Advances in Applied Mathematics 应用数学进展, 2021, 10(11), 3796-3801
Published Online November 2021 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2021.1011402

Hans汉斯

加权非凸非光滑低秩矩阵填充

尚紫微，张　军

辽宁师范大学，辽宁　大连

收稿日期：2021年10月11日；录用日期：2021年11月1日；发布日期：2021年11月16日

摘　要

本文利用矩阵奇异值上的ℓ_0范数的非凸替代族来逼近秩函数，提出一种新的加权非凸非光滑最小化问题，并使用迭代加权核范数(IRNN)算法来求解该问题。实验结果表明，该方法能够很好地处理非凸非光滑问题，实现图像去噪。

关键词

加权非凸低秩最小化，迭代加权核范数算法，超梯度，图像去噪

Weighted Nonconvex Nonsmooth Low-Rank Matrix Completion

Ziwei Shang, Jun Zhang

Liaoning Normal University, Dalian Liaoning

Received: Oct. 11th, 2021; accepted: Nov. 1st, 2021; published: Nov. 16th, 2021

Abstract

In this paper, we propose a new weighted nonconvex nonsmooth minimization problem and use Iteratively Reweighted Nuclear Norm (IRNN) algorithm to solve the problem. It is worth mentioning that we use the nonconvex substitution family of ℓ_0-norm on the singular value of the matrix to approximate the rank function. Experimental results show that this method can deal with nonconvex nonsmooth problems well and realize image denoising.

Keywords

Weighted Nonconvex Low Rank Minimization, Iteratively Reweighted Nuclear Norm Algorithm, Super Gradient, Image Denoising

文章引用: 尚紫微, 张军. 加权非凸非光滑低秩矩阵填充[J]. 应用数学进展, 2021, 10(11): 3796-3801.
DOI: 10.12677/aam.2021.1011402

1. 引言

低秩矩阵逼近是一种从退化的观测中恢复低秩矩阵的方法，近年来受到了计算机视觉和机器学习界的广泛关注，并已经成功应用于各个领域，如矩阵填充、背景建模和运动分割等。

众所周知，秩极小化问题很难求解。因此，秩函数通常由凸核范数代替，这类凸问题可以由许多已知的求解器[1] [2] [3]有效地求解。然而，核范数是秩函数的松散近似，由核范数近似秩函数得到的解通常是次优的。为了更好地逼近秩函数，本文利用矩阵奇异值上的 ℓ_0 范数的非凸替代族来逼近秩函数，从而提出了一种新的加权非凸非光滑最小化问题，并使用迭代加权核范数(IRNN)算法对该问题进行求解。

2. 预备知识

在这一节中，我们介绍了 Lipschitz 连续的定义，并引入了超梯度的概念，它将在求解非凸非光滑问题中使用。我们知道，次梯度是凸函数在非光滑点的梯度的扩展，实际上超梯度则是凹函数在非光滑点的梯度的扩展。

定义 1 (Lipschitz 连续)称 f 的梯度 ∇f 为 Lipschitz 连续，如果对于任意 $X,Y\in\mathbb{R}^{m\times n}, L(f)>0$，有

$$\left\|\nabla f(X)-\nabla f(Y)\right\|_F\le L(f)\left\|X-Y\right\|_F \tag{1}$$

称 $L(f)$ 为 ∇f 的 Lipschitz 常数。

定义 2 (次梯度)若 g 是凸的、非光滑的，则其在 x 处的次梯度为 u，则下列不等式成立

$$g(x)+\langle u,y-x\rangle\le g(y) \tag{2}$$

定义 3 (超梯度)设 g 是凹的，向量 v 是 g 在 x 点的超梯度，如果对于每一个 y，有下面的不等式成立

$$g(x)+\langle v,y-x\rangle\ge g(y) \tag{3}$$

非光滑点处的超梯度可能不是唯一的，g 在 x 处的所有超梯度称为 g 在 x 处的超微分，并表示为 $\partial g(x)$。如果 g 在 x 处可微，那么 $\nabla g(x)$ 是唯一的超梯度，即 $\partial g(x)=\{\nabla g(x)\}$。一些常见凹函数的超梯度如表 1 所示。

