河北经贸大学出版基金资助
河北省重点学科应用统计学资助
河北省自然科学基金项目（项目编号：15215327）资助

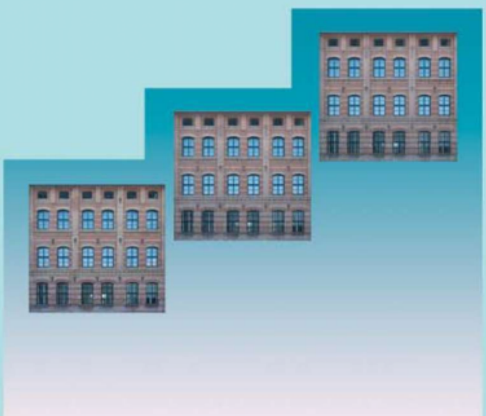
低秩矩阵与张量完整化问题的算法研究

耿 娟 著

Algorithms for Low-rank Matrix and Tensor Completion Problems

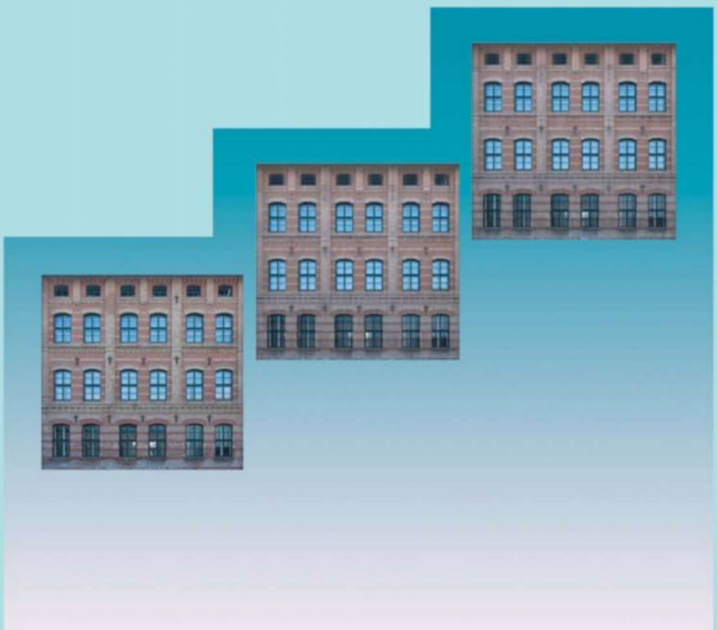


河北科学技术出版社



耿娟，中国农业大学管理学博士，博士期间参与导师主持的国家自然科学基金项目“低秩矩阵恢复算法及其在图像处理中的应用”的研究。现任河北经贸大学数学与统计学学院教师。近年来讲授的课程主要有数据挖掘、多元统计分析、计量经济学、金融计量经济学等。主持两项省级自然科学基金项目，主要研究领域为矩阵与张量优化方法与应用、数据挖掘方法等。近年来在《Numerical Algorithms》《International Journal of Machine Learning and Cybernetics》《International Journal of Signal Processing, Image Processing & Pattern Recognition》等国外三大索引的权威杂志发表论文5篇。

责任编辑 韩 翊 封面设计 东 升 美术编辑 刘英君



图书在版编目（CIP）数据

低秩矩阵与张量完整化问题的算法研究／耿娟著
. -- 石家庄：河北科学技术出版社，2016.6
ISBN 978-7-5375-8370-1

Ⅰ．①低…　Ⅱ．①耿…　Ⅲ．①矩阵-算法-研究②张量分析-算法-研究　Ⅳ．①O151.21②O183.2

中国版本图书馆 CIP 数据核字（2016）第 110615 号

低秩矩阵与张量完整化问题的算法研究

耿　娟　著

出版发行　河北科学技术出版社
地　　址　石家庄市友谊北大街 330 号（邮编：050061）
印　　刷　石家庄燕赵创新印刷有限公司
开　　本　787×1092　1/16
印　　张　10.5
字　　数　250 千字
版　　次　2016 年 7 月第 1 版
　　　　　2016 年 7 月第 1 次印刷
定　　价　32.00 元

