

© 2015

Mohammad Ramezanali

ALL RIGHTS RESERVED

EXTRACTING SPARSE SIGNALS FROM
HIGH-DIMENSIONAL DATA:
A STATISTICAL MECHANICS APPROACH

By

MOHAMMAD RAMEZANALI

A dissertation submitted to the
Graduate School—New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
Graduate Program in Physics and Astronomy

written under the direction of
Professor Anirvan M. Sengupta
and approved by

New Brunswick, New Jersey

October, 2015

ABSTRACT OF THE DISSERTATION

Extracting Sparse Signals from High-dimensional Data: A Statistical Mechanics Approach

By MOHAMMAD RAMEZANALI

Dissertation Director:

Professor Anirvan M. Sengupta

Sparse reconstruction algorithms aim to retrieve high-dimensional sparse signals from a limited amount of measurements under suitable conditions. As the number of variables go to infinity, these algorithms exhibit sharp phase transition boundaries where the sparse retrieval breaks down. Several sparse reconstruction algorithms are formulated as optimization problems. Few of the prominent ones among these have been analyzed in the literature by statistical mechanical methods. The function to be optimized plays the role of energy. The treatment involves finite temperature replica mean-field theory followed by the zero temperature limit. Although this approach has been successful in reproducing the algorithmic phase transition boundaries, the replica trick and the non-trivial zero temperature limit obscure the underlying reasons for the failure of the algorithms. In this thesis, we employ the “cavity method” to give an alternative derivation of the phase transition boundaries, working directly in the zero-temperature limit. This approach provides insight

into the origin of the different terms in the mean field self-consistency equations. The cavity method naturally generates a local susceptibility which leads to an identity that clearly indicates the existence of two phases. The identity also gives us a novel route to the known parametric expressions for the phase boundary of the Basis Pursuit algorithm and to the new ones for the Elastic Net. These transitions being continuous (second order), we explore the scaling laws and critical exponents that are uniquely determined by the nature of the distribution of the density of the nonzero components of the sparse signal. Not only is the phase boundary of the Elastic Net different from that of the Basis Pursuit, we show that the critical behavior of the two algorithms are from different universality classes.

Acknowledgments

In the process of my graduate studies, I got inspired and motivated to work on many interesting problems for which I received constant guidance and encouragement from the people whom I worked with and learned from. I have also been fortunate to receive support and relief of graduate-student-related tension from the friends and family over the past six years. Thank you all for everything.

The biggest influence on the work in this dissertation and support during tough times in the Ph.D. pursuit, has been from my advisor Anirvan Sengupta. His broad and deep knowledge of physics, penetrating insight, charismatic curiosity, and inimitable work ethic have been a constant source of inspiration for me since I joined his group. Everything in this thesis is a direct result of preliminary ideas from Anirvan, our discussions on the white board, and the comments he made at times when I went to his office to discuss a problem or we walked to grab a coffee at the student center. In this process I learned tremendously from Anirvan on how to think, describe and analyze problems intuitively is my prized possession. It has been an absolute pleasure and honor to be his student, and I am immensely grateful to him for everything.

It has been a pleasure to collaborate with and get to know Partha Mitra. I look forward to continuing our discussions in the coming years. I would like to thank all the members in my thesis committee, Sanghyuk Lee, Swagatam Mukhopadhyay, Emil Yuzbashyan and Alexander Zamolodchikov, for reading my thesis, exchanging several ideas, and providing constructive feedback. I would also like to express my sincere gratitude to all the professors at Rutgers university from whom I learned about physics, mathematics, writing, and much more.

A very special thank to Viji Nagaraj, Anirvan's wife, whose kind heart has made the past few years immeasurably better and I never felt the absence of not seeing my mother as much. The friends I have made here are some of the best people I have known. Thanks to my fellow students in BioMaps group, Ted Malliaris, Michael Manhart, Pasha Khromov, George Locke, Dave Hassan, Razvan Chereji, and Willow Kion-Crosby as well my friends in physics department, Bryan Leung, Mike Park, Simon Knapen, Pietro Longhi, Michael Solway, and Victor Alexandrov. Thank you all for making these years count. A bigger thanks to my close friend outside of physics, Conor O'Malley who I have always enjoyed talking with him about philosophy, classical music and life during our trips to Rutgers garden and Princeton and my best housemates, Carlos Cocovi and Trushar Rathod for all the good time we had together and their support during the time when I was writing my dissertation.

Finally, I would like to thank my family members: My father, who sadly passed away during my Ph.D. pursuit, and my mother, I am truly fortunate to have you who have always encouraged me in my academic pursuits, no matter how obscure they may have seemed to them, and have always believed in me as a scientist and as a person. It is only because of your encouragement, love, and prayers that I could do this or anything else. Many thanks to my sisters and brothers for their consistent love and support.

Dedication

Dedicated to the memory of my Father.

Table of Contents

Abstract	ii
Acknowledgments	iv
Dedication	vi
List of Tables	x
List of Figures	xi
0.1. Notation	1
1. Introduction	2
1.1. Main Idea	2
1.1.1. Phase Transition	4
1.2. Applications	5
1.2.1. Biomedical Imaging	6
1.2.2. Systems Biology	7
1.2.3. Group Testing	8
1.2.4. Face Recognition	8
1.3. Contribution of this Thesis	9
1.4. Outline of the Thesis	10
2. An Overview of Compressed Sensing	12
2.1. Problem Formulation	12

2.1.1.	Compressed Sensing	12
2.1.2.	Restricted Isometry Properties	14
2.1.3.	Signal Reconstruction Algorithms	16
2.1.4.	ℓ_2 -norm Minimization	17
2.1.5.	ℓ_0 -norm Minimization	18
2.1.6.	ℓ_1 -norm Minimization	20
2.1.7.	Conversion of (P_1) to Linear Programming (L_P)	22
2.1.8.	Geometric Interpretation	23
3.	Statistical Sparse Reconstruction Methods	25
3.1.	Formulation of the Regularized Least-Squares	25
3.2.	Replica Approach	28
3.3.	Finite Temperature Cavity Method	33
4.	Cavity Method at Zero Temperature	40
4.1.	Susceptibility and Conjugate Variables	40
4.2.	Zero-temperature Cavity Method: Removing a Variable Node	47
4.3.	Zero Temperature Cavity Method: Removing a Constraint Node	53
5.	Phase Transition in Sparse Reconstruction	59
5.1.	Ridge Regression	59
5.2.	Basis Pursuit: ℓ_1 -norm Minimization	63
5.3.	Understanding the Extremely Sparse Limit	67
5.4.	Critical Exponents	71
5.4.1.	Into the Error-prone Regime ($\vartheta \rightarrow 0$ & $\sigma_\zeta^2 = 0$)	74
5.4.2.	Role of an Additive Noise ($\vartheta \rightarrow 0$ & $\sigma_\zeta^2 \neq 0$)	75
5.4.3.	ϑ Trade-off in the Noisy system ($\vartheta \neq 0$ & $\sigma_\zeta^2 \neq 0$)	75
5.5.	Elastic Net	77

5.6. Numerical Experiment	80
6. Conclusion	83
Appendix A. Correlated Measurement Matrices	84
Bibliography	86

List of Tables

5.1. This table presents input parameters used to explain effective individual op- timization	60
5.2. This table shows the comparison of critical exponents for Basis Pursuit and Elastic Net near phase transition.	80

List of Figures

1.1. The phase transition phenomenon in compressed sensing.	5
1.2. A head radiograph	6
2.1. A plot of $J_p(x)$	16
2.2. ℓ_2 penalty shrinks the estimates of the regression coefficients towards zero, however	19
2.3. ℓ_1 penalty shrinks the estimates of the regression coefficients towards zero with only few non-zero regression coefficient	21
2.4. A schematic illustration of how different regularization terms lead to sparse and non-sparse solutions	23
4.1. The diagrammatic expansion of susceptibility.	43
4.2. Planar diagrams contributing to the self-energy.	45
4.3. The leading planar diagrams in covariance computation	46
4.4. Bipartite graph	48
4.5. The $(N - 1, M)$ cavity system	50
4.6. The $(N - 1, M - 1)$ cavity system.	53
5.1. The soft thresholding function	65
5.2. The red curve is the theoretical phase boundary	67
5.3. The transition point for ℓ_1 -norm minimization is	73
5.4. The transition boundary is where the red and blue curves meet	77
5.5. Comparison of MSE for different λ_2	81
5.6. Following the trends where the curves merge, we can find the critical exponent	82

5.7. Varying λ sweeps out entire optimal tradeoff curves	82
--	----

0.1 Notation

Throughout this thesis, for matrices, we use boldface capital letters like \mathbf{H} , and we use \mathbf{H}^T , $\text{tr}(\mathbf{H})$, to denote the transpose and trace, respectively.

For vectors, we use boldface small letters like \mathbf{x} with x_a representing the a^{th} element of \mathbf{x} . We use $[\dots]_{\text{vars}}^{\text{av}}$ to denote quenched averages, with the relevant quenched variables indicated in the subscript. In particular, this average depends on two random variables \mathbf{x}_0 and \mathbf{H} that are drawn from distribution $P_0(\mathbf{x}_0)$ and $\mathcal{P}(\mathbf{H})$. For a Gaussian random variable \mathbf{x} with mean μ and variance v , we write the pdf as $\mathcal{N}(x; \mu, v)$ and, for the special case of $\mathcal{N}(x; 0, 1)$, we abbreviate the pdf as $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ and write the cdf as $\Phi(x) = \int_x^\infty dz \phi(z)$. Dirac's delta function is written as $\delta(x)$ and δ_{mn} is the Kronecker delta symbol. The following table summarizes the most commonly used symbols:

Notation		
Symbol	Ambient Space	Description
N	$\mathbb{N} = \{0, 1, 2, \dots\}$	Signal dimension
\mathbf{x}	\mathfrak{R}^N	Signal
M	\mathbb{N}	Measurement dimension
\mathbf{y}	\mathfrak{R}^M	Measurement
K	\mathbb{N}	Number of non-zero coefficients (Sparsity) of a signal
\mathbf{H}	$\mathfrak{R}^{M \times N}$	Measurement (sensing) matrix
$ \Omega $	\mathbb{N}	Cardinality of a set Ω
$\ \cdot\ _0$	\mathbb{N}	ℓ_0 pseudo-norm: $ \{x_a \neq 0\} $
$\ \cdot\ _p$	\mathfrak{R}	ℓ_p -norm: $(\sum_a^N x_a ^p)^{1/p}$
MSE	\mathfrak{R}	Mean Squared Error

Chapter 1

Introduction

1.1 Main Idea

Recent technological revolutions have enabled sampling, analyzing, storage, and transmission of high-dimensional signals in both variety and quantity. However, most of the time, the underlying structure of associated signal, a vector of values, has much lower information content in a particular basis. For instance, a high resolution natural image might only have a few dominant components in an appropriately chosen basis.

In these cases, the first step to acquire a signal is using sensing devices to make a number of measurements related to the signal by using a procedure suitable for the application. A naive approach would be to measure individual components of the signal vector, making the number of measurements equal to the superficial dimension of the signal. However, due to limitations of time, cost of storage and other constraints, one is often restricted in the amount of data that can be collected. In particular, in presence underlying structure (e.g. sparsity) in the signal, the resulting measurements are highly redundant. This observation raises an important question: since large portion of the sampled data has little impact on the signal quality, why not to recover the desired signal by acquiring the ‘optimal’ amount of data in an efficient sampling process?

Compressed sensing began with the seminal work of Candés and Donoho [1, 2] as a mathematical theory and set of techniques that aim to improve reconstruction quality from a given data set by using the underlying structure of the unknown object. The main idea is to efficiently sample the signal up to obtaining the necessary information content by taking

suitably designed linear measurements and reconstruct it by running some sophisticated and efficient recovery algorithms.

The usual setting of the sparse retrieval problem is a linear equation, $\mathbf{y} = \mathbf{H}\mathbf{x}$, where \mathbf{y} is an M -dimensional measurement vector, \mathbf{H} is an $M \times N$ measurement matrix, and \mathbf{x} is an N -dimensional unknown sparse parameter vector. The goal is to reconstruct the signal \mathbf{x} by taking a vector \mathbf{y} of dimension $M \ll N$. The recovery of the signal requires solving a non-trivial inverse problem. We know from linear algebra that in general it is not possible to reconstruct an arbitrary N -dimensional signal from smaller set of linear measurements. Thus, we need to use our knowledge of the signal structure. In particle, we consider the signal \mathbf{x} to be K -sparse, meaning the vector has at most K non-zero components.

It is well-known that many signals such as real-world signals are sparse in particular bases [3]. Of course, one would think that since sparse signals lie in a much lower dimensional space, they may be reconstructed from a few linear measurements. Even though, this is true, the difficulty is to determine in which lower dimensional subspace such a signal lies. That is, we may know that the signal has few non-zero components, but we do not know which components those are. Theoretically, if a signal has K non-zero components, then, typically, recovery is possible with just $K + 1$ measurements. However, recovery using only this property would require searching through exponentially large set of $\binom{N}{K}$ possible lower dimensional subspaces, which, in practice, is not computationally feasible.

Here is the key point that distinguishes the compressed sensing problem from a general inverse problem: In the compressed sensing problem one has the flexibility of taking more measurements than required ‘in principle’, so as to solve the inverse problem efficiently. In practice, one can extract the full information content of the signal at the cost of moderately higher number of measurements. Recent works in compressed sensing and related areas precisely address questions concerning minimum amount of measurements required for the recovery of sparse signals in a fast, accurate and robust (low sensitivity to additive noise and to incomplete knowledge of the measurement matrix) manner.

1.1.1 Phase Transition

When the measurements are noiseless, our goal is to reconstruct the original signal as perfectly as possible by driving the error probability to zero as size of the signal N grows. One of the most important and popular reconstruction algorithms for compressed sensing is the Basis Pursuit algorithm [4], namely the constrained ℓ_1 -norm minimization algorithm. In particular, for measurement matrices that have independent and identically distributed (iid) Gaussian entries, it is shown that Basis Pursuit requires as low as $M > O(K \log(N/K))$ measurements for perfect reconstruction [5]. Rigorous results, based on the so called Restricted Isometry Property [6] of the measurement matrix, provide parameter regions where the algorithm is guaranteed to work. However, these results do not give the exact threshold where the procedure breaks down. However, it turns out that for many different measurement matrices and input signals, there exists such threshold that the performance failure of such recovery algorithms with polynomial time complexity occurs at a sharp boundary [7–9] distinguishing where the correct unique solution is found, and where the algorithms fail (see Fig. 1.1). Such a phenomenon is known as a phase transition in statistical physics. This is an algorithmic phase transition with a zero-one law, where the probability of correct reconstruction (suitably defined) jumps from zero to one at the transition boundary. Analytical formulae, based on the message-passing method and the replica formalism borrowed from statistical physics, exists for the phase transition boundary in the literature, but the corresponding derivations are not transparent. Particularly, it does not address the critical behavior and scaling laws near the boundary of the phase transition.

The work presented in this thesis is motivated by exploring these transition boundaries using mean field “cavity” approach borrowed from statistical physics. The cavity mean field equations are named after a physical context in spin systems in solid state physics [10, 11]. The goal is to take into account the non-trivial dependencies by estimating the reaction of all the other “spin” variables when a single spin is removed from the system, thereby leaving a “cavity”. The method has since been applied to a wider class of problems including the

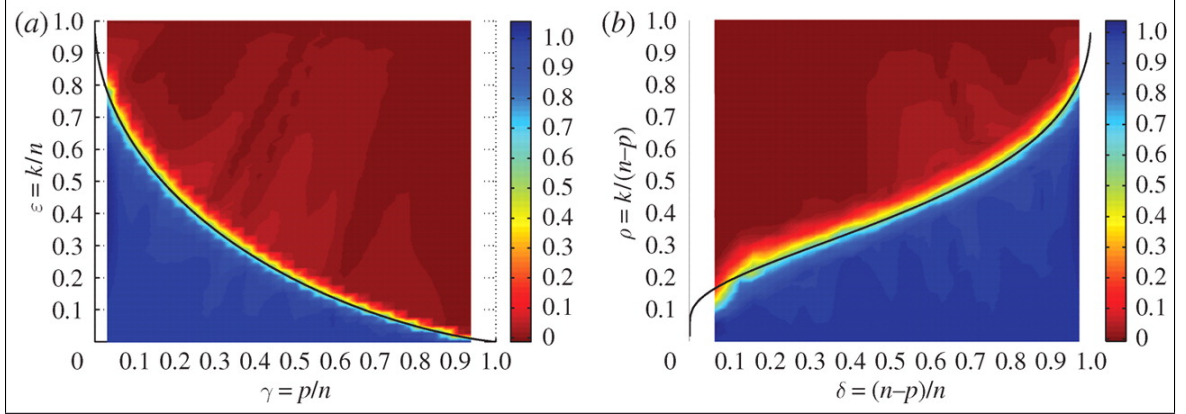


Figure 1.1: The phase transition phenomenon in compressed sensing. These graphs show the empirical probability that the ℓ_1 -norm minimization method successfully recovers the original signal. The brightness of each point reflects the observed probability of success, ranging from certain failure (red) to certain success (blue). Here, p is the number of measurements and n is the size of the signal. (The graph is reproduced from [7])

satisfiability problem [12, 13] and Hopfield neural networks [14]. The cavity method leads to the same results as obtained by replica trick [11] and is closely related to the message-passing algorithm in graphical models [15]. However, we find that for the problem at hand, the cavity method leads to a considerable simplification by utilizing the fact that the system of variables are fully connected and the so-called local susceptibility matrix plays a key role in the system [16]. Finally, this approach is not only different from the replica method but also from the method based on iterations in a message-passing algorithm [17]. Our method gives a clearer picture of the success and the failure of Basis Pursuit as well as provides insight into a more general class of sparse signal recovery algorithms.

1.2 Applications

Besides its deep roots in fundamental theories, compressed sensing has a broad range of applications, from communication technology to business informatics to systems biology. It has already advanced the state of the arts in MRI (magnetic resonance imaging) imaging, genotype-disease relation in GWAS (genome-wide associations studies), and face recognition. Several examples of its applications are listed below.

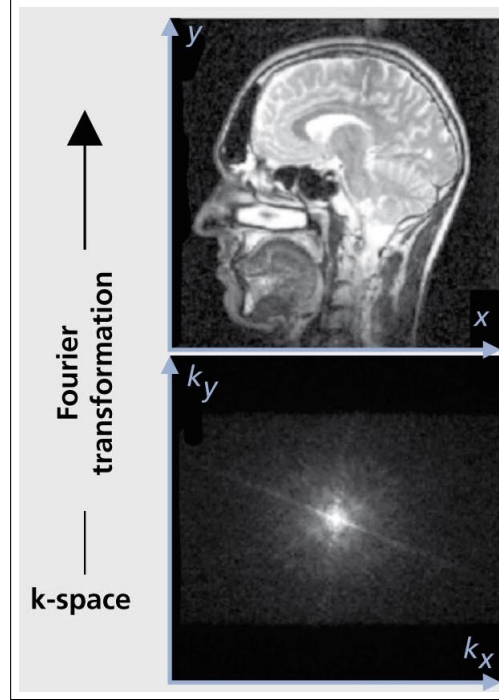


Figure 1.2: A head radiograph is illustrated in the top with its 2-D Fourier spectrum in the bottom. (The graph is reproduced from [21])

1.2.1 Biomedical Imaging

A fundamental assumption in digital image processing is that natural images are piecewise smooth in the pixel basis. That is, there are very few edges in the image, and therefore, the differences between the values of adjacent pixels are usually zero or almost zero. The Fourier transform can be used to map images from the pixel domain to the Fourier domain in which they have sparse (or approximately sparse) representations. For example, Fig. 1.2(top) shows the representation of a natural image in the pixel domain, and Fig. 1.2(bottom) shows the representation of the same image in the Fourier domain. As you can see from the figure, there are very few significant (light) coefficients in the Fourier representation of this image, whereas most Fourier coefficients are almost zero (black). As a result compressed sensing can be used to significantly decrease the number measurements without reducing the accuracy of the MRI image [18–20].

This has an important application, in particular, in dynamic imaging (MRI) that full

signal acquisition is often impossible and only a small number of the so-called K-space measurements can be acquired in a short time [22]. Slow imaging speed in MRI poses challenges for dynamic imaging, in which during the dynamic sequence, the patient may move. Thus, to reduce scan time, the acquisition process is accelerated by under-sampling the K-space (i.e., 2-D Fourier coefficients) [23, 24].

1.2.2 Systems Biology

Following recent advances in technology, the investigation of massive biomedical data with growing scale, diversity, and complexity has taken a center stage in modern data analysis. A more modern example of compressed sensing application idea in biology is the the genome-wide association studies (GWAS). GWAS is one of the most powerful methods for the studies of common diseases and complex traits which have revealed numerous associations between diseases and genetic variants in the last decade [25–28].

Nevertheless, there are still great challenges to intelligently and automatically extract useful information from data and synthesize this knowledge routinely in the clinic. In particular because of the regulatory mechanism in the human genome is complex, it is believed that complex traits are typically caused by multiple genetic variants. Moreover, instead of individual variant at a specific locus, combinations of several variants at different loci can have a significant affect on a phenotype. For instance, an evidence of such higher level multi-variant interactions in complex diseases is seen, such as type-2 diabetes [29]. Identifying such multi-locus or “epistatic” interactions arises as an important problem that has taken a center stage in modern data analysis. Even though the number of genetic variants and their multi-locus interactions often greatly exceeds the sample size, the underlying representations are often sparse [30]. For example, for a certain disease, even though humans have tens of thousands of genes, only a few genes are relevant to the disease; a gene network is sparse in a proper basis since a regulatory pathway involves only a small number of genes. Using methods and ideas from compressed sensing, one can develop a cost-effective

genotyping protocol to detect these variants for severe genetic disorders [31].

1.2.3 Group Testing

Group testing, where groups of individuals are tested for the presence or absence of a disease (or some other binary characteristic), is a common procedure that is used to reduce the costs of screening a large number of individuals [32]. The group testing applications range from the blood testing problem which was used in World War II to test soldiers for syphilis [33], to the problem of testing the impacts of new drugs on human genes [34]. In group testing the purpose is to avoid individual testing of all candidates by repeatedly pooling up a subgroup of multiple individuals and testing this subgroup instead [35]. It is often assumed that there are only a few people sharing some specified property, and the goal is to design an $M \times N$ test matrix describing the M subgroup tests, so that it is possible to efficiently recover the sparse special members of the group from the tests. Therefore, compressed sensing can also be used for efficient and low-cost sensing in the applications of group testing.

1.2.4 Face Recognition

In past decades, face recognition has been studied extensively, in which given a picture, the goal is to predict where his/her face is located at and who he/she is. A key assumption in face classification is that all faces of most human-beings lie in a low dimensional subspace, and much fewer degrees of freedom (compared to the total number of pixels), govern the structure of all possible faces. As a result, given a sufficiently rich training set of faces for a particular person, any new (test) face can be represented by a linear combination of her training faces, and therefore by a sparse linear combination of all training faces of all people in the training repository.

Therefore, sparse approximation can be used to identify the person whose training faces form the largest contribution in approximating the test face [36]. Using compressed sensing

and sparse approximation has provided significant improvements over the existing state-of-the-art methods that use support vector machines [37], or principal component analysis [38] A similar approach has also been used in speaker identification and speech recognition applications [39]

1.3 Contribution of this Thesis

In previous sections, we explained the main essence of the compressed sensing problem and all the benefits that it provides compared with the traditional sampling and data acquisition. We also indicated the existence of a phase transition related to the performance of compressed sensing. This thesis describes our efforts to understand the essential nature of such algorithmic phase transitions and the critical phenomena around them. Although we envisage many practical applications of our results, in this work, our focus remains firmly on theoretical aspects.

The main contributions of this thesis can be classified into the following three categories:

- Obtaining analytical expressions and set of mean-field equations for a generalized least squared optimization problem by employing two-step cavity method at zero temperature and understanding the underlying nature of local susceptibility in such systems
- Using the important role of local susceptibility to rigorously analyze the phase transition boundaries of the Basis Pursuit and the critical behavior and scaling laws near the phase transition boundary
- Obtaining new and different phase boundary for new class of sparse reconstruction algorithm called Elastic Net and showing that the critical behavior of this phase boundary and the one for Basis Pursuit are in different universality classes
- Generalizing the asymptotic relation for the phase boundary, in the extremely sparse limit, to the case the correlated measurement matrix

In the following chapters, we outline our reasons to go beyond the conventional replica approach. In particular, we emphasize the inherent role of local susceptibility as an important measure for the robustness of the signal recovery and indicator of the existence of different phases in sparse reconstruction algorithms. In the end, we provide a simple picture of the compressed sensing phase transition that could be made explicit in the sparse limit.

1.4 Outline of the Thesis

In this section, we briefly overview the structure of the thesis by explaining each chapter separately.

Chapter 1

In this chapter, we concentrate on defining the compressed sensing problem carefully, and set the stage for its answers in later chapters.

Chapter 2

In this chapter, we first briefly review a finite noise/finite temperature formulation of the problem, and a recapitulation of replica approach to its solution, presented previously in the literature. We do this for a generalized regularization or penalty function. Then we show that the proposed method is equivalent to a cavity method at finite temperature. This allows us to intuitively understand the role of two steps in the cavity method that we will employ in the next chapter.

Chapter 3

In this chapter, we introduce a susceptibility matrix associated with this problem and explore its structure. Then, we derive the self-consistent mean-field equations for the mean square error of estimation via a two-step cavity method at zero temperature. We will emphasize the role of local susceptibility in the system that will turn out to carry important information about the underlying structure and performance-guarantees of any sparse reconstruction. This approach is not only different from the replica method but also from the method based on iterations in a message-passing algorithm.

Chapter 4

In this chapter, we exploit local susceptibility obtained in the previous chapter to treat the simple case of Ridge Regression and then to find a simple way to arrive at the two phases and the phase boundary known for the Basis Pursuit. We show that this transition is actually continuous (second order) and find a variety of critical behaviors, including scaling laws and critical exponents that are uniquely determined by the universality class of the phase transition. In the extremely sparse limit, we also generalize the asymptotic limit of these mean-field equations to the case of correlated measurement matrices. In the end, we obtain new results for the Elastic-Net in the high-dimensional limit from a completely different perspective.

Chapter 5

In this chapter, we end by pointing out potential arenas of application where our two-step cavity approach at zero temperature is more natural than conventional replica theory.

Chapter 2

An Overview of Compressed Sensing

In previous sections, we explained the main essence of the compressed sensing problem and all the benefits that it provides compared with the traditional sampling.

Compressed sensing is a simple and efficient signal acquisition technique that collects a few measurements about the signal of interest and later uses optimization techniques for reconstructing the original signal from what appears to be an incomplete set of measurements [1, 2].

The CS technique relies on two fundamental principals: 1) sparse representation of the signal of interest in some basis and 2) incoherence between the measurement matrix and the signal representation. We will define these terms in the next section.

2.1 Problem Formulation

The following subsections present the formal definitions of some terminologies that will be used in this chapter.

