学校编码：10384　　分类号______密级______

学号：19020141152627　　UDC______

厦门大学

硕　士　学　位　论　文

# 基于$S_{\frac{1}{2}}$建模的低秩矩阵填充及其快速近端梯度算法的研究

# The Low-rank Matrix Completion Based on $S_{\frac{1}{2}}$ Modeling and the Research of its APG Algorithm

侯 晓 宇

指导教师姓名：　谭　忠　教　授

专 业 名 称：　应　用　数　学

论文提交日期：　2017 年　　月

论文答辩时间：　2017 年　　月

学位授予日期：

答辩委员会主席：__________

评　　阅　　人：__________

2017 年　4　月

# 厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（　　　　　　　　　　　　）课题（组）的研究成果，获得（　　　　　　　　　）课题（组）经费或实验室的资助，在（　　　　　　　）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年　　月　　日

# 厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（　　　）1.经厦门大学保密委员会审查核定的保密学位论文，于　　年　　月　　日解密，解密后适用上述授权。

（　　　）2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年　　月　　日

# 摘要

低秩矩阵填充问题（Low-rank Matrix Completion）是指对于有部分位置上元素未知的矩阵，在假设矩阵低秩的前提下，可以通过优化算法来将其填充成一个完整的矩阵。低秩矩阵填充在机器学习、图像处理、推荐系统等领域发挥着重要的作用，是现今处理海量、高维数据的有力分析工具。本文首先介绍了低秩矩阵填充模型的理论发展，再分别根据将原模型进行凸松弛和非凸松弛后的改进模型综述了目前主要的算法，其中包括凸松弛的SVT算法,APG算法,ALM算法和非凸松弛的WMMN模型，并分析说明了不同算法在不同的领域，针对不同的模型有着各自的优势。

目前主要用于解决低秩矩阵填充的模型是用矩阵核范数来逼近目标函数中的秩函数，然而在实际数据质量较差的情形下，由于凸松弛过多使得最终求得的解不够精确。源自信号处理领域中的压缩感知原理，我们了解到利用$L_{\frac{1}{2}}$正则化方法可以得出更稀疏的解，本文将一维信号的稀疏性扩展到二维矩阵的低秩性，依据徐宗本提出的矩阵$S_{\frac{1}{2}}$范数，将原目标函数进行非凸松弛，提出基于$S_{\frac{1}{2}}$建模的无约束矩阵秩最小化模型和低秩-稀疏分解模型，并利用加速近端梯度算法来进行求解。最后在本文的数值模拟中，证实了新模型求解低秩矩阵填充问题时可以得出更精确的解。

本文的创新之处在于：一、为了更好地诱导低秩矩阵填充问题中矩阵的低秩性，结合加权核范数思想与$L_{\frac{1}{2}}$正则化方法，提出了基于矩阵$S_{\frac{1}{2}}$范数建模的无约束惩罚模型，并在低秩近似、低秩-稀疏分解、这两种情形下推演出其梯度下降算法的迭代过程；二、利用加速近端梯度算法求解基于$S_{\frac{1}{2}}$建模的低秩近似和低秩-稀疏分解问题，并在数值实验中验证了新算法对比原APG算法具有更好的计算精确性，在低秩矩阵填充问题中有更好的计算精度和计算效率。

**关键词：** 低秩矩阵填充;$L_{\frac{1}{2}}$正则化;加速近端梯度法

# Abstract

Low-rank matrix completion refers to problem that use optimization algorithm to fill a matrix which have unknown elements into a complete matrix.In general,we assume the incomplete matrix is a low-rank matrix.This methods have played an important role in areas such as machine learning, image processing, recommendation systems and so on,it's a powerful analysis tool of high-dimensional data.In this article we introduce the model development of low-rank matrix completion.According to the convex relaxation and non convex relaxation of original model, we summarize the classical algorithm such as SVT,APG,ALM and non convex relaxation WMMN model.

