論文 / 著書情報
Article / Book Information

題目(和文)	1ビット圧縮センシングの統計力学的解析
Title(English)	Statistical mechanics approach to 1-bit compressed sensing
著者(和文)	許インイン
Author(English)	yingying xu
出典(和文)	学位:博士（理学）, 学位授与機関:東京工業大学, 報告番号:甲第10102号, 授与年月日:2016年3月26日, 学位の種別:課程博士, 審査員:樺島 祥介,渡邊 澄夫,高安 美佐子,石井 秀明,青西 亨
Citation(English)	Degree:, Conferring organization: Tokyo Institute of Technology, Report number:甲第10102号, Conferred date:2016/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

TOKYO INSTITUTE OF TECHNOLOGY

DOCTORAL THESIS

Statistical mechanics approach to 1-bit compressed sensing

Author:
Yingying XU

Supervisor:
Prof. Yoshiyuki KABASHIMA

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Science

in the

Yoshiyuki Kabashima's laboratory
Department of Computational Intelligence and Systems Science

February 24, 2016

Declaration of Authorship

I, Yingying XU, declare that this thesis titled, "Statistical mechanics approach to 1-bit compressed sensing" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Name: Yingying Xu

Date: February 16, 2016

TOKYO INSTITUTE OF TECHNOLOGY

Abstract

Department of Computational Intelligence and Systems Science

Doctor of Science

Statistical mechanics approach to 1-bit compressed sensing

by Yingying XU

The 1-bit compressed sensing framework enables the recovery of a sparse vector from the sign information of each entry of its linear transformation. Discarding the amplitude information can significantly reduce the amount of data, which is highly beneficial in practical applications. For simplicity, we consider the case that the measuring matrix has i.i.d entries, and the measurements are noiseless.

First, we analyze the typical performance of an l_1-norm based signal recovery scheme for the 1-bit compressed sensing using statistical mechanics methods. We also develop another approximate recovery algorithm inspired by the cavity method of statistical mechanics.

To further develop the study of l_1-norm based signal recovery scheme for 1-bit compressed sensing, we suggest a strategy that captures scale information by introducing a threshold parameter to the quantization process. For practical use, we develop a heuristic that adaptively tunes the threshold parameter based on measurement results.

Besides l_1-norm minimization, we present a Bayesian approach to signal reconstruction for 1-bit compressed sensing, which provides the optimal bound of the recovery performance of 1-bit measurements. Utilizing the replica method of statistical mechanics to analyze the typical performance, we show that the Bayesian approach enables better reconstruction than the l_1-norm minimization approach, asymptotically saturating the performance obtained when the non-zero entries positions of the signal are known under an appropriate condition. We also test a message passing algorithm for signal reconstruction on the basis of belief propagation. The results of numerical experiments are consistent with those of the theoretical analysis.

Acknowledgements

This thesis is a summary of my graduate study from March 2011 to March 2016. First of all, I would like to express my sincerest gratitute to my supervisor Professor Yoshiyuki Kabashima who has guided me into the scientific world. With great patience, continuous encouragement and wisdom of leaving space to my own, he has watched me trying and changing during the five years of study. His great passion to science has effected me a lot to enjoy the process of wonder and discovery. I also greatly acknowledge Doctor Lenka Zdeborová for her inspiring discussions and guidance of my work. I am very thankful to the whole research group of her, for a very warm reception and to all members for helping me during my stay in France. I am also greatful to all the organizers of workshops, conferences and summer schools where I had participated.

I have recieved financial support from JSPS Research Fellowships DC2 (March 2014– March 2016) and scholarship granted by Rotary Yoneyama Memorial Foundation, Inc (March 2012– March 2013). Thanks to the JSPS Core-to-Core Program "Non-equilibrium dynamics of soft matter and information," to support my stay in France in 2014. Without those supports I would not have been able to concentrate on my study.

List of publications

- Statistical mechanics approach to 1-bit compressed sensing
  Yinging Xu and Yoshiyuki Kabashima
  Journal of Statistical Mechanics: Theory and Experiment
  doi:10.1088/1742-5468/2013/02/P02041
  2013

- Bayesian signal reconstruction for 1-bit compressed sensing
  Yingying Xu, Yoshiyuki Kabashima, Lenka Zdeborová
  Journal of Statistical Mechanics: Theory and Experiment
  doi:10.1088/1742-5468/2014/11/P11015
  2014

- The contents in Chapter 3 are summarized as paper "Statistical mechanics analysis of thresholding 1-bit compressed sensing" (Yinging Xu and Yoshiyuki Kabashima), which has been submitted to the IEEE Transactions on Signal Processing for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

Contents

List of Figures

Chapter 1

Introduction

A big picture of the compressed sensing problem is presented in this chapter, including the historical background rooted in linear algebra, the modern focus on the sparse nature of signal in appropriate domain and the breakthrough of the conventional sampling rate limit, which has been tremendously effected many fields in science and technology. Important results from the theoretical and practical sides on compressed sensing are also introduced. Then the aim and outline of this thesis are provided.

1.1 Linear Algebra – the foundamental

The history of mathematics can be seen as an ever-increasing series of abstractions in human development. As the concept of equation appears, a foundmental form $ax + b = 0$ came in to the stage. The simple equation is an ancient question worked on by people from all walks of lives.

1.1.1 History[8]

Around 4000 years ago, the people of Babylon knew how to solve a simpe 2×2 system of linear equations with two unknowns. Around 200 BC, the Chinese publishes that "Nine Chapters of the Mathematical Art" which displayed the ability to solve a 3×3 system of equations. However, the power and progress in Linear Algebra did not come to fruition until the late 17^{th} century.

Interestingly, Linear Algebra has become more relevant since the emergence of calculus. In the late 17^{th} century, Leibnits brought up the study of determinants. Lagrange came out his work of Lagrange multipliers. Cramer presented his idea of solving systems of linear equations based on determinants. Euler then suggested the idea that a system of equations doesn't neccessarily have to have a solution. And many discussions around the concept of unique solution have been made. In 19^{th} century, Gauss introduced a procedure which termed Gaussian elimination to be used for solving a system of liear equations. This method combining , swapping, or multiplying rows with each other in order to eliminate variables from certain equations. In 1848, a proper notation of describing the process was suggested by J. J. Sylvester who introduced the term "matrix", the Latin word for womb, as a name for an array of numbers. Arthur Cayley defined matrix multiplication or matrix algebra, and he used the letter "A" to represent a matrix, which is a tradition till now.

Matrices at the end of the 19^{th} century were heavily connected with Physics issues and for mathematicians. For a time, however, interest in a

lot of linear algebra slowed until the end of World War II brought on the development of computers.Instead of having to break down an enormous $n \times n$ matrix, computers could quickly and accurately solve there systems of linear algera. Regardless of the technology though. Gaussian elimination still proves the best way known to solve a system of linear equations.

The influence of Linear Algebra in the mathematical world is spread wide because it provides and important base to many of the principles and practices. For example, some of the applications are to solve systems of linear format, to find least-square best fit lines of the data points to predict future outcomes, and the use of the Fourier series expansion as a means to solving partial differential equations. Even for broader topics like to solve questions of energy in quantum mechanics or everyday household games Sudoku, Linear Algebra supplied the basics. In this chapter, some detailed examples will be shown.

Technology continues to push the use further and further, but the history of Linear Algebra continues to provide the foundation.

1.1.2 General formulation

Expressing Linear Equations in a general form is

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{1.1}$$

where, $\boldsymbol{A}$ is a $M \times N$ dimensional matrix, $\boldsymbol{x}$ is a $N \times 1$ dimensionl unknown vector, and $\boldsymbol{y}$ is a $M \times 1$ dimensionl observed vector. Mordern Linear algebra is the branch of mathematics concerning vector spaces and linear mappings between such spaces. It includes the study of lines, planes, and subspaces, but is also concerned with properties common to all vector spaces. Techniques from linear algebra are also used in analytic geometry, engineering, physics, natural sciences, computer science, computer animation, and the social sciences (particularly in economics). Because linear algebra is such a well-developed theory, nonlinear mathematical models are sometimes approximated by linear models.

1.2 A story from signal processing point of view [25]

A typical example of the powerful use of Linear Algebra is in the field of signal processing, which provides the core technology of modern human lives. In signal processing, we sample a contiuous-time signal (often called ”analog signal") N_s times then send the discrete-time signal (often called ”digital signal"). The reciever needs to recover the original signal from the sampled data. This process can be called Analog-Digital conversion.

Assume that the analog signal $y(t)$ is periodic with period T and can be expressed as Fourier series

$$y(t) = a_0 + \sum_{k=1}^{k_m} \left(a_k \cos \frac{2\pi kt}{T} + b_k \sin \frac{2\pi kt}{T} \right) \tag{1.2}$$

where k_m is the maximum wave number and the corresponding maximum frequency in the Fourier domain is $f_m = k_m/T$ (finite region of frequencies).

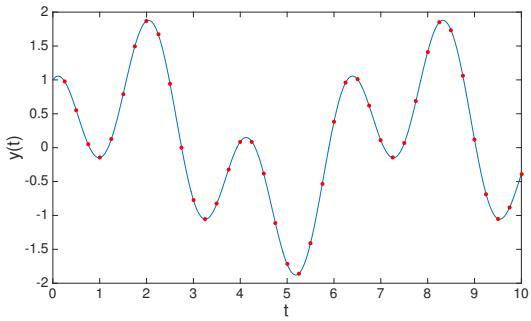

FIGURE 1.1: An example of analog-signal in time. The line represents the signal $y(t)$, and the red dots correspond to the sampled values by $y(nT_s)$.

To perfectly recover the analog signal, we need to know all the coefficients, which are $2k_m + 1$ unknown variables $a_0, \{a_k\}, \{b_k\}$.

Sampling of a fixed period T_s provides a set of independent linear equations

$$y(nT_s) = a_0 + \sum_{k=1}^{k_m} \left(a_k \cos\left(\frac{2\pi k n T_s}{T}\right) + b_k \sin\left(\frac{2\pi k n T_s}{T}\right)\right) \tag{1.3}$$

for determining the Fourier coefficients in 1.2. Here, $T_s = T/N_s$, and $n = 1, 2, \ldots, N_s$.

When $N_s = 2k_m + 1$, we can express the problem in matrix as

$$\begin{pmatrix} y\left(\frac{T}{N_s}\right) \\ y\left(\frac{2T}{N_s}\right) \\ y\left(\frac{3T}{N_s}\right) \\ y\left(\frac{4T}{N_s}\right) \\ \vdots \\ y\left(T\right) \end{pmatrix} = \begin{pmatrix} 1 & \cos\left(2\pi\frac{1}{N_s}\right) & \sin\left(2\pi\frac{1}{N_s}\right) & \cos\left(4\pi\frac{1}{N_s}\right) & \cdots & \sin\left(2k_m\pi\frac{1}{N_s}\right) \\ 1 & \cos\left(2\pi\frac{2}{N_s}\right) & \sin\left(2\pi\frac{2}{N_s}\right) & \cos\left(4\pi\frac{2}{N_s}\right) & \cdots & \sin\left(2k_m\pi\frac{2}{N_s}\right) \\ 1 & \cos\left(2\pi\frac{3}{N_s}\right) & \sin\left(2\pi\frac{3}{N_s}\right) & \cos\left(4\pi\frac{3}{N_s}\right) & \cdots & \sin\left(2k_m\pi\frac{3}{N_s}\right) \\ 1 & \cos\left(2\pi\frac{4}{N_s}\right) & \sin\left(2\pi\frac{4}{N_s}\right) & \cos\left(4\pi\frac{4}{N_s}\right) & \cdots & \sin\left(2k_m\pi\frac{4}{N_s}\right) \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \cos\left(2\pi\frac{N_s}{N_s}\right) & \sin\left(2\pi\frac{N_s}{N_s}\right) & \cos\left(4\pi\frac{N_s}{N_s}\right) & \cdots & \sin\left(2k_m\pi\frac{N_s}{N_s}\right) \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ b_1 \\ a_2 \\ \vdots \\ b_{k_m} \end{pmatrix} \tag{1.4}$$

This is the exactly same form of high demensional Linear Algebra 1.1, where the unknown matrix $\boldsymbol{x}$ is $(a_0, a_1, b_1, a_2, b_2, \cdots, b_{k_m})^{\mathrm{T}}$. Therefore, an AD conversion problem can be seen as finding solution for a linear equation set.

To determining a unique solution for the unknown variables, the relation between the equation number M and the number of unknown variables N is intensely discussed. It is known that condition for getting a unique solution is that $M \geq N$. Therefore, for perfectly determine the analog signal, the condition is

$$N_s \geq 2k_m + 1. \tag{1.5}$$

Devide the both side by T, and take $T \to \infty$ as considering arbitrary signals, we obtain

$$f_s \geq 2f_m. \tag{1.6}$$

This is the result of sampling theorem, also termed as Nyquist-Shannon sampling theorem.

Theorem 1.2.1. Sampling theorem [45]
If a function $x(t)$ contains no frequencies higher than B hertz, it is completely determined by giving its ordinates at a series of points spaced $1/(2B)$ seconds apart. A sufficient sample-rate is therefore $2B$ samples per second, or anything larger. Equivalently, for a given sample rate f_s, perfect reconstruction is guaranteed possible for a bandlimit $B < \frac{f_s}{2}$.

Here, $2B$ and $f_s/2$ are respectively called the Nyquist rate and Nyquist frequency. In the field of digital signal processing, the sampling theorem, establishes a sufficient condition for a sample rate that permits a discrete sequence of samples to capture all the information from a continuous-time signal of finite bandwidth, and it was a important conventional limit found in the late 20^{th} century. Much effort has been made in the mordern technology to achieve this limit in the appearance of noise.

Note that the sampling theorem is the worst case condition. If additional condition for the signal is available, one can solve the problem by a fewer samples than that sampling theorem require. Next section, we will introduce an well noticed condition, sparsity, in modern information science. By noticing sparsity of the signal representation, we could break through the conventional limit of Nyquist rate.

1.3 Sparse representation

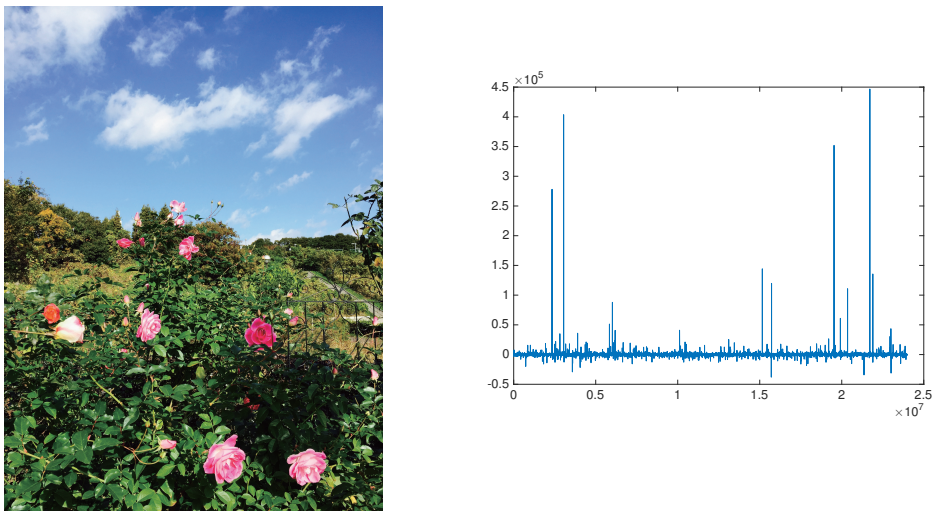

FIGURE 1.2: Left: Original JPG picture. Right: Plot of the coefficients of Haar wavelet conversion of the left picture (arranged in random order for enhanced visibility).

Typically, smooth signals, such as natural images and communications signals, can be represented by a sparsity-inducing basis, such as a Fourier or wavelet basis [17, 48]. Here is a example of sparse representation of a known picture, fig. 1.2. From the plot of the coefficients we can see that the biggest coefficiets which is only few percentatge of the whole number

of coefficients, are carrying most of the information of the original picture. Therefore, we can almost reconstruct the picture by only the value of few largest coeffients and seeing other small enough coeffients as zero.

The purpose of CS is to enhance signal processing performance by utilizing the notion of the *sparsity* of signals [16]–[48].

1.4 Compressed sensing

Compressed (or compressive) sensing (CS) is a technique for recovering a high-dimensional signal from lower-dimensional data, whose components represent partial information about the signal, by utilizing prior knowledge on the sparsity of the signal [7]. For the last decade, CS has received considerable attention as a novel technology in information science. It has been intensively investigated from the theoretical point of view [10, 6, 30, 19, 33], and has also been used in various engineering fields [9]. Many studies in CS research have shown that the sparsity of signals makes it possible to perfectly reconstruct the signal at a viable computational cost, even in the region of $\alpha = M/N < 1$ [6]–[33]. This results in time, cost, and precision advantages. It has led to the hardware-level realization of accurate signal reconstruction that had hitherto been regarded as out of reach due to limitations on sampling rates [51] and/or the number of sensors [15].

1.4.1 Problem setting

Mathematically, the compressed sensing problem can be expressed as follows: a sparse vector $\boldsymbol{x^0} \in \mathbb{R}^N$, many components of which are zero, is linearly transformed into vector $\boldsymbol{y} \in \mathbb{R}^M$ by an $M \times N$ measurement matrix $\boldsymbol{A}$, where

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x^0}. \tag{1.7}$$

The observer is free to choose the measurement protocol. For a given pair of $\boldsymbol{A}$ and $\boldsymbol{y}$, the reconstruction of $\boldsymbol{x^0}$ is required [7]. The goal of CS is to find a solution of 1.7 in the region of

$$M < N. \tag{1.8}$$

When $M < N$, due to the loss of information, the inverse problem has infinitely many solutions. However, when it is guaranteed that $\boldsymbol{x^0}$ has only $K < M$ nonzero entries in some convenient basis (i.e., when the signal is sparse enough) and the measurement matrix is incoherent with that basis, there is a high probability that the inverse problem has a unique and exact solution. Considerable efforts have been made to clarify the condition for the uniqueness and correctness of the solution, and to develop practically feasible signal reconstruction algorithms [10, 6, 30, 19, 33].

According to sparsity, the idea of CS is only sample or measure the signal M-times, in which $M < N$, even we do not know the position of the zeros we can still reconstruct the signal. Different from the traditional data compression, which usually obtain the full data then compress it to a smaller amount, CS compress the sampling precess itself to reduce the cost greatly.