2.1.1 Compressed Sensing

In this subsection we present a formal description of “compressed sensing”. Sensing of signal \mathbf{s} is defined as the process of collecting some measurements about \mathbf{s} by correlating \mathbf{s} with some sensing matrix Φ , i.e.,

$$\mathbf{y} = \Phi \mathbf{s} \tag{2.1}$$

where the sensing matrix $\Phi \in \Re^{M \times N}$, and \mathbf{s} in \Re^N . Based on this model, compressed sensing is defined as the sensing process for which the number M of available measurements is much smaller than the dimension N of the signal \mathbf{s} . Therefore, compressed sensing is formulated as an underdetermined linear systems of equations. The problem associated with it is that we have either no solution, if \mathbf{y} is not in the span of the columns of the matrix Φ , or infinitely many solutions. To take care of former problem, we must assume that the measurement matrix Φ is full-rank. And to obtain a single solution, it is necessary to impose additional criteria on the candidate solution to identify which of these candidate solutions is the desired one.

A powerful constraint that can be used in this regard is the “sparsity” of the solution vector. A vector is called K -sparse if it has at most K non-zero entries. As stated before, an underdetermined system of linear equations has infinite candidate solutions of the form $\hat{\mathbf{s}} = \mathbf{s} + \mathbf{z}$ where \mathbf{s} is any vector that satisfies relation $\mathbf{y} = \Phi\mathbf{s}$ with $\mathbf{z} \in \mathcal{N}(\Phi)$ is a vector in the the null space of Φ . As we will see later, if the candidate solution vector is known to be K -sparse under certain conditions on the sensing matrix Φ , the solution vector can be determined uniquely using an optimization technique. This is also applies to non-sparse vectors that can be sparsely represented in a suitably selected basis $\Psi \in \Re^{N \times N}$, i.e.

$$\mathbf{s} = \Psi\mathbf{x} \tag{2.2}$$

where the coefficient vector \mathbf{x} is sparse. Clearly, \mathbf{x} and \mathbf{s} are equivalent representations of the signal, with \mathbf{x} in the time or space domain and \mathbf{x} in the Ψ domain. Combining Eqs. (2.1), (2.2) and taking into consideration the case of noisy measurements, the sensing process can be written as

Definition 1 (Measurement Process)

$$\mathbf{y} = \Phi\Psi\mathbf{x} + \zeta = \mathbf{H}\mathbf{x} + \zeta \quad (2.3)$$

$\mathbf{H} \in \Re^{M \times N}$, \mathbf{x} and $\zeta \in \Re^N$.

Assuming that the coefficient vector \mathbf{x} is K -sparse, then \mathbf{x} , and hence $\mathbf{s} = \Psi\mathbf{x}$, can only be estimated from \mathbf{y} if the matrices \mathbf{H} satisfies the properties described in the next subsection.

2.1.2 Restricted Isometry Properties

The sparsity of the solution vector, or its representation in some basis, is a necessary but not sufficient condition for finding a unique solution to an underdetermined system of linear equations. One of the fundamental problems in CS is to investigate whether a given matrix \mathbf{H} is a “good” compressed sensing matrix in the sense that it can guarantee the vector \mathbf{x} can be exactly recovered from \mathbf{y} under the condition $M \ll N$. To this end, in addition to the sparsity principle, CS relies on another key property of the matrix \mathbf{H} called Restricted Isometry Property (RIP). RIP is a powerful measure that was introduced by Candès and Tao [40] and has proved to be very useful in studying the general robustness of CS.

Definition 2 (Restricted Isometry Property (RIP)) *We say that an $M \times N$ matrix \mathbf{H} obeys the RIP with parameters K and δ if*

$$(1 - \delta_K)\|\mathbf{x}\|_2^2 \leq \|\mathbf{H}\mathbf{x}\|_2^2 \leq (1 + \delta_K)\|\mathbf{x}\|_2^2 \quad (2.4)$$

holds true for any choice of K -sparse vectors \mathbf{x} .

The key idea in the above definition is the claim that any subset of K columns from \mathbf{H} behave like an orthogonal transform (the columns of \mathbf{H} cannot be exactly orthogonal since we have more columns than rows) that approximately preserves the Euclidean length of K -sparse signals. Therefore, \mathbf{H} obeys the RIP of order K if δ_K is not too close to 1.

When the RIP holds, it is known that certain convex optimization programs, in particular, ℓ_1 -norm minimization accurately recover all signals \mathbf{x} with at most K nonzero elements. However, testing for the RIP is generally a hard combinatorial problem and the only deterministic measurement matrices which are known to satisfy the RIP do so only under extremely strong conditions.

To this end, there are some known random measurement matrices that satisfy the RIP (with high probability), i.e. matrices with column vectors taken from arbitrary subsets being nearly orthogonal. One well-known case is when \mathbf{H} is formed by sampling i.i.d. entries from the normal distribution with mean zero and variance $1/M$, or other sub-Gaussian distribution [41].

Remark 1 (RIP Bound) *Random measurement matrices, with high probability, obey the restricted isometry property provided that*

$$M \geq CK \log(N/K) \tag{2.5}$$

where C is some constant depending on each instance [1].

To reconstruct the original signal from the M measurements in the vector \mathbf{x} , an optimization technique must be used, and this is the topic of the next section.

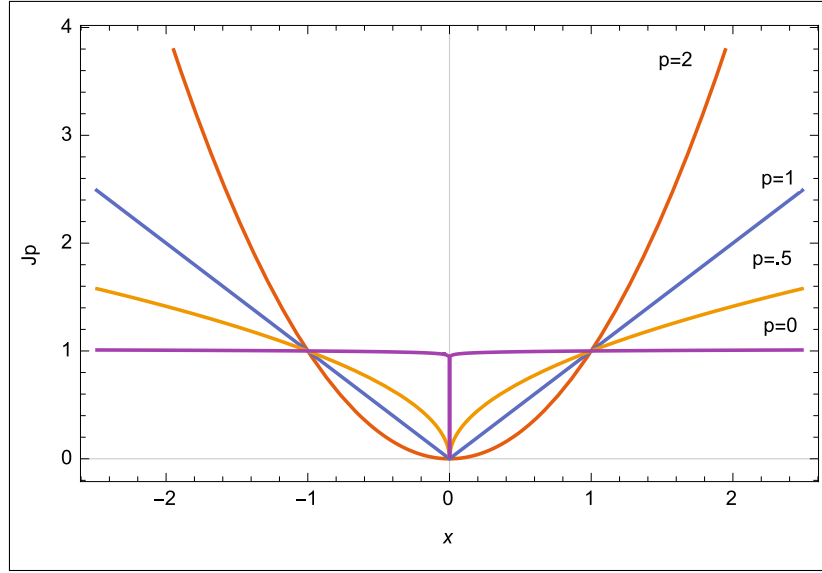


Figure 2.1: A plot of $J_p(x)$ for some values of $0 \leq p \leq 2$.

2.1.3 Signal Reconstruction Algorithms

The goal of signal reconstruction algorithm is to obtain an estimate of the original signal \mathbf{x}_0 given only observation vector \mathbf{y} and measurement matrix \mathbf{H} . Or, equivalently solving the following linear underdetermined system of equations

$$\mathbf{y} = \mathbf{H}\mathbf{x}_0 \quad (2.6)$$

where, once more, \mathbf{H} in $\Re^{M \times N}$, $\mathbf{x}_0 \in \Re^N$ is a k -sparse vector, and $M < N$. Like we mentioned before, there are infinitely many solutions to Eq. (2.6) of the form $\hat{\mathbf{x}} = \mathbf{x}_0 + \mathbf{z}$, where $\hat{\mathbf{x}}$ is any vector that satisfies Eq. (2.6).

Since the original vector \mathbf{x}_0 is sparse, a familiar way to do this is via regularization [42] where a objective function $J(\mathbf{x})$ which is an appropriate measure of sparsity, i.e. evaluates the desirability of the solution \mathbf{x} , is to maximize (minimize) while simultaneously satisfying the constraints defined by Eq. (2.6), respectively.

Definition 3 (Regularized Penalty Optimization Problem) *Defining the general optimization problem (P_J) , this can be expressed mathematically as*

$$(P_J) : \quad \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} J(\mathbf{x}) \quad \text{subject to} \quad \mathbf{y} = \mathbf{H}\mathbf{x} \quad (2.7)$$

where minimization of the objective function $J(\cdot)$ encourages sparsity in the solution. We consider special interest for $J(\cdot)$ to be a convex function of the form (See in Fig. 2.1.3):

$$J_p(\mathbf{x}) = \|\mathbf{x}\|_p^p = \sum_a |x_a|^p \quad (2.8)$$

In particular, we will have a special interest when $p = 1$ or the so called ℓ_1 -norm due to its tendency to promote sparsity of the solution. In the remaining of this chapter, we briefly discuss issues relating to solving

2.1.4 ℓ_2 -norm Minimization

The well known choice of $J(\mathbf{x})$ is the ℓ_2 -norm given by the squared Euclidean norm $\|\mathbf{x}\|_2^2$. This objective function has a closed form solution and can be calculated by introducing the Lagrangian

$$\mathbf{L}(\mathbf{x}) = \lambda^T (\mathbf{H}\mathbf{x} - \mathbf{y}) + \|\mathbf{x}\|_2^2 \quad (2.9)$$

with λ being the Lagrange multipliers for the constraint. Taking a derivative of $\mathbf{L}(\mathbf{x})$ with respect to \mathbf{x} , we get

$$\frac{\partial \mathbf{L}(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{H}^T \lambda + 2\mathbf{x} \quad (2.10)$$

and thus, the solution is given as

$$\hat{\mathbf{x}} = -\frac{1}{2} \mathbf{H}^T \lambda \quad (2.11)$$

plugging this solution into $\mathbf{y} = \mathbf{H}\mathbf{x}$ results in

$$\mathbf{H}\hat{\mathbf{x}} = -\frac{1}{2}\mathbf{H}\mathbf{H}^T\lambda = \mathbf{y} \Rightarrow \lambda = -2(\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{y} \quad (2.12)$$

which eventually gives the well-known unique pseudo-inverse solution

$$\hat{\mathbf{x}} = -\frac{1}{2}\mathbf{H}^T\lambda = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{y} \quad (2.13)$$

The fact that it gives a unique solution can also be seen from the Fig. (2.1.3) due the strictly convexity of the objective function. This feature and its mathematical simplicity make the use of ℓ_2 extensive in various fields of science and engineering. However, since ℓ_2 -norm penalty imposed by $J(x_a)$ on small nonzero coefficients of the solution vector is small, even though it shrinks the estimates of the regression coefficients towards zero, tends to result in all small but many non-zero (non-sparse) regression coefficient. This is illustrated in Fig. 2.1.5. Therefore, ℓ_2 minimization is not appropriate for finding a K -sparse solution.

2.1.5 ℓ_0 -norm Minimization

Since the ℓ_2 -norm does not lead to sparsity of the signal, one can consider the ℓ_0 -norm that counts the number of non-zero entries in \mathbf{x} . The optimization problem of Eq. (2.7) in the case of $J(\mathbf{x}) = J_0(\mathbf{x}) = \|\mathbf{x}\|_0$ becomes

Definition 4 (ℓ_0 -norm Minimization)

$$(P_0) \quad \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{y} = \mathbf{H}\mathbf{x} \quad (2.14)$$

This optimization looks like the minimum ℓ_2 -norm problem (P_2), but referring to Fig.

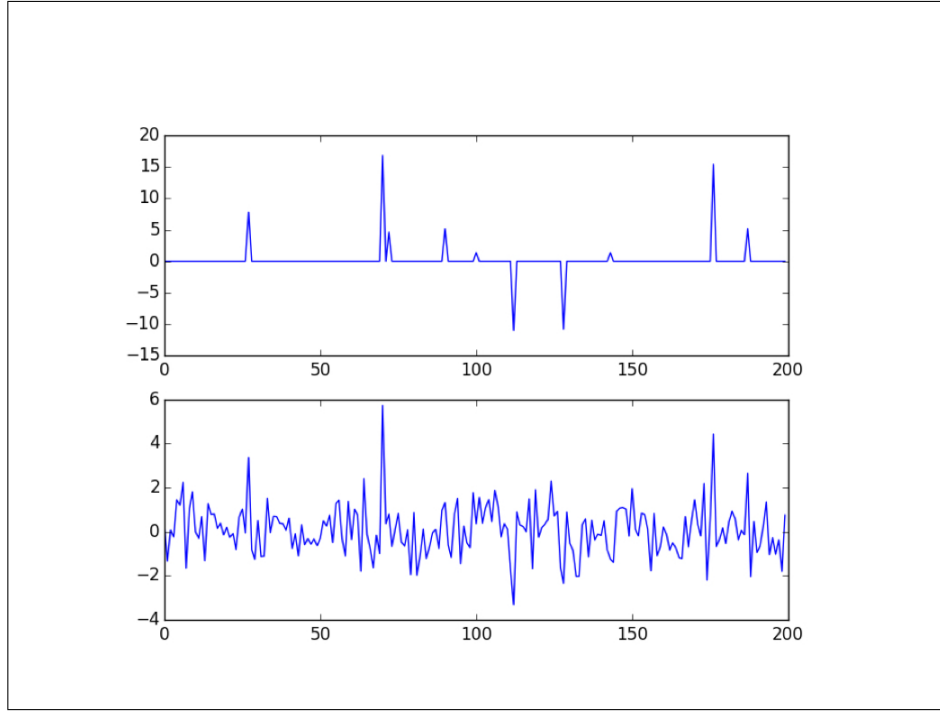


Figure 2.2: ℓ_2 penalty shrinks the estimates of the regression coefficients towards zero, however, result in all small but many non-zero regression coefficient. The top panel: x-axis is the original signal with the size $N = 100$ and y-axis is the $K = 10$ non-zero coordinates. The bottom panel: reconstruction via ℓ_2 -norm minimization method with $M = 60$ measurements. The x-axis is the estimated signal and y-axis is the estimated non-zero coordinates.

2.1.3, we observe that $J_0(\mathbf{x})$ is flat over all values of x except at $x = 0$, which implies that any gradient descent technique will fail to converge to the sparse solution. One can look at this problem as a combinatorial search problem equivalent to selecting K vectors of the measuring matrix \mathbf{H} that best represents the measurement vector \mathbf{y} , the solution vector to (P_0) can be obtained by searching over the $\binom{N}{K}$ possible ways in which the basis sets can be chosen to find the best solution. But, the complexity of exhaustive search is exponential in N and it has been proven that (P_0) is, in general, NP-Hard. In addition, it is shown that (P_0) yields a solution which is not robust to the noise.

These difficulties pose many challenges that have restricted its studies and applications. On the other hand, these limitations have become a motivation for researchers to find other means to solve this problem efficiently. One way would be to replace $J_0(\mathbf{x})$ by other functions

that are robust to noise and can be solved efficiently (such as $J_2(\mathbf{x})$), but nevertheless offer sparse solutions (such as $J_0(\mathbf{x})$). In the framework of compressed sensing, there have been computationally efficient relaxation algorithms for this computationally NP-Hard problem.

2.1.6 ℓ_1 -norm Minimization

One major approach, Basis Pursuit, relaxes the ℓ_0 -norm minimization problem to an ℓ_1 -norm minimization problem with the choice of $J(\mathbf{x}) = \|\mathbf{x}\|_1$ [4]:

Definition 5 (Basis Pursuit)

$$(P_1) \quad \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{H}\mathbf{x} \quad (2.15)$$

Since $p = 1$ is the smallest value of p for which $J_1(\mathbf{x})$ is convex, ℓ_1 -minimization and the equivalence between the solutions of (P_1) and (P_0) has been studied extensively in the context of sparse solutions for many years [2, 43–45]. One key observation is that the ℓ_1 -norm has a tendency to prefer sparse solutions.

This naively can be seen by the fact that the penalty imposed by ℓ_1 -norm on values of $0 \leq |x| < 1$ is greater than the least squares solution imposed by ℓ_2 -norm (See Fig. 2.1.6).

It was with the LASSO algorithm [46] proposed as a method in statistics for sparse model selection, the application areas for ℓ_1 -norm minimization began to broaden. Basis Pursuit [4] was proposed in computational harmonic analysis for extracting a sparse signal representation from highly over-complete dictionaries. It then came as a breakthrough in a remarkable result by Candés and Tao [1] and separately Donoho [2] that (P_1) method was shown to be able to recover sparse signals with a linear fraction of non-zero elements. Certainly this requires some conditions on the measurement matrix \mathbf{H} .

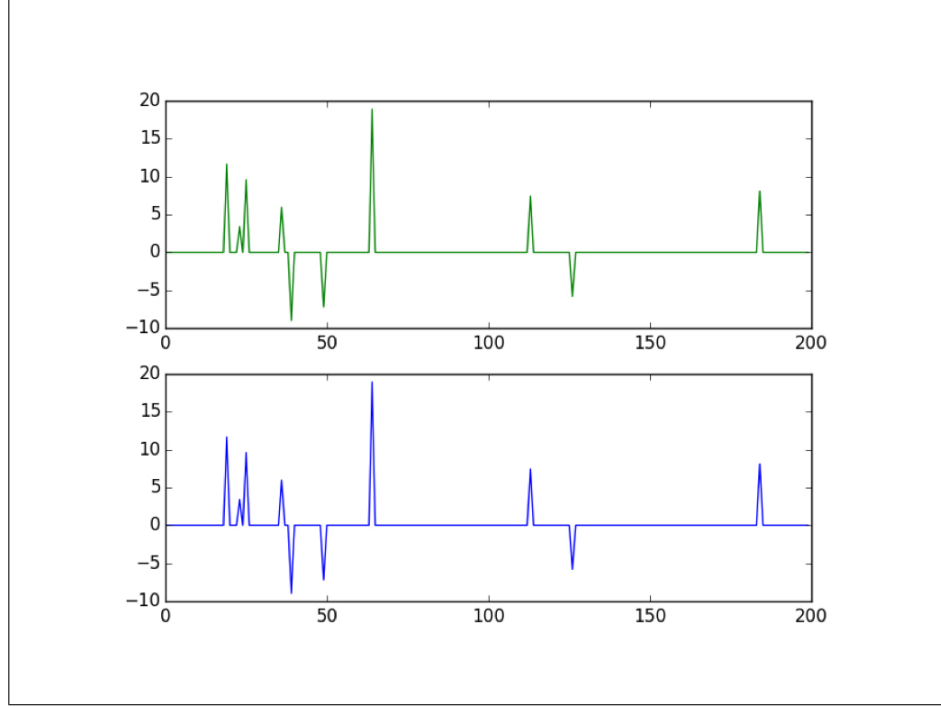


Figure 2.3: ℓ_1 penalty shrinks the estimates of the regression coefficients towards zero with only few non-zero regression coefficient. The top panel: x-axis is the original signal with the size $N = 100$ and y-axis is the $K = 10$ non-zero coordinates. The bottom panel: shows the perfect reconstruction of the original signal via ℓ_1 -norm minimization method with $M = 60$ measurements. The x-axis is the estimated signal and y-axis is the estimated non-zero coordinates.

For example, the restricted isometry property (RIP) conditions were given in [1, 40] to guarantee that ℓ_1 -norm minimization accurately recovers sparse or compressible signals. It is now known that many kinds of matrices satisfy these conditions with the number of measurements $M = CK \log(N/K)$ in (2.5) for some constant C , which depends on the desired probability of success. In any case, C tends to one as $N \rightarrow \infty$. Therefore, the cost of replacing (P_0) by (P_1) is that more measurements are required, depending logarithmically on N . Sharp reconstruction thresholds have been computed by Donoho and Tanner [47] so that for any choice of sparsity K and signal size N , the required number of measurements

M for (P_1) to recover \mathbf{x}_0 with high probability can be determined precisely. Their results replace $\log(N/K)$ with $\log(N/M)$, i.e. $M \geq CK \log(N/M)$. However, M appears in both sides of this inequality, and this can be adjusted to compute a threshold of the sparsity $K \leq MC \log(N/M)$ for a given number of measurements M .

2.1.7 Conversion of (P_1) to Linear Programming (L_P)

As we can see from Eq. (2.15), the objective function is non-differentiable when $x_a = 0$ for any x_a . However, by doubling the number of variables as it will be shown, the non-differentiable objective function can be converted into a set of linear functions, and thus being able to find its intersection with the feasible region of a polyhedron [46]. This approach supposes that in (P_1) , the unknown \mathbf{x} can be represented as the difference between two non-negative variables corresponding to the positive and negative components of \mathbf{x} :

$$\mathbf{x} = \mathbf{x}^+ - \mathbf{x}^- \quad (2.16)$$

where $\mathbf{x}^+, \mathbf{x}^- \in \mathfrak{R}^N$ are both non-negative vectors. If \mathbf{x} is positive, then $\mathbf{x}^+ = \mathbf{x}$ and $\mathbf{x}^- = 0$, if \mathbf{x} is negative, then $\mathbf{x}^- = \mathbf{x}$ and $\mathbf{x}^+ = 0$, and if $\mathbf{x} = 0$ then $\mathbf{x}^+ = 0$, $\mathbf{x}^- = 0$. It is easy to see that by denoting $\mathbf{w} = [\mathbf{x}^+, \mathbf{x}^-]$, $\|\mathbf{x}\|_1 = \mathbf{x}^+ + \mathbf{x}^- = \mathbf{w}$ and $\mathbf{H}\mathbf{x} = \mathbf{H}[\mathbf{x}^+ - \mathbf{x}^-] = [\mathbf{H}, -\mathbf{H}]\mathbf{w}$. We can use this representation to create an equivalent form of Eq. (2.15)

Definition 6 (Linearized Basis Pursuit Optimization) *Linearized Basis Pursuit has $2N$ variables with $2N$ constraints enforcing non-negativity and one constraint enforcing that the positive and negative components satisfy the original constraint.*

$$\arg \min_{\mathbf{w}} \mathbf{w}^T \quad \text{subject to } \mathbf{y} = [\mathbf{H}, -\mathbf{H}]\mathbf{w} \quad \text{and } \mathbf{w} \geq 0 \quad (2.17)$$

The equivalence between P_1 and L_P makes the optimization of Basis Pursuit much more

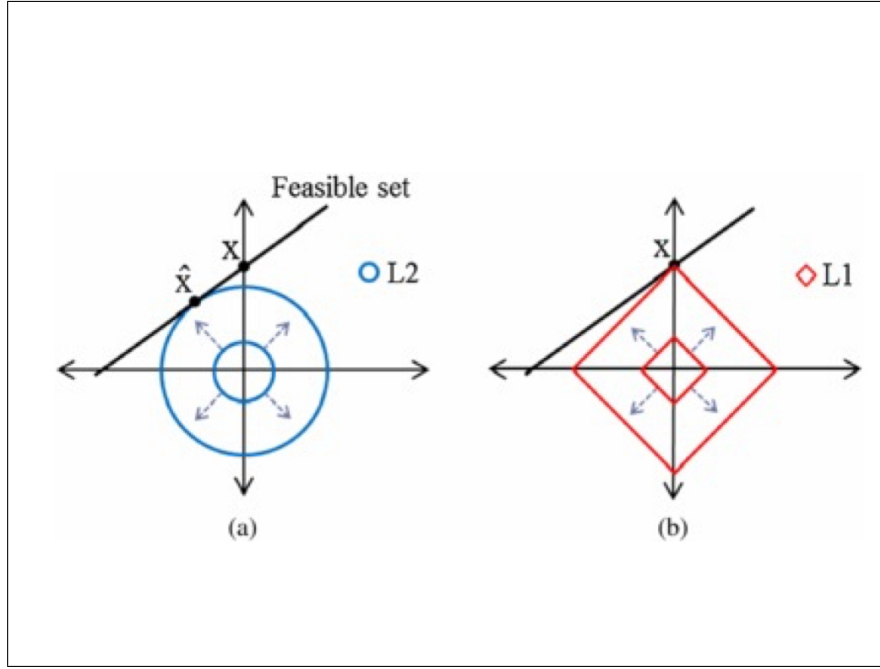


Figure 2.4: A schematic illustration of how different regularization terms lead to sparse and non-sparse solutions. The x-axis and y-axis represent the dimension of the signal (here $N=2$). The feasible set is equivalent to all the \mathbf{x} that is the solution of $\mathbf{y} = \mathbf{H}\mathbf{x}$ a) ℓ_1 regularization corresponds to the diamond shaped ball centered around the origin. b) ℓ_2 regularization corresponds to the spherical ball centered around the origin.

efficient for solving large-scale problems by applying linear programming method such as Simplex algorithm [48].

2.1.8 Geometric Interpretation

In this section we present a geometric interpretation of the performance of the previously discussed objective functions, e.g. ℓ_p -norm where $0 < p \leq 2$, in estimating sparse solutions. This geometric interpretation helps visualize why ℓ_2 -norm reconstruction fails to find the sparse solution that can be identified by ℓ_1 -norm.

For the sake of illustration, consider the simple 2-D example in Fig. 2.1.8. The coordinate axes in this figure are x_1, x_2 . In this figure, the exact and the estimated solution vectors are represented by \mathbf{x} and $\hat{\mathbf{x}}$, respectively. The linear set of equations forming the constraint $\mathbf{H}\hat{\mathbf{x}} = \mathbf{H}\mathbf{x}$ define a feasible set of solutions that are on an affine subspace $\in R^2$.

Geometrically speaking, solving (P_p) is done by “blowing up” the ℓ_p ball centered around the origin, and stopping its inflation when it first touches the feasible set. As it is represented by the circle in Fig. 2.1.8(a), the ℓ_2 minimizer is the point on the feasible set closest to the origin. Due to the randomness of the entries of the sensing matrix \mathbf{H} , the line representing feasible set is oriented at random angle. Therefore, with high probability, the closest point $\hat{\mathbf{x}}$ will be away from the coordinate axes and thus will be neither sparse nor close to the correct answer \mathbf{x} . In contrast, the ℓ_1 ball in Fig. 2.1.8(b) has points aligned with the coordinate axes. Therefore, depending on the orientation of the line, there are two possible cases. In the first case, when the ℓ_1 ball is blown up, it will contact the feasible line precisely at where the sparse vector \mathbf{x} is located as shown in Fig. 2.1.8(b). This is the tendency to sparsity we are referring to as opposed to ℓ_2 . However, in a highly unlikely, it is possible that the angle of the affine subspace will be such that it contacts the line $\mathbf{H}\mathbf{x} = \mathbf{H}\mathbf{x}_0$ at a point far from the exact solution vector.

Put more generally, the ℓ_p ball is “curved outward” for all $p > 1$, and expected that an intersection of an affine subspace and an ℓ_p ball to not take place on the axes. On the other hand, ℓ_p for $p = 1$ has a “diamond” shape and for ℓ_q -norm, where $0 < q < 1$, this diamond is “curved inward”. Therefore, it is expected that an intersection of an affine subspace and an the ℓ_p ball occurs on the axes and thus lead to a sparse solution.

Chapter 3

Statistical Sparse Reconstruction Methods

In this chapter, we employ techniques from the statistical physics of disordered systems to compute the typical behavior of CS as a function of the signal sparsity and measurement density. We first briefly review a finite noise/finite temperature formulation of the problem and a replica approach to its solution, presented previously in the literature [8, 9]. We do this for a generalized regularization or penalty function V . Then, we take an alternative approach and use two-step cavity method [16] to obtain the self-consistent mean field equations for the mean square error of estimation that we obtained via replica method. This will give a better insight into the problem of cavity method at zero temperature we will obtain in the next chapter.

3.1 Formulation of the Regularized Least-Squares

Here, we set up the general framework for investigating the regularized least-squares (LS) based reconstruction algorithms. We assume that the data $\mathbf{y} = \mathbf{H}\mathbf{x}_0 + \boldsymbol{\zeta}$ are generated by a probability distribution $p(\mathbf{y}|\mathbf{x}_0, H)$, given an (unknown) sparse signal \mathbf{x}_0 and a (known) matrix \mathbf{H} , and an (unknown) Gaussian noise vector $\boldsymbol{\zeta}$ whose components are i.i.d. samples from $\mathcal{N}(0, \sigma_\zeta^2)$. The vector \mathbf{x}_0 is considered to be a random sample from a distribution $P_0(\mathbf{x}_0) = \prod_a p_0(x_{a0})$.

