ON LOW-RANK PLUS SPARSE MATRIX SENSING

SIMON VARY
Exeter College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
January 2021

pre Mamu

PREFACE

This thesis comprises research undertaken at the Mathematical Institute, University of Oxford between October 2016 and December 2020 under the supervision of Prof. Jared Tanner. The research involves collaboration with Leonardo MW Ltd. and was funded by the EPSRC I-CASE studentship in partnership with our collaborators and by the Alan Turing Institute.

No part of the thesis has been accepted or is currently being submitted for any degree, diploma, certificate or other qualification, apart from the degree of Doctor of Philosophy at the University of Oxford. I also confirm that the thesis I am submitting is wholly my own work.

An article based on the content of Chapter 2 of this thesis, apart from §2.7 and §2.9, has been published in the *SIAM Journal on Mathematics of Data Science* , co-authored by J. Tanner and A. Thompson.

Content that forms Chapters 3 and 4 is under review in the journal of *Applied and Computational Harmonic Analysis*, co-authored by J. Tanner.

Chapter 5 is based on the content of an article that was published as part of proceedings of the *IEEE Data Science Workshop*, co-authored by A. Giancarlo, D. Humphreys, R. A. Lamb, J. Piper, and J. Tanner.

Oxford, January 2021

The up-to-date version of the thesis can be found at simonvary.github.io/thesis.pdf.

ACKNOWLEDGEMENTS

First and foremost, I am deeply grateful to my supervisor, Jared Tanner, for his mentorship and his thoughtful guidance of my first steps in academia. Thank you for showing me how to approach unanswered questions and how to seek new ones worth answering. I am also greatly indebted to Andrew Thompson, who generously gave me so much of his time, especially at the beginning of the project, and also for instructing me how to write. The profound care and respect they have for their professional craft has been inspiring.

I have also very much enjoyed the collaboration with Leonardo MW Ltd, thanks to my industrial supervisors Robert Lamb and David Humphreys. Their unparalleled insight into optics and computational spectral imaging was both fascinating and integral to the applied part of the thesis.

I am thankful to Mike Davies and Yuji Nakatsukasa, who acted as the examiners for my viva, and whose careful reading of the work led to its great improvement.

I gratefully acknowledge the Alan Turing Institute for the multiple research and travel grants.

I am incredibly thankful for having shared the journey of doctoral studies with my office mates, Abi, Bogdan, Julien, and Thomas, in the S2.33 Matlab fanbase, and also with my good friends Yuki and Clint. I am also very grateful for friendships in Bratislava, in particular, my thanks goes to Maťo, Jakub, Lukáš, Matúš, Michal, and Marek, in the Fire & Blood group. I extend my thanks also to Mišo, Lidka, Myko, and Marko, for their continuing support and inspiration throughout many years. I wish to thank Oto Grošek for encouraging me to pursue mathematics since my childhood. The value of friendships cannot be overstated—they gave me stability in turbulent times and the comfort of feeling at home while living between places.

Finally, I have to thank my parents, Maroš and Barbora, for their selfless love, unending support, and many personal sacrifices through which I was able to grow at this incredible place. My sister Žofka has been always there for me, whether to offer an opinion on writing or to give words of encouragement. Your strength and kindness has become an inspiration.

ABSTRACT

Expressing a matrix as the sum of a low-rank matrix plus a sparse matrix is a flexible model capturing global and local features in data, and is the foundation of robust principal component analysis (Candès et al., 2011). This thesis is concerned with *low-rank plus sparse matrix sensing*—the problem of recovering a matrix that is formed as the sum of a low-rank and a sparse matrix, and the two components, from a number of measurements far smaller than the dimensionality of the matrix.

It is well-known, that inverse problems over low-rank matrices, such as robust principal component analysis and matrix completion, require a low coherence between the low-rank matrix and the canonical basis. However, in this thesis, we demonstrate that the well-posedness issue is even more fundamental; in some cases, both robust principal component analysis and matrix completion can fail to have any solutions due to the fact that the set of low-rank plus sparse matrices is not closed. As a consequence, the lower restricted isometry constants (RICs) cannot be upper bounded for some low-rank plus sparse matrices unless further restrictions are imposed on the constituents. We close the set of low-rank plus sparse matrices by posing an additional bound on the Frobenius norm of the low-rank component, and ensure the optimisation is well-posed and that the RICs can be bounded. We show that constraining the incoherence of the low-rank component also closes the set provided $\mu < \sqrt{mn}/\left(r\sqrt{s}\right)$ and satisfies a certain additivity property necessary for the analysis of recovery algorithms.

Compressed sensing, matrix completion, and their variants have established that data satisfying low complexity models can be efficiently measured and recovered from a number of measurements proportional to the model complexity rather than the ambient dimension (Foucart and Rauhut, 2013). This thesis develops similar guarantees showing that $m \times n$ matrices that can be expressed as the sum of a rank-r matrix and a s-sparse matrix can be recovered by computationally tractable methods from $O(r(m+n-r)+s)\log(mn/s)$ linear measurements. More specifically, we establish that the RICs for the aforementioned matrices remain bounded independent of problem size provided p/mn, s/p, and $r(m+n-r)/p$ remain fixed. Additionally, we show that semidefinite programming and two hard threshold gradient descent algorithms, NIHT and NAHT, converge to the measured matrix provided the measurement operator's RICs are sufficiently small. The convex relaxation and NAHT also provably solve Robust PCA with the optimal order of the number of corruptions $s = O\left(mn/\left(\mu^2 r^2\right)\right)$. Numerical experiments illustrating these results are shown for synthetic problems, dynamic-foreground/static-background separation, and multispectral imaging.

CONTENTS

1 INTRODUCTION 1
- 1.1 Motivation 1
- 1.2 Problem description and scope 8
- 1.3 Objectives and structure of the thesis 11

2 MATRIX RIGIDITY & THE ILL-POSEDNESS OF MATRIX RECOVERY 14
- 2.1 Introduction 14
- 2.2 Incoherence and other matrix recovery assumptions 16
- 2.3 Simple example of the lack of existence 18
- 2.4 Matrix rigidity is not lower-semicontinuous 19
- 2.5 The set of low-rank plus sparse matrices is not closed 20
- 2.6 Numerical examples of divergent matrix recovery 29
- 2.7 Closing the set of low-rank plus sparse matrices 32
- 2.8 Summary and discussion 36
- 2.9 Supporting lemmata 37

3 RESTRICTED ISOMETRY CONSTANTS FOR THE SET OF LOW-RANK PLUS SPARSE MATRICES 40
- 3.1 Introduction 40
- 3.2 Nearly isometrically distributed maps 42
- 3.3 Connection to the Johnson-Lindenstrauss lemma 43
- 3.4 Restricted isometry constants for the LS set 44
- 3.5 Summary and discussion 50
- 3.6 Supporting lemmata 51

4 ALGORITHMS FOR LOW-RANK PLUS SPARSE MATRIX SENSING 57
- 4.1 Introduction 57
- 4.2 Relation to prior algorithmic work 60
- 4.3 Recovery by convex relaxation 64
- 4.4 Recovery by non-convex algorithms 69
- 4.5 Empirical average case performance 77
- 4.6 Applications 82
- 4.7 Summary and discussion 85
- 4.8 Supporting lemmata 86

5 LOW-RANK MODELS FOR MULTISPECTRAL IMAGING 92
- 5.1 Introduction 92
- 5.2 Direct interpolation methods 93

5.3 Sparse approximation inpainting 94
5.4 Low-rank matrix completion inpainting 95
5.5 Numerical simulations 95
5.6 Summary and discussion 100

6 SUMMARY & FINAL REMARKS 101
6.1 Summary of main results 101
6.2 Open problems and future work 103

BIBLIOGRAPHY 105

1 INTRODUCTION

Parsimony concepts of low-rank and sparsity have a wide range of applications in mathematical modelling, statistics, and computation. Their combined additive structure forms a flexible model capturing global and local features in data, and is the foundation of robust principal component analysis. This thesis is dedicated to the study of mathematical theory and algorithms that allow for exact recovery of low-rank plus sparse matrices in an information restrained setting.

1.1 MOTIVATION

More than half a century ago, Tukey (1962) coined the term *data analysis*, and with it, made a case for a new scientific discipline in a practical pursuit of learning from data. Like statistics, this new science also seeks to infer from the particular to the general, but unlike statistics, data analysis also includes, among other things: techniques for interpreting data, ways of planning gathering of data, and all the machinery needed for analysing data.

Since Tukey's paper, there have been major advances in information theory, digitisation and computing. The last fifty years have witnessed an explosion of digital devices and sensors that generate an unprecedented amount of information. It is estimated that in 2020 there were around 30 billion devices communicating through the internet—a figure projected to double in the next few years (Nordrum, 2016).

Hand in hand with the scale comes the immense cost of acquiring, storing, processing and transmission of information in ever-increasing volumes. In order to perform such tasks, it is necessary to perform computation over the data and analyse it in an interpretable way.

The large scale pushes the traditional frame of thought in computing and statistics. Optimisation problems arising in data-oriented applications often have millions of parameters and require billions of floating-point operations. Modern statistics has to deal with high-dimensional datasets, i.e. when the number of samples might be lower than its dimension, and to explain the generalisation of highly overparameterised models.

Fortunately, the actual information content in datasets is often far lower than the ambient dimension would suggest. This principle of a latent simpler structure is essential and is most easily apparent in the success of *compression*—a process which exploits redundancies in data in the search of simpler representations of datasets. Hidden simplicities of data allow for fast transfer across the network and lower storage requirements of many compressed data formats.

However, the idea of detecting simple structures in datasets has far-reaching implications beyond just compression: it offers advantages in *interpretability*, *cheaper acquisition*, and *faster computation*.

(i) Interpretability based on sparsity and rank is the core principle of statistical techniques such as LASSO estimation, which finds a selection of important factors out of many (Tibshirani, 1996), or PCA, which finds a lower-dimensional subspace that preserves the maximal amount of statistical variance (Pearson, 1901; Hotelling, 1933). Additionally, the simpler model is often more correct. The principle of *Occam's razor* tells us that: "Among competing explanations for a phenomenon, the simplest one is the best", and in mathematics, simplicity is often expressed through the notion of sparsity or low-rank.

(ii) Acquisition of data can be prohibitively expensive in many applications, either in terms of time or money. Moreover, often the data is compressed right after the acquisition, i.e. simplified using sparsity or low-rank, to be subsequently stored or transmitted. *Compressed sensing* is a novel sampling paradigm, which states that it is sufficient to take a fewer number of measurements proportional to the complexity of the data in its compressed form provided the signal comes from a low-order model, e.g. it has a sparse (Candès and Recht, 2009) or a low-rank representation (Recht et al., 2010).

(iii) Computation required for modelling of large physical systems or optimisation problems in engineering sciences can be stupendously costly in terms of time and hardware. Many of the successful numerical algorithms are based on clever tricks utilising a low-order structure, such as low-rank or sparsity, arising in such problems. For example, Greengard and Rokhlin (1997) exploit the fact that matrices that appear in the simulation of interacting particle systems are well approximated by a hierarchical low-rank matrices leading to a method with a linear instead of a quadratic complexity. Halko et al. (2011) design an algorithm that can significantly speed up computing singular value decomposition through randomised sketching if used on matrices approximable by low-rank matrices. Burer and Monteiro (2003) use a low-rank factorisation for parameterisation of semidefinite programming optimisation making the computation manageable even for large problem sizes. The list goes on.

Interpretability, computation, the information content in data, and the interplay between those, are at the heart of data analysis, and thus, are at the centre of the focus of this thesis.

It will be necessary to evaluate algorithms based on how much resources they require, in terms of computation and information. To this end, we will

employ the notion of the *computational complexity* and the *sample complexity* of an algorithm. While the former reflects the amount of resources needed for an algorithm to run, the latter refers to the amount of information, i.e. samples, it requires. The computational complexity can be further broken down into the *time complexity*, which refers to the number of elementary operations required for a specific task, and the *space complexity*, which reflects the maximal amount of memory needed at any single time.

In the remaining of this section, we review two major applications of sparsity and of low-rank, *compressed sensing* and *principal component analysis* (PCA). In §1.2, we describe the common meeting point of the two, and thus, delineate the topic addressed by this work. The chapter is concluded by §1.3, where we lay out the central objectives and the structure of the rest of the thesis.

1.1.1 Sparse approximation and compressed sensing

The classical result of Shannon (1948), which laid the foundations of information theory, is the observation that continuous signals whose highest frequency is upper bounded, can be exactly represented digitally using finitely many bits if sampled at the Nyquist-Shannon rate: the number of samples must be proportional to twice the highest frequency contained in the signal.

Nowadays, we frequently encounter sensors generating high-dimensional data that make achieving the Nyquist-Shannon rate difficult or entirely infeasible. Compressed sensing is a novel sampling paradigm, which goes against the common practice in data acquisition, and states that it is possible to lower the sampling rate provided the signal contains redundancies. While Shannon (1948) defined the signal complexity based on its maximal frequency, in compressed sensing, the complexity is expressed through the notion of sparsity, e.g. in the frequency domain of the signal.

Sparse data models were being successfully implemented long before the theory of compressed sensing, for example in the context of subset selection (Garside, 1965), seismology (Levy and Fullagar, 1981; Santosa and Symes, 1986), or medical ultrasound (Papoulis and Chamzas, 1979). In statistics, sparse approximation was proposed for the overdetermined case of least squares regression by Tibshirani (1996) to perform both variable selection and regularization in order to enhance the prediction accuracy and interpretability and is referred to as Least Absolute Shrinkage and Selection (LASSO). In signal processing, Mallat and Zhang (1993) defined the term *dictionary* as an overcomplete set of *atoms* whose linear combinations can be used to represent signals, but can cause difficulties in terms of non-uniqueness of decomposition and overfitting. To resolve such issues, Mallat and Zhang (1993) also introduced the *Matching Pursuit* algorithm which finds the sparsest representation through an iterative greedy process that selects the best

matching elements in the dictionary to the given signal.

At its core, compressed sensing seeks a sparse solution to an underdetermined system of linear equations

$$Ax = b, \qquad \text{s.t.} \quad \|x\|_0 \leq k, \tag{1.1}$$

where $A \in \mathbb{R}^{p \times n}$ is the measurement operator with $n > p$, $b \in \mathbb{R}^p$ is the vector of measurements, and $x \in \mathbb{R}^n$ is the signal we wish to recover. The $\|x\|_0$ is the ℓ_0-norm which gives the number of non-zero entries of x. Solving the *minimum set cover* problem can be reduced to that of finding a solution to (1.1), and therefore, solving the problem in (1.1) is NP-hard in general (Foucart and Rauhut, 2013, §2.3).

The seminal work of compressed sensing began with Candès and Tao (2006), Candès et al. (2006a) and Donoho (2006a), who showed that the optimal s-sparse representation of a signal can be efficiently recovered via

$$\min_{x \in \mathbb{R}^n} \|x\|_1, \qquad \text{s.t.} \quad b = Ax, \tag{1.2}$$

from p measurements $b \in \mathbb{R}^p$ taken by a linear subsampling operator $A \in \mathbb{R}^{p \times n}$ provided $p = O(s \log n)$, and $\|x\|_1$ is the ℓ_1-norm which is computed as the sum of absolute values of x. This is referred to as *Basis Pursuit* and can be solved by linear programming.

Candès and Tao (2005) and Donoho (2006b) were able to lower the suboptimal $\log(n)$-factor for matrices $A \in \mathbb{R}^{m \times n}$ sampled from the uniform spherical ensemble and the random Gaussian ensemble

$$p = O\left(s \log\left(\frac{n}{s}\right)\right). \tag{1.3}$$

This is a significant improvement because it asymptotically reduces to the optimal rate $p = O(s)$ as

$$\frac{p}{n} \to \delta \in (0,1) \quad \text{and} \quad \frac{s}{p} \to \rho \in (0,1) \quad \text{as} \quad s, p, n \to \infty, \tag{1.4}$$

where the constant δ is known as the *undersampling ratio* and corresponds to the number of samples and the constant ρ is the *sparsity ratio* and expresses how sparse is the signal.

An important part of the proof by Candès and Tao (2005) are the *restricted isometry constants* (RICs)

$$(1 - \Delta)\|x\|_2 \leq \|Ax\|_2 \leq (1 + \Delta)\|x\|_2, \qquad (\forall x) : \|x\|_0 \leq s, \tag{1.5}$$

which control the degree to which the linear mapping $A \in \mathbb{R}^{p \times n}$ acts as an approximate isometry when restricted to the set of sparse vectors x. It can be shown that the measurement operators obeying concentration of measure inequalities have their RICs upper bounded (Baraniuk et al., 2008).

Candès and Tao (2006) proved that RICs of Gaussian matrices satisfy $\Delta \leq 0.6246$, which in turn guarantees the ℓ_0/ℓ_1 equivalence in the sense that

the solution of (1.2) solves also the sparsity constrained problem in (1.1). Moreover, the recovery by ℓ_1-norm minimisation can be also extended to the case when there is an additive noise and/or model mismatch, i.e. the signal is not exactly sparse (Candès et al., 2006a). Another important result is that Gaussian matrices can be multiplied with a fixed ortohogonal matrix $Q \in \mathbb{R}^{n\times n}$ and $A \in \mathbb{R}^{p\times n}$ with $p = O\left(s \log\left(\frac{n}{s}\right)\right)$, then $AQ \in \mathbb{R}^{p\times n}$ will also have a sufficiently small RICs (Foucart and Rauhut, 2013). This greatly generalises the results and enables recovery of signals that are sparse in any orthonormal basis Q.

The successes described above motivated the research into the limits of ℓ_1-norm minimisation. Donoho (2006c) gave a lower bound on the sparsity stating that the ℓ_0/ℓ_1 equivalence holds for A sampled uniformly from the Grassmannian manifold when $\rho < \rho_S(\Delta)$ with probability $1 - o(1)$ and $\rho < \rho_W(\Delta)$ with overwhelming probability. It was proved by Donoho and Tanner (2009b) that the bounds are precise, i.e. having $\rho > \rho_S(\Delta)$ or $\rho > \rho_W(\Delta)$ implies the ℓ_0/ℓ_1 equivalence fails to hold. As a consequence, there is an abrupt *phase transition* in which for a given subsampling ratio δ, the probabilty of a successfull recovery drastically changes from one to zero as ρ grows beyond $\rho_W(\delta)$ or $\rho_S(\delta)$. Donoho and Tanner (2009a) conjectured that a similarly abrupt phase transition in the ℓ_1-norm minimisation happens also for measurement operators sampled from other random matrix ensembles.

There is a wide range of iterative algorithms for solving compressed sensing problems which we review in §4.2.1. For a comprehensive overview of compressed sensing, see (Eldar and Kutyniok, 2012) and (Foucart and Rauhut, 2013).

1.1.2 *Low-rank approximation, PCA, and its extensions*

Large datasets are becoming progressively common and are often hard to interpret. In order to be able to analyse such data, methods are needed to reduce their dimensionality in an interpretable way. There now exist many techniques for the task, but principal component analysis (PCA) is one of the oldest and most widely used (Goodall and Jolliffe, 1988). The core idea of PCA is to reduce the dimensionality of a dataset, while preserving the maximal variance, i.e. statistical information, as possible.

The earliest work on PCA dates to Pearson (1901) and Hotelling (1933), but its widespread use began only with the advent of computing. Computation of PCA is equivalent to finding the closest rank-r approximation to a covariance matrix M of mean centered data in the Frobenius norm

$$\min_{X\in\mathbb{R}^{n\times n}} \|X - M\|_F, \qquad \text{s.t.} \quad \operatorname{rank}(X) \leq r. \tag{1.6}$$

For an early history of SVD computation, see (Stewart, 1993).

The minimisation in (1.6) has a closed-form solution that is by Eckart-Young-Mirsky theorem expressed through truncating the *singular value decomposition* (SVD). The development of SVD algorithms by Golub and Kahan

(1965) allowed PCA to come into popularity. Golub and Reinsch (1970) subsequently formulated the algorithm that have been the standard for computing SVD until now.

The standard algorithm for computing SVD has cubic time complexity, which is sufficient for moderately sized covariance matrices, but becomes prohibitively costly for analysing large datasets. Recent advances in randomised numerical methods led to fast sketching algorithms suitable for large problem sizes with good convergence properties when there exists a close enough low-rank approximation (Halko et al., 2011; Drineas et al., 2006; Woodruff, 2014). Randomised sketching methods compute only the k-leading singular values and their singular vectors with time complexity $O\left(n^2 \log(k) + nk^2\right)$.

Modern datasets are not only big, but often also messy; significant parts of the data can be corrupted or even missing. To account for this, over the last decade PCA has been extended to allow for missing data—*matrix completion*, subsampled measurement of data—*matrix sensing*, or data with few entries grossly corrupted or inconsistent with the low-rank model—*Robust PCA*.

Missing data: Matrix completion

Matrix completion can be equated with computing PCA with missing entries. It is the task of filling in the missing entries of a partially observed low-rank matrix

$$P_\Omega(X) = b, \qquad \text{s.t.} \quad \operatorname{rank}(X) \leq r, \tag{1.7}$$

where $b \in \mathbb{R}^p$ is a vector of p observed entries and $P_\Omega : \mathbb{R}^{n\times n} \to \mathbb{R}^p$ is an entry-wise subsampling operator that keeps entries of X at indices $\Omega \subset [n] \times [n]$.

This problem arises in many applications including collaborative filtering in recommender systems, e.g. for Netflix prize or MovieLens (Koren et al., 2009), traffic sensing (Du et al., 2015), multispectral imaging (Antonucci et al., 2019), integrated radar and communications (Liu et al., 2013), multi-task learning (Obozinski et al., 2010; Argyriou et al., 2008), and localization of Internet of Things (IoT) devices (Nguyen et al., 2019).

Although the rank minimisation is NP-hard in general (Harvey et al., 2006; Hardt et al., 2014), Candès and Recht (2009) were the first to show that, under certain conditions, the entries of X can be filled in exactly by solving an optimisiation problem that is a convex relaxation

$$\min_{X\in\mathbb{R}^{n\times n}} \|X\|_*, \qquad \text{s.t.} \quad P_\Omega(X) = b, \tag{1.8}$$

where $\|X\|_*$ is the sum of singular values of X, often referred to as the *nuclear norm* or the *Schatten*-1 *norm*. The sufficient conditions on the exact recovery by Candès and Recht (2009) are that the singular vectors of X are sufficiently spread out through the notion of *incoherence*, the indices of the observed entries are chosen uniformly at random, and $p \geq O\left(n^{1.2} r \log(n)\right)$. This result has been improved upon by Candès and Tao (2010) by lowering the sample complexity bound to $O\left(nr \log(n)\right)$ which is the optimal order of the sample complexity times a logarithmic factor.

Incoherence, regularisation and assumptions on matrix recovery are discussed in §2.2.

The convex relaxation can be readily solved by *semidefinite programming* (SDP), but has the time complexity $O(n^6)$ (Nesterov, 2004), making it infeasible even for moderately large matrices. This led to the development of a range of gradient descent optimisation techniques that have lower time complexity [see the discussion in §4.2.2] and non-convex optimisation methods that solve

$$\min_{X \in \mathbb{R}^{n \times n}} \|P_\Omega(X) - b\|_F, \qquad \text{s.t.} \quad \operatorname{rank}(X) \leq r. \tag{1.9}$$

Subsampled data: Matrix sensing

Matrix sensing is a generalisation of the matrix completion in the sense that the subsampling is applied through a general linear map $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$ and therefore it is not limited to an entry-wise sampling. Recht et al. (2010) proved an analogous result to compressed sensing, that a low-rank matrix X_0 can be recovered by a convex relaxation

$$\min_{X \in \mathbb{R}^{m \times n}} \|X\|_*, \qquad \text{s.t.} \quad \mathcal{A}(X) = b, \tag{1.10}$$

provided $\mathcal{A}(\cdot)$ has its RICs in respect to the set of low-rank matrices sufficiently bounded. Noteable difference between provable recovery in matrix sensing and matrix completion is that the former does not require the singular vectors to be sufficiently spread out in terms of coherence.

Grossly corrupted data: Robust PCA

Robust PCA is a related problem that arises when we are presented a covariance matrix with some of its entries corrupted, but without the knowledge of their locations. The goal is to find an additive decomposition of a covariance matrix $M \in \mathbb{R}^{n \times n}$ such that

$$M = L + S, \qquad \text{s.t.} \quad \operatorname{rank}(L) \leq r, \quad \|S\|_0 \leq s. \tag{1.11}$$

Allowing the addition of a sparse matrix to the low-rank matrix can be viewed as modelling globally correlated structure in the low-rank component while allowing local inconsistencies, innovations, or corruptions. Exemplar applications of this model include image restoration (Gu et al., 2014), hyperspectral image denoising (Gogna et al., 2014; Chen et al., 2017; Wei et al., 2015), face detection (Luan et al., 2014; Wright et al., 2009a), acceleration of dynamic MRI data acquisition (Otazo et al., 2015; Xu et al., 2017), analysis of medical imagery (Baete et al., 2018; Gao et al., 2011), separation of moving objects in an otherwise static scene (Bouwmans et al., 2017), and target detection (Oreifej et al., 2013; Sabushimike et al., 2016) .

Motivated by the successes of ℓ_1 and ℓ_* norm relaxations Candès et al. (2011) and Chandrasekaran et al. (2011) independently showed that the convex relaxation

$$\min_{L,S \in \mathbb{R}^{m \times n}} \|L\|_* + \lambda \|S\|_1, \qquad \text{s.t.} \quad L + S = M, \tag{1.12}$$

recovers L and S under two assumptions: one on the low-rank component L and one on the sparse component S. The first one, assumed by both works alike, is the same assumption as in matrix completion, that the singular vectors of L cannot be concentrated only in few of its entries, but must be

spread out by being sufficiently incoherent. The other assumption, on the sparse component, differs in both works. Candès et al. (2011) assumes that the support set of the sparse component is drawn uniformly at random from all support sets with s entries, while Chandrasekaran et al. (2011) works with a deterministic model and posed that there is an upper bound on the fraction of corrupted entries in each column.

Again, (1.12) can be solved by SDP, but the computation is prohibitively expensive. A number of gradient descent algorithms have been developed for either the convex case or to directly solve Robust PCA in its non-convex formulation

$$\min_{L,S\in\mathbb{R}^{m\times n}} \|L+S-M\|_F, \qquad \text{s.t.} \quad \operatorname{rank}(L)\leq r, \quad \|S\|_0\leq s. \tag{1.13}$$

For a review of Robust PCA algorithms see §4.2.3.

1.2 PROBLEM DESCRIPTION AND SCOPE

The main object of interest of this thesis is the additive combination of low-rank and sparse matrices. That is, we work with matrices $M\in\mathbb{R}^{m\times n}$ expressible as

$$M = L+S, \tag{1.14}$$

where the low-rank component L is of rank at most $r\in\mathbb{N}$ and the sparse component S has at most $s\in\mathbb{N}$ non-zero entries.

1.2.1 *Well-posedness of optimisation over low-rank plus sparse matrices*

The set formed by low-rank plus sparse matrices is captured in the following definition.

Definition 1.1 (The set of low-rank plus sparse matrices)**.** *Denote the set of $m\times n$ real matrices that are the sum of a rank r matrix and a s sparse matrix as*

$$\mathrm{LS}_{m,n}(r,s) = \{L+S\in\mathbb{R}^{m\times n} : \operatorname{rank}(L)\leq r,\ \|S\|_0\leq s\}. \tag{1.15}$$

The $\mathrm{LS}_{m,n}(r,s)$ set plays an important role in Robust PCA where it is the set of viable decompositions in (1.11) and the set of feasible solutions of the non-convex optimisation in (1.12). It can be also seen as a generalisation of the feasible sets for non-convex optimisation in compressed sensing, when $\mathrm{LS}_{m,n}(0,s)$, and matrix completion/sensing, when $\mathrm{LS}_{m,n}(r,0)$.

Despite the importance of the low-rank plus sparse matrix set, some of its properties have not been investigated yet, but have been implicitly assumed in the literature. The study of the rank function in respect to sparsity has been studied from a theoretical standpoint in the complexity theory by Valiant (1977) through the notion of *matrix rigidity*; see §2.4, page 19. Most of the theory of Robust PCA focuses on convergence analysis of algorithms and

identifiability results, which by controlling the correlation of the low-rank and the sparse component guarantee that there exists a *unique* solution.

On the other hand, the *existence* of a solution in matrix completion and Robust PCA is implicitly assumed. This is a reasonable assumption since both, the set of sparse vectors and the set of low-rank matrices, are closed sets, and if the objective function is bounded, a global minimum must exist and can be attained. Indeed, since a solution is guaranteed to exists for both PCA and compressed sensing, one would also expect the same to hold in the case of matrix completion and Robust PCA.

However, $\mathrm{LS}_{m,n}(r,s)$ is constructed as the Minkowski sum of the low-rank set and the sparse set, and therefore it is not guaranteed to be closed, thus possibly jeopardising the existence of a solution and the convergence analysis of iterative algorithms. A sufficient condition for the low-rank plus sparse matrix set to be closed is that the norm of one of the components is proportionally upper bounded by the norm of the matrix sum; see Lemma 2.8 on page 33. This motivates the following definition in which the low-rank component has an additional constraint on its Frobenius norm.

Definition 1.2 (The set of bounded low-rank plus sparse matrices)**.** *Denote the set of $m \times n$ real matrices that are the sum of a rank-r matrix and a sparsity-s matrix as*

$$\mathrm{LS}^{\tau}_{m,n}(r,s) = \{X = L + S \in \mathbb{R}^{m\times n} : \operatorname{rank}(L) \leq r,\ \|S\|_0 \leq s,\ \|L\|_F \leq \tau \|X\|_F\}\,, \tag{1.16}$$

where the rank-r matrix has its Frobenius norm upper bounded by τ times the Frobenius norm of the matrix sum.

The advantage of the set $\mathrm{LS}^{\tau}_{m,n}(r,s)$ compared to the set $\mathrm{LS}_{m,n}(r,s)$ lies in the fact that the former is guaranteed to be closed, the optimisation over it is well-posed, and if normalized, it has a finite covering number. Due to the upper bound on the Frobenius norm of the rank-r component being proportional to the Frobenius norm of the matrix sum, the set also remains conic.

Nevertheless, the set $\mathrm{LS}^{\tau}_{m,n}(r,s)$ in Definition 1.2 has a critical limitation hindering the analysis of optimisation algorithms: it is difficult to guarantee, in general, that the matrix sum of two low-rank plus sparse matrices $X_1, X_2 \in \mathrm{LS}^{\tau}_{m,n}(r,s)$ lies in the set $\mathrm{LS}^{\tau'}_{m,n}(2r,2s)$ for some $\tau' > 0$, see the discussion in §2.7.

The lack of additivity is overcome by the notion of *incoherence*, discussed in detail in §2.2, which controls the correlation between the singular vectors of the low-rank component and the canonical basis.

Definition 1.3 (The set of incoherent low-rank plus sparse matrices)**.** *Denote the set of $m \times n$ real matrices that are the sum of a rank-r matrix and a sparsity-s*

matrix as

$$\mathrm{LS}_{m,n}(r,s,\mu) = \left\{ L+S \in \mathbb{R}^{m\times n} : \begin{array}{c} \mathrm{rank}(L) \le r,\ \|S\|_0 \le s \\ \max_{i\in\{1,\dots,m\}} \|U^T e_i\|_2 \le \sqrt{\frac{\mu r}{m}} \\ \max_{i\in\{1,\dots,n\}} \|V^T f_i\|_2 \le \sqrt{\frac{\mu r}{n}} \end{array} \right\}, \quad (1.17)$$

where $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ are the left and the right singular vectors of L respectively, $e_i \in \mathbb{R}^m$, $f_j \in \mathbb{R}^n$ are the canonical basis vectors, and $\mu \in \left[1, \sqrt{mn}/r\right]$ controls the incoherence of L.

1.2.2 *Sensing of low-rank plus sparse matrices*

While Robust PCA deals with grossly corrupted data, matrix completion and matrix sensing recover a low-rank matrix from incomplete and subsampled data, respectively. In the past decade, Robust PCA, matrix sensing, and matrix completion have been extensively studied and successfully implemented in a wide range of engineering problems.

However, much less focus has been given to the combined case, i.e. the recovery of a grossly corrupted low-rank matrix in an information restrained setting. Mathematically speaking, the goal is to recover an unknown matrix $X_0 \in \mathrm{LS}_{m,n}(r,s,\mu)$, possibly in the form of two components $X_0 = L_0 + S_0$, from a vector of p subsampled measurements $b = \mathcal{A}(X_0)$ made by a linear subsampling operator $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$.

The task is a mixture of *compressed sensing* and *low-rank matrix sensing* with the additional challenge of distinguishing between the low-rank and the sparse component which can become correlated.

This problem has been investigated by (Waters et al., 2011) who design an iterative algorithm called SpaRCS that is based on the widely popular CoSaMP (Needell and Tropp, 2009) and ADMiRA (Lee and Bresler, 2010) designed for compressed sensing and matrix sensing respectively and solve the following non-convex optimisation problem

$$\min_{X\in\mathbb{R}^{m\times n}} \|\mathcal{A}(X) - b\|_2, \qquad \text{s.t.} \quad X \in \mathrm{LS}_{m,n}(r,s). \quad (1.18)$$

While Waters et al. (2011) prove the convergence of their method, the analysis is based on the assumption that $\mathcal{A}(\cdot)$ has its RICs bounded in respect to the low-rank plus sparse matrix set $\mathrm{LS}_{m,n}(r,s)$, which has not been proved yet and has remained an open question.

The derivation of RICs for the set of low-rank matrices by Recht et al. (2010) is based on an ε-covering argument of the set and was inspired by the analogous proof made by Baraniuk et al. (2008) for the set of sparse vectors. The complication with RICs for the low-rank plus sparse matrix set $\mathrm{LS}_{m,n}(r,s)$ comes from the fact that the set does not need to be closed and therefore might not have an ε-covering.

Another unresolved question is if, analogous to compressed sensing and matrix sensing, it is possible to recover a low-rank plus sparse matrix by

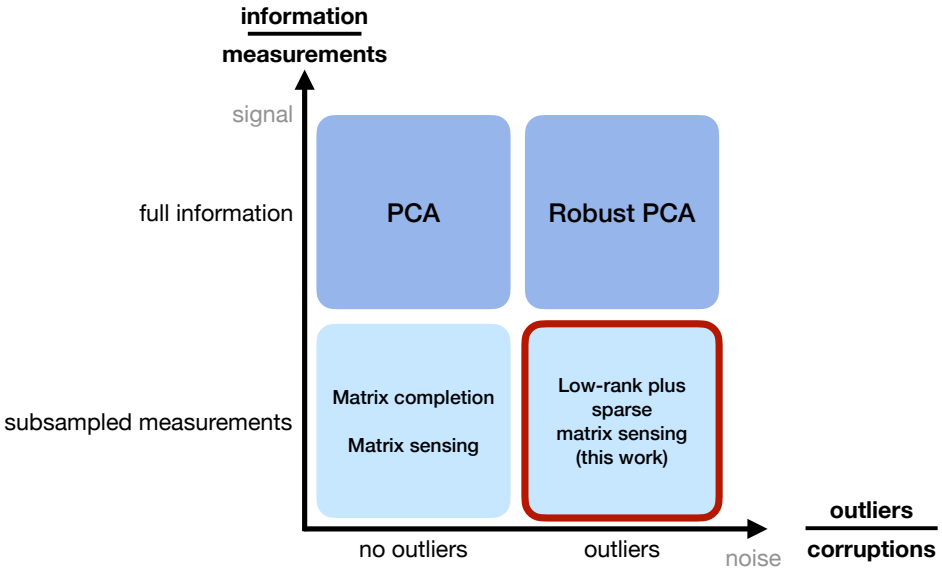


Figure 1.1: Schematic diagram that outlines the different problems based on low-rank matrix approximation. The main problem which is considered in this work is annotated in red.

solving the convex relaxation

$$\min_{X \in \mathbb{R}^{m \times n}} \|L\|_* + \lambda \|S\|_1 , \qquad \text{s.t.} \quad \mathcal{A}(L+S) = b, \tag{1.19}$$

which is solved by *semidefinite programming* techniques.

The main focus of this work is the study of low-rank plus sparse matrix recovery and the corresponding optimisation problems formulated in (1.18) and (1.19). The surrounding context of the work presented here is best explained by the diagram shown in Figure 1.1.

1.3 Objectives and structure of the thesis

In the previous section, we highlighted several unresolved questions about the well-posedness of low-rank plus sparse matrix optimisation and the task of recovering such matrices. This thesis is dedicated to addressing these issues. We shall systematically study the low-rank plus sparse sets and the optimisation problems defined on them. The work can be divided between the theoretical contributions (Chapters 2–4) and the practical contributions on multispectral imaging (Chapter 5).

Chapter 2: Matrix rigidity & the ill-posedness of matrix recovery

In Chapter 2, we discuss the *well-posedness* and *regularisation* of low-rank matrix recovery problems. We begin by reviewing the list of common assumptions posed on Robust PCA and matrix completion in the literature. By constructing a simple 3×3 matrix, we demonstrate that there is another source of ill-posedness in Robust PCA and matrix completion that is not considered by the existing assumptions. Optimisation related to Robust PCA and matrix completion can fail to have any solutions due to the set of low-rank plus sparse matrices not being closed, which in turn is equivalent to the notion of the *matrix rigidity* function not being *lower semicontinuous*. By constructing infinite families of matrices, we derive bounds $n \geq (r +$

1)(s + 2) and $n \geq (r+2)^{3/2}s^{1/2}$ on the matrix size, rank and sparsity for which the set of low-rank plus sparse matrices $\mathrm{LS}_{n,n}(r,s)$ is not closed, see Theorem 2.1. We also demonstrate numerically that a wide range of nonconvex algorithms for both Robust PCA and matrix completion have diverging components when applied to the specifically constructed matrices. We conjecture the best attainable bound is achieved at $n \geq r + \sqrt{s+1}$ using bounds on maximum matrix rigidity, see Conjecture 1 in §2.5.2. The problem is resolved by restraining the Frobenius norm of the low-rank component, thus closing the set, and motivating the set of *bounded* low-rank plus sparse matrices $\mathrm{LS}^{\tau}_{m,n}(r,s)$ in Definition 1.2. However, the sum of two matrices in the set $\mathrm{LS}^{\tau}_{m,n}(r,s)$ does not need to be a sufficiently bounded low-rank plus sparse matrix in general. We conclude the chapter by constraining the incoherence of the low-rank component leading to the notion of the set of *incoherent* low-rank plus sparse matrices $\mathrm{LS}_{m,n}(r,s,\mu)$ in Definition 1.3, which satisfies the additive property [Lemma 2.9] and is also a subset of $\mathrm{LS}^{\tau}_{m,n}(r,s)$ when $\mu < \sqrt{mn}\,/\,\left(r\sqrt{s}\right)$ [Lemma 2.1, Lemma 2.10].

Chapter 3: Restricted isometry constants for low-rank plus sparse matrix set

By closing the set we are able to show that random linear maps obeying certain *concentration of measure inequalities* act as approximate isometries when restricted to the set of low-rank plus sparse matrices. The foundational analytical tool for our results are the *restricted isometry constants* (RICs) for $\mathrm{LS}_{m,n}(r,s,\mu)$, which as for other RICs (Baraniuk et al., 2008; Recht et al., 2010), follows from balancing the *covering number* for the set $\mathrm{LS}_{m,n}(r,s,\mu)$ and the measurement operator being a near isometry as described in Definition 3.2. Random linear maps which have sufficient concentration of measure phenomenon can overcome the dimensionality of $\mathrm{LS}_{m,n}(r,s,\mu)$ to achieve RICs which are bounded by a fixed value independent of dimension size provided the number of measurements is proportional to the number of degrees of freedom of a rank-r plus sparsity-s matrix:

$$p \geq O\left((r(m+n-r)+s)\log\left(\left(1-\mu^2\frac{r^2 s}{mn}\right)^{-1/2}\frac{mn}{s}\right)\right), \tag{1.20}$$

provided $\mu < \sqrt{mn}/(r\sqrt{s})$, see Theorem 3.1. Examples of random linear maps which satisfy these bounds include random Gaussian matrices, random Bernoulli matrices, and the Fast Johnson-Lindenstrauss transform.

Chapter 4: Algorithms for low-rank plus sparse matrix sensing

We devise several computationally tractable methods for the recovery of incoherent low-rank plus sparse matrices from subsampled measurements. We show in Theorem 4.1 that an upper bound on the RICs of the measurement operator implies uniqueness of the solution to the recovery problem. Additionally, we prove that semidefinite programming that solves the optimisation in (1.19) [Theorem 4.2] and two gradient descent algorithms that solve (1.18), NIHT and NAHT [Theorem 4.3 & Theorem 4.4], converge to the subsampled matrix provided the RICs of the measurement operator are sufficiently small. In addition, the convex relaxation and NAHT also provably solve Robust PCA with the asymptotically optimal number of corruptions

$s = O\left(mn/\left(\mu^2 r^2\right)\right)$ when the sensing operator is the identity, and therefore an isometry. We perform numerical experiments illustrating these results for synthetic problems, dynamic-foreground/static-background separation, and multispectral imaging.

Chapter 5: Low-rank models for multispectral imaging

Finally, we apply low-rank matrix completion and compressed sensing in the context of reconstruction of multispectral imagery. In particular, we investigate a reconstruction problem that arises in the acquisition of multispectral images by snapshot mosaic cameras. We show that the missing entries in the images can be accurately imputed despite the severe snapshot undersampling using non-convex techniques from sparse approximation and matrix completion initialised with classical demosaicing algorithms. We observe the peak signal-to-noise ratio can typically be improved by 2 dB to 5 dB over the current state-of-the-art methods when simulating a 16-band mosaic sensor measuring both high and low altitude urban and rural scenes as well as ground-based scenes.

2

MATRIX RIGIDITY & THE ILL-POSEDNESS OF MATRIX RECOVERY

SYNOPSIS

In this chapter, we show that the set of matrices, which can be expressed as the sum of a low-rank and a sparse matrix, is not closed for a range of ranks, sparsities, and matrix dimensions; see Theorem 2.1. Moreover there are a number of algorithms that when given a matrix of a specific form and with constraints on the rank and sparsity, seek such a decomposition where the constituents diverge while at the same time the sum of the matrices converges to a matrix outside of the feasible set. We thereby highlight a previously unknown issue practitioners might experience using these techniques. The situation is analogous to the lack of closedness for Tensor CP decomposition rank (Hitchcock, 1928, 1927) which motivates the notions of multilinear rank approximation (De Lathauwer et al., 2000).

2.1 INTRODUCTION

Many problems in data science take the form of *an inverse problem* – a process of calculating from a set of observations the causal factors that produce them. Typically, there might be a limited number of observations and many possible ways of explaining them. Take as an example the problem of imputing missing entries of a matrix. Clearly, without further constraints, it is impossible to determine the missing entries of a matrix. There are too many possible solutions and the problem is not well defined.

In mathematics, the term *well-posed problem* comes from a definition by Hadamard (1902). According to Hadamard, a well-posed mathematical model of a physical phenomenon should posses three properties

(i) a solution exists,

(ii) the solution is unique,

(iii) the solution depends continuously on the initial conditions.

These are referred to as *existence*, *uniqueness*, and *stability*. Problems that do not satisfy these conditions are said to be *ill-posed*.

In this chapter we will observe that inverse problems formulated as a recovery of an $m \times n$ matrix with restricted rank, sparsity, or a combination of the two, can encounter several types of the aforementioned sources of ill-posedness.

Adding rank and sparsity regularisation restricts the set of suitable solutions.

At first sight, inverse matrix recovery problems, such as Robust PCA and matrix completion, might appear as impossible tasks. In both cases, the limited number of data observations does not uniquely specify a solution to the underlying model—the system is underdetermined. In matrix completion, this is due to observing only p measurements while we wish to recover a matrix with mn entries, with $p < mn$. In Robust PCA, we are given a matrix with mn entries, and the goal is to decompose it into two matrices that have together $2mn$ entries. By posing an additional regularisation condition on the solution in the form of a rank or a sparsity constraint we lower the number of degrees of freedom of the model such that it becomes smaller in comparison to the number of observations.

Matrices can be both low-rank and sparse leading to multiple solutions of Robust PCA and matrix completion. Low-coherence ensures *uniqueness* of a solution, see §2.2.

However, in general, adding the regularisation does not guarantee well-posedness of Robust PCA and matrix completion. This is because some matrices can be both low-rank and sparse. Such matrices have a non-unique low-rank plus sparse decompositions and an entry-wise subsampling P_Ω can miss important information about them. This is a well-known ambiguity in matrix completion and Robust PCA that arises when a low-rank matrix is highly correlated with a sparse matrix. A crucial assumption in these cases is the *incoherence* that ensures *uniqueness* of the solution.

The set of low-rank plus sparse matrices is not closed, and such there can be an issue with *existence* of a solution, see §2.5.

The main result of this chapter highlights the presence of a more fundamental difficulty that leads to ill-posedness of Robust PCA and matrix completion in terms of *existence*. There are matrices for which Robust PCA and matrix completion have no solution in that iterative algorithms that attempt to solve them can generate sequences of iterates (L^t, S^t) for which $\lim_{t\to\infty} \|M - (L^t + S^t)\|_F = 0$ and $L^t + S^t \in \mathrm{LS}_{m,n}(r, s)$ for all t, but $M^* = \lim_{t\to\infty} L^t + S^t \notin \mathrm{LS}_{m,n}(r, s)$. This is not because of the ambiguity between possible solutions or lack of information about the matrix, but instead because $\mathrm{LS}_{m,n}(r, s)$ is not a closed set. Moreover, this is not an isolated phenomenon, as sequences of $\mathrm{LS}_{m,n}(r, s)$ matrices converging outside of the set can be constructed for a wide range of ranks, sparsities and matrix sizes.

The structure of the chapter is as follows. We first define in §2.2 the *incoherence property* of matrices – a standard assumption for a provable solution of Robust PCA and matrix completion. In §2.3, we demonstrate a simple example of a matrix for which Robust PCA is ill-posed and which is not covered by the incoherence condition. The following §2.4 introduces the notion of *matrix rigidity* function, which is the central theoretical object of the chapter, and shows how it relates to the low-rank plus sparse matrix sets not being closed. In §2.5, we generalise the simple example to matrices of arbitrary sizes and a range of ranks and sparsities. In §2.6 shows that many non-convex algorithms follow the diverging path when given specifically

constructed matrices. Finally, we discuss in §2.7 how the problem of the set of low-rank plus sparse matrices not being closed can be resolved by imposing an upper bound on the norm of one of the components, or alternatively, by controlling the coherence of the low-rank component.

