Doctoral theses at NTNU, 2016:182

Solomon Abedom Tesfamicael

Compressive Sensing in Signal Processing: Performance Analysis and Applications

ISBN 978-82-326-1703-6 (printed ver.)
ISBN 978-82-471-1702-9 (electronic ver.)
ISSN 1503-8181

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology,
Mathematics and Electrical Engineering
Department of Electronics and
Telecommunications

Doctoral theses at NTNU, 2016:182

NTNU

Solomon Abedom Tesfamicael

Compressive Sensing in Signal Processing: Performance Analysis and Applications

Thesis for the Degree of Philosophiae Doctor

Trondheim, June 2016

Norwegian University of Science and Technology
Faculty of Information Technology,
Mathematics and Electrical Engineering
Department of Electronics and Telecommunications

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology, Mathematics and Electrical Engineering
Department of Electronics and Telecommunications

ISBN 978-82-326-1703-6 (printed ver.)
ISBN 978-82-471-1702-9 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2016:182

Printed by NTNU Grafisk senter

To the three wonderful people in my life:
my mother Silas Tesfahun,
my wife Genet Gebru and
my daughter Abigail Solomon.
May God bless you all!

Abstract

This thesis deals with an emerging area of signal processing, called Compressive Sensing (CS), that allows the reconstruction of sparse or compressible signals from fewer measurements than are used in traditional schemes. Like traditional signal representation schemes, CS follows a similar framework: encoding, transmission/storing, and decoding. The encoding part is done using random projection (RP) or random sensing, and the decoding is done via nonlinear reconstruction algorithms from a reduced amount of measurements. The performance of the reconstruction schemes used and the application of such paradigm are the two main focuses of the thesis. It has three parts: the introduction, performance analysis of recovery algorithms in CS and some applications of CS.

The introductory part provides the background for the following four chapters. It begins by defining the basic concepts used in CS theory and presents the Bayesian framework. Further, an analytical tool from statistical mechanics for performance analysis of physical systems is introduced applied on a non-noisy CS problem. The Bayesian framework is given ample emphasis in the thesis for two reasons. First, it serves as a bridge between the recovery algorithms used in CS and a tool from the statistical mechanics, called the replica method. Second, it is used as main framework to incorporate different prior signal information, like sparsity and clusteredness. Furthermore, a short description of CS applications is given before the introduction concludes by presenting the scope and the contribution of the thesis.

The second part of the thesis deals with the study of the performance of recovery methods in CS systems using the Replica Method. At the beginning, the study of the performance of the recovery algorithms in CS was focused on the ratio of the amount of measurement used, the sparsity level, and how accurate the recovered information is. However, there was a luck of universal performance analysis. The Replica method provides this by considering large size systems. This thesis contributes to such analysis via the Bayesian framework. In this work noisy CS systems are considered and the recovering algorithms are reinterpreted as a maximum a posteriori (MAP) estimator. It, therefore, provides replica analysis including one step replica breaking ansatz for CS systems as an extension of similar analysis done for other systems like multiple input/multiple output (MIMO).

The CS application is the third part of the thesis. The theory of CS has been applied in many signal processing fields such as image processing, communication, networks and so on. There are hundreds, if not thousands of articles on this subject at present. In this thesis, there are novel results that contribute to the application of CS theory. The theory is applied to limited feedback in temporally correlated MIMO channels, where the sparsity property was used to reduce feedback overhead significantly while delivering the same performance. Further, including another assumption, i.e., more structure among the sparse entries, to the sparsity of signals, and modeling it as a modified Laplacian prior in Bayesian setting, a novel way of compressive sensing is presented in this thesis. It can have potential impact on medical imaging processing, especially to magnetic resonance imaging (MRI).

Preface

This thesis is submitted to the Norwegian University of Science (NTNU) for partial fulfilment of the requirements for the degree of Philosophiae Doctor (PhD). The research was carried out at the Department of Electronics and Telecommunications (IET), NTNU, Trondheim, Norway.

The work has been carried out from Sept. 2009 to June 2016 at IET, first with Professor Ralf R. Müller and later Professor Lars Lundheim as the main supervisor. The research began as an extended part of the project Statistical Physics of Advanced Multiuser Communications (SPAM) to Compressive Sensing (CS) systems. At the beginning it was self financed and later from Jan, 2011 to June, 2012 the department of IET has supported it through course works. Funding for attending conferences and publication in proceedings and journals was from IET, and FLT (former HiST). The PhD study includes coursework of 59 credits and two years of teaching duties in graduate courses.

Acknowledgements

First of all I would like to give all the glory and praises to God almighty, who is the source of everything and everyone in this journey. Yes, He gives the desires of my heart (Psalm 37:5). I would like to thank also my supervisors Professor Lars Lundheim and Professor Ralf R. Müller for their support, guidance, and fruitful discussions. I am grateful for Professor Müller for giving me the opportunity to begin this PhD study and for his excellent academic contribution. My deepest gratitude goes to Professor Lundheim for his assistance, encouragement, practical and academic assistance from the beginning to the end of this study. It was through assisting his courses that I was supported financially in the department. And of course this work has come to an end by his commitment in the work.

Dr. Bruhtesfa Godana began his PhD work almost at the same time with me. This timing was one of the good things that happened for me in my study, as it is said in Proverbs 17:17 " $\cdots$ *a brother is born for a time of adversity.*" I have benefited from our deep conversations and discussions during the course work. I will always be grateful for fruitful collaboration and cheerful friendship together with your beloved Aida. You will always be remembered my friend! I am also thankful to Faraz Barzideh for collaboration, his coding help in MATLAB, and proofreading my thesis. I would like to express my gratitude for Dr. Anteneh Biru also for proofreading the introduction of my thesis.

All friends, PhD colleagues, professors and other staff at the Signal Processing group and IET department deserve much gratitude for the conducive environment and hospitality during my studies. My special thanks goes to Professor Ragnar Hergum, Professor Geir Eøien, Randi Hostad, and Bente Seem Lindquist for facilitating the financial support at the department and Kristen Ekseth and others at the office for providing necessary official assistance. My thanks also go to Professor Kimmo Kansanen for involving me in the Signal Processing Seminar Series. I am grateful to the previous PhD colleagues Dr. Hieu T. Nguyen, Dr Nurul Huda, Dr. Hessam Moussavinik, Dr. Rodrigo de Miguel, Dr. Vesna Gardasevic and others to their assistance, advice and fellowship during my studies.

I would like to express my special gratitude to all colleagues at FLT- NTNU (former ALT-HiST) for their encouragement and support, especially, the leaders for the financial help to participate in conferences and to publication in proceedings and journals, and also giving me leave to prepare for the defence.

I would like to thank also all the wonderful people at Betel, Ebenezer, Peniel and other churches here in Norway and back home in Ethiopia for their prayers, support, encouragement, and motivation. I would like to extend my gratitude also to the Ethiopian and Eritrean community here in Trondheim for their encouragement and support.

My most gratitude and appreciation goes to my families here and back home. Especially, to my wife Genet Gebru and my sweet daughter Abigail for their unfailing support, love, care, encouragement, understanding, patience, and perseverance through all the years which enabled me to begin and to finish my work. Special thanks to " Tante Abigail " and to all my brothers back home for prayers, support, patience and understanding. I love you all!

Trondheim, June 2016
Solomon A. Tesfamicael

Organization of the Thesis

This thesis has three main parts:

Part I is the introduction to the thesis and it contains only one chapter (Chapter 1).

Part II deals with performance analysis of large size CS system using Statistical Mechanics tools. It consists of one journal paper included as chapter 2.

Part III is about CS application. It contains adapted versions of three conference papers (chapter 3, 4 and 5) included as three chapters in the thesis.

In addition, appendices is included as part IV, providing proofs, mainly for chapter 1 and 2, and abstracts of secondary papers.

Contents

Abbreviations

AT	Almedia-Thouless
ADRQ	Adaptive Rotative Quantization
AMP	Approximate Message Passing
BM3D	Block_Matching and 3D Filtering
BPDN	Basis Persuit Denoising
CCS	Clustered Compressed Sensing
CCSD	Clustered Compressed Sensing Denoising
CDMA	Code Division Multiple Access
CLASSO	Clustered Least Absolute Shrinkage and Selection Operator
CluSS-MCMC	Clustered Sparse Signal via Markov Chain Monte Carleo
CluSS-VB	Clustered Sparse Signal via Variational Bayes Approach
CoSaMP	Compressive Sampling Matching Pursuit
CPT	Classical Probablity Theory
CS	Compressive or Compressed(Compact) Sensing
CSI	Channel State Information
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
FBMP	Fast Bayesian Matching Pursuit
FDD	Frequency Division Duplex
FFT	Fast Furier Transform
fMRI	functional Magnetic Resonance Imaging
GP	Greedy Pursuits
GLM	General Linear Model
IEEE	Institute of Electrical and Electronics Engineers
IHT	Iterative Hard Thresholding

i.i.d.	Independent and identically distributed
IJSSST	International Journal of Simulation Systems, Science and Technology
LASSO	Least Absolute Shrinkage and Selection Operator
LMMSE	Linear Minimum Mean Square Error
L1-LS	L1-norm Regularized Least Squares
LS	Least Squares
LP	Linear Programming
MAP	Maximum a Posterior
MAR	Minimum-to-Average Ratio
MC	Mutual Coherence
MCMC	Markov Chain Monte Carlo
MIMO	Multiple Input Multiple Output
MMSE	Minimum Mean Square Error
MRFs	Markov Random Fields
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
MSE_LASSO	Mean Squared Error of LASSO
MSE_OP	Mean Squared Error of the Oracle Projector
NP	Non deterministic Polynomial
OMP	Orthogonal Matching Pursuit
P	Percentage
PSNR	Pick Signal-to-Noise Ratio
QCPL	Quadratic Constrained Linear Programing
QP	Quadratic Programing
RIP	Restricted Isometry Property

RMT	Random Matrix Theory
RP	Randomly Project
RS	Replica Symmetry
RSB	Replica Symmetry Breaking
1RSB	One-step Replica Symmetry Breaking
SNR	Signal-to-Noise Ratio
SNR_m	Measurment Signal-to-Noise Ratio
SNR_r	Recovered Signal Signal-to-Noise Ratio
SK model	Sherrington-Kirkpatrick Model
SP	Subspace Pursuit
SVD	Singular Value Decompostion
VB	Variational Bayesian
WSN	Wireless Sensor Network
WSNs	Wireless Sensor Networks

List of Nomenclature

$\mathbb{N}$	The set of natural numbers
$\mathbb{R}$	The set of real numbers
$\mathbb{C}$	The set of complex numbers
$\mathbf{x}^0$	Original sparse vector
$\mathbf{x}$	Sparse vector
$\mathbf{y}$	Measurment vector
$\mathbf{A}$	Random matrix
$\mathbf{J}$	Cross Correlation matrix
Ψ	Sensing matrix
Φ	Representation matrix
Φ	Representation matrix
$\mathbf{w}, \mathbf{n}$	Noise vector vectors
$\hat{\mathbf{x}}$	Estimated signal
k	The number of non-zero elements in a sparse vector
M	Dimension of measurement vector
N	Dimension of the sparse vector
ρ	Sparsity ratio
α	Measurement ratio
λ, γ	System parameters
σ_0	True noise variance
σ_u, σ	Assumed noise variance
δ_k	Constant of radius for the RIP-condition
α_c	Critical compression rate
$\|\cdot\|$	Absolute value
$\langle\cdot\rangle$	Inner product of two vectors

$\begin{pmatrix} \cdot \\ \cdot \end{pmatrix}$	Binomial coefficient
$\|\cdot\|_p$	l_p-norm for $p = 0, 1, 2$.
$\mathcal{B}_p$	Unit balls corresponding to l_p-norm
$\mathcal{C}_p$	The typical value of the minimized l_p-norm
$(\cdot)^{-1}$	The inverse operator
$(\cdot)^T$	The transpose operator
$\equiv$	Equal by definition
$\ll$	Much less than
$\gg$	Much greater than
$p(\cdot)$	Probability and probability density for discrete and continuous random variables, respectively
$p(\cdot\|\cdot)$	Conditional probability density
$h(\cdot)$	Entropy function
$\mu(\cdot)$	Coherence of the matrix
$\mathbb{E}(\cdot)$	The expectation operator
$\mathbb{E}(\cdot)_\star$	The expectation operator with respect to the distribution of $\star$
$\text{extr}_\star(\cdot)$	The extremization of a function $\cdot$ with respect to $\star$
$Tr(\cdot)$	The trace of a matrix
$R(\cdot)$	R-transform
$R'(\cdot)$	Derivative of R-transform
$\Re\{\cdot\}$	Real part of $\{\cdot\}$
$G(\cdot)$	Green's function
J_0	Zero order Bessel function of the first kind
f_d	The Doppler frequency
$\mathcal{Z}$	Partition function
$\mathcal{S}$	Set of ground states of a physical system

$\mathcal{H}$	Hamiltonian of a physical system
$\mathcal{E}$	Energy of a physical system
$\mathcal{P}$	Entropy of a physical system
$\bar{\mathcal{E}}$	Average energy of a physical system
$\mathcal{F}$	Free energy of a physical system
$\bar{\mathcal{F}}$	Average free energy of a physical system
β	Inverse of the temperature of a physical system
$\mathbf{Q}$	Correlation matrix of the interacting states
$\tilde{\mathbf{Q}}$	Correlation matrix of the interacting states in the Fourier domain
$\mathbf{x}^a$	Replicated sparse vector for $a = 1, 2, \cdots, n$,
$\mathcal{Z}_a$	Partition function over replicated vectors
$\mathbf{Q}_{ab}$	Covariance of replicated vectors
$q_i, b_i, e_i,$ f_i, p_1, μ_1 $Q, \chi, m, \hat{Q}, \hat{\chi}, \hat{m}$	Microscopic variables of a physical system for $i = 0, 1$
$\mathbf{D}_i$	Denote the set of closest neighbours of x_i
$\mathbf{D}_i(\cdot)$	Clustering function x_i
$\hat{\mathbf{x}}_{\mathrm{LASSO}}$	Estimator based on LASSO
$\hat{\mathbf{x}}_{\mathrm{CLASSO}}$	Estimator based on modified LASSO
$\mathbf{H}[n]$ n	Channel matrix with channel feedback time index
η	Temporal correlation coefficient
N_t	Number of transmit antennas
N_r	Number of receive antennas
$\mathbf{U}[\cdot]$	Decoder matrix (a unitary matrice)
$\mathbf{V}[\cdot]$	Precoder matrix (a unitary matrice)
$\hat{\mathbf{V}}[\cdot]$	Quantized version
$\mathbf{R}[\cdot]$	Generalized receiver matrix

Σ	Diagonal matrices
$\mathbf{T}[\cdot], \hat{\mathbf{T}}[\cdot]$	A matrix of singular vectors and its estimate, respectively
$\Theta[\cdot], \hat{\Theta}[\cdot]$	A rotation matrix and its estimate, respectively
$vec(\cdot)$	Column vectorization of a matrix
$\mathbf{I}[\cdot]$	Identity matrix
$\mathcal{N}()$	Normal distribution
$\mathcal{CN}()$	Circularly-symmetric normal distribution
$exp(\cdot)$	Exponential function
$\delta(\cdot)$	Dirac delta function function
min	Minimum operator
max	Maximum operator

List of Figures

List of Tables

Part I

Introduction

Chapter 1

Introduction

This thesis deals with Compressive Sensing (CS), a new methodology to capture signals at a lower rate than the Nyquist sampling rate when the signals are sparse or sparse in some domains [1, 2, 3]. CS has recently gained a lot of attention due to its exploitation of signal sparsity. Sparsity, an inherent characteristic of many natural signals, enables the signal to be stored in a few samples and subsequently be recovered accurately [4]. In CS, encoding is done via random projection (RP), while at the decoding part nonlinear reconstruction algorithms are used to reconstruct the original signal from a reduced number of measurements. i.e., if the length of the original signal is N and one tries to recover it from M measurements, the traditional wisdom suggests that $M \geq N$. This principle is the basis for most devices used in current technology, such as analog-to-digital conversion, medical imaging, radar, and mobile communication [5].

However, if $M \ll N$, then classical linear algebra indicates that there are infinitely many solutions. It is impossible to recover the original signal (i.e., uniquely) from the reduced amount of information without additional information. This problem of a limited number of measurements can occur in multiple scenarios, e.g. when we have limitations on the number of data capturing devices, measurements are very expensive or slow to capture such as magnetic resonance imaging (MRI). Therefore, the study on the performance of the reconstruction schemes and on the possible application of CS is a hot topic. This chapter introduces the main concepts, theoretical frameworks and tools used in the thesis. In addition, the background, the scope and the contribution of the thesis are provided [4], [5].

In Section 1.1 CS is introduced. This section begins by defining the basic concepts and terms used in CS theory. It focuses on the reconstruction algorithms used and their performance due to the emphasis given to it later in this thesis. In Section 1.2, the Bayesian framework is introduced. This framework helps to connect CS and a tool from statistical physics, called the replica method (RM), presented in Section 1.3. This provides an elegant tool to analyze the performance analysis of large size physical systems. Specifically, a non-noisy CS problem is considered and its performance is provided analytically using the replica method. The Bayesian framework also serves as a basis for the later chapters, where the special structure found in sparse entries is included into the inference as additional prior information. In Section 1.4, a short description of CS applications is given. Finally, in Section 1.5, the scope and the contribution of the thesis are provided.

1.1 Compressed Sensing

In traditional signal processing, it is customary to represent a signal after sampling it at the Nyquist rate so that information can be recovered exactly. The process of sampling, which is the process of converting a continuous signal into a numeric sequence, is most demanding and can be expensive at times. Think about taking pictures or images using a high megapixel camera, which take so many samples of the object. According to the present signal-processing paradigm, the sampled data are compressed and stored on a storage device (sent through a medium) and later the data are decompressed and processed further, see Figure 1.1. During compression most of the data are then discarded and only some are stored (or transmitted).

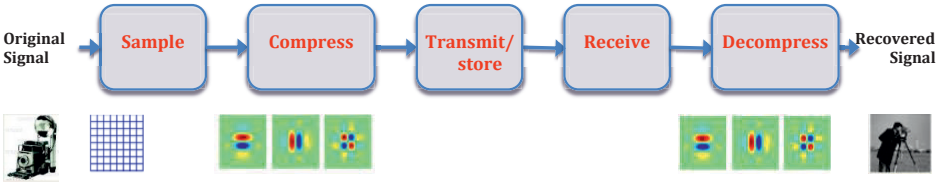

Figure 1.1: Traditional Signal Processing Paradigm.

However, recently questions such as *why go to so much effort to acquire all the data when most of what we get will be thrown away? Can we just directly measure the part that will not end up being thrown away?* that were posed by Donoho and others, triggered a new way of sampling (sensing) called compact ("compressed") sampling (sensing) (CS) [1, 2, 3]. CS combines the sampling and compression into one step, as depicted in Figure 1.2, by measuring few samples that contain maximum information about the signal: this eliminates the need to acquire and store a large number of samples [4].

Actually, CS has been used by Claerbout and Muir in Seismology [6], Rudin et al. in denoising [7] and many others before 2004, during which a revival of CS is witnessed. Since then there are many hundreds of papers produced in a short period of time (see http://www.compressedsensing.com or http://dsp.rice.edu/cs).

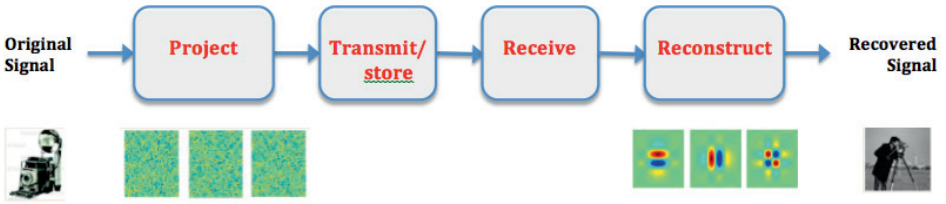

Figure 1.2: Compressed Sensing Paradigm.

CS builds upon the fundamental fact that we can represent many signals using only a few non-zero coefficients in a suitable basis or dictionary [8]. It assumes that the dimension of signals that are known to be sparse in some domain/basis can be reduced and that it is possible to accurately recover the original signal from the compressed data with high probability and efficient algorithms

[9]. Image and sound signals are some of the natural signals that can be well approximated using a sparse representation in some domain.

In order to elaborate on this, let us take a known example, Lena's image, an image typically used in image analysis, a two-dimensional signal. We first transform the image using discrete cosine transform (DCT) and Fast Fourier transform (FFT). Then by using only some percentages (P) of coefficients in the new basis, we reconstruct back the original image and the result is shown in Figure 1.3 for different P. Thus, most of the coefficients of a typical image can be thrown away without a significant loss in the quality of the image. In other words, the image is nearly sparse in these domains.

Figure 1.3: Transforming Lena's Image using DCT and FFT and approximating the image obtained by keeping only the largest P=30%, P=20%, and P=10% of the DCT and Fourier transform coefficients.

1.1.1 Fundamental Concepts in CS

In a CS framework signals are acquired or measured (sensed) using linear functionals. Consider a natural signal $f(t)$. We can obtain information about this signal using inner products of functions

$$y_i = \int f(t)\psi_i(t)dt, \qquad i = 1, \cdots, M, \tag{1.1.1}$$

where $\psi_i(t)$ are the basis functions or sensing waveforms which are usually suitable to a particular application and y_i are called measurements. It is customary in modern signal processing to model signals as vectors living in an appropriate vector space. If for example $f(t)$ is a band limited signal with period T, then we describe it as a discrete-time signal $\mathbf{f} = [f_1, f_2, \cdots, f_N]^T$, where $f_n \equiv f(nT)$. Hence, $\mathbf{f} \in \mathbb{R}^N$ represents discrete-time signals. In this thesis we consider signals

as vectors and focus on under-sampled situations in which the number of measurements, M, is smaller than the dimension of the signal $\mathbf{f}$, $M \ll N$. In such cases, different questions can arise [10]:

- *Is it possible to reconstruct the signal* $\mathbf{f}$ *accurately from* $M \ll N$ *measurements?*
- *Is it possible to design* $M \ll N$ *sensing waveforms to capture almost all the information about* $\mathbf{f}$*?*
- *And how can one approximate* $\mathbf{f}$ *from this information?*

These questions will be addressed partly in this subsection. Note that this is a brief introduction and if one needs more information, then reference [8] can be a good start.

Some signals can be sparse in nature, while others can have sparse representation where most of their energy is concentrated in only a small number of significant coefficients. Let $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ be a *sensing matrix* such that $\mathbf{y} = \mathbf{\Phi f}$ and further let $\mathbf{\Psi} \in \mathbb{R}^{N\times N}$ be a sparse *representation matrix/basis* such that $\mathbf{f} = \mathbf{\Psi x}$, where $\mathbf{x}$ is a sparse vector of coefficients. Putting these together we get $\mathbf{y} = \mathbf{\Phi f} = \mathbf{\Phi\Psi x}$.

Denoting the product of the sensing and the representation matrix by $\mathbf{A} = \mathbf{\Phi\Psi}$, we call it a *measurement matrix*. Then the measurement equation is given by

$$\mathbf{y} = \mathbf{Ax}, \tag{1.1.2}$$

where the matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$ is a rank deficient with $M \ll N$ which may cause loss of information in such frameworks. Note that, the three variables are central in CS: the sparse vector $\mathbf{x}$, the measurement matrix $\mathbf{A}$, and the measurement vector $\mathbf{y}$.

There are two fundamental concepts underlying the CS paradigm. The first is *sparsity*. If $\mathbf{x}$ is a sparse vector, i.e. if only $k \ll N$ of its entries are non-zero, then we call such vectors as a k-sparse vector. Its zero norm is defined as follows.
Definition 1. *(Zero Norm)*

$$\|\mathbf{x}\|_0 = k \equiv \#\{i \in \{1, 2, \cdots, N\} | x_i \neq 0\}. \tag{1.1.3}$$

Basically, (1.1.3) *is not a true norm since it does not satisfy triangle inequality.*

The second is *inchoherence*. Consider the two matrices above $\mathbf{\Phi}$, and $\mathbf{\Psi}$. $\mathbf{\Phi}$ is used as sensing the object $\mathbf{f}$ and $\mathbf{\Psi}$ is used to represent $\mathbf{f}$ [1], [10]. Then the coherence of the two matrices is defined as follows.

Definition 2. *(Coherence)*
The coherence between the sensing matrix $\mathbf{\Phi}$ *and the representation matrix* $\mathbf{\Psi}$ *is*

$$\mu(\mathbf{\Phi}, \mathbf{\Psi}) = \sqrt{N} \cdot \max_{1\leq j,k\leq N} |\langle \mathbf{\Phi}_j, \mathbf{\Psi}_k \rangle|, \tag{1.1.4}$$

where $\mathbf{\Phi}_j$ *and* $\mathbf{\Psi}_k$ *are the* j^{th} *and* k^{th} *columns of the sensing and the representation matrices, respectively.*

Coherence is therefore a measure of the largest correlation between any two vectors of the matrices. And low coherence pairs are suitable for CS. One particular example for such pairs is if $\mathbf{\Phi}$ is fixed and $\mathbf{\Psi}$ is a random matrix, like independent identically distributed (i.i.d) (example Gaussian, binary) entries exhibit a very low coherence with high probability with any fixed representation $\mathbf{\Psi}$. The reader is encouraged to refer to [11] for more examples of such pairs of bases. With this concepts at hand, let us define the CS problem formally as follows:

Definition 3. *(The standard CS problem)*
Find the k-sparse signal vector $\mathbf{x} \in \mathbb{R}^N$ provided the measurement vector $\mathbf{y} \in \mathbb{R}^M$, the measurement matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ and the under-determined set of linear equations as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}, \tag{1.1.5}$$

where $\mathbf{w} \in \mathbb{R}^M$ represents some measurement noise and generally, $k \ll M \ll N$.

One can ask again two of the questions raised, on page 4, in relation to the standard CS problem. First, how should we design the matrix $\mathbf{A}$ to ensure that it preserves the information in the signal $\mathbf{x}$? Second, how can we recover the original signal $\mathbf{x}$ from measurements $\mathbf{y}$ [8]? To address the first question, the solution for the CS problem presented here is dependent on the design of $\mathbf{A}$. This matrix can be considered as a transformation of the signal from the signal space to the measurement space, Figure 1.4 [9]. In the seminal paper [12], the authors defined the sufficient condition that matrix $\mathbf{A}$ should satisfy for the reconstruction of the signal $\mathbf{x}$. It is called the Restricted Isometric Property (RIP) and it is defined next.

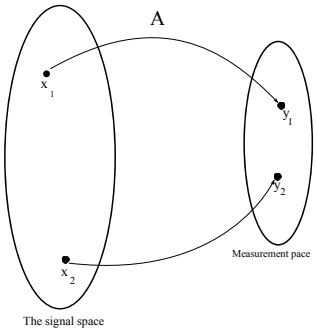

Figure 1.4: Transformation from the signal-space to the measurement-space.

Definition 4. *(Restricted Isometry Property)*
For all $\mathbf{x} \in \mathbb{R}^N$ so that $\|\mathbf{x}\|_0 \leq k$, if there exists $0 \leq \delta_k < 1$ such that

$$(1 - \delta_k)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_k)\|\mathbf{x}\|_2^2 \tag{1.1.6}$$

is satisfied, then $\mathbf{A}$ fulfills RIP of order k with radius δ_k.

An equivalent description of the RIP is to say that all subsets of k columns taken from $\mathbf{A}$ are nearly orthogonal (the columns of $\mathbf{A}$ cannot be exactly orthogonal since we have more columns than rows) [11]. For example, if a matrix $\mathbf{A}$ satisfies the RIP of order $2k$, then we can interpret (1.1.6) as saying that $\mathbf{A}$ approximately preserves the distance between any pair of k-sparse vectors.

For random matrix $\mathbf{A}$ the following theorem is one of the results in relation to RIP for the noiseless CS problem given by the researchers who revived the CS paradigm, provided that the entries of the random matrix $\mathbf{A}$ are drawn from some distributions which are given later in this subsection.

Theorem 1. *(Perfect Recovery Condition, Candes and Tao [2])*
If $\mathbf{A}$ satisfies the RIP of order $2k$ with radius δ_{2k}, then for any k-sparse signal $\mathbf{x}$ sensed by $\mathbf{y} = \mathbf{A}\mathbf{x}$, $\mathbf{x}$ is with high probability perfectly recovered by the ideal program

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{arg\ min}\ \|\mathbf{x}\|_0 \quad subject\ to\ \ \mathbf{y} = \mathbf{A}\mathbf{x} \tag{1.1.7}$$

and it is unique.

The complete proof of most of the theorems in this section are given in the literature mentioned with respect to each of them and an interested reader is referred to them. In order to ease the introduction only a few proofs are considered. Noticing this, let us return to the concept at hand, RIP. If $\mathbf{A}$ satisfies the RIP of order k with radius δ_k, then for any $k' < k$, $\mathbf{A}$ satisfies the RIP of order k' with constant $\delta_{k'} < \delta_k$ [13]. Another important implication of RIP can be exposed considering the concept of stability (Definition 5 and theorem 2 given below).

Definition 5. *(Definition 1.4 of [8])*
Let $\mathbf{A}$: $\mathbb{R}^M \to \mathbb{R}^N$ denote a measurement matrix and $\Delta : \mathbb{R}^M \to \mathbb{R}^N$ denote a recovery algorithm. We say that the pair ($\mathbf{A}$, Δ) is a C-stable if for any k-sparse $\mathbf{x}$ and any error $\mathbf{w} \in \mathbb{R}^M$ we have that

$$\|\Delta(\mathbf{A}\mathbf{x} + \mathbf{w}) - \mathbf{x}\|_2 \leq C\|\mathbf{w}\|_2. \tag{1.1.8}$$

Theorem 2. *(Stability) (Theorem 1.3 of [8])*
If the pair ($\mathbf{A}, \Delta$) is C-stable, then

$$\frac{1}{C}\|\mathbf{x}\|_2 \leq \|\mathbf{A}\mathbf{x}\|_2 \tag{1.1.9}$$

for all k-sparse $\mathbf{x}$.

Proof. Since the role of such constants is visible in the introduction part of the thesis the proof of this theorem is provided in Appendix A.1. □

According to definition 5 if we add a small amount of noise to the measurements, then the impact of this on the recovered signal cannot be arbitrarily large, whereas theorem 2 asserts that the existence of a stable decoding algorithm requires that the RIP lower bound in (1.1.6) is satisfied by $\mathbf{A}$ with a constant determined by C [8].

Further, one may ask how one can construct sensing matrices that satisfy the RIP property. There are many matrices that fulfil this condition if $M = O(k^2 \log N)$ number of measurements is available [10], which in a real life problem can be large and unacceptable. Fortunately random sensing enters into the picture here. In fact, random matrices $\mathbf{A}$ of size $M \times N$ whose entries are independent and identically distributed (i.i.d) will satisfy the RIP with high probability if the entries are chosen according to a Gaussian, Bernoulli, or more generally any sub-Gaussian distribution [8]. Let us take particular examples:

1. If $\mathbf{A}$ is drawn from a Gaussian distribution with i.i.d. entries and normalized columns or if $\mathbf{A}$ is formed by sampling i.i.d. entries from a symmetric Bernoulli distribution with the number of measurements
$$M \geq C_1 k \log(\frac{N}{k}), \tag{1.1.10}$$
then with high probability, $\mathbf{A}$ satisfies the RIP condition with radius δ_k[10].

2. If a matrix $\mathbf{A}$ is chosen according to a sub-Gaussian distribution with
$$M = O\Big(k^2 \log(N/k)/\delta_k^2\Big), \tag{1.1.11}$$
then $\mathbf{A}$ satisfies the RIP of order $2k$ with probability at least $1 - 2\exp(-C_2\delta_k^2 M)$,

where C_1 and C_2 are some constants dependent only on δ_k.

While the RIP provides guarantee for the recovery of k-sparse signals, verifying that a matrix $\mathbf{A}$ satisfies this property has a combinatorial computational complexity, since one must consider $\begin{pmatrix} N \\ k \end{pmatrix}$ sub-matrices [8]. Instead, one can use another property of $\mathbf{A}$ called, coherence of a matrix, the definition is below as in [14] , which has far less complexity than verifying that $\mathbf{A}$ fulfils RIP and it can provides guarantee of recovery as shown below.

Definition 6. *(Coherence of a matrix, Donoho and Huo [14])*
The coherence of a matrix $\mathbf{A}$*,* $\mu(\mathbf{A})$*, is the largest absolute inner product between any two columns, which are assumed to be normalized,* $\mathbf{a}_i$*,* $\mathbf{a}_j$ *of* $\mathbf{A}$*:*

$$\mu(\mathbf{A}) = \max_{1 \leq k,j \leq N} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle|. \tag{1.1.12}$$

In fact it is possible to relate the RIP property and the coherence by the following lemma.

Lemma 1. *(Lemma 1.5 of [8])*
If $\mathbf{A}$ *has a unit-norm columns and coherence* $\mu = \mu(\mathbf{A})$*, then* $\mathbf{A}$ *satisfies the RIP of order* k *with radius* $\delta_k = (k-1)\mu$ *for all* $k < 1/\mu$*.*

Note that, there are other properties of $\mathbf{A}$ that can guarantee the recovery of k-sparse signals. For example, the spark of a matrix, the null space property of a matrix, which is discussed well in literature [8], but not considered here for the sake of minimizing the scope of the introduction. Instead the RIP is chosen to serve as a springboard for the next chapter in the thesis.

It should also be mentioned that building a hardware that does random sensing is an on going research activity and so far only some results are reported, like the random demodulator [15], random filtering [16], random convolution[17], the compressive multiplexer [18]. Therefore, the work in this thesis is rather focused on theoretical results.

1.1.2 Signal Reconstruction

So far we have focused on the conditions that guarantee the recovery but not on the recovery itself which is the second question posed earlier, and one of the central concerns of the thesis. As

a matter of fact, it is the most difficult part of CS-based signal processing. In this pradgim one transfers most of the complexity from encoding to the decoding part. The reason is that, in CS the encoding part is very simple, i.e., one just multiplies the sparse vector $\mathbf{x}$ by random matrix $\mathbf{A}$ of size $M \times N$, where $M \ll N$ and later in the process noise $\mathbf{w}$ is added to it. However, the decoding part is very cumbersome but possible to deal with, with some degree of precision and complexity.

Starting with the noiseless CS problem, i.e., based on measurement matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ and measurement vector $\mathbf{y} \in \mathbb{R}^M$ with $M \ll N$, the objective is to re-construct or recover the sparse vector $\mathbf{x} \in \mathbb{R}^N$, which is related through the linear transform $\mathbf{y} = \mathbf{A}\mathbf{x}$. This can be clarified easily by the diagrammatic scheme as shown in Figure 1.5. If one tries to solve this problem using basic linear algebra knowledge, then the solutions are infinitely many since matrix $\mathbf{A}$ is rank deficient. However, if we assume that $\mathbf{x}$ is k-sparse vector with $k \ll N$ and use this additional information in solving the problem, we can get a unique solution.

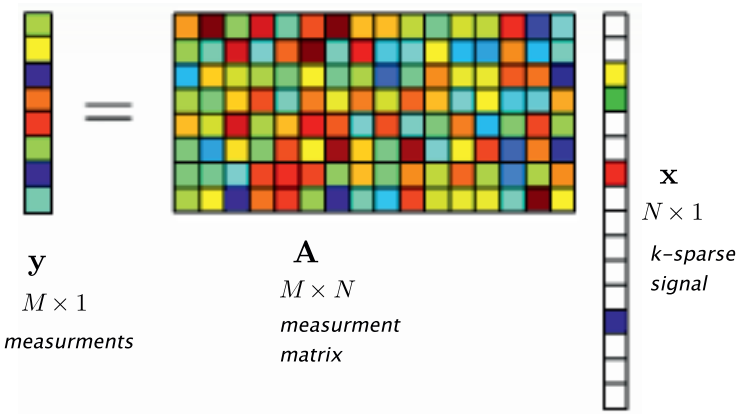

Figure 1.5: Multiplication of a random 'rectangular' matrix $\mathbf{A}$ with a sparse vector $\mathbf{x}$ resulting a measurement vector $\mathbf{y}$ [19].

The best solution for this problem will be the sparsest vector that satisfies the condition $\mathbf{y} = \mathbf{A}\mathbf{x}$. As defined before in (1.1.3), the mathematical representation that counts the non zero entries of the vector $\mathbf{x}$ is the l_0-norm,$\|\mathbf{x}\|_0$. The best solution can thus be found using this norm. Actually, the solution has been mentioned in Theorem 1 before, in connection with RIP.

Recently, there has been revival in the CS literature due the emergence of two categories of algorithms to approximate the solution for the CS problem. These are called the *convex relaxation* methods and *greedy pursuits* (GP). Among the GP are algorithms like Subspace Pursuit (SP) [20], Orthogonal Matching Pursuit (OMP) [21], Compressive Sampling Matching Pursuit (CoSaMP)[13], Fast Bayesian Matching Pursuit (FBMP) [22]. All GP algorithms are iterative signal recovery methods. One calculates the support of the signal and it makes the locally optimal choice at each time to build up an approximation. This is repeated until the criterion is fulfilled. Another of the same kind is Iterative Hard Thresholding(IHT). IHT initializes with zero vector and iterates a gradient descent step followed by hard thresholding until a convergence criterion is met [23]. These algorithms are very fast but they have limitations in stability.

In this thesis the focus is rather on the first categories or families of methods of approximating the l_0-norm based solution. Therefore, the convex relaxation method is presented starting from the noiseless CS problem solution.

Let us first summarize this solution. The noiseless CS problem reduces to finding the solution for the optimization problem:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0, \quad \text{such that } \mathbf{y} = \mathbf{A}\mathbf{x} \tag{1.1.13}$$

However, minimizing l_0-norm is a non-convex optimization problem which is NP-hard [24]. By relaxing the objective function to convexity, it is possible to get good approximation. That is, replacing the l_0-norm by the l_1-norm, one can find a problem which is tractable. Note that it is also possible to use other l_p-norms to relax the condition given by l_0. However, keeping our focus on l_1-norm, consider the minimization problem instead of (1.1.13).

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1, \quad \text{such that } \mathbf{y} = \mathbf{A}\mathbf{x} \tag{1.1.14}$$

The solution of the relaxed problem (1.1.14) gives the same as that of (1.1.13) and this equivalence was provided by Donoho and Huo in [14].

Theorem 3. *($l_0 - l_1$ Equivalence)*
If $\mathbf{A}$ *satisfies the RIP of order* $2k$ *with radius* $\delta_{2k} < \sqrt{2} - 1$*, then*

$$\hat{\mathbf{x}} = \underset{x}{arg\,min} \|\mathbf{x}\|_1 \quad \textit{subject to } \mathbf{y} = \mathbf{A}\mathbf{x} \tag{1.1.15}$$

is equivalent to (1.1.7) and will find the same unique $\hat{\mathbf{x}}$*.*

Justified by this theorem, (1.1.14) is an optimization problem which can be solved in polynomial time and the fact that it gives the exact solution for the problem (1.1.13) under some circumstance has been one of the main reasons for the recent developments in CS. There is a simple geometric intuition on why such an approach gives good approximations. A unit ball in l_p-space of dimension N can be defined as

$$\mathcal{B}_p \equiv \{\mathbf{x} \in \mathbb{R}^N \colon \|\mathbf{x}\|_p \leq 1\}. \tag{1.1.16}$$

Unit balls corresponding to $p = 0$, $p = 1/2$, $p = 1$, $p = 2$, $p = \infty$, and $N = 2$, the balls are shown in Figure 1.6. Among the l_p-norms that can be used in the construction of CS related optimization problems, only those which are convex give rise to a convex optimization problem which is more feasible than the non-convex counter parts. That means l_p-norms with only $p \geq 1$ satisfy such a condition. On the other hand, l_p-norms with $p > 1$ do not favour sparsity, for example l_2-norm minimization tends to spread reconstruction across all coordinates even if the true solution is sparse. But l_1-norm is able to enforce sparsity. The intuition is that l_1-minimization solution is most likely to occur at corners or edges, not faces [4], [25]. That is why l_1-norm became famous for CS.

In CS literature, convex relaxation is presented as either l_2-penalized l_1-minimization called Basis Pursuit De-nosing (BPDN) [26] or l_1-penalized l_2-minimization called LASSO (Least Absolute Shrinkage and Selection Operator) [25], which are equivalent and effective in estimating a high-dimensional data. The later is given more emphasis in this thesis.

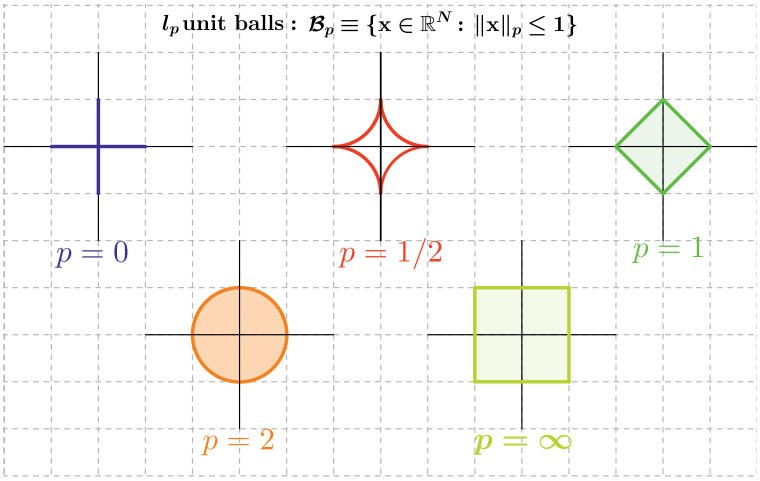

Figure 1.6: Different l_p-balls in different l_p-spaces for $N = 2$, only balls with $p \geq 1$ are convex.

Noisy Signal Recovery

So far recovery of signals from noiseless systems is discussed. It is great that these approximation methods work well for the non-noisy system. What happens to the convex relaxation methods discussed when noise is added to the observations? Are these algorithms or solutions robust enough to handle noise? Surprisingly, the relaxation method discussed before works well under some conditions [2]. Usually real world systems are contaminated with noise, $\mathbf{w}$, and in this thesis the focus is on such problems as given in Definition 3. Henceforth, our focus is on signal recovery in noisy systems. The noisy recovery problem becomes a simple extension of (1.1.14),

$$\min_x \|\mathbf{x}\|_1, \quad \text{such that} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon \tag{1.1.17}$$

where ϵ is a bound on $\|\mathbf{w}\|_2$. The real problem for (1.1.17) is stability. Introducing small changes in the observations should result in small changes in the recovery. We can visualize this using the balls shown in Figure 1.7. Both the l_0 and l_1-norms give exact solutions for the noise-free CS problem while giving a close solution for the noisy problem. However, the l_2-norm gives worst approximation in both cases compared to the other l_p-norms with $p < 2$ (see Figure 1.7). Here we should mention one to include a result from literature that describes the noisy signal recovery via l_1 minimization in connection with RIP condition as in [8]. In fact, the theorem gives a bound to the worst-case performance for uniformly bounded noise.

Theorem 4. *(Theorem 1.9 of [8] as in [11])*
Suppose that $\mathbf{A}$ *satisfies the RIP of order 2k with* $\delta_{2k} < \sqrt{2} - 1$ *and let* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$ *where* $\|\mathbf{w}\|_2 \leq \epsilon$. *Then the solution* $\hat{\mathbf{x}}$ *to (1.1.17) obeys*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C_0 \frac{\sigma_k(\mathbf{x})}{\sqrt{k}} + C_2\epsilon, \tag{1.1.18}$$

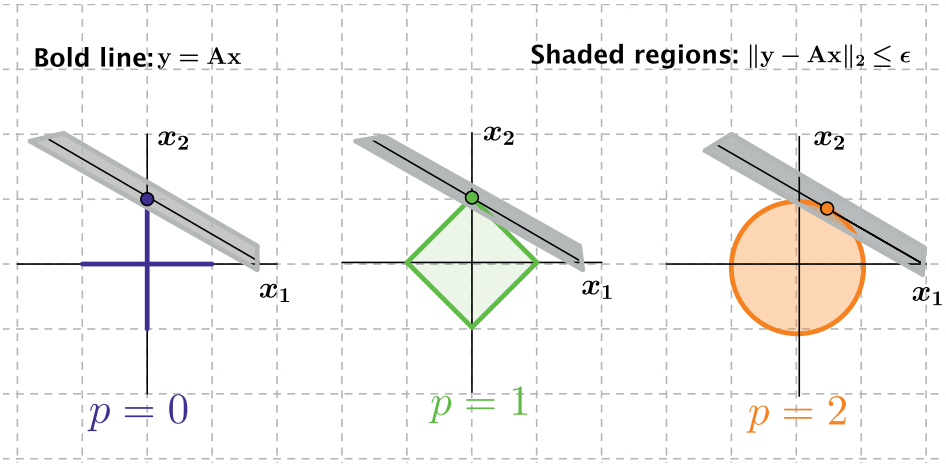

Figure 1.7: l_p-norm approximations: The constraints for the noise-free CS problem is given by the bold line while the shaded region is for the noisy one.

where

$$C_0 = 2\frac{1-(1-\sqrt{2})\delta_{2k}}{1-(1+\sqrt{2})\delta_{2k}}, \quad C_2 = 4\frac{\sqrt{1+\delta_{2k}}}{1-(1+\sqrt{2})\delta_{2k}} \tag{1.1.19}$$

Proof. Since the solution to the noisy CS problem is central in this thesis, the proof of this theorem is provided in Appendix A.2 . □

Further, (1.1.17) is equivalent to an unconstrained quadratic programming problem as

$$\min_x \ \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \gamma\|\mathbf{x}\|_1, \tag{1.1.20}$$

as it will be shown later as LASSO, where γ is a tuning parameter. The equivalency of (1.1.17) and (1.1.20) is shown in [22], [27]. In this thesis, the generalized form of the minimization problem in (1.1.20) with different l_p-norm regularization is considered, that is,

$$\min_x \ \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \gamma\|\mathbf{x}\|_p. \tag{1.1.21}$$

How well the algorithms, specifically in this thesis the convex relaxation methods, perform to solve the CS problem is another important investigation in CS literature. This is discussed next.