前　言

随着信息技术和信息产业的迅猛发展,数字通信技术和计算机网络在社会各个领域都有广泛的应用,随之而来的是各种纷繁复杂的数据以及如何对这些数据进行分析和处理. 在许多实际的应用中,经常会遇到数据缺失或污染的问题,例如著名的 Netflix 问题. Netflix 公司是一家在线影片租赁提供商,他利用顾客对其观看过的 DVD 之评分来推测顾客的喜好从而进行影片推荐. 如果将每位顾客对其观看过电影所打的分数作为行,每部电影获得不同顾客对它的评分作为列,就形成了一个由观众评分构成的矩阵,但这个矩阵是不完整的,原因是每位顾客只对很少的几部影片进行了评分. 接下来的问题就是如何将这样一个含有缺失值的矩阵补充完整,进而预测顾客的喜好. 这看似是一件不可能的事情,但我们一般认为只有少数因素影响顾客的偏好,这就导致这个评分矩阵是低秩或近似低秩的. 在这个低秩的前提下就可以通过一些合适的方法把矩阵的缺失部分补充完整,这就是近几年来备受人们关注的低秩矩阵完整化问题.

低秩矩阵完整化问题研究的是已知一个低秩矩阵的部分元素,如何将其他不知道的元素补充完整. 对于这个问题是否可解,很多人持怀疑的态度,Candès 等严格证明了大部分低秩矩阵都可以在已知少数元素的情况下,以很大的概率近乎完美地恢复出完整的矩阵,而这个恢复过程只需要求解一个简单的凸优化问题. 继这一突破性的工作之后,近几年来许多学者致力于这方面的研究,设计了各种巧妙、高效的算法来求解,本书是笔者在博士论文基础上加以补充完成,将详细介绍一些现有的理论和算法. 与本书的成果并驾齐驱的是低秩矩阵完整化的应用,目前,它已成功地应用于机器学

习、数据挖掘和人工智能等各个领域. 比如将带有混合噪声的视频去噪问题转换为矩阵完整化问题,再利用凸规划的方法来求解;矩阵完整化方法还可以解决在大量特征丢失情况下的多标签分类问题;除此之外,矩阵完整化方法在诸多问题中受到人们的重视,如维数约减、传感网络等.

当数据更复杂的时候,仅仅用矩阵来表示这些数据往往会丢失一些数据本身的结构信息,这时就需要借助张量来表示这种多目标之间的关系. 高阶张量作为向量(一阶张量)与矩阵(二阶张量)概念的推广在很多领域都有很重要的应用,如化学计量方法、心理测量学、信号处理等. 在高维的情形中很多问题都可以转化为求解一个可以很好的描述原始数据的低秩张量,而这一类问题就叫做低秩张量完整化问题. 近年来,这一问题也得到了广泛的应用,如在计算机视觉和图像处理等领域.

总之,在这个信息爆炸的时代,矩阵完整化与张量完整化方法已经成为人们处理复杂数据的新方法和途径,这些问题的研究能够推动机器学习与人工智能等多个领域的发展. 近年来,许多学者致力于建立合理的模型与快速有效的算法,但这并未减少大家对于此问题的研究热情. 针对已有模型和算法的优缺点进行分析和改进仍是一项有意义的研究内容. 本书主要针对矩阵完整化问题和张量完整化问题的模型建立和求解算法进行探讨,其主要内容和具体安排如下:

第一章主要介绍了矩阵完整化问题与张量完整化问题基本知识、各个模型建立的理论保证以及目前已有的一些模型和求解算法.

