

基于监控视频背景建模的一类非凸矩阵优化算法研究

贾凯贤, 陈小英, 孟书羽, 程越, 张弦, 彭定涛*

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2023年5月22日; 录用日期: 2023年7月29日; 发布日期: 2023年8月3日

摘要

为了实现稳健的低秩 - 稀疏矩阵分解, 本文考虑非凸非光滑优化模型, 其中矩阵的秩函数采用奇异值的 Capped- l_1 松弛, 矩阵的 l_0 范数采用矩阵的 l_1 范数松弛。首先, 给出了适用于矩阵问题的 Capped- l_1 阈值算子。其次, 提出了交替方向乘子法求解我们的非凸非光滑矩阵优化模型, 并分析了算法的收敛性。最后, 通过大量数值实验表明: 在有噪声和无噪声的情况下, 所提出的算法都能有效、稳健地分解出低秩 - 稀疏矩阵。并将提出的算法应用于监控视频的前景和背景分离问题, 发现所提出的算法对于该问题有良好的性能, 这说明该算法能够解决相关实际问题。

关键词

稀疏 - 低秩矩阵分解, Capped- l_1 正则, 交替方向乘子法, 前景和背景分离

Research on a Class of Nonconvex Matrix Optimization Algorithm Based on Surveillance Video Background Modeling

Kaixian Jia, Xiaoying Chen, Shuyu Meng, Yue Cheng, Xian Zhang, Dingtao Peng*

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: May 22nd, 2023; accepted: Jul. 29th, 2023; published: Aug. 3rd, 2023

Abstract

In order to achieve robust low-rank sparse matrix decomposition, this paper considers a nonconvex and nonsmooth optimization model, in which the ranking function of the matrix is relaxed by

*通讯作者。

文章引用: 贾凯贤, 陈小英, 孟书羽, 程越, 张弦, 彭定涛. 基于监控视频背景建模的一类非凸矩阵优化算法研究[J]. 运筹与模糊学, 2023, 13(4): 2817-2830. DOI: 10.12677/orf.2023.134282

Capped- l_1 regularization of the singular values, and l_0 norm of the matrix is relaxed with the l_1 norm of the matrix. First, we provide the closed-form thresholding operator for the matrix-Capped- l_1 function. Secondly, the alternating direction method of multipliers is proposed to solve our nonconvex and nonsmooth optimization model, and the convergence analysis is provided. Finally, a large number of numerical experiments show that the proposed algorithm can effectively and robustly decompose the low-rank and sparse matrices in the case of noise and noise. The proposed algorithm is applied to the background separation problem of surveillance video, and it is indicated that the proposed algorithm has good performance for this problem, which shows that the algorithm can solve the relevant practical problems.

Keywords

Sparse and Low-Rank Matrices Decomposition, Capped- l_1 Regularization, Alternating Direction Method of Multipliers, Foreground and Background Separation

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在大数据时代,随着计算机和信息技术的普及与应用,各种各样的数据信息呈爆炸性增长,时刻都会产生大量的、多样化的、结构复杂的、冗余的、高维的海量数据。研究表明无法通过常规手段直接处理使用这类信息,那么如何从这些高维且冗余的数据中提取有价值的信息便成为大数据科学问题的关键课题之一。大量已有的研究工作表明稀疏性是研究这一课题的重要手段。

在不同实际应用领域中出现的复杂系统和复杂模型的矩阵,其通常可以具有由低秩分量和稀疏分量组成的特征,而这种复杂系统的可分性为理解复杂系统的行为和性质提供了可能。实现这种矩阵表示的可分性可以描述为如下的矩阵低秩稀疏分解(Low Rank and Sparse Decomposition, LRSD) [1] [2] [3]:

$$\min_{L,S \in \mathbb{R}^{m \times n}} \text{rank}(L) + \lambda \|S\|_{l_0}, \quad \text{s.t. } D = L + S, \quad (1)$$

其中参数 $\lambda > 0$, $\text{rank}(\cdot)$ 是矩阵的秩函数, $\|\cdot\|_{l_0}$ 是矩阵的 l_0 范数,表示矩阵中所有元素的非零个数和,其本质上是寻找数据最具价值的低维结构及最佳投影,通过去除数据中的冗余信息,从而找到数据中有价值的信息。

因为 LRSD 方法克服了主成分分析(Principal Component Analysis, PCA) [4] [5] 方法对非高斯样本或异常值的敏感性在实践中往往使其不够稳健的缺陷,因此该方法又被称为稳定主成分分析法(Robust Principal Component Analysis, RPCA)。LRSD 问题的基本出发点在于很多的实际数据矩阵 D 往往是具有先验信息——近似低秩特性和稀疏特性的。问题(1)中的参数 λ 则用于平衡矩阵 L 的低秩程度和矩阵 S 的稀疏程度,和向量 l_0 极小化问题一样,问题(1)显然是一个 NP-hard 的问题,求解的复杂度极高,甚至在实际的应用中可能会出现难以求解的情况,不利于该模型在实际问题中的应用。

为了高效求解 LRSD 问题, Candès 和 Li [6] 等提出在一定条件下,最小化 l_1 范数求得的解等价于最小化 l_0 范数求得的解,而秩函数则可以通过核范数进行近似表达,该方法即是被称为著名的主成分追踪法(Principal Component Pursuit, PCP),它的思想是将 LRSD 的非凸问题进行凸松弛,转化为凸优化问题进行求解。

在求解模型(1)时, PCP 方法的出发点在于矩阵的秩等于非零奇异值的个数以及 l_1 范数是 l_0 范数的最紧凸松弛, 进而将非凸问题(1)进行松弛得到如下最紧凸松弛问题进行求解:

$$\min_{L,S \in \mathbb{R}^{m \times n}} \|L\|_* + \lambda \|S\|_1, \quad s.t. \quad D = L + S \quad (2)$$

其中 $\|\cdot\|_*$ 表示矩阵奇异值之和, 称为矩阵的核范数, $\|\cdot\|_1$ 表示矩阵所有元素绝对值之和。实际应用中该模型的平衡参数 λ 通常设置为 $1/\max\{m,n\}$, m 和 n 分别表示的是矩阵 D 的行数和列数。

在求解模型(1)时, 使用的 PCP 方法[6]会出现对低秩部分和稀疏部分刻画不准确的问题。因此, 很多研究者一直在寻求秩函数和稀疏度函数的更加精确的表达方法。在许多经典的 LRSD 方法中主要是利用核范数对秩函数进行凸近似。虽然核范数已经在低秩矩阵的近似表达中被广泛使用, 但是核范数表示的是非零奇异值的和, 并不是秩函数的最佳刻画。为了解决该问题, 文献[7]中提出了截断核范数(Truncated Nuclear Norm, TNN), 实验结果表明该方法在一定程度上能更准确地诱导低秩部分, 后续中许多研究基于 TNN 的方法被广泛应用到低秩稀疏分解中[8] [9] [10]。