2.2 INCOHERENCE AND OTHER MATRIX RECOVERY ASSUMPTIONS

Adding an additional regularisation constraint in the form of a low rank or a low sparsity lowers the number of degrees of freedom of the underlying model. However, having a model with fewer degrees of freedom than the number of measurements still does not guarantee the well-posedness of Robust PCA and matrix completion. This is because matrices which are both low-rank and sparse can have non-unique low-rank plus sparse decompositions and thus violate the *uniqueness* criterion.

Trivial examples of matrices with non-unique low-rank plus sparse decompositions in $\mathrm{LS}_{m,n}(r,s)$ include any matrix with two nonzero entries in differing rows and columns as they are in $\mathrm{LS}_{m,n}(r,s)$ for any r and s such that $r+s=2$ with the entries of the matrix assigned to the sparse or low-rank components selected arbitrarily. Moreover, matrix completion of a low-rank matrix is impossible for sampling patterns P_Ω that are disjoint from the support of the matrix M, which can be likely for matrices that have few nonzeros.

Both of the aforementioned problems are controlled by the *incoherence* parameter which ensures the singular vectors of the low-rank matrix have most entries being nonzero and can be used to ensure recovery guarantees of matrix completion (Candès and Tao, 2010) and Robust PCA (Chandrasekaran et al., 2011; Candès et al., 2011).

Definition 2.1 (Incoherence μ of the low-rank component L)**.** *For a matrix $L \in \mathbb{R}^{m\times n}$ define its incoherence μ as the smallest $\mu \in \left[1, \sqrt{mn}/r\right]$ such that*

$$\max_{i\in\{1,\dots,r\}} \left\|U^T e_i\right\|_2 \le \sqrt{\frac{\mu r}{m}}, \qquad \max_{i\in\{1,\dots,r\}} \left\|V^T e_i\right\|_2 \le \sqrt{\frac{\mu r}{n}}, \tag{2.1}$$

where $L = U\Sigma V^T$ is the singular value decomposition of the rank r component L of size $m\times n$. Matrices with $\mu = 1$ are called maximally incoherent and matrices with $\mu = \sqrt{mn}/r$ are called maximally coherent.

The incoherence condition for small values of μ ensures that left and right singular vectors are well spread out and not sparse. In the case of matrix completion, this prevents from the subsampling operator P_Ω missing important information about the singular vectors. For Robust PCA, a low value of μ of the low-rank component prevents ambiguity of the low-rank and the sparse component.

The following lemma reveals the usefulness of incoherence in controlling the correlation between *incoherent low-rank* and *sparse* matrices.

Lemma 2.1 (The rank-sparsity correlation bound)**.** *Let $L, S \in \mathbb{R}^{m\times n}$ and $L = U\Sigma V^T$ be the singular value decomposition of L, then*

$$|\langle L, S\rangle| \leq \left\|\operatorname{abs}(U)\operatorname{abs}\left(V^T\right)\right\|_\infty \sigma_{\max}(L)\,\|S\|_1\,, \tag{2.2}$$

where abs$(\cdot)$ *denotes the entry-wise absolute value of a matrix, the matrix norms are vectorised entry-wise ℓ_p-norms, and $\sigma_{\max}(L)$ is the largest singular value of L.*

As a consequence, if L is a rank-r matrix that is μ-incoherent and S is an s-sparse matrix

$$|\langle L, S\rangle| \leq \mu\frac{r\sqrt{s}}{\sqrt{mn}}\,\|L\|_F\,\|S\|_F\,, \tag{2.3}$$

where we define $\gamma^{\mu}_{r,s} := \mu\frac{r\sqrt{s}}{\sqrt{mn}}$ to be the (r, s, μ)-rank-sparsity correlation coefficient.

The proof is presented on page 38 as part of §2.9.

Remark 2.1. *Note that the above Lemma 2.1 is informative only when the rank-sparsity correlation coefficient $\gamma^{\mu}_{r,s} < 1$, i.e. when $\mu < \frac{\sqrt{mn}}{r\sqrt{s}}$. As a consequence, we will derive asymptotically optimal low-rank plus sparse matrix recovery rates in terms of incoherence in algorithms presented in Chapter 4.*

Another assumption often made in Robust PCA aims to prevent the case of all of the nonzero entries of S occuring in a single column or in few columns. Suppose for example, that one column of S is the opposite of that of L, and that all the other columns of S are zero. Clearly, we will not be able to recover L and S by any method as $M = L + S$ would have a column space equal to, or included in the column space of L. To avoid these situations, we have to pose an additional assumption on the support set of S.

The assumption on the support set given by Chandrasekaran et al. (2011) is deterministic and upper bounds the number of corruptions in columns and rows of S.

Definition 2.2 (Sparsity ratio of the sparse component S)**.** *The support set of the sparse corruptions matrix S must be sufficiently spread out. We require that for $S \in \mathbb{R}^{m\times n}$, there exists $\alpha \in [0, 1)$, such that $S \in S_\alpha$, where*

$$S_\alpha = \left\{A \in \mathbb{R}^{m\times n} : \|A^T e_i\|_0 \leq \alpha n,\ \|Ae_j\|_0 \leq \alpha m,\ \forall (i, j) \in [m]\times[n]\right\}. \tag{2.4}$$

A consequence of this assumption is also an upper bound on the sparsity $\|S\|_0 \leq \alpha^2 mn$. We refer to the parameter α as the sparsity ratio of matrix S.

A stochastic variant of the requirement is used by Candès et al. (2011), where it is only required that the support set of S is uniformly distributed among all sets of cardinality $\left|\operatorname{supp}(S)\right| = \alpha^2 mn$. The deterministic assumption of Chandrasekaran et al. (2011) is more prevalent in the Robust PCA literature.

Now that we reviewed the main assumptions posed on Robust PCA and matrix completion for ensuring *uniqueness* of the solution, we are prepared to investigate another source of ill-posedness: Robust PCA and matrix completion can cease to have any solution at all.

2.3 SIMPLE EXAMPLE OF THE LACK OF EXISTENCE

Consider solving for the optimal $\mathrm{LS}_{3,3}(1,1)$ approximation to the following 3×3 matrix, which is a special case of construction given by Kumar et al. (2014) in the context of the matrix rigidity function not being lower semicontinuous.

$$\min_{X \in \mathbb{R}^{3\times 3}} \|X - M\|_F, \qquad \text{s.t.} \quad X \in \mathrm{LS}_{3,3}(1,1),$$
$$M = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \tag{2.5}$$

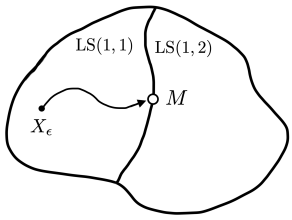
Recall Definition 1.1:

$$\mathrm{LS}_{m,n}(r,s) = \left\{ \begin{array}{c} X = L + S : \\ \mathrm{rank}(L) \le r, \\ \|S\|_0 \le s \end{array} \right\}$$

We have the following sequence of matrices X_ε

$$X_\varepsilon = \begin{pmatrix} 0 & 1 & 1 \\ 1 & \varepsilon & \varepsilon \\ 1 & \varepsilon & \varepsilon \end{pmatrix} \in \mathrm{LS}_{3,3}(1,1)$$
$$= \underbrace{\begin{pmatrix} 1/\varepsilon & 1 & 1 \\ 1 & \varepsilon & \varepsilon \\ 1 & \varepsilon & \varepsilon \end{pmatrix}}_{L_\varepsilon} + \underbrace{\begin{pmatrix} -1/\varepsilon & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}}_{S_\varepsilon},$$

which can decrease the objective function $\|X_\varepsilon - M\|_F = 2\varepsilon$ to zero as $\varepsilon \to 0$, but at the cost of the constituents L_ε and S_ε diverging with unbounded energy. Moreover, the sequence which minimizes the error converges to a matrix M lying outside of the feasible set $\mathrm{LS}_{3,3}(1,1)$ and is in the set $\mathrm{LS}_{3,3}(1,2)$ instead. By the fact that $M \notin \mathrm{LS}_{3,3}(1,1)$, we have that zero objective value cannot be attained and therefore one cannot construct sequences that yield the desired solution. Therefore Robust PCA as posed in (2.5) does not have a global minimum. As the objective function is decreased towards zero, the energy of both the low-rank and the sparse components diverge to infinity.



Sequence X_ε converges outside of the feasible set $\mathrm{LS}_{3,3}(1,1)$.

Likewise, we could consider an instance of the matrix completion problem in (1.9) in which the top left entry of M is missing and a rank 1 approximation is sought. We see that a rank 1 solution cannot be obtained as there does not exist a choice for the top left entry that would reduce the rank of M to 1. However, the sequence L_ε decreases the objective arbitrarily close to zero while the energy of the iterates grows without bounds, $\|L_\varepsilon\|_F \to \infty$.

The underlying mathematical issue is that the matrix rigidity function is not lower semicontinuous as we discuss in the following section.

2.4 MATRIX RIGIDITY IS NOT LOWER-SEMICONTINUOUS

Robust PCA is closely related to the notion of the *matrix rigidity* function which was originally introduced in complexity theory by Valiant (1977) and refers to the minimum number of entries of M that must be changed in order to reduce it to rank r or lower.

$$\text{Rig}(M, r) = \min_{S \in \mathbb{R}^{m \times n}} \|S\|_0, \qquad \text{s.t.} \quad \text{rank}(M - S) \leq r.$$

Note that the original definition by Valiant (1977) works with $\text{rank}(M + S) \leq r$. Here, we change the sign to be consistent with Robust PCA notation, $M = L + S$ and $\text{rank}(L) \leq r$.

Lower bounds on matrix rigidity are motivated by their applications in complexity theory. Specifically, Valiant (1977) showed that lower bounds of the form $\text{Rig}(A, \varepsilon n) = n^{1+\delta}$ for some constants $\varepsilon, \delta > 0$ imply that the linear transform defined by A cannot be computed by an arithmetic circuit with complexity of order $O\left(n \log(n)\right)$. Using the terminology of low-rank plus sparse matrix sets: if a matrix can not be well expressed as a sum of a matrix with sufficiently low rank and a sufficiently sparse matrix, there does not exist an arithmetic circuit that would perform multiplication with the matrix in super-linear or lower complexity. Numerous other applications of lower bounds on rigidity have been found in circuit complexity, communication complexity, and learning complexity (Forster et al., 2001; Lokam, 2001; Linial and Shraibman, 2009), for a comprehensive survey on applications of matrix rigidity see (Codenotti, 2000).

Matrix rigidity is upper bounded for any $M \in \mathbb{R}^{n \times n}$ and rank r as

$$\text{Rig}(M, r) \leq (n - r)^2. \tag{2.6}$$

due to elementary matrix properties. Matrices which achieve this upper bound for every r are referred to as *maximally rigid*. Despite the fact that Valiant (1977) showed that most matrices satisfy such a strong condition, that is they are maximally rigid, it was only recently proved by (Kumar et al., 2014, Theorem 7) how to construct them explicitly. Explicit construction of maximally rigid matrices was a long-standing open question originally posed by Valiant (1977).

Moreover, Kumar et al. (2014) also provide an example of the rigidity function not being *lower-semicontinuous*, which intuitively means that for any point, there exists a small neighbourhood in which the function is nondecreasing.

Definition 2.3 (Semicontinuity)**.** *Let Y be a topological space. A function $\phi : Y \to Z$ is (lower) semicontinuous if, for each $c \in Z$, the set $\left\{y \in Y : \phi(y) \leq c\right\}$ is a closed subset of Y, that is, for each y, there is a neighborhood U of y such that*

$$\forall y' \in U, \quad \phi(y') \geq \phi(y). \tag{2.7}$$

For example, the rank and the sparsity of a matrix, are lower continuous functions on the space of all $m \times n$ matrices. However, the matrix rigidity

For ease of reference:

$$M = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$$X_\varepsilon = \begin{pmatrix} 0 & 1 & 1 \\ 1 & \varepsilon & \varepsilon \\ 1 & \varepsilon & \varepsilon \end{pmatrix}$$

function does not possess this desirable property. Indeed, making an arbitrarily small perturbation does not suffice for a decrease of rank or sparsity. In order to see the matrix rigidity function is not lower-semicontinuous, revisit the simple 3×3 matrix M defined in (2.5) for which $\mathrm{Rig}(M, 1) = 2$. For any neighborhood U, there exists $\varepsilon > 0$ small enough such that $X_\varepsilon \in U$ and $\mathrm{Rig}(X_\varepsilon, 1) = 1 < \mathrm{Rig}(M, 1)$. We see that matrix rigidity is indeed not lower-semicontinuous and that this in turn implies the set $\mathrm{LS}_{3,3}(1, 1)$ is not closed.

Both rank and sparsity are lower-semicontinuous, and as such the optimisation problems over the sets of low-rank and sparse matrices are well-defined. On the other hand, matrix rigidity and CP-rank are not lower-semicontinuous, therefore the corresponding optimisation problems do not need to be well defined (Tanner et al., 2019; de Silva and Lim, 2008).

With the formal definitions of the matrix rigidity and the semicontinuity in place, we are now ready to present the main mathematical result of the chapter. We show that the above simple example of a 3×3 matrix can be generalized for a wide range of ranks and sparsities.

2.5 THE SET OF LOW-RANK PLUS SPARSE MATRICES IS NOT CLOSED

Here we generalize the simple example presented above and show that the set of low-rank plus sparse matrices $\mathrm{LS}_{m,n}(r, s)$ is not closed. Consequently, in some cases, both matrix completion as in (1.9) and Robust PCA as in (1.13) can fail to have any solutions at all. This is equivalent to the notion of the *matrix rigidity* function not being lower semicontinuous as observed in trivial cases by Kumar et al. (2014) and described in the preceding sections.

Robust PCA as in (1.13)

$$\min_{X \in \mathbb{R}^{m \times n}} \|X - M\|_F \quad \text{s.t.} \quad X \in \mathrm{LS}_{m,n}(r, s).$$

Theorem 2.1 ($\mathrm{LS}_{n,n}(r, s)$ is not closed)**.** *The set of low-rank plus sparse matrices $\mathrm{LS}_{n,n}(r, s)$ is not closed for $r \geq 1$, $s \geq 1$ provided $(r + 1)(s + 2) \leq n$, or provided $(r + 2)^{3/2} s^{1/2} \leq n$ where s is of the form $s = p^2 r$ for an integer $p \geq 1$.*

Proof. By Theorem 2.2 and Theorem 2.3. □

Matrix completion as in (1.9)

$$\min_{X \in \mathbb{R}^{m \times n}} \|P_\Omega X - b\|_F \quad \text{s.t.} \quad \mathrm{rank}(X) \leq r.$$

Theorem 2.1 implies that there are matrices M such that the Robust PCA problem in (1.13) and matrix completion in (1.9) are ill-posed in that the objective can be decreased to zero with the sequence of iterates converging to a matrix outside of the feasible set $\mathrm{LS}_{m,n}(r, s)$. Moreover, the proof of Theorem 2.2 and Theorem 2.3 is constructive, and achieved by designing constituents L and S of the sequence diverge with unbounded energy. The problem size bounds in Theorem 2.1 allow for matrices with $r = O(n^\ell)$ to have number of corruptions of order $s = O(n^{2-3\ell})$ for $\ell \in [0, 1/2]$, which for constant rank allows s to be quadratic in n, and for $\ell \in (1/2, 1]$ to have the number of corruptions of order $s = O(n^{(1-\ell)})$.

We extend the example of $\mathrm{LS}_{3,3}(1,1)$ with $M_3 \in \mathbb{R}^{3\times 3}$ given in (2.5) by constructing $M_n, N_n \notin \mathrm{LS}_{n,n}(r,s)$ and yet for which there exists a sequence of matrices $M_n^{(i)}(\varepsilon)$ which are in $\mathrm{LS}_{n,n}(r,s)$ and $\lim_{\varepsilon\to 0} \|M_n^{(i)} - M_n^{(i)}(\varepsilon)\|_F = 0$. Matrix $M_n(\varepsilon)$ as in (2.12) demonstrates that $\mathrm{LS}_{n,n}(r,s)$ is not closed for $r \le s$ (Lemma 2.3) and matrix $N_n(\varepsilon)$ as in (2.18) is constructed for $r > s$ (Lemma 2.4). In both cases we require n to be sufficiently large in terms of r and s.

For the case $r \le s$, consider M_n and $M_n(\varepsilon)$ of the following general form

$$M_n = \begin{pmatrix} 0_{r,r} & A \\ B & 0_{n-r,n-r} \end{pmatrix}, \qquad M_n(\varepsilon) = \begin{pmatrix} 0_{r,r} & A \\ B & \varepsilon BA \end{pmatrix}, \tag{2.8}$$

where $A, B^T \in \mathbb{R}^{r\times(n-r)}$ and $0_{k,k}$ denotes the $k\times k$ matrix with all zero entries. These constructed matrices satisfy the following properties.

Lemma 2.2 (General form of M_n)**.** *Let M_n and $M_n(\varepsilon)$ be as defined in (2.8). Then $M_n(\varepsilon) \in \mathrm{LS}_{n,n}(r,r)$. Furthermore*

$$\lim_{\varepsilon\to 0} \|M_n(\varepsilon) - M_n\|_F = 0. \tag{2.9}$$

Proof. We can write $M_n(\varepsilon)$ in the form

$$\begin{pmatrix} \frac{1}{\varepsilon} I_r \\ B \end{pmatrix} \begin{pmatrix} I_r & \varepsilon A \end{pmatrix} + \begin{pmatrix} -\frac{1}{\varepsilon} I_r & 0 \\ 0 & 0 \end{pmatrix}, \tag{2.10}$$

which shows that $M_n(\varepsilon) \in \mathrm{LS}_{n,n}(r,r)$. It also follows trivially from the definition (2.8) that (2.9) is satisfied. □

Remark 2.2 (Nested property of $\mathrm{LS}_{m,n}(r,s)$ sets)**.** *Note that $\mathrm{LS}_{m,n}(r,s)$ sets form a partially ordered set*

$$\mathrm{LS}_{m,n}(r,s) \subseteq \mathrm{LS}_{m,n}(r',s'), \tag{2.11}$$

for any $r' \ge r$ and $s' \ge s$.

As a consequence $M_n(\varepsilon) \in \mathrm{LS}_{n,n}(r,r)$ implies that also $M_n(\varepsilon) \in \mathrm{LS}_{n,n}(r,s)$ for $s \ge r$.

With Lemma 2.2 we give the general form of M_n and $M_n(\varepsilon)$ such that $M_n(\varepsilon) \in \mathrm{LS}_{n,n}(r,s)$ for $s \ge r$. It remains to show that, for a more specific choice of A and B, we also have $M_n \notin \mathrm{LS}_{n,n}(r,s)$. In particular, we construct

M_n and $M_n(\varepsilon)$ as follows.

$$M_n = \begin{pmatrix} 0_{r,r} & \beta & A^{(1)} & \dots & A^{(\ell)} \\ \alpha^T & 0_{k,k} & \dots & \dots & 0_{k,r} \\ B^{(1)} & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ B^{(\ell)} & 0_{r,k} & \dots & \dots & 0_{r,r} \end{pmatrix},$$
$$M_n(\varepsilon) = \begin{pmatrix} 0_{r,r} & \beta & A^{(1)} & \dots & A^{(\ell)} \\ \alpha^T & \varepsilon\alpha^T\beta & \dots & \dots & \varepsilon\alpha^T A^{(l)} \\ B^{(1)} & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ B^{(\ell)} & \varepsilon B^{(\ell)}\beta & \dots & \dots & \varepsilon B^{(\ell)} A^{(\ell)} \end{pmatrix}, \tag{2.12}$$

where $\alpha, \beta \in \mathbb{R}^{r\times k}$ are matrices with all non-zero entries, $A^{(i)}, B^{(i)} \in \mathbb{R}^{r\times r}$ are arbitrary non-singular matrices which may, but need not, be the same, $0_{a,b}$ and $\mathbb{1}_{a,b}$ denote $a \times b$ matrices with all entries equal to zero or one respectively, and we set $\ell = \lceil (s+1)/2 \rceil$, $k = \lceil \ell/r \rceil$.

By construction, the matrix size is $n = r(\ell+1) + k$, due to the ℓ matrices $A^{(i)}$ and $B^{(i)}$ for $i = 1, \dots, \ell$ each being of size $r \times r$, the top left $r \times r$ zero matrix and k columns of α and β.

Lemma 2.3. *$\mathrm{LS}_{n,n}(r,s)$ is not closed for $1 \le r \le s$ provided*

$$n \ge r\left(\left\lceil \frac{s+1}{2} \right\rceil + 1\right) + \left\lceil \frac{\lceil (s+1)/2 \rceil}{r} \right\rceil. \tag{2.13}$$

Proof. Take M_n as in (2.12). By Lemma 2.2 there exists a matrix sequence $M_n(\varepsilon) \in \mathrm{LS}_{n,n}(r,r)$ such that $\|M_n(\varepsilon) - M_n\|_F \to 0$ as $\varepsilon \to 0$. Since for $r \le s$ we have $\mathrm{LS}_{n,n}(r,r) \subseteq \mathrm{LS}_{n,n}(r,s)$, it follows also that $M_n(\varepsilon) \in \mathrm{LS}_{n,n}(r,s)$.

It remains to prove that $M_n \notin \mathrm{LS}_{n,n}(r,s)$, which is equivalent to showing $\mathrm{Rig}(M_n, r) > s$. We show that having a sparse component $\|S\|_0 \le s$ is insufficient for $\mathrm{rank}(M_n - S) \le r$, because for any choice of such S with at most s non-zero entries, the matrix $M_n - S$ must have a $(r+1)\times(r+1)$ minor with nonzero determinant implying $\mathrm{rank}(M_n - S) \ge r+1$.

In order to establish $\mathrm{rank}(M_n - S) \ge r+1$ we consider 2ℓ minors of M_n each of size $(r+1)\times(r+1)$. For ℓ of these we select minors that include $A^{(i)}$, $i = 1, \dots, \ell$, along with an additional column from the first r columns and an additional row entry from row index $r+1$ to $k+r$ from M_n; and for the remaining ℓ minors we similarly choose a $B^{(i)}$ and an additional row and column as before.

These minors are of the form C_i as shown in (2.14) where the α_i, β_i are chosen to be different entries from α, β for each $i = 1, \dots, \ell$. This requires α, β to be of size $r \times k$ for $k = \lceil \ell/r \rceil$. Recall that, by construction of M_n, the α, β have no zero entries and $A^{(i)}, B^{(i)}$ are each full rank. The C_i are

constructed as

$$C_i = \begin{cases} \begin{pmatrix} 0_{r,1} & A^{(i)} \\ \alpha_i & 0_{1,r} \end{pmatrix}, & i = 1, \ldots, \ell, \\ \begin{pmatrix} 0_{1,r} & \beta_{i-\ell} \\ B^{(i-\ell)} & 0_{r,1} \end{pmatrix}, & i = \ell+1, \ldots, 2\ell, \end{cases} \tag{2.14}$$

where $0_{u,v}$ denotes the $u \times v$ matrix with all entries equal to zero.

Note that matrices C_i do not have disjoint supports as they have some elements from the top left $r \times r$ submatrix of M_n in common. These are the left r zero entries in the first row of C_i for $i = 1, \ldots, \ell$ and the top r zero entries in the first column of C_i for $i = (\ell+1, \ldots, 2\ell)$. We refer to these entries as the *intersecting part* of C_i.

We now consider the possible S such that $\operatorname{rank}(M_n - S) = r$ and show that any such S must have at least 2ℓ nonzeros, thus $\operatorname{Rig}(M_n, r) \geq 2\ell$. This follows by noting that although the C_i have intersecting portions, S restricted to the i^{th} subminor associated with C_i will have at least one distinct nonzero per i. Consider the C_i for $i = 1, \ldots, \ell$ associated with α_i and $A^{(i)}$ and let S_i be the corresponding $(r+1) \times (r+1)$ sparsity mask of S. It follows that S_i must have at least one entry in the non-intersecting set otherwise the determinant of $C_i + S_i$ is of the form

$$|C_i + S_i| = \begin{vmatrix} | & & \\ s_i & & A^{(i)} \\ | & & \\ \alpha_i & 0 & \ldots \quad 0 \end{vmatrix} = \alpha_i |A^{(i)}| \neq 0, \tag{2.15}$$

which is insufficient for the rank of C_i to become rank deficient; similarly for $i = \ell+1, \ldots, 2\ell$.

Having shown $\operatorname{Rig}(M_n, r) \geq 2\ell$ we set $\ell = \lceil (s+1)/2 \rceil$, which then implies that $M_n \notin \operatorname{LS}_{n,n}(r, s)$. By the construction of M_n in this argument we have

$$n \geq r(\ell + 1) + k \tag{2.16}$$

due to the ℓ matrices $A^{(i)}$ and $B^{(i)}$ each of size $r \times r$, the top left $r \times r$ matrix $0_{r,r}$ and k columns of β or rows of α respectively, and by zero padding of the matrix we can arbitrarily increase its size. Substituting $\ell = \lceil (s+1)/2 \rceil$ and $k = \lceil \ell / r \rceil$, we conclude that $\operatorname{LS}_{n,n}(r, s)$ is not a closed set for $s \geq r \geq 1$ provided

$$n \geq r \left(\left\lceil \frac{s+1}{2} \right\rceil + 1 \right) + \left\lceil \frac{\lceil (s+1)/2 \rceil}{r} \right\rceil. \tag{2.17}$$

□

Turning to the $r > s$ case, we now build upon Lemma 2.4 by constructing matrices N_n and $N_n(\varepsilon)$ as

$$N_n = \begin{pmatrix} \widehat{M}_{n'} & 0 & \dots & 0 \\ 0 & E^{(1,1)} & \dots & E^{(1,s+1)} \\ \vdots & \vdots & \ddots & \\ 0 & E^{(s+1,1)} & & E^{(s+1,s+1)} \end{pmatrix} = \begin{pmatrix} \widehat{M}_{n'} & 0_{n',(s+1)(r-s)} \\ 0_{(s+1)(r-s),n'} & E \end{pmatrix}, \quad (2.18)$$

$$N_n(\varepsilon) = \begin{pmatrix} \widehat{M}_{n'}(\varepsilon) & 0_{n',(s+1)(r-s)} \\ 0_{(s+1)(r-s),n'} & E \end{pmatrix}$$

where $E^{(i,j)} \in \mathbb{R}^{(r-s)\times(r-s)}$ are identical full rank matrices and

$$\widehat{M}_{n'} = \begin{pmatrix} 0_{s,s} & \beta & A^{(1)} & \dots & A^{(\ell)} \\ \alpha^T & 0 & \dots & \dots & 0_{1,s} \\ B^{(1)} & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ B^{(\ell)} & 0_{s,1} & \dots & \dots & 0_{s,s} \end{pmatrix},$$

$$\widehat{M}_{n'}(\varepsilon) = \begin{pmatrix} 0_{s,s} & \beta & A^{(1)} & \dots & A^{(\ell)} \\ \alpha^T & \varepsilon\alpha^T\beta & \dots & \dots & \varepsilon\alpha^T A^{(\ell)} \\ B^{(1)} & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ B^{(\ell)} & \varepsilon B^{(\ell)}\beta & \dots & \dots & \varepsilon B^{(\ell)} A^{(\ell)} \end{pmatrix}, \quad (2.19)$$

have the same structure as in (2.12) but with r replaced by s and as a result $A^{(i,j)}, B^{(i,j)} \in \mathbb{R}^{s\times s}$, $\alpha, \beta \in \mathbb{R}^s$, $\ell = \lceil (s+1)/2 \rceil$, so $\widehat{M}_{n'} \notin \mathrm{LS}_{n',n'}(s,s)$ while $\widehat{M}_{n'}(\varepsilon) \in \mathrm{LS}_{n',n'}(s,s)$.

By construction, the size of $\widehat{M}_{n'}$ is $n' = s(\ell+1)+1$ and the size of N_n is $n = n' + (s+1)(r-s)$.

Lemma 2.4. *$\mathrm{LS}_{n,n}(r,s)$ is not closed for $1 \le s < r$ provided*

$$n \ge s\left(\left\lceil \frac{s+1}{2} \right\rceil + 1\right) + 1 + (s+1)(r-s). \quad (2.20)$$

Proof. Consider N_n and $N_n(\varepsilon)$ from (2.18). By additivity of rank for block diagonal matrices, $\mathrm{rank}\,(E) = (r-s)$ and $\widehat{M}_{n'}(\varepsilon) \in \mathrm{LS}_{n',n'}(s,s)$, we have that $N_n(\varepsilon) \in \mathrm{LS}_{n,n}(r,s)$.

It remains to show that $N_n \notin \mathrm{LS}_{n,n}(r,s)$ by proving that $\mathrm{Rig}(N_n, r) > s$. We show that having a sparse component $\|S\|_0 \le s$ is insufficient for $\mathrm{rank}(N_n - S) \le r$, because for any such S, matrix $(N_n - S)$ must have at least one $(r+1)\times(r+1)$ minor with non-zero determinant, implying $\mathrm{rank}(N_n - S) \ge r+1$.

We consider minors D_i of size $(r+1)\times(r+1)$ by diagonally appending a minor $\widehat{C}_i \in \mathbb{R}^{(s+1)\times(s+1)}$ of $\widehat{M}_{n'}$ of a similar structure as in (2.14) and the whole i^{th} diagonal block $E^{(i,i)} \in \mathbb{R}^{(r-s)\times(r-s)}$

$$D_i = \begin{pmatrix} \widehat{C}_i & 0 \\ 0 & E^{(i,i)} \end{pmatrix}, \qquad i = 1, \dots, s+1. \quad (2.21)$$

Due to matrices $E^{(i,i)}$ being picked from the block diagonal, the intersecting parts of supports between D_i are only the intersecting parts between individual $\widehat{C}_i$ as explained in (2.14) in the proof of Lemma 2.3. We will ensure that in order for $\operatorname{rank}(D_i) \leq r$ we require S_i to have at least one non-zero in a part of D_i that is disjoint from D_j for $j \neq i$. Either S_i has at least one non-zero on a zero block or $E^{(i,j)}$ or $\widehat{C}_i$. If the non-zero is in a zero block or $E^{(i,j)}$, then these are disjoint which implies at least $s+1$ non-zero entries. On the other hand, if the non-zero is in $\widehat{C}^{(i)}$ then at least one entry of E must be changed in the non-intersecting part of $\widehat{C}_i$ as argued following equation (2.14). Therefore for every D_i at least one distinct entry per i must be changed using the corresponding sparsity component S_i, and since $i = 1, \ldots, s+1$, we must also change at least $s+1$ entries of N_n. We thus have $\operatorname{Rig}(N_n, r) \geq s+1$.

By the construction of N_n in this argument we have

$$n \geq \underbrace{s(\ell+1)+1}_{n', \text{ size of } \widehat{M}_{n'}} + \underbrace{(s+1)(r-s)}_{\text{size of } \mathbb{1}_{s+1} \otimes N}, \tag{2.22}$$

where the size of $\widehat{M}_{n'}$ comes from ℓ times repeating the matrices $A^{(i)}$ and $B^{(i)}$ each of size $s \times s$, the top left $s \times s$ matrix $0_{s,s}$, the β column and α row respectively and $s+1$ times repeating matrix E of size $(r-s)$. By zero padding of the matrix we can arbitrarily increase its size. Substituting $\ell = \lceil (s+1)/2 \rceil$ gives that $\mathrm{LS}_{n,n}(r,s)$ is not a closed set for $r > s$ provided

$$n \geq s\left(\left\lceil \frac{s+1}{2} \right\rceil + 1\right) + 1 + (s+1)(r-s). \tag{2.23}$$

□

The following theorem gives a sufficient lower bound on the matrix size such that both size requirements derived in Lemma 2.3 and Lemma 2.4 are met, thus unifying both results.

Theorem 2.2. *The low-rank plus sparse set* $\mathrm{LS}_{n,n}(r,s)$ *is not closed provided* $n \geq (r+1)(s+2)$ *and* $r \geq 1$, $s \geq 1$.

Proof. Suppose $n \geq (r+1)(s+2)$. We show that this is a sufficient condition for the matrix size requirements in (2.13) in Lemma 2.3 and (2.20) in Lemma 2.4 to hold.

We first obtain a sufficient condition on the matrix size in (2.13) in Lemma 2.3, bounding

$$\begin{aligned}
&r\left(\left\lceil \frac{s+1}{2} + 1 \right\rceil\right) + \left\lceil \frac{\lceil (s+1)/2 \rceil}{r} \right\rceil \\
&\leq r\left(\frac{s+1}{2} + 2\right) + \left(\frac{1}{r}\right)\left(\frac{s+1}{2} + 1\right) + 1 \\
&\leq r\left(\frac{s+5}{2}\right) + \left(\frac{s+5}{2}\right) = (r+1)\left(\frac{s+5}{2}\right) \\
&\leq (r+1)(s+2),
\end{aligned} \tag{2.24}$$

where the first inequality in (2.24) comes from an upper bound on the ceiling function $\lceil x \rceil \leq x + 1$, the second inequality follows from $r \geq 1$ and the last inequality holds for $s \geq 1$.

We also obtain a sufficient bound condition on the matrix size in (2.20) in Lemma 2.4 of the form

$$\begin{aligned} & s\left(\left\lceil \frac{s+1}{2} + 1 \right\rceil\right) + 1 + (s+1)(r-s) \\ \leq\ & s\left(\frac{s+1}{2} + 2\right) + (s+1)(r-s) = -\frac{s^2}{2} + \frac{3}{2} + rs + 1 \\ \leq\ & (r+1)(s+1) \leq (r+1)(s+2). \end{aligned} \tag{2.25}$$

The first inequality in (2.25) comes from an upper bound on the ceiling function and the second inequality holds for $s \geq 1$.

Combining (2.24), (2.25) with Lemma 2.3 and Lemma 2.4 gives that $\mathrm{LS}_{n,n}(r,s)$ is not a closed set for $n \geq (r+1)(s+2)$ and $r \geq 1, s \geq 1$. □

2.5.1 *Quadratic sparsity*

Note that the condition $n \geq (r+1)(s+1)$ limits the order of r and s; in particular if $r = O(n^\ell)$ then $s = O(n^{1-\ell})$ which for $\ell \geq 0$ constrains s to be at most linear in n, $s = O(n)$. In Lemma 2.5 and Lemma 2.6, we extend the result so that for $r = O(n^\ell)$ and $\ell \leq 1/2$ we obtain $s = O(n^{2-3\ell})$ which for constant rank, $\ell = 0$, allows s to be quadratic $O(n^2)$.

Lemma 2.5 establishes a lower bound on the rigidity of block matrices in terms of the rigidity of a single block. Lemma 2.6 shows that the sequence $K(\varepsilon)$ converging to K is an example of $\mathrm{LS}_{n,n}(r, p^2 r)$ not being closed provided $n \geq p\left(r\left(\left\lceil \frac{r+1}{2} \right\rceil + 1\right) + 1\right)$. Let

$$K = \begin{pmatrix} \widehat{M}_{n'}^{(1,1)} & \cdots & \widehat{M}_{n'}^{(1,p)} \\ \vdots & \ddots & \vdots \\ \widehat{M}_{n'}^{(p,1)} & \cdots & \widehat{M}_{n'}^{(p,p)} \end{pmatrix}, \quad K(\varepsilon) = \begin{pmatrix} \widehat{M}_{n'}^{(1,1)}(\varepsilon) & \cdots & \widehat{M}_{n'}^{(1,p)}(\varepsilon) \\ \vdots & \ddots & \vdots \\ \widehat{M}_{n'}^{(p,1)}(\varepsilon) & \cdots & \widehat{M}_{n'}^{(p,p)}(\varepsilon) \end{pmatrix} \tag{2.26}$$

where matrices $\widehat{M}_{n'}^{(i,j)}(\varepsilon) \in \mathrm{LS}_{n',n'}(r,r)$ and $\widehat{M}_{n'}^{(i,j)} \notin \mathrm{LS}_{n',n'}(r,r)$ are of the same structure as in (2.19) and $\lim_{\varepsilon \to 0} K(\varepsilon) = K$ where $K \in \mathbb{R}^{(n'p)\times(n'p)}$ is constructed by repeating $\widehat{M}_{n'}$ in p row and column blocks.

Lemma 2.5. *For K as in (2.26)*

$$\mathrm{Rig}(K, r) \geq p^2 \mathrm{Rig}(\widehat{M}_{n'}, r). \tag{2.27}$$

Proof. Let S be the sparsity matrix corresponding to $\mathrm{Rig}(K, r)$, such that

$$\begin{aligned} & \mathrm{rank}(K - S) \leq r, \qquad \|S\|_0 = \mathrm{Rig}(K, r), \\ & \text{and} \qquad S = \begin{pmatrix} \widehat{S}^{(1,1)} & \cdots & \widehat{S}^{(1,p)} \\ \vdots & \ddots & \vdots \\ \widehat{S}^{(p,1)} & \cdots & \widehat{S}^{(p,p)} \end{pmatrix}, \end{aligned} \tag{2.28}$$

where $\widehat{S}^{(i,j)} \in \mathbb{R}^{n' \times n'}$ denotes the sparsity matrix used in the place of the $\widehat{M}_{n'}^{(i,j)}$ block. A necessary condition for $\operatorname{rank}(K - S) \leq r$ is that also the rank of individual blocks is less than or equal to r, that is

$$\operatorname{rank}(\widehat{M}_{n'} - \widehat{S}^{(i,j)}) \leq r, \qquad \forall i, j \in \{1, \dots, p\}. \tag{2.29}$$

By definition of the rigidity function as the minimal sparsity of S such that $\operatorname{rank}(\widehat{M}_{n'} - S) \leq r$, we have that

$$\|\widehat{S}^{(i,j)}\|_0 \geq \operatorname{Rig}(\widehat{M}_{n'}, r). \tag{2.30}$$

Summing over all blocks $i, j \in \{1, \dots, p\}$ yields the result

$$\|S\|_0 = \sum_{i,j}^{p,p} \|\widehat{S}^{(i,j)}\|_0 \geq \sum_{i,j}^{p,p} \operatorname{Rig}(\widehat{M}_{n'}, r), \tag{2.31}$$

and consequently that

$$\operatorname{Rig}(K, r) \geq p^2 \operatorname{Rig}(\widehat{M}_{n'}, r). \tag{2.32}$$

□

Lemma 2.6. *The low-rank plus sparse set* $\mathrm{LS}_{n,n}(r, p^2 r)$ *is not closed provided*

$$n \geq p \left(r \left(\left\lceil \frac{r+1}{2} \right\rceil + 1 \right) + 1 \right)$$

and $r \geq 1$, $p \geq 1$.

Proof. Consider K and $K(\varepsilon)$ as in (2.26). Repeating $\widehat{M}_{n'} \in \mathrm{LS}_{n',n'}(r, r)$ p times in row and column blocks does not increase the rank, so $\operatorname{rank}(K(\varepsilon)) = r$ and by additivity of sparsity we have that $K(\varepsilon) \in \mathrm{LS}_{n,n}(r, p^2 r)$. By Lemma 2.5 and $\operatorname{Rig}(\widehat{M}_{n'}, r) > r$ we have the strict lower bound on the rigidity of K

$$\operatorname{Rig}(K, r) \geq p^2 \operatorname{Rig}(\widehat{M}_{n'}, r) > p^2 r, \tag{2.33}$$

which implies that $K \notin \mathrm{LS}_{n,n}(r, p^2 r)$ while $K(\varepsilon) \in \mathrm{LS}_{n,n}(r, p^2 r)$.

Recall that the size of $\widehat{M}_{n'}$ as defined in (2.19) is $n' = r(\ell + 1) + 1$ and, since $\widehat{M}_{n'}$ is repeated p times, we obtain

$$n \geq p\,(r(\ell + 1) + 1) = p \left(r \left(\left\lceil \frac{r+1}{2} \right\rceil + 1 \right) + 1 \right), \tag{2.34}$$

where the inequality comes from zero padding of the matrix to arbitrarily expand its size. □

Theorem 2.3. *The low-rank plus sparse set* $\mathrm{LS}_{n,n}(r, s)$ *is not closed provided*

$$n \geq (r+2)^{3/2} s^{1/2}$$

and $r \geq 1$, *and* s *is of the form* $s = p^2 r$ *for an integer* $p \geq 1$.

Proof. We weaken the condition of Lemma 2.6 and show that it suffices to have $n \geq (r+2)^{3/2} s^{1/2}$ for $\mathrm{LS}_{n,n}(r,s)$ not closed by substituting $s = p^2 r$

$$p\left(r\left(\left\lceil \frac{r+1}{2} \right\rceil + 1\right) + 1\right) = \left(\frac{s}{r}\right)^{\frac{1}{2}} \left(r\left(\left\lceil \frac{r+1}{2} \right\rceil + 1\right) + 1\right) \tag{2.35}$$

$$\leq s^{1/2}\left(r^{1/2}\left(\frac{r+5}{2}\right) + 1\right) = s^{1/2}\left(\frac{r^{3/2}}{2} + 2r^{1/2} + r^{-1/2}\right) \tag{2.36}$$

$$\leq s^{1/2}\left(\frac{r^{3/2}}{2} + 2r^{1/2} + \frac{3}{2}r^{-1/2}\right) = s^{1/2}\frac{(r+1)(r+2)}{2\sqrt{r}} \tag{2.37}$$

$$\leq s^{1/2}(r+2)^{3/2}, \tag{2.38}$$

where in the first line we substitute $s = p^2 r$, the first inequality comes from an upper bound on the ceiling function, the second inequality follows from $r^{-1/2} \leq \frac{3}{2} r^{-1/2}$, and the last inequality holds for $r \geq 1$. □

2.5.2 *Almost maximally rigid examples of non-closedness*

We would like to prove non-closedness of $\mathrm{LS}_{n,n}(r,s)$ sets for as high ranks r and sparsities s as possible. There cannot be a maximally rigid sequence converging outside $\mathrm{LS}_{n,n}\left(r,(n-r)^2\right)$ because $\mathrm{LS}_{n,n}\left(r,(n-r)^2\right)$ corresponds to the set of all $\mathbb{R}^{n\times n}$ matrices. Similarly, it is necessary that both $r \geq 1$ and $s \geq 1$ hold since sets of rank r matrices $\mathrm{LS}_{n,n}(r,0)$ and sets of sparsity s matrices $\mathrm{LS}_{n,n}(0,s)$ are both closed. As a consequence, the highest possible rank and sparsity for which we may hope to prove that $\mathrm{LS}_{n,n}(r,s)$ is not closed corresponds to one strictly less than the maximal rigidity bound, i.e. $\mathrm{LS}_{n,n}\left(r,(n-r)^2-1\right)$ for $r \geq 1$ and also $s = (n-r)^2 - 1 \geq 1$.

It is shown by Kumar et al. (2014) that the matrix rigidity function might not be semicontinuous even for maximally rigid matrices. This translates into the set $\mathrm{LS}_{3,3}(1,3)$ not being closed as we have $M(\varepsilon) \in \mathrm{LS}_{3,3}(1,3)$ which converges to $M \notin \mathrm{LS}_{3,3}(1,3)$ by choosing

$$M = \begin{pmatrix} a & b & c \\ d & e & 0 \\ g & 0 & i \end{pmatrix} \quad \text{and} \quad M(\varepsilon) = \begin{pmatrix} a & b & c \\ d & e & \varepsilon cd \\ g & \varepsilon bg & i \end{pmatrix}. \tag{2.39}$$

It is easy to check that for a general choice of $\{a,\ldots,i\}$, M is maximally rigid with $\mathrm{Rig}(M,1) = 4$. However, $\mathrm{Rig}\left(M(\varepsilon),1\right) = 3$ since $M(\varepsilon)$ can be expressed in the following way

$$M(\varepsilon) = \begin{pmatrix} \varepsilon^{-1} & b & c \\ d & \varepsilon bd & \varepsilon cd \\ g & \varepsilon bg & \varepsilon cg \end{pmatrix} + \begin{pmatrix} a-\varepsilon^{-1} & 0 & 0 \\ 0 & e-\varepsilon bd & 0 \\ 0 & 0 & i-\varepsilon cg \end{pmatrix}. \tag{2.40}$$

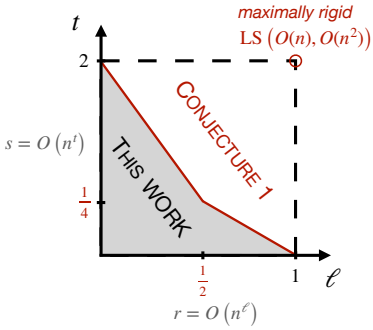
We therefore have that $\mathrm{LS}_{3,3}(1,3)$ is not a closed set, which is the optimal result with the highest possible sparsity for sets of rank 1 matrices of size 3×3.

We pose the question as to whether this result can be generalized and the following conjecture holds.

Conjecture 1 (Almost maximally rigid non-closedness). *The low-rank plus sparse set* $\mathrm{LS}_{n,n}(r,s)$ *is not closed provided*

$$n \geq r + (s+1)^{1/2}, \tag{2.41}$$

for $s \in [1,(n-1)^2-1]$ *and* $r \in [1,n-2]$.



A diagram depicting the results of this work, Conjecture 1, and the maximally rigid matrices in the $\mathrm{LS}(r,s)$ space.

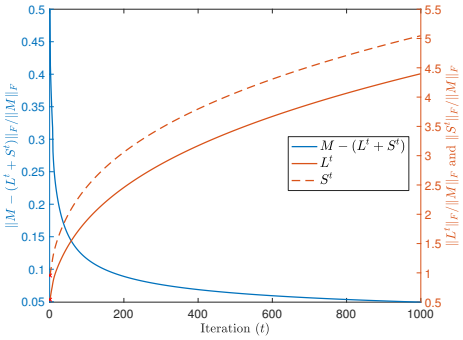
2.6 NUMERICAL EXAMPLES OF DIVERGENT MATRIX RECOVERY

Theorem 2.1 and the constructions in Section 2.5 indicate that there are matrices for which Robust PCA and matrix completion, as stated in (1.13) and (1.9) respectively, are not well defined. In particular, the objective can be driven to zero while the components diverge with unbounded norms. Herein we give examples of two simple matrices which are of a similar construction to M in (2.5),

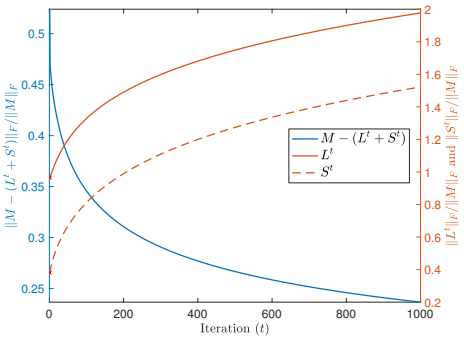
$$M^{(1)} = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad M^{(2)} = \begin{pmatrix} 1 & -2 & -2 \\ -2 & 0 & 0 \\ -2 & 0 & 0 \end{pmatrix},$$

which are not in $\mathrm{LS}_{3,3}(1,1)$, but can be approximated by an arbitrarily close $M^{(1)}_\varepsilon, M^{(2)}_\varepsilon \in \mathrm{LS}_{3,3}(1,1)$, and for which popular Robust PCA and matrix algorithms exhibit this divergence. This is analogous to the problem of diverging components for CP-rank decomposition of higher-order tensors which is especially pronounced for algorithms employing alternating search between individual components, see (de Silva and Lim, 2008) and references therein.