Table 1. Popular nonconvex surrogate functions and their supergradients
表 1. 常见的非凸替代函数和它们的超梯度

罚函数	公式 $g_\lambda(\theta),\theta\ge 0,\lambda>0$	超梯度 $\partial g_\lambda(\theta)$
L_p	$\lambda\theta^p$	$\begin{cases}+\infty, & 当\theta=0,\\ \lambda p\theta^{p-1}, & 当\theta>0.\end{cases}$
SCAD	$\begin{cases}\lambda\theta, & 当\theta\le\lambda,\\ \dfrac{-\theta^2+2\gamma\lambda\theta-\lambda^2}{2(\gamma-1)}, & 当\lambda<\theta\le\gamma\lambda,\\ \dfrac{\gamma^2(\gamma+1)}{2}, & 当\theta>\gamma\lambda.\end{cases}$	$\begin{cases}\lambda, & 当\theta\le\lambda,\\ \dfrac{\gamma\lambda-\theta}{\gamma-1}, & 当\lambda<\theta\le\gamma\lambda,\\ 0, & 当\theta\le\gamma\lambda.\end{cases}$

Continued

Logarithm	$\frac{\lambda}{\log(\gamma+1)}\log(\gamma\theta+1)$	$\frac{\gamma\lambda}{(\gamma\theta+1)\log(\gamma+1)}$
MCP	$\begin{cases}\lambda\theta-\frac{\theta^2}{2\gamma}, & \text{当}\,\theta<\gamma\lambda,\\ \frac{1}{2}\gamma\lambda^2, & \text{当}\,\theta\ge\gamma\lambda.\end{cases}$	$\begin{cases}\lambda-\frac{\theta}{\gamma}, & \text{当}\,\theta<\gamma\lambda,\\ 0, & \text{当}\,\theta\ge\gamma\lambda.\end{cases}$
ETP	$\frac{\lambda}{1-\exp(-\gamma)}\left(1-\exp(-\gamma\theta)\right)$	$\frac{\lambda\gamma}{1-\exp(-\gamma)}\exp(-\gamma\theta)$

对于凹函数 $g,-g$ 是凸的，反之亦然。根据这一事实，g 的超梯度与 $-g$ 的次梯度之间存在以下关系。

引理 1 设 $g(x)$ 是凹的，$h(x)=-g(x)$。对于任何 $v\in\partial g(x)$，$u=-v\in-\partial h(x)$ 成立，反之亦然。

引理 1 给出了超梯度与次梯度的关系，从而得到了超梯度的一些性质。众所周知，对于任何 $u_1\in\partial h(x)$，$u_2\in\partial h(y)$，凸函数 h 的次微分称为单调算子，即

$$\langle u_1-u_2,x-y\rangle\ge 0 \tag{4}$$

凹函数的超微分则具有如下相反的性质。

引理 2 对于任何 $v_1\in\partial g(x)$，$v_2\in\partial g(y)$，凹函数 g 的超微分称为反单调算子，即

$$\langle v_1-v_2,x-y\rangle\le 0 \tag{5}$$

根据引理 2，假设 $g(x)$ 是凹的。如果 $x\le y$，那么对于任意 $v_1\in\partial g(x)$ 和 $v_2\in\partial g(y)$ 都有 $v_1\ge v_2$。

引理 3 [4] 对于任意 $\lambda>0,Y\in\mathbb{R}^{m\times n}$ 和 $0\le w_1\le w_2\le\cdots\le w_s$ $\left(s=\min(m,n)\right)$，下列问题的全局最优解

$$\min\lambda\sum_{i=1}^{s}w_i\sigma_i(X)+\frac{1}{2}\|X-Y\|_F^2 \tag{6}$$

由加权奇异值阈值(WSVT)给出

$$X^*=US_{\lambda w}(\Sigma)V^{\mathrm{T}} \tag{7}$$

其中 $Y=U\Sigma V^{\mathrm{T}}$ 是 Y 的 SVD，并且 $S_{\lambda w}(\Sigma)=\mathrm{Diag}\left\{(\Sigma_{ii}-\lambda w_i)_+\right\}$。

