Compressive sensing algorithms for recovery of sparse and low rank signals

Mukund Sriram Narasimhan

2020

Mukund Sriram Narasimhan. (2020). Compressive sensing algorithms for recovery of sparse and low rank signals. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/145980>

<https://doi.org/10.32657/10356/145980>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 16:42:56 SGT

Compressive Sensing Algorithms for Recovery of Sparse and Low Rank Signals

Mukund Sriram Narasimhan

School of Electrical & Electronic Engineering

**A thesis submitted to the Nanyang Technological
University in partial fulfillment of the requirement for
the Degree of Master of Engineering**

2020

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarized materials, and has not been submitted for a higher degree to any other University or Institution.

[Enter Date in dd-mm-yy]

25/07/2020

.....
Date

N. Mukund Sriram

[Provide Student's Signature Here]

Mukund Sriram Narasimhan

.....
[Key in Student's Name Here]

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

25/7/2020

.....
Date

.....
Anamitra Makur

Authorship Attribution Statement

This thesis contains material from ONE paper published in the following peer-reviewed journal(s) / from papers accepted at conferences in which I am listed as an author.

Chapter 4 is published as: N. Mukund Sriram, S. S. Narayanan and A. Makur; 'Joint Recovery of Sparse Signal Ensemble using Rank-Sparsity Decomposition'. 2019 International Symposium on Signals, Circuits and Systems (ISSCS), Iasi, Romania, 2019, pp. 1-4.

The contributions of the co-authors are as follows:

- I prepared the manuscript drafts. The manuscript was revised with inputs from Dr Anamitra Makur.
- The core Idea was formulated by me. I derived the BP formulation and analyzed the algorithm.
- Dr. S. Sathiya Narayanan helped in conducting the experiments, providing valuable inputs for refining the presentation.
- Dr. Anamitra Makur, my advisor helped in structuring the paper. He also provided inputs on analysis of the algorithm and interpreting the simulation results.

[Enter Date in dd-mm-yy]

25/07/2020

.....
Date

N. Mukund Sriram

[Provide Student's Signature Here]

Mukund Sriram Narasimhan

.....
[Key in Student's Name Here]

Acknowledgement

I would like to sincerely thank Dr. Anamitra Makur for his guidance and support. I am also grateful to him for providing me the opportunity to work in the project where much of the research was carried out. I thank Dr. S. Sathiya Narayanan for his suggestions and help in this work. My parents have always stood by me in times of difficulty. I thank them for their moral support and encouragement throughout the academic period.

I would like to thank Sujith Kumar Sahoo, Abirami Srinivasan and my colleagues at Media lab-Infinitus, school of EEE -NTU for their co-operation in carrying out the day to day work. Last but not the least, I would like to thank my friends whose company I have enjoyed during my stay at NTU.

Compressive Sensing Algorithms for Recovery of Sparse and Low Rank Signals

Mukund Sriram Narasimhan (G1503371H)

MEngg student, School of EEE , NTU

Supervisor: Prof. Anamitra Makur, NTU

Abstract

Many natural signals are atomic, i.e., the signals may be represented in some low dimensional space due to their inherent structure. Two most common atomic structures are sparsity and low rank. A sparse signal (vector/matrix) has very few nonzero entries. A low rank matrix has very small rank in comparison to its dimensions.

These structures were first exploited in compressive sensing (CS) and matrix completion respectively. CS is a sampling scheme which aims to recover a sparse signal from reduced number of measurements using an appropriate recovery algorithm. In matrix completion we have a low rank matrix with missing entries, and the aim is to find the missing entries. The concepts were extended to de-mixing (separating sparse and low rank components from a mixture) and phase retrieval (recovery of signal from magnitude measurements alone). In these applications, the data exhibits both low rank and sparse properties.

We propose to study new signal models to which these concepts can be applied and to develop corresponding recovery algorithms. Towards this goal, we have worked on the following problems.

Distributed compressive Sensing (DCS) is an extension of CS from single vector to multiple signal vectors. We consider DCS in the context of joint sparse model (JSM). It consists of an ensemble of correlated signals (from different sensors in a network) that can be decomposed into a common component and an innovation component (both sparse). We write this model as a sum of a sparse and low rank matrix and develop an algorithm to recover the components from the compressed measurements. This model is further recast as a basis pursuit problem for ease of computation. The proposed method requires less number of measurements and recovery accuracy is found to be better than the existing algorithms for JSM.

Hyper-spectral data are collection of images that are acquired simultaneously from a scene in a few hundred adjacent frequency bands. It can be represented as a matrix which is both sparse and low rank. Compressive acquisition of this type of signals can result in substantial savings in terms of sensor cost and data storage. In order to recover the signal from CS type measurements, we need to solve a convex optimization problem that has both sparse and low rank penalty in the objective. Inspired by iterative re-weighted least squares, we develop an approach that involves finding a quadratic approximation for the objective. It is an iterative two step procedure where we solve a quadratic program (QP) and use the result to update the weight matrix of the objective. This approach simplifies computational complexity of recovery process and allows the use of efficient QP solvers which are widely available.

Contents

1	Introduction	6
1.1	Contributions	7
2	Compressive Sensing and Matrix Completion	9
2.1	Compressive sensing of sparse signals	9
2.2	Sampling low-rank matrix and matrix completion	11
3	Applications combining Sparsity and Low Rank	14
3.1	Phase retrieval by convex programming	14
3.2	Convexity in source separation	18
4	Joint Recovery of Sparse Signal Ensemble using Rank-Sparsity Decomposition	22
4.1	Prior work and motivation	23
4.2	Proposed joint recovery approach for JSM-1	23
4.2.1	Modeling JSM-1 as low-rank plus sparse structure	23
4.2.2	Optimization using BP	24
4.3	Analysis of the proposed method	25
4.3.1	Model based recovery versus direct combined recovery	25
4.3.2	Mutual coherence of the measurement matrix	25
4.4	Experimental results	26
5	CS for Low Rank and Sparse Signals	31
5.1	Quadratic approximation to convex objectives	32
5.1.1	Iterative re-weighted approximation for SLR model	33
5.1.2	Extensions and improvements to the approach	34
5.2	Simulation and experimental results	35

6 Conclusion and Future Directions	39
Author's publications	41

List of Figures

3.1	Reconstructed frequency domain magnitude for $m < n \log n$	16
3.2	Reconstructed time domain magnitude for $m < n \log n$	17
3.3	Reconstructed frequency domain magnitude for $m \approx n \log n$	17
3.4	Reconstructed time domain magnitude for $m \approx n \log n$	18
3.5	Composite image	20
3.6	Low rank part of the image	20
3.7	Sparse part of the image	21
4.1	Average MSE versus measurement ratio ($\frac{m}{n}$)	28
4.2	Average MSE versus common support ratio ($\frac{k_c}{k}$)	28
4.3	Average MSE versus sparsity ratio ($\frac{k}{n}$)	29
4.4	Average MSE versus ensemble size (J)	29
4.5	Average MSE versus ensemble size (J) for small J	30
4.6	Average MSE versus measurement ratio ($\frac{m}{n}$) for different λ	30
5.1	RMSE versus measurement ratio (M/N) for real valued data and fixed sparsity	36
5.2	RMSE versus measurement ratio (M/N) for integer data and fixed sparsity	37
5.3	RMSE versus measurement ratio (M/N) for real valued data and fixed rank	37
5.4	RMSE versus measurement ratio (M/N) for integer data and fixed rank	38
5.5	Reconstruction error versus iteration number for signal with $n \neq J$	38

Chapter 1

Introduction

In many engineering applications, we encounter signals which are atomic, i.e., they have structures that can be decomposed into a small number of atoms in a low-dimensional space. The focus is on recovering low dimensional signal/objects from limited number of linear measurements. Two most common low-dimension structures are sparsity and low-rank (LR). They occur in applications involving image/video processing, magnetic resonance imaging (MRI), hyperspectral images (HSI), coherent diffraction imaging (CDI), collaborative filtering, etc.

Many signals encountered in nature are sparse in a specific orthogonal basis. The main principle of compressive sensing (CS) is that a sparse signal can be recovered from far fewer samples than that required by the Nyquist theorem. CS suggests a new framework for simultaneous sampling and compression, in contrast to the common framework of first collecting data and then discarding the redundant data by digital compression techniques. At the heart of compressive sensing lies the fact that it is possible to reconstruct a sparse signal accurately from an under-determined linear system of equations by a computationally efficient approach via convex programming. In order for such an approach to succeed, the sensing matrix must be such that the measurements are incoherent with respect to each other. The restricted isometry property (RIP) and the coherence of a matrix are to date the most widely used tools to derive and characterize such conditions.

A problem which is closely related to CS is the recovery of a data matrix from incomplete or corrupted information. The problem in its simplest form is to recover a matrix from a small sample of its entries. We refer to the problem of inferring the missing entries as the matrix completion problem. This task is possible only if the underlying matrix has low rank (LR). By extension, inferring a matrix from just a few linear functionals is called the low-rank matrix recovery problem. The recovery is accomplished through nuclear-norm minimization, the tightest convex relaxation of the NP-hard rank minimization problem.

