中国科学: 数学 2021 年 第 51 卷 第 4 期: 549～560
SCIENTIA SINICA Mathematica
自然科学基金项目进展专栏 论 文

求解低秩矩阵恢复问题的非单调交替梯度方向法

闫喜红[1], 王川龙[1*], 李超[1], 薛靖婷[2]

1. 太原师范学院数学系, 晋中 030619;
2. 北京工业大学应用数理学院, 北京 100124
E-mail: xihong1@e.ntu.edu.sg, clwang1964@163.com, l_chao1989@163.com, 15804847171@139.com

收稿日期: 2019-07-15; 接受日期: 2020-08-04; 网络出版日期: 2020-12-29; * 通信作者
国家自然科学基金 (批准号: 11901424 和 91630202) 和山西省自然科学基金 (批准号: 201801D121022) 资助项目

摘要 低秩矩阵恢复问题作为一类在图像处理和信号数据分析等领域中都十分重要的问题已被广泛研究. 本文在交替方向算法的框架下, 应用非单调技术, 提出一种求解低秩矩阵恢复问题的新算法. 该算法在每一步迭代过程中, 首先利用一步带有变步长梯度算法同时更新低秩部分的两块变量, 然后采用非单调技术更新稀疏部分的变量. 在一定的假设条件下, 本文证明了新算法的全局收敛性. 最后通过解决随机低秩矩阵恢复问题和视频前景背景分离的实例验证新算法的有效性, 同时也显示非单调技术极大改善了算法的效率.

关键词 低秩矩阵恢复 交替方向法 非单调技术

MSC (2020) 主题分类 15A18, 90C20, 90C25

1 引言

低秩矩阵恢复问题, 也被称为低秩稀疏分解问题, 又被称为鲁棒主成分分析问题, 可描述为从一个已知矩阵 $D \in \mathbb{R}^{m\times n}$ 恢复出一个低秩矩阵 $L \in \mathbb{R}^{m\times n}$ 和一个稀疏矩阵 $S \in \mathbb{R}^{m\times n}$. 低秩矩阵恢复问题在人脸识别、视频图像处理和计算机视觉等领域有着广泛的应用 [1–4]. 例如, 在核磁共振图像的配准中, 将同一个人的多幅图像进行列拉长处理, 形成一个待恢复的矩阵, 利用低秩矩阵恢复可将该矩阵恢复出一个低秩的信号矩阵部分和一个稀疏的噪声矩阵部分, 即核磁共振图像问题就是一个低秩矩阵恢复问题 [5].

通过以上的描述, 低秩矩阵恢复问题的数学模型可表示为如下形式:

$$\min_{L,S} \operatorname{rank}(L) + \|S\|_0, \quad \text{s.t.} \quad D = L + S, \tag{1.1}$$

其中 $\operatorname{rank}(L)$ 表示矩阵 L 的秩, $\|S\|_0$ 表示矩阵 S 非零元的个数. 由于 $\operatorname{rank}(L)$ 和 $\|S\|_0$ 都是非凸非线性的组合优化函数, 模型 (1.1) 是非凸 NP- 难问题, 直接求解很难. Candès 等 [6] 在一定条件下将其松

英文引用格式: Yan X H, Wang C L, Li C, et al. A non-monotone alternating directional method for matrix recovery problem (in Chinese). Sci Sin Math, 2021, 51: 549–560, doi: 10.1360/SSM-2019-0183

弛成如下的凸优化问题进行求解:

$$\min_{L,S} \|L\|_* + \lambda\|S\|_1, \quad \text{s.t.} \quad D = L + S, \tag{1.2}$$

其中 $\|L\|_*$ 表示矩阵 L 的核范数, 其值等于矩阵 L 的奇异值的和; $\|S\|_1$ 表示矩阵 S 的 l_1- 范数, 其值等于矩阵 S 所有元素绝对值的和; λ 是参数. 基于凸模型 (1.2), 已经有很多算法被提出, 其中包括增广 Lagrange 算法 (augmented Lagrange multiplier, ALM) [7]、加速的近似梯度法 (accelerated proximal gradient, APG) [8] 和交替方向法 (alternating direction method of multipliers, ADMM) [9]. 这些基于凸优化松弛模型建立的算法具有很好的理论结果, 但是用于求解实际问题时, 效果并不理想. 为了更加有效求解实际问题中的低秩矩阵恢复问题, Zhou 和 Tao [10] 和 Netrapalli 等 [11] 设计出交替投影非凸算法, 此算法根据求解如下的非凸模型而得:

$$\min_{L,S} \frac{1}{2}\|D - L - S\|_F^2, \quad \text{s.t.} \quad \text{rank}(L) \leqslant r^*, \quad \|S\|_0 \leqslant |\Omega|, \tag{1.3}$$

其中 r^* 是控制低秩矩阵的秩的参数, Ω 是稀疏矩阵的支撑集, $|\Omega|$ 表示集合 Ω 元素的个数. 此非凸算法的基本思想是, 低秩矩阵和稀疏矩阵交替更新. 此算法的每一步迭代, 首先采用奇异值分解算法将矩阵 $D - S$ 投影到秩为 r 的矩阵组成的集合上得到下一步的低秩矩阵; 然后, 将 $D - L$ 通过硬阈值方法投影到稀疏矩阵构成的集合上. 2019 年, Cai 等 [12] 基于非凸模型提出了一种加速的交替投影法. 这类交替投影算法不仅理论上能保证收敛性, 而且相较之前基于凸优化的算法来说也大大提高了求解实际问题的效率. 但是这类方法每一步处理低秩矩阵时需要进行奇异值分解, 随着矩阵规模的增大, 奇异值的计算量也随之变得很大, 其计算成本越来越高. 因此, 这类算法在求解大规模问题时, 效率并不是很高.