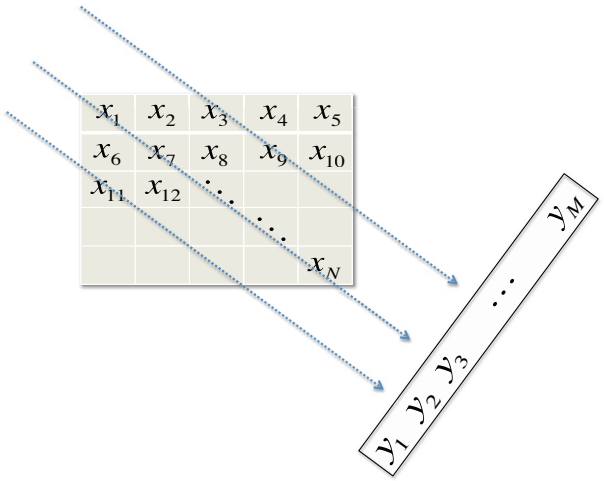

FIGURE 1.3: A CT projection example. Blue doted lines represent the process of CT scan.

1.4.2 Applications

Compressed sensing has been tremendously effected many fields in science and technology, such as audio and visual electronics, medical imaging devices, and astronomical observations, etc.[9]. Here we introduce some examples of the applications of compressed sensing.

- **CT scan [31]**
  A CT scan, also called X-ray computed tomography (X-ray CT) or computerized axial tomography scan (CAT scan), makes use of computer-processed combinations of many X-ray images taken from different angles to produce cross-sectional (tomographic) images (virtual 'slices') of specific areas of a scanned object, allowing the user to see inside the object without cutting.

  In fig. 1.3, $\boldsymbol{x} = (x_1, x_2, \cdots, x_N)$ represents the X-ray absorption coefficients vector. $\boldsymbol{y} = (y_1, y_2, \cdots, y_M)$ represents the projection data. $\boldsymbol{A}$ carrys the projecting angle information. Assume that X-ray absorption coefficients are the same in the same parts of the body, they only change in the boundary of the parts. The gradient of the slice image can be expressed as a sparse vector. Utilizing this sparsity, we can apply the technology of CS to reduce the needed sample number, therefore, to keep the damage to the patient's body as small as possible. Also, we can design smaller scan devices to save the space and make it convenient to carry. The similar application also works for other medical imaging technology, for example MRI.

- **Astronomical observations [22]**
  In order to proof the existence of black hole in the universe, an international cooperated research project about black hole direct imaging is builted. Such a big scale experiment cost a huge amount of effort to get the sampling data, and still the data is not fully enough for directly solving the basic equations. In this situation, CS technology can help to solve the ill-posed problem and reconstruct the image of black hole.

- **Wireless channel estimation [21]**
  Channel estimation is one of the most important techniques in wireless communications systems, because a lot of modern communication technologies assume availability of channel state information. Reduction of the required amount of training signals while keeping a sufficient estimation accuracy has been one of the main scopes of the study on channel estimation. It s known that the impulse response of wirless channel tends to be sparse for larger bandwidth. Therefore, several works have tried to utilize the saprsity for the channel estimation. CS is one of the most resent example. Here, we briefly review a simple approach to sparse channel estimation with a naive assumption that the channel impulse response itself is sparse in time domain. Let $\boldsymbol{a} = [a_1, a_2, \cdots, a_p]^{\mathrm{T}} \in \mathbb{R}^P$ denote a vector of training signals for the channel estimation, which is inserted between data signals, and $\boldsymbol{x} = [x_1, x_2, \cdots, x_L]^{\mathrm{T}} \in \mathbb{R}^L$ be a vector of finite channel impulse response wth $||\boldsymbol{x}||_0 << L$. Assuming $P > L$, the corresponding received signal vector $\boldsymbol{y} = [y_1, y_2, \cdots, y_{P-L+1}]^{\mathrm{T}}$, which is not contaminated by data signals. Definding the Toeplitz channel matrix $\boldsymbol{T}$ of size $(P - L + 1) \times P$ as

$$\boldsymbol{T} = \begin{pmatrix} x_L & \cdots & x_1 & 0 & \cdots & 0 \\ 0 & x_L & \cdots & x_1 & \cdots & \vdots \\ \vdots & & \ddots & & \cdots & 0 \\ 0 & \cdots & 0 & x_L & \cdots & x_1 \end{pmatrix}, \tag{1.9}$$

  $\boldsymbol{y}$ can be written as

$$\boldsymbol{y} = \boldsymbol{T}\boldsymbol{a} + \boldsymbol{v}, \tag{1.10}$$

  where $\boldsymbol{v}$ is an additive white noise vector. By using the channel impulse response vector $\boldsymbol{x}$, $\boldsymbol{y}$ can be alternatively be written as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v}, \tag{1.11}$$

  where $\boldsymbol{A}$ is a Toeplitz matrix of size $(P - L + 1) \times L$ defined as

$$\boldsymbol{A} = \begin{pmatrix} a_L & a_{L-1} & \cdots & a_1 \\ a_{L+1} & a_L & \cdots & a_2 \\ \vdots & \vdots & & \vdots \\ a_P & a_{P-1} & \cdots & a_{P-l+1} \end{pmatrix}. \tag{1.12}$$

  Thus, by regarding $\boldsymbol{A}$ to be a sensing matrix and assuming $P - L + 1 < L$ in order to achieve higher spectral efficiency, the problem to obtain $\boldsymbol{x}$ from $\boldsymbol{y}$ in 1.11 can be considered as a problem of CS with observation noise.

1.5 Algorithms for compressed sensing[17]

Intuitively, we see the additional condition to $\boldsymbol{x}$ as a sparse vector make this problem solvable. However, how to solve it is not trivial. Here are two questions we need to ask:

- Does the problem have solutions? If yes, is the solution unique or not?
- How to make sure that the solution is sparse?

The following part of this section will answer these two qustions.

1.5.1 Regularization

The system 1.7 with $M < N$ has more unknowns than equations, which is termed underdetermined linear system. Thus it has either no solution, if $\boldsymbol{y}$ is not in the span of the columns of the matrix $\boldsymbol{A}$, or infinitely many solutions. In order to avoid the anomany of having no solution, we shall hereafter assume that $\boldsymbol{A}$ is a full-rank matrix, impling that its columns span the entire space $\mathbb{R}^M$.

In most underdetermined linear system, we desire a signle solution, however, the fact that there are infinitely many of those stands as a major obstacle. In order to narrow the solution space, additional criteria are needed. A familiar way to do this is regularization which also called as the penalty term, where a function $\boldsymbol{J}(\boldsymbol{x})$ that evaluates a candidate solution $\boldsymbol{x}$, with smaller values being preferred. The linear equaltion problem then is transformed to a optimization problem

$$\min_{\boldsymbol{x}} \boldsymbol{J}(\boldsymbol{x}) \text{ subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}. \tag{1.13}$$

l_2-norm

First, let us look at a historically widely used example of norm penalty, the l_2-norm. The optimization problem becomes

$$\min_{\boldsymbol{x}} ||\boldsymbol{x}||_2 \text{ subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \tag{1.14}$$

where $||\boldsymbol{x}||_2 = \sqrt{x_1^2 + \cdots + x_N^2}$.

Here is the process to solve 1.14. Introducing Lagrange multipliers λ for the constraint set, we define the Lagrangian

$$\mathcal{L}(\boldsymbol{x}) = ||\boldsymbol{x}||_2^2 + \lambda^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}). \tag{1.15}$$

To solve the optimization problem, we require $\frac{\partial \mathcal{L}(\boldsymbol{x})}{\partial \boldsymbol{x}} = 0$. Thus the solution is obtained as $\hat{\boldsymbol{x}}_{opt} = -\frac{1}{2}\boldsymbol{A}^T\lambda$. Plugging this solution into the constraint $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ leads to

$$\lambda = -2(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{y}. \tag{1.16}$$

Assigning this to $\hat{\boldsymbol{x}}_{opt}$ gives the pseudo-inverse solution

$$\hat{\boldsymbol{x}}_{opt} = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{y} = \boldsymbol{A}^+\boldsymbol{y}. \tag{1.17}$$

Note that since we have assumed that $\boldsymbol{A}$ is full-rank, the matrix $\boldsymbol{A}\boldsymbol{A}^T$ is positive-definite and thus invertible.

Due to its simplicity as manifested by the above closed-form and unique solution, the l_2-norm is widespread in various fields of engineering, such as signal and image processing. However, this is by no means a declaration that l_2 is the truly best choice for the various problems. In many cases,

the mathematical simplicity of l_2 is a misleading fact that diverts engineers from a making better choice for the regularization.

l_0-norm

The most direct way to put sparsity in the problem is using the l_0-norm as the penalty

$$\min_{\boldsymbol{x}} ||\boldsymbol{x}||_0 \text{ subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \tag{1.18}$$

where $||\boldsymbol{x}||_0 = \lim_{p\to 0} ||\boldsymbol{x}||_p^p = \#\{i : x_i \neq 0\}$, which means the number of non-zerol entries in $\boldsymbol{x}$. However, due to the decrete and discontinuous nature of l_0-norm, the standard convex analysis ideas do not apply. Moreover, many of the basic questions about 1.18 remain unclear, such as the existence of the unique solution and the way to varify a solution is actually the global minimizer of 1.18 or not. Besides these conceptual issues, the computational difficulty cannot be underestimated. It is well know that 1.18 is NP-hard in general. Assume we know that the sparsest solution to 1.18 has K non-zeros. The number of possible solutions is C_N^K, which is a extreamly huge number in the as the dimention increases. For example, when $N = 500, K = 20$, the combination number $C_N^K \approx 3.9 \times 10^{47}$. And for each test we need to solve the linear equation as well. Suppose each test of such system cost 1×10^{-9} seconds, the whole search will take more than 1.2×10^{31} years. Therefore, we need to find a replacement penalty function to solve the sparse signal optimization problems.

1.5.2 Convex relaxation

The idea of convexity can help us to find the proper penalty function. Let us recall the definition of convexity for sets and for functions

Definition 1.5.1. A set Ω is **convex if** $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \Omega$ **and** $\forall t \in [0, 1]$**, the convex combination** $\boldsymbol{x} = t\boldsymbol{x}_1 + (1-t)\boldsymbol{x}_2$ **is also in** Ω**.**

Definition 1.5.2. A function $f(\boldsymbol{x}) : \Omega \to \mathbb{R}$ is **convex if** $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \Omega$ **and** $\forall t \in [0, 1]$**, the convex combination point** $\boldsymbol{x} = t\boldsymbol{x}_1 + (1-t)\boldsymbol{x}_2$ **satisfies**

$$f(t\boldsymbol{x}_1 + (1-t)\boldsymbol{x}_2) \leq tf(\boldsymbol{x}_1) + (1-t)f(\boldsymbol{x}_2). \tag{1.19}$$

If $f(\cdot)$ is twice continuously differentialble, its derivatives can be used for alternative definitions of convexity:

Definition 1.5.3. A function $f(\boldsymbol{x}) : \Omega \to \mathbb{R}$ is **convex if** $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \Omega$ **if and only if**

$$f(\boldsymbol{x}_2) \geq f(\boldsymbol{x}_1) + \nabla f(\boldsymbol{x}_1)^T(\boldsymbol{x}_2 - \boldsymbol{x}_1), \tag{1.20}$$

or alternatively, if and only if $\nabla^2 f(\boldsymbol{x}_1)$ is positive-definite.

We can verfy the set $\Omega = \{\boldsymbol{x}|\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}\}$ is convex using the definition of 1.5.1. Thus, the feasible set of solutions of the problem 1.13 is convex. In order for this optimization problem to be convex as a whole, the penalty need to be convex as well.

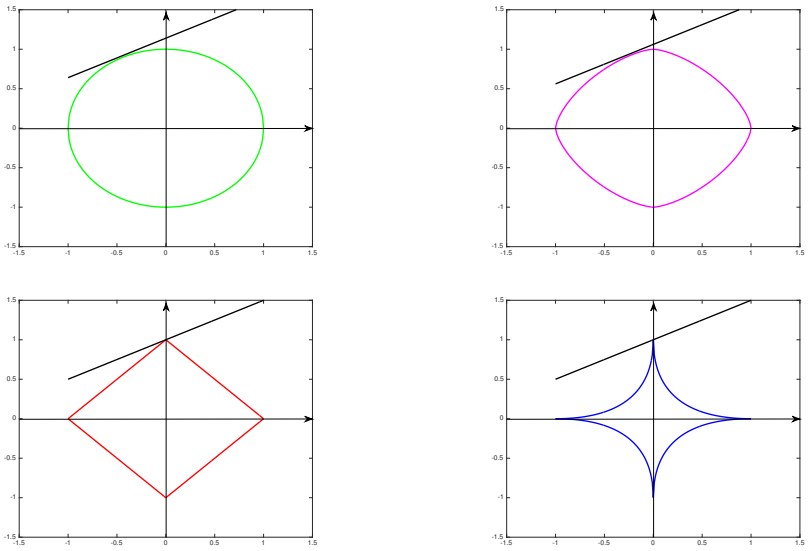

FIGURE 1.4: A demonstration of intersection for two dimensional $\boldsymbol{x}$ and several p values: 2, 1.5, 1, and 0.5, for l_p-norm minimization problem. Black solid lines represent the linear constraint.

An intensively discussed form of penalty is l_p-norms

$$||\boldsymbol{x}||_p \equiv \left(\sum_i^N |x_i|^p \right)^{1/p}. \tag{1.21}$$

For any $p \in \mathbb{N}$, the l_p-norm are convex. This can be proved by the definition 1.5.3. Take l_2-norm as an example, using the definition of 1.5.3, since $\nabla^2 ||\boldsymbol{x}||_2^2 = 2\boldsymbol{I} \succeq \boldsymbol{0}$ for all $\boldsymbol{x}$, the l_2-norm is strictly convex. Therefore, the costraint-set and the penalty are both strictly convex in optimization problem 1.14, a unique slution is guaranteed.

Besides the requirment of convexity, the penalty function should have a tendency to promote the sparsity of $\boldsymbol{x}$. Geometrical way of looking at the constrained l_p-norm minimization problem

$$\min_{\boldsymbol{x}} ||\boldsymbol{x}||_p \text{ subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \tag{1.22}$$

where we exand $p \in \mathbb{R}$, may help us. The linear equation constraint defined a set of solutions that appears as a hyperplane of dimensionality $\mathbb{R}^{N-M}$ embedded in the $\mathbb{R}^N$ space. With in this space, we seek the minimum of the l_p-norm. This search is done by "blowing" an l_p balloon centered around the origin, and stopping its inflation when it first touches the feasible set of the constraint-set. Fig.1.4 present a simple demonstration of this process for two dimentional $\boldsymbol{x}$ and several p values: 2, 1.5, 1, and 0.5. One can see that norms with $p \leq 1$ tend to give that the intersection point on the on the axes, which is a sparse solution. On the other hand, l_2 and $l_{1.5}$ give non-sparse solutions. Note that such l_p for $p < 1$ are no longer norms, as the traiangle inequality is no longer satisfied. Nevertheless, we shall use the term norm for these funtions a well.

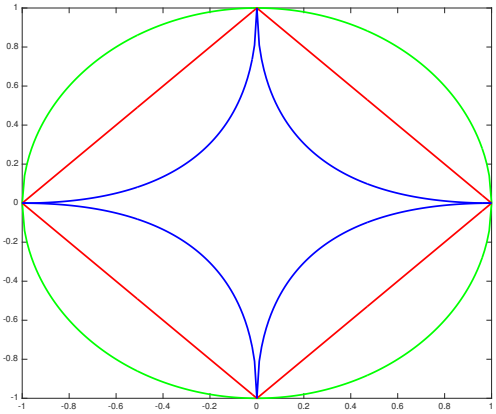

FIGURE 1.5: Demonstrating the fact that a unit-length l_p-norm vector becomes shortest in l_q ($p > q$) when it is the sparsest possible. The blue, red and green lines represent $p = 0.5, p = 1, p = 2$ unit-length l_p-norm respectively.

Another way to show this is fig.1.5, which is demonstrating the fact that a unit-length l_p-norm vector becomes shortest in l_q ($p > q$) when it is the sparsest possible. The blue, red and green lines represent $p = 0.5, p = 1, p = 2$ unit-length l_p-norm respectively, in two dimentional world.

Therefore, the widely used l_2-norm is not suitable for CS as it fails to give us sparse solution. The region of $0 < P \leq 1$ becomes our concern of interst. Unfortunately, each choice $0 < p < 1$ leads to a non-convexs optimization problem, which raises many difficulties. Naturally, l_1-norm become our first choice of prefernce to promoting sparsity and holding convexity in the same time.

l_1-norm

Let us look at l_1-norm more closely. The choice $f(\boldsymbol{x}) = ||\boldsymbol{x}||_1$ is convex but not strictly so. Thus the problem

$$\min_{\boldsymbol{x}} ||\boldsymbol{x}||_1 \text{ subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{1.23}$$

may have more than one solutions. Nevertheless, we can claim that all solution are nearby. In other words, these solutions are gathered in a set that is bounded and convex [17].

The cardinality function is a non-decreasing submodular function. It has been proven that, the l_1-norm is the convex envelop of the l_0-norm [3]. Therefore, l_1-norm is a reasonable choice of convex relaxation of combinatorial penalty. Typical l_1 based algorithm are Basis Persuit algorithm and Lasso, which have been widely used in engineering.

1.5.3 Greedy algorithms

Another way to avoid the exhaustive search in l_0-norm minimization is a type of algorithm that called greedy algorithms. The basic idea is that to see the sample vector $\boldsymbol{y}$ as a linear combination of columns $\boldsymbol{a}_i$ ($i = 1, 2 \cdots N$) of

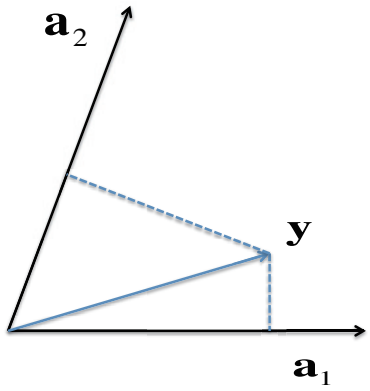

FIGURE 1.6: Sample vector $\boldsymbol{y}$ projects on column vector. Here we only show $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ as examples. Dashed lines represent vertical projections. The largest projection will be choosen as a support.

matrix $\boldsymbol{A}$,

$$\boldsymbol{y} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_N\boldsymbol{a}_N. \tag{1.24}$$

The goal is to construct an appropriate set of columns whose coefficients are non-zero, which is termed **support** S, by updating the support set one by one in each iteration.