第二章针对以往求解核范数极小化模型中需要对正则化参数进行讨论这一特点,提出了指示函数与核范数的极小化模型. 该模型是一个无约束优化问题,其中不含正则化参数,其目标函数是矩阵的核范数与 $\mathbf{R}^m$ 中一个凸集的指示函数之和. 在求解模型之前,首先介绍了次微分、Moreau - Yosida 正则、迫近点映射与迫近算子等相关概念和性质. 随后,利用迫近算子推导出原问题最优解所满足的一组不动点方程,并设计出求解方程不动点的迭代算法. 接下来,分析提出算法的收敛性,并证明了当参数适当选取时,该算法能够收敛到原问题的极小解. 最后,对随机生成的人工数据和真实的图像

恢复问题进行实验,测试了提出算法的有效性与鲁棒性. 实验结果表明,新提出的模型和算法能够有效地求解矩阵完整化问题,并且恢复效果优于所比较的算法.

第三章用含有参数的双曲正切函数来逼近秩函数,从而代替秩函数来求解矩阵完整化问题. 选用双曲正切函数的优势在于它的可微性质,因此在求解模型时更加方便. 但需要注意的是模型中双曲正切函数的可微性是针对矩阵的奇异值向量的,而非针对矩阵本身,因此需要利用复合函数的微分得到函数关于矩阵本身的微分. 接下来利用函数的梯度方向作为下降方向设计了梯度投影算法来求解光滑函数模型. 大量的实验结果表明本章提出的模型和算法能够获得较好的恢复效果.

第四章基于核范数与秩函数之间的差别在于矩阵奇异值的大小,提出了加权核范数正则化模型. 权向量取为上一步迭代得到最优解的奇异值之倒数,这样有效地缩小了核范数与秩函数之间的差别. 但该模型是非凸优化模型,这给问题的求解带来了困难. 首先利用极大极小方法构造一个满足条件的并且比较容易求解的辅助函数,然后通过迭代地求解辅助函数极小化问题来得到原优化模型的最优解. 最后,利用大量的实验来说明算法的有效性,这些实验包括:对人工随机合成的矩阵进行测试,对真实的 Jester 数据进行测试和对真实图像恢复问题的测试.

第五章用非凸函数来代替秩函数建立了一个求解矩阵完整化问题的非凸模型框架,在这个框架下,前面第三、四章以及一些已有的非凸模型都可以归到所提出的框架中. 由于模型属于非凸优化问题,因此采用求解非凸问题的 DC 规划与 DC 算法来求解. 此外,本章中还提出了另外一个指数型的函数来代替秩函数求解矩阵完整化问题. 数值实验中,将几个例子连同指数型函数模型进行对比,对人工合成矩阵、真实世界的矩阵完整化问题和图像恢复问题进行实验,数值结果表明了指数型函数模型的恢复效果优于其他算法.

第六章主要考虑的是硬阈值类的算法. 以往的硬阈值类算法用的都是梯度投影方法,利用梯度方向作为搜索方向,也有许多文献在步长的选取方

面做工作,本章利用半迭代方法改进了搜索方向,使算法具有更快的收敛速度. 同样对人工随机合成的矩阵与真实的图像恢复问题来测试算法的恢复效果. 实验结果表明,半迭代硬阈值算法比未改进下降方向的算法收敛速度要快很多倍.

第七章主要探讨张量完整化问题,由于它是矩阵完整化问题的直接推广,因此也备受人们关注. 已有不少文献将求解矩阵完整化问题的凸松弛模型推广到张量的情况,但对于非凸类型的模型在张量情况下的讨论还比较缺乏. 本章将加权核范数模型推广到张量的问题中,并利用极大极小方法来求解得到了求解张量完整化问题的极大极小加权软阈值算法. 实验仍然对人工随机合成的张量和对真实图像以及 MRI 图像的恢复问题进行测试,得到了很好的恢复效果.

第八章主要研究硬阈值算法在张量完整化问题中的应用. Rauhut 等将硬阈值算子引入到张量的问题中,提出了求解张量完整化问题的硬阈值方法. 本章将此方法中的下降方向用半迭代方法进行改进,得到了求解张量完整化问题的半迭代硬阈值算法. 在实验部分,仍然对人工随机合成的张量和真实图像进行测试,将本章提出的算法与未改进的 TIHT 算法比较,同时也与两种基于软阈值算子的方法 FP - LATC 和 HaLRTC 进行比较,实验结果表明,改进后的算法在计算精度与计算时间方面都有很大提升.