矩阵分解模型是处理大规模矩阵恢复问题的一种非常有效的方法, 其基本思想是, 将一个秩为 r 的 $m \times n$ 矩阵分解成一个 $m \times r$ 矩阵和 $r \times n$ 矩阵相乘从而满足低秩的要求, 即 $L = UV$, 其中 U 为 $m \times r$ 矩阵, V 为 $r \times n$ 矩阵. 利用上述的矩阵分解, 模型 (1.3) 可以表示成如下形式:

$$\min_{U,V,S} \frac{1}{2}\|D - UV - S\|_F^2 \quad \text{s.t.} \quad \|S\|_0 \leqslant |\Omega|. \tag{1.4}$$

根据模型 (1.4), Gu 等 [13] 提出了交替最优算法, 其将模型 (1.4) 看作是三块变量 U、V 和 S 的优化问题, 对三块变量进行 Gauss 形式的迭代, 并且利用求解每一块变量对应的最小二乘子问题求得新的相应块变量. 但是, 精确求解一个子问题往往代价很大. 相比精确求解子问题, 易实现的非精确方法更快速、有效. Yi 等 [14] 采用非精确求解子问题的思想提出了交替梯度下降法求解模型 (1.4). 该算法在每一步迭代过程中采用一步带有固定步长的梯度法同时更新变量 U 和 V, 然后利用软阈值方法更新稀疏矩阵 S. 众所周知, 步长很大程度地决定一个算法的效率. 因此, 本文针对模型 (1.4) 采用一步带有变步长的梯度法同时更新矩阵变量 U 和 V, 然后采用非单调技术更新稀疏矩阵 S. 我们知道, 对于求解优化问题, 非单调技术具有很好的数值表现, 并且有助于算法的快速收敛 [15]. 基于此, 本文提出一种新的非单调交替梯度法求解低秩矩阵恢复模型 (1.4), 并且证明这种新算法的全局收敛性.

本文余下内容的结构如下: 第 2 节描述新算法, 并分析新算法的一些性质; 第 3 节证明新算法的收敛性; 第 4 节给出一些初步的数值实验验证新算法的有效性; 最后, 给出本文的总结.

符号说明 为了描述方便, 首先定义

$$f(U,V,S) = \frac{1}{2}\|D - UV - S\|_F^2.$$

当 V 和 S 固定时, $f(U,V,S)$ 只看作是 U 的函数, 记为 $f_{V,S}(U)$; 同理, 用 $f_{U,S}(V)$ 表示 $f(U,V,S)$ 只看作是 V 的函数. 显然, 目标函数 $f(U,V,S)$ 关于变量 U 和 V 的梯度分别为

$$\nabla f_{V,S}(U) = -(D-UV-S)V^{\mathrm{T}},$$
$$\nabla f_{U,S}(V) = -U^{\mathrm{T}}(D-UV-S).$$

$\mathrm{tr}(A)$ 表示矩阵 A 的迹, $\lambda_{\min}(A)$ 表示矩阵 A 的最小特征值, $\lambda_{\max}(A)$ 表示矩阵 A 的最大特征值. $|\Omega|$ 表示集合 Ω 元素的个数. A_{ij} 表示矩阵 A 的 (i,j) 元.

2 算法

本节提出一种新的算法求解低秩矩阵恢复问题 (1.4), 其基本思想是, 该算法把模型 (1.4) 的变量分成三块变量 U、V 和 S, 其中 U 和 V 是与低秩部分相关的两块变量, S 是与稀疏部分相关的块变量. 在新算法的每一步迭代过程中, 首先采用一步带有变步长的梯度法并行地更新矩阵变量 U 和 V; 然后根据新得到的矩阵变量 U 和 V, 利用非单调技术更新稀疏部分变量 S. 基于以上的讨论, 新算法的具体步骤总结如下:

算法 1 非单调交替梯度方向法

初始数据: $D\in\mathbb{R}^{m\times n}$, $U_0\in\mathbb{R}^{m\times r}$, $V_0\in\mathbb{R}^{r\times n}$, $S_0\in\mathbb{R}^{m\times n}$, 稀疏度 γ, 参数 $0<\beta\leqslant 1$, $0<\rho<1$, 整数 $l\geqslant 1$; 重复以下步骤:

1. 计算

$$\nabla f_{V_k,S_k}(U_k) = -R_kV_k^{\mathrm{T}},$$
$$\nabla f_{U_k,S_k}(V_k) = -U_k^{\mathrm{T}}R_k,$$