Here, we inroduce a typical greedy algorithm, Orthogonal Matching Pursuit (OMP), for the CS problem. In the first iteration of OMP, we evaluate the quality of approximation of the column $\boldsymbol{a}_i$ by

$$\epsilon(i) = \min_{x_i} |\boldsymbol{y} - x_i\boldsymbol{a}_i|^2 = |\boldsymbol{y}|^2 - \frac{(\boldsymbol{a}_i \cdot \boldsymbol{y})^2}{|\boldsymbol{a}_i|^2}. \tag{1.25}$$

Then we select the column that has the smallest ϵ_i to put in to the support set S. This can be also seen as the column that has the largest projection from $\boldsymbol{y}$, see fig.1.6. The solution of this approximation by the support set of the moment can be obtained by

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}_S}{\operatorname{argmin}} |\boldsymbol{y} - \boldsymbol{A}_S\boldsymbol{x}_S|^2, \tag{1.26}$$

where $\boldsymbol{A}_S$ represents matrix only keep the columns in the support set and $\boldsymbol{x}_S$ stands for a vector that only contains the elements that corresponding to elements in S. Remove the approximated part in $\{\boldsymbol{a}_i\}$ and continue the same procedure iteratively to the residual

$$\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{A}_S\hat{\boldsymbol{x}}. \tag{1.27}$$

In other words, we use $\boldsymbol{r}$ instead of $\boldsymbol{y}$ in 1.25 in later iterations. We set a stopping threshold ϵ_0 to measure the norm of the residual till

$$|\boldsymbol{r}| < \epsilon_0 \tag{1.28}$$

holds.

A summery of the OMP algorithm is shown in fig.1.7.

Algorithm 1: ORTHOGONAL MATCHING PURSUIT ALGORITHM($\boldsymbol{x}^0, \boldsymbol{r}^0, S^0, \epsilon_0$)

1) **Initialization** :
   Solution seed : $\boldsymbol{x}^0 \leftarrow \boldsymbol{0}$
   Residual seed : $\boldsymbol{r}^0 \leftarrow \boldsymbol{y}$
   Support set seed : $S^0 \leftarrow \emptyset$
   Counter : $k \leftarrow 0$
2) **Counter increase** :
   $k \leftarrow k + 1$
3) **Sweep** :
   Compute
   $\epsilon(i) = \min_{x_i} |\boldsymbol{r}^{k-1} - x_i \boldsymbol{a}_i|^2 = |\boldsymbol{r}^{k-1}|^2 - \frac{\left(\boldsymbol{a}_i \cdot \boldsymbol{r}^{k-1}\right)^2}{|\boldsymbol{a}_i|^2}$
   for all $i \notin S^{k-1}$
4) **Update Support** :
   $i_0 = \underset{i \notin S^{k-1}}{\text{argmin}}\{\epsilon(i)\}, S^k = S^{k-1} \cup \{i_0\}$
5) **Update Solution** :
   $\hat{\boldsymbol{x}}^k = \underset{\boldsymbol{x}_{S^k}}{\text{argmin}} |\boldsymbol{y} - \boldsymbol{A}_{S^k} \boldsymbol{x}_{S^k}|^2$
6) **Update Residual** :
   $\boldsymbol{r}^k = \boldsymbol{y} - \boldsymbol{A}_{S^k} \hat{\boldsymbol{x}}^k$
7) **Stopping Rule** :
   Stop if $|\boldsymbol{r}^k| < \epsilon_0$ holds.
   Otherwise, apply another iteration.

FIGURE 1.7: OMP algorithm. A typical greedy algorithm.

There are also many varieties of greedy algorithms of the similer spirit. For example,

- Matching Pursuit (MP)
  Get lazy in approximating solution by using privious solution

$$\hat{\boldsymbol{x}}^k = \hat{\boldsymbol{x}}^{k-1} + \frac{\left(\boldsymbol{a}_{i_0} \cdot \boldsymbol{r}^{k-1}\right)}{|\boldsymbol{a}_{i_0}|^2}\boldsymbol{a}_{i_0} \tag{1.29}$$

  for saving cost.

- Week-MP
  In sweep, get lazy in optimizing $\epsilon(i)$. Instead of compute for all i, stop the sweep when the condition

$$\frac{\left(\boldsymbol{a}_i \cdot \boldsymbol{r}^{k-1}\right)}{|\boldsymbol{a}_i|^2} \geq t|\boldsymbol{r}^{k-1}|^2 \tag{1.30}$$

  where $r \in [0, 1]$, is satisfied.

- Least Squars OMP (LS-OMP)
  In sweep, approximation error ϵ_i is evaluated for $S^{k-1} \cup \{i\}$, which can provide more accurate evaluation than OMP although cost increases.

- Thresholding algorithm
  Given a sopport size T and fix Support by the result of the first quality evaluation. In other words, S is a set of indices of T lowest values of ϵ_i.

Although greedy algorithms only lead to a local minimum in general instead of the optimal solution for 1.18, it may outperform the l_1 optimization in some cases.

There are another branch of approach using probabilistic inference to this reconstruction problem. We will discuss it in later chapters as the main part of this thesis.

1.6 Performance guarantee for compressed sensing algorithms

1.6.1 The worst-case study

For the underdetermined linear system of equations, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ (a full-rank matrix $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ with $M < N$), we pose the following questions:

- When can we claim the uniqueness of the sparsest solution?
- Can a candidate solution be tested to verify its global optimality?

In this section, we show theoretical studies that answer these questions.

A key property for the study of uniqueness is the *spark* of the matrix $\boldsymbol{A}$, a term defined by Donoho and Elad in 2003. The definition of *spark* is

Definition 1.6.1. The spark of a given matrix $\boldsymbol{A}$ is the smallest number of columns from $\boldsymbol{A}$ that are linearly-dependent.

By definition, the vectors in the null-space of the matrix $\boldsymbol{Ax} = \boldsymbol{0}$ must satisfy $||\boldsymbol{x}||_0 \geq spark(\boldsymbol{A})$. Let us suppose that the underdetermined linear equation $\boldsymbol{y} = \boldsymbol{Ax}$ has two solutions $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. This implies that $\boldsymbol{x}_1 - \boldsymbol{x}_2$ must be in the null space of $\boldsymbol{A}$, i.e., $\boldsymbol{A}(\boldsymbol{x}_1 - \boldsymbol{x}_2) = \boldsymbol{0}$. By employing the triangle inqueality and the definition of *spark*, we obtain

$$||\boldsymbol{x}_1||_0 + ||\boldsymbol{x}_2||_0 \geq ||\boldsymbol{x}_1 - \boldsymbol{x}_2||_0 \geq spark(\boldsymbol{A}). \tag{1.31}$$

This means that the underdetermined linear equation cannot have two solutions of $||\boldsymbol{x}||_0 \leq \frac{\text{spark}(\boldsymbol{A})}{2}$ simultaneously, which gives a simple criterion for uniqueness of sparse solutions:

Theorem 1.6.1. *If a system of linear equations $\boldsymbol{Ax} = \boldsymbol{y}$ has a solution $\boldsymbol{x}$ obeying*

$$||\boldsymbol{x}||_0 < \frac{spark(\boldsymbol{A})}{2}, \tag{1.32}$$

this solution is necessarily the sparsest possible.

When we have a solution satisfying this condition, we can conclude that any alternative solution necessarily has more than $spark(\boldsymbol{A})/2$ non-zeros.

This seemingly simple result is wonderful and powerful. We have a complete answer for the two questions we have posed at the begining of this section. Namely, we can certainly claim uniqueness for sparse enough solutions, and once such a solution is found, we can immediately claim its global optimality. Notice that l_0 optimization is a highly complicated optimization task of cominatorial flavor. In general combinatorial optimization problems, a given solution can at best be verified as being locally optimal, that no simple modificatons gives a better result. Here, we have the ability to verify its optimality globally.

Clearly, the value of *spark* is very informative and larger values are more useful. By definition, *spark* must be in the range

$$2 \leq spark(\boldsymbol{A}) \leq M + 1. \tag{1.33}$$

When elements of $\boldsymbol{A}$ are random independent and identically distributed from standard Gaussian distribution, then we have

$$spark(\boldsymbol{A}) = M + 1, \tag{1.34}$$

implying that no M columns are linearly-dependent. Similarly, for the two-ortho identity-Fourier matrix $[\boldsymbol{I}, \boldsymbol{F}]$, the *spark* is $2\sqrt{M}$.

Unfortunately, in genneral, the computation of *spark* of a given matrix $\boldsymbol{A}$ is difficult, as it calls for a combinatorial search over all possible subset of columns from $\boldsymbol{A}$. For practically overcoming this difficulty, lower bounding spark by *mutual coherence* is proposed. The definition of mutual cohenrence is as follows:

Definition 1.6.2. The mutual-coherence of a given matrix $\boldsymbol{A}$ is the largest absolute normalized inner product between different columns from $\boldsymbol{A}$.Denoting the k-th column in $\boldsymbol{A}$ by $\boldsymbol{a}_k$, the mutual-coherence is given by

$$\mu(\boldsymbol{A}) = \min_{1 \leq i,j \leq N, i \neq j} \frac{|\boldsymbol{a}_i^T \boldsymbol{a}_j|}{||\boldsymbol{a}_i||_2 ||\boldsymbol{a}_j||_2}. \tag{1.35}$$

Mutual-coherence is a way to characterize the dependence between columns of the matrix $\boldsymbol{A}$. The evaluation of 1.35 is computationally feasible, and as such, it allows us to lower-bound the *spark*.

Lemma 1.6.2. *For any matrix $\boldsymbol{A} \in \mathbb{R}^{M\times N}$, the following relationship holds:*

$$spark(\boldsymbol{A}) \geq 1 + \frac{1}{\mu(\boldsymbol{A})}. \tag{1.36}$$

The proof of this lemma can be found in book [17]. Therefore, We have the following theorem on uniqueness based on mutual-coherence.

Theorem 1.6.3. *If a system of linear equations $\boldsymbol{Ax} = \boldsymbol{y}$ has a solution $\boldsymbol{x}$ obeying*

$$||\boldsymbol{x}||_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(\boldsymbol{A})}\right), \tag{1.37}$$

this solution is necessarily the sparsest possible.

Unfortunately, this replacement of the bound generally loosens the assessment accuracy considerably. When each entry of $\boldsymbol{A}$ is chosen from a Gaussian distribution of zero mean and a fixed variance, the typical value of the mutual-coherence can never be smaller than $1/\sqrt{M}$, while the *spark* can easily be as large as M.

These theoretical studies provide the guaratees of many algorithms. For example,

Theorem 1.6.4. *(**Equivalence - Orthogonal Greedy Algorithm**): For a system of linear equations $\boldsymbol{Ax} = \boldsymbol{y}$ ($\boldsymbol{A} \in \mathbb{R}^{M\times N}$ full-rank with $M < N$), if a solution $\boldsymbol{x}$ exists obeying*

$$||\boldsymbol{x}||_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(\boldsymbol{A})}\right), \tag{1.38}$$

OMP run with threshold parameter $\epsilon_0 = 0$ is guaranteed to find it exactly.

Theorem 1.6.5. *(**Equivalence - Basis Pursuit**): For a system of linear equations $\boldsymbol{Ax} = \boldsymbol{y}$ ($\boldsymbol{A} \in \mathbb{R}^{M\times N}$ full-rank with $M < N$), if a solution $\boldsymbol{x}$ exists obeying*

$$||\boldsymbol{x}||_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(\boldsymbol{A})}\right), \tag{1.39}$$

that solution is both the unique solution of l_1-norm minimization problem and the unique solution of l_0-norm minimization problem.

The proof of these theorem can be found in book [17].

The above-discribed theorems motivate using OMP and Basis Persuit to approximate the solution of l_0-norm minimization problem. However, note that, these results are the worst-case study, they are week and lead to overly pessimistic prediction of performance. In other words, they guarantee success only when the object is extremely sparse, with fewer than $\sqrt{M}$ non-zeros out of M. In many cases such sparsity is out of the question.

Actually, in the case of random matrices $\boldsymbol{A}$, much more favorable results are possible. Empirical evidence shows that OMP and Basis Persuit perform well even in situations violating the above bounds. Naturally, considering a probabilistic viewpont that yields reasonable and more optiistic bounds are needed.

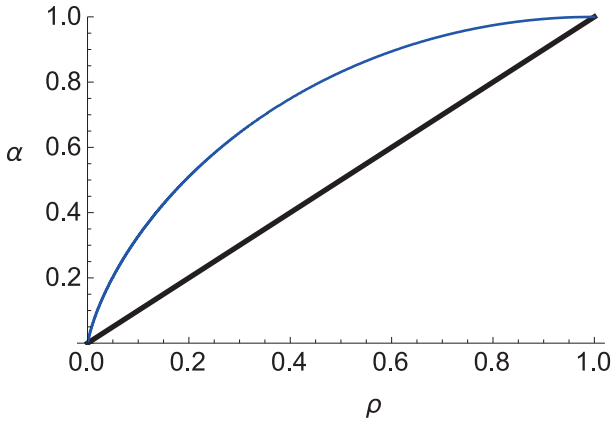

FIGURE 1.8: Comparison of typical reconstruction limits of the l_p-reconstruction for $p = 0, 1$. The blue curve represents $p = 1$ and the black line represents $p = 0$.

1.6.2 The statistical evaluation

In the 19^{th} century, the discovery of atoms and molecules has opened the door of the microscopic world. A theory to connect the microscopic and the macroscopic worlds was needed, which gave birth to statistical mechanics. Statistical mechanics was the first foundational physical theory in which probabilistic concepts and probabilistic explanation played a fundamental role. A common use of statistical mechanics is in explaining the thermodynamic behaviour of large systems.The benefit of using statistical mechanics is that it provides exact methods to connect thermodynamic quantities (such as heat capacity) to microscopic behavior. Since the framework of statistical mechanics is general, its application domains are wide-ranging. Nowadays, it starts to come in to the world of information, which is abstract and different from the traditional physical object. The connection between microscopic and macroscopic behaviors is the core of the so-called "big data" research, which is also the essence of statistical mechanics. The common goal of physics and information science is to infer the behaviour of the system from the observed data. Realizing the same essence, can crash down the barrier between different research areas and enhance the ability of the whole. This thesis is an example of how method developed in physics can empower the theory and algorithm in information science.

To statistically evaluate the performance, we need to assume the probabilistic distribution for signal $\boldsymbol{x}$ and measurement matrix $\boldsymbol{A}$. The reconstruction performance avaraged by the probabilistic distribution of the measurement matrix and the signal is called the *typical* preformance.

A typical reconstruction limit for CS with standard Gaussian signals and Guassian random measurement matrices based on l_p-norm minimization is presented in [30]. In fig.1.8, each curve represents the Replica Symmetric estimate of the typical critical compression rate $\alpha_c(\rho)$ for the signal density ρ of the original signal for the l_p-reconstruction scheme. Correct reconstruction is typically possible for $\alpha > \alpha_c(\rho)$. After submitting letter [30], the authors noticed that the typical criticality of the l_1-reconstruction was explored in [11, 13] for compression matrices consisting of i.i.d. zero-mean Gaussian random column vectors utilizing techniques of combinatorial geometry. It turns out that [11, 13]'s weak threshold corresponds to [30]'s

result for $\alpha_c(\rho)$ with $p = 1$. In view of [11, 13]'s criterion of reconstruction success, in which no errors are allowed, [30]'s result implies that the criticality of l_1-reconstruction is 'tight' in the sense that it does not change irrespective of whether or not we allow small errors which are vanishing asymptotically as $N \to \infty$. The connection between [30]'s analysis and [11, 13]'s further suggests a possibility of wide application of statistical mechanics tools to problems in large-dimensional random combinatorial geometry.

1.7 1-bit compressed sensing

Quantizing continuous data is unavoidable in most real-world applications, particularly those in which the measurement is accompanied by digital information transmission [34]. In the signal processing context, the CS framework eases the burden on analog-to-digital converters (ADCs) by reducing the sampling rate required to acquire and recover sparse signals. However, in practice, ADCs not only sample, but also quantize each measurement to a finite number of bits; moreover, there is an inverse relationship between achievable sampling rate and bit depth. Therefore, many discussions on CS have shifted emphasis from sampling rate to number of bits per measurement [43, 18]. Also, there are many researches about how to find a satisfactory quantized representation of real number measurement and requires preprocessing based on real number measurements. However, in this thesis, we are more interested of directly minimize MSE instead of minimize the residual between quantized representation and real number measurement.

The extream case of quantization is to only keep the sign information. Addressing the practical relevance of CS in such operation, Boufounos and Baraniuk proposed and examined a CS scheme, often called *1-bit compressed sensing (1-bit CS)*, in which the signal is recovered from only the sign data of the linear measurements

$$\boldsymbol{y} = \text{sign}(\boldsymbol{A}\boldsymbol{x}^{\mathbf{0}}), \tag{1.40}$$

where $\text{sign}(x) = x/|x|$ for $x \neq 0$ operates for vectors in the component-wise manner [49]. Thus, the measurement operator is a mapping from $\mathbb{R}^N$ to the Boolean cube $B^M := \{-1, 1\}^M$. Since the linear constraint becomes a sign information, when $y_i = 1$, we have

$$\sum_j A_{ij} x_j > 0, \tag{1.41}$$

and $y_i = -1$ means the opposite. In fig.1.4, black solid lines represent the linear constraint in CS for two dimensional case, instead of that, the 1-bit CS measurement 1.40 is not giving us a line, but a half space which cutted from the line. Therefore, we need more measurements to tell us a small region of consist set of all the 1-bit constraints.

Note that, here what we have compressed is the measurement data, which is different form reduce the measuring process in CS that we have talked so on. In other words, for 1-bit CS problem, we are working in the region of $M > N$, and the application is different from CS. The reason why I started from introducing CS is because of the historical reason and the

background that 1-bit CS can rely on. The scheme of 1-bit CS is considered practical relevant in situations where measurements are inexpensive and precise quantization is expensive, for example wirless communication, in which the cost of measurements should be quantified by the number of total bits needed to store the data instead of by the number of measurements. Discarding the amplitude information can significantly reduce the amount of data that needs to be stored and/or transmitted. This is highly advantageous when perfect recovery is not required. In addition, quantization to the 1-bit (sign) information is appealing in hardware implementations because 1-bit quantizer takes the form of a comparator to zero and does not suffer from dynamic range issues.