1.1.3 Performance of Recovery Algorithms

In subsection 1.1.2 we have mentioned two categories of algorithms that are used for recovery of CS signals. The first are the convex relaxation recovery algorithms while the others are the greedy pursuits (GP) algorithms. For an algorithm to be implemented in applications, its performance should be known clearly. In the literature there have been some general characteristics that describe the performance of an algorithm. Among the desired characteristics are such attributes

as stability, efficiency, accuracy and robustness. By stability, we mean that when the observed data are perturbed slightly by noise, recovery using the algorithm should still be approximately accurate. Efficiency is about how much information, in CS setting how many measurements, the algorithm needs to give a good approximation. While accuracy of a measurement system is the degree of closeness of measurements of a quantity to that quantity's true value. i.e., it should be free of inherent, systematic error. The reconstruction algorithm is called robust if the noise sensitivity is finite, and reconstruction is robust only when the measurement ratio (M/N), which is the ratio of the number of measurements to the signal length, exceeds a certain threshold.

Authors of the different algorithms proposed solutions to the CS problem argue that their algorithm is better than the others using some of these criteria. Some used mutual coherence as a measure of performance, others used RIP in connection to the ratio of the measurement M, and the sparsity k to the dimension of the signal, N, being estimated. The bounds vary from algorithm to algorithm. In that regard, there is no universal way of comparing the performance of the algorithms. In the following example, estimation of a sparse signal using LASSO and its performance evaluation via RIP and mutual coherence is illustrated.

Example 1. *(***Performance of LASSO** *[28])*
Consider the signal reconstruction problem presented in (1.1.17), that is the noisy CS problem as illustrated by Figure 1.8. Let the entries of $\mathbf{A}$ *be i.i.d. Gaussian with zero mean and variance* $1/M$ *and let the entries of* $\mathbf{w}$ *be i.i.d. Gaussian. If* $k = 10$ *and* $N = 10000$*, then how does LASSO perform?*

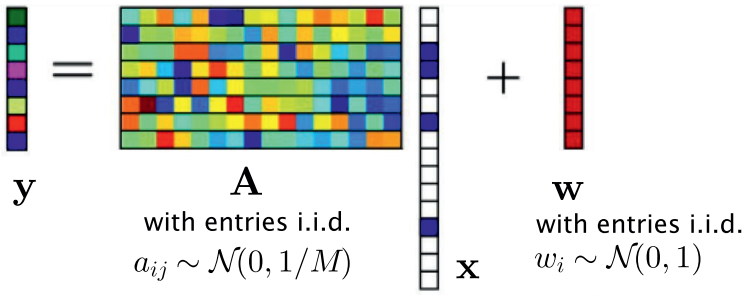

Figure 1.8: Noisy sparse signal estimation problem as given in the example (1) .

Solution 1. *(***RIP based performance analysis***)*
Based on the results from [12], for RIP-based analysis to hold we want $\delta_{2k} < \sqrt{2} - 1$ *and a sufficient condition for this is given by*

$$1 + \sqrt{2k/M} + \sqrt{(2N/M)h(2k/N)} < 2^{1/4}, \tag{1.1.22}$$

where $h(q)$ *is the entropy function* $h(q) := -q \log q - (1 - q) \log(1 - q)$ *defined for* $0 < q < 1$*. For example, for* $k = 10$ *and* $N = 10000$ *it is sufficient to have* $M \geq 7595$ *for recovery. There are other improvements on* M*, example* $M \geq 1801$ *in [29].*

Under-sampling is not justified for any $k/N > 0.0010$, in this particular example. Actually this also leads us to the concept of *phase transition* in CS literature. It is another central aspect in the CS paradigm [30, 31, 32].

Solution 2. *(**Coherence based performance analysis**)*
Based on the results from [33], $\mu(\mathbf{A})$ concentrates to at least $(2\log N)/M$ when $k < 1/(3\mu)$. Then the mean square error (MSE) of the recovered sparse signal follows

$$MSE_{lasso} \leq (\sqrt{3} + 3\sqrt{2\log(N-k)})^2 MSE_{OP}, \tag{1.1.23}$$

where MSE_{OP} is performance of oracle projector which is orthogonal projection to correct k-dimensional subspace. That is, for $k = 10$ and $N = 100$, $MSE_{lasso} \leq 115 \cdot MSE_{OP}$. This bound does not apply for $M < 277$.

It is possible to find many such examples of performance analysis in the literature, this is like the tip of an iceberg. Table 1.1 shows how performance comparison is done, using the RIP condition and complexity of an algorithm with SNR $\to \infty$, as $N \to \infty$, and minimum to avarage ratio (MAR) is defined by $\frac{\min |x_j|^2}{||\mathbf{x}||^2/k}$ [33]. Later in this thesis, other comparisons of algorithms using parameters like mean square error (MSE), signal to noise ratio (SNR) versus sparsity level or amount of measurement are given (see 1.3, and 4.5).

Table 1.1: Performance comparison of Algorithms [34], [33], [12] .

Algorithm	RIP condition for reliable reconstruction	Complexity
Maximum Likelihood	$M > k$	Very hard
LASSO	$M > 2k\ln(N-k)$	Moderate
OMP	$M > 2k\ln(N-k)$	Easy
IHT	$M > \frac{8}{\text{MAR}} k\ln(N-k)$	Very easy

Such kinds of analysis are not able to give us one optimal constant factor as a benchmark for assessing performance. Instead, as an option, a universal performance analysis for a given recovery algorithm in CS system is given by Gua et al. [35] by considering a large size system. They develop a single letter characterisation of the posterior of each individual entry of the sparse vector in the large system limit. They argue that their results often offer a better approximation for the performance of finite-size systems than the other existing bounds. This is done via tools which analyze complex systems in physics. In Chapter 2 performance analysis of CS system estimators using tools from statistical mechanics, specifically a tool called replica method, is presented. In Section (1.3), a simple introduction of replica method with an example from literature for the noiseless CS problem. Further, this method is used to analyze the performance of the estimators for the noisy CS problem is given in Chapter 2. To facilitate the use of such tools in a CS Bayesian framework a connector is needed between the CS estimators and the replica tools. The Bayesian framework is presented in the next section.

1.2 Bayesian Framework

There are two schools of thought in the field of statistics, called the classical or the frequentist and the Bayesian. The term "Bayesian" comes from the usage of Bayes' theorem, which was named after Reverend Thomas Bayes, an eighteenth century Presbyterian minister. The basic

difference between these two schools of thought arises from the basic definition of probability. Frequentists define P(A) as a long-run relative frequency with which an event A occurs in identical repeats of an experiment. A Bayesian defines P(A|B) as a real number measure of the probability of a proposition A, given the truth of the information represented by proposition B. So under the Bayesian school, probability is considered as an extension of logic and it can represent the investigator's degree of belief. Hence it is perceived as subjective according to the frequentists where probabilities are kept to be objective. To add on the differences, under the classical inference parameters are not random variable, they are fixed and prior information is absent. However, under the later parameters can be random variables, and prior information is an integral part, since for the Bayesian, inference is impossible with out assumption [36, 37, 38, 39, 40, 41, 42, 43].

The Bayesian framework has been used in different fields, including CS systems [41, 44, 45, 46]. In this thesis the Bayesian approach is used frequently as a main statistical framework. It acts as a bridge between CS theory and the Replica Method (the topic of the next subsection and Chapter 2), and it is also used as a main framework for analyzing CS system applications later in the thesis. This short introduction is therefore to facilitate its use in connection to CS systems.

Suppose we are interested in estimating a sparse vector $\mathbf{x}$ from measured data $\mathbf{y}$ by using a statistical model described by a probability density function (pdf), $p(\mathbf{y}|\mathbf{x})$. Bayesian philosophy states that $\mathbf{x}$ cannot be estimated exactly and for that the uncertainty about the parameter can be expressed using probability distributions. Further, under Bayesian inference the following three distributions are very crucial: (i) distribution of $p(\mathbf{x})$, which is known as the *prior distribution* or just *prior*. The prior expresses our beliefs about the parameter that we are going to estimate. (ii) the probabilities of the evidence that we get from the measurement data, for the CS case given by $\mathbf{y}$ can be represented by a statistical model $p(\mathbf{y}|\mathbf{x})$ to describe the distribution of $\mathbf{y}$ given $\mathbf{x}$, also called *the likelihood function* of $\mathbf{x}$, and (iii) to update our beliefs about $\mathbf{x}$ by combining information from the prior distribution and the data, we get a new distribution called *posterior distribution*, $p(\mathbf{x}|\mathbf{y})$, using Bayes' theorem and product rule:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\int p(\mathbf{y}|\tilde{\mathbf{x}})p(\tilde{\mathbf{x}})d\tilde{\mathbf{x}}}, \tag{1.2.1}$$

where $p(\mathbf{y})$ is called the normalizing constant of the posterior distribution. The posterior is directly proportional to the product of the likelihood function and the prior distribution:

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}). \tag{1.2.2}$$

Bayes' theorem helps to update the existing information about the variable we are interested in and Equation (1.2.2) is called *updating rule* [37, 47, 48], in which the data allows us to update our prior views about $\mathbf{x}$. As a result we get the posterior which combines both the data and non-data information of $\mathbf{x}$. As an example, for binomial trials, considering a beta distribution as a prior, gamma distribution for the data, we get posterior distribution which is also a beta distribution. Figure 1.9 shows that the posterior density is taller and narrower than the prior density. It therefore favours strongly a smaller range of $\mathbf{x}$ values, reflecting the fact that we now have less uncertainty about $\mathbf{x}$.

Theoretically, inference in the Bayesian method begins from the posterior distribution [35, 49] which is simple, though the calculation of it may be cumbersome. However, today the computational ability has increased tremendously and there are many algorithms (one example is the use of Markov Chain Monte Carlo (MCMC) methods) that can help to approximate it. It is thus possible

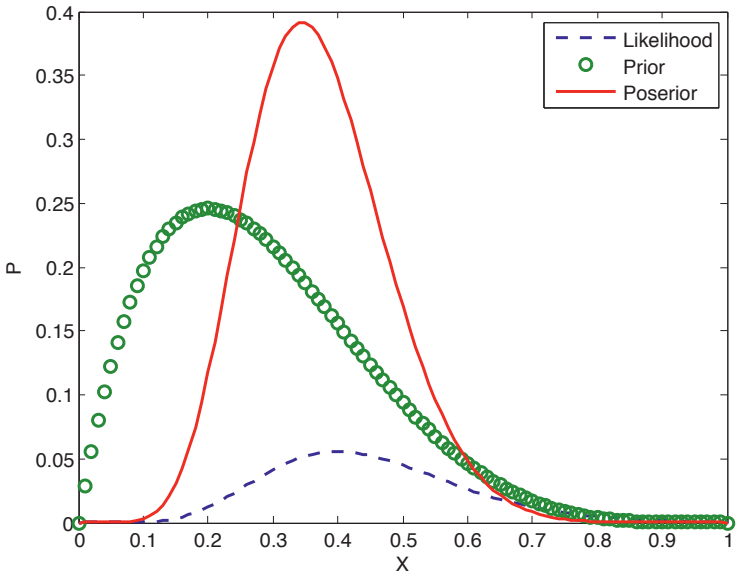

Figure 1.9: Figure showing the updating rule: The posterior synthesizes and compromises by favouring values between the maximum of the prior density and likelihood. The prior distribution we had is challenged to shift by the arrival of a little amount of data.

to use the Bayesian framework to approximate or analyze signals of our interest using the posterior distribution. Inference made using this distribution is indeed optimal [35].

In this thesis the usage of the posterior distribution has been crucial. Especially, we have used the maximum of this distribution called maximum a posterior (MAP) as a strategy to generalize the convex optimization-based solutions of the CS problem. Therefore, let us focus back on the CS problem given by definition 3 and represent the CS problems solutions given by (1.1.21) as MAP estimators [49].

Sparse Prior

We can represent our knowledge of the sparse signal $\mathbf{x}$ as prior knowledge [43, 48, 49]. That is, one can use a probability distribution $p(\mathbf{x})$ to model the sparsity of $\mathbf{x}$. Throughout the thesis, the prior knowledge of sparsity is modeled by pdfs that incorporate the l_p-norms. We restrict our attention where the prior pdf of a random variable $\mathbf{x}$ is given by

$$p(\mathbf{x}) = \frac{e^{-uf(\mathbf{x})}}{\int_{\mathbf{x}\in\mathbb{R}^N} e^{-uf(\mathbf{x})} d\mathbf{x}} \tag{1.2.3}$$

where $f : \chi \to \mathbb{R}$ is some scalar-valued, non-negative function with $\chi \subseteq \mathbb{R}$ called a regularizing function, which can be expanded to a vector argument by

$$f(\mathbf{x}) = \sum_{i=1}^{N} f(x_i), \tag{1.2.4}$$

such that for sufficiently large u, $\int_{\mathbf{x}\in\mathbb{R}^N} \exp(-uf(\mathbf{x}))d\mathbf{x}$ is finite. Further, let the assumed variance of the noise be given by

$$\sigma^2 = \frac{\lambda}{u},$$

where λ is the system parameter. Note that the prior, (1.2.3), is defined in such a way that it can incorporate the different estimators by assuming different penalizing terms via $f(\mathbf{x})$ [49]. Further, in (1.1.5), let us assume that the pdf of the noise $\mathbf{w}$ is Gaussian, such that the conditional pdf of $\mathbf{y}$ given $\mathbf{x}$ is

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \mid \mathbf{x}) = \frac{1}{(2\pi\sigma)^{N/2}} e^{-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2}. \tag{1.2.5}$$

Using (1.2.1), (1.2.3) and (1.2.5), the posterior pdf becomes

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}; \mathbf{A}) = \frac{e^{-u(\frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2+\lambda f(\mathbf{x}))}}{(2\pi\sigma)^{N/2}\int_{\mathbf{x}\in\mathbb{R}^N} e^{-u(\frac{1}{2\lambda}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2+\lambda f(\mathbf{x}))}d\mathbf{x}}. \tag{1.2.6}$$

Further, we find the maximum of the posterior distribution (MAP) which provides the most probable value for the parameters, denoted as $\hat{\mathbf{x}}_{MP}$, as

$$\hat{\mathbf{x}}_{MP} = \underset{x\in R^N}{\arg\min} \frac{1}{2} \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel_2^2 + \lambda f(\mathbf{x}). \tag{1.2.7}$$

(Detailed proof is given in A.3). Now, as we choose different regularizing function, we get the various types of CS estimators defined earlier in (1.1.21) [49]. These are

1. Linear Estimators: when $f(\mathbf{x}) = \parallel \mathbf{x} \parallel_2^2$, (1.2.7) reduces to

$$\hat{\mathbf{x}}_{\text{Linear}} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T + \lambda\mathbf{I})^{-1}\mathbf{y}, \tag{1.2.8}$$

   which is known as linear minimum mean square error (LMMSE) estimator.

2. LASSO Estimator: when $f(\mathbf{x}) = \parallel \mathbf{x} \parallel_1$ (1.2.7) reduces to

$$\hat{\mathbf{x}}_{\text{LASSO}} = \underset{x\in R^N}{\arg\min} \frac{1}{2} \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel_2^2 + \lambda \parallel \mathbf{x} \parallel_1, \tag{1.2.9}$$

   we get the LASSO estimator, which is the same as (1.1.20).

3. Zero-Norm regularization estimator: when $f(\mathbf{x}) = \|\mathbf{x}\|_0$,(1.2.7) reduces to

$$\hat{\mathbf{x}}_{\text{Zero-Norm}} = \underset{x\in R^N}{\arg\min} \frac{1}{2} \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel_2^2 + \lambda \parallel \mathbf{x} \parallel_0, \tag{1.2.10}$$

   we get the Zero-Norm regularization estimator.

The regularizing functions used in estimators 2 and 3 above enforce sparsity into the vector $\mathbf{x}$. However, the regularizing functions in estimator 1 have a smooth nature as shown in Figure 1.6 and it is hard to find a single, best solution for the CS problem. Therefore, l_2- regularization does not provide good approximation for the CS problem. The best solution for the CS minimization problem is given by (1.2.10), but it is NP-complete. However, it can be approximated by Equation (1.2.9) [3, 4, 9]. The performance of these estimators for the CS problem, assuming large size systems, $N, M \to \infty$, can be analyzed using tools from statistical mechanics. The next section 1.3 is devoted for such purpose.

Clustered Prior

It is also possible that the entries to the sparse vector $\mathbf{x}$ can have some special structure (clusteredness) among themselves. This in turn can be included in the prior pdf. Later in the thesis another prior distribution is given by

$$q(\mathbf{x}) = \frac{e^{-(\lambda\|\mathbf{x}\|_1 + \gamma D(\mathbf{x}))}}{\Delta}, \tag{1.2.11}$$

where λ and γ are tuning parameters for sparsity and clusteredness, respectively given that $D(\mathbf{x})$ is a function modelling the special structure among the sparse entries of $\mathbf{x}$, and Δ is a normalizing constant for the pdf $q(\mathbf{x})$.

1.3 CS Performance Analysis via Statistical Mechanics

High dimensional data can be difficult to model and process. One way to solve such problem is to reduce the dimensions of the data. However, how can we reduce it without losing too much important information? The recent advance in dimensionality reduction, that is, by randomly projecting (RP) the data into a low dimensional space has provided solution [50]. If a signal is sparse (compressible), then it is possible to RP it into lower dimensional data. The central idea of compressed sensing, discussed in Section 1.1, is that a sparse high dimensional signal can be recovered from a random projection down to a surprisingly low dimension by solving a computationally tractable convex optimization problem. Nonetheless, there is a limit to such a reduction of dimensionality. This means there will be a phase transition in the performance of CS, i.e., for any given level of signal sparsity, there is a critical lower bound on the dimensionality of a RP that is required to accurately recover the signal; this critical dimension decreases with increasing sparsity [51]. In this subsection, we are going to verify it using tools from statistical mechanics.

Statistical mechanics is a powerful field with elegant tools like *replica method*, *cavity method*, *gauge theory*, to analyze large size systems. In recent times, these tools have been applied to other fields, but only the replica method [2] is considered in this thesis. This choice was due to the project called Statistical Physics in Advanced Multiuser Systems (SPAM), in which excellent work [53, 54, 55] was done in analyzing large size MIMO systems using the replica method and the plan

[2] The Replica Method is a powerful approach for analyzing the statistical mechanics of systems with quenched disorder. However, some of the assumptions taken in the calculation are not yet rigorously proved [51, 52].

was to extend the results in that project to CS systems. The replica method is developed under the context of the Sherrington-Kirkpatrick (SK) model to analyze the random interactions between magnetic moments or spin glasses [36, 50, 51, 52, 56, 57]. Recently this method has been used in error correcting codes [58, 59, 60], CDMA [61, 62], MIMO systems [60, 63], and CS systems [35, 64, 65]. The central concepts and quantities involved in the replica method, that are relevant in the thesis, are defined in Appendix A.4. An interested reader is referred to [53, 54, 55, 66, 67] in addition to the references mentioned before.

Let us briefly discuss some results from the literature [51, 68, 69] on CS performance analysis based on statistical mechanics approach for large size(scale) systems, i.e., when $N, M \to \infty$ while keeping the measurement ratio M/N fixed. Consider the l_p-reconstruction schemes for the noiseless CS problem (deferring the noisy one for Chapter 2),

$$\min_x \|\mathbf{x}\|_p, \quad \text{such that } \mathbf{y} = \mathbf{A}\mathbf{x}, \tag{1.3.1}$$

for $p = 0, 1$ and 2. In order to give asymptotic analysis of the performance of such schemes, let us proceed with probabilistic approach to the CS problem, as was introduced in Section 1.2. Given a measurement vector $\mathbf{y}$ and a random measurement matrix $\mathbf{A}$ arising from a linear observation process expressed as $\mathbf{y} = \mathbf{A}\mathbf{x}^0$, the task in CS is to get approximations to the original sparse vector $\mathbf{x}^0$. Inferring probabilistically, it is equivalent to finding the estimation of $p(\mathbf{x}|\mathbf{y})$ given probabilistic information about $\mathbf{y}$ and $\mathbf{A}$. Further, if there is prior information about the original sparse vector, say $p(\mathbf{x})$, it is possible to use Bayes' theorem to estimate $p(\mathbf{x}|\mathbf{y})$ via $p(\mathbf{x}|\mathbf{y}) = (p(\mathbf{y}|\mathbf{x})p(\mathbf{x}))/\mathcal{Z}$, where $\mathcal{Z}$ is the *normalizing constant*. As an example, let us assume that the prior distribution of the entries of the sparse vector follow a distribution,

$$P(x) = (1-\rho)\delta(x) + \rho\xi(x), \tag{1.3.2}$$

where ρ is the sparsity ratio and $\xi(x)$ is the distribution of x. The posterior, $p(\mathbf{x}|\mathbf{y})$, becomes

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}) &= \frac{1}{\mathcal{Z}} p(\mathbf{x}) p(\mathbf{y}|\mathbf{x}) \\ &= \frac{1}{\mathcal{Z}} \prod_{i=1}^{N} p(x_i) \prod_{\nu=1}^{M} \delta(y_\nu - \sum_{i=1}^{N} A_{\nu i} x_i) \\ &= \frac{1}{\mathcal{Z}} e^{-\sum_{i=1}^{N} \log p(x_i) - \log\left(\delta\left(y_\nu - \sum_{i=1}^{N} A_{\nu i} x_i\right)\right)}, \end{aligned} \tag{1.3.3}$$

where x_i, y_ν, and $A_{\nu i}$ are entries of $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{A}$, respectively. Equation (1.3.3) can be interpreted as a mean-field disordered statistical mechanics problem, where the entries of $\mathbf{A}$, $A_{\nu i}$, give the disordered interactions (the connectivity) in the system, $\mathcal{Z}$ is the *partition function*, and the argument of the exponential term is the energy function or the *Hamiltonian* of the system [36, 51, 52, 56, 64, 68]. When $\xi(x) \sim \mathcal{N}(0, 1)$ and the l_p-reconstruction is considered the Hamiltonian, $\mathcal{H}(\mathbf{x})$, becomes

$$\mathcal{H}(\mathbf{x}) = -\log(\delta(\mathbf{y} - \mathbf{A}\mathbf{x})) + \|\mathbf{x}\|_p. \tag{1.3.4}$$

The Boltzman-Gibbs distribution of such a system with Hamiltonian, $\mathcal{H}(\mathbf{x})$, is given by

$$P_\beta(\mathbf{x}|\mathbf{y}) = \frac{1}{\mathcal{Z}} e^{-\beta \mathcal{H}(\mathbf{x})}, \tag{1.3.5}$$

where β is the *inverse of the temperature* of a physical system and it reflects sources of noise. For the CS system with the Hamiltonian given in (1.3.4), Equation (1.3.5) becomes

$$P_\beta(\mathbf{x}|\mathbf{y}) = \frac{e^{-\beta\|\mathbf{x}\|_p}\delta(\mathbf{y}-\mathbf{A}\mathbf{x})}{\mathcal{Z}(\beta;\mathbf{y})}, \tag{1.3.6}$$

where the *partition function*, $\mathcal{Z}(\beta;\mathbf{y})$, is given by

$$\mathcal{Z}(\beta;\mathbf{y}) = \int e^{-\beta\|\mathbf{x}\|_p}\delta(\mathbf{y}-\mathbf{A}\mathbf{x})d\mathbf{x}. \tag{1.3.7}$$

As $\beta \to \infty$, i.e., in the low-temperature limit of a physical system, Boltzman-Gibbs distribution converges to a uniform distribution of the ground states of such systems [68, 69]. That is,

$$\lim_{\beta\to\infty} P_\beta(\mathbf{x}|\mathbf{y}) \to \begin{cases} \frac{1}{|\mathcal{S}|}, & if \;\; \mathbf{x}\in\mathcal{S} \\ 0, & \text{otherwise,} \end{cases} \tag{1.3.8}$$

where $\mathcal{S}$ denotes the set of ground states such as the solutions of the optimization problem given in (1.3.1).

The statistical structure of high probability (low energy) is the main interest here. And it is done by understanding the statistical properties of the Boltzman-Gibbs distribution at a low temperature. Further, in order to understand the statistical mechanics terms which are used in the remaining parts of this section, the reader is encouraged to refer to Appendix A.4 and the literature mentioned therein. Some of these terminologies are *self averaing*, *free energy*, *replica method*, and *replica symmetry*. But, first let us introduce the performance metrics that will be calculated using these statistical mechanics tools.

Reconstruction Limit

Let $\hat{\mathbf{x}}$ be the recovered vector using the l_p-recovery schemes for the problem given in (1.3.1). Let the Mean Squared Error (MSE) per element defined by

$$\text{MSE}(\mathbf{x}^0, \mathbf{A}) = \frac{1}{N}\sum_{\mathbf{x}\in\mathcal{S}} \frac{1}{|\mathcal{S}|}|\mathbf{x}-\mathbf{x}^0|^2. \tag{1.3.9}$$

In turn, this MSE expression can be re-written as

$$\begin{aligned} \text{MSE}(\mathbf{x}^0, \mathbf{A}) &= \frac{1}{N}\langle|\hat{\mathbf{x}}-\mathbf{x}^0|^2\rangle \\ &= \frac{1}{N}\langle|\hat{\mathbf{x}}|^2\rangle - \frac{2}{N}\langle\hat{\mathbf{x}}\cdot\mathbf{x}^0\rangle + \frac{1}{N}\langle|\mathbf{x}^0|^2\rangle, \end{aligned} \tag{1.3.10}$$

where $\langle\cdot\rangle$ is averaging with respect to the Boltzmann-Gibbs distribution (1.3.5). That means there are three order parameters that determine the MSE. As $N \to \infty$, in the third term we have the sparsity ratio, $\rho = \frac{1}{N}|\mathbf{x}^0|^2$. Denoting the remaining two as

$$Q = \frac{1}{N}|\hat{\mathbf{x}}|^2, \quad and \quad m = \frac{1}{N}\hat{\mathbf{x}}\cdot\mathbf{x}^0, \tag{1.3.11}$$

the MSE is given in short as

$$\text{MSE}(\mathbf{x}^0, \mathbf{A}) = Q - 2m + \rho. \tag{1.3.12}$$

Note that if $\hat{\mathbf{x}} = \mathbf{x}^0$ [1] , then $Q = m = \rho$ and $\text{MSE}(\mathbf{x}^0, \mathbf{A}) = 0$.

[1]Note also our notation: $\mathbf{x}^0$ is the original sparse signal, $\mathbf{x}$ is its statistical realization, and $\hat{\mathbf{x}}$ is the recovered signal.

In this subsection since we do asymptotic performance analysis, it is assumed that $\alpha = M/N$ is fixed though $N, M \to \infty$. Now, let us define the performance metric based on the measurement ratio α and the sparsity ratio ρ called *reconstruction limit* [68, 70, 71].

Definition 7. *(Reconstruction limit)*
For a given ρ, a reconstruction limit $\alpha_c(\rho)$ is defined as

- *If* $\alpha \geq \alpha_c(\rho)$,
$$\mathbb{E}(\textit{MSE}(\mathbf{x}^0, \mathbf{A}))_{\mathbf{A},\mathbf{x}^0} = 0 \;\; \Leftrightarrow \;\; Q = m = \rho; \tag{1.3.13}$$
- *If* $\alpha < \alpha_c(\rho)$,
$$\mathbb{E}(\textit{MSE}(\mathbf{x}^0, \mathbf{A}))_{\mathbf{A},\mathbf{x}^0} = \textit{Constant} \;\; \Leftrightarrow \;\; Q \neq m \neq \rho, \tag{1.3.14}$$

where $\mathbb{E}(\cdot)_{\mathcal{X}}$ denotes averaging over the distributions of $\mathcal{X}$.

This critical compression rate, $\alpha_c(\rho)$, which will be calculated via tools from statistical mechanics soon, will help to find the *phase transition* for each l_p-reconstruction scheme. Hence, in the rest part of this subsection follows the replica analysis, where details are provided in Appendix A.4 including definitions of terminologies and relevant assumptions.

Replica Method

The Boltzman-Gibbs distribution $P_\beta(\mathbf{x}|\mathbf{y})$, and therefore its free energy $-\beta^{-1}\log \mathcal{Z}$ and the average error (MSE) are all dependent on the measurement matrix $\mathbf{A}$ and the sparse signal $\mathbf{x}^0$. Since both $\mathbf{A}$ and $\mathbf{x}^0$ are random variables, calculating these quantities is a difficult task. However, many interesting quantities (such as the free energy, MSE), which involve averages over all interacting entries, are *self averaging*. Hence, their fluctuations across different realizations vanish in the large system ($N \to \infty$, $M \to \infty$ and $\alpha = M/N$). In this thesis, we consider the entries of $\mathbf{A}$, $A_{ij} \sim \mathcal{N}(0, 1)$ and the entries of $\mathbf{x}^0$ as given by (1.3.2) and take their realization. Thus, $\mathbf{A}$ and $\mathbf{x}^0$ play the role of quenched disorder for the CS system at hand which is described by the distribution $P_\beta(\mathbf{x}|\mathbf{y})$.

Now, since the free energy is a self averaging quantity, it is given by

$$\mathcal{F} = -\beta^{-1}\mathbb{E}(\log \mathcal{Z}), \tag{1.3.15}$$

where the $\mathbb{E}(\cdot)$ operator is done over $\mathbf{A}$ and $\mathbf{x}^0$. However, calculating the expectation or average of the logarithm function is a difficult task and this lead us to the next important concept, the replica method. In order to ease the calculation we take n replicas of the partition function and apply an identity. That is

$$\mathbb{E}(\log \mathcal{Z}) = \lim_{n\to 0} \frac{\partial}{\partial n} \log \mathbb{E}(\mathcal{Z}^n). \tag{1.3.16}$$

After applying Equation (1.3.16) the avarage free energy is given by

$$\bar{\mathcal{F}} = -\lim_{n\to 0} \frac{\partial}{\partial n} \lim_{N\to\infty} \frac{1}{N\beta} \log \mathbb{E}(\mathcal{Z}^n). \tag{1.3.17}$$

For the l_p-reconstruction schemes of the noiseless CS problem, the key quantity that plays the role of a free energy in statistical mechanics and which represents the typical value (per element) of the minimized cost function (1.3.1) can be defined by

$$\mathcal{C}_p = -\lim_{\beta\to\infty} \frac{1}{\beta N} \mathbb{E}(\log \mathcal{Z}(\beta; \mathbf{y})), \tag{1.3.18}$$

as in [68, 70, 71]. Substituting (1.3.16) into (1.3.18) and interchanging the limits we get

$$\mathcal{C}_p = -\lim_{\beta\to\infty}\lim_{n\to 0}\frac{\partial}{\partial n}\lim_{N\to\infty}\frac{1}{\beta N}\log\mathbb{E}(\mathcal{Z}^n(\beta;\mathbf{y})). \tag{1.3.19}$$

In (1.3.19), we are taking the expectation of the product of n identical replicas of $\mathcal{Z}(\beta;\mathbf{y})$, i.e.,

$$\mathbb{E}(\mathcal{Z}^n(\beta;\mathbf{y})) = \mathbb{E}\Big(\prod_{a=1}^{n} Z_a(\beta;\mathbf{y})\Big). \tag{1.3.20}$$

and inserting it in (1.3.19), we get

$$\mathcal{C}_p = -\lim_{\beta\to\infty}\lim_{n\to 0}\frac{\partial}{\partial n}\lim_{N\to\infty}\frac{1}{\beta N}\log\mathbb{E}\Big(\prod_{a=1}^{n} Z_a(\beta;\mathbf{y})\Big), \tag{1.3.21}$$

where $Z_a(\beta;\mathbf{y})$ refers to the calculation of $\mathcal{Z}(\beta;\mathbf{y})$ using the replicated sparse vector $\mathbf{x}^a$. Further, to simplify (1.3.21) another assumption is needed, replica symmetry.

Replica Symmetry

The replica ansatz are arbitrary and often assumed to be independent random variables. This makes calculating the expectation of the right-hand side of (1.3.21) hard. However, it is possible to assume symmetry among the replica ansatz, called *replica symmetry* (RS) ansatz here. The RS assumption on the ansatz has been shown to be sufficient if the optimization problem under consideration is convex [36, 52, 53, 54, 55, 66, 67]. And the CS minimization problems, given by (1.3.1) for $p = 2$ and $p = 1$, is a convex optimization problem.

For any fixed realization of the quenched disordered $\mathbf{A}$, the replicated sparse vector $\mathbf{x}^a$ is independent and marginalizing (integrating) it over the disorder introduces attractive interactions between replicas [51]. The interactions between replicas depends only on the overlap matrix $\mathbf{Q}$, with entries given by

$$Q_{ab} = \frac{1}{N}\mathbf{x}^a\cdot\mathbf{x}^b, \tag{1.3.22}$$

where $a, b = 0, 1, 2, \cdots, n$. The landscape described by the overlap matrix $\mathbf{Q}$ is very complicated. It can however be simplified by assuming RS ansatz. One possible assumption can be $Q_{00} = \rho$, $Q_{0b} = \mathbf{x}^0\cdot\mathbf{x}^a = Nm$, $Q_{ab} = \mathbf{x}^a\cdot\mathbf{x}^b = Nq$ for $a \neq b$, and $Q_{aa} = \mathbf{x}^a\cdot\mathbf{x}^a = NQ$. These replica symmetry assumption together with their conjugate variables further simplify the calculation of $\mathcal{C}_p$, (1.3.21), and we get

$$\mathcal{C}_p = \text{extr}_\Theta\left\{\frac{\alpha(Q-2m+\rho)}{2\chi} + m\hat{m} - \frac{\hat{Q}Q}{2} + \frac{\hat{\chi}\chi}{2} + \Psi(Q,\chi,m,\hat{Q},\hat{\chi},\hat{m})\right\}, \tag{1.3.23}$$

where $\text{extr}_\Theta\{\mathcal{G}(X)\}$ denotes the extremization of a function $\mathcal{G}(X)$ with respect to X, $\Theta = \{Q,\chi,m,\hat{Q},\hat{\chi},\hat{m}\}$. For the detailed derivation of these quantities the reader is referred to Appendix A.4 and references [68] and [70]. The macroscopic quantities, $Q,\chi,m,\hat{Q},\hat{\chi},\hat{m}$, are all calculated with the saddle point integration.

The comparison of the reconstruction of the different l_p-norms is shown in Figure 1.10 using the critical compression rate $\alpha_c(\rho)$, definition (7), versus the signal density ρ.

Phase transition behaviour is exhibited as shown in Figure 1.10: the l_p -reconstruction schemes behave differently above and below the respective curves. In the region between the green and the

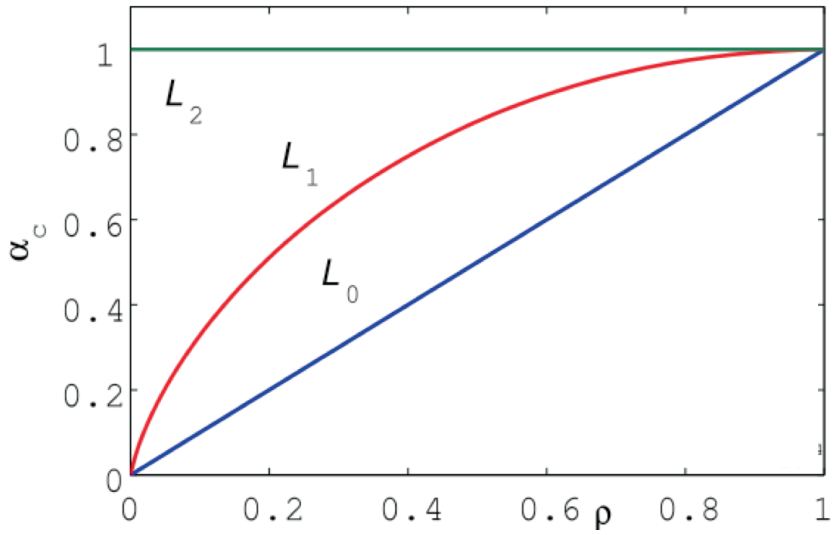

Figure 1.10: Performance of the l_p reconstruction schemes for $p = 0, 1$ and 2 using critical compression rate $\alpha_c(\rho)$ versus the signal density ρ [68, 70] .

red lines, it is possible to use the l_1 -reconstruction schemes. However in the region below the red line this scheme fails to provide good recovery, whereas in the region between the red and the blue line the l_0-reconstruction scheme provides the best solution. The region below the blue line is the region where reconstruction of the sparse signal is completely impossible. For further discussion on phase transition, [30] and [72] can be referred to.

Replica Symmetry Breaking

The assumption of replica symmetry ansatz is a very simplified model for some problems. If the local stability of the RS saddle point is lost against the perturbations that break it, then the solution fails to realize. The local instability of the RS solution is verified by the de Almeida-Thouless (AT) condition [52]. Therefore, replica symmetry breaking ansatz has to be used. This corresponds to the physical picture in which there are many free energy landscapes with many valleys. A good figurative explanation is provided in [51] (see Figure 1.11). For example, if we consider the CS minimization problem given by (1.3.1) for $p = 0$, it is a non-convex optimization problem, and the expression for the minimized l_p-norm per element, $\mathcal{C}_p$, cannot be calculated using the replica symmetry ansatz. In such cases the symmetry breaks. Therefore one has to use the replica symmetry breaking ansatz.

In chapter 2, we have used a one step replica symmetry breaking (1RSB) to analyze the performance of the l_p-norm reconstructing schemes for the noisy CS problem, which is different from the example we described above. This is done by transforming these schemes as MAP estimators using Bayesian framework. In general, the performance of MAP estimators in large systems can be analyzed well using 1RSB ansatz [55, 73, 74]. This was the main motivation to do the analytical work using 1RSB in Chapter 2 [65].

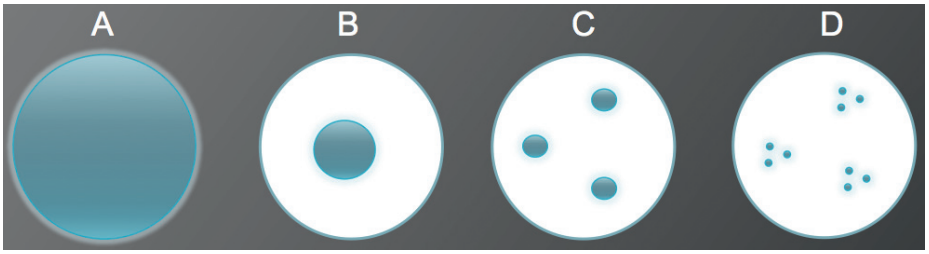

Figure 1.11: Probability lumps in free energy valleys. The larger circles represent the space of all possible sparse vectors. The shaded regions represent the space of configurations (free energy valleys) with non-negligible probability under the Boltzman-Gibbs distribution. (A) At high temperature all configurations are explored by this distribution. (B) The replica symmetric ansatz for a low temperture phase. The interaction between replicas freeze into small set of configurations. (C) One possible ansatz for replica symmetry breaking (RSB) in which the replica overlap matrix Q is characterized by two order parameters. (D) There exists a series of k-step RSB schemes describing scenarios in which the distribution decomposes into a nested hierarchy of lumps of depth k. Here k=2. (Provided by Advani et al. [50])

1.4 Compressive Sensing Applications

Compressive sensing is an exciting, rapidly growing field which has attracted considerable attention in electrical engineering, applied mathematics, statistics, and computer science [75]. This paradigm has been applied to many signal processing areas such as image processing, communication and networks. The signals of interest in these application areas have sparse representation in some signal representation. However, at this time building the hardware that can translate the CS theory into practical use is very limited. Nonetheless, the demand for cheaper, faster and efficient devices will motivate the use of the CS paradigm in real-time systems in the near future. So far we have few examples like the single-pixel cameras built from Rice University shown in Figure 1.12 [76], in magnetic resonance imaging (MRI) [77], and in few other applications.

Even though the hardware to apply the CS theory is in its early stage, the theory has been used in several fields in signal processing and a few of them are summarized here.

1.4.1 Compressive Imaging

In imaging, CS theory has a remarkable impact than other application areas in signal processing. It reduces the number of measurements, which in turn can reduce power consumption, computational complexity and storage without significant degradation in the spatial resolution. For example the single-pixel camera mentioned above [76], is comprised of optical computer, with a single photon detector, that computes random linear measurements. The camera design reduces the required size, complexity, and cost of the photon detector array down to a single unit, which enables the use of exotic detectors that would be impossible in a conventional digital camera. The random CS measurements also enable a tradeoff between space and time during image acquisition. In

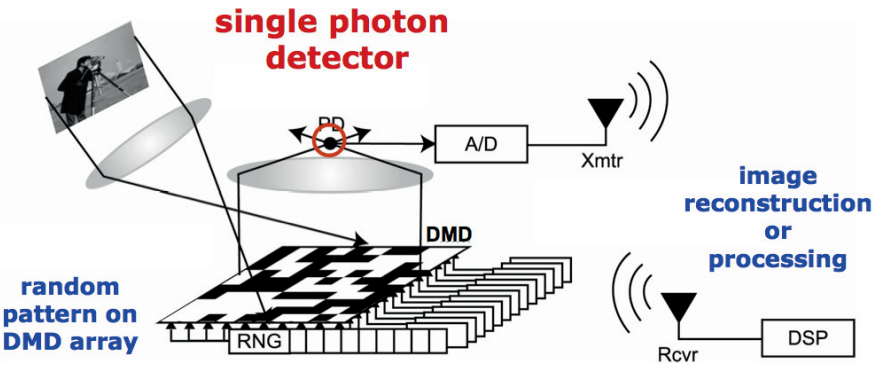

Figure 1.12: Rice Single-Pixel Camera Project, http://dsp.rice.edu/cscamera.

addition, since the camera compresses as it images, it has the capability to efficiently and scalably handle high-dimensional data sets from applications like video and hyperspectral imaging [78].

1.4.2 CS and Medical Imaging

CS has received a great deal of attention in medical imaging, especially in Magnetic Resonance Imaging (MRI), which is a costly and time consuming process because of its data collection process. However, the introduction of CS based techniques has improved the image quality through reduction in the number of collected measurements and by taking advantage of their implicit sparsity [4],[77]. Magnetic Resonance (MR) images, like angiograms, have sparsity properties, in domains such as Fourier or wavelet basis. The transform sparsity of MR images and the coded nature of MR acquisition are two key properties enabling CS in MRI. Figure 1.13 illustrates these elements, making MRI a natural CS system [4]. MRI is still a hot research area for CS theory [77], [79].

1.4.3 Compressive Radar

Radar imaging seems to be a very promising application of CS techniques for simplifying hardware design and to obtain a high resolution image [80, 81]. Standard methods for radar imaging actually also use the sparsity assumption, but only at the very end of the signal processing procedure in order to clean up the noise in the resulting image. Using sparsity systematically from the very beginning by exploiting compressive sensing methods is therefore a natural approach [75], [80].

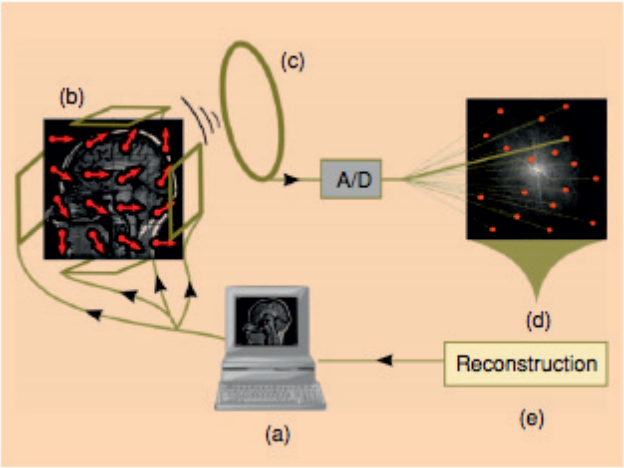

Figure 1.13: MRI as a compressed sensing system. The user controls the gradient and RF waveforms that, in turn, control the phase of the pixels/voxels in the image. An RF coil receives the signal in an encoded form samples in k-space. Careful crafting of the gradient waveforms allows for incoherent measurements of k-space. With an appropriate nonlinear reconstruction enforcing sparsity, an image can be reconstructed [77].

1.4.4 CS in Communication

In the communications community the application of compressive sensing has been mainly on sparse channel estimation for various types of channels, with extensions to multiuser and cognitive radio systems [82], [83]. A new channel estimation technique based on CS was also proposed to exploit the "delay-Doppler sparsity" of wireless channels for a reduction of the number of pilots required for channel estimation within multi-carrier systems [84]. CS-based sparse channel estimation has also been shown to achieve much less reconstruction error while utilizing significantly less energy and, in some cases, less latency and bandwidth as well [85]. The estimation of underwater acoustic channels, which are inherently sparse, through CS technique yields better results than the conventional 'Least Square Estimator' [86, 87].

1.4.5 CS in Wireless Sensor Networks

Wireless sensor networks (WSNs) are resources constrained by limited power supply, memory, processing performance and communication bandwidth. Due to their limited power supply, energy consumption is a key issue in the design of protocols and algorithms for WSNs. Compressed sensing and distributed compressed sensing have been adopted as potential approaches to provide energy efficient sensing in wireless sensor networks [75]. In another study, CS theory to sensor data gathering for large scale wireless sensor networks (WSNs) was first developed by considering the scenario in which a large number of sensor nodes are densely deployed and sensor readings are spatially correlated. This CS data gathering is able to reduce the global scale communication cost without introducing intensive computation or complicated transmission control [88]. By using

distributed compressed sensing for compression of the data in the network, the communication cost (bandwidth usage) to the sink was shown to decreased at the expense of delay induced by the local communication [89]. One can mention other applications of CS in WSNs like [90], [91], [92].

A good CS application literature review is provided in [4], which basically is the summary of the bulk of literature given at http://dsp.rice.edu/cs. The list of application areas include: analog-to-information conversion, computational biology, geophysical data analysis, hyperspectral imaging, astronomy, remote sensing, and computer engineering. It does not stop here, however note that, so far the clear progress in applying CS theory is in imaging, specifically in MRI imaging. This thesis also contributes to the list in the areas of imaging and communication [48, 93, 94, 95].