其中 $R_k = D-U_kV_k-S_k$;

2. 更新

$$U_{k+1} = U_k - t_k\nabla f_{V_k,S_k}(U_k),$$
$$V_{k+1} = V_k - t_k\nabla f_{U_k,S_k}(V_k),$$
$$L_{k+1} = U_{k+1}V_{k+1},$$

其中

$$t_k = \beta\frac{\langle R_k, P_k\rangle}{\|P_k\|_F^2},$$

$P_k = R_kV_k^{\mathrm{T}}V_k + U_kU_k^{\mathrm{T}}R_k$;

3. 令 $\bar{S}_k = \mathrm{argmin}_{|S|\leqslant\gamma mn}\|D-L_{k+1}-S\|_F^2$, 然后更新

$$(S_{k+1})_{ij} = \begin{cases}(\bar{S}_k)_{ij}+\tau_k, & 若\ (\bar{S}_k)_{ij}<-\tau_k,\\ (\bar{S}_k)_{ij}-\tau_k, & 若\ (\bar{S}_k)_{ij}>\tau_k,\\ 0, & 否则,\end{cases}$$

其中

$$\tau_k = \rho\sqrt{\frac{\max_{i=0}^{l}\{\|D-L_{k-i+1}-S_{k-i}\|_F^2\} - \|D-L_{k+1}-\bar{S}_k\|_F^2}{\gamma mn}}.$$

直到停机准则满足.

注 2.1 上述算法中, 步 2 定义的步长 t_k 是用来更新 U_{k+1} 和 V_{k+1} 的共同步长, 其值通过非精

确求解如下的优化问题 (求解其二次模型) 计算而得:

$$\min_t \frac{1}{2}\|D-(U_k-t\nabla f_{V_k,S_k}(U_k))(V_k-t\nabla f_{U_k,S_k}(V_k))-S_k\|_F^2.$$

算法的停机准则为

$$\frac{\|D-U_{k+1}V_{k+1}-S_{k+1}\|_F^2}{\|D\|_F^2}<\epsilon.$$

注 2.2 算法 1 中, 步 3 采用非单调搜索技术更新稀疏矩阵 S, 即

$$\|D-L_{k+1}-S_{k+1}\|_F^2\leqslant\max_{i=0}^{l}\{\|D-L_{k-i+1}-S_{k-i}\|_F^2\}. \tag{2.1}$$

证明如下:

$$\begin{aligned}
&\|D-U_{k+1}V_{k+1}-S_{k+1}\|_F^2\\
&\quad\leqslant\|D-L_{k+1}-\bar{S}_k\|_F^2+\gamma mn\tau_k^2\\
&\quad\leqslant\|D-L_{k+1}-\bar{S}_k\|_F^2\\
&\qquad+\gamma mn\frac{\max_{i=0}^{l}\{\|D-L_{k-i+1}-S_{k-i}\|_F^2\}-\|D-L_{k+1}-\bar{S}_k\|_F^2}{\gamma mn}\\
&\quad=\max_{i=0}^{l}\{\|D-L_{k-i+1}-S_{k-i}\|_F^2\}.
\end{aligned} \tag{2.2}$$

以往很多的研究工作 [15,16] 都验证了非单调搜索技术在求解优化问题时有着很好的数值效果. 因此, 在上述算法中采用非单调搜索技术从而提高算法的效率.

3 收敛性分析

本节分析新算法的收敛性. 为了方便后面的分析, 首先给出一些记号和相关的基本知识. 记 $\{U_k, V_k, S_k\}$ 是新算法产生的迭代序列, 并且

$$\begin{aligned}
R_k&=D-U_kV_k-S_k,\\
R_{k+\frac{1}{2}}&=D-U_{k+1}V_{k+1}-S_k,\\
P_k&=R_kV_k^{\mathrm{T}}V_k+U_kU_k^{\mathrm{T}}R_k.
\end{aligned} \tag{3.1}$$

引理 3.1 设 A 和 B 都是对称正定矩阵, 且 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$ 是矩阵 A 的特征值, 则有

$$\lambda_n\|B\|_F\leqslant\|AB\|_F\leqslant\lambda_1\|B\|_F.$$

证明 由于 A 是对称正定矩阵, 因此存在正交矩阵 U 使得 $U^{\mathrm{T}}AU=\Lambda=\mathrm{diag}(\lambda_i)$, 从而,

$$\begin{aligned}
\|AB\|_F&=\|U^{\mathrm{T}}AUU^{\mathrm{T}}BU\|_F\\
&=\|\Lambda U^{\mathrm{T}}BU\|_F\\
&\geqslant\lambda_n\|U^{\mathrm{T}}BU\|_F\\
&=\lambda_n\|B\|_F.
\end{aligned}$$

同理可得 $\|AB\|_F\leqslant\lambda_1\|B\|_F$. □

下面给出一些引理说明新算法每一迭代步都会使得目标函数 $f(U,V,S)$ 有所下降.

