DOI：10.13718/j.cnki.xdzk.2022.05.022

基于加权核范数的低秩矩阵补全算法研究

石莹[1]，　黄华[2]，　王智[3]，　高超[4]

1. 西南大学 信息化建设办公室，重庆 400715；2. 重庆工程职业技术学院，现代教育技术中心，重庆 402260；
3. 西南大学 计算机与信息科学学院，重庆 400715；4. 西北工业大学 光电与智能研究院，西安 710072

摘要：利用加权核范数去松弛原始低秩极小化问题，基于 Soft-Impute 算法思想提出 WNNM-Impute 算法. 通过引入不精确的近邻算子极大地降低 WNNM-Impute 算法的时间复杂度，从而使得算法收敛更快. 同时，在算法中引入 Nesterov 加速策略，使得算法的总体迭代次数进一步减少. 大量的实验结果表明，所提算法能得到更精确的解且拥有比 Soft-Impute 和大多数对比算法更快的收敛速率.

关　键　词：低秩矩阵补全；Soft-Impute 算法；Nesterov 优化理论

中图分类号：TN911.72　　**文献标志码**：A

文 章 编 号：1673－9868(2022)05－0192－11　　开放科学(资源服务)标识码(OSID)：

Low-Rank Matrix Completion with Weighted Nuclear Norm Regularizer

SHI Ying[1]，　HUANG Hua[2]，　WANG Zhi[3]，　GAO Chao[4]

1. *Office of Information Technology*，*Southwest University*，*Chongqing 400715*，*China*；
2. *Modern Education Technology Center*，*Chongqing Vocational Institute of Engineering*，*Chongqing 402260*，*China*；
3. *College of Computer and Information Science*，*Southwest University*，*Chongqing 400715*，*China*；
4. *School of Artificial Intelligence*，*Northwestern Polytechnical University*，*Xi'an 710072*，*China*

Abstract：Low rank matrix completion is a most basic problem in machine learning and data analysis. It plays a key role in solving many important problems，such as collaborative filtering，dimensionality reduction，multi-task learning and pattern recognition. In this work，we employ the weighted nuclear norm to relax the rank function. Inspired by the idea of Soft-Impute，we proposed WNNM-Impute algorithm by using inexact proximal operator and Nesterov's rule. Experiments delivered very encouraging results in terms of the quality of achieved solution and the processing time required.

Key words：low rank matrix completion；Soft-Impute；Nesterov's rule

收稿日期：2021－10－29
基金项目：国家自然科学基金项目(61976181).
作者简介：石莹，高级工程师，主要从事机器学习、网络优化等研究.
通信作者：黄华，硕士，高级实验师.

在诸如协同过滤[1-2]、降维处理[3]、子空间聚类[4]、图像和信号处理[5-6]等机器学习或数据分析领域，人们通常需要求解如下优化问题

$$\min_{\boldsymbol{X}\in\mathbb{R}^{m\times n}} \operatorname{rank}(\boldsymbol{X}) \text{ s. t. } \mathcal{P}_{\Omega}(\boldsymbol{X})=\mathcal{P}_{\Omega}(\boldsymbol{M}) \tag{1}$$

其中：$\boldsymbol{M}\in\mathbb{R}^{m\times n}$ 为仅知道部分观察元素的待恢复矩阵，$\boldsymbol{X}\in\mathbb{R}^{m\times n}$ 为目标矩阵，Ω 表示已知的观察元素的位置，$\mathcal{P}_{\Omega}$ 为正交投影算子，定义为

$$[\mathcal{P}_{\Omega}(\boldsymbol{M})]_{ij}=\begin{cases}\boldsymbol{M}_{ij} & (i,j)\in\Omega\\ 0 & (i,j)\notin\Omega\end{cases} \tag{2}$$

求解优化问题(1)的主要任务是寻找一个最小秩矩阵 $\boldsymbol{X}^{*}\in\mathbb{R}^{m\times n}$，使其满足 $\mathcal{P}_{\Omega}(\boldsymbol{X}^{*})=\mathcal{P}_{\Omega}(\boldsymbol{M})$ 且使得秩函数 $\operatorname{rank}(\boldsymbol{X})$ 最小. 利用正则化方法，问题(1)可等价地转换为如下非约束最小化问题

$$\min_{\boldsymbol{X}\in\mathbb{R}^{m\times n}} \frac{1}{2}\|\mathcal{P}_{\Omega}(\boldsymbol{X}-\boldsymbol{M})\|_{F}^{2}+\lambda\operatorname{rank}(\boldsymbol{X}) \tag{3}$$

这里 $\|\cdot\|_{F}$ 表示 F 范数，$\lambda>0$ 为正则化参数.