Although the probability distribution of \mathbf{H} , in general, $\mathcal{P}(\mathbf{H})$ could be a non-Gaussian distribution, at this point we consider it to be Gaussian with

Definition 7 (Mean and Variance of $P(\mathbf{H})$)

$$[H_{ia}]^{\text{av}} = 0 \quad (3.1)$$

$$[H_{ia}H_{jb}]^{\text{av}} = \frac{1}{M}\delta_{ij}\delta_{ab} \quad (3.2)$$

We study the performance of an estimator of \mathbf{x}_0 , namely the location $\hat{\mathbf{x}}$ of the value of the minimum of a cost function

$$\mathcal{E}_0(\mathbf{x}) = \frac{(\mathbf{y} - \mathbf{H}\mathbf{x})^2}{2\sigma^2} + V(\mathbf{x}). \quad (3.3)$$

As we reformulate this exercise in estimation as a statistical mechanics problem, the cost function will play the role of energy. We assume the penalty/potential term $V(\mathbf{x})$ is such that there is a unique minimum of \mathcal{E}_0 . Note that $\hat{\mathbf{x}} = \arg \min_x \mathcal{E}_0(\mathbf{x})$ depends on \mathbf{y}, \mathbf{H} , meaning that it can be written as a function $\hat{\mathbf{x}} = \mathbf{g}(\mathbf{x}_0, \mathbf{H}, \boldsymbol{\zeta})$, using the fact that $\mathbf{y} = \mathbf{H}\mathbf{x}_0 + \boldsymbol{\zeta}$. Since we set up an ensemble of problem instances by specifying the probability distribution of the variables $\mathbf{x}_0, \mathbf{H}, \boldsymbol{\zeta}$, we could study the performance of the estimator over this distribution. For example, we could study the distribution of the estimation error $\hat{\mathbf{x}} - \mathbf{x}_0$ for this problem instance ensemble.

In order to make a connection between the optimizations problem and statistical mechanics, one could choose a probability distribution of \mathbf{x} parametrized by β , playing the role of inverse temperature,

$$\begin{aligned} p_\beta(\mathbf{x}|\mathbf{y}, \mathbf{H}) &= \frac{1}{Z} \exp(-\beta \mathcal{E}_0(\mathbf{x})) \\ &= \frac{1}{Z(\beta, \mathbf{y}, \mathbf{H})} \exp \left\{ -\beta \left(\frac{(\mathbf{y} - \mathbf{H}\mathbf{x})^2}{2\sigma^2} + V(\mathbf{x}) \right) \right\} \end{aligned} \quad (3.4)$$

with normalization factor $Z = Z(\beta, \mathbf{y}, \mathbf{H})$, namely the partition function, being given by

$$\begin{aligned} Z(\beta, \mathbf{y}, \mathbf{H}) &= \int d^N \mathbf{x} \frac{1}{Z} \exp(-\beta \mathcal{E}_0(\mathbf{x})) \\ &= \int d^N \mathbf{x} \exp \left\{ -\beta \left(\frac{(\mathbf{y} - \mathbf{H}\mathbf{x})^2}{2\sigma^2} + V(\mathbf{x}) \right) \right\}. \end{aligned} \quad (3.5)$$

If we send β to ∞ , equivalent to sending the temperature to zero, the probability gets concentrated at the minimum of the cost/energy function. Keep in mind that we define β to be dimensionless.

We will consider averages of functions of the form $f(\mathbf{x}, \mathbf{x}_0)$ containing both the original sparse signal and the variable related to the estimate. For example, we are interested in the estimation error, which can be quantified by a suitable norm placed on the difference between the original and the estimated signal. The average of the function $f(\mathbf{x}, \mathbf{x}_0)$ is given by

$$\langle f(\mathbf{x}, \mathbf{x}_0) \rangle = \frac{\int d^N \mathbf{x} f(\mathbf{x}, \mathbf{x}_0) \exp\{-\beta \frac{(\mathbf{y} - \mathbf{H}\mathbf{x})^2}{2\sigma^2} - \beta V(\mathbf{x})\}}{\int d^N \mathbf{x} \exp\{-\beta \frac{(\mathbf{y} - \mathbf{H}\mathbf{x})^2}{2\sigma^2} - \beta V(\mathbf{x})\}}. \quad (3.6)$$

This ‘thermal’ average, represented by $\langle \dots \rangle$, depends on the random variables \mathbf{x}_0 , \mathbf{H} and $\boldsymbol{\zeta}$. Note that in the limit $\beta \rightarrow \infty$, this average should become $f(\hat{\mathbf{x}}, \mathbf{x}_0)$, for continuous f . Averaging the result of this calculation over the random instances of \mathbf{x}_0 , \mathbf{H} and $\boldsymbol{\zeta}$ is a technical challenge related to quenched averages in disordered systems.

The function $f(\mathbf{x}, \mathbf{x}_0) = \frac{1}{N}(\mathbf{x} - \mathbf{x}_0)^2$ plays an important role in our analysis. Its average corresponds to the mean squared estimation error.

Definition 8 (Mean Squared Error (MSE))

$$\text{MSE} \equiv \frac{1}{N} \sum_{a=1}^N [\langle x_a - x_{a0} \rangle^2]_{\mathbf{x}_0, \mathbf{H}, \boldsymbol{\zeta}}^{\text{av}} = \frac{1}{N} [\langle (\mathbf{x} - \mathbf{x}_0)^2 \rangle]_{\mathbf{x}_0, \mathbf{H}, \boldsymbol{\zeta}}^{\text{av}} \quad (3.7)$$

We will use $[\dots]_{\text{vars}}^{\text{av}}$ to denote quenched averages, with the relevant quenched variables

indicated in the subscript, when necessary. We use the notation $\mathbf{u} = \mathbf{x} - \mathbf{x}_0$ to indicate the estimation error vector. The size of the vector \mathbf{u} provides a measure of the inaccuracy of the reconstruction.

In the context of penalized regression, the penalty function is often chosen to be a sum of potentials involving single variables, namely, $V(\mathbf{x}) = \sum_a U(x_a)$. We will focus on $V(\mathbf{x})$ of this nature. An important special case for example is in compressed sensing with sparsity promoting regularizing potential $U(x) = \lambda|x|$.

For $\zeta = 0$, we will be interested in the result of the constrained optimization problem of minimizing $V(\mathbf{x})$ subject to the constraint $\mathbf{y} = \mathbf{H}\mathbf{x}$. In the $M, N \rightarrow \infty$ limit, this problem may exhibit a phase transition from a perfect reconstruction phase to an error-prone phase, with the MSE, mentioned above, as the order parameter. This constrained optimization could be studied in more than one equivalent ways. After taking $\beta \rightarrow \infty$ limit, we will take the route $\sigma \rightarrow 0$ to enforce the equality $\mathbf{y} = \mathbf{H}\mathbf{x}$.

3.2 Replica Approach

In this section, we review the replica approach to the problem [8, 9], presenting the mean field equations in terms of a distribution of asymptotically independent single-variable problems with a set of self-consistency conditions. In order to calculate quantities like the MSE, we need to compute quenched averages of the form $[\langle f(\mathbf{x}, \mathbf{x}_0) \rangle]^{\text{av}}$, which is complicated by the presence of the denominator in Eq. (3.6). Formally, the denominator is handled by introducing n non-interacting replicas of the system and taking $n \rightarrow 0$, as shown below. In the noiseless case, $\mathcal{E}_0(\mathbf{x})$ depends on \mathbf{x} as well as on \mathbf{x}_0, H . To emphasize those additional

dependences, we write $\mathcal{E}_0(\mathbf{x})$ as $\mathcal{E}_0(\mathbf{x}_\mu, \mathbf{x}_0, \mathbf{H})$ in the next few equations.

$$\begin{aligned}
\langle f(\mathbf{x}, \mathbf{x}_0) \rangle_{\mathbf{x}} &= \frac{\int d^N \mathbf{x} f(\mathbf{x}, \mathbf{x}_0) \exp(-\beta \mathcal{E}_0(\mathbf{x}, \mathbf{x}_0, \mathbf{H}))}{\int d^N \mathbf{x} \exp(-\beta \mathcal{E}_0(\mathbf{x}, \mathbf{x}_0, \mathbf{H}))} \\
&= \lim_{n \rightarrow 0} \left(\int d^N \mathbf{x} \exp(-\beta \mathcal{E}_0(\mathbf{x}, \mathbf{x}_0, \mathbf{H})) \right)^{n-1} \\
&\quad \int d^N \mathbf{x} f(\mathbf{x}, \mathbf{x}_0) \exp(-\beta \mathcal{E}_0(\mathbf{x}, \mathbf{x}_0, \mathbf{H})) \\
&= \lim_{n \rightarrow 0} \int f(\mathbf{x}_1, \mathbf{x}_0) \prod_{\mu=1}^n \{d^N \mathbf{x}_\mu \exp(-\beta \mathcal{E}_0(\mathbf{x}_\mu, \mathbf{x}_0, \mathbf{H}))\}. \tag{3.8}
\end{aligned}$$

Averaging over the quenched variables \mathbf{x}_0 and H , we get

$$\begin{aligned}
&[\langle f(\mathbf{x}, \mathbf{x}_0) \rangle_{\mathbf{x}}]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}} \\
&= \lim_{n \rightarrow 0} \left[\int \prod_{\mu=1}^n \{d^N \mathbf{x}_\mu\} f(\mathbf{x}_1, \mathbf{x}_0) \exp(-\beta \sum_{\mu} \mathcal{E}_0(\mathbf{x}_\mu, \mathbf{x}_0, \mathbf{H})) \right]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}} \tag{3.9}
\end{aligned}$$

Using $\mathbf{y} = \mathbf{H}\mathbf{x}_0$ in the noiseless case, the energy function for the n -th replica would be

$$\begin{aligned}
\mathcal{E}_0(\mathbf{x}_\mu, \mathbf{x}_0, \mathbf{H}) &= \frac{(\mathbf{y} - \mathbf{H}\mathbf{x}_\mu)^2}{2\sigma^2} + V(\mathbf{x}_\mu) \\
&= \frac{(\mathbf{H}\mathbf{x}_0 - \mathbf{H}\mathbf{x}_\mu)^2}{2\sigma^2} + V(\mathbf{x}_\mu) = \frac{\mathbf{H}\mathbf{u}_\mu^2}{2\sigma^2} + V(\mathbf{u}_\mu + \mathbf{x}_0), \tag{3.10}
\end{aligned}$$

rewritten in terms of the error variables $\mathbf{u}_\mu = \mathbf{x}_\mu - \mathbf{x}_0, \mu = 1, \dots, n$. Thus, we are interested in average quantities in the replicated ensemble whose partition functions is given by

$$\begin{aligned}
&[Z^n]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}} \\
&= \left[\int \prod_{\mu=1}^n d\mathbf{u}_\mu \exp \left[-\beta \left\{ \sum_{\mu=1}^n \frac{(\mathbf{H}\mathbf{u}_\mu)^2}{2\sigma^2} + V(\mathbf{u}_\mu + \mathbf{x}_0) \right\} \right] \right]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}}. \tag{3.11}
\end{aligned}$$

In order to average over $\mathcal{P}(\mathbf{H})$, the only quantity that needs to be computed is

$$\begin{aligned}
& \left[\exp \left(- \sum_{\mu=1}^n \frac{\beta}{2\sigma^2} (\mathbf{H}\mathbf{u}_\mu)^2 \right) \right]_{\mathbf{H}}^{\text{av}} \\
&= \frac{1}{Z_0} \int d\mathbf{H} \exp \left[- \frac{M}{2} \text{Tr} \left(\mathbf{H}^\top \mathbf{H} \right) - \frac{\beta}{2\sigma^2} \sum_{\mu=1}^n \mathbf{u}_\mu^\top \mathbf{H}^\top \mathbf{H} \mathbf{u}_\mu \right] \\
&= \left\{ \frac{1}{\det \left(\mathbf{I}_n + \frac{\beta}{\alpha\sigma^2} \mathbf{Q} \right)} \right\}^{(-M/2)} = \exp \left[- \frac{M}{2} \text{Tr} \log \left(\mathbf{I}_n + \frac{\beta}{\alpha\sigma^2} \mathbf{Q} \right) \right] \quad (3.12)
\end{aligned}$$

where Z_0 is the normalization term for the Gaussian distribution of \mathbf{H} , and $\alpha = M/N$ is the sampling ratio. The elements of the $n \times n$ matrix \mathbf{Q} are defined by $Q_{\mu\nu} = \frac{1}{N} \mathbf{u}_\mu^\top \mathbf{u}_\nu$.

Using the Fourier representation of the δ function

$$\delta(\mathbf{u}_\mu^\top \mathbf{u}_\nu - NQ_{\mu\nu}) = \frac{1}{2\pi} \int dR_{\mu\nu} \exp(-iR_{\mu\nu}(\mathbf{u}_\mu^\top \mathbf{u}_\nu - NQ_{\mu\nu})) \quad (3.13)$$

and inserting this delta function with an integral over $\mathbf{Q}_{\mu\nu}$ in Eq. (3.11), we get

$$[Z^n]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}} = \int \prod_{\mu \leq \nu} dQ_{\mu\nu} dR_{\mu\nu} \exp[-S(\mathbf{Q}, \mathbf{R})] \quad (3.14)$$

$$\begin{aligned}
S[\mathbf{Q}, \mathbf{R}] &= \frac{M}{2} \text{Tr} \log \left(\mathbf{I}_n + \frac{\beta}{\alpha\sigma^2} \mathbf{Q} \right) - iN \text{Tr}(\mathbf{R}\mathbf{Q}) \\
&- \log \left[\int \prod_{\mu=1}^n d\mathbf{u}_\mu \exp \left[-i \sum_{\mu, \nu} R_{\mu\nu} \mathbf{u}_\mu^\top \mathbf{u}_\nu + \sum_{\mu} V(\mathbf{u}_\mu + \mathbf{x}_0) \right] \right]_{\mathbf{x}_0}^{\text{av}} \quad (3.15)
\end{aligned}$$

This integral over \mathbf{Q}, \mathbf{R} can be evaluated using the saddle point method [8, 9] when $M, N \rightarrow \infty$, holding $\alpha = \frac{M}{N}$ fixed. The saddle point $\mathbf{Q} = \bar{\mathbf{Q}}, \mathbf{R} = -i\bar{\mathbf{R}}$ satisfies the conditions:

$$\bar{Q}_{\mu\nu} = \frac{1}{N} \langle \langle \mathbf{u}_\mu^\top \mathbf{u}_\nu \rangle \rangle \quad (3.16)$$

$$\bar{\mathbf{R}} = \frac{\beta}{2\sigma^2} [\mathbf{I}_n + \frac{\beta}{\alpha\sigma^2} \mathbf{Q}]^{-1} \quad (3.17)$$

obtained by differentiating $S(\mathbf{Q}, \mathbf{R})$ with respect to the elements of \mathbf{Q}, \mathbf{R} . The expectation $\langle \langle \mathbf{u}_\mu^\top \mathbf{u}_\nu \rangle \rangle$ depends on $\bar{\mathbf{R}}$ via

$$\langle\langle \mathbf{u}_\mu^\top \mathbf{u}_\nu \rangle\rangle = \beta \frac{\partial F(\bar{\mathbf{R}})}{\partial \bar{R}_{\mu\nu}} \quad (3.18)$$

with $\exp\{-\beta F(\bar{\mathbf{R}})\}$

$$= \left[\int \prod_{\mu=1}^n \{d^N \mathbf{u}_\mu\} \exp \left[- \sum_{\mu,\nu} \bar{R}_{\mu\nu} \mathbf{u}_\mu^\top \mathbf{u}_\nu - \beta \sum_{\mu} V(\mathbf{u}_\mu + \mathbf{x}_0) \right] \right]_{\mathbf{x}_0}^{\text{av}} \quad (3.19)$$

If $U(x)$ is a convex function, we expect a unique state and a replica symmetric solution for \mathbf{Q}, \mathbf{R} . This implies $\bar{Q}_{\mu\nu} = (Q - q)\delta_{\mu\nu} + q$ and $\bar{R}_{\mu\nu} = (R - r)\delta_{\mu\nu} + r$. With that ansatz,

$$\begin{aligned} & \int \prod_{\mu=1}^n \{d^N \mathbf{u}_\mu\} \exp \left[- \sum_{\mu,\nu} \bar{R}_{\mu\nu} \mathbf{u}_\mu^\top \mathbf{u}_\nu - \beta \sum_{\mu} V(\mathbf{u}_\mu + \mathbf{x}_0) \right] \\ &= \int \prod_{\mu=1}^n \{d^N \mathbf{u}_\mu\} \exp \left[- (R - r) \sum_{\mu} \mathbf{u}_\mu^2 \right. \\ & \quad \left. - r \left(\sum_{\mu} \mathbf{u}_\mu \right)^2 - \beta \sum_{\mu} V(\mathbf{u}_\mu + \mathbf{x}_0) \right] \\ &= \int \frac{d^N \boldsymbol{\xi}}{(2\pi\sigma_\xi^2)^{N/2}} \exp\left(-\frac{\boldsymbol{\xi}^2}{2\sigma_\xi^2}\right) \int \prod_{\mu=1}^n \{d^N \mathbf{u}_\mu\} \\ & \quad \exp \left[- \frac{\beta}{2\sigma_{\text{eff}}^2} \sum_{\mu} \mathbf{u}_\mu^2 + \frac{\beta}{\sigma_{\text{eff}}^2} \boldsymbol{\xi}^\top \left(\sum_{\mu} \mathbf{u}_\mu \right) - \beta \sum_{\mu} V(\mathbf{u}_\mu + \mathbf{x}_0) \right] \end{aligned} \quad (3.20)$$

identifying $R - r \equiv \frac{\beta}{2\sigma_{\text{eff}}^2}$ and $r \equiv -\frac{\beta^2\sigma_\xi^2}{2\sigma_{\text{eff}}^4}$. We have used

$$\begin{aligned} & \int \frac{d^N \boldsymbol{\xi}}{(2\pi\sigma_\xi^2)^{N/2}} \exp\left(-\frac{\boldsymbol{\xi}^2}{2\sigma_\xi^2}\right) \exp\left[\frac{\beta}{\sigma_{\text{eff}}^2} \boldsymbol{\xi}^\top \left(\sum_{\mu} \mathbf{u}_\mu \right)\right] \\ &= \exp\left[\frac{\beta^2\sigma_\xi^2}{2\sigma_{\text{eff}}^4} \left(\sum_{\mu} \mathbf{u}_\mu \right)^2\right] \end{aligned} \quad (3.21)$$

to decouple the item replica coupling in the $(\sum_{\mu} \mathbf{u}_\mu)^2$ term, at the cost of introducing another quenched variable $\boldsymbol{\xi}$. Note that we require $R - r > 0$ and $r < 0$ for this approach to work. These inequalities follow from (3.17) and from $Q - q > 0$ and $q > 0$. The conditions on Q and q would be obvious once we look at interpretation of these quantities described below.

For $V(\mathbf{x}) = \sum_a U(x_a)$ we can simplify further. Remembering that we also need to do the quenched average over \mathbf{x}_0 ,

$$\begin{aligned}
& \left[\int \prod_{\mu=1}^n \{d^N \mathbf{u}_\mu\} \exp \left[- \sum_{\mu, \nu} \bar{R}_{\mu\nu} \mathbf{u}_\mu^\top \mathbf{u}_\nu - \beta \sum_{\mu} V(\mathbf{u}_\mu + \mathbf{x}_0) \right] \right]_{\mathbf{x}_0}^{\text{av}} \\
&= \left[\int \prod_{\mu=1}^n \{d^N \mathbf{u}_\mu\} \exp \left[- \beta \left\{ \frac{1}{2\sigma_{\text{eff}}^2} \sum_{\mu} (\mathbf{u}_\mu^2 - \boldsymbol{\xi}^\top \mathbf{u}_\mu) \right. \right. \right. \\
&\quad \left. \left. \left. + \sum_{\mu} V(\mathbf{u}_\mu + \mathbf{x}_0) \right\} \right] \right]_{\boldsymbol{\xi}, \mathbf{x}_0}^{\text{av}} \\
&= \prod_a \left[\int \prod_{\mu=1}^n \{du_{\mu a}\} \exp \left[- \beta \left\{ \frac{1}{2\sigma_{\text{eff}}^2} \sum_{\mu} (u_{\mu a}^2 - \xi_a u_{\mu a}) \right. \right. \right. \\
&\quad \left. \left. \left. + \sum_{\mu} U(u_{\mu a} + x_{0a}) \right\} \right] \right]_{\xi_a, x_{0a}}^{\text{av}} \tag{3.22}
\end{aligned}$$

Thus, in the saddle point approximation, each of the N components of \mathbf{u} become effectively independent and the saddle point conditions reduce to a self-consistent problem for each component $a = 1, \dots, N$. Since this self-consistent problem is similar for each index, we suppress the subscript a in $u_{\mu a}$ and in x_{0a} . For each a , we have the integral of the form

$$\left[\int \prod_{\mu=1}^n du_{\mu} \exp \left[- \beta \left\{ \frac{1}{2\sigma_{\text{eff}}^2} \sum_{\mu} (u_{\mu}^2 - \xi u_{\mu}) + \sum_{\mu} U(u_{\mu} + x_0) \right\} \right] \right]_{\xi, x_0}^{\text{av}}$$

The replica problem corresponds to a collection of effectively independent optimization.

Definition 9 (Effective Individual Optimization (Replica Approach)) *Single variable u is drawn from the distribution:*

$$P_{\text{eff}}(u | x_0, \xi) = \frac{1}{Z(x_0, \xi)} e^{-\beta \mathcal{E}_{\text{eff}}(u; x_0, \xi)}, \tag{3.23}$$

with an effective mean-field Hamiltonian

$$\mathcal{E}_{\text{eff}}(u; x_0, \xi) = \frac{1}{2\sigma_{\text{eff}}^2} (u^2 - 2\xi u) + U(u + x_0) \tag{3.24}$$

which depends on two quenched variables x_0 and ξ . The variable x_0 has the probability distribution $p_0(x_0)$, whereas ξ is distributed according to a Gaussian distribution with mean zero and variance σ_ξ^2 . The two parameters σ_{eff}^2 and σ_ξ^2 are given by the following set of self-consistency conditions.

$$q = [\langle u \rangle^2]_{x_0, \xi}^{\text{av}}, \quad \Delta Q \equiv Q - q = [\langle (u - \langle u \rangle)^2 \rangle]_{x_0, \xi}^{\text{av}} \quad (3.25)$$

$$\sigma_{\text{eff}}^2 = \sigma^2 + \frac{\beta \Delta Q}{\alpha}, \quad \sigma_\xi^2 = \frac{q}{\alpha} \quad (3.26)$$

where the thermal averages $\langle \dots \rangle$ over u are performed in the P_{eff} ensemble and the so-called quenched average $[\dots]_{x_0, \xi}^{\text{av}}$ is over variables ξ and x_0 .

3.3 Finite Temperature Cavity Method

In this section, we solve the finite temperature problem formulated in the above via the cavity method. With the cost function written in terms of \mathbf{u} as

$$\mathcal{E}(\mathbf{u}) = \frac{1}{2\sigma^2}(\mathbf{H}\mathbf{u})^2 + V(\mathbf{u} + \mathbf{x}_0) \quad (3.27)$$

we define the Boltzmann distribution $P(\mathbf{u}|\mathbf{H}, \mathbf{x}_0)$:

$$P(\mathbf{u}|\mathbf{H}, \mathbf{x}_0) = \frac{1}{Z(\beta|\mathbf{H}, \mathbf{x}_0)} e^{-\beta \mathcal{E}} \quad (3.28)$$

with the normalization factor/partition function given by

$$Z(\beta|\mathbf{H}, \mathbf{x}_0) = \int d\mathbf{u} e^{-\beta \mathcal{E}} \quad (3.29)$$

We now apply the first step of the two-step cavity method. First, we rewrite \mathcal{E} as an

interaction between variable u_a and the rest of the variables

$$\mathcal{E}(\mathbf{u}) = \frac{1}{2\sigma^2} \mathbf{h}_a^2 u_a^2 + \frac{1}{\sigma^2} u_a \mathbf{h}_a \cdot \sum_{b \neq a} \mathbf{h}_b u_b + U(u_a + x_{0a}) + \mathcal{E}_{\setminus a}(\mathbf{u}_{\setminus a}) \quad (3.30)$$

By defining

$$\eta_a \equiv -\frac{\mathbf{h}_a \cdot \sum_{b \neq a} \mathbf{h}_b u_b}{\mathbf{h}_a^2} \quad (3.31)$$

and using $\mathbf{h}_a^2 = 1 + O(\frac{1}{\sqrt{M}})$ we have

$$\mathcal{E} = \frac{1}{2\sigma^2} (u_a^2 - 2u_a \eta_a) + U(u_a + x_{0a}) + \mathcal{E}_{\setminus a}(\mathbf{u}_{\setminus a}) \quad (3.32)$$

with subscript $\setminus a$ indicates that we leave out the node “a”, therefore, $\mathbf{u}_{\setminus a}$, $\mathcal{E}_{\setminus a}$. Equation (3.32) indicates that the variable u_a interacts with all the others only through η_a . Therefore, we rewrite the marginal distribution $P(u_a)$ as an integral over the joint distribution of η_a and u_a , $P(u_a, \eta_a)$.

$$P(u_a) = \frac{1}{Z} \int d\mathbf{u}_{\setminus a} e^{-\beta \mathcal{E}} = \int d\eta_a P(u_a, \eta_a) \quad (3.33)$$

where

$$P(u_a, \eta_a) = \frac{1}{Z} \int d\mathbf{u}_{\setminus a} \delta(\eta_a + \mathbf{h}_a \cdot \sum_{b \neq a} \mathbf{h}_b u_b) e^{-\beta \mathcal{E}} \quad (3.34)$$

for all $a = 1, \dots, N$. Now we introduce a cavity “field” distribution of η_a at the removed node a as

$$P_{\setminus a}(\eta_a) = \frac{1}{Z_{\setminus a}} \int d\mathbf{u} \delta(\eta_a + \mathbf{h}_a \cdot \sum_{b \neq a} \mathbf{h}_b u_b) e^{-\beta \mathcal{E}_{\setminus a}}. \quad (3.35)$$

By comparing (3.34) and (3.35), we get

$$P(u_a) = \frac{\int d\eta_a \exp \left[-\beta \left\{ \frac{(u_a^2 - 2u_a \eta_a)}{2\sigma^2} + U(u_a + x_{0a}) \right\} \right] P_{\setminus a}(\eta_a)}{\int du_a d\eta_a \exp \left[-\beta \left\{ \frac{(u_a^2 - 2u_a \eta_a)}{2\sigma^2} + U(u_a + x_{0a}) \right\} \right] P_{\setminus a}(\eta_a)} \quad (3.36)$$

The assumption of continuity of the global ground state, even in the presence of the

cavity after removing node a , is equivalent to the replica symmetric (RS) hypothesis. This is a valid assumption when the penalty function V is convex. Therefore, in the limit of $N \rightarrow \infty$, even if the nodes of $(N-1, M)$ system are weakly correlated, η_a is still a sum of many variables and $P(\eta_a)_{\setminus a}$ can well be approximated by a Gaussian distribution.