引理 3.2　令 $\sigma_{\min}^{(k)}=\min\{\lambda_{\min}(U_kU_k^{\mathrm{T}}),\ \lambda_{\min}(V_k^{\mathrm{T}}V_k)\}$, $\sigma_{\max}^{(k)}=\max\{\lambda_{\max}(U_kU_k^{\mathrm{T}}),\ \lambda_{\max}(V_k^{\mathrm{T}}V_k)\}$, 则有

$$\cos\alpha_k=\frac{\langle R_k,P_k\rangle}{\|R_k\|_F\|P_k\|_F}\geqslant\frac{\sigma_{\min}^{(k)}}{\sigma_{\max}^{(k)}}.$$

并且, 若

$$\sigma_{\min}=\min_k\{\sigma_{\min}^{(k)}\}>0,\quad \sigma_{\max}=\max_k\{\sigma_{\max}^{(k)}\}<+\infty,$$

则存在 α 使得

$$\cos\alpha_k\geqslant\cos\alpha=\frac{\sigma_{\min}}{\sigma_{\max}}.$$

证明　根据 P_k 的定义及引理 3.1 的结论可得

$$\begin{aligned}\langle R_k,P_k\rangle&=\langle R_k,R_kV_k^{\mathrm{T}}V_k+U_kU_k^{\mathrm{T}}R_k\rangle\\&=\mathrm{tr}(R_k^{\mathrm{T}}R_kV_k^{\mathrm{T}}V_k+R_k^{\mathrm{T}}U_kU_k^{\mathrm{T}}R_k)\\&=\mathrm{tr}(R_k^{\mathrm{T}}R_kV_k^{\mathrm{T}}V_k)+\mathrm{tr}(R_k^{\mathrm{T}}U_kU_k^{\mathrm{T}}R_k)\\&=\mathrm{tr}(R_k^{\mathrm{T}}R_kV_k^{\mathrm{T}}V_k)+\mathrm{tr}(U_kU_k^{\mathrm{T}}R_kR_k^{\mathrm{T}})\\&\geqslant\lambda_{\min}(V_k^{\mathrm{T}}V_k)\|R_k\|_F^2+\lambda_{\min}(U_kU_k^{\mathrm{T}})\|R_k\|_F^2\\&\geqslant 2\sigma_{\min}^{(k)}\|R_k\|_F^2.\end{aligned}$$

类似地, 有

$$\|P_k\|_F=\|R_kV_k^{\mathrm{T}}V_k+U_kU_k^{\mathrm{T}}R_k\|_F\leqslant 2\sigma_{\max}^{(k)}\|R_k\|_F.$$

从而,

$$\cos\alpha_k=\frac{\langle R_k,P_k\rangle}{\|R_k\|_F\|P_k\|_F}\geqslant\frac{\sigma_{\min}^{(k)}}{\sigma_{\max}^{(k)}}.$$

结合 $\sigma_{\min}$ 和 $\sigma_{\max}$ 的定义及条件 $\sigma_{\min}>0$ 和 $\sigma_{\max}<+\infty$, 可得

$$\cos\alpha_k\geqslant\cos\alpha=\frac{\sigma_{\min}}{\sigma_{\max}}.$$

证毕. □

引理 3.3　设引理 3.2 的条件成立, 若 $\alpha_k=\arg\langle P_k,R_k\rangle\leqslant\alpha<\frac{\pi}{2}$, 则存在常数 $0<c<1$ 使得

$$\|R_{k+\frac{1}{2}}\|_F\leqslant c\|R_k\|_F.$$

证明　根据算法 1 步 2 及 R_k 和 $R_{k+\frac{1}{2}}$ 的定义, 可得

$$\begin{aligned}\|R_{k+\frac{1}{2}}\|_F^2&=\|D-U_{k+1}V_{k+1}-S_k\|_F^2\\&=\|D-(U_k+t_kR_kV_k^{\mathrm{T}})(V_k+t_kU_k^{\mathrm{T}}R_k)-S_k\|_F^2\\&=\|R_k-t_kR_kV_k^{\mathrm{T}}V_k-t_kU_kU_k^{\mathrm{T}}R_k-t_k^2R_kV_k^{\mathrm{T}}U_k^{\mathrm{T}}R_k\|_F^2.\end{aligned}$$

由 $\beta \ll 1$ 可得 $t_k \ll 1$, 从而,

$$\|R_{k+\frac{1}{2}}\|_F^2 \approx \|R_k - t_k R_k V_k^{\mathrm{T}} V_k - t_k U_k U_k^{\mathrm{T}} R_k\|_F^2 = \|R_k - t_k P_k\|_F^2.$$

根据 $t_k = \beta \frac{\langle R_k, P_k\rangle}{\|P_k\|_F^2}$, 可得

$$\|R_{k+\frac{1}{2}}\|_F^2 \leqslant \|R_k\|_F^2 - (2\beta - \beta^2)\frac{\langle R_k, P_k\rangle^2}{\|P_k\|_F^2}.$$