3. 加权非凸非光滑最小化问题

在这一部分中，我们利用矩阵奇异值上的 ℓ_0 范数的非凸替代族来逼近秩函数，提出了加权非凸非光滑最小化问题，并利用迭代加权核范数(IRNN)算法对该问题进行求解。

3.1. 模型的建立

本文为了更好地逼近秩函数，将表 1 中的 ℓ_0 范数的非凸替代族扩展到矩阵的奇异值上，建立了下列一般加权非凸非光滑低秩极小化模型。

$$\min_{X\in\mathbb{R}^{m\times n}}F(x)=\sum_{i=1}^{m}s_i g\left(\sigma_i(X)\right)+f(X) \tag{8}$$

其中 $m\ll n$，$\sigma_i(X)$ 是 $X\in\mathbb{R}^{m\times n}$ 的第 i 个奇异值，s 是非递减、非负的权重向量，$s=sort\left(\left(\max(\sigma-1,0)+1\right)/\left(\max(\sigma+1,0)+1\right)\right)$，$g$ 是罚函数，f 是损失函数。它们分别满足以下假设：

假设 1　$g:\mathbb{R}^+\to\mathbb{R}^+$ 在 $[0,\infty)$ 上是单调递增的、连续的、凹的，并且可以是非光滑的。

假设 2　$f:\mathbb{R}^{m\times n}\to\mathbb{R}^+$ 是 $C^{1,1}$ 型光滑函数，即它的梯度是 Lipschitz 连续的。

可以看到表 1 中所有 ℓ_0 范数的非凸替代都满足假设 1，因此 $\sum_{i=1}^{m} g\left(\sigma_i\left(X\right)\right)$ 是秩函数的非凸替代，对于假设 2 中的损失函数 f，最广泛使用的是 $\frac{1}{2}\left\|X-Y\right\|_F^2$。

3.2. 模型的求解

为了符号简便，我们把 X 的奇异值表示为 $\sigma_1,\sigma_2,\cdots,\sigma_m$，并且是非递增的。$X^k$ 表示第 k 次迭代的变量 X，$\sigma_i\left(X^k\right)$ 是 X^k 的第 i 个奇异值，记为 σ_i^k。

由假设 g 是凹函数，根据定义 3 超梯度的概念，对于 $w_i^k\in\partial g\left(\sigma_i^k\right)$，我们有

$$g\left(\sigma_i\right)\le g\left(\sigma_i^k\right)+w_i^k\left(\sigma_i-\sigma_i^k\right) \tag{9}$$

又由于奇异值是非递减且不小于 0 的，即 $\sigma_1^k\ge\sigma_2^k\ge\cdots\ge\sigma_m^k\ge0$，由引理 2 超梯度是反单调算子，我们有

$$0\le w_1^k\le w_2^k\le\cdots\le w_m^k \tag{10}$$

因此，我们可以解决下列松弛问题来更新 X^{k+1}：

$$\begin{aligned} X^{k+1} &= \arg\min_X \sum_{i=1}^{m} s_i g\left(\sigma_i^k\right)+s_i w_i^k\left(\sigma_i-\sigma_i^k\right)+f\left(X\right) \\ &= \arg\min_X \sum_{i=1}^{m} s_i w_i^k\sigma_i+f\left(X\right) \end{aligned} \tag{11}$$

进一步地，我们在 X^k 处对 $f\left(X\right)$ 进行线性化，并添加一个近端项：

$$f\left(X\right)\approx f\left(X^k\right)+\left\langle\nabla f\left(X^k\right),X-X^k\right\rangle+\frac{\mu}{2}\left\|X-X^k\right\|_F^2 \tag{12}$$