$$P_{\setminus a}(\eta_a) \propto e^{-\frac{(\eta_a - \langle \eta_a \rangle_{\setminus a})^2}{2\langle \delta \eta_a^2 \rangle_{\setminus a}}} \quad (3.37)$$

Then (3.36) becomes

$$P(u_a) = \frac{\exp\{-\frac{\beta}{2\sigma^2}\left(1 - \frac{\beta}{\sigma^2}\langle \delta \eta_a^2 \rangle_{\setminus a}\right)u_a^2 + \frac{\beta u_a}{\sigma^2}\langle \eta_a \rangle_{\setminus a} - \beta U(u_a + x_{0a})\}}{\int du_a \exp\{-\frac{\beta}{2\sigma^2}\left(1 - \frac{\beta}{\sigma^2}\langle \delta \eta_a^2 \rangle_{\setminus a}\right)u_a^2 + \frac{\beta u_a}{\sigma^2}\langle \eta_a \rangle_{\setminus a} - \beta U(u_a + x_{0a})\}} \quad (3.38)$$

Therefore, only the thermal averages $\langle \eta_a \rangle_{\setminus a}$ and the thermal fluctuation strength $\langle \delta \eta_a^2 \rangle_{\setminus a} = \langle (\eta_a - \langle \eta_a \rangle_{\setminus a})^2 \rangle_{\setminus a}$ of the field η_a for the distribution $P_{\setminus a}(\eta_a)$ are left to be computed. In that process the effects of (weak) correlation between the u_a 's have to be accounted for. To do so, we define

$$v_i = \sum_{b \neq a} H_{ib} u_b \quad (3.39)$$

and utilize our definition,

$$\eta_a = - \sum_i H_{ia} v_i \quad (3.40)$$

then we arrive at

$$\langle \eta_a \rangle_{\setminus a} = - \sum_i H_{ia} \langle v_i \rangle \quad (3.41)$$

and

$$\begin{aligned} \langle \delta \eta_a^2 \rangle_{\setminus a} &= \sum_{ij} H_{ia} H_{ja} \langle \delta v_i \delta v_j \rangle \\ &\approx \sum_{ij} \frac{1}{M} \delta_{ij} \langle \delta v_i \delta v_j \rangle = \frac{1}{M} \sum_i \langle \delta v_i^2 \rangle \end{aligned} \quad (3.42)$$

Having done that we need to compute $\langle v_i \rangle$ and $\langle \delta v_i^2 \rangle$. To do so, this time in addition to site

a we exclude site i . Hence from (3.32) we get

$$\mathcal{E}_{\setminus a}(\mathbf{u}_{\setminus a}) = \frac{1}{2\sigma^2}v_i^2 + \mathcal{E}_{\setminus ai}(\mathbf{u}_{\setminus a}) \quad (3.43)$$

After carrying out the same computation as in (3.33), (3.34), and (3.36) for the marginal distribution $Q_{\setminus a}(v_i)$, we arrive at

$$Q_{\setminus a}(v_i) = \frac{\exp\left\{-\frac{\beta}{2\sigma^2}v_i^2 - \frac{(v_i - \langle v_i \rangle_{\setminus i})^2}{2\langle \delta v_i^2 \rangle_{\setminus i}}\right\}}{\int dv_i \exp\left\{-\frac{\beta}{2\sigma^2}v_i^2 - \frac{(v_i - \langle v_i \rangle_{\setminus i})^2}{2\langle \delta v_i^2 \rangle_{\setminus i}}\right\}} \quad (3.44)$$

Therefore

$$Q_{\setminus a}(v_i) = \frac{\exp\left\{-\frac{\beta}{2\sigma^2}\left(1 + \frac{\sigma^2}{\beta\langle \delta v_i^2 \rangle_{\setminus i}}\right)\left(v_i - \frac{\langle v_i \rangle_{\setminus i}}{1 + \frac{\sigma^2}{\beta\langle \delta v_i^2 \rangle_{\setminus i}}}\right)^2\right\}}{\int dv_i \exp\left\{-\frac{\beta}{2\sigma^2}\left(1 + \frac{\sigma^2}{\beta\langle \delta v_i^2 \rangle_{\setminus i}}\right)\left(v_i - \frac{\langle v_i \rangle_{\setminus i}}{1 + \frac{\sigma^2}{\beta\langle \delta v_i^2 \rangle_{\setminus i}}}\right)^2\right\}} \quad (3.45)$$

and then $\langle \delta v_i^2 \rangle$ is

$$\langle \delta v_i^2 \rangle = \frac{1}{\frac{\beta}{\sigma^2}\left(1 + \frac{\sigma^2}{\beta\langle \delta v_i^2 \rangle_{\setminus i}}\right)} = \frac{\langle \delta v_i^2 \rangle_{\setminus i}}{1 + \frac{\beta\langle \delta v_i^2 \rangle_{\setminus i}}{\sigma^2}} \quad (3.46)$$

and $\langle v_i \rangle$ is at

$$\langle v_i \rangle = \frac{\langle v_i \rangle_{\setminus i}}{1 + \frac{\beta\langle \delta v_i^2 \rangle_{\setminus i}}{\sigma^2}}. \quad (3.47)$$

Notice how both these moments for the $(N-1, M)$ system is scaled down by the same factor, when compared to the moments for the $(N-1, M-1)$ system. Using arguments similar to the fluctuation-dissipation [49] theorem, we could show that the change in $\langle v_i \rangle$ due a change in $\langle v_i \rangle_{\setminus i}$, susceptibility of sorts, is closely related to $\langle \delta v_i^2 \rangle_{\setminus i}^{-2}$ times $\langle \delta v_i^2 \rangle$, with the first term of the product playing the role of temperature.

Carrying on, we get

$$\begin{aligned}
\langle \delta v_i^2 \rangle_{\setminus i} &= \sum_{b,c \neq a} H_{ib} H_{ic} \langle \delta u_b \delta u_c \rangle_{\setminus ai} \\
&= \sum_{b,c \neq a} \frac{1}{M} \delta_{bc} \langle \delta u_b \delta u_c \rangle_{\setminus ai} + O\left(\frac{N}{M^{3/2}}, \frac{N^{1/2}}{M}\right) \\
&\approx \frac{1}{M} \sum_{b \neq a} \langle \delta u_b^2 \rangle_{\setminus ai}
\end{aligned} \tag{3.48}$$

since the $(N-1, M-1)$ system, indicated by the subscript ‘ $\setminus ai$ ’, is independent of H_{ib} and H_{ic} , $H_{ib}H_{ic} = \frac{1}{M}\delta_{bc} + O(\frac{1}{M})$ fluctuations, and $\langle \delta u_b \delta u_c \rangle_{\setminus ai} \sim O(\frac{1}{\sqrt{M}}, \frac{1}{\sqrt{N}})$ when $b \neq c$, indicating that nodes are only weakly correlated.

To make connection with the notation in Sec. 3.2, let us introduce ΔQ

$$\Delta Q \equiv \frac{1}{N} \sum_a \langle \delta u_a^2 \rangle \approx \frac{1}{N-1} \sum_{b \neq a} \langle \delta u_b^2 \rangle_{\setminus ai}, \tag{3.49}$$

the second approximate equality becoming exact in the thermodynamic limit. Then, we have

$$\langle \delta v_i^2 \rangle_{\setminus i} = \Delta Q / \alpha. \tag{3.50}$$

Therefore from (3.42), (3.46), and (3.60)

$$\frac{\beta}{\sigma^2} \langle \delta \eta_a^2 \rangle_{\setminus a} = \frac{1}{(1 + \frac{\sigma^2}{\beta \Delta Q / \alpha})} \tag{3.51}$$

and from (3.41), (3.47), and (3.60)

$$\langle \eta_a \rangle_{\setminus a} = \frac{\sum_i H_{ia} \sum_{b \neq a} H_{ib} \langle u_b \rangle_{\setminus ai}}{(1 + \frac{\beta \Delta Q}{\alpha \sigma^2})}. \tag{3.52}$$

Moreover, we define

$$\xi_a \equiv \sum_i H_{ia} \sum_{b \neq a} H_{ib} \langle u_b \rangle_{\setminus ai} \tag{3.53}$$

which has variance $\sigma_\xi^2 = q/\alpha$ with q

$$q = \frac{1}{N} \sum_a \langle u_a \rangle^2 \approx \frac{1}{N-1} \sum_b \langle u_b \rangle_{\setminus ai}^2 \quad (3.54)$$

being the mean squared error. Therefore, by plugging (3.52) and (3.53) into Eq. (3.38), the marginal distribution for single variable u_a becomes

$$P(u_a) = \frac{\exp\{-\frac{\beta}{2\sigma_{\text{eff}}^2}(u_a^2 - 2u_a\xi_a) - \beta U(x_{0a} + u_a)\}}{\int du_a \exp\{-\frac{\beta}{2\sigma_{\text{eff}}^2}(u_a^2 - 2u_a\xi_a) - \beta U(x_{0a} + u_a)\}} \quad (3.55)$$

with $\sigma_{\text{eff}}^2 = \sigma^2(1 + \frac{\beta\Delta Q}{\alpha\sigma^2})$, and the effective cost function for the individual node is

$$\mathcal{E}(u_a) = \frac{1}{2\sigma_{\text{eff}}^2}(u_a^2 - 2u_a\xi_a) + U(x_{0a} + u_a) \quad (3.56)$$

Therefore, with \mathcal{E} replaced by a set of effectively decoupled nodes, and the sum over index a replaced by a quenched average over ξ_a, x_{0a} . As a result, the self-consistency conditions for the MSE

$$q = \frac{1}{N} \sum_{a=1}^N \langle u_a \rangle^2 \quad (3.57)$$

and for

$$\Delta Q = \frac{1}{N} \sum_{a=1}^N \langle \delta u_a^2 \rangle \quad (3.58)$$

reduce to

$$q = [\langle u \rangle_{\text{eff}}^2]_{\xi, x_0}^{\text{av}} \quad (3.59)$$

and

$$\Delta Q = [\langle \delta u^2 \rangle_{\text{eff}}]_{\xi, x_0}^{\text{av}} \quad (3.60)$$

where the thermal average $\langle \dots \rangle_{\text{eff}}$ is performed with respect to the effective individual node distribution (3.23) and $[\dots]_{\xi, x_0}^{\text{av}}$ is the quenched average over variables ξ, x_0 , with ξ drawn

from $\mathcal{N}(0, q/\alpha)$ and signal x_0 drawn independently from a distribution $P(x_0)$. These self-consistency equations are exactly the same those from the RS ansatz in Sec. 3.2.

Chapter 4

Cavity Method at Zero Temperature

In the previous chapter, we saw that in order to study the LS based reconstruction, we need to take the limits $\beta \rightarrow \infty$, and then $\sigma \rightarrow 0$. A nontrivial aspect of the zero temperature limit ($\beta \rightarrow \infty$) is the quantity $\beta\Delta Q$ in Eq. (3.26) that behaves differently in different phases of reconstruction. Using Eq. (3.25), this quantity is just β times the thermal fluctuation in u . The fluctuation-dissipation relation [49] implies that this quantity may be interpreted as a local susceptibility. In the following chapter based on the zero temperature cavity method, we formally introduce a susceptibility and use its properties to give a more transparent derivation of the same equations.

4.1 Susceptibility and Conjugate Variables

The optimization problem associated with the regularized least-squared based reconstruction problem involves minimizing the energy function $\mathcal{E}_0(\mathbf{x}) = \frac{(\mathbf{y} - \mathbf{H}\mathbf{x})^2}{2\sigma^2} + V(\mathbf{x})$. For the noise free case, using $\mathbf{y} = \mathbf{H}\mathbf{x}_0$, the energy to be optimized may be rewritten as

$$\mathcal{E}(\mathbf{u}) = \frac{1}{2\sigma^2} \mathbf{u}^T \mathbf{H}^T \mathbf{H} \mathbf{u} + V(\mathbf{u} + \mathbf{x}_0). \quad (4.1)$$

where $\mathbf{u} = \mathbf{x} - \mathbf{x}_0$. Note that, unlike the function $\mathcal{E}_0(\mathbf{x})$, which is parametrized by known quantities (the data \mathbf{y} and the measurement matrix \mathbf{H}) and can therefore be empirically optimized with respect to its argument, the closely related function $\mathcal{E}(\mathbf{u}) = \mathcal{E}_0(\mathbf{u} + \mathbf{x}_0)$ depends on the knowledge of the original signal \mathbf{x}_0 . The purpose of dealing with this function is *not* to provide an algorithm to estimate this signal given measured data, but

to study the *statistical* behavior of this function and its minima over the distribution of problem instances, namely, input signals and the measurement matrices. For example, we can calculate the distribution of each component of the estimation error vector \mathbf{u} , given the distributions of \mathbf{x}_0 and \mathbf{H} . We will be working with $\mathcal{E}(u)$, although the susceptibility for a particular problem instance, to be defined below, could be defined completely in terms of $\mathcal{E}_0(\mathbf{x})$.

In case this cost function reproduces the correct answer, the function $\mathcal{E}(\mathbf{u})$ minimizes at $\mathbf{u} = 0$. Looking at the structure of $\mathcal{E}(u)$ near zero tells us about potential “flat” directions in error space, along which the cost function fails to constrain errors. This failure could be quantified in terms of a susceptibility to error, the dependence of which on the problem parameters allows us to characterize the phase transition boundaries.

As usual, let us consider a general regularization function $V(\mathbf{x})$ for which there is a unique minimum to the cost function. Let the minimum of $\mathcal{E}(\mathbf{u})$ be at $\mathbf{u} = \hat{\mathbf{u}}$. We introduce an augmented cost function

$$\mathcal{E}(\mathbf{u}; \mathbf{f}) = \frac{1}{2\sigma^2} \mathbf{u}^T \mathbf{H}^T \mathbf{H} \mathbf{u} + V(\mathbf{u} + \mathbf{x}_0) - \mathbf{f} \cdot \mathbf{u}. \quad (4.2)$$

with the variables \mathbf{f} , which are conjugate to \mathbf{u} . Optimizing $\mathcal{E}(\mathbf{u}; \mathbf{f})$ will produce an \mathbf{f} dependent answer $\mathbf{u} = \hat{\mathbf{u}}(\mathbf{f})$. For small \mathbf{f} we expect

$$\hat{\mathbf{u}}(\mathbf{f}) = \hat{\mathbf{u}} + \boldsymbol{\chi} \mathbf{f} + \dots \quad (4.3)$$

defining the susceptibility matrix $\boldsymbol{\chi}$.

If \mathcal{E} is differentiable, the optimum $\hat{\mathbf{u}}(\mathbf{f})$ is the solution of

$$\mathbf{f} = \nabla_{\mathbf{u}} \mathcal{E}(\mathbf{u}) \quad (4.4)$$

Where $\mathbf{f} = \mathbf{0}$, \mathbf{u} is at its optimal value $\hat{\mathbf{u}}$. If perturbation \mathbf{f} is small and \mathcal{E} is differentiable

to higher orders, we can expect $\delta \mathbf{u}$ to be small and, therefore, Taylor expand $\mathcal{E}(\mathbf{u} + \delta \mathbf{u})$ around $\mathbf{u} = \hat{\mathbf{u}}$

$$\mathcal{E}(\hat{\mathbf{u}} + \delta \mathbf{u}) = \mathcal{E}(\hat{\mathbf{u}}) + \frac{1}{2} \sum_{ab} \delta u_a \delta u_b \left. \frac{\partial^2 \mathcal{E}}{\partial u_a \partial u_b} \right|_{\mathbf{u}=\hat{\mathbf{u}}} + \dots \quad (4.5)$$

From (4.4) and (4.5), we can identify the inverse susceptibility $(\chi^{-1})_{ab} = \left. \frac{\partial^2 \mathcal{E}}{\partial u_a \partial u_b} \right|_{\mathbf{u}=\hat{\mathbf{u}}}$ and can show that

$$\min_{\mathbf{u}} \mathcal{E}(\mathbf{u}; \mathbf{f}) = \mathcal{E}(\hat{\mathbf{u}}) - \hat{\mathbf{u}}^T \mathbf{f} - \frac{1}{2} \mathbf{f}^T \chi \mathbf{f} + \dots \quad (4.6)$$

This expansion will appear several times in our derivations.

We will now discuss the structure of the susceptibility matrix χ . It is simplest to study the properties of χ , when the potential $U(x)$ has continuous second derivatives.

If we could Taylor expand around the solution $\mathbf{u} = \hat{\mathbf{u}}$, we will have

$$\mathcal{E}(\hat{\mathbf{u}} + \delta \mathbf{u}; \mathbf{f}) = \mathcal{E}(\hat{\mathbf{u}}; 0) + \frac{1}{2} \delta \mathbf{u}^T \left[\frac{\mathbf{H}^T \mathbf{H}}{\sigma^2} + \mathbf{W}(\mathbf{x}) \right] \delta \mathbf{u} - \mathbf{f} \cdot (\hat{\mathbf{u}} + \delta \mathbf{u}) + \dots \quad (4.7)$$

where $W_{ab}(\mathbf{x}) = U''(\hat{u}_a + x_{0a})\delta_{ab}$. Optimizing over $\delta \mathbf{u}$, we see that the susceptibility matrix would be given by

Definition 10 (Susceptibility Matrix)

$$\chi(\mathbf{x}, \mathbf{H}) = \left[\frac{\mathbf{H}^T \mathbf{H}}{\sigma^2} + \mathbf{W}(\mathbf{x}) \right]^{-1}. \quad (4.8)$$

When \mathbf{H} is a large random matrix, we can make asymptotic estimates of the mean and the variance of different components of the susceptibility matrix χ . One way to approach this problem is to formally expand the RHS of Eq. (4.8) in powers of $\frac{\mathbf{H}^T \mathbf{H}}{\sigma^2}$ (see Fig. 4.1)

$$\chi_{aa} = a \text{---} a - a \text{---} \text{---} a + \dots$$

$W_{aa}^{-1} \quad W_{aa}^{-1} H_{ai}^T H_{ia} W_{aa}^{-1} \quad W_{aa}^{-1} H_{ai}^T H_{ib} W_{bb}^{-1} H_{bj}^T H_{ja} W_{aa}^{-1}$

Figure 4.1: The diagrammatic expansion of susceptibility.

and compute moments by averaging over H_{ia} diagrammatically. Namely, we expand

$$\begin{aligned} \chi &= \mathbf{W}^{-1} - \frac{1}{\sigma^2} \mathbf{W}^{-1} \mathbf{H}^T \mathbf{H} \mathbf{W}^{-1} \\ &+ \frac{1}{\sigma^4} \mathbf{W}^{-1} \mathbf{H}^T \mathbf{H} \mathbf{W}^{-1} \mathbf{H}^T \mathbf{H} \mathbf{W}^{-1} - \dots \end{aligned} \quad (4.9)$$

and we associate a double-line digram with each of the terms in (4.9) and then compute moments of the form $[\chi^{a_1 b_1} \chi^{a_2 b_2} \dots \chi^{a_k b_k} H_{i_1 c_1} \dots H_{i_l c_l}]_{\mathbf{H}}^{\text{av}}$ using Wick's theorem, since \mathbf{H} distribution is Gaussian with mean and covariance specified by Eq. (3.1) and Eq. (3.2), respectively.

Following works on singular values of random matrices [50, 51], we use the fact that, in the large M, N limit, only the planar diagrams survive. $[\chi(\mathbf{x}, \mathbf{H})]_{\mathbf{H}}^{\text{av}}$ could be written as $[\mathbf{W}(\mathbf{x}) - \Sigma(\mathbf{x}) \mathbf{I}_N]^{-1}$, where $\Sigma(\mathbf{x})$ is a self-energy term. The planar contributions to the self-energy are shown in Fig. 4.1 and can be re-summed as

$$\Sigma(\mathbf{x}) = -\frac{1}{\sigma^2} \frac{1}{1 + \frac{1}{M\sigma^2} \sum_a \chi^{aa}(\mathbf{x})} \quad (4.10)$$

Hence, the mean susceptibility (holding x_a 's fixed but averaging over \mathbf{H}) is given by

$$\chi^{\text{av}}(\mathbf{x}) \equiv [\chi(\mathbf{x}, \mathbf{H})]_{\mathbf{H}}^{\text{av}} = \left[\mathbf{W}(\mathbf{x}) + \frac{M}{M\sigma^2 + \text{Tr}[\chi^{\text{av}}(\mathbf{x})]} \mathbf{I}_N \right]^{-1}. \quad (4.11)$$

Moreover, one can show that the variance of each element of the χ matrix is of the order of $1/M$ and vanishes in the $N, M \rightarrow \infty$ limit (for $\alpha = M/N$ held fixed). Covariance of χ , $[\chi^{ab}(\mathbf{x}, \mathbf{H})\chi^{cd}(\mathbf{x}, \mathbf{H})]_{\mathbf{H}}^{\text{av}}$ could be computed using the diagrams in Fig. 4.1 and they are suppressed in the large M, N limit, since their contributions are $O(\frac{1}{M}, \frac{1}{N})$.

Note that only the diagonal terms χ^{aa} have non-trivial means, whereas the off-diagonal terms average to zero. For diagonal terms, namely local susceptibilities, we get the following equations:

$$[\chi^{aa}(\mathbf{x})]_{\mathbf{H}}^{\text{av}} = \left[W_{aa}(x_a) + \frac{1}{\sigma^2 + \frac{\bar{\chi}(\mathbf{x})}{\alpha}} \right]^{-1} \quad (4.12)$$

$$\bar{\chi}(\mathbf{x}) \equiv \frac{1}{N} \sum_a [\chi^{aa}(\mathbf{x})]_{\mathbf{H}}^{\text{av}}. \quad (4.13)$$

Thus, given the distribution of W_{aa} , which themselves depends on the distribution of x_a (where $x_a = u_a + x_{0a}$), we can determine the distribution of $[\chi^{aa}]_{\mathbf{H}}^{\text{av}}$, self-consistently, from these equations.

One should note, although $[\chi^{ab}]_{\mathbf{H}}^{\text{av}} = 0$ for $a \neq b$, for a particular choice of \mathbf{H} , χ^{ab} is a \mathbf{H} -dependent number of the order $1/\sqrt{M}$. Even if these off-diagonal terms are small compared to the self-averaging diagonal terms, they have an important effect on the self-consistency equations via the so-called Onsager reaction term [52]. In particular, we will need the correlation of χ^{ab} with the corresponding matrix elements of $\mathbf{H}^T \mathbf{H}$. Using the identity $[\mathbf{W} + \frac{\mathbf{H}^T \mathbf{H}}{\sigma^2}] \chi = \mathbf{I}_N$, we can prove a useful corollary of the result in Eq. (4.11).

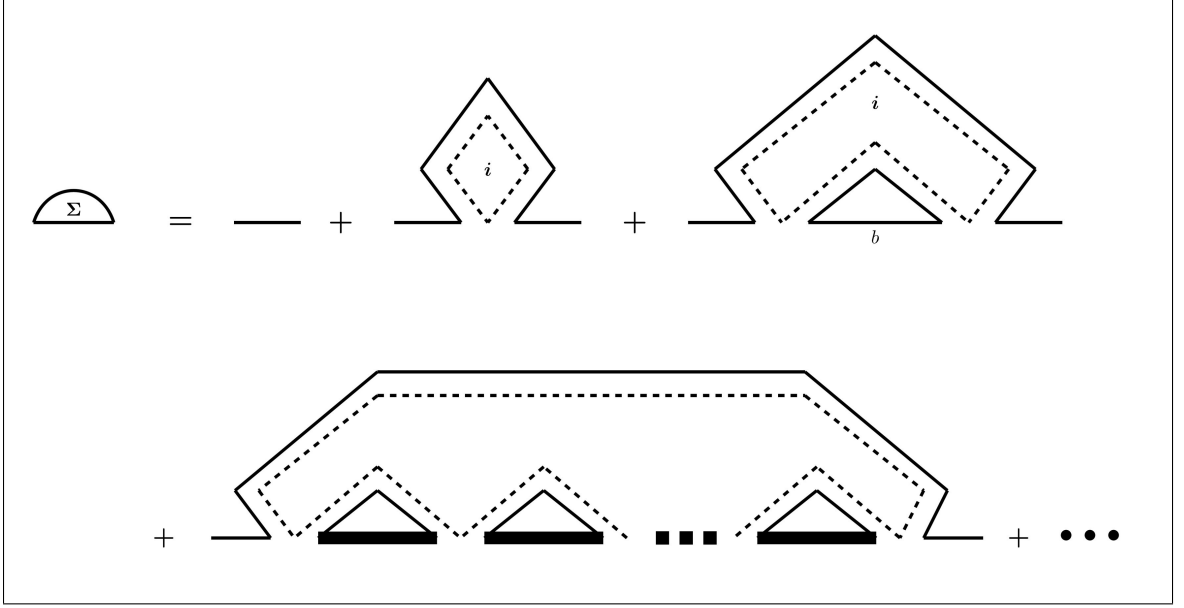


Figure 4.2: Planar diagrams contributing to the self-energy.

$$\begin{aligned}
\frac{1}{\sigma^2} [\mathbf{H}^T \mathbf{H} \boldsymbol{\chi}(\mathbf{x}, \mathbf{H})]_{\mathbf{H}}^{\text{av}} &= \mathbf{I}_N - \mathbf{W}(\mathbf{x}) \boldsymbol{\chi}^{\text{av}}(\mathbf{x}) \\
&= \mathbf{I}_N - \mathbf{W}(\mathbf{x}) [\mathbf{W}(\mathbf{x}) + \frac{M}{M\sigma^2 + \text{Tr}[\boldsymbol{\chi}^{\text{av}}(\mathbf{x})]} \mathbf{I}_N]^{-1} \\
&= \frac{M}{M\sigma^2 + \text{Tr}[\boldsymbol{\chi}^{\text{av}}(\mathbf{x})]} [\mathbf{W} + \frac{M}{M\sigma^2 + \text{Tr}[\boldsymbol{\chi}^{\text{av}}(\mathbf{x})]} \mathbf{I}_N]^{-1} \\
&= \frac{M \boldsymbol{\chi}^{\text{av}}(\mathbf{x})}{M\sigma^2 + \text{Tr}[\boldsymbol{\chi}^{\text{av}}(\mathbf{x})]} \\
&= \frac{\alpha \boldsymbol{\chi}^{\text{av}}(\mathbf{x})}{\alpha\sigma^2 + \bar{\chi}(\mathbf{x})}
\end{aligned} \tag{4.14}$$

In particular, Eq. (4.14) implies

$$[\text{Tr}(\mathbf{H}^T \mathbf{H} \boldsymbol{\chi}(\mathbf{x}, \mathbf{H}))]_{\mathbf{H}}^{\text{av}} = \frac{M\sigma^2 \bar{\chi}(\mathbf{x})}{\alpha\sigma^2 + \bar{\chi}(\mathbf{x})} \tag{4.15}$$

which will be a useful identity in the next section.