1.8 Aim and outline of this thesis

The aim of this thesis is to clarify the typical performance of several approaches to 1-bit compressed sensing and to develop fast and practically feasible signal reconstruction algorithms for the problem.

In chapter 2, we analyze the typical performance of an l_1-norm based signal recovery scheme for the 1-bit compressed sensing using statistical mechanics methods. We also develop another approximate recovery algorithm inspired by the cavity method. In chapter 3, to further develop the study of l_1-norm based signal recovery scheme for 1-bit compressed sensing, we suggest a strategy that captures scale information by introducing a threshold parameter to the quantization process. For practical use, we develop a heuristic that adaptively tunes the threshold parameter based on measurement results. In chapter 4, we present a Bayesian approach to signal reconstruction for 1-bit compressed sensing, which is a statistical lower bound of the recovery performance of 1-bit measurements. We also test a message passing algorithm for signal reconstruction on the basis of belief propagation. Conclusion, appendix and bibliography are stated in the end.

Chapter 2

l_1-norm minimization approach

The purpose of this chapter is to explore the abilities and limitations of a 1-bit CS scheme utilizing statistical mechanics methods. In [49] an approximate signal recovery algorithm based on minimization of the l_1-norm $||\boldsymbol{x}||_1 = \sum_{i=1}^{N} |x_i|$ under the constraint of $\mathrm{sign}\left(\boldsymbol{\Phi x}\right) = \boldsymbol{y}$ was proposed and its utility was shown by numerical experiments. Quantization to the sign information, however, leads to the loss of the convexity of the resulting optimization problem, which makes it difficult to mathematically examine how well the obtained solution approximates the correct solution. Comparing (in terms of the mean square error) the results of numerical experiments with the theoretical prediction evaluated by the replica method [14], we will show that the performance of the approximate algorithm is nearly as good as that potentially achievable by the l_1-based scheme. We will also develop another approximate algorithm inspired by the cavity method [38, 37] and will show that when the density of nonzero entries of the original signal is relatively high the new algorithm offers better recovery performance with much lower computational cost.

This chapter is organized as follows. The next section sets up the problem that we will focus on when explaining the 1-bit CS scheme. Section 3 uses the replica method to examine the signal recovery performance achieved by the scheme. In section 4 an approximate signal recovery algorithm based on the cavity method is developed and evaluated, and the final section is devoted to a summary.

The result in this chapter is published in [56].

2.1 Problem setup

Let us suppose that entry x_i^0 $(i = 1, 2, \ldots, N)$ of N-dimensional signal (vector) $\boldsymbol{x}^0 \in \mathbb{R}^N$ is independently generated from an identical sparse distribution:

$$P\left(x\right) = \left(1-\rho\right)\delta\left(x\right) + \rho\tilde{P}\left(x\right), \tag{2.1}$$

where $\rho \in [0, 1]$ represents the density of nonzero entries in the signal and $\tilde{P}(x)$ is a distribution function of $x \in \mathbb{R}$ that does not have finite mass at $x = 0$. In 1-bit CS the measurement is performed as

$$\boldsymbol{y} = \mathrm{sign}\left(\boldsymbol{\Phi x^0}\right), \tag{2.2}$$

where for simplicity we assume that each entry of $M \times N$ measurement matrix $\boldsymbol{\Phi}$ is provided as an independent sample from an identical Gaussian distribution of zero mean and variance N^{-1}.

Given $\boldsymbol{y}$ and $\boldsymbol{\Phi}$, the signal reconstruction is carried out by searching for a sparse vector $\boldsymbol{x} = (x_i) \in \mathbb{R}^N$ under the constraint of $\mathrm{sign}\,(\boldsymbol{\Phi x}) = \boldsymbol{y}$. For this task the authors of [49] proposed a scheme of

$$\min_{\boldsymbol{x}} \{||\boldsymbol{x}||_1\} \text{ subj. to } \mathrm{sign}\,(\boldsymbol{\Phi x}) = \boldsymbol{y} \text{ and } ||\boldsymbol{x}||_2 = \sqrt{N}, \tag{2.3}$$

based on the l_1-recovery method widely used and studied for standard CS problems [7]. Here $||\boldsymbol{x}||_1 = \sum_{i=1}^N |x_i|$ and $||\boldsymbol{x}||_2 = |\boldsymbol{x}| = \sqrt{\sum_{i=1}^N x_i^2}$ denote the l_1- and l_2-norms of $\boldsymbol{x}$, respectively. The measurement process of (2.2) completely erases the information of length $|\boldsymbol{x}^0|$, which makes it impossible to recover the signal uniquely. We therefore introduce an extra normalization constraint $|\hat{\boldsymbol{x}}| = \sqrt{N}$ for the recovered signal $\hat{\boldsymbol{x}}$, and we consider the recovery successful when the direction cosine $\boldsymbol{x}^0 \cdot \hat{\boldsymbol{x}}/(|\boldsymbol{x}^0||\hat{\boldsymbol{x}}|)$ is sufficiently large.

Unlike the standard CS problem, finding a solution of (2.3) is non-trivial because the norm constraint $|\boldsymbol{x}| = \sqrt{N}$ keeps it from being a convex optimization problem (Figures 2.1 (a) and (b)). The authors of [49] also developed, as a practically feasible solution, a double-loop algorithm called Renormalized Fixed Point Iteration (RFPI) that combines a gradient descent method and enforcement to a sphere of a fixed radius. It is summarized in Figure 2.2.

The practical utility of RFPI was shown by numerical experiments, but how good solutions are actually obtained is unclear because in general the algorithm can be trapped at various local optima. One of our main concerns is therefore to theoretically evaluate the typical performance of the global minimum solution of (2.3) for examining the possibility of performance improvement.

2.2 Performance assessment by the replica method

The partition function

$$Z\left(\beta; \Phi, \boldsymbol{x}^0\right) = \int d\boldsymbol{x} \delta\left(|\boldsymbol{x}|^2 - N\right) e^{-\beta||\boldsymbol{x}||_1} \prod_{\mu=1}^M \Theta\left((\boldsymbol{\Phi x}^0)_\mu (\boldsymbol{\Phi x})_\mu\right), \tag{2.4}$$

where $\Theta(x) = 1$ and 0 for $x > 0$ and $x < 0$, respectively, offers the basis for our analysis. As β tends to infinity, the integral of (2.4) is dominated by the correct solution of (2.3). One therefore can evaluate the performance of the solution by examining the macroscopic behavior of equation (2.4) in the limit of $\beta \to \infty$.

A characteristic feature of the current problem is that (2.4) depends on the predetermined random variables $\boldsymbol{\Phi}$ and $\boldsymbol{x^0}$, which requires us to assess the average of free energy density $f \equiv -(\beta N)^{-1} \left[\ln Z(\beta; \boldsymbol{\Phi}, \boldsymbol{x}^0)\right]_{\boldsymbol{\Phi}, \boldsymbol{x}^0}$ when evaluating the performance for typical samples of $\boldsymbol{\Phi}$ and $\boldsymbol{x}^0$. Here, $[\cdots]_{\boldsymbol{\Phi}, \boldsymbol{x}^0}$ denotes the configurational average concerning $\boldsymbol{\Phi}$ and $\boldsymbol{x^0}$. Because directly averaging the logarithm of the partition function is technically difficult, we here resort to the replica method [14].

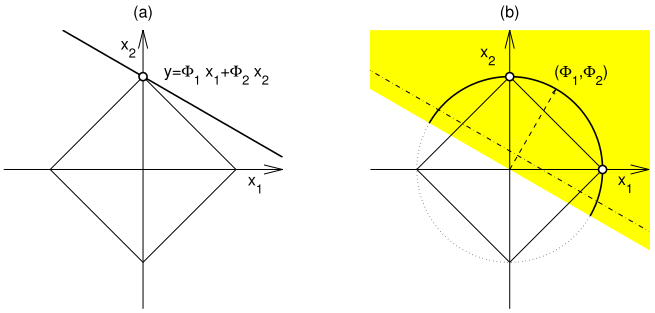

FIGURE 2.1: Graphical representations of (a) standard and (b) 1-bit CS problems in the case of $N = 2$, $M = 1$, and $K = \rho N = 1$. (a): A thick line and a square of thin lines represent a measurement result $y = \Phi_1 x_1 + \Phi_2 x_2$ and a contour of l_1-norm $|x_1| + |x_2|$, respectively. The optimal solution denoted by a circle is uniquely determined since both the set of feasible solutions $y = \Phi_1 x_1 + \Phi_2 x_2$ and the cost function $|x_1| + |x_2|$ are convex. (b): The shaded area $y \times (\Phi_1 x_1 + \Phi_2 x_2) > 0$ represents the region that is compatible with the sign information of the linear measurement $y = \Phi_1 x_1 + \Phi_2 x_2$ (dotted broken line). This and the l_2-norm constraint $x_1^2 + x_2^2 = 2$ yield the set of feasible solutions as a semicircle (thick curve), which is not a convex set. As a consequence, the constraint optimization problem of (2.3) generally has multiple solutions (two circles). This graph is cited from [56]®IOP Publishing Ltd.

For this we first evaluate n-th moment of the partition function $\left[Z^n\left(\beta;\boldsymbol{\Phi},\boldsymbol{x}^0\right)\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0}$ for $n = 1, 2, \ldots \in \mathbb{N}$, using the formula

$$Z^n\left(\beta;\boldsymbol{\Phi},\boldsymbol{x}^0\right) = \int \prod_{a=1}^{n} \left(d\boldsymbol{x}^a \delta\left(|\boldsymbol{x}^a|^2 - N\right) \times e^{-\beta||\boldsymbol{x}^a||_1} \right) \times \prod_{a=1}^{n} \prod_{\mu=1}^{M} \Theta\left((\boldsymbol{\Phi}\boldsymbol{x}^0)_\mu (\boldsymbol{\Phi}\boldsymbol{x}^a)_\mu\right), \tag{2.5}$$

which holds only for $n = 1, 2, \ldots \in \mathbb{N}$. Here, $\boldsymbol{x}^a$ $(a = 1, 2, \ldots, n)$ denotes a-th replicated signal. Averaging (2.5) with respect to $\boldsymbol{\Phi}$ and $\boldsymbol{x}^0$ results in the saddle point evaluation concerning macroscopic variables $q_{0a} = q_{a0} \equiv N^{-1}\boldsymbol{x}^0 \cdot \boldsymbol{x}^a$ and $q_{ab} = q_{ba} \equiv N^{-1}\boldsymbol{x}^a \cdot \boldsymbol{x}^b$ $(a, b = 0, 1, 2, \ldots, n)$. Although (2.5) holds only for $n \in \mathbb{N}$, the expression of $N^{-1} \ln \left[Z^n\left(\beta;\boldsymbol{\Phi},\boldsymbol{x}^0\right)\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0}$ obtained by the saddle point evaluation under a certain assumption concerning the permutation symmetry with respect to the replica indices $a, b = 1, 2, \ldots n$ is obtained as an analytic function of n, which is likely to also hold for $n \in \mathbb{R}$. Therefore, we next utilize the analytic function for evaluating the average of the logarithm of the partition function as

$$N^{-1} \ln \left[\ln Z(\beta;\boldsymbol{\Phi},\boldsymbol{x}^0)\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0} \lim_{n\to 0} N^{-1} \ln \left[Z^n\left(\beta;\boldsymbol{\Phi},\boldsymbol{x}^0\right)\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0}. \tag{2.6}$$

Algorithm 1: RENORMALIZED FIXED POINT ITERATION(δ, λ)

1) **Initialization** :
   Seed : $\hat{\mathbf{x}}_0$ s.t $||\hat{\mathbf{x}}_0||_2 = \sqrt{N}$,
   Descent step size : δ
   Counter : $k \leftarrow 0$
2) **Counter Increase** :
   $k \leftarrow k + 1$
3) **One-sided quadratic gradient** :
   $\overline{\mathbf{f}}_k \leftarrow (\mathrm{Y}\Phi)^{\mathrm{T}} f'(\mathrm{Y}\Phi\hat{\mathbf{x}}_{k-1})$
4) **Gradient projection on sphere surface** :
   $\tilde{\mathbf{f}}_k \leftarrow \overline{\mathbf{f}}_k - \langle\overline{\mathbf{f}}_k, \hat{\mathbf{x}}_{k-1}\rangle\hat{\mathbf{x}}_{k-1}/N$
5) **One-sided quadratic gradient descent** :
   $\mathbf{h} \leftarrow \hat{\mathbf{x}}_{k-1} - \delta\tilde{\mathbf{f}}_k$
6) **Shrinkage (l_1-gradient descent)** :
   $(\mathbf{u})_i \leftarrow \mathrm{sign}((\mathbf{h})_i)\max\{|(\mathbf{h})_i| - \frac{\delta}{\lambda}, 0\}$ for all i
7) **Normalization** :
   $\hat{\mathbf{x}}_k \leftarrow \sqrt{N}\frac{\mathbf{u}}{||\mathbf{u}||_2}$
8) **Iteration** : Repeat from 2) until convergence.

FIGURE 2.2: Pseudocode for the inner loop of the Renormalized Fixed Point Iteration (RFPI) proposed in [49]. The function $f'(x)$ in step 3 is defined as $f'(x) = x$ for $x \leq 0$ and 0, otherwise, and it operates on a vector in a component-wise manner. In the original expression in [49] the normalization constraint is introduced as $||\hat{\mathbf{x}}_k||_2 = 1$, but we here use $||\hat{\mathbf{x}}_k||_2 = \sqrt{N}$ for convenience in considering the large system limit of $N \to \infty$. RFPI is a double-loop algorithm. In the outer loop the parameter λ is increased as $\lambda_n = c\lambda_{n-1}$, where $c > 1$ and n are a certain constant and the counter of the outer loop, respectively. The convergent solution of $i-1$th outer loop is used for the initial state of the inner loop of the ith outer loop. The algorithm terminates when difference between the convergent solutions of two successive outer loops become sufficiently small.This figure is cited from [56]®IOP Publishing Ltd.

In particular, under the replica symmetric (RS) ansatz where the dominant saddle point is assumed to be of the form of

$$q_{ab} = q_{ba} = \begin{cases} \rho & (a = b = 0) \\ m & (a = 1, 2, \ldots, n;\ b = 0) \\ 1 & (a = b = 1, 2, \ldots, n) \\ q & (a \neq b = 1, 2, \ldots, n) \end{cases}, \tag{2.7}$$

when the distribution of nonzero entries in (2.1) is given as the standard Gaussian $\tilde{P}(x) = \exp(-x^2/2)/\sqrt{2\pi}$, the above procedure offers an expression of the average free energy density as

$$\begin{aligned} \bar{f} &= \underset{\omega}{\mathrm{extr}} \Bigg\{ \left[\phi\left(\sqrt{\hat{q}} z + \hat{m} \boldsymbol{x}^0; \hat{Q} \right) \right]_{\boldsymbol{x}^0, z} - \frac{1}{2}\hat{Q} + \frac{1}{2}\hat{q}\chi + \hat{m}m \\ &\quad + \frac{\alpha}{2\pi\chi} \left(\arctan\left(\frac{\sqrt{\rho - m^2}}{m} \right) - \frac{m}{\rho}\sqrt{\rho - m^2} \right) \Bigg\} \end{aligned} \tag{2.8}$$

in the limit of $\beta \to \infty$. Here $\alpha = M/N$, $\mathrm{extr}_X\{g(X)\}$ denotes extremization of a function $g(X)$ with respect to X, $\omega = \{\chi, m, \hat{Q}, \hat{q}, \hat{m}\}$, $\mathrm{D}z = \mathrm{d}z\exp(-z^2/2)/\sqrt{2\pi}$ is a Gaussian measure, and

$$\begin{aligned} \phi\left(\sqrt{\hat{q}} z + \hat{m} \boldsymbol{x}^0; \hat{Q} \right) &= \min_x \left\{ \frac{\hat{Q}}{2} x^2 - \left(\sqrt{\hat{q}} z + \hat{m} x^0 \right) x + |x| \right\} \\ &= -\frac{1}{2\hat{Q}} \left(\left| \sqrt{\hat{q}} z + \hat{m} \boldsymbol{x}^0 \right| - 1 \right)^2 \Theta\left(\left| \sqrt{\hat{q}} z + \hat{m} \boldsymbol{x}^0 \right| - 1 \right). \end{aligned} \tag{2.9}$$

The derivation of (2.8) is provided in A.

