非连续, 这导致求解上述优化问题是十分困难的。众所周知, 由于核范数 (矩阵奇异值向量上的  $L_1$  范数, 即为矩阵奇异值之和) 是秩函数的一个凸包, 故将上述问题转化为下面问题

$$\min_X \|X\|_*, \quad \text{s.t.} \quad \|\pi_\Omega(Y) - \pi_\Omega(X)\|_F^2 \leq \epsilon, \quad (2-4)$$

其中  $\|\cdot\|_*$  为核范数。

然而, 对于数据被随机的强噪声污染时的场景, 经典主成分分析和矩阵填充问题往往难以处理, 甚至失效。为了更好的解决这一问题, Chandrasekaran 等人 [8] 和 Wright 等人 [6] 于 2009 年同时提出了鲁棒主元分析 (Robust principal component analysis, RPCA), 并将其应用于解决在数据被稀疏强噪声污染的情况下如何去恢复数据潜在的低秩结构的问题:

$$\min_{L,E} \text{rank}(L) + \alpha \|E\|_0, \quad \text{s.t.} \quad P = L + E, \quad (2-5)$$

其中  $\|E\|_0$  为  $E$  的  $L_0$  范数 (即为矩阵  $E$  的非零元素个数),  $\alpha$  为权衡参数以用于平衡数据保真性和低秩性。由于 (2-5) 中的  $\|E\|_0$  和  $\text{rank}(L)$  都是非线性非凸, 故 (2-5) NP 难。人们转而求解下述凸优化问题 [6]:

$$\min_{L,E} \|L\|_* + \alpha \|E\|_1, \quad \text{s.t.} \quad P = L + E, \quad (2-6)$$

其中  $\|E\|_1$  为  $E$  的所有元素绝对值之和。该问题也被称为主成分追踪问题 (Principal component pursuit, PCP)。理论证明了 (2-6) 的精确可恢复性 [8] [6]。然而正如 [13] 所言, 在实际应用中, 观测值常常受到小噪声的干扰, 或者潜在的数据只是近似低秩。为了使主成分追踪问题得到更广泛应用, [13] 提出稳健主成分追踪模型 (Stable principal component pursuit, SPCP):

$$\min_{L,E} \|L\|_* + \alpha \|E\|_1, \quad \text{s.t.} \quad \|P - L - E\|_F^2 \leq \epsilon \quad (2-7)$$

并在理论上证明了 (2-7) 的鲁棒可恢复性。

在视觉应用中, 数据样本通常会被列向量化然后重排成矩阵进行处理。因此, 当少量样本被噪声污染或是存在少量的异常值时 (如被标签错误的样本), 数据矩阵  $D$  可以表示为  $D = A_0 + E_0$ 。其中  $A_0$  是表示为数据潜在的干净部分 (其往往具有低秩特性),  $E_0$  表示的是数据潜在的噪声部分 (往往具有列稀疏的特性)。Xu 等

人利用  $L_{2,0}$  范数 (矩阵非零列向量的个数) 来取代 RPCA 模型中的  $L_0$  范数, 得到如下离群点追踪模型 (Outlier persuit, OP) [23],

$$\begin{aligned} (\text{OP}) \quad & \min_{L,E} \quad \text{rank}(L) + \lambda \|E\|_{2,0} \\ \text{s.t.} \quad & A + E = D. \end{aligned} \quad (2-8)$$

类似地, OP 相应的凸松弛问题被给出用于问题的有效求解:

$$\min_{L,E} \|L\|_* + \alpha \|E\|_{2,1}, \quad \text{s.t.} \quad P = L + E, \quad (2-9)$$

其中  $\|E\|_{2,1}$  为  $E$  的  $L_{2,1}$  范数 (即矩阵  $E$  的所有列向量的模长之和)。类似于 PCP 上的结果, Xu 等人在理论上证明了 (2-9) 的精确可恢复性 [23]。

如上所述, 由于  $L_0$  范数和秩函数的非凸非连续性, 原始的低秩恢复模型通常都是 NP-难的。这种问题的一种常见解决方法是利用  $L_1$  范数和核范数来替换优化问题中  $L_0$  范数和秩函数, 即将原始的离散优化问题转为凸优化问题的进行求解。并且对于上文提到多数低秩恢复模型 (如 MC, PCP, OP) 的精确可恢复性都已经在理论上被证明, 即当某些条件被满足时, 通过求解凸优化问题能够得到真实的解。然而在实际应用中, 这些严格的前提条件往往难以被满足。这使得由凸优化得到的解可能会严重偏离问题的真实解。

## 2.2 基于非局部相似块的图像去噪算法介绍

在这一部分中, 我们对正文第二章中所提到的传统的基于非局部相似块的图像去噪算法进行了更为详细的叙述。

在介绍具体的方法之前, 我们首先给出这一节将会用到的定义与概念。我们定义  $I \in \mathbb{R}^{w \times h}$  是一个噪声图像, 其潜在的干净图片为  $I_c$ 。  $I(i)$  和  $I_c(i)$  为第  $i$  个位置上的像素值。当图片为彩色图像时,  $I(i)$  和  $I_c(i)$  为 RGB 三通道上的值所构成的向量。定义  $p(i) \in \mathbb{R}^{a \times b}$  ( $a < w, b < h$ ) 为以像素点  $i$  为中心的像素块。

### 2.2.1 非局部均值算法

给定噪声图片  $I$ , 相应的由 NLM 得到的去噪后的图片为  $I'$ :

$$I'(i) = \sum_{j \in I} w(i, j) I(j), \quad (2-10)$$

其中  $w(i, j)$  表示像素  $i$  和像素  $j$  之间的相似性。正如之前所言, 为了使 NLM 方法更加鲁棒, 图像块  $p(i)$  和图像块  $p(j)$  被用于计算权重, 其中  $p(i) \in \mathbb{R}^{a \times b}$  ( $a < w, b < h$ ) 和  $p(j) \in \mathbb{R}^{a \times b}$  ( $a < w, b < h$ ) 分别为以像素点  $i$  和像素点  $j$  为中心的像素块。

$$w(i, j) = \frac{1}{c(i)} \exp(-d(p(i), p(j))/h^2) \quad (2-11)$$

其中  $d(p(i), p(j))$  为  $p(i)$  和  $p(j)$  之间的距离:  $d(p(i), p(j)) = \|p(i) - p(j)\|_{F, \sigma}^2$ ,  $\sigma > 0$  为高斯核的标准差。  $h$  和  $c(i)$  分别为指数函数的衰变因子和标准化参数。

对于每个  $i$ , 我们首先列向量化所有的像素块, 并乘以相应的权重系数:  $P = [w(i, 1)\text{vec}(p(1)) \ w(i, 2)\text{vec}(p(2)) \ \cdots]$ , 其中  $\text{vec}(p(i))$  为对像素块  $p(i)$  列向量化后所得到的向量。最后我们得到下面低秩恢复问题:

$$\min_L \|P - L\|_F^2, \quad \text{s.t.} \quad \text{rank}(L) = 1. \quad (2-12)$$

通过这种方式, 非局部方法被转化为一个低秩恢复问题。

### 2.2.2 基于块的 SVD 的图像去噪方法

基于块的 SVD 的图像去噪方法 [24] 具体步骤如下:

- 将噪声图像  $I$  划分成一组重叠的图像块, 并在其中取一个样本块  $p$  (这里, 块的中心像素的指标被省略以简化符号)。在给定的区域中找到  $k$  个与样本块最为相似的块  $\{p_i\}_{i=1}^k$ 。将这些所有块列向量化, 并将它们构成矩阵  $P = [p \ p_1 \ p_2 \ \cdots \ p_k]$ 。
- 对  $P$  进行 SVD 分解得到  $P = U\Sigma V^T$ 。如果  $\Sigma$  中的奇异值低于固定阈值  $\tau$ , 则将其截断, 以获得矩阵  $\Sigma'$ 。
- 重构去噪后的矩阵  $L = U\Sigma'V^T$  得到去噪后图像  $I'$ 。对于重叠的块中的像素点, 取估计点的平均值做为最终去噪的点。

可以注意到基于块的 SVD 的图像去噪方法中的第二步能够被等价地转换为下面模型的求解,

$$\min_L \frac{1}{2} \|P - L\|_F^2 + \alpha \text{rank}(L), \quad (2-13)$$

通过对权衡系数  $\alpha$  的合理选择, 则由上述问题的所得到的结果与基于块的 SVD 的图像去噪方法相同。

### 2.2.3 基于同步稀疏编码方法

给定一组类似的补丁  $p = [p \ p_1 \ p_2 \ \cdots \ p_k]$ , 同步稀疏编码模型 (Simultaneous sparse coding based method, SSCB) [25] 定义如下。

$$\min_{D, A} \frac{1}{2} \|P - DA\|_F^2 + \alpha \|A^T\|_{2,1}, \quad (2-14)$$

其中,  $D$  是从块  $P$  中学习得到的字典,  $A$  是编码系数矩阵。  $A$  的第  $i$  列  $A_i$  为对字典  $D$  上的第  $i$  个块  $p_i$  的编码为:  $p_i = DA_i$ 。其中  $A^T$  为  $A$  的转置矩阵,  $\|A^T\|_{2,1} = \sum_{i=1}^m \left( \sqrt{\sum_{j=1}^n A_{ij}^2} \right)$ 。

当我们定义  $\sigma_i = \sqrt{\sum_{j=1}^n A_{ij}^2}$ , 则  $\|A^T\|_{2,1}$  和  $A$  由下式获得:

$$\|A^T\|_{2,1} = \sum_{i=1}^m \sigma_i, \quad A = \Sigma V^T,$$

其中  $\Sigma = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_m\}$ ,  $V$  的每个列向量被定义为  $v_i = (A^i)^T / \sigma_i$ 。这里的  $A^i$  表示为  $A$  的第  $i$  个行向量。则问题 (2-14) 转为

$$\min_{D, \Sigma, V} \frac{1}{2} \|P - D\Sigma V^T\|_F^2 + \alpha \sum_{i=1}^m \sigma_i. \quad (2-15)$$

在论文 [26] 中, 作者认为优化问题 (2-15) 与问题 (2-16) 等价, 而后者能够利用  $P$  的 SVD 分解和软阈值得到。

$$\min_L \frac{1}{2} \|P - L\|_F^2 + \alpha \|L\|_*, \quad (2-16)$$

然而, 实验结果显示这两个优化问题的解十分相似, 但并不是严格等价的。尽管在这两种情况下都得到了很好的结果。



## 第三章 不动点算法

### 3.1 不动点算法及其理论分析

正如第一章中所提到的，尽管算法1收敛到该问题 (1-5) 的全局最优解 [15]，但是正如下面所将要讨论的，算法1至多为线性收敛。因此，为了提高算法收敛速度，一个新的迭代函数被给出：

$$\psi(x) = \varphi(x) - \frac{(\varphi(\varphi(x)) - \varphi(x))(\varphi(x) - x)}{\varphi(\varphi(x)) - 2\varphi(x) + x},$$

其中  $\varphi(x) = y - \lambda g'(x)$ 。基于这个迭代函数，我们得到一个新的加速算法 GAI(具体细节见算法2)。

---

#### 算法 2 GAI 算法

---

输入：  $y > 0$ ,  $\tau > 0$  以及最大迭代次数  $\kappa > 0$ 。

- 1: 令  $a_0 = \max\{x | \lambda g''(x) = -1 \text{ 或 } x = 0\}$ ，若  $f'_y(a_0) < 0$ ，则执行第 2 步。否则返回  $\hat{x}_y = a_0$  且执行第 5 步。
- 2: 初始化  $x_G^{(0)} = y$  和  $k = 0$ 。继续执行第 3 步。
- 3: 若  $|\varphi(\varphi(x_G^{(k)})) - 2\varphi(x_G^{(k)}) + x_G^{(k)}| > \tau$  &  $k < \kappa$ ，返回  $x_G^{(k+1)} = \psi(x_G^{(k)})$  且执行第 4 步。否则返回  $\hat{x}_y = \varphi(x_G^{(k)})$  且执行第 5 步。
- 4:  $k = k + 1$  且返回第 3 步。
- 5: 比较  $f_y(0)$  和  $f_y(\hat{x}_y)$  得到最终的最优解  $x_G^*$ 。

输出：问题 (1-5) 的最优解  $x_G^*$ 。

---

下面我们将会证明 GAI 超线性收敛于问题 (1-5) 的全局最优解。为了方便起见，我们定义符号如下

$$\begin{aligned}\bar{x}_y &= \max\{x | f'_y(x) = 0, 0 \leq x \leq y \text{ or } x = 0\}, \\ \mathcal{S} &= \{x | f'_y(x) = 0, 0 \leq x \leq y\}, \\ g''(0) &= \lim_{x \rightarrow 0^+} g''(x), \psi(\bar{x}_y) = \lim_{x \rightarrow \bar{x}_y^+} \psi(x), \\ a_0 &= \max\{x | \lambda g''(x) = -1 \text{ or } x = 0\}.\end{aligned}$$

首先回顾 [15] 中几个结论，并给出一些关于  $f_y(x)$ ,  $\varphi(x)$  和  $\psi(x)$  性质的结论，以用于 GAI 收敛性的证明。

定义 1 对于  $g(\cdot) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ , 定义假设 **A1-A3** 如下:

**A1**  $g(x)$  严格凹且严格增,  $g(0) = 0$ 。

**A2**  $g'(x)$  严格凸。

**A3**  $g''(x)$  在区间  $(0, +\infty)$  上连续。

引理 1 [15] 给定满足假设 **A1-A3** 的  $g$ 。当  $S \neq \emptyset$  时, 则有

- (i) 假设  $\{x_L^{(k)}\}$  是算法 1 得到序列, 即  $x_L^{(k+1)} = \varphi(x_L^{(k)})$ , 其中初始值  $x_L^{(0)} = y$ 。则该序列收敛于  $\bar{x}_y$ ;
- (ii) 对于任意的  $\forall x \in (\bar{x}_y, y]$ , 则  $\bar{x}_y < \varphi(x) < x$ ;
- (iii)  $\bar{x}_y$  是  $\varphi(x)$  的一个不动点, 即,  $\varphi(\bar{x}_y) = \bar{x}_y$ ;
- (iv)  $\arg \min_{x \in \{0, \bar{x}_y\}} f_y(x)$  为优化问题 (1-5) 的全局最优解。

引理 2 给定满足假设 **A1-A3** 的  $g$ 。当  $S \neq \emptyset$  时, 则

- (i)  $0 \leq a_0 < y$ ;
- (ii)  $\varphi(x)$  严格递增, 而  $\varphi'(x)$  在区间  $(0, +\infty)$  严格递减;
- (iii) 对于  $\forall x \in (a_0, \infty)$  有  $-1 < \lambda g''(x) < 0$  成立;
- (iv) 方程  $f'_y(x) = 0$  在区间  $(a_0, y]$  内最多有一个根。

引理 3 给定满足假设 **A1-A3** 的  $g$ 。当  $S \neq \emptyset$  时, 则下面结论成立:

- (i)  $\varphi'(\bar{x}_y) \leq 1$  成立;
- (ii) 若  $f'_y(a_0) < 0$ , 则  $a_0 < y$ , 并且方程  $f'_y(x) = 0$  在区间  $(a_0, y]$  内有唯一解;
- (iii) 若  $f'_y(a_0) \geq 0$ , 若  $\bar{x}_y \notin (0, a_0) \cup (a_0, y]$  (若  $a_0 \geq y$ , 定义  $(a_0, y] = \emptyset$ ; 当  $a_0 = y$  时, 定义  $(0, a_0] = \emptyset$ )。
- (iv)  $f_y(x)$  在区间  $(0, +\infty)$  内至多有一个非零极小值点。

引理 4 给定满足假设 **A1-A3** 的  $g$ 。当  $S \neq \emptyset$  且  $f'_y(a_0) < 0$ , 则

- i) 对于任意的  $x \in (\bar{x}_y, y]$ , 都有  $\psi(x) < x$  成立;
- ii)  $\bar{x}_y$  是  $\psi(x)$  的不动点, 即,  $\psi(\bar{x}_y) = \bar{x}_y$ 。



**引理 5** 给定满足假设 **A1-A3** 的  $g$ 。当  $S \neq \emptyset$  且  $f'_y(a_0) < 0$ ，则  $\psi(x)$  是  $(\bar{x}_y, y]$  区间上的一个严格增函数。

基于上述结论，下面给出两个定理来说明 GAI 超线性收敛于优化问题 (1-5) 的全局最优解。

**定理 1** 给定满足假设 **A1-A3** 的  $g$ 。当  $S \neq \emptyset$  且  $f'_y(a_0) < 0$  成立，则由  $x_G^{(k+1)} = \psi(x_G^{(k)})$  ( $x_G^{(0)} = y$ ) 得到的序列  $\{x_G^{(k)}\}$  收敛于  $\bar{x}_y$ 。

**证明** 下面利用归纳法证明对于  $\forall k$ ,  $x_{\text{GAI}}^{(k)} > \bar{x}_y$ 。当  $k = 1$ ,  $x_G^{(1)} = \psi(y) > \psi(\bar{x}_y) = \bar{x}_y$  其中第二个不等式和最后一个等式分别由引理 5 和引理 4 (ii) 得到。当  $k = n$  时,  $x_G^{(k)} > \bar{x}_y$  成立, 则  $x_G^{(n+1)} = \psi(x_G^{(n)}) > \psi(\bar{x}_y) = \bar{x}_y$ 。从另一方面, 由引理 4 (i) 可以得到  $x_G^{(k+1)} < x_G^{(k)} \forall k$ 。因此  $\{x_G^{(k)}\}$  收敛, 并且可以假设  $\{x_G^{(k)}\}$  收敛于  $x_G$ 。显然,  $x_G \geq \bar{x}_y$ 。