Before we leave the section, we should mention that many observations made here are

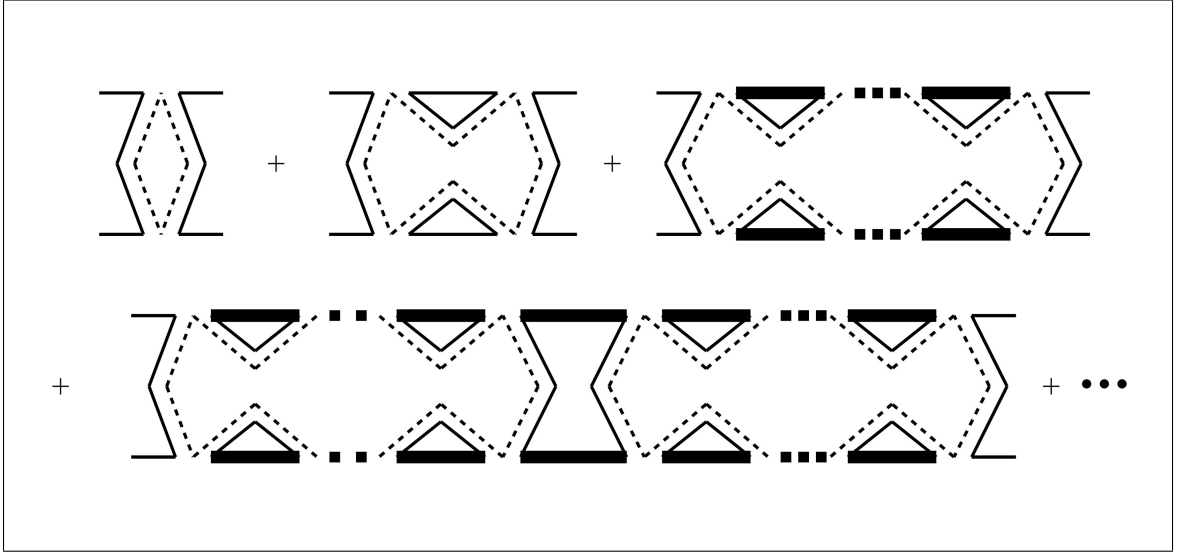


Figure 4.3: The leading planar diagrams in covariance computation are of the order $O(\frac{1}{M}, \frac{1}{N})$, as can be seen from counting a factor of M or N for appropriate index loop, and counting a factor of $\frac{1}{M}$ for each double-line contraction coming from averaging over the matrix elements.

independent of the assumption that $U(x)$ has a continuous second derivative. For example, in the case of compressed sensing with $U(x) = \lambda|x|$, we could define a second-differentiable function $U_\epsilon(x)$ such that $\lim_{\epsilon \rightarrow 0} U_\epsilon(x) = U(x)$, for example, $U_\epsilon(x) = \sqrt{x^2 + \epsilon^2}$ or $U_\epsilon(x) = \frac{1}{\epsilon} \log(2 \cosh(\epsilon x))$. If $x_a = x_{0a} + u_a$ goes to zero as ϵ vanishes, then the corresponding $W_{aa} = U''_\epsilon(x_a)$ diverges. However, the corresponding local susceptibility, χ^{aa} , just becomes zero in this limit. Therefore, as $\epsilon \rightarrow 0$, the idea of using effective single variable optimization problems and determining the self-consistent distribution of x_a and χ^{aa} remains meaningful. We just need to separate out the set of variables x_a for which W_{aa} diverges and treat this set carefully. As a consequence of χ remaining well-defined in the $\epsilon \rightarrow 0$ limit, many relations derived in this section, such as Eqs. (4.6), (4.14) and (4.15), remain valid even if the potential $U(x)$ becomes singular at some values of x .

4.2 Zero-temperature Cavity Method: Removing a Variable Node

As a motivation for introduction of the cavity method [10], notice that Eq. (4.12) could be written as $[\chi^{aa}]_{\mathbf{H}}^{\text{av}} = 1/(W_{aa} + \frac{1}{\sigma_{\text{eff}}^2})$. This expression for local susceptibility would make sense if we could break this problem into effectively independent optimization problems for individual variables in the following manner: For the variable u_a , the function to be optimized is $\mathcal{E}_{\text{eff}}(u_a; x_{0a}, \xi_a) = \frac{1}{2\sigma_{\text{eff}}^2} (u_a^2 - 2\xi_a u_a) + U(u_a + x_{0a})$, as mentioned in Eq. (3.24), along with identifying σ_{eff}^2 as $\sigma^2 + \frac{\bar{\chi}}{\alpha}$. We will prove these assertions in this section. Cavity method essentially codifies how each individual variable u_a interacts with the rest of the system, and how, in the asymptotic limit of large M, N , this interaction becomes simple.

To perform any concrete calculation, we will also need the distribution of W_{aa} , which, in turn means that we need the joint distribution of x_{0a}, u_a . Given a particular x_{0a} , the optimum u_a depends on the influence of other variables through the quantity ξ_a which appears as a parameter in the expression of \mathcal{E}_{eff} . For a system with large M, N , the distribution of ξ_a can be approximated by a Gaussian distribution. Cavity method provides a way of characterizing this distribution.

For the ensuing discussion, it is useful to visualize the problem in terms of a bipartite graph (see Fig. 4.2), where the variables x_a are represented by circular nodes and the ‘constraints’ arising from each y_i (namely, the terms $\frac{1}{2\sigma^2}(y_i - \sum_a H_{ia}x_a)^2 = \frac{1}{2\sigma^2}(\sum_a H_{ia}u_a)^2$ in the cost function) are represented by squares. Had we stuck to a finite temperature description, this graph would be the factor graph [53]. If we insist on satisfying the condition $\mathbf{y} = \mathbf{H}\mathbf{x}$, this graph could be thought of as a Tanner graph [54], with the circles being the variable nodes and the squares being the ‘check’ nodes. The system with N variables (circles) and M data constraints (squares) would be represented as the (N, M) system. The task is to relate properties of the (N, M) system to $(N - 1, M)$ system and obtain self consistency conditions based on quantities that converge in the thermodynamic limit, $N, M \rightarrow \infty$.

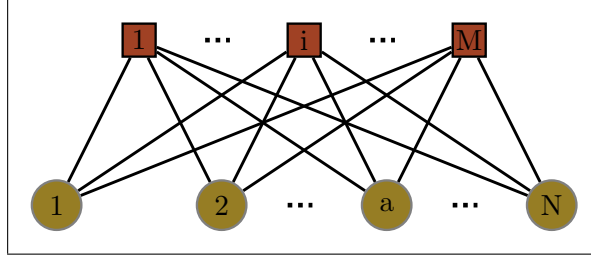


Figure 4.4: Bipartite graph with variable nodes (circles) and constraint nodes (squares).

We pick a particular node a and partition the cost function into a contribution purely from the node, a term representing the interaction of the node variable with the rest of the system, and, lastly, the cost function of the $(N - 1, M)$ system:

$$\mathcal{E}(\mathbf{u}) = \frac{1}{2\sigma^2} \mathbf{u}^T \mathbf{H}^T \mathbf{H} \mathbf{u} + V(\mathbf{u} + \mathbf{x}_0) \quad (4.16)$$

$$\begin{aligned} \mathcal{E}(\mathbf{u}) = & \frac{1}{2\sigma^2} u_a^2 + U(u_a + x_{0a}) + \frac{1}{\sigma^2} u_a \mathbf{h}_a \cdot \sum_{b \setminus a} \mathbf{h}_b u_b \\ & + \frac{1}{2\sigma^2} \left(\sum_{b \setminus a} \mathbf{h}_b u_b \right)^2 + \sum_{b \setminus a} U(u_b + x_{0b}) \end{aligned} \quad (4.17)$$

Here, the a -th column of the \mathbf{H} matrix is being represented by the vector \mathbf{h}_a , and, the subscript $\setminus a$ indicates that we leave out the node a . Moreover, we approximated \mathbf{h}_a^2 by its average value

$$[\mathbf{h}_a^2]_{\mathbf{H}}^{\text{av}} = \sum_i [H_{ia}^2]_{\mathbf{H}}^{\text{av}} = \sum_{i=1}^M \frac{1}{M} = 1, \quad (4.18)$$

since \mathbf{h}_a^2 is a sum of M terms and is self-averaging. The typical fluctuation of \mathbf{h}_a^2 from its average value 1 asymptotically vanishes as $O(1/\sqrt{M})$.

The system without node ‘ a ’, i.e., the system with a ‘cavity’ (see Fig. 4.2), will have its own optimum values $u_b = \hat{u}_b$, for all $b \neq a$. The variable u_a interacts with the rest of the system through the quantity $\mathbf{h}_a \cdot \sum_{b \setminus a} \mathbf{h}_b u_b$. The program of cavity method is to characterize the distribution of this quantity in terms of some parameters relating to the

$(N - 1, M)$ system, and then use the fact that node ‘ a ’ is statistically the same as every other node to relate these parameters to the distribution of u_a . We summarize the above in the following figure.

Cavity Mean-field Approach

- Look at the optimization as a systems of **N variable nodes (green circles)** and **M constraint nodes (brown squares)**.

- The goal is to integrate out the rest of the system but variable node “ a ” and find an individual optimization in terms of only u_a .
- Expansion of cost function in terms of variable u_a leads to

$$\min_{\mathbf{u}} \left\{ \frac{1}{2\sigma^2} u_a^2 + U(u_a + x_{0a}) + \frac{1}{\sigma^2} u_a \mathbf{h}_a \cdot \sum_{b \neq a} \mathbf{h}_b u_b + \mathcal{E}_{(N-1, M)} \right\}$$
- Cavity method is to characterize the distribution of $\mathbf{h}_a \cdot \sum_{b \neq a} \mathbf{h}_b u_b$

One may be tempted to carry on this program, replacing $\sum_{b \neq a} \mathbf{h}_b u_b$ by $\sum_{b \neq a} \mathbf{h}_b \hat{u}_b$, treating it as a Gaussian quenched variable (because is a sum of many random vector \mathbf{h}_b) and estimating its variance over many realizations of \mathbf{H} as $\sum_{b \neq a} [(\mathbf{h}_a \cdot \mathbf{h}_b)^2]^{\text{av}} [\hat{u}_b^2]^{\text{av}} = [\hat{\mathbf{u}}_{\setminus a}^2]^{\text{av}} / M$. We use the notation $\mathbf{u}_{\setminus a}$ for the vector for leaving out the a component and $\hat{\mathbf{u}}_{\setminus a}$ for its optimal value in the $(N - 1, M)$ system. Had this approach been correct, the asymptotic equality between $[u_a^2]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}}$ and $[\hat{\mathbf{u}}_{\setminus a}^2]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}} / N$ would have led to a simple self-consistency condition: namely the variance of the Gaussian quenched variable is $[u_a^2]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}} / \alpha$.

Unfortunately, such a ‘straightforward’ approach is wrong on two counts. First, the components of $\mathbf{u}_{\setminus a}$, are influenced by u_a and *vice versa*. Therefore, $\mathbf{u}_{\setminus a}$ cannot be replaced by $\hat{\mathbf{u}}_{\setminus a}$. Second, the components of the optimum vector $\hat{\mathbf{u}}_{\setminus a}$ are correlated with the vectors

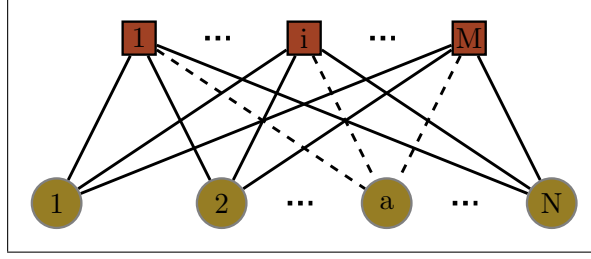


Figure 4.5: The $(N - 1, M)$ cavity system. Node a has been removed from the system by removing the links to it.

\mathbf{h}_b , making the variance estimate wrong.

Correcting the first error requires introducing the Onsager reaction terms [52]. One can start by rewriting Eq. (4.17) as follows

$$\mathcal{E}(\mathbf{u}) = \frac{1}{2\sigma^2}u_a^2 + U(u_a + x_{0a}) + \mathcal{E}_{\setminus a}(\mathbf{u}_{\setminus a}) - \mathbf{u}_{\setminus a}^T \mathbf{f}_{\setminus a}. \quad (4.19)$$

The cost function of the $(N - 1, M)$ system is $\mathcal{E}_{\setminus a}(\mathbf{u}_{\setminus a})$. We identify $(\mathbf{f}_{\setminus a})_b = -\frac{1}{\sigma^2}\mathbf{h}_b \cdot \mathbf{h}_a u_a$ to be the local force exerting on each node u_b , due to presence of node u_a . Since we are looking for the ground state, we minimize the expression in Eq. (4.19)

$$\begin{aligned} \min_{\mathbf{u}} \mathcal{E}(\mathbf{u}) = & \min_{u_a, \mathbf{u}_{\setminus a}} \left\{ \frac{1}{2\sigma^2}u_a^2 + U(u_a + x_{0a}) \right. \\ & \left. + \mathcal{E}_{\setminus a}(\mathbf{u}_{\setminus a}) - \mathbf{u}_{\setminus a}^T \mathbf{f}_{\setminus a} \right\}. \end{aligned} \quad (4.20)$$

Given that $\mathbf{h}_b \cdot \mathbf{h}_a$ is of the order $1/\sqrt{M}$, $\mathbf{f}_{\setminus a}$ is small, and we can invoke the definition of susceptibility $\chi_{\setminus a}$ for the $(N - 1, M)$ system and use expansion of the minimized cost function (4.6).

$$\begin{aligned} \min_{\mathbf{u}} \mathcal{E}(\mathbf{u}) = & \min_{u_a} \left\{ \frac{1}{2\sigma^2}u_a^2 + U(u_a + x_{0a}) \right. \\ & \left. + \mathcal{E}_{\setminus a}(\hat{\mathbf{u}}_{\setminus a}) - \hat{\mathbf{u}}_{\setminus a}^T \mathbf{f}_{\setminus a} - \frac{1}{2}\mathbf{f}_{\setminus a}^T \chi_{\setminus a} \mathbf{f}_{\setminus a} \right\} \end{aligned} \quad (4.21)$$

and plugging in $(\mathbf{f}_{\setminus a})_b = -\frac{1}{\sigma^2} \mathbf{h}_b \cdot \mathbf{h}_a u_a$, we get

$$\begin{aligned} \min_{\mathbf{u}} \mathcal{E}(\mathbf{u}) = \min_{u_a} \{ & \frac{1}{2\sigma_{\text{eff}}^2} u_a^2 + U(u_a + x_{0a}) \\ & + \frac{1}{\sigma^2} u_a \mathbf{h}_a \cdot \sum_{b \neq a} \mathbf{h}_b \hat{u}_b + \mathcal{E}_{\setminus a}(\hat{\mathbf{u}}_{\setminus a}) \} \end{aligned} \quad (4.22)$$

where

$$\frac{1}{\sigma_{\text{eff}}^2} = \frac{1}{\sigma^2} \left(1 - \frac{1}{\sigma^2} \sum_{b,c \neq a} (\mathbf{h}_a \cdot \mathbf{h}_b)(\mathbf{h}_a \cdot \mathbf{h}_c) \chi_{\setminus a}^{bc} \right) \quad (4.23)$$

The quantity $\sum_{b,c \neq a} (\mathbf{h}_b)_i (\mathbf{h}_c)_j \chi_{\setminus a}^{bc}$, is independent of \mathbf{h}_a . As a result, $\sum_{b,c \neq a} (\mathbf{h}_a \cdot \mathbf{h}_b)(\mathbf{h}_a \cdot \mathbf{h}_c) \chi_{\setminus a}^{bc}$ can be replaced by $\sum_{b,c \neq a} (\mathbf{h}_b) \cdot (\mathbf{h}_c) \chi_{\setminus a}^{bc} / M$, thanks to the self-averaging of $(\mathbf{h}_a)_i (\mathbf{h}_a)_j$. Using Eq. (4.15) for the $(N-1, M)$ system,

$$\sum_{b,c \neq a} \mathbf{h}_b \cdot \mathbf{h}_c \chi_{\setminus a}^{bc} = \frac{M\sigma^2 \bar{\chi}_{\setminus a}}{\alpha\sigma^2 + \bar{\chi}_{\setminus a}} \approx \frac{M\sigma^2 \bar{\chi}}{\alpha\sigma^2 + \bar{\chi}} \quad (4.24)$$

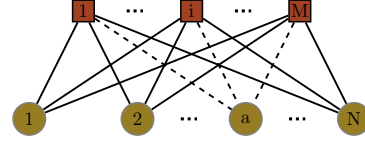
with the last step having to do with $\bar{\chi}$ becoming independent of N, M asymptotically. Using this last relation Eq. (4.24), in Eq. (4.23) we get

$$\sigma_{\text{eff}}^2 = \sigma^2 + \frac{\bar{\chi}}{\alpha} \quad (4.25)$$

The first step of cavity is summarized as the following.

Cavity Step 1

- Characterize the distribution of $\sum_{b \neq a} \mathbf{h}_b u_b$ in a “cavity system” of $N - 1$ variable nodes obtained by removing node u_a from the system, thereby leaving “cavity”.



- This optimization becomes equivalent to the minimization of a system with the cavity and Onsager term.

$$\min_{u_a} \left\{ \frac{1}{2\sigma^2} u_a^2 + U(u_a + x_{0a}) - \frac{1}{2\sigma^2} \left(\frac{\bar{\chi}}{\alpha\sigma^2 + \bar{\chi}} \right) u_a^2 + \frac{1}{\sigma^2} u_a \mathbf{h}_a \cdot \sum_{b \neq a} \mathbf{h}_b \hat{u}_b + \mathcal{E}_{(N-1, M)} \right\}$$

- Onsager term $\frac{1}{2\sigma^2} \left(\frac{\bar{\chi}}{\alpha\sigma^2 + \bar{\chi}} \right) u_a^2$ is appeared as a reaction term toward the variable u_a due to the adjustment of the other nodes after optimizing over them and holding u_a .

Looking at Eq. (4.22), the node variable u_a is coupled to the rest of the system, via the vector $\mathbf{l} = \sum_{b \neq a} \mathbf{h}_b \hat{u}_b$, which is independent of \mathbf{h}_a and depends solely on the $(N - 1, M)$ system. Since u_a is coupled to $\mathbf{h}_a \cdot \mathbf{l}$, we need to know the moments of this dot product. Because \mathbf{l} is independent of \mathbf{h}_a , for mean, we have

$$[\mathbf{h}_a \cdot \mathbf{l}]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}} = [\mathbf{h}_a]_{\mathbf{H}}^{\text{av}} \cdot [\mathbf{l}]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}} = 0 \quad (4.26)$$

and for variance.

$$\begin{aligned} [(\mathbf{h}_a \cdot \mathbf{l})^2]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}} &= \sum_{ij} [H_{ia} H_{ja}]_{\mathbf{H}}^{\text{av}} [v_i v_j]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}} \\ &= \frac{1}{M} \sum_i [v_i^2]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}}. \end{aligned} \quad (4.27)$$

The order k cumulants go as $M^{1-k/2}$ and, for $k > 2$, they tend to zero as M goes to

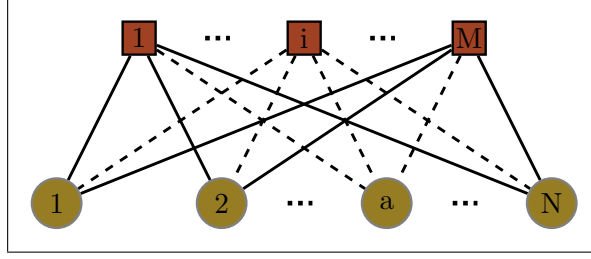


Figure 4.6: The $(N - 1, M - 1)$ cavity system. Node a and constraint i have been removed from the system by removing the links to them.

infinity. Therefore, we will stop with the variance and treat $\mathbf{h}_a \cdot \mathbf{l}$ as a zero-mean Gaussian variable. We still need the variance, for which we need a condition to determine $[v_i^2]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}}$. This requires a second step of the cavity method.

4.3 Zero Temperature Cavity Method: Removing a Constraint Node

The subtlety in determining $[v_i^2]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}}$ involves accounting for correlation between matrix elements H_{ib} and the optimal values \hat{u}_b of the $(N - 1, M)$ system. To do this, we need to set up an $(N - 1, M - 1)$ system with the constraint ‘ i ’ removed. (see Fig. 4.3).

Such a two-stage cavity method has been used in the context of Hopfield neural networks before [14].

To find $v_i = \sum_{b \neq a} H_{ib} \hat{u}_b$, we break up the minimization over $\mathbf{u}_{\setminus a}$ into two steps:

$$\min_{\mathbf{u}_{\setminus a}} \mathcal{E}_{\setminus a}(\mathbf{u}_{\setminus a}) = \min_{v_i} \left\{ \min_{\substack{\mathbf{u}_{\setminus a} \\ \text{s.t. } \sum_{b \neq a} H_{ib} u_b = v_i}} \{ \mathcal{E}_{\setminus ai}(\mathbf{u}_{\setminus a}) \} + \frac{1}{2\sigma^2} v_i^2 \right\} \quad (4.28)$$

the first minimization being a constrained one for the $(N - 1, M - 1)$ system, subject to $\sum_{b \neq a} H_{ib} u_b = v_i$, and the second one being over v_i . The cost function for the system without nodes a, i is represented by $\mathcal{E}(\mathbf{u}_{\setminus a})_{\setminus i}$. The term $\frac{1}{2\sigma^2} v_i^2$ represents the constraint coming from the i -th observation. Had we done an unconstrained optimization of $\mathcal{E}_{\setminus ai}(\mathbf{u}_{\setminus a})$, the optimum $\hat{\mathbf{u}}_{\setminus a}$ would be independent of H_{ib} . Trying to keep v_i small perturbs this solution

by a small amount and induces correlation with H_{ib} . Our strategy would be to compute the effect of perturbation in terms of the system susceptibility. This part is summarized in the following.

Cavity Step 2

- Now that u_a is coupled to $\sum_{b \neq a} h_b \hat{u}_b$, one needs to characterize the distribution of components of this vector: $v_i = \sum_{b \neq a} H_{ib} \hat{u}_b$
- v_i is related to constraint i .
- To figure out the distribution of v_i , we should extend the optimization of the system with one further cavity step by removing the constraint i .
- This leads to the optimization over

$$\min_{v_i} \left\{ \min_{\text{s.t. } \sum_{b \neq a} H_{ib} u_b = v_i} \mathcal{E}_{(N-1, M-1)} + \frac{1}{2\sigma^2} v_i^2 \right\}$$

In order to do constrained minimization, we use the Lagrange multiplier method

$$\begin{aligned} \min_{\mathbf{u}_{\setminus a}} \mathcal{E}_{\setminus a}(\mathbf{u}_{\setminus a}) = & \max_{\gamma_i} \min_{\mathbf{u}_{\setminus a}, v_i} \{ \mathcal{E}_{\setminus ai}(\mathbf{u}_{\setminus a}) \\ & + \frac{1}{2\sigma^2} v_i^2 - \gamma_i (v_i - \sum_{b \neq a} H_{ib} u_b) \}. \end{aligned} \quad (4.29)$$

Minimizing Eq. (4.29) with respect to v_i we get $v_i = \sigma^2 \gamma_i$, and making that substitution for v_i into the cost function we get

$$\min_{\mathbf{u}_{\setminus a}} \mathcal{E}_{\setminus a}(\mathbf{u}_{\setminus a}) = \max_{\gamma_i} \min_{\mathbf{u}_{\setminus a}} \{ \mathcal{E}_{\setminus a}(\mathbf{u}_{\setminus a}) - \frac{1}{2} \sigma^2 \gamma_i^2 - \mathbf{u}_{\setminus a}^T \mathbf{g} \} \quad (4.30)$$

$$= \max_{\gamma_i} \{ -\frac{1}{2} \sigma^2 \gamma_i^2 + \mathcal{E}_{\setminus i}^*(\mathbf{g}) \} \quad (4.31)$$

with $g_b = -\gamma_i H_{ib}$ and with $\mathcal{E}_{\setminus i}^*(\mathbf{g})$ is defined as

$$\mathcal{E}_{\setminus i}^*(\mathbf{g}) = \min_{\mathbf{u}_{\setminus a}} \{ \mathcal{E}_{\setminus ai}(\mathbf{u}_{\setminus a}) - \mathbf{u}_{\setminus a}^T \mathbf{g} \} \quad (4.32)$$

where the presence of \mathbf{g} alters the optimal $\mathbf{u}_{\setminus a}$ from the unconstrained optimum $\mathbf{u}'_{\setminus a}$. Since each component of \mathbf{g} is small ($O(1/\sqrt{M})$), we can expand around $\mathbf{u}'_{\setminus a}$ using $\chi_{\setminus ai}$, the susceptibility of the $(N-1, M-1)$ system, as in Eq. (4.6). Therefore, $\mathcal{E}_{\setminus i}^*(\mathbf{g})$ can be written as

$$\mathcal{E}_{\setminus i}^*(\mathbf{g}) = \mathcal{E}_{\setminus i}(\mathbf{u}'_{\setminus a}) - \mathbf{u}_{\setminus a}^T \mathbf{g} - \frac{1}{2} \mathbf{g}^T \chi_{\setminus ai} \mathbf{g} + \dots \quad (4.33)$$

Now, Eq. (4.31) becomes

$$\min_{\mathbf{u}_{\setminus a}} \mathcal{E}_{\setminus a}(\mathbf{u}_{\setminus a}) = \min_{\gamma_i} \left\{ -\frac{1}{2} \sigma^2 \gamma_i^2 + \mathcal{E}_{\setminus ai}(\mathbf{u}'_{\setminus a}) - \mathbf{u}_{\setminus a}^T \mathbf{g} - \frac{1}{2} \mathbf{g}^T \chi_{\setminus ai} \mathbf{g} \right\} \quad (4.34)$$

The quadratic term $\mathbf{g}^T \chi_{\setminus ai} \mathbf{g} = \gamma_i^2 \sum_{ij} H_{ib} H_{ic} \chi_{\setminus ai}^{bc}$ can be simplified because of self-averaging. We have

$$\sum_{ij} [H_{ib} H_{ic}]_{\mathbf{H}}^{\text{av}} \chi_{\setminus ai}^{bc} = \frac{1}{M} \sum_b \chi_{\setminus ai}^{bb} \approx \frac{\bar{\chi}}{\alpha}, \quad (4.35)$$

once more using the fact that the average local susceptibility $\bar{\chi}$ is nearly the same for the (N, M) system and the $(N-1, M-1)$ system.

Putting everything together

$$\min_{\mathbf{u}_{\setminus a}} \mathcal{E}_{\setminus a}(\mathbf{u}_{\setminus a}) = \max_{\gamma_i} \left\{ -\frac{\sigma^2}{2} \left(1 + \frac{\bar{\chi}}{\alpha \sigma^2} \right) \gamma_i^2 + \gamma_i \sum_{b \neq a} H_{ib} u'_b \right\}, \quad (4.36)$$

maximizing with respect to γ_i and then using $v_i = \sigma^2 \gamma_i$ gives us

$$v_i = \frac{1}{1 + \frac{\bar{\chi}}{\alpha \sigma^2}} \sum_{b \neq a} H_{ib} u'_b. \quad (4.37)$$

The denominator $(1 + \frac{\bar{\chi}}{\alpha \sigma^2})$ ‘scales down’ the unconstrained answer $\sum_{b \neq a} H_{ib} u'_b$. It is the same

factor that relates σ^2 to σ_{eff}^2 . The summary of the rest of the second cavity step is given below.

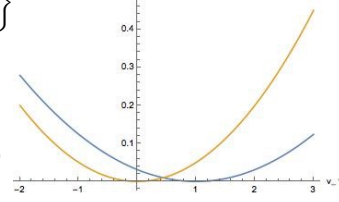
Cavity Step 2 Cont.

- One can show that this optimization is equivalent to

$$\min_{v_i} \left\{ \frac{\alpha}{2\bar{\chi}} (v_i - v'_i)^2 + \frac{1}{2\sigma^2} v_i^2 \right\}$$

where $\min_{\substack{\text{s.t. } \sum_{b \neq a} H_{ib} u_b = v_i}} \mathcal{E}_{(N-1, M-1)}$

is optimized at $v'_i = \sum_{b \neq a} H_{ib} u'_b$ (yellow curve) and $\frac{1}{2\sigma^2} v_i^2$ at zero (blue curve).