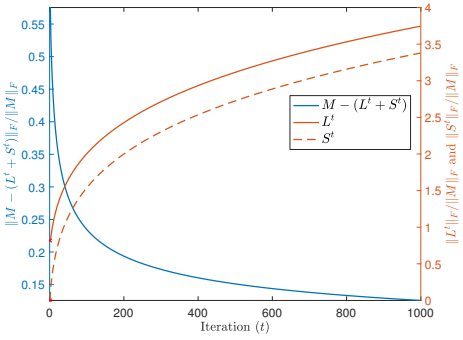
Non-convex algorithms for solving the Robust PCA problem (1.13) are typically observed to be faster than convex relaxations of the problem and often are able to recover matrices with higher ranks than possible by solving the convex relaxation (1.12). We explore the performance of four widely considered non-convex Robust PCA algorithms: Fast Robust PCA via Gradient Descent (FastGD) (Yi et al., 2016), Alternating Minimization (AltMin) (Gu and Wang, 2016), Alternating Projection (AltProj) (Dutta et al., 2018), and Go Decomposition (GoDec) (Zhou and Tao, 2011) applied to $M^{(1)}$ or $M^{(2)}$ with algorithm parameters set to rank $r = 1$ and sparsity $s = 1$. The matrices $M^{(1)}$ and $M^{(2)}$ have values chosen so that the algorithm default initialization causes divergence. However, we would not wish to claim this result is typical in that we do not typically observe divergence for randomly sampled instance of α, β in (2.12) unless the initialization of the algorithm is adjusted to search the diverging sequence. In each case Figure 2.1 shows the convergence of the residual $\min_{X \in \mathbb{R}^{m\times n}} \|X - M\|_F$ to zero while the norms of the constituents of $M = L + S$ diverge.



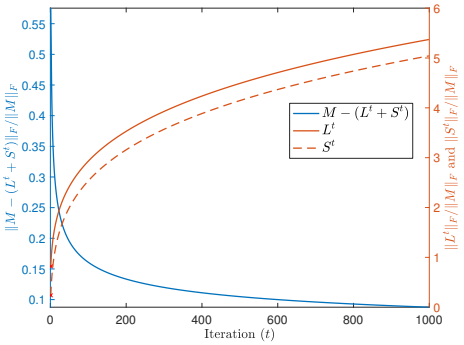
(a) FastGD (Yi et al., 2016) applied to $M^{(1)}$.



(b) AltMin (Gu and Wang, 2016) applied to $M^{(2)}$.



(c) AltProj (Dutta et al., 2018) applied to $M^{(2)}$.

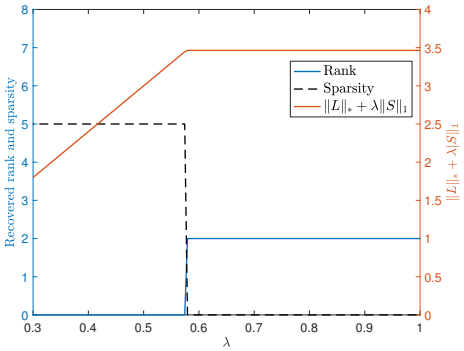


(d) GoDec (Zhou and Tao, 2011) applied to $M^{(2)}$.

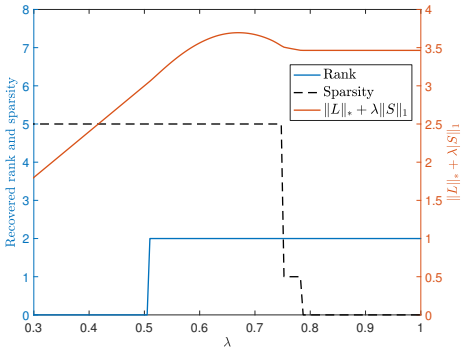
Figure 2.1: Solving for an $\mathrm{LS}_{3,3}(1,1)$ approximation to $M^{(1)}$ and $M^{(2)}$ using four non-convex Robust PCA algorithms. Despite the norm of the residual $\|M-(L^t+S^t)\|_F$ converging to zero, norms of the constituents L^t, S^t diverge. We set algorithms parameters $r=1, s=1$ where possible. For FastGD we set $\lambda=3.23$ and stepsize $\eta=1/6$ which corresponds to choosing $s=1$. For GoDec we set the low-rank projection precision parameter to be 10.

A line of work suggests adding a regularization term to the objective (Gu and Wang, 2016; Ge et al., 2017; Zhang et al., 2018). This leads to bounding the energy of components resulting in the optimization problem to have a global minimum with bounded energies of the constituents. However, the issue of ill-posedness is a more fundamental one; the best rank-r and sparsity-s approximation still has no solution. We observe in Figure 2.2 that energy regularizers result in solutions that are not in the desired space $\mathrm{LS}_{3,3}(r,s)$ for values of (r,s) where the unregularized solution has unbounded energy of its constituents.

Convex relaxations of Robust PCA such as posed in (1.12) do not suffer from the divergence of constituents as shown in Figure 2.1 due to their explicit minimization of their norms. However, they suffer from sub-optimal performance. Figure 2.2 depicts recovered ranks, sparsities and their convex relaxations based on choice for λ of $M^{(1)}$ for Principal Component Pursuit by Alternating Directions (PCP) (Candès et al., 2011) and Inexact Augmented Lagrangian Method (IALM) (Lin et al., 2010). For both PCP and IALM, as the regularization parameter λ is increased from near zero it first produced a solution with $r=0$ and $s=5$, then at approximately $\lambda=1/2$ transitions to solutions with overspecified degrees of freedom $r=2$ and $s=5$, and then for large values of λ gives solutions with $r=2$ and $s=0$. It is interesting to note that for these convex relaxations of Robust PCA there were no values of λ that would produce solutions with $r=1$ and $s=1$ which are the parameters for which the non-convex Robust PCA algorithms diverge. In contrast, the aforementioned non-convex algorithms for Robust PCA applied to $M^{(1)}$ converge to zero residual with bounded constituents for the rank



(a) PCP (Candès et al., 2011) applied to $M^{(1)}$.



(b) IALM (Lin et al., 2010) applied to $M^{(1)}$.

Figure 2.2: Recovered ranks and sparsities using two convex Robust PCA algorithms applied to $M^{(1)}$ with varying choice of λ. Both PCP and IALM do not recover the $r = 1, s = 1$ solution for any λ. IALM recovers solutions with overspecified degrees of freedom $r = 2, s = 5$ for λ roughly $1/2$.

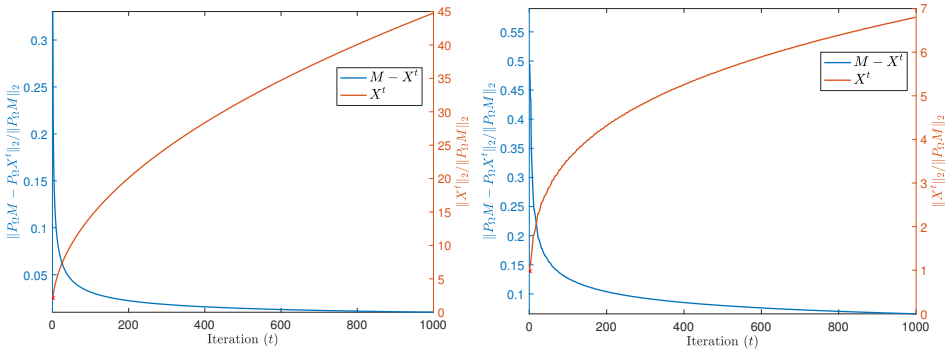
and sparsity parameters generated by PCP and IALM.

The diverging constituents in Figure 2.1 follow the selected (r, s) for which $M^{(1)}, M^{(2)} \notin \mathrm{LS}_{3,3}(r, s)$ but produce a sequence $L^t + S^t \in \mathrm{LS}_{3,3}(r, s)$ and $\lim_{t\to\infty} \|M^{(i)} - (L^t + S^t)\| = 0$ but $\|L^t\|_F$ and $\|S^t\|_F$ diverge. This phenomenon does not occur for these matrices if we allow other choices of (r, s). In particular, Alternating Projection method by Netrapalli et al. (2014) has the rank constraint prescribed and the sparsity constraint is chosen adaptively based on the parameter β and the largest singular value of the low-rank component. Such methods, that do not prescribe both r and s, are less susceptible to the diverging constituents problem. Methods such as the Alternating Projection (Netrapalli et al., 2014) typically have a parameter which controls values of (r, s) and can be selected, such that when applied to $M^{(1)}$ it gives a local minimum in $\mathrm{LS}_{3,3}(1, 1)$.

Similar to the divergence of the non-convex Robust PCA algorithms, non-convex matrix completion algorithms applied to $M^{(1)}$ with only the top left, index $(1, 1)$, entry missing can diverge. Figure 2.3 depicts the residual error converging to zero and energy of the recovered low-rank matrix diverging for four exemplar non-convex algorithms: Power Factorization (PF) (Haldar and Hernando, 2009), Low-Rank Matrix Fitting (LMaFit) (Wen et al., 2012), Conjugate Gradient Iterative Hard Thresholding (CGIHT) (Blanchard et al., 2015) and Alternating Steepest Descent (ASD) (Tanner and Wei, 2016).

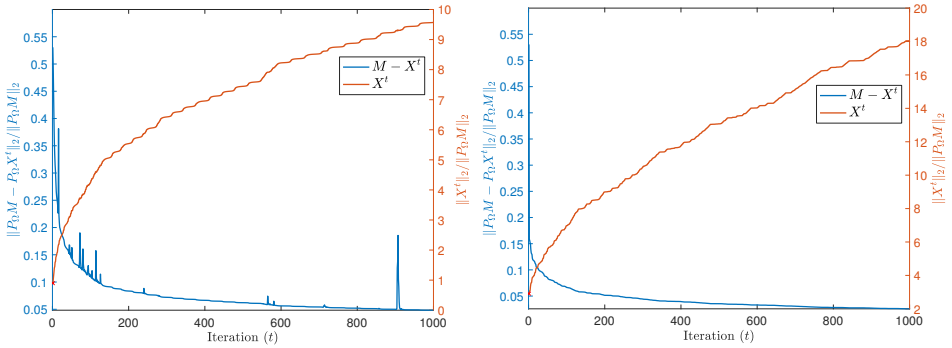
It is required to provide the algorithm with an initial guess that does not have 0 as the top left entry.

The diverging sequences of low-rank plus sparse matrices constructed in §2.5, and followed by iterative Robust PCA and matrix completion algorithms, are pathological in the sense that the low-rank and the sparse component must become highly negatively correlated. In other words, their energy and the magnitude of their inner product must diverge to infinity. In the following section, we discuss how the issue of the diverging components can be overcome and the low-rank plus sparse matrix set be closed.



(a) PF (Haldar and Hernando, 2009) applied to $M^{(1)}$.

(b) LMaFit (Wen et al., 2012) applied to $M^{(1)}$.



(c) CGIHT with restarts (Blanchard et al., 2015) applied to $M^{(1)}$.

(d) ASD (Tanner and Wei, 2016) applied to $M^{(1)}$.

Figure 2.3: Recovery of $M^{(1)}$ given a rank 1 constraint by four non-convex matrix completion algorithm. Despite the norm of the residual $\|y - P_\Omega(X^t)\|_F$ converging to zero, the norm of the recovered matrix X^t diverges.

2.7 CLOSING THE SET OF LOW-RANK PLUS SPARSE MATRICES

The set of low-rank plus sparse matrices $\mathrm{LS}_{m,n}(r,s)$ is constructed as the Minkowski sum of two closed sets: the set of low-rank matrices and the set of sparse matrices, and therefore, it is not guaranteed to be closed by construction. A well-known sufficient condition for the sum of two sets to be also a closed set is that the addition is between a *closed set* and a *closed compact set*, see for example (Aliprantis and Border, 2006, Lemma 5.3) restated here.

Lemma 2.7 (The sum of a closed set and a closed compact set is closed)**.** *The Minkowski sum of a closed compact set $A \subseteq V$ and a closed set $B \subseteq V$ in a normed vector space V*

$$A + B = \{a + b : a \in A, b \in B\}, \tag{2.42}$$

is a closed set.

Proof. Take a sequence $c_n = a_n + b_n \in A + B$ that converges to some $c_n \to c \in V$. Since A is compact, there exist a subsequence $a_{n_i} \xrightarrow{i\to\infty} a \in A$, thus:

$$b_{n_i} = c_{n_i} - a_{n_i} \to c - a. \tag{2.43}$$

From B being closed, we have that $c - a \in B$, which gives:

$$c = a + (c - a) \in A + B. \tag{2.44}$$

□

The above lemma suggests that in order to resolve the issue of non-closedness of $\mathrm{LS}_{m,n}(r, s)$ we should restrict the norm of one of the components, e.g. the low-rank component as $\|L\|_F \leq \tau$ for some $\tau > 0$. The upper bound on the Frobenius norm makes the set the set of low-rank matrices *closed and compact*, and as such, its sum with the *closed set* of sparse matrices is guaranteed to be closed by Lemma 2.7.

However, by bounding the norm of the low-rank component as $\|L\|_F \leq \tau$, the set is no longer conic, and the problem became scale-dependent. To retain the conic property of the set, we instead bound the norm of the low-rank component in proportion to the norm of the matrix sum as $\|L\|_F \leq \tau\|X\|_F$ for some $\tau > 0$. This leads to Definition 1.2 of the set of bounded low-rank plus sparse matrices $\mathrm{LS}^{\tau}_{m,n}(r, s)$.

A small modification in the proof of Lemma 2.7 leads to the following result that shows the constraint $\|L\|_F \leq \tau\|X\|_F$ is sufficient for the set to be closed.

Lemma 2.8 (The set of bounded low-rank plus sparse matrices is closed)**.** *The set of low-rank plus sparse matrices, whose low-rank component has its norm proportionally bounded to the norm of the matrix sum*

$$\mathrm{LS}^{\tau}_{m,n}(r, s) = \{X = L + S \in \mathbb{R}^{m\times n} : \mathrm{rank}(L) \leq r,\ \|S\|_0 \leq s,\ \|L\|_F \leq \tau\|X\|_F\},$$

is a closed set.

The proof of the lemma is given on page 37 in §2.9 and follows from taking subsequences of matrices X_i from a sufficiently large $i_0 \in \mathbb{N}$, for which $\|L_i\|_F \leq C$ for all $i \geq i_0$ and some $C > 0$, and then applying the same arguments as in the proof of Lemma 2.7.

However, the sum of two matrices in the set $\mathrm{LS}^{\tau}_{m,n}(r, s)$ does not need to be a sufficiently bounded low-rank plus sparse matrix in general. In other words, it is difficult to ensure that for $X_1, X_2 \in \mathrm{LS}^{\tau}_{m,n}(r, s)$ that their sum $X_1 + X_2 \in \mathrm{LS}^{\tau'}_{m,n}(2r, 2s)$ for some $\tau' > 0$.

This limitation is overcome in Definition 1.3 of the set of *incoherent* low-rank plus sparse matrices

$$\mathrm{LS}_{m,n}(r, s, \mu) = \left\{ L + S \in \mathbb{R}^{m\times n} : \begin{array}{r} \mathrm{rank}(L) \leq r,\ \|S\|_0 \leq s \\ \max_{i\in\{1,\dots,m\}} \|U^T e_i\|_2 \leq \sqrt{\frac{\mu r}{m}} \\ \max_{i\in\{1,\dots,n\}} \|V^T f_i\|_2 \leq \sqrt{\frac{\mu r}{n}} \end{array} \right\},$$

which satisfies the additive property that the sum of two incoherent low-rank plus sparse matrices is a low-rank plus sparse matrix with the same incoherence μ.

Lemma 2.9 (Addition preserves incoherence)**.** *The sum of two incoherent low-rank plus sparse matrices $X_1, X_2 \in \mathrm{LS}_{m,n}(r, s, \mu)$ is also an incoherent low-rank plus sparse matrix $X_1 + X_2 \in \mathrm{LS}_{m,n}(2r, 2s, \mu)$, and consequently*

$$\mathrm{LS}_{m,n}(r, s, \mu) + \mathrm{LS}_{m,n}(r, s, \mu) = \mathrm{LS}_{m,n}(2r, 2s, \mu), \tag{2.45}$$

where the plus sign denotes the Minkowski sum of two sets.

Proof. Let $X_1, X_2 \in \mathrm{LS}_{m,n}(r, s, \mu)$ with $X_1 = L_1 + S_1$, $X_2 = L_2 + S_2$, and U_1, U_2 and V_1, V_2 being the left and the right singular vectors of L_1 and L_2 respectively.

Construct the sum $X = L + S$, where $L = L_1 + L_2$, $S = S_1 + S_2$, and U, V are the left and right singular vectors of the newly constructed L. Since the column space of U is a subspace of the column space of the concatenated matrix $[U_1\, U_2]$ we have that the projection on U must have a smaller or equal norm than the projection on $[U_1\, U_2]$

$$\begin{aligned} \left\|U^T e_i\right\|_2^2 &\leq \left\|[U_1\, U_2]^T\, e_i\right\|_2^2 && (2.46)\\ &= e_i^T\, [U_1\, U_2]\, [U_1\, U_2]^T\, e_i && (2.47)\\ &= \left\|U_1^T e_i\right\|_2^2 + \left\|U_2^T e_i\right\|_2^2 \leq 2\frac{\mu r}{m}, && (2.48) \end{aligned}$$

where in the third line we use the definition of incoherence. Since the rank of the matrix doubled, the inequality yields the desired result $\|U^T e_i\| \leq \sqrt{\frac{\mu 2r}{m}}$. The argument can be followed *mutatis mutandis* for the upper bound on the right singular vectors V. □

We now show that if $X \in \mathrm{LS}_{m,n}(r, s, \mu)$ with $\mu < \sqrt{mn}/(r\sqrt{s})$ then also $\|L\|_F \leq \tau \|X\|_F$ for $\tau = \left(1 - \mu^2 r^2 s/(mn)\right)^{-1/2}$

Lemma 2.10. *Let $X = L + S \in \mathrm{LS}_{m,n}(r, s, \mu)$ and $\mu < \sqrt{mn}/(r\sqrt{s})$, then we can upper bound the Frobenius norm of the low-rank and the sparse component as*

$$\|L\|_F \leq \frac{1}{\sqrt{1-\gamma^2}} \|X\|_F = \left(1 - \mu^2 \frac{r^2 s}{mn}\right)^{-1/2} \|X\|_F \tag{2.49}$$

$$\|S\|_F \leq \frac{1}{\sqrt{1-\gamma^2}} \|X\|_F = \left(1 - \mu^2 \frac{r^2 s}{mn}\right)^{-1/2} \|X\|_F\,, \tag{2.50}$$

where $\gamma := \mu \frac{r\sqrt{s}}{\sqrt{mn}}$ is the rank-sparsity correlation coefficient as defined also in Lemma 2.1. Consequently

$$\mathrm{LS}_{m,n}(r, s, \mu) \subset \mathrm{LS}^{\tau}_{m,n}(r, s) \tag{2.51}$$

for $\tau = \left(1 - \mu^2 r^2 s/(mn)\right)^{-1/2}$.

Proof. Let $X = L + S \in \mathrm{LS}_{m,n}(r, s, \mu)$. We denote $\gamma = \mu \frac{r\sqrt{s}}{\sqrt{mn}}$ to be the rank-sparsity correlation bound as defined in Lemma 2.1, and by $\mu < \sqrt{mn}/(r\sqrt{s})$, we have that $\gamma < 1$. By conicity of $\mathrm{LS}^{\tau}_{m,n}(r, s)$ and $\mathrm{LS}_{m,n}(r, s, \mu)$, we can assume without loss of generality $\|X\|_F = 1$. The rank-sparsity correlation bound in Lemma 2.1 states

$$\gamma \geq \frac{|\langle L, S\rangle|}{\|L\|_F \|S\|_F}, \tag{2.52}$$

which combined with the rearranged terms of the identity $\|X\|_F^2 = 1 = \|L\|_F^2 + \|S\|_F^2 + 2\langle L, S\rangle$ yields

$$\gamma \geq \frac{|\langle L, S\rangle|}{\|L\|_F \|S\|_F} = \frac{1}{2}\left|\frac{1}{\|L\|_F \|S\|_F} - \frac{\|S\|_F}{\|L\|_F} - \frac{\|L\|_F}{\|S\|_F}\right|. \tag{2.53}$$

The proof follows by showing that the inequality in (2.53) implies an upper bound on $\|L\|_F$ and $\|S\|_F$. For ease of notation, we denote $x = \|L\|_F$ and $y = \|S\|_F$, and multiply the inequality in (2.53) by $\|L\|_F \|S\|_F$

$$2\gamma xy \geq |1 - x^2 - y^2|. \tag{2.54}$$

where we used that $\|L\|_F, \|S\|_F$ are strictly positive.

The case of $1 - x^2 - y^2 \geq 0$ implies that $x \leq 1$ and $y \leq 1$, and thus concludes the proof.

In the other case of $1 - x^2 - y^2 \leq 0$, the inequality in (2.54) is equivalent to

$$2\gamma xy \geq x^2 + y^2 - 1, \tag{2.55}$$

which has two roots $y = -cx \pm \sqrt{c^2x^2 - x^2 + 1}$. Since x, y denote the Frobenius norm of L and S respectively, we seek only the real roots, for which to exist we need $c^2x^2 - x^2 + 1 \geq 0$, and because $c < 1$, we can rearrange the terms as

$$x \leq \frac{1}{\sqrt{1 - \gamma^2}}, \tag{2.56}$$

which is equivalent to

$$\|L\|_F \leq \frac{1}{\sqrt{1 - \gamma^2}} = \left(1 - \mu^2 \frac{r^2 s}{mn}\right)^{-1/2}. \tag{2.57}$$

By $\|X\|_F = 1$, the matrix sum is also in the desired set $X \in \mathrm{LS}^{\tau}_{m,n}(r, s)$ when $\tau = \left(1 - \mu^2 r^2 s/(mn)\right)^{-1/2}$. □

Corollary 2.1. *The set of incoherent low-rank plus sparse matrices* $\mathrm{LS}_{m,n}(r, s, \mu)$ *is closed when* $\mu < \sqrt{mn}/(r\sqrt{s})$.

Proof. By Lemma 2.8 and by the upper bound $\|L\|_F \leq \tau \|X\|_F$ for $\tau = \left(1 - \mu^2 r^2 s/(mn)\right)^{-1/2}$ in Lemma 2.10. □

2.8 SUMMARY AND DISCUSSION

In this chapter, we discussed regularisation and sources of ill-posedness in matrix recovery problems. In particular, we brought into attention an overlooked issue in Robust PCA and matrix completion: that both problems can be ill-posed because the set of low-rank plus sparse matrices is not closed without further conditions being set on the constituent matrices.

It remains to be determined what fraction of the set $\mathrm{LS}_{m,n}(r,s)$ is open, or similarly what fraction has constituents whose norm exceeds a prescribed threshold to ensure well conditioning; it should be noted that in the case of Tensor CP rank the fraction of the space of tensors with unbounded constituent energy is a positive measure (de Silva and Lim, 2008). Numerical experiments confirm that Robust PCA and matrix completion algorithms used on specifically constructed matrices follow the diverging sequences, however, it remains to be seen if this problem arises in a practical setting.

It also remains to determine what is the maximal matrix size n, as a function of (r,s), such that the set $\mathrm{LS}_{n,n}(r,s)$ is open. We give lower bound of $n(r,s) \geq (r+1)(s+2)$ and $n(r,s) \geq (r+2)^{(3/2)}s^{1/2}$ in Theorem 2.1 and conjecture the best attainable bound is achieved at $n(r,s) \geq r+(s+1)^{1/2}$ using bounds on maximum matrix rigidity, see Conjecture 1. Moreover, we note that there are references in the literature (Gu and Wang, 2016; Waters et al., 2011) which reference the use of restricted isometry constants (RIC) for $\mathrm{LS}_{m,n}(r,s)$ in order to prove recovery of Robust PCA using non-convex algorithms. A consequence of our result is that the lower RIC bound is not satisfied for some $M \in \mathrm{LS}_{m,n}(r,s)$ unless further restrictions are imposed on the constituents, such as bounds on the Frobenius norm of one of the components that form M, as done in Definition 1.2:

$$\mathrm{LS}_{m,n}(r,s,\tau) = \{X = L + S \in \mathbb{R}^{m\times n} : \operatorname{rank}(L) \leq r,\ \|S\|_0 \leq s,\ \|L\|_F \leq \tau\|X\|_F\}\,.$$

However, the set of bounded low-rank plus sparse matrices $\mathrm{LS}^{\tau}_{m,n}(r,s)$ does not need to satisfy the additive property in general. The problem is overcome in Definition 1.3 of the set of incoherent low-rank plus sparse matrices which satisfy the additive property as shown in Lemma 2.9. In addition, Lemma 2.10 shows that $\mathrm{LS}_{m,n}(r,s,\mu) \subseteq \mathrm{LS}^{\tau}_{m,n}(r,s)$ for $\tau = \left(1-\mu^2r^2s/(mn)\right)^{-1/2}$ when $\mu < \sqrt{mn}/(r\sqrt{s})$.

In the following Chapter 3, which deals with the restricted isometry constants (RICs) for random linear maps constrained to the set of low-rank plus sparse matrices and requires the set to be closed, we consider $\mathrm{LS}^{\tau}_{m,n}(r,s)$ from Definition 1.2. In Chapter 4, which deals with the recovery of low-rank plus sparse matrices and requires the additive property shown in Lemma 2.9, we use the set of incoherent low-rank plus sparse matrices $\mathrm{LS}_{m,n}(r,s,\mu)$ from Definition 1.3. A consequence of Lemma 2.10 is that the RICs developed in Chapter 3 also apply for $\mathrm{LS}_{m,n}(r,s,\mu)$ when $\mu < \sqrt{mn}/\left(r\sqrt{s}\right)$.

2.9 SUPPORTING LEMMATA

This section contains proofs of lemmata used in the chapter.

Lemma 2.8 (The set of bounded low-rank plus sparse matrices is closed)**.** *The set of low-rank plus sparse matrices, whose low-rank component has its norm proportionally bounded to the norm of the matrix sum*

$$\mathrm{LS}^{\tau}_{m,n}(r,s) = \{X = L + S \in \mathbb{R}^{m\times n} : \operatorname{rank}(L) \le r,\ \|S\|_0 \le s,\ \|L\|_F \le \tau\|X\|_F\}\,,$$

is a closed set.

Proof. Take a sequence $X_i = L_i + S_i \in \mathrm{LS}^{\tau}_{m,n}(r,s)$ that converges to a matrix $X \in \mathbb{R}^{m\times n}$ as $i \to \infty$. Since, also $\|X_i\|_F \to \|X\|_F$, we have that for any $\varepsilon > 0$, there exists $i_0 \in \mathbb{N}$ such that

$$\forall i > i_0 : \qquad \|X\|_F - \varepsilon \le \|X_i\|_F \le \|X\|_F + \varepsilon, \tag{2.58}$$

which, combined with $\|L_i\|_F \le \tau\,\|X_i\|_F$, implies that $\|L_i\|_F \le \tau\,\|X\|_F + \tau\varepsilon$ for all $i \ge i_0$.

Denote the closed set of rank-r matrices whose Frobenius norm is bounded by $\gamma > 0$ as

$$\mathcal{L}_{m,n}(r,\gamma) = \{Y \in \mathbb{R}^{m\times n} : \quad \operatorname{rank}(Y) \le r, \quad \|Y\|_F \le \gamma\}\,, \tag{2.59}$$

which is also compact by being closed and bounded.

We have that $L_i \in \mathcal{L}_{m,n}(r, \|X\|_F + \tau\varepsilon)$ for all $i \ge i_0$. Since the set is compact and closed, we can assume, by passing to a subsequence, that $L_i \xrightarrow{i\to\infty} L \in \mathcal{L}_{m,n}(r, \tau\,\|X\|_F + \tau\varepsilon)$ as $i \ge i_0$.

Additionally, since $\tau > 0$ is fixed, the bound on the Frobenius norm of the low-rank component $\|L_i\|_F \le \tau\,\|X_i\|_F$ must also hold in the limit $\|L\|_F \le \tau\,\|X\|_F$.

By the set of s-sparse matrices being closed, we have that the limit point

$$S_i = X_i - L_i \to X - L, \tag{2.60}$$

is also an s-sparse matrix, thus

$$X = L + (X - L) \in \mathrm{LS}^{\tau}_{m,n}(r,s), \tag{2.61}$$

proving that $\mathrm{LS}^{\tau}_{m,n}(r,s)$ is closed. □

What follows is the proof of Lemma 2.1 that controls the correlation between an incoherent low-rank matrix and a sparse matrix.

Lemma 2.1 (The rank-sparsity correlation bound)**.** *Let $L, S \in \mathbb{R}^{m\times n}$ and $L = U\Sigma V^T$ be the singular value decomposition of L, then*

$$|\langle L, S\rangle| \le \left\|\operatorname{abs}(U)\operatorname{abs}\left(V^T\right)\right\|_{\infty}\,\sigma_{\max}(L)\,\|S\|_1\,, \tag{2.2}$$

where $\mathrm{abs}(\cdot)$ *denotes the entry-wise absolute value of a matrix, the matrix norms are vectorised entry-wise* ℓ_p*-norms, and* $\sigma_{\max}(L)$ *is the largest singular value of* L*.*

As a consequence, if L *is a rank-*r *matrix that is* μ*-incoherent and* S *is an* s*-sparse matrix*

$$|\langle L, S\rangle| \leq \mu \frac{r\sqrt{s}}{\sqrt{mn}} \|L\|_F \|S\|_F \,, \tag{2.3}$$

where we define $\gamma^{\mu}_{r,s} := \mu \frac{r\sqrt{s}}{\sqrt{mn}}$ *to be the* (r, s, μ)*-rank-sparsity correlation coefficient.*

Proof. For $L, S \in \mathbb{R}^{m\times n}$ and $L = U\Sigma V^T$ being the singular value decomposition of L, we have

$$\begin{aligned}
|\langle L, S\rangle| &= \left| \sum_{(i,j)\in[m]\times[n]} S_{i,j}\, L_{i,j} \right| = \left| \sum_{(i,j)\in[m]\times[n]} S_{i,j}\, e_i^T U\Sigma V^T f_j \right| && (2.62)\\
&\leq \sum_{(i,j)\in[m]\times[n]} \left|S_{i,j}\right| \left| \left(U^T e_i\right)^T \Sigma \left(V^T f_j\right) \right| && (2.63)\\
&= \sum_{(i,j)\in[m]\times[n]} \left|S_{i,j}\right| \left| \sum_{k=1}^{r} \sigma_k \left(U^T e_i\right)_k \left(V^T f_j\right)_k \right| && (2.64)\\
&\leq \sum_{(i,j)\in[m]\times[n]} \left|S_{i,j}\right| \sum_{k=1}^{r} \sigma_k \,\mathrm{abs}\left(U^T e_i\right)_k \mathrm{abs}\left(V^T f_j\right)_k && (2.65)\\
&\leq \sigma_{\max}(L) \sum_{(i,j)\in[m]\times[n]} \left|S_{i,j}\right| \mathrm{abs}\left(U^T e_i\right)^T \mathrm{abs}\left(V^T f_j\right) && (2.66)\\
&\leq \sigma_{\max}(L) \|S\|_1 \left\|\mathrm{abs}\left(U\right) \mathrm{abs}\left(V^T\right)\right\|_\infty && (2.67)
\end{aligned}$$

where in the first line in (2.62) we denote $e_i \in \mathbb{R}^m$, $f_i \in \mathbb{R}^n$ to be the canonical basis vectors of $\mathbb{R}^m$ and $\mathbb{R}^n$, the inequality in the second line (2.63) comes from the subadditivity of the absolute value, in the third line (2.64) we write out the inner product as a sum, in the fourth line (2.65) we use the subadditivity and multiplicativity of the abslolute value and denote $\mathrm{abs}(\cdot)$ as the entry-wise absolute value of a vector, the fifth line (2.66) comes from $\sigma_{\max}(L)$ being the largest singular value of L, and the final line in (2.67) comes from the entry-wise ℓ_∞-norm bounding the absolute value of all entries of $\left(\mathrm{abs}\left(U\right) \mathrm{abs}\left(V^T\right)\right)$.

If the low-rank component L is also μ-incoherent, we further have

$$\begin{aligned}
\left\|\mathrm{abs}\left(U\right) \mathrm{abs}\left(V^T\right)\right\|_\infty &= \max_{(i,j)\in[m]\times[n]} \mathrm{abs}\left(U^T e_i\right)^T \mathrm{abs}\left(V^T f_j\right) && (2.68)\\
&\leq \left\|U^T e_i\right\|_2 \left\|V^T f_j\right\|_2 && (2.69)\\
&\leq \mu \frac{r}{\sqrt{mn}}, && (2.70)
\end{aligned}$$

where the first upper bound comes from the Cauchy-Schwarz inequality on the entry-wise absolute values of $U^T e_i$ and $V^T f_j$, and the second upper bound comes from Definition 2.1 of incoherence.

Combining (2.70) with (2.67), the fact that $\|S\|_1 \leq \sqrt{s}\,\|S\|_F$ for s-sparse matrices, and that $\sigma_{\max} \leq \|L\|_F$ yields the result in (2.3)

□

3

RESTRICTED ISOMETRY CONSTANTS FOR THE LOW-RANK PLUS SPARSE MATRIX SET

SYNOPSIS

In this chapter, we show that random measurement operators obeying the concentration of measure inequalities act as approximate isometries when restricted to the set of low-rank plus sparse matrices. These random linear maps can overcome the dimensionality of the $\mathrm{LS}_{m,n}(r, s, \mu)$ set and have their restricted isometry constants in respect to the incoherent low-rank plus sparse matrix set upper bounded as long as the number of measurements is proportional to the number of degrees of freedom of the set. Examples of random linear maps which satisfy these bounds include random Gaussian matrices, random Bernoulli matrices, and the Fast Johnson-Lindenstrauss transform. We discuss the essential properties of the suitable class of random maps that imply the upper bounds on restricted isometry constants and their connection to the widely known Johnson-Lindenstrauss lemma.

3.1 INTRODUCTION

At the heart of compressed sensing is the fact that many signal classes have a low-dimensional structure compared to the high-dimensional ambient space. Remarkably, matrices and projections drawn from certain random distributions allow for a compressed measurement of signals with such structure, while preserving enough information to enable their exact recovery in the original high-dimensional space.

The principle of employing random projections in order to overcome the high dimensionality of signals, referred to as *sketching*, has found extensive applications far beyond compressed sensing. Many high-dimensional problems can be projected to a lower-dimensional space where they can be solved more efficiently. In recent years, sketching has been successfully used in a range of numerical linear algebra problems leading to algorithms

with time and space complexity that is orders of magnitude lower compared to their deterministic counterparts. Exemplar cases include solving least-squares problems (Drineas et al., 2011; Dhillon et al., 2013), low-rank matrix factorization (Halko et al., 2011), semidefinite programming (Yurtsever et al., 2019), and tensor decompositions (Battaglino et al., 2018; Jin et al., 2020) – for an overview of randomized algorithms in numerical linear algebra see (Martinsson and Tropp, 2020).

The fundamental mathematical innovation that allows the aforementioned applications is the concentration of measure phenomenon. The study of the concentration of measure is an important subject in probability and analysis, for a systematic study see the monograph of Ledoux (2001), and has many applications, e.g. the random matrix theory (Tao, 2012). Here we focus on the fact that it provides a way to reduce the dimensionality of high-dimensional data without losing information about its geometry.

In compressed sensing, the main tool for the analysis of recovery guarantees and convergence of algorithms are the *restricted isometry constants* (RICs) introduced in the seminal work of Candès and Tao (2005). If a measurement operator has its RICs bounded, it acts as an approximate isometry when restricted to signals from a certain set. Consequently, Candès and Tao (2005) show that if the RICs of a measurement operator is sufficiently upper-bounded, then there exists a unique sparse solution to the compressed sensing problem that can be provably recovered by solving a convex optimisation problem. An equivalent result for low-rank plus sparse matrix sensing was proved by Recht et al. (2010).

While this chapter extends the previous results of bounded restricted isometry constants to the case of low-rank plus sparse sets, the following Chapter 4 shows that these bounds imply an exact recovery by computationally tractable methods such as a convex relaxation or by gradient descent algorithms.

The main object of interest in this chapter are the RICs for measurement operators $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ when restricted to the set of low-rank plus sparse matrices $\mathrm{LS}_{m,n}(r, s, \mu)$ as defined in Definition 1.3, page 9. The natural generalization of the RIC definition from sparse vectors and low-rank matrices, defined by Candès and Tao (2005) and Recht et al. (2010) respectively, to the set of low-rank plus sparse matrices $\mathrm{LS}_{m,n}(r, s, \mu)$ is given in Definition 3.1.

Definition 3.1 (RIC for $\mathrm{LS}_{m,n}(r, s, \mu)$)**.** *Let $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ be a linear map. For every pair of integers (r, s) and every $\mu \geq 1$, define the (r, s, μ)-restricted isometry constant to be the smallest $\Delta_{r,s,\mu}(\mathcal{A}) > 0$ such that*

$$\big(1 - \Delta_{r,s,\mu}(\mathcal{A})\big)\,\|X\|_F^2 \leq \|\mathcal{A}(X)\|_2^2 \leq \big(1 + \Delta_{r,s,\mu}(\mathcal{A})\big)\,\|X\|_F^2, \qquad (3.1)$$

for all matrices $X \in \mathrm{LS}_{m,n}(r, s, \mu)$.

In general, verifying that the RICs of a measurement operator are bounded is practically an impossible task. In the case of the RICs for s-sparse vectors,

For ease of reference:

$$\mathrm{LS}_{m,n}(r, s, \mu) = \left\{ \begin{array}{c} X = L + S \,: \\ \mathrm{rank}(L) \leq r, \\ \|S\|_0 \leq s, \\ \|U^T e_i\|_2 \leq \sqrt{\frac{\mu r}{m}}, \\ \|V^T e_i\|_2 \leq \sqrt{\frac{\mu r}{n}} \end{array} \right\}$$

An alternative term is the *restricted isometry property* (RIP). A measurement operator is said to satisfy the RIP if its RICs remain bounded as p and mn go to infinity at a fixed rate $\delta := p/(mn)$.

this property requires the condition number of all $s \times s$ submatrices to be bounded. Thus, there are $\binom{mn}{s}$ matrices whose spectral norm we have to compute, which is computationally infeasible. The situation is even more dire if we want to computationally verify an upper bound on the RICs for low-rank matrices. While there is a finite number of support sets of s-sparse vectors, in the case of low-rank matrices there is an infinite number of left and right singular vectors.

Fortunately, as will be discussed in this chapter, many random projection operators have their RICs bounded. For example, it has been shown that RICs are bounded for Gaussian random matrices both in respect to the set of sparse vectors (Candès and Tao, 2005; Baraniuk et al., 2008) and low-rank matrices (Recht et al., 2010). Moreover, Krahmer and Ward (2011) proved that the Fast Johnson-Lindenstrauss transform (FJLT) satisfies the concentration of measure inequalities which also implies an upper bound on its RICs.

The chapter is organised as follows. The following §3.2 characterises the class of suitable random linear maps $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ that obey the concentration of measure inequalities. In §3.3, we state the Johnson-Lindenstrauss (JL) lemma and its connection to the RICs. In §3.4, we give the main result that the RICs of random linear transforms satisfying the conditions given in §3.2 and restricted to $\mathrm{LS}_{m,n}(r, s, \mu)$ remain bounded independent of a problem size provided the number of measurements p is proportional to $O\left(r(m + n - r) + s\right)$ times a logarithmic factor. Finally, §3.5 summarizes the chapter and §3.6 gives supporting lemmata used throughout the chapter.

3.2 NEARLY ISOMETRICALLY DISTRIBUTED MAPS

We begin by defining the class of random projections $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ which have a sufficient concentration of measure phenomenon and are able to project high dimensional points into a random lower-dimensional subspace while preserving their pointwise distances. A suitable class of random linear maps is captured in the following definition.

Definition 3.2 (Nearly isometrically distributed map)**.** *Let $\mathcal{A}$ be a random variable that takes values in linear maps $\mathbb{R}^{m\times n} \to \mathbb{R}^p$. We say that $\mathcal{A}$ is nearly isometrically distributed if, for $\forall X \in \mathbb{R}^{m\times n}$,*

$$\mathrm{E}\left[\|\mathcal{A}(X)\|^2\right] = \|X\|_F^2 \tag{3.2}$$

and for all $\varepsilon \in (0, 1)$, we have

$$\Pr\left(\left|\|\mathcal{A}(X)\|_2^2 - \|X\|_F^2\right| \geq \varepsilon\|X\|_F^2\right) \leq 2\exp\left(-\frac{p}{2}\left(\varepsilon^2/2 - \varepsilon^3/3\right)\right), \tag{3.3}$$

and there exists some constant $\gamma > 0$ such that for all $t > 0$, we have

$$\Pr\left(\|\mathcal{A}\| \geq 1 + \sqrt{\frac{mn}{p}} + t\right) \leq \exp\left(-\gamma p t^2\right). \tag{3.4}$$

There are two crucial properties for a random map to be nearly isometric. Firstly, it needs to be isometric in expectation as in (3.2), and exponentially concentrated around the expected value as in (3.3). Secondly, the probability of large distortions of length must be exponentially small as in (3.4). This ensures that even after taking a union bound over an exponentially large covering number for $\mathrm{LS}_{m,n}(r, s, \mu)$, see Lemma 3.4, the probability of distortion remains small (Baraniuk et al., 2008; Recht et al., 2010).

Examples of random ensembles of $\mathcal{A}$ which satisfy the conditions of Definition 3.2 include random Gaussian ensemble which acquires the information about the matrix X through p linear measurements of the form

$$b_\ell := \mathcal{A}(X)_\ell = \langle A^{(\ell)}, X \rangle \quad \text{for} \quad \ell = 1, 2, \ldots, p, \tag{3.5}$$

where the p distinct sensing matrices $A^{(\ell)} \in \mathbb{R}^{m \times n}$ are the sensing operators defining $\mathcal{A}$ and have entries sampled from the Gaussian distribution as $A^{(\ell)}_{i,j} \sim \mathcal{N}(0, 1/p)$. Other notable examples include symmetric Bernoulli ensembles, and Fast Johnson-Lindenstrauss Transform (FJLT) introduced by Ailon and Chazelle (2009). Krahmer and Ward (2011) proved the concentration inequalities for FJLT which are weaker compared to the Gaussian sensing operator.

Note that applying a dense linear transform such as the one described in (3.5) to an $n \times n$ matrix is extremely costly with the complexity $O\left(pn^2\right)$ that is quartic in n when $p = O(n^2)$. Devising fast transforms such as FJLT that have complexity of $O\left(n^2 \log(n^2) + p\right)$ is crucial for applications with high-dimensional data such as large matrices or tensors (Jin et al., 2020). Another avenue of research are sparse sketching transforms that allow for fast iterative updates (Kane and Nelson, 2014) and are generally used in streaming applications.

3.3 Connection to the Johnson-Lindenstrauss lemma

The Johnson-Lindenstrauss (JL) lemma (Johnson and Lindenstrauss, 1984) answers the following problem. We are given a set $Q \subset \mathbb{R}^n$ of N points with the dimension n typically large. We would like to embed these points into a lower-dimensional Euclidean space $\mathbb{R}^p$ while approximately preserving the relative distances between any of these points. The question we are interested in is: How small can we make the dimension p of the projected points while approximately preserving their distances?

The JL lemma answers this question by stating that with high probability the geometry of a point cloud is disturbed by some Lipschitz mapping onto a space of dimension logarithmic in the number of points.

Lemma 3.1 (Johnson-Lindenstrauss lemma)**.** *Let $\varepsilon \in (0, 1)$ be given. For every set Q of N points in $\mathbb{R}^n$, if p is a positive integer such that $p > p_0 = O\left(\log(N)/\varepsilon^2\right)$,*

there exists a Lipschitz mapping $f : \mathbb{R}^n \to \mathbb{R}^p$ *such that*

$$(1-\varepsilon)\,\|u-v\|_2^2 \leq \|f(u)-f(v)\|_2^2 \leq (1+\varepsilon)\,\|u-v\|_2^2, \tag{3.6}$$

for all $u, v \in Q$.

There have been various improvements to the proof of this lemma (Indyk and Motwani, 1998; Achlioptas, 2003; Dasgupta and Gupta, 2003; Donoho, 2006a). In particular, there now exist simple proofs of the lemma which show the mapping f can be represented by an $p \times n$ matrix whose entries are randomly drawn from some i.i.d. probability distribution. In fact, Achlioptas (2003) showed that any random variable satisfying certain moment conditions is suitable and will satisfy the conditions of the lemma with a non-zero probability.

In a nutshell, the proof of the JL lemma relies on showing that transforms corresponding to suitable random matrices satisfy the conditions of *nearly isometrical random maps* described in Definition 3.2. Subsequently, the conditions of Definition 3.2 in combination with the union bound on the set of all pair-wise differences of points $u, v \in Q$ prove the result.

In the context of compressed sensing, Baraniuk et al. (2008) showed that the same properties of *nearly isometrical random maps* that allow for the proof of the JL lemma can be also used to verify an upper bound on the RICs for sparse vectors. This is done by constructing an ε-net covering for the set of s-sparse vectors and then applying the union bound. The work of Recht et al. (2010) applied the same technique combined with ε-covering of the Grassmannian manifold by Szarek (1983) to prove that RICs of *nearly isometrical random maps* are upper bounded also when restricted to the set of low-rank matrices.

We are now ready to combine the two arguments of Baraniuk et al. (2008) and Recht et al. (2010), with a modification arising from the non-closedness property discussed in Chapter 2, and prove that the RICs of *nearly isometrical random maps* given in Definition 3.2 are also upper bounded when restricted to matrices from the low-rank plus matrix set.

3.4 Restricted isometry constants for the set of low-rank plus sparse matrices

Linear maps $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ which have a sufficient concentration of measure phenomenon can overcome the dimensionality of $\mathrm{LS}_{m,n}(r, s, \mu)$ to achieve $\Delta_{r,s,\mu}$ given in Definition 3.1 which is bounded by a fixed value independent of dimension size provided the number of measurements p is proportional to the degrees of freedom of a rank-r plus sparsity-s matrix $r(m+n-r)+s$.