Next, we explore applications which make use of both low rank and signal sparsity property. Robust principle

component analysis (RPCA) has been proposed to remove the effect of sparse gross errors on data. For a given data matrix, RPCA seeks to decompose it into two parts, a LR matrix and a sparse matrix. That is, RPCA assumes that the given matrix $\mathbf{M}$ is a superposition of a sparse matrix and a LR matrix. As a result, the gross errors will be captured by the sparse matrix, and the low-rank matrix can still approximate $\mathbf{M}$ well. RPCA not only provide a low-dimensional approximation which is robust to outliers, but it also finds vast applications in a variety of real applications such as computer vision, image alignment, subspace recovery, clustering, and so on.

In many applications in science and engineering, we are given the modulus square of the Fourier transform of an unknown signal and are then tasked with solving an inverse problem, known as phase retrieval. Phase retrieval (PR) is the problem of recovering a signal from the magnitude of its Fourier transform. It is an ill-posed problem since the Fourier phase information is lost, and prior information is needed in order to recover the signal. In the phase-lift approach, the quadratic measurement constraints are lifted to a higher dimension by re-defining the signal through a rank one matrix. The problem can then be recast as a semi-definite program which involves minimization of rank of a lifted matrix subject to recovery constraints. For phase retrieval of sparse signals, the rank one signal matrix is also sparse. Hence, we can add a sparse penalty term to phase-lift. By leveraging sparsity in this fashion, we can reduce the necessary number of Fourier modulus measurements for recovery of the sparse signal. This approach is referred to as compressive phase retrieval via lifting (CPRL).

Chapter 2 presents the background material on CS and LR matrix recovery. In chapter 3, we discuss PR and RPCA in detail and lay the groundwork for application of these concepts to new problems.

1.1 Contributions

Distributed compressive sensing (DCS) is an extension of compressive sensing from a single measurement vector to multiple signal vectors. The theory rests on the concept of joint sparsity of a signal ensemble which enables algorithms that exploit both intra and inter-signal correlation. We consider DCS in the context of joint sparse model one (JSM-1). JSM-1 consists of an ensemble of correlated signals that can be decomposed into a sparse common component and a sparse innovation component. In chapter 4, we propose a new DCS recovery method for JSM-1. We recast JSM-1 as the sum of a sparse and a low rank matrix and develop an algorithm to recover the signal from the compressed measurements using a rank-sparsity decomposition. The problem is transformed to a basis pursuit (BP) problem for ease of computation. The proposed method requires less number of measurements and provides better recovery accuracy compared to the existing algorithms.

Hyper-spectral data are collection of images that are acquired simultaneously from a scene in a few hundred adjacent frequency bands. It can be represented as a matrix which is both sparse and low rank. Similar structure

is exhibited by multi-view image data where a scene is captured by cameras located at different positions/view-points. Compressive acquisition of these signals can result in substantial savings in terms of sensor cost and data storage. In order to recover the signal from CS measurements, we need to solve a convex optimization problem that has both sparse and low rank penalty in the objective. Inspired by iterative re-weighted least squares, we develop an approach in chapter 5 that involves finding a quadratic approximation for this objective. This is accomplished by finding a suitable weight matrix for each of the penalty functions. A quadratic approximation of the objective is constructed using the weight matrix. The resulting quadratic program (QP) is solved, and the result is used to update the weight matrices to be used in the next iteration. This process is repeated till a sufficiently accurate solution is found. The quadratic program step can be split across the bands to simplify computation and to perform parallel updates . The weight matrices are updated jointly after pooling the results.

Chapter 2

Compressive Sensing and Matrix Completion

Sparse signal recovery from subsampled measurements and the low-rank matrix completion problems are mathematically similar in that they both involve reconstruction of an atomic signal from a small number of linear measurements. In both problems, we need to use the signal structure to solve an under-determined system of equations. In this chapter, we review the theory underlying the two problems.

2.1 Compressive sensing of sparse signals

Compressive sensing (CS) [1, 2] is a framework for acquisition of a signal, which is known to be sparse in a particular basis, from a small number of linear measurements. To be specific, for $\mathbf{x} \in \mathbb{R}^n$ with k nonzero basis components, it has been shown that $m \geq ck \log(\frac{n}{k})$ random linear measurements are sufficient to reconstruct the signal for an appropriate constant c . The most common reconstruction procedure is basis pursuit (BP) [1],

$$\arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 : \mathbf{y} = \mathbf{A}\mathbf{x}, \quad (2.1)$$

where $\mathbf{y} \in \mathbb{R}^m$ is the measurement, $\mathbf{A} \in \mathbb{R}^{m \times n}$ is called the measurement matrix and $\|\mathbf{x}\|_1 = \sum_i |x_i|$ (l_1 norm). The l_1 norm is a convex relaxation of l_0 norm which measures the number of non-zero entries in $\mathbf{x}$. Hence, equation (2.1) is an approximation to

$$\arg \min_{\mathbf{x}} \|\mathbf{x}\|_0 : \mathbf{y} = \mathbf{A}\mathbf{x}. \quad (2.2)$$

We can interpret the CS problem as finding a sparse atomic decomposition (basis vectors are the atoms) [3] for the given signal in an appropriate basis ¹.

$$\mathbf{y} = \sum_{i \in \Omega} x_i \mathbf{a}_i, \quad |\Omega| = k < n. \quad (2.3)$$

The set of nonzero entry locations Ω is called the support of $\mathbf{x}$. If $\mathbf{x}$ is sparse in a basis $\Theta \in \mathbb{R}^{n \times n}$ (wavelet, Fourier or identity), then we can express the measurements as $\mathbf{y} = \mathbf{A}\Theta\alpha$, where α is sparse. Equation (2.1) then becomes

$$\arg \min_{\alpha} \|\alpha\|_1 : \mathbf{y} = \mathbf{A}\Theta\alpha. \quad (2.4)$$

An advantage of CS is that a knowledge of the basis Θ is not necessary to perform compression at the transmitter, unlike other data compression methods which require the computation of an orthogonal transform of the signal at the transmitter. In practice, the compressive measurements are corrupted by additive noise, $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, where $\mathbf{n}$ is white Gaussian noise. To account for the noise, a modified form of (2.1), termed basis pursuit denoising (BPDN) [3], is used ²:

$$\arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1. \quad (2.5)$$

In order to recover the unique sparse solution using (2.1), the measurement matrix $\mathbf{A}$ must satisfy the restricted isometry property (RIP) [2] of order k ; a matrix is said to satisfy RIP if

$$(1 - \delta_k) \|\mathbf{z}\|_2^2 \leq \|\mathbf{A}\mathbf{z}\|_2^2 \leq (1 + \delta_k) \|\mathbf{z}\|_2^2, \quad (2.6)$$

where, $\mathbf{z}$ is any k sparse vector ³. $\delta_k \in (0, 1)$ is called the restricted isometry constant of order k .

It has been shown that random sub Gaussian matrices satisfy this property with high probability (Gaussian and Bernoulli random matrices belong to this class) when $m \geq ck \log(\frac{n}{k})$ ([4]). The term 'with high probability' means that for prescribed δ and $m \geq c_1 k \log(\frac{n}{k})$, RIP 2.6 holds with probability $\geq 1 - 2\exp(-c_2 m)$ for some positive constants c_1, c_2 .

The following theorem is an example for recovery guarantee using RIP.

Theorem 1. [1] Suppose that $k \geq 1$ such that $\delta_{3k} + \delta_{4k} < 2$ and $\mathbf{y} = \mathbf{A}\mathbf{x}$ has a k -sparse solution $\mathbf{x}^*$. Then, $\mathbf{x}^* = \arg \min \{\|\mathbf{x}\|_1 : \mathbf{y} = \mathbf{A}\mathbf{x}\}$.

Since the RIP constant is difficult to compute, an alternative measure called coherence is often used to study recovery properties.

¹ x_i is the i -th component of $\mathbf{x}$ and $\mathbf{a}_i \in \mathbb{R}^m$ is the i -th column vector of $\mathbf{A}$

² $\|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2}$ (l_2 norm)

³ A k sparse vector has k non-zero entries/components

Definition 1. [1] The coherence of a matrix $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \dots \mathbf{a}_n]$ where $\mathbf{a}_i \in \mathbb{R}^m$, is defined as

$$\mu(\mathbf{A}) = \max_{i \neq j} \frac{|\langle \mathbf{a}_i, \mathbf{a}_j \rangle|}{\|\mathbf{a}_i\| \|\mathbf{a}_j\|}. \quad (2.7)$$

Note that $0 \leq \mu(\mathbf{A}) \leq 1$ ⁴. Also, $\mu(\mathbf{A}) = 0$ iff columns of $\mathbf{A}$ are orthogonal, and $\mu(\mathbf{A}) = 1$ iff $\exists i, j, i \neq j \ni \mathbf{a}_i = c\mathbf{a}_j$ for some $|c| = 1$. RIP constant and coherence are related by the bound $\delta_k \leq \mu(\mathbf{A})(k-1)$. Results based on coherence are less tight than those obtained using RIP, but analysis is easier for coherence based results. An example is the following result connecting uniqueness and coherence [5].

Theorem 2. [6] If $\mathbf{x}^*$ is a solution to $\mathbf{Ax} = \mathbf{b}$ and $\|\mathbf{x}^*\|_0 = k < \frac{1}{2}(1 + \frac{1}{\mu(\mathbf{A})})$, then $\mathbf{x}^*$ is the unique k -sparse solution.

The above result suggests that for given fixed dimensions, smaller values of $\mu(\mathbf{A})$ allows recovery of less sparse/more dense vectors.