结合等式 $\langle R_k, P_k\rangle^2 = \cos^2\alpha_k \|R_k\|^2 \|P_k\|^2$ 及引理的条件, 则上式可表示为如下形式:

$$\|R_{k+\frac{1}{2}}\|_F^2 \leqslant (1 - (2\beta - \beta^2)\cos^2\alpha_k)\|R_k\|_F^2 \leqslant (1 - (2\beta - \beta^2)\cos^2\alpha)\|R_k\|_F^2.$$

令 $c = \sqrt{1 - (2\beta - \beta^2)\cos^2\alpha} < 1$, 则可得

$$\|R_{k+\frac{1}{2}}\|_F \leqslant c\|R_k\|_F.$$

证毕. □

引理 3.4 假设引理 3.3 的条件都成立. 对于算法 1, τ_k 是存在的. 并且, 存在一个常数 $0 < c < 1$ 使得

$$\|R_{k+1}\|_F \leqslant c \cdot \max\{\|R_k\|_F, \ldots, \|R_{k-l}\|_F\}. \tag{3.2}$$

证明 根据算法 1 中 $\bar{S}_k$ 的定义及引理 3.3, 可得

$$\|D - L_{k+1} - \bar{S}_k\|_F \leqslant \|R_{k+\frac{1}{2}}\|_F \leqslant c\max\{\|R_k\|_F^2, \ldots, \|R_{k-l}\|_F^2\},$$

因此, τ_k 是存在的. 再根据 (2.2) 和引理 3.3, 则有

$$\begin{aligned}
\|R_{k+1}\|_F^2 &= \|D - L_{k+1} - S_{k+1}\|_F^2 \\
&\leqslant \max\{\|R_{k+\frac{1}{2}}\|_F^2, \ldots, \|R_{k-l+\frac{1}{2}}\|_F^2\} \\
&= c\max\{\|R_k\|_F^2, \ldots, \|R_{k-l}\|_F^2\}.
\end{aligned}$$

下面分析新算法产生的迭代序列的性质. 由以上的讨论可知, 非单调线搜索技术使得算法 1 产生的迭代序列满足 $\|R_{k+1}\|_F \leqslant c \cdot \max\{\|R_k\|_F, \ldots, \|R_{k-l}\|_F\}$. 因此, 新算法产生的序列 $\{L_k\}$ 和 $\{S_k\}$ 包含单调子序列. 不失一般性, 假设 $\{L_{lk+i}\}$ 和 $\{S_{lk+i}\}$ $(i = 0, 1, \ldots, l-1)$ 为单调子序列, 满足

$$\|R_{l(k+1)+i}\|_F \leqslant c\|R_{lk+i}\|_F. \tag{3.3}$$

接下来给出定理显示这样的单调子序列是 Cauchy 序列. □

定理 3.1 设引理 3.3 的假设条件都成立, 则序列 $\{L_{lk+i}\}$ 和 $\{S_{lk+i}\}$ $(i = 0, 1, \ldots, l-1)$ 是 Cauchy 序列, 即

$$\lim_{k\to\infty} L_{lk+i} = L_*, \quad i = 0, 1, \ldots, l-1,$$

$$\lim_{k\to\infty} S_{lk+i} = S_*, \quad i = 0, 1, \ldots, l-1,$$

且

$$L_* + S_* = D.$$

证明 由 (3.3) 和引理 3.3 的结论, 以及 R_k 和 $R_{k+\frac{1}{2}}$ 的定义, 有

$$\begin{aligned}\|L_{lk+i}-L_{l(k+1)+i}\|_F &= \|L_{lk+i}-L_{lk+i+1}+\cdots+L_{lk+l+i-1}-L_{lk+l+i}\|_F\\ &\leqslant \sum_{j=0}^{l-1}\|L_{lk+i+j}-L_{lk+i+j+1}\|_F=\sum_{j=0}^{l-1}\|R_{lk+i+j}-R_{lk+i+j+\frac{1}{2}}\|_F\\ &\leqslant 2\sum_{j=0}^{l-1}\|R_{lk+i+j}\|_F\leqslant 2\sum_{j=0}^{l-1}c^k\|R_{i+j}\|_F\\ &\leqslant 2lc^k\max_{0\leqslant j\leqslant l-1}\|R_j\|_F,\end{aligned}$$

从而,

$$\begin{aligned}\|L_{l(k+p)+i}-L_{lk+i}\|_F &\leqslant \sum_{j=0}^{p-1}\|L_{l(k+j)+i}-L_{l(k+1+j)+i}\|_F\\ &\leqslant 2l\max_{0\leqslant j\leqslant l-1}\|R_j\|_F\left(\sum_{j=0}^{p-1}c^{k+j}\right)\\ &\leqslant \frac{2lc^k}{1-c}\max_{0\leqslant j\leqslant l-1}\|R_j\|_F.\end{aligned}$$

由 $0<c<1$ 可得 $\{L_{lk+i}\}$ $(i=0,1,\ldots,l-1)$ 是 Cauchy 序列. 同理可得 $\{S_{lk+i}\}$ $(i=0,1,\ldots,l-1)$ 也是 Cauchy 序列. 又因为

$$\lim_{k\to\infty}R_{lk+i}=0,\quad i=0,1,\ldots,l-1,$$

所以一定存在唯一的 L_* 和 S_* 使得

$$\begin{aligned}&\lim_{k\to\infty}L_{lk+i}=L_*,\\ &\lim_{k\to\infty}S_{lk+i}=S_*,\\ &L_*+S_*=D.\end{aligned}$$