然而，由于秩函数 $\operatorname{rank}(\boldsymbol{X})$ 具有非凸和不连续的特性，优化问题(1)和(3)是 NP 难的[7]，无法在多项式时间内求得有效解. 为了解决这一难题，一个广泛使用的策略是将秩函数 $\operatorname{rank}(\boldsymbol{X})$ 极小化问题松弛为核范数极小化问题. 当满足一定的条件[8]，如 $|\Omega|\geqslant O(N^{1.2}r\log N)(N=\max(m,n),r=\operatorname{rank}(\boldsymbol{X}))$ 时，通过求解核范数极小化问题可以高概率近乎完美地恢复原始矩阵 $\boldsymbol{M}$ 中缺失的元素. 一般来说，核范数正则化问题可以被看作为半定规划问题(semi-definite program，SDP)[9]. 但是，这种策略仅对 $m,n\leqslant 100$ 的矩阵有效. 为了有效地求解这一问题，大量的一阶梯度算法被提出. 这些方法包括：SVT (singular value thresho-lding) 算法[10] 和 APGL (accelerated proximal gradient with linesearch) 算法[11]. SVT 和 APGL 算法在通常情况下能够得到较为满意的解且有严格的理论保证，但是它们在每次迭代过程中需要进行费时的 SVD (singular value decomposition) 分解，这限制了它们在大规模矩阵上的应用. 为了缓解这一缺陷，FPCA (fixed point continuation with approximate) 算法[12] 被提出求解和 APGL 算法相同的优化问题. 由于在 FPCA 算法中引入了快速蒙特卡罗算法来近似求解 SVD，它的效率得到了极大的提升. 另一个被广泛关注的算法是 Soft-Impute 算法[13]，该方法利用求解过程存在“低秩 + 稀疏”的特性，同时结合 SVT 算法，达到在每次迭代中快速进行 SVD 分解的目的. 最近，通过引入 Nesterov 优化理论，使得 Soft-Impute 算法的计算时间复杂度由 $O\left(\frac{1}{T}\right)$ 提升为 $O\left(\frac{1}{T^{2}}\right)$，其中 T 为迭代次数[14].

核范数正则化方法作为一个强大的工具已广泛地应用于求解低秩矩阵补全问题，然而正如文献[15]所述，该方法同等对待目标矩阵中的所有奇异值，从而导致过度惩罚较大奇异值. 换句话说，核范数正则化方法在大多数时得到的解严重偏离真实解. 为了使得较大奇异值得到更少的惩罚，大量的非凸函数被提出用于替代秩函数 $\operatorname{rank}(\boldsymbol{X})$，例如 Capped-$l_1$，LSP(log-sum penalty)，TNN(truncated nuclear norm)，SCAD(smoothly clipped absolute deviation) 和 MCP(minimax concave penalty). 在文献[16]中，基于 Capped-l_1 的目标优化函数利用多步的凸松弛策略得以求解；在文献[17]中，理论和实验表明，基于 TNN 的低秩矩阵补全模型能更佳地逼近目标矩阵的秩函数；在文献[18]中，一个快速、连续和精确的算法被提出用于求解高维线性回归问题. 另一类在低秩矩阵补全领域得到广泛关注的策略是，利用 Schatten-q ($0<q\leqslant 1$) 拟范数来逼近目标矩阵的秩函数 $\operatorname{rank}(\boldsymbol{X})$. 然而，基于该策略得到的非凸 l_q 正则化优化问题较难被求解. 为了解决这一问题，l_qPG(l_q-Proximal-Gradient)算法被提出用于求解该非凸 l_q 正则化优化问题. 在 l_qPG 算法的每次迭代过程中包含不精确的近似求解和费时的 SVD，这严重制约了算法的应用范围. 基于此，文献[19]和文献[20]分别提出利用 Schatten-$\frac{1}{2}$ 和 Schatten-$\frac{2}{3}$ 拟范数来逼近目标矩阵的秩函数

rank($\boldsymbol{X}$). 同时，提出基于半阈值算子和$\frac{2}{3}$阈值算子的不动点算法. 对比 l_qPG 算法，基于 Schatten-$\frac{1}{2}$ 和 Schatten-$\frac{2}{3}$ 拟范数的阈值函数具有闭式解，因此得到的解更精确. 所有实验结果表明，相比核范数正则化方法，这些非凸正则化方法性能更佳.

基于非凸正则化方法的优良性能，本文对基于加权核范数正则化模型进行进一步的分析和探讨，并借鉴 Soft-Impute 算法的思想提出一个拥有更快收敛速率和能够得到更高精度解的低秩矩阵补全算法. 在本文提出 WNNM-Impute(weighted nuclear norm minimization impute) 算法的迭代过程中，需要进行费时的 SVD 分解. 针对这一问题，通过引入不精确的近邻算子极大地降低 WNNM-Impute 算法的时间复杂度，从而使得算法收敛更快. 同时，在算法中引入 Nesterov 加速策略，使得算法的总体迭代次数进一步减少. 本文提出的算法基于 Soft-Impute 算法，但是实验结果表明，它比 Soft-Impute 算法收敛更快且得到的解的精度更高. 因此，本文提出的算法适合用于求解大规模低秩矩阵补全问题.

1 基础知识

当使用核范数去松弛秩函数 rank($\boldsymbol{X}$) 时，数学上可将低秩矩阵补全问题建模为

$$\min_{\boldsymbol{X}\in\mathbb{R}^{m\times n}} \frac{1}{2}\| \mathcal{P}_{\Omega}(\boldsymbol{X}-\boldsymbol{M})\|_F^2+\lambda\|X\|_* \tag{4}$$

这里 $\|\boldsymbol{X}\|_*$ 表示核范数，定义为

$$\|\boldsymbol{X}\|_*=\sum_{i=1}^{r}|\sigma_i(\boldsymbol{X})| \tag{5}$$

其中 $r=\min(m,n)$，$\sigma_i(\boldsymbol{X})$ 为矩阵 $\boldsymbol{X}$ 的第 i 个奇异值.

模型(4) 能够被 SVT 等算法快速求解. 在 SVT 算法中，需要求解如下奇异值阈值算子

$$\mathcal{S}_{\lambda}(\boldsymbol{Y})=\min_{\boldsymbol{X}\in\mathbb{R}^{m\times n}} \frac{1}{2}\|\boldsymbol{X}-\boldsymbol{Y}\|_F^2+\lambda\|\boldsymbol{X}\|_* \tag{6}$$

该算子的求解需要如下引理.