下面利用反证法证明  $x_G = \bar{x}_y$ 。假设  $x_G > \bar{x}_y$ 。因为  $\psi(x)$  在区间  $(\bar{x}_y, y]$  上连续, 故  $\{\psi(x_G^{(k)})\}$  收敛于  $\{\psi(x_G)\}$ 。又因为  $\{x_G^{(k+1)} = \psi(x_G^{(k)})\}$  收敛于  $x_G$ , 故  $\psi(x_G) = x_G$ 。这与由引理 4 (i) 得到的  $\psi(x_G) < x_G$  相矛盾。证明完成。□

**定理 2** 若  $g$  满足假设 **A1-A3**,  $f'_y(a_0) < 0$ , 则有

- (i) 由算法 1 得到的序列  $\{x_L^{(k)}\}$  线性收敛;
- (ii) 由 GAI 得到的序列  $\{x_G^{(k)}\}$  超线性收敛。
- (iii) 对任意  $k > 0$ , 存在  $0 < \rho < 1$  使得  $|x_G^{(k)} - \bar{x}_y| = O(\rho^k)$  成立。

**证明** (i) 当  $f'_y(x) \neq 0$ ,  $f'_y(a_0) < 0$ , 则可以由引理 3 (ii) 推出  $S \neq \emptyset$ 。由引理 2 (iii) 容易得到

$$\begin{aligned} \lim_{k \rightarrow +\infty} \frac{x_L^{(k+1)} - \bar{x}_y}{x_L^{(k)} - \bar{x}_y} &= \lim_{k \rightarrow +\infty} \frac{\varphi(x_L^{(k)}) - \varphi(\bar{x}_y)}{x_L^{(k)} - \bar{x}_y} \\ &= \lim_{\xi_0 \rightarrow \bar{x}_y^+} \varphi'(\xi_0) = \varphi'(\bar{x}_y) > 0, \end{aligned}$$

其中  $\xi_0 \in (\bar{x}_y, x_L^{(k)})$ 。又因为由  $\bar{x}_y > a_0$  可以得到  $\varphi'(\bar{x}_y) < 1$ , 故序列  $\{x_L^{(k)}\}$  是线性收敛。由引理 5 的证明, 可以得到

(ii)

$$\begin{aligned} \psi'(x) &= \varphi'(x) - \frac{1}{(\varphi'(\xi) - 1)^2} \{ \varphi'(\xi)^2 \varphi'(x) - \varphi'(\xi)^2 \\ &\quad - \varphi'(x) \varphi'(\varphi(x)) + \varphi'(x) \}, \end{aligned}$$

表 3-1:  $L_0$  范数的一些非凸替代Table 3-1: Some surrogate functions of  $\ell_0$ 

Name	$g'(x)$	$g''(x)$	$a_0$
$L_p$	$px^{p-1}, 0 < p < 1$	$p(p-1)x^{(p-2)}$	$\max((\lambda p(1-p))^{\frac{1}{2-p}}, 0)$
LSP	$\frac{1}{\gamma+x}$	$-\frac{1}{(\gamma+x)^2}$	$\max(\sqrt{\lambda} - \gamma, 0)$
Laplace	$\frac{1}{\gamma} \exp(\frac{-x}{\gamma})$	$-\frac{1}{\gamma^2} \exp(\frac{-x}{\gamma})$	$\max(-\gamma \log(\frac{\lambda^2}{\lambda}), 0)$
LOG	$\frac{1}{\gamma+x}$	$-\frac{1}{(\gamma+x)^2}$	$\max(\sqrt{\lambda} - \gamma, 0)$
Logarithm	$\frac{\gamma}{(\gamma x+1) \log(\gamma+1)}$	$-\frac{\gamma^2}{(\gamma x+1)^2 \log(\gamma+1)}$	$\max(\frac{\sqrt{\frac{\gamma^2 \lambda}{\log(\gamma+1)}} - 1}{\gamma}, 0)$
ETP	$\frac{\gamma \exp(-\gamma x)}{1 - \exp(-\gamma)}$	$\frac{-\gamma^2 \exp(-\gamma x)}{1 - \exp(-\gamma)}$	$\max(\frac{\log(\frac{1 - \exp(-\gamma)}{\lambda \gamma^2})}{-\gamma}, 0)$

其中  $\xi \in (\varphi(x), x) \subseteq (\bar{x}_y, y)$ 。因为  $\varphi'(\bar{x}_y) \neq 1$ ，则

$$\begin{aligned}
\lim_{x \rightarrow \bar{x}_y^+} \psi'(x) &= \varphi'(\bar{x}_y) - \frac{1}{(\varphi'(\bar{x}_y) - 1)^2} \left\{ \varphi'(\bar{x}_y)^2 \varphi'(\bar{x}_y) \right. \\
&\quad \left. - \varphi'(\bar{x}_y)^2 - \varphi'(\bar{x}_y) \varphi'(\bar{x}_y) + \varphi'(\bar{x}_y) \right\} \\
&= \varphi'(\bar{x}_y) - \varphi'(\bar{x}_y) = 0.
\end{aligned}$$

定义  $e_k = x_G^{(k)} - \bar{x}_y$ ，则

$$\begin{aligned}
e_{k+1} &= x_G^{(k+1)} - \bar{x}_y = \psi(x_G^{(k)}) - \psi(\bar{x}_y) \\
&= \psi'(\xi_0)(x_G^{(k)} - \bar{x}_y) = \psi'(\xi_0)e_k,
\end{aligned}$$

其中  $\xi_0 \in (\bar{x}_y, x_G^{(k)})$ 。因此我们可以推出

$$\lim_{k \rightarrow +\infty} \frac{e_{k+1}}{e_k} = \lim_{\xi_0 \rightarrow \bar{x}_y^+} \psi'(\xi_0) = 0,$$

其中由定理1可得对于  $\forall k$ ，有  $e_k > 0$  成立。

(iii) 对任意的  $\forall x \in (\bar{x}_y, y)$ ，因为  $\psi(x) = \varphi(x) - \frac{\varphi(\varphi(x)) - \varphi(x)}{\varphi'(\xi_1) - 1} \leq \varphi(x)$  (其中  $\xi_1 \in (\varphi(x), x)$ ) 以及  $\psi(\bar{x}_y) = \varphi(\bar{x}_y) = \bar{x}_y$ ，可以得到

$$\psi(x) - \bar{x}_y \leq \varphi(x) - \varphi(\bar{x}_y) = \varphi'(\xi_2)(x - \bar{x}_y) < \varphi'(\bar{x}_y)(x - \bar{x}_y),$$

其中  $\xi_2 \in (\bar{x}_y, x)$ 。令  $\rho = \varphi'(\bar{x}_y) (< 1)$ ，有  $\psi(x) - \bar{x}_y \leq \rho(x - \bar{x}_y) \forall x \in (\bar{x}_y, y)$  成立。由定理 1 (ii) 可知， $\varphi(x) \in (\bar{x}_y, y)$ ， $\forall x \in (\bar{x}_y, y)$ 。因此， $\psi$  是一个压缩常数为  $\rho < 1$  的压缩映射。故  $|x_G^{(k)} - \bar{x}_y| = O(\rho^k)$  成立。

□

由引理 3(iii) 和 (v), 可以得到当  $f'_y(a_0) \geq 0$  时, 问题 (1-5) 的最优解属于  $\{0, a_0\}$ 。另外由  $\bar{x}_y$  的定义, 引理 3 (iii) 以及 (iv) 的证明过程可以得到当  $f'_y(a_0) < 0$  时, 问题 (1-5) 的最优解属于  $\{0, \bar{x}_y\}$ 。故 GAI 算法收敛于问题 (1-5) 的最优解。

另外由定理 2 可以看出, 当  $f'_y(y) \neq 0$ ,  $f'_y(a_0) < 0$  时, GAI 的收敛速度要比算法 1 更快。当定理 2 的条件不被满足时, GAI 通过比较  $f_y(0)$  和  $f_y(x_G)$  可直接得到最优解 (此时  $x_G = a_0$ )。而算法 1 在比较  $f_y(0)$  和  $f_y(\bar{x}_y)$  之前, 仍需迭代步骤以得到  $\bar{x}_y$ 。故无论在何种情况下 GAI 收敛速度比算法 1 更快。

下面讨论 GAI 的停止准则。由中值定理可以得到

$$\begin{aligned} & \left| \varphi(\varphi(x_G^{(k)})) - 2\varphi(x_G^{(k)}) + x_G^{(k)} \right| \\ &= \left( \frac{1}{\varphi'(\xi)} - 1 \right) \left| \varphi(x_G^{(k)}) - \varphi(\varphi(x_G^{(k)})) \right|, \end{aligned}$$

其中由引理 2 (ii) 可以得到  $\xi \in (\varphi(x), x)$ ,  $\frac{1}{\varphi'(\xi)} - 1 > 0$  成立。令  $c_0 = \frac{1}{\frac{1}{\varphi'(\bar{x}_y)} - 1} > 0$ , 故由引理 2 (i) 可以得到

$$c_0 |\varphi(\varphi(x_G^{(k)})) - 2\varphi(x_G^{(k)}) + x_G^{(k)}| \geq |\varphi(x_G^{(k)}) - \varphi(\varphi(x_G^{(k)}))|.$$

因此设  $|\varphi(\varphi(x_G^{(k)})) - 2\varphi(x_G^{(k)}) + x_G^{(k)}| < \tau$  为 GAI 的停止条件有以下两点好处:

- $f'_y(\varphi(x_G^{(k)})) = |\varphi(x_G^{(k)}) - \varphi(\varphi(x_G^{(k)}))|$  的值能够由  $c_0\tau$  所控制。
- 它保证了 GAI 能够在实际情况中有效的应用。因为当 GAI 的迭代函数  $\psi(x)$  里面的分母  $|\varphi(\varphi(x_G^{(k)})) - 2\varphi(x_G^{(k)}) + x_G^{(k)}|$  在小于  $\tau$  时, 算法被停止。

### 3.2 数值实验与分析

为了对 GAI 算法和算法 1 做一个更直观的分析, 我们利用仿真实验对 GAI 算法和算法 1 的收敛速度的进行对比。考虑下面问题:

$$\arg \min_X \mathbf{G}(X) = \frac{1}{2} \|Y - X\|_F^2 + \lambda \sum_{i=1}^m \sum_{j=1}^n g(|X_{ij}|), \quad (3-1)$$

其中  $X, Y \in \mathbb{R}^{m \times n}$ ,  $X_{ij}$  是矩阵  $X$  第  $i$  行第  $j$  列的元素。

在我们的实验中, 矩阵  $Y \in \mathbb{R}^{100 \times 100}$  中的元素独立同分布于标准正态分布  $\mathcal{N}(0, 1)$ ,  $\lambda$  被设置为 1。对于每一个非凸替代函数, 我们将算法 1 中停止条件为  $\kappa = 1000$ ,  $\tau = 10^{-20}$  时, 所得到的输出作为真实解的一个逼近, 并记为  $X^*$ 。利用绝对误差  $\text{Absolute Error} = |\mathbf{G}(X^*) - \mathbf{G}(\hat{X}_k)|$  来衡量不同方法的有效性, 其中  $\hat{X}_k$  为不同算法第  $k$  次迭代所得到的结果。显然, 对于相同的  $k$ , Absolute Error 越小表示算法的收敛速度越快。

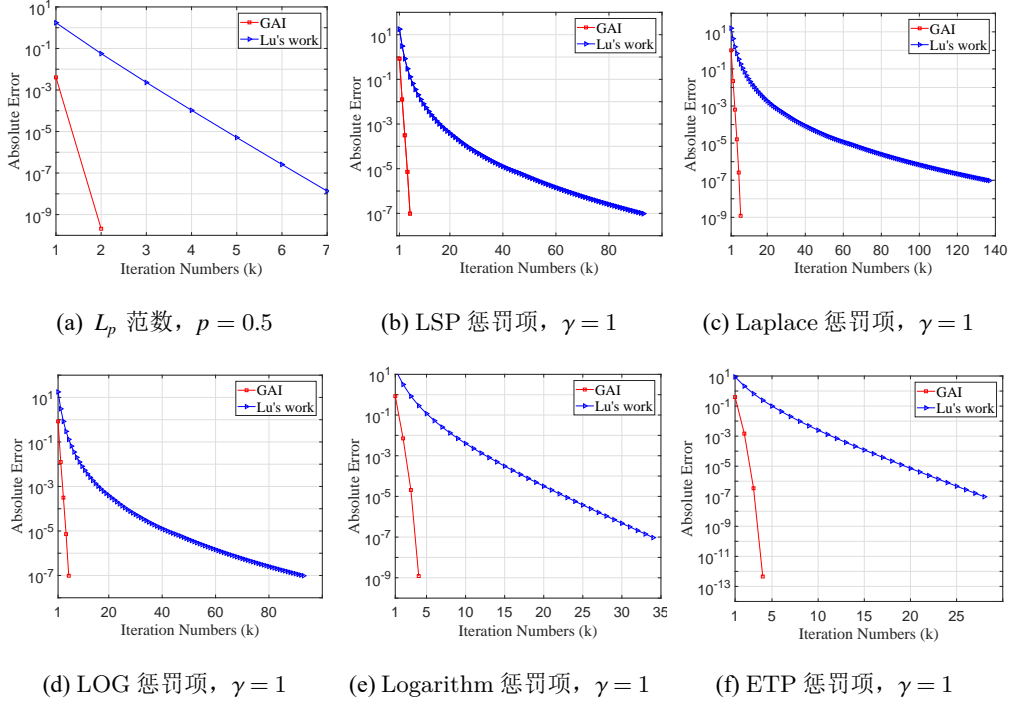


图 3-1:  $g$  为 (a)  $L_p$  范数, (b) LSP 惩罚项, (c) Laplace 惩罚项, (d) LOG 惩罚项, (e) Logarithm 惩罚项以及 (f) ETP 惩罚项时, 算法1 和 GAI 的收敛曲线。

Figure 3-1: The convergence curves of Algorithm 1 and GAI with the surrogate  $g$  being (a)  $L_p$  norm, (b) LSP penalty, (c) Laplace penalty, (d) LOG penalty, (e) Logarithm penalty and (f) ETP, respectively.

由图 3-1 我们可以观察到: 由 GAI 算法得到曲线下降较为迅速。最多经历 6 次迭代, 相应的绝对误差就能降到  $10^{-7}$ 。而相对的, 由算法1 得到曲线下降较为缓慢。在有些例子中, 算法1 的迭代次数甚至超过 100 次, 其相应的绝对误差才能达到  $10^{-7}$ 。

## 第四章 第二类低秩矩阵恢复框架的解法以及应用

### 4.1 第二类低秩恢复框架

为了使更多的  $L_0$  范数的非凸替代能够被更有效的应用于更多的低秩恢复问题中，以提升传统低秩恢复算法的性能。在这一章，我们将给出一个一般的解法用于求解第二类低秩恢复框架 (1-6)。

为了方便讨论，我们引入下面模型：

$$\begin{aligned} \min_{A,E} \quad & \text{rank}(A) + \alpha \|E\|_{\mathcal{G},0}, \\ \text{s.t.} \quad & \|P - A - E\|_F^2 \leq \epsilon, \end{aligned} \quad (4-1)$$

其中  $\|E\|_{\mathcal{G},0} = \sum_{i=1}^h |\sqrt{\sum_{(i,j) \in \mathcal{G}_i} E_{ij}^2}|^0$ ，且集族  $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \dots, \mathcal{G}_h\}$  满足

$$\begin{cases} \cup_{i=1}^h \mathcal{G}_i = \{(i,j) | 1 \leq i \leq m, 1 \leq j \leq n\}; \\ \cap_{i=1}^h \mathcal{G}_i = \emptyset. \end{cases}$$

由 (4-1) 的定义容易看出，对于第一章所提到的低秩恢复模型，包括 PCA, RPCA 和 OP，都能够看作是 (4-1) 的一个特殊情况。然而，由于 (4-1) 是一个 NP 难问题，因此不能被直接求解。为了解决这个问题，我们利用  $L_0$  范数的非凸替代函数  $g$  来取代模型中的  $L_0$  范数，故得到下面框架 (第二类低秩恢复框架)

$$\begin{aligned} \min_{L,E} \quad & \sum_{i=1}^{\max(m,n)} g(\sigma_i(A)) + \alpha \|E\|_{\mathcal{G},g}, \\ \text{s.t.} \quad & \|P - A - E\|_F^2 \leq \epsilon, \end{aligned} \quad (4-2)$$

其中  $\|E\|_{\mathcal{G},g} = \sum_{i=1}^h g(\sqrt{\sum_{(i,j) \in \mathcal{G}_i} E_{ij}^2})$ 。我们可以通过对该模型的求解，来得到框架 (4-1) 的一个近似解。这里的  $g$  泛指  $L_0$  范数的某类非凸替代函数而非某个具体的非凸范数。

### 4.2 块坐标下降算法的一种变种算法及其理论分析

我们首先将框架 (4-2) 转化为以下无约束优化问题：

$$\min_{A,E} \mathbf{F}_1(A, E) = \alpha \sum_{i=1}^r g(\sigma_i(A)) + \beta \sum_{\mathcal{G}_i \in \mathcal{G}} g(\|E_{\mathcal{G}_i}\|_F) + \frac{1}{2} \|D - A - E\|_F^2. \quad (4-3)$$