在求解优化问题(2)时, 文献[11]基于目标函数的一阶信息提出了迭代阈值算法(Iterative Thresholding, IT), 虽然算法的迭代方式简单也收敛, 但其每次迭代都要进行一次奇异值分解且收敛速度很慢, 也难以选取合适的步长, 无法快速求解大型实际问题。此后, 文献[12]提出了求解问题(2)的加速临近梯度算法(Accelerated Proximal Gradient, APG)和求解相应对偶问题的梯度上升算法, 并通过实验说明这两种算法在处理 1000×1000 维的矩阵时相较于迭代阈值算法加速了 50 倍。进一步文献[13]提出了增广拉格朗日乘子法(Augmented Lagrange Multiplier, ALM), 其中 ALM 算法相比 IT 算法和 APG 算法具有更快的计算速度且所需存储空间更小等优点, 在显著提高计算精度的同时其计算速度快于 APG 算法。然而当目标函数含有两个及以上的变量时, 由于 ALM 将乘子函数中的所有变量同时最小化, 使该算法具有较高的计算复杂度, 给求解带来了困难。因此, 交替方向乘子法(Alternating Direction Method of Multipliers, ADMM)算法[14] [15] [16]应运而生, ADMM 算法采用分而治之的思想将不易求解的问题转化为容易求解的子问题进行求解, 使其同时兼具了乘子法和交替最小化算法的优点。此外, 该算法的复杂度较低, 有利于求解, 并且该算法具有更高的计算速度, 因此受到研究者的广泛关注。

另一方面向量 $l_{1/2}$ 正则化理论表明: 相较于常用的 l_1 范数, $l_{1/2}$ 范数可进一步诱导向量的稀疏性, 在稀疏信号重建中对噪声具有更强的稳健性。为得到矩阵 D 更为精确的稀疏低秩表示, 有学者将向量的 $l_{1/2}$ 范数推广到矩阵, 分别用 $S_{1/2}$ 范数[17] [18] (矩阵奇异值构成向量的 $l_{1/2}$ 范数)和 $l_{1/2}$ 范数来刻画矩阵的秩与稀疏度。与此同时, 近年来越来越受到重视的折叠凹正则理论[19] [20]表明, 相比于凸正则, 某些特殊的非凸正则则在诱导低秩性和稀疏性方面具有更大的优势。受 $l_{1/2}$ 正则化理论的良好表现及拓展, 本文将 Capped- l_1 正则理论应用于矩阵问题中。事实上, 我们的数值实验表明了相较于 RPCA 方法中的核范数正则, Capped- l_1 正则可进一步诱导低秩矩阵的低秩性, 即拟解决的具体优化模型如下:

$$\min_{L,S \in \mathbb{R}^{m \times n}} \Phi(L) + \lambda \|S\|_1, \quad s.t. \quad D = L + S \quad (3)$$

其中, $\Phi(L) = \sum_{i=1}^{\min\{m,n\}} \min\left\{1, \frac{\sigma_i(L)}{\gamma}\right\}$, γ 是给定的参数, $\sigma_i(L)$ 表示对低秩矩阵 L 进行奇异值分解(Singular Value Decomposition, SVD)后每一个奇异值所对应的项。

考虑到 Capped- l_1 模型的目标函数与约束条件关于低秩矩阵与稀疏矩阵是可分的, 本文基于文献[14]中的交替方向乘子法(Alternating Direction Method of Multipliers, ADMM)算法思想提出交替迭代阈值算法。该算法利用增广 Lagrange 乘子技术, 在迭代过程中采用交替投影的思想逐个更新低秩矩阵、稀疏矩阵和 Lagrange 乘子。低秩矩阵与稀疏矩阵的更新需要求解两个线性非凸优化子问题, 之后通过将作用于

向量的 Capped- l_1 算子推广到矩阵情形, 给出其子问题最优解的显式形式, 避免每次迭代都要对矩阵进行奇异值分解的弊端, 这从很大程度上保证了所设计算法的高精度与低时间代价特性。

与 ALM 算法相比, 大量的模拟实验说明本文所提出的交替阈值迭代算法具有以下优点: 达到收敛所需迭代次数与时间代价大幅降低、分解出的低秩矩阵的秩更接近于真实值并且同时算法的可靠性对矩阵 L 的低秩程度依赖更少。另外, 在监控视频背景建模这一实际应用中, 交替阈值迭代算法能得到更为稀疏的背景矩阵, 具有更为稳健的恢复性能且时间代价相较于 ALM 算法降幅高达一个数量级, 这对实际中海量视频数据的快速处理具有重要意义。

2. 交替方向乘子法

在本节中, 我们提出交替方向乘子法来解决问题(3), 并给出该算法的收敛性分析。

2.1. 交替方向乘子法的迭代框架

本为了给出交替方向乘子法的迭代框架, 我们首先给出问题(3)的增广拉格朗日函数:

$$\mathcal{L}_\mu(L, S, Y) = \Phi(L) + \lambda \|S\|_{\ell_1} + \langle Y, D - L - S \rangle + \frac{\mu}{2} \|D - L - S\|_F^2,$$

其中 $Y \in R^{m \times n}$ 是 Lagrange 乘子, $\mu > 0$ 表示违反线性约束的惩罚参数。

给定 (L_k, S_k, Y_k) , 那么用于解决问题(3)的交替方向乘子法可以表示为求解如下三个子问题:

$$\begin{cases} S_{k+1} = \arg \min_S \mathcal{L}_\mu(L_k, S, Y_k) \\ L_{k+1} = \arg \min_L \mathcal{L}_\mu(L, S_{k+1}, Y_k) \\ Y_{k+1} = Y_k + \mu(D - L_{k+1} - S_{k+1}). \end{cases} \quad (4)$$

很容易可以看出(4)式中的第一个子问题可以表示为如下形式:

$$\begin{aligned} S_{k+1} &= \arg \min_S \lambda \|S\|_{\ell_1} + \langle Y_k, D - L_k - S \rangle + \frac{\mu}{2} \|D - L_k - S\|_F^2 \\ &= \arg \min_S \lambda \|S\|_{\ell_1} + \frac{\mu}{2} \left\| S - \left(D - L_k + \frac{1}{\mu} Y_k \right) \right\|_F^2 \\ &= \arg \min_S \lambda \|S\|_{\ell_1} + \frac{\mu}{2} \|S - \hat{Y}\|_F^2, \end{aligned} \quad (5)$$

其中 $\hat{Y} = D - L_k + \frac{1}{\mu} Y_k$ 。显然(4)式问题有如下的闭式解

$$[S_{k+1}]_j = \text{sign}(\hat{Y}_{ij}) \cdot \max \left\{ \left\| \hat{Y}_{ij} \right\| - \frac{\lambda}{\mu}, 0 \right\}. \quad (6)$$

根据求解(4)中第一个子问题的思想, (4)式中的第二个子问题可以等价的写为

$$\begin{aligned} L_{k+1} &= \arg \min_L \Phi(L) + \langle Y_k, D - L - S_{k+1} \rangle + \frac{\mu}{2} \|D - L - S_{k+1}\|_F^2 \\ &= \arg \min_L \Phi(L) + \frac{\mu}{2} \left\| L - \left(D - S_{k+1} + \frac{1}{\mu} Y_k \right) \right\|_F^2 \\ &= \arg \min_L \Phi(L) + \frac{\mu}{2} \|L - \bar{Y}\|_F^2, \end{aligned} \quad (7)$$