The extremization problem of (2.8) yields the following saddle point equations:

$$\hat{q} = \frac{\alpha}{\pi\chi^2} \left(\arctan\left(\frac{\sqrt{\rho - m^2}}{m} \right) - \frac{m}{\rho}\sqrt{\rho - m^2} \right), \tag{2.10}$$

$$\hat{m} = \frac{\alpha}{\pi\chi\rho}\sqrt{\rho - m^2}, \tag{2.11}$$

$$\begin{aligned} \hat{Q}^2 &= 2 \Bigg\{ (1 - \rho) \left[(\hat{q} + 1)\, \mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}}} \right) - \sqrt{\frac{\hat{q}}{2\pi}} e^{-\frac{1}{2\hat{q}}} \right] \\ &\quad + \rho \left[\left(\hat{q} + \hat{m}^2 + 1 \right) \mathcal{Q}\left(\frac{1}{\sqrt{\hat{q} + \hat{m}^2}} \right) - \sqrt{\frac{\hat{q} + \hat{m}^2}{2\pi}} e^{-\frac{1}{2(\hat{q} + \hat{m}^2)}} \right] \Bigg\}, \end{aligned} \tag{2.12}$$

$$\chi = \frac{2}{\hat{Q}} \left[(1 - \rho)\, \mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}}} \right) + \rho \mathcal{Q}\left(\frac{1}{\sqrt{\hat{q} + \hat{m}^2}} \right) \right], \tag{2.13}$$

$$m = \frac{2\rho\hat{m}}{\hat{Q}} \mathcal{Q}\left(\frac{1}{\sqrt{\hat{q} + \hat{m}^2}} \right), \tag{2.14}$$

where $\mathcal{Q}(x) = \int_x^{+\infty} \mathrm{D}z$. The value of m determined by these equations physically means the typical overlap $N^{-1} \left[\boldsymbol{x}^0 \cdot \hat{\boldsymbol{x}} \right]_{\boldsymbol{\Phi}, \boldsymbol{x}^0}$ between the original signal $\boldsymbol{x}^0$ and the solution $\hat{\boldsymbol{x}}$ of (2.3). Therefore the typical value of the direction cosine between $\boldsymbol{x}^0$ and $\hat{\boldsymbol{x}}$, which serves as a performance measure

of the current recovery problem, is evaluated as $\left[(\boldsymbol{x}^0 \cdot \hat{\boldsymbol{x}})/|\boldsymbol{x}^0||\hat{\boldsymbol{x}}|\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0} = Nm/(\sqrt{N\rho} \times \sqrt{N}) = m/\sqrt{\rho}$. Alternatively, we may also use as a performance measure the mean square error (MSE) between the normalized vectors:

$$\mathrm{MSE} = \left[\left| \frac{\hat{\boldsymbol{x}}}{|\hat{\boldsymbol{x}}|} - \frac{\boldsymbol{x}^0}{|\boldsymbol{x}^0|} \right|^2\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0} = 2\left(1 - \frac{m}{\sqrt{\rho}}\right). \tag{2.15}$$

We solved the saddle point equations for various sets of α and ρ. The curves in figures 2.3 (a)–(d) show the theoretical prediction of MSE evaluated by (2.15) plotted against the measurement bit ratio $\alpha = M/N$ for $\rho = 1/32, 1/16, 1/8$, and $1/4$. To examine the validity of the RS ansatz, we also evaluated the local stability of the RS solutions against the disturbances that break the replica symmetry [2], which offers

$$\begin{aligned} &\frac{\alpha}{\pi(\hat{Q}\chi)^2} \arctan\left(\frac{\sqrt{\rho - m^2}}{m}\right) \\ &\times 2\left((1-\rho)\mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}}}\right) + \rho\mathcal{Q}\left(\frac{1}{\sqrt{\hat{q} + \hat{m}^2}}\right)\right) - 1 < 0, \end{aligned} \tag{2.16}$$

as the stability condition. A brief sketch of the derivation of this condition is shown in B. Unfortunately, (2.16) is not satisfied for any regions in figures 2.3 (a)–(d). This is presumably because the optimization problem for (2.3) has many local optima reflecting the fact that the constraint of $||\boldsymbol{x}||_2 = \sqrt{N}$ loses the convexity. This indicates that taking the replica symmetry breaking (RSB) into account is necessary for evaluating the exact performance of the signal recovery scheme defined by (2.3).

We nonetheless think that the RS analysis offers considerably accurate approximates of the exact performance in terms of MSE. The ($\times$) symbols in figures 2.3 (a)–(d) stand for MSE experimentally achieved by RFPI, which were assessed as the arithmetic averages over 1000 samples for each condition of $N = 128$ systems. Excellent consistency between the curves and symbols suggests that even if (2.3) has many local optima, they are close to one another in terms of the l_2-norm yielding similar values of MSE. This also implies that RFPI, which is guaranteed to find one of the local optima, performs nearly saturates as well (as measured by the MSE) as the signal recovery scheme based on (2.3).

Of course, we have to keep in mind that the consistency between the theory and experiments depends highly on the performance measure used. Figures 2.4 (a)–(d) show the probabilities of wrongly predicting sites of nonzero and zero entries, which are sometimes referred to as false positive (FP) and false negative (FN), respectively. These indicate that there are considerably large discrepancies between the theory and experiments in terms of these performance measures, which is probably due to the influence of RSB. Nevertheless, the RS-based theoretical predictions are still qualitatively consistent with the experimental results in the way that the probability of a FP remains finite even when the measurement bit ratio $\alpha = M/N$ tends to infinity for any values of ρ. This implies that the l_1-based scheme is intrinsically unable to correctly identify sites of nonzero and zero entries.

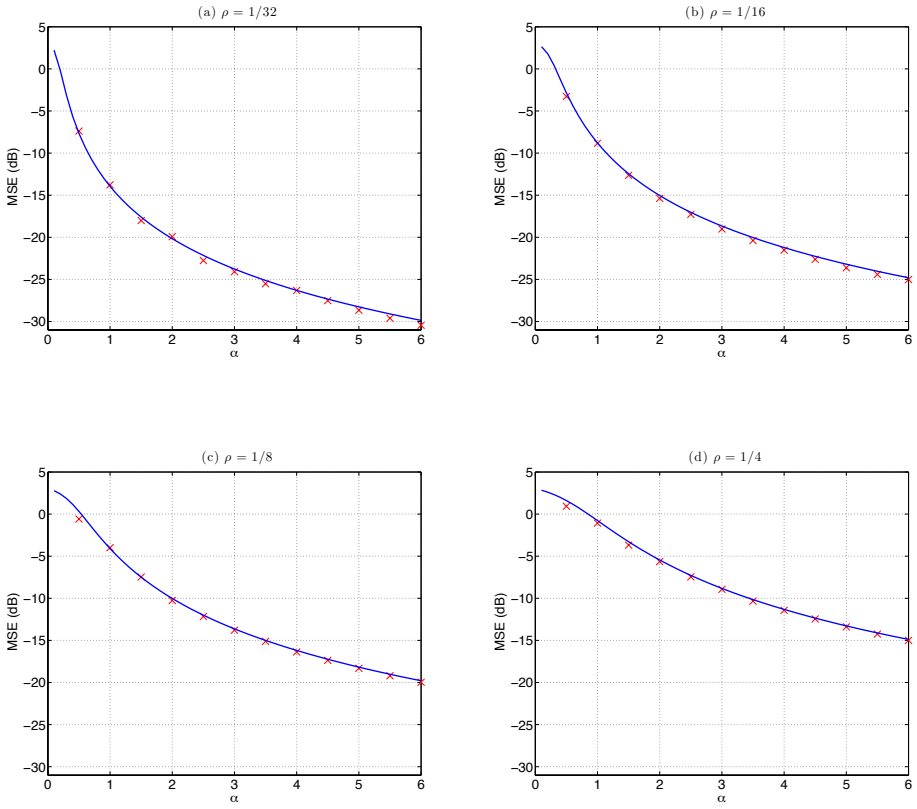

FIGURE 2.3: MSE versus the measurement bit ratio α for the signal recovery scheme using (2.3). (a), (b), (c), and (d) correspond to the $\rho = 1/32, 1/16, 1/8$, and $1/4$ cases, respectively. Curves represent the theoretical prediction evaluated by the RS solution, which is locally unstable for disturbances that break the replica symmetry for all regions of (a)–(d). Each symbol ($\times$) stands for the experimental estimate obtained for RFPI in [49] from 1000 experiments with $N = 128$ systems. This figure is cited from [56]®IOP Publishing Ltd.

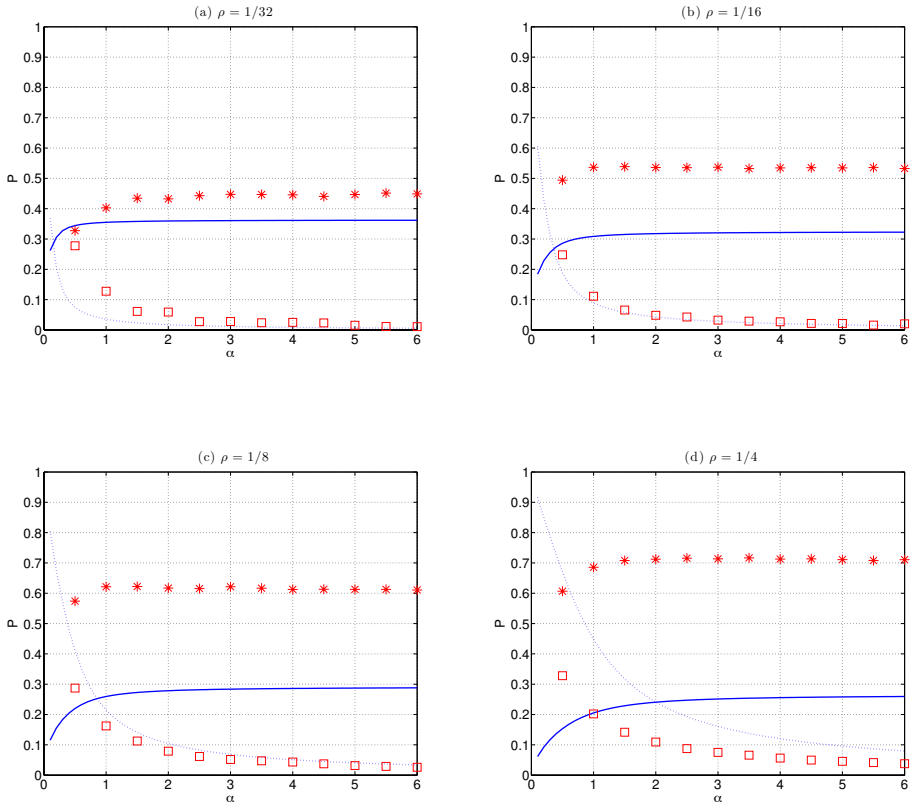

FIGURE 2.4: FP and FN probabilities versus the measurement bit ratio $\alpha = M/N$. (a), (b), (c), and (d) corresponds to the $\rho = 1/32, 1/16, 1/8$, and $1/4$ cases, respectively. Solid and dashed curves represent theoretical predictions obtained by the RS solution for FP and FN, respectively. Asterisks and squares denote experimental results for FP and FN, respectively. The experimental results were obtained by RFPI from 1000 samples for each condition of $N = 128$ systems. This figure is cited from [56]®IOP Publishing Ltd.

2.3 Cavity-inspired signal recovery algorithm

The analysis so far indicates that the performance of RFPI is good enough in the sense that there is little room for improvement in achievable MSE. RFPI requires tuning of two parameters δ and λ, however, which is rather laborious. In addition, the convergence of the inner loop of Figure 2.2 is relatively slow, which may limit its application range to systems of relatively small sizes. We therefore developed another recovery algorithm following the framework of the cavity method of statistical mechanics [38, 37], or equivalently, the belief propagation of probabilistic inference [35, 36, 26].

For simplicity of notations, let us first convert all the measurement results to $+1$ by multiplying y_μ $(\mu = 1, 2, \ldots, N)$ to each row of the measurement matrix $\mathbf{\Phi} = (\Phi_{\mu i})$ as $(\Phi_{\mu i}) \to (y_\mu \Phi_{\mu i})$, and newly denote the resultant matrix as $\mathbf{\Phi} = (\Phi_{\mu i})$. In the new notation, introduction of Lagrange multipliers $\boldsymbol{a} = (a_\mu)$ and surplus variables $\boldsymbol{z} = (z_\mu)$ converts (2.3) to an unconstrained optimization problem:

$$\min_{\boldsymbol{x},\boldsymbol{z}>0} \max_{\boldsymbol{a},\Lambda} \left\{ \sum_{i=1}^{N} |x_i| + \sum_{\mu=1}^{M} a_\mu \left(\sum_{i=1}^{N} \Phi_{\mu i} x_i - z_\mu \right) + \frac{\Lambda}{2} \left(\sum_{i=1}^{N} x_i^2 - N \right) \right\}$$
$$= \min_{\boldsymbol{x},\boldsymbol{z}>0} \max_{\boldsymbol{a},\Lambda} \left\{ \sum_{i=1}^{N} \left(\frac{\Lambda}{2} x_i^2 + |x_i| \right) - \sum_{\mu=1}^{M} a_\mu z_\mu + \sum_{\mu,i} \Phi_{\mu i} a_\mu x_i - \frac{N\Lambda}{2} \right\}, \quad (2.17)$$

where $\boldsymbol{z} > 0$ means that each entry of $\boldsymbol{z}$ is restricted to be positive.

Coupling terms $\sum_{\mu i} \Phi_{\mu i} a_\mu x_i$ make the optimization of (2.17) a nontrivial problem. In statistical mechanics, a standard approach to resolving such a difficulty is to approximate (2.17) with a bunch of optimizations for single-body cost functions parameterized as

$$\mathcal{L}_i(x_i) = \frac{A_i}{2} x_i^2 - H_i x_i + |x_i|, \quad (2.18)$$

and

$$\mathcal{L}_\mu(a_\mu, z_\mu) = -\frac{B_\mu}{2} a_\mu^2 + K_\mu a_\mu - z_\mu a_\mu, \quad (2.19)$$

where A_i, B_μ, H_i, and K_μ are parameters to be determined in a self-consistent manner.

In the cavity method this is done by introducing virtual systems that are defined by removing a single variable x_i or a single pair of variables (a_μ, z_μ) from the original system [38]. When N is sufficiently large, the law of large numbers allows us to assume that the values of A_i and B_μ are constant independently of their indices; that is, that A and B are constants. Under

this simplification, this method yields a set of self-consistent equations:

$$K_\mu = \sum_{i=1}^{N} \Phi_{\mu i}\hat{x}_i - B\hat{a}_\mu, \tag{2.20}$$

$$\hat{a}_\mu = -\frac{1}{B} f'(K_\mu), \tag{2.21}$$

$$H_i = \sum_{\mu=1}^{M} \Phi_{\mu i}\hat{a}_\mu + \Gamma\hat{x}_i, \tag{2.22}$$

$$\hat{x}_i = \frac{1}{A} g'(H_i), \tag{2.23}$$

where $\mu = 1, 2, \ldots, M$, $i = 1, 2, \ldots, N$, $f(u) \equiv (u^2/2)\Theta(-u)$, and $g(u) \equiv \left((|u|-1)^2/2\right)\Theta(|u|-1)$. Γ is evaluated using $\{K_\mu\}$ and B as

$$\Gamma = B^{-1}\left(N^{-1}\sum_{\mu=1}^{M} f''(K_\mu)\right) = B^{-1}\left(N^{-1}\sum_{\mu=1}^{M}\Theta(|K_\mu|-1)\right). \tag{2.24}$$

A is determined so that $\sum_{i=1}^{N}\hat{x}_i^2 = N$ holds in (2.23), which provides B as

$$B = A^{-1}\left(N^{-1}\sum_{i=1}^{N} g''(H_i)\right) = A^{-1}\left(N^{-1}\sum_{i=1}^{N}\Theta(|H_i|-1)\right). \tag{2.25}$$

$\Gamma\hat{x}_i$ on the right-hand side of (2.22) is often referred to as the *Onsager reaction term* [50, 46]. Equation (2.23) offers the recovered signal. The derivations of these equations are provided in C.

A distinctive feature of the above set of equations is that they are free from tuning parameters such as λ and δ in RFPI, which is highly beneficial in practical use. It is therefore unfortunate that in most cases the naive iterations of (2.20)→(2.21), (2.24)→(2.22)→(2.23), (2.25)→(2.20)$\cdots$ hardly converge, which is considered a consequence of RSB [24], while a similar approach offers successful results for various other problems of compressed sensing [41, 4].

We found, however, that instead of updating B by (2.25) at each iteration, handling B as a parameter to be controlled in the outer loop, in conjunction with modifying (2.20) and (2.21) to

$$\hat{a}_\mu = \hat{a}_\mu - \frac{1}{B} f'\left(\sum_{i=1}^{N}\Phi_{\mu i}\hat{x}_i\right), \tag{2.26}$$

results in a fairly good approximate signal recovery algorithm.

The necessity of controlling B in the outer loop, which is essential for having good convergence in the inner loop, means that our algorithm still requires one tuning parameter. Nonetheless, the reduction in the number of the tuning parameters from two to one is considerably advantageous for practical use. In practice, the initial value of B should be set so that only a single entry becomes nonzero. This is easily done by the Binary Iterative Hard Thresholding algorithm [23], which requires the number of nonzero entries as extra prior knowledge. After the initial value is set, B is reduced

Algorithm 2: CAVITY-INSPIRED SIGNAL RECOVERY($\mathrm{B}, \mathbf{x}^*, \mathbf{H}^*$)

1) **Initialization** :
   X seed : $\hat{\mathbf{x}}_0 \leftarrow \hat{\mathbf{x}}^*$
   H seed : $\mathbf{H}_0 \leftarrow \mathbf{H}^*$
   Counter : $k \leftarrow 0$
2) **Counter increase** :
   $k \leftarrow k + 1$
3) **One-sided quadratic gradient descent** :
   $\mathbf{H}_k \leftarrow \mathbf{H}_{k-1} - \mathrm{B}^{-1}(\mathrm{Y}\Phi)^{\mathrm{T}} f'\left(\mathrm{Y}\Phi\hat{\mathbf{x}}_{k-1}\right)$
4) **Assessment of Onsager coefficient** :
   $\Gamma \leftarrow (\mathrm{NB})^{-1}\mathbf{1}^{\mathrm{T}} f''\left(\mathrm{Y}\Phi\hat{\mathbf{x}}_{k-1}\right)$
5) **Self-feedback cancellation** :
   $\tilde{\mathbf{H}}_k \leftarrow \mathbf{H}_k + \Gamma\hat{\mathbf{x}}_{k-1}$
6) **Shrinkage (l_1-gradient descent)** :
   $(\mathbf{u})_i \leftarrow \mathrm{sign}((\tilde{\mathbf{H}})_i)\max\{|(\tilde{\mathbf{H}})_i| - 1, 0\}$ for all i
7) **Normalization** :
   $\hat{\mathbf{x}}_k \leftarrow \sqrt{N}\frac{\mathbf{u}}{||\mathbf{u}||_2}$
8) **Iteration** : Repeat from 2) until convergence.

FIGURE 2.5: Pseudocode for the inner loop of the cavity-inspired signal recovery (CISR) algorithm. $\mathbf{x}^*$ and $\mathbf{H}^*$ are the convergent vectors of $\hat{\mathbf{x}}_k$ and $\hat{\mathbf{H}}_k$ obtained by the previous outer loop. The $\mathbf{1}$ in step 4) is the N-dimensional vector all entries of which are unity. If $(\mathbf{u})_i = 0$ eventually holds for $\forall i$ in step 6), B is reduced so that only $\max_i\{|(\mathbf{u})_i|\}$ becomes nonzero, and the procedure is restarted from step 3).
This figure is cited from [56]®IOP Publishing Ltd.

as $B_n = rB_{n-1}$ with an appropriate constant $0 < r < 1$, where n is the counter of the outer loop. The algorithm terminates when the difference between the convergent solutions of two successive outer loops is sufficiently small.

The resultant algorithm is somewhat similar to RFPI as the combination of (2.22) and (2.26) roughly acts as the **One-sided quadratic gradient descent** step in Figure 2.2. However, as the length of $\boldsymbol{H} = (H_i) = \boldsymbol{\Phi}^{\mathrm{T}}\hat{\boldsymbol{a}}$ is not restricted to a fixed value, the current algorithm does not need a small step size δ for the convergence. Another significant difference from RFPI is the existence of the Onsager reaction term in (2.22). This term effectively cancels the self-feedback effects included in H_i of (2.22), and this is expected to accelerate the convergence of the algorithm. A pseudocode for the inner loop is summarized in Figure 2.5.