1.5 Scope and Contribution of the Thesis

This research work is about a new paradigm in signal processing called compressive sensing (CS) using the Bayesian framework. It consists of two main categories: performance analysis and application, where one paper, given in Chapter 2, in addition to Section 1.1.3, is dedicated to analytical performance analysis to CS recovery schemes and the other three chapters, 3, 4, and 5 provide applications of the CS theory in imaging (Chapter 3 and Chapter 4) and in communication (Chapter 5). Figure 1.14 shows the overview of the thesis.

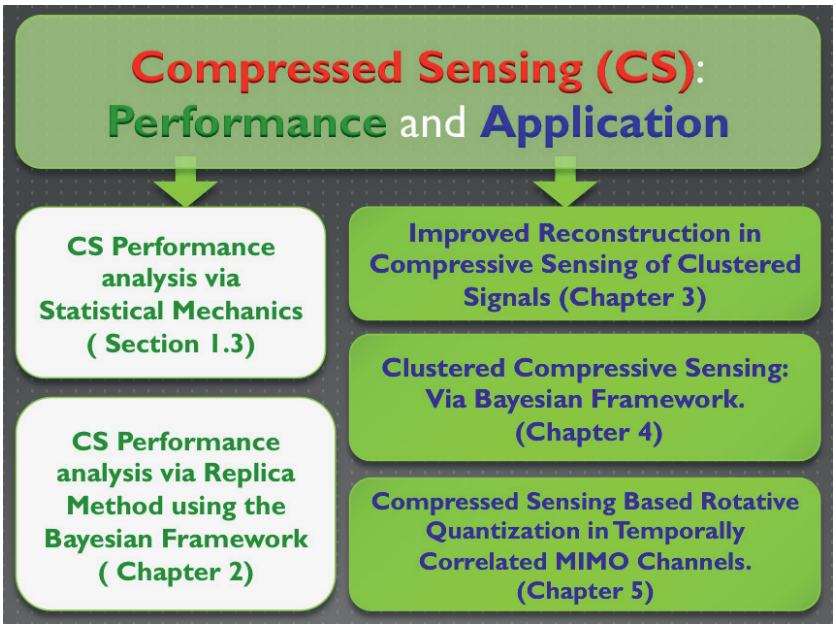

Figure 1.14: The thesis overview.

To deal with performance analysis of CS recovery algorithms, the key idea applied in this thesis can be described by the Venn diagram given in Figure 1.15. As discussed in Section 1.1.3 (also

in Chapter 2), performance analysis of the CS recovery schemes is done by applying tools from statistical mechanics by using the posterior distribution, which lies at the intersection of the three circles in the Venn diagram.

As presented in [35], the posterior distribution of each element of the sparse vector conditioned on the measurements is a sufficient statistic to infer about an individual element of the sparse signal based on measurements. Hence, due to its advantages more emphasis is given to the posterior distribution in this thesis. As presented in Section 1.2, the minimization problem of CS was equivalently represented using the maximum of a posterior distribution (MAP). In addition, the replica analysis of statistical mechanics begins at the Boltzman-Gibbs distribution in (1.3.6), which is also shown to be the posterior distribution. This triplet face of the posterior enables us to use the methods from one field to another, in our case tools developed from statistical mechanics to the asymptotic analysis of the CS reconstruction schemes. One finds several such works if the CS circle is replaced by neural systems [51], CDMA [61], MIMO [96] and so on.

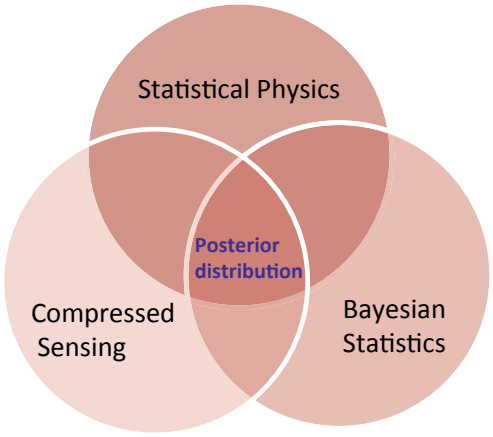

Figure 1.15: A Venn diagram that illustrates how statistical mechanics, Bayesian statistics and compressed sensing systems are connected, via the posterior and/or the partition function (evidence or normalizing factor).

The performance analysis part of the thesis (Section 1.1.3 and Chapter 2) lies at the center while the application Chapters (3,4 and 5) totally reside in the region of the intersection of CS and Bayesian Statistics circles of the venn diagram. The four papers that are published in different conference proceedings and journals are taken as separate chapters. Next, a short summary of the contribution of each paper is given as follows:

Chapter 2

Solomon A. Tesfamicael "Compressed Sensing Performance Analysis via Replica Method Using Bayesian Framework," *International Journal of Simulation Systems, Science & Technology (IJSSST), vol. 16, no. 3, 2016.* An earlier version is also in the Proceedings of IEEE UKSIM2015-AMSS 17th International Conference on Modelling and Simulation, and the analytical framework was presented at the 2012 IEEE European School of Information Theory in Antalya, Turkey between the 16 and 20 April.

In this chapter, the performance of CS estimators is analyzed using a tool from statistical mechanics, especially the replica method. It is an efficient tool to analyze large size systems in general. By considering the dimension of the signal N and the amount of measurements M to be large, i.e., N, $M \to \infty$, $N/M < \infty$, the performance of estimators used in CS like LASSO (the Least Absolute Shrinkage and Selection Operator) estimator and Zero-Norm regularizing estimator are analyzed, as a special case of maximum a posteriori (MAP) estimator by using a Bayesian framework. We use both replica symmetric (RS) ansatz and one step replica symmetry breaking (1RSB) ansatz, clamming the latter is efficient when the problem is not convex. This work is more analytical in its form. It is deferred for the next step to focus on the numerical results. A similar work has been done to analyze communication systems like Code Division Multiple Access (CDMA) and Multiple Input Multiple Output (MIMO) with large size. Actually, this work can be considered as an extension to CS systems.

Chapter 3

Solomon A. Tesfamicael, Faraz Barzideh, Lars Lundheim, "Improved Reconstruction in Compressive Sensing of Clustered Signals," in the proceedings of *IEEE AFRICON 2015*, vol., no., pp.1-7, 14-17 Sept. 2015.

In this chapter, a new method of compressive sensing reconstruction is presented by considering the signal to be estimated as both sparse and clustered. These two properties are modelled as a modified Laplacian prior in a Bayesian setting, resulting in two penalizing terms in the corresponding unconstrained minimization problem. Applying the algorithm on images with noisy observations show a significant gain when including the clustered assumption compared to the traditional Least Absolute Shrinkage and Selection Operator (LASSO) approach only penalizing for sparsity. The method was compared to other methods, which consider the two properties, and our approach is particularly well suited to clustered signals with little or no variation within the clustered regions, such as two-level images or other binary signals. However, there is a limitation to the result when there is variation, indicating that one has to use a more informed clustering function for better recovery.

Chapter 4

Solomon A. Tesfamicael and Faraz Barzideh, "Clustered Compressed Sensing via Bayesian Framework," in the proceedings of *IEEE UKSIM2015 17th International Conference on Modelling and Simulation. Selected as best paper for further journal publication.*

The results are also partially published in the *International Journal of Information and Electronics Engineering vol. 4, no. 2, pp. 74-80, 2014,* and presented at ICCEI 2014 held in Melbourne, Australia Jan .2-3, 2014 and awarded as the best oral presentation. In addition some parts were presented at the 2014 Joint National PhD Conference in Medical Imaging and MedViz Conference held in Bergen, Norway, on 17-18 June 2014.

This chapter provides clustered compressive sensing (CCS) based signal processing using

the Bayesian framework. By incorporating the different prior information such as sparsity and the special structures that are found among the sparse entries of different types of signals. The method is applied on synthetic and medical images like angiogram image, phantom and functional MRI (fMRI) images. The results show that applying the clustered compressive sensing out performs the non-clustered but only sparse counter parts when it comes to Mean Square Error (MSE), pick signal to noise ratio (PSNR) and other performance metrics.

Chapter 5

Solomon A. Tesfamicael and Lars Lundheim, "Compressed Sensing Based Rotative Quantization in Temporally Correlated MIMO Channels," in the proceedings of *Recent Developments on Signal Processing (RDSP 2013), September 2013.* Also it can be found at http://dsp.rice.edu/cs.

In this chapter, CS methods together with rotative quantization are used to compress and feedback channel state information for multiple input multiple output (MIMO) systems so that it reduces feedback overhead and improve performance. Using simulation, it is shown that the CS-based method reduces feedback overhead while delivering the same performance as the direct quantization scheme.

Part II

Performance Analysis

Chapter 2

Compressive Sensing Performance Analysis via Replica Method using the Bayesian Framework

Solomon A. Tesfamicael

Adapted from the International Journal of Simulation Systems, Science & Technology (IJSSST), vol. 16, no. 3, 2016.

2.1 Abstract

Compressive sensing (CS) is a new methodology to capture signals at lower rate than the Nyquist sampling rate when the signals are sparse or sparse in some domain. Studying the performance of such novel paradigms is an interesting subject. In this paper, the performance of CS estimators is analyzed using tools from statistical mechanics, especially the replica method via the Bayesian framework. This method has been used to analyze communication systems like Code Division Multiple Access (CDMA) and multiple input multiple output (MIMO) systems with large size. Replica analysis, partially proved to be rigorous, is an efficient tool to analyze large systems in general. Specifically, we analyze the performance of some of the estimators used in CS like LASSO (the Least Absolute Shrinkage and Selection Operator) estimator and Zero-Norm regularizing estimator as a special case of maximum a posteriori (MAP) estimator by using the Bayesian framework to connect the CS estimators and replica method. We use both replica symmetric (RS) ansatz and one-step replica symmetry breaking (1RSB) ansatz, claiming the latter is efficient when the problem is not convex. This work is analytical. It is deferred for next step to focus on the numerical results.

2.2 Introduction

Recently questions like, *why go to so much effort to acquire all the data when most of what we get will be thrown away?Can we just directly measure the part that will not end up being thrown away?* that were paused by Donoho and others [1, 2, 3, 12] triggered a new way of sampling or sensing called compact ("compressed") sensing (CS).

The CS paradigm in signal processing requires three important ingredients [4]. First, the desired signal should have a sparse representation in a known transform domain, i.e., it should be compressible. If the signal is sparse spatially, for example consider an image which is sparse in the pixels, then the transform domain can be the identity. Second, the aliasing artefacts due to undersampling should be incoherent in the transform domain. This creates a noise-like structure. This measurement noise then can be modelled using white Gaussian noise. Third, a nonlinear reconstruction scheme should be used to enforce sparsity and consistency with the data [77]. Recently, this recovery using CS has been shown to be mathematically exact [1, 2, 3, 12]. As a signal processing scheme, CS follows a similar framework: encoding, transmission/storing, and decoding. A block diagram is given in Figure 2.1 focusing on the encoding and decoding of such a system for noisy measurement.

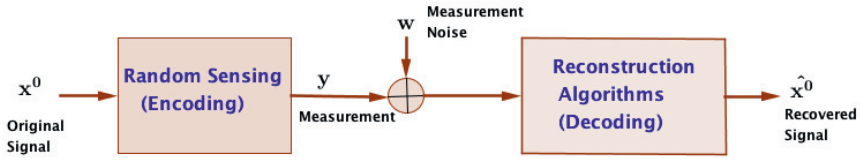

Figure 2.1: Blockdiagram for CS-based reconstruction.

In CS the task is to estimate or recover a sparse or compressible vector $\mathbf{x}^0 \in \mathbb{R}^N$ from a measurement vector $\mathbf{y} \in \mathbb{R}^M$. These are related through the linear transform $\mathbf{y} = \mathbf{A}\mathbf{x}^0$. Here, $\mathbf{x}^0$ is a sparse vector and $M \ll N$. In the seminal papers [1, 2, 3], $\mathbf{x}^0$ is estimated from $\mathbf{y}$, by solving a convex optimization problem [26, 97]. Others have used greedy algorithms, like subspace pursuit (SP)[20], orthogonal matching pursuit (OMP) [21] to solve the problem. In this chapter, the focus is rather on the convex optimization methods. We consider the noisy measurement system and the linear relation becomes

$$\mathbf{y} = \mathbf{A}\mathbf{x}^0 + \sigma_0 \mathbf{w}, \tag{2.2.1}$$

where $\mathbf{y}$ and $\mathbf{x}^0$ are as above where as the noise term, $\mathbf{w} \sim \mathcal{N}(0, \mathbf{I})$. There exists a large body of work on how to efficiently obtain an estimate for $\mathbf{x}^0$. The performance of such estimators are measured using metrics like Restricted Isometric Property (RIP) [12], Mutual Coherence (MC) [14], yet there is apparently no consensus on the bound determining how many measurements M are needed to approximate the sparse signal

with length N and sparsity k by using such metrics. The tool used in this chapter gives performance bounds of large size CS systems [67] using these system parameters.

Generally the linear model (2.2.1) is used to describe a multitude of linear systems like code division multiple access (CDMA) and multiple antenna systems like MIMO, to mention just a few. Tools from statistical mechanics have been employed to analyze large CDMA [61] and MIMO systems [53, 63], and in this work the same wisdom is applied to analyze the performance of estimators used in CS. Guo et al. in [67] used a Bayesian framework for statistical inference with noisy measurements and characterize the posterior distribution of individual elements of the sparse signal by describing the mean square erorr (MSE) exactly. To do so, they consider (2.2.1) in a large system and applied the decoupling principle using tools from statistical mechanics.

One can also find work that has used the tools from statistical mechanics to analyze CS system performance. To mention some, in [67] as stated above, Guo et al. used the tools to describe the minimum mean square error (MMSE) estimator, in [49] Rangan et al. used the maximum a posterior(MAP) estimator of CS systems. These are referred as Replica MMSE claim and Replica MAP claim in [49]. In [32, 98, 99] and [100] authors have used Belief propagation and message passing algorithms for probabilistic reconstruction in CS using replica methods including RS. Especially, in [101] one finds excellent work about phase diagrams in CS systems while [102] generalizes replica analysis using free random matrices. Recently, authors in [103] have proposed a turbo compressed sensing algorithm with partial discrete Fourier transform (DFT) sensing matrices. They claim that algorithm outperforms the well-known approximate message passing (AMP) algorithm when a partial DFT sensing matrix is involved. Kabashima et. al in [68], Ganguli and Sompolinsky in [104] and Takeda and Kabashima [70, 105, 106] have shown statistical mechanical analysis of the CS by considering the noiseless recovery problem and they indicated that RSB analysis is needed in the phase regimes where the RS solution is not stable. In this work, the performance of these CS estimators, considered as MAP estimator, is shown for the noisy problem by using the replica method including RS and RSB as in [54, 55, 96], where the RSB ansatz gives a better solution when the replica symmetry (RS) solution is unstable. This work is a kind of extension of [55], from MIMO systems to the CS systems.

The chapter is organized as follows. In Section 2.3, the estimator in the CS system is presented and redefined using the Bayesian framework, and based on that we present our basis of analysis in Section 2.4 which is the replica method from the statistical physics and apply it on the different CS estimators which are presented generally as a MAP estimator. In Section 2.5, we show our analysis using a particular example, and Section 2.6 presents conclusion and of future work.

2.3 Bayesian Framework for Sparse Estimation

Beginning with a given vector of measurements $\mathbf{y} \in \mathbb{R}^M$ and measurement matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, we assume noisy measurement with $\mathbf{w} \in \mathbb{R}^M$ being i.i.d. Gaussian random variables with zero mean and covariance matrix $\mathbf{I}$, estimating the sparse vector $\mathbf{x}^0 \in \mathbb{R}^N$ is the problem that we are considering where these variables are related by the linear model (2.2.1).

2.3.1 Sparse Signal Estimation

Various methods for estimating $\mathbf{x}^0$ may be used. The classical approach to solving inverse problems of such type is by the least squares (LS) estimator in which no prior information is used and its closed form is

$$\hat{\mathbf{x}^0} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}, \tag{2.3.1}$$

which performs very badly for the CS estimation problem we are considering since it does not find the sparse solution. Another approach to estimate $\mathbf{x}^0$ is via the solution of the unconstrained optimization problem

$$\hat{\mathbf{x}^0} = \min_{\mathbf{x}\in\mathbb{R}^N}\left[\frac{1}{2} \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel_2^2 + uf(\mathbf{x})\right], \tag{2.3.2}$$

where $uf(\mathbf{x})$ is a regularizing term, for some non-negative u. By taking $f(\mathbf{x}) = \parallel \mathbf{x} \parallel_p$, emphasis is made on a solution with lp-norm, and $\parallel \mathbf{x} \parallel_p$ is defined as a penalizing norm. When $p = 2$, we get

$$\hat{\mathbf{x}^0} = \min_{\mathbf{x}\in\mathbb{R}^N}\left[\frac{1}{2} \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel_2^2 + u \parallel \mathbf{x} \parallel_2\right], \tag{2.3.3}$$

This is penalizing the least square error by the l_2-norm and this performs badly as well, since it does not introduce sparsity into the problem. When $p = 0$, we get the l_0-norm, which is defined as

$$\|\mathbf{x}\|_0 = k \equiv \#\{i \in \{1, 2, \cdots, N\} | x_i \neq 0\},$$

the number of the non-zero entries of $\mathbf{x}$, which actually is a partial norm since it does not satisfy the triangle inequality property, but can be treated as norm by defining it as in [49], and get the l_0-norm regularizing estimator

$$\hat{\mathbf{x}^0} = \min_{\mathbf{x}\in\mathbb{R}^N}\left[\frac{1}{2} \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel_2^2 + u \parallel \mathbf{x} \parallel_0\right], \tag{2.3.4}$$

which gives the best solution for the problem at hand since it favors sparsity in $\mathbf{x}$. Nonetheless, it is an NP- hard combinatorial problem. Instead, it has been a practice to approximate it using the l_1- penalizing norm to get the estimator

$$\hat{\mathbf{x}^0} = \min_{\mathbf{x}\in\mathbb{R}^N} \left[\frac{1}{2} \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel_2^2 + u \parallel \mathbf{x} \parallel_1 \right], \tag{2.3.5}$$

which is a convex approximation to the l_0-penalizing solution 2.3.4. The best solution for estimating the sparse vector $\mathbf{x}$ is given by the Zero-Norm regularized estimator which is a hard combinatorial problem. These estimators, (2.3.3) - (2.3.5), can equivalently be presented as solutions to the constrained optimization problem [1, 2, 3]. This constrained optimization version of (2.3.5) is known as the l_1-penalized l_2-minimization called LASSO (Least Absolute Shrinkage and Selection Operator) or BPDN(Basis Persuit Denoising), which can be set as Quadratic Programming (QP) and Quadratic Constrained Linear Programming (QCPL) optimization problems. [1] In the following subsection the above estimators are presented as a MAP estimator in a Bayesian framework.

2.3.2 Bayesian Framework for Sparse Signals

Equivalently, the estimator of $\mathbf{x}^0$ in (2.3.2) can generally be presented as MAP estimator under the Bayesian framework. Assume a prior probability distribution for $\mathbf{x}$ to be

$$p_u(\mathbf{x}) = \frac{e^{-uf(\mathbf{x})}}{\int_{\mathbf{x}\in\chi^N} e^{-uf(\mathbf{x})} d\mathbf{x}}, \tag{2.3.6}$$

where the cost function $f : \chi \to \mathbb{R}$ is some scalar-valued, non-negative function with $\chi \subseteq \mathbb{R}$ and

$$f(\mathbf{x}) = \sum_{i=1}^{N} f(x_i). \tag{2.3.7}$$

such that for sufficiently large u, $\int_{\mathbf{x}\in\chi^N} \exp(-uf(\mathbf{x}))d\mathbf{x}$ is finite as in [49]. Let the assumed variance of the noise be given by

$$\sigma_u^2 = \frac{\gamma}{u}$$

where γ is system parameter which can be taken as $\gamma = \sigma_u^2 u$ where σ_u^2 is the assumed variance for each component of $\mathbf{w}$. Note that we incorporate the sparsity in the prior pdf via $f(\mathbf{x})$. By (2.2.1) the probability density function of $\mathbf{y}$ given $\mathbf{x}$ is given by

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \mid \mathbf{x}; \mathbf{A}) = \frac{1}{(2\pi\sigma_u^2)^{N/2}} e^{-\frac{1}{2\sigma_u^2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2}, \tag{2.3.8}$$

and prior distribution of $\mathbf{x}$ by (4.3.5), the posterior distribution for the measurement channel (2.2.1) according to Bayes' law is

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}; \mathbf{A}) = \frac{e^{-u(\frac{1}{2\gamma}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + f(\mathbf{x}))}}{\int_{\mathbf{x}\in\chi^N} e^{-u(\frac{1}{2\gamma}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + f(\mathbf{x}))} d\mathbf{x}}. \tag{2.3.9}$$

[1] In this chapter, we consider the former and leave the later as they are equivalent algorithms.

Then the MAP estimator can be shown to be

$$\hat{\mathbf{x}}^{MAP} = \arg\min_{\mathbf{x}\in\chi^N} \left[\frac{1}{2\gamma} \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel_2^2 + f(\mathbf{x}) \right]. \tag{2.3.10}$$

Now, as we choose a different penalizing function in (2.3.10) we get the different estimators defined above in Equations (2.3.3), (2.3.4), and (2.3.5) but this time under the framework of Bayesian framework as a MAP estimator [49].

1. Linear Estimators : when $f(\mathbf{x}) = \parallel \mathbf{x} \parallel_2^2$ (2.3.10) reduces to

$$\hat{\mathbf{x}}_{Linear}^{MAP} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T + \gamma\mathbf{I})^{-1}\mathbf{y}, \tag{2.3.11}$$

   which is the LMMSE estimator.

2. LASSO Estimator: when $f(\mathbf{x}) = \parallel \mathbf{x} \parallel_1$ we get the LASSO estimator and (2.3.10) becomes

$$\hat{\mathbf{x}}_{Lasso}^{MAP} = \arg\min_{\mathbf{x}\in\chi^N} \left[\frac{1}{2\gamma} \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel_2^2 + \parallel \mathbf{x} \parallel_1 \right]. \tag{2.3.12}$$

3. Zero-Norm regularization estimator: when $f(\mathbf{x}) = \|\mathbf{x}\|_0$, we get the Zero-Norm regularization estimator and (2.3.10) becomes

$$\hat{\mathbf{x}}_{Zero}^{MAP} = \arg\min_{\mathbf{x}\in\chi^N} \left[\frac{1}{2\gamma} \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel_2^2 + \parallel \mathbf{x} \parallel_0 \right]. \tag{2.3.13}$$

Whether these minimization problems are solvable or not, the replica analysis results can provide the asymptotic performances of all the above estimators via the replica method as given in [35, 49, 68, 70, 104]. We apply RS ansatz as used by Müller et al. in [35] and RSB ansatz as used by Zaidel et al. [55] on vector precoding for MIMO. Actually, this work is an extension of the RSB analysis to MIMO systems done in [55] to the CS system.

2.4 Statistical Analysis

The performance of the Bayesian estimators like MMSE and MAP can be done by determining the pdf of the error vector. The error is random and it should be centered about zero for the estimator to perform well. Kay showed in that way (see Section 11.6 in [107]) the performance analysis of MMSE estimator. We believe in general that inference for the asymptotic performance of MAP estimators is best done with statistical mechanical tools including RSB assumption. The outline of the replica analysis is done in this section. The reader is referred to [35, 108] and [55], for deeper understanding of replica method which is a central idea in this section.

We begin our analysis from the posterior distribution (2.3.9), which is sufficient statistics to estimate $\mathbf{x}^0$ [35] and the denominator is called the normalizing factor or evidence

in Bayesian inference according to [47] and Partition function in statistical mechanics [108]. Actually, it is this connection which provides the ground to apply the tools which are used in statistical mechanics. So the task of evaluating the above estimators for the sparse vector $\mathbf{x}^0$ can be translated to that of statistical physics task. Let us justify first how the analysis using statistical mechanical tool is able to do it.

The Boltzmann-Gibbs distribution is defined as

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\mathcal{Z}} e^{-\beta \mathcal{H}(\mathbf{x})} \tag{2.4.1}$$

where β is a constant known as the inverse temperature in the terminology of physical systems. For small β, the prior probability becomes flat, and for large β, the prior probability has sharp modes. $\mathcal{H}$, which is an expression of the total energy of the system, is called the Hamiltonian in physics literature and $\mathcal{Z}$ is the partition function given by

$$\mathcal{Z} = \int_{\chi^N} e^{-\beta \mathcal{H}(\mathbf{x})} d\mathbf{x}. \tag{2.4.2}$$

Often the Hamiltonian can be given by a quadratic form like

$$\mathcal{H}(\mathbf{x}) = \mathbf{x}^T \mathbf{J} \mathbf{x}, \tag{2.4.3}$$

with $\mathbf{J}$ being a random matrix of dimension $N \times N$. The energy of the system is given by

$$\mathcal{E} = \int_{\mathbf{x} \in \chi^N} p_{\mathbf{x}}(\mathbf{x}) \mathcal{H}(\mathbf{x}) d\mathbf{x}, \tag{2.4.4}$$

and the entropy(disorder) of the system is defined as

$$\mathcal{S} = -\int_{\mathbf{x} \in \chi^N} p_{\mathbf{x}}(\mathbf{x}) \log p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}. \tag{2.4.5}$$

The free energy can be calculated using

$$\mathcal{F} \equiv \mathcal{E} - \frac{\mathcal{S}}{\beta}. \tag{2.4.6}$$

At thermal equilibrium, the energy of the system being preserved, according to the second law of thermodynamics the entropy of the system is maximized when the free energy is minimized, where β, the inverse temperature, is the Lagrange multiplier in the maximization of (2.4.5), subject to the mean energy constraint. Therefore, at equilibrium, the free energy can be expressed as

$$\mathcal{F} = -\frac{1}{\beta} \log \mathcal{Z}. \tag{2.4.7}$$

The minimum average energy per component of $\mathbf{x}$ can be given by

$$\mathcal{E} = \frac{1}{N} \min_{\mathbf{x} \in \chi^N} \mathcal{H}(\mathbf{x}) \tag{2.4.8}$$

For our system that we address, which is given by (2.3.10) or equivalently by (2.3.2), the Hamiltonian becomes

$$\mathcal{H}(\mathbf{x}) = \frac{1}{2\sigma_u^2}(\mathbf{y} - \mathbf{A}\mathbf{x})^T(\mathbf{y} - \mathbf{A}\mathbf{x}) + uf(\mathbf{x}). \tag{2.4.9}$$

Compared to (2.4.3), the Hamiltonian in (2.4.9) has regularizing term in addition to the quadratic form in which the regularizing term $f(\mathbf{x})$ contributes to the additional terms involved in CS. After inserting (2.4.2) and (2.4.9) in (2.4.1) this gives information about the solution to (2.3.10) or to (2.3.2) in general, since they are equivalent problems. Therefore, one can use tools from statistical mechanics which helps to analyze the performance of such estimators. For example in this chapter, we infer about the minimum energy of each component of $\mathbf{x}$ using these tools.

Let $\mathbf{x}^0$ and $\mathbf{x}$ being drawn from the same set. The partition function of the posterior distribution given in (2.4.1) becomes

$$\mathcal{Z} = \int_{\mathbf{x}\in\chi^N} e^{-\beta\left[\frac{1}{2\sigma_u^2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + uf(\mathbf{x})\right]} d\mathbf{x}, \tag{2.4.10}$$

by using (2.4.2) and (2.4.9). The posterior distribution (2.3.9) depends on the predetermined random variables $\mathbf{y}$ and $\mathbf{A}$ called quenched states in physics literature [105, 106]. That is, we use fixed states $\mathbf{y} = \mathbf{A}\mathbf{x}^0 + \mathbf{w}$ instead of $\mathbf{y}$ for the large system limit, as $N, M \to \infty$, while maintaining N/M fixed. We then calculate the n^{th} moment of the partition function $\mathcal{Z}$ with respect to the predetermined variables, n replicas, hence this is where the name replica method came from. The replicated partition function is then given by

$$\mathcal{Z}^n = \int_{\{\mathbf{x}^a\}} e^{-\beta\left[\frac{1}{2\sigma_u^2}\sum\limits_{a=1}^{n}\left(\|\mathbf{y}-\mathbf{A}\mathbf{x}^a\|_2^2\right) + \frac{\gamma}{\sigma_u^2}\sum\limits_{a=1}^{n} f(\mathbf{x}^a)\right]} \prod_{a=1}^{n} d\mathbf{x}^a, \tag{2.4.11}$$

where $\int_{\{\mathbf{x}^a\}} = \int_{\mathbf{x}^1\in\chi^N} \cdots \int_{\mathbf{x}^n\in\chi^N}$. And after substituting $\mathbf{y}$, it becomes

$$\mathcal{Z}^n = \int_{\{\mathbf{x}^a\}} e^{-\beta\left[\frac{1}{2\sigma_u^2}\sum\limits_{a=1}^{n}\left(\|\mathbf{A}(\mathbf{x}^0-\mathbf{x}^a)+\mathbf{w}\|_2^2\right) + \frac{\gamma}{\sigma_u^2}\sum\limits_{a=1}^{n} f(\mathbf{x}^a)\right]} \prod_{a=1}^{n} d\mathbf{x}^a. \tag{2.4.12}$$

Averaging (2.4.12) over the noise $\mathbf{w}$, we get

$$\int_{\mathbb{R}^M} \frac{d\mathbf{w}}{\pi^M} e^{-\frac{1}{2\sigma_0}(\mathbf{w}^T\mathbf{w})} \mathcal{Z}^n = \alpha^{N/2} \int_{\{\mathbf{x}^a\}} e^{-\beta\left[\frac{1}{2}Tr\mathbf{J}\mathbf{L}(n) + \frac{\gamma}{\sigma_u^2}\sum\limits_{a=1}^{n} f(\mathbf{x}^a)\right]} \prod_{a=1}^{n} d\mathbf{x}^a, \tag{2.4.13}$$

where $\alpha = \frac{\sigma_u^2}{\sigma_u^2 + n\sigma_0^2}$, $\mathbf{J} = \mathbf{A}^T\mathbf{A}$ which is assumed to decompose into

$$\mathbf{J} = \mathbf{O}\mathbf{D}\mathbf{O}^{-1}, \tag{2.4.14}$$

where $\mathbf{D}$ is a diagonal matrix while $\mathbf{O}$ is $N \times N$ orthogonal matrix assumed to be drawn randomly from the uniform distribution defined by the Haar measure on the orthogonal group. For more clarity on this one can refer to in [105, 109]. $\mathbf{L}(n)$ is given by

$$\mathbf{L}(n)=-\frac{1}{\sigma_u^2}\sum_{a=1}^{n}(\mathbf{x}^0-\mathbf{x}^a)(\mathbf{x}^0-\mathbf{x}^a)^T+\frac{\sigma_0^4}{\sigma_u^2(\sigma_u^2+n\sigma_0^2)}\left(\sum_{a=1}^{n}(\mathbf{x}^0-\mathbf{x}^a)\right)\left(\sum_{b=1}^{n}(\mathbf{x}^0-\mathbf{x}^b)\right)^T. \tag{2.4.15}$$

Further averaging what we get on the right hand side of (2.4.13) over the cross correlation matrix $\mathbf{J}$, by assuming the eigenvalue spectrum of $\mathbf{J}$ to be self-averaging, we get

$$\begin{aligned}\underset{w,J}{E}\{\mathcal{Z}^n\} &= \underset{J}{E}\left(\alpha^{N/2}\int_{\{\mathbf{x}^a\}} e^{-\beta\left[\frac{1}{2}Tr\mathbf{JL}(n)+\frac{\gamma}{\sigma_u^2}\sum_{a=1}^{n}f(\mathbf{x}^a)\right]}\prod_{a=1}^{n}d\mathbf{x}^a\right)\\ &= \alpha^{N/2}\int_{\{\mathbf{x}^a\}} e^{\frac{-\beta\gamma}{\sigma_u^2}\sum_{a=1}^{n}f(\mathbf{x}^a)}\underset{J}{E}\left(e^{-\beta\left[\frac{1}{2}Tr\mathbf{JL}(n)\right]}\right)\prod_{a=1}^{n}d\mathbf{x}^a,\end{aligned} \tag{2.4.16}$$

The inner expectation in (2.4.16) is the Harish -Chandra -Itzykoson-Zuber integeral (again see in [55, 96] and the references therein). The plan here is to evaluate the fixed-rank matrices $\mathbf{L}(n)$ as $N \to \infty$. Further following the explanation in [55] (2.4.16) becomes

$$\underset{w,J}{E}\{\mathcal{Z}^n\} = \alpha^{N/2}\int_{\{\mathbf{x}^a\}} e^{\frac{-\beta\gamma}{\sigma_u^2}\sum_{a=1}^{n}f(\mathbf{x}^a)}e^{-N\sum_{a=1}^{n}\int_0^{\lambda_a}R(-v)dv+o(N)}\prod_{a=1}^{n}d\mathbf{x}^a \tag{2.4.17}$$

where $R(v)$ is the R-transform of the limiting eigenvalue distribution of the matrix J(see, definition 1 in [96] of R-transform or in [53] for better understanding of R-transform) and $\{\lambda_a\}$ denote the eigenvalues of $\mathbf{L}(n)$ as explained in [55, 96, 109].

After applying the replica trick, the average free energy is given by

$$\begin{aligned}\beta\bar{\mathcal{F}} &= -\lim_{N\to\infty}\frac{1}{N}\underset{w,J}{E}\{\log\ \mathcal{Z}\}\\ &= -\lim_{N\to\infty}\frac{1}{N}\lim_{n\to 0}\frac{\partial}{\partial n}\log\underset{w,J}{E}\{(\mathcal{Z})^n\}.\end{aligned} \tag{2.4.18}$$

and then we calculate the cost function of the L_p-norm reconstruction using the quenched average of the free energy [69, 70] as

$$\begin{aligned}\bar{\mathcal{E}} &= \lim_{\beta\to\infty}\frac{1}{\beta}\bar{\mathcal{F}}\\ &= -\lim_{\beta\to\infty}\frac{1}{\beta}\lim_{N\to\infty}\tfrac{1}{N}\underset{w,J}{E}\{\log\ \mathcal{Z}\}\\ &= -\lim_{\beta\to\infty}\frac{1}{\beta}\lim_{n\to 0}\frac{\partial}{\partial n}\underbrace{\lim_{N\to\infty}\tfrac{1}{N}\log\underset{w,J}{E}\{(\mathcal{Z})^n\}}_{\Xi n}.\end{aligned} \tag{2.4.19}$$

where we get (2.4.19) by using one of the assumptions used in replica calculations, after interchanging the order of the limits we assumed we get the same result. Further, for Ξ_n we have

$$\Xi_n = \lim_{N\to\infty} \tfrac{1}{N} \log\left(\alpha^{N/2} \int_{\{\mathbf{x}^a\}} e^{\frac{-\beta\gamma}{\sigma_u^2} \sum_{a=1}^{n} f(\mathbf{x}^a)} e^{\sum_{a=1}^{n} \int_0^{\lambda_a} R(-v)dv} \prod_{a=1}^{n} d\mathbf{x}^a \right). \tag{2.4.20}$$

Since the additive exponential terms of order $\circ(N)$ have no effect on the results when taking saddle point integration in the limiting regime as $N \to \infty$ due to the factor $\frac{1}{N}$ outside the logarithm in (2.4.20) any such terms are dropped further for notational simplicity as in [55].

In order to find the summation in (2.4.20) we employed the procedure in [55] and the nN dimensional space spanned by the replicas is split into subshells, defined through $n \times n$ matrix $\mathbf{Q}$

$$S(\mathbf{Q}) = \{\mathbf{x}^1, ..., \mathbf{x}^n \mid (\mathbf{x}^0 - \mathbf{x}^a)^T (\mathbf{x}^0 - \mathbf{x}^b) = NQ_{ab}\}. \tag{2.4.21}$$

The limit $N \to \infty$ enables us to use saddle point integration. Hence we can have the following general result as similar to [55] but extended in this work with the term which pertains to CS.

Proposition 1. *The energy $\mathcal{E}$ from* (2.4.8)*, for any inverse temperature β, any structure of $\mathbf{Q}$ consistent with* (2.4.21)*, and R-transform $R(\cdot)$, such that $R(\mathbf{Q})$ is well-defined, is given by*

$$\bar{\mathcal{E}} = \lim_{n\to 0} \frac{1}{n} Tr[\mathbf{Q}R(-\beta\mathbf{Q})], \tag{2.4.22}$$

where $\mathbf{Q}$ is the solution to the saddle point equation

$$\mathbf{Q} = \int \frac{\int_{\{\tilde{\mathbf{x}}\in\chi^n\}} \psi_2(\tilde{\mathbf{x}}) e^{\psi_1(\tilde{\mathbf{x}}) + \frac{1}{2}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2} f(\tilde{\mathbf{x}})} d\tilde{\mathbf{x}}}{\int_{\{\tilde{\mathbf{x}}\in\chi^n\}} e^{\psi_1(\tilde{\mathbf{x}}) + \frac{1}{2}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2} f(\tilde{\mathbf{x}})} d\tilde{\mathbf{x}}} dF_{X^0}(x^0), \tag{2.4.23}$$

where $dF_{X^0}(x^0)$ is a probablity measure of x^0, $\psi_1(\tilde{\mathbf{x}}) = (x^0\mathbf{1} - \tilde{\mathbf{x}})^T \tilde{\mathbf{Q}} (x^0\mathbf{1} - \tilde{\mathbf{x}})$, $\psi_2(\tilde{\mathbf{x}}) = (x^0\mathbf{1} - \tilde{\mathbf{x}})(x^0\mathbf{1} - \tilde{\mathbf{x}})^T$, and $\tilde{\mathbf{x}}$ is vector of dimension n .

Proof. See Appendix B.2. □

Further, to simplify the result in (2.4.22), we assume a simple structure on to the $n \times n$ cross correlation matrix $\mathbf{Q}$ at the Saddel point. So we assume two different assumptions for the entries of $\mathbf{Q}$ called ansatz: Replica Symmetry(RS) and one step Replica Symmetric Breaking (1RSB) ansatz. For example given the convexity of the energy function (2.3.12), the replica symmetric ansatz for the saddle point is reasonable. Whereas the energy function in (2.3.13) is not convex and assuming 1RSB assumption can suffice. Since the CS recovery estimators applied in this work are redefined as MAP estimators and generally it is believed that using 1RSB ansatz is enough to analyse such estimators if the local stability of the RS saddle point is lost against the perturbations [55], [74].

Showing the derivation of (2.4.22) analytically and further simplifying it using additional assumptions given under, is the central purpose of this pchapter as it was done in [96] and [55] for MIMO systems, but for a different purposes.

As in [96] and [55] we assume that the ansatz are given by:

1. replica symmetry ansatz:
$$\mathbf{Q} = q_0 \mathbf{1}_{n\times n} + \frac{b_0}{\beta}\mathbf{I}_{n\times n} \tag{2.4.24}$$

2. one replica symmetry breaking ansatz:
$$\mathbf{Q} = q_1 \mathbf{1}_{n\times n} + p_1 \mathbf{I}_{\frac{n\beta}{\mu_1}\times\frac{n\beta}{\mu_1}} \otimes \mathbf{1}_{\frac{\mu_1}{\beta}\times\frac{\mu_1}{\beta}} + \frac{b_1}{\beta}\mathbf{I}_{n\times n} \tag{2.4.25}$$

Applying these ansatzs we found the results given in the following subsections. In the first subsection the RS ansatz and in the second subsection the RSB ansatz results are presented.

2.4.1 LASSO Estimator with RS Ansatz

Consider the LASSO estimator given in (2.3.12), which is equivalent to the solution of the main unconstrained optimization problem (2.3.2) in l_1-penalized sense. Its performance is calculated via energy per component, $\bar{\mathcal{E}}$, using two macroscopic variables q_0 and b_0 given by

$$q_0 = \int_{\mathbb{R}}\int_{\mathbb{C}} \Big| x^0 - \Psi_1(x) \Big|^2 Dz dF_{X^0}(x^0), \tag{2.4.26}$$

$$b_0 = \frac{1}{f_0}\int_{\mathbb{R}}\int_{\mathbb{C}} \Re\Big\{ x^0 - \Psi_1(x) z^* \Big\} Dz dF_{X^0}(x^0), \tag{2.4.27}$$

where

$$\Psi_1(x) = \arg\min_{x\in\chi} \left[\Big| -z f_0 + 2e_0(x^0 - x) - \frac{\gamma}{\sigma_u^2} \Big| \right], \tag{2.4.28}$$

$$e_0 = \frac{1}{\sigma_u^2} R\Big(\frac{-b_0}{\sigma_u^2}\Big), \tag{2.4.29}$$

$$f_0 = \sqrt{2\frac{q_0}{\sigma_u^4} R'\Big(\frac{-b_0}{\sigma_u^2}\Big)}, \tag{2.4.30}$$

where $dF_{X^0}(x^0)$ is a probability measure of x^0 where as Dz refers to integration over Gaussian measure. $R(\cdot)$ and $R'(\cdot)$ are the R-transform and its derivative, respectively (See Appendix B.1.2). Under RS ansatz assumptions we then get the following statement.

Proposition 2. *Given the LASSO estimator in* (2.3.12) *and the macroscopic variables* q_0 *and* b_0*, in addition given the conditions in proposition 1, the typical value (per element) of the minimized cost function* (2.4.22) *simplifies to*

$$\bar{\mathcal{E}}_{rs}^{lasso} = \frac{q_0}{\sigma_u^2} R\Big(\frac{-b_0}{\sigma_u^2}\Big) - \frac{b_0 q_0}{\sigma_u^4} R'\Big(\frac{-b_0}{\sigma_u^2}\Big) \tag{2.4.31}$$

Proof. See Appendix B.3. □

2.4.2 LASSO Estimator with 1RSB Ansatz

Moving further to the RSB ansatz instead of assuming RS ansatz, we get more macroscopic parameters involved in the calculation of the quantity given by (2.4.22): b_1, p_1, q_1, and μ_1. These are given by the following fixed point equations as $n \to 0$ and $\beta \to \infty$, and using the compact notation as in [55]. Let

$$\Delta(y,z) \equiv e^{-\mu_1 \min_{x\in\chi} -2\Re\{(x^0-x)(f_1 z^* + g_1 y^*)\} + e_1(x^0-x)^2 - \frac{\gamma}{\sigma_u^2}|x|}, \quad (y,z) \in \Re^2 \tag{2.4.32}$$

and its normalized version

$$\tilde{\Delta}(y,z) = \frac{\Delta(y,z)}{\int_{\mathbb{C}} \Delta(\tilde{y},z) d\tilde{y}} \tag{2.4.33}$$

$$b_1 + p_1\mu_1 = \frac{1}{f_1} \int_{\mathbb{R}} \int_{\mathbb{C}^2} \Re\Big\{ \Big(x^0 - \Psi_2\Big) z^* \Big\} \tilde{\Delta}(y,z) Dy Dz dF_{X^0}(x^0) \tag{2.4.34}$$

$$b_1 + (q_1 + p_1)\mu_1 = \frac{1}{g_1} \int_{\mathbb{R}} \int_{\mathbb{C}^2} \Re\Big\{ \Big(x^0 - \Psi_2\Big) y^* \Big\} \tilde{\Delta}(y,z) Dy Dz dF_{X^0}(x^0) \tag{2.4.35}$$

$$q_1 + p_1 = \frac{1}{g_1} \int_{\mathbb{R}} \int_{\mathbb{C}^2} |\Psi_2|^2 \tilde{\Delta}(y,z) Dy Dz dF_{X^0}(x^0) \tag{2.4.36}$$

and

$$\begin{aligned} \int_{\frac{b_1}{\sigma_u^2}}^{\frac{b_1+\mu_1 p_1}{\sigma_u^2}} R(-v)dv = -R\Big(-\frac{b_1+\mu_1 p_1}{\sigma_u^2}\Big) - \mu_1^2 \Big((q_1+p_1) g_1^2 + p_1 f_1^2\Big) \\ + \int_{\mathbb{R}} \int_{\mathbb{C}} \log\Big(\int_{\mathbb{C}} \Delta(y,z) Dy \Big) Dz dF_{X^0}(x^0), \end{aligned} \tag{2.4.37}$$

where

$$\Psi_2 = \arg\min_{x\in\chi} \Big| 2\Re\{(x^0-x)(f_1 z^* + g_1 y^*)\} - e_1 |(x^0-x)|^2 - \frac{\gamma}{\sigma_u^2}|x| \Big|$$

and the other variables e_1, f_1, and g_1, are given by

$$e_1 = \frac{1}{\sigma_u^2} R(\frac{-b_1}{\sigma_u^2}), \tag{2.4.38}$$

$$g_1 = \sqrt{\frac{1}{\mu_1 \sigma_u^2}\left[R(\frac{-b_1}{\sigma_u^2}) - R(\frac{-b_1 - \mu_1 p_1}{\sigma_u^2})\right]}, \tag{2.4.39}$$

$$f_1 \overset{n\to 0}{\longrightarrow} \frac{1}{\sigma_u^2}\sqrt{q_1 R'(\frac{-b_1 - \mu_1 p_1}{\sigma_u^2})} \tag{2.4.40}$$

and where $dF_{X^0}(x^0)$ is a probability measure of x^0 where as Dy and Dz refers to integration over Gaussian measure. All these equations are shown in Appendix B.4. The use of RSB ansatz for a convex optimization problem like LASSO can be unnecessary if the RS ansatz provide the global minimum solution. However, the use of the RSB ansatz becomes crucial for the CS problem, since the best solution is provided by the Zero-Norm regularizing estimator, which is a non-convex problem. The next two proposition are provided as an extension of the proposition in [55] to CS problems.

Proposition 3. *Given the LASSO estimator in* (2.3.12) *and suppose the random matrix* **J** *satisfies the decomposability property* (2.4.14)*. Then under some technical assumptions, including one-step replica symmetry breaking, and the macroscopic variables given by the above fixed point equations, the effective typical value of the minimized cost function per component converges in probability as* N, $M \to \infty$, $N/M < \infty$ *, to*

$$\begin{aligned}\bar{\mathcal{E}}_{\text{1RSB}}^{\text{LASSO}} = &\frac{1}{\sigma_u^2}(q_1 + p_1 + \frac{b_1}{\mu_1})R(\frac{-b_1 - \mu_1 p_1}{\sigma_u^2}) - \frac{b_1}{\mu_1 \sigma_u^2}R(-\frac{b_1}{\sigma_u^2}) \\ &- q_1(\frac{b_1 + \mu_1 p_1}{\sigma_u^4})R'(\frac{-b_1 - \mu_1 p_1}{\sigma_u^2})\end{aligned} \tag{2.4.41}$$

Proof. See Appendix B.4.