引理 1 若记矩阵 $\boldsymbol{Y}$ 的 SVD 分解为 $\boldsymbol{Y}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$，则(6) 式的最优解为

$$\boldsymbol{X}^*=\mathcal{S}_{\lambda}(\boldsymbol{Y})=\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}} \tag{7}$$

其中 $\mathcal{S}_{\lambda}(\cdot)$ 为软阈值函数，$D_{ii}=\max(\Sigma_{ii}-\lambda,0)$.

从上述引理可以看出，主要的计算集中在求解矩阵 $\boldsymbol{Y}$ 的 SVD 分解，其时间复杂度为 $O(mn^2)$，因此它不适合求解大规模矩阵补全问题. 针对这一问题，文献[13] 注意到

$$\mathcal{P}_{\Omega}(\boldsymbol{M})+\mathcal{P}_{\Omega^{\perp}}(\boldsymbol{X}_k)=\mathcal{P}_{\Omega}(\boldsymbol{M}-\boldsymbol{X}_k)+\boldsymbol{X}_k \tag{8}$$

其中：$\boldsymbol{X}_k$ 为第 k 次迭代值，$\mathcal{P}_{\Omega^{\perp}}(\cdot)$ 表示 Ω 的补集中的元素. 显然，(8) 式右边的第一部分具有稀疏的特性，第二部分具有低秩的特性. 因此，利用 SVT 算法的思想提出了 Soft-Impute 算法，在第 $k+1$ 次迭代时有

$$\boldsymbol{X}_{k+1}=\mathcal{S}_{\lambda}(\mathcal{P}_{\Omega}(\boldsymbol{M})+\mathcal{P}_{\Omega^{\perp}}(\boldsymbol{X}_k)) \tag{9}$$

该算法的详细步骤如算法 1 所示.

算法 1 Soft-Impute 算法[13]

输入：部分观察矩阵 $\boldsymbol{M}$，参数 $\lambda>0$

输出：秩为 r 的矩阵 $\boldsymbol{X}$

1) 初始化 $\boldsymbol{X}_1=0$；

2) for $k=1,2,\cdots$ do

3)　$\boldsymbol{Y}=\mathcal{P}_{\Omega}(\boldsymbol{M})+\mathcal{P}_{\Omega^{\perp}}(\boldsymbol{X}_k)$

4)　$\boldsymbol{X}_{k+1}=\mathcal{S}_{\lambda}(\boldsymbol{Y})$

5) end for

6) return $\boldsymbol{X}_{k+1}$

然而，在核范数极小化模型中，它同等对待目标矩阵的所有奇异值，这导致该模型不能很好地刻画矩阵的低秩结构. 为了解决这一困难，近年来加权核范数[21] 被提出来用于更好地刻画目标矩阵的低秩结构. 对目标低秩矩阵，加权核范数模型尽可能地对较大奇异值进行较小惩罚，而对较小奇异值进行较大惩罚. 对于任一矩阵 $\boldsymbol{X}\in\mathbb{R}^{m\times n}$，加权核范数可定义为

$$g(\boldsymbol{X})=\|\boldsymbol{X}\|_{w,*}=\sum_{i=1}^{r}w_i\sigma_i(\boldsymbol{X}) \tag{10}$$

其中 $\boldsymbol{w}=[w_1,w_2,\cdots,w_r]^{\mathrm{T}}$ 和 $\boldsymbol{w}_i\geqslant 0$ 为分配给 $\sigma_i(\boldsymbol{X})$ 的非负加权. 核范数极小化方法为了刻画矩阵的低秩结构同等对待目标矩阵的奇异值，而(10) 式考虑对较小奇异值进行更多惩罚，而较大的奇异值得到更少惩罚. 基于此，利用(10) 式去松弛秩函数 rank($\boldsymbol{X}$)，将原始低秩矩阵补全问题建模为

$$\min_{\boldsymbol{X}\in\mathbb{R}^{m\times n}}\frac{1}{2}\|\mathcal{P}_{\Omega}(\boldsymbol{X}-\boldsymbol{M})\|_F^2+\lambda\|\boldsymbol{X}\|_{w,*} \tag{11}$$

模型(11) 比模型(4) 能够更好地刻画目标矩阵的低秩结构，但是由于加权核范数的引入，得到非凸、非光滑的优化问题，使得对该问题的求解变得非常困难. 传统的针对核范数正则化模型的求解方法不能直接用于求解非凸正则化模型. 因此，如何设计出快速、稳健、可扩展和可靠的算法是一个至关重要的挑战. 本文针对这一问题，利用 Soft-Impute 的求解思想，融入不精准近邻算法和 Nesterov 优化理论，提出一个快速高效的低秩矩阵补全算法用于求解模型(11)，理论和实验结果都表明所提出的算法能够得到更快的收敛速率和更高精度的解.

2　WNNM-Impute 算法

受到 Soft-Impute 算法的启发，本节将融入不精准近邻算法和 Nesterov 优化理论，构建用于求解优化模型(11) 的一阶梯度算法. 然而，由于加权核范数的引入，优化模型(11) 是非凸和非光滑的. 因此，为了得到该模型的闭式解，需要求解如下加权核范数近邻算子.

$$\min_{\boldsymbol{X}\in\mathbb{R}^{m\times n}}\|\boldsymbol{X}-\boldsymbol{M}\|_F^2+\|\boldsymbol{X}\|_{w,*} \tag{12}$$

下面的定理表明，加权核范数近邻问题(12) 可转化为线性约束的二次规划问题，从而得到它的闭式解.