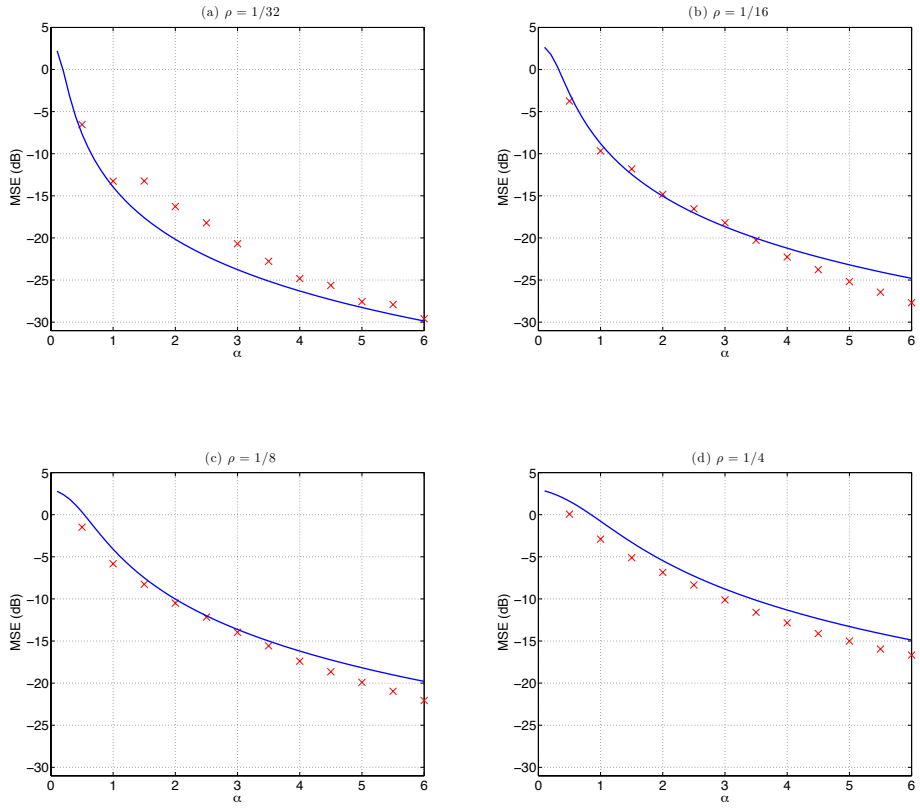

FIGURE 2.6: MSE versus measurement bit ratio α for the cavity-inspired signal recovery (CISR) algorithm. Experimental conditions are the same as in Figures 2.3 (a)–(d).This figure is cited from [56]®IOP Publishing Ltd.

The MSE results obtained in numerical experiments with the cavity-inspired signal recovery (CISR) algorithm are shown in Figures 2.6 (a)–(d). They indicate that except in the case in which the nonzero density ρ of the original signals is significantly low, CISR provides MSE values almost equal to or lower than those of RFPI. Figures 2.7 (a)–(d) show the FP and FN probabilities for CISR. The discrepancies from the theoretical prediction are not unexpected because the modification of (2.21) to (2.26) means that CISR is no longer based on (2.3) or (2.17). The FN probabilities for CISR are higher than those for RFPI, while the FP probabilities are lower. This implies that CISR has a capability of yielding sparser signals than RFPI, which is presumably because parameter B of CISR is initially set so that only a single entry of $\hat{\boldsymbol{x}}$ is nonzero while such a tuning is not taken into account in RFPI.

The run times actually required for performing the experiments in a MATLAB® environment for the cases of $N = 128$ and $M = 3N = 384$ are listed in Table 2.1. Although the run times of RFPI may be reduced by optimally tuning the descent step size δ, CISR is several hundreds of times faster than RFPI. This shows the significant computational efficiency of CISR. The NORT values in Table 2.1 are the run times when the Onsager reaction term in (2.22) was removed from CISR. Their being 1.13–2.37

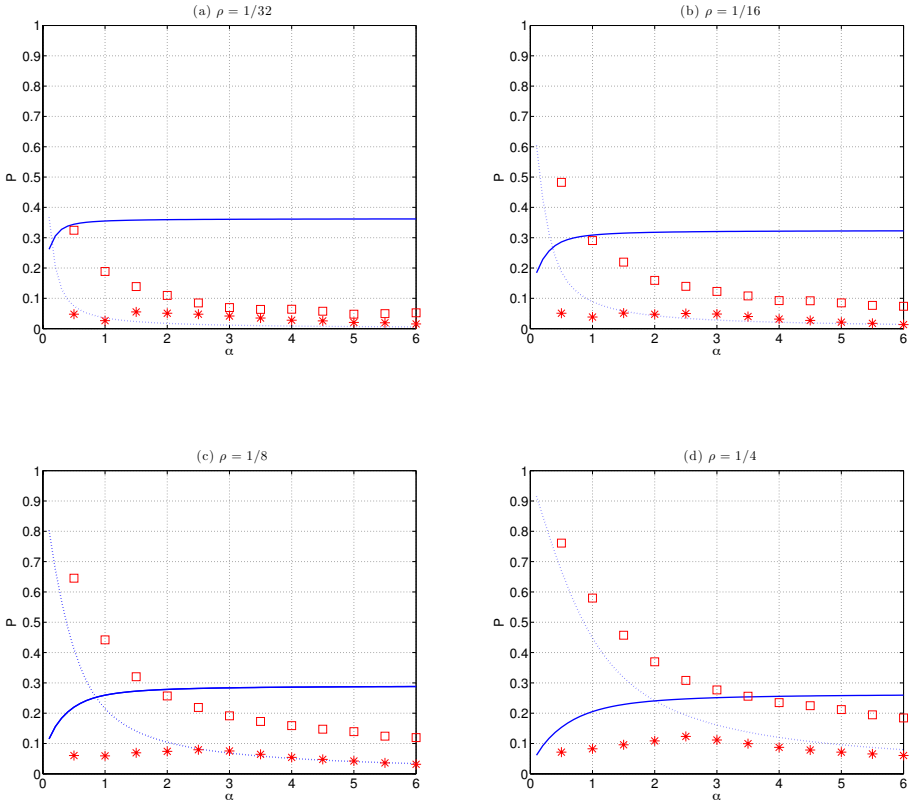

FIGURE 2.7: FP and FN probabilities of versus measurement bit ratio α for the CISR. Experimental conditions are the same as in Figures 2.4 (a)–(d). This figure is cited from [56]®IOP Publishing Ltd.

times longer than those for CISR indicates that the cancellation of the self-feedback effects by adding the Onsager reaction term speeds the convergence of CISR significantly.

TABLE 2.1: Comparison of computational costs for the $N = 128$ and $M = 3N = 384$ cases. The values listed here are the average run times (in seconds) evaluated in 1000 experiments, and the numbers in parentheses are the standard deviations. In RFPI, δ was roughly tuned as 0.01, and λ was enlarged as $\lambda_n = 2\lambda_{n-1}$ with the initial value $\lambda_0 = 0.005$ in the outer loop. On the other hand, B of CISR was reduced as $B_n = 0.9B_{n-1}$. The NORT values are the run times required for performing the same experiments when the Onsager reaction term was removed from CISR. In all cases the algorithms terminated when the difference per entry, in terms of l_1-norm, between the convergent solutions of two successive outer loops was less than 10^{-8}. This table is cited from [56]®IOP Publishing Ltd.

	$K = 4$	$K = 8$	$K = 16$	$K = 32$
RFPI	$25.7636(10.0799)$s	$27.8293(3.3566)$s	$33.3552(3.2914)$s	$35.4574(3.3869)$s
CISR	$0.0385(0.0583)$s	$0.0705(0.1058)$s	$0.0245(0.0346)$s	$0.0247(0.0207)$ s
NORT	$0.0557(0.0889)$s	$0.0795(0.1095)$s	$0.0581(0.0566)$s	$0.0369(0.0316)$s

2.4 Summary

In summary, we have examined typical properties of 1-bit compresses sensing (CS) proposed in [49] utilizing methods of statistical mechanics. Signal recovery based on the l_1-norm minimization is a standard approach in CS research. Unlike the normal CS scheme, however, the l_1-based signal recovery cannot be formulated as a convex optimization problem, which makes practically performing it nontrivial.

We have shown that the theoretical prediction of the performance of the l_1-based scheme, which is obtained by the replica method under the replica symmetric (RS) ansatz, exhibits a fairly good accordance (in terms of MSE) with experimental results obtained using for an approximate signal recovery algorithm, RFPI, proposed in [49]. The replica symmetry of the RS solution turned out to be broken, however, which implies that there are many local optima for the optimization problem of the signal recovery. Our results suggest that the local optima, which can be searched by RFPI, yield similar values of MSE representing the potential performance limit of l_1-based recovery scheme.

We have also developed an approximate signal recovery algorithm utilizing the cavity method. Naive iterations of self-consistent equations derived directly from the cavity method hardly converge in most cases, which can be regarded as a consequence of the replica symmetry breaking. However, we have shown that modification of one equation in an appropriate manner, in conjunction with controlling a macroscopic variable in the outer loop, results in a fairly good signal recovery algorithm. Compared with

RFPI, the resultant algorithm is beneficial in that the number of tuning parameters is reduced from two to one. Numerical experiments have also shown that whenever the density of nonzero entries of the original signal is not considerably small the cavity-inspired algorithm performs as well as or better than RFPI (in terms of MSE) and has a lower computational cost.

We here focused on the l_1-based recovery scheme since it was proposed and examined in the seminal paper on 1-bit CS [49]. However, the significance of the l_1-based scheme may be rather weak for 1-bit CS because the loss of convexity it entails keeps it from leading to the development of mathematically guaranteed and practically feasible algorithms. Therefore, much effort should be devoted to developing recovery algorithms following various principles. In the next chapter, we will suggest a strategy to solve the convexity problem. And in chapter 4, we will propose a idea based on the Bayesian inference for 1-bit CS.

Chapter 3

Thresholding l_1-norm minimization

It is obvious that the scale (absolute amplitude) of the signal is lost in 1-bit CS measurements. To compensate for this, past studies, as in the previous chapter, have proposed the imposition of an additional constraint whereby the l_2-norm of the signal is normalized to a fixed constant [49, 56]. In other words, this can only reconstruct the directional information but not the true scale information of the signal. Moreover, it yields another drawback such that solving the reconstruction problem becomes nontrivial, since the problem is no longer formulated as a convex optimization. In order to address these issues, we propose introducing a set of finite thresholds $\boldsymbol{\lambda} = (\lambda_\mu)$ $(\mu = 1, 2, \ldots, M)$ to the quantizer as

$$\boldsymbol{y} = \text{sign}(\boldsymbol{\Phi x} + \boldsymbol{\lambda}). \tag{3.1}$$

Combining the knowledge of the thresholds and frequencies of the binary outputs allows us to estimate the scale of the signal. Furthermore, as the feasible set provided by the constraint of (3.1) for given measurements $\boldsymbol{y}$ is a convex region of $\boldsymbol{x}$, one can reconstruct a sparse signal in polynomial time by solving the l_1-norm minimization problem

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x} \in \mathbb{R}^N}{\text{argmin}} \, ||\boldsymbol{x}||_1 \text{ subject to } \boldsymbol{y} = \text{sign}(\boldsymbol{\Phi x} + \boldsymbol{\lambda}) \tag{3.2}$$

by using versatile convex optimization algorithms [20].

A lingering, natural question is how we should set the values of λ_μ. To partially answer this, we compare two strategies: one involves fixing the thresholds at a constant value $\lambda_\mu = \lambda$ for all measurements, and the other consists of independently selecting λ_μ from an identical Gaussian distribution. We will show that the fixing-value strategy yields better mean squared error (MSE) performance than the random strategy when adjustable parameters are optimally tuned using the replica method [14] of statistical mechanics.

Unfortunately, the value of the optimal threshold depends on the statistical property of the target signal. However, experimental results indicate that MSEs are typically minimized when the frequency of positive (or negative) output, which can be statistically estimated from the measurements, is placed in a relatively narrow range. By relying on this empirical observation, we will develop an online learning algorithm to tune λ_μ based on the results of past measurements. Numerical experiments show that our algorithm exhibits satisfactory comparable to that achieved by the optimally tuned threshold.

The rest of this chaper is organized as follows: In Section 3.1, we formulate the problem to be addressed in this study. In Section 3.2, we evaluate the performance of the reconstruction method of (3.2). Section 3.3 is devoted to a description of our learning algorithm to tune the threshold value, whereas last section summarizes our work in this chapter.

3.1 Problem set up

Let us suppose a situation where entry x_i^0 $(i = 1, 2, \ldots, N)$ of N-dimensional signal (vector) $\boldsymbol{x}^0 \in \mathbb{R}^N$ is independently generated from an identical sparse distribution:

$$P(x) = (1-\rho)\,\delta(x) + \rho\tilde{P}(x), \tag{3.3}$$

where $\rho \in [0, 1]$ represents the density of nonzero entries in the signal, and $\tilde{P}(x)$ is a distribution function of $x \in \mathbb{R}$ that does not have finite mass at $x = 0$. In the thresholding 1-bit CS, the measurement is performed as

$$\boldsymbol{y} = \mathrm{sign}\left(\boldsymbol{\Phi}\boldsymbol{x}^{\mathbf{0}} + \boldsymbol{\lambda}\right), \tag{3.4}$$

where we assume that each entry of the $M \times N$ measurement matrix $\boldsymbol{\Phi}$ is provided as an independent sample from a Gaussian distribution of mean zero and variance N^{-1}.

We consider two strategies for setting the thresholding vector $\boldsymbol{\lambda} = (\lambda_\mu)$. Case 1: entry $\lambda_\mu = \lambda$ is fixed for all $\mu = 1, 2, \ldots, M$. Case 2: λ_μ is independently sampled from a Gaussian distribution $\mathcal{N}(0, \sigma_\lambda^2)$. For both cases, the feasible set consistent with given outputs $\boldsymbol{y}$ is provided by a set of inequalities

$$y_\mu \left(\sum_{i=1}^{N} \Phi_{\mu i} x_i + \lambda_\mu \right) > 0 \tag{3.5}$$

$(\mu = 1, 2, \ldots, M)$, which defines a convex region of $\boldsymbol{x}$. Therefore, a sparse signal is reconstructed by the l_1-norm minimization (3.2) utilizing a certain convex optimization algorithm.

3.2 Analysis

3.2.1 Method

The key to finding the statistical properties of reconstruction (3.2) is the average free energy density

$$\bar{f} \equiv -\lim_{\beta, N\to\infty} \frac{1}{\beta N} \left[\ln Z(\beta; \boldsymbol{\Phi}, \boldsymbol{x}^0, \boldsymbol{\lambda})\right]_{\boldsymbol{\Phi}, \boldsymbol{x}^0, \boldsymbol{\lambda}}, \tag{3.6}$$

where

$$Z\left(\beta;\boldsymbol{\Phi},\boldsymbol{x}^{\mathbf{0}},\boldsymbol{\lambda}\right) = \int d\boldsymbol{x} e^{-\beta||\boldsymbol{x}||_1} \prod_{\mu=1}^{M} \Theta\left((\boldsymbol{\Phi}\boldsymbol{x}^0+\boldsymbol{\lambda})_\mu(\boldsymbol{\Phi}\boldsymbol{x}+\boldsymbol{\lambda})_\mu\right) \tag{3.7}$$

is the partition function. Here, $\Theta(x) = 1$ and 0 for $x > 0$ and $x < 0$, respectively, offers the basis for our analysis. $[\cdots]_X$ generally denotes the operation of the average with respect to the random variable X. As β tends to infinity, the integral of (3.7) is dominated by the correct solution of (3.2). One can therefore evaluate the performance of the solution by examining the macroscopic behavior of (3.7) in the limit of $\beta \to \infty$. Because directly averaging the logarithm of the partition function is difficult, we employ the replica method [14], which allows us to calculate the average free energy density as

$$\bar{f} = -\lim_{n\to+0}\frac{\partial}{\partial n}\lim_{\beta,N\to\infty}\frac{1}{\beta N}\ln\left[Z^n(\beta;\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda})\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}}. \tag{3.8}$$

For this, we first evaluate the n-th moment of the partition function $\left[Z^n\left(\beta;\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}\right)\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}}$ for $n = 1,2,\ldots \in \mathbb{N}$ by using the formula

$$Z^n\left(\beta;\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}\right) = \int \prod_{a=1}^{n}\left(d\boldsymbol{x}^a e^{-\beta||\boldsymbol{x}^a||_1}\right) \times \prod_{a=1}^{n}\prod_{\mu=1}^{M}\Theta\left((\boldsymbol{\Phi}\boldsymbol{x}^0+\boldsymbol{\lambda})_\mu(\boldsymbol{\Phi}\boldsymbol{x}^a+\boldsymbol{\lambda})_\mu\right), \tag{3.9}$$

which holds only for $n = 1,2,\ldots \in \mathbb{N}$. Here, $\boldsymbol{x}^a$ $(a = 1,2,\ldots,n)$ denotes the a-th replicated signal. Averaging (3.9) with respect to $\boldsymbol{\Phi}$ and $\boldsymbol{x}^0$ results in the saddle point evaluation concerning macroscopic variables $q_{0a} = q_{a0} \equiv N^{-1}\boldsymbol{x}^0\cdot\boldsymbol{x}^a$ and $q_{ab} = q_{ba} \equiv N^{-1}\boldsymbol{x}^a\cdot\boldsymbol{x}^b$ $(a,b = 1,2,\ldots,n)$. Although (3.9) holds only for $n \in \mathbb{N}$, the expression $(\beta N)^{-1}\ln\left[Z^n\left(\beta;\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}\right)\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}}$ obtained by the saddle point evaluation, under a certain assumption concerning the permutation symmetry with respect to the replica indices a,b, is obtained as an analytic function of n, which is likely to also hold for $n \in \mathbb{R}$. Therefore, we utilize the analytic function to evaluate the average of the logarithm of the partition function to obtain $\bar{f}$.