由于同时求解 (4-3) 中的  $A$  和  $E$  难度较大。本文结合坐标下降法和邻近算法对 (4-3) 的  $A$  和  $E$  进行迭代求解。这种相结合的方法被称为块坐标下降算法的变种算法 (The combining the coordinate descent with the proximal algorithm, PBCD)。

(1) 给定  $A^{(k)}$  和  $E^{(k)}$ , 得到  $A^{(k+1)}$  :

$$\begin{aligned} & \arg \min_A \alpha \sum_{i=1}^r g(\sigma_i(A)) + \frac{1}{2} \|D - A - E^{(k)}\|_F^2 + c \|A - A^{(k)}\|_F^2 \\ &= \arg \min_A \frac{\alpha}{2(\frac{1}{2} + c)} \sum_{i=1}^r g(\sigma_i(A)) + \frac{1}{2} \|A - \frac{\frac{1}{2}(D - E^{(k)}) + cA^{(k)}}{\frac{1}{2} + c}\|_F^2. \end{aligned} \quad (4-4)$$

(2) 给定  $A^{(k+1)}$  和  $E^{(k)}$ , 得到  $E^{(k+1)}$ :

$$\begin{aligned} & \arg \min_E \beta \sum_{\mathcal{G}_i \in \mathcal{G}} g(\|E_{\mathcal{G}_i}\|_F) + \frac{1}{2} \|D - A^{(k+1)} - E\|_F^2 + c \|E - E^{(k)}\|_F^2 \\ &= \arg \min_E \frac{\beta}{2(\frac{1}{2} + c)} \sum_{\mathcal{G}_i \in \mathcal{G}} g(\|E_{\mathcal{G}_i}\|_F) + \frac{1}{2} \|E - \frac{\frac{1}{2}(D - A^{(k+1)}) + cE^{(k)}}{\frac{1}{2} + c}\|_F^2. \end{aligned} \quad (4-5)$$

**定理 3** 对于任意的  $c > 0$ , 假设  $\{(A^{(k)}, E^{(k)})\}$  是由 PBCD 算法所得到的一个序列, 则  $\{(A^{(k)}, E^{(k)})\}$  满足下面性质:

- 1)  $F_1(A^{(k)}, E^{(k)})$  单调递减, 并且有  $F_1(A^{(k+1)}, E^{(k+1)}) - F_1(A^{(k)}, E^{(k)}) \geq c(\|A^{(k+1)} - A^{(k)}\|_F^2 + \|E^{(k+1)} - E^{(k)}\|_F^2)$  成立;
- 2)  $F_1(A^{(1)}, E^{(1)}) \geq c \sum_{k=1}^{+\infty} (\|A^{(k+1)} - A^{(k)}\|_F^2 + \|E^{(k+1)} - E^{(k)}\|_F^2)$ , 这意味着  $\lim_{k \rightarrow +\infty} (A^{(k)}, E^{(k)}) - (A^{(k+1)}, E^{(k+1)}) = \mathbf{0}$ 。

**证明** 因为  $A^{(k+1)}$  是 (4-4) 的最优解, 故

$$\begin{aligned} & \alpha \sum_{i=1}^r g(\sigma_i(A^{(k+1)})) + \frac{1}{2} \|D - A^{(k+1)} - E^{(k)}\|_F^2 + \eta \|A^{(k+1)} - A^{(k)}\|_F^2 \\ & \leq \alpha \sum_{i=1}^r g(\sigma_i(A^{(k)})) + \frac{1}{2} \|D - A^{(k)} - E^{(k)}\|_F^2 + \eta \|A^{(k)} - A^{(k)}\|_F^2, \end{aligned} \quad (4-6)$$

故得到

$$\begin{aligned} & F_1(A^{(k)}, E^{(k)}) - F_1(A^{(k+1)}, E^{(k)}) \\ &= \alpha \sum_{i=1}^r g(\sigma_i(A^{(k)})) + \frac{1}{2} \|D - A^{(k)} - E^{(k)}\|_F^2 - \alpha \sum_{i=1}^r g(\sigma_i(A^{(k+1)})) \\ & \quad - \frac{1}{2} \|D - A^{(k+1)} - E^{(k)}\|_F^2 \geq \eta \|A^{(k+1)} - A^{(k)}\|_F^2. \end{aligned} \quad (4-7)$$

类似地, 由  $E^{(k+1)}$  是 (4-5) 的最优解, 得

$$\begin{aligned} & \beta \sum_{\mathcal{G}_i \in \mathcal{G}} g(\|E_{\mathcal{G}_i}^{(k+1)}\|_F) + \frac{1}{2} \|D - A^{(k+1)} - E^{(k+1)}\|_F^2 + \eta \|E^{(k+1)} - E^{(k)}\|_F^2 \\ & \leq \beta \sum_{\mathcal{G}_i \in \mathcal{G}} g(\|E_{\mathcal{G}_i}^{(k)}\|_F) + \frac{1}{2} \|D - A^{(k+1)} - E^{(k)}\|_F^2 + \eta \|E^{(k)} - E^{(k)}\|_F^2, \end{aligned} \quad (4-8)$$

因而

$$\mathbf{F}_1(A^{(k+1)}, E^{(k)}) - \mathbf{F}_1(A^{(k+1)}, E^{(k+1)}) \geq \eta \|E^{(k+1)} - E^{(k)}\|_F^2. \quad (4-9)$$

合并 (4-7) 和 (4-9), 得到

$$\begin{aligned} & \mathbf{F}_1(A^{(k)}, E^{(k)}) - \mathbf{F}_1(A^{(k+1)}, E^{(k+1)}) \\ & \geq \eta (\|A^{(k+1)} - A^{(k)}\|_F^2 + \|E^{(k+1)} - E^{(k)}\|_F^2). \end{aligned} \quad (4-10)$$

因此  $\mathbf{F}_1(A^{(k)}, E^{(k)})$  单调递减。对  $k = 1, 2, \dots, t_0$  情况下时的 (4-10) 求和, 有

$$\mathbf{F}_1(A^{(1)}, E^{(1)}) - \mathbf{F}_1(A^{(t_0+1)}, E^{(t_0+1)}) \geq \eta \sum_{k=1}^{t_0} (\|A^{(k+1)} - A^{(k)}\|_F^2 + \|E^{(k+1)} - E^{(k)}\|_F^2). \quad (4-11)$$

因此由  $\mathbf{F}_1(A, E)$  非负可以得到

$$\mathbf{F}_1(A^{(1)}, E^{(1)}) \geq \eta \sum_{k=1}^{\infty} (\|A^{(k+1)} - A^{(k)}\|_F^2 + \|E^{(k+1)} - E^{(k)}\|_F^2). \quad (4-12)$$

故结论成立。  $\square$

如上所述, 我们可以利用 PBCD 算法将问题 (4-3) 分解成两个子问题 (4-4) 和 (4-5) 进行解决。由于当函数  $g(x)$  有下界时, (4-4) 能够被转化为下面问题进行求解 [15]:

$$\arg \min_{x \geq 0} f_y(x) = \frac{1}{2} (y - x)^2 + \lambda g(x), \quad (4-13)$$

其中  $y \geq 0$ 。因此在这一节的剩余部分, 我们主要讨论子问题 (4-5)。另外, 由于问题 (4-5) 中的各个  $E_{\mathcal{G}_i}$  都是独立的, 我们只需要考虑如下等价问题

$$\min_x \frac{1}{2} \|y - x\|_2^2 + \lambda g(\|x\|_2) \quad (4-14)$$

其中  $y$  和  $x$  为向量。在下面的部分中, 我们研究了 (4-14) 一种更一般的形式

$$\min_x \mathbf{F}(x) = \frac{1}{p} \|y - x\|_p^p + \lambda g(\|x\|_p), \quad (4-15)$$

其中  $p \geq 1$  和  $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$ ,  $x_i$  为向量  $x$  第  $i$  个位置上的元素。当  $y = \mathbf{0}$  时, 容易得到 (4-15) 的最优解为  $\mathbf{0}$ 。

因此, 我们集中解决问题 (4-15) 在  $y \neq \mathbf{0}$  时的情况, 并且定义

$$\begin{cases} \mathcal{A} = \arg \min_x F(x) = \frac{1}{p} \|y - x\|_p^p + \lambda g(\|x\|_p), \\ \mathcal{B} = \left\{ x^* \frac{y}{\|y\|_p} \mid x^* \in \arg \min_x f(x) = \frac{1}{p} \|y\|_p - |x|^p + \lambda g(|x|) \right\}. \end{cases}$$

为了方便起见, 我们分别定义  $F(x)$  和  $f(x)$  的最小值为  $\alpha = \min_x F(x)$  和  $\beta = \min_x f(x)$ 。则我们有以下结论。

**引理 6** 如果  $y \neq \mathbf{0}$ ,  $p \geq 1$ ,  $g(\cdot)$  是下半连续函数, 且对于任意的  $x \in \mathbb{R}$  都满足  $g(|x|) \geq 0$ , 则  $\alpha = \beta$ 。

基于引理 6, 我们得到我们第一个主要的结果如下。

**定理 4** 若  $p \geq 1$ ,  $g(\cdot)$  是一个下半连续函数,  $g(0) = 0$ , 且对任意的  $x \in \mathbb{R}$  都有  $g(|x|) \geq 0$  成立, 则

$$\arg \min_x F(x) = \begin{cases} \{\mathbf{0}\}, & \text{if } y = \mathbf{0}; \\ \mathcal{B}, & \text{if } y \neq \mathbf{0}. \end{cases}$$

**证明** 显然, 当  $y = \mathbf{0}$  时,  $\mathcal{A} = \{\mathbf{0}\}$ 。因此只需证明当  $y \neq \mathbf{0}$  时有  $\mathcal{A} = \mathcal{B}$  成立即可。从一方面来讲, 对于任意的  $x^* \frac{y}{\|y\|_p} \in \mathcal{B}$ , 由引理 6 都能得到

$$F\left(x^* \frac{y}{\|y\|_p}\right) = \frac{1}{p} \|y\|_p - |x^*|^p + \lambda g(|x_0|) = \beta = \alpha, \quad (4-16)$$

由此得到  $x^* \frac{y}{\|y\|_p} \in \mathcal{A}$ 。因此,  $\mathcal{B} \subseteq \mathcal{A}$ 。

从另一方面来讲, 令  $x^* \in \mathcal{A}$ , 我们证明  $x^* \in \mathcal{B}$ 。当  $x^* = \mathbf{0}$  时, 则由引理 6 我们可以得到  $\beta = \alpha = F(x^*) = \frac{1}{p} \|y\|_p^p = f(0)$ , 即,  $\mathbf{0} \in \mathcal{B}$ 。否则, 我们得到

$$\begin{aligned} \alpha = \beta &\leq \frac{1}{p} \|y\|_p - \|x^*\|_p^p + \lambda g(\|x^*\|_p) \\ &\leq \frac{1}{p} \|y - x^*\|_p^p + \lambda g(\|x^*\|_p) = \alpha, \end{aligned} \quad (4-17)$$

这意味着

$$\|y\|_p - \|x^*\|_p = \|y - x^*\|_p, \quad (4-18)$$

$$\beta = \frac{1}{p} \|y\|_p - \|x^*\|_p^p + \lambda g(\|x^*\|_p). \quad (4-19)$$

由 (4-18), 存在  $c_0 \geq 0$  使得  $x^* = c_0 \frac{y}{\|y\|_p}$  成立。容易得到  $c_0 = \|x^*\|_p$ 。由 (4-19) 可知,  $f(x)$  在  $x = c_0$  处取得最小值。因此  $x^* = \|x^*\|_p \frac{y}{\|y\|_p} \in \mathcal{B}$ 。证明完成。  $\square$

**引理 7** 对于  $y > 0$ , 定义

$$\min_x \frac{1}{p} |y - x|^p + \lambda g(|x|) \quad (4-20)$$

的解为  $x^*$ , 则  $x^* \geq 0$ 。



由定理 4 和引理 7, 问题 (4-14) 被转为下面一维问题进行解决

$$\arg \min_{x \geq 0} f_y(x) = \frac{1}{2}(y - x)^2 + \lambda g(x), \quad (4-21)$$

其中  $y \geq 0$ 。

将上述结论和 GAI 算法相结合可以得到下面问题的一个一般解法:

$$\arg \min_X \frac{1}{2} \|Y - X\|_F^2 + \lambda \sum_{\mathcal{G}_i \in \mathcal{G}} g(\|X_{\mathcal{G}_i}\|_2). \quad (4-22)$$

算法 3 总结了问题 (4-22) 的完整求解过程。由此, 我们得到第二类低秩恢复框架的一个通用算法。

---

**算法 3** 对于问题 (4-22) 的一个优化过程

---

输入:  $Y \in \mathbb{R}^{m \times n}$  和  $\mathcal{G}$ 。

1: 对  $i = 1, 2, \dots, |\mathcal{G}|$  执行:

    若  $\|Y_{\mathcal{G}_i}\|_F = 0$ , 令  $C_i = \{\mathbf{0}\}$ ;

    否则  $C_i = \{x^* \frac{Y_{\mathcal{G}_i}}{\|Y_{\mathcal{G}_i}\|_F} | x^* \in \arg \min_{x \geq 0} \frac{1}{2} (\|Y_{\mathcal{G}_i}\|_F - x)^2 + \lambda g(x)\}$ 。

2: 令  $X_{\mathcal{G}_i}^* \in C_i$  对于  $i = 1, 2, \dots, |\mathcal{G}|$ 。

输出: 问题 (4-22) 的最优解  $X^*$ 。

---

## 4.3 实验结果与分析

### 4.3.1 异常识别

为了证明 GAI 和算法 3 在异常值追踪问题中的有效性, 我们在 MNIST 手写数字数据集和 Hopkins 155 数据集上对不同的方法 (包括  $L_1$ 、 $L_p$ 、 $L_p(\text{Wang})$ 、LSP、Laplace、LOG、Logarithm 和 ETP) 进行比较, 其中对于被比较方法具体解释如下:

- 框架 (4-2) 中的  $\|\cdot\|_{\mathcal{G},g}$  被取为  $\sum_{i=1}^n g(\|E_i\|_2)$ 。
- $L_1$ : 算法 3 和软阈值被用于求解问题 (4-21)。
- $L_p(\text{Wang})$ : 广义低秩稀疏分解模型中的  $g$  被取为  $L_p$  范数。[22] 中的命题 2 与 [27] 中的广义软阈值 (Generalized Soft-Thresholding, GST) 分别被用于求解 (4-5) 和 (4-4)。
- 当广义低秩稀疏分解模型中的  $g$  分别为  $L_p$  范数, LSP 惩罚项, Laplace 惩罚项, LOG 惩罚项, Logarithm 惩罚项以及 ETP 惩罚项时, 利用算法 3 和 GAI 求解模型的方法被分别简记为  $L_p$ , LSP, Laplace, LOG, Logarithm 以及 ETP。

我们重复实验 20 次，Ham 和 F1 度量 [28] 的平均值用于评价不同方法的有效性。其中

- Ham: Ham 度量的具体公式为  $\text{Ham} = \frac{100-h}{100}$ ，其中  $h$  为  $A_0$  的列支撑集和  $A^*$  的支撑集之间的汉明距离 [29]，其中  $A_0$  为真实值， $A^*$  为由求解 (4-2) 所得到的解。
- F1 度量: F1 度量的具体公式为  $F1 = \frac{2 \times P \times R}{P+R}$ ，其中  $P = \frac{TP}{TP+FN}$ ， $R = \frac{TN}{TN+FP}$ 。这里的  $TP$ ， $FP$ ， $TN$  以及  $FN$  分别表示真阳性，假阳性，真阴性以及假阴性。

#### 4.3.1.1 在 MNIST 手写数字数据集上的实验结果

表 4-1: 在 MNIST 手写数字数据集上，不同方法的 Ham 结果比较

Table 4-1: Comparison of Ham for anomalies identification on MNIST digit dataset.

Video Clip	c=5	c=10	c=15	c=20
$L_1$	0.9680	0.9460	0.9060	0.8800
$L_p$	0.9760	0.9480	0.9220	0.8840
$L_p$ (Wang)	0.9780	0.9460	0.9160	0.8940
LSP	0.9420	0.8960	0.8440	0.7980
Laplace	0.9680	0.9480	0.8960	0.8620
LOG	0.9840	0.9620	0.9260	0.8940
Logarithm	0.9780	0.9560	0.9180	0.8980
ETP	0.9480	0.9020	0.8560	0.7960

表 4-2: 在 MNIST 手写数字数据集上，不同方法的 F1 度量结果比较

Table 4-2: Comparison of F1 measurement for anomalies identification on MNIST digit dataset.