- Therefore, $v_i = \frac{1}{1 + \frac{\bar{\chi}}{\alpha\sigma^2}} v'_i$
- Since in this system of (N-1, M-1), u'_b , even though can have non-trivial correlation with each other, are independent of H_{ib} , we can compute their moments.

Given that this result is true for any i 's, (4.22) becomes

$$\min_{\mathbf{u}} \mathcal{E}(\mathbf{u}) = \min_{u_a} \left\{ \frac{1}{2\sigma_{\text{eff}}^2} u_a^2 - \frac{\xi}{\sigma^2(1 + \frac{\bar{\chi}}{\alpha\sigma^2})} u_a + U(u_a + x_{0a}) \right\} \quad (4.38)$$

with

$$\xi \equiv - \sum_i H_{ia} \sum_{b \neq a} H_{ib} u'_b \quad (4.39)$$

being a random Gaussian variable with mean zero and variance

$$\begin{aligned}
\sigma_\xi^2 &\equiv [\xi^2]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}} = \sum_{i,j} [H_{ia} H_{ja}]_{\mathbf{H}}^{\text{av}} \sum_{b,c \neq a} [H_{ib} H_{jc}]_{\mathbf{H}}^{\text{av}} [u'_b u'_c]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}} \\
&= \sum_{i,j} \frac{\delta_{ij}}{M} \sum_{b,c \neq a} \frac{\delta_{ij} \delta_{bc}}{M} [u'_b u'_c]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}} \\
&= \frac{1}{M} \sum_{b \neq a} [u_b'^2]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}} = \frac{q}{\alpha}
\end{aligned} \tag{4.40}$$

thanks to H_{ja} and H_{ib} being independent for $a \neq b$, as well as u'_b 's being independent of those matrix elements. The quantity $q \equiv \frac{1}{N-1} \sum_{b,c \neq a} [u'^2]_{\mathbf{x}_0, \mathbf{H}}^{\text{av}}$ is the MSE for the $(N-1, M-1)$ system. Insisting that q is also the MSE of the (N, M) system is one of the self-consistency conditions.

In summary, the zero temperature problem boils down to a collection of independent single variable optimization

$$\min_{u_a} \left\{ \frac{1}{2\sigma_{\text{eff}}^2} (u_a^2 - 2\xi_a u_a) + U(u_a + x_{0a}) \right\} \tag{4.41}$$

which has the same effective cost function as Eq. (3.24) in Sec. 3.2. The variables ξ_a are chosen independently from $\mathcal{N}(0, \sigma_\xi^2)$. With x_{0a}, ξ_a chosen randomly, we can calculate a distribution of u_a and χ^{aa} . The self-consistency require u_a^2 average q to satisfy $\sigma_\xi^2 = q/\alpha$ and χ^{aa} average $\bar{\chi}$ to be related to σ_{eff}^2 by $\sigma_{\text{eff}}^2 = \sigma^2 + \bar{\chi}/\alpha$.

For the future convenience, we put these together in the following proposition:

Proposition 1 (Effective Individual Optimization)

$$\hat{u}_a = \min_{u_a} \left\{ \frac{1}{2\sigma_{\text{eff}}^2} (u_a^2 - 2\xi_a u_a) + U(u_a + x_{0a}) \right\} \quad (4.42)$$

$$\xi_a \in \mathcal{N}(\xi; 0, \sigma_\xi^2) \text{ with } \sigma_\xi^2 \equiv \sigma_\zeta^2 + \frac{q}{\alpha} \quad (4.43)$$

$$q \equiv \sum_a [\hat{u}_a^2]_{x_0, \xi}^{\text{av}} \quad (4.44)$$

$$\sigma_{\text{eff}}^2 \equiv \sigma^2 + \frac{\bar{\chi}}{\alpha} \quad (4.45)$$

$$\bar{\chi} \equiv \frac{1}{N} \sum_a \chi^{aa} \quad (4.46)$$

The quantity q is the sum of the squared of error residuals, i.e. MSE. In addition, local susceptibility is obtained via $\hat{u}_a(f) - \hat{u}_a(0) = \chi^{aa} f_a$ with $f_a \rightarrow 0$ and $\hat{u}_a(f)$ is carried out by minimizing $\min_{u_a} \left\{ \frac{1}{2\sigma_{\text{eff}}^2} (u_a^2 - 2\xi_a u_a) + U(u_a + x_{0a}) - f_a u_a \right\}$. In the end, summing over χ^{aa} 's for all the instances of measurement matrix and then taking average over all nodes yields to the average local susceptibility, $\bar{\chi}$, and thus σ_{eff}^2 .

In this new formulation, we do not need to invoke temperature. One can consider $\bar{\chi}$ to be equivalent to the $\beta\Delta Q$ in the replica approach. As we will see in the next chapter, we could use $\bar{\chi}$ to distinguish phases around the zero-temperature transition described by Donoho and Tanner [7].

Chapter 5

Phase Transition in Sparse Reconstruction

In the previous chapter, we adapted the zero-temperature cavity method in two steps to derive self-consistency conditions on MSE when the regularization term $V(\mathbf{x})$ is such that a unique solution exists. This led to a considerable simplification by utilizing the fact that the system of variables are fully connected and the so-called local susceptibility matrix χ plays a key role in the system. In particular, in the asymptotic limit of large M and N , we showed that the optimization of Eq. (4.1) breaks down into a collection of effectively independent optimization as the followings. As we will see in this chapter, we solve this effective individual optimization (see Proposition 1) for the penalty function of the form $\lambda|x|^q$ with $q = 1, 2$ and the combination of ℓ_1 and ℓ_2 norms, i.e. $V(\mathbf{x}) = \lambda_1 \|\mathbf{x}\|_1 + \frac{\lambda_2}{2} \|\mathbf{x}\|_2^2$, known as Elastic Net [55]. We point out the importance of $\bar{\chi}$ over $\beta\Delta Q$, obtained by [8, 9] with the replica approach, to distinguish phases around the zero-temperature transition.

To facilitate such adoption, we summarize the symbols used in the next section in the Table 5.

5.1 Ridge Regression

We start by considering the simplest form of regularization with $U(x) = \frac{\lambda}{2}x^2$, a penalty function that does not impose sparsity on the solutions. This is just a noise-free ridge regression with Tikhonov regularization [56]

$$\hat{\mathbf{x}}(\vartheta = \lambda\sigma^2) = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2} (\mathbf{H}(\mathbf{x} - \mathbf{x}_0))^2 + \frac{\lambda}{2} \mathbf{x}^2 \right\}. \quad (5.1)$$

Symbols	
Symbol	Description
u_a	Measure of residual error $x_a - x_{0a}$
q	Mean squared error (MSE)
α	Measure for the number of constraints, $\frac{M}{N}$
ρ	Measure for the sparsity, $\frac{K}{N}$
λ_1	ℓ_1 -norm regression coefficient
λ_2	ℓ_2 -norm regression coefficient
σ^2	Error variance on the constraint $\mathbf{y} = \mathbf{H}\mathbf{x}$
ϑ	$\lambda\sigma^2$
σ_{eff}^2	Effective σ^2 given in the asymptotic limit of large M, N
θ	$\lambda\sigma_{\text{eff}}^2$
σ_{ξ}^2	$\frac{q}{\alpha}$
σ_{ζ}^2	Variance of external noise
τ	$\frac{\theta}{\sigma_{\xi}}$

Table 5.1: This table presents input parameters used to explain effective individual optimization

We could explicitly minimize \mathbf{x} and proceed with our analysis using random matrix theory; however, we will apply first the self-consistency formalism we have developed (Proposition 1).

$$\min_u \left\{ \frac{1}{2\sigma_{\text{eff}}^2} (u^2 - 2\xi u) + \frac{\lambda}{2} (u + x_0)^2 \right\} \quad (5.2)$$

Recalling that $u = x - x_0$ and identifying $\theta = \lambda\sigma_{\text{eff}}^2$, minimization of Eq. (5.2) gives

$$x = \frac{x_0 + \xi}{1 + \theta} \quad (5.3)$$

This result can be used to determine σ_{ξ}^2 in Eq. (4.43)

$$\sigma_{\xi}^2 = \frac{q}{\alpha} = \frac{1}{\alpha} [u^2]_{x_0, \xi}^{\text{av}} = \frac{\sigma_{\xi}^2 + \theta^2 \rho [x_0^2]_{x_0}^{\text{av}}}{\alpha(1 + \theta)^2}. \quad (5.4)$$

where $[\dots]_{x_0}^{\text{av}}$ means average over $\pi(x_0)$. On the other hand, to determine $\theta (= \lambda\sigma_{\text{eff}}^2)$, we look at the local susceptibility in Proposition 1. One can see that with the ridge regression

penalty function, local susceptibility is the same everywhere:

$$\bar{\chi} = \left[\lambda + \frac{1}{\sigma^2 + \frac{\bar{\chi}}{\alpha}} \right]^{-1} \implies \theta = \left(\frac{1}{\lambda \bar{\chi}} - 1 \right)^{-1}. \quad (5.5)$$

In particular in the $\vartheta \rightarrow 0$ limit, i.e. the minimal ℓ_2 norm subject to linear constraints $\mathbf{H}\mathbf{x} = \mathbf{H}\mathbf{x}_0$, with $\lambda \bar{\chi} = 1 - \alpha$ gives $\theta = \alpha^{-1} - 1$. With the knowledge of θ , the Eqs. (5.3), and (5.4) lead us to

$$x = \alpha(x_0 + \xi) \quad (5.6)$$

$$\sigma_\xi^2 = \frac{(1 - \alpha)\rho}{\alpha} [x_0^2]_{x_0}^{\text{av}}. \quad (5.7)$$

Remark 2 *The estimated x can be seen as a Gaussian variable, with αx_0 as its mean and $(1 - \alpha)\alpha\rho[x_0^2]_{x_0}^{\text{av}}$ as its variance. When $x_0 = 0$, we expect the fluctuation of x around zero to be of the order $((1 - \alpha)\alpha\rho[x_0^2]_{x_0}^{\text{av}})^{1/2}$. We could have a simple threshold θ so that if $|x| < \theta$ we set x_0 to zero. We could compute the false positive and false negative rates of such a procedure. When $\rho \ll \alpha/(1 - \alpha)$, it is possible to choose a threshold θ such that $((1 - \alpha)\alpha\rho[x_0^2]_{x_0}^{\text{av}})^{1/2} \ll \theta \ll \alpha([x_0^2]_{x_0}^{\text{av}})^{1/2}$. With such a threshold, both error rates would be small.*

For the sake of completeness, we derive the above Eqs. (5.6) and (5.7) from a formal singular value decomposition point of view. Elementary derivation leads us to an explicit expression:

$$\hat{\mathbf{x}} = \frac{\mathbf{H}^T \mathbf{H}}{\sigma^2} \left[\frac{\mathbf{H}^T \mathbf{H}}{\sigma^2} + \lambda \mathbf{I}_N \right]^{-1} \mathbf{x}_0 = \sum_{i=1}^M \frac{s_i^2}{s_i^2 + \lambda \sigma^2} \mathcal{V}_i (\mathcal{V}_i^T \mathbf{x}_0). \quad (5.8)$$

where we use the singular vector basis of the matrix \mathbf{H} , with \mathbf{s}_i being the non-zero singular values, and \mathcal{V}_i the corresponding right singular vectors. When we take the limit of vanishing

σ^2 , we just have a projection of the N dimensional vector \mathbf{x}_0 to an M -dimensional projection spanned by \mathcal{V}_i 's. In other words

$$x_a = \sum_{b=1}^N \sum_{i=1}^M \mathcal{V}_{ia} \mathcal{V}_{ib} x_{0a} = \sum_{a=1}^N P_{ab} x_{0a} \quad (5.9)$$

\mathbf{P} being the projection matrix. For random \mathbf{H} , \mathcal{V}_i 's are just a random choice of M orthonormal vectors. Thus, the properties of the estimate depends on the statistics of the projection matrix to a random M -dimensional subspace.

$$[P_{ab}]_{\mathbf{H}}^{\text{av}} = \sum_{i=1}^M [\mathcal{V}_{ia} \mathcal{V}_{ib}]_{\mathbf{H}}^{\text{av}} = \sum_{i=1}^M \frac{\delta_{ab}}{N} = \alpha \delta_{ab} \implies [\hat{x}_a]_{\mathbf{H}}^{\text{av}} = \alpha x_{0a} \quad (5.10)$$

For variance, we need to think of second order moments of the matrix elements of \mathbf{P} , particularly, $[P_{ab}P_{ac}]_{\mathbf{H}}^{\text{av}}$. We could parametrize $[P_{ab}P_{ac}]_{\mathbf{H}}^{\text{av}} = A\delta_{bc} + B\delta_{ab}\delta_{bc}$. Since \mathbf{P} is a projection operator, $\mathbf{P}^2 = \mathbf{P}$ and it is a symmetric matrix. Hence,

$$\sum_a [P_{ab}P_{ac}]_{\mathbf{H}}^{\text{av}} = \sum_a [P_{ba}P_{ac}]_{\mathbf{H}}^{\text{av}} = [P_{bc}]_{\mathbf{H}}^{\text{av}} = \alpha \delta_{bc}. \quad (5.11)$$

In the limit of $M, N \rightarrow 0$ with α fixed, the distribution of P_{aa} gets highly concentrated around the mean α . As a result,

$$[P_{aa}P_{aa}]_{\mathbf{H}}^{\text{av}} \approx \left(\sum_a [P_{aa}]_{\mathbf{H}}^{\text{av}} \right)^2 = \alpha^2. \quad (5.12)$$

Using the two constraints, represented by Eqs. (5.11) and (5.12), we can determine A and B , in the large M, N limit, leading to,

$$[P_{ab}P_{ac}]_{\mathbf{H}}^{\text{av}} \approx \frac{\alpha(1-\alpha)}{N} \delta_{bc} + \alpha^2 \delta_{ab}\delta_{bc}. \quad (5.13)$$

The variance is now given by,

$$\begin{aligned}
& [\hat{x}_{0a}^2]_{\mathbf{H}}^{\text{av}} - ([\hat{x}_{0a}]_{\mathbf{H}}^{\text{av}})^2 \\
&= \sum_a \left[\frac{\alpha(1-\alpha)}{N} \delta_{bc} + \alpha^2 \delta_{ab} \delta_{bc} \right] x_{0b} x_{0c} - (\alpha x_{0a})^2 \\
&= (1-\alpha) \alpha \rho [x_0^2]_{x_0}^{\text{av}}
\end{aligned} \tag{5.14}$$

which is the same conclusion from a formal singular value decomposition point of view.

5.2 Basis Pursuit: ℓ_1 -norm Minimization

In this section, we reconsider the much-analyzed case where the penalty function is the ℓ_1 norm of \mathbf{x} [2, 7]. The reconstructed sparse solution is given by

$$\hat{\mathbf{x}}(\vartheta) = \min_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2} (\mathbf{H}(\mathbf{x} - \mathbf{x}_0))^2 + \lambda \|\mathbf{x}\|_1 \right\}. \tag{5.15}$$

Like in the case of ridge regression, we aim to solve the above optimization problem self-consistently. To determine θ , once again we look at the local susceptibility in proposition 1. In this case $U''(x)$ is zero everywhere except at $x = 0$, where it is formally infinite. Consequently,

$$\begin{aligned}
\chi^{aa} &= 0, \text{ if } x_a = 0 \\
\chi^{aa} &= \sigma_{\text{eff}}^2, \text{ otherwise.}
\end{aligned} \tag{5.16}$$

We define $\hat{\rho}$ to be the estimated sparsity, i.e. the fraction of x_a 's that are non-zero. Therefore

$$\bar{\chi} = \hat{\rho} \sigma_{\text{eff}}^2 \quad (\lambda \bar{\chi} = \hat{\rho} \theta) \text{ and}$$

$$\sigma_{\text{eff}}^2 = \sigma^2 + \frac{\bar{\chi}}{\alpha} = \sigma^2 + \frac{\hat{\rho} \sigma_{\text{eff}}^2}{\alpha} \tag{5.17}$$

implying

$$\theta \left(1 - \frac{\hat{\rho}}{\alpha}\right) = \vartheta \tag{5.18}$$

Remark 3 *The equation $\theta(1 - \frac{\hat{\rho}}{\alpha}) = \vartheta$ is central to understanding the $\vartheta \rightarrow 0$ limit and the associated phase transition. When ϑ goes to zero, we either have $\theta = 0$ ($\hat{\rho} \neq \alpha$) or $\hat{\rho} = \alpha$ ($\theta \neq 0$). These two conditions correspond to the two phases of the system, the first being the perfect reconstruction phase and the second, the non-zero error regime. In terms of average local susceptibility, the first phase has $\bar{\chi} = \hat{\rho}\theta = 0$, while the second one has $\bar{\chi} \neq 0$.*

Now we can set up the notation for the single variable optimization problem to find the value for σ_ξ^2 (\propto MSE) in these two regimes. More precisely, by searching for the solutions to

$$\min_u \left\{ \frac{1}{2\sigma_{\text{eff}}^2} (u^2 - 2\xi u) + \lambda |u + x_0| \right\} \quad (5.19)$$

we arrive at the following soft-thresholding function for the estimated value of \hat{x} that we will denote by $\eta_{\text{soft}}(t; \theta)$, with the variable $t = x_0 + \xi$ (illustrated in Fig. a).

Definition 11 (Soft Thresholding Function)

$$\eta_{\text{soft}}(t; \theta) = \begin{cases} t - \theta & \text{if } \theta \leq t, \\ 0 & \text{if } -\theta \leq t \leq \theta, \\ t + \theta & \text{if } t < -\theta. \end{cases} \quad (5.20)$$

According to remark 3, the perfect reconstruction regime which ends to the phase boundary from above is the case where, as ϑ becomes small, θ becomes small as well. From Eq. (5.20), there are three sources of error that can contribute to σ_ξ^2 in this regime:

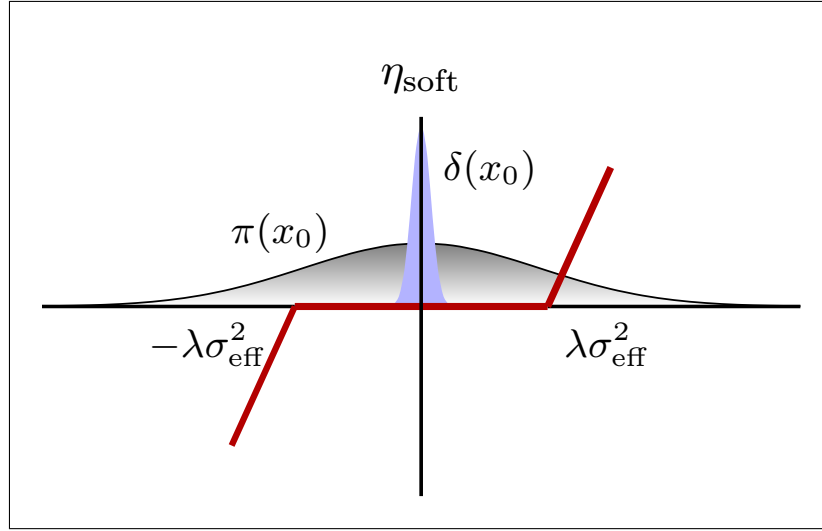


Figure 5.1: The soft thresholding function (in red) defined in (5.20). The non-zero entries of the sparse vector \mathbf{x}_0 drawn from random distribution is represented by π (in grey) and the zero components are represented by delta function (in blue).

a) $(x_0 \neq 0 \rightarrow \hat{x} = 0)$

Here x_0 was initially non-zero, but the estimated \hat{x} , due to the shift by ξ , has fallen into the $[-\theta, \theta]$ interval and then been truncated to zero. One can see that since θ is small, the probability of this event can be ignored for the time being¹.

b) $(x_0 \neq 0 \rightarrow \hat{x} \neq x_0)$

For non-zero x_0 that does not get set to zero, the contribution to MSE is

$$\rho[(\hat{x} - x_0)^2]_{x_0, \xi}^{\text{av}} = \rho[(\xi - \theta \text{sgn}(\hat{x}))^2]_{x_0, \xi}^{\text{av}} = \rho(\sigma_\xi^2 + \theta^2) \quad (5.21)$$

c) $(x_0 = 0 \rightarrow \hat{x} \neq 0)$

Another source of error is the event when the x_0 is zero but \hat{x} has fallen outside the interval $[-\theta, \theta]$ and has been estimated to be non-zero. In this case, the contribution to

¹Under this circumstance, if ξ remains of order one, then the error is dominated by ξ , i.e. $q(\text{MSE}) = \sigma_\xi^2$. However, this is not consistent with $\sigma_\xi^2 = q/\alpha$, unless $\sigma_\xi^2 = 0$. Hence in this regime, we need to consider a σ_ξ^2 that is comparable to θ . Therefore, as $\vartheta \rightarrow 0$, we will have $\sigma_\xi^2 \rightarrow 0$ and $q \rightarrow 0$, making the reconstruction perfect, i.e. the limit when $\vartheta, \theta, \sigma_\xi^2 \rightarrow 0$ with $\frac{\theta}{\sigma_\xi}$ of order one.

MSE is

$$\begin{aligned}
(1 - \rho)[\hat{x}^2]_{x_0, \xi}^{\text{av}} &= 2(1 - \rho) \int_{\theta}^{\infty} d\xi \frac{1}{\sqrt{2\pi\sigma_{\xi}^2}} e^{-\frac{\xi^2}{2\sigma_{\xi}^2}} (\xi - \theta)^2 \\
&= 2\sigma_{\xi}^2(1 - \rho) \{(1 + \tau^2)\Phi(\tau) - \tau\phi(\tau)\}
\end{aligned} \tag{5.22}$$

with $\tau = \frac{\theta}{\sigma_{\xi}}$. Adding up these contributions from Eq. (5.21) and (5.22), we get the total MSE, q (i.e. $\alpha\sigma_{\xi}^2$). Therefore using Eq. (4.43), $\sigma_{\xi}^2 = q/\alpha$, and the knowledge of $\theta = 0$ lead to the first parametric expression for the perfect reconstruction phase:

$$\alpha = 2(1 - \rho) \{(1 + \tau^2)\Phi(\tau) - \tau\phi(\tau)\} + \rho(1 + \tau^2). \tag{5.23}$$

To determine $\hat{\rho}$, one can notice that if $x_0 = 0$, we have to have $|\xi| > \theta$ to lead to a non-zero x . On the other hand, since θ is small, a non-zero x_0 remains non-zero with probability approaching one. Counting all sources of the non-zero \hat{x} 's, then we have ²

$$\hat{\rho} = 2(1 - \rho)\Phi(\tau) + \rho. \tag{5.24}$$

Recall that in the error-prone phase $\hat{\rho} = \alpha$ (Remark 3). This is due to the fact that q, σ_{ξ}^2 and therefore θ need to be non-zero in this regime. If the transition happens continuously, the condition for the phase boundary is $\alpha = \hat{\rho} = 2(1 - \rho)\Phi(\tau) + \rho$. Hence the relation between α and ρ at the phase boundary is obtained by solving and eliminating τ from

$$\alpha = 2(1 - \rho) \{(1 + \tau^2)\Phi(\tau) - \tau\phi(\tau)\} + \rho(1 + \tau^2) \tag{5.25}$$

$$\alpha = 2(1 - \rho)\Phi(\tau) + \rho \tag{5.26}$$

Alternatively, Eq. (5.25) and (5.26) can be solved for α and ρ at the phase boundary and

²Note that $\hat{\rho} > \rho$, even in the perfect reconstruction phase. That is because a fraction of x_a 's remain non-zero as long as $\vartheta > 0$, and vanish only in the $\vartheta \rightarrow 0$ limit.

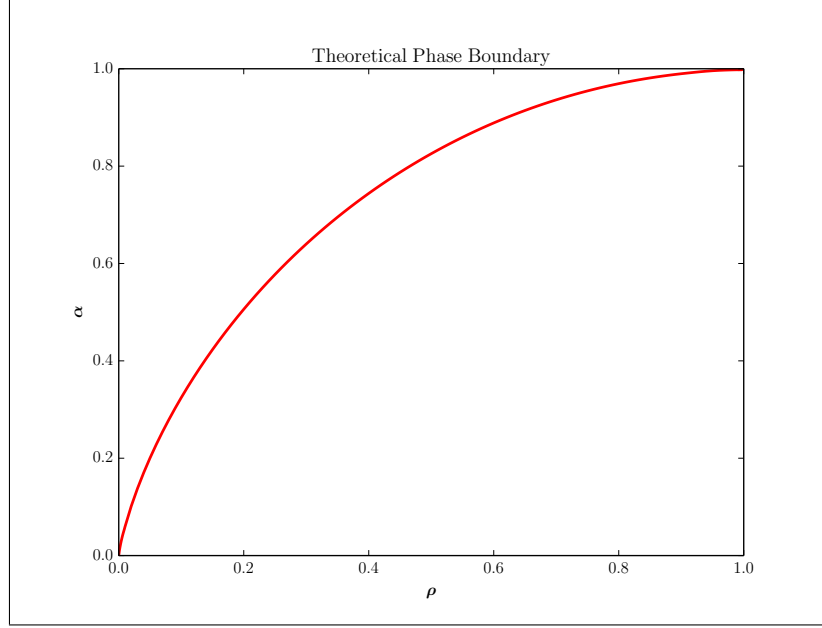


Figure 5.2: The red curve is the theoretical phase boundary obtained by solving Eq. (5.25) and (5.26). As $\rho \rightarrow 0$ this boundary is of the form $\alpha_c(\rho) = 2\rho \log(\frac{1}{\rho})$ shown by the blue curve. The recovery occurs above the curves.

expressed parametrically as a function of τ :

$$\alpha = \frac{2\phi(\tau)}{\tau + 2(\phi(\tau) - \tau\Phi(\tau))} \quad (5.27)$$

$$\rho/\alpha = 1 - \frac{\tau\Phi(\tau)}{\phi(\tau)} \quad (5.28)$$

This leads to the phase diagram showing the transition from absolute success to absolute failure depicted in Fig. 5.2

5.3 Understanding the Extremely Sparse Limit

In the extremely sparse limit, $\rho \ll 1$, one can obtain a more explicit asymptotic relation between α and ρ . In this limit τ is large, and the dominant contributions are the second term, $\rho(1 + \tau^2)$, from Eq. (5.25) and the first term, $2(1 - \rho)\Phi(\tau)$, from Eq. (5.26).

Consequently,

$$\begin{aligned}\alpha &\approx \sqrt{\frac{2}{\pi} \frac{e^{-\frac{\tau^2}{2}}}{\tau}} \\ \rho &\approx \sqrt{\frac{2}{\pi} \frac{e^{-\frac{\tau^2}{2}}}{\tau^3}}.\end{aligned}\tag{5.29}$$

Therefore, in this extreme-sparsity limit, we have $\alpha_c(\rho) \sim 2\rho \log \frac{1}{\rho}$, a result that has been known for some time and comes close to the bounds from the restricted isometry property [6].

We can look at the above example as an approximate but a simpler picture of the transition as we go to the very sparse limit. We saw that for a single site, an effective random noise, ξ , arises which really comes from the error in the estimation of all the other variables. This error could be associated with the error in estimating genuine non-zero x_0 or with x_0 that are zero and has become non-zero.