其中 $\bar{Y} = D - S_{k+1} + \frac{1}{\mu} Y_k$, 那么可得到问题(7)的解

$$L_{k+1} = T_{Capped\ell_1}(\bar{Y}) := UT_{Capped\ell_1}(\Sigma)V^T, \tag{8}$$

这里 $U\Sigma V^T = \bar{Y}$ 是 \bar{Y} 的奇异值分解, $T_{Capped\ell_1}(\Sigma)$ 是矩阵 Σ 的阈值算子, 即

$$T_{Capped\ell_1}(\Sigma)_{ii} = \begin{cases} \max\left\{\Sigma_{ii} - \frac{1}{\gamma\mu}, 0\right\}, & \Sigma_{ii} \leq \gamma + \frac{1}{2\gamma\mu}, \\ \Sigma_{ii}, & \Sigma_{ii} > \gamma + \frac{1}{2\gamma\mu}. \end{cases} \tag{9}$$

上述交替方向乘子法的关键是在每次迭代中交替更新低秩矩阵、稀疏矩阵和 Lagrange 乘子, 直到满足算法的收敛条件。因此, 算法整体的计算精度与时间代价在很大程度上取决于式(4)中两个子问题的求解。(6)与(8)给出了(4)中两个子问题的显式解, 因此我们提出了如下交替方向乘子法解问题(3):

算法 2.1 交替方向乘子法(ADMM)

初始步: 初始化 $\{Y_0, L_0, S_0, \mu_0\}$

步 1: 计算:

$$\begin{aligned} S_{k+1} &= \arg \min_S \frac{\mu_k}{2} \left\| S - \left(D - L_k + \frac{1}{\mu_k} Y_k \right) \right\|_F^2 + \lambda \|S\|_{\ell_1} \\ L_{k+1} &= \arg \min_L \frac{\mu_k}{2} \left\| L - \left(D - S_{k+1} + \frac{1}{\mu_k} Y_k \right) \right\|_F^2 + \Phi(L) \\ Y_{k+1} &= Y_k + \mu_k (D - L_{k+1} - S_{k+1}) \end{aligned} \tag{10}$$

步 2: 设置 $\mu_{k+1} = \rho\mu_k$ 。

步 3: 判断 $\frac{\|D - L_{k+1} - S_{k+1}\|_F}{\|D\|_F} < \text{xtol}$ 是否成立, 是则终止迭代; 否则令 $k = k + 1$, 转步 1。

输出 L_k, S_k

2.2. 收敛性分析

本小结给出算法的收敛性分析, 针对算法(2.1)有如下结论:

定理 2.1 令 $\{(L_k, S_k, Y_k)\}$ 是由 ADMM 算法生成的迭代序列, 则以下结论成立:

- i) 序列 $\{(L_k, S_k, Y_k)\}$ 是有界的;
- ii) $\lim_{k \rightarrow \infty} \|S_{k+1} - S_k\|_F = 0, \lim_{k \rightarrow \infty} \|L_{k+1} - L_k\|_F = 0$;
- iii) $\lim_{k \rightarrow \infty} \|D - S_{k+1} - L_{k+1}\|_F = 0$, 且 $\{(L_k, S_k, Y_k)\}$ 的聚点是问题的可行解。

证明. i) 令 $U_k \Lambda_k V_k^T$ 为矩阵 $\bar{Y}_k := D - S_{k+1} + \frac{1}{\mu_k} Y_k$ 在 $k+1$ 次迭代中的奇异值分解, 根据 L_{k+1} 在(8)中的定义, 有

$$L_{k+1} = U_k T_{Capped\ell_1}(\Lambda_k) V_k^T. \tag{11}$$

再由(10)中的第三个等式和(11), 可得

$$\begin{aligned} \|Y_{k+1}\|_F &= \|Y_k + \mu_k (D - L_{k+1} - S_{k+1})\|_F \\ &= \mu_k \left\| \frac{1}{\mu_k} Y_k + D - L_{k+1} - S_{k+1} \right\|_F \\ &= \mu_k \left\| U_k \Lambda_k V_k^\top - U_k T_{Capped\ell_1}(\Lambda_k) V_k^\top \right\|_F \\ &= \mu_k \left\| \Lambda_k - T_{Capped\ell_1}(\Lambda_k) \right\|_F \\ &\leq \frac{\sqrt{m}}{\gamma} < \infty, \end{aligned}$$

这里 $T_{Capped\ell_1}(\cdot)$ 定义如(9)式所示, 因此 $\{Y_k\}$ 是有界的。

接下来, 我们考虑序列 $\{S_k\}$ 的有界性。由(10), 我们得到

$$\mathcal{L}(L_{k+1}, S_{k+1}, Y_k, \mu_k) \leq \mathcal{L}(L_{k+1}, S_k, Y_k, \mu_k) \leq \mathcal{L}(L_k, S_k, Y_k, \mu_k), \tag{12}$$

联合迭代公式 $Y_{k+1} = Y_k + \mu_k (D - L_{k+1} - S_{k+1})$, 可得

$$\begin{aligned} \mathcal{L}(L_{k+1}, S_{k+1}, Y_k, \mu_k) &\leq \mathcal{L}(L_k, S_k, Y_k, \mu_k) \\ &= \mathcal{L}(L_k, S_k, Y_{k-1}, \mu_{k-1}) + \langle Y_k - Y_{k-1}, D - L_k - S_k \rangle + \frac{\mu_k - \mu_{k-1}}{2} \|D - L_k - S_k\|_F^2 \\ &= \mathcal{L}(L_k, S_k, Y_{k-1}, \mu_{k-1}) + \frac{\mu_k - \mu_{k-1}}{2} \|D - L_k - S_k\|_F^2 + \left\langle Y_k - Y_{k-1}, \frac{1}{\mu_{k-1}} (Y_k - Y_{k-1}) \right\rangle \tag{13} \\ &= \mathcal{L}(L_k, S_k, Y_{k-1}, \mu_{k-1}) + \frac{\mu_k + \mu_{k-1}}{2\mu_{k-1}^2} \|Y_k - Y_{k-1}\|_F^2 \\ &= \mathcal{L}(L_k, S_k, Y_{k-1}, \mu_{k-1}) + \frac{\rho + 1}{2\mu_{k-1}} \|Y_k - Y_{k-1}\|_F^2. \end{aligned}$$

通过(12)、(13)和 $\rho > 1$, 立即可得

$$\mathcal{L}(L_{k+1}, S_{k+1}, Y_k, \mu_k) \leq \mathcal{L}(L_1, S_1, Y_0, \mu_0) + \frac{(\rho + 1)\Theta}{2\mu_0} \sum_{k=1}^{\infty} \left(\frac{1}{\rho}\right)^{k-1} < \infty,$$

这里 Θ 是 $\|Y_k - Y_{k-1}\|_F^2$ 的上界。上式也表明 $\mathcal{L}(L_{k+1}, S_{k+1}, Y_k, \mu_k)$ 是有上界的。另外我们有

$$\begin{aligned} \Phi(L_k) + \lambda \|S_k\|_{\ell_1} &= \mathcal{L}(L_k, S_k, Y_{k-1}, \mu_{k-1}) - \langle Y_{k-1}, D - L_k - S_k \rangle - \frac{\mu_{k-1}}{2} \|D - L_k - S_k\|_F^2 \\ &= \mathcal{L}(L_k, S_k, Y_{k-1}, \mu_{k-1}) + \left\langle Y_{k-1}, \frac{1}{\mu_{k-1}} (Y_k - Y_{k-1}) \right\rangle - \frac{1}{2\mu_{k-1}} \|Y_k - Y_{k-1}\|_F^2 \tag{14} \\ &= \mathcal{L}(L_k, S_k, Y_{k-1}, \mu_{k-1}) - \frac{1}{2\mu_{k-1}} (\|Y_k\|_F^2 - \|Y_{k-1}\|_F^2) < \infty. \end{aligned}$$