Theorem 3.1 (RICs for $\mathrm{LS}_{m,n}(r, s, \mu)$)**.** *For a given* $m, n, p \in \mathbb{N}$, $\Delta \in (0, 1)$, $\mu \in \left[1, \frac{\sqrt{mn}}{r\sqrt{s}}\right)$, *and a random linear transform* $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ *satisfying the*

concentration of measure inequalities in Definition 3.2, there exist constants $c_0, c_1 > 0$ *such that the RIC for* $\mathrm{LS}_{m,n}(r, s, \mu)$ *is upper bounded with* $\Delta_{r,s,\mu}(\mathcal{A}) \leq \Delta$ *provided*

$$p > c_0 \left(r(m+n-r)+s\right) \log\left(\left(1-\mu^2 \frac{r^2 s}{mn}\right)^{-1/2} \frac{mn}{s}\right), \tag{3.7}$$

with probability at least $1-\exp\left(-c_1 p\right)$*, where* c_0, c_1 *are constants that depend only on* Δ*.*

Our proof of Theorem 3.1 follows from proving the alternative form of (3.1) defined without the squared norms by

$$\left(1-\bar{\Delta}_{r,s,\mu}(\mathcal{A})\right) \|X\|_F \leq \|\mathcal{A}(X)\|_2 \leq \left(1+\bar{\Delta}_{r,s,\mu}(\mathcal{A})\right) \|X\|_F, \tag{3.8}$$

which we denote as $\bar{\Delta}$. The discrepancy between (3.8) and (3.1) is due to (3.8) being more direct to derive and (3.1) allowing for more concise derivation of recovery results by computationally tractable methods in Chapter 4.

The two definitions are related, since $\bar{\Delta}$ satisfiying the inequalities in (3.8) also implies

$$\left(1-\bar{\Delta}\right)^2 \|X\|_F^2 \leq \|\mathcal{A}(X)\|_2^2 \leq \left(1+\bar{\Delta}\right)^2 \|X\|_F^2, \tag{3.9}$$

which in turn ensures that Δ in Definition 3.1 is $\Delta = 2\bar{\Delta} - \bar{\Delta}^2$ and strictly smaller than one for $\Delta \leq \sqrt{2}-1$. For $\bar{\Delta} > \sqrt{2}-1$, the squared RICs remain bounded, but become asymetrical, with the constant in the lower bound becoming $\Delta_L = 2\bar{\Delta} + \bar{\Delta}$ while the constant in the upper bound remaining $\Delta_U = 2\bar{\Delta} - \bar{\Delta}^2$.

The proof of Theorem 3.1 begins with the derivation of an RIC for a single subspace $\Sigma_{m,n}(V, W, T, \mu)$ of $\mathrm{LS}_{m,n}(r, s, \mu)$ when the column space of $\mathrm{Col}\,(\mathrm{L})$ is restricted in the subspace V, the row space $\mathrm{Col}\left(\mathrm{L}^{\mathrm{T}}\right)$ in the subspace W and the sparse component S is in the subspace T

$$\Sigma_{m,n}\left(V, W, T, \mu\right) = \left\{ X = L + S \in \mathbb{R}^{m\times n} : \begin{array}{c} \mathrm{Col}\,(\mathrm{L}) \subseteq V,\ \mathrm{Col}\left(\mathrm{L}^{\mathrm{T}}\right) \subseteq W, \\ \mathrm{supp}\,(S) \subseteq T, \\ \forall i \in [m] :\ \|\mathrm{P}_V e_i\|_2 \leq \sqrt{\frac{\mu r}{m}}, \\ \forall i \in [n] :\ \left\|\mathrm{P}_W f_i\right\|_2 \leq \sqrt{\frac{\mu r}{n}} \end{array} \right\}, \tag{3.10}$$

where P_V and P_W denote the orthogonal projection on the subspace V and W respectively, and $e_i \in \mathbb{R}^m$ and $f_i \in \mathbb{R}^n$ are the canonical basis vectors. Note, that the conditions of $\|\mathrm{P}_V e_i\|_2 \leq \sqrt{\frac{\mu r}{m}}$ and $\left\|\mathrm{P}_W f_i\right\|_2 \leq \sqrt{\frac{\mu r}{n}}$ combined with $\mathrm{Col}\,(\mathrm{L}) \subseteq V$ and $\mathrm{Col}\left(\mathrm{L}^{\mathrm{T}}\right) \subseteq W$ imply that the coherence of the singular vectors of the matrix L is bounded by μ.

Following the proof of the RIC for a single subspace, we show that the isometry constant of $\mathcal{A}$ is robust to a perturbation of the column and the row subspaces (V, W) of the low-rank component. Finally, we use a covering

argument over all possible column and row subspaces (V, W) of the low-rank component and count over all possible sparsity subspaces T of the sparse component to derive an exponentially small probability bound for the event that $\mathcal{A}(\cdot)$ satisfies RIC with constant $\bar{\Delta}$ for sets

$$\mathrm{LS}_{m,n}(r, s, \mu) = \left\{ \Sigma_{m,n}(V, W, T, \mu) \; : \; \begin{array}{c} V \in \mathcal{G}(m, r),\, W \in \mathcal{G}(n, r), \\ T \in \mathcal{V}(mn, s) \end{array} \right\}, \quad (3.11)$$

where $\mathcal{G}(m, r)$ is the Grassmannian manifold—the set of all r-dimensional subspaces of $\mathbb{R}^m$, and $\mathcal{V}(mn, s)$ is the set of all possible supports sets of an $m \times n$ matrix that has s elements. Thus proving RIC for sets of low rank plus sparse matrices given the bound on the Frobenius norm of the low-rank component L.

The following result describes the behavior of $\mathcal{A}$ when constrained to a single fixed column and a row space (V, W) and a single sparse matrix space T.

Lemma 3.2 (RICs for a fixed LS subspace $\Sigma_{m,n}(V, W, T, \mu)$)**.** *Let $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$ be a nearly isometric random linear map from Definition 3.2 and $\Sigma_{m,n}\left(V, W, T, \mu\right)$ as defined in (3.10) is fixed for some (V, W, T), and $\mu \in \left[1, \frac{\sqrt{mn}}{r\sqrt{s}}\right)$. Then for any $\bar{\Delta} \in (0, 1)$*

$$\forall X \in \Sigma_{m,n}\left(V, W, T, \mu\right) : \quad (1 - \bar{\Delta})\|X\|_F \leq \|\mathcal{A}(X)\| \leq (1 + \bar{\Delta})\|X\|_F, \quad (3.12)$$

with probability at least

$$1 - 2\left(\frac{24}{\bar{\Delta}}\tau\right)^{\dim V \cdot \dim W}\left(\frac{24}{\bar{\Delta}}\tau\right)^{\dim T} \exp\left(-\frac{p}{2}\left(\frac{\bar{\Delta}^2}{8} - \frac{\bar{\Delta}^3}{24}\right)\right), \quad (3.13)$$

where $\tau = 1 \Big/ \sqrt{1 - \mu^2 \frac{r^2 s}{mn}}$.

The proof follows the same argument as the one for sparse vectors (Baraniuk et al., 2008, Lemma 5.1) and for low-rank matrices in (Recht et al., 2010, Lemma 4.3). Our variant of the proof for low-rank plus sparse matrices is presented in §3.6, page 51.

To establish the impact of a perturbation of the subspaces (V, W) on the $\bar{\Delta}$ in Lemma 3.2 we define a metric $\rho(\cdot, \cdot)$ on $\mathcal{G}(D, d)$ as follows

$$U_1, U_2 \in \mathcal{G}(D, d) : \quad \rho(U_1, U_2) := \|\mathrm{P}_{U_1} - \mathrm{P}_{U_2}\|, \quad (3.14)$$

where $\|\mathrm{P}_{U_1} - \mathrm{P}_{U_2}\|$ denotes the spectral norm of $\mathrm{P}_{U_1} - \mathrm{P}_{U_2}$. The Grassmannian manifold $\mathcal{G}\left(D, d\right)$ combined with distance $\rho(\cdot, \cdot)$ as in (3.14) defines a metric space $\left(\mathcal{G}\left(D, d\right), \rho\left(\cdot, \cdot\right)\right)$, where P_U denotes an orthogonal projection associated with the subspace U. Let us also denote a set of matrices whose column and row space is a subspace of V and W respectively

$$(V, W) = \left\{X \; : \; \mathrm{Col}\left(\mathrm{X}\right) \subseteq V,\, \mathrm{Col}\left(\mathrm{X}^{\mathrm{T}}\right) \subseteq W\right\}, \quad (3.15)$$

and $\mathrm{P}_{(V,W)}$ is an orthogonal projection that ensures that the column space and row space of $\mathrm{P}_{(V,W)}X$ lies within V and W. The distance between $\Sigma_1 := \Sigma_{m,n}\left(V_1, W_1, T, \mu\right)$ and $\Sigma_2 := \Sigma^{\tau}_{m,n}\left(V_2, W_2, T, \mu\right)$ that have a fixed T is given by

$$\rho\left(\left(V_1, W_1\right), \left(V_2, W_2\right)\right) = \|\mathrm{P}_{(V_1,W_1)} - \mathrm{P}_{(V_2,W_2)}\|. \tag{3.16}$$

Lemma 3.3 (Variation of $\bar{\Delta}$ in RIC in respect to a perturbation of (V, W))**.** *Let $\Sigma_1 := \Sigma_{m,n}(V_1, W_1, T, \mu)$ and $\Sigma_2 := \Sigma_{m,n}(V_2, W_2, T, \mu)$ be two low-rank plus sparse subspaces with the same fixed sparse subspace T and $\mu \in \left[1, \frac{\sqrt{mn}}{r\sqrt{s}}\right)$. Suppose that for $\bar{\Delta} > 0$, the linear operator $\mathcal{A}$ satisfies*

$$\forall X \in \Sigma_1 : \quad (1 - \bar{\Delta})\|X\|_F \leq \|\mathcal{A}(X)\| \leq (1 + \bar{\Delta})\|X\|_F. \tag{3.17}$$

Then

$$\forall Y \in \Sigma_2 : \quad (1 - \bar{\Delta}')\|Y\|_F \leq \|\mathcal{A}(Y)\| \leq (1 + \bar{\Delta}')\|Y\|_F, \tag{3.18}$$

with $\bar{\Delta}' := \bar{\Delta} + \tau\rho\left(\left(V_1, W_1\right), \left(V_2, W_2\right)\right)\left(1 + \bar{\Delta} + \|\mathcal{A}\|\right)$ with ρ as defined in (3.14) and $\tau = 1\Big/\sqrt{1 - \mu^2\frac{r^2 s}{mn}}$.

The proof is similar to the line of argument made in (Recht et al., 2010, Lemma 4.4), see the proof in §3.6, page 54. The notable exception is the term τ appearing in the expression for $\bar{\Delta}'$, which is the result of the set $\mathrm{LS}_{m,n}(r, s)$ not being closed without the constraint $\|L\|_F \leq \tau\|X\|_F$ (Tanner et al., 2019, Theorem 1.1) but which is here guaranteed by Lemma 2.10 provided $\mu < \frac{\sqrt{mn}}{r\sqrt{s}}$.

To establish the proof of Theorem 3.1 we combine Lemma 3.2 and Lemma 3.3 with an ε-covering of the subspaces of $\mathrm{LS}_{m,n}(r, s)$ which also contains the subspaces of the set $\mathrm{LS}_{m,n}(r, s, \mu)$, where ε will be picked to control the maximal allowed perturbation $\rho\left(\left(V_1, W_1\right), \left(V_2, W_2\right)\right)$ between the subspaces. The *covering number* $\mathfrak{R}(\varepsilon)$ of the subspaces of $\mathrm{LS}_{m,n}(r, s)$ at resolution ε is the smallest number of subspaces (V_i, W_i, T_i) such that, for any triple of $V \in \mathcal{G}(m, r), W \in \mathcal{G}(n, r), T \in \mathcal{V}(mn, s)$ there exists i with $\rho\left(\left(V, W\right), \left(V_i, W_i\right)\right) \leq \varepsilon$ and $T = T_i$. The following Lemma 3.4 gives an upper bound on the cardinality of the ε-covering.

Lemma 3.4 (Covering number of the subspaces of $\mathrm{LS}_{m,n}(r, s)$)**.** *The covering number $\mathfrak{R}(\varepsilon)$ of the subspaces of the set $\mathrm{LS}_{m,n}(r, s)$ is bounded above by*

$$\mathfrak{R}(\varepsilon) \leq \binom{mn}{s}\left(\frac{4\pi}{\varepsilon}\right)^{r(m+n-2r)}. \tag{3.19}$$

Consequently, by $\mathrm{LS}_{m,n}(r, s, \mu) \subset \mathrm{LS}_{m,n}(r, s)$, we also obtained an upper bound on the covering number for the subspaces of the set of incoherent low-rank plus sparse matrices $\mathrm{LS}_{m,n}(r, s, \mu)$.

The proof comes by counting the possible support sets with cardinality s and by Theorem 8 of Szarek (1998) on ε-covering of the Grassmannian, for completeness the proof is given in §3.6, page 53.

Bounds on the RICs for the set of low-rank plus sparse matrices then follow a proof technique that uses the covering number argument in combination with the concentration of measure inequalities as done by Baraniuk et al. (2008) for sparse vectors and subsequently for low-rank matrices by Recht et al. (2010).

We now have the required theoretical background in place to prove the main result of the chapter in the form of Theorem 3.1 restated here.

Theorem 3.1 (RICs for $\mathrm{LS}_{m,n}\left(r,s,\mu\right)$)**.** *For a given $m,n,p \in \mathbb{N}$, $\Delta \in (0,1)$, $\mu \in \left[1, \frac{\sqrt{mn}}{r\sqrt{s}}\right)$, and a random linear transform $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ satisfying the concentration of measure inequalities in Definition 3.2, there exist constants $c_0, c_1 > 0$ such that the RIC for $\mathrm{LS}_{m,n}(r,s,\mu)$ is upper bounded with $\Delta_{r,s,\mu}(\mathcal{A}) \leq \Delta$ provided*

$$p > c_0\left(r(m+n-r)+s\right)\log\left(\left(1-\mu^2\frac{r^2 s}{mn}\right)^{-1/2}\frac{mn}{s}\right), \tag{3.7}$$

with probability at least $1-\exp\left(-c_1 p\right)$, where c_0, c_1 are constants that depend only on Δ.

Proof. By linearity of $\mathcal{A}$ and conicity of $\mathrm{LS}_{m,n}(r,s,\mu)$ assume without loss of generality $\|X\|_F = 1$ and consequently also $\|L\|_F \leq \tau$ and $\|S\|_F \leq \tau$ with $\tau := 1\Big/\sqrt{1-\mu^2\frac{r^2 s}{mn}}$ by Lemma 2.10 and by $\mu < \frac{\sqrt{mn}}{r\sqrt{s}}$.

Let (V_i, W_i, T_i) be an ε-covering of the subspaces of $\mathrm{LS}_{m,n}(r,s)$, that is also an ε-covering of the subspaces of $\mathrm{LS}_{m,n}(r,s,\mu)$, with the covering number $\mathfrak{R}(\varepsilon)$ bounded by Lemma 3.4. For every triple (V_i, W_i, T_i) define a subset of matrices

$$\mathcal{B}_i = \left\{X \in \Sigma_{m,n}\left(V, W, T_i, \mu\right) \,:\, \rho\left((V,W),(V_i,W_i)\right) \leq \varepsilon\right\}. \tag{3.20}$$

By (V_i, W_i, T_i) being an ε-covering of the subspaces and by the relation between $\Sigma_{m,n}(V,W,T,\mu)$ and $\mathrm{LS}_{m,n}(r,s,\mu)$ in (3.11), we have that $\mathrm{LS}_{m,n}(r,s,\mu) \subseteq \bigcup_i \mathcal{B}_i$. Therefore, if for all $\mathcal{B}_i$ the following holds

$$(\forall X \in \mathcal{B}_i): \quad (1-\bar{\Delta})\|X\|_F \leq \|\mathcal{A}(X)\| \leq (1+\bar{\Delta})\|X\|_F, \tag{3.21}$$

then necessarily $\bar{\Delta}_{r,s,\mu} \leq \bar{\Delta}$, proving that

$$\Pr(\bar{\Delta}_{r,s,\mu} \leq \bar{\Delta}) = \Pr\left(\forall X \in \mathrm{LS}_{m,n}(r,s,\mu) :\, (1-\bar{\Delta})\|X\|_F \leq \|\mathcal{A}(X)\| \leq (1+\bar{\Delta})\|X\|_F\right) \tag{3.22}$$

$$\geq \Pr\left((\forall i), (\forall X \in \mathcal{B}_i) :\, (1-\bar{\Delta})\|X\|_F \leq \|\mathcal{A}(X)\| \leq (1+\bar{\Delta})\|X\|_F\right), \tag{3.23}$$

where the inequality comes from the fact that $\mathrm{LS}_{m,n}(r,s,\mu)$ is a subset of $\bigcup_i \mathcal{B}_i$ and therefore the statement holds with less or equal probability. It remains to derive a lower bound on the probability in the equation (3.23) which in turn proves the theorem.

In the case that $\|\mathcal{A}\| \leq \frac{\bar{\Delta}}{2\tau\varepsilon} - 1 - \frac{\bar{\Delta}}{2}$, which we show later in (3.26) occurs with probability exponentially converging to one, rearranging the terms yields

$$\tau\varepsilon(1 + \bar{\Delta}/2 + \|\mathcal{A}\|) \leq \bar{\Delta}/2. \tag{3.24}$$

If the RIC holds for a fixed (V_i, W_i, T_i) with $\bar{\Delta}/2$, then by Lemma 3.3 in combination with (3.24) yields

$$(\forall X \in \mathcal{B}_i) : (1 - \bar{\Delta})\|X\|_F \leq \|\mathcal{A}\| \leq (1 + \bar{\Delta})\|X\|_F. \tag{3.25}$$

Therefore, using the probability union bound on (3.23) over all i's and the probability of $\|\mathcal{A}\|$ satisfying the bound $\varepsilon \leq \bar{\Delta}/(2\tau\,(1 + \|\mathcal{A}\|))$.

$$\Pr\left((\forall i), (\forall X \in \mathcal{B}_i) : (1 - \bar{\Delta})\|X\|_F \leq \|\mathcal{A}(X)\| \leq (1 + \bar{\Delta})\|X\|_F\right) \tag{3.26}$$

$$\geq 1 - \sum_i \Pr\left(\exists Y \in \Sigma_{m,n}(V_i, W_i, T_i, \mu) : \begin{array}{c} \|\mathcal{A}(Y)\| < (1 - \bar{\Delta}/2) \\ \text{or} \quad \|\mathcal{A}(Y)\| > (1 + \bar{\Delta}/2) \end{array}\right) \tag{3.27}$$

$$- \Pr\left(\|\mathcal{A}\| \geq \frac{\bar{\Delta}}{2\tau\varepsilon} - 1 - \frac{\bar{\Delta}}{2}\right). \tag{3.28}$$

The probability in (3.27) is bounded from above as

$$\sum_i \Pr\left(\exists Y \in \Sigma_{m,n}(V_i, W_i, T_i, \mu) : \begin{array}{c} \|\mathcal{A}(Y)\| < (1 - \bar{\Delta}/2) \\ \text{or} \quad \|\mathcal{A}(Y)\| > (1 + \bar{\Delta}/2) \end{array}\right) \tag{3.29}$$

$$\leq 2\Re(\varepsilon)\left(\frac{48}{\bar{\Delta}}\tau\right)^{r^2}\left(\frac{48}{\bar{\Delta}}\tau\right)^{s}\exp\left(-\frac{p}{2}\left(\frac{\bar{\Delta}^2}{32} - \frac{\bar{\Delta}^3}{192}\right)\right) \tag{3.30}$$

$$\leq 2\binom{mn}{s}\left(\frac{4\pi}{\varepsilon}\right)^{r(m+n-2r)}\left(\frac{48}{\bar{\Delta}}\tau\right)^{r^2+s}\exp\left(-\frac{p}{2}\left(\frac{\bar{\Delta}^2}{32} - \frac{\bar{\Delta}^3}{192}\right)\right), \tag{3.31}$$

where in the first inequality we used Lemma 3.2 and in the second inequality the bound on the ε-covering of the subspaces by Lemma 3.4.

In order to complete the lower bound in (3.26) it remains to upper bound (3.28) which we obtain by selecting the covering resolution ε sufficiently small so that the $\Pr\left(\|\mathcal{A}\| \geq \frac{\bar{\Delta}}{2\tau\varepsilon} - 1 - \frac{\bar{\Delta}}{2}\right)$ is exponentially small with the exponent proportional to the bound in (3.31). From condition (3.4) of Definition 3.2 we have that the random linear map satisfies

$$(\exists \gamma > 0) : \quad \Pr\left(\|\mathcal{A}\| \geq 1 + \sqrt{\frac{mn}{p}} + t\right) \leq \exp\left(-\gamma p t^2\right), \tag{3.32}$$

in particular

$$\Pr\left(\|\mathcal{A}\| \geq \frac{\bar{\Delta}}{2\tau\varepsilon} - 1 - \frac{\bar{\Delta}}{2}\right) \leq \exp\left(-\gamma p\left(\frac{\bar{\Delta}}{2\tau\varepsilon} - \frac{\bar{\Delta}}{2} - \sqrt{\frac{mn}{p}} - 2\right)^2\right). \tag{3.33}$$

Selecting the covering resolution ε

$$\varepsilon < \frac{\bar{\Delta}}{4\tau\left(\sqrt{mn/p} + 1 + \bar{\Delta}/4\right)}, \tag{3.34}$$

obtains the following exponentially small upper bound

$$\Pr\left(\|\mathcal{A}\| \geq \frac{\bar{\Delta}}{2\tau\varepsilon} - 1 - \frac{\bar{\Delta}}{2}\right) \leq \exp\left(-\gamma mn\right). \tag{3.35}$$

Returning to the inequality (3.26), combined with the bound on the first term in (3.31), and setting $\varepsilon = \bar{\Delta} \Big/ \left(4\tau\left(\sqrt{mn/p} + 1 + \bar{\Delta}/4\right)\right)$ in the second term of (3.31), such that (3.34) is satisfied, we have that

$$2\left(\frac{emn}{s}\right)^s \left(\frac{16\pi(\sqrt{mn/p} + 1 + \bar{\Delta}/4)}{\bar{\Delta}}\tau\right)^{r(m+n-2r)} \left(\frac{48}{\bar{\Delta}}\tau\right)^{r^2+s} \cdot \exp\left(-\frac{p}{2}\left(\frac{\bar{\Delta}^2}{32} - \frac{\bar{\Delta}^3}{192}\right)\right) \tag{3.36}$$

$$= \exp\Bigg(- pa(\bar{\Delta}) + r\left(m + n - 2r\right)\log\left(\sqrt{\frac{mn}{p}} + 1 + \frac{\bar{\Delta}}{4}\right) + r\left(m + n - 2r\right)\log\left(\frac{16\pi}{\bar{\Delta}}\tau\right) + \left(r^2 + s\right)\log\left(\frac{48}{\bar{\Delta}}\tau\right) + s\log\left(\frac{emn}{s}\right) + \log(2)\Bigg), \tag{3.37}$$

where we used the inequality $\binom{mn}{s} \leq \left(\frac{emn}{s}\right)^s$ and we define $a(\bar{\Delta}) := \bar{\Delta}^2/64 - \bar{\Delta}^3/384$. The 2^{nd}, 3^{rd} and 4^{th} terms in (3.37) can be bounded as

$$(\exists c_2 > 0): \quad 2^{nd} + 3^{rd} + 4^{th} \leq \left(c_2/a(\bar{\Delta})\right) r(m + n - r)\log\left(\frac{mn}{p}\tau\right), \tag{3.38}$$

and the 5^{th} and 6^{th} term of (3.37) as

$$(\exists c_3 > 0): \quad 5^{th} + 6^{th} \leq \left(c_3/a(\bar{\Delta})\right) s\log\left(\frac{mn}{s}\tau\right), \tag{3.39}$$

where c_2 and c_3 are dependent only on $\bar{\Delta}$. Therefore there exists a positive constant c_0 dependent only on $\bar{\Delta}$ such that if $p \geq c_0\left(r(m + n - r) + s\right)\log\left(\frac{mn}{s}\tau\right)$, then RICs are upper bounded by the constant $\bar{\Delta}$ with probability at least $e^{-c_0 p}$. By the inequality in (3.9) on page 45 and the discussion therein, the result also implies an upper bound on RICs with the squared norms Δ in Definition 3.1. □

3.5 SUMMARY AND DISCUSSION

In this chapter, we studied properties of random measurement operators obeying concentration of measure inequalities when applied to low-rank plus sparse matrices from the set $\mathrm{LS}_{m,n}(r, s, \mu)$. We discussed the connection of the concentration of measure inequalities to the Johnson-Lindenstrauss lemma, which allows for a reduction of the dimensionality of a point cloud without distorting the relative distances between individual points. We defined the restricted isometry constants (RICs) of an operator, which translate to a notion of an operator acting as an approximate isometry when restricted

to the set of incoherent low-rank plus sparse matrices. The main result of this chapter is Theorem 3.1 which says that measurement operators with a sufficient concentration of measure phenomenon have their RICs bounded independent of a problem size provided the number of measurements p is proportional to $O\left(r\left(m+n-r\right)+s\right)$ times a logarithmic factor provided $\mu < \frac{\sqrt{mn}}{r\sqrt{s}}$. The theorem has important consequences in Chapter 4 where we propose computational methods to recover low-rank plus sparse matrices $X_0 \in \mathrm{LS}_{m,n}(r,s,\mu)$ from subsampled measurements $b = \mathcal{A}(X_0)$ for operators $\mathcal{A}$ whose RICs are bounded. Results of this chapter also illustrate how RICs can be developed for more complex additive data models.

3.6 SUPPORTING LEMMATA

This section gives additional lemmata used in this chapter. What follows is the proof of Lemma 3.2 that uses similar arguments as Lemma 5.1 by Baraniuk et al. (2008) and Lemma 4.3 by Recht et al. (2010) with the exception that here we consider two subsets, one for the low rank and another for the sparse component.

Lemma 3.2 (RICs for a fixed LS subspace $\Sigma_{m,n}(V,W,T,\mu)$)**.** *Let $\mathcal{A}:\mathbb{R}^{m\times n}\to\mathbb{R}^{p}$ be a nearly isometric random linear map from Definition 3.2 and $\Sigma_{m,n}\left(V,W,T,\mu\right)$ as defined in (3.10) is fixed for some (V,W,T), and $\mu\in\left[1,\frac{\sqrt{mn}}{r\sqrt{s}}\right)$. Then for any $\bar{\Delta}\in(0,1)$*

$$\forall X\in\Sigma_{m,n}\left(V,W,T,\mu\right):\quad (1-\bar{\Delta})\|X\|_F\le\|\mathcal{A}(X)\|\le(1+\bar{\Delta})\|X\|_F, \tag{3.12}$$

with probability at least

$$1-2\left(\frac{24}{\bar{\Delta}}\tau\right)^{\dim V\cdot\dim W}\left(\frac{24}{\bar{\Delta}}\tau\right)^{\dim T}\exp\left(-\frac{p}{2}\left(\frac{\bar{\Delta}^2}{8}-\frac{\bar{\Delta}^3}{24}\right)\right), \tag{3.13}$$

where $\tau=1\Big/\sqrt{1-\mu^2\frac{r^2s}{mn}}$.

Proof. By the linearity of $\mathcal{A}(\cdot)$ and conicity of $\Sigma_{m,n}(V,W,T,\mu)$ we can assume without loss of generality that $\|X\|_F=1$. By Lemma 2.10 with $\|X\|_F=1$ and $\mu<\frac{\sqrt{mn}}{r\sqrt{s}}$, we can bound the Frobenius norm of the low-rank and the sparse component as $\|L\|_F\le\tau$ and $\|S\|_F\le\tau$, where $\tau=1\Big/\sqrt{1-\mu^2\frac{r^2s}{mn}}$.

There exist two finite $(\bar{\Delta}/8)$-coverings of the two matrix sets with bounded norms

$$\left\{L\in\mathbb{R}^{m\times n}\ :\ \mathrm{Col}\left(\mathrm{L}\right)\subseteq V,\ \mathrm{Col}\left(\mathrm{L}^{\mathrm{T}}\right)\subseteq W,\ \|L\|_F\le\tau\right\} \tag{3.40}$$

$$\left\{S\in\mathbb{R}^{m\times n}\ :\ S\subseteq T,\ \|S\|_F\le\tau\right\}, \tag{3.41}$$

that we denote Λ^L,Λ^S and by (Lorentz et al., 1996, Chapter 13) they are subsets of the two sets in (3.40) and (3.41), and their covering numbers are

upper bounded as

$$\left|\Lambda^L\right| \leq \left(\frac{24}{\bar{\Delta}}\tau\right)^{\dim V \cdot \dim W} \qquad \left|\Lambda^S\right| \leq \left(\frac{24}{\bar{\Delta}}\tau\right)^{\dim T}. \tag{3.42}$$

Let $\Lambda := \left\{Q^L + Q^S : Q^L \in \Lambda^L,\, Q^S \in \Lambda^S\right\}$ be the set of sums of all possible pairs of the two coverings. The set Λ is a $(\bar{\Delta}/4)$-covering of the set $\Sigma_{m,n}\left(V, W, T, \mu\right)$ since for all $X \in \Sigma_{m,n}\left(V, W, T, \mu\right)$ there exists a pair $Q \in \Lambda$ such that

$$\|X - Q\|_F = \left\|L + S - \left(Q^L + Q^S\right)\right\|_F \tag{3.43}$$

$$\leq \left\|L - Q^L\right\|_F + \left\|S - Q^S\right\|_F \leq \frac{\bar{\Delta}}{8} + \frac{\bar{\Delta}}{8}, \tag{3.44}$$

where in the first line we used the fact that X can be expressed as $L + S$, and in the second line we applied the triangular inequality combined with the Q^L, Q^S being $(\bar{\Delta}/8)$-coverings of the matrix sets for the low-rank component and the sparse component respectively.

Applying the probability union bound on concentration of measure of $\mathcal{A}$ as in (3.3) with $\varepsilon = \bar{\Delta}/2$ gives that

$$(\forall Q \in \Lambda) : \left(1 - \frac{\bar{\Delta}}{2}\right)\|Q\|_F \leq \|\mathcal{A}(Q)\|_2 \leq \left(1 + \frac{\bar{\Delta}}{2}\right)\|Q\|_F, \tag{3.45}$$

holds with the probability at least

$$1 - 2\left(\frac{24}{\bar{\Delta}}\tau\right)^{\dim V \cdot \dim W}\left(\frac{24}{\bar{\Delta}}\tau\right)^{\dim T} \exp\left(-\frac{p}{2}\left(\frac{\bar{\Delta}^2}{8} - \frac{\bar{\Delta}^3}{24}\right)\right). \tag{3.46}$$

By $\Sigma_{m,n}(V, W, T, \mu)$ being a closed set, the maximum

$$M = \max_{Y \in \Sigma_{m,n}(V,W,T,\mu),\, \|Y\|_F = 1} \|\mathcal{A}(Y)\|_2, \tag{3.47}$$

is attained. Then there exists $Q \in \Lambda$ such that

$$\|\mathcal{A}(X)\|_2 \leq \|\mathcal{A}(X)\|_2 + \|\mathcal{A}(X - Q)\|_2 \leq 1 + \frac{\bar{\Delta}}{2} + M\frac{\bar{\Delta}}{4}, \tag{3.48}$$

where the first inequality comes from applying the triangle inequality to X and $Q - X$ and in the second inequality we used (3.45) to upper bound $\|\mathcal{A}(X)\|_2$ since $(X - Q) \in \Sigma_{m,n}(V, W, T, \mu)$ by Lemma 2.9 and the upper bound of $\|X - Q\|_F$ comes from $Q \in \Lambda$ combined with Λ being a $(\bar{\Delta}/4)$-covering. Note that the inequality (3.48) holds for all $X \in \Sigma_{m,n}(V, W, T, \mu)$ whose Frobenius norm $\|X\|_F = 1$ and thus also for a matrix $\widehat{X}$ for which the maximum in (3.47) is attained. The inequality in (3.48) applied to the matrix that attains the maximum $\widehat{X}$ yields

$$M \leq 1 + \frac{\bar{\Delta}}{2} + M\frac{\bar{\Delta}}{4} \quad \Longrightarrow \quad M \leq 1 + \bar{\Delta}. \tag{3.49}$$

The lower bound follows from the reverse triangle inequality

$$\|\mathcal{A}(X)\|_2 \geq \|\mathcal{A}(Q)\|_2 - \|\mathcal{A}(X-Q)\|_2 \geq \left(1 - \frac{\bar{\Delta}}{2}\right) - (1+\bar{\Delta})\frac{\bar{\Delta}}{4} \geq 1 - \bar{\Delta} \quad (3.50)$$

where the second inequality comes from $\|\mathcal{A}(X-Q)\|_2 \leq M \|X-Q\|_F \leq (1+\bar{\Delta})\frac{\bar{\Delta}}{4}$ by (3.47) combined with Q being an element of a $(\bar{\Delta}/4)$-covering.

Combining (3.48) with the bound on M in (3.49) gives the upper bound and (3.50) gives the lower bound on $\|\mathcal{A}(X)\|_2$ completing the proof. □

The following result by Szarek (1998) gives a covering number for the Grassmanian. For proof see (Szarek, 1998, Theorem 8).

Lemma 3.5 (ε-covering of the Grassmannian)**.** *Let $\big(\mathcal{G}(D,d), \rho(\cdot,\cdot)\big)$ be a metric space on a Grassmannian manifold $\mathcal{G}(D,d)$ with the metric ρ as defined in (3.14). Then there exists ε-covering $\mathcal{G}(D,d)$ with $\Lambda = \{U_i\}_{i=1}^N \subset \mathcal{G}(D,d)$ such that*

$$\forall U \in \mathcal{G}(D,d): \quad \min_{\widehat{U}\in\Lambda} \rho(U,\widehat{U}) \leq \varepsilon, \quad (3.51)$$

and $N \leq \left(\frac{C_0}{\varepsilon}\right)^{d(D-d)}$ with C_0 independent of ε, bounded by $C_0 \leq 2\pi$.

The above bound on the covering number of the Grassmannian is used to bound the covering number of the set $\mathrm{LS}^{\tau}_{m,n}(r,s)$ in the proof of the following Lemma 3.4.

Lemma 3.4 (Covering number of the subspaces of $\mathrm{LS}_{m,n}(r,s)$)**.** *The covering number $\mathfrak{R}(\varepsilon)$ of the subspaces of the set $\mathrm{LS}_{m,n}(r,s)$ is bounded above by*

$$\mathfrak{R}(\varepsilon) \leq \binom{mn}{s}\left(\frac{4\pi}{\varepsilon}\right)^{r(m+n-2r)}. \quad (3.19)$$

Consequently, by $\mathrm{LS}_{m,n}(r,s,\mu) \subset \mathrm{LS}_{m,n}(r,s)$, we also obtained an upper bound on the covering number for the subspaces of the set of incoherent low-rank plus sparse matrices $\mathrm{LS}_{m,n}(r,s,\mu)$.

Proof. By Lemma 3.5 there exist two finite $(\varepsilon/2)$-coverings $\Lambda_1 := \{V_i\}_{i=1}^{|\Lambda_1|} \subseteq \mathcal{G}(m,r)$ and $\Lambda_2 := \{W_i\}_{i=1}^{|\Lambda_2|} \subseteq \mathcal{G}(n,r)$, with their covering numbers upper bounded as

$$|\Lambda_1| \leq \left(\frac{4\pi}{\varepsilon}\right)^{r(m-r)} \qquad |\Lambda_2| \leq \left(\frac{4\pi}{\varepsilon}\right)^{r(n-r)}, \quad (3.52)$$

as given in (Recht et al., 2010, (4.18)) that uses (Szarek, 1998, Theorem 8). By Λ_1, Λ_2 being $(\varepsilon/2)$-coverings

$$\forall V \in \mathcal{G}(m,r): \quad \exists V_i \in \Lambda_1, \quad \rho(V,V_i) \leq \varepsilon/2, \quad (3.53)$$

$$\forall W \in \mathcal{G}(n,r): \quad \exists W_i \in \Lambda_2, \quad \rho(W,W_i) \leq \varepsilon/2. \quad (3.54)$$

Let $\Lambda_3 = \mathcal{V}(mn, s)$ where $\mathcal{V}(mn, s)$ is the set of all possible support sets of an $m \times n$ matrix that has s elements. Thus the cardinality of Λ_3 is $\binom{mn}{s}$.

Construct $\Lambda = (\Lambda_1 \times \Lambda_2 \times \Lambda_3)$ where $\times$ denotes the Cartesian product. Choose any $V \in \mathcal{G}(m, r), W \in \mathcal{G}(n, r)$ and $T \in \mathcal{V}(mn, s)$ for which we now show there exists $\left(\widehat{V}, \widehat{W}, \widehat{T}\right) \in \Lambda$ such that $\rho\left((V, W), \left(\widehat{V}, \widehat{W}\right)\right) \leq \varepsilon$ and $T = \hat{T}$, thus showing that the set Λ is an ε-covering of $\mathrm{LS}_{m,n}(r, s, \mu)$.

Satisfying $T = \widehat{T}$ comes from $\Lambda_3 = \mathcal{V}(mn, s)$ containing all support sets with at most s entries. The projection operator onto the pair (V, W) can be written as $P_{(V,W)} = P_V \otimes P_W$, so for the two pairs of subspaces (V, W) and $(\widehat{V}, \widehat{W})$ we have the following

$$\begin{aligned}
\rho\left((V, W), \left(\widehat{V}, \widehat{W}\right)\right) &= \|P_{(V,W)} - P_{(\widehat{V},\widehat{W})}\| && (3.55)\\
&= \|P_V \otimes P_W - P_{\widehat{V}} \otimes P_{\widehat{W}}\| && (3.56)\\
&= \| \left(P_V - P_{\widehat{V}}\right) \otimes P_W + P_{\widehat{V}} \left(P_W - P_{\widehat{W}}\right) \| && (3.57)\\
&\leq \|P_V - P_{\widehat{V}}\| \|P_W\| + \|P_{\widehat{V}}\| \|P_W - P_{\widehat{W}}\| && (3.58)\\
&= \rho\left(V, \widehat{V}\right) + \rho\left(W, \widehat{W}\right). && (3.59)
\end{aligned}$$

By Λ_1 and Λ_2 being $(\varepsilon/2)$-coverings, we have that for any V, W exist $\widehat{V} \in \Lambda_1$ and $\widehat{W} \in \Lambda_2$, such that $\rho\left((V, W), \left(\widehat{V}, \widehat{W}\right)\right) \leq \rho\left(V, \widehat{V}\right) + \rho\left(W, \widehat{W}\right) \leq \varepsilon$. Using the bounds on the cardinality of Λ_1, Λ_2 in (3.52) combined with $|\Lambda_3| = \binom{mn}{s}$ yields that the cardinality of Λ is bounded above by

$$\Re(\varepsilon) = |\Lambda_1| \, |\Lambda_2| \, |\Lambda_3| \leq \binom{mn}{s} \left(\frac{4\pi}{\varepsilon}\right)^{r(m+n-2r)}. \tag{3.60}$$

□

Lemma 3.3 (Variation of $\bar{\Delta}$ in RIC in respect to a perturbation of (V, W))**.** *Let $\Sigma_1 := \Sigma_{m,n}(V_1, W_1, T, \mu)$ and $\Sigma_2 := \Sigma_{m,n}(V_2, W_2, T, \mu)$ be two low-rank plus sparse subspaces with the same fixed sparse subspace T and $\mu \in \left[1, \frac{\sqrt{mn}}{r\sqrt{s}}\right)$. Suppose that for $\bar{\Delta} > 0$, the linear operator $\mathcal{A}$ satisfies*

$$\forall X \in \Sigma_1 : \quad (1 - \bar{\Delta})\|X\|_F \leq \|\mathcal{A}(X)\| \leq (1 + \bar{\Delta})\|X\|_F. \tag{3.17}$$

Then

$$\forall Y \in \Sigma_2 : \quad (1 - \bar{\Delta}')\|Y\|_F \leq \|\mathcal{A}(Y)\| \leq (1 + \bar{\Delta}')\|Y\|_F, \tag{3.18}$$

with $\bar{\Delta}' := \bar{\Delta} + \tau\rho\left((V_1, W_1), (V_2, W_2)\right)\left(1 + \bar{\Delta} + \|\mathcal{A}\|\right)$ with ρ as defined in (3.14) and $\tau = 1 \Big/ \sqrt{1 - \mu^2 \frac{r^2 s}{mn}}$.

Proof. Recall the notation used in Lemma 3.3 that there are sets $\Sigma_1 := \Sigma^{\tau}_{m,n}(V_1, W_1, T)$ and $\Sigma_2 := \Sigma^{\tau}_{m,n}(V_2, W_2, T)$ which are subsets of $\mathrm{LS}^{\tau}_{m,n}(r, s)$ with a shared support T of the sparse component.

Let $Y \in \Sigma_2$, so we can write $Y = L + S$ such that $\text{supp}(S) = T, \text{Col}\,(\text{L}) \subseteq V_2, \text{Col}\left(\text{L}^{\text{T}}\right) \subseteq W_2$ and $\|L\|_F \leq \tau \|X\|_F$. By linearity of $\mathcal{A}$ assume without loss of generality $\|Y\|_F = 1$ and therefore $\|L\|_F \leq \tau$. Denote $U_1 = (V_1, W_1)$ and $U_2 = (V_2, W_2)$ and let P_{U_i} be an orthogonal projection onto the space of matrices whose column and row space is defined by V_i, W_i such that left and right singular vectors of $P_{U_i} Y$ lie in V_i respectively W_i. Then

$$\begin{aligned}
\|\mathcal{A}(Y)\| &= \|\mathcal{A}(L+S)\| = \|\mathcal{A}\left(P_{U_1} L + S - (P_{U_1} L - P_{U_2} L)\right)\| && (3.61)\\
&\leq \|\mathcal{A}\left(P_{U_1} L + S\right)\| + \|\mathcal{A}\left(\left[P_{U_1} - P_{U_2}\right] L\right)\| && (3.62)\\
&\leq (1+\bar{\Delta}) \|P_{U_1} L + S\| + \|\mathcal{A}\| \rho\left(U_1, U_2\right) \|L\| && (3.63)\\
&= (1+\bar{\Delta}) \|P_{U_2} L + S + \left[P_{U_1} - P_{U_2}\right] L\| + \|\mathcal{A}\| \rho\left(U_1, U_2\right) \|L\| && (3.64)\\
&\leq (1+\bar{\Delta}) \left(\|Y\|_F + \rho(U_1, U_2) \|L\|\right) + \|\mathcal{A}\| \rho\left(U_1, U_2\right) \|L\| && (3.65)\\
&\leq \|Y\|_F \left(1 + \bar{\Delta} + \tau \rho(U_1, U_2) \left(1 + \bar{\Delta} + \|\mathcal{A}\|\right)\right), && (3.66)
\end{aligned}$$

where in the first line (3.61) we use the fact that $P_{U_2} L = L$, the second line (3.62) follows by the triangle inequality and linearity of $\mathcal{A}$, and in the third inequality we bound the effect of $\mathcal{A}$ on $(P_{U_1} L + S)$ using the RICs of $\mathcal{A}$ combined with the definition of ρ in (3.14). We proceed in (3.64) and (3.65) by projecting L to space U_2 and again bounding the effect of $\mathcal{A}$ on $(P_{U_2} L + S)$. Finally, in (3.66) we use $\|L\|_F \leq \tau$. We obtain a similar lower bound using the reverse triangular inequality

$$\begin{aligned}
\|\mathcal{A}(Y)\| &= \|\mathcal{A}\left(P_{U_1} L + S - (P_{U_1} L - P_{U_2} L)\right)\| && (3.67)\\
&\geq \|\mathcal{A}\left(P_{U_1} L + S\right)\| - \|\mathcal{A}\left(\left[P_{U_1} - P_{U_2}\right] L\right)\| && (3.68)\\
&\geq \left(1 - \bar{\Delta}\right) \|P_{U_1} L + S\| - \|\mathcal{A}\| \rho(U_1, U_2) \|L\|_F && (3.69)\\
&= \left(1 - \bar{\Delta}\right) \|P_{U_2} L + S - \left[P_{U_2} - P_{U_1}\right] L\| - \|\mathcal{A}\| \rho(U_1, U_2) \|L\|_F && (3.70)\\
&\geq \left(1 - \bar{\Delta}\right) \left(\|Y\|_F - \rho(U_1, U_2) \|L\|_F\right) - \|\mathcal{A}\| \rho(U_1, U_2) \|L\|_F && (3.71)\\
&\geq \|Y\|_F \left(1 - \bar{\Delta} - \tau \rho(U_1, U_2)(1 - \bar{\Delta} + \|\mathcal{A}\|)\right). && (3.72)
\end{aligned}$$

Combining (3.66) and (3.72) yields

$$\forall Y \in \Sigma_2 : \quad (1 - \bar{\Delta}')\|Y\|_F \leq \|\mathcal{A}(Y)\| \leq (1 + \bar{\Delta}')\|Y\|_F, \tag{3.73}$$

with $\bar{\Delta}' = \bar{\Delta} + \tau\rho(U_1, U_2)\left(1 + \bar{\Delta} + \|\mathcal{A}\|\right)$. □

4 ALGORITHMS FOR LOW-RANK PLUS SPARSE MATRIX SENSING

SYNOPSIS

In this chapter, we show that incoherent low-rank plus sparse matrices can be recovered by computationally tractable methods with sample complexity $p > O(r(m + n - r) + s)$ times a logarithmic factor when the number of corruptions is less than $mn/(\mu^2 r^2)$, which is equivalent to the optimal order of the number of corruptions in the Robust PCA literature. We show that an upper bound on the restricted isometry constants (RICs) of an operator implies uniqueness of the solution to the problem of incoherent matrix recovery. Additionally, we show that semidefinite programming and two gradient descent algorithms, NIHT and NAHT, converge to the measured matrix provided the RICs of the measurement operator are sufficiently small. The convex relaxation and NAHT also provably solve Robust PCA with the optimal sparsity bound when the sensing operator is chosen to be the identity. We perform numerical experiments in which we observe a phase transition in the space of parameters for which the methods succeed. We also provide two exemplar applications on dynamic-foreground/static-background separation and multispectral imaging.

4.1 INTRODUCTION

While the preceding chapters have focused on the theoretical properties of low-rank plus sparse matrix sets and random linear maps acting on them, the central objective of this chapter is practical: Our goal is to devise a computationally practical way of recovering an unknown low-rank plus sparse matrix X_0 given a vector of measurements $b = \mathcal{A}(X_0)$ and a subsampling operator $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$.