Video Clip	c=5	c=10	c=15	c=20
$L_1$	0.7600	0.7200	0.7133	0.7150
$L_p$	0.8000	0.7300	0.7200	0.7250
$L_p$ (Wang)	0.7800	0.7200	0.7133	0.7050
LSP	0.4600	0.4100	0.4800	0.4800
Laplace	0.7800	0.6900	0.6800	0.6600
LOG	0.8200	0.7900	0.7600	0.7350
Logarithm	0.7800	0.7800	0.7267	0.7450
ETP	0.4200	0.5400	0.4800	0.5150

我们选取  $100 - c_0$  个数字“0”作为潜在样本，从数字“1-9”中随机选取  $c_0$  个样本作为异常值。观测矩阵由所有这些样本数据构成。

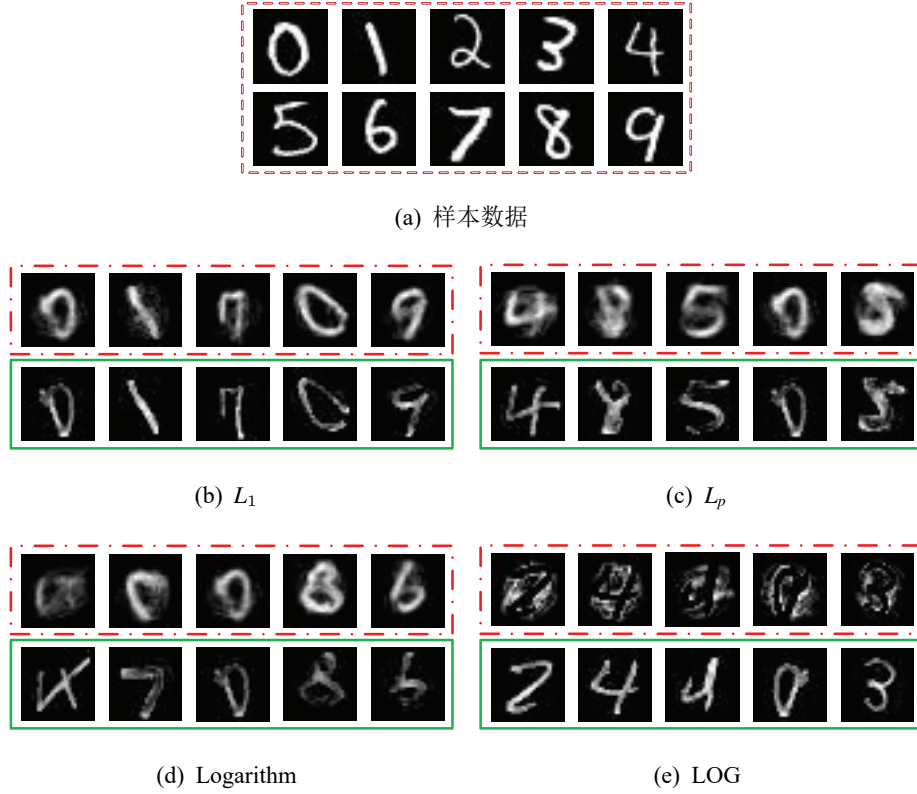


图 4-1: 由不同方法得到的异常值识别结果。由红线和蓝线包围的部分分别为低秩部分  $X^*$  和稀疏部分  $E^*$  的视觉表示。

Figure 4-1: Identified outliers by different surrogates. The parts surrounded by red lines and blue lines are the visual representation of the low rank part  $X^*$  and sparse part  $E^*$ , respectively.

由表格4-1和表格4-2可以看出多数非凸方法通常要比  $L_1$  更有效 (特别是 LOG)。其原因在于：相比较  $L_1$  而言，基于非凸替代方法的模型更接近  $L_0$ 。因此更容易得到我们所期望的低秩解和稀疏解，从而能够更准确地检测出异常值。图 4-1 比较了当  $c_0 = 5$  时，由不同方法所得到的低秩稀疏分解结果。因为  $L_p$ 、LOG 和 Logarithm 在 Ham 值和 F1 度量中都有较好的表现，所以我们只在图中呈现由这三种方法以及  $L_1$  所得到的结果。从图4-1可以看出，由 LOG 提取的稀疏噪声最为清晰明显。这直观地证明了非凸方法的有效性。

#### 4.3.1.2 在 Hopkins 155 数据集上的实验结果

在这部分，我们利用数据集 Hopkins 155 中的视频序列“1R2RC”，“1R2RCR”和“three-cars” (见图4-2) 对不同方法进行测试，其中对测试数据的详细介绍如下：

- (1) “1R2RC”：视频序列包含三种不同类型的运动，包括由两个物体和相机本身所产生的运动。1R2RC 表示第一个对象和第二个对象旋转，相机固定。

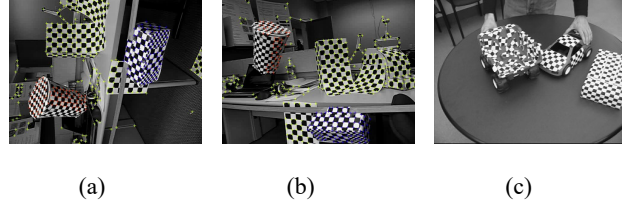


图 4-2: Hopkins 155 中的三个例子: (a) “1R2RC”; (b) “1R2RCR”; (c) “three-cars”。

Figure 4-2: Examples in Hopkins 155: (a) “1R2RC”; (b) “1R2RCR”; (c) “three-cars”。

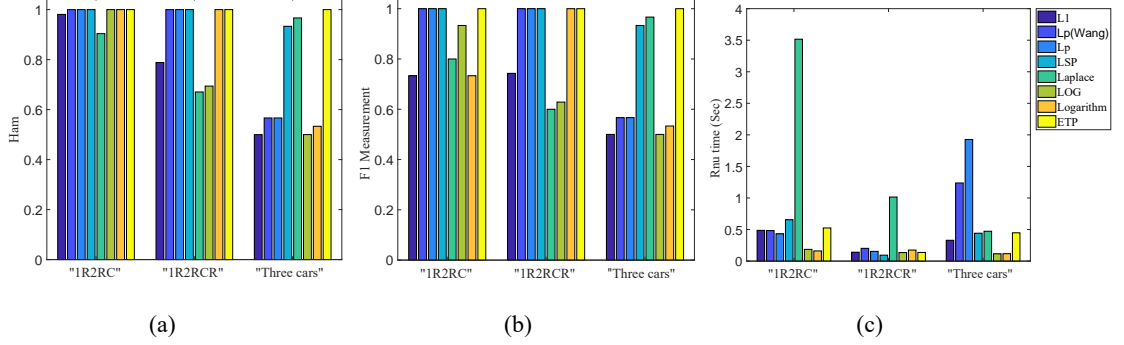


图 4-3: 不同范数在 (a) Ham 结果; (b) F1 度量; (c) 运行时间 (秒) 三个方面的比较。

Figure 4-3: Comparison of (a) Ham; (b) F1 measurement; (c) Running time (seconds) for anomalies identification on three datasets.

- (2) “1R2RCR”: 视频序列包含三种不同类型的运动, 包括由两个物体和相机本身所产生的运动。“1R2RCR”表示两个对象和相机都旋转。
- (3) “three-cars”: 视频序列 “three-cars” 包含两辆玩具车的和一个由固定摄像机拍摄的在平面 (桌子) 上移动的盒子所产生的三个动作。

如 [30] 所述, 每个移动对象上的角点跨越多个帧的轨迹所构成的向量分布于一个低维空间之中。而不同运动对象上角点轨迹所得向量分别属于不同的低维子空间。因此, 在这个部分, 我们将每个视频序列的第一个运动对象中选择  $c_1$  个角点的轨迹所得向量 (每个角点的轨迹所构成的是  $m_0$  维的向量) 作为潜在样本, 然后采用第二个运动对象的  $c_2$  个角点的轨迹所得向量作为异常值。在实验中, 对于不同的视频序列 (包括 “1R2RC”, “1R2RCR” 和 “three-cars”), 参数  $(c_1, c_2, m_0)$  分别被设置为  $(89, 15, 59)$ ,  $(50, 35, 49)$  和  $(30, 30, 31)$ 。所有的实验结果都被呈现在图 4-3。

从图 4-3(a) 和 (b) 可以看出, 非凸方法在大多数情况下都要优于  $L_1$ 。其中, Laplace 和 ETP 在 Ham 和 F1 两个方面都达到了显著的效果, 其中 ETP 的 F1 度量和 Ham 甚至达到 1。从运行时间上看 (见图 4-3(c)), 尽管 ETP 在求解子问题 (4-21) 的过程中使用了迭代算法 GAI, 但仍是取得了较好的结果 (超过  $L_1$ )。这也

说明了 GAI 的高效性。



## 第五章 基于非凸低秩恢复模型的非局部图像去噪方法

随着多媒体与信息技术的发展,在计算机视觉与图像处理领域,图像去噪一直以来都是一个重要且基础的课题。图像去噪的主要任务是在去除图片中各类噪声的同时还能保留图片中的细节和纹理信息。在过去近十年的时间里,已有大量文章提出了较为优良的图像去噪方法。现有的图像去噪方法主要分为局部方法 [31–37] 和非局部方法 [38,39]。局部方法通常通过在图片的局部空间域核卷积运算获得。由于局部方法没有用到全局信息和结构,这常常会使得去噪后的图片出现模糊和细节损失。而非局部图像去噪方法由于利用了自然图片的自相似特性,往往会达到更好的去噪效果。其中非局部均值算法 (Nonlocal mean method, NLM) 的提出大大促进了图像去噪这个领域的发展。NLM 充分的利用自然图像的自相似性,以像素点为中心的图像块之间的欧式距离来确定图像块与块之间的相似性程度,对这些相似块进行简单的加权平均后就能估计出参考块的中心点的值,其去噪效果优于以往的其他算法。

受 NLM 的非局部去噪思想的启发,后又出现了大量基于非局部去噪方法。基于非局部相似块的图像去噪方法的主要思想是利用整张图片 (或是在一个相当大的领域内) 的不同相似块之间的相似性来改进图像去噪的效果。另外,一个潜在的假设是如果将那些相似块列向量化后所构成的矩阵近似的被看作是低秩的。这种低秩假设被直接或间接地用于基于非局部相似块的图像去噪算法中,例如 [25,26,38,40,41]。Guo 等人 [24] 提出了一种有效处理高斯噪声的两步去噪方法: (1) 对由非局部图像块组成的矩阵应用奇异值分解 (SVD); (2) 截断小奇异值并重构去噪后的图像块。在文献 [25] 中,图像去噪是通过同时学习非局部图像块的字典稀疏编码来实现的。基本假设是非局部图像块应该在字典中共享相同的稀疏原子。Dabov 等人 [40] 提出了一种协同图像去噪算法,该算法对相似的图像块进行稀疏 3D 变换,并在变换域中使用协同维纳滤波器去除图像噪声。Dong 等人 [26] 提出了一种基于去噪图像方差估计的空间自适应迭代奇异值阈值图像去噪算法。Peng 等人 [42] 提出了一种有效的迭代重加权方案来解决加权  $L_1$  最小化问题,以提升图像去噪算法的性能。Gu 等人 [41] 则提出了一种基于加权核范数极小化的图像去噪算法,并通过加权奇异值阈值算子得到不动点解。

虽然上述基于低秩恢复的去噪方法已经取得了很大的成功,但是它们依然存在以下局限性: 因为所有的去噪方法都是针对零均值高斯噪声设计的,因此它们对异常值 (如脉冲噪声) 非常敏感。即使图片中只是出现单个异常值,也可能会使

去噪结果出现较大的误差。而事实上，在实际的图像去噪应用中，零均值高斯噪声的假设并不令人满意。相比较零均值高斯噪声而言，零均值高斯噪声和脉冲噪声的混合模型是实际图像去噪应用中经常遇到的一种更为合理的模型。为了解决这些问题，我们将鲁棒主成分分析模型应用于图像去噪问题，并利用参数化方法给出了  $L_0$  范数的非凸替代用以取代模型中的  $L_0$  范数和秩函数。

### 5.1 基于 $L_0$ 范数的参数化非凸替代的低秩恢复模型

因为传统的混合噪声去除方法 [43–47] 通常需要先检测脉冲噪声的位置，再利用中值滤波器去除脉冲噪声，最后再处理零均值高斯噪声。然而当脉冲噪声类型为随机脉冲噪声时，通常是难以被检测的。这使得传统混合噪声去除方法的去噪效果在随机脉冲噪声情况下被大大减弱。

为了解决这个问题，我们将 RPCA 引入到高斯脉冲混合噪声去除问题中去：

$$\min_{L, E} \frac{1}{2} \|P - L - E\|_F^2 + \alpha \text{rank}(L) + \beta \|E\|_0, \quad (5-1)$$

其中  $\|\cdot\|_F^2$  和  $\|\cdot\|_0$  被分别用来约束零均值高斯噪声和脉冲噪声。当权衡系数  $\beta = +\infty$ ，则上述模型将退化为 (2-2)。因此，该模型可以同时处理零均值高斯噪声、脉冲噪声以及其它任何可以通过混合这两种噪声来近似的噪声。

因为秩函数和  $L_0$  范数都是非凸非连续的，所以上述优化问题难以被直接求解。由于秩函数能被看作是矩阵奇异值的  $L_0$  范数。因此为了得到一个可解的优化问题，一种常见方法是将原问题中的  $L_0$  范数和秩函数松弛为相应的凸包络  $L_1$  范数和核范数进行求解。因此问题被转换为

$$\min_{L, E} \frac{1}{2} \|P - L - E\|_F^2 + \alpha \|L\|_* + \beta \|E\|_1. \quad (5-2)$$

由图5-1可见  $L_0$  范数和  $L_1$  范数两者之间的差异。当优化问题中的正则项为  $\|x\|_0$  时，向量  $x$  中所有的非零元素都得到由正则系数所决定的相同的惩罚。而当正则项取  $\|x\|_1$  时，向量  $x$  中的非零元素受到的惩罚与其绝对值成正比，即元素绝对值越小受到相应的惩罚越小。这迫使出现优化问题中最优解的很多元素绝对值可能很小却不为零的情况。另一方面正如 [48] 所言，只有当某些严格的条件被满足时，(5-2)才能精确恢复真实解，或是得到与真实值的近似解。因此在很多实际问题中，凸模型 (5-2) 可能无法得到理想的稀疏解和低秩解。为了解决这一问题，我们利用参数化方法给出一个 (5-1) 的非凸近似。

首先，重新构造  $L_0$  函数如下，

$$H(x) = \begin{cases} 1, & \text{if } x > 0; \\ 0, & \text{if } x = 0. \end{cases} \quad (5-3)$$



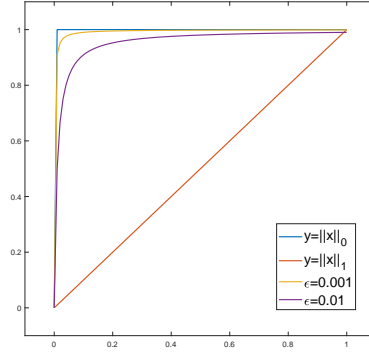

 图 5-1:  $L_1$  函数,  $L_0$  函数以及参数化范数在不同的参数  $\epsilon$  下的图像。

 Figure 5-1: Curves of  $L_1$  function,  $L_0$  function and the parameterized norm with respect to different  $\epsilon$ .

为了构造一个接近  $L_0$  函数曲线的连续可微函数, 我们将参数化部分引入公式 (5-3),

$$\mathbf{H}_\epsilon(x) = \begin{cases} 1 + h(x, \epsilon), & \text{若 } x > 0; \\ 0, & \text{若 } x = 0. \end{cases}, \epsilon > 0, \quad (5-4)$$

并使  $\mathbf{H}_\epsilon(x)$  满足定义2。

定义 2 对于  $x \in [0, +\infty)$ ,

- 1)  $0 \leq \mathbf{H}_\epsilon(x) \leq 1$ ;
- 2)  $\mathbf{H}_\epsilon(x)$  对变量  $x$  连续且可导;
- 3)  $\lim_{\epsilon \rightarrow 0} \mathbf{H}_\epsilon(x) = \mathbf{H}(x)$ 。

为了满足这些条件, 我们定义  $h(x, \epsilon) = -\frac{\epsilon}{\epsilon + x}, x > 0, \epsilon > 0$ 。因此, 本文基于  $\mathbf{H}_\epsilon(x)$  给出如下定义,

$$\|X\|_{PN}^\epsilon := \sum_{i=1}^m \sum_{j=1}^n \left(1 - \frac{\epsilon}{\epsilon + |X_{ij}|}\right), \epsilon > 0,$$

其中  $X_{ij}$  是矩阵  $X \in \mathbb{R}^{m \times n}$  (假设  $m < n$ ) 中位于第  $i$  行第  $j$  列的元素,  $PN$  用来表示参数化范数。根据上述定义, 有下面一组等式成立。

$$\begin{cases} \lim_{\epsilon \rightarrow 0} \left(1 - \frac{\epsilon}{\epsilon + |X_{ij}|}\right) = 1, X_{ij} \neq 0, \\ \lim_{\epsilon \rightarrow \infty} \epsilon \left(1 - \frac{\epsilon}{\epsilon + |X_{ij}|}\right) = |X_{ij}|. \end{cases}$$

显而易见, 如果  $\epsilon \rightarrow 0$ ,  $\|X\|_{PN}^\epsilon \rightarrow \|X\|_0$ 。如果  $\epsilon \rightarrow \infty$ ,  $\epsilon \|X\|_{PN}^\epsilon \rightarrow \|X\|_1$ 。

类似的，可以定义参数化核范数  $\|\cdot\|_{PNN}^\epsilon$  如下，

$$\|L\|_{PNN}^\epsilon = \sum_{i=1}^{\min\{m,n\}} \left(1 - \frac{\epsilon}{\epsilon + |\sigma_i(L)|}\right),$$

其中  $\sigma_i(L)$  是  $L \in \mathbb{R}^{m \times n}$  中的第  $i$  个奇异值， $PNN$  用来表示参数化范数。

因此，利用参数化核范数和参数化范数来分别替代 (5-1) 中的核范数和  $L_1$  范数，得到

$$\min_{L,E} \frac{1}{2} \|P - L - E\|_F^2 + \alpha \|L\|_{PNN}^{\epsilon_1} + \beta \|E\|_{PNN}^{\epsilon_2}, \quad (5-5)$$