At present,the main algorithm solving low-rank matrix completion is a convex relaxation of original model with nuclear norm.However, it is so much convex relaxation of original model that nuclear norm minimization could not have a more accurate solution.Derived from compressed sensing principle in the field of signal processing, we know that one can take advantage of the $L_{\frac{1}{2}}$ regularization to draw a more sparse solution.Thus in this paper, we propose a new model based on $S_{\frac{1}{2}}$ matrix norm,which is the $l_{\frac{1}{2}}$ norm of matrix singular value vector. On the basis of $S_{\frac{1}{2}}$ modeling,we proposed low rank approximation model, low rank - sparse matrix factorization and maximum margin matrix factorization futher. And we confirmed the accuracy and efficiency of modified APG algorithm in numerical simulation.

There are two innovation of this paper, first, on the basis of $L_{\frac{1}{2}}$ regularization,we proposed three new models including low rank approximation model and low rank - sparse matrix factorization, and also analysis the convergence of the algorithm.Then we modified the accelerated proximal gradient algorithm solving low rank approximation model and low rank - sparse matrix factorization based on $S_{\frac{1}{2}}$ modeling,and verified it has better accuracy in the low-rank condition.

**Key Words:** low-rank matrix completion;$L_{\frac{1}{2}}$ regularization; APG algorithm

# 目 录

# Contents

# 第一章　绪论

## 1.1 研究背景

低秩矩阵填充（Low-rank matrix completion）作为海量高维数据的有力分析工具，在机器学习、图像处理、推荐系统等领域扮演着重要的角色。低秩矩阵填充问题研究的是在假设矩阵低秩的条件下，通过合理的算法将一个含有未知元素的矩阵填充成为一个完整的矩阵。其一个经典应用场景就是著名的Netflix 电影推荐问题，通过填充用户对电影的评分矩阵的方式，将用户可能最喜欢的电影推荐给用户。受到信号处理领域中压缩感知原理的启发，Candès [1]将一维信号向量的稀疏性拓展到二维数据矩阵的低秩性上，将低秩矩阵填充问题转化为一个凸优化问题，从而利用目前的凸优化算法来求解。并且证明了在矩阵满足一定的条件下，可以以一个接近于1 的概率合理精确的填充一个低秩矩阵。

在数据日益递增的信息时代，我们面对的海量信息中有很多都是冗余数据，而实际中很多数据均可表示成为矩阵的形式，因此这些矩阵都是低秩的。矩阵的低秩性本质上是利用高维空间的低维结构，寻找一个合适的低秩矩阵近似原复杂矩阵，达到维数约简的作用，可以有效地降低所需处理的数据量。低秩矩阵填充理论能够有效的解决高维数据的分析与处理，在当今的社会生活与科学研究中占据着越来越重要的地位。

低秩矩阵填充问题希望能利用有效而少量已知数据信息来推断出普遍形势。在实际生活中，以Netflix 电影推荐问题为例，将每个用户对电影的评分作为行，每部电影作为列，由于每个用户所看电影有限，则用户对电影的评分数据形成一个不完整的矩阵，人们希望可以通过这些已知的的矩阵元素来估计出所有的评分信息，从而可以将各个用户可能评分最高的电影推荐给该用户。低秩矩阵填充问题的低秩假设条件也是矩阵可填充的前提，实际中的数据矩阵往往也是低秩的，例如Netflix问题中，影响用户对电影评分的喜好因素只有少数几个。除此之外，对于视频监控的背景恢复以及受损图像的修复等方面，低秩矩阵填充也发挥着重要的作用。