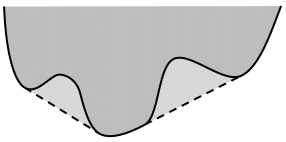
The task is a mixture of a *compressed sensing* and a *low-rank matrix sensing* problem, which deal with the recovery of sparse and of low-rank matrix respectively, from a limited number of measurements. The additional challenge of the combined problem of recovering a matrix with the additive structure of a low-rank and a sparse matrix comes in the difficulty of distinguishing between the two components.

Solving compressed sensing and matrix sensing problems is generally NP-hard, see (Foucart and Rauhut, 2013, §2.3) and (Harvey et al., 2006; Hardt et al., 2014) respectively. However, it is well known that in many cases both compressed sensing and matrix sensing can be efficiently solved. Notably, if the measurement operator has its restricted isometry constants (RIC) sufficiently upper bounded as discussed in Chapter 3, the non-convexity can be overcome by computationally tractable methods.

There is now a wide body of literature of compressed sensing and matrix sensing algorithms many of which have provable convergence guarantees and come with fast software implementations. Most of the methods fall into one of the two categories:

(i) *Convex relaxation approach* is based on formulating a convex optimisation problem that shares the same global minimum with the non-convex problem. The convex relaxation approach relies on approximating the non-convex function with its *convex envelope*. The *convex envelope* of a (possibly non-convex) function $f : C \to \mathbb{R}$ is defined as the largest convex function $g : C \to \mathbb{R}$ such that $g(x) \leq f(x)$ for all $x \in C$. This means, that if g can be conveniently evaluated, it serves as an approximation to f that can be minimized efficiently. Subsequently, a number of algorithms provably converge to a global minimum of the convex problem.

(ii) *Non-convex approach* solves the recovery problem directly in its non-convex formulation. Non-convex approaches are computationally faster and in some cases are empirically observed to be able to recover matrices of higher ranks than the convex methods (Blanchard et al., 2015).



The convex envelope of a non-convex function.

The challenge in designing an algorithm for recovery of a sparse or a low-rank solution is two-fold: (i) overcoming the non-convex geometry and (ii) proving it converges to a minimum. While in the convex relaxation approach, these two aspects are dealt with separately, the non-convex techniques tackle them jointly, as a part of a single convergence analysis.

The task of recovering a matrix that is formed as the sum of a low-rank and a sparse matrix comes with the additional difficulty of distinguishing between the two components which can become correlated. Recall the central topic of Chapter 2, that there can be sequences of low-rank plus sparse matrices converging outside of the feasible set while both components become correlated and their norm diverges. The issue can be alleviated by closing the set by restraining the Frobenius norm of the low-rank component as stated in Definition 1.1, page 8. The task is then to find X such that

$$\mathcal{A}(X) = b \qquad \text{and} \qquad X \in \mathrm{LS}_{m,n}(r, s, \mu), \tag{4.1}$$

where $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$ is the subsampling operator and $b \in \mathbb{R}^p$ is the vector of measurements.

For ease of reference:

$$\mathrm{LS}_{m,n}(r, s, \mu) = \left\{ \begin{array}{c} X = L + S : \\ \mathrm{rank}(L) \leq r, \\ \|S\|_0 \leq s, \\ \|U^T e_i\|_2 \leq \sqrt{\frac{\mu r}{m}}, \\ \|V^T f_i\|_2 \leq \sqrt{\frac{\mu r}{n}} \end{array} \right\}$$

where $L = U\Sigma V^T$.

$$\forall X \in \mathrm{LS}_{m,n}(r, s, \mu): \quad (1 - \Delta)\|X\|_F^2 \leq \|\mathcal{A}(X)\|_2^2 \leq (1 + \Delta)\|X\|_F^2,$$

where $\Delta := \Delta_{r,s,\mu}(\mathcal{A})$ is the RIC of $\mathcal{A}$ and $\mu < \frac{\sqrt{mn}}{r\sqrt{s}}$.

For a linear transform $\mathcal{A}$ which has its RIC suitably upper bounded and a given vector of samples $b = \mathcal{A}(X_0)$, the matrix X_0 is the only matrix in the set $\mathrm{LS}_{m,n}(r, s, \mu)$ that satisfies the linear constraint.

Theorem 4.1 (Existence of a unique solution for $\mathcal{A}$ with RIC)**.** *Suppose that* $\Delta_{2r,2s,\mu}(\mathcal{A}) < 1$ *for some integers* $r, s \geq 1$ *and* $\mu < \sqrt{mn}/\left(r\sqrt{s}\right)$*. Let* $b = \mathcal{A}(X_0)$*, then* X_0 *is the only matrix in the set* $\mathrm{LS}_{m,n}(r, s, \mu)$ *satisfying* $\mathcal{A}(X) = b$*.*

Proof. Assume, on the contrary, that there exists a matrix $X \in \mathrm{LS}_{m,n}(r, s, \mu)$ such that $\mathcal{A}(X) = b$ and $X \neq X_0$. Then $Z := X_0 - X$ is a non-zero matrix and $Z \in \mathrm{LS}_{m,n}(2r, 2s, \mu)$ by Lemma 2.9 and $\mathcal{A}(Z) = 0$. But then by the RIC we would have $0 = \|\mathcal{A}(Z)\|_2^2 \geq (1 - \Delta_{2r,2s,\mu}) \|Z\|_F^2 > 0$, which is a contradiction. □

The unique low-rank plus sparse matrix X_0 can be recovered by the non-convex optimisation

$$\min_{X \in \mathbb{R}^{m \times n}} \|\mathcal{A}(X) - b\|_F, \qquad \text{s.t.} \quad X \in \mathrm{LS}_{m,n}(r, s, \mu). \tag{4.2}$$

Same as in compressed sensing and matrix sensing, the problem in (4.2) is NP-hard. This follows from the non-convexity of the feasible set of low-rank plus sparse matrices. However, if the RIC of $\mathcal{A}$ are sufficiently upper-bounded when restricted to matrices in $\mathrm{LS}_{m,n}(r, s, \mu)$, then the solution can be obtained by two iterative gradient descent algorithms, Normalized Iterative Hard Thresholding (NIHT) and Normalized Alternating Hard Thresholding (NAHT), which provably converge to a global minimum of (4.2). Alternatively, X_0 can be recovered by solving the convex relaxation

$$\min_{X = L + S \in \mathbb{R}^{m \times n}} \|L\|_* + \lambda \|S\|_1, \qquad \text{s.t.} \quad \|\mathcal{A}(L + S) - b\|_2 \leq \varepsilon_b, \tag{4.3}$$

where $\|\cdot\|_*$ is the Schatten 1-norm and $\|\cdot\|_1$ is the sum of the absolute value of the entries and ε_b is the model misfit.

Our use of $\|\cdot\|_1$ as the sum of the modulus of the entries of a matrix differs from the vector induced 1-norm of a matrix.

The rest of this chapter is organised as follows. Firstly, we review the most popular algorithmic approaches in compressed sensing and low-rank matrix sensing in §4.2. In §4.3, we prove that low-rank plus sparse matrices can be recovered by solving the convex optimisation in (4.3). In §4.4, we introduce two fast non-convex algorithms NIHT and NAHT, which are natural extensions of algorithms developed for compressed sensing by Blumensath and Davies (2010) and matrix completion by Tanner and Wei (2013), and prove their convergence to the global minimum of (4.2). In §4.5, we empirically study the recovery of the average case on synthetic data by solving convex optimisation and by the proposed gradient descent methods and observe a phase transition in the space of parameters for which the methods succeed. We give an example of two practical applications of the low-rank plus sparse matrix recovery in the form of a subsampled dynamic-foreground/static-background video separation and robust recovery of multispectral imagery in §4.6.

4.2 RELATION TO PRIOR ALGORITHMIC WORK

Herein we provide the reader with a brief overview of algorithmic approaches in compressed sensing, low-rank matrix sensing, and Robust PCA.

4.2.1 *Compressed sensing algorithms*

In compressed sensing, the convex envelope of the ℓ_0-norm is the ℓ_1-norm leading to the following *linear programming* (LP) problem

$$\min_{x \in \mathbb{R}^n} \|x\|_1, \qquad \text{s.t.} \quad \mathcal{A}(x) = b, \tag{4.4}$$

refered to as *Basis Pursuit* (BP) (Chen et al., 2001). The compressed sensing relaxation in (4.4) can be solved by the simplex algorithm (Dantzig, 1963) which has a cubic complexity for the average case (Borgwardt, 1987) despite its worst-case complexity being exponential (Dantzig and Thapa, 1998). Cubic complexity is impractical in many applications where the signal dimension is large. Chen and Donoho (1994) proposed the use of the *interior point method* which solves a sequence of problems which converge to the solution of (4.4). The interior point method has been improved upon (Candès and Romberg, 1995; Kim et al., 2007) but is still costly when the $\mathcal{A}(\cdot)$ is dense which is often the case in compressed sensing.

The most successful methods for ℓ_1-minimization are gradient methods that alternate between the competing goals of satisfying the linear constraint and minimizing the sparsity promoting ℓ_1-norm. Many iterative soft-thresholding methods have been proposed for solving (4.4) under different names, such as iterative thresholding (Daubechies et al., 2004), forward-backward splitting (Combettes and Wajs, 2005), fixed-point iteration (Hale et al., 2008), and sparse reconstruction by separable approximation (SpaRSA) (Wright et al., 2009b). Convergence analysis of these methods often relies on taking small gradient steps, whereas in practice, the optimal performance is attained by taking large steps, for example as the Barzilai-Borwein criterion (Wright et al., 2009b). The convergence can be sped up by taking into the account the information from the previous steps, which leads to two-step IST (TwIST) (Bioucas-Dias and Figueiredo, 2007), and even faster global convergence rate is observed and proved by the fast IST algorithm (FISTA) (Beck and Teboulle, 2009).

A number of greedy algorithms that solve directly the non-convex optimisation over the feasible set of sparse vectors have been proposed. Convergence analysis of non-convex optimisation is hindered by the fact that algorithms might encounter local minima and saddle points. Despite the non-convexity, many iterative gradient schemes can be showed to provably converge to a global minimum if the measurement operator RICs are sufficiently upper bounded. An early example of such an algorithm is Compressive Sampling Matching Pursuit (CoSaMP), which solves a sequence of least-squares problems supported on the indices of the largest residual

absolute values (Needell and Tropp, 2009). Blumensath and Davies (2009) show that a simple iterative hard thresholding algorithm (IHT), that alternates between minimizing the objective and projecting the estimates on the non-convex constraint, provably converges to a global minimum with fixed step sizes and if the RICs are sufficiently bounded. The ideas of CoSAMP and IHT were combined by Foucart (2011) who designs Hard Thresholding Pursuit (HTP) which has faster empirically convergence and more relaxed bound on the RICs guaranteeing the convergence.

The choice of the step size in hard-thresholding algorithms is crucial: if the step size is too small, the algorithm converges very slowly, while a step size too big, might not converge at all or hinder the convergence analysis. An important innovation was made by Blumensath and Davies (2010), who suggested choosing a step size based on the assumption that the support set of iterates does not change too much in subsequent iterations. Tanner and Wei (2013) streamlined the convergence analysis and also proposed an analogous method for matrix completion and matrix sensing. The work of Blanchard et al. (2015) extends this approach further and designs a hard-thresholding version of a conjugate gradients method to compute not only the optimal stepsize but also the optimal descent direction.

4.2.2 *Matrix sensing/completion algorithms*

Fazel (2002) showed that the Schatten-1 norm, also referred to as the nuclear norm, is the convex envelope of the rank function. The convex relaxation of the low-rank matrix sensing is formulated in the following *semidefinite programming* (SDP) problem

$$\min_{X \in \mathbb{R}^{m \times n}} \|X\|_*, \qquad \text{s.t.} \quad \mathcal{A}(X) = b \tag{4.5}$$

and can be readily solved by a multitude of algorithms and software packages such as SDPT3 (Toh et al., 1999) and SeDuMi (Sturm, 1999) as part of the modelling framework CVX (Grant and Boyd, 2014, 2008) for Matlab. However, to recover an $n \times n$ matrix, we have to solve an SDP with $2\left(n^4\right)$ variables that can have complexity as large as $O\left(n^6\right)$ making it infeasible for even moderately large matrices (Nesterov, 2004).

A computationally faster way to solve the optimisation in (4.5) is to employ gradient descent algorithms. Cai et al. (2010) introduced the Singular Value Thresholding algorithm that alternates between a gradient descent steps and a soft-thresholding operation on the singular values of the matrix iterates. This method has been improved upon by Lin et al. (2009) and Toh and Yun (2010) who develop an acceleration strategy for the gradient step. The work of Ma et al. (2011) speeds up the soft-thresholding operation by computing only an approximate SVD. However, all of the algorithms for (4.5) apply the soft-thresholding operator which involves a costly computation of the full SVD in every iteration.

The computationally fastest methods to the matrix sensing/completion solve directly the non-convex formulation. Many of the algorithms are analogous to their compressed sensing counterparts. These include the Matrix ALPS family (Kyrillidis and Cevher, 2014) and Atomic Decomposition for Minimum Rank Approximation (ADMiRA) (Lee and Bresler, 2010) which is the matrix extension of the compressed sensing algorithm CoSaMP (Needell and Tropp, 2009). There are numerous other hard-thresholding algorithms based on the fact that a set of fixed-rank matrices forms the Grassmannian manifold (Keshavan et al., 2010; Meyer et al., 2011; Vandereycken, 2013). Tanner and Wei (2013) proposed a hard-thresholding algorithm that adaptively finds the optimal step-size based the singular vectors of previous matrix iterates. Blanchard et al. (2015) extend this approach and design a hard-thresholding version of a conjugate gradients method to compute not only the optimal stepsize but also the optimal descent direction. Another popular approach is to parameterise the set of low-rank matrices through a rank enforcing factorisation, which avoids the need to compute an SVD altogether. Examples of such algorithms include Power Factorization (Haldar and Hernando, 2009), Low-Rank Matrix Fitting (Wen et al., 2012), and Alternating Steepest Descent (Tanner and Wei, 2016).

See the review by Davenport and Romberg (2016) for a comprehensive survey of low-rank matrix completion and sensing algorithms.

4.2.3 *Robust PCA algorithms*

Robust PCA is closely related to the low-rank plus sparse matrix sensing problem. The only difference is that while in the former we have access to the full matrix X_0, in the latter we have access only to the subsampled measurements $b = \mathcal{A}(X_0)$.

The early algorithms for Robust PCA were based on various heuristics to identify the support of the corrupted entries that form the matrix S, however these methods lack convergence guarantees (Torre and Black, 2001, 2003). The advances in compressed sensing and matrix sensing/completion motivated the work of Candès et al. (2011) and Chandrasekaran et al. (2011), who analysed the convex relaxation of the original problem as posed in (1.13) on page 8,

$$\min_{X=L+S\in\mathbb{R}^{m\times n}} \|L\|_* + \lambda \|S\|_1 , \qquad \text{s.t.} \qquad L + S = M, \tag{4.6}$$

and proved guarantees for exact recovery of L and S provided the *incoherence* parameter μ of the low-rank component is sufficiently small and with some additional assumptions on the *sparsity ratio* parameter α for the sparse component.

See §2.2, Definition 2.1 of the *incoherence* parameter μ, Definition 2.2 of the *sparsity ratio* parameter α and the discussion on regularisation therein.

Naively solving the convex relaxation (4.6) by semidefinite programming methods is too costly. Candès et al. (2011) proposed the Principal Component Pursuit (PCP) algorithm for solving the convex relaxation of Robust

PCA problem as formulated in (4.6). Although, the convex approach provably decomposes the low-rank plus sparse matrix, PCP comes with two main shortcomings. Firstly, it does not account for an error that might be present in every entry of the matrix, e.g. Gaussian noise. Secondly, PCP is computationally expensive, requiring computing Singular Value Decomposition in every iteration, leading to the time complexity of $O(m^2 n)$ per iteration and will typically require $O(1/\varepsilon)$ iterations (Candès et al., 2011), where ε is the required tolerance on the relative error of the solution.

The limitations of the convex relaxation and the PCP algorithm motivated work on computationally faster methods. Zhou et al. (2010) extend the result of Candès et al. (2011) to account for Gaussian noise present in all entries of the matrix. A general framework for minimization of a sum of two convex functions is proposed by Goldfarb et al. (2013) who are able to improve convergence by reducing the number of iterations to $O(1/\sqrt{\varepsilon})$ instead of $O(1/\varepsilon)$. Wang et al. (2013) propose using Alternating Direction Method of Multipliers (ADMM) for solving a two block-separable convex minimization problem and numerically demonstrate that their ADMM outperforms PCP in practice, but it lacks theoretical recovery guarantees. On the other hand, work of Mu et al. (2011) is able to minimize the nuclear norm of a randomly sketched matrix L, resulting in cheaper computation of SVD for smaller, projected matrix, while preserving theoretical guarantees for exact recovery with high-probability.

More recently, a number of faster non-convex algorithms with convergence guarantees have been proposed (Gu and Wang, 2016; Netrapalli et al., 2014; Yi et al., 2016; Chen and Wainwright, 2015) for the non-convex optimisation problem formulated in (1.13) on page 8.

Gu and Wang (2016) reformulate the problem using a low-rank factorisation with $U \in \mathbb{R}^{m\times r}, V \in \mathbb{R}^{r\times n}$ plus a sparse matrix $S \in \mathbb{R}^{m\times n}$

$$\min_{U\in\mathbb{R}^{m\times r},\, V\in\mathbb{R}^{n\times r},\, S\in\mathbb{R}^{m\times n}} \left\|M - (UV^T + S)\right\|^2, \qquad \text{s.t.} \quad \|S\|_0 \le s, \tag{4.7}$$

which they minimise in an alternating fashion and prove exact recovery for sparsity ratio $\alpha = O\left(1/\left(\mu^{2/3} r^{1/3} n\right)\right)$ where μ is the incoherence parameter of L. The result is improved by Yi et al. (2016), who show recovery guarantee for nearly the optimal sparsity $\alpha = O(1/(\mu r^{1.5}))$ in a deterministic model, with the optimal sparsity ratio being $\alpha = O(1/(\mu r))$ (Chen and Wainwright, 2015). Their algorithm requires only one imprecise computation of SVD per iteration, which is then followed by a two gradient steps due to the splitting $L = UV$ and converges linearly in a noise-free setting. Step-size is chosen extremely conservative in the proof section, but in numerical section a different, less conservative, step-size is used which results into faster convergence rate. The optimal sparsity $\alpha = O(1/(\mu r))$ for deterministic model was achieved by Netrapalli et al. (2014), whose *Alternating Projection* (AltProj) has an approximate complexity of $O\left(\mu r^2 n \log(n)^2 \log(1/\varepsilon)^2\right)$. Their

approach performs alternating projections on the set of low-rank and sparse matrices performed by the truncated SVD and hard thresholding.

See (Bouwmans et al., 2017) for a detailed survey on Robust PCA and publicly available implementation of Robust PCA algorithms.

4.2.4 *Summary*

To summarize, the convex relaxation approach is a powerful modelling framework. Its strength lies in its interpretative ability to approximate the non-convex geometry of the original optimisation problem. Its main disadvantage comes from the computation, namely that the convex envelope can be computationally costly to evaluate and that we are forced to optimise over a larger feasible set, e.g. all matrices instead of only low-rank matrices. There have been recent advances in fast algorithms for solving SDPs such as (4.5) based on sketching and low-rank parametrizations by Yurtsever et al. (2019) but these have not been yet explored in the context of compressed sensing and low-rank matrix sensing.

4.3 RECOVERY BY CONVEX RELAXATION

This section contains the proofs of our first main algorithmic contribution that a low-rank plus sparse matrix $X_0 \in \mathrm{LS}_{m,n}(r, s, \mu)$ can be robustly recovered from subsampled measurements taken by a linear mapping $\mathcal{A}(\cdot)$ which satisfies given bounds on its RIC by solving the convex relaxation in (4.3).

Let $X_* = L_* + S_*$ be the solution of the convex optimization problem formulated in (4.3). Here it is shown that if the RICs of the measurement operator $\mathcal{A}(\cdot)$ are sufficient small, then $X_* = X_0$ when the linear constraint in the convex optimization problem (4.3) is satisfied exactly, or alternatively that $\|X_* - X_0\|_F$ is proportional to $\|\mathcal{A}(X^*) - b\|_2$.

Theorem 4.2 (Guaranteed recovery by the convex relaxation)**.** *Let $b = \mathcal{A}(X_0)$ and suppose that $r, s \in \mathbb{N}$ and $\mu < \sqrt{mn}/\left(4r\sqrt{3s}\right)$ are such that the restricted isometry constant $\Delta_{4r,3s,\mu}(\mathcal{A}) \leq \frac{1}{5} - \frac{5}{3}\gamma_{4r,3s,\mu}$. Let $X_* = L_* + S_*$ be the solution of (4.3) with $\lambda = \sqrt{r/s}$, then $\|X_* - X_0\|_F \leq 67\varepsilon_b$.*

Proof. Let $R = X_* - X_0 = (L_* - L_0) + (S_* - S_0) = R^L + R^S$ be the residual split into the low-rank component $R^L = L_* - L_0$ and the sparse component $R^S = S_* - S_0$. We treat R^L and R^S separately, combining the method of proof used in the context of compressed sensing by Candès et al. (2006b) and its extension for the low-rank matrix recovery by Recht et al. (2010).

By Lemma 4.1 on page 86 there exist matrices $R_0^L, R_c^L \in \mathbb{R}^{m\times n}$ such that $R^L = R_0^L + R_c^L$ and

$$\operatorname{rank}(R_0^L) \leq 2r \tag{4.8}$$

$$L_0(R_c^L)^T = 0_{m\times m} \quad \text{and} \quad L_0^T R_c^L = 0_{n\times n}. \tag{4.9}$$

Similarly, by the argument made in the proof of Theorem 1 by Candès et al. (2006b), which we state in §4.8 on page 86 as Lemma 4.2, there exist matrices $R_0^S, R_c^S \in \mathbb{R}^{m\times n}$ such that $R^S = R_0^S + R_c^S$ and

$$\left\|R_0^S\right\|_0 \leq s \tag{4.10}$$

$$\operatorname{supp}(S_0) \cap \operatorname{supp}\left(R_c^S\right) = \emptyset. \tag{4.11}$$

By (L_*, S_*) being a minimum and X_0 being feasible of the convex optimization problem (4.3)

$$\begin{aligned}
\|L_0\|_* + \lambda \|S_0\|_1 &\geq \|L_*\|_* + \lambda \|S_*\|_1 && (4.12)\\
&= \left\|L_0 + R_0^L + R_c^L\right\|_* + \lambda \left\|S_0 + R_0^S + R_c^S\right\|_1 && (4.13)\\
&\geq \left\|L_0 + R_c^L\right\|_* - \left\|R_0^L\right\|_* + \lambda \left\|S_0 + R_c^S\right\|_1 - \lambda \left\|R_0^S\right\|_1 && (4.14)\\
&= \|L_0\|_* + \left\|R_c^L\right\|_* - \left\|R_0^L\right\|_* + \lambda \|S_0\|_1 + \lambda \left\|R_c^S\right\|_1 - \lambda \left\|R_0^S\right\|_1, && (4.15)
\end{aligned}$$

where the second line comes from $L_* - L_0 = R_0^L + R_c^L$ and $S_* - S_0 = R_0^S + R_c^S$, the inequality in the third line comes from the reverse triangle inequality, and the fourth line comes from the construction of R_c^L and R_c^S combined with (Recht et al., 2010, Lemma 2.3), restated as Corollary 4.1, and by $\operatorname{supp}(R_c^S) \cap \operatorname{supp}(R_0^S) = \emptyset$. Subtracting $\|L_0\|_*$ and $\|S_0\|_1$ from both sides of (4.15) and rearranging terms yields

$$\left\|R_c^L\right\|_* + \lambda \left\|R_c^S\right\|_1 \leq \left\|R_0^L\right\|_* + \lambda \left\|R_0^S\right\|_1. \tag{4.16}$$

We proceed by decomposing the remainder terms R_c^L and R_c^S as sums of matrices with decreasing energy as was done by Recht et al. (2010) for low-rank matrices and by Candès et al. (2006b) for sparse vectors. Let $R_c^L = U\operatorname{diag}(\sigma)V^T$ be the singular value decomposition of R_c^L and split the indices of the singular values into sets of size M_r as

$$I_i := \{(i-1)M_r + 1, \ldots, iM_r\}. \tag{4.17}$$

Constructing $R_i^L := U_{I_i}\operatorname{diag}(\sigma_{I_i})V_{I_i}^T$ decomposes R_c^L into a sum $R_c^L = R_1^L + R_2^L + \ldots$ such that

$$\operatorname{rank}\left(R_i^L\right) \leq M_r, \qquad \forall i \geq 1 \tag{4.18}$$

$$R_i^L \left(R_j^L\right)^T = 0_{m\times m} \quad \text{and} \quad \left(R_i^L\right)^T R_j^L = 0_{n\times n}, \quad \forall i \neq j \tag{4.19}$$

$$\sigma_k \leq \frac{1}{M_r} \sum_{j\in I_i} \sigma_j, \qquad \forall k \in I_{i+1} \tag{4.20}$$

where the inequality (4.20) implies that $\left\|R_{i+1}^L\right\|_F^2 \leq \frac{1}{M_r}\left\|R_i^L\right\|_*^2$.

Similarly, order the indices of R_c^S as $v_1, v_2, \ldots, v_{mn} \in [m] \times [n]$ in decreasing order of magnitude of the entries of R_c^S and split the indices of the entries into sets of size M_s as

$$T_i := \{v_\ell : (i-1)M_s \leq \ell \leq iM_s\}, \tag{4.21}$$

Constructing $R_i^S := \left(R_c^S\right)_{T_i}$ decomposes R_c^S into a sum $R_c^S = R_1^S + R_2^S + \ldots$ such that

$$\left\|R_i^S\right\|_0 \leq M_s, \qquad \forall i \geq 1 \tag{4.22}$$

$$\emptyset = T_i \cap T_j, \qquad \forall i \neq j \tag{4.23}$$

$$\left|R_c^S\right|_{(v)} \leq \frac{1}{\sqrt{M_s}} \sum_{j \in T_i} \left|R_i^S\right|_{(j)}, \qquad \forall v \in T_{i+1} \tag{4.24}$$

where the inequality (4.24) implies that $\left\|R_{i+1}^S\right\|_F^2 \leq \frac{1}{M_s}\left\|R_i^S\right\|_1^2$. Combining the two decompositions of R_c^L and R_c^S gives the following bound

$$\sum_{j \geq 2} \left\|R_j^L + R_j^S\right\|_F \leq \sum_{j \geq 2} \left\|R_j^L\right\|_F + \sum_{j \geq 2} \left\|R_j^S\right\|_F \tag{4.25}$$

$$\leq \sqrt{\frac{1}{M_r}} \sum_{j \geq 1} \left\|R_j^L\right\|_* + \sqrt{\frac{1}{M_s}} \sum_{j \geq 1} \left\|R_j^S\right\|_1 \tag{4.26}$$

$$= \sqrt{\frac{1}{M_r}} \left\|R_c^L\right\|_* + \sqrt{\frac{1}{M_s}} \left\|R_c^S\right\|_1 \tag{4.27}$$

$$\leq \sqrt{\frac{1}{M_r}} \left(\left\|R_0^L\right\|_* + \sqrt{\frac{M_r}{M_s}} \left\|R_0^S\right\|_1 \right) \tag{4.28}$$

$$\leq \sqrt{\frac{2r}{M_r}} \left\|R_0^L\right\|_F + \sqrt{\frac{s}{M_s}} \left\|R_0^S\right\|_F, \tag{4.29}$$

where the inequality in the first line comes from the triangle inequality, the second inequality comes as a consequence of (4.20) and (4.24), the third line comes from (4.19) combined with (Recht et al., 2010, Lemma 2.3), restated as Corollary 4.1, and from (4.23), the fourth inequality comes from (4.16) with $\lambda = \sqrt{M_r/M_s}$, and the last fifth line is a property of ℓ_1 and Schatten-1 norms. Choosing $M_r = r$ and $M_s = s$ in (4.29) gives

$$\sum_{j \geq 2} \left\|R_j^L + R_j^S\right\|_F \leq \sqrt{2}\left\|R_0^L\right\|_F + \left\|R_0^S\right\|_F, \tag{4.30}$$

and also that $\lambda = \sqrt{r/s}$ as stated in the theorem.

By feasibility of X^* and linearity of $\mathcal{A}$ we have

$$\varepsilon_b \geq \|\mathcal{A}(X^*) - b\|_2 = \|\mathcal{A}(X^* - X_0)\|_2 = \|\mathcal{A}(R)\|_2. \tag{4.31}$$

Let $\Delta := \Delta_{4r,3s,\mu}$ be the RIC with squared norms for $\mathrm{LS}_{m,n}(4r, 3s, \mu)$ and $\gamma := \gamma_{4r,3r,\mu}$ be the rank-sparsity correlation coefficient defined in Lemma 2.1

on page 17. Then

$$(1-\Delta)\|R_0^L+R_1^L\|_F^2 \le \left\|\mathcal{A}\left(R_0^L+R_1^L\right)\right\|_2^2 = \left|\left\langle \mathcal{A}(R_0^L+R_1^L),\, \mathcal{A}(R_0^L+R_1^L-R+R)\right\rangle\right| \tag{4.32}$$

$$= \left|\left\langle \mathcal{A}(R_0^L+R_1^L),\, \mathcal{A}(R_0^L+R_1^L-R)\right\rangle + \left\langle \mathcal{A}(R_0^L+R_1^L),\, \mathcal{A}(R)\right\rangle\right| \tag{4.33}$$

$$\le \left|\left\langle \mathcal{A}\left(R_0^L+R_1^L\right),\, \mathcal{A}\left(-R_0^S-R_1^S-\sum_{j\ge 2}R_j\right)\right\rangle\right| + \left|\left\langle \mathcal{A}(R_0^L+R_1^L),\, \mathcal{A}(R)\right\rangle\right| \tag{4.34}$$

$$\le \left(\Delta+\frac{2\gamma}{1-\gamma^2}\right)\left\|R_0^L+R_1^L\right\|_F\left(\left\|R_0^S+R_1^S\right\|_F+\sum_{j\ge 2}\left\|R_j\right\|_F\right) + \left\|\mathcal{A}\left(R_0^L+R_1^L\right)\right\|_2\|\mathcal{A}(R)\|_2, \tag{4.35}$$

where the inequality in the first line comes from $R_0^L+R_1^L \in \mathrm{LS}_{m,n}(4r,3s,\mu)$ by Lemma 2.9 satisfying the RICs, the second line is a consequence of feasibility in (4.31), and the third line comes from Lemma 4.3 and by sums of individual pairs in the inner product being in $\mathrm{LS}_{m,n}(4r,3s,\mu)$ by Lemma 2.9.

The first term in (4.35) can be bounded as

$$\left(\Delta+\frac{2\gamma}{1-\gamma^2}\right)\left\|R_0^L+R_1^L\right\|_F\left(\left\|R_0^S+R_1^S\right\|_F+\sum_{j\ge 2}\left\|R_j\right\|_F\right) \tag{4.36}$$

$$\le \left(\Delta+\frac{2\gamma}{1-\gamma^2}\right)\left\|R_0^L+R_1^L\right\|_F\left(\left\|R_0^S+R_1^S\right\|_F+\sqrt{2}\left\|R_0^L\right\|_F+\left\|R_0^S\right\|_F\right) \tag{4.37}$$

$$\le \left(\Delta+\frac{2\gamma}{1-\gamma^2}\right)\left\|R_0^L+R_1^L\right\|_F\left(2\left\|R_0^S+R_1^S\right\|_F+\sqrt{2}\left\|R_0^L+R_1^L\right\|_F\right) \tag{4.38}$$

where the second line comes as a consequence of optimality in (4.30) with $M_r = r$ and $M_s = s$, and the third line comes from $\left\|R_0^L\right\|_F \le \left\|R_0^L+R_1^L\right\|_F$ and $\left\|R_1^L\right\|_F \le \left\|R_0^L+R_1^L\right\|_F$. The upper bound of the second term in (4.35) is a consequence of feasibility bound in (4.31) and of the RIC for $R_0^L+R_1^L \in \mathrm{LS}_{m,n}(4r,3s,\mu)$ by Lemma 2.9

$$\left\|\mathcal{A}\left(R_0^L+R_1^L\right)\right\|_2\|\mathcal{A}(R)\|_2 \le \varepsilon_b(1+\Delta)\|R_0^L+R_1^L\|_F. \tag{4.39}$$

Combining inequality (4.38) and (4.39) yields an upper bound of (4.35)

$$(1-\Delta)\left\|R_0^L+R_1^L\right\|_F^2 \le \left(\Delta+\frac{2\gamma}{1-\gamma^2}\right)\left\|R_0^L+R_1^L\right\|_F\left(2\left\|R_0^S+R_1^S\right\|_F+\sqrt{2}\left\|R_0^L+R_1^L\right\|_F\right) + \varepsilon_b\left\|R_0^L+R_1^L\right\|_F(1+\Delta), \tag{4.40}$$

which after dividing both sides by $(1-\Delta)\left\|R_0^L + R_1^L\right\|_F$ gives

$$\left\|R_0^L + R_1^L\right\|_F \leq \frac{1}{1-\Delta}\left(\Delta + \frac{2\gamma}{1-\gamma^2}\right)\left(2\left\|R_0^S + R_1^S\right\|_F + \sqrt{2}\left\|R_0^L + R_1^L\right\|_F\right) + \varepsilon_b \frac{1+\Delta}{1-\Delta}. \tag{4.41}$$

Mutatis mutandis, the same argument applies to $\left\|R_0^S + R_1^S\right\|_F$ (for details, see Remark 4.1)

$$\left\|R_0^S + R_1^S\right\|_F \leq \frac{1}{1-\Delta}\left(\Delta + \frac{2\gamma}{1-\gamma^2}\right)\left(\left(1+\sqrt{2}\right)\left\|R_0^L + R_1^L\right\|_F + \left\|R_0^S + R_1^S\right\|_F\right) + \varepsilon_b \frac{1+\Delta}{1-\Delta}. \tag{4.42}$$

Adding (4.41) and (4.42) together

$$\left\|R_0^L + R_1^L\right\|_F + \left\|R_0^S + R_1^S\right\|_F \leq \frac{1}{1-\Delta}\left(\Delta + \frac{2\gamma}{1-\gamma^2}\right)\left(\left(1+2\sqrt{2}\right)\left\|R_0^L + R_1^L\right\|_F + 3\left\|R_0^S + R_S^L\right\|_F\right) + 2\varepsilon_b \frac{1+\Delta}{1-\Delta}. \tag{4.43}$$

For $\Delta < \frac{1}{5} - \frac{5}{3}\gamma$ the prefactor $\frac{1}{1-\Delta}\left(\Delta + \frac{2\gamma}{1-\gamma^2}\right) < \frac{1}{4}$ and therefore also $\frac{\Delta}{1-\Delta} < \frac{1}{4}$, resulting into (4.43) being upper bounded as

$$\left\|R_0^L + R_1^L\right\|_F + \left\|R_0^S + R_1^S\right\|_F \leq \frac{1+2\sqrt{2}}{4}\left\|R_0^L + R_1^L\right\|_F + \frac{3}{4}\left\|R_0^S + R_1^S\right\|_F + 3\varepsilon_b, \tag{4.44}$$

The maximum of $\left\|R_0^L + R_1^L\right\|_F + \left\|R_0^S + R_1^S\right\|_F$ over the constraints given by the inequality in (4.44) is attained when

$$\left\|R_0^L + R_1^L\right\|_F + \left\|R_0^S + R_1^S\right\|_F = \frac{3\varepsilon_b}{3-2\sqrt{2}}. \tag{4.45}$$

By orthogonality from construction in (4.19) and (4.23) we have that

$$\begin{aligned}
\frac{3\varepsilon_b}{3-2\sqrt{2}} &\geq \left\|R_0^L + R_1^L\right\|_F + \left\|R_0^S + R_1^S\right\|_F && (4.46)\\
&\geq \left\|R_0^L\right\|_F + \left\|R_0^S\right\|_F && (4.47)\\
&\geq \frac{1}{\sqrt{2}}\left(\sum_{j\geq 2}\left\|R_j\right\|_F\right) && (4.48)\\
&\geq \frac{1}{\sqrt{2}}\left\|\sum_{j\geq 2} R_j\right\|_F = \frac{1}{\sqrt{2}}\left\|R_c - R_1\right\|_F, && (4.49)
\end{aligned}$$

where the inequality in the third line comes from (4.30) and the inequality in the fourth line comes from triangle inequality. Applying the triangle inequality on $R = \left(R_0^L + R_1^L\right) + \left(R_0^S + R_1^S\right) + \left(R_c - R_1\right)$ and using the bounds in (4.46) and (4.49) concludes the proof

$$\begin{aligned}
\|R\|_F &\leq \left\|R_0^L + R_1^L\right\|_F + \left\|R_0^S + R_1^S\right\|_F + \|R_c - R_1\|_F && (4.50)\\
&\leq 3\varepsilon_b \frac{1+\sqrt{2}}{3-2\sqrt{2}} \leq 67\varepsilon_b. && (4.51)
\end{aligned}$$

□

Alternatively, X_0 can be obtained from its compressed measurements $\mathcal{A}(X_0)$ by iterative gradient descent methods that are guaranteed to converge to a global minimizer of the non-convex optimization problem

$$\min_{X=L+S\in\mathbb{R}^{m\times n}} \|\mathcal{A}(X)-b\|_2\,, \qquad \text{s.t.} \quad X\in \mathrm{LS}_{m,n}(r,s,\mu). \tag{4.52}$$

We introduce two natural extensions of the simple yet effective Normalized Iterative Hard Thresholding (NIHT) for compressed sensing (Blumensath and Davies, 2010) and matrix completion (Tanner and Wei, 2013) algorithms, here called NIHT and Normalized Alternative Hard Thresholding (NAHT) for low-rank plus sparse matrices, Algorithm 1 and Algorithm 2 respectively. In both cases we establish that if the measurement operator has suitably small RICs then NIHT and NAHT provably converge to the global minimum of the non-convex problem formulated in (4.52) and recover $X_0\in \mathrm{LS}_{m,n}(r,s,\mu)$ for which $b=\mathcal{A}(X_0)$.

4.4.1 *Normalized Iterative Hard Thresholding*

We begin by analysing NIHT presented in Algorithm 1. The hard thresholding projection in Algorithm 1 is performed by computing Robust PCA which is solved to an accuracy proportional to ε_p as given by (4.55). The RPCA projection of a matrix $W\in\mathbb{R}^{m\times n}$ on the set of $\mathrm{LS}_{m,n}(r,s,\mu)$ with precision ε returns a matrix $X\in\mathrm{LS}_{m,n}(r,s,\mu)$ such that

$$(L,S)\leftarrow \mathrm{RPCA}_{r,s,\mu}(W,\varepsilon_p) \qquad \text{s.t.} \qquad \left\|(L+S)-W_{\mathrm{rpca}}\right\|_F\le\varepsilon_p, \tag{4.53}$$

where $W_{\mathrm{rpca}}:=\arg\min_{Y\in\mathrm{LS}_{m,n}(r,s,\mu)}\|Y-W\|_F$ is the optimal projection of the matrix W on the set $\mathrm{LS}_{m,n}(r,s,\mu)$.

In order to achieve the recovery with the asymptotically optimal bound on sparsity $s=O\left(mn/\left(\mu^2r^2\right)\right)$, it is necessary to choose RPCA subroutine that has similiar guarantees, e.g. the Alternating Projection algorithm (AltProj) by Netrapalli et al. (2014) or the Accelerated Alternating Projection algorithm (AccAltProj) by Cai et al. (2019), both of which have provable global linear convergence when $s=O\left(mn/\left(\mu^2r^2\right)\right)$ and high robustness in practice. For a discussion on Robust PCA algorithms see §4.2.3 or (Bouwmans et al., 2017).

Note that the projection used in computing the stepsize is defined as $\mathrm{Proj}_{(U^j,\Omega^j)}\left(R^j\right):=P_{U^j}R^j+\mathbb{1}_{\Omega^j}\circ(R^j-P_{U^j}R^j)$, where $P_{U^j}:=U^j\left(U^j\right)^*$, $\mathbb{1}_{\Omega^j}$ is a matrix with ones at indices Ω^j, and $\circ$ denotes the entry-wise Hadamard product. This corresponds to first projecting the left singular vectors of R^j on the subspace spanned by columns of U^j and then setting entries at indices Ω^j to be equal to the entries of R^j at indices Ω^j. One can repeat this process to achieve a more precise projection of R^j in the low-rank plus sparse matrix set defined by $\left(U^j,\Omega^j\right)$.

The proof of NIHT follows the same line of thought as the one for low-rank matrix completion by Tanner and Wei (2013), with the only difference

Algorithm 1 Normalized Iterative Hard Thresholding (NIHT) for LS recovery

Input: $b = \mathcal{A}(X_0)$, $\mathcal{A}$, r, s, and termination criteria
Set: $(L^0, S^0) = \text{RPCA}_{r,s,\mu}(\mathcal{A}^*(b), \varepsilon_p)$, $X^0 = L^0 + S^0$, $j = 0$,
$\Omega^0 = \text{supp}(S^0)$ and U^0 as the top r left singular vectors of L^0

1: **while** not converged **do**
2: Compute the residual $R^j = \mathcal{A}^*\left(b - \mathcal{A}(X^j)\right)$
3: Compute the stepsize: $\alpha_j = \left\|\text{Proj}_{(U^j,\Omega^j)}\left(R^j\right)\right\|_F^2 / \left\|\mathcal{A}\left(\text{Proj}_{(U^j,\Omega^j)}\left(R^j\right)\right)\right\|_2^2$
4: Set $W^j = X^j + \alpha_j R^j$
5: Compute $(L^{j+1}, S^{j+1}) = \text{RPCA}_{r,s,\mu}(W^j, \varepsilon_p)$ and set $X^{j+1} = L^{j+1} + S^{j+1}$
6: Let $\Omega^{j+1} = \text{supp}(S^{j+1})$ and U^{j+1} be the top r left singular vectors of L^{j+1}
7: $j = j + 1$
8: **end while**

Output: X^j

of the hard thresholding projection, in the form of RPCA, being an imprecise projection with accuracy ε_p as stated in (4.53). The theorem provides full guarantees for the practical algorithm only if the Robust PCA subroutine is guaranteed to solve the projection within the ε_p optimality of (4.53) with the number of corruptions $s = O\left(mn/\left(\mu^2 r^2\right)\right)$. The proof consists of deriving an inequality where $\left\|X^{j+1} - X_0\right\|_F$ is bounded by a factor multiplying $\left\|X^j - X_0\right\|_F$, and then showing that this multiplicative factor is strictly less then one if $\mathcal{A}$ satisfies RIC with $\Delta_3 := \Delta_{r,s,\mu}(\mathcal{A}) < 1/5$.

Theorem 4.3 (Guaranteed recovery by NIHT)**.** *Suppose that $r, s \in \mathbb{N}$ and $\mu < \sqrt{mn} \,/\, \left(3r\sqrt{3s}\right)$ are such that the restricted isometry constant $\Delta_3 := \Delta_{3r,3s,\mu}(\mathcal{A}) < \frac{1}{5}$. Then NIHT applied to $b = \mathcal{A}(X_0)$ as described in Algorithm 1 will linearly converge as*

$$\|X^{j+1} - \widehat{X}\|_F \leq 8\frac{\Delta_3\left(1 - 3\Delta_3\right)}{\left(1 - \Delta_3\right)^2}\|X^j - \widehat{X}\|_F + \frac{1 - 5\Delta_3}{1 - \Delta_3}\varepsilon_p, \tag{4.54}$$

where ε_p is the accuracy of the Robust PCA oblique projection that performs projection on the set of incoherent low-rank plus sparse matrices $\text{LS}_{m,n}(r, s, \mu)$ and $\widehat{X}$ is a matrix in proximity of X_0

$$\|\widehat{X} - X_0\|_F \leq \varepsilon_p \frac{1 - \Delta_3}{1 - 5\Delta_3}. \tag{4.55}$$

Proof. Let $b = \mathcal{A}(X_0)$ be the vector of measurements of the matrix $X_0 \in \text{LS}_{m,n}(r, s, \mu)$ and $W^j = X^j - \alpha_j^L \mathcal{A}^*\left(\mathcal{A}(X^j) - b\right)$ to be the update of X^j before the oblique Robust PCA projection step $X^{j+1} = \text{RPCA}_{r,s,\mu}(W^j, \varepsilon_p)$. By X^{j+1} being within an ε_p distance in the Frobenius norm of the optimal

RPCA projection $X^{j+1}_{\text{rpca}} := \text{RPCA}_{r,s,\mu}(W^j, 0)$ defined in (4.53)

$$\left\|W^j - X^{j+1}\right\|_F^2 = \left\|W^j - X^{j+1}_{\text{rpca}} + X^{j+1}_{\text{rpca}} - X^{j+1}\right\|_F^2 \tag{4.56}$$

$$\leq \left(\left\|W^j - X^{j+1}_{\text{rpca}}\right\|_F + \left\|X^{j+1} - X^{j+1}_{\text{rpca}}\right\|_F\right)^2 \tag{4.57}$$

$$\leq \left(\left\|W^j - X_0\right\|_F + \varepsilon_p\right)^2, \tag{4.58}$$

where in the second line we used the triangle inequality, and the third line comes from X^{j+1}_{rpca} being the optimal projection thus being the closest matrix in $\text{LS}_{m,n}(r,s,\mu)$ to W^j in the Frobenius norm and by X^{j+1} being within ε_p distance of X^{j+1}_{rpca}. By expansion of the left hand side of (4.56)

$$\left\|W^j - X^{j+1}\right\|_F^2 = \left\|W^j - X_0 + X_0 - X^{j+1}\right\|_F^2 \tag{4.59}$$

$$= \left\|W^j - X_0\right\|_F^2 + \left\|X_0 - X^{j+1}\right\|_F^2 + 2\left\langle W^j - X_0,\, X_0 - X^{j+1}\right\rangle \tag{4.60}$$

$$= \left(\left\|W^j - X_0\right\|_F + \varepsilon_p\right)^2 \leq \left\|W^j - X_0\right\|_F^2 + 2\varepsilon_p \left\|W^j - X_0\right\|_F + \varepsilon_p^2 \tag{4.61}$$

where the last line (4.61) follows from the inequality in (4.58). Subtracting $\|W^j - X_0\|_F^2$ from both sides of (4.61) gives

$$\left\|X^{j+1} - X_0\right\|_F^2 \leq 2\left\langle W^j - X_0,\, X^{j+1} - X_0\right\rangle + 2\varepsilon_p \left\|W^j - X_0\right\|_F + \varepsilon_p^2. \tag{4.62}$$

The matrix W^j in the inner product on the right hand side of (4.62) can be expressed using the update rule $W^j = X^j - \alpha_j \mathcal{A}^*\left(\mathcal{A}\left(X^j\right) - b\right)$

$$2\left\langle W^j - X_0,\, X^{j+1} - X_0\right\rangle$$

$$= 2\langle X^j - X_0, X^{j+1} - X_0\rangle - 2\alpha_j \left\langle \mathcal{A}^*\mathcal{A}\left(X^j - X_0\right),\, X^{j+1} - X_0\right\rangle \tag{4.63}$$

$$= 2\left\langle X^j - X_0,\, X^{j+1} - X_0\right\rangle - 2\alpha_j \left\langle \mathcal{A}\left(X^j - X_0\right),\, \mathcal{A}\left(X^{j+1} - X_0\right)\right\rangle \tag{4.64}$$

$$\leq 2\left\|I - \alpha_j A_Q^* A_Q\right\|_2 \left\|X^j - X_0\right\|_F \left\|X^{j+1} - X_0\right\|_F, \tag{4.65}$$

where in the first line we use $b = \mathcal{A}(X_0)$ and linearity of $\mathcal{A}$, in the second line we split the inner product into two inner products by linearity of $\mathcal{A}$, and the inequality in the third line is a consequence of Lemma 4.4, stated in §4.8 on page 90.