□

If we only have the RS-ansatz instead of the RSB-ansatz, that is, with $p_1 = 0$, $\mu_1 = 1$, $b_1 = b_0$, and $q_1 = q_0$, then

$$\bar{\mathcal{E}}_{\text{1RSB}}^{\text{LASSO}} \to \bar{\mathcal{E}}_{\text{RS}}^{\text{LASSO}}. \tag{2.4.42}$$

2.4.3 Zero-Norm Regularizing Estimator with 1RSB Ansatz

The LASSO estimation is considered as the convex relaxation of the the Zero-Norm regularizing estimation. Since the latter is a non-convex problem its performance is better evaluated when we use RSB ansatz. So extending proposition (3) to this estimator we get the following statement.

Proposition 4. *Given the Zero-Norm regularizing estimator in* (2.3.13) *and suppose the random matrix* $\mathbf{J}$ *satisfies the decomposability property* (2.4.14)*. Then under some technical assumptions, including one-step replica symmetry breaking, the effective energy penalty per component converges in probability as* N, $M \to \infty$, $N/M < \infty$ *, to*

$$\begin{aligned}\bar{\mathcal{E}}^{\textit{Zero-Norm}}_{\textit{1RSB}} = &\frac{1}{\sigma_u^2}(q_1 + p_1 + \frac{b_1}{\mu_1})R(\frac{-b_1 - \mu_1 p_1}{\sigma_u^2}) - \frac{b_1}{\mu_1 \sigma_u^2}R(-\frac{b_1}{\sigma_u^2}) \\ &- q_1(\frac{b_1 + \mu_1 p_1}{\sigma_u^4})R'(\frac{-b_1 - \mu_1 p_1}{\sigma_u^2}) \end{aligned} \tag{2.4.43}$$

Proof. The proof is similar to that of Proposition 3 given in Appendix B.4. The adjustment needed is to change the regularizing term from L1 to L0, that is, $\| \mathbf{x} \|_1$ to $\| \mathbf{x} \|_0$. □

Note that even though (2.4.41) and (2.4.43) have the same expression they are different since the macroscopic variables are calculated from different prior distributions for $\mathbf{x}$ and $\mathbf{x}^0$.

2.5 Particular Example: Bernoulli-Gaussian Mixture Distribution

Assume the original vector $\mathbf{x}^0 \in \mathbb{R}^N$ follows a Bernoulli-Gaussian mixture distribution. So following the Bayesian framework analysis in Section 2.4, let $\mathbf{x}$ be composed of random variables with each component obeying the pdf

$$p(x) \sim \begin{cases} \mathcal{N}(0,1) & \text{with probablity } \rho \\ 0 & \text{with probablity } 1-\rho, \end{cases} \tag{2.5.1}$$

where $\rho = k/N$, with k being the number of non-zero entries of $\mathbf{x}$. Without loss of generality, let $\rho = 0.1$, M/N vary between 0.2 and 0.6. Also let us assume that the entries of the measurement matrix $\mathbf{A}$ follow i.i.d. Gaussian random variable of mean zero and variance $1/M$. In addition, w.l.o.g., let σ_u^2 be such that the signal to noise ratio is $10dB$ and $30dB$.

2.5.1 Replica Symmetry Analysis

Considering the macroscopic variables given by (B.5.3) and (B.5.6) and inserting the assumed distributions above and simplifying, the fixed point equations become

$$q_0 = \frac{(1-\rho)}{\pi}\int_{\mathbb{C}} \Psi_3(z)dz + \frac{\rho}{\sqrt{2}\pi^{3/2}}\int_{\mathbb{C}}\int_{\mathbb{R}} \Psi_4(z, x^0)dx^0 dz, \tag{2.5.2}$$

$$b_0 = \frac{(1-\rho)}{\pi f_0}\int_{\mathbb{C}} \Psi_5(z)dz + \frac{\rho}{\sqrt{2}\pi^{3/2} f_0}\int_{\mathbb{C}}\int_{\mathbb{R}} \Psi_6(z,x^0)dx^0 dz \tag{2.5.3}$$

where e_0 and f_0 are given by (2.4.29) and (2.4.30), respectively, also

$$\Psi_3(z) = \Big|\frac{zf_0 + \frac{\gamma}{\sigma_u^2}}{2e_0}\Big|^2 e^{-|z|^2}, \qquad \Psi_4(z,x^0) = \Big|\frac{zf_0 + \frac{\gamma}{\sigma_u^2}}{2e_0}\Big|^2 e^{-\left(\frac{(x^0)^2}{2}+|z|^2\right)}, \tag{2.5.4}$$

$$\Psi_5(z) = \Re\left\{\left(\frac{zf_0 + \frac{\gamma}{\sigma_u^2}}{2e_0}\right)z^*\right\}e^{-|z|^2}, and \tag{2.5.5}$$

$$\Psi_6(z,x^0) = \Re\left\{x^0(1-z^*) + \left(\frac{zf_0 + \frac{\gamma}{\sigma_u^2}}{2e_0}\right)z^*\right\}e^{-\left(\frac{(x^0)^2}{2}+|z|^2\right)}, \tag{2.5.6}$$

(See Appendix B.5) . With the defined values for the parameters given in this subsection

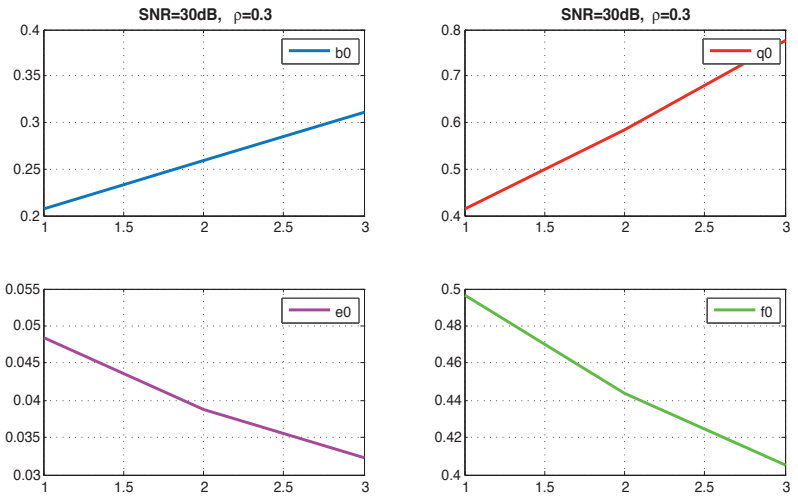

Figure 2.2: The macroscopic variables of the RS ansatz for LASSO versus the measurement ratio M/N.

above we first plot these macroscopic variables and this is shown in Figure 2.2. Using these macroscopic variables, we plot the minimized cost function per component versus measurement ratio M/N in Figure 2.3 which is given under proposition 2 for different sparsity ratios ($\rho = 0.1$, $\rho = 0.3$).

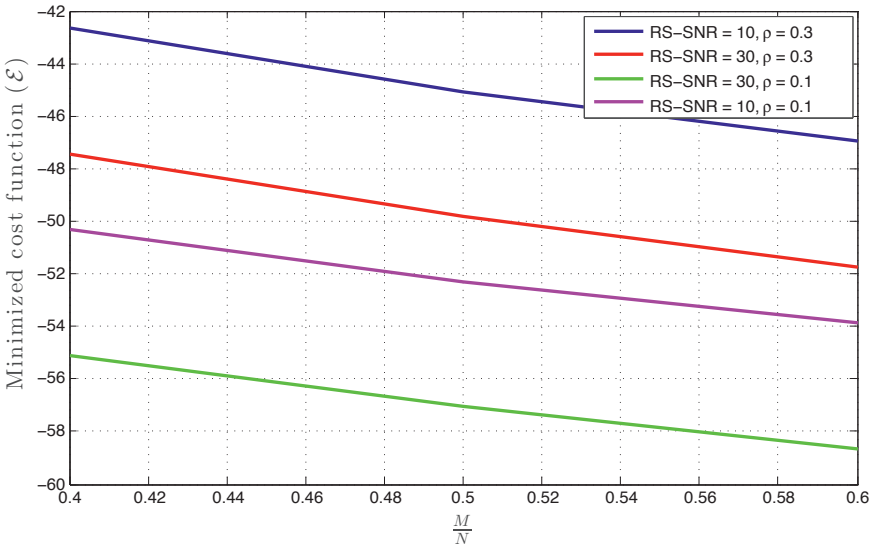

Figure 2.3: The effective typical value of the minimized cost function per component against measurement ratio M/N for different sparsity.

2.5.2 Replica Symmetry Breaking Analysis

Considering the same Bernouli-Gaussian mixture distribution (2.5.1) assumed in this section we consider the macroscopic variables which arises from 1RSB ansatz. Then the effective typical value of the minimized cost function per component as $M \to \infty, N \to \infty$, while M/N is finite, which are given by (2.4.41) and (2.4.43) are dependent up on four macroscopic variables given by (2.4.34)-(2.4.37).

It is possible to simplify these equations further and give numerical results. But this is deferred for further work. We expect that the free energy from the RSB ansatz to be greater than the free energy from the RS ansatz for the Zero-Norm regularizing, which can be seen from the analytical terms which have more parameters in (2.4.43). However, for LASSO these free energies, hence the typical value of the minimized cost function, will be quite similar since for convex minimization problems there is one global minimum and RS ansatz is sufficient enough to produce the solution.

2.6 Conclusion

In this chapter, we have used the replica method to analyze the performance of the estimators used in compressed sensing in which we generalized them as MAP estimators. The performance of MAP estimators can be shown using the replica method. 1RSB ansatz can be enough to analyze such estimators [55]. We have only shown here one particu-

lar example for the CS problem, i.e. for Bernoulli-Gaussian distribution. One may be interested to verify it using different examples. In addition we have only compared the performance of the estimators based on the free energy, but one can also use other metrics such as comparing the input/output distribution using replica analysis, as it is done in [55]. The main result of this chapter is analytical analysis for the performance of the estimators used in CS. These issues and others, like doing the numerical analysis for the 1RSB, are left for future work.

2.7 Acknowlegment

We are greatful to Ralf R. Müller, Lars Lundheim, Faraz Barzideh, Rodrigo Vicente de Miguel, the late Bruhtesfa E. Godana and Benjamin M. Zaidel for interesting discussions and suggestions.

Part III

CS Application

Chapter 3

Improved Reconstruction in Compressive Sensing of Clustered Signals

Solomon A. Tesfamicael, Faraz Barzideh and Lars Lundheim

Adapted from the proceedings of IEEE AFRICON 2015, vol., no., pp.1-7, 14-17 Sept. 2015.

3.1 Abstract

A new method of compressive sensing reconstruction is presented. The method assumes that the signal to be estimated is both sparse and clustered. These properties are modelled as a modified Laplacian prior in a Bayesian setting, resulting in two penalizing terms in the corresponding unconstrained minimization problem. In the implementation an equivalent constrained minimization problem is solved using quadratic programming. Experiments on images with noisy observations show a significant gain when including the clustered assumption compared to the traditional Least Absolute Shrinkage and Selection Operator (LASSO) approach only penalizing for sparsity.

Comparison with other methods highlights that our approach is particularly well suited to clustered signals with little or no variation within the clustered regions, such as two-level images or other binary signals.

3.2 Introduction

Compressive Sensing denotes a method of representing a signal with far fewer samples than what would be required by traditional Nyquist sampling [1, 4, 12]. Theoretical results show under what conditions reconstruction from the measurements are possible [2, 3], and an important requirement is that the signal in question is sparse in some suitable base.

Whereas existence of a solution of the reconstruction problem can be proved mathematically, actual reconstruction algorithms may be hard to find and complex to execute. During the last decade many reconstruction methods have been suggested using results from optimization theory [2, 3, 12].

Useful insight in the problem has been provided when the reconstruction can be regarded as a MAP estimator with a certain prior pdf. This was first suggested by Rangan et al. in [49].

Reconstruction performance can be improved by assuming signal properties in addition to sparsity. One such property is clusteredness. This has been investigated in [110, 111, 112, 113, 114, 115]. In [110], the authors extend the theory of CS to include signals that are concisely represented in terms of a graphical model, and used Markov Random Fields (MRFs) to represent sparse signals whose non-zero coefficients are clustered. A structured sparsity model is also used in [111] and random projections are applied to recover signals from fewer measurements. In [112] a robust recovery of signals from a structured union of subspaces is applied. While in [113] and [114] a hierarchical Bayesian generative model for sparse signals is found in which they have applied full Bayesian analysis by assuming prior distributions to each parameter appearing in the analysis, in [115] the authors provide an algorithm inspired by sparse subspace clustering (SSC) to cluster noisy data, and develop a novel theory demonstrating its correctness.

After presenting the problem and the conventional LASSO approach [6], [25] in Section 4.3, we show how both sparsity and clusteredness can be modelled by assuming a modified Laplacian prior pdf along the lines of [113]. In Section 3.4, we present a numerical method for solving the ensuing modified LASSO optimization problem. Some experimental results follow in Section 4.5 where we see that an SNR improvement of 10 dB can be gained by the new method. Conclusions and suggestions for further research are given in Section 3.6.

3.3 Problem formulation

3.3.1 Compressive Sensing Reconstruction

The reconstruction problem in compressive sensing can be regarded as estimating $\mathbf{x} = [x_1, x_2, \dots x_N]^\top \in \mathbb{R}^N$ from observations $\mathbf{y} \in \mathbb{R}^M$ where the two are related by

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}. \tag{3.3.1}$$

Here $\mathbf{A} \in \mathbb{R}^{M \times N}$ is a measurement matrix and $\mathbf{w} \in \mathbb{R}^M$ is a zero mean, white Gaussian noise vector with covariance matrix $\sigma^2\mathbf{I}$.

In classical regression, it is usually assumed that $M >> N$. In compressed sensing, the situation is reversed, i.e. $N >> M$. On the other hand, $\mathbf{x}$ is sparse, meaning that $\| \mathbf{x} \|_0 = k << N$ where the "Zero-Norm" $\| \mathbf{x} \|_0 = \#\{x_i | x_i \neq 0\}$ denotes the number of non-zero elements in $\mathbf{x}$.

Even without noise, solving (4.3.1) for $\mathbf{x}$ is an underdetermined problem, and estimates are often sought by the minimization

$$\hat{\mathbf{x}} = \min_{x \in R^N} \frac{1}{2} \| \mathbf{y} - \mathbf{A}\mathbf{x} \|_2^2 + \lambda f(\mathbf{x}), \tag{3.3.2}$$

where $\lambda \geq 0$ and $f(\mathbf{x})$ a regularizing function penalizing non-sparse solutions. The CS literature contains many solutions to this problem. A large class fall under the LASSO approach [1, 2, 3, 4, 12] using $f(\mathbf{x}) = \| \mathbf{x} \|_1$ resulting in the estimator

$$\hat{\mathbf{x}}_{\text{LASSO}} = \min_{x \in R^N} \frac{1}{2} \| \mathbf{y} - \mathbf{A}\mathbf{x} \|_2^2 + \lambda \| \mathbf{x} \|_1 \tag{3.3.3}$$

This will also be our point of departure, but in addition we introduce an assumption of clusteredness as described in the next paragraph.

3.3.2 Bayesian Argument for Modified LASSO

As pointed out in [49] the minimization problem (4.3.3) can be interpreted as an equivalent MAP estimator

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}) p(\mathbf{x})$$

with

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \mid \mathbf{x}) = \frac{1}{(2\pi\sigma)^{N/2}} e^{-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2}, \tag{3.3.4}$$

and a prior pdf

$$p(\mathbf{x}) = \frac{e^{-\lambda f(\mathbf{x})}}{\int_{\mathbf{x} \in \mathbb{R}^N} e^{-\lambda f(\mathbf{x})} d\mathbf{x}}. \tag{3.3.5}$$

Under the assumption of x_i i.i.d., it is seen that $f(\mathbf{x}) = \| \mathbf{x} \|_2^2$ leads to a Gaussian $p(\mathbf{x})$ whereas $f(\mathbf{x}) = \| \mathbf{x} \|_1$ (the LASSO approach) leads to a Laplacian.

By assuming a sparse and clustered $\mathbf{x}$, the x_i are no longer independent. In the following we show how assuming a local dependence between individual samples of x_i leads to a modified Laplacian prior and, in turn, a modified LASSO estimator.

Let $\mathbf{D}_i$ denote the set of closest neighbours of x_i. The contents of $\mathbf{D}_i$ will depend on the nature of the signal in question. For a one-dimensional time series,

$$\mathbf{D}_i = \{x_{i-1}, x_{i+1}\}. \tag{3.3.6}$$

If $\mathbf{x}$ is a vectorized two-dimensional image, $\mathbf{D}_i$ will consist of the neighbour pixels in horizontal and vertical directions, possibly including diagonal neighbours as well. We now assume that x_i is statistically dependent on its neighbours. We also assume that far apart variables are only indirectly dependent as in a Markov chain, such that for an arbitrary sample set A,

$$p(x_i|\mathbf{D}_i \cup A) = p(x_i|\mathbf{D}_i). \tag{3.3.7}$$

Now, by the chain rule, we get

$$\begin{aligned} p(\mathbf{x}) &= p(x_1|x_2, x_3, \cdots, x_N) \cdot p(x_2|x_3, x_3, \cdots, x_N) \cdots \\ &\quad p(x_{N-1}|x_N) \cdot p(x_N) \\ &= p(x_1|\mathbf{D}_i^+) \cdot p(x_2|\mathbf{D}_2^+) \cdots (x_{N-1}|\mathbf{D}_{N-1}^+) \cdot p(x_N) \end{aligned} \tag{3.3.8}$$

where we have used the notation $\mathbf{D}_i^+ = \{x_j \in \mathbf{D}_i | j > i\}$.

Local dependence can now be modelled by a modified Laplacian

$$p(x_i|\mathbf{D}_i^+) = \frac{1}{c} e^{-(\lambda_1|x_i| + \lambda_2 \sum_{x \in \mathbf{D}_i^+}^N |x_i - x|)} \tag{3.3.9}$$

with an appropriate normalization c. This prior is seen to favour sparse signals and signals where neighbouring elements have a small difference. Combining (3.3.8) and (3.3.9) we finally obtain

$$p(\mathbf{x}) = e^{-\lambda_1 \|\mathbf{x}\|_1 - \lambda_2 D(\mathbf{x})}, \tag{3.3.10}$$

where

$$D(\mathbf{x}) = \sum_{i=1}^{N-1} |x_i - x_{i+1}|. \tag{3.3.11}$$

From here it follows that the MAP estimator is given by

$$\hat{\mathbf{x}}_{\text{CLASSO}} = \arg\min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel_2^2 + \lambda_1 \parallel \mathbf{x} \parallel_1 + \lambda_2 D(\mathbf{x}), \tag{3.3.12}$$

which we now recognize as a modified LASSO estimator where an additional regularizing term $\lambda_2 D(\mathbf{x})$ has been included. By appropriate choice of the constants λ_1, λ_2 sparsity and clusteredness will be properly taken into account under the minimization of the squared error.

3.4 Implemented Algorithm

3.4.1 Reformulation for Quadratic Programming

With a traditional (3.3.3) or modified (3.3.12) LASSO approach, several numerical methods can be used in order to solve the minimization [4, 23, 116]. A common scheme is to deal with the equivalent constrained problem, that is expressed as

$$\hat{\mathbf{x}}_{\text{LASSO}} = \arg\min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \text{ subject to } \|\mathbf{x}\|_1 \leq t_1$$

for LASSO (3.3.3) and

$$\hat{\mathbf{x}}_{\text{CLASSO}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 \text{ subject to } \|\mathbf{x}\|_1 \leq t_1;\ D(\mathbf{x}) \leq t_2 \tag{3.4.1}$$

for the modified version (3.3.12), where t_1 and t_2 are appropriately chosen thresholds. For this kind of constrained minimization problem, quadratic programming is a natural choice for solution. A challenge is then to reformulate the (non linear) constraints to a set of linear ones. For the problem at hand (4.4.2) a linear constraint can be formulated by introducing two auxiliary variables $\mathbf{q}_1 \in \mathbf{R}^N$ and $\mathbf{q}_2 \in \mathbf{R}^N$. We compose an extended variable $\bar{\mathbf{x}} \in \mathbf{R}^{3N}$, a matrix $\bar{\mathbf{A}} \in \mathbf{R}^{N\times 3N}$ and constrains vector $\mathbf{b} \in \mathbf{R}^{4N+2}$ by

$$\bar{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ \mathbf{q}_1 \\ \mathbf{q}_2 \end{bmatrix},\ \bar{\mathbf{A}} = [\mathbf{A}\ \mathbf{0}_{N\times 1}\ \mathbf{0}_{N\times 1}],\ \mathbf{b} = \begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{0}_{N\times 1} \\ t_1 \\ \mathbf{0}_{N\times 1} \\ \mathbf{0}_{N\times 1} \\ t_2 \end{bmatrix}, \tag{3.4.2}$$

and solve the extended quadratic programming problem

$$\min_{\bar{\mathbf{x}}\in\mathbb{R}^{3N}} \|\bar{\mathbf{A}}\bar{\mathbf{x}}-\mathbf{y}\|_2^2 \text{ subject to } \mathbf{C}\bar{\mathbf{x}} \leq \mathbf{b} \tag{3.4.3}$$

where $\mathbf{C}$ is a $(4N+2)\times 3N$ matrix composed as follows:

$$\mathbf{C} = \left[\begin{array}{c|c|c} \mathbf{I} & -\mathbf{I} & \mathbf{0}_{N\times N} \\ \hline -\mathbf{I} & -\mathbf{I} & \mathbf{0}_{N\times N} \\ \hline \mathbf{0}_{1\times N} & \mathbf{1}_{1\times N} & \mathbf{0}_{1\times N} \\ \hline \mathbf{D} & \mathbf{0}_{N\times N} & -\mathbf{I} \\ \hline -\mathbf{D} & \mathbf{0}_{N\times N} & -\mathbf{I} \\ \hline \mathbf{0}_{1\times N} & \mathbf{1}_{1\times N} & 0\quad \mathbf{1}_{1\times(N-1)} \end{array}\right].$$

The inequality sign "$\leq$" in (3.4.3) is meant to be understood elementwise for the vector quantities concerned. Moreover, $\mathbf{I}$ is the $N\times N$ identity matrix, and the notation $\mathbf{0}_{m\times n}$ and $\mathbf{1}_{m\times n}$ signifies $m\times n$ sub matrices consisting of zeros and ones respectively.

$\mathbf{D}$ is an $N\times N$ matrix performing partial sums in the operator $D(\mathbf{x})$ in (3.3.11):

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & \cdots & & 0 \\ -1 & 1 & 0 & & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}.$$

3.4.2 Parameter Tuning

Good reconstruction requires proper choice of the thresholds t_1 and t_2 [115, 117] (or, equivalently, of λ_1, λ_2). This choice may be both signal and application dependent. Often an automatic tuning will be required, but situations where manual tuning are appropriate, can also be imagined. In the present work, no particular method for parameter choice is assumed. We have found experimentally that using $t_1 = \|\mathbf{x}\|_1$ and $t_2 = \|D\mathbf{x}\|_1$ give close to optimal results. These numbers will of course be unknown in practical use of the method, so the results in the next section might be regarded as optimistic.

3.5 Experimental Results

3.5.1 Noise Robustness

A signal's degree of clusteredness depends both on the signal's origin and on the base in which it is represented. To reveal some of the basic properties of the suggested method, we first present an experiment with a family of binary signals with sparsity k/N ranging from 0 to 1 defined by $\mathbf{x}_k = [x_{k,1}, x_{k,2}, \ldots x_{k,N}]^\top$ with

$$x_{k,1} = \begin{cases} 1 & \text{if } l \leq k \\ 0 & \text{otherwise} \end{cases}$$

These signals are "ultimately clustered", as all non zero values (which are all equal) are placed together in a single connected block. For the measurement matrix $\mathbf{A}$ we have used a random matrix with i.i.d. zero mean, unit variance Gaussian entries, further we have scaled $\mathbf{A}$ row wise ($\frac{1}{\sqrt{M}} \times randn(M, N)$) and then scaled column wise so that we get $||\mathbf{Ax}||= ||\mathbf{x}||$. With noise as defined in (4.3.1) this gives a *measurement SNR*

$$\text{SNR}_m = \mathbf{E}[y_i^2]/\mathbf{E}[w_i^2] = k/\sigma^2. \tag{3.5.1}$$

The reconstruction quality can be measured by the *reconstruction SNR*

$$\text{SNR}_r = \frac{1}{N} \parallel \mathbf{x} \parallel_2^2 / \mathbf{E}[(\hat{x}_i - x_i)^2] = \frac{k/N}{\mathbf{E}[(\hat{x}_i - x_i)^2]}. \tag{3.5.2}$$

The reconstruction SNR was estimated for both LASSO and modified LASSO by averaging over 20 instances where both the noise vector and measurement matrix were randomly chosen for each instance. The experiment was executed with signal length N= 300 and observation length $M = 200$. The sparsity parameter varied in the range $1 \leq k \leq N$. In Figure 3.1 reconstruction SNR is plotted against k/M.

The top set of curves show the result when the measurement noise $\mathbf{w} = \mathbf{0}$. Then $\text{SNR}_m = \infty$, and LASSO outperforms the modified versions when the signal is clearly sparse. (This corresponds to $t_2 = 0$, showing that the chosen threshold value is evidently not optimal in this case.) When the signal is less sparse, the performance gradually decreases until it

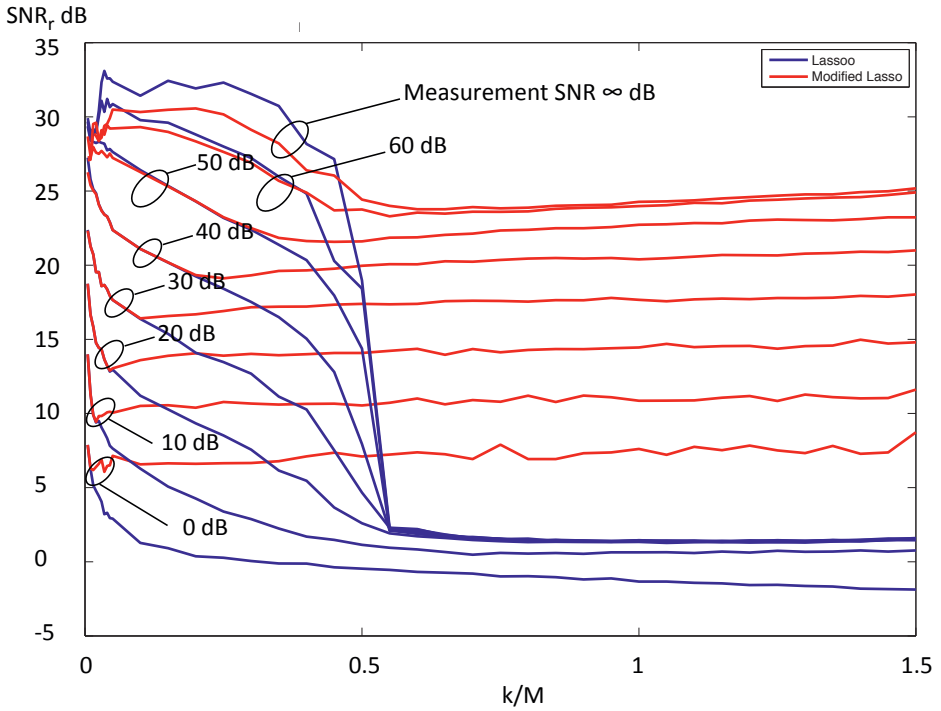

Figure 3.1: SNR$_r$ vs. k/M for different SNR$_m$

breaks down around $k/M = 0.5$ as expected, due to the phase transition property [30]. It is worth noting that the modified version does not have this break-down behaviour, but obtains a stable performance around SNR$_r$ =24 dB.

For reduced SNR_m this pattern repeats itself. Modified and unmodified LASSO have similar performance for sparse signals, but when LASSO deteriorates with decreasing sparseness, modified LASSO stabilizes at an SNR_r level roughly proportional to the measurement SNR. The noise robustness advantage of the modified version is clearly increased with low values of SNR_m.

As mentioned in Section 3.4.2 we have not done a thorough investigation of the choice of thresholds t_1, t_2 in this work. More detailed methods for choosing these values might improve the performance.

3.5.2 Comparison with other Methods

Our approach has been compared with two other CS algorithms that also take clusteredness into account. We will call these methods CluSS-MCMC and CluSS-VB respectively.

Both methods apply a hierarchical Bayesian model to model both the sparse prior and

cluster prior simultaneously by taking into account the cluster structure property of sparse signals, of which the non zero coefficients appear in clustered blocks. Their basic difference is in the signal recovery method. CluSS-MCMC [113] applies Markov Chain Monte Carlo (MCMC) sampling, whereas CluSS-VB [114] uses a variational Bayes approach.

In the comparison two types of signals, Type I and Type II, have been used. Both types of signals have length $N = 100$ and k non-zero values concentrated in two clusters whose positions are randomly (uniformly) chosen.

For Type I signals, $k = N/4 = 25$ and the first cluster has a size k_1 uniformly distributed in the range 1 to $k - 1$ so that the other one gets length $k_2 = k - k_1$. Signal values within each cluster are zero mean Gaussian variable with variance 1. An example is shown in Figure 3.2.

For Type II signals, both clusters have length $k = N/4 = 25$. Signal values are constant within each cluster with values 1 for the first one and 1/2 for the second one.

An indication of algorithm performance with $M = N/2 = 50$ measurements is given in Figures 3.2–3.3.

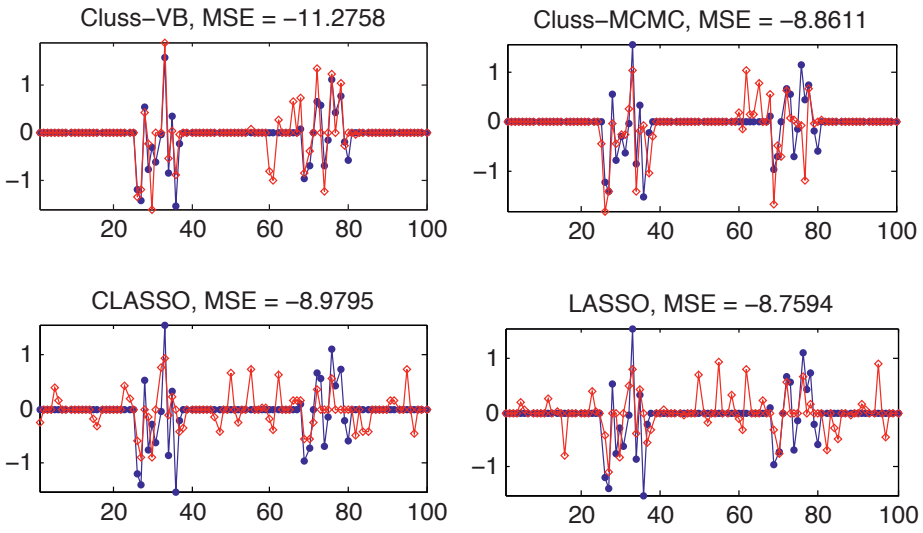

Figure 3.2: Performance of algorithms using signal Type I, where the original signal is represented in blue and the recovery by the respective algorithms are in red.

For signals of Type I we see that the methods CluSS-MCMC an CluSS-VB are better at localizing the clustered regions and generating zero values where no signals are present. For Signals of type II, where signal values are constant within each cluster, our method clearly outperforms the rest.

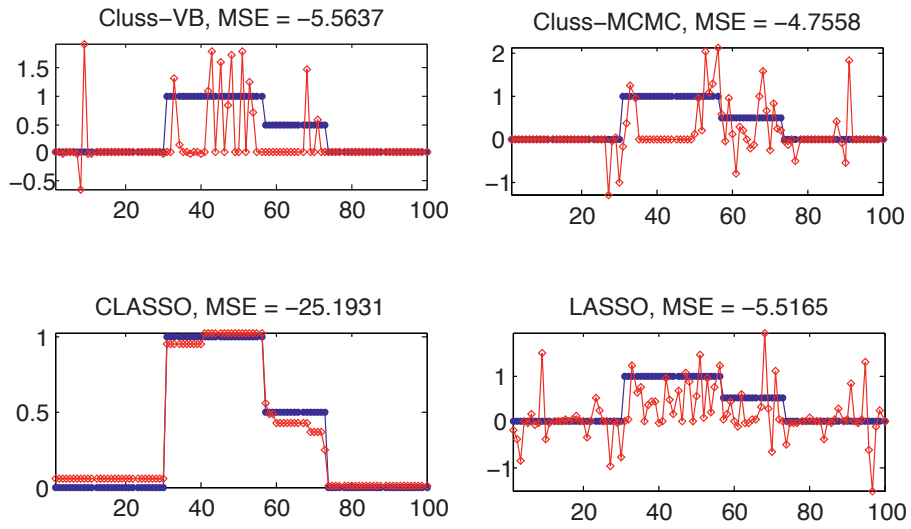

Figure 3.3: Performance of algorithms using signal Type II, where the original signal is represented in blue blue colour and the recovery by the respective algorithms are in red.

A more quantitative comparison can be based on MSE computations and is shown in Figures 3.4–3.5 for Type I and Type II signals, respectively. The figures have been generated by averaging over 10 simulations for each parameter combination.

For Type I signals, the performance difference between the methods is within 1-3 dB for all tested measurement ratios. CLASSO shows only a slight improvement over LASSO for these signals. The alternative methods CluSS-MCMC and CluSS-VB are inferior at low M/N but improvs with increasing M/N where CluSS-VB seems to be the best method among the tested ones.

For Type II signals, CLASSO is almost 10-15 dB better than the others over the whole range of measurement ratios.

As a rough measure of computational complexity, the run time of Matlab implementations of the different methods has been measured. This has shown for signals of Type I in Figure 3.6 Similar results were obtained for Type II.

As expected, our method is somewhat more demanding than LASSO. The complexity of the other two algorithms is one to two orders of magnitude larger.

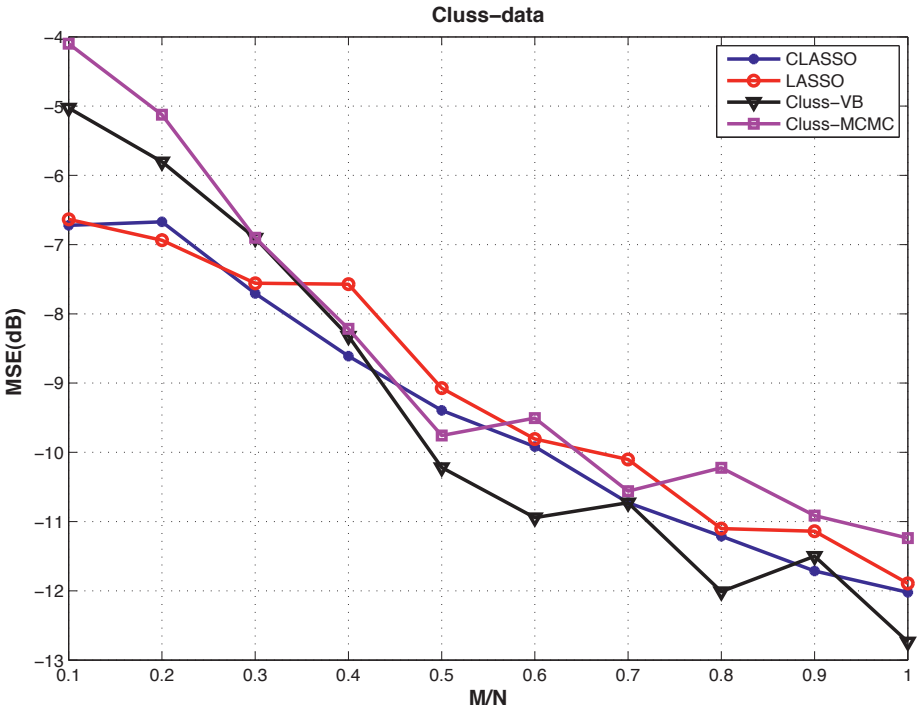

Figure 3.4: MSE performance versus measurement ratio M/N with signal type I.

3.5.3 Performance on a Naturally Sparse Image

Some images (e.g. printed text) are sparse in their native domain. As an example of such an image we have chosen a picture containing six black alphabetic characters on a white background as shown in Figure 3.7. The image consists of 300 columns, each of length $N = 300$. For simplicity, all processing is made columnwise, i.e. only the simple neighbour model (3.3.6) has been assumed. The columns of the image will mostly contain zero pixels, but may have one or more segments consisting of clustered non-zero values. Thus, the performance of the method should be expected to follow the same pattern as the one investigated in Section 3.5.1 above. The number of non zero elements per column varies from $k = 0$ to $k = 122$ distributed among up to six segments. To secure satisfactory performance of LASSO, the number of observations per column should be larger than $2k$, and we have chosen to use $M = 244$. The entries of the A matrix follow the same distribution as used in Section 3.5.1. As an example of performance, we have used a noise variance $\sigma^2 = 0.03$ for this particular image. Now k, and consequently $\| \mathbf{x} \|_2^2$ and $\mathbf{E}[y_i^2]$, varies from column to column, giving a maximum measurement SNR SNR_{m} of 6.4 dB by (3.5.1). The average number of non zero elements is $\overline{k} = 21.6$ resulting in an average SNR_{m} of 0.5364 dB.

For zero-valued columns the SNR definition (3.5.2) is not very useful, so we have found

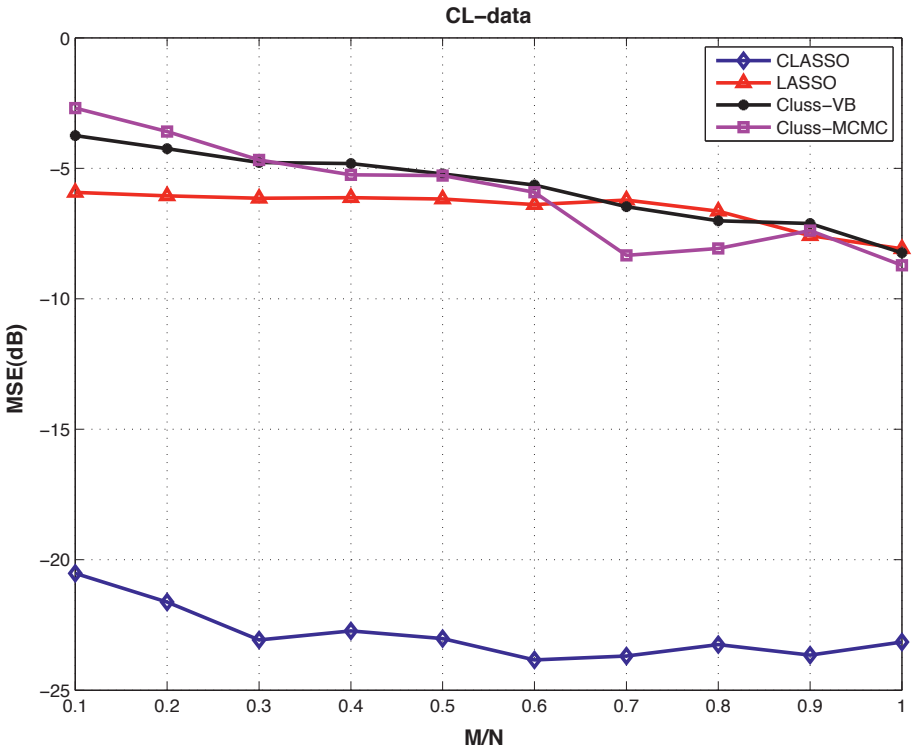

Figure 3.5: MSE performance versus measurement ratio M/N with signal type II.

it natural to use PSNR for comparison, which is also a much used performance metric in image processing. With $\max x_i = 1$ we have

$$PSNR_{\mathrm{r}} = 1/\mathbf{E}[(\hat{x}_i - x_i)^2].$$

The reconstruction PSNR was calculated by averaging over the 300 columns of the image. This resulted in reconstruction PSNR values is given in Table 3.1. Visually, the advantage of taking appropriate clustering into account is visible from Figure 3.7 and Table 3.1.

Again we see that CLASSO clearly outperforms the other methods when the signal shows no or little variability within the clustered regions.

3.5.4 Sparsification

Natural images are most often not sparse in themselves, with a transform to a suitable domain, however, their representation will often be sparse or nearly sparse. Such transformation with a subsequent thresholding resulting in $k << N$ non-zero values may be termed sparsification.

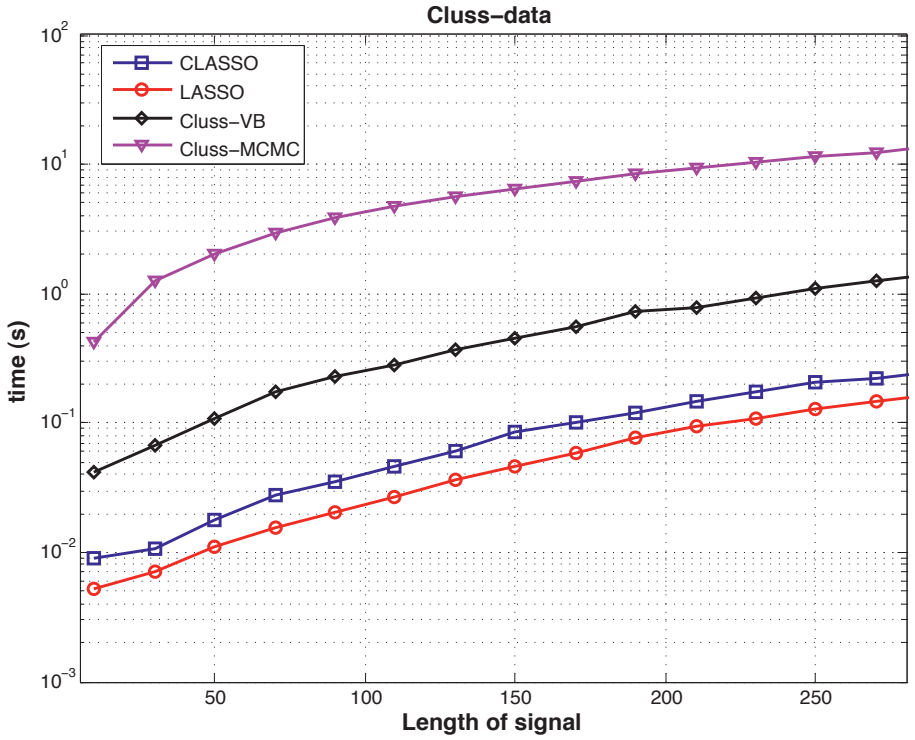

Figure 3.6: Execution time for the algorithms applied to signal Type I: LASSO, CLASSO, Cluss-BV, Cluss-MCMC.

Table 3.1: Performance comparison for ABCDEF-Image

Algorithm	Avarage $PSNR_{\mathrm{r}}$ in dB
LASSO	0.4675
CLASSO	5.485
CluSS	1.21
Cluss-VB	1.847

To demonstrate our method on a sparsified image, we have chosen the *Shepp-Logan phantom,* a much used test image for medical application, as shown in Figure 3.8).

This image consists of 200 columns, each of length $N = 200$. For sparsification, we used a discrete cosine transform (DCT) on each column, obtaining $\mathbf{x}' = DCT(\mathbf{x})$. Each

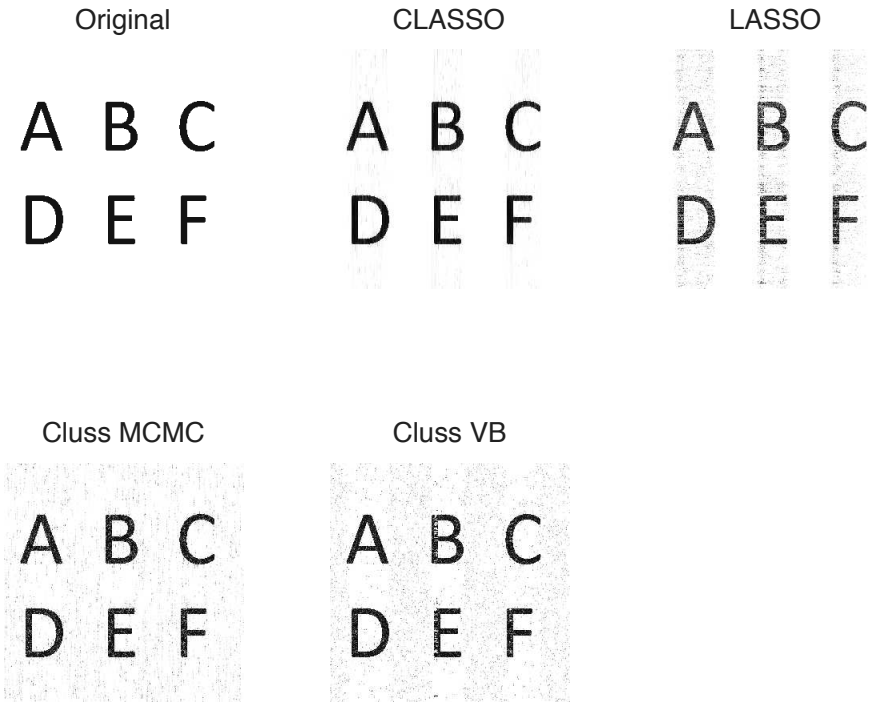

Figure 3.7: Reconstruction of naturally sparse image. a) Original, b) LASSO, c) Modified LASSO

element of $\mathbf{x}'$ with an absolute value less than a threshold was then set to zero, resulting in a sparsified vector $\mathbf{x}''$. The threshold was chosen such that the maximum number of non zero values $k = 56$. The sparsified vector was then subjected to CS measurement by

$$\mathbf{y} = \mathbf{A}\mathbf{x}'' + \mathbf{w}. \tag{3.5.3}$$

Finally, sparsified reconstructions $\hat{\mathbf{x}}''$ were produced by the investigated methods and a final reconstructed signal by taking the inverse cosine transform

$$\hat{\mathbf{x}} = DCT^{-1}(\hat{\mathbf{x}}'').$$

For experiments, $M = 112$ measurements were taken, where the measurement matrix $\mathbf{A}$ was generated with the same statistical properties as in the previous example. All processing was performed columnwise as in Section 3.5.3 with maximum measurement SNR ($PSNR_{\mathrm{m}}$) of 13.44 dB. The mean square error of the reconstruction is given in Table 3.2.