## 1.2 低秩矩阵填充理论发展

低秩矩阵填充问题的求解启发于信号处理中的压缩感知理论，是从一维向量到

二维矩阵的拓展延伸，核心都是对原目标函数松弛以便于求解。在低秩矩阵填充问题的原始模型基础上，学者们做出了一系列模型优化，并证明了改进模型的解与原问题解的等价性以及改进模型的解的存在性，不同模型有着对应可行的高效求解算法。为了能够以一个尽可能低秩的矩阵完美近似于原复杂矩阵，低秩矩阵填充问题的模型在研究中不断优化。

### 1.2.1 压缩感知

在信号处理领域，压缩感知（Compressed Sensing）原理[2][3]是利用现实中的信号具有稀疏性这一特点，降低了经典Nyquist采样定理所需的较高的信号采样率，也可以将原信号准确重构出来，大大提高了信号还原效率[4]。

从数学上讲，搜集到的信号数据可表示成向量形式$x \in \mathbb{R}^n$，信号的稀疏性就可理解为该向量中非零元的个数，即向量x的$l_0$范数。压缩感知的含义就是通过测量较少的样本$y \in \mathbb{R}^m, (m \ll n)$，来还原出原始信号x，即求解下列问题：

$$\begin{aligned} &\min_{X} \ \|x\|_0 \\ &s.t. \quad y = Ax \end{aligned} \tag{1.1}$$

其中矩阵$A \in \mathbb{R}^{m \times n}$可理解成为信号采样形式。

由于目标函数x的$l_0$范数是非连续非凸的，该问题的求解是NP-hard问题。求解该非凸问题可以用贪婪算法，迭代硬阈值算法等，但这些算法均不能保证全局收敛性。文献[3] 中提出将原目标函数进行凸松弛，转而求解如下问题：

$$\begin{aligned} &\min_{X} \ \|x\|_1 \\ &s.t. \quad y = Ax \end{aligned} \tag{1.2}$$

其中$\|x\|_1 = \sum\limits_{i=1}^{n} |x_i|$，这样就将原始问题转化为带约束条件的凸优化问题。

这里我们还需知道，并不是所有信号x都可以通过求解(1.2)而得以还原，这与信号的采样方式，也就是矩阵A的结构有关。文献[3]中给出了矩阵的约束等距性条件（Restricted isometry property, RIP）的定义，并证明了在矩阵A 满足RIP 条件的约束下，(1.2) 的解等价于问题(1.1)的解，并且$l_1$ 范数最小化模型(1.2)存在唯一最优解。

而$l_1$ 范数最小化模型(1.2)，是易于求解的凸优化问题。

### 1.2.2 低秩近似

矩阵的低秩性可看成是一维向量的稀疏性在二维矩阵上的一个扩展，要填充一个含有未知元素的低秩矩阵M，我们要求解出的是这样一个矩阵X，X在M矩阵已知元素的位置上的元素与M相同，并且秩尽可能的小，相当于求解一个矩阵秩最小化问题：

$$\begin{aligned} &\min_X \ rank(X) \\ &s.t. \quad X_{ij} = M_{ij} \quad (i,j) \in \Omega \end{aligned} \tag{1.3}$$

其中，$M \in \mathbb{R}^{m\times n}$是待填充矩阵，M中已知的矩阵元素集合为$\{M_{i,j}, (i,j) \in \Omega\}$。

引入正交投影算子$P_\Omega : \mathbb{R}^{m\times n} \to \mathbb{R}^p$

$$P_\Omega(M) = \begin{cases} M_{i,j}, (i,j) \in \Omega \\ 0, \quad (i,j) \notin \Omega \end{cases}$$

则问题(1.3)转化为：

$$\begin{aligned} &\min_X \ rank(X) \\ &s.t. \quad P_\Omega(X) = P_\Omega(M) \end{aligned} \tag{1.4}$$

优化问题(1.4)的求解也是NP-hard的，受压缩感知理论启发，文献[5]中提出以易于求解的凸优化问题来替代：

$$\begin{aligned} &\min_X \ \|X\|_* \\ &s.t. \quad P_\Omega(X) = P_\Omega(M) \end{aligned} \tag{1.5}$$