Here it would be possible to extend the result to be stable under measurement error ε_b as done in Theorem 4.2 by adding an error term in (4.63).

The matrix W^j can be expressed using the update rule $W^j = X^j - \alpha_j \mathcal{A}^*\left(\mathcal{A}\left(X^j\right) - b\right)$ in the second term of the right hand side of (4.62) and upper bounded by Lemma 4.4

$$\left\|W^j - X_0\right\|_F = \left\|X^j - X_0 + \alpha_j \mathcal{A}^*\left(\mathcal{A}\left(X^j - X_0\right)\right)\right\|_2 \tag{4.66}$$

$$\leq \left\|I - \alpha_j A_Q^* A_Q\right\|_2 \left\|X^j - X_0\right\|_F. \tag{4.67}$$

By Lemma 4.4, the eigenvalues of $\left(I - \alpha_j A_Q^* A_Q\right)$ are bounded by

$$1 - \alpha_j(1 + \Delta_3) \leq \lambda\left(I - \alpha_j A_Q^* A_Q\right) \leq 1 + \alpha_j(1 - \Delta_3), \tag{4.68}$$

where $\Delta_3 := \Delta_{3r,3s,\mu}$.

Consider the stepsize computed in Algorithm 1, Line 3 inspired by the previous work on NIHT in the context of compressed sensing (Blumensath and Davies, 2010) and low-rank matrix sensing (Tanner and Wei, 2013)

$$\alpha_j = \frac{\left\| \mathrm{Proj}_{(U^j,\Omega^j)}\left(R^j\right)\right\|_F}{\left\| \mathcal{A}\left(\mathrm{Proj}_{(U^j,\Omega^j)}\left(R^j\right)\right)\right\|_2} \tag{4.69}$$

where the projection $\mathrm{Proj}_{(U^j,\Omega^j)}\left(R^j\right)$ ensures that the residual R^j is projected onto the set $\mathrm{LS}_{m,n}(r,s,\mu)$. Then we can bound α_j using the RIC of $\mathcal{A}$ as

$$\frac{1}{1+\Delta_1} \le \alpha_j \le \frac{1}{1-\Delta_1}, \tag{4.70}$$

where $\Delta_1 := \Delta_{r,s,\mu}$. Combining (4.68) with (4.70) gives

$$1 - \frac{1+\Delta_3}{1-\Delta_1} \le \lambda\left(I - \alpha_j A_Q^* A_Q\right) \le 1 - \frac{1-\Delta_3}{1+\Delta_1}. \tag{4.71}$$

Since $\Delta_3 \ge \Delta_1$, the magnitude of the lower bound in (4.71) is greater than the upper bound. Therefore

$$\eta := 2\left(\frac{1+\Delta_3}{1-\Delta_1} - 1\right) \ge 2\left\|I - \alpha_j A_Q^* A_Q\right\|_2, \tag{4.72}$$

where the constant η is strictly smaller than one if $\Delta_3 < 1/5$.

Finally, the error in (4.62) can be upper bounded by (4.65) combined with (4.67) with η being the upper bound on the operator norm in (4.72)

$$\left\|X^{j+1} - X_0\right\|_F^2 \le \eta \left\|X^j - X_0\right\|_F \left\|X^{j+1} - X_0\right\|_F + \eta\varepsilon_p \left\|X^j - X_0\right\|_F + \varepsilon_p^2. \tag{4.73}$$

It remains to show the inequality (4.73) implies the update rule contracts the error and the iterates X^j converge to a matrix $\hat{X}$ that is within a small in the Frobenius norm from X_0 depending on the precision ε_p of the RPCA. Rewrite (4.73) using the notation $e^j := \|X^j - X_0\|_F$

$$\left(e^{j+1}\right)^2 \le \eta\, e^j\, e^{j+1} + \eta\, \varepsilon_p\, e^j + \varepsilon_p^2 \tag{4.74}$$

$$e^{j+1} \le \eta\, e^j + \eta\, \varepsilon_p\, \frac{e^j}{e^{j+1}} + \frac{\varepsilon_p^2}{e^{j+1}}. \tag{4.75}$$

In the following, assume $\frac{\varepsilon_p}{1-\eta} \le e^{j+1}$ and upper bound the right hand side of (4.75) as

$$\begin{aligned} e^{j+1} &\le \eta\, e^j + \eta\, \varepsilon_p\, \frac{e^j}{e^{j+1}} + \frac{\varepsilon_p^2}{e^{j+1}} && (4.76)\\ &= \eta\, e^j + \eta(1-\eta)e^j \frac{\varepsilon_p}{(1-\eta)e^{j+1}} + \varepsilon_p(1-\eta)\frac{\varepsilon_p}{(1-\eta)e^{j+1}} && (4.77)\\ &\le \eta\, e^j + \eta(1-\eta)e^j + (1-\eta)\varepsilon_p && (4.78)\\ &< e^j\eta, && (4.79) \end{aligned}$$

where the inequality (4.78) is a consequence of $e^{j+1} \geq \varepsilon_p/(1-\eta)$ and the inequality (4.79) holds if $e^j > \varepsilon_p/(1-\eta)$. Therefore, if $e^j > \varepsilon_p/(1-\eta)$ then the error sequence is contractive because $\eta < 1$ by $\Delta_3 < 1/5$ and has a fixed point $e^* = \varepsilon_p/(1-\eta) = \varepsilon_p \frac{1-\Delta_3}{1-5\Delta_3}$. Moreover, by $\Delta_3 \geq \Delta_1$, the equation in (4.78) is upper bounded as

$$e^{j+1} \leq 8\frac{\Delta_3(1-3\Delta_3)}{(1-\Delta_3)^2}e^j + \frac{1-5\Delta_3}{1-\Delta_3}\varepsilon_p, \tag{4.80}$$

which gives an upper bound on the rate of convergence. □

4.4.2 *Normalized Alternating Hard Thresholding*

One of the shortcomings of the previously discussed NIHT is the need to compute a hard-thresholding operation on $\mathrm{LS}_{m,n}(r,s,\mu)$ in the form of an imprecise Robust PCA subroutine. We now present NAHT which overcomes this issue by alternating between projecting of the low-rank and the sparse component. NAHT is also stable to error ε_b, but we omit the stability analysis for clarity in the proofs.

It is also possible to modify the algorithm to perform the two hard-thresholding projections in parallel. This would be especially beneficial in the case when both of the thresholding operations are expensive to compute.

Same as in the case of NIHT, the projection used in computing the stepsize is defined as $\mathrm{Proj}_{(U^j,\Omega^j)}\left(R^j\right) := P_{U^j}R^j + \mathbb{1}_{\Omega^j} \circ (R^j - P_{U^j}R^j)$, where $P_{U^j} := U^j\left(U^j\right)^*$, $\mathbb{1}_{\Omega^j}$ is the matrix with ones at indices Ω^j, and $\circ$ denotes the entry-wise Hadamard product.

In addition, NAHT also provably solves Robust PCA with the optimal order of the number of corruptions $s = \mathcal{O}\left(mn/\left(\mu^2 r^2\right)\right)$ when the sensing operator is the identity, and therefore satisfies the RICs with $\Delta = 0$.

Theorem 4.4 (Guaranteed recovery by NAHT)**.** *Suppose that $r, s \in \mathbb{N}$ and $\mu < \sqrt{mn}\,/\left(3r\sqrt{3s}\right)$ are such that the restricted isometry constant $\Delta_3 := \Delta_{3r,3s,\mu}(\mathcal{A}) < \frac{1}{9} - \gamma_2$. Then NAHT applied to $b = \mathcal{A}(X_0)$ as described in Algorithm 2 will linearly converge to $X_0 = L_0 + S_0$ as*

$$\left\|L^{j+1} - L_0\right\|_F + \left\|S^{j+1} - S_0\right\|_F \leq \frac{6\Delta_3 + \frac{9}{2}\gamma_2}{1 - 3\Delta_3 - \frac{9}{2}\gamma_2}\left(\left\|L^j - L_0\right\|_F + \left\|S^j - S_0\right\|_F\right). \tag{4.81}$$

Proof. Let $b = \mathcal{A}(X_0)$ be the vector of measurements of the matrix $X_0 \in \mathrm{LS}_{m,n}(r,s,\mu)$ and $V^j = L^j - \alpha_j^L \mathcal{A}^*\left(\mathcal{A}(X^j) - b\right)$ to be the update of L^j before the rank r projection $L^{j+1} = \mathrm{HT}_r(V^j)$. As a consequence of L^{j+1} being the closest rank r matrix to V^j in the Frobenius norm we have that

Again, it is possible to extend the result to the case when there is a measurement error ε_b as done in Theorem 4.2 by having $b = \mathcal{A}(X_0) + e$, with $\|e\|_2 \leq \varepsilon_b$.

$$\begin{aligned}\left\|V^j - L_0\right\|_F^2 &\geq \left\|V^j - L^{j+1}\right\|_F^2 = \left\|V^j - L_0 + L_0 - L^{j+1}\right\|_F^2 \\ &= \left\|V^j - L_0\right\|_F^2 + \left\|L_0 - L^{j+1}\right\|_F^2 + 2\left\langle V^j - L_0,\, L_0 - L^{j+1}\right\rangle. \end{aligned} \tag{4.82}$$

Algorithm 2 Normalized Alternating Hard Thresholding (NAHT) for LS recovery

Input: $b = \mathcal{A}(X_0)$, $\mathcal{A}$, r, s, and termination criteria
Set: $L^0 = \mathrm{HT}(\mathcal{A}^*(b); r)$, $S^0 = \mathrm{HT}(\mathcal{A}^*(b) - L^0; s)$, $X^0 = L^0 + S^0$, $j = 0$, $\Omega^0 = \mathrm{supp}(S^0)$ and U^0 as the top r left singular vectors of L^0

while not converged **do**
  Compute the residual $R_L^j = \mathcal{A}^*\left(\mathcal{A}(X^j) - b\right)$
  Compute the stepsize $\alpha_j^L = \left\|\mathrm{Proj}_{(U^j,\Omega^j)}\left(R^j\right)\right\|_F^2 \Big/ \left\|\mathcal{A}\left(\mathrm{Proj}_{(U^j,\Omega^j)}\left(R^j\right)\right)\right\|_2^2$
  Set $V^j = L^j - \alpha_j^L R_L^j$
  Set $L^{j+1} = \mathrm{HT}(V^j; r)$ and let U^{j+1} be the left singular vectors of L^{j+1}
  Set $X^{j+\frac{1}{2}} = L^{j+1} + S^j$
  Compute the residual $R_S^j = \mathcal{A}^*\left(\mathcal{A}(X^{j+\frac{1}{2}}) - b\right)$
  Compute the stepsize $\alpha_j^S = \left\|\mathrm{Proj}_{(U^{j+1},\Omega^j)}\left(R^j\right)\right\|_F^2 \Big/ \left\|\mathcal{A}\left(\mathrm{Proj}_{(U^{j+1},\Omega^j)}\left(R^j\right)\right)\right\|_2^2$
  Set $W^j = S^j - \alpha_j^S R_S^j$
  Set $S^{j+1} = \mathrm{HT}(W^j; s)$ and let $\Omega^{j+1} = \mathrm{supp}(S^{j+1})$
  Set $X^{j+1} = L^{j+1} + S^{j+1}$
  $j = j + 1$
end while

Output: $X^j = L^j + S^j$

Subtracting $\left\|V^j - L_0\right\|_F^2$ from both sides of (4.82) and rearranging terms gives

$$\begin{aligned}
\left\|L_0 - L^{j+1}\right\|_F^2 &\le 2\left\langle V^j - L_0,\, L^{j+1} - L_0\right\rangle && (4.83)\\
&= 2\left\langle L^j - \alpha_j^L \mathcal{A}^*\left(\mathcal{A}\left(X^j - X_0\right)\right) - L_0,\, L^{j+1} - L_0\right\rangle && (4.84)\\
&= 2\left\langle L^j - L_0 - \alpha_j^L \mathcal{A}^*\left(\mathcal{A}\left(L^j - L_0 + S^j - S_0\right)\right),\, L^{j+1} - L_0\right\rangle && (4.85)\\
&= 2\left\langle L^j - L_0,\, L^{j+1} - L_0\right\rangle - 2\,\alpha_j^L \left\langle \mathcal{A}\left(L^j - L_0\right),\, \mathcal{A}\left(L^{j+1} - L_0\right)\right\rangle \\
&\qquad - 2\,\alpha_j^L \left\langle \mathcal{A}\left(S^j - S_0\right),\, \mathcal{A}\left(L^{j+1} - L_0\right)\right\rangle && (4.86)\\
&\le 2\left\|I - \alpha_j^L A_Q^* A_Q\right\| \left\|L^j - L_0\right\|_F \left\|L^{j+1} - L_0\right\|_F \\
&\qquad + 2\,\alpha_j^L\, \rho_2 \left\|S^j - S_0\right\|_F \left\|L^{j+1} - L_0\right\|_F, && (4.87)
\end{aligned}$$

where in the second line we expanded V^j using the update rule $V^j = L^j - \alpha_j^L \mathcal{A}\left(\mathcal{A}(X^j) - b\right)$ and $b = \mathcal{A}(X_0)$, in the third line we expanded $X^j = L^j + S^j$, in the fourth line we split the inner product into two inner products by linearity of $\mathcal{A}$, and in the last line the inequality comes from Lemma 4.4 bounding the first two terms and Lemma 4.3 bounding the third term with $\rho_2 := \left(\Delta_2 + \frac{2\gamma_2}{1-\gamma_2^2}\right)$ where $\Delta_2 := \Delta_{2r,2s,\mu}$ and $\gamma_2 := \gamma_{2r,2s,\mu}$ since $\left(L^{j+1} - L_0 + S^j - S_0\right) \in \mathrm{LS}_{m,n}(2r, 2s, \mu)$. Dividing both sides of (4.87)

by $\left\|L_0 - L^{j+1}\right\|_F$ gives

$$\left\|L_0 - L^{j+1}\right\|_F \leq 2\left\|I - \alpha_j^L A_Q^* A_Q\right\| \left\|L^j - L_0\right\|_F + 2\,\alpha_j^L\ \rho_2 \left\|S^j - S_0\right\|_F . \quad (4.88)$$

Let $W^j = S^j - \alpha_j^S\ \mathcal{A}^*\left(\mathcal{A}(X^{j+\frac{1}{2}}) - b\right)$ be the subsequent update of S^j before the s-sparse projection $S^{j+1} = \mathrm{HT}_s(W^j)$. By S^{j+1} being the closest s sparse matrix to W^j in the Frobenius norm and by $\|S_0\|_0 \leq s$, it follows that

$$\begin{aligned}\left\|W^j - S_0\right\|_F^2 &\geq \left\|W^j - S^{j+1}\right\|_F^2 = \left\|W^j - S_0 + S_0 - S^{j+1}\right\|_F^2 \\ &= \left\|W^j - S_0\right\|_F^2 + \left\|S_0 - S^{j+1}\right\|_F^2 + 2\left\langle W^j - S_0,\ S_0 - S^{j+1}\right\rangle . \quad (4.89)\end{aligned}$$

Subtracting $\left\|W^j - S_0\right\|_F^2$ from both sides in (4.89) and rearranging terms gives

$$\begin{aligned}\left\|S_0 - S^{j+1}\right\|_F^2 &\leq 2\left\langle W^j - S_0,\ S^{j+1} - S_0\right\rangle & (4.90)\\ &= 2\left\langle S^j - \alpha_j^S\ \mathcal{A}^*\left(\mathcal{A}\left(X^{j+\frac{1}{2}} - X_0\right)\right) - S_0,\ S^{j+1} - S_0\right\rangle & (4.91)\\ &= 2\left\langle S^j - S_0 - \alpha_j^S\ \mathcal{A}^*\left(\mathcal{A}\left(L^{j+1} - L_0 + S^j - S_0\right)\right),\ S^{j+1} - S_0\right\rangle & (4.92)\\ &= 2\left\langle S^j - S_0,\ S^{j+1} - S_0\right\rangle - 2\,\alpha_j^S\left\langle \mathcal{A}\left(S^j - S_0\right),\ \mathcal{A}\left(S^{j+1} - S_0\right)\right\rangle \\ &\qquad - 2\,\alpha_j^S\left\langle \mathcal{A}\left(L^{j+1} - L_0\right),\ \mathcal{A}\left(S^{j+1} - S_0\right)\right\rangle & (4.93)\\ &\leq 2\left\|I - \alpha_j^S\ A_Q^* A_Q\right\| \left\|S^j - S_0\right\|_F \left\|S^{j+1} - S_0\right\|_F \\ &\qquad + 2\,\alpha_j^S\ \rho_2 \left\|L^{j+1} - L_0\right\|_F \left\|S^{j+1} - S_0\right\|_F , & (4.94)\end{aligned}$$

where in the second line we express W^j using the update rule $W^j = S^j - \alpha_j^S\ \mathcal{A}\left(\mathcal{A}(X^{j+\frac{1}{2}}) - b\right)$ and $b = \mathcal{A}(X_0)$, in the third line we expanded $X^{j+\frac{1}{2}} = L^{j+1} + S^j$, in the fourth line we split the inner product into two inner products by linearity of $\mathcal{A}$, and the inequality in the last line comes from Lemma 4.4 bounding the first two terms and Lemma 4.3 bounding the third term with $\rho_2 := \left(\Delta_2 + \frac{2\gamma_2}{1-\gamma_2^2}\right)$ where $\Delta_2 := \Delta_{2r,2s,\mu}$ and $\gamma_2 := \gamma_{2r,2s,\mu}$ since $\left(L^{j+1} - L_0 + S^{j+1} - S_0\right) \in \mathrm{LS}_{m,n}(2r, 2s, \mu)$. Dividing both sides of (4.94) by $\left\|S_0 - S^{j+1}\right\|_F$ gives

$$\left\|S_0 - S^{j+1}\right\|_F \leq 2\left\|I - \alpha_j^S\ A_Q^* A_Q\right\| \left\|S^j - S_0\right\|_F + 2\,\alpha_j^S\ \rho_2 \left\|L^{j+1} - L_0\right\|_F . \quad (4.95)$$

Adding together (4.88) and (4.95)

$$\begin{aligned}\left\|L_0 - L^{j+1}\right\|_F + \left\|S_0 - S^{j+1}\right\|_F \leq \\ 2\left\|I - \alpha_j^L\ A_Q^* A_Q\right\| \left\|L^j - L_0\right\|_F + 2\,\alpha_j^L\ \rho_2 \left\|S^j - S_0\right\|_F \\ +2\left\|I - \alpha_j^S\ A_Q^* A_Q\right\| \left\|S^j - S_0\right\|_F + 2\,\alpha_j^S\ \rho_2 \left\|L^{j+1} - L_0\right\|_F , \quad (4.96)\end{aligned}$$

which after rearranging terms in (4.96) becomes

$$\begin{aligned}\left(1 - 2\,\alpha_j^S\ \rho_2\right)\left\|L_0 - L^{j+1}\right\|_F + \left\|S_0 - S^{j+1}\right\|_F \\ \leq 2\left\|I - \alpha_j^L\ A_Q^* A_Q\right\| \left\|L^j - L_0\right\|_F \\ +2\left(\left\|I - \alpha_j^S\ A_Q^* A_Q\right\| + \alpha_j^L\ \rho_2\right)\left\|S^j - S_0\right\|_F \quad (4.97)\end{aligned}$$

and because $\alpha_j^S, \alpha_j^L, \Delta_2 \geq 0$ and $\gamma_2 \in (0,1)$, subtracting $2\,\alpha_j^S\,\rho_2 \|S_0 - S^{j+1}\|_F$ on the left does not increase the left hand side while adding $2\,\alpha_j^L\,\rho_2 \left\|L^j - L_0\right\|_F$ on the right does not decrease the right hand side of (4.97), therefore

$$\begin{aligned}\left(1 - 2\,\alpha_j^S\,\rho_2\right)\left(\left\|L_0 - L^{j+1}\right\|_F + \left\|S_0 - S^{j+1}\right\|_F\right) \\ \leq 2\left(\left\|I - \alpha_j^S\,A_Q^* A_Q\right\| + \alpha_j^L\,\rho_2\right)\left(\left\|L^j - L_0\right\|_F + \left\|S^j - S_0\right\|_F\right), \quad (4.98)\end{aligned}$$

Dividing both sides of (4.98) by $\left(1 - 2\,\alpha_j^S\,\rho_2\right)$ simplifies to

$$\begin{aligned}\left\|L_0 - L^{j+1}\right\|_F + \left\|S_0 - S^{j+1}\right\|_F \\ \leq 2\frac{\left\|I - \alpha_j^S\,A_Q^* A_Q\right\| + \alpha_j^L\,\rho_2}{1 - 2\,\alpha_j^S\,\rho_2}\left(\left\|L^j - L_0\right\|_F + \left\|S^j - S_0\right\|_F\right). \quad (4.99)\end{aligned}$$

By Lemma 4.4, the eigenvalues of $\left(I - \alpha_j\,A_Q^* A_Q\right)$ can be bounded as

$$1 - \alpha_j\,(1 + \Delta_3) \leq \lambda\left(I - A_Q^* A_Q\right) \leq 1 + \alpha_j\,(1 - \Delta_3), \quad (4.100)$$

with $\Delta_3 := \Delta_{3r,3s,\mu}$ being the RIC of $\mathcal{A}$. By α_j^L and α_j^S being the normalized stepsizes as introduced in (Blumensath and Davies, 2010; Tanner and Wei, 2013)

$$\alpha_j^L = \frac{\left\|\mathrm{Proj}_{(U^j,\Omega^j)}\left(R^j\right)\right\|_F^2}{\left\|\mathcal{A}\left(\mathrm{Proj}_{(U^j,\Omega^j)}\left(R^j\right)\right)\right\|_2^2} \quad \text{and} \quad \alpha_j^S = \frac{\left\|\mathrm{Proj}_{(U^{j+1},\Omega^j)}\left(R^j\right)\right\|_F^2}{\left\|\mathcal{A}\left(\mathrm{Proj}_{(U^{j+1},\Omega^j)}\left(R^j\right)\right)\right\|_2^2} \quad (4.101)$$

where the projection $\mathrm{Proj}_{(U^j,\Omega^j)}\left(R^j\right), \mathrm{Proj}_{(U^{j+1},\Omega^j)}\left(R^{j+\frac{1}{2}}\right)$ ensures that the residual R^j and $R^{j+\frac{1}{2}}$ is projected into the set $\mathrm{LS}_{m,n}(r,s,\mu)$. Then, it follows from the RIC for $\mathcal{A}$ that the stepsizes α_j^L, α_j^S can be bounded as

$$\frac{1}{1 + \Delta_1} \leq \alpha_j^{L/S} \leq \frac{1}{1 - \Delta_1}, \quad (4.102)$$

where $\Delta_1 := \Delta_{r,s,\mu}$. Putting (4.100) and (4.102) together

$$1 - \frac{1 + \Delta_3}{1 - \Delta_1} \leq \lambda\left(I - \alpha_j^{L/S}\,A_Q^* A_Q\right) \leq 1 - \frac{1 - \Delta_3}{1 + \Delta_1}. \quad (4.103)$$

Since $\Delta_3 \geq \Delta_1$ we have that the magnitude of the lower bound in (4.103) is greater than the upper bound. Therefore

$$\frac{1 + \Delta_3}{1 - \Delta_1} - 1 \geq \left\|I - \alpha_j^{L/S}\,A_Q^* A_Q\right\|_2. \quad (4.104)$$

Finally, the constant on the right hand side of (4.99) can be upper bounded

$$\eta := 2\frac{\left\|I - \alpha_j^S A_Q^* A_Q\right\| + \alpha_j^L \rho_2}{1 - 2\,\alpha_j^S \rho_2} \tag{4.105}$$

$$\leq 2\frac{\left(\frac{1+\Delta_3}{1-\Delta_1} - 1\right) + \frac{1}{1-\Delta_1}\left(\Delta_2 + \frac{2\gamma_2}{1-\gamma_2^2}\right)}{1 - 2\frac{1}{1-\Delta_1}\left(\Delta_2 + \frac{2\gamma_2}{1-\gamma_2^2}\right)} = 2\frac{\Delta_3 + \Delta_1 + \Delta_2 + \frac{2\gamma_2}{1-\gamma_2^2}}{1 - \Delta_1 - 2\Delta_2 - \frac{4\gamma_2}{1-\gamma_2^2}} \tag{4.106}$$

$$\leq \frac{6\Delta_3 + \frac{4\gamma_2}{1-\gamma_2^2}}{1 - 3\Delta_3 - \frac{4\gamma_2}{1-\gamma_2^2}} \tag{4.107}$$

$$\leq \frac{6\Delta_3 + \frac{9}{2}\gamma_2}{1 - 3\Delta_3 - \frac{9}{2}\gamma_2} \tag{4.108}$$

where the inequality in the second line in (4.106) comes from upper bounds in (4.104) and in (4.102), the third line in (4.107) follows from $\gamma_3 \geq \gamma_2 \geq \gamma_1$, and the last inequality in (4.108) follows from $\frac{4\gamma_2}{1-\gamma_2^2} < \frac{9}{2}$ when $\gamma_2 < 9$.

To ensure that $\eta < 1$, it suffices to show that the right-hand side in (**??**) is smaller than one, which translates to

$$6\Delta_3 + \frac{9}{2}\gamma_2 < 1 - 3\Delta_3 - \frac{9}{2}\gamma_2. \tag{4.109}$$

Rearranging of the terms in the above inequality in (4.109) results into

$$\Delta_3 \leq \frac{1}{9} - \gamma_2, \tag{4.110}$$

which is satisfied only when $\gamma_2 < 1/9$, i.e. the RICs are positive $\Delta_3 > 0$.

For $\Delta_{3r,3s,\mu} < \frac{1}{9} - \gamma_{2r,2s,\mu}$ the inequality in (4.99) implies contraction of the error

$$\left\|L_0 - L^{j+1}\right\|_F + \left\|S_0 - S^{j+1}\right\|_F \leq \eta\left(\left\|L^j - L_0\right\|_F + \left\|S^j - S_0\right\|_F\right), \tag{4.111}$$

because $\eta < 1$ which guarantees linear convergence of iterates L^j and S^j to L_0 and S_0 respectively. □

4.5 EMPIRICAL AVERAGE CASE PERFORMANCE

Figure 4.1 presents empirically observed phase transitions, which indicate the values of model complexity r, s and measurements p for which recovery is possible. Figure 4.2 and 4.3 gives examples of convergence rates for NIHT, NAHT, and SpaRCS, including contrasting different methods to implement the projection NIHT, step 5 of Algorithm 1. An additional phase transition simulation for the convex relaxation is given in Figure 4.4.

Synthetic matrices $X_0 = L_0 + S_0 \in \mathrm{LS}_{m,n}(r, s, \mu)$ are generated using the experimental setup proposed in the Robust PCA literature (Netrapalli et al., 2014; Yi et al., 2016; Cai et al., 2019). The low-rank component is formed as

$L_0 = V^T$, where $U \in \mathbb{R}^{m\times r}, V \in \mathbb{R}^{n\times r}$ are two random matrices having their entries drawn i.i.d. from the standard Gaussian distribution. The support set of the sparse component S_0 is generated by sampling a uniformly random subset of $[m] \times [n]$ indices of size s and each non-zero entry $(S_0)_{i,j}$ is drawn from the uniform distribution over $\left[-\mathrm{E}\left(|(L_0)_{i,j}|\right), \mathrm{E}\left(|(L_0)_{i,j}|\right)\right]$.

Each synthetic matrix is measured using linear operators $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$. The random Gaussian measurement operators are constructed by p matrices $A^{(\ell)} \in \mathbb{R}^{m\times n}$ whose entries are sampled from Gaussian distribution $A^{(\ell)}_{i,j} \sim \mathcal{N}(0, 1/p)$ where p is the number of measurements. The Fast Johnson-Lindenstrauss Transform is implemented as

$$\mathcal{A}_{\mathrm{FJLT}}(X) = RHD \operatorname{vec}(X), \tag{4.112}$$

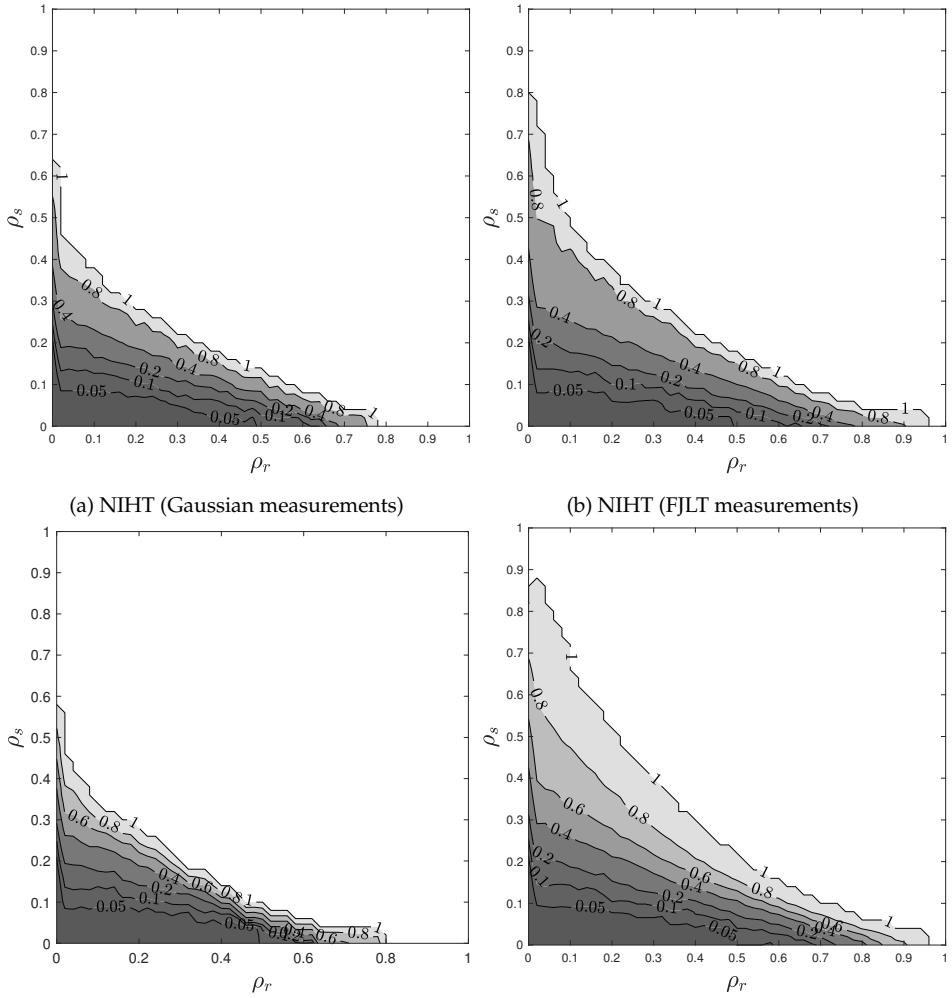
where $R \in \mathbb{R}^{p\times mn}$ is a restriction matrix constructed from a $mn \times mn$ identity matrix with p rows randomly selected, $H \in \mathbb{R}^{mn\times mn}$ is discrete cosine transform matrix, $D \in \mathbb{R}^{mn\times mn}$ is a diagonal matrix whose entries are sampled independently randomly from $\{-1, 1\}$, and $\operatorname{vec}(X) \in \mathbb{R}^{mn}$ is the vectorized matrix $X \in \mathbb{R}^{m\times n}$.

Theorem 4.2, Theorem 4.3, and Theorem 4.4 indicate that recovery of X_0 from $\mathcal{A}(X_0)$ depends on the problem dimensions through the ratios of the number of measurements p with the ambient dimension mn, and the minimum number of measurements, $r(m + n - r) + s$, through an undersampling and two oversampling ratios

$$\delta = \frac{p}{mn} \quad \text{and} \quad \rho_r = \frac{r(m+n-r)}{p}, \quad \rho_s = \frac{s}{p}. \tag{4.113}$$

The matrix dimensions m and n are held fixed, while p, r and s are chosen according to varying parameters δ, ρ_r and ρ_s. For each pair of $\rho_r, \rho_s \in \{0, 0.02, 0.04, \ldots, 1\}$ where $\rho_r + \rho_s \leq 1$, with the sampling ratio restricted to values $\delta \in \{0.02, 0.04, \ldots, 1\}$, 20 simulated recovery tests are conducted and we compute the critical subsampling ratio δ^* above which more than half of the experiments succeeded. For the linear transform $\mathcal{A}$ drawn from the (dense) Gaussian distribution, the highest per iteration cost in NIHT and NAHT comes from applying $\mathcal{A}$ to the residual matrix, which requires pmn scalar multiplications which scales proportionally to $(mn)^2$. For this reason, our tests are restricted to the matrix size of $m = n = 100$ in the case of NIHT and NAHT, and to a smaller size $m = n = 30$ for testing the recovery by solving the convex relaxation (4.3) with semidefinite programming (Toh et al., 1999) that has $O\left((mn)^2\right)$ variables which is more computationally demanding[1] compared to the hard thresholding gradient descent methods. Algorithms are terminated at iteration ℓ when either: the relative residual

[1] As an example, a low-rank plus sparse matrix with $m = n = 100$ with $\rho_r = \rho_s = 0.1$ undersampled and measured with Gaussian matrix with $\delta = 0.5$ takes 2.5 seconds and 2.3 seconds to recover using NIHT and NAHT respectively, while the recovery using the convex relaxation takes over 7 hours.



(a) NIHT (Gaussian measurements)

(b) NIHT (FJLT measurements)

(c) NAHT (Gaussian measurements)

(d) NAHT (FJLT measurements)

Figure 4.1: Phase transition level curves denoting the value of δ^* for which values of ρ_r and ρ_s *below which* are recovered for at least half of the experiments for δ, ρ_r, and ρ_s as given by (4.113). NIHT is observed to recover matrices of higher ranks and sparsities from FJLT than from Gaussian measurements, while the phase transitions for NIHT and NAHT are comparable. The RPCA projection in NIHT, step 5 in Alg. 1, is performed by AccAltProj (Cai et al., 2019).

error is smaller than 10^{-6}, that is when $\left\|\mathcal{A}(X^\ell) - b\right\|_2 / \|b\|_2 \leq 10^{-6} \|b\|_2$, or the relative decrease in the objective is small

$$\left(\frac{\left\|\mathcal{A}(X^\ell) - b\right\|_2}{\|\mathcal{A}(X^{\ell-15}) - b\|_2} \right)^{1/15} > 0.999, \tag{4.114}$$

or the maximum of 300 iterations is reached. An algorithm is considered to have successfully recovered $X_0 \in \mathrm{LS}_{m,n}(r, s, \mu)$ if it returns a matrix $X^\ell \in \mathrm{LS}_{m,n}(r, s, \mu)$ that is within 10^{-2} of X_0 in the relative Frobenius error, $\left\|X^\ell - X_0\right\|_F \leq 10^{-2} \|X_0\|_F$.

Figure 4.1 depicts the phase transitions of δ above which NIHT and NAHT successfully recovers X_0 in more than half of the experiments. For example, the level curve 0.4 in Figure 4.1 denotes the values of ρ_r and ρ_s *below which* recovery is possible for at least half of the experiments for $p = 0.4mn$ and ρ_r, ρ_s as given by (4.113). Note that the bottom left portion of Figure 4.1 corresponds to smaller values of model complexity (r, s) and are correspondingly easier to recover than larger values of (r, s). Both algorithms are observed to recover matrices with prevalent rank structure,

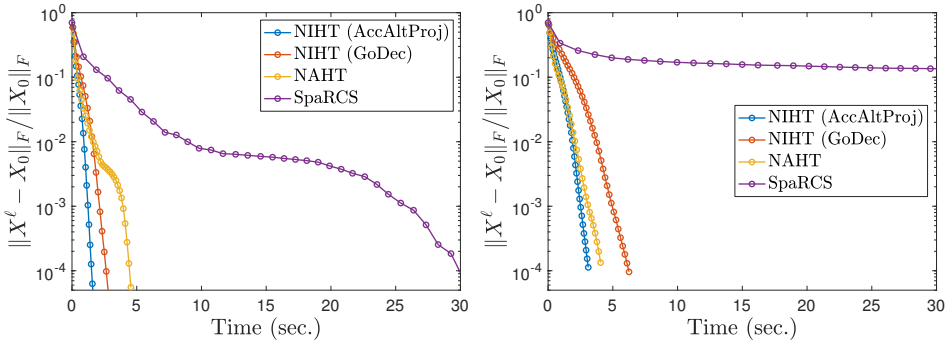
$\rho_r \leq 0.6$, even from very few measurements as opposed to matrices with prevalent sparse structure requiring in general more measurements for a successful recovery. Phase transitions corresponding to the sparse-only ($\rho_r = 0$) and to the rank-only ($\rho_s = 0$) cases are roughly in agreement with phase transitions that have been observed for non-convex algorithms in compressed sensing (Blanchard and Tanner, 2015) and matrix completion literature (Tanner and Wei, 2013; Blanchard et al., 2015). We observe that NAHT achieves almost identical performance to NIHT in terms of possible recovery despite not requiring the computationally expensive Robust PCA projection in every iteration. For both algorithms we see that the successful recovery is possible for matrices with higher ranks and sparsities in the case of FJLT measurements compared to Gaussian measurements.

Equivalent experiments are conducted for the convex relaxation (4.3), but with smaller matrix size 30×30 and limited to 10 simulations for each set of parameters due to the added computational demands. The convex optimization is formulated using the CVX modeling framework developed by Grant and Boyd (2014, 2008) and solved in Matlab by the semidefinite programming optimization package SDPT3 from Toh et al. (1999). We observe that recovery by solving the convex relaxation is successful for somewhat lower ranks and sparsities and requiring a larger sampling ratio δ compared to the non-convex algorithms.

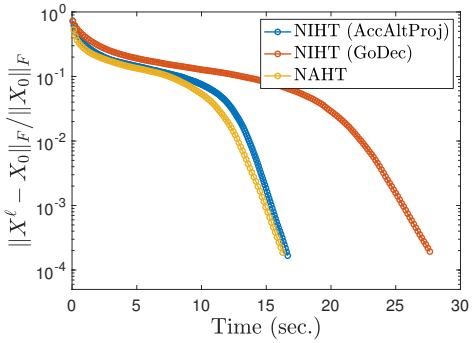
The observed phase transitions of the convex relaxation alongside phase transitions for $m = n = 30$ experiments with NIHT and NAHT are depicted in Figure 4.4. Comparing the phase transitions of the non-convex algorithms in Figure 4.1 and Figure 4.4 show that with the increased problem size, the phase transition are independent of the dimension with only small differences due to the finite-dimensional effects of the smaller problem size in the case of $m = n = 30$.

Figure 4.2 presents convergence timings of Matlab implementations of the three non-convex algorithms used for recovery of matrices with $m = n = 100$ from $p = (1/2)10^2$ ($\delta = 1/2$) measurements and three values of $\rho_r = \rho_s = \{0.05, 0.1, 0.2\}$. The convergence results are presented for two variants of NIHT with different Robust PCA algorithms Accelerated Alternating Projection (AccAltProj) by Cai et al. (2019) and Go Decomposition (GoDec) by Zhou and Tao (2011) in the projection step 5 of Algorithm 1. Both NIHT and NAHT converge at a much faster rate than the existing non-convex algorithm for low-rank plus sparse matrix recovery SpaRCS by Waters et al. (2011). All the algorithms take longer to recover a matrix for increased rank r and/or sparsity s.

The computational efficacy of NIHT compared to NAHT depends on the cost of computing the Robust PCA calculation in comparison to the cost of applying $\mathcal{A}$. NAHT computes two step sizes in each iteration which results into computing $\mathcal{A}$ twice per iteration in comparison to just one such computation per iteration in the case of NIHT. On the other hand, NIHT involves

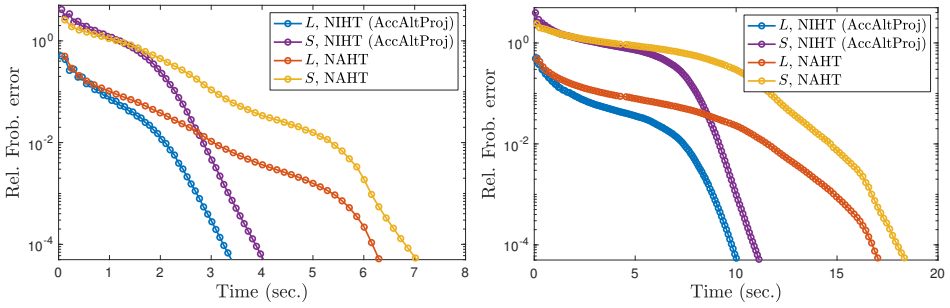


(a) $\rho_r = \rho_s = 0.05$ (b) $\rho_r = \rho_s = 0.1$



(c) $\rho_r = \rho_s = 0.2$

Figure 4.2: Relative error in the approximate X^ℓ as a function of time for synthetic problems with $m = n = 100$ and $p = (1/2)100^2$, $\delta = 1/2$, for Gaussian linear measurements $\mathcal{A}$. In (b), SpaRCS converged in 171 sec. (45 iterations), and in (c), SpaRCS did not converge.



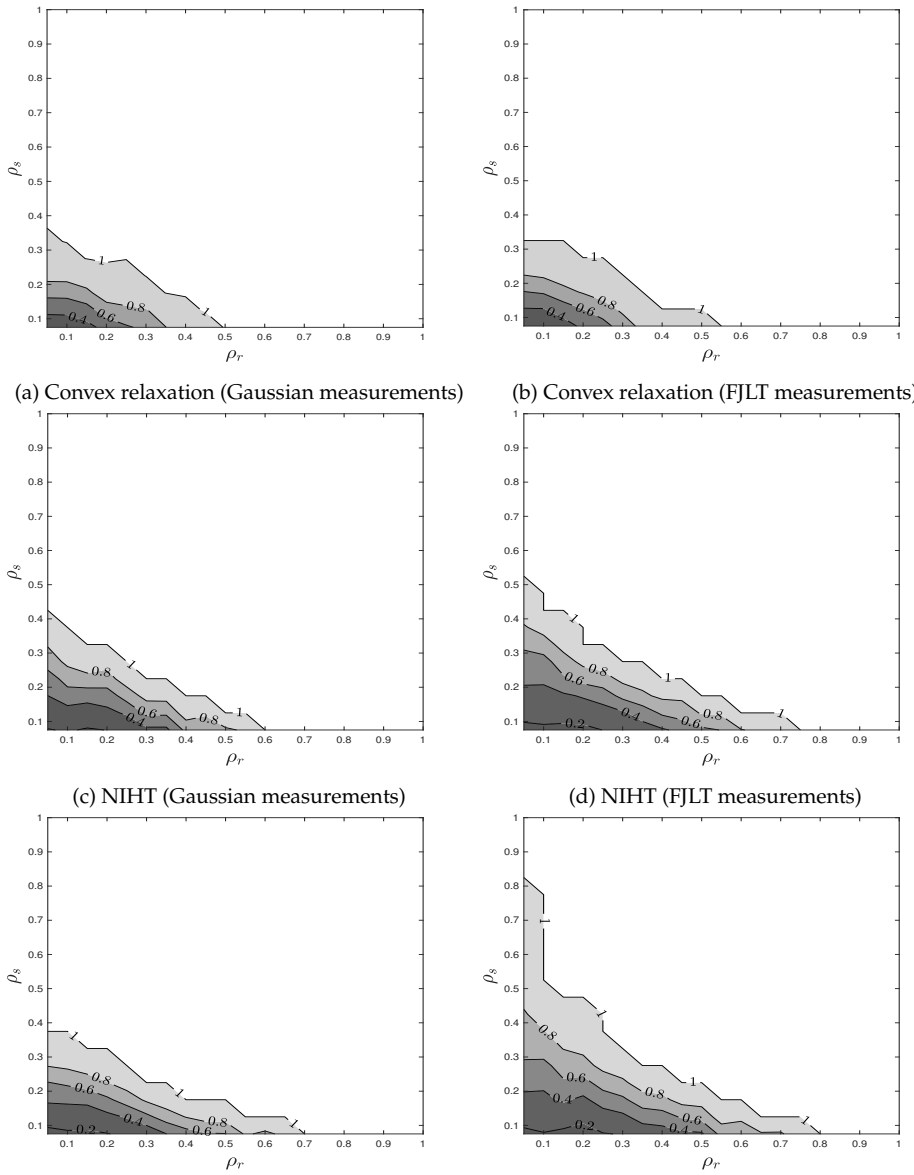
(a) $\delta = 0.5, \rho_r = 0.05, \rho_s = 0.15$ (b) $\delta = 0.5, \rho_r = 0.05, \rho_s = 0.25$

Figure 4.3: Error between the approximate recovered low-rank and sparse components L^ℓ and S^ℓ and the true low-rank and sparse components L_0 and S_0. Error is plotted as a function of recovery time for synthetic problems with $m = n = 100$ and $p = (1/2)100^2$, $\delta = 1/2$, for Gaussian linear measurements $\mathcal{A}$.

solving Robust PCA in every iteration for the projection step whereas NAHT performs computationally cheaper singular value decomposition (SVD) and sparse hard thresholding projection.

Figure 4.3 illustrates the convergence of the individual low-rank and sparse components $\|L^\ell - L_0\|_F$ and $\|S^\ell - S_0\|_F$ as a function of time. The algorithms are observed to approximate the low-rank factor more accurately than the sparse component and that the computational time increases for larger values of sparsity fraction ρ_s. Moreover, for both NIHT and NAHT the relative error of both components decreases together.