Apart from BP, several sparse approximation algorithms have been proposed, the most popular class being greedy pursuits. Greedy pursuits are iterative algorithms which refine a sparse solution by successively picking one or more atoms in each iteration, which best improves the solution quality [5]. Examples of this approach are orthogonal matching pursuit (OMP) [7] and compressive sampling matching pursuit (CoSaMP) [8].

2.2 Sampling low-rank matrix and matrix completion

The CS approach can be extended to recovery of low rank (LR) matrices [9]. The rank of a matrix is defined as the maximum number of linearly independent rows/columns present in it. For a matrix $\mathbf{X} \in \mathbb{R}^{n \times l}$, its rank $r \leq \min(n, l)$. A matrix is said to have low rank if $r \ll \min(n, l)$. A LR matrix (LRM) $\mathbf{X} \in \mathbb{R}^{n \times l}$ can be expanded as a sparse sum of rank one matrices

$$\mathbf{X} = \sum_{k=1}^r \sigma_k \mathbf{u}_k \mathbf{v}_k^T, \quad (2.8)$$

where $\mathbf{u}_i \in \mathbb{R}^n$, $\mathbf{v}_i \in \mathbb{R}^l$ are the left and right singular vectors. σ_i is the corresponding singular value. This expansion is popularly known as the singular value decomposition (SVD). We can view (2.8) as an atomic decomposition where the rank one matrices form the atoms.

In many applications, the LRM cannot be fully sampled or directly observed. A popular example is the netflix challenge where we have a $n \times l$ array of movie ratings from various users. The matrix is low rank due to the correlation between ratings of various users. Only a few entries of the above matrix are revealed to us. Yet,

⁴ $\langle \mathbf{x}, \mathbf{y} \rangle$ denotes inner product of $\mathbf{x}$ and $\mathbf{y}$

we would like to correctly infer all ratings for the purpose of prediction. Completion of a partially observed low rank matrix from measurements/samples $\mathbf{y} \in \mathbb{R}^m$ can be accomplished by minimizing the nuclear norm, which is the convex envelope of the rank [10].

$$\begin{aligned} \arg \min_{\mathbf{X}} \quad & \|\mathbf{X}\|_* \\ \text{s.t.} \quad & \mathbf{y} = \mathcal{A}(\mathbf{X}). \end{aligned} \quad (2.9)$$

$\|\mathbf{X}\|_* = \sum_{i=1}^r \sigma_i(\mathbf{X})$, where r is the rank of $\mathbf{X}$ and $\sigma_i(\mathbf{X})$ are the non-zero singular values of $\mathbf{X}$. $\mathcal{A}$ is a linear measurement operator, which in the case of matrix completion is the subset of known matrix elements.

We can generalize $\mathcal{A}$ to measurements of the form⁵

$$\mathcal{A} : \{y_i = \langle \mathbf{X}, \mathbf{A}_i \rangle\}_{i=1}^m = \{\text{Tr}(\mathbf{A}_i^T \mathbf{X})\}_{i=1}^m, \quad (2.10)$$

where $\mathbf{A}_i \in \mathbb{R}^{n \times l}$ and $\mathbf{y}_i$ is the i 'th component of $\mathbf{y}$. In order to determine the number of measurements sufficient to reconstruct the unique LRM $\mathbf{X}$ from the measurements (2.10), we can use matrix RIP which is an extension of (2.6) [11]. $\mathcal{A}$ is said to satisfy RIP of order r if⁶

$$(1 - \delta_r) \|\mathbf{Z}\|_F^2 \leq \|\mathcal{A}(\mathbf{Z})\|_F^2 \leq (1 + \delta_r) \|\mathbf{Z}\|_F^2, \quad (2.11)$$

where $\mathbf{Z}$ is any matrix of rank r and $\delta_r \in (0, 1)$. If we consider $\mathbf{A}_i$ in (2.10) to be Gaussian random matrices, we have the following theorem.

Theorem 3. [11] *If the entries of $\mathbf{A}_i$ are i.i.d Gaussian entries $N(0, \frac{1}{m})$, then $\mathcal{A}$ defined in (2.10) satisfies r -RIP with RIP constant δ_r with high probability, if $m \geq (n + l)r/\delta_r^2$.*

Similar to theorem 1, for recovery of LRM, we have the following result.

Theorem 4. [11] *Suppose that $r \geq 1$ such that $\delta_{5r} < 1/10$ and $\mathbf{y} = \mathcal{A}(\mathbf{X})$ has a rank r solution $\mathbf{X}^*$. Then, $\mathbf{X}^* = \arg \min \{\|\mathbf{X}\|_* : \mathbf{y} = \mathcal{A}(\mathbf{X})\}$.*

For the matrix completion problem where a subset of entries are observed, this approach is not effective for the purpose of analysis. An alternate approach uses the concept of matrix coherence [9], which measures the alignment between the row/column spaces of a matrix and the standard basis vectors.

Definition 2. *If $\mathbf{U} \in \mathbb{R}^{n \times r}$ ($r < n$) is a full rank matrix, let $\mathbf{P}_U$ be the orthogonal projection onto the column*

⁵Trace of a matrix, $\text{Tr}(\mathbf{A}) = \sum_i a_{i,i}$

⁶ $\|\mathbf{X}\|_F = \sqrt{\sum_{i,j} x_{i,j}^2}$ (Frobenius norm)

space of $\mathbf{U}$. The matrix coherence of $\mathbf{U}$ defined as

$$\mu(\mathbf{U}) = \frac{n}{r} \max_{1 \leq i \leq n} \|\mathbf{P}_U \mathbf{e}_i\|_2^2, \quad (2.12)$$

where $\mathbf{e}_i$ is the i^{th} standard basis vector. For a general matrix $\mathbf{X} \in \mathbb{R}^{n \times l}$ of rank r with SVD $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, the matrix coherence is defined as $\mu(\mathbf{X}) = \max\{\mu(\mathbf{U}), \mu(\mathbf{V})\}$.

The following theorem gives us the condition under which (2.9) can recover the true matrix from the samples:

Theorem 5. Suppose that each entry of $\mathbf{X}$ is observed independently with probability $p \in (0, 1)$. If p satisfies $p > C \frac{\mu(\mathbf{X})r(\log n)^2}{n}$ for some large enough constant $C > 0$ and $n = \max(n, l)$, then nuclear norm minimization (2.9) recovers the unique optimal solution [10].

We have mostly focussed on the convex relaxation based approach to LRM sensing. However, to overcome the limitations posed by this approach with respect to computational complexity and storage, several alternative approaches have been proposed. Examples are gradient descent (Procrustes flow) [12], alternating minimization (AltMin) [13] and singular value projection (SVP) [14] .

Chapter 3

Applications combining Sparsity and Low Rank

Sparse and low-rank signals are atomic since they can be represented as a weighted sum of a small number of simple atoms. A sparse signal (vector/matrix) $\mathbf{x} \in \mathbb{R}^n$ has a few (say, k) nonzero entries (with unknown locations) compared to the signal size ($k < n$). A LR matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$ has a small rank $r < \min(M, N)$ in comparison to its dimension. In case of the sparse signals, the atomic set consists of vectors which span $\mathbb{R}^n$. For low rank matrices, the atomic set contains rank one matrices which span $\mathbb{R}^{M \times N}$. The rank of a matrix is defined as the maximum number of linearly independent rows/columns in the matrix. A k sparse signal has k degrees of freedom, and a rank r matrix has $r(M + N - r)$ degrees of freedom. Atomic norms, which promote atomic structures, enable us to find sparse and LR solutions by convex relaxation technique. Nuclear norm and l_1 norm are commonly used to find low rank and sparse solutions respectively.

This chapter surveys some applications where sparsity and low rank occur jointly, describes existing recovery algorithms for these scenarios, and presents simulation results to illustrate the performance of the existing recovery algorithms.

3.1 Phase retrieval by convex programming

Phase retrieval (PR) [15] finds application in areas such as coherent diffraction imaging, x-ray crystallography, optical coherence tomography and astronomical imaging. Here, we need to reconstruct a signal $\mathbf{x} \in \mathbb{R}^n$ from the magnitude of its linear measurements $y_i = |\langle \mathbf{a}_i, \mathbf{x} \rangle|^2$, $\forall i = 1, 2, \dots, m$. It can be written as a quadratic

optimization problem,

$$\arg \min_{\mathbf{x}} \sum_{i=1}^m (|\langle \mathbf{a}_i, \mathbf{x} \rangle|^2 - y_i)^2. \quad (3.1)$$

The measurements are usually the magnitude of Fourier transform of the signal in standard applications. The problem is called phase retrieval, since knowledge of the phase of $\langle \mathbf{a}_i, \mathbf{x} \rangle$ will reduce the problem to (2.1) and will allow the recovery of $\mathbf{x}$. PR is not a well posed problem since any solution to (3.1) multiplied by an arbitrary phase factor is still a solution. Hence, PR solution is unique only up to a global phase factor.