因为 $\|S_k\|_{\ell_1}$ 是具有强制性的, 即 $\lim_{S \rightarrow \infty} \|S\|_{\ell_1} = +\infty$, 且 $\Phi(L_k)$ 非负的, 因此由(14)可得到序列 $\{S_k\}$ 有界。

最后考虑序列 $\{L_k\}$ 的有界性。特别地, 由 μ_k 的更新准则知 $\lim_{k \rightarrow +\infty} \mu_k = +\infty$, 以及 $\{Y_k\}$ 的有界性, 因此有如下结果

$$\lim_{k \rightarrow \infty} \|D - L_{k+1} - S_{k+1}\|_F = \lim_{k \rightarrow \infty} \frac{1}{\mu_k} \|Y_{k+1} - Y_k\|_F = 0, \tag{15}$$

此外还有

$$\begin{aligned} \|L_{k+1}\|_F &= \|D - S_{k+1} - (D - L_{k+1} - S_{k+1})\|_F \\ &\leq \|D\|_F + \|S_{k+1}\|_F + \|D - L_{k+1} - S_{k+1}\|_F. \end{aligned} \tag{16}$$

那么由(15)、(16)以及 $\{S_{k+1}\}$ 的有界性, 可知 $\{L_{k+1}\}$ 有界。综上可得序列 $\{(L_k, S_k, Y_k)\}$ 有界。

ii) 我们首先证明 $\lim_{k \rightarrow +\infty} \|S_{k+1} - S_k\|_F = 0$ 。由(10)的第三个更新准则以及(5)可知

$$\begin{aligned} \lim_{k \rightarrow +\infty} \|S_{k+1} - S_k\|_F &= \lim_{k \rightarrow +\infty} \|S_{k+1} - (D - L_k - \mu_{k-1}^{-1}Y_k + \mu_{k-1}^{-1}Y_{k-1})\|_F \\ &= \lim_{k \rightarrow +\infty} \|S_{k+1} - (D - L_k + \mu_k^{-1}Y_k) - (\mu_{k-1}^{-1}Y_{k-1} - \mu_k^{-1}Y_k - \mu_{k-1}^{-1}Y_k)\|_F \\ &\leq \lim_{k \rightarrow +\infty} \|S_{k+1} - (D - L_k + \mu_k^{-1}Y_k)\|_F + \|\mu_{k-1}^{-1}Y_{k-1} - \mu_k^{-1}Y_k - \mu_{k-1}^{-1}Y_k\|_F \\ &\leq \lim_{k \rightarrow +\infty} \left[\left(\sum_{i,j} [S_{k+1}]_{ij} - [D - L_k + \mu_k^{-1}Y_k]_{ij} \right)^2 \right]^{1/2} + \|\mu_{k-1}^{-1}Y_{k-1} - \mu_k^{-1}Y_k - \mu_{k-1}^{-1}Y_k\|_F \\ &\leq \lim_{k \rightarrow +\infty} \left[\frac{\lambda \sqrt{mn}}{\mu_k} + \mu_{k-1}^{-1} \|Y_k - Y_{k-1}\|_F + \mu_k^{-1} \|Y_k\|_F \right] \\ &= 0, \end{aligned} \tag{17}$$

这里最后等式成立是因为 $\|Y_k - Y_{k-1}\|_F$ 和 $\|Y_k\|_F$ 有界且 $\lim_{k \rightarrow +\infty} \mu_k = +\infty$ 。再由(10)的第三个更新准则以及(8)、(9)和(11)可知

$$\begin{aligned} \lim_{k \rightarrow +\infty} \|L_{k+1} - L_k\|_F &= \lim_{k \rightarrow +\infty} \|L_{k+1} - (D - S_k + \mu_{k-1}^{-1}Y_{k-1} - \mu_{k-1}^{-1}Y_k)\|_F \\ &= \lim_{k \rightarrow +\infty} \|\mathbb{T}_{Capped \ell_1}(D - S_{k+1} + \mu_k^{-1}Y_k) - (D - S_{k+1} + \mu_k^{-1}Y_k) - (S_{k+1} - S_k) - (\mu_{k-1}^{-1}Y_{k-1} - \mu_k^{-1}Y_k - \mu_{k-1}^{-1}Y_k)\|_F \\ &\leq \lim_{k \rightarrow +\infty} \|\Lambda_k - T_{Capped \ell_1}(\Lambda_k)\|_F + \|S_{k+1} - S_k\|_F + \mu_k^{-1} \|Y_k\|_F + \mu_{k-1}^{-1} \|Y_k - Y_{k-1}\|_F \\ &\leq \lim_{k \rightarrow +\infty} \frac{m}{\gamma \mu_k} + \|S_{k+1} - S_k\|_F + \mu_k^{-1} \|Y_k\|_F + \mu_{k-1}^{-1} \|Y_k - Y_{k-1}\|_F \\ &= 0. \end{aligned} \tag{18}$$

最后由式子(15)、(17)和(18)可知结论 ii) 成立。

iii) 由(i)可知, 因为序列 $\{(L_k, S_k, Y_k)\}$ 有界, 所以存在聚点 (L^*, S^*, Y^*) 和一个收敛子序列 $(L_{k_j}, S_{k_j}, Y_{k_j})$ 使得 $\lim_{j \rightarrow +\infty} (L_{k_j}, S_{k_j}, Y_{k_j}) = (L^*, S^*, Y^*)$, 根据(10)式中的第三个等式, 可知

$$L_{k_j+1} + S_{k_j+1} - D = \frac{1}{\mu_{k_j}} (Y_{k_j} - Y_{k_j+1}) \tag{19}$$

于是令 $j \rightarrow +\infty$, 有 $L^* + S^* - D = \frac{1}{\mu^*} (Y^* - Y^*) = 0$, 故聚点为问题的可行解。

3. 数值实验

本节我们将对所提出的 ADMM 算法与以下三种算法进行比较: Augmented Lagrange Multiplier (ALM) 算法[13], MoG [21]算法和 Weighted Nuclear Norm Minimization (WNNM) [22]算法。对于待分解的矩阵 $D \in \mathbb{R}^{m \times m}$, 用矩阵中非零元素的比例(记作 pe)来刻画其稀疏成分 S 的稀疏程度, 用矩阵秩 r 与维数 m 的比值(记作 pr)来刻画其低秩成分 L 的低秩程度。我们采用与相关研究[6] [12] [23]似的模式来生成矩阵 D 。

对于三种比较算法的参数, 我们依然选取他们原始的取值。例如 ALM 算法的原始参数 $\rho = 1.1$, $\alpha = 1$, $\beta = 0.2$ 。MoG 算法中的参数 $\mu_0 = 0$, $\beta_0 = 1e-6$, $\alpha_0 = 1e-6$ 。WNNM 算法的参数 $\lambda = 1/\sqrt{m}$, $\rho = 1.05$ 。