## 5.2 优化算法

由于同时求解 (5-5) 中的  $L$  和  $E$  较为困难，我们利用块坐标下降算法 (The block coordinate descent method, BCD) [49] 来对模型中的  $L$  和  $E$  进行迭代求解，

$$L^{t+1} = \arg \min_L \frac{1}{2} \|P - L - E^t\|_F^2 + \alpha \|L\|_{PNN}^{\epsilon_1} \quad (5-6)$$

$$E^{t+1} = \arg \min_E \frac{1}{2} \|P - L^{t+1} - E\|_F^2 + \beta \|E\|_{PNN}^{\epsilon_2} \quad (5-7)$$

而这个优化过程的关键是找到以下两个子问题的闭式解，

$$\min_X \frac{1}{2} \|Y - X\|_F^2 + \lambda \|X\|_{PNN}^\epsilon \quad (5-8)$$

$$\min_X \frac{1}{2} \|Y - X\|_F^2 + \lambda \|X\|_{PNN}^\epsilon. \quad (5-9)$$

由于子问题 (5-8) 和 (5-9) 中的  $X$  和  $Y$  的各个元素可分离的，故可将 (5-8) 转为下述问题进行求解：

$$\begin{aligned} & \arg \min_x \frac{1}{2} (y - x)^2 + \lambda \left(1 - \frac{\epsilon}{\epsilon + |x|}\right) \\ &= \arg \min_x \frac{1}{2} (y - x)^2 - \frac{\lambda \epsilon}{\epsilon + |x|}. \end{aligned} \quad (5-10)$$

因为在  $\|X\|_{PNN}^\epsilon$  的定义中涉及到  $X$  的奇异值，故相比较 (5-8) 而言，子问题 (5-9) 更难被解决。为了克服这个困难，子问题 (5-9) 被重新定义为，

$$\arg \min_X \frac{1}{2} \|Y - X\|_F^2 + \lambda \sum_{i=1}^m b(\sigma_i(X)), \quad (5-11)$$

其中  $\sigma_i(X)$  表示  $X$  的第  $i$  个奇异值， $b(x) = 1 - \frac{\epsilon}{\epsilon + |x|}$ 。如论文 [15] 所述，当函数  $b(x)$  有下界时，下面结论成立。

$$\arg \min_X \frac{1}{2} \|Y - X\|_F^2 + \lambda \sum_{i=1}^m b(\sigma_i(X))$$

$$= U \text{Diag}(\text{Prox}_b(\sigma(Y), \lambda)) V^T, \quad (5-12)$$

其中  $U$ 、 $\sigma(Y)$  和  $V$  由  $Y$  的 SVD 分解  $Y = U \text{Diag}(\sigma(Y)) V^T$  得到,  $\text{Prox}_b(\cdot)$  定义如下

$$\text{Prox}_b(y) = \arg \min_x \frac{1}{2}(y - x)^2 + \lambda b(x) \quad (5-13)$$

此形式与 (5-10) 完全相同。由此容易看出, 当问题 (5-10) 被解决时, 则子问题 (5-8) 和 (5-9) 都能够被解决。

为了进一步简化 (5-10), 我们又给出以下结论。

**结论 1**

$$\arg \min_x \frac{1}{2}(y - x)^2 - \frac{\lambda \epsilon}{\epsilon + |x|} = \text{sign}(y) \arg \min_{x \geq 0} \left( \frac{1}{2}(|y| - x)^2 - \frac{\lambda \epsilon}{\epsilon + x} \right). \quad (5-14)$$

记  $d(x) := \frac{1}{2}(|y| - x)^2 - \frac{\lambda \epsilon}{\epsilon + x}$ 。如论文 [15] 所述有  $\arg \min_{x \geq 0} d(x) = \arg \min_{x \in \{0, \hat{x}_y\}} d(x)$  成立, 其中  $\hat{x}_y = \max\{x | d'(x) = 0, 0 \leq x \leq |y|\} \cup \{0\}$ 。因为

$$d'(x) = (x - |y|) + \frac{\lambda \epsilon}{(\epsilon + x)^2} = 0 \Rightarrow (x - |y|)(\epsilon + x)^2 + \lambda \epsilon = 0,$$

利用结论 1 和卡丹公式, 我们得到 (5-10) 的闭式解。证明详情请参阅附录。

**结论 2** 令  $\text{Prox}_{H_\epsilon}(y) = \arg \min_x \frac{1}{2}(y - x)^2 - \frac{\lambda \epsilon}{\epsilon + |x|}$ 。利用  $h_0$ ,  $\Delta$  和  $x_1$  的定义 (见表格 5-1), 我们下面结论

(1) 当  $\Delta > 0$  时,

$$\text{Prox}_{H_\epsilon}(y) = \begin{cases} \frac{x_1 + |x_1|}{2}, & \text{if, } y \geq \sqrt{2(d(\frac{x_1 + |x_1|}{2}) + \lambda)}; \\ 0, & \text{if, } |y| < \sqrt{2(d(\frac{x_1 + |x_1|}{2}) + \lambda)}; \\ -\frac{x_1 + |x_1|}{2}, & \text{if, } y \leq -\sqrt{2(d(\frac{x_1 + |x_1|}{2}) + \lambda)}. \end{cases} \quad (5-15)$$

(2) 当  $\Delta \leq 0$  时,

$$\text{Prox}_{H_\epsilon}(y) = \begin{cases} \frac{h_0 + |h_0|}{2}, & \text{if, } y \geq \sqrt{2(g(\frac{h_0 + |h_0|}{2}) + \lambda)}; \\ 0, & \text{if, } |y| < \sqrt{2(d(\frac{h_0 + |h_0|}{2}) + \lambda)}; \\ -\frac{h_0 + |h_0|}{2}, & \text{if, } y \leq -\sqrt{2(d(\frac{h_0 + |h_0|}{2}) + \lambda)}. \end{cases} \quad (5-16)$$

为了进一步验证参数化范数的有效性, 我们首先回顾以下一维形式的  $L_0$  最小化问题,

$$\text{Prox}_H(y) = \arg \min_x \frac{1}{2}(y - x)^2 + \lambda H(x). \quad (5-17)$$

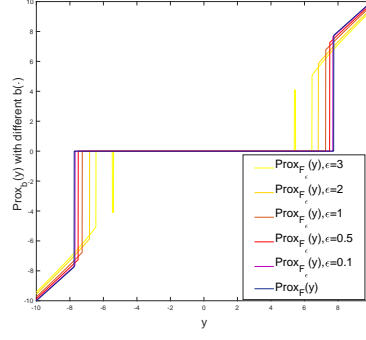


图 5-2:  $\text{Prox}_b(y) = \arg \min_x \frac{1}{2}(y-x)^2 + \lambda b(x)$ , 其中  $\lambda = 30$ 。

Figure 5-2:  $\text{Prox}_b(y) = \arg \min_x \frac{1}{2}(y-x)^2 + \lambda b(x)$ , where  $\lambda = 30$ .

表 5-1:  $h_0, x_1, x_2, x_3, \Delta, p$  和  $q$  的定义。

Table 5-1: Definitions of  $h_0, x_1, x_2, x_3, \Delta, p$  and  $q$ .

一些特殊参数的定义
$h_0 = \frac{1}{2}( \frac{ x_1-x_2 +x_1+x_2}{2} - x_3  + \frac{ x_1-x_2 +x_1+x_2}{2} + x_3)$
$x_1 = \sqrt[3]{-\frac{q}{2} + \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}} - \frac{2\epsilon- y }{3}$
$x_2 = \frac{-1 + \sqrt{3}i}{2} \sqrt[3]{-\frac{q}{2} + \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}} + (\frac{-1 + \sqrt{3}i}{2})^2 \sqrt[3]{-\frac{q}{2} - \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}} - \frac{2\epsilon- y }{3}$
$x_3 = (\frac{-1 + \sqrt{3}i}{2})^2 \sqrt[3]{-\frac{q}{2} + \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}} + \frac{-1 + \sqrt{3}i}{2} \sqrt[3]{-\frac{q}{2} - \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}} - \frac{2\epsilon- y }{3}$
$\Delta = (\frac{q}{2})^2 + (\frac{p}{3})^3$
$p = -\frac{(\epsilon +  y )^2}{3}$
$q = \lambda\epsilon - \frac{2}{27}(\epsilon +  y )^3$

上述问题的闭式解如下：

$$\text{Prox}_H(y) = \begin{cases} |y|, & \text{if } y \geq \sqrt{2\lambda}; \\ 0, & \text{if } |y| < \sqrt{2\lambda}; \\ -|y|, & \text{if } y \leq -\sqrt{2\lambda}. \end{cases} \quad (5-18)$$

由图5-2所示， $\epsilon$  值越小， $\text{Prox}_{H_\epsilon}(y)$  越接近  $\text{Prox}_H(y)$ 。下面给出该结论的严格证明。

**结论 3**  $\forall |y|, \forall \epsilon_0 > 0, \exists \delta_{\epsilon_0} > 0$  使得，当  $|\epsilon| < \delta_{\epsilon_0}$ ， $|\text{Prox}_{H_\epsilon}(y) - \text{Prox}_H(y)| < \epsilon_0$ 。

综上所述，一旦得到 (5-8)和 (5-9)的闭式解，就可以利用块坐标下降算法对低秩恢复模型 (5-5)进行迭代求解。现对该算法的理论分析如下。

**结论 4** 令

$$C(E, L) := \frac{1}{2} \|P - L - E\|_F^2 + \alpha \|L\|_{PNN}^{\epsilon_1} + \beta \|E\|_{PN}^{\epsilon_2},$$

则

$$C(E^{k+1}, L^{k+1}) \leq C(E^k, L^k).$$

**证明** 由  $L^{k+1}$  的定义, 得到

$$\frac{1}{2} \|P - L^{k+1} - E^k\|_F^2 + \alpha \|L^{k+1}\|_{PNN}^{\epsilon_1} \leq \frac{1}{2} \|P - L^k - E^k\|_F^2 + \alpha \|L^k\|_{PNN}^{\epsilon_1}$$

因此,

$$C(E^k, L^{k+1}) = \frac{1}{2} \|P - L^{k+1} - E^k\|_F^2 + \alpha \|L^{k+1}\|_{PNN}^{\epsilon_1} + \beta \|E^k\|_{PN}^{\epsilon_2} \leq C(E^k, L^k)$$

由  $E^{k+1}$  的定义得到

$$\frac{1}{2} \|P - L^{k+1} - E^{k+1}\|_F^2 + \beta \|E^{k+1}\|_{PN}^{\epsilon_2} \leq \frac{1}{2} \|P - L^{k+1} - E^k\|_F^2 + \beta \|E^k\|_{PN}^{\epsilon_2}$$

因此  $C(E^{k+1}, L^{k+1}) \leq C(E^k, L^{k+1}) \leq C(E^k, L^k)$ 。因此, 定理成立。  $\square$

又因为函数  $C(E, L) \geq 0$  有下界, 故问题 (5-5) 可以由该迭代算法进行求解。

### 5.3 基于非凸低秩恢复的非局部图像去噪算法

基于模型 (5-6), 本文给出一个新的图像去噪 (见算法4)。

---

#### 算法 4 去噪框架

---

输入: 一系列由相似块得到的矩阵  $P = [p, p_1, p_2 \cdots p_k]$ ,  $\alpha, \beta, \epsilon_1, \epsilon_2$ 。

- 1: 初始化稀疏噪声项  $E^{(0)} = 0$ , 并使  $t = 0$ 。
- 2: 计算  $L^{(t+1)}$ : 利用结论 2 中的结论求解子问题 (5-6)。
- 3: 计算  $E^{(t+1)}$ : 利用 (5-12) 和结论 2 求解子问题 (5-7)。
- 4: 令  $t = t + 1$ 。若算法不收敛, 则返回第 2 步。

输出:  $L = L^{(t+1)}$ ,  $E = E^{(t+1)}$ 。

---

### 5.4 实验结果

为了进一步验证算法 4 的有效性, 我们在 Lansel 标准数据集 [50] 上分别对零均值高斯噪声和零均值高斯脉冲混合噪声两种情况进行实验。

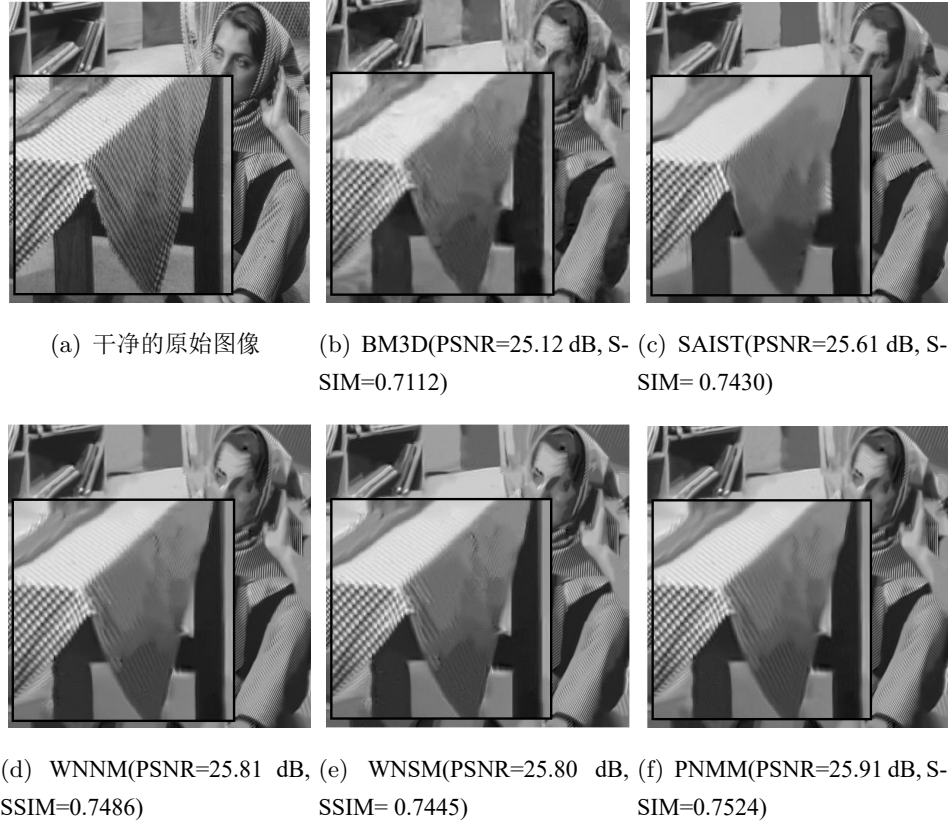


图 5-3: 当噪声水平  $\sigma = 75$  时, 不同方法在图像“barbara”上的实验结果。“SSIM”表示为结构相似性。

Figure 5-3: Results on the image ‘barbara’ with different methods (noise level  $\sigma = 75$ ). ‘SSIM’ stands for structural similarity.

#### 5.4.1 零均值高斯噪声

在这一部分中, 我们将算法4(当  $\beta = +\infty$ , 该算法被称为参数化核范数最小化方法, 简称为 PNMM) 与四种最新方法进行了比较。这些方法包括块匹配三维滤波 [40](BM3D)、空间自适应迭代奇异值阈值化 [26](SAIST)、加权核范数最小化 [41](WNNM) 和加权 Schatten p 范数最小化 [51](WNSM)。具体实验细节介绍如下: (1) 分别在零均值高斯噪声标准差  $\sigma = \{30, 50, 60, 75, 85, 100\}$  的六种噪声水平上对不同的方法进行测试; (2) 算法 4 中的参数  $\epsilon_1$  被设置为 0.01。 (3) 峰值信噪比 (Peak signal to noise ratio, PSNR) 被用来评估不同方法的去噪效果。PSNR 的数值越大则去噪效果越好。 (4) 所有被比较方法的去噪结果是由其作者提供的源代码所得到。

五种不同去噪算法的 PSNR 结果在表 5-2 中被给出。由表 5-2 可以观察到, 在大多数情况下, PNMM 的去噪效果都是优于其它方法的。当  $\sigma$  较小时, BM3D、SAIST、WNNM、WNSM 和 PNMM 的 PSNR 值非常接近。但随着噪声水平的增加, 两者差异逐渐变得显著。另外, 我们还在视觉角度上对各种算法进行了比较。



图 5-4: 当噪声水平  $\sigma = 100$  时, 不同方法在图像“hill”上的实验结果。“SSIM”表示为结构相似性。

Figure 5-4: Results on the image ‘hill’ with different methods (noise level  $\sigma = 100$ ). ‘SSIM’ stands for structural similarity.