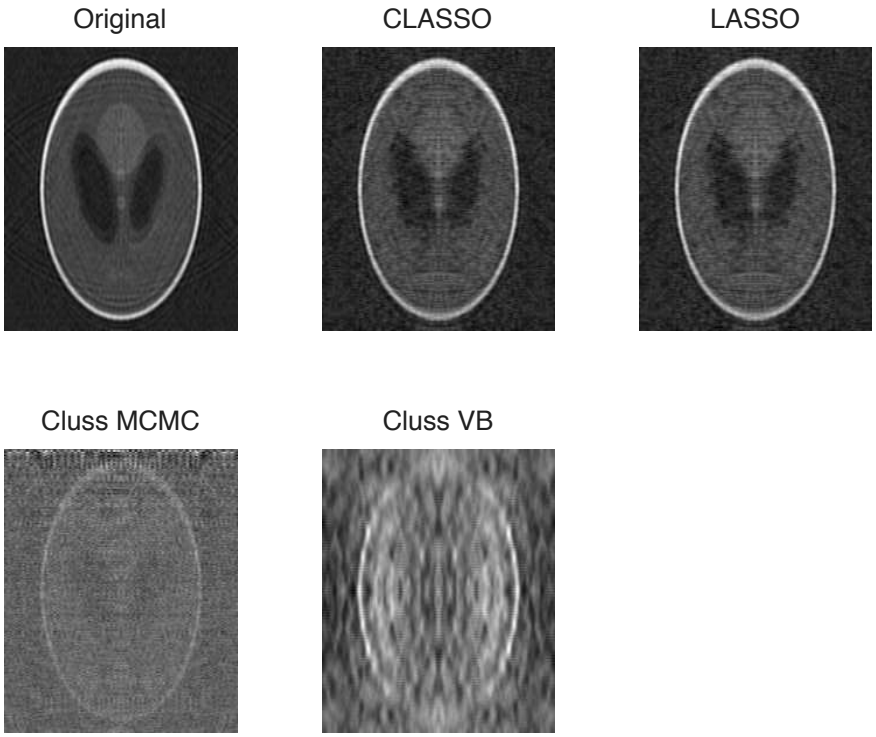

Figure 3.8: Reconstruction of Phantom image using the algorithms CLASSO, LASSO, Cluss-VB and CluSS.

Table 3.2: Performance comparison for Phantom-Image

Algorithm	MSE in dB
LASSO	23.38
CLASSO	23.03
CluSS	37.79
Cluss-VB	32.52

The small difference between LASSO and CLASSO is deemed insignificant, as can also be seen from Figure 3.8. Experiments with noisy measurements showed no further significant difference between the two methods. Investigations of the transformed image reveals that the representation is far from clustered. This can be ascribed to the rather strong

symmetry of the original image. Symmetry in the spatial domain results in signal energy being concentrated in even numbered coefficients in the DCT domain. Consequently, clusteredness is not present, and the advantage over LASSO disappears. However, as is visible from Figure 3.8 and Table 3.2 CLASSO performs better than CluSS, and Cluss-VB. Actually, the two methods break down worst than the proposed method when there is no structured sparsity in the signal, while our method is more robust, since it performs approximately as LASSO.

3.6 Conclusion

We have suggested to model sparsity and clusteredness by a modified Laplacian prior distribution, resulting in a MAP estimator corresponding to a modified LASSO procedure for compressed sensing reconstruction. For a naturally sparse image, an SNR improvement of 10 dB over LASSO under noisy observations has been demonstrated. The method involves the choice of two parameters, and the procedure for optimal choice of these has not been addressed. In experiments values have been chosen based on knowledge of the original signal. It is assumed that the measurements contain sufficient information that appropriate parameter tuning can be performed.

Experiments with sparsification of non-sparse images turned out not to give the same robustness gain. This shows the importance of choosing a representation base in which the signal is both sparse and clustered.

Comparison with other methods highlights that our approach is particularly well suited to clustered signals with little or no variation within the clustered regions, such as two-level images or other binary signals. It should also be noted that, in contrast to methods that *require* clusteredness in order to succeed, our method is fairly robust also with non-clustered signals.

Acknowledgment

The authors give thanks to Professor Ralf R. Müller for valuable suggestions in the early parts of the work.

Chapter 4

Clustered Compressive Sensing via a Bayesian Framework

Solomon A. Tesfamicael and Faraz Barzideh

Adapted from the Proceedings of IEEE UKSIM2015-AMSS 17th International Conference on Modelling and Simulation.

4.1 Abstract

This chapter provides clustered compressive sensing (CCS) based signal processing using a Bayesian framework. Images like magnetic resonance images (MRI) are usually very weak due to the presence of noise and due to the weak nature of the signal itself. The compressive sensing (CS) paradigm can be applied in order to boost such signal recoveries. We applied CS paradigm via a Bayesian framework. That is incorporating the different prior information such as sparsity and the special structure that can be found in such sparse signal improves signal recovery. The method is applied on synthetic and medical images including MRI images. The results show that applying the clustered compressive sensing outperforms the non-clustered but only sparse counter parts when it comes to mean square error(MSE), pick signal to noise ratio (PSNR) and other performance metrics.

4.2 Introduction

Compressive Sensing (CS) is a paradigm to capture information at lower rate than the Nyquist-Shannon sampling rate when signals are sparse in some domain. Actually, it is a revived research topic by authors like Donoho, Candes, Romberg and Tao [1, 2, 3, 12], which was used first in 1970s in Seismology by Claerbout and Muir [6]. It is a paradigm, which tries to find sparse solutions to underdetermined linear systems, and reconstruct signals from far fewer samples than is possible using the Nyquist paradigm. The problem

of limited number of samples or measurements can occur in multiple scenarios, e.g. when we have limitations on the number of data capturing devices, measurements are very expensive or slow to capture such as in magnetic resonance imagining (MRI) [4, 8, 77].

The CS paradigm in signal processing requires three important ingredients [4]. First, the desired signal should have a sparse representation in a known transform domain, i.e., it should be compressible. If the signal is sparse spatially, for example consider an image which is sparse in the pixels, then the transform domain can be the identity. Second, the aliasing artefacts due to undersampling should be incoherent in the transform domain. This creates a noise-like structure. This measurement noise then can be modelled using white Gaussian noise. Third, a nonlinear reconstruction scheme should be used to enforce sparsity and consistency with the data [77]. Recently, this recovery using CS has been shown to be mathematically exact [3, 12].

As a signal processing scheme, CS follows a similar framework: encoding, transmission/storing, and decoding. Focusing on the encoding and decoding of such a system with noisy measurement the block diagram is given in Figure 4.1. At the encoding side, CS combines the sampling and compression stages of a traditional signal processing into one step by measuring few samples that contain maximum information about the signal. This measurement/sampling is done by linear projections using random sensing transformations as shown in the landmark papers by the authors mentioned above. Under conventional sensing paradigm the dimension of the original signal and the measurement should be at least equal. But in CS, the measurement vector can be far less than the original.

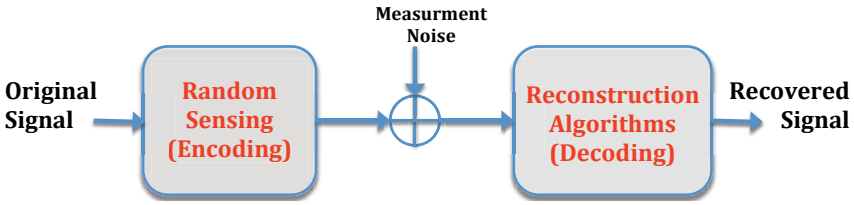

Figure 4.1: Block diagram for CS-based reconstruction.

While at the decoding side, reconstruction is done using nonlinear schemes. Eventually, the reconstruction is more cumbersome than the encoding which was only projections from a large space to a smaller space. On the other hand, finding a unique solution that satisfies the constraint that the signal itself is sparse or sparse in some domain is complex. Fortunately, there are many algorithms to solve the CS problem, such as *convex relaxations* [2, 12], *greedy iterative algorithms* [118], *iterative thresholding algorithms* [23, 25, 119]. However, in this paper, the focus is merely on the convex relaxation methods. We consider a noisy measurement and applied convex relaxation algorithms for robust reconstruction. This is done using a Bayesian framework.

Among the two schools of thought called the classical (also called the frequentist) and the Bayesian in the statistical world, the later is emphasized in this work. The basic difference between the two schools arises from the basic definition of probability. Frequentists define

$P(x)$ as a long-run relative frequency with which x occurs in identical repeats of an experiment. Whereas Bayesian defines $P(x|y)$ as a real number measure of the probability of a proposition x, given the truth of the information represented by proposition y. So under Bayesian theory, probability is considered as an extension of logic. Probabilities represent the investigator's degree of belief- hence it is subjective. That belief or prior information is an integral part of the inference done by the Bayesian [37, 38, 40, 47, 113, 120, 121]. For its flexibility and robustness this work focuses on Bayesian approach. Specifically the prior information like sparsity and clusteredness (or structures on the patterns of sparsity) of signals as two different priors. Therefore, our contribution in this work is to use the Bayesian framework and incorporate two different priors in order to recover the original signals using reconstructing algorithms and in addition we compare different algorithms. Actually, this work is a generalization of the recent work [48, 122, 123].

Therefore, the chapter is organized as follows. In Section 4.3, we present the CS problem and redefine it under Bayesian framework and provide the convex recovery methods. Section 4.4 discusses how the algorithms are implied. Results are shown in section 4.5 for synthetic signals and in Section 4.6 for medical images. Finally, in Section 4.7, conclusions and future work are presented.

4.3 Compressive Sensing-Based Recovery

Beginning with a given vector of measurements $\mathbf{y} \in \mathbb{R}^M$ and measurement matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$, assuming noisy measurement with $\mathbf{w} \in \mathbb{R}^M$ being i.i.d. Gaussian random variables with zero mean and covariance matrix $\sigma^2\mathbf{I}$, reconstructing the sparse vector $\mathbf{x} \in \mathbb{R}^N$ is the problem that we are considering given the linear model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}. \tag{4.3.1}$$

Here $N \gg M$ and $N \gg k$, where k is the number of non-zero entries in $\mathbf{x}$. Applying CS reconstructions using different algorithms we recover the estimate of the original signal $\mathbf{x}$, say $\hat{\mathbf{x}}$. The measurement noise is reduced simultaneously with the reconstruction of the true image data using nonlinear reconstruction schemes.

Various methods for reconstructing $\mathbf{x}$ may be used. We have the least square (LS) estimator in which no prior information is applied:

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}, \tag{4.3.2}$$

which performs very badly for the CS reconstruction problem considered here. Another approach to reconstruct $\mathbf{x}$ is via the solution of the unconstrained optimization problem

$$\hat{\mathbf{x}} = \min_{x\in R^N} \frac{1}{2} \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel_2^2 + uf(\mathbf{x}), \tag{4.3.3}$$

where $uf(\mathbf{x})$ is a regularizing term, for some non-negative u. Emphasis is done on $f(\mathbf{x}) = \|\mathbf{x}\|_p$ as a penalizing norm. In this work, we shall consider when $p = 0, 1$ and 2, which gives us different estimators which we define them here using a Bayesian framework.

4.3.1 Bayesian Framework

Under Bayesian inference consider two random variables $\mathbf{x}$ and $\mathbf{y}$ with probability density function (pdf) $p(\mathbf{x})$ and $p(\mathbf{y})$, respectively. Using Bayes' theorem it is possible to show that the posterior distribution, $p(\mathbf{x}|\mathbf{y})$, is proportional to the product of the likelihood function, $p(\mathbf{y}|\mathbf{x})$, and the prior distribution, $p(\mathbf{x})$,

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}). \tag{4.3.4}$$

Equation (4.3.4) is called the Updating Rule in which the data allows us to update our prior views about $\mathbf{x}$. And as a result we get the posterior which combines both the data and non-data information of $\mathbf{x}$ [37, 47, 48]. Further, the Maximum a posterior (MAP), $\hat{\mathbf{x}}_{MP}$, is given by

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

To proceed further, we assume two prior distributions on $\mathbf{x}$.

4.3.2 Sparse Prior

The reconstruction of $\mathbf{x}$ resulting from the estimator (4.3.3) for the sparse problem, consider in this work, can be presented as a maximum a posteriori (MAP) estimator under the Bayesian framework as in [49]. We show this by defining a prior probability distribution for $\mathbf{x}$ on the form

$$p(\mathbf{x}) = \frac{e^{-uf(\mathbf{x})}}{\int_{\mathbf{x}\in\mathbb{R}^N} e^{-uf(\mathbf{x})} d\mathbf{x}}. \tag{4.3.5}$$

Further, the likelihood function, $p(\mathbf{y}|\mathbf{x})$, can be shown to be

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \mid \mathbf{x}) = \frac{1}{(2\pi\sigma)^{N/2}} e^{-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_{\mathbf{2}}^{\mathbf{2}}}, \tag{4.3.6}$$

the posterior, $p(\mathbf{x}|\mathbf{y})$,

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}; \mathbf{A}) = \frac{e^{-u(\frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2+\lambda f(\mathbf{x}))}}{(2\pi\sigma)^{N/2}\int_{\mathbf{x}\in\mathbb{R}^N} e^{-u(\frac{1}{2\lambda}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2+\lambda f(\mathbf{x}))} d\mathbf{x}}$$

and the MAP estimator is then given by

$$\hat{\mathbf{x}}_{\text{MAP}} = \underset{x\in R^N}{\arg\min} \frac{1}{2} \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel_2^2 + \lambda f(\mathbf{x}), \tag{4.3.7}$$

as shown in [48]. Now, as we chose different regularizing function which enforces sparsity into the vector $\mathbf{x}$, we get different estimators listed below [49].

1. Linear Estimators: when $f(\mathbf{x}) = \| \mathbf{x} \|_2^2$ (4.3.7) reduces to

$$\hat{\mathbf{x}}_{\text{Linear}} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T + \lambda\mathbf{I})^{-1}\mathbf{y}, \tag{4.3.8}$$

   which is the LMMSE estimator. But this estimator is not good enough for sparsity problem since it does not enforce sparsity well. Instead, the following two estimators are more interesting for CS problems.

2. LASSO Estimator: when $f(\mathbf{x}) = \| \mathbf{x} \|_1$ we get the LASSO estimator and (4.3.7) becomes,

$$\hat{\mathbf{x}}_{\text{LASSO}} = \arg\min_{x \in R^N} \frac{1}{2} \| \mathbf{y} - \mathbf{A}\mathbf{x} \|_2^2 + \lambda \| \mathbf{x} \|_1 . \tag{4.3.9}$$

3. Zero-Norm regularization estimator: when $f(\mathbf{x}) = \|\mathbf{x}\|_0$, we get the Zero-Norm regularization estimator and (4.3.7) becomes

$$\hat{\mathbf{x}}_{\text{Zero-Norm}} = \arg\min_{x \in R^N} \frac{1}{2} \| \mathbf{y} - \mathbf{A}\mathbf{x} \|_2^2 + \lambda \| \mathbf{x} \|_0, \tag{4.3.10}$$

This is the best solution for reconstruction of the sparse vector $\mathbf{x}$, but is NP-complete. The worst solution among these l_p-penalizing forms for the sparse problem considered is the l_2-regularization solution given by (4.3.8). However, the best approximation is given by equation (4.3.9) and its equivalent forms such as l_1-norm regularized least-squares ($l_1 - LS$) and others [1, 3, 12].

4.3.3 Clustering Prior

Building on the Bayesian philosophy, we can further assume another prior distribution for clustering. The entries of the sparse vector $\mathbf{x}$ may have some dependency. We define the dependency or structure among the sparse entries as follows: given D_i as neighbor hood of x_i, typically

$$D_i \equiv \{x_{i-1}, x_{i+1}\}, \quad \text{and}$$
$$D_i^+ \equiv \{x_j \in D_i | j > i\}.$$

$x_j \in D_i$ implies that x_j and x_i are dependent. This dependence can be described as a Markov type dependence as

$$p(x_i | D_i \cup \mathbf{E}) = p(x_i | D_i) \tag{4.3.11}$$

for some set $\mathbf{E}$. By chain the rule:

$$\begin{aligned} p(\mathbf{x}) &= p(x_1|x_2, x_3, \cdots, x_N) \cdot p(x_2|x_3, x_4, \cdots, x_N) \cdot \\ &\quad p(x_3|x_4, x_5, \cdots, x_N) \cdot \cdots p(x_{N-1}|x_N) \cdot p(x_N) \\ &= p(x_1|D_i^+) \cdot p(x_2|D_2^+) \cdots (x_{N-1}|D_{N-1}^+) \cdot p(x_N) \end{aligned} \tag{4.3.12}$$

while

$$p(x_i|D_i^+) = e^{-\left[-\lambda|x_i|+\gamma\sum_{x\in D_i^+}^{N}|x_i-x|\right]}. \tag{4.3.13}$$

Now, by (4.3.11) and (4.3.13)

$$\begin{aligned} p(\mathbf{x}) &= e^{-\left[-\lambda\sum_{j=1}^{N-1}|x_i|+\gamma\sum_{j=1}^{N-1}\sum_{x\in D_i^+}^{N}|x_i-x|\right]} \\ &= e^{-\lambda\|\mathbf{x}\|_1|-\gamma\sum_{i=1}^{N-1}|x_i-x_{i+1}|}, \end{aligned} \tag{4.3.14}$$

and by using this new prior and similar arguments as used in 4.3.2 we arrive at the Clustered LASSO estimator

$$\begin{aligned} \hat{\mathbf{x}}_{\text{CLASSO}} = \arg\min_{x\in R^N} &\frac{1}{2} \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel_2^2 + \lambda \parallel \mathbf{x} \parallel_1 \\ &+ \gamma \sum_{i=2}^{N} |x_i - x_{i-1}|, \end{aligned} \tag{4.3.15}$$

where λ, γ are our tuning parameters for the sparsity in $\mathbf{x}$ and the way the entries are clustered, respectively.

4.4 Implementation of the Analysis

In this chapter, we have considered the convex optimization algorithms for solving the CS recovery problem and we extended these algorithms by involving clusteredness among the sparse entries of the original signal. Given relatively few measurements $\mathbf{y}$ contaminated by measurement noise $\mathbf{w}$ and measurement matrix $\mathbf{A}$ we reconstruct the original signal $\mathbf{x}$. For recovery we used LMMSE, LASSO and clustered LASSO given by equations (4.3.8), (4.3.9),(4.3.15) respectively. In the Equations (4.3.8) , (4.3.9), and (4.3.15) we have parameters like λ and γ. As we have based our analysis in the Bayesian framework we could have assumed some prior distributions on each of them, and build hierarchical Bayesian compressive sensing. Instead we have used them as a tuning parameter for the constraint and we have used them in the optimal way.

The optimal λ value for the LMMSE in (4.3.8), that is $\lambda = 1e-07$. In implementing (4.3.9), that is least square optimization with L1 regularization, we have used the Quadrature programming with constraints similar to Tibshirani [8, 25] by solving

$$\begin{aligned} \hat{\mathbf{x}} &= \min_x \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \\ &\text{subject to } \|\mathbf{x}\|_1 \leq t \end{aligned} \tag{4.4.1}$$

instead of (4.3.9). t and λ are inversely related. In addition, Equation (4.3.15) is implemented similar to LASSO with additional term on the constraint as follows:

$$\begin{aligned} \hat{\mathbf{x}} &= \min_x \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \\ &\text{subject to } \|\mathbf{x}\|_1 \leq t \text{ and } D(\mathbf{x}) \leq d. \end{aligned} \tag{4.4.2}$$

d and γ are inversely related. This is done to put a constrain on the neighbouring elements. Since we have vectorized the image for the sake of efficiency of the algorithm, the penalizing terms are applied columnwise. Other ways of implementing (constraining) are also possible. But we differ it for future work. In our simulations we have used optimal values of these constraints. Figure 4.2 and 4.3 show the respective optimal values for one of the simulations in the next section.

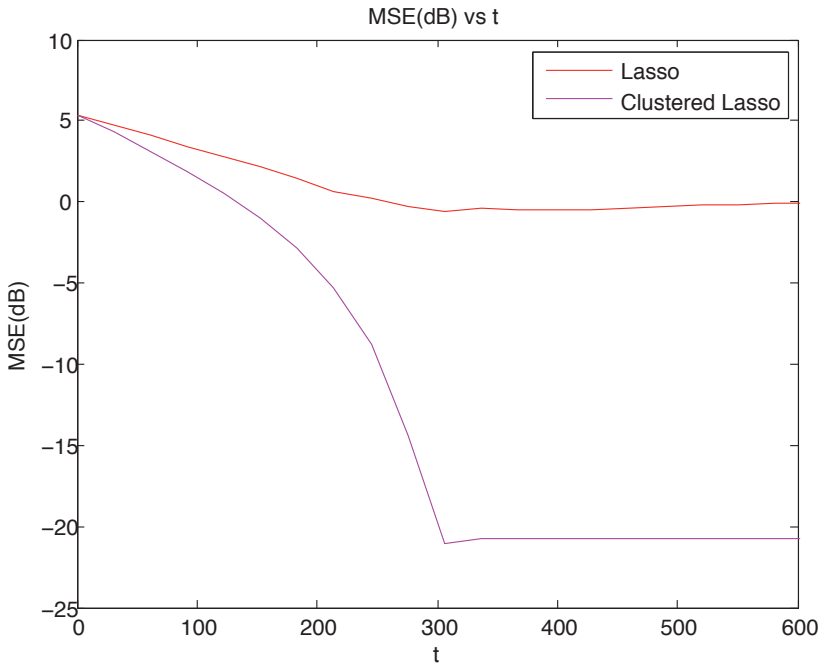

Figure 4.2: This figure shows the MSE of LASSO and clustered LASSO for different values of t for Figure 4.4. It can be seen that there is only one optimal value.

4.5 Results on Synthetic Data

The main focus of this work is to give a practical application of clustered compressed sensing. In order to verify the theory we have selected different medical related images. We used LS, LMMSE, LASSO and clustered LASSO given by Equations (4.3.2), (4.3.8), (4.3.9), and (4.3.15), respectively, to reconstruct from a noisy measurement and compare their performances. We apply our analysis to angiogram, phantom and then to functional MRI (fMRI) images.

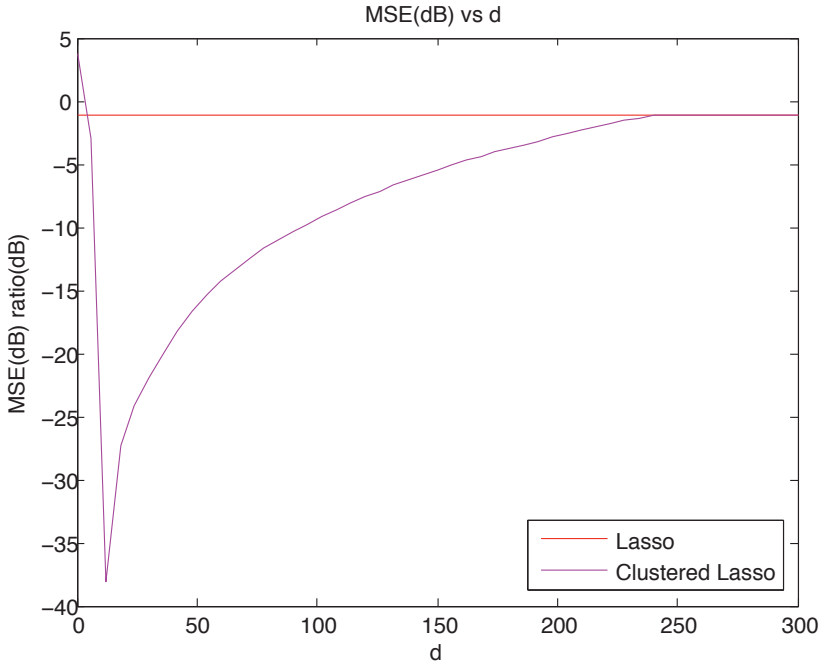

Figure 4.3: This figure shows the MSE of LASSO and clustered LASSO for different values of d for Figure 4.4. It can be seen that there is only one optimal value d and by loosening the constraint clustered LASSO will converge to LASSO.

4.5.1 First Set of Synthetic Data

In order to demonstrate the performance of reconstruction of the sparse signal presented in the chapter we have used synthetic data. The first set of data is a vector with entries that are sparse and clustered in the spatial domain. We have applied Gaussian noise with mean zero and variance $\sigma^2 = 1$ and random matrix $\mathbf{A}$ with Gaussian entries with variance $\sigma^2 = 1$. For LMMSE we used $\lambda = 0$ in our simulations. However, we have used equivalent constraints for λ and γ for the LASSO and clustered LASSO. The original signal is $\mathbf{x}$ of length $N = 300$ and we added noise to it. By taking 204 measurements, that is $\mathbf{y}$ is of length $M = 204$, and maximum number of non-zero elements $k = 102$, we applied different reconstruction techniques such as LMMSE, LASSO and clustered LASSO. The results are shown in Figure 4.4 and Table 4.1.

4.5.2 Second Set of Synthetic Data

The second set of data is an image with a single English letter, where the image itself is sparse and clustered in the spatial domain(in the pixel). We have applied Gaussian noise

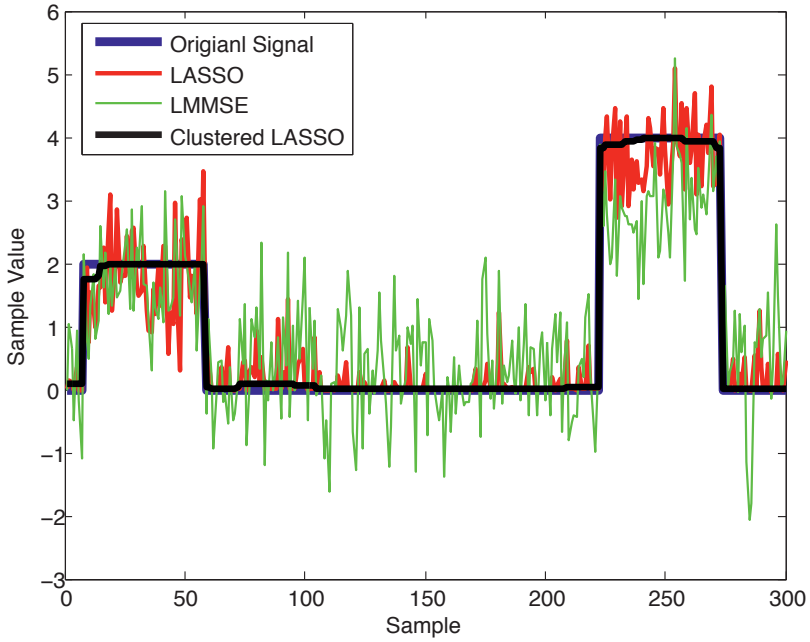

Figure 4.4: Application of different reconstruction techniques discussed in the paper (in their vertical order: Original signal, LMMSE, LASSO, clustered LASSO).

Table 4.1: Performance comparison for the first set of Synthetic Data

Algorithm	SNR in dB
LMMSE	5.0208
LASSO	19.3526
Clustered LASSO	33.8917

with mean zero and variance $\sigma^2 = 0.25$ and random matrix $\mathbf{A}$ with Gaussian entries with variance $\sigma^2 = 1$. For LMMSE we used $\lambda = 1e - 07$ in our simulations. However, we have used equivalent constraints for λ and γ for the LASSO and clustered LASSO.

The original signal after vectorization is $\mathbf{x}$ of length $N = 300$ and we added noise to it. By taking 185 measurements, that is $\mathbf{y}$ is of length $M = 185$, and minimum number of non-zero elements $k = 0$ maximum number of non-zero elements $k = 84$, and average $k = 28.71$, we applied different reconstruction techniques such as LMMSE, LASSO and clustered LASSO. The results are shown in Figure 4.5 and Table 4.2.

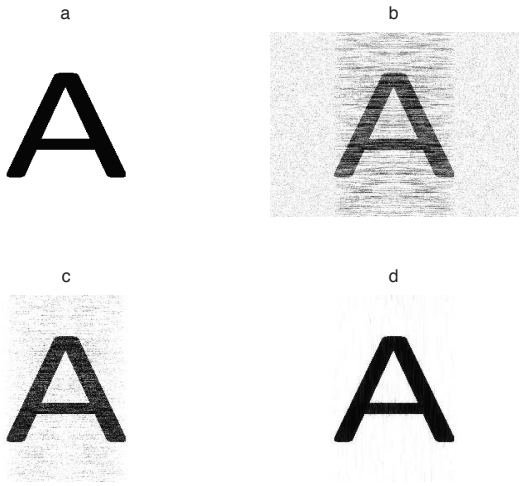

Figure 4.5: Comparison of reconstruction schemes: a) Original image $\mathbf{x}$ b) LMMSE c) LASSO d) Clustered LASSO.

Table 4.2: Performance comparison for the second set of Synthetic Data

Algorithm	MSE in dB
LMMSE	-25.2224
LASSO	-37.0353
Clustered LASSO	-63.9963

4.5.3 Third Set of Synthetic Data

The third set of data is an image with a several English letters, where the image itself is sparse and clustered in the spatial domain but more evolving in terms of the sparseness and clusteredness in the data than the second one. We have applied Gaussian noise with mean zero and variance $\sigma^2 = 0.25$ and random matrix $\mathbf{A}$ with Gaussian entries with variance $\sigma^2 = 1$. For LMMSE we used $\lambda = 1e-07$ in our simulations. However, we have used equivalent constraints for λ and γ for the LASSO and clustered LASSO.

The original signal after vectorization is $\mathbf{x}$ of length $N = 300$ and we added noise to it. By taking 269 measurements, that is $\mathbf{y}$ is of length $M = 269$, and minimum number of non-zero elements $k = 0$ maximum number of non-zero elements $k = 122$, and average $k = 21.56$, and applied the reconstructing algorithms used above. Note that, we needed more measurements in this case than the previous because we have more non-zero elements. The results are shown in Figure 4.6 and Table 4.3. The results in Figures 4.4-4.10

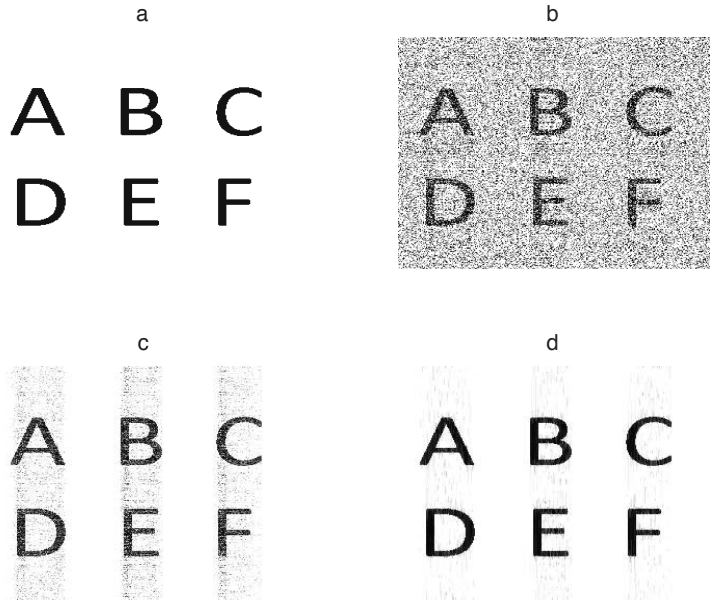

Figure 4.6: Comparison of reconstruction schemes: a) Original image x b) Noisy measurement c) LASSO d) Clustered LASSO.

Table 4.3: Performance comparison for the third set of Synthetic Data

Algorithm	MSE in dB
LMMSE	-15.3121
LASSO	-36.6692
Clustered LASSO	-52.7103

and Tables 4.1-4.3 showed that reconstruction using LASSO is much better than LS and LMMSE algorithms for the sparse reconstruction problem. Further, clustered LASSO outperforms LASSO since it uses more accurate information about the structure of the sparsity.

4.6 Results on Medical Images

To give another practical application of clustered compressed sensing method, it is applied on different medical related images: Angiogram, phantom and then to functional MRI (fMRI) images.

4.6.1 Angiogram Image

The first one is an angiogram image taken from University Hospital Rechts der Isar, Munich, Germany [124]. Angiogram images are already sparse in the pixel representation. In general MRI images are sparse (and even clustered) in the spatial and the transformed domain. The image we took is also clustered as well. The original signal after vectorization is $\mathbf{x}$ of length $N = 960$. By taking 746 measurements, and maximum number of non-zero elements $k = 373$, we applied the different reconstruction schemes and the results are shown in Figure 4.7 and Table 4.4.

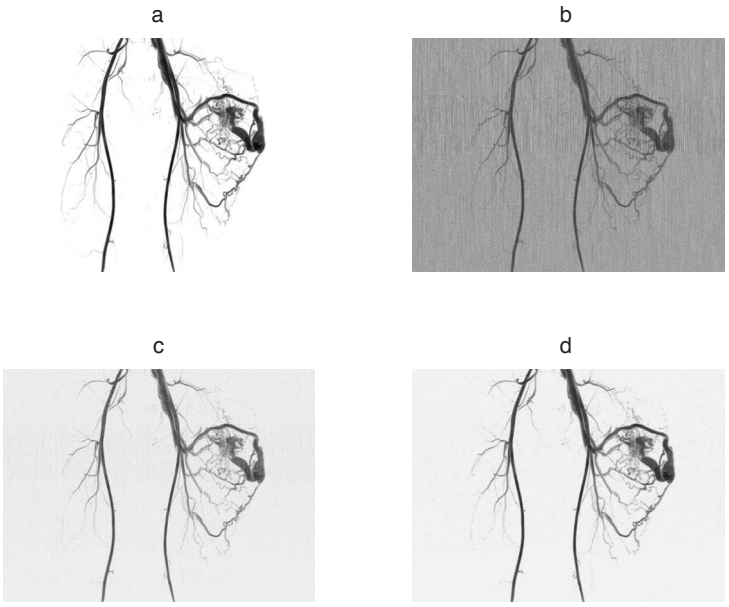

Figure 4.7: Comparison of reconstruction schemes: a) Original image $\mathbf{x}$ b) LMMSE c) LASSO d) Clustered LASSO.

Table 4.4: Performance comparison using MSE

Algorithm	MSE in dB
LMMSE	-35.1988
LASSO	-53.6195
Clustered LASSO	-63.6889

4.6.2 Phantom Image

Consider the Shepp-Logan phantom which is not sparse in the spatial domain but can be sparsified in K-space by zeroing out small coefficients. We then measured the sparsified image and added noise. The original signal after vectorization is $\mathbf{x}$ of length $N = 200$. By taking 94 measurements, that is $\mathbf{y}$ is of length $M = 94$, and maximum number of non-zero elements $k = 47$, we applied the different reconstruction algorithms used above. The result shows clustered LASSO does well compared to the others as can be seen in Figure 4.8 and Table 4.5.

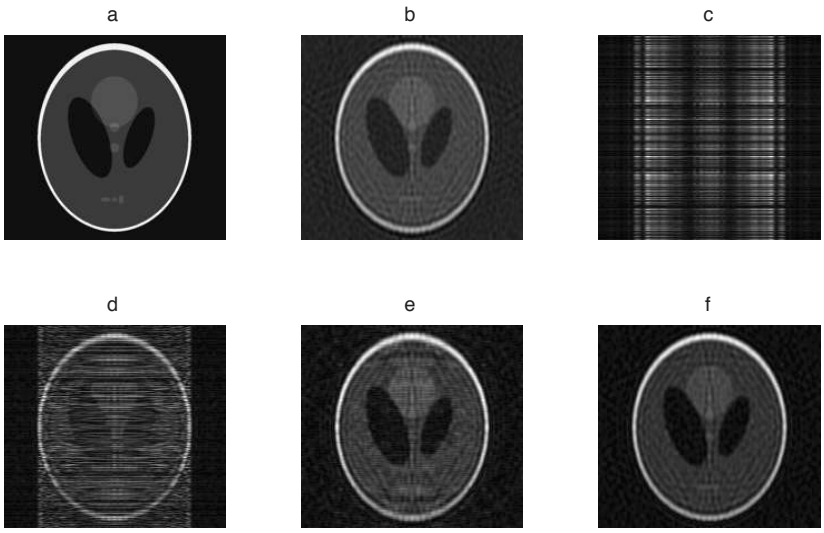

Figure 4.8: Comparison of reconstruction schemes: a) Original image $\mathbf{x}$ b) sparsified image c) Least Square (LS) d) LMMSE e) LASSO f) Clustered LASSO.

Table 4.5: Performance comparison of Phantom Image

Algorithm	MSE in dB
LS	-21.3304
LMMSE	-27.3876
LASSO	-37.9978
Clustered LASSO	-40.0068

4.6.3 fMRI Image

Another example to apply the clustered LASSO based image reconstruction using Bayesian framework to medical images is a functional MRI (fMRI) image. Taken from the slices of fMRI image, as it is shown in Figure 4.9 in the transform domain, the fMRI image in Figure 4.10 has presented the sparsity and clusteredness properties. That gives grounds to apply the framework and the procedure used here. The performance of the different reconstruction schemes is visible from Figure 4.10.

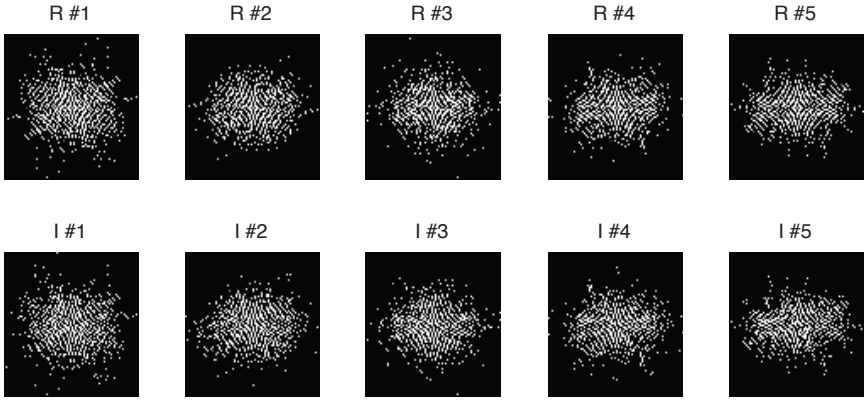

Figure 4.9: The five column images represent the real and imaginary part of the Fourier transform representation of the data set we have chosen to present further, which in general shows that the fMRI image have sparse and clustered representation.

In addition, for synthetic data we have compared the different recovery techniques by using pick signal to noise ratio (PSNR) versus measurement ratio (M/N) and the result is shown in Figure 4.11. Generally, reconstruction using LASSO is much better than the LS and LMMSE algorithms for the sparse reconstruction problem. Further, clustered

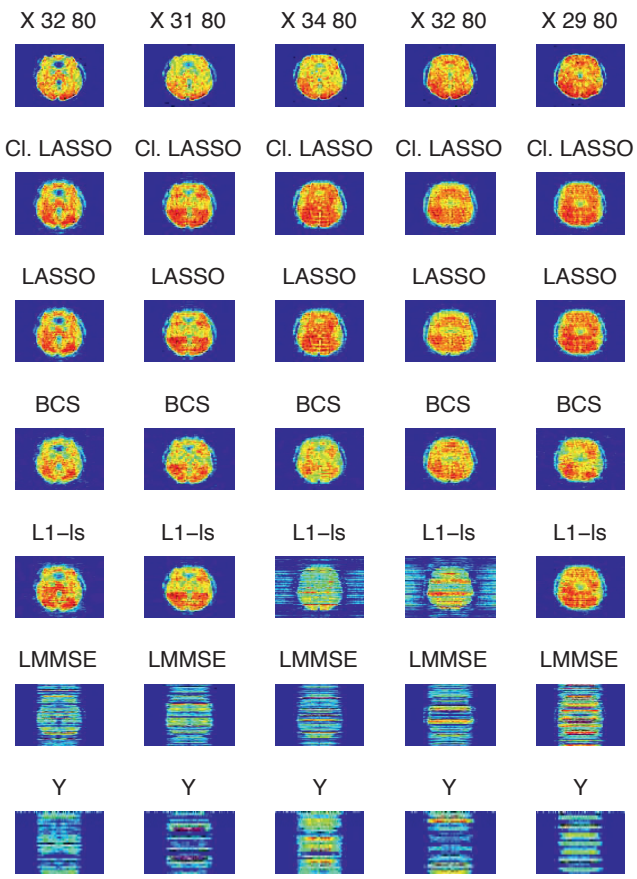

Figure 4.10: Application of sparse and cluster prior on a fMRI data analysis:$N = 80$ and $k > 50$

LASSO outperforms LASSO since it uses more accurate information about the structure of the sparsity. Furthermore, we see the impact of sparsity ratio on the performances of the reconstruction schemes in Figures 4.12 and 4.13 using the reconstruction ratio versus the sparsity ratio, k/N and M/N, respectively.

In addition, we show the robustness of the LASSO and clustered LASSO reconstruction schemes where we have applied different noise levels. The result, as given in Figure 4.14, reveals that both are robust but clustered LASSO is better.

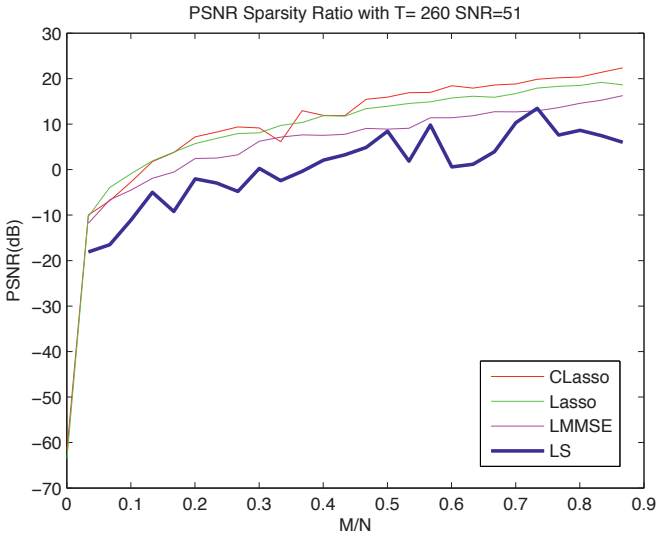

Figure 4.11: Here LS is included in the figure in addition to LMMSE, LASSO and CLASSO.

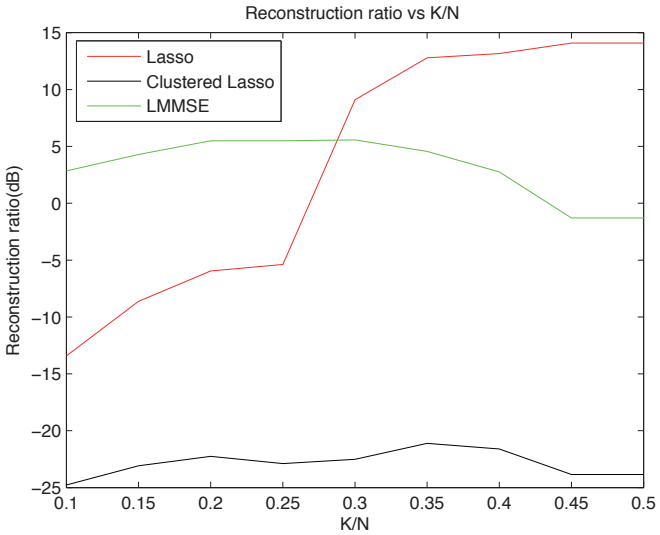

Figure 4.12: Comparison of different reconstruction algorithms using the metric reconstruction ratio versus and sparsity ratio, k/N.

4.7 Conclusions

In this chapter, clustered compressive sensing using a Bayesian framework is presented. Our emphasis in this work is to incorporate prior information like sparseness and clus-

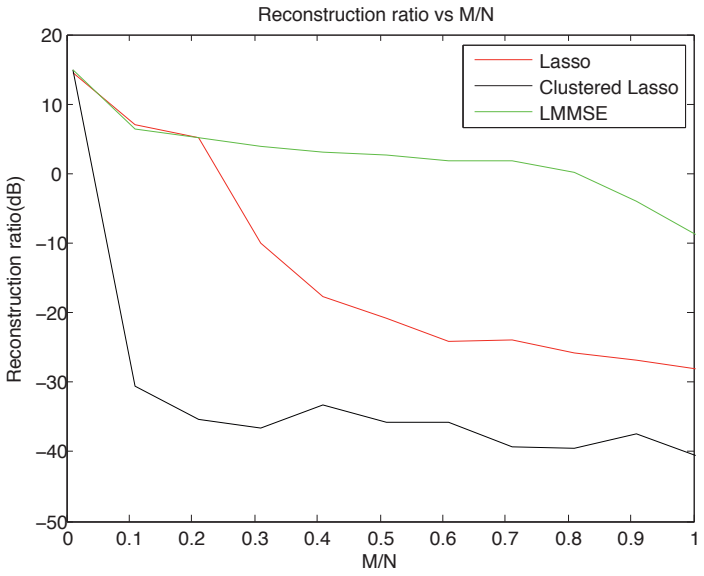

Figure 4.13: Comparison of different reconstruction algorithms using the metric reconstruction ratio versus and sparsity ratio, M/N.