其中函数$\|X\|_* = \sum_{i=1}^{r} \sigma_i(X)$是矩阵X的核范数，为矩阵X的所有奇异值之和，$\sigma_1(X) \geqslant \sigma_2(X) \geqslant ... \geqslant \sigma_r(X)$ 是矩阵X的所有奇异值。为充分理解核范数的概念，这里引进矩阵的奇异值分解定理：

**定理 1.1 (奇异值分解定理)** 设$A \in \mathbb{R}^{m\times n}$，$rank(A)=r$，则矩阵A的奇异值分解为：

$$A = USV^T = U\begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} V^T$$

其中$U \in \mathbb{R}^{m\times m}, V \in \mathbb{R}^{n\times n}$是正交矩阵，对角矩阵$\Sigma_r = diag(\sigma_1, \sigma_2, ..., \sigma_r) \in \mathbb{R}^{r\times r}$，并且对角线元素满足$\sigma_1 \geqslant \sigma_2 \geqslant ... \geqslant \sigma_r > 0$。

记$U=(U_1, U_2, ..., U_m), V=(V_1, V_2, ..., V_n)$，则上式可以重新表示为

$$A = \sum_{i=1}^{r} \sigma_i U_i V_i^T$$

$U_i, V_i$分别为奇异值$\sigma_i$对应的左奇异向量和右奇异向量。

由奇异值分解定理，我们可知，一个矩阵的主要数据信息都包含在矩阵的奇异值中，矩阵核范数是矩阵秩函数的一个最优凸近似，在文献[5]中，Candès 给出了问题(1.4) 等价于问题(1.5)更详细的证明。

和压缩感知理论相同，上述核范数最小化模型还需要考虑求解的可行性，以及求解最终能完美填充矩阵的概率有多大。文献[5]将压缩感知中的RIP条件推广到低秩矩阵填充中，为叙述RIP条件，需先了解矩阵的一致性：

**定义 1.1 (一致性)** 设U是$\mathbb{R}^n$的一个r维子空间，$P_U$为正交投影算子，可以将$\mathbb{R}^n$上任意一个向量投影到子空间U中。子空间U的一致性（coherence）定义为：

$$\mu(U) = \frac{n}{r}\max_{1\leqslant i\leqslant n}\|P_U e_i\|^2$$

其中$e_i$为单位标准基向量。

设矩阵$A \in \mathbb{R}^{m\times n}$的秩为r，它的奇异值分解为$A = U_r\Sigma_r V_r^T$，根据定义，可以求出矩阵A的列空间的一致性为：

$$\mu(U) = \frac{m}{r}\max_{1\leqslant i\leqslant m}\|e_i^T U_r\|^2$$

矩阵A的行空间的一致性为：

$$\mu(V)=\frac{n}{r}\max_{1\leqslant i\leqslant n}\|e_i^TV_r\|^2$$

对待填充矩阵A做出假设：

1.存在正数$\mu_0$，使得$\max(\mu(U),\mu(V))\leqslant\mu_0$

2.对于矩阵$UV^T\in\mathbb{R}^{m\times n}$的每一个元素，存在正数$\mu_1$，使得$\sum\limits_{1\leqslant k\leqslant r}u_kv_k^T\leqslant\mu_1\sqrt{\frac{r}{mn}}$