由图5-3和5-4可以看出, 与 BM3D 和 SAIST 相比, WNNM、WNSM 和 PNMM 保留了更多关于图像的细节信息, 如图中桌布和裤子的纹理。此外, 比较 WNNM、WNSM 和 PNMM 三种方法可以发现, 由 WNNM 和 WNSM 得到去噪后图像出现了较多光斑。其原因是相比较加权核范数和加权 Schatten p 范数而言, PNN 能够得到更精确的低秩解, 因此对噪声和奇异点具有更强的鲁棒性。总之, 无论是从数值角度还是视觉角度, 相比较其他方法而言, PNMM 都达到了比较好的去噪效果。

#### 5.4.2 零均值高斯脉冲混合噪声

在这一部分中, 我们将算法 4(被简单记为 PNN+PN) 与 Cai 的工作 [43], Xiao 的工作 [46] 以及基于凸模型 (5-2) 的去噪方法 (被简单记为 NN+ $L_1$ )。这里, NN+ $L_1$  和 PNN+PN 的唯一不同之处在于 NN+ $L_1$  是基于 (5-2) 进行算法 4 中的低秩恢复步骤的。由于随机脉冲噪声比椒盐噪声更为一般且更难处理, 故在此部分, 我们选择随机值脉冲噪声和零均值高斯噪声的混合噪声来对这些方法进行测试。噪声参数设置与 [43] 中相同: 零均值高斯噪声标准差为  $\sigma = \{5, 10, 15\}$  和随机值脉冲噪声密度为  $r = \{10\%, 30\%, 50\%, 100\%\}$ 。与上个实验相同, 峰值信噪比 (PSNR) 被用



表 5-2: 不同方法在零均值高斯噪声下的 PSNR 结果

Table 5-2: Denoising results (PSNR) of different methods in case of zero mean Gaussian noise

Image	$\sigma = 30$					Image	$\sigma = 50$				
	BM3D	SAIST	WNNM	WNSM	PNMM		BM3D	SAIST	WNNM	WNSM	PNMM
lena512	31.26	31.27	31.43	31.44	<b>31.47</b>	lena512	29.05	29.01	29.25	29.24	<b>29.29</b>
monarch	28.36	28.68	28.91	<b>28.92</b>	<b>28.92</b>	monarch	25.82	25.89	26.32	26.33	<b>26.34</b>
barbara	29.81	30.09	30.31	30.33	<b>30.39</b>	barbara	27.23	27.54	27.79	27.79	<b>27.87</b>
boat	29.12	28.93	29.23	29.25	<b>29.26</b>	boat	26.78	26.66	26.97	26.98	<b>27.02</b>
cameraman	28.64	28.46	28.78	<b>28.80</b>	28.78	cameraman	26.13	26.19	26.42	26.44	<b>26.45</b>
couple	28.87	28.68	28.97	28.98	<b>29.01</b>	couple	26.46	26.30	26.64	26.67	<b>26.70</b>
fingerprint	26.83	26.98	26.96	27.00	<b>27.02</b>	fingerprint	24.53	24.55	<b>24.66</b>	24.60	24.65
hill	29.16	28.99	29.23	29.25	<b>29.26</b>	hill	27.19	26.94	27.33	27.33	<b>27.39</b>
house	32.09	32.39	<b>32.55</b>	32.54	32.54	house	29.69	30.20	30.33	<b>30.37</b>	30.35
man	28.86	28.74	28.98	29.00	<b>29.01</b>	man	26.81	26.66	26.94	26.94	<b>26.97</b>
peppers256	29.28	29.33	<b>29.49</b>	<b>29.49</b>	29.44	peppers256	26.68	26.76	26.91	<b>26.93</b>	26.85
straw	24.94	25.46	25.44	25.48	<b>25.52</b>	straw	22.41	22.85	22.93	22.92	<b>22.99</b>
average	28.93	29.00	29.19	29.21	<b>29.22</b>	average	26.56	26.63	26.87	26.88	<b>26.91</b>

Image	$\sigma = 60$					Image	$\sigma = 75$				
	BM3D	SAIST	WNNM	WNSM	PNMM		BM3D	SAIST	WNNM	WNSM	PNMM
lena512	28.27	28.26	28.41	28.40	<b>28.47</b>	lena512	27.26	27.23	27.54	27.49	<b>27.59</b>
monarch	24.97	24.82	25.45	25.47	<b>25.51</b>	monarch	23.91	23.75	24.31	24.36	<b>24.38</b>
barbara	26.28	26.73	26.78	26.79	<b>26.87</b>	barbara	25.12	25.61	25.81	25.80	<b>25.91</b>
boat	26.02	25.94	26.17	26.19	<b>26.24</b>	boat	25.12	25.05	25.30	25.31	<b>25.36</b>
cameraman	25.32	25.22	25.63	25.66	<b>25.67</b>	cameraman	24.33	24.34	24.55	<b>24.59</b>	24.55
couple	25.66	25.60	25.79	25.81	<b>25.85</b>	couple	24.70	24.60	24.86	24.88	<b>24.91</b>
fingerprint	23.75	23.85	<b>23.82</b>	23.77	23.81	fingerprint	22.83	22.85	<b>23.03</b>	22.89	22.99
hill	26.52	26.38	26.64	26.65	<b>26.70</b>	hill	25.68	25.51	25.88	25.87	<b>25.94</b>
house	28.74	29.24	29.44	29.48	<b>29.50</b>	house	27.51	28.17	28.24	<b>28.34</b>	28.23
man	26.14	26.03	26.24	26.24	<b>26.30</b>	man	25.32	25.17	25.42	25.42	<b>25.46</b>
peppers256	25.81	25.80	26.07	<b>26.09</b>	26.07	peppers256	24.73	24.76	24.92	<b>24.99</b>	24.90
straw	21.63	22.14	22.03	22.03	<b>22.08</b>	straw	20.72	21.03	<b>21.21</b>	21.17	21.18
average	25.76	25.83	26.04	26.05	<b>26.09</b>	average	24.77	24.84	25.09	25.09	<b>25.12</b>

Image	$\sigma = 85$					Image	$\sigma = 100$				
	BM3D	SAIST	WNNM	WNSM	PNMM		BM3D	SAIST	WNNM	WNSM	PNMM
lena512	26.71	26.64	26.97	26.90	<b>27.04</b>	lena512	25.95	25.93	26.21	26.11	<b>26.31</b>
monarch	23.30	23.17	23.72	23.77	<b>23.81</b>	monarch	22.52	22.41	22.95	23.00	<b>23.06</b>
barbara	24.47	24.98	25.18	25.16	<b>25.29</b>	barbara	23.62	24.13	24.37	24.34	<b>24.48</b>
boat	24.62	24.56	24.78	24.79	<b>24.85</b>	boat	23.97	23.90	24.11	24.11	<b>24.19</b>
cameraman	23.79	23.86	24.02	<b>24.05</b>	24.02	cameraman	23.08	23.22	23.36	<b>23.38</b>	23.36
couple	24.16	24.05	24.28	24.30	<b>24.36</b>	couple	23.51	23.35	23.56	23.56	<b>23.66</b>
fingerprint	22.30	22.29	<b>22.50</b>	22.36	22.47	fingerprint	23.61	21.59	21.82	21.70	<b>21.83</b>
hill	25.20	24.98	25.40	25.39	<b>25.48</b>	hill	24.58	24.27	24.76	24.74	<b>24.88</b>
house	26.80	27.57	27.58	<b>27.65</b>	27.54	house	25.87	26.75	26.66	<b>26.69</b>	26.64
man	24.84	24.68	24.97	24.95	<b>25.02</b>	man	24.22	24.02	24.36	24.33	<b>24.42</b>
peppers256	24.15	24.14	24.29	<b>24.35</b>	24.29	peppers256	23.39	23.32	23.45	23.47	<b>23.49</b>
straw	20.23	20.40	<b>20.65</b>	20.59	20.61	straw	19.59	19.54	19.90	19.82	<b>19.92</b>
average	24.21	24.28	24.53	24.52	<b>24.56</b>	average	23.49	23.54	23.79	23.77	<b>23.85</b>



表 5-3: 在零均值高斯脉冲混合噪声下, 不同方法在图片“Barbara”上的结果。

Table 5-3: Results on Barbara with Gaussian noise and random-valued impulse noise.

$\sigma$	$r(\%)$	Cai's work	Xiao's work	NN+ $L_1$	PNN+PN
5	10	25.33	30.48	30.65	<b>33.74</b>
	20	24.72	27.76	28.78	<b>31.59</b>
	30	24.10	25.92	26.63	<b>29.17</b>
10	10	24.46	28.42	29.23	<b>31.66</b>
	20	23.99	26.59	27.64	<b>30.25</b>
	30	23.66	25.34	25.92	<b>28.19</b>
15	10	23.80	27.31	27.89	<b>29.33</b>
	20	23.60	25.69	26.64	<b>28.91</b>
	30	23.27	24.55	24.85	<b>27.04</b>

来评估不同方法的去噪效果。PNN+PN 中的参数  $\epsilon_1$ 、 $\epsilon_2$  都被设置为 0.01。

由于寻找相似块的过程极易受到脉冲噪声的干扰, 因此首先使用自适应中心加权中值滤波器 (ACWMF) 检测随机值脉冲噪声, 然后利用所检测到的脉冲噪声位置确定权重矩阵, 其中相对于脉冲噪声的位置的权重值被设置为 0.3, 否则设置为 0.7。该矩阵可用于在优化过程中对  $\|\cdot\|_{p_N}$  项进行加权。

不同方法的去噪结果被呈于表 5-3。从表中可以看出 PNN+PN 的去噪效果明显优于 Cai、Xiao 和 NN+ $L_1$  其原因是: (1) 相比较 NN+ $L_1$  而言, PNN+PN 能够得到的更精确的低秩解和稀疏解, 因而对噪声和异常值具有更强的鲁棒性。(2) 一方面由 ACWMF 得到的脉冲噪声检测结果为新模型提供了先验信息, 另一方面该模型中的稀疏误差项  $\|\cdot\|_{p_N}$  以及权重矩阵的引入可以有效地修正那些没有被 ACWMF 准确检测到的脉冲噪声位置的值。



## 第六章 论文小结与展望

### 6.1 论文小结

近年来，低秩恢复方法由于其广泛的实际应用背景以及对强噪声和缺失数据的鲁棒性而受到广泛的关注。由于原始低秩模型都是离散优化问题，因此很多模型都是 NP-难的。这对求解它们造成了很大的困难。为了克服这个困难，一个常用的办法是把它们近似成凸优化问题，即利用  $L_1$  范数和核范数来模型中的  $L_0$  范数和秩函数。由此就能将原始的 NP 难问题转化为凸问题进行解决。然而尽管对于大部分的低秩恢复模型都已经被证明：当某些条件被满足时，通过利用凸方法能够得到真实解。但是在实际应用中，这些严格的前提是难以被满足的。这意味着由凸方法得到的解可能严重偏离问题的真实解。为了解决这个问题，一些非凸替换方法被给出，即利用一些非凸函数  $g(x)$  来取代  $\|x\|_0 = \sum_i |x_i|^0$  (或是  $\text{rank}(X) = \sum_i |\delta_i(X)|^0$ ) 中的  $|\cdot|^0$  以权衡问题的可解性和有效性。这些非凸函数  $g(x)$  往往介于  $|\cdot|^0$  和  $|\cdot|^1$  之间。

为了更有效的解决由这些非凸替代所得到的优化问题，本文首先给出超线性收敛不动点算法 GAI 用于 (1-5) 的求解。文章分别从理论分析与仿真实验两个方面证明了：相比较 Lu 的工作 (算法1) 而言，算法 GAI 以更快的速度收敛到问题 (1-5) 的全局最优。于是，我们能够结合 GAI 算法，GPG 算法以及结论 (1-4) 得到形式为 (1-1) 的第一类低秩恢复问题的一个更有效的解法。对于形式为 (1-6) 的第二类低秩恢复框架，本文基于块坐标下降算法的一种变种算法 (PBCD)，定理4以及 GAI 算法给出了相应的一个有效解法。文章从理论和实验出发证明了该解法的有效性。最后，文章基于参数化方法给出了一种新的非凸替代方法 (即利用参数化方法给出参数化范数和参数化核范数)。利用这两个新非凸范数，我们给出了一个新的非凸低秩恢复模型。为了验证非凸替代方法在图像去噪问题中的有效性，我们将该模型应用于图像去噪问题，得到一种新的非局部图像去噪方法。该方法能够同时处理零均值高斯噪声和高斯脉冲混合噪声。在实验部分，文章分别在零均值高斯噪声和零均值高斯脉冲混合噪声两种不同情况下，对不同方法进行比较。实验结果表明在多数情况下，基于非凸替代的非局部图像去噪算法相比较基于凸方法的非局部图像去噪算法能够得到更精确的低秩解，相对传统的去噪算法对噪声和奇异点也具有更强的鲁棒性。

## 6.2 展望

近年来，联合稀疏表示已经被广泛地提出和研究，其目的是假设相同类别的信号在同一个字典下的稀疏系数具有相同的稀疏性分布。该假设可以转化为具有最小非零行数的系数矩阵  $A$ ，即  $A^T$  的  $L_{2,0}$  范数最小化。而求解一个基于  $L_{2,0}$  范数的联合稀疏表示问题通常是 NP 难的。一种自然的方法是将  $L_{2,0}$  范数凸松弛为  $L_{2,1}$  范数以获得一个可行的解决方案。然而，类似于  $L_1$  范数的情形， $L_{2,1}$  范数也可能导致过惩罚的情况出现。故我们同样可以考虑引入非凸方法解决这个问题。

另外，注意到模型 (1-6) 中的  $\mathcal{G}$  是被事先给定的。这在一定程度上限制了该模型的应用。未来，我们将考虑一种有效的算法用于最佳  $\mathcal{G}$  求解。

## 参考文献

- [1] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2009, 31(2):210–227
- [2] Donoho D L. For most large underdetermined systems of linear equations the minimal  $\ell_1$  norm solution is also the sparsest solution[J]. Communications on Pure & Applied Mathematics, 2006, 59(7):907–934
- [3] Wright J, Ma Y, Mairal J, et al. Sparse representation for computer vision and pattern recognition[J]. Communications on pure and applied mathematics, 2010, 98(6):1031–1044
- [4] Recht B, Fazel M, Parrilo P A. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization[J]. SIAM Review, 2010, 52(3):471–501
- [5] Emmanuel J Candès, Recht B. Exact matrix completion via convex optimization[J]. Foundations of Computational Mathematics, 2009, 9(6):717–772
- [6] Candès Emmanuel J, Li Xiaodong, Ma Yi, et al. Robust principal component analysis ?[J] Journal of the ACM, 2011, 58(3):1-37
- [7] Emmanuel J Candès, Plan Y. Matrix completion with noise[J]. Proceedings of the IEEE, 2010, 98(6):925-936
- [8] Chandrasekaran V, Sanghavi S, Parrilo P A, et al. Sparse and low-rank matrix decompositions[C]. In: Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2009. 1493-1498
- [9] Eckart C, Young G. The approximation of one matrix by another of lower rank[J]. Psychometrika, 1936, 1(3):211-218
- [10] Svante Wold, Kim Esbensen, Paul Geladi. Principal component analysis[J]. Chemometrics & Intelligent Laboratory Systems, 1987, 2(1):37-52
- [11] Wright John, Ganesh Arvind, Rao Shankar, et al. Robust principal component analysis: Exact recovery of corrupted low-rank matrices[C]. In: Advances in Neural Information Processing Systems, 2009. 2080-2088
- [12] Xu H, Caramanis C, Sanghavi S. Robust PCA via outlier pursuit[J]. IEEE Transactions on Information Theory, 2012, 58(5):3047-3064

- [13] Zhou Zihan, Li Xiaodong, Wright John, et al. Stable principal component pursuit[C]. In: IEEE international symposium on information theory, 2010. 1518-1522
- [14] Lu C, Tang J, Yan S, et al. Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm[J]. IEEE Transactions on Image Processing, 2016, 25(2):829–839
- [15] Lu C, Zhu C, Xu C, et al. Generalized singular value thresholding[C]. In: AAAI Conference on Artificial Intelligence, 2015:1805-1811
- [16] Lu C, Tang J, Yan S, et al. Generalized nonconvex nonsmooth low-rank minimization[C]. In: IEEE Conference on Computer Vision and Pattern Recognition, 2014. 4130-4137
- [17] Mazumder R, Friedman J H, Hastie T. SparseNet: Coordinate descent with nonconvex penalties[J]. Journal of the American Statistical Association, 2010, 106(495):1125-1138
- [18] Gong P, Zhang C, Lu Z, et al. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems[C]. In: International Conference on Machine Learning, 2013. 37-45
- [19] Massimo Fornasier, Holger Rauhut. Iterative thresholding algorithms[J]. Applied & Computational Harmonic Analysis, 2008, 25(2):187-208
- [20] Liu G, Lin Z, Yan S, 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
- [21] Xie Y, Zhang W, Qu Y, et al. Hyper-laplacian regularized multilinear multiview self-representations for clustering and semisupervised learning[J]. IEEE Transactions on Cybernetics, 2018, 1–15
- [22] Wang D, Zhang X, Fan M, et al. Semi-supervised dictionary learning via structural sparse preserving[C]. In: AAAI Conference on Artificial Intelligence, 2016. 2137-2144
- [23] Zhang Hongyang, Lin Zhouchen, Zhang Chao, et al. Exact recoverability of robust PCA via outlier pursuit with tight recovery bounds[C]. In: AAAI Conference on Artificial Intelligence, 2015. 3143-3149
- [24] Guo Q, Zhang C, Zhang Y, et al. An efficient SVD-based method for image denoising[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(5):868-880
- [25] Mairal J, Bach F, Ponce J, et al. Non-local sparse models for image restoration[C], In: International Conference on Computer Vision, 2009. 2272-2279
- [26] Dong W, Shi G, Li X. Nonlocal image restoration with bilateral variance estimation: a low-rank approach[J]. IEEE Transactions on Image Processing, 2013, 22(2):700-711
- [27] Zuo W, Meng D, Zhang L, et al. A generalized iterated shrinkage algorithm for non-convex sparse coding[C]. In: IEEE International Conference on Computer Vision, 2013. 217-224