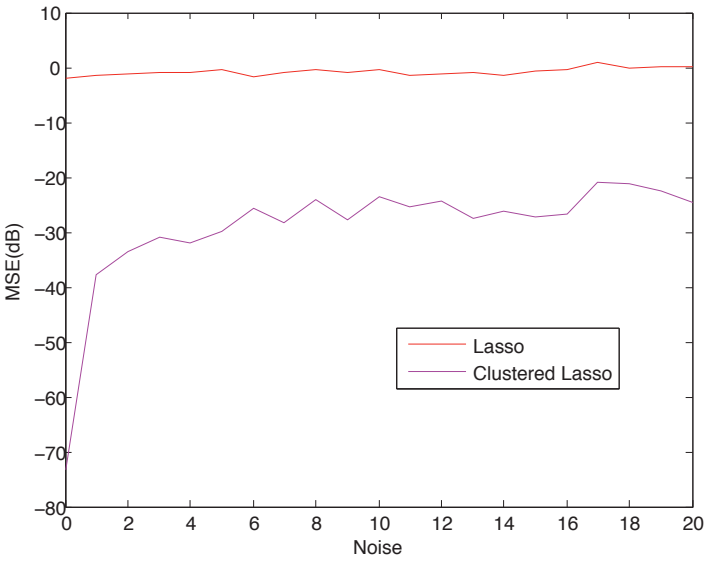

Figure 4.14: Robustness of LASSO and Clustered LASSO reconstruction schemes.

teredness in the reconstruction of signals from fewer measurements. We apply it on different medical related images. Clustered LASSO recovery does well in terms of PSNR and MSE than LASSO (using only sparse prior), LMMSE and LS. In addition, in this work we have shown comparison of the performance of different reconstruction algorithms for different amounts of measurement ratio versus PSNR and MSE. In addition sparsity ratio and measurement ratio versus reconstruction ratio is provided to see how the schemes behave with the amount of sparsity and measurement. For future work we plan to apply different forms of clustering depending on the prior informations of images or geometry of clusteredness and compare them to state of the art algorithms in the area.

Chapter 5

Compressed Sensing-Based Rotative Quantization in Temporally Correlated MIMO Channels

Solomon A. Tesfamicael and Lars Lundheim

Adapted from the Proceedings of Recent Developments on Signal Processing (RDSP), September 2013.

5.1 Abstract

Channel adaptive transmission requires knowledge of channel state information at the transmitter. In temporally correlated MIMO channels, the correlation can be utilized to reduce feedback overhead and improve performance. In this paper, Compressed Sensing (CS) methods and rotative quantization are used to compress and feedback channel state information for MIMO systems as an extension work of [125]. Using simulation, it is shown that the CS-based method reduces feedback overhead while delivering the same performance as the direct quantization scheme.

5.2 Introduction

In modern wireless communications multiple-input multiple-output (MIMO) systems are integrated due to their advantage in improving performance with respect to many performance metrics. One of the advantages is the ability to transmit multiple streams using spatial multiplexing [126]. However, one needs channel state information (CSI) at the transmitter in order to get optimal system performance [127]. In frequency division duplexing (FDD) MIMO, a dedicated feedback channel of limited capacity is usually

assumed. Several limited feedback strategies are proposed using codebooks which are known to the transmitter and receiver [128, 129, 130, 131, 132, 133].

Temporal correlation of wireless channels can be used to reduce the feedback requirement in limited feedback systems. One technique to reduce the feedback requirement in temporally correlated channels is to quantize the rotative change of singular vectors. For instance, differential rotation feedback is proposed in [128].

In [125], scalar quantization using adaptive range is used to utilize the temporal correlation. This paper, extends the work done in [125] by introducing the concept of compressed sensing (CS) for rotative quantization methods. The near-sparse nature of the rotation matrices is used to reduce the feedback requirement by using compressed sensing based coding and decoding.

Recently questions like, why go to so much effort to acquire all the data when most of what we get will be thrown away? Can we not just directly measure the part that will not end up being thrown away? that were posed by Donoho [1, 3] and others, triggered a new way of sampling or sensing called compact ("compressed") sensing (CS). In CS the task is to estimate or recover a sparse or compressible vector $\mathbf{x} \in \mathbb{R}^N$ from a measurement vector $\mathbf{y} \in \mathbb{R}^M$ [2, 3]. These are related through the linear transform $\mathbf{y} = \mathbf{A}\mathbf{x}$. Here, $\mathbf{x}$ is a sparse vector and $M \ll N$. In the seminal papers [1] and [3], $\mathbf{x}$ is estimated from $\mathbf{y}$, by the algorithm: min $||\mathbf{x}||_0$ such that $\mathbf{y} = \mathbf{A}\mathbf{x}$, This is a non-convex NP-complete. The usual wisdom is to solve it using approximation with min $||\mathbf{x}||_1$ such that $\mathbf{y} = \mathbf{A}\mathbf{x}$, which is a convex optimization problem [30], [49]. One of the most famous approaches is the l_1-regularized least square or LASSO and we used this estimator as a recovery algorithm in this paper.

The concept of applying compressed sensing for limited (compressed) feedback of parameters of the channel is well known [134, 135, 136]. In this work though, we combine rotative quantization and CS to reduce the feedback overhead in temporally correlated MIMO channels.

This chapter is organized as follows. First we give the MIMO system model that we work with in Section 5.3. In Section 5.4, we review the concept of rotative quantization. In Section 5.5, it is shown how rotative quantization can be combined with CS to reduce the feedback overhead. Then in Section 5.6, the performance of the proposed method is shown using simulations. The last section gives conclusions and recommendations for future work.

5.3 System Model

Considering a frequency division duplex (FDD) MIMO system consisting of N_t transmit and N_r receive antennas we assume that the channel is a flat-fading, temporally correlated channel denoted by a matrix $\mathbf{H}[n] \in \mathbb{C}^{N_r \times N_t}$ where n indicates a channel feedback time index with block fading assumed during the feedback interval. Applying Singular Value Decomposition (SVD) of $\mathbf{H}[n]$ gives $\mathbf{H}[n] = \mathbf{U}[n]\mathbf{\Sigma}[n]\mathbf{V}^H[n]$, where $\mathbf{U} \in \mathbb{C}^{N_r \times r}$ and

$\mathbf{V} \in \mathbb{C}^{N_t \times r}$ are unitary matrices and $\mathbf{\Sigma} \in \mathbb{C}^{r \times r}$ is a diagonal matrix consisting of $r = \min(N_t, N_r)$ singular values.

In the presence of perfect channel state information (CSI) a MIMO system model can be given by the equation

$$\tilde{\mathbf{y}} = \mathbf{U}^H[n]\mathbf{H}[n]\mathbf{V}[n]\tilde{\mathbf{x}} + \mathbf{U}^H[n]\mathbf{n} \tag{5.3.1}$$

where $\tilde{\mathbf{x}} \in \mathbb{C}^{r \times 1}$ is transmitted vector, $\mathbf{V}[n]$ is used as precoder at the transmitter, $\mathbf{U}^H[n]$ is used as decoder at the reciver, $\mathbf{n} \in \mathbb{C}^{N_r \times 1}$ denotes a noise vector whose entries are i.i.d. and distributed according to $\mathcal{CN}(0, 1)$ and $\tilde{\mathbf{y}} \in \mathbb{C}^{N_r \times 1}$ is the received vector.

In practice, partial channel state information is available at the transmitter, hence we assume that only a quantized version $\hat{\mathbf{V}}[n]$ is available, Further, assuming a generalized receiver $\mathbf{R}[n]$, (5.3.1) becomes:

$$\tilde{\mathbf{y}} = \mathbf{R}^H[n]\mathbf{H}[n]\hat{\mathbf{V}}[n]\tilde{\mathbf{x}} + \mathbf{U}^H[n]\mathbf{n}. \tag{5.3.2}$$

In this chapter we consider two different alternatives for $\mathbf{R}[n]$. Assuming a minimum mean square error (MMSE) approach we get

$$\mathbf{R}[n] = \left[(\mathbf{H}[n]\hat{\mathbf{V}}[n])^H\mathbf{H}[n]\hat{\mathbf{V}}[n] + r\sigma_n^2\mathbf{I}\right]^{-1}(\mathbf{H}[n]\hat{\mathbf{V}}[n])^H.$$

Alternatively, a Matched Filter (MF) receiver gives

$$\mathbf{R}[n] = (\mathbf{H}[n]\hat{\mathbf{V}}[n])^H.$$

The receiver estimates the channel from pilot symbols, computes SVD, quantizes and then feedbacks $\tilde{\mathbf{V}}[n]$.

Further we assume a first-order Gauss-Markov process to model the channel variation in a channel with temporal correlation as used in [125] and[126] given by the equation

$$\mathbf{H}[n] = \eta\mathbf{H}[n-1] + \sqrt{1-\eta^2}\mathbf{G}[n] \tag{5.3.3}$$

where η is the temporal correlation and $\mathbf{G}[n] \in \mathbb{C}^{N_r \times N_t}$ denotes the innovation process having i.i.d. entries distributed according to $\mathcal{CN}(0, 1)$. η is given by $\eta = J_0(2\pi f_d \tau)$, where $J_0(\cdot)$ is the zero order Bessel function of the first kind, τ is the channel feedback interval, f_d is the Doppler frequency and $f_d\tau$ is the normalized Doppler frequency.

5.4 Rotation-Based Quantization

In the temporally correlated environment we apply a rotation-based limited feedback system for a Rayleigh flat fading MIMO channels as in [128]. The differential rotation of the precoder matrix at a time index n, $\mathbf{V}[n]$, compared to the one at a previous time instant,$\mathbf{V}[n-1]$ is quantized. This has advantage of reducing the feedback overhead.

The precoder matrix $\mathbf{V}[n]$ can be represented equivalently as

$$\mathbf{V}[n] = \mathbf{T}[n]\tilde{\mathbf{I}}_{N_t \times r} \tag{5.4.1}$$

where, $\mathbf{T}[n] \in \mathbb{C}^{N_t \times N_t}$ is a matrix containing all singular vectors and $\tilde{\mathbf{I}}[n]$ is a matrix composed of the first m columns of the identity matrix $\mathbf{I}_{N_t \times N_t}$. As in [128] we assume that both receiver and transmitter has a common estimate $\hat{\mathbf{T}}[n]$ of $\mathbf{T}[n]$. Then we can define a rotation matrix

$$\mathbf{\Theta}\,[n] = \hat{\mathbf{T}}^H[n-1]\mathbf{V}[n] \tag{5.4.2}$$

where, $\mathbf{\Theta}\,[n] \in \mathbb{C}^{N_t \times r}$. A quantized version $\hat{\mathbf{\Theta}}[n]$ is then fed back to the transmitter where an estimate

$$\hat{\mathbf{V}}[n] = \hat{\mathbf{T}}[n-1]\hat{\mathbf{\Theta}}[n] \tag{5.4.3}$$

can be reconstructed.

In order to quantize $\mathbf{\Theta}\,[n]$ various methods can be used. Vector quantization of this matrix is used in [128], parameterization using Givens rotations is also used in [125]. Before introducing our new method, we present a baseline approach for comparison. This is Algorithm 1 below, where $\mathbf{\Theta}\,[n]$ is vectorized and vector quantization of the resulting vector is used. The codebook used for vector quantization consists of vectors uniformly distributed in a $N_t \times r$ dimensional unit hypersphere. We have assumed that the feedback channel is error free and the only inaccuracy comes from quantization error.

Algorithm 1 Rotative Quantization feedback using CS for temporally correlated MIMO channels

---

1. **Initialization:** (both transmitter and receiver)
Set $\tilde{\mathbf{T}}[0] = \mathbf{I}_{N \times N}$.

2. For each time index $n \geq 1$:
Receiver:
Obtain $\mathbf{\Theta}\,[n] = \hat{\mathbf{T}}^H[n-1]\mathbf{V}[n]$.
Update $\hat{\mathbf{T}}[n]$.
Perform vector quantization of $\hat{\mathbf{\Theta}}[n]$.

Transmitter: Re-construct $\hat{\mathbf{\Theta}}[n]$ from received parameters.
Obtain the estimate of the current singular vector matrix using $\hat{\mathbf{V}}[n] = \hat{\mathbf{T}}[n-1]\hat{\mathbf{\Theta}}[n]$.

To simplify the analysis we assume $N_r = N_t = N$, for the rest of the chapter.

5.5 Quantization using Compressed Sensing

We will now modify Algorithm 1 above by using CS instead of direct quantization of $\hat{\mathbf{\Theta}}[n]$. The approach is summarized in Algorithm 2 and Figure5.1 below. The first step

is to arrange the entries of $\mathbf{\Theta}\,[n]$ in a column vector $\mathbf{x}$. We denote this operator by $\mathbf{x} = \text{vec}(\mathbf{\Theta}\,[n])$. Assuming strong correlation, $\hat{\mathbf{\Theta}}[n]$ will be close to diagonal and $\mathbf{x}[n]$ sparse. Next, we apply a random fat matrix $\mathbf{A}$, which is known both to the transmitter and the receiver, to get another vector $\mathbf{y} = \mathbf{A}\mathbf{x}$, which has much less dimension than the orignal vector $\mathbf{x}$. A quantized version $\hat{\mathbf{y}}$ is sent through the feedback channel. The receiver reconstructs $\hat{\mathbf{x}}$ from the received vector $\hat{\mathbf{y}}$ and the matrix $\mathbf{A}$ using LASSO:

$$\hat{\mathbf{x}} = \underset{x}{argmin}\ \ ||\hat{\mathbf{y}} - \mathbf{A}\mathbf{x}||_2^2 + ||\mathbf{x}||_1. \tag{5.5.1}$$

Then we apply the reverse process of the vectorization, unvec$(\hat{\mathbf{x}})$, and we get $\hat{\mathbf{\Theta}}[n]$, which is an estimate of $\mathbf{\Theta}\,[n]$. Finally, since we are interested in estimating $\mathbf{V}[n]$, we can derive it from $\hat{\mathbf{V}}[n] = \hat{\mathbf{T}}[n-1]\hat{\mathbf{\Theta}}[n]$.

Algorithm 2 Rotative Quantization feedback using CS for temporally correlated MIMO channels

---

For each time index $n \geq 1$:

Receiver:

Vectorize $\mathbf{\Theta}\,[n]$ obtained from algorithm 1, $\mathbf{x} = \text{vec}(\mathbf{\Theta}\,[n])$.

CS encoding $\mathbf{y} = \mathbf{A}\mathbf{x}$.

Quantize $\mathbf{y}$ and feedback the resulting $\hat{\mathbf{y}}$.

Transmitter:

Recover using LASSO, $\hat{\mathbf{x}} = \underset{x}{argmin}\ \ ||\hat{\mathbf{y}} - \mathbf{A}\mathbf{x}||_2^2 + ||\mathbf{x}||_1$.

Unvectorize $\hat{\mathbf{x}}$ to get $\hat{\mathbf{\Theta}}[n]$.

Obtain $\hat{\mathbf{V}}[n] = \mathbf{T}[n-1]\hat{\mathbf{\Theta}}[n]$ (as in algorithm 1).

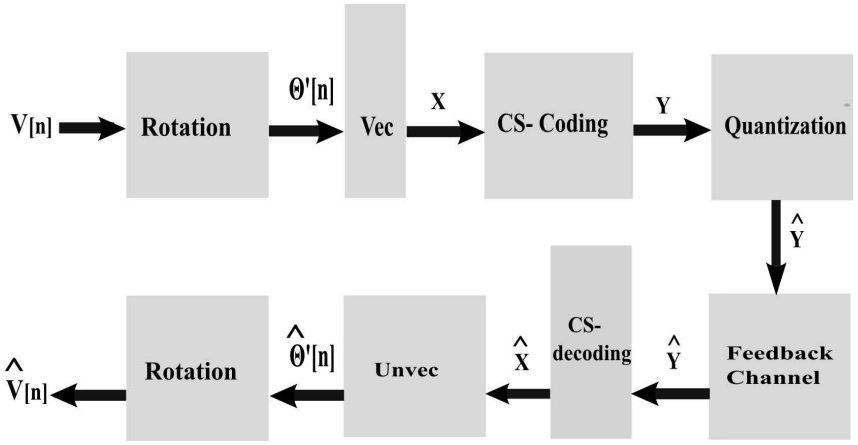

Figure 5.1: Algorithm 2

5.6 Results

In order to verify the proposed algorithm, we consider a 2×2 MIMO channel with a temporal correlation of 0.98. Spatial streams are assumed to be transmitted with equal power allocation in the 2×2 MIMO system. Unitary precoding is applied based on the feedback, and matched filter or MMSE equalizers are applied at the receiver.

Three methods are compared in the simulations in Figure 5.2, Figure5.3 and Figure 5.4. The first is the perfect channel state information scenario. The second is using Algorithm 1. The third method is using Algorithm 2.

In Figure 5.2, and Figure 5.3, sum rates are compared against signal-to-noise-ratio (SNR); result using both matched filter and MMSE receivers is shown. In the second method, a total feedback bits $B = 10$ are used. On the other hand, the CS method uses half the number of bits, $B = 5$. We can observe that the performance of the CS method is almost equal to that of the method without using CS while saving half the number of bits.

The advantage of CS can also be confirmed from the result in Figure5.5, where the bit-error-rate is plotted but in this case the CS and without CS using same number of bits. In this case, we observe that the CS method has a better bit error rate performance. These two figures demonstrate the clear advantage of using CS in feedback of singular vectors in the rotative-based method.

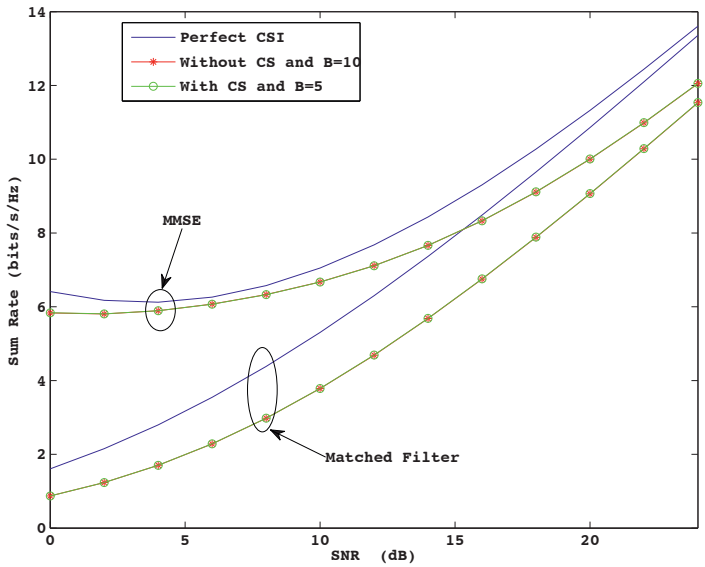

Figure 5.2: Sum rate vs. SNR for a 2×2 MIMO system with and without CS with two streams,

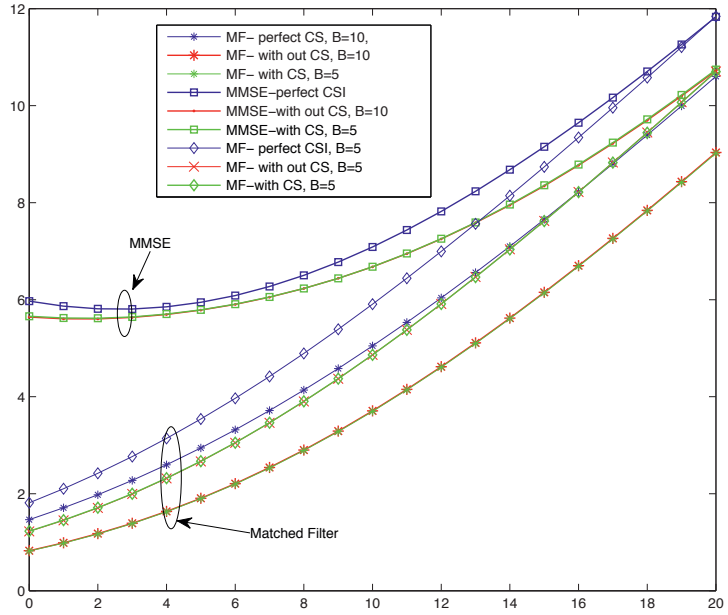

Figure 5.3: Sum rate vs. SNR for a 2×2 MIMO system with and without CS with two streams,

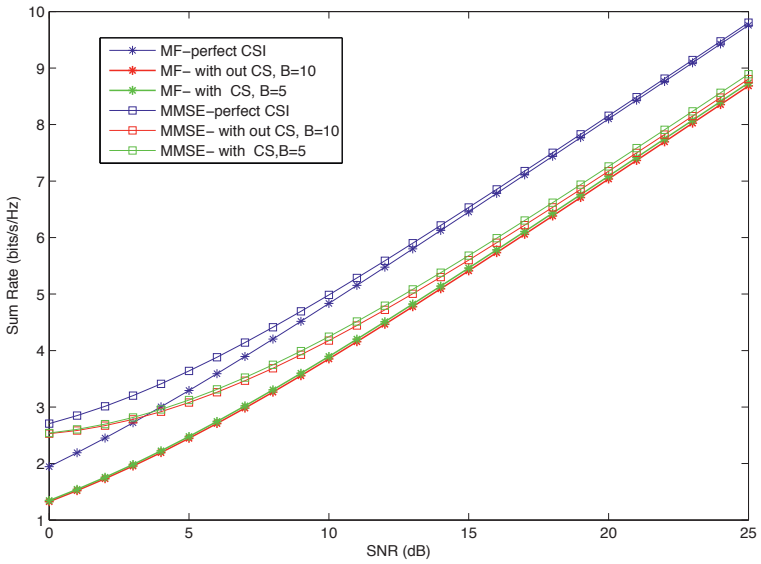

Figure 5.4: Sum rate vs. SNR for a 2×2 MIMO system with and without CS with one stream,

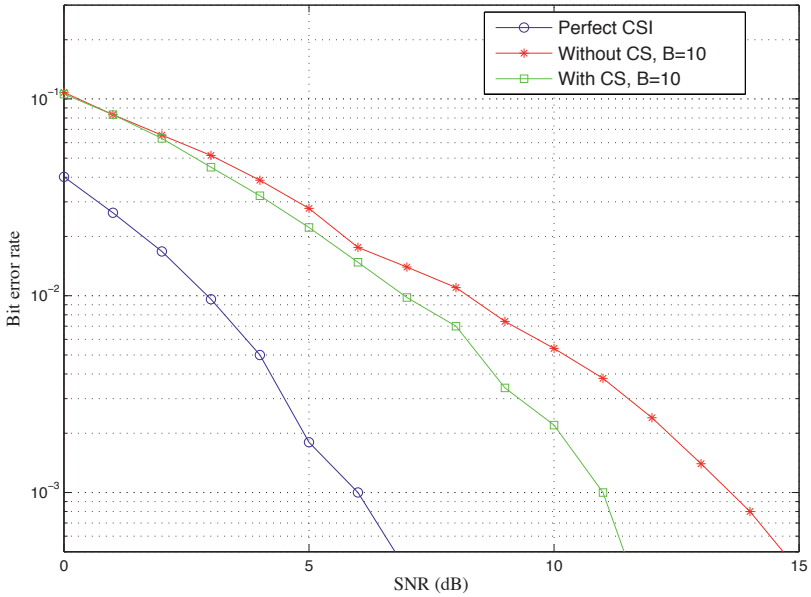

Figure 5.5: Bit error rate vs. SNR using matched filter receiver for a 2×2 MIMO system with one stream,

5.7 Conclusion

In this chapter, the concept of compressed sensing (CS) is applied to limited feedback in temporally correlated MIMO channels. The near-sparse nature of the rotation matrices is utilized to combine techniques from CS with rotative quantization for reducing feedback overhead. Simulations show that the use of CS reduces feedback overhead significantly while delivering the same performance. On the other hand, CS-based feedback improves performance as compared to direct quantization for the same feedback overhead. CS-based limited feedback is in general a promising method that can be applied in various scenarios taking advantage of the sparse nature of the elements to be quantized.

5.8 Acknowlegment

We are grateful to Bruhtesfa Godana for interesting discussions and suggestions.

Chapter 6

Conclusions and Future work

6.1 Summary and Discussion

In Chapter 1, a new paradigm in signal processing, called compressive sensing (CS), is presented. The fundamental concepts, the standard CS problem, together with how reconstructions are done, were explained. The conditions for perfect recovery of sparse signals were described using concepts like coherence and restricted isometric properties of matrices. Two major categories of recovery algorithms were also identified: convex and greedy algorithms. The use of convex relaxation methods were adhered to rather than the greedy algorithms, due to the analytical tools that came later in the research. Examples of how recovery algorithms performance is done in the literature are given.

In addition, Bayesian theory was introduced briefly since much emphasis was given to inference based on prior and posterior distribution of signals in the thesis. Based on the Bayesian inference a very powerful analytical tool borrowed from statistical mechanics, called the replica method, was also presented in this chapter. Considering a non-noisy CS problem, an example of performance analysis via the replica method was provided showing the phase transition when different convex relaxation-based recovery algorithms, like l_p-reconstruction schemes, were used. For convex optimization algorithms like, the l_1-reconstruction, the replica symmetry ansatz provide gives good results. However, for l_0-reconstruction, which is non-convex, the need for the use of the replica symmetry breaking ansatz is motivated in the chapter. Further, application of compressed sensing theory in imaging and communication is also presented briefly as an extension of research done elsewhere before concluding the chapter with the scope and the contribution of the thesis.

In Chapter 2, the performance of the estimators used in compressed sensing which are generalized as MAP estimators were done analytically using the statistical mechanics tool. This employed the replica method, to extend results from the SPAM project to CS systems for large systems with M/N fixed, where N and M are the original and measured signal dimensions, respectively. Beginning from the posterior distribution of the CS system and redefining it as Boltzmann-Gibbs distribution in the context of statistical mechanics terms, the free energy of the system is calculated via the partition function. From the

free energy other important quantities like the energy of the system can be calculated. The analysis is done for the LASSO and Zero-Norm regularizing estimator, where replica symmetry is sufficient for the former while the later requires at least one-step replica breaking ansatz (1RSB). This is shown only for one example, i.e. for Bernoulli-Gaussian Mixture Distribution due to time and financial constraints in the project.

In Chapter 3, a new method of compressive sensing reconstruction was presented assuming the signal to be estimated as both sparse and clustered. Under the Bayesian framework the two priors are modelled as a modified Laplacian prior distribution, resulting a modified LASSO reconstruction algorithm. The results were compared with other contemporary methods that consider clusteredness and this method gives better performance for some sparse and structured signals. While in Chapter 4, a similar algorithm was called clustered LASSO and it was applied on a different synthetic data and medical images such as angiogram, phantom and functional magnetic resonance images (fMRI). It was shown that the performance of clustered LASSO is much better in terms of the metrics PSNR and MSE than the traditional LASSO and other equivalent algorithms.

In Chapter 5, an application of CS on communication was provided. The near-sparse nature of the rotation matrices in temporally correlated MIMO channels has given the opportunity to apply CS theory to reduce feedback overhead with out reducing the performance of the system. This indicates that the CS method can have a great impact in future communication systems.

6.2 Contributions of the Thesis

This thesis deals with the contemporary signal processing technique called CS. As a signal processing scheme, it follows a similar framework: encoding, transmission/storing, and decoding, where the encoding part is done using random projection (RP) or random sensing, and the decoding is done via nonlinear reconstruction algorithms from fewer measurements. The performance of the reconstruction schemes used and the application of such paradigm is the focus in this thesis. Saying that, the thesis contributes the following particular results:

- Analytical performance analysis of noisy CS systems is done using the replica method including one step replica breaking ansatz (1RSB) by interpreting the convex relaxation-based recovery algorithms as MAP estimators as proposed by [49]. The idea here is that for non-convex recovery algorithms, where the replica symmetry ansatz may not give a precise result, and in such cases 1RSB ansatz may be enough to provide us a better analysis as shown for the MIMO case in [55]. However, it was only shown for one particular example for the CS problem, i.e., for Bernoulli-Gaussian distribution.
- A new method of compressive sensing was suggested to model sparsity and clusteredness by a modified Laplacian prior distribution, resulting in a MAP estimator corresponding to a modified LASSO procedure for compressed sensing reconstruction. For a naturally sparse image, an SNR improvement of 10 dB over LASSO

under noisy observations has been demonstrated.

- Additional information such as the structure of the sparse entries can be exploited in order to boost signal recovery. This has been demonstrated for synthetic and known medical images using Bayesian framework. This can have great impact in reducing scanning time and improving reconstructed images for better diagnosis in medical imaging. Using the performance metrics like PSNR and MSE it was demonstrated that clustered compressive sensing (CSS) performs much better than the non-clustered sparse counterparts.

- The concept of compressed sensing applied to limited feedback in temporally correlated MIMO channels can reduce the feedback overhead. A novel algorithm was proposed in the Rotative Quantization feedback using CS for temporally correlated MIMO channels. The near-sparse nature of the rotation matrices is utilized to combine techniques from CS with rotative quantization. Simulations show that the use of CS reduces the feedback overhead significantly while delivering the same performance. On the other hand, CS-based feedback improves performance as compared to direct quantization for the same feedback overhead. CS-based limited feedback is in general a promising method that can be applied in various scenarios taking advantage of the sparse nature of the elements to be quantized.

6.3 Future Work

- Though analytical performance analysis using replica method using the replica symmetry (RS) and one step replica breaking ansatz (1RSB) is done. However, the numerical simulation of it was far from simple to show it due to lack of funds and shortage of time. The idea of 1RSB analysis to MAP estimators was done for other systems like MIMO and CDMA prior to 2011 when this analysis was done. Extending it to CS was the task and it gives a better understanding if numerical simulations of different distributions are done at that time since in this thesis it is done only for one particular example, i.e., Bernoulli-Gaussian distribution. It may be interesting to compare estimators based on other metrics such as the input/output distribution as done in [55], other than using the free energy. In addition, it is also possible to see such analysis for the greedy algorithms.

- To study the connection of other tools from statistical mechanics, such as *cavity method*, *gauge theory* and others to CS systems as replica method are applied in CS and other systems.

- The method modified LASSO for compressed sensing reconstruction presented in the thesis was dependent on the choice of two parameters. The procedure for optimal choice of these parameters has not been addressed. This can be a bottleneck for the performance of the algorithm if there is a lack of prior knowledge about the

data. Building on a robust parameter tuning algorithm overcoming this can also be an interesting problem to work with, especially, in the medical imaging area.

- In clustered LASSO recovery, simple clustering function was used and one can develop other clustering functions that will model effectively the particular signal type. Therefore, for future work one can apply different forms of clustering depending on the prior information of images or geometry of clusteredness.

- It is interesting to see CS-based feedback coupled with other types of feedback schemes for example with Adaptive Rotative Quantization (ADRQ) proposed by [125].

Part IV

Appendices

Appendix A

Appendices to Chapter 1

A.1 Proof of Theorem 2

Consider an two arbitrary k-sparse vectors $\mathbf{x}$ and $\mathbf{z}$ and define

$$\theta_x = \frac{\mathbf{A}(\mathbf{z}-\mathbf{x})}{2} \quad and \quad \theta_z = \frac{\mathbf{A}(\mathbf{x}-\mathbf{z})}{2}, \tag{A.1.1}$$

and note that

$$\mathbf{A}\mathbf{x} + \theta_x = \mathbf{A}\mathbf{x} + \theta_z = \frac{\mathbf{A}(\mathbf{x}+\mathbf{z})}{2}. \tag{A.1.2}$$

Let $\hat{\mathbf{x}} = \Delta(\mathbf{A}\mathbf{x} + \theta_x) = \Delta(\mathbf{A}\mathbf{z} + \theta_z)$. From triangle inequality and the definition of C-stability, we have that

$$\begin{aligned}
||\mathbf{x}-\mathbf{z}||_2 &= ||\mathbf{x}-\hat{\mathbf{x}}+\hat{\mathbf{x}}-\mathbf{z}||_2 && \text{(A.1.3)}\\
&\leq ||\mathbf{x}-\hat{\mathbf{x}}||_2 + ||\hat{\mathbf{x}}-\mathbf{z}||_2 && \text{(A.1.4)}\\
&\leq C||\mathbf{w}_x||_2 + C||\mathbf{w}_z||_2 && \text{(A.1.5)}\\
&= 2C||\frac{\mathbf{A}(\mathbf{x}-\mathbf{z})}{2}||_2 && \text{(A.1.6)}\\
&= C||\mathbf{A}\mathbf{x}-\mathbf{A}\mathbf{z}||_2. && \text{(A.1.7)}
\end{aligned}$$

Since this holds for any k-sparse vectors $\mathbf{x}$ and $\mathbf{z}$, the results holds. Completes the proof (see also [137]).

A.2 Proof of Theorem 1.1.18

Let $\Lambda \subset \{1, 2, 3, \cdots n\}$ be a subset of indices and $\mathcal{B}(\mathbf{y}) = \{\mathbf{x} : ||\mathbf{A}\mathbf{x}-\mathbf{y}||_2 \leq \epsilon\}$. We are interested in bounding $||\mathbf{h}||_2 = ||\hat{\mathbf{x}}-\mathbf{x}||_2$. Since $||\mathbf{w}||_2 \leq \epsilon$, $\mathbf{x} \in \mathcal{B}(\mathbf{y})$, and therefore we know that $||\hat{\mathbf{x}}|| \leq ||\mathbf{x}||_1$. Thus we apply the result in Lemma 1.6 in [8] for the noiseless CS

problem, where $\mathbf{h}$ is bounded

$$||\mathbf{h}||_2 \leq C_0 \frac{\sigma_k(\mathbf{x})}{\sqrt{k}} + C_1 \frac{|\langle \mathbf{A}\mathbf{h}_\Lambda, \mathbf{A}\mathbf{h}\rangle|}{||h_\Lambda||_2} \tag{A.2.1}$$

where

$$C_0 = 2\frac{1-(1-\sqrt{2})\delta_{2k}}{1-(1+\sqrt{2k}}, \quad C_1 = \frac{2}{1-(1+\sqrt{2k}} \tag{A.2.2}$$

Now it remains to bound $|\langle \mathbf{A}\mathbf{h}_\Lambda, \mathbf{A}\mathbf{h}\rangle|$. In order to do that

$$\begin{aligned} ||\mathbf{A}\mathbf{h}||_2 &= ||\mathbf{A}(\hat{\mathbf{x}} - \mathbf{x})||_2 && \text{(A.2.3)} \\ &= ||\mathbf{A}\hat{\mathbf{x}} - \mathbf{y} + \mathbf{y} - \mathbf{A}\mathbf{x})||_2 && \text{(A.2.4)} \\ &\leq ||\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}||_2 + ||\mathbf{y} - \mathbf{A}\mathbf{x})||_2 \leq 2\epsilon. && \text{(A.2.5)} \end{aligned}$$

where the last inequality follows since $\mathbf{x}$, $\hat{\mathbf{x}} \in \mathcal{B}(y)$. Combining this with the RIP (1.1.6) and the Cauchy-Schwarz inequality we obtain

$$|\langle \mathbf{A}\mathbf{h}_\Lambda, \mathbf{A}\mathbf{h}\rangle| \leq ||\mathbf{A}\mathbf{h}_\Lambda||_2 ||\mathbf{A}\mathbf{h}||_2 \leq 2\epsilon\sqrt{1+\delta_{2k}}||\mathbf{h}_\Lambda||_2. \tag{A.2.6}$$

Thus,

$$\begin{aligned} ||\mathbf{h}||_2 &\leq C_0 \frac{\sigma_k(\mathbf{x})}{\sqrt{k}} + C_1 2\epsilon\sqrt{1+\delta_{2k}} && \text{(A.2.7)} \\ &= C_0 \frac{\sigma_k(\mathbf{x})}{\sqrt{k}} + C_2\epsilon. && \text{(A.2.8)} \end{aligned}$$

Completes the proof.

A.3 Proof of Equations 1.2.6 and 1.2.7.

The posterior distribution is calculated from the likelihood and the prior pdf as follows:

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}; \mathbf{A}) = \frac{p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\int p_{\mathbf{y}|\tilde{\mathbf{x}}}(\mathbf{y}|\tilde{\mathbf{x}})p(\tilde{\mathbf{x}})d\tilde{\mathbf{x}}}. \tag{A.3.1}$$

After plugging the prior pdf, $p(\mathbf{x})$,

$$p(\mathbf{x}) = \frac{e^{-uf(\mathbf{x})}}{\int_{\mathbf{x}\in\mathbb{R}^N} e^{-uf(\mathbf{x})}d\mathbf{x}} \tag{A.3.2}$$

and the likelihood function

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \mid \mathbf{x}) = \frac{1}{(2\pi\sigma)^{N/2}} e^{-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2}. \tag{A.3.3}$$

in A.3.1, the posterior pdf becomes

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}; \mathbf{A}) = \frac{\left(e^{-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_{\mathbf{2}}^{\mathbf{2}}}\right)\left(e^{-uf(\mathbf{x})}\right)}{(2\pi\sigma)^{N/2}\int_{\tilde{\mathbf{x}}\in\mathbb{R}^N} e^{-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{A}\tilde{\mathbf{x}}\|_{\mathbf{2}}^{\mathbf{2}}} e^{-uf(\tilde{\mathbf{x}})} d\tilde{\mathbf{x}}}$$
$$= \frac{e^{-u(\frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2+\lambda f(\mathbf{x}))}}{(2\pi\sigma)^{N/2}\int_{\tilde{\mathbf{x}}\in\mathbb{R}^N} e^{-u(\frac{1}{2\lambda}\|\mathbf{y}-\mathbf{A}\tilde{\mathbf{x}}\|_2^2+\lambda f(\tilde{\mathbf{x}}))} d\tilde{\mathbf{x}}}. \tag{A.3.4}$$

Further, we find the maximum a posterior (MAP) of the distribution which gives the most probable value for the parameters, denoted as $\hat{\mathbf{x}}_{MP}$, is calculated as

$$\begin{aligned}\hat{\mathbf{x}}_{MP} &= \arg\max_{\mathbf{x}} \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\int_{\tilde{\mathbf{x}}} p(\mathbf{y}|\tilde{\mathbf{x}})p(\tilde{\mathbf{x}})d\tilde{\mathbf{x}}} \\ &\overset{(a)}{=} \arg\max_{\mathbf{x}}\ e^{-u(\frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2+\lambda f(\mathbf{x}))} \\ &\overset{(b)}{=} \arg\min_{\mathbf{x}} \frac{1}{2} \parallel \mathbf{y}-\mathbf{A}\mathbf{x} \parallel_2^2 +\lambda f(\mathbf{x}).\end{aligned} \tag{A.3.5}$$

Note that at (a) we used the fact that the denominator is constant and finding the $\mathbf{x}$ that maximises is not dependent on it. Where as at (b), maximising the exponential function is the same as finding the maximum of the argument and further, maximising the negative argument is the as minimizing the opposite. Hence, 1.2.6 and 1.2.7 are proved here as it is shown by A.3.4 and A.3.5 , respectively.

A.4 Basic Concepts from Statistical Mechanics

Macroscopic versus Microscopic properties

Large size systems can be characterized by physical terms like "macroscopic" and "microscopic" properties. Let us take different examples of such physical terms. The magnetization of a magnet is a macroscopic property but the interactions of the magnetic spins can be a microscopic property. Temperature or pressure of a gas is a macroscopic property whereas the trajectories of the particles in the gas can be a microscopic property. A particular macroscopic property in the context of the thesis can be the norm of a sparse vector $\mathbf{x}$ or the posterior distribution $\mathbf{x}$ given $\mathbf{y}$. In general, whatever inference we do on the vector will be on the macro level, however if one attempts to see the combination of entries of the sparse vector minimizing some other macroscopic characteristics of the given system, then one is at micro level [66], [53], [67].

Self-Averaging

Obtaining the macroscopic characterizations of large disordered systems on the basis of their microscopic specifications is a central purpose in large size systems study. One fundamental property that comes in between these two characteristics of a large system is *self-averaging*. Large systems that can be approximated by self-averaging are called *thermodynamic systems*. When a system is at a state of thermodynamic equilibrium, it means that the system can be described by using deterministic value. That is, the probabilistic value converges to the deterministic one. In other words, by assuming self-averaging on the microscopic properties we can describe the macroscopic properties of a system. Actually the replica method is a tool that helps to infer about the macroscopic properties by using self averaging over the microscopic properties. In fact, to apply replica method the system is assumed to be at thermodynamic equilibrium [66], [53], [67].

Partition Function

The statistical mechanics of the Boltzman-Gibbs distributuion is given by

$$P_\beta(\mathbf{x}|\mathbf{y}) = \frac{1}{\mathcal{Z}} e^{-\beta \mathcal{H}(\mathbf{x})}, \tag{A.4.1}$$

as in (1.3.5) for the CS system under consideration, where $\mathcal{Z}(\beta; \mathbf{y})$ is the *partition function*,

$$\mathcal{Z}(\beta; \mathbf{y}) = \int e^{-\beta \|\mathbf{x}\|_p} \delta(\mathbf{y} - \mathbf{A}\mathbf{x}) d\mathbf{x}. \tag{A.4.2}$$

This function is an important quantity from which other parameters of a given physical system can be computed (see below).

Free Energy

One of the important macroscopic quantities in statistical mechanics which is also a self averaging quantity is *free energy* [56], [36], [66], [53], [67] [55] and [54], defined as

$$\mathcal{F} = \mathcal{E} - \beta^{-1} \mathcal{P}, \tag{A.4.3}$$

where $\mathcal{E}$ and $\mathcal{P}$ are the energy and the entropy of the system, respectively. Further, these three quantities: the free energy, the energy, and the entropy, can be derived from the partition function $\mathcal{Z}$ as follows:

$$\mathcal{F} = -\beta^{-1} \log \mathcal{Z} \tag{A.4.4}$$

$$\mathcal{E} = \partial(\mathcal{F}\beta\,)/\partial\beta \quad and \quad \mathcal{P} = -\partial\mathcal{F}/\partial(\frac{1}{\beta}). \tag{A.4.5}$$

Since free energy is a self averaging quantity it can be re-written as

$$\mathcal{F} = -\beta^{-1} \mathbb{E}(\log \mathcal{Z}). \tag{A.4.6}$$

To calculate $\mathbb{E}(\log \mathcal{Z})$ we need the replica method.

Replica Method

Let us as assume that we are interested in evaluating $\mathbb{E}(\ln \mathcal{Z})$, in relation to the average free energy as shown above in Equation (A.4.6). In order to ease the difficulty the following identity is used:

$$\log(\mathcal{Z}) = \lim_{n\to 0} \frac{\partial}{\partial n} \log \mathcal{Z}^n. \tag{A.4.7}$$

Using this identity (A.4.7), $\mathbb{E}(\log \mathcal{Z})$ can be expressed as

$$\mathbb{E}(\log \mathcal{Z}) = \lim_{n\to 0} \frac{\partial}{\partial n} \log \mathbb{E}(\mathcal{Z}^n). \tag{A.4.8}$$

At this stage a number of assumptions are made: It is assumed that the limit and the expectation can be interchanged giving the opportunity to to deal with $\mathbb{E}(\mathcal{Z}^n)$ whenever dealing with $\mathbb{E}(\log \mathcal{Z})$ is difficult task. If n is a natural number then first we calculate $\mathcal{Z}^n$, i.e., n *replicas* of the random variable. Then in order to calculate the derivative and the limit, n has to be continuous variable. This is another critical assumption! These assumptions together with (A.4.8) are called *replica trick*, and this is where the name replica and the method, *replica method*, has evolved from. It is only recently that some of the assumptions made here have been rigorously proved [52]. Therefore, the replica method is now a standard technique to study the free energy of disordered systems. Applying the replica method the avaraged free energy given in (A.4.6) becomes

$$\begin{aligned}
\bar{\mathcal{F}} &= -\lim_{N\to\infty} \mathbb{E}(\frac{1}{N\beta} \log \mathcal{Z}) \\
&= -\lim_{N\to\infty} \frac{1}{N\beta} \lim_{n\to 0} \frac{\partial}{\partial n} \log \mathbb{E}(\mathcal{Z}^n) \qquad \text{(A.4.9)} \\
&= -\lim_{n\to 0} \frac{\partial}{\partial n} \lim_{N\to\infty} \frac{1}{N\beta} \log \mathbb{E}(\mathcal{Z}^n). \qquad \text{(A.4.10)}
\end{aligned}$$

Note that here interchanging the order of the limits $\lim_{N\to\infty}$ and $\lim_{n\to 0}$ is another assumption made in the replica calculation. Further, taking the expectation of the product of n identical replicas of $\mathcal{Z}(\beta; \mathbf{y})$, i.e.,

$$\mathbb{E}(\mathcal{Z}^n(\beta; \mathbf{y})) = \mathbb{E}\Big(\prod_{a=1}^{n} Z_a(\beta; \mathbf{y})\Big). \tag{A.4.11}$$

Replica Symmetry

The replica ansatz are generally arbitrary and often assumed to be independent random variables. Hence calculating the expectation of the right hand side of (A.4.11) is not an easy matter. Another assumption is needed in order to simplify it further. First, it is possible to assume symmetry among the replica ansatz, which is called *replica symmetry*

(RS), which is sufficient for the convex CS optimization problem at hand [36], [52] [66], [53], [67] [55] and [54]. Therefore, it is possible to assume RS ansatz in (A.4.11).