Figure 4.4 depicts the phase transitions of δ above which NIHT, NAHT and solving the convex relaxation problem in (4.3) successfully recovers X_0 in more than half of the experiments. Comparing Figure 4.4 to Figure 4.1 we see that the phase transitions roughly occur for the same parameters ρ_r, ρ_s with only small differences due to the finite-dimensional effects of



(a) Convex relaxation (Gaussian measurements)

(b) Convex relaxation (FJLT measurements)

(c) NIHT (Gaussian measurements)

(d) NIHT (FJLT measurements)

(e) NAHT (Gaussian measurements)

(f) NAHT (FJLT measurements)

Figure 4.4: Phase transition level curves denoting the value of δ^* for which values of ρ_r and ρ_s *below which* are recovered for at least 5 out of 10 experiments for δ, ρ_r, and ρ_s as given by (4.113). The convex optimization problem is solved by SDPT3 (Toh et al., 1999). NIHT and NAHT are observed to recover matrices of higher ranks and sparsities compared to solving the convex relaxation.

the smaller problem size being more pronounced when $m = n = 30$. We also observe that non-convex algorithms perform better than the convex relaxation in that they are able to recover higher ranks and sparsities from fewer samples in addition to also taking less time to converge.

4.6 APPLICATIONS

We now present exemplary applications of the low-rank plus sparse matrix recovery in dynamic-foreground/static-background and computational multispectral imaging. Software to reproduce the experiments of this section is publicly available at: `github.com/SimonVary/lrps-recovery`.

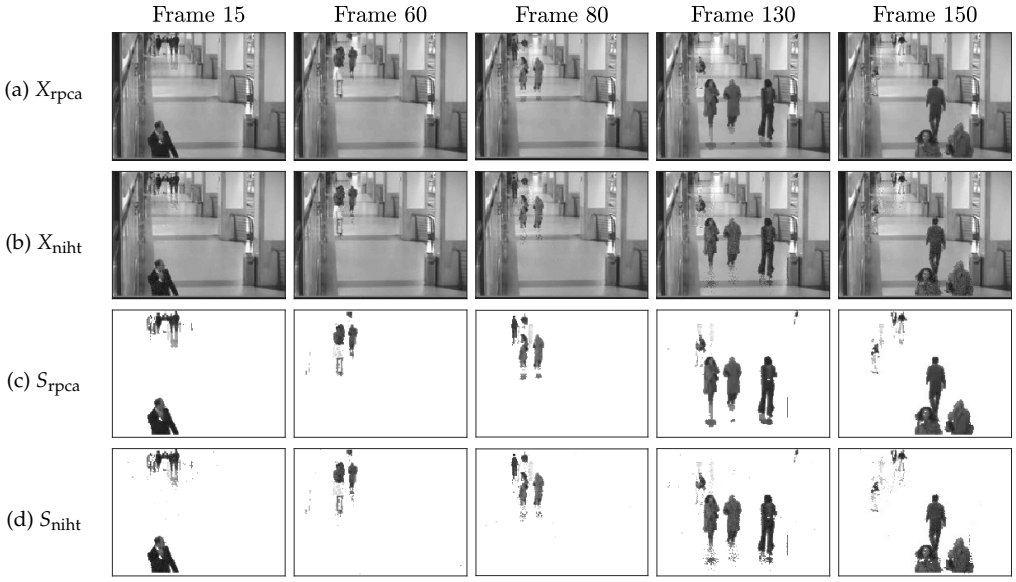
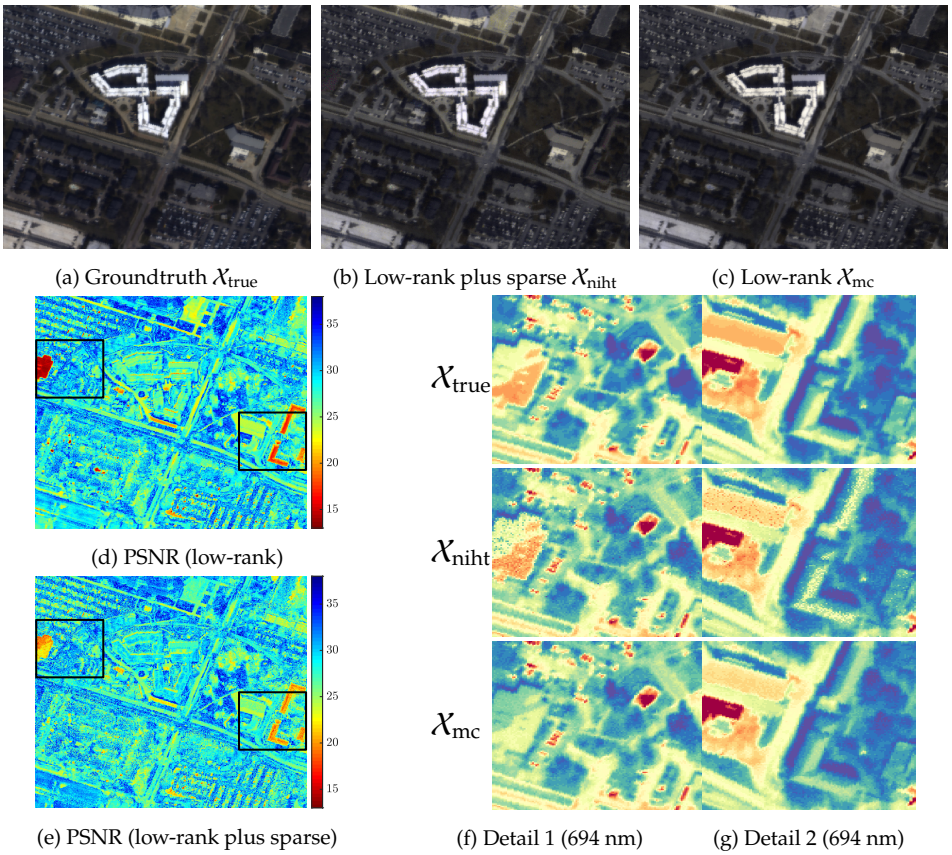


Figure 4.5: NIHT recovery results of a $190 \times 140 \times 150$ video sequence compared to the approximation of the complete video sequence by Robust PCA (AccAltProj (Cai et al., 2019)). The video sequence is reshaped into a $26\,600 \times 150$ matrix and either recovered from FJLT measurements with $\delta = 0.33$ using rank $r = 1$ and sparsity $s = 197\,505$ or approximated from the full video sequence by computing RPCA by AccAltProj with the same rank and sparsity parameters. Recovery by NIHT from subsampled information achieves PSNR of 34.5 dB whereas the Robust PCA approximation from the full video sequence achieves PSNR of 35.5 dB.

4.6.1 *Dynamic-foreground/static-background video separation*

Background/foreground separation is the task of distinguishing moving objects from the static-background in a time series, e.g. a video recording. A widely used approach is to arrange frames of the video sequence into an $m \times n$ matrix, where m is the number of pixels and n is the number of frames of the recording and apply Robust PCA to decompose the matrix into the sum of a low-rank and a sparse component which model the static background and dynamic foreground respectively (Bouwmans et al., 2017). Herein we consider the same problem but with the additional challenge of recovering the video sequence from subsampled information (Waters et al., 2011) analogous to compressed sensing.

We apply NIHT, Algorithm 1, to the well studied shopping mall surveillance introduced by Li et al. (2004) which is a $190 \times 140 \times 150$ video sequence. The video sequence is rearranged into a matrix of size $26\,600 \times 150$ and measured using subsampled FJLT (4.112) with one third as many measurements as the ambient dimension, $\delta = 0.33$. The static-background is modeled with a rank-r matrix with $r = 1$ and the dynamic-foreground by an s-sparse matrix with $s = 197\,505$ ($\rho_r = 0.02$, $\rho_s = 0.15$). Figure 4.5 displays the reconstructed image X_{niht} and its sparse component S_{niht} alongside the results obtained from applying Robust PCA (AccAltProj by Cai et al. (2019)) which makes use of the fully sampled video sequence rather than the one-third measurements available to NIHT. NIHT accurately estimates the video sequence achieving PSNR of 34.5 dB while also separating the low-rank background from the sparse foreground. The results are of a similar visual quality to the case of Robust PCA that achieves PSNR of 35.5 dB which requires access to the full video sequence.



(a) Groundtruth $\mathcal{X}_{\text{true}}$ (b) Low-rank plus sparse $\mathcal{X}_{\text{niht}}$ (c) Low-rank $\mathcal{X}_{\text{mc}}$

(d) PSNR (low-rank)

(e) PSNR (low-rank plus sparse)

(f) Detail 1 (694 nm) (g) Detail 2 (694 nm)

Figure 4.6: Recovery by NIHT from FJLT measurements with $\delta = 0.33$ using low-rank model ($\rho_r = 0.25$, $\rho_s = 0$) compared to the low-rank plus sparse model ($\rho_r = 0.25$, $\rho_s = 0.05$). Figure 4.6a - 4.6b show the color renderings of the original multispectral image and the two recovered images. Figure 4.6d and Figure 4.6e show the spatial PSNR of the recovery from the low-rank only model (overall PSNR of 38.9 dB) and the low-rank plus sparse model (overall PSNR of 39.1 dB) respectively. Figure 4.6f and Figure 4.6g show two details of size 128×128 in the 694 nm band.

4.6.2 *Computational multispectral imaging*

A multispectral image captures a wide range of light spectra generating a vector of spectral responses at each image pixel thus acquiring information in the form of a third-order tensor. The low-rank model has a vital role in multispectral imaging in the form of linear spectral mixing models that assume the spectral responses of the imaged scene are well approximated as a linear combination of spectral responses of only few core materials referred to as *endmembers* (Dimitris et al., 2003). As such, the low-rank structure can be exploited by computational imaging systems which acquire the image in a compressed from and use computational methods to recover a high-resolution image (Cao et al., 2016; Degraux et al., 2015; Antonucci et al., 2019). However, when different materials are in close proximity the resulting spectrum can be a highly nonlinear combination of the *endmembers* resulting in anomalies of the model (Stein et al., 2002). Herein we propose the low-rank plus sparse matrix recovery as a way to model the spectral anomalies in the low-rank structure.

We employ NIHT on a $512 \times 512 \times 48$ airborne hyperspectral image from the GRSS 2018 Data Fusion contest (Xu et al., 2019) that is rearranged into a matrix of size $262\,144 \times 48$ and subsampled using FJLT with $\delta = 0.33$. Figure 4.6 demonstrates recovery by NIHT using rank $r = 3$ and sparsity $s = 150\,995$ ($\rho_r = 0.25$, $\rho_s = 0.05$) in comparison with the the low-rank

model with rank $r = 3$ and $s = 0$ ($\rho_r = 0.25$, $\rho_s = 0$). Both methods recover the image well but the low-rank plus sparse recovery achieves slightly higher PSNR of 39.1 dB compared to the low-rank recovery that has PSNR of 38.9 dB and slightly better fine details. Figure 4.6d and Figure 4.6e depict the localization of the error in terms of PSNR and shows that adding the sparse component improves PSNR of a few localized parts. Although the overall gain in the PSNR is small compared to the low-rank model, the differences in the localized regions of the image can be potentially impactful when further analyzed in practical applications such as semantic segmentation (Kemker et al., 2018).

4.7 SUMMARY AND DISCUSSION

This chapter began with the aim of recovering an unknown low-rank plus sparse matrix X_0 given a vector of measurements $b = \mathcal{A}(X_0)$ and a subsampling operator $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$.

We first showed in Theorem 4.1 that if the RICs of $\mathcal{A}$ are bounded, there exists a unique matrix $X_0 \in \mathrm{LS}_{m,n}(r, s, \mu)$ such that $\mathcal{A}(X_0) = b$. We then reviewed the most popular algorithms for compressed sensing, matrix sensing, and Robust PCA.

In §4.3, we proved Theorem 4.2 that if the RICs are sufficiently bounded, solving a convex optimisation problem recovers X_0 robustly even in the presence of the measurement error and/or model mismatch.

We proposed in §4.4 two gradient descent methods: Normalized Iterative Hard Thresholding (NIHT) and Normalized Alternating Hard Thresholding (NAHT). The first one, NIHT is based on the analogous compressed sensing and matrix sensing algorithms developed by (Blumensath and Davies, 2010) and (Tanner and Wei, 2013) respectively. The main disadvantage of NIHT for recovery of low-rank plus sparse matrices is that in every iteration it requires to perform an oblique projection on the set $\mathrm{LS}_{m,n}(r, s, \mu)$ by solving a Robust PCA problem. Solving Robust PCA is expensive because it usually involves an iterative process that solves SVD in every iteration. This hindrance is overcome by NAHT, which updates the low-rank and the sparse components in an alternating fashion and applies the projections separately and is therefore faster.

We prove, that if RICs are sufficiently bounded, both NIHT (Theorem 4.3) and NAHT (Theorem 4.4) converge to the unique minimizer, or in the case of NIHT to a matrix sufficiently close to the minimizer depending on the accuracy of the used oblique projection method. The non-convex algorithms are also stable to error ε_b, but we omit the stability analysis for clarity in the proofs.

Both the convex formulation and NAHT, also provably solve Robust PCA with the optimal order of the number of corruptions $s = O\left(mn/\left(\mu^2 r^2\right)\right)$ when the sensing operator is the identity, and therefore satisfies the RICs with

$\Delta = 0$.

We implemented the methods in MATLAB and performed a range of numerical experiments. Firstly, in §4.5, we empirically observe a phase transition in the parameter space for which the recovery is successful by convex relaxation and the two non-convex methods. Numerical experiments on synthetic data empirically demonstrate phase transitions in the parameter space for which the recovery is possible. Experiments for dynamic-foreground/static-background video separation show that the segmentation of moving objects can be obtained with similar error from only one third as many measurements as compared to the entire video sequence.

The contributions here open up the possibility of other algorithms in compressed sensing and low-rank matrix completion/sensing to be extended to the case of low-rank plus sparse matrix recovery, e.g. more efficient algorithms such as those employing momentum (Kyrillidis and Cevher, 2014; Wei, 2015) or minimising over increasingly larger subspaces (Blanchard et al., 2015).

4.8 SUPPORTING LEMMATA

In the proof of Theorem 4.2 we make use of the following Lemma 4.1 and Corollary 4.1 from Recht et al. (2010) which we restate here for completeness.

Lemma 4.1 ((Recht et al., 2010, Lemma 3.4)). *Let A and B be matrices of the same dimensions. Then there exist matrices B_1 and B_2 such that*

1. $B = B_1 + B_2$,
2. $\text{rank}B_1 \leq 2\text{rank}A$,
3. $AB_2^T = 0$ *and* $A^T B_2 = 0$,
4. $\langle B_1, B_2 \rangle = 0$.

Corollary 4.1 ((Recht et al., 2010, Lemma 2.3)). *Let A and B be matrices of the same dimensions. If $AB^T = 0$ and $A^T B = 0$, then $\|A + B\|_* = \|A\|_* + \|B\|_*$.*

Lemma 4.2 (Decomposing $R^S = R_0^S + R_c^S$). *Let* $\text{supp}\, S_0 = \Omega_0$ *and construct a matrix R_0^S that has the entries of R^S at indices Ω_0*

$$(R_0^S)_{i,j} = \begin{cases} (R^S)_{i,j} & \text{if} \quad (i,j) \in \Omega_0, \\ 0 & \text{if} \quad (i,j) \notin \Omega_0, \end{cases} \tag{4.115}$$

and a matrix $R_c^S = R^S - R_0^S$ that has the entries of R^S at the indices of the complement of Ω_0. Then

1. $\|R_0^S\|_0 \leq \|S_0\|_0 = s \quad$ *(by $|\Omega_0| = s$),*
2. $\|S_0 + R_c^S\|_1 = \|S_0\|_1 + \|R_c^S\|_1 \quad$ *(by* $\text{supp}(R_0^S) \cap \text{supp}(R_c^S) = \emptyset$*),*
3. $\langle R_0^S, R_c^S \rangle = 0 \quad$ *(by* $\text{supp}(R_0^S) \cap \text{supp}(R_c^S) = \emptyset$*).*

Proof. It can be easily verified that R_0^S and R_c^S constructed as in (4.115) satisfy the conditions (1)-(3). □

Lemma 4.3 (Upper bound on $\langle \mathcal{A}(\cdot), \mathcal{A}(\cdot) \rangle$)**.** *For an operator $\mathcal{A}(\cdot)$ whose RICs are upper bounded by $\Delta_2 := \Delta_{2r,2s,\mu}$ and two incoherent low-rank plus sparse matrices $X_1 = L_1 + S_1 \in \mathrm{LS}_{m,n}(r, s, \mu)$, $X_2 = L_2 + S_2 \in \mathrm{LS}_{m,n}(r, s, \mu)$ that have orthogonal components $\langle L_1, L_2 \rangle = 0$, $\langle S_1, S_2 \rangle = 0$ and have bounded the rank-sparsity coefficient $\gamma_2 := \gamma_{2r,2s,\mu} < 1$, we have that*

$$\left| \langle \mathcal{A}(X_1), \mathcal{A}(X_2) \rangle \right| \leq \left(\Delta_2 + \frac{2\gamma_2}{1 - \gamma_2^2} \right) \|X_1\|_F \, \|X_2\|_F, \tag{4.116}$$

where $\gamma_2 = \mu \frac{2r\sqrt{2s}}{\sqrt{mn}}$ is the rank-sparsity correlation coefficient as defined in Lemma 2.3 on page 17.

Proof. By $\mathcal{A}(\cdot)$ being a linear transform, bilinearity of the inner-product, and conicity of $\mathrm{LS}_{m,n}(r, s, \mu)$, we can assume without loss of generality that $\|X_1\|_F = 1$ and $\|X_2\|_F = 1$. The parallelogram law applied to $\|\mathcal{A}(X_1)\|_2$ and $\|\mathcal{A}(X_2)\|_2$ yields

$$2 \left(\|\mathcal{A}(X_1)\|_2^2 + \|\mathcal{A}(X_2)\|_2^2 \right) = \|\mathcal{A}(X_1) + \mathcal{A}(X_2)\|_2^2 + \|\mathcal{A}(X_1) - \mathcal{A}(X_2)\|_2^2 . \tag{4.117}$$

Subtracting $2 \|\mathcal{A}(X_1) - \mathcal{A}(X_2)\|_2^2$ from both sides of (4.117)

$$4 \langle \mathcal{A}(X_1), \mathcal{A}(X_2) \rangle = \|\mathcal{A}(X_1) + \mathcal{A}(X_2)\|_2^2 - \|\mathcal{A}(X_1) - \mathcal{A}(X_2)\|_2^2 . \tag{4.118}$$

We can expand the equality in (4.118) to bound its right-hand side using the RICs as

$$\begin{aligned}
|\langle \mathcal{A}(X_1), \mathcal{A}(X_2) \rangle| &= \frac{1}{4} \left| \|\mathcal{A}(X_1 + X_2)\|_F^2 - \|\mathcal{A}(X_1 - X_2)\|_F^2 \right| && (4.119) \\
&\leq \frac{1}{4} \left| (1 + \Delta_2) \|X_1 + X_2\|_F^2 - (1 - \Delta_2) \|X_1 - X_2\|_F^2 \right| && (4.120) \\
&\leq \frac{1}{4} \Big| (1 + \Delta_2) \left(\|X_1\|_F^2 + 2\langle X_1, X_2 \rangle + \|X_2\|_F^2 \right) \\
&\qquad - (1 - \Delta_2) \left(\|X_1\|_F^2 - 2\langle X_1, X_2 \rangle + \|X_2\|_F^2 \right) \Big| && (4.121) \\
&= \left| \frac{\Delta_2}{2} \left(\|X_1\|_F^2 + \|X_2\|_F^2 \right) + \langle X_1, X_2 \rangle \right| = \left| \Delta_2 + \langle X_1, X_2 \rangle \right| \\
& && (4.122)
\end{aligned}$$

where the inequality in the second line in (4.120) comes from the RICs of $\mathcal{A}(\cdot)$ and by $X_1 + X_2$ and $X_1 - X_2$ being in the set $\mathrm{LS}_{m,n}(2r, 2s, \mu)$ combined with Lemma 2.9, the equality in the third line in (4.121) is the result of expanding the inner products, and finally, the last equality in (4.121) comes from elementary operations and using the fact that $\|X_1\| = 1$ and $\|X_2\| = 1$.

Moreover, by X_1 and X_2 being component-wise orthogonal $\langle L_1, L_2 \rangle = 0$ and $\langle S_1, S_2 \rangle = 0$, we can upper-bound the magnitude of the correlation

between X_1 and X_2 as

$$|\langle X_1, X_2\rangle| = |\langle L_1, L_2\rangle + \langle L_1, S_2\rangle + \langle L_2, S_1\rangle + \langle S_1, S_2\rangle| \tag{4.123}$$
$$= |\langle L_1, S_2\rangle + \langle L_2, S_1\rangle| \tag{4.124}$$
$$\leq \gamma_2 \Big(\|L_1\|_F \|S_2\|_F + \|L_2\|_F \|S_1\|_F \Big) \tag{4.125}$$
$$\leq \frac{2\gamma_2}{1-\gamma_2^2}, \tag{4.126}$$

where in the first equality in (4.123) we expanded the inner-product, the second equality in (4.124) is the consequence of the components being orthogonal, the inequality in the third line in (4.125) is the consequence of Lemma 2.1, and the last inequality in (4.126) comes from the upper-bound of the norms $\|L_1\|_F, \|L_2\|_F, \|S_1\|_F, \|S_2\|_F$ from Lemma 2.8 and by $\|X_1\|_F = 1$ and $\|X_2\|_F = 1$.

We can now further upper bound (4.122) using the bound in (4.122) combined with the triangle on the absolute value

$$\Big|\langle \mathcal{A}(X_1), \mathcal{A}(X_2)\rangle\Big| \leq \Delta_2 + \frac{2\gamma_2}{1-\gamma_2^2}, \tag{4.127}$$

when $\|X_1\|_F = 1$ and $\|X_2\|_F = 1$ which translates into the bound in (4.116) in the general case.

Note that the bound can be lowered for specific matrices X_1, X_2 such that the matrices of their sums $X_1 + X_2$ and $X_1 - X_2$ are in $\text{LS}(r, s, \mu)$ sets with smaller ranks or sparsities. □

Remark 4.1 (Bounding the residual of the sparse component)**.** *Herein we derive inequality in* (4.42) *as was done in* (4.32) *to* (4.35) *for the low-rank component of the error.*

Proof. Let $\Delta := \Delta_{4r,3s,\mu}$ be an RIC with squared norms for $\text{LS}_{m,n}(4r, 3s, \mu)$ and $\gamma := \gamma_{4r,3r,\mu}$ be the rank-sparsity correlation coefficient defined in Lemma 2.1 on page 17. Then let R_0^L, R_1^L and R_0^S, R_1^S be defined as above equation (4.18)

and (4.22) respectively

$$(1-\Delta)\|R_0^S + R_1^S\|_F^2 \leq \|\mathcal{A}(R_0^S + R_1^S)\|_2^2 = \left|\left\langle \mathcal{A}(R_0^S + R_1^S),\, \mathcal{A}(R_0^S + R_1^S - R + R)\right\rangle\right| \tag{4.128}$$

$$= \left|\left\langle \mathcal{A}(R_0^S + R_1^S),\, \mathcal{A}(R_0^S + R_1^S - R)\right\rangle + \left\langle \mathcal{A}(R_0^S + R_1^S),\, \mathcal{A}(R)\right\rangle\right| \tag{4.129}$$

$$\leq \left|\left\langle \mathcal{A}\left(R_0^S + R_1^S\right),\, \mathcal{A}\left(-R_0^L - R_1^L - \sum_{j\geq 2} R_j\right)\right\rangle\right| + \left|\left\langle \mathcal{A}(R_0^S + R_1^S),\, \mathcal{A}(R)\right\rangle\right| \tag{4.130}$$

$$\leq \left(\Delta + \frac{2\gamma}{1-\gamma^2}\right)\|R_0^S + R_1^S\|_F \left(\|R_0^L + R_1^L\|_F + \sum_{j\geq 2}\|R_j\|_F\right) + \left\|\mathcal{A}\left(R_0^S + R_1^S\right)\right\|_2 \|\mathcal{A}\left(R\right)\|_2, \tag{4.131}$$

where the inequality in the first line comes from $R_0^S + R_1^S \in \mathrm{LS}_{m,n}(0, 2s, 0) \subset \mathrm{LS}_{m,n}(4r, 3s, \mu)$ satisfying the RIC, the second line is the consequence of feasibility in (4.31), the third line comes from Lemma 4.3 and by sums of individual pairs in the inner product being in $\mathrm{LS}_{m,n}(4r, 3s, \mu)$ by Lemma 2.9.

The first term in (4.131) can be bounded as

$$\left(\Delta + \frac{2\gamma}{1-\gamma^2}\right)\|R_0^S + R_1^S\|_F \left(\|R_0^L + R_1^L\|_F + \sum_{j\geq 2}\|R_j\|_F\right) \leq \left(\Delta + \frac{2\gamma}{1-\gamma^2}\right)\|R_0^S + R_1^S\|_F \left(\|R_0^L + R_1^L\|_F + \sqrt{2}\|R_0^L\|_F + \|R_0^S\|_F\right) \tag{4.132}$$

$$\leq \left(\Delta + \frac{2\gamma}{1-\gamma^2}\right)\|R_0^S + R_1^S\|_F \left((1+\sqrt{2})\|R_0^L + R_1^L\|_F + \|R_0^S + R_1^S\|_F\right) \tag{4.133}$$

where the first inequality comes as a consequence of optimality in (4.30) with $M_r = r$ and $M_s = s$, and the second inequality comes from $\left\|R_0^L\right\|_F \leq \left\|R_0^L + R_1^L\right\|_F$ and $\left\|R_1^L\right\|_F \leq \left\|R_0^L + R_1^L\right\|_F$. Having bounded the first term in (4.131) we now move to upper bounding the second term in (4.131) which comes as a consequence of feasibility bound in (4.31) and of the RIC for $R_0^L + R_1^L \in \mathrm{LS}_{m,n}(4r, 3s, \mu)$

$$\left\|\mathcal{A}\left(R_0^S + R_1^S\right)\right\|_2 \|\mathcal{A}\left(R\right)\|_2 \leq \varepsilon_b(1+\Delta)\left\|R_0^S + R_1^S\right\|_F. \tag{4.134}$$

We now bound (4.131) by combining the upper bounds of its constituents in (4.133) and (4.134)

$$(1-\Delta)\left\|R_0^S + R_1^S\right\|_F^2 \leq \left(\Delta + \frac{2\gamma}{1-\gamma^2}\right)\left\|R_0^S + R_1^S\right\|_F \left((1+\sqrt{2})\left\|R_0^L + R_1^L\right\|_F + \left\|R_0^S + R_1^S\right\|_F + \varepsilon_b\frac{1+\Delta}{\Delta}\right), \tag{4.135}$$

which after dividing both sides by $(1-\Delta)\left\|R_0^S + R_1^S\right\|_F$ yields the inequality in (4.42). □

Lemma 4.4. *Let X^j, X^{j+1}, X_0 be any matrices in the set $\mathrm{LS}_{m,n}(r,s,\mu)$ with $\mu < \sqrt{mn} \,/\, \left(3r\sqrt{3s}\right)$, $\alpha_j \geq 0$, and $\mathcal{A}(\cdot)$ be an operator whose RICs are sufficiently upper bounded, then the following two inequalities hold*

$$\begin{aligned}\langle X^j - X_0, X^{j+1} - X_0\rangle - \alpha_j \langle \mathcal{A}(X^j - X_0), \mathcal{A}(X^{j+1} - X_0)\rangle \\ \leq \|I - \alpha_j A_Q^T A_Q\|_2 \|X^j - X_0\|_F \|X^{j+1} - X_0\|_F,\end{aligned} \tag{4.136}$$

and

$$\|X^j - X_0 - \alpha_j \mathcal{A}^*\left(\mathcal{A}\left(X^j - X_0\right)\right)\|_F \leq \|I - \alpha_j A_Q^T A_Q\|_2 \|X^j - X_0\|_F, \tag{4.137}$$

where the spectrum of the matrix $\left(I - \alpha_j A_Q^T A_Q\right) \in \mathbb{R}^{mn\times mn}$ is bounded as

$$1 - \alpha_j\left(1 + \Delta_{3r,3s,\mu}\right) \leq \lambda\left(I - A_Q^T A_Q\right) \leq 1 + \alpha_j\left(1 - \Delta_{3r,3s,\mu}\right), \tag{4.138}$$

which gives an upper bound on the norm $\|I - \alpha_j A_Q^T A_Q\|_2 \leq \left|1 - \alpha_j\left(1 + \Delta_{3r,3s,\mu}\right)\right|$ as the lower bound in (4.138) is larger then the upper bound.

Proof. We vectorize the matrices on the left hand side of (4.136) using a mapping $\mathrm{vec}(\cdot) : \mathbb{R}^{m\times n} \to \mathbb{R}^{mn}$ that stacks columns of a given matrix into a vector and a mapping $\mathrm{mat}(\cdot)$ from the space of linear operators $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ to the space of matrices of size $p \times mn$

$$\begin{aligned} x_0 = \mathrm{vec}\left(X_0\right),\ x^j = \mathrm{vec}\left(X^j\right),\ x^{j+1} = \mathrm{vec}\left(X^{j+1}\right) \in \mathbb{R}^{mn} \\ A = \mathrm{mat}\left(\mathcal{A}\right) = \begin{bmatrix} — & \mathrm{vec}\left(A_1\right)^T & — \\ & \vdots & \\ — & \mathrm{vec}\left(A_p\right)^T & — \end{bmatrix} \in \mathbb{R}^{p\times mn}. \end{aligned} \tag{4.139}$$

Let $X_0 = U^0\Sigma^0 V^0 + S^0$, $X^j = U^j\Sigma^j V^j + S^j$, $X^{j+1} = U^{j+1}\Sigma^{j+1}V^{j+1} + S^{j+1}$ be the singular value decompositions where the matrices of the left singular vectors are $U^j \in \mathbb{R}^{m\times r}$ and their sparse components are supported at indices $\Omega^j = \mathrm{supp}\left(S^j\right)$. Consider the union of the index sets $\Omega := \left\{\Omega^0, \Omega^j, \Omega^{j+1}\right\}$ and construct the following frame

$$Q = \begin{bmatrix} I_n \otimes U & E \end{bmatrix} = \left[\begin{array}{cccc|ccc} U & 0_{n,3r} & \dots & 0_{n,3r} & & & \\ 0_{n,3r} & U & \dots & 0_{n,3r} & & & \\ \vdots & & \ddots & & e_{\Omega_1} & \dots & e_{\Omega_{3s}} \\ 0_{n,3r} & \dots & \dots & U & & & \end{array}\right] \in \mathbb{R}^{mn\times 3(nr+s)}, \tag{4.140}$$

where $U \in \mathbb{R}^{m\times 3r}$ is formed by concatenating U^0, U^j, U^{j+1} and e_{Ω_i} is a vector corresponding to a vectorized matrix with a single entry 1 at the index Ω_i. Note that $P_Q = Q\left(Q^TQ\right)^{-1}Q^T$ is an orthogonal projection matrix on the low-rank plus sparse subspace defined by the matrix U and the index set Ω. Note

that by Q being formed by the low-rank plus sparse bases of X_0, X^j, X^{j+1} we have that the projection does not change the vectorized matrices

$$P_Q x_0 = x_0, \quad P_Q x^j = x^j, \quad P_Q x^{j+1} = x^{j+1}. \tag{4.141}$$

To establish the bound in (4.136) we write the left hand side in its vectorized form

$$\left(x^j - x_0\right)^T \left(x^{j+1} - x_0\right) - \alpha_j \left(A(x^j - x_0)\right)^T \left(A(x^{j+1} - x_0)\right), \tag{4.142}$$

and replacing A with $A_Q = AP_Q$ in (4.142) using the identities in (4.141) simplifies the term as

$$\begin{aligned}
&\left(x^j - x_0\right)^T \left(x^{j+1} - x_0\right) - \alpha_j \left(A_Q(x^j - x_0)\right)^T \left(A_Q(x^{j+1} - x_0)\right) && (4.143)\\
&\quad = \left(x^j - x_0\right)^T \left((x^{j+1} - x_0) - \alpha_j A_Q^* A_Q (x^{j+1} - x_0)\right) && (4.144)\\
&\quad = \left(x^j - x_0\right)^T \left((I - \alpha_j A_Q^* A_Q)(x^{j+1} - x_0)\right) && (4.145)\\
&\quad \leq \|I - \alpha_j A_Q^* A_Q\|_2 \, \|x^j - x_0\|_2 \, \|x^{j+1} - x_0\|_2 && (4.146)\\
&\quad = \|I - \alpha_j A_Q^* A_Q\|_2 \, \|X^j - X_0\|_F \, \|X^{j+1} - X_0\|_F, && (4.147)
\end{aligned}$$

where $\|I - \alpha_j A_Q^* A_Q\|_2$ is the ℓ_2 operator norm of an $mn \times mn$ matrix.

Similarly we now establish the bound in (4.137)

$$\begin{aligned}
\left\|X^j - X_0 - \alpha_j \mathcal{A}^* \left(\mathcal{A}\left(X^j - X_0\right)\right)\right\|_F &= \left\|x^j - x_0 + \alpha_j A^T A \left(x_0 - x^j\right)\right\|_2 && (4.148)\\
&= \left\|\left(I - \alpha_j A^T A\right)\left(x^j - x_0\right)\right\|_2 && (4.149)\\
&\leq \left\|I - \alpha_j A_Q^* A_Q\right\|_2 \left\|X^j - X_0\right\|_F, && (4.150)
\end{aligned}$$

where we just vectorized the matrices and the linear operator $\mathcal{A}(\cdot)$ and upper bounded the expression using ℓ_2-operator norm $\|I - \alpha_j A_Q^* A_Q\|_2$. Matrix A_Q acts on a subspace of $\mathrm{LS}_{m,n}(3r, 3s, \mu)$ and is self-adjoint, as such its eigenvalues can be bounded using the RICs as done by Tanner and Wei (2013) and by Blanchard et al. (2015)

$$1 - \alpha_j \left(1 + \Delta_{3r,3s,2\mu}\right) \leq \lambda \left(I - A_Q^* A_Q\right) \leq 1 + \alpha_j \left(1 - \Delta_{3r,3s,2\mu}\right). \tag{4.151}$$

□

5 LOW-RANK MODELS FOR MULTISPECTRAL IMAGING

SYNOPSIS

In this chapter, we apply low-rank matrix completion and compressed sensing in the context of the reconstruction of multispectral imagery. Snapshot mosaic multispectral imagery acquires an undersampled data cube by acquiring a single spectral measurement per spatial pixel. Sensors which acquire p frequencies, therefore, suffer from severe $1/p$ undersampling of the full data cube. We show that the missing entries can be accurately imputed using non-convex techniques from sparse approximation and matrix completion initialised with traditional demosaicing algorithms. In particular, we observe the peak signal-to-noise ratio can typically be improved by 2 dB to 5 dB over current state-of-the-art methods when simulating a $p = 16$ mosaic sensor measuring both high and low altitude urban and rural scenes as well as ground-based scenes.

5.1 INTRODUCTION

Multispectral imaging is the process of recording 2D arrays of information at multiple spectra – light frequencies. Having access to such a rich, three-dimensional data cube allows different materials to be distinguished due to their differing spectral emission profiles. As a result, multispectral imaging is used in applications ranging from landmine detection, precision agriculture, and medical diagnosis to name but a few of its application domains; for a partial survey see the January 2014 special issue of IEEE Signal Processing Magazine (Ma et al., 2014). The increased sensor size and acquisition time are some of the central obstacles to the more widespread use of multispectral imagery.

Snapshot mosaic multispectral sensors allow for a compact video-rate multispectral imaging by acquiring only a fraction of the multispectral cube. For example, the IMEC SNm4x4 records 16 bands at a rate of 340 frames per second on a spatial two-dimensional 2048×1088 pixel domain by only acquiring a single spectrum per pixel; specifically this is achieved by tiling the two-dimensional domain by 4×4 pixel supercells where each supercell acquires the spectra independently. This chapter illustrates the architecture through the IMEC sensor, but note there are numerous similar sensors such

as the S137 system by CUBERT. The undersampled snapshot data cube can either then be viewed directly as 16 separate 512×272 pixel arrays, or more typically the missing multispectral data cube values can be interpolated to give approximate images on the full 2048×1088 pixel spatial domain.

Herein we demonstrate the efficacy of multiple methods for interpolating the missing values in the above snapshot mosaic data cube by simulating the undersampling from complete three-dimensional data cubes provided by DSTL as well as AVIRIS (Vane et al., 1993), Stanford SCIEN (Skauli and Farrell, 2013), Nascimento (Nascimento et al., 2002), Foster (Foster et al., 2004), IEEE GRSS Data Fusion Contest (Le Saux et al., 2018).

The rest of the chapter is organised as follows. In addition to reviewing the existing state-of-the-art interpolation methods in §5.2, we demonstrate sparse approximation and matrix completion methods in §5.3 and 5.4 respectively, which we observe to substantially outperform the prior state-of-the-art. Over the above diverse data sets, we observe that non-convex compressed sensing and matrix completion methods initialised with traditional interpolations methods typically improve the peak signal-to-noise ratio by 2 dB to 5 dB, see Table 5.1.

Demosaicing is the process by which the undersampled three-dimensional multispectral data cube has the missing entries approximated so as to simulate a full data acquisition. While most three-dimensional interpolations methods would be directly applicable, we consider a few methods previously used by the multispectral community, such as direct interpolation as described in §5.2 as well as sparse approximation regularisation methods in §5.3 and low-rank structure as presented in §5.4.

5.2 DIRECT INTERPOLATION METHODS

Brauers and Aach (2006) developed methods to estimate the missing values in the multispectral cube based on extending a spatial bilinear interpolation of the measured values to include any spectral correlation. The weighted bilinear interpolation (WB) for the 4×4 pixel regular mosaic filter follows by padding the missing entries with zeros and convolving with the cartesian product of a discrete seven-pixel width filter $\frac{1}{4}[1\ 2\ 3\ 4\ 3\ 2\ 1]$. Then, the spectral correlation is included in the spectral difference (SD) method by a) taking the output of WB to independently compute, for each band, say k, the difference between the values of the measured pixels for spectrum k and the WB interpolated values of every other band restricted to the support of the measured pixels of spectrum k, then b) applying WB to these spectral differences c) to form an approximation of the full spectrum k by adding the output of step (b) to the difference with l at the location of the measured pixels for spectrum l.

Mihoubi et al. (2015) extended the SD approach to consider alternative ways to build correlations between the bands. In intensity difference (ID)

they build spectral correlation by first constructing a spatial intensity map whose value at a pixel is the measurement for whichever spectra was measured at that spatial location, then this intensity map is averaged using a weighting based on the number of pixels per spectra contained in the averaging width. See §5.5.2 for details on the choice of averaging used here and Mihoubi et al. (2015) for a more general discussion. Hence, the difference between this averaged intensity map and the measurements for each spectrum is computed, the unknown values zero-padded, and each band averaged such as in WB.

Interpolation methods designed in transform domains have been considered by Miao et al. (2006) in the binary tree-based edge-sensing (BTES) method, which has the additional benefit of allowing for variable sampling densities per frequency band. However, we observe it is inferior to SD and ID described above in the setting of snapshot imaging. Pseudo-panchromatic image difference (PPID) (Mihoubi et al., 2017) builds upon BTES and ID. However, due to the applicability of PPID to only some specific mosaic arrangements we leave comparison with our algorithms for a later time.

5.3 SPARSE APPROXIMATION INPAINTING

Sparse approximation inpainting allows one to easily consider the interpolation of the under-complete snapshot data cube in transforms more general than the linear interpolation of (WB). In particular, one can assume that the image is well approximated by a sparse representation in a suitable image domain and exploit this structure to reconstruct it from undersampled measurements (King et al., 2013; Dong et al., 2012; Elad et al., 2005), e.g. by solving

$$\min_x \left\| y - P_\Omega \Psi^{-1} x \right\|_2 , \quad \text{s.t. } \|x\|_0 \leq k , \tag{5.1}$$

where Ψ represents the transform in which the data is known to be compressible and y is the full data cube projected by P_Ω to the undersampled locations. Degraux et al. (2015) apply this model to a reconstruction of multispectral imagery acquired by mosaic snapshot cameras.

The primary challenge lies in two aspects: (i) the significant subsampling ratio of $1/K$, where K is the number of spectral bands, and (ii) the selection of the suitable transform Ψ. The first problem can be overcome by initialising the state-of-the-art sparse approximation algorithms for solving (5.1) with sufficiently accurate initial estimates, such as those from the classical interpolation methods described in §5.2. As it was pointed out in (Degraux et al., 2015), we find that, for natural scenes captured by snapshot multispectral imaging, a Kronecker product of 2D wavelet transform spatially and the discrete cosine transform for the spectral dimension is an effective choice for the representation Ψ. In particular, the 2D wavelet transform spatially includes elements of nearly global support to allow broad correlations as

well as local elements to express fine detail and the discrete cosine transform models the slowly varying values in the spectral dimension.

5.4 LOW-RANK MATRIX COMPLETION INPAINTING

Rather than using local correlations, matrix (and tensor) completion exploit the correlation in the data cube through a low-rank structure, e.g. by solving

$$\min_X \left\| y - P_\Omega X \right\|_2 , \quad \text{s.t.} \quad \operatorname{rank}(X) \leq r, \tag{5.2}$$

where y is the observed data, X is a matrix corresponding to an unfolding of the complete three-dimensional data cube, and P_Ω is a restriction to the measured values as before.

Although the low-rank matrix completion problem is NP-hard in general (see a recent survey by Davenport and Romberg (2016)) there is a number of computationally fast solvers for the problem with provable convergence guarantees (Wen et al., 2012; Tanner and Wei, 2013; Blanchard et al., 2015). In fact, matrix completion has been previously applied to the reconstruction of subsampled multispectral imagery (Gelvez et al., 2015) by the Coded Aperture Snapshot Spectral Imager (CASSI) (Gehm et al., 2007). Here we show that matrix completion can be used also in the case of a more severe subsampling by snapshot mosaic camera designs if provided with suitable initialisation.

While there are many non-convex methods for matrix completion, we showcase two exemplary cases but expect that other non-convex methods would perform similarly well. We apply conjugate gradient iterative hard thresholding (CGIHT) (Blanchard et al., 2015) and alternating steepest descent (ASD) (Tanner and Wei, 2016) to solve (5.2), providing them with an initial guess from either SD or ID. We show that both CGIHT and ASD can improve on the classical interpolation methods. This differs substantially from prior work both in terms of using more recent algorithms for matrix completion which have been shown to be more effective and initialising them with prior state-of-the-art interpolation methods, and moreover in that unlike (Gelvez et al., 2015) which treat each spectral band separately with 30% undersampling, we vectorise the spatial dimensions to create a matrix of size $262,144$ by 16 with $1/16$ undersampling. We observe that this unfolding which allows full correlation between the spectral information is particularly effective, often resulting in reconstructions which are visually indistinguishable from fully acquired data.

5.5 NUMERICAL SIMULATIONS

In this section, we show and explain the numerical results obtained by applying the methods discussed above.

5.5.1 Data sets

We consider the efficacy of the algorithms for demosaicing on the following data sets:

- *High altitude airborne images* from the AVIRIS (Vane et al., 1993) and 2018 IEEE GRSS Data Fusion contest (Le Saux et al., 2018). AVIRIS line-scan captures 224 spectral bands in the 380 nm to 2500 nm and the GRSS images have 48 spectral bands in the range of 380 nm to 1050 nm.
- *Low altitude airborne images* acquired at DSTL Porton Down, in August 2014, from which we selected 10 representative radiance images of fields from a HySpex VNIR-1600 line-scan camera in the range 400 nm to 1000 nm.
- *Ground-based images* from the Stanford SCIEN (Skauli and Farrell, 2013), Nascimento (Nascimento et al., 2002) and Foster (Foster et al., 2004). The Stanford SCIEN images come from the line-scan HySpex VNIR-1600 camera.

We processed these data sets to simulate the spectrum measures by the IMEC SNm4x4 snapshot sensor with access to the complete data cube. Then, we undersampled the data cube following the sensor sampling pattern and the following simulations conducted.

5.5.2 Simulation setup

We implement and test recovery by two interpolation methods ID and SD, two matrix completion methods ASD and CGIHT and a compressed sensing version of CGIHT with a sparsifying transform as a Kronecker product of 2D Daubechies wavelets (W2) in the spatial and 1D Discrete Cosine Transform (DCT) in the spectral domain which we reference as W2×DCT.

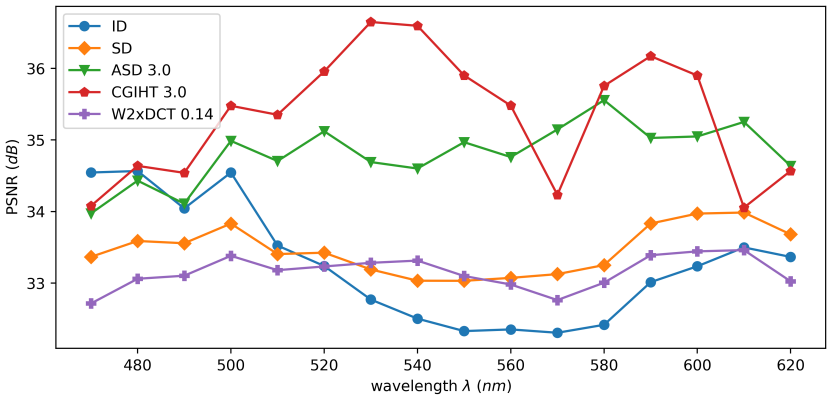
The iterative algorithms are terminated once the error in iteration t is $\left\|P_\Omega X^{(t)} - y\right\|_2 / \left\|y\right\|_2 \leq 10^{-7}$ or at the 500^{th} iteration.