对于我们的 ADMM 算法, 设置参数 $\rho=1.1$, $\gamma=0.25$, $\lambda=1/\sqrt{m}$, $\text{maxiter}=1000$, $\text{xtol}=1e-7$ 。

我们算法的所有结果都是在台式电脑(CPU 为 3.40 GHz, 内存为 8.00 GB)上使用软件 Matlab R2018b 完成的。

3.1. 随机实验模拟

本节将在有、无噪声干扰的情况下, 比较 ALM、MoG、WNNM 和 ADMM 这四种算法实现稀疏 - 低秩矩阵的精度与速度, 并研究算法对于低秩成分的低秩程度与稀疏成分的稀疏程度的稳健性。

3.1.1. 无噪声情况下的实验结果

在本结中, 首先生成待分解矩阵 $D \in \mathbb{R}^{m \times m}$, 具体生成如下: 1) 独立生成随机矩阵 $U, V \in \mathbb{R}^{m \times r}$, 这里 $r = pr \times m$, $U = \text{randn}(m, r)$ 和 $V = \text{randn}(m, r)$ 。令 $L = UV'$, 得到秩为 r 的低秩矩阵。2) 均匀随机选取稀疏矩阵 $S \in \mathbb{R}^{m \times m}$ 的 $pe \times m^2$ 个非零元的位置, 这些位置上元素的取值独立服从 $[0, 1]$ 区间上的均匀分布。3) 令 $D = L + S$, 得到待分解矩阵 D 。其次, 设置 $m = 500:500:3000$, $pr = 0.01$, $pe = 0.05$ 。对于以上问题, 我们运行 10 次并记录以下结果: 算法达到收敛所需的时间(CPU Time)、迭代数(Iterations)和相对误差, 其中相对误差的定义如下:

$$\text{rel_err}L = \frac{\|L - L^*\|_F}{\|L\|_F}, \text{rel_err}S = \frac{\|S - S^*\|_F}{\|S\|_F}.$$

计算结果见表 1 和图 1 所示。

Table 1. The numeric results of ALM, MoG, WNNM and ADMM algorithms for sparse low-rank matrix factorization problems of different scales ($\sigma = 0$)

表 1. ALM, MOG, WNNM 和 ADMM 算法对不同规模的稀疏低秩矩阵分解问题的数值结果($\sigma = 0$)

Promblem	Algorithm	rel_errL	rel_errS	Iterations	Time
$m = n = 500$	ADMM	8.69e-09	1.06e-06	100	0.87
	WNNM	9.85e-09	1.09e-06	138	1.35
	ALM	6.53e-02	1.12e+00	51	22.27
	MoG	1.37e+00	0.00e+00	16	1.17
$m = n = 1000$	ADMM	1.39e-09	2.23e-06	103	5.72
	WNNM	1.41e-08	2.25e-06	188	10.30
	ALM	4.17e-02	1.01e+00	50	133.66
	MoG	2.22e+00	0.00e+00	16	3.89
$m = n = 1500$	ADMM	1.38e-08	2.82e-06	105	17.03
	WNNM	1.39e-08	2.83e-06	190	32.39
	ALM	3.77e-01	1.13e+01	35	640.07
	MoG	2.81e+00	0.00e+00	15	9.24
$m = n = 2000$	ADMM	1.37e-08	3.19e-06	106	94.52
	WNNM	1.46e-08	3.34e-06	192	73.97
	ALM	6.62e-01	2.29e+01	31	1802.02
	MoG	3.32e+00	0.00e+00	15	16.78

Continued

$m = n = 2500$	ADMM	1.39e-08	3.62e-06	107	86.27
	WNNM	1.43e-08	3.73e-06	193	73.97
	ALM	7.35e-01	2.84e+01	29	3816.63
	MoG	3.77e+00	0.00e+00	15	27.38
$m = n = 3000$	ADMM	1.37e-08	3.95e-06	107	131.47
	WNNM	1.42e-08	4.05e-06	196	250.85
	ALM	7.83e-01	3.31e+01	28	4467.64
	MoG	4.14e+00	0.00e+00	15	205.32

由表 1 可以看出: 在无噪声情况下, ADMM 算法虽然 Iterations 高于 ALM 和 MoG 算法, 但是 ADMM 算法的 rel_errL、rel_errS 和 CPU Time 都是小于其他算法的, 并且随着矩阵维数 m, n 的增加, 同样可以得到 ALM 与 MoG 算法比 ADMM 算法的 rel_errL、rel_errS 和 CPU Time 指标值高, 即 ADMM 算法分解得到的 rel_errL、rel_errS、Iterations 都优于 ALM、MoG 算法。对于 WNNM 算法而言, 无论是 rel_errL、rel_errS、Iterations 和 CPU Time 都高于 ADMM 算法。因此, ADMM 算法总体上来看优于其他三个算法, 能够准确地分解稀疏 - 低秩矩阵。

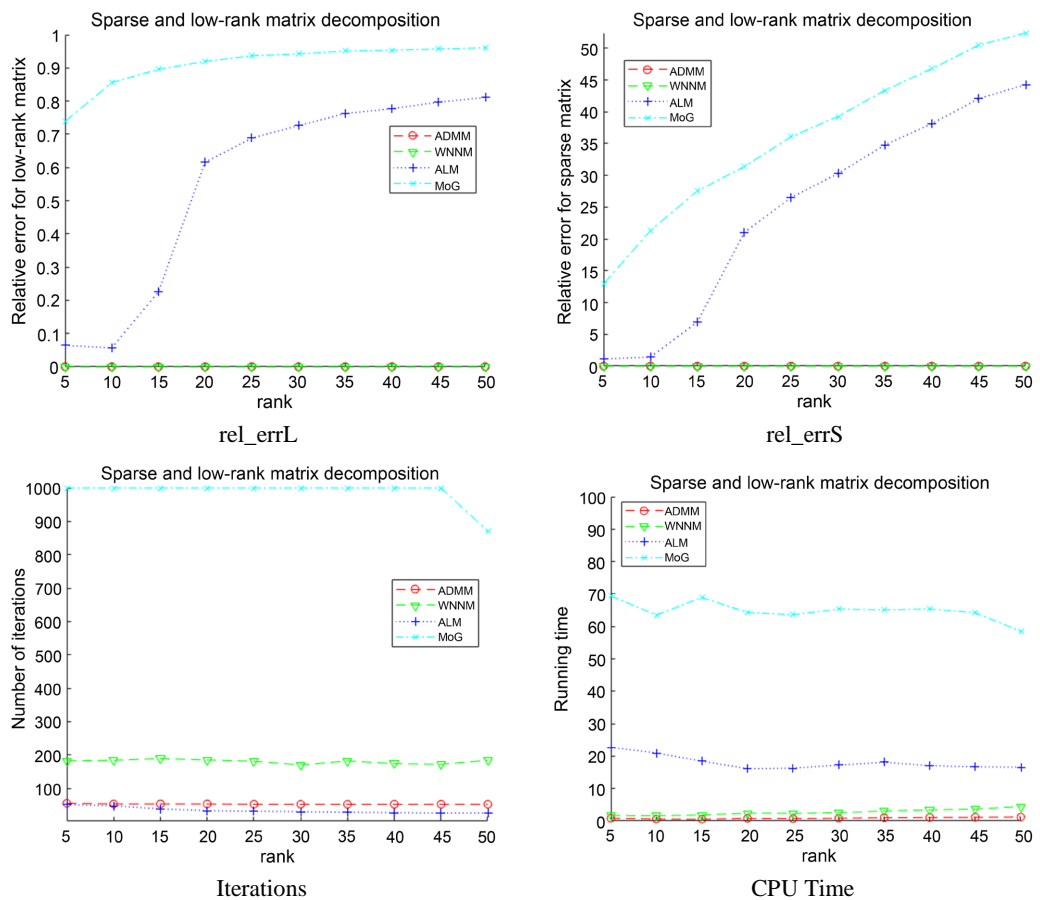


Figure 1. The numeric results for sparse low-rank matrix factorization problems of different scales ($\sigma = 0$)