若 $\lim_{k\to\infty}\cos\alpha_k=0$, 即定理 3.1 的条件不满足, 则我们在更弱的条件下证明了算法 1 的收敛性. □

定理 3.2 如果

$$\sum_{k=0}^{\infty}\cos\alpha_k=+\infty,\quad \sum_{k=1}^{\infty}\cos\alpha_k\|R_k\|_F<+\infty,$$

且存在常数 M_1 和 M_2 使得

$$\begin{cases}\dfrac{\|R_kV_k^{\mathrm{T}}\|_F}{\|P_k\|_F}\leqslant M_1, & k=1,2,\ldots,\\ \dfrac{\|U_k^{\mathrm{T}}R_k\|_F}{\|P_k\|_F}\leqslant M_2, & k=1,2,\ldots,\end{cases}$$

则算法 1 产生的序列 $\{L_k\}$ 和 $\{S_k\}$ 收敛于问题 (1.4) 的解.

证明 根据 U_k 的定义可得

$$U_k=U_{k-1}+t_{k-1}R_{k-1}V_{k-1}^{\mathrm{T}}=\sum_{i=1}^{k-1}t_iR_iV_i^{\mathrm{T}}+U_0.$$

再根据步长 t_k 的定义及条件 $\frac{\|R_kV_k^{\mathrm{T}}\|_F}{\|P_k\|_F}\leqslant M_1$ 可推出

$$\begin{aligned}\|t_iR_iV_i^{\mathrm{T}}\|_F^2&=t_i^2\mathrm{tr}(R_iV_i^{\mathrm{T}}V_iR_i^{\mathrm{T}})\\&=\beta\frac{\langle R_i,P_i\rangle^2}{\|P_i\|_F^4}\mathrm{tr}(R_iV_i^{\mathrm{T}}V_iR_i^{\mathrm{T}})\\&=\beta\cos^2\alpha_i\frac{\|R_i\|_F^2}{\|P_i\|_F^2}\|R_iV_i^{\mathrm{T}}\|_F^2\\&\leqslant\beta M_1^2\cos^2\alpha_i\|R_i\|_F^2.\end{aligned}$$

从而有

$$\|U_k\|_F\leqslant\sum_{i=1}^{\infty}\|t_iR_iV_i^{\mathrm{T}}\|_F+\|U_0\|_F\leqslant\sum_{i=1}^{\infty}\sqrt{\beta}M_1\cos\alpha_i\|R_i\|_F+\|U_0\|_F<+\infty.$$

类似地可得

$$\|V_k\|_F<+\infty.$$

从而 $\{L_k\}$ 和 $\{S_k\}$ 是有界序列. 由条件 $\sum_{k=0}^{\infty}\cos\alpha_k=+\infty$ 和 $\sum_{k=1}^{\infty}\cos\alpha_k\|R_k\|_F<+\infty$ 可推出

$$\lim_{k\to\infty}\|R_k\|_F=0.$$

因此, 每个有界的单调递减子序列 $\{L_{lk+i}\}$ 和 $\{S_{lk+i}\}$ $(i=0,1,\dots,l-1)$ 都满足

$$\begin{aligned}&\lim_{k\to\infty}L_{lk+i}=L_*,\quad i=0,1,\dots,l-1,\\&\lim_{k\to\infty}S_{lk+i}=S_*,\quad i=0,1,\dots,l-1,\\&D=L_*+S_*,\end{aligned}$$

即 $\{L_*,S_*\}$ 是问题 (1.4) 的解. □

4 数值实验

本节通过数值实验验证新算法的有效性. 数值实验中所有的程序都用 Matlab 2011b 编写, 运行环境为 Intel Xeon E5 处理器 2.5 GHz、内存 20 GB 的台式机. 本节分为两部分. 首先, 将新算法应用于求解一些随机产生的低秩矩阵恢复问题, 并且与文献 [14] 的算法进行了比较, 程序来源于该文献作者的原程序; 其次, 将新算法应用于求解实际问题从而说明新算法的有效性.

4.1 随机低秩矩阵恢复数值实验

数值实验中, 考虑的矩阵的维数, 记为 Size(X), 分别为 1000×1000, 2000×2000, 3000×3000, 4000×4000, 5000×5000, 6000×6000, 7000×7000, 8000×8000, 9000×9000, 10000×10000. 对于每个维数的矩阵, 低秩矩阵的秩记为 $r(L)$, 分别设为 50 和 100, 稀疏矩阵的稀疏度 γ 设为 5% 和 10%. 总之, 对于每个维数的问题, 我们运行 4 个例子. 每个例子的矩阵都是随机生成的. 数值结果都记录在表

1 和 2 中, 其中 IT 表示算法的迭代次数, CPU 表示 CPU 的运行时间, rel.error 表示算法得到的矩阵与给定矩阵的相对误差, 其计算公式如下:

$$\text{rel.error} = \frac{\|D - L_k - S_k\|_F}{\|D\|_F}.$$

本实验的停机精度为 10^{-4}, $l = 2$.