PR can be framed as a LR matrix recovery (convex optimization) problem by using an approach called lifting [16, 15].

$$\begin{aligned} |\langle \mathbf{a}_i, \mathbf{x} \rangle|^2 &= \mathbf{x}^T \mathbf{a}_i \mathbf{a}_i^T \mathbf{x}, \\ &= |\langle \mathbf{a}_i \mathbf{a}_i^T, \mathbf{x} \mathbf{x}^T \rangle|, \\ &= \langle \mathbf{X}, \mathbf{A}_i \rangle, \end{aligned}$$

where $\mathbf{X} = \mathbf{x} \mathbf{x}^T$ and $\mathbf{A}_i = \mathbf{a}_i \mathbf{a}_i^T$ are rank one symmetric matrices. Thus, we can solve PR if we find an $\mathbf{X}$ which satisfies

$$\begin{aligned} \mathbf{y} &= \mathcal{A}(\mathbf{X}), \\ \mathbf{X} &\succeq \mathbf{0}, \\ \text{rank}(\mathbf{X}) &= 1. \end{aligned}$$

We can drop the rank one requirement and instead minimize rank of $\mathbf{X}$, subject to the remaining constraints¹. As discussed earlier, nuclear norm minimization is a convex relaxation of the rank. In case of positive semi-definite matrices, nuclear norm is same as the trace of the matrix.

Putting these observations together results in a formulation known as phase-lift [16]:

$$\begin{aligned} \arg \min_{\mathbf{X}} \text{Tr}[\mathbf{X}] \\ \text{s.t. } \mathbf{y} &= \mathcal{A}(\mathbf{X}), \\ \mathbf{X} &\succeq \mathbf{0}. \end{aligned} \quad (3.2)$$

It has been shown that for an appropriate constant c , $m \geq cn \log n$ are sufficient for accurate recovery of $\mathbf{x}$ from its magnitude measurements, upto a constant phase factor [16].

¹ $\mathbf{X} \succeq \mathbf{0}$ denotes that $\mathbf{X}$ is a positive semi-definite matrix

The signals encountered in PR are usually sparse in an appropriate basis. It has been shown that $\mathcal{O}(k^2 \log(\frac{4n}{k^2}))$ random Fourier measurements [17] are sufficient to recover the phase of any k sparse signal. We note that if $\mathbf{x}$ is k sparse, then $\mathbf{X} = \mathbf{x}\mathbf{x}^T$ is k^2 sparse. We can leverage the sparsity to reduce the number of phaseless measurements using the following variation to phaselift:

$$\begin{aligned} \arg \min_{\mathbf{X}} \|\mathbf{X}\|_1 + \lambda \text{Tr}[\mathbf{X}] \\ \text{s.t. } \mathcal{A}(\mathbf{X}) = \mathbf{y}, \\ \mathbf{X} \succeq \mathbf{0}, \end{aligned} \quad (3.3)$$

where the l_1 norm enforces sparsity². Using (3.3), $m \geq ck^2 \log(n)$ measurements are sufficient to solve PR [18]. The approach (3.3) is referred to as compressive phase retrieval via lifting (CPRL) [18]. Both (3.2) and (3.3) can be rewritten in semi-definite program (SDP) form.

Figure 3.1: Reconstructed frequency domain magnitude for $m < n \log n$

To compare the performance of phaselift and CPRL, we conduct a simple experiment. Consider a length $n = 32$ complex signal with $k = 3$ frequency components of equal strength. We attempt to reconstruct the signal from varying numbers of linear magnitude measurements $\mathbf{A} = \mathbf{R}\mathbf{F}$, where $\mathbf{F} \in \mathbb{R}^{n \times n}$ is the DFT matrix and $\mathbf{R} \in \mathbb{R}^{m \times n}$ is a complex Gaussian random matrix.

For $m = 30$, in Fig 3.1 we see that CPRL correctly estimates the magnitudes and the positions of the

²Here, l_1 norm is defined for a matrix in the same way as for a vector

Figure 3.2: Reconstructed time domain magnitude for $m < n \log n$

Figure 3.3: Reconstructed frequency domain magnitude for $m \approx n \log n$

three component sinusoids, while phaselift is not able to localize the correct components and has a lot of spurious frequency components in the spectrum. This point is confirmed by Fig 3.2 where the magnitude of the reconstructed signal from the estimated phase using both approaches is plotted.

Figure 3.4: Reconstructed time domain magnitude for $m \approx n \log n$

We see clearly that the signal estimated using CPRL matches the original while the estimate obtained using phaselift does not. This is consistent with the fact stated earlier that $m < n$ measurements are sufficient for reconstruction using CPRL, while phaselift needs $n \log(n)$ measurements.

For $m = 160$, we observe from Fig 3.3 that both approaches correctly estimate the three frequency components present in the signal. The same can be inferred from Fig 3.4 where the reconstructed signal magnitude from the recovered phase using both approaches.

3.2 Convexity in source separation

Source separation or de-mixing is the process of extracting multiple components which have been combined together in a given signal [19]. A basic model for mixed signal $\mathbf{z}_0 \in \mathbb{R}^n$ is the superposition model $\mathbf{z}_0 = \mathbf{x}_0 + \mathbf{y}_0$, where we wish to determine the component signals $\mathbf{x}_0$ and $\mathbf{y}_0$. Sparse and LR structure of components can be used to accomplish demixing. A simple example is the separation of $\mathbf{x}$ and $\mathbf{y}$ which are sparse in different dictionaries (spikes and sine bases)

$$\begin{aligned} \min \|\mathbf{x}\|_1 + \lambda \|\mathbf{D}\mathbf{y}\|_1 \\ \text{s.t. } \mathbf{x} + \mathbf{y} = \mathbf{z}_0, \end{aligned} \tag{3.4}$$

where $\mathbf{D}$ is a DCT matrix. Robust principal component analysis (RPCA) [20] is a more interesting example which involves separation of a sparse matrix and a LR matrix

$$\begin{aligned} \min \|\mathbf{X}\|_* + \lambda \|\mathbf{Y}\|_1 \\ \mathbf{X}, \mathbf{Y} \in \mathbb{R}^{M \times N} \\ \text{s.t. } \mathbf{X} + \mathbf{Y} = \mathbf{Z}_0. \end{aligned} \tag{3.5}$$

Texture in-painting is a useful application. We consider a digital image $\mathbf{Z}_0 \in \mathbb{R}^{M \times N}$, formed by the texture component $\mathbf{X}_0$ which is low rank and an occlusion/object $\mathbf{Y}_0$ which is sparse, resulting in a rank-sparsity decomposition. To investigate the performance of RPCA for texture separation, we conduct a simple experiment [19]. Consider Fig 3.5 where we have a chessboard with a few pawns on it. The image is grayscale and its size is 300×300 and its rank is 260. The chessboard is a repetitive texture pattern of low-rank, while the few chess pieces on the board form the sparse occlusions. Performing RPCA on this image yields a separation of the board and pieces as shown in Fig 3.6 and Fig 3.7 respectively. The low-rank part Fig 3.6 has rank 220.

Figure 3.5: Composite image

Figure 3.6: Low rank part of the image

Figure 3.7: Sparse part of the image

Chapter 4

Joint Recovery of Sparse Signal Ensemble using Rank-Sparsity Decomposition

In many real word problems, we deal with an ensemble of signals from sensors distributed in the environment. The signal at each sensor is sparse in some basis and correlated with signals from other sensors. Each sensor independently encodes its signal and transmits a few of the resulting coefficients to a single collection point, where the signals are jointly decoded. This setup is referred to as distributed compressive sensing (DCS) [21].

Joint sparse model (JSM) of a signal ensemble reflects the correlation between signals acquired by different sensors. Three different models for joint sparsity are discussed in literature [21]. In the JSM-1 model [22], known as sparse common plus innovation model, all signals share a common sparse component while each individual signal of the ensemble contains a sparse innovation component.

$$\mathbf{x}_j = \mathbf{z}_c + \mathbf{z}_j, j = \{1, 2, \dots, J\}. \quad (4.1)$$

$\mathbf{z}_c$ has sparsity k_c and is common to all $\mathbf{x}_j \in \mathbb{R}^n$. $\mathbf{z}_j$ has sparsity k_i and is the unique portion of each $\mathbf{x}_j$ with respect to both support (nonzero locations) and values. J is the number of signals in the ensemble. The total sparsity of each signal in the ensemble is $k = k_c + k_i$.

In JSM-2 model, all signals of the ensemble share the same support but the values at the support location differ for each signal $\mathbf{x}_j$ [22]. This model has been studied extensively and is also referred to as multiple measurement vector (MMV) problem [23]. In the JSM-3 model [21], signals of the ensemble consist of common

part and innovation part similar to JSM-1. However, the main difference is that the common component $\mathbf{z}_c$ is not sparse.

4.1 Prior work and motivation

Our focus is on developing an effective algorithm for DCS recovery of JSM-1 signal ensemble. Texas hold em (THoEm) is a recently proposed algorithm for DCS recovery of JSM-1 signals that is scalable both in terms of bandwidth and complexity [24]. The procedure separates the recovery of the common and innovation components into two stages. The measurements used to recover the common component are obtained by averaging, while the innovation component at each sensor is recovered independently.

Robust-MSBL (RMSBL) is a modified version of multiple sparse Bayesian learning (MSBL) algorithm [25]. The algorithm was proposed for MMV sparse signal recovery when data contains outlier points which do not fit the shared sparsity structure. It modifies the least trimmed squares method, that is used for robust linear regression, to allow outliers. These two algorithms will be used to gauge the performance of our proposed approach.