式中：$\mu>L\left(f\right)$。结合(9)和(10)，我们可以更新 X^{k+1} 通过求解下式：

$$\begin{aligned} X^{k+1} &= \arg\min_X \sum_{i=1}^{m} s_i g\left(\sigma_i^k\right)+s_i w_i^k\left(\sigma_i-\sigma_i^k\right) \\ &\quad +f\left(X^k\right)+\left\langle\nabla f\left(X^k\right),X-X^k\right\rangle+\frac{\mu}{2}\left\|X-X^k\right\|_F^2 \\ &= \arg\min_X \sum_{i=1}^{m} s_i w_i^k\sigma_i+\frac{\mu}{2}\left\|X-\left(X^k-\frac{1}{\mu}\nabla f\left(X^k\right)\right)\right\|_F^2 \\ &= \arg\min_X \frac{1}{\mu}\sum_{i=1}^{m} s_i w_i^k\sigma_i+\frac{1}{2}\left\|X-\left(X^k-\frac{1}{\mu}\nabla f\left(X^k\right)\right)\right\|_F^2 \end{aligned} \tag{13}$$

因此，根据引理 3，(13)的全局最优解可以由加权奇异值阈值(WSVT)给出

$$X^*=US_{sw/\mu}\left(\Sigma\right)V^{\mathrm{T}} \tag{14}$$

其中，$Y=U\Sigma V^{\mathrm{T}}$ 是 Y 的 SVD，$S_{sw/\mu}\left(\Sigma\right)=\mathrm{Diag}\left\{\left(\Sigma_{ii}-sw/\mu\right)_+\right\}$。

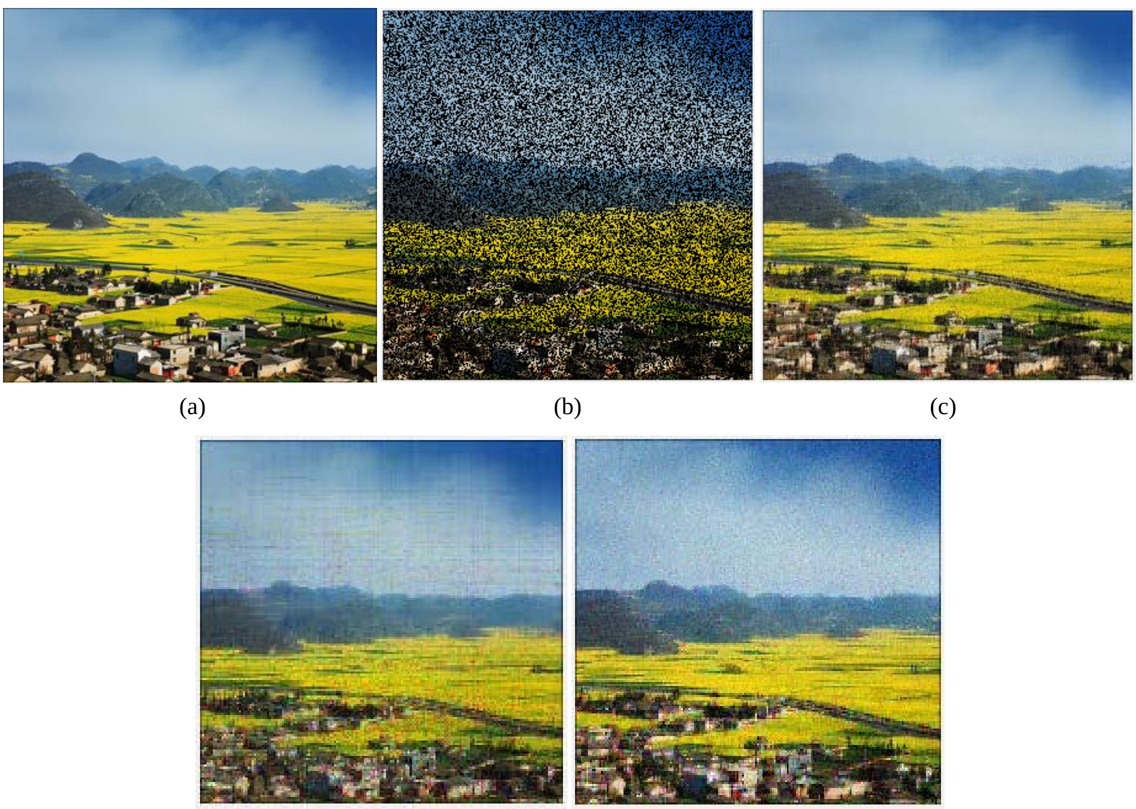
从引理 3 可以看出，(13)在求解(11)的过程中起着重要的作用，并且该方法对于满足假设 1 的所有 g