定理 1　对任意给定矩阵 $\boldsymbol{M}\in\mathbb{R}^{m\times n}$ 且其 SVD 分解为 $\boldsymbol{M}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$，其中 $\boldsymbol{\Sigma}=\begin{pmatrix}\mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)\\0\end{pmatrix}\in\mathbb{R}^{m\times n}$. 则加权核范数近邻问题(12) 的全局最优解为

$$\boldsymbol{X}^{*}=\boldsymbol{U}\hat{\boldsymbol{D}}\boldsymbol{V}^{\mathrm{T}} \tag{13}$$

其中 $\hat{\boldsymbol{D}}=\begin{pmatrix}\mathrm{diag}(d_1,d_2,\cdots,d_r)\\0\end{pmatrix}$，$(d_1,d_2,\cdots,d_r)$ 是如下凸优化问题的最优解

$$\min_{d_1,d_2,\cdots,d_r}\sum_{i=1}^{r}(\sigma_i-d_i)^2+w_id_i \tag{14}$$

$$\text{s.t. } d_1\geqslant d_2\geqslant\cdots\geqslant d_r\geqslant 0$$

该定理的证明见文献[21]. 基于定理 1，如取 $w_1\leqslant w_2\leqslant\cdots\leqslant w_r$，可以得到如下推论.

推论 1　对任意给定 $\lambda>0$，$\boldsymbol{Y}\in\mathbb{R}^{m\times n}$ 且其 SVD 分解为 $\boldsymbol{Y}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$，$w_1\leqslant w_2\leqslant\cdots\leqslant w_r$，$r$ 为矩阵 $\boldsymbol{Y}$ 的秩，则极小化问题

$$\min_{X\in\mathbb{R}^{m\times n}} \frac{1}{2}\|\boldsymbol{X}-\boldsymbol{Y}\|_F^2+\lambda\sum_{i=1}^{r}w_i\sigma_i(\boldsymbol{X}) \tag{15}$$

的最优解为

$$\boldsymbol{X}^*=\mathcal{S}_\lambda(\boldsymbol{Y})=\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}} \tag{16}$$

其中 $D_{ii}=\max(\Sigma_{ii}-\lambda w_i,0)$.

该推论的证明见文献[21]. 该推论表明，虽然极小化问题(15)是非凸和非光滑的，但是它仍然存在全局最优闭式解. 利用定理1和推论1的结论，我们提出如下算法用于求解优化模型(11).

算法2 WNNM-Impute算法

输入：部分观察矩阵 $\boldsymbol{M}$，参数 $\lambda>0$；给定 $\mu\in\left(0,\frac{1}{L}\right)$，$\bar{\lambda}=10^{-6}$，$\delta>0$ 和 $\eta\in(0,1)$

输出：秩为 r 的矩阵 $\hat{\boldsymbol{X}}$

1) 初始化 $\boldsymbol{V}_1,\boldsymbol{V}_2\in\mathbb{R}^{m\times n}$ 为随机的高斯矩阵，$\boldsymbol{X}_0=\boldsymbol{X}_1=0$，$\boldsymbol{Y}_1=\boldsymbol{X}_0$，$\lambda_0=\eta\|\mathcal{P}_\Omega(\boldsymbol{M})\|_2$，$\theta_0=1$；

2) for $k=1,2,\cdots,T$ do

3) $\quad\boldsymbol{R}_k=QR([\boldsymbol{V}_k,\boldsymbol{V}_{k-1}])$//$QR()$ 表示 QR 分解

4) $\quad\boldsymbol{G}=\boldsymbol{Y}_k-\mu(\mathcal{P}_\Omega(\boldsymbol{Y}_k-\boldsymbol{M}))$

5) $\quad\boldsymbol{Q}=\mathrm{PowerMethod}(\boldsymbol{G},\boldsymbol{R}_k)$

6) $\quad[\boldsymbol{U},\boldsymbol{\Sigma},\boldsymbol{V}]=SVD(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{G})$

7) $\quad\boldsymbol{U}=\{u_k\mid\Sigma_{ii}>\lambda w_i\}$

8) $\quad\boldsymbol{V}=\{v_k\mid\Sigma_{ii}>\lambda w_i\}$

9) $\quad\boldsymbol{D}=\{d_{ii}\mid\max(\Sigma_{ii}-\lambda_{k-1}w_i,0)\}$

10) $\quad\hat{\boldsymbol{Z}}=\boldsymbol{Q}\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}}$

11) $\quad$if $F(\hat{\boldsymbol{Z}})\leqslant F(\boldsymbol{X}_k)-\frac{\delta}{2}\|\hat{\boldsymbol{Z}}-\boldsymbol{X}_k\|_F^2$

12) $\quad\quad\boldsymbol{X}_{k+1}=\hat{\boldsymbol{Z}}$

13) $\quad$else

14) $\quad\quad\boldsymbol{X}_{k+1}=\boldsymbol{Y}_k$

15) $\quad$end if

16) $\quad\theta_k=\frac{1+\sqrt{1+4\theta_{k-1}^2}}{2}$

17) $\quad\boldsymbol{Y}_{k+1}=\boldsymbol{X}_{k+1}+\left(\frac{\theta_{k-1}}{\theta_k}\right)(\hat{\boldsymbol{Z}}-\boldsymbol{X}_{k+1})+\left(\frac{\theta_{k-1}-1}{\theta_k}\right)(\boldsymbol{X}_{k+1}-\boldsymbol{X}_k)$