第九章主要针对计算中的关键问题,设计和实现了几种求解张量完整化问题的求解算法并进行比较,这些算法包括 TC - SV 算法、TC - MM 算法和 TC - MWST 算法. 对人工合成张量和真实图像恢复问题的实验结果表明这三种算法在计算精度和计算速度上都有很大的优势.

第十章总结了本书的主要工作并提出了以后的研究方向.

本书是以作者在博士期间的研究成果为中心展开讨论的. 我们默认读者具有最优化理论的基础知识,并有兴趣了解矩阵和张量完整化问题的理论和应用. 所以,针对上述读者群,本书的特点如下:

1. 依据最优化理论和方法,对书中的部分算法进行收敛性分析

尽管本书的研究重点是相关问题的模型建立和求解算法的设计,但为

了使研究结果更加全面,我们对部分算法进行了收敛性分析并进行了严格的数学证明.

2. 在每章的最后一部分,加入了算法的应用

每章的大概结构是问题引入、模型的提出、相应算法的设计和算法的效果检验. 在最后一部分每种算法的实现中加入了真实的数据和图像处理问题,目的有两方面,一是能够使读者看到研究算法的精度和速度,另一方面是给出一些具体的应用.

本书的出版获得了河北经贸大学出版基金资助、河北省重点学科应用统计学资助. 在此一并致谢! 另外,在本书出版之际,首先要感谢中国农业大学的王来生教授,在本书的研究工作中给予指导,并花费了大量的时间仔细审阅和修改了全书的理论和方法部分. 还要感谢与作者一起读博士的各位同学,给作者的研究提出了许多宝贵的意见,使得本书的质量得以保证.

本书在撰写过程中,参阅了大量国内外相关文献,已在本书之后予以注明,谨在此对这些文献的作者表示衷心感谢!

由于我的水平有限,错误和不当之处在所难免,诚恳地欢迎同行、专家、前辈和读者批评指正,并提出宝贵的意见.

耿　娟

2016 年 6 月

Preface

With the rapid development of information society, digital communication technology and computer network has been widely used in various fields of the society. Accordingly, there are all sorts of diverse and unpredictable data and how to carry on the analysis and processing of these data is the first problem. In many practical applications, we often encounter the problem of missing data, such as the famous Netflix problem. Netflix is an online movie rental provider. He used the score which is the customers' rating for the DVD they have seen to predict customers' preferences and then recommending the film to the customer accurately. We can form a matrix where the rows index each individual and the columns index the questions. We collect data to fill out this table but unfortunately, many questions are left unanswered. The question, then, is how to make an educated guess about what the missing answers should be. This seems to be an impossible job. However, in generally, the data matrix of all user – ratings may be approximately low – rank because it is commonly believed that only a few factors contribute to an individual's tastes or preferences. In this case, we can recover the matrix from only a few of elements through some suitable methods. This problem is the now famous low – rank matrix completion problem.

The low – rank matrix completion problem involves recovering an unknown low rank or approximately low rank matrix from very limited number of known entries. A lot of people have much skepticism about whether this problem can be solved. Candès prove strictly that one can perfectly recover most low – rank matrices from what appears to be an incomplete set of entries by solving a simple convex optimization program. Following this breakthrough work, many scholars have devoted themselves to the research of this area in the past few years. In the first chapter, we will introduce the existing theories and algorithms in detail.

Together with the development of the theoretical results above, the matrix completion problem has been successfully applied in many fields such as machine learning, data mining and artificial intelligence. For example, we can convert the video denoising problem with mixed noises to the matrix completion problem and then using convex programming method to solve it. Matrix completion method can also solve the problem of multi label classification in a large number of missing cases. In addition, matrix completion method has been paid more attention to in many problems, such as dimension reduction, sensing network and so on.