In particular, under the replica symmetric (RS) ansatz, where the dominant saddle point is assumed to be of the form

$$q_{ab} = q_{ba} = \begin{cases} Q_0 & (a = b = 0) \\ m & (a = 1,2,\ldots,n;\ b = 0) \\ Q & (a = b = 1,2,\ldots,n) \\ q & (a \neq b = 1,2,\ldots,n) \end{cases}. \tag{3.10}$$

For simplicity, we hereafter assume that x^0 is distributed from (3.3) with $\tilde{P}(x) = \mathcal{N}(0,\sigma_0^2)$; therefore $Q_0 = \rho\sigma_0^2$.

3.2.2 Resulting equations

The above procedure (3.10) offers an expression of the average free-energy density as

$$\begin{aligned}\bar{f} &= \underset{\omega}{\text{extr}}\Bigg\{\int \mathrm{D}z P(x^0)\phi\left(\sqrt{\hat{q}}z+\hat{m}x^0;\hat{Q}\right)\\ &\quad -\frac{1}{2}\hat{Q}q+\frac{1}{2}\hat{q}\chi+\hat{m}m\sigma_0^2\\ &\quad +\frac{\alpha}{2\chi}\Bigg[\mathcal{Q}\left(-\frac{\frac{m}{\sqrt{q}}t+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)\left(\sqrt{q}t+\lambda\right)^2\Theta\left(-\sqrt{q}t-\lambda\right)\\ &\quad +\mathcal{Q}\left(\frac{\frac{m}{\sqrt{q}}t+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)\left(\sqrt{q}t+\lambda\right)^2\Theta\left(\sqrt{q}t+\lambda\right)\Bigg]_{t,\lambda}\Bigg\} \end{aligned} \tag{3.11}$$

in the limit of $\beta\to\infty$. Here, $\alpha=M/N$, $\text{extr}_X\{g(X)\}$ denotes the extremization of function $g(X)$ with respect to X, $\omega=\{\chi,m,q,\hat{Q},\hat{q},\hat{m}\}$, $\mathcal{Q}(x)=\int_x^{+\infty}\mathrm{D}z$, $\mathrm{D}z=\mathrm{d}z\exp(-z^2/2)/\sqrt{2\pi}$ is a Gaussian measure, t and z are independent and identically distributed (i.i.d) random variables from $\mathcal{N}(0,1)$. The function $\phi(h;\hat{Q})$ is defined as

$$\begin{aligned}\phi(h;\hat{Q}) &= \min_x\left\{\frac{\hat{Q}}{2}x^2-hx+|x|\right\}\\ &= -\frac{1}{2\hat{Q}}\left(|h|-1\right)^2\Theta\left(|h|-1\right).\end{aligned} \tag{3.12}$$

The derivation of (3.11) is provided in Appendix D.

For Case 1, which fixes the threshold for all measurements to a constant λ as $\lambda_\mu=\lambda$ $(\mu=1,2,\ldots,M)$, the extremization of (3.11) is reduced to the following saddle point equations:

$$\hat{q} = \frac{\alpha}{\chi^2}\left\{\left[\mathcal{Q}\left(-\frac{\frac{mt}{\sqrt{q}}+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)u\left(-\sqrt{q}t-\lambda\right)\right]_t + \left[\mathcal{Q}\left(\frac{\frac{mt}{\sqrt{q}}+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)u\left(\sqrt{q}t+\lambda\right)\right]_t\right\}, \tag{3.13}$$

$$\hat{Q} = \frac{\alpha}{\chi}\left\{\left[\mathcal{Q}\left(-\frac{\frac{mt}{\sqrt{q}}+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)u''\left(-\sqrt{q}t-\lambda\right)\right]_t + \left[\mathcal{Q}\left(\frac{\frac{mt}{\sqrt{q}}+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)u''\left(\sqrt{q}t+\lambda\right)\right]_t\right\}, \tag{3.14}$$

$$\hat{m} = \frac{\alpha}{\chi\sqrt{2\pi\left(\rho\sigma_0^2-\frac{m^2}{q}\right)}}\left[\exp\left(-\frac{\left(\frac{mt}{\sqrt{q}}+\lambda\right)^2}{2\left(\rho\sigma_0^2-\frac{m^2}{q}\right)}\right) \times\left(u'\left(\sqrt{q}t+\lambda\right)-u'\left(-\sqrt{q}t-\lambda\right)\right)\right]_t, \tag{3.15}$$

$$q = \frac{2}{\hat{Q}^2}\left\{(1-\rho)\left(\left(\hat{q}+1\right)\mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}}}\right)-\sqrt{\frac{\hat{q}}{2\pi}}e^{-\frac{1}{2\hat{q}}}\right) +\rho\left(\left(\hat{q}+\hat{m}^2\sigma_0^2+1\right)\mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}+\hat{m}^2\sigma_0^2}}\right) -\sqrt{\frac{\hat{q}+\hat{m}^2\sigma_0^2}{2\pi}}e^{-\frac{1}{2\left(\hat{q}+\hat{m}^2\sigma_0^2\right)}}\right)\right\}, \tag{3.16}$$

$$\chi = \frac{2}{\hat{Q}}\left\{(1-\rho)\,\mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}}}\right)+\rho\mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}+\hat{m}^2\sigma_0^2}}\right)\right\}, \tag{3.17}$$

$$m = \frac{2\rho\hat{m}\sigma_0^2}{\hat{Q}}\mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}+\hat{m}^2\sigma_0^2}}\right), \tag{3.18}$$

where $u(x)=x^2\Theta(x)$, and t obeys the standard normal distribution $\mathcal{N}(0,1)$.

On the other hand, for Case 2, where λ_μ is sampled independently from $\mathcal{N}(0,\sigma_\lambda^2)$ for $\mu=1,2,\ldots,M$, the saddle point equations of $\hat{q}$, $\hat{Q}$, and $\hat{m}$ are

modified to

$$\hat{q} = \frac{2\alpha}{\chi^2}\left[\mathcal{Q}\left(\frac{\frac{mt}{\sqrt{q}}+\sigma_\lambda r}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)u\left(\sqrt{q}t+\sigma_\lambda r\right)\right]_{r,t}, \tag{3.19}$$

$$\hat{Q} = \frac{2\alpha}{\chi}\left[\mathcal{Q}\left(\frac{\frac{mt}{\sqrt{q}}+\sigma_\lambda r}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)u''\left(\sqrt{q}t+\sigma_\lambda r\right)\right]_{r,t}, \tag{3.20}$$

$$\hat{m} = \frac{2\alpha}{\chi\sqrt{2\pi\left(\rho\sigma_0^2-\frac{m^2}{q}\right)}}\left[\exp\left(-\frac{\left(\frac{mt}{\sqrt{q}}+\sigma_\lambda r\right)^2}{2\left(\rho\sigma_0^2-\frac{m^2}{q}\right)}\right)\right. \\ \left.\times\, u'\left(\sqrt{q}t+\sigma_\lambda r\right)\right]_{r,t}, \tag{3.21}$$

where r is a variable sampled from the standard normal distribution $\mathcal{N}(0,1)$. The remaining equations for q, χ, and m are identical to (3.16), (3.17), and (3.18), respectively.

3.2.3 Simulations and observations

The value of m determined by these equations physically represents the typical overlap $N^{-1}\left[\boldsymbol{x}^0\cdot\hat{\boldsymbol{x}}\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}}$ between the original signal $\boldsymbol{x}^0$ and the solution $\hat{\boldsymbol{x}}$ of (3.2). Therefore, the typical value of MSE between $\boldsymbol{x}^0$ and $\hat{\boldsymbol{x}}$, which serves as the performance measure of the reconstruction problem, is evaluated as

$$\mathrm{MSE} = N^{-1}\left[\left|\hat{\boldsymbol{x}}-\boldsymbol{x}^0\right|^2\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}} = q+\rho\sigma_0^2-2m. \tag{3.22}$$

Note that in past studies on 1-bit CS, reconstruction performance was evaluated through directional MSE, which is defined by $|\frac{\hat{\boldsymbol{x}}}{|\hat{\boldsymbol{x}}|}-\frac{\boldsymbol{x}^0}{|\boldsymbol{x}^0|}|^2$ as scale information is lost.

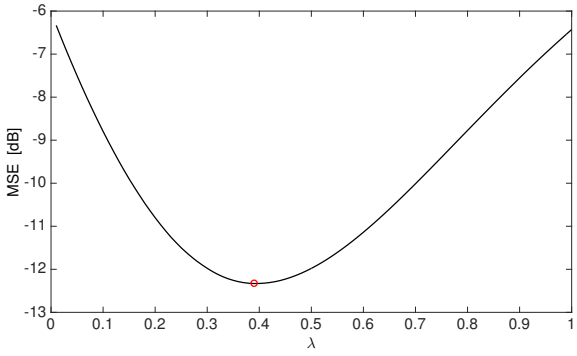

FIGURE 3.1: Replica prediction of MSE (in decibel) versus fixed threshold λ for signal distribution $\rho = 0.25, \sigma_0^2 = 1$, and ratio $\alpha = 3$.

We solved the saddle point equations for signal sparsity $\rho = 0.25$ and variance $\sigma_0^2 = 1$ when ratio $\alpha = 3$. The curve in Fig. 3.1 denotes the theoretical predictions of MSE as evaluated by (3.13)–(3.18) (strategy 1) and (3.22)

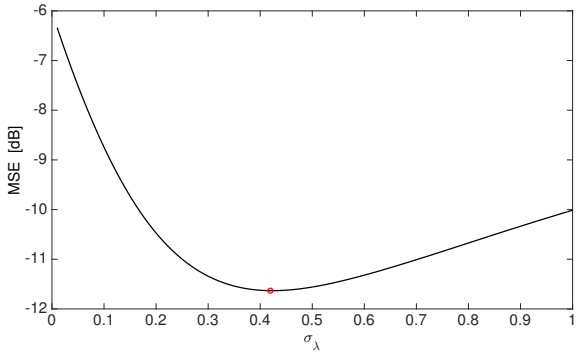

FIGURE 3.2: Replica prediction of MSE (in decibel) versus σ_λ for signal $\rho = 0.25, \sigma_0^2 = 1$, and ratio $\alpha = 3$.

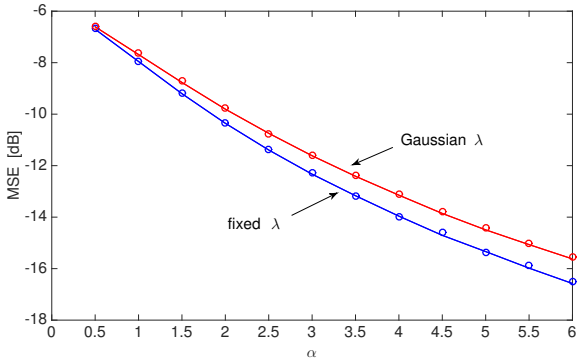

FIGURE 3.3: Lowest MSE [dB] (envelop) for each ratio α of signal $\rho = 0.25$, $\sigma_0^2 = 1$. The blue and red curves represent threshold strategies 1 and 2, respectively. The circles stand for the experimental estimate obtained using the CVX algorithm [20] averaged over $1,000$ experiments with signal size $N = 128$ for each parameter set.

plotted against the threshold λ. Fig. 3.2 represents the theoretical predictions of MSE evaluated by (3.19)–(3.21), (3.16)–(3.18) (strategy 2), and (3.22) plotted against the standard deviation σ_λ of the threshold. Figures 3.1 and 3.2 show that there is an optimal threshold distribution (red circle symbol) that minimizes MSE for each set of parameters. Similar features hold for various sets of values of α, ρ, σ_0^2 for both strategies 1 and 2.

To compare the optimal MSE (changing threshold distribution) of strategy 1 and strategy 2, we plot the optimal MSE for the same signal distribution in Fig. 3.1 and Fig. 3.2 against α in Fig. 3.3, which is referred to the envelope curve of MSE. The blue and red curves represent the envelope curves for strategies 1 and 2, respectively. From Fig. 3.3, we can see that strategy 1 outperforms strategy 2 when parameters are optimally tuned. Therefore, we hereafter focus on strategy 1, for which the thresholds are fixed.

The optimal value of λ depends on ρ and σ_0^2, which are not necessarily available in practice. To cope with such situations, we focus here on the distribution of binary output y, which indirectly conveys the information of $\rho\sigma_0^2$ and can be estimated from measurements. Fig. 3.4 shows the relation between the optimal MSE and $P(y = +1)$ for eight signal distributions. For

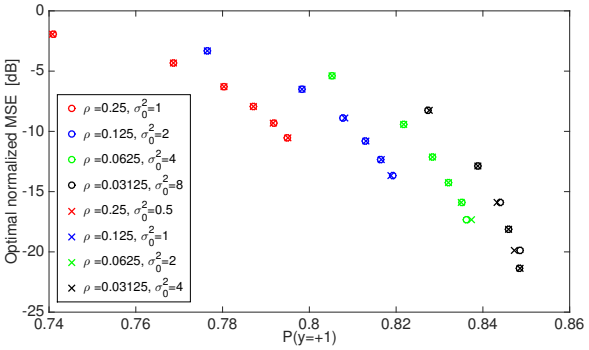

FIGURE 3.4: Optimal MSE: MSE$/(\rho\sigma_0^2)$ in decibel versus the probability of $+1$ in $\boldsymbol{y}$ for fixed threshold 1-bit CS model. Different colors represent varying signal sparsity. For each color, from left to right, the plot represents the result for $\alpha = 1, 2, \ldots, 6$.

given λ, the probability of positive output $y = +1$ is evaluated as

$$\begin{aligned} P(y=+1) &= \frac{1}{M}\left[\prod_{\mu=1}^{M}\Theta\left(\boldsymbol{\Phi}\boldsymbol{x_0}+\boldsymbol{\lambda}\right)_\mu\right]_{\boldsymbol{\Phi},\boldsymbol{x_0}} \\ &= \left[\Theta\left(\boldsymbol{\Phi}\boldsymbol{x_0}+\boldsymbol{\lambda}\right)_\mu\right]_{\boldsymbol{\Phi},\boldsymbol{x_0}} \\ &= \int_{-\infty}^{\infty}\mathrm{D}t\Theta\left(\sqrt{\rho}\sigma_0 t+\lambda\right) \\ &= \mathcal{Q}\left(-\frac{\lambda}{\sqrt{\rho}\sigma_0}\right). \end{aligned} \tag{3.23}$$

The horizontal axis in Fig. 3.4 is calculated from (3.23) by inserting the optimal value of λ. MSE is normalized by $\rho\sigma_0^2$ in order to eliminate its dependence on the scale of the original signal. For each color, from left to right, the plot represents the result for $\alpha = 1, 2, \ldots, 6$, respectively. Different colors represent varying signal sparsity. For each value of signal sparsity, we tested two values of signal variance. All circles have identical values of $\rho\sigma_0^2$, and all crosses have different values of $\rho\sigma_0^2$, which can be seen as the "power" of the signal. From these results, we can see that the normalized MSE, MSE$/(\rho\sigma_0^2)$, is the same when signal sparsity is the same. Besides, the data indicate that when the signal is sparser, the corresponding $P(y=+1)$ is greater. The data also show that the $P(y=+1)$ that yields the optimal MSE monotonically increases with α when the signal distribution is fixed. Although the optimal MSE depends on all system parameters ρ, σ_0^2, and compression rate α, we can see that the corresponding $P(y=+1)$ is always placed in the range of $0.75 \sim 0.85$ for modest values of $1 \leq \alpha \leq 6$. In addition, the data imply that although the optimal value of $P(y=+1)$ monotonically increases as α grows, it tends to converge to a value close to 0.85.

Algorithm 1: ADAPTIVE THRESHOLDING($\gamma, \delta, \lambda_0, U_0, V_0$)

1) **Initialization** :
   λ seed : λ_0
   Useed : $U_0 \leftarrow 0$
   Vseed : $V_0 \leftarrow 0$
   Counter : $k \leftarrow 0$
2) **Counter increase** :
   $k \leftarrow k + 1$
3) **Measurement of signal** :
   $y_k = \text{sign}\left(\sum_i \mathbf{\Phi}_{ki} \boldsymbol{x}_{\mathbf{0}i} + \lambda_k\right)$
4) **Update** T_k :
   $U_k \leftarrow (y_k > 0) + \gamma U_{k-1}$
   $V_k \leftarrow 1 + \gamma V_{k-1}$
   $T_k \leftarrow U_k / V_k$
5) **Update** λ :
   $\lambda_k \leftarrow \lambda_{k-1} + \delta \text{sign}(T - T_k)$
6) **Iteration** : Repeat from 2) until $k = M$.

FIGURE 3.5: Pseudocode for adaptive thresholding of 1-bit CS measurements. Here, y_k and λ_k for $k = 1, 2, ..., M$ represent each element of vector $\boldsymbol{y}$ and $\boldsymbol{\lambda}$, respectively. Signal reconstruction can be carried out by versatile convex optimization algorithms.

3.3 Learning algorithm for threshold

The results of the last section suggest that for each parameter set, the optimal threshold that minimizes MSE is loosely characterized by the value of $P(y = +1)$, which can be statistically estimated from the outputs of measurements. This property may be utilized to adaptively tune the threshold for each measurement based on the results of previous measurements.

A few studies have been conducted in the past on adaptive tuning of the threshold to improve signal reconstruction performance. For example, in [32], given past measurements, a threshold value was determined to partition the consistent region along its centroid computed by generalized approximate message passing [41, 57]. However, in many realistic situations, precise knowledge of the prior distribution is unavailable, even if we might reasonably expect the signal to be sparse. Therefore, we will here develop a learning algorithm that can be executed without knowledge of the prior distribution of the signal. There is another general adaptive algorithm called $\Sigma\Delta$ quantization [5]. However, its goal is to find a satisfactory quantized representation of real number measurement and requires preprocessing based on real number measurements. Instead, the algorithm we develop aims to directly minimize MSE, and needs no preprocessing.