- [28] Zhou Z. Machine Learning. Beijing: Tsinghua University Press, 2016. 30-33
- [29] Zhang H, Lin Z, Zhang C, et al. Exact recoverability of robust pca via outlier pursuit with tight recovery bounds[C]. In: AAAI Conference on Artificial Intelligence, 2015. 3143-3149
- [30] Tron R, Vidal R. A benchmark for the comparison of 3-d motion segmentation algorithms. In: IEEE Conference on Computer Vision and Pattern Recognition, 2007. 1-8
- [31] Tomasi C, Manduchi R. Bilateral filtering for gray and color images[C]. In: International Conference on Computer Vision, 1998. 839-846
- [32] Donoho D L. De-noising by soft-thresholding[J]. IEEE Transactions on Information Theory, 1995, 41(3):613-627
- [33] Moulin P, Liu J. Analysis of multiresolution image denoising schemes using generalized gaussian and complexity priors[J]. IEEE Transactions on Information Theory, 1999, 45(3):909-919
- [34] Chang S G, Yu B, Vetterli M. Adaptive wavelet thresholding for image denoising and compression[J]. IEEE Transactions on Image Processing, 2000, 9(9):1532-1546
- [35] Portilla J, Strela V, Wainwright M J, et al. Image denoising using scale mixtures of gaussians in the wavelet domain[J]. IEEE Transactions on Image Processing, 2003, 12(11):1338-1351
- [36] Cunha A L D, Zhou J, Do M N. The nonsubsampled contourlet transform: theory, design, and applications[J]. IEEE Transactions on Image Processing, 2006, 15(10):3089-3101
- [37] Foi A, Katkovnik V, Egiazarian K. Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images[J]. IEEE Transactions on Image Processing, 2007, 16(5):1395-1411
- [38] Buades A, Coll B, Morel J M. A non-local algorithm for image denoising[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2005. 60-65
- [39] Awate S P, Whitaker R T. Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering[C]. In: IEEE Conference on Computer Vision and Pattern Recognition, 2005:44-51
- [40] Dabov K, Foi A, Katkovnik V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering[J]. IEEE Transactions on Image Processing, 2007, 16(8):2080-2095
- [41] Gu S, Xie Q, Meng D, et al. Weighted nuclear norm minimization and its applications to low level vision[J]. International Journal of Computer Vision, 2017, 121(2):183-208
- [42] Peng Y, Suo J, Dai Q, et al. Reweighted low-rank matrix recovery and its application in image restoration[J]. IEEE Transactions on Cybernetics, 2014, 44(12):2418-2430
- [43] Cai J F, Chan R H, Nikolova M. Two-phase methods for deblurring images corrupted by impulse plus gaussian noise[J]. Inverse Problems & Imaging, 2012, 2(2):187-204

- [44] Garnett R, Huegerich T, Chui C, et al. A universal noise removal algorithm with an impulse detector[J]. IEEE Transactions on Image Processing, 2005, 14(11):1747-1754
- [45] Lin C H, Tsai J S, Chiu C T. Switching bilateral filter with a texture/noise detector for universal noise removal[J]. IEEE Transactions on Image Processing, 2010, 19(9):2307-2320
- [46] Xiao Y, Zeng T, Yu J, et al. Restoration of images corrupted by mixed Gaussian-impulse noise via  $l_1$ - $l_0$  minimization[J]. Pattern Recognition, 2011, 44(8):1708-1720
- [47] Xiong B, Yin Z. A universal denoising framework with a new impulse detector and nonlocal means[J]. IEEE Transactions on Image Processing, 2012, 21(4):1663-1675
- [48] Candès Emmanuel J, Li Xiaodong, Ma Yi, et al. Robust principal component analysis ?[J] Journal of the ACM, 2011, 58(3):1-37
- [49] Tseng P. Convergence of a block coordinate descent method for nondifferentiable minimization[J]. Journal of Optimization Theory and Applications, 2001, 109(3):475-494
- [50] S. Lancel and DenoiseLab, <http://www.stanford.edu/slansel/DenoiseLab/documentation.htm>, 2006
- [51] Xie Y, Gu S, Liu Y, et al. Weighted Schatten  $p$ -norm minimization for image denoising and background subtraction [J]. IEEE Transactions on Image Processing, 2016, 25(10):4842-4857
- [52] 王松桂, 贾忠贞. 矩阵论中的不等式 [M]. 安徽: 安徽教育出版社, 1994. 304-305



## 附录 A

**引理 1** 如果  $y \neq 0$ ,  $p \geq 1$ ,  $g(\cdot)$  是下半连续函数, 且对于任意的  $x \in \mathbb{R}$  都满足  $g(|x|) \geq 0$ , 则  $\alpha = \beta$ 。

**证明** 因为  $g(|x|) \geq 0$  ( $\forall x \in \mathbb{R}$ ) 且  $g(\cdot)$  下半连续, 所以  $\arg \min_x f(x) = \frac{1}{2} \|y\|_p - |x|^p + \lambda g(|x|)$  存在有一个有限解, 记为  $x_0$ 。故

$$\begin{aligned} \beta = f(x_0) &\leq f(\|x\|_p) = \frac{1}{2} \|y\|_p - \|x\|_p^p + \lambda g(\|x\|_p) \\ &\leq \frac{1}{2} \|y - x\|_p^p + \lambda g(\|x\|_p) = F(x), \end{aligned}$$

其中第三个不等式由 Minkowski 不等式得到 [52]。因此不等式  $\alpha \geq \beta$  成立。

另一方面

$$\begin{aligned} \alpha \leq F\left(x_0 \frac{y}{\|y\|_p}\right) &= \frac{1}{2} \left\| y - x_0 \frac{y}{\|y\|_p} \right\|_p^p + \lambda g\left(\left\| x_0 \frac{y}{\|y\|_p} \right\|_p\right) \\ &= \frac{1}{2} \|y\|_p - x_0^p + \lambda g(|x_0|) = \beta. \end{aligned}$$

因此, 我们有  $\alpha = \beta$ 。证明完成。  $\square$

**引理 2** 对于  $y > 0$ , 定义

$$\min_x \frac{1}{2} |y - x|^p + \lambda g(|x|) \quad (\text{A-1})$$

的解为  $x^*$ , 则  $x^* \geq 0$ 。

**证明** 设  $x^* \in \arg \min_x \frac{1}{2} |y - x|^p + \lambda g(|x|)$ , 则

$$\frac{1}{2} |y - x^*|^p + \lambda g(|x^*|) \leq \frac{1}{2} |y - (-x^*)|^p + \lambda g(|-x^*|),$$

这意味着  $(y - x^*)^2 \leq (y + x^*)^2$  成立。因此  $x^* y \geq 0$ 。故由  $y > 0$  可以得到  $x^* \geq 0$ 。  $\square$

**引理 4** 给定满足假设 **A1-A3** 的  $g$ 。当  $\mathcal{S} \neq \emptyset$  时, 则

- (i)  $0 \leq a_0 < y$ ;
- (ii)  $\varphi(x)$  严格递增, 而  $\varphi'(x)$  在区间  $(0, +\infty)$  严格递减;
- (iii) 对于  $\forall x \in (a_0, \infty)$  有  $-1 < \lambda g''(x) < 0$  成立;

(iv) 方程  $f'_y(x) = 0$  在区间  $(a_0, y]$  内最多有一个根。

**证明** (i) 由  $a_0$  的定义可以得到  $a_0 \geq 0$ 。因为  $f'_y(y) = \lambda g'(y) > 0$ ，故由引理1(ii) 可得  $\varphi(y) - \bar{x}_y < y - \bar{x}_y$ 。结合  $\bar{x}_y$  的定义可以得到对于任意的  $\forall y > 0$  有  $\bar{x}_y < y$ 。故

$$1 > \frac{\varphi(y) - \bar{x}_y}{y - \bar{x}_y} = \frac{\varphi(y) - \varphi(\bar{x}_y)}{y - \bar{x}_y} = \varphi'(\xi_0), \quad (\text{A-2})$$

其中  $\xi_0 \in (\bar{x}_y, y)$ 。这里第二个等式成立是由于  $\varphi(\bar{x}_y) = \bar{x}_y$ 。当方程  $\varphi'(x) = 1$  有非负根时， $\varphi'(a_0) = 1$ 。结合 (A-2)，有  $\varphi'(\xi_0) < \varphi'(a_0)$ 。故由  $\varphi'(x)$  严格递减得到  $y > \xi_0 > a_0$ 。而当  $\varphi'(x) = 1$  无非负根时， $a_0 = 0$ 。故自然得到  $y > a_0$  成立。

(ii) 由于  $g(x)$  在  $(0, +\infty)$  上严格凹，所以  $g'(x)$  是严格递减， $\varphi(x) = y - \lambda g'(x)$  严格递增。因为  $g'(x)$  在  $(0, \infty)$  上严格凸， $g''(x)$  是严格增。因此  $\varphi'(x) = -\lambda g''(x)$  是严格递减。

(iii) 如 (i) 中的证明所示， $g'(x)$  是严格递减，因此  $g''(x) < 0$ 。因为  $g''(x)$  是严格增，且对任意的  $x \in (a_0, +\infty)$  都有  $\lambda g''(x) > \lambda g''(a_0) \geq -1$ 。故  $-1 < \lambda g''(x) < 0$  成立。

(iv) 假设  $f'_y(x) = 0$  在区间  $(a_0, y]$  上至少有两个根，则由罗尔定理， $\lambda g''(x) + 1 = 0$  在  $(a_0, y)$  上至少有一个根。这和  $a_0$  的定义矛盾。  $\square$

**引理 5** 给定满足假设 A1-A3 的  $g$ 。当  $S \neq \emptyset$  时，则下面结论成立：

(i)  $\varphi'(\bar{x}_y) \leq 1$  成立；

(ii) 若  $f'_y(a_0) < 0$ ，则  $a_0 < y$ ，并且方程  $f'_y(x) = 0$  在区间  $(a_0, y]$  内有唯一解；

(iii) 若  $f'_y(a_0) \geq 0$ ，若  $\bar{x}_y \notin (0, a_0) \cup (a_0, y]$  (若  $a_0 \geq y$ ，定义  $(a_0, y] = \emptyset$ ；当  $a_0 = y$  时，定义  $(0, a_0] = \emptyset$ )。

(iv)  $f_y(x)$  在区间  $(0, +\infty)$  内至多有一个非零极小值点。

**证明** (i) 对于任意的  $k$ ，由均值定理可以得到存在  $\xi_k \in (\bar{x}_y, x_L^{(k)})$  使得  $\varphi'(\xi_k)(x_L^{(k)} - \bar{x}_y) = \varphi(x_L^{(k)}) - \varphi(\bar{x}_y) < x_L^{(k)} - \bar{x}_y$  (其中第二个不等式由引理 1 (ii) 和 (iii) 得到)。因此  $\varphi'(\xi_k) < 1$ 。因为由1(ii) 可得序列  $\{x_L^{(k)}\}$  收敛到  $\bar{x}_y$ ，故  $\{\xi_k\}$  收敛到  $\bar{x}_y$ 。由  $\varphi'(x)$  的连续性可以得到  $\varphi'(\bar{x}_y) = \lim_{k \rightarrow \infty} \varphi'(\xi_k) \leq 1$ 。

(ii) 由  $f'_y(a_0) = (a_0 - y) + \lambda g'(a_0) < 0$ ， $g'(a_0) > 0$  可以得到  $a_0 < y$ ，故  $(a_0, y) \neq \emptyset$ 。因为  $g(x)$  严格增，所以  $f'_y(y) = \lambda g'(y) + (y - y) > 0$ 。因为  $f'_y(a_0) < 0$ ，故由零点定理可以得到  $f'_y(x) = 0$  在  $(a_0, y)$  上至少有一个根。结合引理 2 (iv)，结论成立。

(iii) 首先证明  $\bar{x}_y \notin (a_0, y]$ 。当  $a_0 \geq y$  时，则  $(a_0, y] = \emptyset$ 。故  $\bar{x}_y \notin (a_0, y]$  成立。当  $a_0 < y$  时，由引理2 (ii) 可以得到  $f''_y(x) = 1 + \lambda g''(x)$  是个严格增函数。故对于

$\forall x > a_0$  都有  $0 \leq f_y''(a_0) < f_y''(x)$  成立。这意味着  $f_y'(x)$  在  $(a_0, +\infty)$  上是一个严格增函数。因此, 若  $f_y'(a_0) \geq 0$ ,  $f_y'(x) = 0$  在  $(a_0, y]$  上无解。由  $\bar{x}_y$  的定义可以得到  $\bar{x}_y \notin (a_0, y]$ 。

下面我们证明  $\bar{x}_y \notin (0, a_0)$ 。显然, 当  $a_0 = 0$  时,  $f_y'(x) = 0$  在  $(0, a_0)$  上无解。当  $a_0 \neq 0$  时, 因为  $g''(x)$  是严格增函数, 所以对于  $\forall x \in (0, a_0)$ , 有  $f_y''(x) = \lambda g''(x) + 1 < f_y''(a_0) = 0$  成立。所以  $\forall x \in (0, a_0)$ , 有  $\varphi'(x) = -\lambda g''(x) > 1$  成立。若  $\bar{x}_y \in (0, a_0)$ ,  $\varphi'(\bar{x}_y) > 1$ 。这与引理3 (i) 里的结论矛盾。

(iv) 当  $a_0 = 0$  时, 因为  $0 < f_y''(x) = 1 - \varphi'(x) < 1 \ \forall x \in (0, +\infty)$  (即  $f_y(x)$  在区间  $(0, +\infty)$  上严格凸), 由引理 3 (ii) 可以得到  $f_y(x)$  在区间  $(0, \infty)$  上存在唯一极小值。当  $a_0 > 0$  时, 有  $f_y''(a_0) = 1 - \varphi'(a_0) = 0$  成立。考虑到  $f_y''(x) = 1 + \lambda g''(x)$  是严格增, 因此  $f_y''(x) < 0 \ \forall x \in (0, a_0)$  且  $f_y''(x) > 0 \ \forall x \in (a_0, +\infty)$ 。由此推出, 当  $f_y(x)$  在区间  $(0, +\infty)$  上存在极小值时, 该极小值大于  $a_0$ 。因为对于任意  $x \in [y, \infty)$  有  $f_y'(x) = (x - y) + \lambda g'(x) > 0$  成立, 故由引理 3 (iii) 和 (iv) 得到方程  $f_y'(x) = 0$  在区间  $[a_0, +\infty)$  上至多一个根。因此, 结论成立。  $\square$

**引理 6** 给定满足假设 **A1-A3** 的  $g$ 。当  $S \neq \emptyset$  且  $f_y'(a_0) < 0$ , 则

i) 对于任意的  $x \in (\bar{x}_y, y]$ , 都有  $\psi(x) < x$  成立;

ii)  $\bar{x}_y$  是  $\psi(x)$  的不动点, 即,  $\psi(\bar{x}_y) = \bar{x}_y$ 。

**证明** (i) 由中值定理和引理 1 (ii) 可以得到, 对于任意的  $x \in (\bar{x}_y, y]$ , 存在  $\xi_0 \in (\varphi(x), x) \subseteq (\bar{x}_y, y] \subseteq (a_0, +\infty)$  使  $\varphi(\varphi(x)) - \varphi(x) = \varphi'(\xi_0)(\varphi(x) - x)$  成立。故  $\psi(x) = \varphi(x) - \frac{\varphi(\varphi(x)) - \varphi(x)}{\varphi'(\xi_0) - 1}$ 。因为  $\varphi'(\xi_0) < 1$  (由引理 2 (ii) 可得),  $\varphi(\varphi(x)) < \varphi(x)$  (由引理 1 (ii) 可得), 我们可以得到  $\frac{\varphi(\varphi(x)) - \varphi(x)}{\varphi'(\xi_0) - 1} > 0$ 。因此, 对于任意的  $x \in (\bar{x}_y, y]$ , 有  $\psi(x) = \varphi(x) - \frac{\varphi(\varphi(x)) - \varphi(x)}{\varphi'(\xi_0) - 1} \leq \varphi(x) < x$  成立。

(ii) 由引理 3, 我们可以得到当  $f_y'(y) \neq 0$ ,  $f_y'(a_0) < 0$  时,  $f_y'(x) = 0$  在  $(a_0, y]$  上有唯一解  $\bar{x}_y$ 。因为由引理2 (iii), 有  $0 < \varphi'(\bar{x}_y) < 1$  成立。故我们可以得到

$$\begin{aligned} \lim_{x \rightarrow \bar{x}_y^+} \psi(x) &= \varphi(\bar{x}_y) - \lim_{x \rightarrow \bar{x}_y^+} \frac{(\varphi(\varphi(x)) - \varphi(x))(\varphi(x) - x)}{\varphi(\varphi(x)) - 2\varphi(x) + x} \\ &= \varphi(\bar{x}_y) - \lim_{x \rightarrow \bar{x}_y^+} \frac{(\varphi(\varphi(x)) - \varphi(x))(\varphi(x) - x)}{(\varphi'(\xi_0) - 1)(\varphi(x) - x)} \\ &= \varphi(\bar{x}_y) - \lim_{x \rightarrow \bar{x}_y^+} \frac{\varphi(\varphi(x)) - \varphi(x)}{\varphi'(\xi_0) - 1} \\ &= \varphi(\bar{x}_y) - \frac{0}{\varphi'(\bar{x}_y) - 1} = \bar{x}_y. \end{aligned} \quad (\text{A-3})$$

其中第二等式由均值定理和  $\xi_0 \in (\varphi(x), x)$  得到, 由  $\bar{x}_y$  是  $\varphi(x)$  的不动点得到倒数第二个等式。  $\square$

**引理 7** 给定满足假设 **A1-A3** 的  $g$ 。当  $S \neq \emptyset$  且  $f'_y(a_0) < 0$ ，则  $\psi(x)$  是  $(\bar{x}_y, y]$  区间上的一个严格增函数。

**证明** 由均值定理可知，存在  $\xi_0 \in (\varphi(x), x)$  使得

$$\varphi'(\xi_0)(\varphi(x) - x) = \varphi(\varphi(x)) - \varphi(x), \quad (\text{A-4})$$