We propose an alternate formulation for JSM-1 motivated by the sparse and low rank decomposition, referred to RPCA [20],

$$\begin{aligned} \min \|\mathbf{X}\|_* + \lambda \|\mathbf{Y}\|_1 \\ \mathbf{X}, \mathbf{Y} \in \mathbb{R}^{M \times N} \\ \text{s.t. } \mathbf{X} + \mathbf{Y} = \mathbf{Z}_0. \end{aligned}$$

In RPCA, we decompose a given matrix $\mathbf{Z}_0$ as a sum of a sparse matrix $\mathbf{Y}$ and a low-rank matrix $\mathbf{X}$.

4.2 Proposed joint recovery approach for JSM-1

4.2.1 Modeling JSM-1 as low-rank plus sparse structure

We can view JSM-1 model as a matrix in which the signals of an ensemble form columns $\mathbf{V} = [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_J]$. According to (4.1), we can decompose the signal matrix into a matrix of the common component and a matrix of innovations. Rank of the matrix of common component is one (low-rank) since all its columns are alike. The matrix of innovations is sparse and almost full rank. We can thus rewrite (4.1) as

$$\mathbf{V} = \mathbf{u}\mathbf{1}^T + \mathbf{X}, \tag{4.2}$$

where $\mathbf{u} = \mathbf{z}_c$, $\mathbf{1}^T \in \mathbb{R}^J$ is a vector of all ones and $\mathbf{X} = [\mathbf{z}_1 \mathbf{z}_2 \dots \mathbf{z}_J] \in \mathbb{R}^{n \times J}$.

In (4.2), we observe that the JSM-1 ensemble has a sparse plus low-rank structure similar to RPCA. However, since we know that $\mathbf{u}$ (or $\mathbf{z}_c$) is sparse and has k_c nonzero entries, we need not minimize the nuclear norm to recover the LR component. Our approach is to jointly recover the ensemble from compressed measurements $\mathbf{Y} = \mathbf{A}\mathbf{V}$, by solving

$$\begin{aligned} \min \quad & \|\mathbf{u}\|_1 + \lambda \|\mathbf{X}\|_1 \\ \text{s.t.} \quad & \mathbf{Y} = \mathbf{A}(\mathbf{u}\mathbf{1}^T + \mathbf{X}). \end{aligned} \tag{4.3}$$

For mathematical simplicity, we assume that the same measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is used on all signals. The value of λ to be used depends on the value of k_c relative to k and the number of signals J ($\lambda \propto \frac{k_i}{k_c J}$).

4.2.2 Optimization using BP

Equation (4.3) is a convex optimization problem which can be rewritten in BP form. This allows us to use efficient BP solvers which are already available. For convenience, we shall refer to this method as joint recovery for JSM-1 (JRJSM1).

Let $\tilde{\mathbf{x}} = \lambda \text{vec}(\mathbf{X})$ and $\tilde{\mathbf{y}} = \text{vec}(\mathbf{Y})$ ¹. Vectorizing (4.3), we get

$$\begin{aligned} \min \quad & \|\mathbf{u}\|_1 + \|\tilde{\mathbf{x}}\|_1 \\ \text{s.t.} \quad & \tilde{\mathbf{y}} = (\mathbf{1} \otimes \mathbf{A})\mathbf{u} + \frac{1}{\lambda}(\mathbf{I} \otimes \mathbf{A})\tilde{\mathbf{x}}. \end{aligned} \tag{4.4}$$

where $\otimes$ denotes Kronecker product, and $\mathbf{I} \in \mathbb{R}^{J \times J}$ is the identity matrix. Now, if we let $\boldsymbol{\alpha} = [\mathbf{u} ; \tilde{\mathbf{x}}]$ and $\boldsymbol{\Lambda} = [(\mathbf{1} \otimes \mathbf{A}) \quad \frac{1}{\lambda}(\mathbf{I} \otimes \mathbf{A})]$ we have the BP form

$$\min \|\boldsymbol{\alpha}\|_1 : \tilde{\mathbf{y}} = \boldsymbol{\Lambda}\boldsymbol{\alpha}. \tag{4.5}$$

Note, $\boldsymbol{\alpha} \in \mathbb{R}^{n+Jn}$ and $\tilde{\mathbf{y}} \in \mathbb{R}^{mJ}$. Thus, solving (4.5) using a BP solver achieves DCS recovery of any JSM-1 signal ensemble.

¹ $\text{vec}()$ denotes vectorization.

4.3 Analysis of the proposed method

4.3.1 Model based recovery versus direct combined recovery

We know from CS theory [1] that we need approximately $k \log(\frac{n}{k})$ linear measurements to recover a k sparse signal $\mathbf{x} \in \mathbb{R}^n$. If we apply the result to (4.5), when J is large, we roughly need $k_i \log(\frac{n}{k_i})$ measurements per signal. The number of measurements needed for direct recovery via

$$\min \|\text{vec}(\mathbf{V})\|_1 : \text{vec}(\mathbf{Y}) = (\mathbf{I} \otimes \mathbf{A})\text{vec}(\mathbf{V}), \quad (4.6)$$

without splitting the signal ensemble into common-sparse and innovation components, is $(k_c + k_i) \log(\frac{n}{k_c + k_i})$ per signal. This simple calculation shows JRJSM1 needs lesser number of measurements compared to direct joint recovery.

4.3.2 Mutual coherence of the measurement matrix

Most analysis of BP assumes that the measurement matrix is a random matrix. However, even if $\mathbf{A}$ in (4.3) is a Gaussian random matrix, the composite matrix $\mathbf{\Lambda}$ in (4.5) is not Gaussian. Hence, it is instructive to examine its coherence. In (2.7), coherence is defined as

$$\mu(\mathbf{A}) = \max_{i \neq j} \frac{|\mathbf{a}_i^T \mathbf{a}_j|}{\|\mathbf{a}_i\| \|\mathbf{a}_j\|},$$

$\mathbf{a}_i \in \mathbb{R}^m$ is a column of $\mathbf{A}$. From the results discussed in [26], we have $\mu(\mathbf{1} \otimes \mathbf{A}) = \mu(\mathbf{A})$ and $\mu(\frac{\mathbf{I} \otimes \mathbf{A}}{\lambda}) = \mu(\mathbf{A})$. Using the definition (2.7), for a concatenated matrix $[\mathbf{X} \ \mathbf{Y}] \in \mathbb{R}^{n_1 \times (m_1 + m_2)}$, we have

$$\mu([\mathbf{X} \ \mathbf{Y}]) = \max\{\mu(\mathbf{X}), \mu(\mathbf{Y}), \max_{i,j} \frac{|\mathbf{x}_i^T \mathbf{y}_j|}{\|\mathbf{x}_i\| \|\mathbf{y}_j\|}\}. \quad (4.7)$$

$\mathbf{x}_i, \mathbf{y}_j$ are columns of $\mathbf{X}$ and $\mathbf{Y}$ respectively. If we set $\mathbf{X} = (\mathbf{1} \otimes \mathbf{A})$ and $\mathbf{Y} = \frac{1}{\lambda}(\mathbf{I} \otimes \mathbf{A})$ ($n_1 = mJ, m_1 = n$ and $m_2 = nJ$), we can calculate

$$\mu(\mathbf{\Lambda}) = \max(\frac{1}{\sqrt{J}}, \mu(\mathbf{A})). \quad (4.8)$$

Notice that $\mu(\mathbf{\Lambda})$ is independent of λ . Thus, for $J > \frac{1}{\mu^2(\mathbf{A})}$, $\mu(\mathbf{\Lambda}) = \mu(\mathbf{A})$ and the recovery error of JRJSM1 will saturate [6]. In other words, increasing the number of signals in the ensemble initially reduces the recovery error for JRJSM1 as mutual coherence is reduced, but increasing J beyond a value will not result in any further improvement.

The sufficient condition for $\mathbf{x}^*$ to be a solution of (2.1) is that [6]

$$\|\mathbf{x}^*\|_0 \leq \frac{1 + \mu(\mathbf{A})^{-1}}{2}.$$

Applying this result to equations (4.5) and (4.6), for $J > \frac{1}{\mu^2(\mathbf{A})}$ we have

$$\begin{aligned} JRJSM1 : k_c + Jk_i &\leq \frac{1 + \mu(\mathbf{A})^{-1}}{2} \\ Direct : J(k_c + k_i) &\leq \frac{1 + \mu(\mathbf{A})^{-1}}{2} \end{aligned} \tag{4.9}$$

Since $k_c + Jk_i < J(k_c + k_i)$, the first inequality is more likely to be satisfied. Thus if the number of measurements is fixed, JRJSM1 is more likely to recover the signal as compared to the direct method.

4.4 Experimental results

We now present experimental results of simulations which illustrate the performance of the proposed approach (4.4). The approach is similar to the one followed in [27] and [24]. Average mean square error (AMSE) is used as the metric to compare the recovery accuracy against THoEm and RMSBL algorithms. MSE is computed over all J signals of the ensemble and averaged over multiple samples. To explore how various parameters which define the ensemble affect average MSE, four sets of experiments were carried out. The parameters of interest are measurement ratio (MR) $\frac{m}{n}$, common support ratio (CSR) $\frac{k_c}{k}$, sparsity ratio (SR) $\frac{k}{n}$ and ensemble size J . Signal length $n = 100$ in all experiments, $\lambda = \frac{1}{J}$ unless otherwise specified, and each graph point is generated by averaging over 250 trials. The measurement matrix is chosen to be a Gaussian random matrix with entries drawn from the distribution $N(0, \frac{1}{\sqrt{n}})$.