18) $\quad\lambda_k=\{\bar{\lambda},\eta\lambda_{k-1}\}$

19) $\quad$if $\lambda_k==\bar{\lambda}$

20) $\quad\quad$break

21) $\quad$end if

22) $\quad\boldsymbol{V}_{k+1}=\boldsymbol{V}$

23) end for

24) return $\boldsymbol{X}_{k+1}$

从定理2和推论1可以看出，在对极小化模型(11)的求解过程中，需要对观察矩阵 $\boldsymbol{Y}\in\mathbb{R}^{m\times n}$ 做SVD分解，它的时间复杂度为 $O(mn^2)$. 因此，当矩阵的规模较大时，这是相当费时的. 为此，在算法2中，我们引

入 PowerMethod 算法得到一个正交矩阵 $\boldsymbol{Q}\in\mathbb{R}^{m\times t}$，其中 $t>r$ 且 $t\ll n$. 这样，对矩阵 $\boldsymbol{G}$ 的 SVD 分解可以转为对矩阵 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{G}$ 的 SVD 分解. 但是，$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{G}$ 的规模仅为 $m\times t$，远小于矩阵 $\boldsymbol{G}$ 的规模 $m\times n$. 从而，时间复杂度降为 $O(mt^2)$. 该过程可概述为如下引理.

引理 2　假定矩阵 $\boldsymbol{G}\in\mathbb{R}^{m\times n}$ 有 r 个奇异值大于 λw，$\boldsymbol{G}=\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{G}$，其中 $\boldsymbol{Q}\in\mathbb{R}^{m\times t}$ 为正交矩阵且 $t>r$，则

1) $\mathrm{range}(\boldsymbol{G})\subseteq\mathrm{range}(\boldsymbol{Q})$；

2) $\mathcal{S}_{\lambda}(\boldsymbol{G})=\boldsymbol{Q}\mathcal{S}_{\lambda}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{G})$.

该引理的证明过程与文献[22]中的引理1类似，这里省略. 需要指出的是，引理2是处理的加权核范数正则子，而非核范数正则子.

由于 PowerMethod 算法[22] 的引入，得到阈值算子的值将是不精确的. 因此，需要判断当前值是否可行. 于是，在算法2中，我们采用如下式子来确保目标函数 F 是单调递减的，从而保证目标函数收敛到某一稳定点：

$$F(\overset{\vee}{\boldsymbol{Z}})\leqslant F(\boldsymbol{X}_k)-\frac{\delta}{2}\|\overset{\vee}{\boldsymbol{Z}}-\boldsymbol{X}_k\|_F^2 \tag{17}$$

为了提升算法的收敛速率，在算法 2 中引入 Nesterov 加速准则，该准则被广泛地应用于机器学习算法中，如 nmAPG，FaNCL-acc，niAPG 等. 基于此，在算法 2 中采用如下加速策略

$$\theta_k=\frac{1+\sqrt{1+4\theta_{k-1}^2}}{2} \tag{18}$$

$$\boldsymbol{Y}_{k+1}=\boldsymbol{X}_{k+1}+\left(\frac{\theta_{k-1}}{\theta_k}\right)(\overset{\vee}{\boldsymbol{Z}}-\boldsymbol{X}_{k+1})+\left(\frac{\theta_{k-1}-1}{\theta_k}\right)(\boldsymbol{X}_{k+1}-\boldsymbol{X}_k) \tag{19}$$

实验部分将进一步表明，由于(18) 和(19) 式加速策略的引入，使得算法 2 的收敛速率得到大幅地提高.

众所周知，正则化参数 λ 在算法中扮演非常重要的角色，因此如何设定该参数将直接影响到算法的准确性. 幸运的是，这一问题在文献[12,20] 中得到很好地研究. 在算法 2 中，引入连续化技术，该技术被广泛地应用在众多一阶优化算法中. 连续化技术的核心思想是引入一个单调递减的序列 λ_k：$\lambda_1>\lambda_2>\cdots>\overline{\lambda}>0$，这里 $\overline{\lambda}$ 为任意小的正实数. 在第 k 次迭代中，令 $\lambda=\lambda_k$. 基于该思想，在算法 2 中，采用如下策略

$$\lambda_k=\{\overline{\lambda},\ \eta\lambda_{k-1}\} \tag{20}$$

这里 $0<\eta<1$.

在算法 2 的迭代过程中，得到的解 $\boldsymbol{X}_k$ 不断地逼近真实解，且目标函数 F 单调递减，直到收敛到某一稳定点. 下面的定理保证了算法 2 的收敛性，该定理的证明参照文献[20].

定理 2　假定 $\{\boldsymbol{X}_k\}$ 为算法 2 在迭代过程中产生的解序列，则

1) $\{\boldsymbol{X}_k\}$ 有界且至少存在一个聚点，即 $\lim\limits_{k\to\infty}\boldsymbol{X}_k=\boldsymbol{X}^*$；

2) 目标函数 F 单调递减，且收敛到 $F(\boldsymbol{X}^*)$，这里 $\boldsymbol{X}^*$ 为序列 $\{\boldsymbol{X}_k\}$ 的任一聚点；

3) $\{\boldsymbol{X}_k\}$ 渐进正则，即 $\lim\limits_{k\to\infty}\|\boldsymbol{X}_{k+1}-\boldsymbol{X}_k\|_F^2=0$.