图 1. 对不同规模的稀疏低秩矩阵分解问题的数值结果($\sigma = 0$)

图 1 记录了四个算法在无噪声情况下，对维度为 $m = n = 500$ ， $pe = 0.05$ ， $r = 5:5:50$ 的稀疏 - 低秩矩阵分解的实验结果。从计算精度上来看，MoG、ALM 算法分解出的稀疏 - 低秩矩阵精度较低相对误差较大，并且随着 r 的不断增大，其分解结果的相对误差也呈现出比较大的递增趋势，而本文的 ADMM 算法分解效果不仅准确，而且也比较稳定。从计算效率上来看，MoG 算法的 Iterations 和 CPU Time 比其他几种算法都多，并且 ALM、WNNM 算法的效率也不高，但是对于 ADMM 算法而言，随着 r 的不断增大，它不仅能够准确地分解稀疏 - 低秩矩阵，而且具有较好的稳定性，综上 ADMM 算法要比 WNNM、ALM 和 MoG 算法能够快速、精确地分解稀疏低秩矩阵。因此 ADMM 算法在无噪声情况下是优于其他算法的。

3.1.2. 有噪声情况下的实验结果

待分解矩阵含噪声的情况下数值结果如表 2 和图 2 所示。

Table 2. The numeric results of ALM, MoG, WNNM and ADMM algorithms for sparse low-rank matrix factorization problems of different scales ($\sigma = 0$)

表 2. ALM, MOG, WNNM 和 ADMM 算法对不同规模的稀疏低秩矩阵分解问题的数值结果($\sigma = 0$)

Promblem	Algorithm	rel_errL	rel_errS	Iterations	Time
$m = n = 500$	ADMM	2.38e-03	2.25e-02	146	3.84
	WNNM	7.64e-04	4.52e-02	146	2.95
	ALM	6.72e-02	1.17e+00	50	19.79
	MoG	1.44e+00	0.00e+00	24	1.83
$m = n = 1000$	ADMM	1.68e-03	2.25e-02	146	26.19
	WNNM	5.65e-04	4.52e-02	146	34.62
	ALM	4.39e-02	1.07e+00	49	103.59
	MoG	2.26e+00	0.00e+00	19	4.56
$m = n = 1500$	ADMM	1.37e-03	2.25e-02	146	75.31
	WNNM	1.17e-03	2.28e-02	146	152.45
	ALM	5.15e-01	1.54e+01	34	389.45
	MoG	2.90e+00	0.00e+00	15	11.38
$m = n = 2000$	ADMM	1.18e-03	2.25e-02	146	156.07
	WNNM	1.07e-03	2.25e-02	146	319.90
	ALM	6.64e-01	2.30e+01	31	1180.51
	MoG	3.39e+00	0.00e+00	25	24.08
$m = n = 2500$	ADMM	1.06e-03	2.25e-02	146	269.40
	WNNM	9.78e-04	2.25e-02	146	659.92
	ALM	7.37e-01	2.87e+01	29	2489.94
	MoG	3.83e+00	0.00e+00	15	39.97

Continued

$m = n = 3000$	ADMM	9.69e-04	2.25e-02	146	446.00
	WNNM	9.07e-04	2.25e-06	146	954.96
	ALM	7.82e-01	3.32e+01	28	4528.83
	MoG	4.22e+00	0.00e+00	20	47.86

由表 2 可以看出：在有噪声的情况下，ADMM 算法在时间上慢于 MoG 算法，但是 MoG 算法不能成功分解稀疏 - 低秩矩阵，其中对于 ALM 算法来说，虽然 Iterations 比 ADMM 算法低，但是 ALM 算法的 CPU Time 比 ADMM 高出一个数量级，并且其计算精度远不如 ADMM 算法。对于 WNNM 算法，通过表中数据看出，ADMM 算法与 WNNM 算法在计算精度和 Iterations 相差不是很大，但是可以明显得到 WNNM 算法的 CPU Time 比 ADMM 算法的所需时间长得多，其次分解得到的低秩矩阵与原始矩阵的误差虽然大于 WNNM 算法，但是得到的稀疏矩阵要更近似原始的稀疏矩阵，并优于其他两个算法。因此，ADMM 算法具有更加准确和稳健地分解稀疏 - 低秩矩阵的性能，且优于其他三种算法。因此我们的算法有能更稳健的恢复性能。