Once J , k , k_c and n are fixed, the sparse data for the simulation is generated using the equation $\mathbf{u}\mathbf{1}^T + \mathbf{X}$. To generate the common component, a set of k_c locations are chosen at random from n locations of u and populated with values sampled from a normal distribution. To generate the innovation component, a set of $J(k - k_c)$ locations (l, m) are chosen at random in the $n \times J$ matrix $\mathbf{X}$ and filled with entries drawn from a normal distribution. The rest of the locations of $\mathbf{X}$ are zero.

Fig 4.1 is a logarithmic plot of AMSE versus MR with $k = 15$, $k_c = 9$ and $J = 20$. Overall, JRJSM1 achieves smaller error than the other two algorithms in all cases. As the number of measurements increase, there is a larger decrease in AMSE for JRJSM1 compared to THoEm and RMSBL. There is a sharp transition in AMSE for MR around 0.5, since as discussed in section 4.3.1, with $m > k_i \log(\frac{n}{k_i})$ the signal ensemble can be recovered accurately according to CS theory. This is also the reason for low AMSE of JRJSM1 for $MR > 0.56$.

In Fig 4.2, we plot AMSE versus CSR with $m = 40$, $k = 15$ and $J = 20$. JRJSM1 outperforms THoEm over the entire range of CSR, and outperforms RMSBL for $\text{CSR} < 0.8$. AMSE for JRJSM1 and THoEm algorithms decrease till $\text{CSR} = 0.6$, after which the AMSE increases. This is due to model mismatch, as $\text{CSR} > 0.6$ implies that the common sparse component is dominant and there are negligible number of entries in the innovation component.

However, RMSBL does not separate the signal into two components, and relies on statistical averaging to estimate the support. Its support estimate improves as CSR increases. Hence, its AMSE steadily decreases with CSR.

Fig 4.3 plots AMSE versus SR for $m = 40$, $k_c = k/2$. AMSE of all three algorithms increase as SR increases, which is consistent with CS theory. JRJSM1 has lower AMSE over the entire range of SR compared to THoEm and RMSBL.

In Fig 4.4 we plot AMSE versus J for $m = 40$, $k = 15$ and $k_c = 9$. JRJSM1 outperforms THoEm and RMSBL for any ensemble size. The ensemble size does not have much effect on AMSE of JRJSM1 for $J > 10$. The AMSE of THoEm steadily decreases with ensemble size as the accuracy of common support estimate improves. The AMSE of RMSBL increases with J for fixed SR and CSR, since the algorithm fails to estimate increased number of entries in the innovation matrix as the ensemble size increases.

Fig 4.5 is a logarithmic plot of AMSE versus J for $m = 40$, $k = 20$ and $k_c = 10$. The aim is to study the performance of THoEm for small J . For clarity, we have also included direct joint recovery (4.6), denoted as block BP. We note that the reconstruction error for JRJSM1 is smaller than that of block BP in accordance with (4.9). There is also a clear drop in AMSE for $J > \frac{1}{\mu^2(\mathbf{A})}$, as predicted in section 4.3.2. This is due to the fact that a low value of coherence of $\mathbf{A}$ in (4.5) is desired for a successful recovery of the signal ensemble. For $J < 8$, $\mu(\mathbf{A}) = \frac{1}{\sqrt{J}}$ and decreases with J . For $J > 8$, $\mu(\mathbf{A}) = \mu(\mathbf{A})$ is independent of J and cannot reduce further and minimum AMSE is reached.

Fig 4.6 is a logarithmic plot of AMSE versus MR for $k = 15$, $k_c = 9$ and $J = 20$ with different values of λ in JRJSM1. Since both $\lambda = 1$ and $\lambda = 10$ have smallest AMSE irrespective of MR, any value between these two can be chosen. It shows that there is a sufficiently wide range of values of λ over which accurate recovery is possible.

JRJSM1 outperforms the other two algorithms due to the following reasons: The algorithm splits the common and innovation parts of the ensemble such that the solution to (4.5) has fewer nonzero entries than in combined signal ensemble. A sparser signal has a better chance of recovery. Splitting also prevents mixing of entries from the two components.

The algorithm does not make use of averaging to estimate the common component. Averaging will provide a poor estimate of the common component if J is small. With large J , averaging may overestimate sparsity of

common component and increase the error in the innovation component.

Figure 4.1: Average MSE versus measurement ratio ($\frac{m}{n}$)

Figure 4.2: Average MSE versus common support ratio ($\frac{k_c}{k}$)

Figure 4.3: Average MSE versus sparsity ratio ($\frac{k}{n}$)

Figure 4.4: Average MSE versus ensemble size (J)

Figure 4.5: Average MSE versus ensemble size (J) for small J

Figure 4.6: Average MSE versus measurement ratio ($\frac{m}{n}$) for different λ

Chapter 5

CS for Low Rank and Sparse Signals

Signals with both sparse and low rank (SLR) structure occur in applications like multiview imaging and hyperspectral images (HSI). HSI is a collection of images acquired from a scene at a few hundred adjacent frequency bands [28]. Multiview images are obtained from a visual sensor network (VSN) whose camera sensors acquire different images of a scene from specific view points [29]. The price to pay for such high spatio-spectral resolution is the large data size. In this regard, many compression methods have been developed to reduce the size of data for transmission and storage.

We can represent each multi-view image $\mathbf{X}_j$ by a column vector $\mathbf{x}_j = \text{vec}(\mathbf{X}_j) \in \mathbb{R}^n$, where n is the image resolution. Then the set of images is a matrix $\mathbf{X} \in \mathbb{R}^{n \times J}$, where J is the number of image viewpoints. Similarly, we can represent HSI by a matrix $\mathbf{X} \in \mathbb{R}^{n \times J}$, where J is the number of spectral bands and n is the resolution of the spatial image per band. There exists spatial (local) correlation in each image of the set $\mathbf{x}^j$, due to which it is sparse in an appropriate basis. The sparsity pattern can be the same or can vary across the images. The case where all signals of the set have the same support is referred to as joint sparse model 2 (JSM-2) [22]. Since the same scene is being acquired at different spectra/views, there is correlation (non-local) across the images of the set, and as a consequence $\mathbf{X}$ has low rank.

Compressive sensing can be used to acquire multiview and hyperspectral images [29, 28]. For joint recovery from an ensemble of random measurements, we can combine the sparsity of each image in an appropriate basis with a low rank prior over the entire set of images as follows,

$$\begin{aligned} \arg \min_{\mathbf{X}} \|\mathbf{X}\|_* + \lambda \|\mathbf{\Gamma X}\|_1 \\ \text{s.t. } \mathbf{Y} = \mathbf{A X}, \end{aligned} \tag{5.1}$$

or if we have JSM-2 structure,

$$\begin{aligned} \arg \min_{\mathbf{X}} \|\mathbf{X}\|_* + \lambda \|\mathbf{\Gamma X}\|_{2,1} \\ \text{s.t. } \mathbf{Y} = \mathbf{A X}, \end{aligned} \quad (5.2)$$

where ¹, $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the measurement matrix, and m is the number of measurements per image. Image $\mathbf{X}^j$ corresponding to band/view j is usually sparse in a 2D basis. In many applications, we know the nature of the sparsity pattern in the relevant basis. For example in 2D-DCT, most of the nonzero coefficients are concentrated at lower frequencies. $\mathbf{\Gamma} = \mathbf{\Phi} \otimes \mathbf{\Psi}$ is the 2D transform (separable) used to represent the individual images $\mathbf{x}^j$. In the presence of noise, the problem in (5.1) can be written as

$$\arg \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{A X}\|_2^2 + \lambda_1 \|\mathbf{X}\|_* + \lambda_2 \|\mathbf{\Gamma X}\|_1 \quad (5.3)$$

In this work we will focus mainly on the noiseless case (5.1).

5.1 Quadratic approximation to convex objectives

The problem of joint reconstruction of an image stack, formulated in (5.1), is a convex optimization problem. It can be transformed into a semi-definite program (SDP) and solved using existing algorithmic packages. This approach can become computationally complex as the image resolution and number of bands increase. We propose an iterative approach based on finding a local quadratic approximation to the objective function. Our approach is inspired by the iterative re-weighted least squares (IRLS) algorithm which has been discussed in the context of sparse vector recovery [30] and LR matrix completion [31]. We will discuss IRLS approach to solving BP (2.1) and IRLS approach to matrix completion (2.9) separately. Then we propose to combine the two approaches to solve (5.1).

In order to form a quadratic approximation of the objective, we need a suitable weight matrix. The solution to this approximate objective is used to update the weight matrix and by extension the objective for the next iteration. For the l_1 norm minimization (2.1), we use the iterative process discussed in [30]:

$$\begin{aligned} \arg \min_{\mathbf{x}} \mathbf{x}_k^T \mathbf{\Gamma}^T \mathbf{W}_{2,k} \mathbf{\Gamma} \mathbf{x}_k \\ \text{s.t. } \mathbf{y} = \mathbf{A x}_k, \end{aligned} \quad (5.4)$$

¹ $\|\mathbf{X}\|_{2,1} = \sum_j \|\mathbf{x}_j\|_1$ ($l_{2,1}$ norm)

where k denotes iteration number², $\mathbf{x}_k \in \mathbb{R}^n$ (since $J = 1$). $\mathbf{W}_{2,k}$ is a diagonal matrix with entries $w_{2,k}^i = \frac{1}{\sqrt{\mathbf{x}_{i,k-1}^2 + \epsilon_2^2}}$, which is updated every iteration (ϵ_2 is a small constant used to prevent any diagonal entry from becoming infinity).