都成立。如果 $g(x)=x$，那么 $\sum_{i=1}^{m} g(\sigma_i)$ 降到凸核范数 $\|X\|_*$。在这种情况下，对于所有 $i=1,\cdots,m$，$w_i^k=1$。加权奇异值阈值(WSVT)就降为传统的奇异值阈值法(SVT) [5]，这是凸低秩优化中的一个重要子程序。这时更新规则(11)降为已知的近端梯度方法[6]。

通过解(13)更新 X^{k+1}，我们再更新权重 $w_i^{k+1} \in \partial g\left(\sigma_i\left(X^{k+1}\right)\right), i=1,2,\cdots,m$。迭代更新 X^{k+1} 及其奇异值对应的权值，得到了迭代加权核范数(IRNN)算法。IRNN 的整个过程如算法 1 所示。如果 Lipschitz 常数 $L(f)$ 未知或不可计算，则可使用回溯规则在每次迭代中估计 μ [6]。其中，对于 L_p 惩罚，如果 $\sigma_i^k=0$，那么 $w_i^k \in \partial g\left(\sigma_i^k\right), i=\{+\infty\}$。根据(11)中 X^{k+1} 的更新规则，我们得到了 $\sigma_i^{k+1}=0$，这保证了序列 $\left\{X^k\right\}$ 的秩是非递增的。

4. 实验

在本节中，我们将本文方法应用于图像恢复，并与 TNNR 和 DW-TNNR 方法作比较。我们选取的图片大小为 300 × 300 × 3，并且噪声等级设置为 50%。通过对比原始图片的恢复情况，可以直观看到去噪模型的恢复效果如图 1 所示。同时，峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)值是评价图像质量的常用标准。我们使用恢复图像的 PSNR 值来评估在相同噪声情况下不同方法的性能。值得注意的是，恢复图像的 PSNR 值越高，代表图片的恢复效果就越好，具体数据如表 2 所示。



(a) (b) (c) (d) (e)

Figure 1. Comparison of image restoration results of different methods: (a) Original images; (b) Adding 50% noise images; (c) IRNN; (d) TNNR-APGL; (e) DW-TNNR

图 1. 不同方法图像恢复结果对比：(a) 原始图片；(b) 添加 50%噪声图片；(c) IRNN；(d) TNNR-APGL；(e) DW-TNNR

Table 2. Comparison of PSNR values and time for denoising by four algorithms
表 2. 四种算法去噪 PSNR 值和时间对比

算法	PSNR	运行时间(s)
IRNN	24.56	12.66
TNNR-APGL	18.86	433.37
DW-TNNR	19.57	18.40

5. 结论

本文提出并求解了一种新的加权非凸非光滑最小化问题。图像去噪的实验结果表明，在相同噪声下，该方法能够有效去噪并获得较好的视觉效果。

参考文献

[1] Toh, K. and Yun, S. (2010) An Accelerated Proximal Gradient Algorithm for Nuclear Norm Regularized Linear Least Squares Problems. *Pacific Journal of Optimization*, **6**, 615-640.

[2] Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. (2011) Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends® in Machine Learning*, **3**, 1-122. https://doi.org/10.1561/2200000016

[3] Mazumder, R., Hastie, T. and Tibshirani, R. (2010) Spectral Regularization Algorithms for Learning Large Incomplete Matrices. *JMLR*, **11**, 2287-2322.

[4] Gaıffas, S. and Lecue, G. (2011) Weighted Algorithms for Compressed Sensing and Matrix Completion. arXiv:1107.1638 [cs.IT]

[5] Cai, J., Candès, E. and Shen, Z. (2010) A Singular Value Thresholding Algorithm for Matrix Completion. *SIAM Journal on Optimization*, **20**, 1956-1982. https://doi.org/10.1137/080738970

[6] Beck, A. and Teboulle, M. (2009) A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, **2**, 183-202. https://doi.org/10.1137/080716542