3　实验仿真

为了测试 WNNM-Impute 算法的效率和准确性，本节将在模拟数据集和真实数据集上进行一系列的对比实验. 在接下来的实验中，将包括如下经典或最新算法.

1) SVT 算法：经典的、被广泛应用于低秩矩阵补全问题中的方法，它将线性 Bregman 迭代用于求解低秩矩阵优化问题；

2) APG 算法：基于核范数的低秩矩阵补全方法，是快速迭代收缩阈值算法的矩阵版本；

3) R1MP算法：该方法是将正交匹配追踪算法用于求解低秩矩阵补全问题；

4) Soft-Impute算法：该方法是在SVT方法的基础上，利用迭代过程存在“低秩+稀疏”的特性，达到快速求解低秩补全问题的目的；

5) WNNM算法：基于加权核范数的低秩矩阵补全方法.

在接下来的实验中，各个算法的参数设置为对应文章提供的推荐设置. WNNM-Impute算法的参数设置为：$\bar{\lambda}=10^{-6}$，$\mu=1.6$，$\eta=0.75$. 此外，当前后两次迭代得到的解没有显著变化，即 $\|\boldsymbol{X}_{k+1}-\boldsymbol{X}_k\|_F^2<10^{-5}$ 时算法停止. 所有算法采用MATLAB R2014编程，实验环境为Intel Xeon E5-2680-v4处理器(3核，2.4GHz)，256GB内存，Windows server 2008操作系统.

本文的第一组实验为在模拟数据集上的低秩补全. 首先，随机生成矩阵 $\boldsymbol{M}_L\in\mathbb{R}^{m\times r}$ 和 $\boldsymbol{M}_R\in\mathbb{R}^{r\times n}$，令 $\boldsymbol{M}=\boldsymbol{M}_L\boldsymbol{M}_R+\boldsymbol{N}$，其中矩阵 $\boldsymbol{M}_L$，$\boldsymbol{M}_R$ 和 $\boldsymbol{N}$ 中的每个元素为独立高斯随机变量，$\boldsymbol{N}$ 为方差为0.1的高斯白噪音矩阵. 不失一般性，这里令 $m=n$ 和 $r=5$. 随后，从观测矩阵 $\boldsymbol{M}$ 中随机均匀采样得到若干观测值，使用 $SR=\dfrac{10m\log m}{m^2}$ 表示采样率. 本文采用如下方式来评估所有算法的性能：

1) 计算

$$NMSE=\frac{\|\mathcal{P}_{\Omega^\perp}(\boldsymbol{X}-\boldsymbol{UV})\|_F}{\|\mathcal{P}_{\Omega^\perp}(\boldsymbol{UV})\|_F}$$

其中：$\boldsymbol{X}$ 为结果矩阵，$\Omega^\perp$ 为除观察值以外的位置；

2) 结果矩阵 $\boldsymbol{X}$ 的秩 r；

3) 运行时间 t.

在实验中，我们对不同规模的矩阵进行了测试，即 $m\in\{500,1\,000,1\,500,3\,000\}$. 每个算法独立运行10次，报告其平均 $NMSE$ 值.

表1 不同算法在不同规模和不同采样率下的矩阵补全结果

	$m=500$			$m=1\,500$			$m=2\,000$			$m=3\,000$		
	$NMSE$	r	t/s	$NMSE$	r	t/s	$NMSE$	r	t/s	$NMSE$	r	t/s
SVT	5.11	5	4.96	4.97	5	33.34	4.75	5	52.96	5.17	5	121.39
APG	4.14	45	5.29	4.05	91	369.17	3.81	104	462.34	3.78	143	2 615.23
R1MP	29.42	142	0.62	27.51	118	1.12	30.02	141	2.84	29.20	156	5.29
Soft-Impute	6.26	52	45.31	9.74	83	201.37	21.30	94	529.48	37.25	88	1 282.21
WNNM	2.01	5	25.70	2.33	5	156.98	2.17	5	429.73	2.30	5	984.29
WNNM-Impute	1.99	5	0.25	1.89	5	0.79	1.84	5	2.03	1.86	5	2.45

实验结果如表1所示. 从表1可以看出，WNNM-Impute算法无论是精度还是效率都要优于其他算法. 具体来说，WNNM-Impute算法在所有的数据集上均得到了最小的NMSE，且运行的平均时间是最短的. 此外，在实验中，SVT，WNNM和WNNM-Impute算法能够得到秩为5的近似解，而APG，R1MP和Soft-Impute算法得到的近似解的秩远高于5. 实验同时揭示了在WNNM-Impute算法中引入不精确的近邻算法和Nesterov加速准则能够极大地提升算法的效率.

第二组实验为在真实图像数据集上的低秩矩阵补全. 实验中所使用的图像数据如图1所示，每张图片的大小为512×512.

在实验中，我们直接对原始图像进行处理. 由于原始图像可能不是严格低秩的，因此使用秩为50的矩阵去近似原始图像. 对于每一张图像，随机地去掉图像中50%的数据，余下的数据作为观察值. 在本组实验中，采用如下方式来评估所有算法的性能：

1) $PSNR$(peak signal-to-noise ratio)；

2) 运行时间 t.

实验结果如表 2 和图 2 所示.