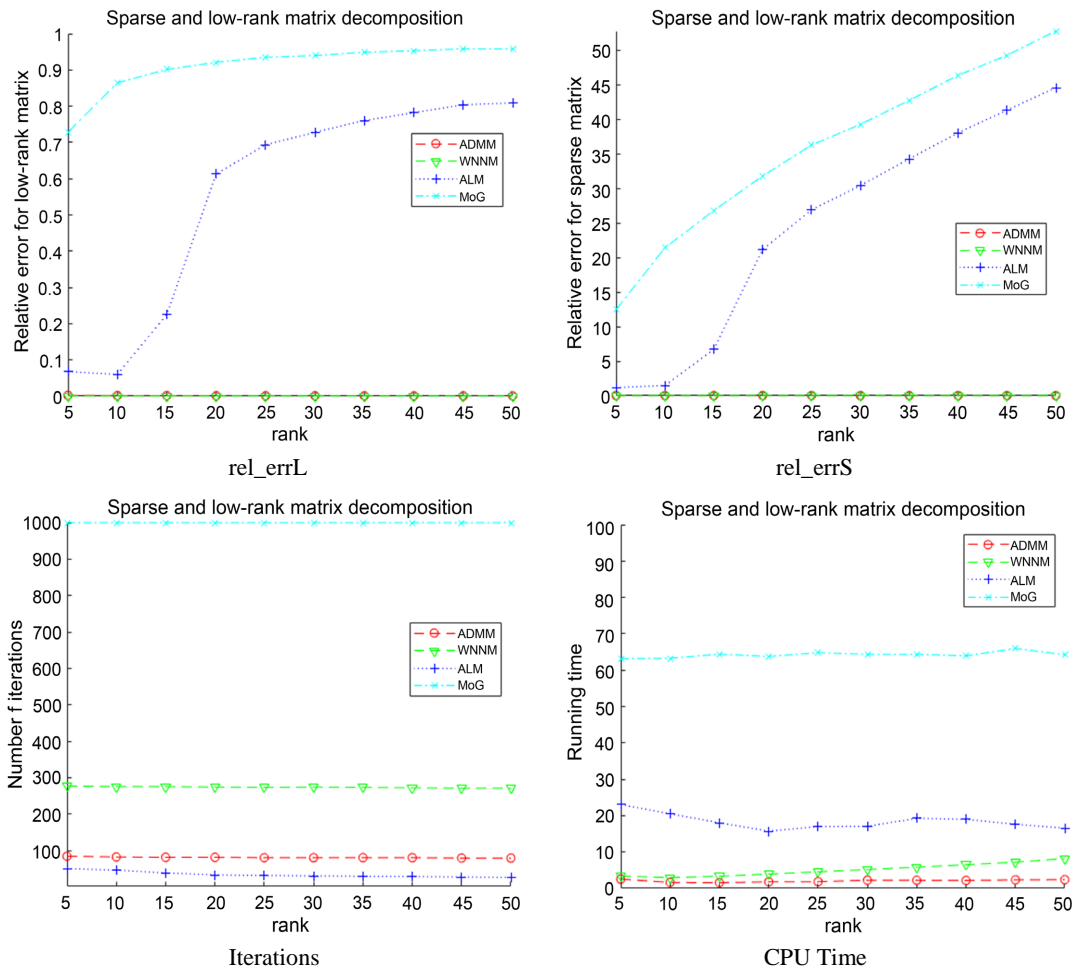


Figure 2. The numeric results for sparse low-rank matrix factorization problems of different scales ($\sigma = 0.01$)

图 2. 对不同规模的稀疏低秩矩阵分解问题的数值结果($\sigma = 0.01$)

图 2 记录了四个算法在有噪声情况下，对维度为 $m = n = 500$ ， $pe = 0.05$ ， $r = 5:5:50$ 的稀疏 - 低秩矩阵分解的实验结果。由图同样可以看出，提出的 ADMM 算法在迭代数上多于 ALM，但比其他算法速度快并且更能精确的分解稀疏低秩矩阵。总之，可以得到我们所提出的算法在有噪声的情况下，都能有效、稳健地分解稀疏 - 低秩矩阵。

3.2. 监控视频应用

在监控视频应用中，最重要的环节是对视频进行背景消除。首先我们从文献[24]的 12R 数据集中选取视频，详情如下：

- Campus 该视频包含 80 个 240×320 的视频帧；
- Lobby 该视频包含 80 个 128×160 的视频帧；
- Escalator 该视频包含 80 个 128×160 的视频帧；

其次，为了定量评估背景消除的结果，我们使用基于四个量的度量：TP、TN、FP 和 FN，其中 TP 是为正确分类为正数的总数——即前景像素被正确分类为前景；TN 是为正确分类为负数的总数——即背景像素被正确分类为背景；FN 是错误分类为负数的总数——即前景像素被错误地分类为背景；FP 是错误分类为正数的总数——即背景像素被错误地分类为前景。那么可得到 recall，precision 和 F-measure 三个指标，如下所示：

$$\text{recall} = \frac{TP}{TP+FN}, \quad \text{precision} = \frac{TP}{TP+FP}, \quad \text{F-measure} = 2 \frac{\text{recall} * \text{precision}}{\text{recall} + \text{precision}}.$$

我们再记录每个算法使用的时间，那么可以说一个更高的召回率、精度和 F-度量值以及更少的时间代表更好的算法性能。最后，所有算法的参数取值依然和随机实验部分一样并将结果展现在表 3 和图 3 中。

Table 3. The numeric results of all the compared algorithm

表 3. 所有比较算法的数值结果

Promblem	Algorithm	precision	recall	F-measure	Time
Campus	ADMM	0.4170	0.6296	0.5017	6.0355
	WNNM	0.1425	0.6866	0.2360	19.2749
	ALM	0.0725	0.6439	0.1303	141.8244
	MoG	0.1427	0.6895	0.2358	61.5361
Lobby	ADMM	0.9735	0.8235	0.8923	5.1696
	WNNM	0.9767	0.8235	0.8936	13.7341
	ALM	0.9963	0.7479	0.8544	153.1680
	MoG	0.9703	0.8235	0.8909	39.4163
Escalator	ADMM	0.2606	0.7742	0.3899	6.6002
	WNNM	0.2438	0.7616	0.3694	19.1113
	ALM	0.1373	0.7652	0.2328	156.0334
	MoG	0.2377	0.7867	0.3651	63.2443

由表 3 可以看出, 我们算法在对视频进行背景消除实验中, 所需要的时间最短, 而且几乎比 WNNM、ALM、MoG 算法有更高的 precision, recall, F-measure 值。因此, 我们的 ADMM 算法比其他算法处理前景背景分离更有优势。

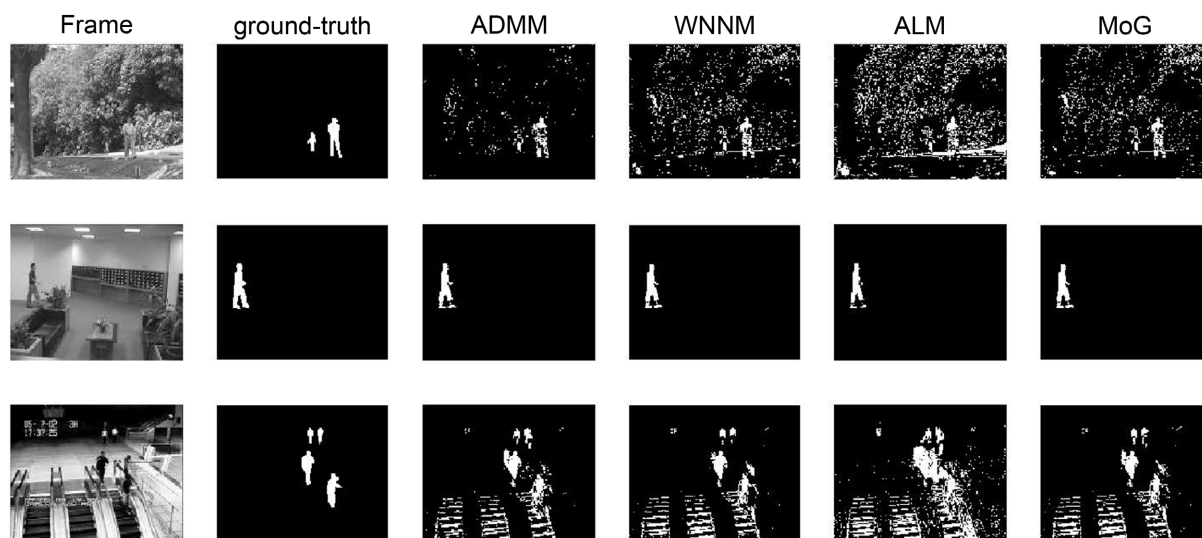


Figure 3. Experimental results on background subtraction

图 3. 背景消除的数值结果

由图 3 中的结果表明, 我们所提出的算法相比于其他三种算法可以更有效、清晰地分离背景和前景。因此, 我们的算法要比其他算法更具优势。

4. 总结

本文将研究向量稀疏优化中的 Capped- l_1 正则松弛理论拓展到矩阵问题中, 进而提出了可适用于矩阵问题的 Capped- l_1 阈值算子, 并将其应用于非凸非光滑优化模型。首先, 我们在交替方向乘子法框架下求解该模型, 并证明我们所提出的算法具有稳定的可行解。其次, 通过大量的数值模拟实验表明: 与其他算法相比, 无论是在有噪声还是无噪声的情况下, 我们的算法都能够快速准确地分解出低秩 - 稀疏矩阵。特别地, 在有噪声的情况下, 我们的算法还具有较强的稳健性。最后, 我们成功地将这种新算法应用于监控视频背景建模这一实际问题, 并且也验证了在处理实际问题时我们的算法能够进行很好地背景消除。