When the data is more complex, just use the matrix to indicate these data tend to lose some structural information of the data itself, then we need to use the tensor to represent the relationship between this multi objective. High order tensor as the extension of the concept of vector (1 – order tensor) and matrix (2 – order tensor) has very important applications in many fields, such as chemical measurement method, psychological measurement, signal processing and so on. We find that, in the case of high dimension, a lot of problems can be translated into a low rank tensor which can describe the original data well. We call this kind of problem the low – rank tensor completion problem. In recent years, this problem has been widely used, such as in the field of computer vision and image processing.

Anyway, in the information explosion era, matrix completion and tensor completion method has become the new approach and way for people to deal with complex data. Research on these problems can promote the development of many fields such as machine learning and artificial intelligence. In recent years, although many scholars are committed to the establishment of a reasonable model and fast and effective algorithm, but this has not reduced the enthusiasm for the study of this problem. The analysis and improvement of the advantages and disadvantages of the existing models and algorithms is still a significant research content. The aim of this book is to discuss the establishment of the model and the solving algorithm for the matrix completion and tensor completion problem. Its main contents and specific arrangements are as follows.

In the first chapter, we mainly introduce the basic concepts and knowledge, the theoretical guarantee of each model and some existing models and algorithms.

When solving the nuclear norm minimization model, we often need to discuss the regularization parameters. To avoid this, in the second chapter we propose a new model for matrix completion problem based on nuclear norm and indicator function. This model is an unconstraint optimization problem, which does not contain the regularization parameter and the objective function of which is a sum of nuclear norm of the matrix and an indicator function on a convex set in. Firstly, we introduce the subdifferential, Moreau – Yosida regularization, proximal point mapping and proximity operator and other related concepts and properties. Subsequently, we obtain a set of fixed point equations by using the proximity operator, the optimal solution of which is the same with the original problem and then we design a proximal point algorithm to solve it. Then the convergence of our algorithm is established strictly. We prove when the parameters are suitable the algorithm can converge to the optimal solution of the original problem. Finally, we test the effectiveness and robustness of the proposed algorithm by experiment on randomly generated artificial data and real image recovery problems. The numerical results suggest that significant improvement can be achieved by our algorithm compared to the other reported methods.

In the third chapter, we use a smoothed function—Hyperbolic Tangent function to approximate the rank function, and then with the smoothed function to solve the matrix completion problem. The advantage of the hyperbolic tangent function is its differential property, and this makes the method more convenient to solve the model. But the need to pay attention to is that the differential property of the hyperbolic tangent function is not for the matrix but the singular vector of the matrix. Therefore, it is needed to use the differential of the composite function to get the function of the differential of the matrix itself. Next, we design the gradient projection algorithm to solve the smooth function model by the gradient direction of the function as the descent direction. A large number of experimental results show that the proposed model and algorithm can obtain a better recovery effect.

In the fourth chapter, in order to decrease the gap between the nuclear norm and the rank function, we add weight vector in the nuclear norm and build a weighted nuclear norm regularization model. The weight vector is the inverse of the singular value of the optimal solution is obtained in the last iteration, by this the gap between the nuclear norm and the rank function is reduced effectively. But the model is a non - convex optimization model, which brings difficulties to the solution of the problem. We first construct an auxiliary function which satisfies the condition and can be easily solved by using the majorization - minimization approach, and then by solving the auxiliary function minimization problem iteratively to get the optimal solution of the original optimization model. Finally, we use a large number of experiments to illustrate the effectiveness of the algorithm, including the test of the matrices generated randomly, the real Jester data image recovery problems.

In the fifth chapter, we propose a unified non - convex model framework for matrix completion by using non - convex function to replace the rank function. Under the framework, many non - convex models and some previously mentioned models can be obtained as special cases of the general framework. Because the model is the non - convex optimization problem, we use Difference of Convex functions (DC) programming and DC Algorithms (DCA) to solve the model. In addition, we give a new non - convex function—exponential type function to instead of rank function. We make some numerical comparison between our algorithms and the state – of – the – art method on randomly generated matrices, real data and image recovery problems. The results suggest that exponential type function model is more effective and promising.