The error from the non-zero variables comes from two sources: 1) those variables affected by this effective noise 2) and those from the shrinkage. Its clear that in the very sparse limit near the transition, the dominant contribution to the error comes from the shrinkage of the non-zeros. Most of the non-zero x_0 are large enough that they don't get set to zero by small additive noise.

We also noticed that the fraction of $x_0 = 0$ that becomes non-zero contributes majorly to $\hat{\rho}$. However, although the number of them is large, they do not contribute in a big way to the error. This is because sum of those are still small, especially, in the good phase as θ goes to zero, the total sum goes to zero. But it turned out that the phase condition is dominated by the number of zero variables that are estimated to be non-zero via $\hat{\rho} = \alpha$.

In addition, in the next section we will see that even though the non-zero variables that has truncated to zero had little role in determining the position of the phase boundary, they are important in determining the critical behaviors and the universality classes near the transition boundary.

As a bonus, we use this point of view to figure out how correlations in the measurement matrices affect the location of the transition boundary. This is important because in real world application, the random measurement matrix must be replaced by a non-trivial correlated matrix that corresponds to the characteristics of feasible sensing device.

In particular, we look at the generalization of the Gaussian $\mathcal{P}(\mathbf{H})$ with the matrix elements correlated in a “factorized” manner, a special case of correlation that appears in many practical problems [50, 51, 57–60]. More precisely, for the mean and variance of $\mathcal{P}(\mathbf{H})$ we have

$$[H_{ia}]^{\text{av}} = 0 \quad (5.30)$$

$$[H_{ia}H_{jb}]^{\text{av}} = \frac{1}{M}C_{ij}D_{ab}. \quad (5.31)$$

We show in the appendix A that such zero-temperature optimization problem reduces to the minimization of the following equation:

Proposition 2 (Effective Optimization for the Correlated Matrices)

$$\min_{\mathbf{u}} \left\{ \frac{1}{2\sigma_{\text{eff}}^2} (\mathbf{u}^\top \mathbf{D} \mathbf{u} - 2\boldsymbol{\xi}^\top \mathbf{D} \mathbf{u}) + \lambda |\mathbf{u} + \mathbf{x}_0| - \mathbf{f} \cdot \mathbf{u} \right\} \quad (5.32)$$

with

$$\frac{1}{\sigma_{\text{eff}}^2} = \frac{1}{M} \text{Tr} \left[\mathbf{C} \left\{ \mathbf{I}_M \sigma^2 + \frac{\text{Tr}(\mathbf{D} \bar{\boldsymbol{\chi}})}{M} \mathbf{C} \right\}^{-1} \right] \quad (5.33)$$

Once more, $\bar{\boldsymbol{\chi}}$ can be computed from the relation $\delta \mathbf{u} = \boldsymbol{\chi} \mathbf{f}$ and the error is given by

$$q = \frac{1}{N} [\mathbf{u}^\top \mathbf{D} \mathbf{u}]_{\mathbf{x}_0, \boldsymbol{\xi}}^{\text{av}}. \quad (5.34)$$

Here, the Gaussian quenched vector ξ has mean zero and the covariance matrix

$$\text{Cov}(\xi, \xi) = \frac{q \mathbf{D}^{-1} \text{Tr} \left[\mathbf{C} \left\{ \mathbf{I}_M \sigma^2 + \frac{\text{Tr}(\mathbf{D} \bar{\chi})}{M} \mathbf{C} \right\}^{-1} \right]^2 / M}{\alpha \left(\text{Tr} \left[\mathbf{C} \left\{ \mathbf{I}_M \sigma^2 + \frac{\text{Tr}(\mathbf{D} \bar{\chi})}{M} \mathbf{C} \right\}^{-1} \right] / M \right)^2} \quad (5.35)$$

Considering the case of the full-rank matrices \mathbf{C}, \mathbf{D} and $\vartheta \rightarrow 0$ limit, we obtain the following relation from the Eqs. (5.33) and (5.35):

$$\sigma_{\text{eff}}^2 = \frac{1}{M} \text{Tr}(\mathbf{D} \bar{\chi}) \quad (5.36)$$

$$\text{Cov}(\xi, \xi) = \frac{q \mathbf{D}^{-1}}{\alpha} \quad (5.37)$$

Like we mentioned above, the major contribution to the error q is coming from the shrinkage of the non-zero variables. This can be obtained by minimizing Eq. (5.32) without invoking ξ . Thus we obtain

$$\mathbf{u} = \lambda \sigma_{\text{eff}}^2 \mathbf{D}^{-1} \text{sgn}(\mathbf{x}_0) \quad \Rightarrow \quad q \equiv \frac{1}{N} \mathbf{u}^{-1} \mathbf{D} \mathbf{u} = \rho \theta^2 \frac{\text{Tr} \mathbf{D}^{-1}}{N} \quad (5.38)$$

Moreover, by plugging Eq. (5.38) into the Eq. (5.37) we get

$$\text{Cov}(\xi, \xi) = \frac{\rho(\theta)^2}{\alpha} \mathbf{D}^{-1} \left(\frac{\text{Tr} \mathbf{D}^{-1}}{N} \right) \quad (5.39)$$

We also noticed that the fraction of $x_0 = 0$ that becomes non-zero contributes majorly to $\hat{\rho}$, and thus, to determine the position of the phase boundary. After minimizing Eq. (5.32), we get $\mathbf{D} \mathbf{u} = \mathbf{D} \xi \pm \theta$. This implies that if $x_0 = 0$, we should have $|(\mathbf{D} \xi)_a| > \theta$ to lead to a non-zero x . It is straightforward to show that the variance of $|\mathbf{D} \xi|$ equals $\sigma_{\xi}^2 \mathbf{D}_{aa}$. Therefore,

counting the zero variables that has become non-zero, we obtain

$$\alpha = \frac{2}{N} \sum_a \Phi_a(\tau/\sqrt{\mathbf{D}_{aa}}) \quad (5.40)$$

Moreover, in the extremely sparse limit, τ is large and Eq. (5.40) becomes:

$$\alpha = \sqrt{\frac{2}{\pi}} \frac{1}{N} \sum_a \frac{e^{-\frac{\tau^2}{2\mathbf{D}_{aa}}}}{\tau/\sqrt{\mathbf{D}_{aa}}} \quad (5.41)$$

Using Eqs. (5.41) and (5.38), in a particular case when all the D_{aa} 's are the same, we can arrive at the following asymptotic relation for the correlated measurement matrix

$$\alpha_c(\rho) \sim 2\rho \ln\left(\frac{1}{\rho}\right) \frac{\text{Tr} \mathbf{D}}{N} \frac{\text{Tr} \mathbf{D}^{-1}}{N} \quad (5.42)$$

5.4 Critical Exponents

To get a better understanding of the nature of this phase transition and characterizing its behavior as one decreases α from above $\alpha_c(\rho)$ to below, we should search for solutions of Eq. (5.43) and (5.44) in the error-prone regime where both θ and σ_ξ^2 remain $\mathcal{O}(1)$. In this case, we have to deal carefully with the possibility that \hat{x} has been set to zero, because $x_0 + \xi$ fell within $\pm\theta$. It is straightforward to show that the self-consistency equation for σ_ξ^2 becomes

$$\begin{aligned} \alpha &= \alpha \frac{\sigma_\xi^2}{\sigma_\xi^2} + 2(1-\rho) \{ (1+\tau^2)\Phi(\tau) - \tau\phi(\tau) \} \\ &+ \rho \left[\tau_0^2 \{ 1 - \Phi(\tau + \tau_0) - \Phi(\tau - \tau_0) \} \right. \\ &+ (1+\tau^2) \{ \Phi(\tau + \tau_0) + \Phi(\tau - \tau_0) \} \\ &\left. - (\tau - \tau_0)\phi(\tau + \tau_0) - (\tau + \tau_0)\phi(\tau - \tau_0) \right]_{x_0}^{\text{av}} \end{aligned} \quad (5.43)$$

where $[\dots]_{x_0}^{\text{av}}$ means average over $\pi(x_0)$ and $\tau_0 = \frac{x_0}{\sigma_\xi}$. The quantity τ and functions $\Phi(\tau)$ and $\phi(\tau)$ are defined as before. In addition, the parametric expression of Eq. (5.26) becomes

$$\alpha = \frac{\vartheta}{\theta} + 2(1 - \rho)\Phi(\tau) + \rho \left[\Phi(\tau + \tau_0) + \Phi(\tau - \tau_0) \right]_{x_0}^{\text{av}} \quad (5.44)$$

One should notice that in Eqs. (5.43), (5.44) we included extra terms $\alpha \frac{\sigma_\zeta^2}{\sigma_\xi^2}$ coming from the additive noise and $\frac{\vartheta}{\theta}$ from not setting ϑ to zero, respectively.

In order to better understand the behavior close to the transition where θ and σ_ξ^2 are small, we rewrite Eqs. (5.43) and (5.44) as ³

$$\alpha = \alpha \frac{\sigma_\zeta^2}{\sigma_\xi^2} + A_2(\rho, \tau) - \rho [\psi_\xi(\tau_0, \tau)]_{x_0}^{\text{av}} \quad (5.45)$$

$$\alpha = \alpha \frac{\vartheta}{\theta} + A_0(\rho, \tau) - \rho [\psi_\theta(\tau_0, \tau)]_{x_0}^{\text{av}} \quad (5.46)$$

where $\psi_\xi(\tau_0, \tau)$, $\psi_\theta(\tau_0, \tau)$ are even functions of τ_0 that falls off quickly as τ_0 becomes much larger than 1. we also introduced A_2 and A_0 as

$$\begin{aligned} A_2(\rho, \tau) &= 2(1 - \rho) \{ (1 + \tau^2)\Phi(\tau) - \tau\phi(\tau) \} + \rho(1 + \tau^2) \\ A_0(\rho, \tau) &= 2(1 - \rho)\Phi(\tau) + \rho \end{aligned} \quad (5.47)$$

Remark 4 *The transition boundary is where these two curves intersect at the point τ_c (see Fig. 5.4). Note that $\frac{dA_2}{d\tau} = 2\frac{A_2 - A_0}{\tau}$. Thus, at the transition point τ_c , $\frac{dA_2}{d\tau} = 0$, i.e. A_2 behaves like $\sim \delta\tau^2$ (A_0 behaves like $\sim -\delta\tau$). As we will see in section 5.5, this relation won't be valid for Elastic Net. Therefore, we expect to have different critical behavior near the transition point for Elastic Net than Basis Pursuit.*

³Note that, when $|\tau_0| = \frac{|x_0|}{\sigma_\xi} \rightarrow \infty$, $\Phi(\tau + \tau_0) + \Phi(\tau - \tau_0) \rightarrow 1$ and $(\tau - \tau_0)\phi(\tau + \tau_0), (\tau + \tau_0)\phi(\tau - \tau_0) \rightarrow 0$. The τ_0 dependent expression inside $[\dots]_{x_0}^{\text{av}}$ in Eq. (5.43) goes from $2\{(1 + \tau^2)\Phi(\tau) - \tau\phi(\tau)\}$ to $1 + \tau^2$ as τ_0 goes from zero to infinity. We wrote this expression as $1 + \tau^2 - \psi_\xi(\tau_0, \tau)$.

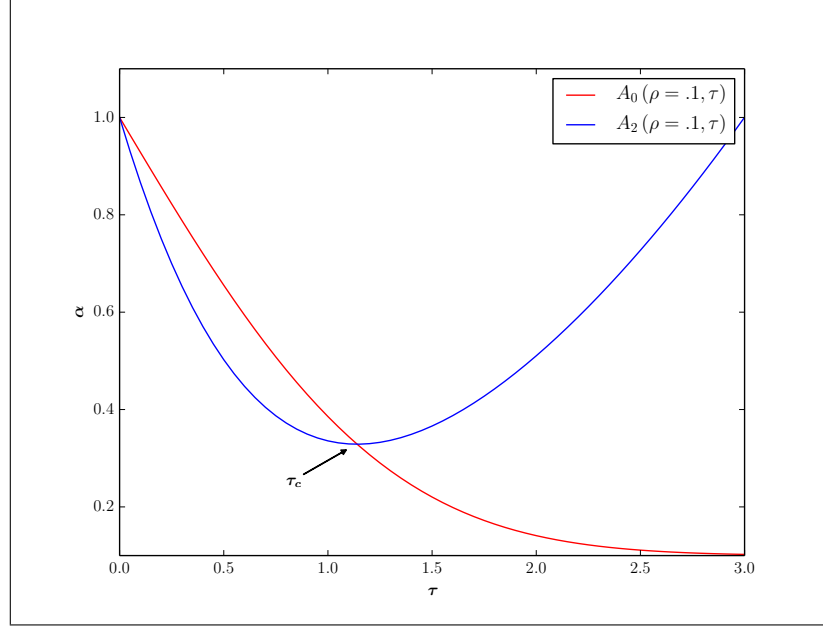


Figure 5.3: The transition point for ℓ_1 -norm minimization is where the red and blue curves meet at the critical τ_c where the derivative of A_2 is zero.

Moreover, we can write

$$\begin{aligned}
 [\psi_\xi(\tau_0, \tau)]_{x_0}^{\text{av}} &= \int dx_0 \pi(x_0) \psi_\xi\left(\frac{x_0}{\sigma_\xi}, \tau\right) \\
 &= \sigma_\xi \int d\tau_0 \pi(\sigma_\xi \tau_0) \psi_\xi(\tau_0, \tau)
 \end{aligned} \tag{5.48}$$

and get the same expression for $\psi_\theta(\tau_0, \tau)$. Thus, the small σ_ξ^2 behavior of these averages depends on how $\pi(x)$ behaves at small x . When $\pi(x) \sim Cx^\gamma$ with $\gamma > -1$:

$$[\psi_\xi(\tau_0, \tau)]_{x_0}^{\text{av}} \approx \sigma_\xi^{\gamma+1} \int d\tau_0 \tau_0^\gamma \psi_\xi(\tau_0, \tau) \sim \sigma_\xi^{\gamma+1} \tag{5.49}$$

Similarly $[\psi_\theta(\tau_0, \tau)]_{x_0}^{\text{av}} \sim \sigma_\xi^{\gamma+1}$. Thus, the perturbations added to phase boundary Eqs. (5.25) and (5.26) are of the order of $\sigma_\xi^{\gamma+1}$. Accordingly, in the case of a gap so that $\pi(x) = 0$ when $|x| < \Delta$ and $\pi(x) = \pi(0)$ for some $|x| > \Delta$ we have:

$$[\psi_\xi(\tau_0, \tau)]_{x_0}^{\text{av}} \approx \sigma_\xi \int_\Delta d\tau_0 \psi_\xi(\tau_0, \tau) \sim e^{-\frac{\Delta^2}{\sigma_\xi^2}} \sigma_\xi \tag{5.50}$$

And $[\psi_\theta(\tau_0, \tau)]_{x_0}^{\text{av}} \sim e^{-\frac{\Delta^2}{\sigma_\xi^2}} \sigma_\xi$

5.4.1 Into the Error-prone Regime ($\vartheta \rightarrow 0$ & $\sigma_\zeta^2 = 0$)

To find an estimate for the mean-squared error by entering into the error-prone regime, we express the phase boundary as $\alpha = \alpha_c(\rho), \tau = \tau_c(\rho)$ by solving Eqs. (5.25), and (5.26). To explore close to the phase boundary, we can write $\alpha = \alpha_c(\rho) - \delta\alpha$ and $\tau = \tau_c(\rho) - \delta\tau$. Since the perturbations to Eqs. (5.25), (5.26) for the case of $\pi(x) \sim Cx^\gamma$ are of the order $\sigma_\xi^{\gamma+1}$, from equation (??) we get

$$\delta\alpha \sim \sigma_\xi^{\gamma+1} = \left(\frac{q}{\alpha}\right)^{\frac{\gamma+1}{2}} \quad (5.51)$$

Therefore, for nonzero terms drawn from a distribution with nonzero density at the origin, Eq. (5.51) tells us that the mean square error rises as

$$q(MSE) \sim (\alpha_c - \alpha)^{\frac{2}{\gamma+1}} \quad (5.52)$$

Similarly, for $\pi(x)$ with a gap, we get a sharp rise for the error:

$$q \sim 1/\ln(\alpha_c - \alpha) \quad (5.53)$$

Remark 5 *The additional insight is that although the phase boundary $\alpha_c(\rho)$ does not depend on the distribution of non-zeros, the rise of the error does and becomes sharper when non-zero components are farther from zero. Moreover, the rise is continuous, i.e. it is a second-order phase transition and its critical exponent depends on the behavior of $\pi(x_0)$ near $x_0 = 0$.*

5.4.2 Role of an Additive Noise ($\vartheta \rightarrow 0$ & $\sigma_\zeta^2 \neq 0$)

To examine the behavior of Eqs. (5.25), and (5.26) close to the phase boundary within the presence of noise, once again, we Taylor expand them around the transition point where $\alpha = \alpha_c(\rho)$ and $\tau = \tau_c(\rho)$. Therefore, for the case of $\pi(x) \sim Cx^\gamma$, Eqs. (5.45), (5.46) in terms of perturbing variables $\delta\alpha$ and $\delta\tau$ become:

$$\delta\alpha = \alpha_c \frac{\sigma_\zeta^2}{\sigma_\xi^2} + C\delta\tau^2 - D\sigma_\xi^{\gamma+1} + \dots \quad (5.54)$$

$$\delta\alpha = -C'\delta\tau - D'\sigma_\xi^{\gamma+1} + \dots \quad (5.55)$$

Where C, D, C' and D' are functions of ρ and τ_c and “...” contains higher order corrections. From Eq. (5.55), we have $\delta\tau = -(D'/C')\sigma_\xi^{\gamma+1}$ which, by substitution to the first Eq. (5.54), gives

$$0 = \alpha_c \frac{\sigma_\zeta^2}{\sigma_\xi^2} + \frac{C D'^2}{C'^2} \sigma_\xi^{2+2\gamma} - D\sigma_\xi^{\gamma+1} \implies \sigma_\xi^2 \propto (\sigma_\zeta^2)^{2/(3+\gamma)}$$

which we arrived at it by taking into account that $\sigma_\xi^2 \rightarrow 0^+$. With a similar calculation for the case of a gap in the distribution $\pi(x)$, we obtain

$$\sigma_\xi^2 \propto 1/\ln(\sigma_\zeta^2) \quad (5.56)$$

5.4.3 ϑ Trade-off in the Noisy system ($\vartheta \neq 0$ & $\sigma_\zeta^2 \neq 0$)

In the previous subsection, we considered the role of additive Gaussian noise in the behavior of the phase boundary near the transition from perfect reconstruction to the error regime. However, one should take into consideration that in most situations noise arises from several sources and there is no good estimation of either the level or distribution of the noise. Therefore, there is often a trade-off between the least squares of the residual and the ℓ_1 norm of the solution. If the regularization is too much, the regularized solution does not fit the given signal properly as the residual error is too large. If the regularization is too small,

the fit will be good but error will be more. One can control this trade-off and the sparsity of the solution by proper selection of the regularization parameter ϑ . Taylor expand of Eqs. (5.25), and (5.26) close to the transition point leads to:

$$\delta\alpha = C\delta\tau^2 - D\sigma_\xi^{\gamma+1} + \dots \quad (5.57)$$

$$\delta\alpha = \alpha_c \frac{\vartheta}{\theta} - C'\delta\tau - D'\sigma_\xi^{\gamma+1} + \dots \quad (5.58)$$

From Eq. (5.57), we have $\delta\tau = (D/C)^{1/2}\sigma_\xi^{\frac{\gamma+1}{2}}$ which by substitution into Eq. (5.58) and by taking into account that $\theta \sim \sigma_\xi$ gives

$$0 = \alpha_c \frac{\vartheta}{\theta} - \frac{C'D^{1/2}}{C^{1/2}}\sigma_\xi^{\frac{\gamma+1}{2}} - D\sigma_\xi^{\gamma+1} \implies \sigma_\xi^2 \propto \vartheta^{\frac{4}{\gamma+3}} \quad (5.59)$$

Another interesting question would be that at what value of ϑ , we will get the minimum error in the presence of noise. By adding noise to the system, Eqs. (5.45) and (5.46) become

$$\alpha = \alpha \frac{\sigma_\zeta^2}{\sigma_\xi^2} + A_2(\rho, \tau) - \rho[\psi_\xi(\tau_0, \tau)]_{x_0}^{\text{av}} \quad (5.60)$$

$$\alpha = \alpha_c \frac{\vartheta}{\theta} + A_0(\rho, \tau) - \rho[\psi_\theta(\tau_0, \tau)]_{x_0}^{\text{av}}. \quad (5.61)$$

This leads to the following relations in terms of perturbing variables $\delta\alpha$ and $\delta\tau$

$$\delta\alpha = \alpha_c \frac{\sigma_\zeta^2}{\sigma_\xi^2} + C\delta\tau^2 - D\sigma_\xi^{\gamma+1} + \dots \quad (5.62)$$

$$\delta\alpha = \alpha_c \frac{\vartheta}{\theta} - C'\delta\tau - D'\sigma_\xi^{\gamma+1} + \dots \quad (5.63)$$

To have a solution, we get $\sigma_\xi^2 \sim \vartheta^{\frac{2}{\gamma+2}}$ and $\sigma_\xi^2 \sim (\sigma_\zeta^2)^{2/3}$. Therefore, by tuning ϑ to $(\sigma_\zeta^2)^{\frac{2}{\gamma+3}}$, the minimum error occurs.

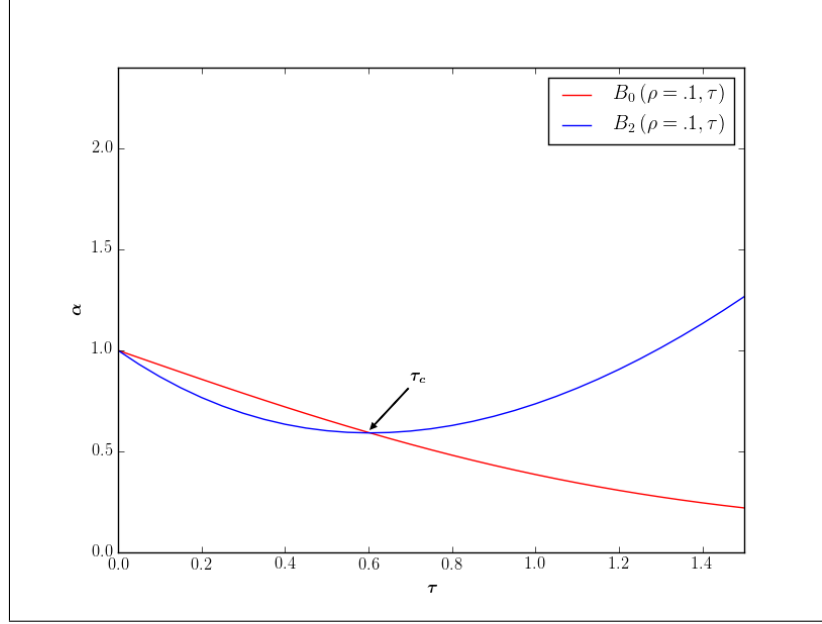


Figure 5.4: The transition boundary is where the red and blue curves meet at the critical τ_c . Unlike ℓ_1 -norm minimization, the slope at this point is not zero and there exists an additional linear term to the B_2 at the critical τ_c for Elastic Net. In the text we will see that this results in different critical behavior for Elastic Net .

5.5 Elastic Net

As a quick application of our zero temperature cavity method, we consider how the phase transition would be affected if we generalize our penalty function $V(\mathbf{x})$ by adding a quadratic term $|\mathbf{x}|^2$ to the ℓ_1 norm. This penalty function is used in the Elastic Net method of variable selection and regularization [55]. The optimization problem becomes

$$\hat{\mathbf{x}}_{\text{EN}} = \min_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{H}\mathbf{x})^2 + \lambda_1 |\mathbf{x}| + \frac{\lambda_2}{2} |\mathbf{x}|^2 \right\} \quad (5.64)$$

In the noiseless reconstruction problem, $\mathbf{y} = \mathbf{H}\mathbf{x}_0$. We take the limit $\sigma^2 \rightarrow 0$ and choose the distribution of \mathbf{H} and \mathbf{x}_0 to be the same as in the previous sections.

Now $U''(x) = \lambda_2$ everywhere except at $x = 0$, where it is formally infinite, leading to

$$\begin{aligned}\chi^{aa} &= 0, \text{ if } x_a = 0 \\ \chi^{aa} &= \frac{\sigma_{\text{eff}}^2}{1 + \lambda_2 \sigma_{\text{eff}}^2}, \text{ otherwise.}\end{aligned}\tag{5.65}$$

Once more we define $\hat{\rho}$ to be fraction of x_a s that are non-zero. Then $\bar{\chi} = \frac{\hat{\rho} \sigma_{\text{eff}}^2}{1 + \lambda_2 \sigma_{\text{eff}}^2}$ and

$$\sigma_{\text{eff}}^2 = \sigma^2 + \frac{\bar{\chi}}{\alpha} = \sigma^2 + \frac{\hat{\rho} \sigma_{\text{eff}}^2}{\alpha(1 + \lambda_2 \sigma_{\text{eff}}^2)}\tag{5.66}$$

implying

$$\sigma_{\text{eff}}^2 \left\{ 1 - \frac{\hat{\rho}}{\alpha(1 + \lambda_2 \sigma_{\text{eff}}^2)} \right\} = \sigma^2\tag{5.67}$$

In the $\sigma^2 \rightarrow 0$ limit, the two phases are given by, $\sigma_{\text{eff}}^2 = 0$ or $\hat{\rho} = \alpha(1 + \lambda_2 \sigma_{\text{eff}}^2)$. Again, the perfect reconstruction phase has $\bar{\chi} = \frac{\hat{\rho} \sigma_{\text{eff}}^2}{1 + \lambda_2 \sigma_{\text{eff}}^2} = 0$ and the error-prone regime has $\bar{\chi} = \frac{\hat{\rho} \sigma_{\text{eff}}^2}{1 + \lambda_2 \sigma_{\text{eff}}^2} = \alpha \sigma_{\text{eff}}^2 \neq 0$.

For the corresponding single variable optimization problem, we can still use the soft-thresholding function described in Eq. (5.20). The estimated value of \hat{x} is once more given by $\eta_{\text{soft}}(t; \theta)$, but with $t = \frac{x_0 + \xi}{1 + \lambda_2 \sigma_{\text{eff}}^2}$ and $\theta = \frac{\lambda_1 \sigma_{\text{eff}}^2}{1 + \lambda_2 \sigma_{\text{eff}}^2}$.

As before, we start in the perfect reconstruction phase, where $\sigma^2, \sigma_{\text{eff}}^2, \sigma_{\xi}^2 \rightarrow 0$ with $\tau = \frac{\lambda_1 \sigma_{\text{eff}}^2}{\sigma_{\xi}}$ of order one. In this phase we ignore the case of non-zero x_0 leading to $\hat{x} = 0$.