By inserting the partition function given in (A.4.2) inside (A.4.11), the expectation of the n-th power of this partition function for the replicated ansatz yields

$$\mathbb{E}(\mathcal{Z}^n(\beta;\mathbf{y})) = \int \prod_{a=1}^{n} e^{-\beta\|\mathbf{x}^a\|_p} \prod_{a=1}^{n} \delta(\mathbf{A}\mathbf{x}^a - \mathbf{y}) \prod_{a=1}^{n} d\mathbf{x}^a, \tag{A.4.12}$$

where the $\mathbf{x}^a$ are the replicated ansatz and $E(\cdot)$ is averaing over the distributions of $\mathbf{A}$ and $\mathbf{x}^0$. Focusing on the inside $\prod_{a=1}^{n} \delta(\cdot)$ expression in (A.4.12), it can be further expressed using an identity

$$\delta(x) = \lim_{\tau \to +0} (2\pi\tau)^{-1} \exp\left(-x^2/(2\tau)\right).$$

and substituting $\mathbf{y} = \mathbf{A}\mathbf{x}^0$ we get,

$$\begin{aligned} \prod_{a=1}^{n} \delta(\mathbf{A}\mathbf{x}^a - \mathbf{y}) &= \lim_{\tau \to +0} \frac{1}{(2\pi\tau)^{nM}} \exp\Big[-\frac{1}{(2\pi\tau)} \sum_{a=1}^{n} |\mathbf{A}(\mathbf{x}^a - \mathbf{x}^0)|^2 \Big] \\ &= \lim_{\tau \to +0} \frac{1}{(2\pi\tau)^{nM}} \exp\Big[-\frac{1}{(2\pi\tau)} \sum_{\mu=1}^{M} \sum_{a=1}^{n} |v_\mu^a - v_\mu^0)|^2 \Big], \end{aligned} \tag{A.4.13}$$

where, $v_\mu^a = \sum_{i=1}^{N} A_{\mu i} x_i^a$, for $\mu = 1, 2, \cdots, M$, and $a = 0, 1, 2, \cdots, n$. The v_μ^a can be assumed to be zero-mean multivariate Gaussian random variable which are characterized by the covariances $[v_\mu^a v_\nu^a]_{\mathbf{A}} = Q_{ab}\delta_{\mu\nu}$, where $[\cdots]_{\mathbf{A}}$ denotes the operation of averaging with respect to $\mathbf{A}$ and

$$Q_{ab} = \frac{1}{N} \mathbf{x}^a \cdot \mathbf{x}^b, \tag{A.4.14}$$

and $\delta_{\mu\nu}$ is unity for $\mu = \nu$ and is zero otherwise. This where the assumption of replica symmetry comes into the calculation of the minimization problem. One possibility of replica symmetry can be

$$\mathbf{Q}_{RS} = \begin{pmatrix} \rho & m & m & m & m & \cdots & m \\ m & Q & q & q & q & \cdots & q \\ m & q & Q & q & q & \cdots & q \\ \vdots & \vdots & \vdots & \ddots & & \vdots & \vdots \\ m & q & q & q & & & q \\ m & q & q & q & \cdots & q & Q \end{pmatrix}.$$

Representing this compactly we get

$$\begin{aligned} &\mathbf{x}^0 \cdot \mathbf{x}^a = Nm \\ &\mathbf{x}^a \cdot \mathbf{x}^b = Nq \text{ for } a \neq b \\ &\mathbf{x}^a \cdot \mathbf{x}^a = NQ. \end{aligned} \tag{A.4.15}$$

Now substituting these quantities in(A.4.13) and averaging over distributions of $\mathbf{A}$, we get

$$\mathbb{E}_{\mathbf{A}}\left[\prod_{a=1}^{n}\delta(\mathbf{A}\mathbf{x}^a-\mathbf{y})\right]=\frac{\alpha}{2}\ln\left(\Big(1-\frac{n(q-2m+\rho)}{Q-q}\Big)(Q-q)^n(2\pi)^2\right)\equiv\tau_n(Q,q,m). \tag{A.4.16}$$

Further averaging over the distribution of $\mathbf{x}^0$, we get

$$\frac{1}{N}\mathbb{E}_{\mathbf{x}^0}\left[\int\prod_{a=1}^{n}d\mathbf{x}^a e^{-\beta\|\mathbf{x}^a\|_p}\mathcal{I}(\{\mathbf{x}^a\}_{a=1}^n;Q,q,m)\right] \tag{A.4.17}$$

$$=\text{extr}_{\hat{Q},\hat{q},\hat{m}}\left\{\frac{n\hat{Q}Q}{2}-\frac{n(n-1)\hat{q}q}{2}-n\hat{m}m+\ln\mathcal{J}(\hat{Q},\hat{q},\hat{m};\beta)\right\} \tag{A.4.18}$$

$$\equiv\mathcal{K}(Q,q,m) \tag{A.4.19}$$

where,

$$\mathcal{I}(\{\mathbf{x}^a\}_{a=1}^n;Q,q,m)=\prod_{a=1}^{n}\delta(|\mathbf{x}^a|^2-NQ)\delta(\mathbf{x}^0\cdot\mathbf{x}^a-Nm)\prod_{a>b}\delta(\mathbf{x}^a\cdot\mathbf{x}^b-Nq) \tag{A.4.20}$$

$$\mathcal{J}(\hat{Q},\hat{\mathbf{q}},\hat{m};\beta)=\lim_{\epsilon\to+0}\int Dz\mathbb{E}_{\mathbf{x}^0}\left[\left(\int dx e^{-\frac{\hat{Q}+\hat{q}}{2}+(\sqrt{\hat{q}}z+\hat{m}x^0)x-\beta|x|^{p+\epsilon}}\right)\right]. \tag{A.4.21}$$

Now as $\beta\to\infty$,

$$\frac{1}{N}\mathbb{E}(\mathcal{Z}^n(\beta;\mathbf{y}))=\mathcal{I}(\{\mathbf{x}^a\}_{a=1}^n;Q,q,m)+\mathcal{J}(\hat{Q},\hat{\mathbf{q}},\hat{m};\beta).$$

and inserting it it in (1.3.21) and contniuing some analytical procedures with new variables $\hat{Q},\hat{\mathbf{q}},\hat{m}$, we get

$$\mathcal{C}_p=\text{extr}_{\Theta}\Bigg\{\frac{\alpha(Q-2m+\rho)}{2\chi}+m\hat{m}-\frac{\hat{Q}Q}{2}+\frac{\hat{\chi}\chi}{2}$$
$$+(1-\rho)\int Dz\phi_p(\sqrt{\hat{\chi}}z;\hat{Q})+\rho\int Dz\phi_p(\sqrt{\hat{\chi}+\hat{m}^2}z;\hat{Q})\Bigg\}, \tag{A.4.22}$$

as shown in [68] and [70], where $\text{extr}_{\Theta}\{\mathcal{G}(X)\}$ denotes the extremization of a function $\mathcal{G}(X)$ with respect to X, $\Theta=\{Q,\chi,m,\hat{Q},\hat{\chi},\hat{m}\}$, also $Dz=dz\exp(-z^2/2)/\sqrt{2\pi}$ is a Gaussian measure and

$$\phi_p(\sqrt{\hat{\chi}}z;\hat{Q})=\lim_{\epsilon\to+0}\left\{\min_x\left\{\frac{\hat{Q}}{2}x^2-hx+|x|^{p+\epsilon}\right\}\right\}. \tag{A.4.23}$$

This proves the expression given by (1.3.23) where the function

$$\Psi(Q,\chi,m,\hat{Q},\hat{\chi},\hat{m})=(1-\rho)\int Dz\phi_p(\sqrt{\hat{\chi}}z;\hat{Q})+\rho\int Dz\phi_p(\sqrt{\hat{\chi}+\hat{m}^2}z;\hat{Q}).$$

Appendix B

Appendices to Chapter 2

B.1 Important Definitions

B.1.1 Green's Function

In classical probability theory (CPT) one is concerned with the densities, moments and comulants of elements of random matrices. Where as in random matrix theory (RMT), also called free random variable calculus, one is engaged in finding the spectral densities, moments and cumilants. As the Fourier transform is the generating function for the moments in CPT, Green's function (also called Stieltjes transform) is the generating function for the spectral moments defined as

$$G(z) \equiv \frac{1}{N}\langle \mathrm{Tr}\frac{1}{z\mathbf{1}_N - \mathbf{X}}\rangle \equiv \int \frac{\rho(\lambda)}{z-\lambda}d\lambda \equiv \sum_{n=0}^{\infty}\frac{1}{z^{n+1}}M_n, \tag{B.1.1}$$

where $\mathbf{X}$ is $N \times N$ random matrix and $\mathbf{1}_N$ is of the same size unit matrix, λ are the eigenvalues, and M_n is the spectral moment. The integral is over the support set of the eigenvalues.

B.1.2 R-Transform

The generating function for the cumulants of the CPT is given by the logarithm of the Fourier transform. In similar manner to the above section we can define the generating function for spectral cumulants. It is called the R-transform (Voiculescu,1986). It is given by

$$R(z) \equiv \sum_{n=1}^{\infty} C_n z^{n-1}, \tag{B.1.2}$$

where C_n are the spectral cumulants of the random matrix $\mathbf{X}$. We can relate R-transform with Greens's function as follows:

$$G(R(z)+\frac{1}{z})=z \quad \text{or} \quad R(G(z))+\frac{1}{G(z)}=z. \tag{B.1.3}$$

The spectral density of the matrix $\mathbf{J}=\mathbf{A}^T\mathbf{A}$ converges almost surely to the Marchenko-Pastur law as $M=\alpha N\to\infty$ [96]. And the R-transform of this matrix is given by

$$R(z)=\frac{1}{1-\alpha z} \tag{B.1.4}$$

and its derivative with respect to z becomes

$$R'(z)=\frac{\alpha}{(1-\alpha z)^2}, \tag{B.1.5}$$

where $\alpha=N/M$ is system load.

B.2 Proof of Propostion 1

The average energy penality can be derived from the average free energy given in (2.4.18)

$$\begin{aligned}\bar{\mathcal{E}}&=\lim_{\beta\to\infty}\frac{1}{\beta}\bar{\mathcal{F}}=-\lim_{\beta\to\infty}\frac{1}{\beta}\lim_{N\to\infty}\frac{1}{N}\underset{w,J}{E}\{\log\ \mathcal{Z}\}\\&=-\lim_{\beta\to\infty}\frac{1}{\beta}\lim_{n\to 0}\frac{\partial}{\partial n}\underbrace{\lim_{N\to\infty}\frac{1}{\mathrm{N}}\log\underset{w,J}{E}\{(\mathcal{Z})^{\mathrm{n}}\}}_{\Xi_{\mathrm{n}}}.\end{aligned} \tag{B.2.1}$$

where Ξ_n is as given by (2.4.20),

$$\Xi_n=\lim_{N\to\infty}\tfrac{1}{N}\log\left(\alpha^{N/2}\int_{\{\mathbf{x}^a\}}e^{\frac{-\beta\gamma}{\sigma_u^2}\sum_{a=1}^{n}f(\mathbf{x}^a)}e^{\sum_{a=1}^{n}\int_0^{\lambda_a}R(-v)dv}\prod_{a=1}^{n}d\mathbf{x}^a\right). \tag{B.2.2}$$

with $\alpha=\frac{\sigma_u^2}{\sigma_u^2+n\sigma_0^2}$. Using (2.4.21) as the splitting of the space, we get

$$\Xi_n=\lim_{N\to\infty}\frac{1}{N}log\int_{\mathbb{R}^{n^2}}e^{N\mathcal{L}}e^{N\mathcal{I}\{\mathbf{Q}\}}e^{-N\mathcal{G}\{\mathbf{Q}\}}D\mathbf{Q} \tag{B.2.3}$$

where

$$D\mathbf{Q}=\prod_{a=1}^{n}dQ_{aa}\prod_{b=a+1}^{n}dQ_{ab} \tag{B.2.4}$$

is the integration measure,

$$\mathcal{G}(\mathbf{Q})=\sum_{a=1}^{n}\int_0^{\frac{\beta\gamma}{\sigma_u^2}\lambda_a(\mathbf{Q})}R(-v)dv \tag{B.2.5}$$

$$=Tr\int_0^{\frac{\beta\gamma}{\sigma_u^2}\mathbf{Q}}R(-v)dv \tag{B.2.6}$$

$$=\int_0^{\frac{\beta\gamma}{\sigma_u^2}}Tr[\mathbf{Q}R(-v\mathbf{Q})]dv \tag{B.2.7}$$

$$\mathcal{L} = \frac{1}{2}\ln\alpha - \frac{\beta\gamma}{2N}\sum_{a=0}^{n} f(\mathbf{x^a}) \quad \textbf{and} \tag{B.2.8}$$

$$e^{N\mathcal{I}\{\mathbf{Q}\}} = \int_{\{\mathbf{x^a}\}} \prod_{a=1}^{n} \delta(\phi_1(\mathbf{x}^{aa}) - NQ_{aa}) \prod_{b=a+1}^{n} \delta(\phi_1(\mathbf{x}^{ab}) - NQ_{ab}) \prod_{a=1}^{n} d\mathbf{x}^a \tag{B.2.9}$$

denotes probability weight of the subshell composed of Dirac-functions in the real line, where $\phi_1(\mathbf{x}^{cd}) = (\mathbf{x}^0 - \mathbf{x}^c)^T(\mathbf{x}^0 - \mathbf{x}^d)$. This procedure is a change of integration variables in multiple dimensions where the integration of an exponential function over the replicas has been replaced by integration over the variables $\mathbf{Q}$. To evaluate $e^{NC}e^{N\mathcal{I}\{Q\}}$ we follow [53], [55] and represent the Dirac measure using the Fourier transform as

$$\delta\Big(\phi_1(\mathbf{x}^{ba}) - NQ_{ab}\Big) = \int_{\mathcal{J}} e^{\tilde{Q}_{ab}\Big(\phi_1(\mathbf{x}^{cd}) - NQ_{ab}\Big)} \frac{d\tilde{Q}_{ab}}{2\pi}, \tag{B.2.10}$$

where $a, b = 1, \cdots, n$ and this gives

$$\begin{aligned} e^{N\mathcal{L}}e^{N\mathcal{I}\{\mathbf{Q}\}} &= \int_{\{\mathbf{x^a}\}} \int_{\mathcal{J}^{n^2}} e^{\sum\limits_{a,b} \tilde{Q}_{ab}\phi_1(\mathbf{x}^{ba}) - NQ_{ab}} e^{\frac{1}{2}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2}\sum\limits_{a=1}^{n} f(\mathbf{x^a})} \tilde{D}\tilde{\mathbf{Q}} \prod_{a=1}^{n} d\mathbf{x}^a \\ &= \int_{\mathcal{J}^{n^2}} e^{-NTr(\tilde{\mathbf{Q}}\mathbf{Q})} \left(\int_{\{\mathbf{x^a}\}} e^{\sum\limits_{a,b} \tilde{Q}_{ab}\phi_1(\mathbf{x}^{ba})} e^{\frac{1}{2}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2}\sum\limits_{a=1}^{n} f(\mathbf{x^a})} \prod_{a=1}^{n} d\mathbf{x}^a \right) \tilde{D}\tilde{\mathbf{Q}} \end{aligned} \tag{B.2.11}$$

where

$$\tilde{D}\tilde{\mathbf{Q}} = \prod_{a=1}^{n} \left(\frac{d\tilde{Q}_{aa}^{(I)}}{2\pi j} \prod_{b=a+1}^{n} \frac{d\tilde{Q}_{ab}^{(I)} d\tilde{Q}_{ab}^{(Q)}}{(2\pi j)^2} \right). \tag{B.2.12}$$

W.l.o.g. [1] assuming $f(\mathbf{x}^a) = \|\mathbf{x}^a\|_1 = \sum\limits_{i=1}^{N} |x_i^a|$, which is the sparsity enforcer as described above in LASSO estimator, and after doing some rearrangements, the inner expectation of (B.2.11) can be given by

$$\int_{\{\mathbf{x^a}\}} e^{\sum\limits_{a,b} \tilde{Q}_{ab}(\mathbf{x}^0 - \mathbf{x}^b)^T(\mathbf{x}^0 - \mathbf{x}^a)} e^{\frac{1}{2}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2}\sum\limits_{a=1}^{n} f(\mathbf{x^a})} \prod_{a=1}^{n} d\mathbf{x}^a \tag{B.2.13}$$

$$= \prod_{i=1}^{N} \int_{\{x_i^a \in \chi\}} e^{(\sum\limits_{a,b} \tilde{Q}_{ab}(x_i^0 - x_i^b)^T(x_i^0 - x_i^a)) + \frac{1}{2}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2}\sum\limits_{a=1}^{n} |x_i^a|} \prod_{a=1}^{n} dx_i^a \tag{B.2.14}$$

Now defining

$$M_i(\tilde{\mathbf{Q}}) = \int_{\{x_i^a \in \chi\}} e^{\left(\sum\limits_{a,b} \tilde{Q}_{ab}(x_i^0 - x_i^b)^T(x_i^0 - x_i^a)\right) + \frac{1}{2}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2}\sum\limits_{a=1}^{n} |x_i^a|} \prod_{a=1}^{n} dx_i^a \tag{B.2.15}$$

[1] It is possible to take the other regularizing terms. In that case, some adjustment are needed. However, the last expression, the general regulirizing term is used.

we can get

$$e^{N\mathcal{L}}e^{N\mathcal{I}\{\mathbf{Q}\}} = \int_{\mathcal{J}^{n^2}} e^{-NTr(\tilde{Q}Q)+\sum_{i=1}^{N}\log M_i(\tilde{Q})}\tilde{D}\tilde{Q}. \tag{B.2.16}$$

Following the i.i.d. assumption for the component of the sparse vector $\mathbf{x}$, and applying the strong law of large numbers as $N \to \infty$ we get

$$\begin{aligned}\log M(\tilde{\mathbf{Q}}) &= \frac{1}{N}\sum_{i=1}^{N}\log M_i(\tilde{\mathbf{Q}}) \\ &\to \int \log \int_{\{x_i^a\in\chi\}} e^{\left(\sum_{a,b}\tilde{Q}_{ab}(x_i^0-x_i^b)^T(x_i^0-x_i^a)\right)+\frac{1}{2}\ln\alpha-\frac{\beta\gamma}{\sigma_u^2}\sum_{a=1}^{n}|x_i^a|}\prod_{a=1}^{n}dx_i^a dF_{X^0}(x^0) \\ &= \int \log \int_{\{\mathbf{x}\in\chi^n\}} e^{(x^0\mathbf{1}-\tilde{\mathbf{x}})^T\tilde{\mathbf{Q}}(x^0\mathbf{1}-\tilde{\mathbf{x}})+\frac{1}{2}\ln\alpha-\frac{\beta\gamma}{\sigma_u^2}\|\tilde{\mathbf{x}}\|_1}d\tilde{\mathbf{x}}dF_{X^0}(x^0) \end{aligned}\tag{B.2.17}$$

where, $dF_{X^0}(x^0)$ is a probability measure of x^0 and $\tilde{\mathbf{x}}$ is vector of dimension n. Next we apply the saddle point integration concept on the remaining part of (B.2.3), i.e., as $N \to \infty$ the integrand will be dominated by the exponential term with maximal exponent. Hence in (B.2.3) only the subshell that corresponds to this extremal value of the correlation between the vectors $\{\mathbf{x}^a\}$ is relevant for the calculation of the integral.

$$\begin{aligned}&\int_{\mathbb{R}^{n^2}} e^{N\mathcal{L}}e^{N\mathcal{I}\{\mathbf{Q}\}}e^{-N\mathcal{G}(\mathbf{Q})}D\mathbf{Q} \\ &= \int_{\mathbb{R}^{n^2}}\left(\int_{\mathcal{J}^{n^2}} e^{-N\mathrm{Tr}(\tilde{\mathbf{Q}}\mathbf{Q})+\sum_{i=1}^{N}\log M_i(\tilde{\mathbf{Q}})}\tilde{D}\tilde{Q}\right)e^{-N\mathcal{G}\{\mathbf{Q}\}}D\mathbf{Q}\end{aligned}\tag{B.2.18}$$

Therefore, at the saddle point we have the following equations with partial derivatives being zero (see the proof in Appendix B of [55]):

$$\frac{\partial}{\partial\mathbf{Q}}\Big[\mathcal{G}(\mathbf{Q})+\mathrm{Tr}(\tilde{\mathbf{Q}}\mathbf{Q})\Big] = \mathbf{0} \qquad and \tag{B.2.19}$$

$$\frac{\partial}{\partial\tilde{\mathbf{Q}}}\Big[\log M(\tilde{\mathbf{Q}})-\mathrm{Tr}(\tilde{\mathbf{Q}}\mathbf{Q})\Big] = \mathbf{0}. \tag{B.2.20}$$

And from the former we get

$$\tilde{\mathbf{Q}} = \beta R(-\frac{\beta\gamma}{\sigma_u^2}\mathbf{Q}) \tag{B.2.21}$$

and from the later, using (B.2.17) we finally get

$$\mathbf{Q} = \int \frac{\int_{\{\tilde{\mathbf{x}}\in\chi^n\}}\psi_2(\tilde{\mathbf{x}})e^{\psi_1(\tilde{\mathbf{x}})+\frac{1}{2}\ln\alpha-\frac{\beta\gamma}{\sigma_u^2}\|\tilde{\mathbf{x}}\|_1}d\tilde{\mathbf{x}}}{\int_{\{\tilde{\mathbf{x}}\in\chi^n\}} e^{\psi_1(\tilde{\mathbf{x}})+\frac{1}{2}\ln\alpha-\frac{\beta\gamma}{\sigma_u^2}\|\tilde{\mathbf{x}}\|_1}d\tilde{\mathbf{x}}}dF_{X^0}(x^0) \tag{B.2.22}$$

with $\psi_1(\tilde{\mathbf{x}}) = (x^0\mathbf{1}-\tilde{\mathbf{x}})^T\tilde{\mathbf{Q}}(x^0\mathbf{1}-\tilde{\mathbf{x}})$, $\psi_2(\tilde{\mathbf{x}}) = (x^0\mathbf{1}-\tilde{\mathbf{x}})(x^0\mathbf{1}-\tilde{\mathbf{x}})^T$, and $\tilde{\mathbf{x}}$ is vector of dimension n .

B.3 Proof of Propostion 2

Taking the same line of thought as we do for $\mathbf{Q}$, we can assume a natural replicated variables for the symmetric correlation matrix $\tilde{\mathbf{Q}}$ and the 1RSB as follows:

1. replica symmetry ansatz :

$$\tilde{\mathbf{Q}} = \frac{\beta^2 f_0^2}{2}\mathbf{1}_{n\times n} - \beta e_0 \mathbf{I}_{n\times n} \tag{B.3.1}$$

2. one replica symmetry breaking ansatz :

$$\tilde{\mathbf{Q}} = \beta^2 f_1^2 \mathbf{1}_{n\times n} + \beta^2 g_1^2 \mathbf{I}_{\frac{n\beta}{\mu_1}\times\frac{n\beta}{\mu_1}} \otimes \mathbf{1}_{\frac{\mu_1}{\beta}\times\frac{\mu_1}{\beta}} - \beta e_1 \mathbf{I}_{n\times n} \tag{B.3.2}$$

The variables q_0, b_0, q_1, p_1,b_1, f_0,e_0,f_1,g_1,e_1, and μ_1 are called the macroscopic variables and they are all functions of n. They all can be calculated from the saddle point equations. First let us prove proposition 2 using the ansatz in (2.4.24) and (B.3.1) using equations (B.2.1), (B.3.3) and (B.2.18) and applying the saddle point integration rule. What matters most becomes the argument of the exponential in (B.2.18): $\mathrm{Tr}(\tilde{\mathbf{Q}}\mathbf{Q})$, $\mathcal{G}(\mathbf{Q})$, $\log M(\mathbf{Q})$. The limiting energy expression is dependent on the macroscopic parameters mentioned above, therefore we also provide the expressions for each macroscopic variable. Hence using (2.4.24) and (B.3.1) we get

$$\mathrm{Tr}(\tilde{\mathbf{Q}}\mathbf{Q}) = n(q_0 + \frac{b_0}{\beta})(\frac{\beta^2 f_0^2}{2} - \beta e_0) + \frac{n(n-1)}{2} q_0 \beta^2 f_0^2 \tag{B.3.3}$$

and using (B.2.15) and (B.3.1) again we get

$$M_i(\tilde{\mathbf{Q}}) = \int_{\{x_i^a \in \chi\}} e^{\left(\sum\limits_{a,b} \tilde{Q}_{ab}(x_i^0 - x_i^b)^T(x_i^0 - x_i^a)\right) + \frac{1}{2}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2}\sum\limits_{a=1}^{n}|x_i^a|} \prod_{a=1}^{n} dx_i^a \tag{B.3.4}$$

$$= \int_{\{x_i^a \in \chi\}} e^{\frac{\beta^2 f_0^2}{2}\left(\sum\limits_{a=1}^{n}(x_i^0 - x_i^a)\right)^2 - e_0\beta\sum\limits_{a=1}^{n}(x_i^0 - x_i^a)^2 + \frac{1}{2}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2}\sum\limits_{a=1}^{n}|x_i^a|} \prod_{a=1}^{n} dx_i^a \tag{B.3.5}$$

$$= \int_{\{x_i^a \in \chi\}} \int_{\mathbb{C}} e^{\beta\sum\limits_{a=1}^{n} f_0 \Re\{(x^0 - x_i^a)z^*\} - e_0(x_i^0 - x_i^a)^2 + \frac{1}{2n\beta}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2}\sum\limits_{a=1}^{n}|x_i^a|} Dz \prod_{a=1}^{n} dx_i^a \tag{B.3.6}$$

$$= \int_{\{x_i^a \in \chi\}} \int_{\mathbb{C}} e^{\beta\sum\limits_{a=1}^{n} f_0 \Re\{(x^0 - x_i^a)z^*\} - e_0(x_i^0 - x_i^a)^2 + \frac{1}{\beta}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2}\sum\limits_{a=1}^{n}|x_i^a|} Dz \prod_{a=1}^{n} dx_i^a \tag{B.3.7}$$

$$= \int_{\mathbb{C}} \left(\int_{\mathbb{R}} e^{\beta f_0 \Re\{(x^0 - x_i)z^*\} + e_0(x_i^0 - x_i)^2 + \frac{1}{2}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2}|x_i|} dx_i \right)^n Dz. \tag{B.3.8}$$

From (B.3.5) to (B.3.7) we apply completing the square on the exponential of the argument and the Hubbard-Stratonovich transform,

$$e^{|x|^2} = \int_{\mathbb{C}} e^{2\Re\{xz^*\}} Dz, \tag{B.3.9}$$

where Dz is Gaussian measure defined as before, to linearize the exponential argument. And we finally transformed the expression to a double integral problem. Moreover, for $N \to \infty$, by the law of large numbers we have

$$\begin{aligned}\log M(e,f) &= \frac{1}{N}\sum_{i=1}^{N} \log M_i(e,f) \\ &= \int_{\mathbb{R}} \log \int_{\mathbb{C}} \left(\int_{\chi} e^{\beta f_0 \Re\{(x^0-x)z^*\} + e_0(x_i^0-x)^2 + \frac{1}{2}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2}|x|} dx \right)^n Dz dF_{X^0}(x^0),\end{aligned} \tag{B.3.10}$$

where, $dF_{X^0}(x^0)$ is a probability measure of x^0 .

To evaluate $\mathcal{G}(\mathbf{Q})$ we should first find the eigenvalues of the matrix L(n). Under the RS ansatz the matrix L(n) has three types of eigenvalues: $\lambda_1 = -(\sigma_u^2 + n\sigma_0^2)^{-1}(b_0 + n\beta q_0)$, $\lambda_2 = -(\sigma_u^2)^{-1}b_0$ and $\lambda_3 = 0$, and the numbers of degeneracy for each are 1, n-1, and N-n, respectively.Thus we get

$$\mathcal{G}(\mathbf{Q}) = \int_0^{\frac{(b_0+n\beta q_0)}{\sigma_u^2+n\sigma_0^2}} R(-v)dv + (n-1)\int_0^{\frac{b_0}{\sigma_u^2}} R(-v)dv \tag{B.3.11}$$

The integral in (B.2.18) is dominated by the maximum argument of the exponential function. Therefore, the derivative of

$$\mathcal{G}(\mathbf{Q}) + \text{Tr}(\tilde{\mathbf{Q}}\mathbf{Q}) \tag{B.3.12}$$

with respect to q_0 and b_0 must vanish as $N \to \infty$. Plugging (B.3.3) and (B.3.11) into (B.3.12) and taking the partial derivatives we get

$$\frac{\beta n}{\sigma_u^2 + n\sigma_0^2} R\Big(\frac{-(b_0+n\beta q_0)}{(\sigma_u^2+n\sigma_0^2)}\Big) + \frac{n(n-1)}{2}\beta^2 f_0^2 + n\beta(\frac{\beta f_0^2}{2} - e_0) = 0 \tag{B.3.13}$$

$$\frac{1}{\sigma_u^2 + n\sigma_0^2} R\Big(\frac{-(b_0+n\beta q_0)}{(\sigma_u^2+n\sigma_0^2)}\Big) + \frac{1}{\sigma_u^2}(n-1)R\Big(\frac{-b_0}{\sigma_u^2}\Big) + n(\frac{\beta f_0^2}{2} - e_0) = 0, \tag{B.3.14}$$

respectively. After algebraic simplification and solving for e_0 and f_0 we get

$$e_0 = \frac{1}{\sigma_u^2} R\Big(\frac{-b_0}{\sigma_u^2}\Big), \tag{B.3.15}$$

$$f_0 = \sqrt{\frac{2}{n\beta}\left[\frac{1}{\sigma_u^2}R\Big(\frac{-b_0}{\sigma_u^2}\Big) - \frac{1}{\sigma_u^2+n\sigma_0^2}R\Big(\frac{-(b_0+n\beta q_0)}{(\sigma_u^2+n\sigma_0^2)}\Big)\right]}. \tag{B.3.16}$$

and with the limit for $n \to 0$

$$f_0 \overset{n\to 0}{\longrightarrow} \sqrt{\frac{2}{\beta}\left[\frac{\sigma_0^2}{\sigma_u^4}R\Big(\frac{-b_0}{\sigma_u^2}\Big) + \frac{\beta q_0\sigma_u^2 + b_0\sigma_0^2}{\sigma_u^6}R'\Big(\frac{-b_0}{\sigma_u^2}\Big)\right]}. \tag{B.3.17}$$

Using (B.3.10), substituting (B.3.3) into (B.2.20), and doing the partial derivative of

$$\begin{aligned}
&\log M(e_0, f_0) - \text{Tr}(\tilde{\mathbf{Q}}\mathbf{Q}) \\
&= \int_{\mathbb{R}} \log \int_{\mathbb{C}} \left(\int_{\chi} e^{\beta f_0 \Re\{(x^0-x)z^*\} + e_0(x_i^0-x)^2 + \frac{1}{2}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2}|x|} dx \right)^n DzdF_{X^0}(x^0) \\
&- \Big(n(q_0 + \frac{b_0}{\beta})(\frac{\beta^2 f_0^2}{2} - \beta e_0) + \frac{n(n-1)}{2} q_0 \beta^2 f_0^2 \Big),
\end{aligned} \tag{B.3.18}$$

with respect to e_0 and f_0 and equating to zero we get,

$$q_0 = -\frac{b_0}{\beta} + \int_{\mathbb{R}} \int_{\mathbb{C}} \frac{\int_{\{x\in\chi\}} (x^0 - x)^2 \psi_3(x) dx}{\int_{\{x\in\chi\}} \psi_3(x) dx} DzdF_{X^0}(x^0) \tag{B.3.19}$$

$$b_0 = -\beta n q_0 + \frac{1}{f_0} \int_{\mathbb{R}} \int_{\mathbb{C}} \frac{\int_{\{x\in\chi\}} \Re\{(x^0 - x_i^a)z^*\} \psi_3(x) dx}{\int_{\{x\in\chi\}} \psi_3(x) dx} DzdF_{X^0}(x^0) \tag{B.3.20}$$

where

$$\psi_3(x) = e^{\beta f_0 \Re\{(x^0 - x_i^a)z^*\} + e_0\beta(x^0-x)^2 + \frac{1}{2}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2}|x|}. \tag{B.3.21}$$

So collecting the macroscopic variables in (B.3.15), (B.3.16), (B.3.19) and (B.3.20) and sending $n \to 0$ we have

$$e_0 = \frac{1}{\sigma_u^2} R\Big(\frac{b_0}{\sigma_u^2}\Big) \tag{B.3.22}$$

$$f_0 \overset{n\to 0}{\longrightarrow} \sqrt{\frac{2}{\beta}\left[\frac{\sigma_0^2}{\sigma_u^4} R\Big(\frac{-b_0}{\sigma_u^2}\Big) + \frac{\beta q_0 \sigma_u^2 + b_0 \sigma_0^2}{\sigma_u^6} R'\Big(\frac{-b_0}{\sigma_u^2}\Big)\right]} \tag{B.3.23}$$

$$q_0 = -\frac{b_0}{\beta} + \int_{\mathbb{R}} \int_{\mathbb{C}} \frac{\int_{\{x\in\chi\}} (x^0 - x)^2 \psi_4(x) dx}{\int_{\{x\in\chi\}} \psi_4(x) dx} DzdF_{X^0}(x^0), \tag{B.3.24}$$

$$b_0 \overset{n\to 0}{\longrightarrow} \frac{1}{f_0} \int_{\mathbb{R}} \int_{\mathbb{C}} \frac{\int_{\{x\in\chi\}} \Re\{(x^0 - x_i^a)z^*\} \psi_4(x) dx}{\int_{\{x\in\chi\}} \psi_4(x) dx} DzdF_{X^0}(x^0). \tag{B.3.25}$$

where

$$\psi_4(x) = e^{\beta f_0 \Re\{(x^0 - x_i^a)z^*\} + e_0\beta(x^0-x)^2 - \frac{\beta\gamma}{\sigma_u^2}|x|}. \tag{B.3.26}$$

And the fixed point equations (B.3.23), (B.3.24) and (B.3.25) further can be simplified via the saddle point integration rule in the limit $\beta \to \infty$ as

$$f_0 = \sqrt{2\frac{q_0}{\sigma_u^4} R'\Big(\frac{-b_0}{\sigma_u^2}\Big)} \tag{B.3.27}$$

$$q_0 = \int_{\mathbb{R}}\int_{\mathbb{C}} \Big| x^0 - \arg\min_{x\in\chi} \Big| -zf_0 + 2e_0(x^0 - x) - \frac{\gamma}{\sigma_u^2}\Big|\Big|^2 DzdF_{X^0}(x^0), \tag{B.3.28}$$

$$b_0 = \frac{1}{f_0}\int_{\mathbb{R}}\int_{\mathbb{C}} \Re\Bigg\{ x^0 - \arg\min_{x\in\chi} \Big| -zf_0 + 2e_0(x^0 - x) - \frac{\gamma}{\sigma_u^2}\Big| z^* \Bigg\} DzdF_{X^0}(x^0). \tag{B.3.29}$$

Putting together the results above we have

$$\begin{aligned}
\Xi_n &= \mathcal{I}\{Q\} + \mathcal{L} - \mathcal{G}(\mathbf{Q}) \\
&= -\mathcal{G}(\mathbf{Q}) + \log \mathbf{M}(\tilde{\mathbf{Q}}) - \mathrm{Tr}(\tilde{\mathbf{Q}}\mathbf{Q}) \qquad \text{(B.3.30)} \\
&= -\int_0^{\frac{(b_0+n\beta q_0)}{\sigma_u^2+n\sigma_0^2}} R(-v)dv - (n-1)\int_0^{\frac{b_0}{\sigma_u^2}} R(-v)dv \\
&+ \log M(e_0, f_0) - \Big(n(q_0 + \frac{b_0}{\beta})(\frac{\beta^2 f_0^2}{2} - \beta e_0) + \frac{n(n-1)}{2} q_0 \beta^2 f_0^2 \Big), \qquad \text{(B.3.31)}
\end{aligned}$$

and the average free energy becomes

$$\beta\bar{\mathcal{F}} = -\lim_{n\to 0}\frac{\partial}{\partial n} \underbrace{\lim_{N\to\infty}\frac{1}{\mathrm{N}} \log \underset{n,J}{E}\{(\mathcal{Z})^{\mathrm{n}}\}}_{\Xi_{\mathrm{n}}} \tag{B.3.32}$$

$$\begin{aligned}
&= \lim_{n\to 0}\frac{\partial}{\partial n}\Bigg\{ \int_0^{\frac{(b_0+n\beta q_0)}{\sigma_u^2+n\sigma_0^2}} R(-v)dv + (n-1)\int_0^{\frac{b_0}{\sigma_u^2}} R(-v)dv \\
&- \log M(e_0, f_0) + (n(q_0 + \frac{b_0}{\beta})(\frac{\beta^2 f_0^2}{2} - \beta e_0) + \frac{n(n-1)}{2} q_0\beta^2 f_0^2) \Bigg\} \qquad \text{(B.3.33)} \\
&= \lim_{n\to 0}\Bigg\{ \Big[\frac{-(b_0+n\beta q_0)}{\sigma_u^2+n\sigma_0^2}\Big] R\Big(\frac{-(b_0+n\beta q_0)}{\sigma_u^2+n\sigma_0^2}\Big) \\
&+ \frac{-(b_0+n\beta q_0)}{(\sigma_u^2+n\sigma_0^2)}\Big[-\frac{\Big(\beta q_0(\sigma_u^2+n\sigma_0^2) - (b_0+n\beta q_0)\sigma_0^2\Big)}{(\sigma_u^2+n\sigma_0^2)^2}\Big] R'\Big(\frac{-(b_0+n\beta q_0)}{(\sigma_u^2+n\sigma_0^2)}\Big) \\
&+ \int_0^{\frac{b_0}{\sigma_u^2}} R(-v)dv - \int_{\mathbb{R}}\int_{\mathbb{C}} \frac{\psi_4(x)^n \ln\psi_4(x)}{\psi_4(x)^n} DzdF_{X^0}(x^0) \Bigg\} \qquad \text{(B.3.34)} \\
&= \frac{-b_0}{\sigma_u^2} R\Big(\frac{-b_0}{\sigma_u^2}\Big) + \frac{b_0(\beta q_0\sigma_u^2 - b_0\sigma_0^2)}{\sigma_u^6} R'\Big(\frac{-b_0}{\sigma_u^2}\Big) \\
&+ \int_0^{\frac{b_0}{\sigma_u^2}} R(-v)dv - \int_{\mathbb{R}}\int_{\mathbb{C}} \ln\psi_4(x) DzdF_{X^0}(x^0). \qquad \text{(B.3.35)}
\end{aligned}$$

Coming back to the main goal, the solution for the main unconstrained optimization problem (4.3.3) is given by the extremum of (2.4.9), it is calculated through the free energy by sending $\beta \to \infty$ as follows

$$\bar{\mathcal{E}}_{\text{rs}}^{\text{lasso}} = -\lim_{\beta\to\infty}\frac{1}{\beta}\lim_{n\to 0}\frac{\partial}{\partial n}\Xi_n \tag{B.3.36}$$

$$= \lim_{\beta\to\infty}\frac{1}{\beta}\Bigg\{\frac{-b_0}{\sigma_u^2}R\Big(\frac{-b_0}{\sigma_u^2}\Big) + \frac{b_0(\beta q_0\sigma_u^2 - b_0\sigma_0^2)}{\sigma_u^6}R'\Big(\frac{-b_0}{\sigma_u^2}\Big) + \int_0^{\frac{b_0}{\sigma_u^2}} R(-w)dw$$
$$- \int_{\mathbb{R}}\int_{\mathbb{C}} \ln\psi_4(x)DzdF_{X^0}(x^0)\Bigg\} \tag{B.3.37}$$

$$= \lim_{\beta\to\infty} R\Big(\frac{-b_0}{\sigma_u^2}\Big)\Big(\frac{q_0}{\sigma_u^2} + \frac{b_0}{\beta\sigma_u^2}\Big) + \frac{b_0 q_0}{\sigma_u^4}R'\Big(\frac{-b_0}{\sigma_u^2}\Big)$$
$$- \lim_{\beta\to\infty}\frac{1}{\beta}\Bigg\{\int_{\mathbb{R}}\int_{\mathbb{C}} \ln\psi_4(x)DzdF_{X^0}(x^0)\Bigg\} \tag{B.3.38}$$

$$= \frac{q_0}{\sigma_u^2}R\Big(\frac{-b_0}{\sigma_u^2}\Big) - \frac{b_0 q_0}{\sigma_u^4}R'\Big(\frac{-b_0}{\sigma_u^2}\Big). \tag{B.3.39}$$

This proves propostion 2.