表 1 和 2 中的数值结果表明了我们提出的新算法的有效性. 新算法能在合理的时间内求解 10000 $\times$10000 维的大规模问题, 而且比文献 [14] 的算法有更好的数值表现. 在这些数值算例中, 相对于文献 [14] 的算法, 新算法的运行时间短, 迭代次数少, 精度上更准确. 这也验证了非单调搜索技术在提高算法效率方面具有一定的优势.

4.2 前景背景分离数值实验

本小节将新算法用于求解实际问题. 考虑来自机场的一个视频[17] (其包含了 200 帧大小为 144×176 的照片) 和来自菜市场的一个视频[18] (其包含了 200 帧大小为 120×160 的照片). 在此实验中, 研究这两个视频中有 10% 的信息已丢失的情形. 图 1 中第一和二列分别给出了两个视频中某一张清晰图片和相应的信息丢失图片的实例. 将我们的算法用于上述视频的背景和前景分离问题, 并且计算相

表 1 小规模问题的数值结果

			文献 [14] 的算法			算法 1		
Size(X)	$r(L)$	γ (%)	rel.error (10^{-5})	CPU (s)	IT	rel.error (10^{-5})	CPU (s)	IT
1000	50	5	8.7456	7.7084	26	8.5508	3.6059	15
1000	50	10	8.5462	8.5366	31	7.3586	3.0679	17
1000	100	5	9.1451	9.6448	36	9.4904	4.7491	18
1000	100	10	8.6815	12.0113	44	8.9906	5.7453	21
2000	50	5	7.6290	13.4377	22	7.5731	10.5744	13
2000	50	10	9.4197	19.6152	26	9.0085	13.3271	14
2000	100	5	9.7533	18.2276	25	5.8623	15.7239	15
2000	100	10	7.9612	21.4509	30	8.4717	15.4689	16
3000	50	5	6.9855	26.9121	20	7.9419	22.6291	12
3000	50	10	9.6145	35.6882	24	8.2970	25.1980	13
3000	100	5	7.9613	31.8869	22	9.3447	29.2059	13
3000	100	10	9.7757	40.4796	26	9.7008	30.9464	14
4000	50	5	7.1879	37.9267	19	5.9375	33.8526	12
4000	50	10	9.1905	55.0031	23	7.1227	39.7223	13
4000	100	5	7.1998	46.8822	21	5.7289	42.3533	13
4000	100	10	9.9943	61.0373	24	5.4515	46.3085	14
5000	50	5	7.9130	55.8443	18	4.9115	49.6132	12
5000	50	10	7.6164	82.6822	23	6.0010	57.5652	13
5000	100	5	8.9144	64.4641	19	8.1014	54.7654	12
5000	100	10	7.9770	99.2382	24	7.9596	59.6463	13

表 2 大规模问题的数值结果

Size(X)	$r(L)$	γ (%)	文献 [14] 的算法 rel.error (10^{-5})	CPU (s)	IT	算法 1 rel.error (10^{-5})	CPU (s)	IT
6000	50	5	7.4499	77.6586	18	9.5999	62.9023	11
6000	50	10	9.1130	112.3149	22	9.3500	74.9630	12
6000	100	5	7.1527	88.4599	19	6.4058	70.4291	12
6000	100	10	8.7420	124.8988	23	6.1849	80.5617	13
7000	50	5	9.2201	99.8282	17	8.5527	82.4295	11
7000	50	10	7.4785	151.2906	22	8.5910	90.2424	12
7000	100	5	9.0895	115.1263	18	5.3850	94.7013	12
7000	100	10	9.8199	165.1208	22	5.4377	105.6243	13
8000	50	5	7.8367	127.6203	17	7.6768	102.6150	11
8000	50	10	9.5746	186.1850	21	7.6291	116.2837	12
8000	100	5	7.4998	142.1498	18	4.4406	119.0636	12
8000	100	10	8.4006	214.2220	22	9.8263	124.3116	12
9000	50	5	7.3708	151.9489	17	7.0939	132.5552	11
9000	50	10	9.2168	231.7159	21	7.4972	144.2785	12
9000	100	5	9.7485	169.0595	17	9.3973	137.1404	11
9000	100	10	9.4503	249.0968	21	8.5033	159.3495	12
10000	50	5	6.8840	193.0562	17	6.6454	158.4349	11
10000	50	10	8.4873	288.2578	21	7.1931	181.9145	12
10000	100	5	9.4217	210.7085	17	9.1307	166.3343	11
10000	100	10	9.3845	300.2361	21	8.1152	190.9680	12

图 1 第一行的四幅图分别表示菜市场视频中的第 42 帧原始图片、信息丢失的图片、背景图和前景图; 第二行的四幅图分别表示机场视频中的第 10 帧原始图片、信息丢失的图片、背景图和前景图

应的 F- 测度, F- 测度的计算公式详见文献 [18]. 数值结果如图 1 的第三列的背景图和第四列的前景图. 针对菜市场和机场的视频, 新算法的 F- 测度分别为 0.6088 和 0.6070. 这些数值结果显示了我们的算法在解决实际问题中的可行性和有效性.