若参数$\mu_0,\mu_1$均$\leqslant\mu$，则称矩阵A满足参数为$\mu$**的强不相关性**，由此可得出低秩矩阵填充的完美恢复定理：

**定理** 1.2 对于秩为r的矩阵$A\in\mathbb{R}^{m\times n}$，其奇异值分解为$A=U_r\Sigma_rV_r^T$，且满足参数$\mu$的强不相干性。不失一般性，令$n=\min(m,n)$，假设矩阵A 中已知元素的个数为$|\Omega|$，观测得到的元素都是独立、均匀随机抽样所得，若存在常数C。使得如果

$$|\Omega|\geqslant C\mu^4n(\log n)^2$$

成立，那么以至少$1-n^{-3}$的概率，低秩矩阵填充问题(1.5)的全局最优解唯一，且解是A。

上述定理表明，若矩阵A满足强不相关性条件，则通过少量的已知元素就可利用核范数最小化模型(1.5)的解来完美恢复出所有元素。而对于任意大小的秩r，已知元素的数量满足

$$|\Omega|\geqslant C\mu^4nr^2(\log n)^2$$

时也有上述定理成立。当r 较小时，已知元素数目只需满足

$$|\Omega|\geqslant Cn^{\frac{6}{5}}r(\log n)$$

模型(1.5)就可以很大概率恢复原矩阵。

事实上，低秩矩阵填充问题都会假设RIP条件成立，再利用不同算法进行求解，本文中的算法也是在该假设条件成立的基础上提出。该定理指出了在一定条件

下，用核范数最小化问题得出的解等价于矩阵秩函数最小化的解，并且指出待填充矩阵的已知元素需要满足随机均匀分布以及已知元素的个数多于一个下界值，矩阵才能够被完美填充。

(1.5)模型为带约束条件的凸优化问题，基于目前有很多良好的求解无约束优化问题的方法，我们可以利用正则化方法将等式约束松弛到目标函数中，则问题可转化为下列形式：

$$\min\ \lambda\|X\|_* + \frac{1}{2}\|P_\Omega(X) - P_\Omega(M)\|_F^2 \tag{1.6}$$

其中$\lambda > 0$是控制复杂度的正则化参数，$\|\cdot\|_F$为矩阵的F范数，定义如下：

**定义 1.2 (Frobenius范数)** 对于矩阵$A = (a_{ij}) \in \mathbb{R}^{m\times n}$，其Frobenius范数为：

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}$$

### 1.2.3 低秩-稀疏矩阵分解

在实际应用中，数据中往往还含有大量噪声，为找到数据的低秩结构，将原矩阵分解成为两个矩阵之和，即$M = X + E$，其中X为低秩部分，E为噪声部分。当矩阵E的元素服从独立同分布的高斯分布时，可利用求解下列最优化问题来获取最优的低秩矩阵X，即经典的主成分分析（PCA）问题：

$$\begin{aligned}&\min_{X,E}\ \|E\|_F\\ &s.t.\ rank(X) \leqslant r, M = X + E\end{aligned}$$

而破坏数据的低维结构的噪声可能不是高斯噪声，当E为较大的稀疏矩阵时，PCA便不再适用，恢复矩阵M成为一个双目标优化问题，要使矩阵X尽可能低秩并且E尽可能稀疏。为此，引入折中因子$\lambda > 0$，将问题转化为：

$$\begin{aligned}&\min_{X,E}\ rank(X) + \lambda\|E\|_0\\ &s.t.\quad M = X + E\end{aligned} \tag{1.7}$$

E为稀疏矩阵指的是其中非零元素的个数极少，$\|E\|_0$是矩阵的拉直向量的$l_0$范

数。则求解该问题也是NP-hard的，类似的，将目标函数中的秩函数凸松弛到核范数，将$l_0$范数凸松弛为$l_1$范数，则问题(1.7) 转化为如下的低秩-稀疏分解模型：

$$\begin{aligned} \min_{X,E} \quad & \|X\|_* + \lambda\|E\|_1 \\ s.t. \quad & M = X + E \end{aligned} \tag{1.8}$$

问题(1.8)又被称为鲁棒主成分分析问题（Robust principal component analysis，RPCA），并在文献[6]中给出了如下定理的证明：