B.4 Proof of Propostions 3

Turning to the LASSO estimator with RSB ansatz we first use (2.4.25) and (B.3.2) to get

$$\text{Tr}(\tilde{\mathbf{Q}}\mathbf{Q}) = n(q_1 + p_1 + \frac{b_1}{\beta})(\beta^2 f_1^2 + \beta^2 g_1^2 - \beta e_1)$$
$$+ n(\frac{\mu_1}{\beta} - 1)(q_1 + p_1)(\beta^2 g_1^2 + \beta^2 f_1^2) + n(n - \frac{\mu_1}{\beta})q_1\beta^2 f_1^2. \tag{B.4.1}$$

To evaluate $\mathcal{G}(q_1, p_1, f_1, \mu_1)$ we should first find the eigenvalues of the matrix $\mathbf{L}(n)$. Under the RSB ansatz the matrix $\mathbf{L}(n)$ has four types of eigenvalues: $\lambda_1 = -(\sigma_u^2 + n\sigma_0^2)^{-1}(b_1 + \mu p_1 + \beta n q_1)$, $\lambda_2 = -(\sigma_u^2)^{-1}(b_1 + \mu p_1)$, $\lambda_3 = -(\sigma_u^2)^{-1}b_1$ and $\lambda_4 = 0$, and the numbers of degeneracy for each are 1, $n\beta/\mu - 1$, $n - n\beta/\mu$, and $N - n$, respectively. Hence

$$\mathcal{G}(q_1, p_1, f_1, \mu_1) = \int_0^{\frac{b_1+\mu_1 p_1+\beta n q_1}{\sigma_u^2+n\sigma_0^2}} R(-v)dv + (\frac{n\beta}{\mu_1} - 1)\int_0^{\frac{b_1+\mu_1 p_1}{\sigma_u^2}} R(-v)dv$$
$$+ (n - \frac{n\beta}{\mu_1})\int_0^{\frac{b_1}{\sigma_u^2}} R(-v)dv \tag{B.4.2}$$

Further with entries of $\tilde{\mathbf{Q}}$ being RSB ansatz (B.2.17) will have more involved terms than

the RS ansatzs. i.e. ,

$$\begin{aligned}
&\log M(q_1, p_1, f_1, \mu_1) \\
&= \int_{\mathbb{R}} \log \int_{\{\mathbf{x}^a \in \chi^n\}} e^{(x^0\mathbf{1}-\mathbf{x}^a)^T \tilde{\mathbf{Q}}(x^0\mathbf{1}-\mathbf{x}^a)+\frac{1}{2}\ln\alpha-\frac{\beta\gamma}{\sigma_u^2}\mathbf{x}^a} \prod_{a=1}^{n} d\mathbf{x}^a dF_{X^0}(x^0) \\
&= \int_{\mathbb{R}} \log \int_{\{\mathbf{x}^a \in \chi^n\}} e^{\psi_5(x)+\frac{1}{2}\ln\alpha-\frac{\beta\gamma}{\sigma_u^2}\sum_{a=1}^{n}|\mathbf{x}^a|} \prod_{a=1}^{n} d\mathbf{x}^a dF_{X^0}(x^0),
\end{aligned} \tag{B.4.3}$$

where $dF_{X^0}(x^0)$ is a probability measure of x^0 and

$$\psi_5(x) = e^{\beta^2 f_1^2 \left|\sum_{a=1}^{n}(x^0-x^a)\right|^2 + \beta^2 g_1^2 \sum_{l=0}^{\frac{n\beta}{\mu}-1}\left|\sum_{a=1}^{\frac{\mu}{\beta}}(x^0-x^{a+\frac{l\mu_1}{\beta}})\right|^2 - \beta e_1 \sum_{a=1}^{n}(x^0-x^a)^2}. \tag{B.4.4}$$

Using the Hubbard-Stratonovich transform (B.3.9) we can express (B.4.3) as in (c.f. [[96], (66)-(70)]) as follows

$$\begin{aligned}
&\log M(q_1, p_1, f_1, \mu_1) \\
&= \int_{\mathbb{R}} \log \int_{\{\mathbf{x}\in\chi^n\}} \int_{\mathbb{C}} e^{\psi_6(x)+\beta^2 g_1^2 \sum_{l=0}^{\frac{n\beta}{\mu}-1}\left|\sum_{a=1}^{\frac{\mu}{\beta}}(x^0-x^{a+\frac{l\mu_1}{\beta}})\right|^2} Dz \prod_{a=1}^{n} dx_i^a dF_{X^0}(x^0) \\
&= \int_{\mathbb{R}} \log \int_{\mathbb{C}} \left[\int_{\mathbb{C}} \left(\int_{\{x\in\chi\}} \mathcal{K}(x, y, z) dx\right)^{\frac{\mu_1}{\beta}} Dy\right]^{\frac{n\beta}{\mu_1}} Dz dF_{X^0}(x^0)
\end{aligned} \tag{B.4.5}$$

where

$$\psi_6(x) = \sum_{a=1}^{n} [2\beta f_1 \Re\{(x^0-x^a)z^*\} - \beta e_1 |(x^0-x^a)|^2 + \frac{1}{2}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2}|x^a|] \tag{B.4.6}$$

$$\mathcal{K}(x, y, z) = e^{2\beta\Re\{(x^0-x)(f_1 z^* + g_1 y^*)\} - \beta e_1 |(x^0-x)|^2 + \frac{1}{2}\ln\alpha - \frac{\beta\gamma}{\sigma_u^2}|x|}. \tag{B.4.7}$$

Applying the partial dervative, as in (B.2.19), on

$$\mathcal{G}(q_1, p_1, f_1, \mu_1) + \text{Tr}(\tilde{\mathbf{Q}}\mathbf{Q}) \tag{B.4.8}$$

with respect to the macroscopic variables q_1, p_1, and b_1 as $N \to \infty$, we get zero by definition of the saddle point approximation. And plugging (B.4.1) and (B.4.2) in (**??**) and calculating the partial derivatives and setting them to zero and after some algebraic manipulation we get the following set of equations

$$0 = n^2\beta^2 f_1^2 + n\beta\mu_1 g_1^2 - n\beta e_1 + \frac{n\beta}{\sigma_u^2 + n\sigma_0^2} R(\frac{-b_1 - \mu_1 p_1 - \beta n q_1}{\sigma_u^2 + n\sigma_0^2}) \tag{B.4.9}$$

$$0 = n\beta\mu_1 b_1^2 + n\beta\mu_1 g_1^2 - n\beta e_1 + \frac{(n\beta - \mu_1)}{\sigma_u^2} R(\frac{-b_1 - \mu_1 p_1}{\sigma_u^2}) + \frac{\mu_1}{\sigma_u^2 + n\sigma_0^2} R(\frac{-b_1 - \mu_1 p_1 - \beta n q_1}{\sigma_u^2 + n\sigma_0^2}) \tag{B.4.10}$$

$$0 = n\beta f_1^2 + n\beta g_1^2 - n e_1 + \frac{(n - \frac{n\beta}{\mu_1})}{\sigma_u^2} R(\frac{-b_1}{\sigma_u^2}) + \frac{(\frac{n\beta}{\mu_1} - 1)}{\sigma_u^2} R(\frac{-b_1 - \mu_1 p_1}{\sigma_u^2}) \tag{B.4.11}$$

$$+ \frac{1}{\sigma_u^2 + n\sigma_0^2} R(\frac{-b_1 - \mu_1 p_1 - \beta n q_1}{\sigma_u^2 + n\sigma_0^2}). \tag{B.4.12}$$

Solving for e_1, g_1, f_1 we get

$$e_1 = \frac{1}{\sigma_u^2} R(\frac{-b_1}{\sigma_u^2}), \tag{B.4.13}$$

$$g_1 = \sqrt{\frac{1}{\mu_1}\left[\frac{1}{\sigma_u^2} R(\frac{-b_1}{\sigma_u^2}) - \frac{1}{\sigma_u^2} R(\frac{-b_1 - \mu_1 p_1}{\sigma_u^2})\right]}, \tag{B.4.14}$$

$$f_1 = \sqrt{\frac{1}{n\beta}\left[\frac{1}{\sigma_u^2} R(\frac{-b_1 - \mu_1 p_1}{\sigma_u^2}) - \frac{1}{\sigma_u^2 + n\sigma_0^2} R(\frac{-b_1 - \mu_1 p_1 - n\beta q_1}{\sigma_u^2 + n\sigma_0^2})\right]}, \tag{B.4.15}$$

and further with the limits $n \to 0$

$$f_1 \overset{n\to 0}{\longrightarrow} \sqrt{\frac{1}{\beta}\left[\frac{\sigma_0^2}{\sigma_u^4} R(\frac{-b_1 - \mu_1 p_1}{\sigma_u^2}) + \frac{(\sigma_u^2\beta q_1 + \sigma_0^2(b_1 + \mu_1 p_1))}{\sigma_u^6} R'(\frac{-b_1 - \mu_1 p_1}{\sigma_u^2})\right]}. \tag{B.4.16}$$

and as $\beta \to \infty$ we can simplify it further as

$$f_1 \overset{n\to 0}{\longrightarrow} \sqrt{\frac{q_1}{\sigma_u^4} R'(\frac{-b_1 - \mu_1 p_1}{\sigma_u^2})}. \tag{B.4.17}$$

Similarly following (B.2.20) the partial derivatives of

$$\log M(q_1, p_1, f_1, \mu_1) - \text{Tr}(\tilde{\mathbf{Q}}\mathbf{Q})$$

with respect to f_1, g_1, and e_1, must also vanish as $N \to \infty$. This produces the following set of equations while taking $n \to 0$.

$$b_1 + p_1\mu_1 = \frac{1}{f_1} \int_{\mathbb{R}} \int_{\mathbb{C}^2} \psi_7(x) \int_{x \in \chi} \Re\{xz^*\} \mathcal{K}(x, y, z)\, dx Dy Dz dF_{X^0}(x^0) \tag{B.4.18}$$

$$b_1 + (q_1 + p_1)\mu_1 = \frac{1}{g_1}\int_{\mathbb{R}}\int_{\mathbb{C}^2}\psi_7(x)\int_{x\in\chi}\Re\{xy^*\}\mathcal{K}(\,x,\,y,\,z)dxDyDzdF_{X^0}(x^0) \tag{B.4.19}$$

$$q_1 + p_1 = -\frac{b_1}{\beta} + \frac{1}{g_1}\int_{\mathbb{R}}\int_{\mathbb{C}^2}\psi_7(x)\int_{x\in\chi}|x|^2\mathcal{K}(\,x,\,y,\,z)dxDyDzdF_{X^0}(x^0), \tag{B.4.20}$$

where

$$\psi_7(x) = \frac{\Big(\int_{x\in\chi}\mathcal{K}(\,x,\,y,\,z)\Big)^{\frac{\mu_1}{\beta}-1}}{\int_{\mathbb{C}}\Big(\int_{x\in\chi}\mathcal{K}(\,x,\,y,\,z)\Big)^{\frac{\mu_1}{\beta}}D\tilde{y}}. \tag{B.4.21}$$

In addition when we take the partial derivative of

$$\mathcal{G}(q_1, p_1, f_1, \mu_1) + \mathrm{Tr}(\tilde{\mathbf{Q}}\mathbf{Q}) - \log M(q_1, p_1, f_1, \mu_1) \tag{B.4.22}$$

with respect of μ_1 is vanishes and yields at the limit as $n \to 0$

$$\begin{aligned}
0 = &\frac{1}{\mu_1^2}\int_{\frac{b_1}{\sigma_u^2}}^{\frac{b_1+\mu_1 p_1}{\sigma_u^2}} R(-v)dv + \frac{p_1}{\mu_1^2}R(-\frac{b_1+\mu_1 p_1}{\sigma_u^2}) + (q_1+p_1)g_1^2 + p_1 f_1^2 \\
&+ \int_{\mathbb{R}}\int_{\mathbb{C}}\Bigg[\frac{1}{\mu_1^2}\log\Big(\int_{\mathbb{C}}\Big(\int_{\{\mathbf{x}\in\chi\}}\mathcal{K}(\,x,\,y,\,z)dx\Big)^{\frac{\mu_1}{\beta}}Dy\Big) \\
&- \int_{\mathbb{C}}\frac{1}{\beta\mu_1^2}\psi_7(x)\cdot\log\Big(\int_{x\in\chi}\mathcal{K}(\,x,\,y,\,z)dx\Big)Dy\Bigg]DzdF_{X^0}(x^0)
\end{aligned} \tag{B.4.23}$$

So as $\beta \to \infty$ these fixed point equations can be simplified as follows:

$$b_1 + p_1\mu_1 = \frac{1}{f_1}\int_{\mathbb{R}}\int_{\mathbb{C}^2}\Re\Big\{\Big(x^0 - \Psi_2\Big)z^*\Big\}\tilde{\Delta}(y,z)DyDzdF_{X^0}(x^0) \tag{B.4.24}$$

$$b_1 + (q_1 + p_1)\mu_1 = \frac{1}{g_1}\int_{\mathbb{R}}\int_{\mathbb{C}^2}\Re\Big\{\Big(x^0 - \Psi_2\Big)y^*\Big\}\tilde{\Delta}(y,z)DyDzdF_{X^0}(x^0) \tag{B.4.25}$$

$$q_1 + p_1 = \frac{1}{g_1}\int_{\mathbb{R}}\int_{\mathbb{C}^2}|\Psi_2|^2\tilde{\Delta}(y,z)DyDzdF_{X^0}(x^0) \tag{B.4.26}$$

where

$$\Psi_2 = \arg\min_{x\in\chi}\Big|2\Re\{(x^0-x)(f_1z^* + g_1y^*)\} - e_1|(x^0-x)|^2 - \frac{\gamma}{\sigma_u^2}|x|\Big|,$$

and

$$\Delta(y,z) \equiv e^{-\mu_1\min_{x\in\chi}-2\Re\{(x^0-x)(f_1z^*+g_1y^*)\}+e_1(x^0-x)^2-\frac{\gamma}{\sigma_u^2}|x|}, \quad (y,z)\in\Re^2 \tag{B.4.27}$$

and its normalized version

$$\tilde{\Delta}(y,z) = \frac{\Delta(y,z)}{\int_{\mathbb{C}} \Delta(\tilde{y},z)d\tilde{y}}. \tag{B.4.28}$$

Putting together the results again as in (B.3.30) and repeating the steps (B.34) to (B.38) for the RSB case

$$\begin{aligned}
\bar{\mathcal{E}}^{\text{lasso}}_{\text{1rsb}} &= -\lim_{\beta\to\infty}\frac{1}{\beta}\lim_{n\to 0}\frac{\partial}{\partial n}\Xi_n && \text{(B.4.29)}\\
&= -\lim_{\beta\to\infty}\frac{1}{\beta}\lim_{n\to 0}\frac{\partial}{\partial n}\{-\mathcal{G}(\mathbf{Q}) - \text{Tr}(\tilde{\mathbf{Q}}\mathbf{Q}) + \log M(\tilde{\mathbf{Q}})\} && \text{(B.4.30)}\\
&= \lim_{\beta\to\infty}\frac{1}{\beta}\lim_{n\to 0}\frac{\partial}{\partial n}\Bigg\{\int_0^{\frac{b_1+\mu_1 p_1+\beta n q_1}{\sigma_u^2+n\sigma_0^2}} R(-v)dv + (\frac{n\beta}{\mu_1}-1)\int_0^{\frac{b_1+\mu_1 p_1}{\sigma_u^2}} R(-v)dv\\
&+ (n-\frac{n\beta}{\mu_1})\int_0^{\frac{b_1}{\sigma_u^2}} R(-v)dv + \Big[n(q_1+p_1+\frac{b_1}{\beta})(\beta^2 f_1^2+\beta^2 g_1^2-\beta e_1)\\
&+ n(\frac{\mu_1}{\beta}-1)(q_1+p_1)(\beta^2 g_1^2+\beta^2 f_1^2) + n(n-\frac{\mu_1}{\beta})q_1\beta^2 f_1^2\Big]\\
&- \log M(q_1,p_1,f_1,\mu_1)\Bigg\}\\
&= \lim_{\beta\to\infty}\frac{1}{\beta}\Bigg\{(\frac{b_1+\mu_1 p_1}{\sigma_u^2})R(\frac{-b_1-\mu_1 p_1}{\sigma_u^2})\\
&+ (\frac{b_1+\mu_1 p_1}{\sigma_u^2})\frac{(-\beta q_1\sigma_u^2-(b_1+\mu_1 p_1)\sigma_0^2)}{\sigma_u^4}R'(\frac{-b_1-\mu_1 p_1}{\sigma_u^2})\\
&+ \frac{\beta}{\mu_1}\int_0^{\frac{b_1+\mu_1 p_1}{\sigma_u^2}} R(-v)dv + (1-\frac{\beta}{\mu_1})\int_0^{\frac{b_1}{\sigma_u^2}} R(-v)dv\\
&+ \Big[b_1(\beta f_1^2+\beta g_1^2-e_1)+\mu_1(q_1+p_1)(\beta f_1^2+\beta g_1^2-\frac{\beta}{\mu_1}e_1)-\mu_1 q_1\beta f_1^2\Big]\\
&- \frac{\beta}{\mu_1}\int \log\int_{\mathbb{C}}\int_{\mathbb{C}}\left(\int_{\{\mathbf{x}\in\chi\}}\mathcal{K}(x,y,z)dx\right)^{\frac{\mu_1}{\beta}} DyDzdF_{X^0}(x^0)\Bigg\}\\
&= \tfrac{-q_1}{\sigma_u^2}(\tfrac{b_1+\mu_1 p_1}{\sigma_u^2})R(\tfrac{-b_1-\mu_1 p_1}{\sigma_u^2}) + \tfrac{1}{\mu_1}\textstyle\int_0^{\frac{b_1+\mu_1 p_1}{\sigma_u^2}} R(-v)dv - \tfrac{1}{\mu_1}\int_0^{\frac{b_1}{\sigma_u^2}} R(-v)dv\\
&+ \Big[(b_1+\mu_1(q_1+p_1))(f_1^2+g_1^2) - e_1(q_1+p_1) - \mu_1 q_1 f_1^2\Big]\\
&- \lim_{\beta\to\infty}\frac{1}{\beta}\Bigg\{\frac{\beta}{\mu_1}\int \log\int_{\mathbb{C}}\int_{\mathbb{C}}\left(\int_{\{\mathbf{x}\in\chi\}}\mathcal{K}(x,y,z)dx\right)^{\frac{\mu_1}{\beta}} DyDzdF_{X^0}(x^0)\Bigg\}
\end{aligned} \tag{B.4.31}$$

Further simplification leads to the 1RSB-ansatz expression given in the propositions.

$$\begin{aligned}\bar{\mathcal{E}}^{\text{lasso}}_{\text{1rsb}} &= \frac{1}{\sigma_u^2}(q_1 + p_1 + \frac{b_1}{\mu_1})R(\frac{-b_1 - \mu_1 p_1}{\sigma_u^2}) - \frac{b_1}{\mu_1 \sigma_u^2}R(-\frac{b_1}{\sigma_u^2}) \\ &\quad - \frac{q_1}{\sigma_u^2}(\frac{b_1 + \mu_1 p_1}{\sigma_u^2})R'(\frac{-b_1 - \mu_1 p_1}{\sigma_u^2}) \end{aligned} \tag{B.4.32}$$

Now, if we only have the RS-ansatz instead of the RSB-ansatz, that is, with $p_1 = 0$, $\mu_1 = 1$, $b_1 = b_0$, and $q_1 = q_0$, then

$$\begin{aligned}\bar{\mathcal{E}}^{\text{lasso}}_{\text{1rsb}} &= \frac{1}{\sigma_u^2}(q_0 + 0 + \frac{b_0}{1})R(\frac{-b_0 - 1 \cdot 0}{\sigma_u^2}) - \frac{b_0}{1 \cdot \sigma_u^2}R(-\frac{b_0}{\sigma_u^2}) \\ &\quad - q_0(\frac{b_1 + 1 \cdot 0}{\sigma_u^2})R'(\frac{-b_0 - 1 \cdot 0}{\sigma_u^2}) \end{aligned} \tag{B.4.33}$$

$$\begin{aligned} &= \frac{q_0}{\sigma_u^2}R(\frac{-b_0}{\sigma_u^2}) + \frac{b_0}{\sigma_u^2}R(\frac{-b_0}{\sigma_u^2}) - \frac{b_0}{\sigma_u^2}R(-\frac{b_0}{\sigma_u^2}) \\ &\quad - \frac{q_1}{\sigma_u^2}(\frac{b_1}{\sigma_u^2})R'(\frac{-b_0}{\sigma_u^2}) \end{aligned} \tag{B.4.34}$$

$$= \bar{\mathcal{E}}^{\text{lasso}}_{\text{rs}} \tag{B.4.35}$$

B.5 Calculations for Section 2.5

Now to evaluate (2.4.31), the macroscopic variables q_0 and b_0 should be simplified further as follows: Beginning from

$$q_0 = \int_{\mathbb{R}}\int_{\mathbb{C}}\left|x^0 - \Psi_1(x)\right|^2 Dz dF_{X^0}(x^0) \tag{B.5.1}$$

$$= \int_{\mathbb{C}}\int_{\mathbb{R}}\left|x^0 - \Psi_1(x)\right|^2 dF_{X^0}(x^0) Dz, \tag{B.5.2}$$

and substituting $dF_{X^0}(x^0) = [(1-\rho)\delta(x^0) + \rho\frac{1}{\sqrt{2\pi}}e^{-\frac{(x^0)^2}{2}}]dx^0$, we get

$$q_0 = \int_{\mathbb{C}}\int_{\mathbb{R}}\left|x^0 - \Psi_1(x)\right|^2[(1-\rho)\delta(x^0) + \rho\frac{1}{\sqrt{2\pi}}e^{-\frac{(x^0)^2}{2}}]dx^0Dz \quad \text{(B.5.3)}$$

$$= (1-\rho)\int_{\mathbb{C}}\int_{\mathbb{R}}\left|x^0 - \Psi_1(x)\right|^2\delta(x^0)dx^0Dz + \rho\frac{1}{\sqrt{2\pi}}\int_{\mathbb{C}}\int_{\mathbb{R}}\left|x^0 - \Psi_1(x)\right|^2e^{-\frac{(x^0)^2}{2}}dx^0Dz \quad \text{(B.5.4)}$$

$$= (1-\rho)\int_{\mathbb{C}}\left|-\Psi_0(x)\right|^2Dz + \rho\frac{1}{\sqrt{2\pi}}\int_{\mathbb{C}}\int_{\mathbb{R}}\left|x^0 - \Psi_1(x)\right|^2e^{-\frac{(x^0)^2}{2}}dx^0Dz, \quad \text{(B.5.5)}$$

and

$$b_0 = \frac{1}{f_0}\int_{\mathbb{R}}\int_{\mathbb{C}}\Re\left\{x^0 - \Psi_1(x)z^*\right\}Dzd F_{X^0}(x^0) \quad \text{(B.5.6)}$$

$$= \frac{1}{f_0}\int_{\mathbb{C}}\int_{\mathbb{R}}\Re\left\{x^0 - \Psi_1(x)z^*\right\}[(1-\rho)\delta(x^0) + \rho\frac{1}{\sqrt{2\pi}}e^{-\frac{(x^0)^2}{2}}]dx^0Dz \quad \text{(B.5.7)}$$

$$= \frac{(1-\rho)}{f_0}\int_{\mathbb{C}}\int_{\mathbb{R}}\Re\left\{x^0 - \Psi_1(x)z^*\right\}\delta(x^0)dx^0Dz + \frac{\rho}{\sqrt{2\pi}f_0}\int_{\mathbb{C}}\int_{\mathbb{R}}\Re\left\{x^0 - \Psi_1(x)z^*\right\}e^{-\frac{(x^0)^2}{2}}dx^0Dz \quad \text{(B.5.8)}$$

$$= \frac{(1-\rho)}{f_0}\int_{\mathbb{C}}\Re\left\{-\Psi_0(x)z^*\right\}Dz + \frac{\rho}{\sqrt{2\pi}f_0}\int_{\mathbb{C}}\int_{\mathbb{R}}\Re\left\{x^0 - \Psi_1(x)z^*\right\}e^{-\frac{(x^0)^2}{2}}dx^0Dz \quad \text{(B.5.9)}$$

where,

$$\Psi_0(x) = \arg\min_{x\in\chi}\left[\left|-zf_0 - 2e_0x - \frac{\gamma}{\sigma_u^2}\right|\right], \quad \text{(B.5.10)}$$

and

$$\Psi_1(x) = \arg\min_{x\in\chi}\left[\left|-zf_0 + 2e_0(x^0 - x) - \frac{\gamma}{\sigma_u^2}\right|\right]. \tag{B.5.11}$$

Solving (B.5.10) and (B.5.11) for x and simplifying further B.5.5 and B.5.9 we get for q_0 and b_0:

$$q_0 = (1-\rho)\int_{\mathbb{C}}\left|\frac{zf_0 + \frac{\gamma}{\sigma_u^2}}{2e_0}\right|^2 Dz + \frac{\rho}{\sqrt{2\pi}}\int_{\mathbb{C}}\int_{\mathbb{R}}\left|\frac{zf_0 + \frac{\gamma}{\sigma_u^2}}{2e_0}\right|^2 e^{-\frac{(x^0)^2}{2}} dx^0 Dz, \tag{B.5.12}$$

$$\begin{aligned} b_0 = \frac{(1-\rho)}{f_0}\int_{\mathbb{C}} \Re\left\{\left(\frac{zf_0 + \frac{\gamma}{\sigma_u^2}}{2e_0}\right)z^*\right\} Dz \\ + \frac{\rho}{\sqrt{2\pi}f_0}\int_{\mathbb{C}}\int_{\mathbb{R}} \Re\left\{x^0(1-z^*) + \left(\frac{zf_0 + \frac{\gamma}{\sigma_u^2}}{2e_0}\right)z^*\right\} e^{-\frac{(x^0)^2}{2}} dx^0 Dz \end{aligned} \tag{B.5.13}$$

Further, replacing $Dz = \frac{1}{\pi}e^{-|z|^2}dz$, we get (2.5.2) and (2.5.3).

Appendix C

Secondary Papers

Clustered Compressed Sensing in fMRI Data Analysis using a Bayesian Framework

Solomon A. Tesfamicael and Faraz Barzideh, *Adapted from: International Journal of Information and Electronics Engineering vol. 4, no. 2, pp. 74-80, 2014.*

Abstract:- This paper provides a Bayesian method of analyzing functional magnetic resonance imaging (fMRI) data. Usually fMRI signals are noisy and need efficient algorithms to estimate or detect the signals accurately. Using a Bayesian framework we have used two different priors: sparsity and clusteredness in the fMRI data by using a general linear model (GLM), which is used as a main tool in fMRI studies. These enhance the effectiveness of the model to help analyze the data better. So in this work we have built the Bayesian framework needed first. Then, we have applied our analysis on synthetic data that we made and that are well known, and the results show that clustered compressive sampling has given better results compared to using of only sparse prior and/ or to the analysis with out considering the two priors. Later we have applied it on fMRI data and the results are much better in terms of signal to noise ratio (SNR) and intensity of images.

Clustered Compressive Sensing: Application to Medical Imaging

Solomon A. Tesfamicael and Faraz Barzideh, *Adapted from: International Journal of Information and Electronics Engineering vol. 5, no. 1, pp. 46-50, 2015.*

Abstract:- This paper provides clustered compressive sensing (CCS) based image pro-

cessing using Bayesian framework applied to medical images. Some images, for example like magnetic resonance images (MRI) are usually very weak due to the presence of noise and due to the weak nature of the signal itself. The compressed sensing (CS) paradigm can be applied in order to boost such signals. We applied the CS paradigm via a Bayesian framework. Using different sparse prior informations and in addition incorporating the special structure that can be found in sparse signal improves image processing. This is shown in the results of this paper. First, we applied our analysis on Angiogram image, then on Shepp-logan phantom finally on another magnetic resonance image (MRI) image. The results show that applying the clustered compressive sensing give better results than the non-clustered version.

Clustered Compressed Sensing-Based Image Denosing Using a Bayesian Framework

Solomon A. Tesfamicael and Faraz Barzideh, *David C. Wyld et al. (Eds) : ITCS, CST, JSE, SIP, ARIA, DMS-2015 pp. 185-197, 2015.*

Abstract:- This paper provides a compressive sensing (CS) method of denoising images using Bayesian framework. Some images, for example magnetic resonance images (MRI) are usually very weak due to the presence of noise and due to the weak nature of the signal itself. So denoising boosts the true signal strength. Under the Bayesian framework, we have used two different priors: sparsity and clusteredness in an image data as prior information to remove noise. Therefore, it is named as clustered compressive sensing based denoising (CCSD). After developing the Bayesian framework, we applied our method on synthetic data, the Shepp-logan phantom and sequences of fMRI images. The results show that applying the CCSD give better results than using only the conventional compressive sensing (CS) methods in terms of Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE). In addition, we showed that this algorithm could have some advantages over the state-of-the-art methods like Block-Matching and 3D Filtering (BM3D).

References

[1] D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289 – 1306, Apr. 2006.

[2] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, no. 12, pp. 5406–5425, 2006.

[3] E. Candes, R. J., and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, no. 2, pp. 489 – 509, Feb. 2006.

[4] S. Qaisar, R. Bilal, W. Iqbal, M. Naureen, and S. Lee, "Compressive sensing: From theory to applications, a survey," *Communications and Networks, Journal of*, vol. 15, no. 5, pp. 443–456, Oct 2013.

[5] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Cambridge University Press, London,UK, 2013.

[6] J. Claerbout and F. Muir, "Robust modeling with erratic data," *GEOPHYSICS*, vol. 38, no. 5, pp. 826–844, 1973. [Online]. Available: http://dx.doi.org/10.1190/1.1440378

[7] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992. [Online]. Available: http://dx.doi.org/10.1016/0167-2789(92)90242-F

[8] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, London,UK, June 29, 2012.

[9] D. Sundman, "Compressed sensing : Algorithms and applications," pp. ix, 35, 2012, qC 20120229.

[10] E. Candes, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, p. 2130, Mar. 2008.

[11] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, May. 2008.

[12] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[13] N. D. and J. J.A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[14] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *Information Theory, IEEE Transactions*, vol. 47, no. 7, pp. 2845 – 2862, Nov. 2001.

[15] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *Information Theory, IEEE Transactions on*, vol. 56, no. 1, pp. 520–544, Jan 2010.

[16] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk, "Random filters for compressive sampling and reconstruction," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 3, May 2006, pp. III–III.

[17] J. Tropp and S. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, June 2010.

[18] J. P. Slavinsky, J. Laska, M. Davenport, and R. Baraniuk, "The compressive multiplexer for multi-channel compressive sensing," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, May 2011, pp. 3980–3983.

[19] M. Davenport, "An introduction to compressive sensing with ieee sps on brainshark." [Online]. Available: http://www.brainshark.com/brainshark/brainshark.net/portal/title.aspx?pid=zCdz10BfTRz0z0

[20] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *Information Theory, IEEE Transactions on*, vol. 55, no. 5, pp. 2230–2249, May. 2009.

[21] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[22] P. Schniter, L. Potter, and J. Ziniel, "Subspace pursuit for compressive sensing signal reconstruction," *Information Theory and Applications Workshop*, pp. 326 – 333, Feb. 2008.

[23] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 8, p. 265274, Nov. 2009.

[24] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, p. 227234, 1995.

[25] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996. [Online]. Available: http://www.jstor.org/stable/2346178

[26] M. A. T. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 4, pp. 586–597, Dec 2007.

[27] F. Teixeira, S. Bergen, and A. Antoniou, "Robust signal recovery approach for compressive sensing using unconstrained optimization," in *Circuits and Systems*

(ISCAS), Proceedings of 2010 IEEE International Symposium on, May 2010, pp. 3521–3524.

[28] A. F. Vivek Goyal and S. Rangan, "The Optimistic Bayesian: Replica method analysis of compressed sensing." [Online]. Available: http://www.math.uiuc.edu/~laugesen/imaha10/goyal_talk.pdf

[29] B. Bah and J. Tanner, "Improved bounds on restricted isometry constants for Gaussian matrices," *CoRR*, vol. abs/1003.3299, 2010.

[30] D. Donoho, A. Maleki, and A. Montanari, "The noise-sensitivity phase transition in compressed sensing," *Information Theory, IEEE Transactions on*, vol. 57, no. 10, pp. 6920–6941, 2011.

[31] Y. Wu and S. Verdu, "Optimal phase transitions in compressed sensing," *Information Theory, IEEE Transactions on*, vol. 58, no. 10, pp. 6241–6263, Oct 2012.

[32] D. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012, pp. 1231–1235.

[33] Z. Ben-Haim, Y. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Transactions on Signal Processinge*, vol. 346, no. 9-10, pp. 589–592, Nov. 2010.

[34] A. Fletcher and S. Rangan, "The optimistic bayesian: Replica method analysis of compressed sensing." [Online]. Available: www.math.uiuc.edu/~laugesen/imaha10/goyal_talk.pdf

[35] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, Sept 2009, pp. 52–59.

[36] Seidenfeld, "R. a. Fisher's fiducial argument and Bayes' theorem," *Institute of Mathematical Statistics*, vol. 7, no. 3, pp. 358–368, Aug., 1992. [Online]. Available: http://www.jstor.org/stable/2246072

[37] E. T. Jaynes., *Probability Theory, the Logic of Science*, G. L. Bretthorst, Ed. Cambridge University Press, London,UK, June 2003.

[38] B. Efron, "Modern science and the Bayesian-frequentist controversy," pp. 1465–1466, 2005. [Online]. Available: http://statistics.stanford.edu/~ckirby/techreports/BIO/BIO%20233.pdfonJun25,2013.

[39] D. P. Ankerst, "Kendall's advanced theory of statistics, vol. 2b: Bayesian inference," *Journal of the American Statistical Association*, vol. 100, no. 472, pp. 1465–1466, 2005. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1198/jasa.2005.s62

[40] G. Gonzlez-Torre, F. Javier, and E. Moreno, "On the frequentist and Bayesian approaches to hypothesis testing," *SORT (Statistics and Operations*

Research Transactions), vol. 30, no. 1, pp. 3–28, 2006. [Online]. Available: http://hdl.handle.net/2099/3789

[41] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *Signal Processing, IEEE Transactions on*, vol. 56, no. 6, pp. 2346–2356, June 2008.

[42] S. 9.2, *Users Guides: Introduction to Bayesian Analysis Procedures*. SAS Institute Inc. , Cary, NC, USA, 2008. [Online]. Available: https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#introbayes_toc.htm

[43] P. Schniter, L. Potter, and J. Ziniel, "Fast Bayesian matching pursuit," in *Information Theory and Applications Workshop, 2008*, Jan 2008, pp. 326–333.

[44] H. Luo and S. Puthusserypady, "A sparse Bayesian method for determination of flexible design matrix for fmri data analysis," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 52, no. 12, pp. 2699–2706, 2005.

[45] F. Merli, G. Vitetta, and X. Wang, "A Bayesian multiuser detector for MIMO-OFDM systems affected by multipath fading, carrier frequency offset and phase noise," in *Signal Processing and Information Technology, 2007 IEEE International Symposium on*, Dec 2007, pp. 1137–1142.

[46] L. Li and B. Jafarpour, "A sparse Bayesian framework for conditioning uncertain geologic models to nonlinear flow measurements," *Advances in Water Resources*, vol. 33, no. 9, pp. 1024 – 1042, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S030917081000117X

[47] D. J. C. MacKay., *Information Theory, Inference and Learning Algorithms*, ser. Technical report (Stanford University. Department of Statistics). Stanford, Calif. : Division of Biostatistics, Stanford University, 2005, no. 2005-19B/233.

[48] S. A. Tesfamicael and F. Barzideh, "Clustered compressed sensing in fMRI data analysis using a Bayesian framework," *International Journal of Information and Electronics Engineerings*, vol. 4, no. 2, pp. 74–80, 2014.

[49] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," *CoRR*, vol. abs/0906.3234, 2009.

[50] S. Ganguli and H. Sompolinsky, "Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis," *Annual Review of Neuroscience*, vol. 35, no. 1, pp. 485–508, 2012, pMID: 22483042. [Online]. Available: http://dx.doi.org/10.1146/annurev-neuro-062111-150410

[51] M. Advani, S. Lahiri, and S. Ganguli, "Statistical mechanics of complex neural systems and high dimensional data," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2013, no. 03, p. P03014, 2013. [Online]. Available: http://stacks.iop.org/1742-5468/2013/i=03/a=P03014

[52] M. Talagrand, "The parisi formula," *Annals of Math.*, vol. 163, pp. 221–263, 2006. [Online]. Available: http://www.jstor.org/stable/20159953

[53] R. R. Müller, G. Alfano, B. M. Zaidel, and R. de Miguel, "Applications of large random matrices in communications engineering," *CoRR*, vol. abs/1310.5479, 2013.

[54] R. De Miguel and R. Muller, "On convex vector precoding for multiuser MIMO broadcast channels," *Signal Processing, IEEE Transactions on*, vol. 57, no. 11, pp. 4497–4508, Nov 2009.

[55] B. Zaidel, R. Muller, A. Moustakas, and R. De Miguel, "Vector precoding for gaussian MIMO broadcast channels: Impact of replica symmetry breaking," *Information Theory, IEEE Transactions on*, vol. 58, no. 3, pp. 1413–1440, March 2012.

[56] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *Journal of Physics F: Metal Physics*, vol. 5, no. 5, p. 965, 1975. [Online]. Available: http://stacks.iop.org/0305-4608/5/i=5/a=017

[57] V. Dotsenko, *Introduction to the Replica Theory of Disordered Statistical Systems*. Cambridge University Press. Copyright, Aug. 2005.

[58] N. Sourlas, "Spin glasses, error-correcting codes and finite-temperature decoding," *EPL (Europhysics Letters)*, vol. 25, no. 3, p. 159, 1994. [Online]. Available: http://stacks.iop.org/0295-5075/25/i=3/a=001

[59] Y. Kabashima and D. Saad, "Statistical mechanics of error-correcting codes," *EPL (Europhysics Letters)*, vol. 45, no. 1, p. 97, 1999. [Online]. Available: http://stacks.iop.org/0295-5075/45/i=1/a=097

[60] A. Montanari and N. Sourlas, "The statistical mechanics of turbo codes," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 18, no. 1, pp. 107–119, 2000. [Online]. Available: http://dx.doi.org/10.1007/PL00011086

[61] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *Information Theory, IEEE Transactions on*, vol. 48, no. 11, pp. 2888–2910, Nov 2002.

[62] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *Journal of Physics A: Mathematical and General*, vol. 36, no. 43, p. 11111, 2003. [Online]. Available: http://stacks.iop.org/0305-4470/36/i=43/a=030

[63] R. Muller and W. Weichselberger, "Minimum probability of error and channel capacity in large dual antenna array systems," in *Telecommunications, 2003. ICT 2003. 10th International Conference on*, vol. 1, Feb 2003, pp. 211–216 vol.1.

[64] S. Ganguli and H. Sompolinsky, "Statistical mechanics of compressed sensing," *Phys. Rev. Lett.*, vol. 104, p. 188701, May 2010. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.104.188701

[65] S. A. Tesfamicael, "Compressed sensing performance analysis via replica method using bayesian framework," *International Journal of Simulation Systems, Science and Technology(IJSSST)*, vol. 16, no. 3, 2016. [Online]. Available: http://ijssst.info/Vol-16/No-3/paper16.pdf

[66] T. Tanaka, "Moment problem in replica method," *Interdisciplinary information sciences*, vol. 13, pp. 17–23, 2007. [Online]. Available: http://hdl.handle.net/10097/17348

[67] D. Guo and S. Verdu, "Randomly spread CDMA: Asymptotics via statistical physics," *Information Theory, IEEE Transactions on*, vol. 51, no. 6, pp. 1983–2010, June 2005.

[68] Y. Kabashima, T. Wadayama, and T. Tanaka, "A typical reconstruction limit for compressed sensing based on lp-norm minimization," *Journal of Statistical Mechanics: Theory and Experiment*, vol. Volume 2009, September 2009.

[69] K. Takeda and Y. Kabashima, "Statistical mechanical assessment of a reconstruction limit of compressed sensing: Toward theoretical analysis of correlated signals," *Europhysics Letters Association*, vol. EPL 95 18006, 2011.

[70] ——, "Statistical mechanical analysis of compressed sensing utilizing correlated compression matrix," *CoRR*, vol. abs/1001.4361, 2010.

[71] S. N. Hong, "Statistical physics of compressed sensing." 2012. [Online]. Available: http://www.isi.edu/~galstyan/courses/Presentations/CSCI599_SNHong.pdf

[72] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Physical Review X*, vol. 2, no. 2, p. 021005, apr 2012.

[73] G. Parisi, "The order parameter for spin glasses: a function on the interval 0-1," *Journal of Physics A: Mathematical and General*, vol. 13, no. 3, p. 1101, 1980. [Online]. Available: http://stacks.iop.org/0305-4470/13/i=3/a=042

[74] T. Uezu, M. Yoshida, T. Tanaka, and M. Okada, "Statistical mechanical analysis of CDMA multiuser detectors: At stability and entropy of the RS solution, and 1RSB solution," *Progress of Theoretical Physics Supplement*, vol. 157, pp. 254–257, 2005. [Online]. Available: http://ptps.oxfordjournals.org/content/157/254.abstract

[75] Abo-Zahhad, H. M.M., A.I., and A. Mohamed, "Compressive sensing algorithms for signal processing applications: A survey." *Int. J. Communications, Network and System Sciences,*, vol. 8, pp. 197–216, 2015. [Online]. Available: http://dx.doi.org/10.4236/ijcns.2015.86021

[76] R. University, "Compressive imaging: A new single-pixel camera." [Online]. Available: http://dsp.rice.edu/cscamera

[77] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid mr imaging," *Magnetic Resonance in*

Medicine, vol. 58, no. 6, pp. 1182–1195, 2007. [Online]. Available: http://dx.doi.org/10.1002/mrm.21391

[78] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 83–91, March 2008.

[79] Y. Huang, J. Paisley, Q. Lin, X. Ding, X. Fu, and X.-P. Zhang, "Bayesian nonparametric dictionary learning for compressed sensing MRI," *Image Processing, IEEE Transactions on*, vol. 23, no. 12, pp. 5007–5019, Dec 2014.

[80] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *Radar Conference, 2007 IEEE*, April 2007, pp. 128–133.

[81] M. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *Signal Processing, IEEE Transactions on*, vol. 57, no. 6, pp. 2275–2284, June 2009.

[82] C. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *Communications Magazine, IEEE*, vol. 48, no. 11, pp. 164–174, November 2010.

[83] Z. Fanzi, Z. Tian, and C. Li, "Distributed compressive wideband spectrum sensing in cooperative multi-hop cognitive networks," in *Communications (ICC), 2010 IEEE International Conference on*, May 2010, pp. 1–5.

[84] G. Taubock and F. Hlawatsch, "A compressed sensing technique for ofdm channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, March 2008, pp. 2885–2888.

[85] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, June 2010.

[86] C. Berger, S. Zhou, J. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *Signal Processing, IEEE Transactions on*, vol. 58, no. 3, pp. 1708–1721, March 2010.

[87] ——, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *Signal Processing, IEEE Transactions on*, vol. 58, no. 3, pp. 1708–1721, March 2010.

[88] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Compressive data gathering for large-scale wireless sensor networks," in *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '09. New York, NY, USA: ACM, 2009, pp. 145–156. [Online]. Available: http://doi.acm.org/10.1145/1614320.1614337

[89] M. Laifenfeld and I. Bilik, "Distributed compressive sensing and communications

in wireless sensor networks," in *Electrical Electronics Engineers in Israel (IEEEI), 2012 IEEE 27th Convention of*, Nov 2012, pp. 1–5.

[90] C. Alippi, G. Anastasi, M. Di Francesco, and M. Roveri, "Energy management in wireless sensor networks with energy-hungry sensors," *Instrumentation Measurement Magazine, IEEE*, vol. 12, no. 2, pp. 16–23, April 2009.

[91] S.-Y. Fu, X.-K. Kuai, R. Zheng, G.-S. Yang, and Z.-G. Hou, "Compressive sensing approach based mapping and localization for mobile robot in an indoor wireless sensor network," in *Networking, Sensing and Control (ICNSC), 2010 International Conference on*, April 2010, pp. 122–127.

[92] Q. Ling and Z. Tian, "Decentralized sparse signal recovery for compressive sleeping wireless sensor networks," *Signal Processing, IEEE Transactions on*, vol. 58, no. 7, pp. 3816–3827, July 2010.

[93] S. A. Tesfamicael and L. Lundheim, "Compressed sensing based rotative quantization in temporally correlated MIMO channels," in *Recent Developments in Signal Processing*, 2013. [Online]. Available: http://conf-scoop.org/RDSP-2013/1_Solomon_RDSP.pdf

[94] S. A. Tesfamicael and F. Barzideh, "Clustered compressed sensing via bayesian framework," in *IEEE UKSim-AMSS 17th International Conference on Computer Modelling and Simulation, UKSim2015 - 19.S.Image, Speech and Signal Processing, Cambridge, United Kingdom*, 2015, pp. 25–27. [Online]. Available: edas.paper-1570090629

[95] S. A. Tesfamicael, F. Barzideh, and L. Lundheim, "Improved reconstruction in compressive sensing of clustered signals," in *AFRICON, 2015*, Sept 2015, pp. 1–7.

[96] R. Muller, D. Guo, and A. Moustakas, "Vector precoding for wireless MIMO systems and its replica analysis," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 3, pp. 530–540, April 2008.

[97] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Measurements vs. bits: Compressed sensing meets information theory," in *in Proc. Allerton Conf. on Comm., Control, and Computing*, 2006.

[98] D. Baron, S. Sarvotham, and R. Baraniuk, "Bayesian compressive sensing via belief propagation," *Signal Processing, IEEE Transactions on*, vol. 58, no. 1, pp. 269–280, Jan 2010.

[99] S. Rangan, "Estimation with random linear mixing, belief propagation and compressed sensing," *CoRR*, vol. abs/1001.2228, 2010.

[100] A. Maleki, L. Anitori, Z. Yang, and R. Baraniuk, "Asymptotic analysis of complex lasso via complex approximate message passing (camp)," *Information Theory, IEEE Transactions on*, vol. 59, no. 7, pp. 4290–4308, July 2013.

[101] S. Rangan, "Probabilistic reconstruction in compressed sensing: algorithms, phase

diagrams, and threshold achieving matrices," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 57, no. 08, Aug 2012.

[102] A. Tulino, G. Caire, S. Verdu, and S. Shamai, "Support recovery with sparsely sampled free random matrices," *Information Theory, IEEE Transactions on*, vol. 59, no. 7, pp. 4243–4271, July 2013.

[103] J. Ma, X. Yuan, and L. Ping, "Turbo compressed sensing with partial DFT sensing matrix," *Signal Processing Letters, IEEE*, vol. 22, no. 2, pp. 158–161, Feb 2015.

[104] S. Ganguli and H. Sompolinsky, "Statistical mechanics of compressed sensing," *Phys. Rev. Lett.*, vol. 104, p. 188701, May 2010. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.104.188701

[105] S. U. K. Takeda and Y. Kabashima, "Analysis of CDMA systems that are characterized by eigenvalue spectrum," *EPL (Europhysics Letters)*, vol. 76, p. 1193, 2010. [Online]. Available: http://iopscience.iop.org/0295-5075/76/6/1193

[106] K. Takeda, A. Hatabu, and Y. Kabashima, "Statistical mechanical analysis of the linear vector channel in digital communication," *CoRR*, vol. abs/0707.3336, 2007.

[107] S. M. Kay, *Fundamentals of statistical signal processing : Estimation theory.*, 1st ed. Prentice-Hall PTR, apr 2010. [Online]. Available: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0133457117

[108] R. Podgornik, "Book review: Statistical physics of spin glasses and information processing. hidetoshi nishimori, oxford science publications, Oxford University Press, 2001," *Journal of Statistical Physics*, vol. 109, no. 1-2, pp. 335–337, 2002. [Online]. Available: http://dx.doi.org/10.1023/A%3A1020091132740

[109] K. Takeda, A. Hatabu, and Y. Kabashima, "Statistical mechanical analysis of the linear vector channel in digital communication," *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 47, p. 14085, 2007. [Online]. Available: http://stacks.iop.org/1751-8121/40/i=47/a=004

[110] V. Cevher, C. Hegde, M. F. Duarte, and R. G. Baraniuk, "Sparse signal recovery using Markov random fields," 2008.

[111] V. Cevher, P. Indyk, C. Hegde, and R. G. Baraniuk, "Recovery of clustered sparse signals from compressive measurements," in *in Int. Conf. on Sampling Theory and Applications (SAMPTA*, 2009.

[112] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theor.*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009. [Online]. Available: http://dx.doi.org/10.1109/TIT.2009.2030471

[113] L. Yu, H. Sun, J. P. Barbot, and G. Zheng, "Bayesian compressive sensing for clustered sparse signals," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, May 2011, pp. 3948–3951.

[114] L. Yu, J.-P. Barbot, G. Zheng, and H. Sun, "Compressive sensing for cluster

structured sparse signals: Variational Bayes approach," March 2011. [Online]. Available: http://hal.archives-ouvertes.fr/file/index/docid/573953

[115] M. Soltanolkotabi, E. Elhamifar, and E. J. Candès, "Robust subspace clustering," *CoRR*, vol. abs/1301.2603, 2013.

[116] E. Candes and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009. [Online]. Available: http://dx.doi.org/10.1007/s10208-009-9045-5

[117] S. Chand, "On tuning parameter selection of lasso-type methods - a monte carlo study," in *Applied Sciences and Technology (IBCAST), 2012 9th International Bhurban Conference on*, Jan 2012, pp. 120–129.

[118] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, pp. 3397–3415, Dec 1993.

[119] D. Donoho, "De-noising by soft-thresholding," *Information Theory, IEEE Transactions on*, vol. 41, no. 3, pp. 613–627, May 1995.

[120] B. Efron., *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, London,UK, October 2003. [Online]. Available: http://www.worldcat.org/title/modern-science-and-the-bayesian-frequentist-controversy/oclc/272554701

[121] K. J. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner, "Classical and Bayesian inference in neuroimaging: Theory," *NeuroImage*, vol. 16, pp. 465–483, 2002.

[122] S. A. Tesfamicael and F. Barzideh, "Clustered compressive sensing: Application on medical imaging," *International Journal of Information and Electronics Engineering*, vol. 5, no. 1, pp. 48–50, 2015. [Online]. Available: http://www.ijiee.org/vol5/499-SA0037.pdf

[123] ——, "Clustered compressive sensing based image denosising using Bayesian framework," in *AIRCC Publishing Corporation*, vol. 5, no. 1, 2015, pp. 185–197. [Online]. Available: http://airccj.org/CSCP/vol5/csit53218.pdf

[124] G. Image:University Hospital Rechts der Isar, Munich, "Depiction of vessel diseases with a wide range of contrast and non-contrast enhanced techniques." [Online]. Available: http://usa.healthcare.siemens.com/magnetic-resonance-imaging/0-35-to-1-5t-mri-scanner/magnetom-avanto/use

[125] B. Godana and T. Ekman, "Rotative quantization using adaptive range for temporally correlated MIMO channels," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, 2013, pp. 1233–1238.

[126] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. K. ., *MIMO-OFDM Wireless Communications with MATLAB*. Wiley-IEEE Press, October, 2010.

[127] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998. [Online]. Available: http://dx.doi.org/10.1023/A%3A1008889222784

[128] T. Kim, D. Love, B. Clerckx, and S. Kim, "Differential rotation feedback MIMO system for temporally correlated channels," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, 2008, pp. 1–5.

[129] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 8, pp. 1341–1365, 2008.

[130] A. Narula, M. Lopez, M. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *Selected Areas in Communications, IEEE Journal on*, vol. 16, no. 8, pp. 1423–1436, 1998.

[131] J. Roh and B. Rao, "Design and analysis of mimo spatial multiplexing systems with quantized feedback," *Signal Processing, IEEE Transactions on*, vol. 54, no. 8, pp. 2874–2886, 2006.

[132] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2562–2579, 2003.

[133] D. Love and R. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *Information Theory, IEEE Transactions on*, vol. 51, no. 8, pp. 2967–2976, 2005.

[134] Hyungjoon, W. Seo, and D. Hong, "Compressive feedback based on sparse approximation for multiuser MIMO systems," *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 2, pp. 1017–1023, 2010.

[135] S. Qaseem, T. Y. Al-Naffouri, and S. Alghadhban, "Compressive sensing for feedback reduction in MIMO broadcast channels," in *Telecommunications (ICT), 2010 IEEE 17th International Conference on*, 2010, pp. 356–361.

[136] R. de Sa Netto and C. Cavalcante, "Quantization and noise impact over feedback reduction of MIMO systems using compressive sensing," in *Wireless Communication Systems (ISWCS), 2012 International Symposium on*, 2012, pp. 396–400.

[137] M. Davenport, *Random observations on random observations: Sparse signal acquisition and processing*. PhD thesis, Rice University, Aug. 2010.