For nuclear-norm minimization problem (2.9), the following approximation from [31] is used:

$$\begin{aligned} \arg \min_{\mathbf{X}} \|\mathbf{W}_{1,k}^{\frac{1}{2}} \mathbf{X}_k\|_F^2 \\ \text{s.t. } \mathbf{Y} = \mathbf{A} \mathbf{X}_k, \end{aligned} \quad (5.5)$$

where $\mathbf{W}_{1,k} = (\mathbf{X}_{k-1} \mathbf{X}_{k-1}^T)^{-\frac{1}{2}}$ and $\mathbf{X}_k \in \mathbb{R}^{n \times J}$. Since (5.4) uses a vector $\mathbf{x}_k$ while (5.5) uses a matrix $\mathbf{X}_k$, (5.5) should be vectorized in order to combine it with (5.4) as discussed in the next subsection.

5.1.1 Iterative re-weighted approximation for SLR model

We need to combine the approaches of (5.4) and (5.5) to solve (5.1). Vectorization is essential to obtain a quadratic approximation to the objective in (5.1),

$$\|\mathbf{X}\|_* + \lambda \|\mathbf{X}\|_1 = \|(\mathbf{I} \otimes \mathbf{W}_{1,k}^{\frac{1}{2}}) \mathbf{x}_k\|_2^2 + \lambda \mathbf{x}_k^T \mathbf{W}_{2,k} \mathbf{x}_k,$$

where $\mathbf{x} = \text{vec}(\mathbf{X}) \in \mathbb{R}^{nJ}$. For ease of presentation, we have dropped Γ . Subscripts are used to distinguish between the weight matrices corresponding to the two norms. Expanding the first term and vectorizing the constraint gives us

$$\begin{aligned} \arg \min_{\mathbf{x}} \mathbf{x}_k^T (\mathbf{I}_J \otimes \mathbf{W}_{1,k} + \lambda \mathbf{W}_{2,k}) \mathbf{x}_k \\ \text{s.t. } \mathbf{y} = (\mathbf{I}_J \otimes \mathbf{A}) \mathbf{x}_k, \end{aligned} \quad (5.6)$$

where $\mathbf{y} = \text{vec}(\mathbf{Y}) \in \mathbb{R}^{mJ}$. The quadratic objective in (5.6) serves as a local approximation to the true objective. The solution to (5.6) is used to update the weight matrices $\mathbf{W}_{1,k+1}$ and $\mathbf{W}_{2,k+1}$. The updated weight matrices are used to obtain an improved solution. Let $\mathbf{W}_k = (\mathbf{I} \otimes \mathbf{W}_{1,k} + \lambda \mathbf{W}_{2,k}) \in \mathbb{R}^{nJ \times nJ}$ and $\hat{\mathbf{A}} = (\mathbf{I}_J \otimes \mathbf{A}) \in \mathbb{R}^{mJ \times nJ}$. Eqn (5.6) reduces to

$$\begin{aligned} \arg \min_{\mathbf{x}} \mathbf{x}_k^T \mathbf{W}_k \mathbf{x}_k \\ \text{s.t. } \mathbf{y} = \hat{\mathbf{A}} \mathbf{x}_k. \end{aligned} \quad (5.7)$$

²In this chapter, we use s to denote signal sparsity

We refer to the approach (5.6) as iterative re-weighted sparse and low-rank solver (IRSLRS). Since it involves a quadratic objective and linear constraints, efficient quadratic programming (QP) solvers that are widely available can be used to solve (5.7).

The steps to solve (5.6) and update the weight matrices is given in algorithm 1 ³.

Algorithm 1 IRSLRS

```

1: while  $k < K_0$  do
2:    $\arg \min_{\mathbf{x}} \mathbf{x}_k^T \mathbf{W}_k \mathbf{x}_k$  s.t.  $\mathbf{y} = \hat{\mathbf{A}} \mathbf{x}_k$ 
3:    $w_{2,k}^i = \frac{1}{\sqrt{\mathbf{x}_{i,k-1}^2 + \epsilon_2^2}}$ ,  $\mathbf{W}_{2,k} = \text{diag}(\mathbf{w}_{2,k})$ 
4:    $\mathbf{X}_k \mathbf{X}_k^T = \mathbf{U}_k \Sigma_k^2 \mathbf{U}_k^T$ ,  $\mathbf{W}_{1,k} = \mathbf{U}_k \Sigma_{k,\epsilon} \mathbf{U}_k^T$ ,  $\Sigma_{k,i,\epsilon} = \frac{1}{\sqrt{\Sigma_{i,k-1}^2 + \epsilon_1^2}}$ 
5:    $\mathbf{W}_k = \mathbf{I} \otimes \mathbf{W}_{1,k} + \lambda \mathbf{W}_{2,k}$ ,  $k = k + 1$ 
6: end while

```

5.1.2 Extensions and improvements to the approach

For simplicity of exposition, we have so far assumed that the same sensing matrix $\mathbf{A}$ is used for sensing all J signals. However, a setup where distinct sensing matrix $\mathbf{A}_j$ is used for each $\mathbf{X}^j$ can be readily incorporated into our reconstruction procedure (5.7) using

$$\hat{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{A}_2 & \cdots & 0 \\ \vdots & \cdots & \cdots & \vdots \\ 0 & \cdots & \cdots & \mathbf{A}_J \end{bmatrix}.$$

An advantage of our approach is that we can split the optimization problem (5.6) into smaller subproblems. Let $\mathbf{y}^j = \mathbf{A}_j \mathbf{x}^j$ be the measurement for the j^{th} band. $\mathbf{W}_{2,k}^j \in \mathbb{R}^{n \times n}$ is the j^{th} diagonal block of $\mathbf{W}_{2,k} \in \mathbb{R}^{nJ \times nJ}$. Then the QP step of the algorithm can be split into smaller QP's

$$\begin{aligned} \arg \min_{\mathbf{x}^j} \mathbf{x}_k^{jT} (\mathbf{W}_{1,k} + \lambda \mathbf{\Gamma}^T \mathbf{W}_{2,k}^j \mathbf{\Gamma}) \mathbf{x}_k^j \\ \text{s.t. } \mathbf{y}^j = \mathbf{A}_j \mathbf{x}_k^j, \end{aligned}$$

this procedure can be carried out on $j = 1, 2, \dots, J$ sequentially or in parallel, followed by update of $\mathbf{W}_{1,k}$ and $\mathbf{W}_{2,k}$. The details of this approach is given in algorithm-2.

³ ϵ_1 is a small constant used to prevent any diagonal entry from becoming infinity

Algorithm 2 Modified IRSRLRS

```
1: while  $k < K_0$  do
2:   for  $j \leftarrow 1$  to  $J$  do
3:      $\mathbf{W}_k^j = (\mathbf{W}_{1,k} + \lambda \Gamma^T \mathbf{W}_{2,k}^j \Gamma)$ 
4:      $\arg \min_{\mathbf{x}^j} \mathbf{x}_k^{jT} \mathbf{W}_k^j \mathbf{x}_k^j$ , s.t.  $\mathbf{y}^j = \mathbf{A}_j \mathbf{x}_k^j$ ,
5:   end for
6:    $w_{2,k}^i = \frac{1}{\sqrt{\mathbf{x}_{i,k-1}^2 + \epsilon_2^2}}$ ,  $\mathbf{W}_{2,k} = \text{diag}(\mathbf{w}_{2,k})$ 
7:    $\mathbf{X}_k \mathbf{X}_k^T = \mathbf{U}_k \Sigma_k^2 \mathbf{U}_k^T$ ,  $\mathbf{W}_{1,k} = \mathbf{U}_k \Sigma_{k,\epsilon} \mathbf{U}_k^T$ ,  $\Sigma_{k,i,\epsilon} = \frac{1}{\sqrt{\Sigma_{i,k-1}^2 + \epsilon_1^2}}$ 
8:   Increment  $k$ 
9: end while
```

5.2 Simulation and experimental results

We study the performance of IRSRLRS in the following experiments. Root mean square error (RMSE) is used as the metric to measure recovery accuracy. The experiments test the performance of IRSRLRS with both integer and real valued data. Depending on the nature of data, we use two types of sensing matrices in the experiments, Gaussian matrix with $a_{ij} \sim N(0, \frac{1}{\sqrt{n}})$ and Rademacher matrix with equiprobable $a_{ij} \in \{\frac{1}{\sqrt{n}}, -\frac{1}{\sqrt{n}}\}$.

The data matrix with $n = 25$ and $J = n$ used in the experiments must have both sparse and LR properties. Hence, it is a row-sparse matrix. Once s and r are decided, the data for simulation is generated as follows. A rank r matrix is created using $\mathbf{S}' = \mathbf{S}\mathbf{S}^T$, $\mathbf{S} \in \mathbb{R}^{n \times r}$. The entries of $\mathbf{S}$ can be chosen from a set of integers (to simulate image data) or from standard normal distribution. A subset of s rows of $\mathbf{S}'$ are chosen at random to be retained and the rest of rows are set to zero to yield an input data matrix $\mathbf{X}_0$ (rank of $\mathbf{X}_0$ is at most r). The RMSE for a given number of measurements is obtained over 70 independent trials. In each of the figures, we plot RMSE against measurement ratio (MR) m/n .