We report the quality of an image approximated by demosaicing by the peak signal-to-noise ratio (PSNR), defined as the log of the ratio between the maximum possible power of (the slide of) an image and the power of corrupting noise that affects the fidelity of its representation, computed in terms of the average squared difference (or mean squared error, MSE) between the reference image and its reconstruction:

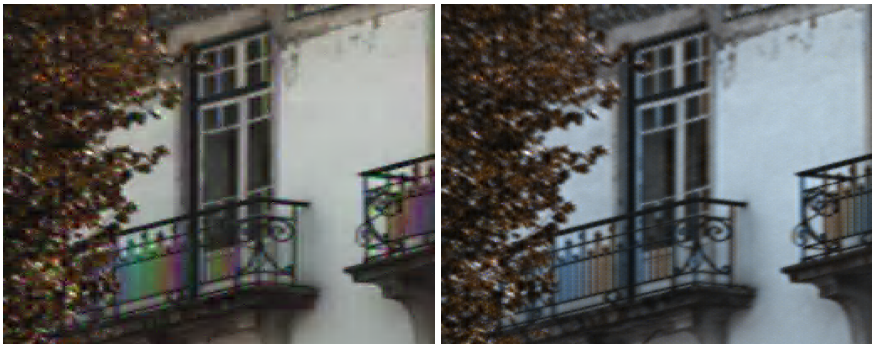
$$\mathrm{PSNR}_k = 10 \log_{10} \left(\frac{(\max_{p \in \mathcal{P}} I_p^k)^2}{\frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} (I_p^k - \hat{I}_p^k)^2} \right), \tag{5.3}$$

where I^k and $\hat{I}^k$ are the k-th band slices of the reference cube and the reconstruction, respectively, and $\mathcal{P}$ denotes the set of all pixels.

We also employ the structural similarity (SSIM) index (Wang et al., 2004), which is a decimal value between −1 and 1, with 1 being reachable only in the case of two identical sets of data. SSIM is a perception-based model that



(a) PSNR throughout spectral bands.



(b) W2×DCT [33.2 dB].

(c) CGIHT [35.3 dB].

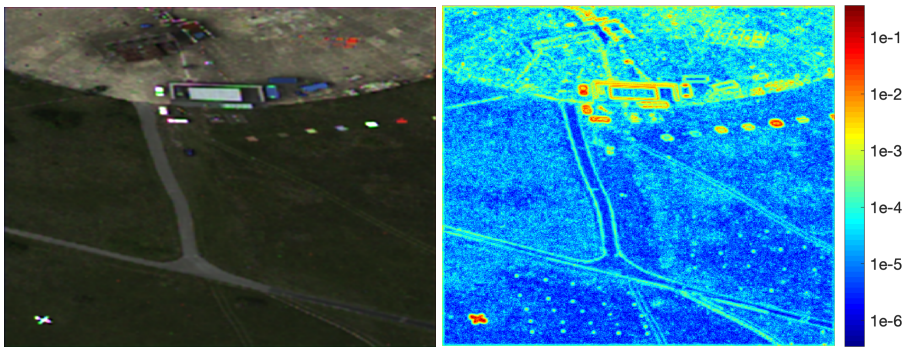
Figure 5.1: Results for reconstruction of `N06`. W2×DCT (CS) and CGIHT (MC) initialised from SD. The sparsity ratio is set to $\rho = 0.14$ for W2×DCT (CS) and the rank is set to $r = 3$ for ASD and CGIHT (MC). Wavelet based CS method smooths out the image while CGIHT MC is able to better preserve sharp edges.

considers image degradation as a perceived change in structural information, while also incorporating important perceptual phenomena, including both luminance masking and contrast masking terms.

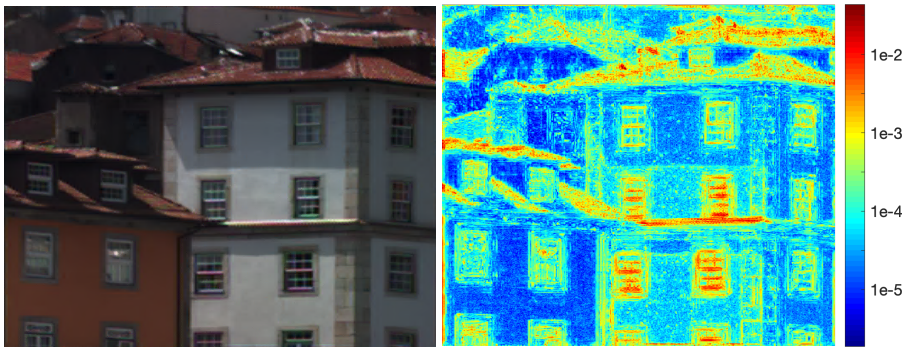
5.5.3 *Results*

Figure 5.1a shows the PSNR of each spectrum given its band centre, for a sample image from Nascimento et al. (2002), for the classical interpolation methods SD and ID as well as the compressed sensing (CS) and matrix completion (MC) techniques initialised with SD. For CS and MC reconstructions, we observe that the sparsity ratio around 0.14, respectively the rank of 3, consistently achieves the best recovery performance in terms of PSNR. Notice that, except for the first band, the matrix completion algorithms outperform SD and ID. On the other hand, the compressed sensing approach does not improve on the interpolation methods. In particular, note the overall incorrect contrast level resulting in yellowing of Figure 5.1b. Moreover, we lose the sharpness of the edges in the balcony when employing CS W2×DCT (Figure 5.1b), while we recover it with the matrix completion variant of CGIHT (Figure 5.1c).

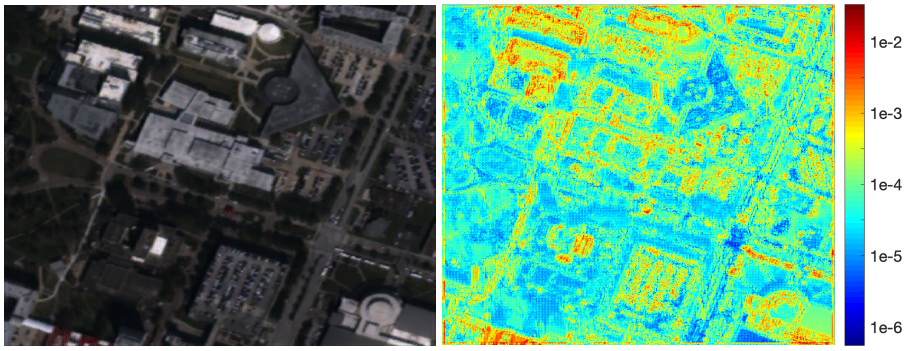
To further emphasise how the different algorithms differ from each other we show the reconstructions PSNR from the corresponding reference images



(a) `D2301`, ID [37.9 dB PSNR and SSIM of 0.99992].



(b) `F06`, W2×DCT from SD [39.3 dB PSNR and SSIM of 0.99955].



(c) `GD`, CGIHT from SD [36.2 dB PSNR and SSIM of 0.99988].

Figure 5.2: Left: Colour renderings of image reconstructions. Right: MSE of reconstructions (log-scale).

in Figure 5.2. Note in Figure 5.2c how ID accurately reconstructs the field image, taking into account the spectral correlation between the bands, but smooths the sharp details. The compressed sensing approach does a better job in identifying the edges (Figure 5.2b), but suffers from the same problem overall. On the other hand, the matrix completion CGIHT outperforms the other methods due to the presence of field-like uniformities.

Finally, as shown in Table 5.1, in the majority of cases the best-performing algorithms are the matrix completion CGIHT and ASD initialised with SD, with improvements over both SD and ID from 2 dB to 5 dB. `StCh` from Stanford SCIEN (Skauli and Farrell, 2013) seems to be the only outlier, with an improvement of just 0.2 dB.

Being directly related to the number of bands, the rank of the spectral unfolding seems to be effective in capturing the spectral information of

IMAGE	INIT		ASD	CGIHT	W2×DCT
D0201	ID	34.5 ± 0.8	36.3 ± 1.1	37.9 ± 1.1	34.3 ± 0.5
	SD	35.9 ± 0.7	38.1 ± 0.9	**39.5 ± 1.2**	35.8 ± 0.6
D0301	ID	36.0 ± 1.8	38.3 ± 2.2	39.2 ± 1.7	35.1 ± 1.1
	SD	37.3 ± 1.3	39.7 ± 1.4	**40.5 ± 1.5**	36.3 ± 1.0
D0303	ID	39.1 ± 1.0	41.2 ± 1.2	43.1 ± 1.4	38.8 ± 0.7
	SD	40.8 ± 0.8	43.4 ± 1.1	**45.5 ± 1.3**	40.5 ± 0.7
D0307	ID	34.1 ± 1.8	36.1 ± 2.3	36.9 ± 2.0	33.2 ± 1.1
	SD	35.2 ± 1.1	37.2 ± 1.2	**37.8 ± 1.5**	34.3 ± 0.8
D0308	ID	37.9 ± 0.5	39.7 ± 0.9	41.5 ± 1.1	37.7 ± 0.4
	SD	39.5 ± 0.5	41.7 ± 1.1	**43.2 ± 1.2**	39.5 ± 0.5
D2301	ID	37.9 ± 0.3	39.4 ± 0.7	41.1 ± 1.3	37.8 ± 0.3
	SD	39.5 ± 0.3	41.5 ± 1.0	**43.6 ± 1.5**	39.7 ± 0.4
AvLF	ID	33.7 ± 0.7	35.6 ± 1.2	36.7 ± 2.3	33.5 ± 0.7
	SD	35.3 ± 0.7	37.8 ± 1.6	**39.3 ± 2.6**	34.7 ± 0.6
StCh	ID	36.2 ± 1.0	36.7 ± 1.7	36.6 ± 1.8	36.0 ± 1.0
	SD	37.3 ± 1.1	**37.5 ± 1.0**	37.2 ± 0.9	37.4 ± 1.0
N04	ID	37.3 ± 3.1	**38.6 ± 2.6**	38.2 ± 2.8	35.3 ± 1.9
	SD	35.6 ± 0.8	36.9 ± 0.8	35.4 ± 2.2	34.9 ± 0.9
N06	ID	33.3 ± 0.8	34.5 ± 0.6	35.2 ± 0.5	32.6 ± 0.4
	SD	33.5 ± 0.3	34.8 ± 0.4	**35.3 ± 0.8**	33.2 ± 0.2
N08	ID	33.5 ± 0.4	34.4 ± 0.6	**35.6 ± 1.0**	32.7 ± 0.2
	SD	33.2 ± 0.4	34.4 ± 0.7	35.5 ± 1.3	32.9 ± 0.3
F05	ID	36.1 ± 2.4	**36.6 ± 1.9**	36.5 ± 1.9	34.0 ± 1.4
	SD	35.0 ± 0.4	36.1 ± 0.8	35.7 ± 1.6	34.1 ± 0.7
F06	ID	40.0 ± 0.9	**40.2 ± 0.8**	39.9 ± 0.9	38.6 ± 0.7
	SD	38.9 ± 0.4	39.6 ± 0.9	39.3 ± 0.9	39.1 ± 0.5
F07	ID	34.9 ± 1.2	**36.3 ± 1.5**	36.1 ± 1.6	34.4 ± 0.9
	SD	34.6 ± 0.8	36.3 ± 1.2	35.1 ± 1.8	34.6 ± 0.8
GB	ID	34.3 ± 1.1	36.0 ± 0.9	36.7 ± 1.1	33.3 ± 0.5
	SD	34.9 ± 0.3	37.2 ± 0.9	**37.5 ± 1.5**	34.7 ± 0.7
GD	ID	34.1 ± 1.4	35.8 ± 1.1	36.2 ± 1.2	33.0 ± 0.5
	SD	34.6 ± 0.4	**36.9 ± 0.9**	36.6 ± 1.1	34.3 ± 0.7
GP	ID	37.9 ± 1.2	39.7 ± 0.9	40.4 ± 1.8	36.9 ± 0.5
	SD	38.6 ± 0.4	40.9 ± 0.9	**41.5 ± 1.2**	38.4 ± 0.7
GR	ID	35.3 ± 1.2	37.0 ± 1.0	37.3 ± 1.2	34.4 ± 0.5
	SD	36.0 ± 0.4	38.0 ± 0.9	**38.1 ± 1.7**	35.7 ± 0.6

Table 5.1: Average PSNR over the 16 bands, with standard deviations. The best results for each image are highlighted in bold.

the analysed datasets. Our results suggest a high correlation between the frequency bands and a low-rank structure of the spectral unfolding of our images, in which most of the information is contained in the first 3 singular values of the spectral unfolding.

Related to the choice of rank is the problem of unmixing of spectral end-members using non-negative matrix factorization (NMF). Here the task is to approximate the matrix that comes from the spectral unfolding of a multispectral image, as a product of two rank-r matrices that are also non-negative. The NMF solution can be interpreted as the set of r end-members in the left matrix and the set of r abundance factors in the right

matrix (Parente and Plaza, 2010). However, in the case of low-rank matrix completion, we do not impose the non-negative constraint, and therefore our solution does not have the same interpretative power.

5.6 SUMMARY AND DISCUSSION

We provide a numerical comparison of multispectral demosaicing by traditional interpolation, sparse approximation and matrix completion methods. Our experiments demonstrate that non-convex matrix completion typically improves reconstruction by 2 dB to 5 dB over the current state-of-the-art methods. This differs substantially from prior work in terms of employing matrix completion on the spectral unfolding of the image in the context of demosaicing, initialising it with classical interpolation methods and using more recent non-convex matrix completion algorithms.

6 SUMMARY & FINAL REMARKS

This thesis develops theory and algorithms for the recovery of low-rank plus sparse matrices from subsampled measurements. In particular, we studied: (i) the well-posedness of optimisation problems over the sets of low-rank plus sparse matrices, (ii) random linear maps whose restricted isometry constants are bounded when acting on the set of low-rank plus sparse matrices, and (iii) algorithms that provably recover low-rank plus sparse matrices from subsampled measurements.

6.1 SUMMARY OF MAIN RESULTS

Our central findings are summarised in the following key areas.

6.1.1 Low-rank plus sparse matrix sets are not closed

Chapter 2: Matrix rigidity and the ill-posedness in matrix recovery

We studied the low-rank plus sparse matrix set $\mathrm{LS}_{m,n}(r,s)$ and the well-posedness of optimisation problems defined over it. We found a curious result in Theorem 2.1 on page 20, which states the set $\mathrm{LS}_{n,n}(r,s)$ is not closed for a range of ranks r and sparsities s satisfying

$$n \geq (r+1)(s+2) \quad \text{or} \quad n \geq (r+2)^{(3/2)}s^{1/2}. \tag{6.1}$$

As a consequence, the corresponding non-convex optimisation problems for low-rank matrix completion and Robust PCA can fail to have any solution. This result might come as a surprise, as in both cases, it has been assumed until now that a solution must exist, and the existing theory have instead focused on guarantees for the uniqueness of the solution. Moreover, we give specific constructions of simple matrices for which we numerically observe the ill-posedness difficulties when applied to state-of-the-art matrix completion and Robust PCA algorithms.

We close the set by restraining the Frobenius norm of the low-rank component, which results into the set of bounded low-rank plus sparse matrices $\mathrm{LS}^{\tau}_{m,n}(r,s)$ in Definition 1.2 that is closed by Lemma 2.8. However, the problem with $\mathrm{LS}^{\tau}_{m,n}(r,s)$ is that it is not possible to guarantee that the sum of two matrices belonging to the set is also a bounded low-rank plus sparse matrix.

The issue with the lack of additivity is overcome in Definition 1.3 of the set of incoherent low-rank plus sparse matrices $\mathrm{LS}_{m,n}(r,s,\mu)$, and which has the property that the sum two matrices in the set retains the incoherence, in other words: $\mathrm{LS}_{m,n}(2r,2s,\mu) = \mathrm{LS}_{m,n}(r,s,\mu) + \mathrm{LS}_{m,n}(r,s,\mu)$ by Lemma 2.9.

Additionally, for $\mu < \sqrt{mn}\,/\,\left(r\sqrt{s}\right)$ the set of incoherent low-rank plus sparse matrices is also a subset of $\mathrm{LS}^{\tau}_{m,n}(r,s)$ as shown in Lemma 2.10.

6.1.2 *RICs for Gaussian measurements and low-rank plus sparse matrices*

Chapter 3: Restricted isometry constants for low-rank plus sparse matrix sets

We developed the theory showing that random linear maps obeying concentration of measure inequalities act as approximate isometries when applied to the set of low-rank plus sparse matrices $\mathrm{LS}_{m,n}(r,s,\mu)$. Theorem 3.1 proves that the restricted isometry constants (RICs) of random linear maps captured in Definition 3.2 in respect to the set $\mathrm{LS}_{m,n}(r,s,\mu)$ are bounded for

$$p = O(r(m+n-r)+s)\log\left(\left(1-\mu^2\frac{r^2 s}{mn}\right)^{-1/2}\frac{mn}{s}\right), \tag{6.2}$$

provided $\mu < \frac{\sqrt{mn}}{r\sqrt{s}}$. This translates into the RICs being bounded independent of the problem size when p/mn, s/p, $r(m+m-r)/p$ and μ remain fixed.

6.1.3 *Methods for provable recovery of low-rank plus sparse matrices*

Chapter 4: Algorithms for low-rank plus sparse matrix sensing

We devised several computationally tractable methods for the recovery of low-rank plus sparse matrices from subsampled measurements.

Firstly, in Theorem 4.1, we proved that an upper bound on the RICs of the measurement operator implies uniqueness of the solution to the recovery problem.

By treating the low-rank and the sparse component individually, we were able to prove in Theorem 4.2 that semidefinite programming that solves the convex optimisation problem in (1.19) robustly recovers the subsampled matrix provided the RICs of the measurement operator are sufficiently small.

Moreover, we proposed two gradient descent algorithms, NIHT in Algorithm 1 and NAHT in Algorithm 2, for solving the non-convex optimisation in (1.18). While NIHT performs an oblique projection on the set $\mathrm{LS}_{m,n}(r,s,\mu)$, NAHT alternates between minimising the objective in the low-rank and in the sparse component. By proving Theorem 4.3 and Theorem 4.4, we show that both of these methods are guaranteed to converge to the subsampled matrix when the RICs of the measurement operator are sufficiently bounded.

The convex relaxation and NAHT also apply to Robust PCA, when the sensing operator is chosen to be the identity $\mathcal{A} = \mathrm{Id}$, which is an isometry with $\Delta = 0$, and achieve global linear convergence in the presence of the number of corruptions of the optimal order $s = O\left(1/\left(\mu^2 r^2\right)\right)$.

We performed numerical experiments illustrating these results. In §4.5, we observed a phase transition in the parameter space for synthetically generated problems. In §4.6, we gave an exemplary applications on dynamic-foreground/static-background separation and multispectral imaging.

6.1.4 *Applications to multispectral imaging*

Chapter 5: Low-rank models for multispectral imaging

We implemented a number of matrix completion and compressed sensing algorithms in the context of the reconstruction of multispectral imagery from snapshot mosaic filters. We are able to accurately recover the missing entries despite the severe 1/16 mosaic undersampling with the typical improvement in the peak signal-to-noise ratio by 2 dB to 5 dB across high and low altitude urban and rural scenes as well as ground-based scenes.

6.2 OPEN PROBLEMS AND FUTURE WORK

This thesis laid down the ground work for low-rank plus sparse matrix sensing opening up the pursuit of several research areas in the future.

6.2.1 *Expansion of the* $\mathrm{LS}_{n,n}(r,s)$ *non-closedness results*

The result of Theorem 2.1, stating the set $\mathrm{LS}_{n,n}(r,s)$ is not closed, can be extended in two ways.

Firstly, it is reasonable to assume that all $\mathrm{LS}_{n,n}(r,s)$ sets that do not simplify to the simple cases of: (i) the sets of sparse matrices $\mathrm{LS}_{n,n}(0,s)$, (ii) the sets of low-rank matrices $\mathrm{LS}_{n,n}(r,0)$, or (iii) the set of all matrices $\mathbb{R}^{n\times n} = \mathrm{LS}_{n,n}(r,(n-r)^2)$, are not closed. This proposition is formally stated in Conjecture 1 and left for future work.

Secondly, the significance of the analogous results of (de Silva and Lim, 2008) for the CP-rank of higher-order tensors, lies also in the fact that the set of the tensors, for which the non-closedness occurs, is of a positive measure. It would be curious to see if this is also the case for low-rank plus sparse matrices, i.e. that the set of matrices for which the non-closedness is an issue has a positive measure.

6.2.2 *Expansion to the sensing of other additive structures*

The proof technique used in Theorem 3.1 of the RICs for the combined additive structure of low-rank plus sparse matrix sets could be extended to additive structures of other sets. The only requirement is that the covering number of the ε-net of these sets is sufficiently upper bounded. To this end, it might be necessary to pose an additional constraint on the energy of one of the components, as done in Definition 1.1.

6.2.3 *Expansion of algorithms for low-rank plus sparse matrix sensing*

The algorithmic contributions presented in this work open up the possibility of other algorithms, developed in compressed sensing and low-rank matrix completion/sensing, to be extended to the case of low-rank plus sparse matrix recovery.

In particular, algorithms involving momentum (Kyrillidis and Cevher, 2014; Cevher, 2011), minimisation over increasingly larger subspaces (Blanchard et al., 2015), or that express the low-rank component in a factorised form (Wen et al., 2012; Haldar and Hernando, 2009; Tanner and Wei, 2016).

BIBLIOGRAPHY

Achlioptas, D. (2003). Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66(4):671–687.

Ailon, N. and Chazelle, B. (2009). The fast Johnson–Lindenstrauss transform and approximate nearest neighbors. *SIAM Journal on Computing*, 39(1):302–322.

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer-Verlag, Berlin/Heidelberg.

Antonucci, G. A., Vary, S., Humphreys, D., Lamb, R. A., Piper, J., and Tanner, J. (2019). Multispectral snapshot demosaicing via non-convex matrix completion. In *2019 IEEE Data Science Workshop (DSW)*, pages 227–231. IEEE.

Argyriou, A., Evgeniou, T., and Pontil, M. (2008). Convex multi-task feature learning. *Machine Learning*, 73(3):243–272.

Baete, S. H., Chen, J., Lin, Y.-C., Wang, X., Otazo, R., and Boada, F. E. (2018). Low rank plus sparse decomposition of ODFs for improved detection of group-level differences and variable correlations in white matter. *NeuroImage*, 174(February):138–152.

Baraniuk, R., Davenport, M., DeVore, R., and Wakin, M. (2008). A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263.

Battaglino, C., Ballard, G., and Kolda, T. G. (2018). A Practical Randomized CP Tensor Decomposition. *SIAM Journal on Matrix Analysis and Applications*, 39(2):876–901.

Beck, A. and Teboulle, M. (2009). A Fast Iterative Shrinkage-Thresholding Algorithm. *Society for Industrial and Applied Mathematics Journal on Imaging Sciences*, 2(1):183–202.

Bioucas-Dias, J. M. and Figueiredo, M. A. (2007). A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image Processing*, 16(12):2992–3004.

Blanchard, J. and Tanner, J. (2015). Performance comparisons of greedy algorithms in compressed sensing. *Numerical Linear Algebra with Applications*, 22(2):254–282.

Blanchard, J. D., Tanner, J., and Wei, K. (2015). CGIHT: Conjugate gradient iterative hard thresholding for compressed sensing and matrix completion. *Information and Inference*, page iav01.

Blumensath, T. and Davies, M. (2010). Normalized iterative hard thresholding: guaranteed stability and performance. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):298–309.

Blumensath, T. and Davies, M. E. (2009). Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274.

Borgwardt, K. H. (1987). *The Simplex Method: A Probabilistic Analysis*, volume 1 of *Algorithms and Combinatorics*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Bouwmans, T., Sobral, A., Javed, S., Jung, S. K., and Zahzah, E.-H. (2017). Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset. *Computer Science Review*, 23:1–71.

Brauers, J. and Aach, T. (2006). A Color Filter Array Based Multispectral Camera. *12 Workshop Farbbildverarbeitung*.

Burer, S. and Monteiro, R. D. (2003). A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357.

Cai, H., Cai, J.-F., and Wei, K. (2019). Accelerated alternating projections for robust principal component analysis. *Journal of Machine Learning Research*, 20(1):685—-717.

Cai, J.-F., Candès, E. J., and Shen, Z. (2010). A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982.

Candès, E., Romberg, J., and Tao, T. (2006a). Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509.

Candès, E. and Tao, T. (2005). Decoding by Linear Programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215.

Candès, E. J., Li, X., Ma, Y., and Wright, J. (2011). Robust principal component analysis? *Journal of the ACM*, 58(3):1–37.

Candès, E. J. and Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772.

Candès, E. J. and Romberg, J. (1995). L1-MAGIC: Recovery of Sparse Signals via Convex Programming. Technical report.

Candès, E. J., Romberg, J. K., and Tao, T. (2006b). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223.

Candès, E. J. and Tao, T. (2006). Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425.

Candès, E. J. and Tao, T. (2010). The power of convex relaxation: near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080.

Cao, X., Yue, T., Lin, X., Lin, S., Yuan, X., Dai, Q., Carin, L., and Brady, D. J. (2016). Computational snapshot multispectral cameras: toward dynamic capture of the spectral world. *IEEE Signal Processing Magazine*, 33(5):95–108.

Cevher, V. (2011). On accelerated hard thresholding methods for sparse approximation. In *Wavelets and Sparsity XIV*, volume 8138, page 813811.

Chandrasekaran, V., Sanghavi, S., Parrilo, P. A., and Willsky, A. S. (2011). Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596.

Chen, S. and Donoho, D. (1994). Basis pursuit. In *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 41–44. IEEE Comput. Soc. Press.

Chen, S. S., Donoho, D. L., and Saunders, M. A. (2001). Atomic decomposition by basis pursuit. *SIAM Review*, 43(1):129–159.

Chen, Y., Guo, Y., Wang, Y., Wang, D., Peng, C., and He, G. (2017). Denoising of hyperspectral images using nonconvex low rank matrix approximation. *IEEE Transactions on Geoscience and Remote Sensing*, 55(9):5366–5380.

Chen, Y. and Wainwright, M. J. (2015). Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees. pages 1–63.

Codenotti, B. (2000). Matrix rigidity. *Linear Algebra and its Applications*, 304(1-3):181–192.

Combettes, P. L. and Wajs, V. R. (2005). Signal Recovery by Proximal Forward-Backward Splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200.

Dantzig, G. (1963). *Linear Programming and Extensions*, volume 53. RAND Corporation.

Dantzig, G. B. and Thapa, M. N. (1998). Linear Programming-1: Introduction. *The Journal of the Operational Research Society*, 49(11):1226.

Dasgupta, S. and Gupta, A. (2003). An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Structures and Algorithms*, 22(1):60–65.

Daubechies, I., Defrise, M., and De Mol, C. (2004). An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457.

Davenport, M. A. and Romberg, J. (2016). An Overview of Low-Rank Matrix Recovery From Incomplete Observations. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):608–622.

De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000). On the best rank-1 and rank-(R1, R2, ..., RN) approximation of higher-order tensors. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1324–1342.

de Silva, V. and Lim, L.-h. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1084–1127.

Degraux, K., Cambareri, V., Jacques, L., Geelen, B., Blanch, C., and Lafruit, G. (2015). Generalized inpainting method for hyperspectral image acquisition. In *2015 IEEE International Conference on Image Processing (ICIP)*, volume 2015-Decem, pages 315–319. IEEE.

Dhillon, P. S., Lu, Y., Foster, D., and Ungar, L. (2013). New subsampling algorithms for fast least squares regression. *Advances in Neural Information Processing Systems*, pages 1–9.

Dimitris, M., Marden, D., and Shaw A., G. (2003). Hyperspectral image processing for automatic target detection applications. *Lincoln Laboratory Journal*, 14(1):79 —- 116.

Dong, B., Ji, H., Li, J., Shen, Z., and Xu, Y. (2012). Wavelet frame based blind image inpainting. *Applied and Computational Harmonic Analysis*, 32(2):268–279.

Donoho, D. (2006a). Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306.

Donoho, D. and Tanner, J. (2009a). Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1906):4273–4293.

Donoho, D. L. (2006b). For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59(6):797–829.

Donoho, D. L. (2006c). High-Dimensional Centrally Symmetric Polytopes with Neighborliness Proportional to Dimension. *Discrete & Computational Geometry*, 35(4):617–652.

Donoho, D. L. and Tanner, J. (2009b). Counting faces of randomly-projected polytopes when the projection radically lowers dimension. *Journal of the American Mathematical Society*, 22(1):1–53.

Drineas, P., Kannan, R., and Mahoney, M. W. (2006). Fast Monte Carlo algorithms for matrices II: computing a low-rank approximation to a matrix. *SIAM Journal on Computing*, 36(1):158–183.

Drineas, P., Mahoney, M. W., Muthukrishnan, S., and Sarlós, T. (2011). Faster least squares approximation. *Numerische Mathematik*, 117(2):219–249.

Du, R., Chen, C., Yang, B., Lu, N., Guan, X., and Shen, X. (2015). Effective urban traffic monitoring by vehicular sensor networks. *IEEE Transactions on Vehicular Technology*, 64(1):273–286.

Dutta, A., Hanzely, F., and Richtárik, P. (2018). A nonconvex projection method for robust PCA. pages 1–24.

Elad, M., Starck, J.-L., Querre, P. L., and Donoho, D. (2005). Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). *Applied and Computational Harmonic Analysis*, 19(3):340–358.

Eldar, Y. C. and Kutyniok, G. (2012). *Compressed sensing: Theory and applications*. Cambridge University Press.

Fazel, M. (2002). *Matrix Rank Minimization with Applications*. PhD thesis, Stanford University.

Forster, J., Krause, M., Lokam, S. V., Mubarakzjanov, R., Schmitt, N., and Simon, H. U. (2001). Relations Between Communication Complexity, Linear Arrangements, and Computational Complexity. In *Lecture Notes in Computer Science*, volume 2245, pages 171–182.

Foster, D. H., Nascimento, S. M. C., and Amano, K. (2004). Information limits on neural identification of colored surfaces in natural scenes. *Visual neuroscience*, 21(3):331–6.

Foucart, S. (2011). Hard Thresholding Pursuit: An Algorithm for Compressive Sensing. *SIAM Journal on Numerical Analysis*, 49(6):2543–2563.

Foucart, S. and Rauhut, H. (2013). *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Springer New York, New York, NY.

Gao, H., Cai, J.-F., Shen, Z., and Zhao, H. (2011). Robust principal component analysis-based four-dimensional computed tomography. *Physics in Medicine and Biology*, 56(11):3181–3198.

Garside, M. J. (1965). The Best Sub-Set in Multiple Regression Analysis. *Applied Statistics*, 14(2/3):196.

Ge, R., Jin, C., and Zheng, Y. (2017). No spurious local minima in nonconvex low rank problems: a unified geometric analysis. *Proceedings of the 34th International Conference on Machine Learning*, 70:1233—-1242.

Gehm, M. E., John, R., Brady, D. J., Willett, R. M., and Schulz, T. J. (2007). Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics Express*, 15(21):14013.

Gelvez, T., Arguello, H., and Rueda, H. (2015). Coded aperture design for hyper-spectral image recovery via Matrix Completion. In *2015 20th Symposium on Signal Processing, Images and Computer Vision (STSIVA)*, pages 1–7. IEEE.

Gogna, A., Shukla, A., Agarwal, H. K., and Majumdar, A. (2014). Split Bregman algorithms for sparse / joint-sparse and low-rank signal recovery: application in compressive hyperspectral imaging. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 1302–1306. IEEE.

Goldfarb, D., Ma, S., and Scheinberg, K. (2013). Fast alternating linearization methods for minimizing the sum of two convex functions. *Mathematical Programming*, 141(1-2):349–382.

Golub, G. and Kahan, W. (1965). Calculating the Singular Values and Pseudo-Inverse of a Matrix. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, 2(2):205–224.

Golub, G. H. and Reinsch, C. (1970). Singular value decomposition and least squares solutions. *Numerische Mathematik*, 14(5):403–420.

Goodall, C. and Jolliffe, I. T. (1988). *Principal Component Analysis*. Springer Series in Statistics. Springer-Verlag, New York.

Grant, M. and Boyd, S. (2008). Graph Implementations for Nonsmooth Convex Programs. In *Recent Advances in Learning and Control*, volume 371, pages 95–110. Springer London, London.

Grant, M. and Boyd, S. (2014). CVX: Matlab software for disciplined convex programming, version 2.1.

Greengard, L. and Rokhlin, V. (1997). A Fast Algorithm for Particle Simulations. *Journal of Computational Physics*, 135(2):280–292.

Gu, Q. and Wang, Z. (2016). Low-rank and sparse structure pursuit via alternating minimization. *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, 51:600—-609.

Gu, S., Zhang, L., Zuo, W., and Feng, X. (2014). Weighted nuclear norm minimization with application to image denoising. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, number 2, pages 2862–2869. IEEE.

Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, pages 49–52.

Haldar, J. and Hernando, D. (2009). Rank-constrained solutions to linear matrix equations using PowerFactorization. *IEEE Signal Processing Letters*, 16(7):584–587.

Hale, E. T., Yin, W., and Zhang, Y. (2008). Fixed-Point Continuation for l1-Minimization: Methodology and Convergence. *SIAM Journal on Optimization*, 19(3):1107–1130.

Halko, N., Martinsson, P. G., and Tropp, J. A. (2011). Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288.

Hardt, M., Meka, R., Raghavendra, P., and Weitz, B. (2014). Computational limits for Matrix Completion. *Journal of Machine Learning Research*, 35:703–725.

Harvey, N. J. A., Karger, D. R., and Yekhanin, S. (2006). The complexity of matrix completion. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm - SODA '06*, pages 1103–1111, New York, New York, USA. ACM Press.

Hitchcock, F. L. (1927). The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189.

Hitchcock, F. L. (1928). Multiple invariants and generalized rank of a p-way matrix or tensor. *Journal of Mathematics and Physics*, 7(1-4):39–79.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441.

Indyk, P. and Motwani, R. (1998). Approximate nearest neighbors. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing - STOC '98*, pages 604–613, New York, New York, USA. ACM Press.

Jin, R., Kolda, T. G., and Ward, R. (2020). Faster Johnson–Lindenstrauss transforms via Kronecker products. *Information and Inference: A Journal of the IMA*, pages 1–30.

Johnson, W. B. and Lindenstrauss, J. (1984). Extensions of Lipschitz mappings into a Hilbert space. In *Conference on Modern Analysis and Probability*, pages 189–206. 26 edition.

Kane, D. M. and Nelson, J. (2014). Sparser Johnson-Lindenstrauss Transforms. *Journal of the ACM*, 61(1):1–23.

Kemker, R., Salvaggio, C., and Kanan, C. (2018). Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145(June 2017):60–77.

Keshavan, R. H., Montanari, A., and Oh, S. (2010). Matrix Completion From a Few Entries. *IEEE Transactions on Information Theory*, 56(6):2980–2998.

Kim, S.-j., Koh, K., Lustig, M., Boyd, S., and Gorinevsky, D. (2007). An Interior-Point Method for Large-Scale -Regularized Least Squares. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):606–617.

King, E. J., Kutyniok, G., and Lim, W.-Q. (2013). Image inpainting: theoretical analysis and comparison of algorithms. In Van De Ville, D., Goyal, V. K., and Papadakis, M., editors, *SPIE Proceedings, Wavelets and Sparsity*, volume XV, page 885802.

Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. *Computer*, 42(8):30–37.

Krahmer, F. and Ward, R. (2011). New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. *SIAM Journal on Mathematical Analysis*, 43(3):1269–1281.

Kumar, A., Lokam, S. V., Patankar, V. M., and Sarma, M. N. J. (2014). Using elimination theory to construct rigid matrices. *computational complexity*, 23(4):531–563.

Kyrillidis, A. and Cevher, V. (2014). Matrix recipes for hard thresholding methods. *Journal of Mathematical Imaging and Vision*, 48(2):235–265.

Le Saux, B., Yokoya, N., Hansch, R., and Prasad, S. (2018). 2018 IEEE GRSS Data Fusion Contest: Multimodal Land Use Classification. *IEEE Geoscience and Remote Sensing Magazine*, 6(1):52–54.

Ledoux, M. (2001). *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, Rhode Island.

Lee, K. and Bresler, Y. (2010). ADMiRA: atomic decomposition for minimum rank approximation. *IEEE Transactions on Information Theory*, 56(9):4402–4416.

Levy, S. and Fullagar, P. K. (1981). Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution. *GEOPHYSICS*, 46(9):1235–1243.

Li, L., Huang, W., Gu, I.-H., and Tian, Q. (2004). Statistical modeling of complex backgrounds for foreground object detection. *IEEE Transactions on Image Processing*, 13(11):1459–1472.

Lin, Z., Chen, M., and Ma, Y. (2010). The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices.

Lin, Z., Ganesh, A., Wright, J., Wu, L., Miniming, C., and Ma, Y. (2009). Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. *Computational Advances in Multi-Sensor Adaptive Processing*, pages 1–18.

Linial, N. and Shraibman, A. (2009). Learning Complexity vs Communication Complexity. *Combinatorics, Probability and Computing*, 18(1-2):227–245.

Liu, Z., Hansson, A., and Vandenberghe, L. (2013). Nuclear norm system identification with missing inputs and outputs. *Systems & Control Letters*, 62(8):605–612.

Lokam, S. V. (2001). Spectral Methods for Matrix Rigidity with Applications to Size–Depth Trade-offs and Communication Complexity. *Journal of Computer and System Sciences*, 63(3):449–473.

Lorentz, G. G., Golitschek, M. V., and Makovoz, Y. (1996). *Constructive approximation: Advanced problems*. Springer-Verlag Berlin Heidelberg.

Luan, X., Fang, B., Liu, L., Yang, W., and Qian, J. (2014). Extracting sparse error of robust PCA for face recognition in the presence of varying illumination and occlusion. *Pattern Recognition*, 47(2):495–508.

Ma, S., Goldfarb, D., and Chen, L. (2011). Fixed point and Bregman iterative methods for matrix rank minimization. *Mathematical Programming*, 128(1-2):321–353.

Ma, W. K., Bioucas-Dias, J. M., Chanussot, J., and Gader, P. (2014). Signal and image processing in hyperspectral remote sensing. *IEEE Signal Processing Magazine*, 31(1):22–23.

Mallat, S. G. and Zhang, Z. (1993). Matching Pursuits With Time-Frequency Dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415.

Martinsson, P.-G. and Tropp, J. A. (2020). Randomized numerical linear algebra: Foundations and algorithms. *Acta Numerica*, 29:403–572.

Meyer, G., Bonnabel, S., and Sepulchre, R. (2011). Linear regression under fixed-rank constraints: A Riemannian approach. *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, pages 545–552.

Miao, L., Qi, H., Ramanath, R., and Snyder, W. E. (2006). Binary tree-based generic demosaicking algorithm for multispectral filter arrays. *IEEE Transactions on Image Processing*, 15(11):3550–3558.

Mihoubi, S., Losson, O., Mathon, B., and Macaire, L. (2015). Multispectral demosaicing using intensity-based spectral correlation. *5th International Conference on Image Processing, Theory, Tools and Applications 2015, IPTA 2015*, pages 461–466.

Mihoubi, S., Losson, O., Mathon, B., and Macaire, L. (2017). Multispectral Demosaicing Using Pseudo-Panchromatic Image. *IEEE Transactions on Computational Imaging*, 3(4):982–995.

Mu, Y., Dong, J., Yuan, X., and Yan, S. (2011). Accelerated low-rank visual recovery by random projection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2609–2616.

Nascimento, S. M. C., Ferreira, F. P., and Foster, D. H. (2002). Statistics of spatial cone-excitation ratios in natural scenes. *Journal of the Optical Society of America A*, 19(8):1484.

Needell, D. and Tropp, J. (2009). CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321.

Nesterov, Y. (2004). *Introductory Lectures on Convex Optimization*, volume 87 of *Applied Optimization*. Springer US, Boston, MA.

Netrapalli, P., Niranjan, U. N., Sanghavi, S., Anandkumar, A., and Jain, P. (2014). Non-convex robust PCA. In *Advances in Neural Information Processing Systems 27*.

Nguyen, L. T., Kim, J., Kim, S., and Shim, B. (2019). Localization of IoT Networks via Low-Rank Matrix Completion. *IEEE Transactions on Communications*, 67(8):5833–5847.

Nordrum, A. (2016). Popular Internet of Things Forecast of 50 Billion Devices by 2020 Is Outdated.

Obozinski, G., Taskar, B., and Jordan, M. I. (2010). Joint covariate selection and joint subspace selection for multiple classification problems. *Statistics and Computing*, 20(2):231–252.

Oreifej, O., Li, X., and Shah, M. (2013). Simultaneous video stabilization and moving object detection in turbulence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):450–462.

Otazo, R., Candès, E., and Sodickson, D. K. (2015). Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of

background and dynamic components. *Magnetic Resonance in Medicine*, 73(3):1125–1136.

Papoulis, A. and Chamzas, C. (1979). Improvement of Range Resolution by Spectral Extrapolation. *Ultrasonic Imaging*, 1(2):121–135.

Parente, M. and Plaza, A. (2010). Survey of geometric and statistical unmixing algorithms for hyperspectral images. In *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, pages 1–4. IEEE.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572.

Recht, B., Fazel, M., and Parrilo, P. A. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501.

Sabushimike, D., Na, S., Kim, J., Bui, N., Seo, K., and Kim, G. (2016). Low-rank matrix recovery approach for clutter rejection in real-time IR-UWB radar-based moving target detection. *Sensors*, 16(9):1409.

Santosa, F. and Symes, W. W. (1986). Linear Inversion of Band-Limited Reflection Seismograms. *SIAM Journal on Scientific and Statistical Computing*, 7(4):1307–1330.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(July 1928):379–423.

Skauli, T. and Farrell, J. (2013). A collection of hyperspectral images for imaging systems research. In Sampat, N. and Battiato, S., editors, *Proc. SPIE 8660, Digital Photography IX*, page 86600C.

Stein, D., Beaven, S., Hoff, L., Winter, E., Schaum, A., and Stocker, A. (2002). Anomaly detection from hyperspectral imagery. *IEEE Signal Processing Magazine*, 19(1):58–69.

Stewart, G. W. (1993). On the Early History of the Singular Value Decomposition. *SIAM Review*, 35(4):551–566.

Sturm, J. F. (1999). Using SeDuMi 1.02, A Matlab toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11(1-4):625–653.

Szarek, S. (1998). Metric entropy of homogeneous spaces. *Banach Center Publications*, 43(1):395–410.

Szarek, S. J. (1983). The finite dimensional basis problem with an appendix on nets of Grassmann manifolds. *Acta Mathematica*, 151(1):153–179.

Tanner, J., Thompson, A., and Vary, S. (2019). Matrix Rigidity and the Ill-Posedness of Robust PCA and Matrix Completion. *SIAM Journal on Mathematics of Data Science*, 1(3):537–554.

Tanner, J. and Wei, K. (2013). Normalized iterative hard thresholding for matrix completion. *SIAM Journal on Scientific Computing*, 35(5):S104–S125.

Tanner, J. and Wei, K. (2016). Low rank matrix completion by alternating steepest descent methods. *Applied and Computational Harmonic Analysis*, 40(2):417–429.

Tao, T. (2012). *Topics in Random Matrix Theory*, volume 123. American Mathematical Society.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the royal statistical society series b-methodological*, 58(1):267–288.

Toh, K. C., Todd, M. J., and Tütüncü, R. H. (1999). SDPT3 — A Matlab software package for semidefinite programming, Version 1.3. *Optimization Methods and Software*, 11(1-4):545–581.

Toh, K. C. and Yun, S. (2010). An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific Journal of Optimization*, 6(3):615–640.

Torre, F. D. and Black, M. J. (2001). Robust principal component analysis for computer vision. *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, 1:362–369.

Torre, F. D. and Black, M. J. (2003). A Framework for Robust Subspace Learning. *International Journal of Computer Vision*, 54(1):117–142.

Tukey, J. W. (1962). The Future of Data Analysis. *The Annals of Mathematical Statistics*, 33(1):1–67.

Valiant, L. G. (1977). Graph-theoretic arguments in low-level complexity. In Gruska, J., editor, *Mathematical Foundations of Computer Science 1977*, volume 53, pages 162–176. Springer Berlin Heidelberg, Berlin, Heidelberg.

Vandereycken, B. (2013). Low-Rank Matrix Completion by Riemannian Optimization. *SIAM Journal on Optimization*, 23(2):1214–1236.

Vane, G., Green, R. O., Chrien, T. G., Enmark, H. T., Hansen, E. G., and Porter, W. M. (1993). The airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sensing of Environment*, 44(2-3):127–143.

Wang, X., Hong, M., Ma, S., and Luo, Z. (2013). Solving multiple-block separable convex minimization problems using two-block alternating direction method of multipliers. Technical report.

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

Waters, A. E., Sankaranarayanan, A. C., and Baraniuk, R. G. (2011). SpaRCS: recovering low-rank and sparse matrices from compressive measurements. In *Advances in Neural Information Processing Systems 24*, number 2, pages 1089—-1097.

Wei, K. (2015). Fast iterative hard thresholding for compressed sensing. *IEEE Signal Processing Letters*, 22(5):593–597.

Wei, W., Zhang, L., Zhang, Y., Wang, C., and Tian, C. (2015). Hyperspectral image denoising from an incomplete observation. In *2015 International Conference on Orange Technologies (ICOT)*, pages 177–180. IEEE.

Wen, Z., Yin, W., and Zhang, Y. (2012). Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. *Mathematical Programming Computation*, 4(4):333–361.

Woodruff, D. P. (2014). Sketching as a tool for numerical linear algebra. *Foundations and Trends in Theoretical Computer Science*, 10(1-2):1–157.

Wright, J., Yang, A., Ganesh, A., Sastry, S., and Yi Ma (2009a). Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227.

Wright, S., Nowak, R., and Figueiredo, M. (2009b). Sparse Reconstruction by Separable Approximation. *IEEE Transactions on Signal Processing*, 57(7):2479–2493.

Xu, F., Han, J., Wang, Y., Chen, M., Chen, Y., He, G., and Hu, Y. (2017). Dynamic magnetic resonance imaging via nonconvex low-rank matrix approximation. *IEEE Access*, 5:1958–1966.

Xu, Y., Du, B., Zhang, L., Cerra, D., Pato, M., Carmona, E., Prasad, S., Yokoya, N., Hansch, R., and Le Saux, B. (2019). Advanced multi-sensor optical remote sensing for urban land use and land cover classification: outcome of the 2018 IEEE GRSS data fusion contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(6):1709–1724.

Yi, X., Park, D., Chen, Y., and Caramanis, C. (2016). Fast algorithms for robust PCA via gradient descent. In *Advances in Neural Information Processing Systems 29*.

Yurtsever, A., Tropp, J. A., Fercoq, O., Udell, M., and Cevher, V. (2019). Scalable Semidefinite Programming. *arxiv:1912.02949*.

Zhang, X., Wang, L., and Gu, Q. (2018). A unified framework for low-rank plus sparse matrix recovery. *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, 84:1097—-1107.

Zhou, T. and Tao, D. (2011). GoDec: Randomized low-rank & sparse matrix decomposition in noisy case. *Proceedings of the 28th International Conference on Machine Learning*, 35(1):33–40.

Zhou, Z., Li, X., Wright, J., Candès, E., and Ma, Y. (2010). Stable principal component pursuit. *IEEE International Symposium on Information Theory - Proceedings*, pages 1518–1522.