基金项目

国家自然科学基金项目(12261020); 贵州省科技计划项目(ZK [2021] 009); 贵州大学 SRT 项目(贵大 SRT 字(2022) 461 号)。

参考文献

- [1] Wright, J., Ganesh, A., Rao, S. and Ma, Y. (2009) Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices. *Advances in Neural Information Processing Systems*, **87**, 3-56.
- [2] Chinh, D. and Hayder, R. (2015) RPCA-KFE: Key Frame Extraction for Video Using Robust Principal Component Analysis. *IEEE Transactions on Image Processing*, **24**, 3742-3753. <https://doi.org/10.1109/TIP.2015.2445572>
- [3] Moradikia, M. and Samadi, S. (2018) SAR Image Formation and Feature Separation via Low Rank Sparse Decomposition. *Iranian Conference on Electrical Engineering (ICEE)*, Mashhad, 8-10 May 2018, 737-742. <https://doi.org/10.1109/ICEE.2018.8472666>
- [4] He, R., Hu, B.G., Zheng, W.S. and Kong, X.W. (2011) Robust Principal Component Analysis Based on Maximum

- Correntropy Criterion. *IEEE Transactions on Image Processing*, **20**, 1485-1494. <https://doi.org/10.1109/TIP.2010.2103949>
- [5] Yuki, Y., Kensuke, T. and Hiroshi, Y. (2019) Constrained Nonmetric Principal Component Analysis. *Behaviormetrika*, **46**, 313-332. <https://doi.org/10.1007/s41237-019-00087-3>
- [6] Candes, E.J., Li, X.D., Ma, Y. and Wright, J. (2011) Robust Principle Component Analysis. *Journal of Association for Computing Machinery*, **58**, 1-37. <https://doi.org/10.1145/1970392.1970395>
- [7] Hu, Y., Zhang, D.B., Ye, J.P., Li, X.L. and He, X.F. (2013) Fast and Accurate Matrix Completion via Truncated Nuclear Norm Regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**, 2117-2130. <https://doi.org/10.1109/TPAMI.2012.271>
- [8] Hong, B., Wei, L., Hu, Y., Cai, D. and He, X.F. (2016) Online Robust Principal Component Analysis via Truncated Nuclear Norm Regularization. *Neurocomputing*, **175**, 216-222. <https://doi.org/10.1016/j.neucom.2015.10.052>
- [9] Kim, H. and Choe, Y. (2017) Background Subtraction via Truncated Nuclear Norm Minimization. *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Kuala Lumpur, 12-15 December 2017, 447-451. <https://doi.org/10.1109/APSIPA.2017.8282073>
- [10] Oh, T.H., Tai, Y.W., Bazin J.C., Kim, H. and Kweon, I.S. (2016) Partial Sum Minimization of Singular Values in Robust PCA: Algorithm and Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**, 744-758. <https://doi.org/10.1109/TPAMI.2015.2465956>
- [11] Massimo, F. and Holger, R. (2008) Iterative Thresholding Algorithms. *Applied and Computational Harmonic Analysis*, **25**, 187-208. <https://doi.org/10.1016/j.acha.2007.10.005>
- [12] Lin, Z.C., Ganesh, A., Wright, J., Wu, L.Q., Chen, M.M. and Ma, Y. (2009) Fast Convex Optimization Algorithms for Exact Recovery of a Corrupted Low-Rank Matrix. *Journal of the Marine Biological Association of the UK*, **56**, 707-722. <https://doi.org/10.1109/CAMSAP.2009.5413299>
- [13] Kim, M.S. and Choi, D.H. (2000) A New Penalty Parameter Update Rule in the Augmented Lagrange Multiplier Method for Dynamic Response Optimization. *KSME International Journal*, **14**, 1122-1130. <https://doi.org/10.1007/BF03185066>
- [14] Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. (2010) Distributed Optimization and Statistical Learning via Thealternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning*, **3**, 1-122. <https://doi.org/10.1561/22000000016>
- [15] Fang, E.X., He, B.S., Liu, H. and Yuan, X.M. (2015) Generalized Alternating Direction Method of Multipliers: New Theoretical Insights and Applications. *Mathematical Programming Computation*, **7**, 149-187. <https://doi.org/10.1007/s12532-015-0078-2>
- [16] Xiao, Y., Chen, L. and Li, D. (2018) A Generalized Alternating Direction Method of Multipliers with Semi-Proximal Terms for Convex Composite Conic Programming. *Mathematical Programming Computation*, **10**, 533-555. <https://doi.org/10.1007/s12532-018-0134-9>
- [17] Peng, D.T., Xiu, N.H. and Yu, J. (2017) $S_{1/2}$ Regularization Methods and Fixed Point Algorithms for Affine Rank Minimization Problems. *Computational Optimization and Applications*, **67**, 543-569. <https://doi.org/10.1007/s10589-017-9898-5>
- [18] Peng, D.T., Xiu, N.H. and Yu, J. (2018) Global Optimality Condition and Fixed Point Continuation Algorithm for Non-Lipschitz Lp Regularized Matrix Minimization. *Science China Mathematics*, **61**, 1139-1152. <https://doi.org/10.1007/s11425-016-9107-y>
- [19] Zhang, T. (2010) Analysis of Multi-Stage Convex Relaxation for Sparse Regularization. *Journal of Machine Learning Research*, **11**, 1081-1107.
- [20] Zhang, X. and Peng, D.T. (2022) Solving Constrained Nonsmooth Group Sparse Optimization via Group Capped- l_1 Relaxation and Group Smoothing Proximal Gradient Algorithm. *Computational Optimization and Applications*, **83**, 801-844. <https://doi.org/10.1007/s10589-022-00419-2>
- [21] Zhao, Q., Meng, D.Y., Xu, Z.B., Zuo, W.M. and Zhang, L. (2014) Robust Principal Component Analysis with Complex Noise. *Proceedings of the 31st International Conference on Machine Learning*, Vol. 32, 55-63.
- [22] Gu, S.H., Xie, Q., Meng, D.Y., Zuo, W.M., Feng, X.C. and Zhang, L. (2017) Weighted Nuclear Norm Minimization and Its Applications to Low Level Vision. *International Journal of Computer Vision*, **121**, 183-208. <https://doi.org/10.1007/s11263-016-0930-5>
- [23] Mackey, L., Talwalkar, A. and Jordan, M.I. (2011) Divide-and-Conquer Matrix Factorization. *Proceedings of the 24th International Conference on Neural Information Processing Systems NIPS'11*, Granada, 12-15 December 2011, 1134-1142.
- [24] Li, L.Y., Huang, W.M., Gu, I.Y.H. and Tian, Q. (2004) Statistical Modeling of Complex Backgrounds for Foreground Object Detection. *IEEE Transactions on Image Processing*, **13**, 1459-1472. <https://doi.org/10.1109/TIP.2004.836169>