图 1　测试图像集

表 2　不同算法在图像数据集上的恢复结果

	Barbara		Bridge		Clown		Couple		Crowd	
	PSNR	*t*/s	*PSNR*	*t*/s	*PSNR*	*t*/s	*PSNR*	*t*/s	*PSNR*	*t*/s
SVT	21.57	13.20	21.49	11.78	25.43	11.39	24.31	11.25	22.96	11.84
APG	23.74	9.39	23.16	7.78	26.36	7.60	25.57	7.67	24.88	7.82
R1MP	18.56	0.41	14.29	0.34	20.11	0.42	19.51	0.47	18.28	0.44
Soft-Impute	23.67	21.30	16.41	1.34	26.35	16.97	16.74	1.35	24.85	20.74
WNNM	23.61	10.32	23.25	8.76	26.69	13.92	26.33	8.71	25.05	14.37
WNNM-Impute	23.92	0.57	23.76	0.43	26.71	0.66	26.30	0.63	25.16	0.58
	Fingerprint		Houses		Girlface		Lena		Kiel	
	PSNR	*t*/s	*PSNR*	*t*/s	*PSNR*	*t*/s	*PSNR*	*t*/s	*PSNR*	*t*/s
SVT	19.47	10.53	19.52	12.75	25.98	11.36	25.61	11.60	21.55	13.52
APG	23.06	7.89	22.34	8.07	26.11	7.97	25.86	7.82	21.59	8.30
R1MP	12.26	0.40	10.97	0.51	21.07	0.47	20.38	0.53	18.37	0.55
Soft-Impute	14.86	1.40	13.72	1.38	26.41	1.31	25.97	1.56	21.71	1.36
WNNM	23.34	10.29	22.76	9.84	26.83	9.17	26.11	9.34	22.17	9.14
WNNM-Impute	23.32	0.51	23.01	0.73	26.82	0.68	26.16	0.75	22.25	0.68

从表 2 可以看出，在大多数测试图像数据上 WNNM-Impute 算法得到更大的 *PNSR* 值. 进一步分析发现，虽然 WNNM-Impute 算法的效率略低于 R1MP 算法，但是本文提出的算法得到的 *PNSR* 值远好于 R1MP 算法. 因此，综合考虑精度和效率，本次实验再一次证实本文提出的 WNNM-Impute 算法的有效性.

为了进一步测试 WNNM-Impute 算法的有效性，接下来将在真实数据集上进行实验. 这些数据集包括：Jester1，Jester2，Jester3，MovieLens-100K，MovieLens-1M 和 MovieLens-10M. 这些数据集的特性如表 3 所示，其中 Jester 数据集来源于一个笑话推荐系统，而 MovieLens 数据集来自于 MovieLens 网站.

1) Jester1：包括 24 983 名用户对 36 个或更多笑话的评价；

2) Jester2：包括 23 500 名用户对 36 个或更多笑话的评价；

3) Jester3：包括 24 983 名用户评价了 15 到 35 个笑话；

4) MovieLens-100K：包含 942 名用户对 1 682 部影片的 100 000 个评价；

5) MovieLens-1M：包含6 040名用户对3 900部影片的1 000 000个评价；

6) MovieLens-10M：包含69 878名用户对10 677部影片的1 000 000个评价.

a. 原始图像 b. 观察图像 c. 恢复结果

图2 WNNM-Impute算法在Lena图像上的恢复结果

表3 协同过滤数据集的特征

dataset	ausers	jokes and movies	ratings
Jester1	24 983	100	10^6
Jester1	23 500	100	10^6
Jester1	24 983	100	6×10^5
MovieLens-100K	943	1 682	10^5
MovieLens-1M	6 040	3 706	10^6
MovieLens-10M	69 878	10 677	10^7

在本组实验中，采用如下方式来评估所有算法的性能：

1) 计算

$$RMSE=\frac{\sqrt{\| \mathcal{P}_{\Omega^{\perp}}(\boldsymbol{X}-\boldsymbol{M})\|_F^2}}{|\Omega^{\perp}|_1}$$

其中：这里$\Omega^{\perp}$表示测试数据集，$\boldsymbol{X}$为结果矩阵；

2) 结果矩阵的秩r；

3) 运行时间t.

在实验中，随机地从每个协调过滤数据集中选取50%的数据作为观察数据，余下的为测试数据. 实验结果如表4和表5所示.

表4 不同算法在Jester笑话数据集上的恢复结果

	Jester1			Jester2			Jester3		
	RMSE	r	t/s	*RMSE*	r	t/s	*RMSE*	r	t/s
SVT	4.441 0	3	60.55	4.475 1	2	59.86	4.923 6	2	78.34
APG	4.522 1	31	122.30	4.416 2	30	118.72	4.956 4	31	267.35
R1MP	5.005 6	6	0.90	4.921 8	5	0.66	5.249 4	5	0.82
Soft-Impute	4.304 8	27	23.52	4.392 6	25	15.93	4.758 1	25	30.10
WNNM	4.335 9	3	20.19	4.527 1	2	14.99	4.882 4	2	28.67
WNNM-Impute	4.301 7	3	1.34	4.405 8	2	0.91	4.742 5	2	3.94

表 5　不同算法在 MovieLens 数据集上的恢复结果

	MovieLens-100K			MovieLens-1M			MovieLens-10M		
	RMSE	r	t/s	*RMSE*	r	t/s	*RMSE*	r	t/s
SVT	0.877 6	2	42.26	0.870 4	5	512.81	—	—	$>10^4$
APG	1.045 2	43	476.26	—	—	$>10^4$	—	—	$>10^4$
R1MP	0.988 4	9	0.24	0.969 3	23	1.67	0.943 5	42	47.19
Soft-Impute	0.961 2	39	4.57	0.912 7	90	95.73	0.850 1	107	1 528.61
WNNM	0.977 4	2	4.22	0.920 8	5	84.27	0.863 3	8	1 186.41
WNNM-Impute	0.913 5	2	0.55	0.910 6	5	5.62	0.842 7	8	118.76

从表 4 和表 5 可以看出，在 6 个协调过滤数据集上 WNNM-Impute 算法几乎总能得到最小的 *RMSE* 值，在效率方面仅比 R1MP 慢一点. 在实验中发现 SVT，WNNM 和 WNNM-Impute 算法能够得到秩更低的解，而 APG，R1MP 和 Soft-Impute 算法不能得到近似低秩矩阵解. 该组实验进一步说了，WNNM-Impute 算法在真实协调过滤数据集上是高效的.