In Fig 5.1 and 5.2 we examine the effect of rank on recovery error. Keeping sparsity fixed at $s = 8$, we plot RMSE vs MR for three different values of matrix rank r . We observe that the RMSE reduces with increasing number of measurements, with perfect recovery for $m = n$. From Fig 5.1 and 5.2, we note that a lower value of rank results in lesser recovery error irrespective of number of measurements.

In Fig 5.3 and 5.4 the rank of the matrix $r = 5$ is fixed and we plot RMSE vs MR for three different values of signal sparsity s . Similar to the previous plots, RMSE reduces as the measurements increase, in line with the theory in literature. From Fig 5.3 and 5.4, we find for a LR matrix, a greater sparsity leads to a lower recovery error. Hence, for the algorithm to recover the signal from reduced number of measurements with acceptable error, the signal matrix must possess both sparse and LR properties simultaneously.

In Fig 5.5, we show the convergence of the algorithm for a single input signal sample with $n = 50$, $J = 40$, $s = 15$ and $r = 8$. We use $m = 30$ Gaussian measurements and plot the reconstruction error ($\|\mathbf{X} - \mathbf{X}_k\|_F$) versus the iteration number.

Separate IRLS approaches for recovery of sparse signal and LR matrix are already available in literature. We intend to combine these approaches to recover signals which have both characteristics. This approach reduces the modeling error incurred by emphasizing only one of the characteristics (overfitting). The combined objective also limits the search space. Unlike SDP based approach, IRLS is iterative, scalable, and can be run to desired accuracy.

Figure 5.1: RMSE versus measurement ratio (M/N) for real valued data and fixed sparsity

Figure 5.2: RMSE versus measurement ratio (M/N) for integer data and fixed sparsity

Figure 5.3: RMSE versus measurement ratio (M/N) for real valued data and fixed rank

Figure 5.4: RMSE versus measurement ratio (M/N) for integer data and fixed rank

Figure 5.5: Reconstruction error versus iteration number for signal with $n \neq J$

Chapter 6

Conclusion and Future Directions

The focus of this thesis is on recovering low dimensional signal/objects from limited measurements and exploiting low dimensional structure in signal processing applications. The two most common low dimension structures are sparsity and low rank. In chapter 2, we discuss CS and matrix completion, which involves recovery by convex programming of low dimensional signal from a small number of linear measurements. In chapter 3, we consider applications where sparse and LR structure occurs together. PR involves recovering a signal from limited number of nonlinear (magnitude only) measurements. The CPRL approach involves using the mechanism of matrix lifting to formulate PR as a low rank matrix recovery problem with sparsity constraint. In the same chapter, we also discuss RPCA, where we intend to separate a sparse signal and a LR signal which have been mixed together. An application to texture separation is presented.

In chapter 4, we proposed JRJSM1, a DCS algorithm for joint reconstruction of JSM-1 signals, which is based on splitting the signal ensemble into common and innovation components. This structure is used to formulate a LR and sparse signal separation problem similar to RPCA. We compare the performance with THoEm and RMSBL algorithms, which were proposed to jointly recover JSM-1 signal ensemble. JRJSM1 requires less number of measurements for successful signal recovery compared to existing methods. Experimental results show that JRJSM1 outperforms THoEm and RMSBL in terms of reconstruction accuracy.

In chapter 5, the focus is on CS of a signal which is jointly sparse and low rank. This type of signal occurs in HSI and multi-view imaging. We develop an iterative re-weighted approach for signal recovery. The algorithm transforms the convex optimization into a sequence of QP's. We also discuss ways to improve computational efficiency. The preliminary experiments performed by us show results in line with the theory established in literature; that the recovery error reduces for greater sparsity, lower rank, and greater measurement ratio.

The work described in chapter 5 may be continued in the following directions.

- 1) Optimizing the algorithm and finding suitable transform so as to apply the method to real world HSI data.
- 2) Work on extending the method to the noisy case.
- 3) Work on simplifying SVD update step of the algorithm.

Author's publications

- N. Mukund Sriram, S. S. Narayanan and A. Makur, 'Joint Recovery of Sparse Signal Ensemble using Rank-Sparsity Decomposition'.2019 International Symposium on Signals, Circuits and Systems (ISSCS), Iasi, Romania, 2019, pp. 1-4.
- N. Mukund Sriram and A. Makur, 'Iterative re-weighted algorithm for recovery of sparse and low-rank signals from compressive measurements '. In preparation

Bibliography

- [1] E. Candes and M. Wakin, “An introduction to compressive sampling,” *IEEE Signal Processing Magazine*, vol. 25, pp. 21–30, March 2008.
- [2] R. Baraniuk, “Compressive sensing [lecture notes],” *IEEE Signal Processing Magazine*, vol. 24, pp. 118–121, July 2007.
- [3] S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [4] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [5] J. A. Tropp, “Greed is good: algorithmic results for sparse approximation,” *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [6] M. Elad, *Sparse and Redundant representations : From theory to applications in signal and image processing*. Technion, Haifa, Israel: Springer, 2010.
- [7] Y. Pati, R. Rezaifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” in *1993 Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers.*, pp. 40–44 vol.1, 1993.
- [8] D. Needell and J. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [9] M. A. Davenport and J. Romberg, “An overview of low-rank matrix recovery from incomplete observations,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, pp. 608–622, June 2016.
- [10] E. J. Candes and Y. Plan, “Matrix completion with noise,” *Proceedings of the IEEE*, vol. 98, pp. 925–936, June 2010.

- [11] B. Recht, M. Fazel, and P. A. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.
- [12] S. Tu, R. Boczar, M. Simchowitz, M. Soltanolkotabi, and B. Recht, “Low-rank solutions of linear matrix equations via procrustes flow,” in *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML16, p. 964973, JMLR.org, 2016.
- [13] P. Jain, P. Netrapalli, and S. Sanghavi, “Low-rank matrix completion using alternating minimization,” *Proceedings of the Annual ACM Symposium on Theory of Computing*, 12 2012.
- [14] P. Jain, R. Meka, and I. S. Dhillon, “Guaranteed rank minimization via singular value projection,” in *Advances in Neural Information Processing Systems 23* (J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, eds.), pp. 937–945, Curran Associates, Inc., 2010.
- [15] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, “Phase retrieval with application to optical imaging: A contemporary overview,” *IEEE Signal Processing Magazine*, vol. 32, pp. 87–109, May 2015.
- [16] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, “Phase retrieval via matrix completion,” *SIAM Review*, vol. 57, no. 2, pp. 225–251, 2015.
- [17] Y. C. Eldar, N. Hammen, and D. G. Mixon, “Recent advances in phase retrieval [lecture notes],” *IEEE Signal Processing Magazine*, vol. 33, pp. 158–162, Sept 2016.
- [18] H. Ohlsson, A. Y. Yang, R. Dong, and S. S. Sastry, “Compressive phase retrieval from squared output measurements via semidefinite programming,” *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 89 – 94, 2012. 16th IFAC Symposium on System Identification.
- [19] M. B. McCoy, V. Cevher, Q. T. Dinh, A. Asaei, and L. Baldassarre, “Convexity in source separation : Models, geometry, and algorithms,” *IEEE Signal Processing Magazine*, vol. 31, pp. 87–95, May 2014.
- [20] E. J. Candes, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis,” *J. ACM*, vol. 58, pp. 11:1–11:37, June 2011.
- [21] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, “An information-theoretic approach to Distributed Compressed Sensing,” in *Proc. 43rd Allerton Conf. Communication, Control, and Computing*, 2005.

- [22] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers, 2005.*, pp. 1537–1541, October 2005.
- [23] Y. Eldar, P. Kuppinger, and H. Bolcskei, "Block-Sparse signals: uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, pp. 3042–3054, June 2010.
- [24] S. R. Schnelle, J. N. Laska, C. Hegde, M. F. Duarte, M. A. Davenport, and R. G. Baraniuk, "Texas hold 'em algorithms for distributed compressive sensing," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2886–2889, March 2010.
- [25] O. Balkan, K. Kreutz-Delgado, and S. Makeig, "Robust joint sparse recovery on data with outliers," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3821–3825, May 2013.
- [26] M. F. Duarte and R. G. Baraniuk, "Kronecker Compressive Sensing," *IEEE Transactions on Image Processing*, vol. 21, pp. 494–504, Feb 2012.
- [27] S. Narayanan, S. K. Sahoo, and A. Makur, "Greedy pursuits assisted basis pursuit for reconstruction of joint-sparse signals," *Signal Processing*, vol. 142, pp. 485–491, 2018.
- [28] M. Golbabaee and P. Vandergheynst, "Hyperspectral image compressed sensing via low-rank and joint-sparse matrix recovery," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2741–2744, March 2012.
- [29] M. H. Kamal and P. Vandergheynst, "Joint low-rank and sparse light field modelling for dense multiview data compression," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3831–3835, May 2013.
- [30] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Gntkr, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.
- [31] M. Fornasier, H. Rauhut, and R. Ward, "Low-rank matrix recovery via iteratively reweighted least squares minimization," *SIAM Journal on Optimization*, vol. 21, no. 4, pp. 1614–1640, 2011.