5 总结

低秩矩阵恢复问题就是从一个已知矩阵恢复出一个低秩矩阵和一个稀疏矩阵. 这个模型可以用来求解信号与图像处理、统计和机器学习等很多领域中的实际问题. 本文针对低秩矩阵恢复问题提出了一种新的带有非单调搜索技术的交替方向法. 该算法的主要思想是, 每一步迭代采用一步梯度算法并行更新低秩部分的两块变量, 然后利用非单调技术更新稀疏部分的变量. 在一定的假设条件下, 我们证明了新算法的全局收敛性. 最后给出的数值实验结果也进一步表明了新算法的有效性, 同时也显示了非单调技术极大地改善了算法的效率.

参考文献

1 Huang P S, Chen S D, Smaragdis P, et al. Singingvoice separation from monaural recordings using robust principal component analysis. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing. Kyoto: IEEE, 2012, 57–60

2 Li L Y, Huang W M, Gu I Y H, et al. Statistical modeling of complex backgrounds for foreground object detection. IEEE Trans Image Process, 2004, 13: 1459–1472

3 Mobahi H, Zhou Z H, Yang A Y, et al. Holistic 3D reconstruction of urban structures from low-rank textures. In: 2011 IEEE International Conference on Computer Vision Workshops. Barcelona: IEEE, 2011, 593–600

4 Tharrault Y, Mourot G, Ragot J, et al. Fault detection and isolation with robust principal component analysis. Int J Appl Math Comput Sci, 2008, 18: 429–442

5 Cui N B. Medical image registration using low-rank matrix recovery (in Chinese). Master Thesis. Beijing: Beijing Jiaotong University, 2015 [崔宁波. 低秩矩阵恢复在医学图像配准中的应用. 硕士学位论文. 北京: 北京交通大学, 2015]

6 Candès E J, Li X D, Ma Y, et al. Robust principal component analysis? J ACM, 2011, 58: 1–37

7 Lin Z, Chen M, Ma Y. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv:1009.5055, 2010

8 Wright J, Ganesh A, Rao S, et al. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In: Proceedings of the 22nd International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2009, 2080–2088

9 Wang X, Hong M, Ma S, et al. Solving multiple-block separable convex minimization problems using two-block alternating direction method of multipliers. Pac J Optim, 2015, 11: 645–667

10 Zhou T, Tao D. GoDec: Randomized lowrank & sparse matrix decomposition in noisy case. In: Proceedings of the 28th International Conference on Machine Learning. Washington: Bellevue, 2011, 33–40

11 Netrapalli P, Niranjan U N, Sanghavi S, et al. Non-convex robust PCA. arXiv:1410.7660, 2014

12 Cai H Q, Cai J F, Wei K. Accelerated alternating projections for robust principal component analysis. J Mach Learn Res, 2019, 20: 1–33

13 Gu Q Q, Wang Z R, Liu H. Low-rank and sparse structure pursuit via alternating minimization. In: Proceedings of the 19th International Conference on Artificial Intelligence and Statistics. Cadiz: PMLR, 2016, 600–609

14 Yi X Y, Park D Y, Chen Y D, et al. Fast algorithms for robust PCA via gradient descent. In: Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems. Barcelona: NIPS, 2016, 4152–4160

15 Grippo L, Lampariello F, Lucidi S. A nonmonotone line search technique for Newton's method. SIAM J Numer Anal, 1986, 23: 707–716

16 Toint P L. An assessment of nonmonotone linesearch techniques for unconstrained optimization. SIAM J Sci Comput, 1996, 17: 725–739

17 He H, Han D. A distributed Douglas-Rachford splitting method for multi-block convex minimization problems. Adv Comput Math, 2016, 42: 27–53

18 Yang L, Pong T K, Chen X. Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction. SIAM J Imag Sci, 2017, 10: 74–110

A non-monotone alternating directional method for matrix recovery problem

Xihong Yan, Chuanlong Wang, Chao Li & Jingting Xue

Abstract The low-rank matrix recovery problem, where the concerned matrix is separable into a low-rank part and a sparse part, has been frequently exploited in the fields of image processing and signal data analysis. In this paper, we propose an alternating directional method equipped with the non-monotone search strategy for solving the low-rank matrix recovery problem, where we apply a single step of the steepest gradient descent method to update variables associated with the low-rank part in parallel and the non-monotone search strategy to update the sparse structure matrix. Theoretically, we prove the global convergence of the proposed algorithm under some mild conditions. The efficiency and effectiveness of the proposed algorithm and superiority of the non-monotone search strategy for improving the performance of algorithms are demonstrated by solving some instances of random matrix recovery problems and background/foreground extraction problems.

Keywords low-rank matrix recovery, alternating directional method, non-monotone search

MSC(2020) 15A18, 90C20, 90C25

doi: 10.1360/SSM-2019-0183