4　结束语

本文利用加权核范数去松弛原始低秩极小化问题，得到一个非凸非光滑的优化问题. 为了高效地求解该问题，文中利用 Soft-Impute 的算法思想，融入不精确近邻算子和 Nesterov 优化理论，提出了一个快速、准确的低秩矩阵补全算法. 实验结果表明，在不同规模和不同采样率的模拟数据集上 WNNM-Impute 算法能够得到更加精确的解且效率更高；在采样率相同和不同规模的协同过滤数据集上 WNNM-Impute 算法能够得到秩更低且更好质量的解. 因此，本文提出的 WNNM-Impute 算法是一个具有较强竞争力的低秩矩阵补全算法.

参考文献：

[1] RENNIE J D M, SREBRO N. Fast Maximum Margin Matrix Factorization for Collaborative Prediction [C] //ICML '05: Proceedings of the 22nd International Conference on Machine Learning. New York: Association for Computing Machinery, 2005: 713-719.

[2] KOREN Y. Factorization Meets the Neighborhood: a Multifaceted Collaborative Filtering Model [C] //KDD '08: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery, 2008: 426-434.

[3] LI X H, WANG Z, GAO C, et al. Reasoning Human Emotional Responses from Large-Scale Social and Public Media [J]. Applied Mathematics and Computation, 2017, 310: 182-193.

[4] PENG X, TANG H J, ZHANG L, et al. A Unified Framework for Representation-Based Subspace Clustering of Out-of-Sample and Large-Scale Data [J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(12): 2499-2512.

[5] LIU Y, NIE F P, GAO Q X. Nuclear-Norm Based Semi-Supervised Multiple Labels Learning [J]. Neurocomputing, 2018, 275: 940-947.

[6] LIU Y, GAO Q X, LI J, et al. Zero Shot Learning via Low-Rank Embedded Semantic AutoEncoder [C] //Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2018: 2490-2496.

[7] 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.

[8] CANDÈS E J, RECHT B. Exact Matrix Completion via Convex Optimization [J]. Foundations of Computational Mathematics, 2009, 9(6): 717-772.

[9] FAZEL M. Matrix Rank Minimization with Applications [D]. Stanford: Stanford University, 2002.

[10] CAI J F, CANDÈS E J, SHEN Z W. A Singular Value Thresholding Algorithm for Matrix Completion [J]. SIAM Journal on Optimization, 2010, 20(4): 1956-1982.

[11] TOH K C, YUN S. An Accelerated Proximal Gradient Algorithm for Nuclear Norm Regularized Linear Least Squares Problems [J]. Pacific Journal of Optimization, 2010, 6(3): 615-640.

[12] MA S Q, GOLDFARB D, CHEN L F. Fixed Point and Bregman Iterative Methods for Matrix Rank Minimization [J]. Mathematical Programming, 2011, 128(1-2): 321-353.

[13] MAZUMDER R, HASTIE T, TIBSHIRANI R. Spectral Regularization Algorithms for Learning Large Incomplete Matrices [J]. Journal of Machine Learning Research, 2010, 11: 2287-2322.

[14] YAO Q M, KWOK J T. Accelerated inexact Soft-Impute for Fast Large-Scale Matrix Completion [C] // Proceedings of the International Joint Conference on Artificial Intelligence. New York: IEEE Computer Society Press: 4002-4008.

[15] FAN J Q, LI R Z. Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties [J]. Journal of the American Statistical Association, 2001, 96(456): 1348-1360.

[16] ZHANG T Z. Analysis of Multi-Stage Convex Relaxation for Sparse Regularization [J]. Journal of Machine Learning Research, 2010, 11(3): 1081-1107.

[17] HU Y, ZHANG D B, YE J P, et al. Fast and Accurate Matrix Completion via Truncated Nuclear Norm Regularization [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(9): 2117-2130.

[18] ZHANG C H. Nearly Unbiased Variable Selection under Minimax Concave Penalty [J]. The Annals of Statistics, 2010, 38(2): 894-942.

[19] PENG D T, XIU N H, YU J. $S_{1/2}$ Regularization Methods and Fixed Point Algorithms for Affine Rank Minimization Problems [J]. Computational Optimization and Applications, 2017, 67(3): 543-569.

[20] WANG Z, WANG W D, WANG J J, et al. Fast and Efficient Algorithm for Matrix Completion via Closed-Form 2/3-Thresholding Operator [J]. Neurocomputing, 2019, 330: 212-222.

[21] GU S H, XIE Q, MENG D Y, et al. Weighted Nuclear Norm Minimization and Its Applications to Low Level Vision [J]. International Journal of Computer Vision, 2017, 121(2): 183-208.

[22] YAO Q M, KWOK J T. Accelerated and Inexact Soft-Impute for Large-Scale Matrix and Tensor Completion [J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 31(9): 1665-1679.

责任编辑 张枸