The contribution to MSE for the non-zero x_0 is slightly different

$$\begin{aligned}\rho[(\hat{x} - x_0)^2]_{x_0, \xi}^{\text{av}} &= \rho \left[\left(\frac{x_0 + \xi - \lambda_1 \sigma_{\text{eff}}^2 \text{sgn}(\hat{x})}{1 + \lambda_2 \sigma_{\text{eff}}^2} - x_0 \right)^2 \right]_{x_0, \xi}^{\text{av}} \\ &\approx \frac{\rho}{(1 + \lambda_2 \sigma_{\text{eff}}^2)^2} \left\{ \sigma_{\xi}^2 + (\lambda_1 \sigma_{\text{eff}}^2)^2 \left(1 + \frac{\lambda_2^2}{\lambda_1^2} [x_0^2]_{x_0}^{\text{av}} + \frac{\lambda_2}{\lambda_1} [|x_0|]_{x_0}^{\text{av}} \right) \right\}\end{aligned}\tag{5.68}$$

The key approximation is that $[x_0 \text{sgn}(\hat{x})]_{x_0, \xi}^{\text{av}} \approx [|x_0|]_{x_0}^{\text{av}}$, since in this limit typically $|\xi| \ll |x_0|$ implying \hat{x} and x_0 have the same sign. The other source of error is the event when the x_0 is zero but \hat{x} has fallen outside the interval $[-\theta, \theta]$ and has been estimated to be

non-zero. In this case, the contribution to MSE is

$$\begin{aligned}
 (1 - \rho)[\hat{x}^2]_{x_0, \xi}^{\text{av}} &= 2(1 - \rho) \int_{\lambda_1 \sigma_{\text{eff}}^2}^{\infty} \frac{d\xi}{\sqrt{2\pi\sigma_{\xi}^2}} e^{-\frac{\xi^2}{2\sigma_{\xi}^2}} \left(\frac{\xi - \lambda_1 \sigma_{\text{eff}}^2}{1 + \lambda_2 \sigma_{\text{eff}}^2} \right)^2 \\
 &= \frac{2\sigma_{\xi}^2(1 - \rho)}{(1 + \lambda_2 \sigma_{\text{eff}}^2)^2} \{ (1 + \tau^2)\Phi(\tau) - \tau\phi(\tau) \}.
 \end{aligned} \tag{5.69}$$

Combining Eq. (5.68) and (5.69) in the self-consistency equation for σ_{ξ}^2 and remembering that $\sigma_{\xi}^2, \sigma_{\text{eff}}^2 \rightarrow 0$ with $\tau = \frac{\lambda_1 \sigma_{\text{eff}}^2}{\sigma_{\xi}}$ order one, we have

$$\begin{aligned}
 \alpha &= 2(1 - \rho) \{ (1 + \tau^2)\Phi(\tau) - \tau\phi(\tau) \} \\
 &\quad + \rho \left\{ 1 + \tau^2 \left(1 + \frac{\lambda_2^2}{\lambda_1^2} [x_0^2]_{x_0}^{\text{av}} + \frac{\lambda_2}{\lambda_1} [|x_0|]_{x_0}^{\text{av}} \right) \right\}.
 \end{aligned} \tag{5.70}$$

The equation for $\hat{\rho}$ remains the same in this limit. The denominator $1 + \lambda_2 \sigma_{\text{eff}}^2$ does not matter for the thresholding condition. As a result once more

$$\hat{\rho} = 2(1 - \rho)\Phi(\tau) + \rho. \tag{5.71}$$

On the other hand, the condition for the phase boundary is $\alpha = \hat{\rho}(1 + \lambda_2 \sigma_{\text{eff}}^2)$. Thus, for the Elastic Net method, the phase boundary is obtained by solving and eliminating τ from the following equations:

Proposition 3 (Theoretical Phase Boundary for Elastic Net)

$$\begin{aligned}
 \alpha \equiv B_0 &= 2(1 - \rho) \{ (1 + \tau^2)\Phi(\tau) - \tau\phi(\tau) \} \\
 &\quad + \rho \left\{ 1 + \tau^2 \left(1 + \frac{\lambda_2^2}{\lambda_1^2} [x_0^2]_{x_0}^{\text{av}} + \frac{\lambda_2}{\lambda_1} [|x_0|]_{x_0}^{\text{av}} \right) \right\}
 \end{aligned} \tag{5.72}$$

$$\alpha \equiv B_2 = 2(1 - \rho)\Phi(\tau) + \rho \tag{5.73}$$

ℓ_1 -norm Minimization	
System	Scaling Law
$\lambda \rightarrow 0, \sigma_\zeta^2 = 0$	$MSE \sim (\alpha_c - \alpha)^2$
$\lambda \rightarrow 0, \sigma_\zeta^2 \neq 0$	$MSE \sim (\sigma_\zeta^2)^{2/3}$
$\lambda \neq 0, \sigma_\zeta^2 = 0$	$MSE \sim \lambda^{4/3}$
$\lambda \neq 0, \sigma_\zeta^2 \neq 0$	$MSE \sim \lambda$
Elastic Net	
$\lambda \rightarrow 0, \sigma_\zeta^2 = 0$	$MSE \sim (\alpha_c - \alpha)^2$
$\lambda \rightarrow 0, \sigma_\zeta^2 \neq 0$	$MSE \sim (\sigma_\zeta^2)^{2/3}$
$\lambda \neq 0, \sigma_\zeta^2 = 0$	$MSE \sim \lambda$
$\lambda \neq 0, \sigma_\zeta^2 \neq 0$	$MSE \sim \lambda$

Table 5.2: This table shows the comparison of critical exponents for Basis Pursuit and Elastic Net near phase transition.

In the case of Gaussian $\pi(x_0)$ with variance $\sigma_{x_0}^2$, the key dimensionless parameter is $\frac{\lambda_2 \sigma_{x_0}}{\lambda_1}$, which determines the relative strength of the quadratic penalty term. It's important to note that unlike the ℓ_1 -norm minimization, the relation $\frac{dA_2}{d\tau} = 2 \frac{A_2 - A_0}{\tau}$ in remark 4 does not hold for Elastic Net. Thus, Taylor expansion of Eq. (5.72) (equivalent to the A_2 term in Eq. (5.47)) near the transition point has linear contribution with positive slope as well as quadratic one (See Fig. 5.5). The theoretical critical exponents can be derived in the same way as described in section 5.4. We only mention the result in the table 5.2 in the next section.

5.6 Numerical Experiment

In this section, we describes some experiments examining critical exponents that we obtained in the previous section (section 5.4) and comparison with the numerical result. Before moving further, we remind the readers by summarizing the critical exponents and scaling laws near the phase transition in the table below (for the case of Gaussian distribution of non-zero components of the signal, i.e. $\gamma = 0$). In order to relate with compressed-sensing literature, we have set $\sigma^2 = 1$, i.e. $\vartheta = \lambda$.

In the first experiment, the goal is to compute MSE for ℓ_1 -norm minimization and Elastic

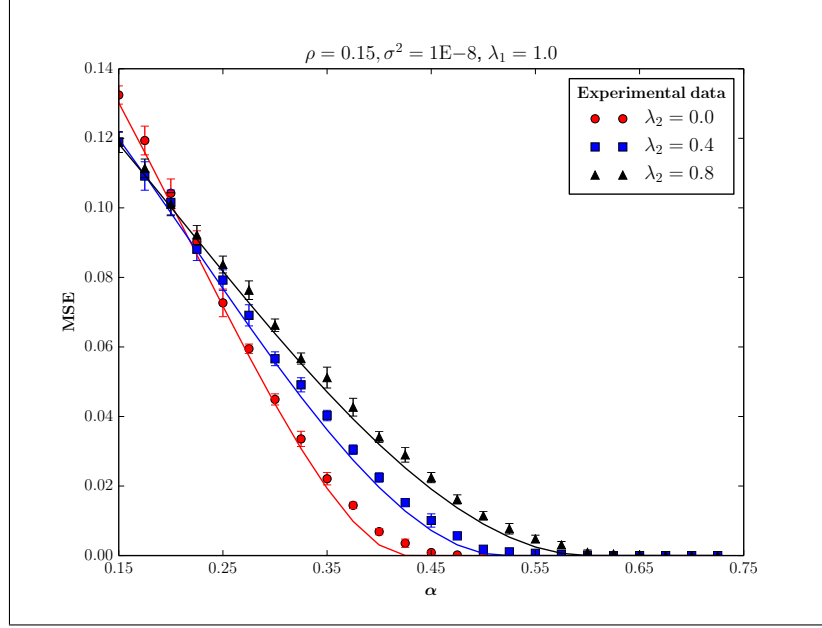


Figure 5.5: Comparison of MSE for different λ_2 . Each solid curve represents the theoretical estimate for MSE as described in Sec. 5.2 and Sec. 5.5. Numerical data for different λ_2 is shown with the markers. We use CVXOPT quadratic cone programming to find MSE for 3 values of λ_2/λ_1 : 0, 0.4, 0.8. Notice that, for the Elastic Net ($\lambda_2 \neq 0$), the transition happens at higher α compared to ℓ_1 -norm minimization.

Net. In this case, the matrix \mathbf{H} is obtained by first filling it with independent samples of a Gaussian distribution with variance $1/M$. In this example, $N = 200$, $K = 30$, the original signal \mathbf{x} contains 30 randomly placed elements driven from a standard Gaussian distribution, i.e. $\gamma = 0$. This experiment is done by using the CVXOPT quadratic cone programming [61] for the case with $\lambda_1 = 1\text{E} - 6$ and $\lambda_2 = 0, .4, .8$ of λ_1 . As it is illustrated in Fig. 5.6, the reconstruction error exhibits a higher mean squared error (MSE) with respect to the theoretical result.

Next, we confirm the exponent in Eq. (5.59) by plotting the theoretical expression in Eqs. (5.45), (5.46). This is shown in Fig. 5.6.

In the end, we consider the important case where the external noise is zero and we are looking for a trade-off for λ where the reconstruction error minimizes. Once more, using Eqs. (5.45), (5.46), this is shown in Fig. 5.6.

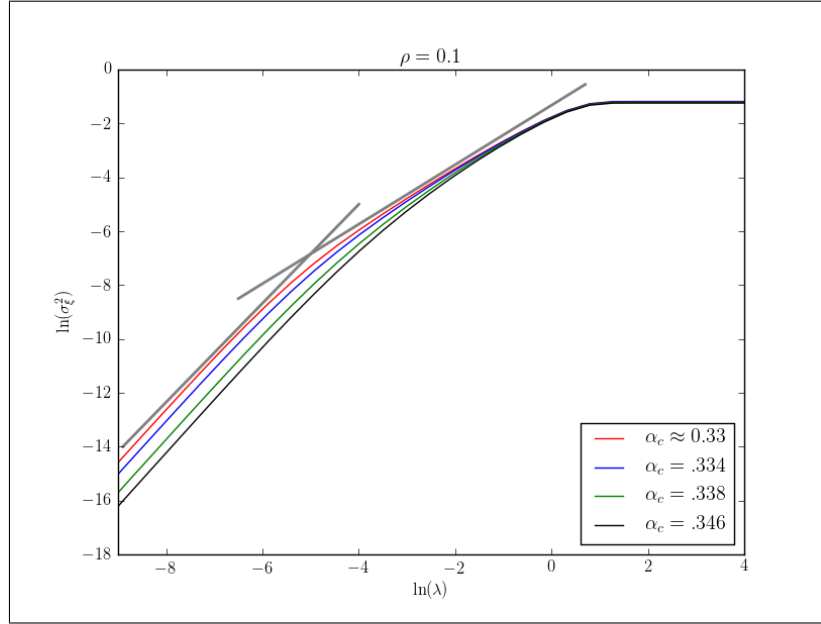


Figure 5.6: Following the trends where the curves merge, we can find the critical exponent near the phase transition. The top grey line has a slope of ~ 1.1 and the bottom one has a slope of ~ 1.9 .

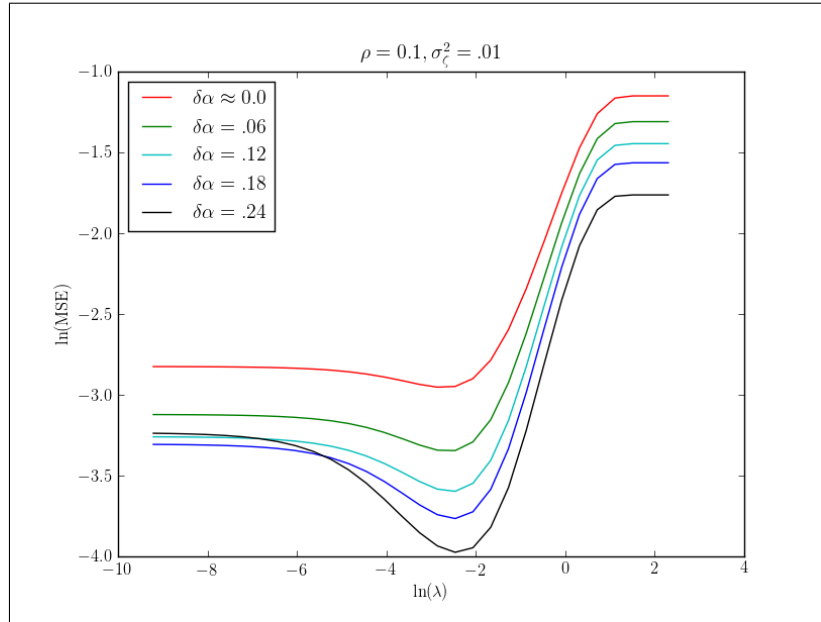


Figure 5.7: Varying λ sweeps out entire optimal tradeoff curves. The theoretical minimum error near phase transition occurs at $\ln(\lambda) = \frac{2}{3} \ln(\sigma_\xi^2)$.

Chapter 6

Conclusion

In this thesis, we presented a different approach to study the statistical properties of compressed sensing problems. In the case of a full random measurement matrix, we directly treat the optimization problem and show how to adapt the cavity method for doing mean field theory in the context. The mean field theory leads to a self-consistency condition on average mean squared error (MSE), since error in estimating one variable affects error in others. Careful derivation of the self-consistency condition involves accounting for subtle correlations in the system. To take care of these correlations, we needed a two-step cavity approach: one step removing a variable and then, another, removing a data constraint.

In the process of this derivation, we realize the key role local susceptibility plays in the system. It turns out that the perfect reconstruction phase corresponds to vanishing average local susceptibility, indicating that the solution of the optimization problem has underlying robustness to perturbations in this phase. We expect that the structure of susceptibility has important information about the reliability of any sparse reconstruction.

The cavity approach looks at the behavior of the system for a particular choice of quenched variables, \mathbf{H} and \mathbf{x}_0 . In contrast, the replica approach centers on immediately averaging those quenched variables away. In the context of compressed sensing, one can imagine many problems where the matrix \mathbf{H} is non-random. Currently there is no obvious way to extend the replica method for such sensing matrices. The cavity method could be a more versatile tool in this regard.

Appendix A

Correlated Measurement Matrices

Here, we present the generalization of our results for Gaussian $\mathcal{P}(\mathbf{H})$ with the mean and variance of $P(\mathbf{H})$ given by

$$[H_{ia}]^{\text{av}} = 0 \quad (\text{A.1})$$

and

$$[H_{ia}H_{jb}]^{\text{av}} = \frac{1}{M}C_{ij}D_{ab}. \quad (\text{A.2})$$

In this case, one could still follow the replica computation in Section 3.2 and show that the finite temperature replica saddle point equations [62] generalize to The saddle point $\mathbf{Q} = \bar{\mathbf{Q}}, \mathbf{R} = -i\bar{\mathbf{R}}$ satisfies the conditions:

$$\bar{Q}_{\mu\nu} = \frac{1}{N} \langle \langle \mathbf{u}_\mu^\top \mathbf{D} \mathbf{u}_\nu \rangle \rangle \quad (\text{A.3})$$

$$\bar{\mathbf{R}} = \frac{\beta}{2\sigma^2} \text{Tr}_M [\mathbf{C} \otimes \mathbf{I}_n (\mathbf{I}_M \otimes \mathbf{I}_n + \frac{\beta}{\alpha\sigma^2} \mathbf{C} \otimes \mathbf{Q})^{-1}] \quad (\text{A.4})$$

obtained by differentiating $S(\mathbf{Q}, \mathbf{R})$ with respect to the elements of \mathbf{Q}, \mathbf{R} . The trace Tr_M is a partial trace only applying to the M dimensional space. The expectation $\langle \langle \mathbf{u}_\mu^\top \mathbf{D} \mathbf{u}_\nu \rangle \rangle$ depends on $\bar{\mathbf{R}}$ via

$$\langle \langle \mathbf{u}_\mu^\top \mathbf{D} \mathbf{u}_\nu \rangle \rangle = \beta \frac{\partial F(\bar{\mathbf{R}})}{\partial \bar{R}_{\mu\nu}} \quad (\text{A.5})$$

with $\exp(-\beta F(\bar{\mathbf{R}}))$

$$= \left[\int \prod_{\mu=1}^n \{d^N \mathbf{u}_\mu\} \exp \left[- \sum_{\mu,\nu} \bar{R}_{\mu\nu} \mathbf{u}_\mu^\top \mathbf{D} \mathbf{u}_\nu - \beta \sum_{\mu} V(\mathbf{u}_\mu + \mathbf{x}_0) \right] \right]_{\mathbf{x}_0}^{\text{av}} \quad (\text{A.6})$$

For $U(x)$ convex, we have replica symmetric ansatz for \mathbf{Q}, \mathbf{R} : $\bar{Q}_{\mu\nu} = (Q - q)\delta_{\mu\nu} + q$ and $\bar{R}_{\mu\nu} = (R - r)\delta_{\mu\nu} + r$. The zero-temperature limit turns out to be a self-consistent problem of coupled nodes in presence of quenched correlated noise. The optimization problem reduces to finding

$$\min_{\mathbf{u}} \left\{ \frac{1}{2\sigma_{\text{eff}}^2} (\mathbf{u}^\top \mathbf{D} \mathbf{u} - 2\boldsymbol{\xi}^\top \mathbf{D} \mathbf{u}) + V(\mathbf{u} + \mathbf{x}_0) - \mathbf{f} \cdot \mathbf{u} \right\} \quad (\text{A.7})$$

with

$$\frac{1}{\sigma_{\text{eff}}^2} = \frac{1}{M} \text{Tr} \left[\mathbf{C} \left\{ \mathbf{I}_M \sigma^2 + \frac{\beta \text{Tr}(\mathbf{D} \Delta \mathbf{Q})}{M} \mathbf{C} \right\}^{-1} \right] \quad (\text{A.8})$$

where the Gaussian quenched vector $\boldsymbol{\xi}$ has mean zero and the covariance matrix given by

$$\frac{q \mathbf{D}^{-1} \text{Tr} \left[\mathbf{C} \left\{ \mathbf{I}_M \sigma^2 + \frac{\beta \text{Tr}(\mathbf{D} \Delta \mathbf{Q})}{M} \mathbf{C} \right\}^{-1} \right]^2 / M}{\alpha \left(\text{Tr} \left[\mathbf{C} \left\{ \mathbf{I}_M \sigma^2 + \frac{\beta \text{Tr}(\mathbf{D} \Delta \mathbf{Q})}{M} \mathbf{C} \right\}^{-1} \right] / M \right)^2} \quad (\text{A.9})$$

where

$$q = \frac{1}{N} [\mathbf{u}^\top \mathbf{D} \mathbf{u}]_{\mathbf{x}_0, \boldsymbol{\xi}}^{\text{av}}. \quad (\text{A.10})$$

Identifying $\beta \Delta \mathbf{Q}$ as the average susceptibility $\bar{\chi}$, we can get the same result if we have done the computation through two-step cavity method.

Bibliography

- [1] Candès EJ, Romberg J, Tao T (2006) Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on* 52:489–509.
- [2] Donoho DL (2006) For most large underdetermined systems of linear equations the minimal. *Communications on pure and applied mathematics* 59:797–829.
- [3] Donoho DL, et al. (2000) High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Challenges Lecture* pp. 1–32.
- [4] Chen SS, Donoho DL, Saunders MA (1998) Atomic decomposition by basis pursuit. *SIAM journal on scientific computing* 20:33–61.
- [5] Donoho DL (2006) Compressed sensing. *Information Theory, IEEE Transactions on* 52:1289–1306.
- [6] Candès E, Romberg J (2007) Sparsity and incoherence in compressive sampling. *Inverse problems* 23:969.
- [7] Donoho DL, Tanner J (2005) Sparse nonnegative solution of underdetermined linear equations by linear programming. *Proceedings of the National Academy of Sciences of the United States of America* 102:9446–9451.
- [8] Kabashima Y, Wadayama T, Tanaka T (2009) A typical reconstruction limit for compressed sensing based on ℓ_p -norm minimization. *Journal of Statistical Mechanics: Theory and Experiment* 2009:L09003.
- [9] Ganguli S, Sompolsky H (2010) Statistical mechanics of compressed sensing. *Physical review letters* 104:188701.
- [10] Mézard M, Parisi G, Virasoro M (1986) Sk model: The replica solution without replicas. *Europhys. Lett* 1:77–82.
- [11] Mézard M, Parisi G, Virasoro MA (1987) *Spin glass theory and beyond*. (World scientific Singapore) Vol. 9.
- [12] Mulet R, Pagnani A, Weigt M, Zecchina R (2002) Coloring random graphs. *Physical Review Letters* 89:268701.
- [13] Mézard M, Parisi G, Zecchina R (2002) Analytic and algorithmic solution of random satisfiability problems. *Science* 297:812–815.
- [14] Shamir M, Sompolsky H (2000) Thouless-anderson-palmer equations for neural networks. *Physical Review E* 61:1839.

- [15] Braunstein A, Mézard M, Zecchina R (2005) Survey propagation: An algorithm for satisfiability. *Random Structures & Algorithms* 27:201–226.
- [16] Ramezanali M, Mitra PP, Sengupta AM (2015) The cavity method for phase transitions in sparse reconstruction algorithms. *arXiv preprint arXiv:1501.03194*.
- [17] Mezard M, Montanari A (2009) *Information, physics, and computation*. (Oxford University Press).
- [18] Megason SG, Fraser SE (2007) Imaging in systems biology. *Cell* 130:784–795.
- [19] Lustig M, Donoho D, Pauly JM (2007) Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic resonance in medicine* 58:1182–1195.
- [20] Lustig M, Donoho DL, Santos JM, Pauly JM (2008) Compressed sensing mri. *Signal Processing Magazine, IEEE* 25:72–82.
- [21] Schad LR (2004) Problems in texture analysis with magnetic resonance imaging. *Dialogues in clinical neuroscience* 6:235.
- [22] Tsao J, Boesiger P, Pruessmann KP (2003) k-t blast and k-t sense: Dynamic mri with high frame rate exploiting spatiotemporal correlations. *Magnetic Resonance in Medicine* 50:1031–1042.
- [23] Jung H, Sung K, Nayak KS, Kim EY, Ye JC (2009) k-t focuss: A general compressed sensing framework for high resolution dynamic mri. *Magnetic Resonance in Medicine* 61:103–116.
- [24] Gamper U, Boesiger P, Kozerke S (2008) Compressed sensing in dynamic mri. *Magnetic Resonance in Medicine* 59:365–373.
- [25] Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 6:95–108.
- [26] Duerr RH et al. (2006) A genome-wide association study identifies il23r as an inflammatory bowel disease gene. *science* 314:1461–1463.
- [27] Hampe J et al. (2007) A genome-wide association scan of nonsynonymous snps identifies a susceptibility variant for crohn disease in atg16l1. *Nature genetics* 39:207–211.
- [28] Ozaki K et al. (2002) Functional snps in the lymphotoxin- α gene that are associated with susceptibility to myocardial infarction. *Nature genetics* 32:650–654.
- [29] Sladek R et al. (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881–885.
- [30] Tomancak P et al. (2002) Systematic determination of patterns of gene expression during drosophila embryogenesis. *Genome Biol* 3:0081–0088.
- [31] Vattikuti S, Lee JJ, Chang CC, Hsu S, Chow CC (2014) Applying compressed sensing to genome-wide association studies. *GigaScience* 3:10.
- [32] Dorfman R (1943) The detection of defective members of large populations. *The Annals of Mathematical Statistics* 14:436–440.

- [33] Sachs H, Klopstock A, Weil A (1925) Die entstehung der syphilitischen blutveraenderung. *DMW-Deutsche Medizinische Wochenschrift* 51:589–592.
- [34] Johnson WO, Gastwirth JL (2000) Dual group screening. *Journal of Statistical Planning and Inference* 83:449–473.
- [35] Du D, Hwang F (1993) *Combinatorial group testing and its applications*. (World Scientific).
- [36] Wright J et al. (2010) Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE* 98:1031–1044.
- [37] Weinberger KQ, Blitzer J, Saul LK (2005) *Distance metric learning for large margin nearest neighbor classification*. pp. 1473–1480.
- [38] Kim KI, Jung K, Kim HJ (2002) Face recognition using kernel principal component analysis. *Signal Processing Letters, IEEE* 9:40–42.
- [39] Kanevsky D, Sainath TN, Ramabhadran B, Nahamoo D (2010) *An analysis of sparseness and regularization in exemplar-based methods for speech classification*. pp. 2842–2845.
- [40] Candès EJ (2008) The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathématique* 346:589–592.
- [41] Baraniuk R, Davenport M, DeVore R, Wakin M (2008) A simple proof of the restricted isometry property for random matrices. *Constructive Approximation* 28:253–263.
- [42] Tikhonov A, Arsenin V (1977) *Solutions of ill-posed problems*. (Vh Winston).
- [43] Amaldi E, Kann V (1998) On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. *Theoretical Computer Science* 209:237–260.
- [44] Donoho DL (2005) Neighborly polytopes and sparse solutions of underdetermined linear equations.
- [45] Candes E, Rudelson M, Tao T, Vershynin R (2005) *Error correction via linear programming*. (IEEE), pp. 668–681.
- [46] Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 267–288.
- [47] Donoho DL, Tanner J (2006) *Thresholds for the recovery of sparse solutions via l1 minimization*. (IEEE), pp. 202–206.
- [48] Nelder JA, Mead R (1965) A simplex method for function minimization. *The computer journal* 7:308–313.
- [49] Kubo R (1966) The fluctuation-dissipation theorem. *Reports on Progress in Physics* 29:255.
- [50] Sengupta A, Mitra PP (1999) Distributions of singular values for some random matrices. *Physical Review E* 60:3389.

- [51] Sengupta AM, Mitra PP (2006) Capacity of multivariate channels with multiplicative noise: Random matrix techniques and large-n expansions (2). *Journal of Statistical Physics* 125:1223–1242.
- [52] Onsager L (1936) Electric moments of molecules in liquids. *Journal of the American Chemical Society* 58:1486–1493.
- [53] Loeliger HA (2004) An introduction to factor graphs. *Signal Processing Magazine, IEEE* 21:28–41.
- [54] Tanner RM (1981) A recursive approach to low complexity codes. *Information Theory, IEEE Transactions on* 27:533–547.
- [55] Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67:301–320.
- [56] Tikhonov AN (1943) On the stability of inverse problems. 39:195–198.
- [57] Moustakas AL, Baranger HU, Balents L, Sengupta AM, Simon SH (2000) Communication through a diffusive medium: Coherence and capacity. *Science* 287:287–290.
- [58] Moustakas AL, Simon SH, Sengupta AM (2003) Mimo capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large n analysis. *Information Theory, IEEE Transactions on* 49:2545–2561.
- [59] Takeda K, Kabashima Y (2011) Statistical mechanical assessment of a reconstruction limit of compressed sensing: Toward theoretical analysis of correlated signals. *EPL (Europhysics Letters)* 95:18006.
- [60] Takeda K, Kabashima Y (2013) Reconstruction algorithm in compressed sensing based on maximum a posteriori estimation. 473:012003.
- [61] Andersen M, Dahl J, Vandenberghe L (2010) CVXOPT: A python package for convex optimization.
- [62] Takeda K, Kabashima Y (2010) Statistical mechanical analysis of compressed sensing utilizing correlated compression matrix. pp. 1538–1542.