其中  $x \in (\bar{x}_y, y]$ ， $\varphi'(\xi_0) < 1$ 。所以， $\varphi(\varphi(x)) - 2\varphi(x) + x = (\varphi'(\xi_0) - 1)(\varphi(x) - x) \neq 0$ 。对  $\psi(x)$  求导得到

$$\psi'(x) = \varphi'(x) - \frac{p(x)}{(\varphi(\varphi(x)) - 2\varphi(x) + x)^2}, \quad (\text{A-5})$$

其中  $p(x)$  为

$$\begin{aligned} & [\varphi'(x)(\varphi'(\varphi(x)) - 1)(\varphi(x) - x) + (\varphi(\varphi(x)) - \varphi(x)) \\ & (\varphi'(x) - 1)][\varphi(\varphi(x)) - 2\varphi(x) + x] - (\varphi(\varphi(x)) - \varphi(x)) \\ & (\varphi(x) - x)[\varphi'(x)\varphi'(\varphi(x)) - 2\varphi'(x) + 1]. \end{aligned} \quad (\text{A-6})$$

由均值定理可知，存在  $\xi_0 \in (\varphi(x), x)$  使得

$$\varphi'(\xi_0)(\varphi(x) - x) = \varphi(\varphi(x)) - \varphi(x). \quad (\text{A-7})$$

将 (A-7) 代入 (A-6)，简化得到

$$\begin{aligned} p(x) = & (\varphi(x) - x)^2 [\varphi'(\xi_0)^2 \varphi'(x) - \varphi'(\xi_0)^2 \\ & - \varphi'(x)\varphi'(\varphi(x)) + \varphi'(x)]. \end{aligned} \quad (\text{A-8})$$

因为  $\varphi'(x)$  严格递减 (由引理 2 (ii) 可知) 且  $\varphi(x) < \xi_0 < x$ ，能够得到  $\varphi'(\varphi(x)) > \varphi'(\xi_0) > \varphi'(x)$ 。因此，

$$\begin{aligned} & \varphi'(\xi_0)^2 \varphi'(x) - \varphi'(\xi_0)^2 - \varphi'(x)\varphi'(\varphi(x)) + \varphi'(x) \\ & < \varphi'(\xi_0)^2 \varphi'(x) - \varphi'(\xi_0)\varphi'(x) - \varphi'(x)\varphi'(\xi_0) + \varphi'(x) \\ & = \varphi'(x)(\varphi'(\xi_0) - 1)^2. \end{aligned} \quad (\text{A-9})$$

合并 (A-8) 与 (A-9)，得到

$$p(x) < (\varphi(x) - x)^2 \varphi'(x)(\varphi'(\xi_0) - 1)^2. \quad (\text{A-10})$$

将 (A-10) 代入 (A-5) 可以得到

$$\begin{aligned} \psi'(x) &= \varphi'(x) - \frac{p(x)}{(\varphi'(\xi_0) - 1)^2 (\varphi(x) - x)^2} \\ &> \varphi'(x) - \varphi'(x) = 0, \end{aligned}$$

因此， $\psi(x)$  在  $(\bar{x}_y, y]$  上严格递减。 □

## 结论 1

$$\arg \min_x \frac{1}{2}(y-x)^2 - \frac{\lambda\epsilon}{\epsilon+|x|} = \text{sign}(y) \arg \min_{x \geq 0} \left( \frac{1}{2}(|y|-x)^2 - \frac{\lambda\epsilon}{\epsilon+x} \right). \quad (\text{A-11})$$

**证明** 方便起见, 记  $f(x) := \frac{1}{2}(y-x)^2 - \frac{\lambda\epsilon}{\epsilon+|x|}$ 。若  $y = 0$ , 则  $\arg \min_x f(x) = \arg \min_x \frac{1}{2}x^2 - \frac{\lambda\epsilon}{\epsilon+|x|} = 0$ 。因此, 当  $y = 0$  时, 定理成立。

接下来, 假设  $y \neq 0$  并设  $x^*$  为  $f(x)$  的最优解。因此有

$$\begin{aligned} \frac{1}{2}(y-x^*)^2 - \frac{\lambda\epsilon}{\epsilon+|x^*|} &\leq \frac{1}{2}(y-(-x^*))^2 - \frac{\lambda\epsilon}{\epsilon+|-x^*|} \\ &= \frac{1}{2}(y^2 + (-x^*)^2 - 2(-x^*)y) - \frac{\lambda\epsilon}{\epsilon+|-x^*|}. \end{aligned}$$

因此  $-x^*y \leq x^*y$ , 这使得  $x^*y \geq 0$  成立, 即

$$\begin{cases} x^* \geq 0, & \text{if } y > 0; \\ x^* \leq 0, & \text{if } y < 0. \end{cases} \quad (\text{A-12})$$

因此

$$\begin{aligned} &\arg \min_x \frac{1}{2}(x-y)^2 - \frac{\lambda\epsilon}{\epsilon+|x|} \\ &= \arg \min_{\text{sign}(y)x \geq 0} \frac{1}{2}(x-y)^2 - \frac{\lambda\epsilon}{\epsilon+\text{sign}(y)x} \end{aligned}$$

当  $y > 0$  时,

$$\arg \min_{\text{sign}(y)x \geq 0} \frac{1}{2}(x-y)^2 - \frac{\lambda\epsilon}{\epsilon+\text{sign}(y)x} = \arg \min_{x \geq 0} \frac{1}{2}(x-|y|)^2 - \frac{\lambda\epsilon}{\epsilon+x}$$

并且当  $y < 0$  时,

$$\begin{aligned} \arg \min_{\text{sign}(y)x \geq 0} \frac{1}{2}(x-y)^2 - \frac{\lambda\epsilon}{\epsilon+\text{sign}(y)x} &= \arg \min_{x \leq 0} \frac{1}{2}(x-y)^2 - \frac{\lambda\epsilon}{\epsilon-x} \\ &= -\arg \min_{t \geq 0} \frac{1}{2}(-t+y)^2 - \frac{\lambda\epsilon}{\epsilon+t} \\ &= -\arg \min_{x \geq 0} \frac{1}{2}(x-|y|)^2 - \frac{\lambda\epsilon}{\epsilon+x} \end{aligned}$$

因此, 当  $y \neq 0$  时, 结论成立。故结论对任意  $y \in \mathbb{R}$  都成立。  $\square$

**结论 2** 令  $\text{Prox}_{H_\epsilon}(y) = \arg \min_x \frac{1}{2}(y-x)^2 - \frac{\lambda\epsilon}{\epsilon+|x|}$ 。利用  $h_0$ ,  $\Delta$  和  $x_1$  的定义 (见表格5-1), 我们下面结论

(1) 当  $\Delta > 0$  时,

$$\text{Prox}_{H_\epsilon}(y) = \begin{cases} \frac{x_1 + |x_1|}{2}, & \text{if } y \geq \sqrt{2(d(\frac{x_1 + |x_1|}{2}) + \lambda)}; \\ 0, & \text{if } |y| < \sqrt{2(d(\frac{x_1 + |x_1|}{2}) + \lambda)}; \\ -\frac{x_1 + |x_1|}{2}, & \text{if } y \leq -\sqrt{2(d(\frac{x_1 + |x_1|}{2}) + \lambda)}. \end{cases} \quad (\text{A-13})$$

(2) 当  $\Delta \leq 0$  时,

$$\text{Prox}_{H_\epsilon}(y) = \begin{cases} \frac{h_0 + |h_0|}{2}, & \text{if } y \geq \sqrt{2(d(\frac{h_0 + |h_0|}{2}) + \lambda)}; \\ 0, & \text{if } |y| < \sqrt{2(d(\frac{h_0 + |h_0|}{2}) + \lambda)}; \\ -\frac{h_0 + |h_0|}{2}, & \text{if } y \leq -\sqrt{2(d(\frac{h_0 + |h_0|}{2}) + \lambda)}. \end{cases} \quad (\text{A-14})$$

**证明** 方便起见, 令  $d(x) := \frac{1}{2}(|y| - x)^2 - \frac{\lambda\epsilon}{\epsilon+x}$ 。由 [15] 可知,  $\arg \min_{x \geq 0} d(x) = \arg \min_{x \in \{0, \hat{x}_y\}} d(x)$ , 其中  $\hat{x}_y = \max\{x | d'(x) = 0, 0 \leq x \leq |y|\} \cup \{0\}$ 。因为

$$\begin{aligned} d'(x) &= (x - |y|) + \frac{\lambda\epsilon}{(\epsilon + x)^2} = 0 \\ \Rightarrow (x - |y|)(\epsilon + x)^2 + \lambda\epsilon &= 0, \end{aligned}$$

由上面的等式可以得到,  $d'(x) = 0$  在区间  $(|y|, +\infty)$  上无解。因此  $\hat{x}_y = \max\{x | d'(x) = 0\} \cup \{0\}$ 。

故可以利用卡丹 (Cardano) 公式得到方程  $(x - |y|)(\epsilon + x)^2 + \lambda\epsilon = 0$  的解。

$$\left\{ \begin{aligned} x_1 &= \sqrt[3]{-\frac{q}{2} + \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}} \\ &\quad - \frac{2\epsilon - |y|}{3}; \\ x_2 &= \frac{-1 + \sqrt{3}i}{2} \sqrt[3]{-\frac{q}{2} + \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}} + (\frac{-1 + \sqrt{3}i}{2})^2 \\ &\quad \sqrt[3]{-\frac{q}{2} - \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}} - \frac{2\epsilon - |y|}{3}; \\ x_3 &= (\frac{-1 + \sqrt{3}i}{2})^2 \sqrt[3]{-\frac{q}{2} + \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}} + \frac{-1 + \sqrt{3}i}{2} \\ &\quad \sqrt[3]{-\frac{q}{2} - \sqrt{(\frac{q}{2})^2 + (\frac{p}{3})^3}} - \frac{2\epsilon - |y|}{3}. \end{aligned} \right. \quad (\text{A-15})$$

其中  $p = -\frac{(\epsilon + |y|)^2}{3}$ ,  $q = \lambda\epsilon - \frac{2}{27}(\epsilon + |y|)^3$ 。

由卡丹公式可知, 若  $\Delta = (\frac{q}{2})^2 + (\frac{p}{3})^3 > 0$ , 则等式  $(x - |y|)(\epsilon + x)^2 + \lambda\epsilon = 0$  有唯一实数解  $x_1$ 。因此  $\hat{x}_y = \frac{x_1 + |x_1|}{2}$ 。通过比较  $d(0)$  和  $d(\hat{x}_y)$  可以得到

$$\arg \min_{x \geq 0} d(x) = \begin{cases} \frac{x_1 + |x_1|}{2}, & \text{若 } |y| \geq \sqrt{2(d(\frac{x_1 + |x_1|}{2}) + \lambda)}; \\ 0, & \text{否则.} \end{cases} \quad (\text{A-16})$$

若  $\Delta = (\frac{q}{2})^2 + (\frac{p}{3})^3 \leq 0$ , 则等式  $(x - |y|)(\epsilon + x)^2 + \lambda\epsilon = 0$  有三个解。因此  $\hat{x}_y = \frac{h_0 + |h_0|}{2}$ , 其中

$$h_0 = \frac{1}{2}(|\frac{|x_1 - x_2| + x_1 + x_2}{2} - x_3| + \frac{|x_1 - x_2| + x_1 + x_2}{2} + x_3).$$

通过比较  $d(0)$  和  $d(\hat{x}_y)$  的大小, 得到

$$\arg \min_{x \geq 0} d(x) = \begin{cases} \frac{h_0 + |h_0|}{2}, & \text{if } |y| \geq \sqrt{2(d(\frac{h_0 + |h_0|}{2}) + \lambda)}; \\ 0, & \text{否则.} \end{cases} \quad (\text{A-17})$$

结合结论 1, 得到结论 2 成立。  $\square$

**结论 3**  $\forall |y|, \forall \epsilon_0 > 0, \exists \delta_{\epsilon_0} > 0$  使得, 当  $|\epsilon| < \delta_{\epsilon_0}$ ,  $|\text{Prox}_{H_\epsilon}(y) - \text{Prox}_H(y)| < \epsilon_0$ 。

**证明** 假设  $\Delta > 0$ 。因为当  $y = 0$  时, 结论成立。因此下面只考虑  $y \neq 0$  的情况。

因为  $\frac{x_1 + |x_1|}{2} = \max(0, x_1)$ ,  $\lim_{x \rightarrow +\infty} d'(x) = +\infty$ , 故当  $x \geq \frac{x_1 + |x_1|}{2}$  时,  $d'(x) \geq 0$ 。因为  $\frac{x_1 + |x_1|}{2} \leq |y|$ ,  $d(\frac{x_1 + |x_1|}{2}) \leq d(|y|) \leq 0$ 。所以当  $y \geq \sqrt{2\lambda}$  时,  $y \geq \sqrt{2\lambda} \geq \sqrt{2(d(\frac{x_1 + |x_1|}{2}) + \lambda)}$ 。当  $y \leq -\sqrt{2\lambda}$  时,  $y \leq -\sqrt{2\lambda} \leq -\sqrt{2(d(\frac{x_1 + |x_1|}{2}) + \lambda)}$ 。

由  $x_1$  的定义可以得到的得到

$$\lim_{\epsilon \rightarrow 0} \frac{x_1 + |x_1|}{2} = |y|, \lim_{\epsilon \rightarrow 0} \sqrt{2(d(\frac{x_1 + |x_1|}{2}) + \lambda)} = \sqrt{2\lambda}, \forall y \neq 0.$$

所以  $\forall y \neq 0, \forall \epsilon_0, \exists \delta_{\epsilon_0}$ , 若  $|\epsilon| \leq \delta_{\epsilon_0}$ ,  $|\text{Prox}_{H_\epsilon}(y) - \text{Prox}_H(y)| < \epsilon_0$ 。其中

$$|\text{Prox}_{H_\epsilon}(y) - \text{Prox}_H(y)| = \begin{cases} |\frac{x_1 + |x_1|}{2} - |y||, & \text{若 } y \geq \sqrt{2\lambda}; \\ |0 - 0|, & \text{若 } |y| < \sqrt{2\lambda}; \\ |-\frac{x_1 + |x_1|}{2} - |y||, & \text{若 } y \leq -\sqrt{2\lambda}. \end{cases} \quad (\text{A-18})$$

对于  $\Delta \leq 0$  的情况, 我们可以得到类似的结论。  $\square$





## 致谢

时光如梭，从 2017 年考取温州大学硕士研究生至今即将三年。三年时间虽然短暂，却学习到很多。在此，致谢在学习以及生活中所有帮助过我的人。

感谢我的导师 -张笑钦老师把我领进学术研究的大门，并在实验能力，学术报告，论文写作等各方面对我展开训练。在学习，竞赛和研究工作上，张老师都给我很多帮助和启迪！

感谢温大的奖学金及助学金政策。虽然不能用丰厚一词形容温大的奖助学金，但是至少使我能够能比较专注于自己的学习和研究工作。

感谢实验室老师以及其他所有教过我的老师。特别感谢王迪老师在学习上给予我的帮助。

感谢我的同学。特别是刘倩倩师姐，唐贵英同学，刘永利同学，蒲青秀同学，韩玲玲，罗碧煊同学，姜贺云同学，蒋润华同学，王涛同学以及康佳源同学。特别感谢他们在日常生活和学习上对我的帮助。

最后，感谢在百忙之中抽出时间来参加我的论文评阅与答辩的各位老师，感谢老师们对本论文的建议和指导。

面对毕业，心中慷慨万分。经过三年的学习与成长，尽管在专业知识学习以及科研能力等各个方面都取得了不同程度的进步，但仍有许多不足需要改进。希望自己能够以更大的热情投入到以后无尽的学习和工作当中去。

郑晶晶

2020 年 5 月



## 攻读硕士期间的科研成果

(1) Xiaoqin Zhang, Jingjing Zheng, Di Wang, Guiying Tang, Zhengyuan Zhou, and Zhouchen Lin. Structured Sparsity Optimization with Non-Convex Surrogates: A Unified Algorithmic Framework[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020. (Submitted)

(2) Zhang Xiaoqin, Zheng Jingjing, Wang Di and Zhao Li. Exemplar-based Denoising: A Unified Low-rank Recovery Framework[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019.

(3) Zhang Xiaoqin, Zheng Jingjing, Yan Yufang, Zhao Li and Jiang Runhua. Joint Weighted Tensor Schatten p-Norm and Tensor lp-Norm Minimization for Image Denoising[J]. IEEE Access, 2019, 7: 20273-20280.

(4) Yan Yufang , Zhang Xiaoqin, Zheng Jingjing and Zhao Li. Weighted Tensor Schatten p-norm Minimization for Image Denoising[C]. Intelligent Systems Conference, 2019:163-172.

(5) Xiaoju Lu, Guiying Tang, Di Wang, Xiaoqin Zhang and Jingjing Zheng (Corresponding Author). Structural Dictionary Learning based on Non-convex Surrogate of  $l_{2,1}$  Norm for Classification[C]. The 2nd International Workshop on Big Media Dataset Construction, Management and Applications, 2019.

(6) 张笑钦, 郑晶晶, 严玉芳, 一种基于新范数的图像噪声去除方法, 专利号: 201810233460.7, 申请日期: 2018.03.21.

(7) 2018 年浙江省大学生科技创新活动计划暨新苗人才计划: 一种新范数及其广义形式在图像分类上的应用研究 (项目代码: 2018R429025), 郑晶晶 (负责人), 路孝举, 唐贵英, 2018-2020。