**定理** 1.3 设矩阵$A \in \mathbb{R}^{m\times n}$，其奇异值分解为$A = U\Sigma V^T$，不失一般性，令$n = \min(m,n)$。设$A = X + E$，且满足如下参数为$\mu$的不相干性条件：

$$\max_i \|U^T e_i\|^2 \leqslant \frac{\mu r}{m}$$

$$\max_i \|V^T e_i\|^2 \leqslant \frac{\mu r}{n}$$

$$\|UV^T\|_\infty \leqslant \sqrt{\frac{\mu r}{mn}}$$

其中$e_i$为单位向量，$r = rank(X)$，范数$\|X\|_\infty = \max|x_{ij}|$。稀疏部分E的支撑集元素为均匀分布的，则当满足

$$rank(A) \leqslant \rho_r n\mu^{-1}(\log m)^{-2}, \|E\|_{l_0} \leqslant \rho_s mn$$

时，存在常数c，使得低秩-稀疏分解模型(1.8)以至少$1 - cm^{-10}$的概率恢复出原矩阵，其中模型(1.8)的参数$\lambda = \frac{1}{\sqrt{\max(m,n)}}$，$\rho_r, \rho_s$为正的常数。

该定理解释了低秩-稀疏矩阵分解问题的可解性。低秩-稀疏分解模型也称为是低秩矩阵恢复问题，与低秩矩阵填充不同的是，该问题中原矩阵是完整的，只是一部分元素受到了稀疏噪声影响，所以需要先判定哪些元素受损，通过分解来找到矩阵的低秩部分，从而达到降维的目的。现广泛应用于视频背景建模，光度立体重建等领域。

## 1.3 本文主要内容

低秩矩阵填充是近年来图像处理、机器学习等领域的研究热点，在实际高维数据分析中具有成功的应用。本文综述了低秩矩阵填充的理论发展，以及多年来研究中的模型变化以及等价的条件证明，阐述了低秩矩阵填充问题的实际意义。在第二章给出了对于矩阵的秩函数最小化的凸松弛和非凸松弛两类模型的经典算法，包括凸松弛模型的奇异值阈值算法（SVT）,加速近端梯度法（APG）,增广Lagrange乘子法（ALM）和非凸松弛的加权核范数（WMMN）模型，并分析说明了不同算法针对不同的模型有着各自的计算优势，由于凸松弛模型对原秩函数进行了过多松弛使得对实际问题的近似不够准确，因此研究非凸松弛模型以提高准确率。

在本文的第三章，启发于信号处理领域的压缩感知原理，得知利用$L_{\frac{1}{2}}$正则子可以得出更稀疏的解，将一维向量的稀疏性扩展到二维矩阵的低秩性，依据徐宗本[10]提出的矩阵$S_{\frac{1}{2}}$范数，将原目标函数进行非凸松弛，提出基于$S_{\frac{1}{2}}$建模的矩阵秩最小化模型、低秩- 稀疏分解模型，并根据模型的形式利用加速近端梯度算法结合加权核范数的思想来进行求解。然而由于新模型中目标函数的非凸性，使得求解变得复杂，所以利用半阈值算子转化为凸松弛的算法求解。在第四章中的数值模拟中，证实了新模型算法的精确性更好，并对小规模的矩阵及秩更低的矩阵有更好的计算精度与效率。最后总结分析了本文新模型的优点与不足，并对低秩矩阵填充问题的算法和应用发展作出展望。

本文的创新之处在于：一、为了更好地诱导低秩矩阵填充问题中矩阵的低秩性，结合加权核范数思想与$L_{\frac{1}{2}}$正则化方法，提出了基于矩阵$S_{\frac{1}{2}}$范数建模的矩阵秩最小化模型、低秩- 稀疏分解模型，分析了利用APG算法求解迭代的收敛性；二、利用加速近端梯度算法求解基于$S_{\frac{1}{2}}$建模的矩阵秩最小化和低秩- 稀疏分解问题，并在数值实验中验证了新算法对比原APG算法具有更好的计算精确性。