In the sixth chapter, we mainly consider algorithm based on the hard threshold. Most algorithms about hard threshold use the negative gradient direction of the object function as search direction. There are some work about the selection of step size. In this chapter, we improve the hard threshold algorithm though the semi – iterative method. Our proposed algorithm has faster convergence rate. In the same way as in the previous chapters, we make some numerical comparison on randomly generated matrices and real image recovery problems. Compared to other iterative shrinkage algorithms, the performances of our proposed algorithm show a clear improvement in computation time.

In the seventh chapter, we mainly research the tensor completion problem. This problem has received a lot of attention, because it is a direct generalization of the matrix completion problem. At present, there are some work which generalize the convex relaxation model for matrix completion to solve the problem of tensor. But there is a lack of discussion about the non – convex model of tensor completion. In this chapter, we introduce the weighted nuclear norm for tensor and develop majorization – minimization weighted soft threshold algorithm to solve tensor completion problems. We still test the recovery effect for the artificial synthesis tensor and real images and MRI images. The experiment results demonstrate the effectiveness of the improved algorithm.

In the eighth chapter, we mainly research the iterative hard threshold algorithm for tensor completion problem. Rauhut introduced the hard threshold operators into the tensor completion problem, and proposed a hard threshold method for tensor completion. We improve the descent direction by semi – iterative method in the hard threshold method, and obtain a semi – iterative hard threshold algorithm for the tensor completion problem. In the experimental part, we still test the algorithms on the artificial synthesis tensor and real images. We compare the proposed algorithm with the original TIHT algorithm and FP – LATC and HaLRTC which are based on soft threshold operators. The experimental results show that the improved algorithm has a great improvement in terms of computational accuracy and computation time.

The ninth chapter, we mainly focus on the key problems in the calculation. We do some work about the design and implementation of several algorithms to solve the tensor completion problem, and then compare them to each other. The algorithms involved in this chapter include TC – SV algorithm, TC – MM algorithm and TC – MWST algorithm. Experimental results on the synthetic tensor and the real image restoration show that these three algorithms have great advantages in both computational accuracy and computational speed.

In the final chapter, we summarize the main work of this book and put forward the future research direction.

This book is mainly based on the author's research results in the period of doctor. We assume that the reader has a basic knowledge of optimization theory, and is interested in the theory and application of matrix and tensor completion problem. Therefore, in view of the above reader group, the characteristics of this book are as follows:

1. According to the theory and method of optimization, the convergence analysis of some algorithms in the book is carried out.

Although our research is focused on issues related to the modeling and algorithm, but in order to make the research more comprehensive, in the part of the algorithm, we analyze the convergence of the algorithm and make a rigorous mathematical proof.

2. In the last part of each chapter, we introduce the application of the algorithm.

The structure of each chapter is the introduction of the problem, the model proposed, the corresponding algorithm and the application effect of algorithm. In the last part, we add the real world data and image processing problems in the numerical experiments. There are two purposes for this, one is to enable readers to see the accuracy and speed of the study algorithm, the other is to give some specific applications.

The publication of this book received the support of the publishing fund of Hebei University of Economics and Business and the key subject applied statistics foundation of Hebei province, here together to thank! In the occasion of the publication of this book, first want to thank Laisheng Wang professor of China Agricultural University, who gave guidance to the research work of this book, and spent a lot of time to carefully review and revise the book's theory and methods. We would also like to thank my doctoral students, they provide a lot of valuable advice, which ensures the quality of the book. Finally, we will thank to the editors of Hebei Science and Technology Publishing House, they have a serious attitude and high quality work efficiency, and left a deep impression on me. We dedicate this book to all those who have given us support and help.

During the writing process, the author has read a large number of domestic and foreign related literature, which has been indicated in the back of this book. I would like to express my heartfelt thanks to the authors of these papers.

Finally, it should be pointed out that due to our limited level, the wrong or inappropriate place is inevitable, so we sincerely welcome peer experts and readers to criticize, and put forward valuable suggestions.

目 录