As shown in Fig. 3.4, MSE is minimized when $P(y = +1)$ takes a value of $0.75 \sim 0.85$ for various sets of parameters. To incorporate this property, we propose a strategy that first fixes a target value of T for $P(y = +1)$, and

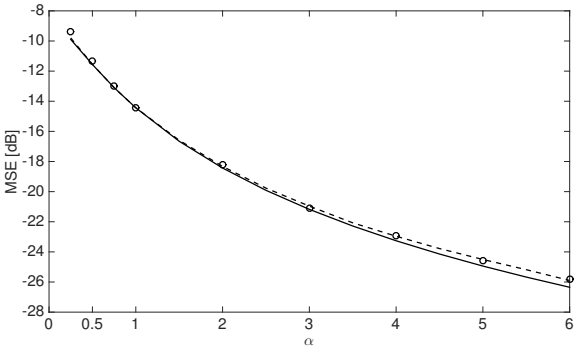

FIGURE 3.6: Experimental result from the adaptive thresholding algorithm for signal $\rho = 0.0625$, $\sigma_0^2 = 2$, and $N = 128$. The circles denote the average of $1,000$ experiments. The parameter settings were $T = 0.8$, $\gamma = 0.8$, $\lambda_0 = 0.5$, and $\delta = 0.01$. The broken line represents the replica prediction when λ is set to offer $P(y = +1) = 0.8$ while the full curve denotes this for optimally tuned λ.

tunes λ so that an empirical distribution of $P(y = +1)$ approaches T. As we see in Fig. 3.4, for larger values of α or sparser signals, we should set T as greater in the relevant range. There are various ways of estimating $P(y = +1)$ from the results of measurements. Of these, we use the damped average

$$T_\mu = \frac{\sum_{n=0}^{\mu-1} \gamma^n \delta_{y_{\mu-n},+1}}{\sum_{n=0}^{k-1} \gamma^n}, \tag{3.24}$$

since it can be computed in an online manner as

$$T_{\mu+1} = \frac{\gamma\left(1-\gamma^{\mu}\right) T_\mu + (1-\gamma)\delta_{y_{\mu+1},+1}}{1-\gamma^{\mu+1}}, \tag{3.25}$$

which does not require referring to the details of previous measurements. Here, the damping factor γ is a parameter that we have to set. In experiments, we set $\gamma = 0.8$; but as long as we tested it, the obtained performance was not particularly sensitive to the choice of this parameter. (3.4) indicates that $P(y = +1)$ monotonically increases as λ_μ grows. This implies that λ_μ should be increased when $T > T_{\mu-1}$, and decreased otherwise. To implement this idea, we design the learning algorithm of λ_μ as

$$\lambda_\mu = \lambda_{\mu-1} + \delta \text{sign}(T - T_{\mu-1}), \tag{3.26}$$

where δ denotes the step size that is also set by users. The pseudocode for adaptive thresholding 1-bit CS measurements is shown in Fig. 3.5. Following measurement, signal reconstruction can be carried out by versatile convex optimization algorithms [20] by solving (3.2).

Since we plan to apply the adaptive algorithm in situations involving a finite number of measurements, the extent to which the initial threshold λ_0 is remote from the optimal threshold λ_{opt}, which is unknown beforehand, and the variation in the step size δ may significantly influence reconstruction performance. In order to set an appropriate value of λ_0, we propose testing it by measuring the signal a few times. If the outputs are limited

almost exclusively to $+1$ or -1, we change the threshold through the bisection method, which involves dividing or multiplying it by 2 until the outputs are adequately mixed with $+1$ and -1. The resulting threshold should yield an appropriate value of λ_0 close to λ_{opt}. Having set λ_0, an appropriate value of δ should be in smaller order in order to tweak it to λ_{opt}.

The results of our numerical experiments are shown in Fig. 3.6 as circles. Each circle denotes the average of $1,000$ experiments for systems where $N = 128$. The parameter settings of the experiments were $T = 0.8$, $\gamma = 0.8$, $\lambda_0 = 0.5$, and $\delta = 0.01$ for signal distribution $\rho = 0.0625$ and $\sigma_0^2 = 2$. The solid line in Fig. 3.6 represents the envelop for MSE (dB) for each α. On the other hand, the dashed curve represents the prediction of MSE (dB) using replica analysis when $P(y = +1) = 0.8$, which was achieved by $\lambda = 0.2976$ according to (3.23). Fig.3.6 shows that the adaptive thresholding algorithm in conjunction with the employment of CVX for signal reconstruction can achieve nearly the same performance in terms of MSE as the statistical prediction for $P(y = +1) = 0.8$, and the result is reasonably close to the envelope MSE.

Note that we have also examed the possibility to develep a message passing algorithm. However, the algorithms doesn't converge well because of the replica symmetric assumption is not stable in the model 3.2. A brief sketch of the derivation of this condition is shown in E.

3.4 Summary

Prevalent schemes of 1-bit compressed sensing can only reconstruct the directional information of signals. While maintaining the advantage of 1-bit measurement, we proposed in this study a method that can reconstruct both the scaling and the directional information of the signal by imposing a threshold prior to quantization. Considering the most general situation, where no detailed prior knowledge of sparse signals is available, we employed the l_1-norm minimization approach. By utilizing the replica method from statistical mechanics, the mean squared error behavior of reconstruction for standard i.i.d measurement matrix and i.i.d Bernoulli-Gaussian signal was derived in the large system size limit. We compared two design strategies for the elements of the threshold vector, which corresponded to setting a fixed or random value as threshold. Our analysis showed that the fixed threshold strategy can achieve lower MSE than the random threshold strategy.

Another observation from the replica results was that there is an optimal threshold that minimizes MSE for a set of signal distributions and measurement ratios. However, in order to evaluate the optimal threshold, we need to know the prior distribution of the signal, which is not necessarily available in practical situations. Therefore, we shifted our focus to the relation between the optimal threshold and the distribution of the binary outputs, which can be empirically evaluated from signal measurements. The replica analysis indicated that the MSE is minimized when $P(y = +1)$ is set in the vicinity of $0.75 \sim 0.85$ for a wide region of system parameters.

On the basis of this observation, an algorithm that adaptively tunes the threshold at each measurement in order to obtain $P(y = +1)$ close to our target value was proposed. Combined with versatile convex optimization

algorithms, the adaptive thresholding algorithm offers a computationally feasible and widely applicable 1-bit CS scheme. Numerical experiments showed that it can yield nearly optimal performance, even when no detailed prior knowledge of sparse signals is available.

Improvements on the adaptive thresholding algorithm as well as the application of the algorithm to practical problems form part of our future research in the area.

Chapter 4

Bayesian inference

The most widely used signal reconstruction scheme in CS is l_1-norm minimization, which searches for the vector with the smallest l_1-norm $||\boldsymbol{x}||_1 = \sum_{i=1}^{N} |x_i|$ under the constraint $\boldsymbol{y} = \boldsymbol{\Phi x}$. This is based on the work of Candès et al. [7]–[6], who also suggested the use of a random measurement matrix $\boldsymbol{\Phi}$ with independent and identically distributed entries. Because the optimization problem is convex and can be solved using efficient linear programming techniques, these ideas have led to various fast and efficient algorithms. The l_1-reconstruction is now widely used, and is responsible for the surge of interest in CS over the past few years. Against this background, l_1-reconstruction was the first technique attempted in the development of the 1-bit CS problem. In [49], an approximate signal recovery algorithm was proposed based on the minimization of the l_1-norm under the constraint $\mathrm{sign}\,(\boldsymbol{\Phi x}) = \boldsymbol{y}$, and its utility was demonstrated by numerical experiments. The capabilities of this method were analyzed, and a new algorithm based on the cavity method was presented[56]. However, as we mentioned before, the significance of the l_1-based scheme may be rather weak for 1-bit CS, because the loss of convexity prevents the development of mathematically guaranteed and practically feasible algorithms.

In this chapter, we propose another approach based on Bayesian inference for 1-bit CS, focused on the case that each entry of $\boldsymbol{\Phi}$ is independently generated from a standard Gaussian distribution, and the output $\boldsymbol{y}$ is noiseless. Although the Bayesian approach is guaranteed to achieve the optimal performance when the actual signal distribution is given, quantifying the performance gain is a nontrivial task. We accomplish this task utilizing the replica method, which shows that the Bayesian optimal inference asymptotically saturates the mean squared error (MSE) performance obtained when the positions of non-zero signal entries are known as $\alpha = M/N \to \infty$. This means that, at least in terms of MSEs, the correct prior knowledge of the sparsity asymptotically becomes as informative as the knowledge of the exact positions of the non-zero entries. Unfortunately, performing the exact Bayesian inference is computationally difficult. This difficulty is resolved by employing the generalized approximate message passing technique, which is regarded as a variation of belief propagation or the cavity method [41, 27].

The content in this chapter is publish in [57].

4.1 Problem setup and Bayesian optimality

Let us suppose that entry x_i^0 $(i = 1, 2, \ldots, N)$ of an N-dimensional signal (vector) $\boldsymbol{x}^0 = (x_i^0) \in \mathbb{R}^N$ is independently generated from an identical

sparse distribution:

$$P(x) = (1-\rho)\,\delta(x) + \rho\tilde{P}(x), \tag{4.1}$$

where $\rho \in [0,1]$ represents the density of nonzero entries in the signal, and $\tilde{P}(x)$ is a distribution function of $x \in \mathbb{R}$ that has a finite second moment and does not have finite mass at $x = 0$. In 1-bit CS, the measurement is performed as

$$\boldsymbol{y} = \mathrm{sign}\left(\boldsymbol{\Phi}\boldsymbol{x}^{\mathbf{0}}\right), \tag{4.2}$$

where $\mathrm{sign}(x) = x/|x|$ operates in a component-wise manner, and for simplicity we assume that each entry of the $M \times N$ measurement matrix $\boldsymbol{\Phi}$ is provided as a sample of a Gaussian distribution of zero mean and variance N^{-1}.

We shall adopt the Bayesian approach to reconstruct the signal from the 1-bit measurement $\boldsymbol{y}$ assuming that $\boldsymbol{\Phi}$ is correctly known in the recovery stage. Let us denote an arbitrary recovery scheme for the measurement $\boldsymbol{y}$ as $\hat{\boldsymbol{x}}(\boldsymbol{y})$, where we impose a normalization constraint $|\hat{\boldsymbol{x}}(\boldsymbol{y})|^2 = N\rho$ to compensate for the loss of amplitude information by the 1-bit measurement. Equations (4.1) and (4.2) indicate that, for a given $\boldsymbol{\Phi}$, the joint distribution of the sparse vector and its 1-bit measurement is

$$P(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\Phi}) = \prod_{\mu=1}^{M}\Theta\left(y_\mu(\boldsymbol{\Phi}\boldsymbol{x})_\mu\right) \times \prod_{i=1}^{N}\left((1-\rho)\,\delta(x_i) + \rho\tilde{P}(x_i)\right), \tag{4.3}$$

where $\Theta(x) = 1$ for $x > 0$, and vanishes otherwise. This generally provides $\hat{\boldsymbol{x}}(\cdot)$ with the mean square error, which is hereafter handled as the performance measure for the signal reconstruction, as follows:

$$\mathrm{MSE}(\hat{\boldsymbol{x}}(\cdot)) = \sum_{\boldsymbol{y}}\int d\boldsymbol{x} P(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\Phi})\left|\frac{\hat{\boldsymbol{x}}(\boldsymbol{y})}{|\hat{\boldsymbol{x}}(\boldsymbol{y})|} - \frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right|^2. \tag{4.4}$$

The following theorem forms the basis of our Bayesian approach.

Theorem 1. $\mathrm{MSE}(\hat{\boldsymbol{x}}(\cdot))$ *is lower bounded as*

$$\mathrm{MSE}(\hat{\boldsymbol{x}}(\cdot)) \geq 2\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{\Phi})\left(1 - \left|\left\langle\frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right\rangle_{|\boldsymbol{y},\boldsymbol{\Phi}}\right|\right), \tag{4.5}$$

where

$$\begin{aligned} P(\boldsymbol{y}|\boldsymbol{\Phi}) &= \int d\boldsymbol{x} P(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\Phi}) \\ &= \int d\boldsymbol{x}\prod_{\mu=1}^{M}\Theta\left(y_\mu(\boldsymbol{\Phi}\boldsymbol{x})_\mu\right) \times \prod_{i=1}^{N}\left((1-\rho)\,\delta(x_i) + \rho\tilde{P}(x_i)\right) \end{aligned} \tag{4.6}$$

is the marginal distribution of the 1-bit measurement $\boldsymbol{y}$ *and*

$$\langle f(\boldsymbol{x})\rangle_{|\boldsymbol{y},\boldsymbol{\Phi}} = \int d\boldsymbol{x} f(\boldsymbol{x}) P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\Phi}) = \int d\boldsymbol{x} f(\boldsymbol{x}) P(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\Phi})/P(\boldsymbol{y}|\boldsymbol{\Phi}) \tag{4.7}$$

generally denotes the posterior mean of an arbitrary function of $\boldsymbol{x}$, $f(\boldsymbol{x})$, given $\boldsymbol{y}$. The equality holds for the Bayesian optimal signal reconstruction

$$\hat{\boldsymbol{x}}^{\mathrm{Bayes}}(\boldsymbol{y}) = \sqrt{N\rho}\left|\left\langle \frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right\rangle_{|\boldsymbol{y},\boldsymbol{\Phi}}\right|^{-1}\left\langle \frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right\rangle_{|\boldsymbol{y},\boldsymbol{\Phi}}. \tag{4.8}$$

Proof. Employing the Bayes formula $P(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\Phi}) = P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\Phi})P(\boldsymbol{y}|\boldsymbol{\Phi})$ in (4.4) yields the expression

$$\begin{aligned}
\mathrm{MSE}(\hat{\boldsymbol{x}}(\cdot)) &= \sum_{\boldsymbol{y}}\int d\boldsymbol{x}P(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\Phi})\left|\frac{\hat{\boldsymbol{x}}(\boldsymbol{y})}{|\hat{\boldsymbol{x}}(\boldsymbol{y})|}-\frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right|^2 \\
&= \sum_{\boldsymbol{y}}\int d\boldsymbol{x}P(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\Phi})P(\boldsymbol{y}|\boldsymbol{\Phi})\left(\left|\frac{\hat{\boldsymbol{x}}(\boldsymbol{y})}{|\hat{\boldsymbol{x}}(\boldsymbol{y})|}\right|^2+\left|\frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right|^2-2\frac{\hat{\boldsymbol{x}}(\boldsymbol{y})\cdot\boldsymbol{x}}{|\hat{\boldsymbol{x}}(\boldsymbol{y})||\boldsymbol{x}|}\right) \\
&= 2\sum_{\boldsymbol{y}}P(\boldsymbol{y}|\boldsymbol{\Phi})\left(1-\frac{\hat{\boldsymbol{x}}(\boldsymbol{y})}{|\hat{\boldsymbol{x}}(\boldsymbol{y})|}\cdot\left\langle\frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right\rangle_{|\boldsymbol{y},\boldsymbol{\Phi}}\right).
\end{aligned} \tag{4.9}$$

Inserting the Cauchy–Schwarz inequality

$$\hat{\boldsymbol{x}}(\boldsymbol{y})\cdot\left\langle\frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right\rangle_{|\boldsymbol{y},\boldsymbol{\Phi}} \le |\hat{\boldsymbol{x}}(\boldsymbol{y})|\left|\left\langle\frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right\rangle_{|\boldsymbol{y},\boldsymbol{\Phi}}\right| \tag{4.10}$$

into the right-hand side of (4.9) yields the lower bound of (4.5), where the equality holds when $\hat{\boldsymbol{x}}(\boldsymbol{y})$ is parallel to $\left\langle\frac{\boldsymbol{x}}{|\boldsymbol{x}|}\right\rangle_{|\boldsymbol{y},\boldsymbol{\Phi}}$. This, in conjunction with the normalization constraint of $\hat{\boldsymbol{x}}(\boldsymbol{y})$, leads to (4.8). □

The above theorem guarantees that the Bayesian approach achieves the best possible performance in terms of MSE. Therefore, we hereafter focus on the reconstruction scheme of (4.8), quantitatively evaluate its performance, and develop a computationally feasible approximate algorithm.

4.2 Performance assessment by the replica method

In statistical mechanics, the macroscopic behavior of the system is generally analyzed by evaluating the partition function or its negative logarithm, free energy. In our signal reconstruction problem, the marginal likelihood $P(\boldsymbol{y}|\boldsymbol{\Phi})$ of (4.6) plays the role of the partition function. However, this still depends on the quenched random variables $\boldsymbol{y}$ and $\boldsymbol{\Phi}$. Therefore, we must further average the free energy as $\bar{f} \equiv -N^{-1}\left[\log P(\boldsymbol{y}|\boldsymbol{\Phi})\right]_{\boldsymbol{y},\boldsymbol{\Phi}}$ to evaluate the typical performance, where $[\cdots]_{\boldsymbol{y},\boldsymbol{\Phi}}$ denotes the configurational average concerning $\boldsymbol{y}$ and $\boldsymbol{\Phi}$.

Unfortunately, directly averaging the logarithm of random variables is, in general, technically difficult. Thus, we resort to the replica method to practically resolve this difficulty [14]. For this, we first evaluate the n-th moment of the marginal likelihood $\left[P^n(\boldsymbol{y}|\boldsymbol{\Phi})\right]_{\boldsymbol{\Phi},\boldsymbol{y}}$ for $n = 1, 2, \ldots \in \mathbb{N}$ using the formula

$$P^n(\boldsymbol{y}|\boldsymbol{\Phi}) = \int\prod_{a=1}^{n}\left(\mathrm{d}\boldsymbol{x}^a P(\boldsymbol{x}^a)\right)\prod_{a=1}^{n}\prod_{\mu=1}^{M}\Theta\left((\boldsymbol{y})_\mu(\boldsymbol{\Phi}\boldsymbol{x}^a)_\mu\right), \tag{4.11}$$

which holds only for $n = 1, 2, \ldots \in \mathbb{N}$. Here, $\boldsymbol{x}^a$ $(a = 1, 2, \ldots, n)$ denotes the a-th replicated signal. Averaging (4.11) with respect to $\boldsymbol{\Phi}$ and $\boldsymbol{y}$ results in the saddle-point evaluation concerning the macroscopic variables $q_{0a} = q_{a0} \equiv N^{-1}\boldsymbol{x}^0 \cdot \boldsymbol{x}^a$ and $q_{ab} = q_{ba} \equiv N^{-1}\boldsymbol{x}^a \cdot \boldsymbol{x}^b$ $(a, b = 1, 2, \ldots, n)$.

Although (4.11) holds only for $n \in \mathbb{N}$, the expression $N^{-1}\log\left[P^n\left(\boldsymbol{y}|\boldsymbol{\Phi}\right)\right]_{\boldsymbol{\Phi},\boldsymbol{y}}$ obtained by the saddle-point evaluation under a certain assumption concerning the permutation symmetry with respect to the replica indices $a, b = 1, 2, \ldots n$ is obtained as an analytic function of n, which is likely to also hold for $n \in \mathbb{R}$. Therefore, we next utilize the analytic function to evaluate the average of the logarithm of the partition function as

$$\bar{f} = -\lim_{n\to 0}(\partial/\partial n)N^{-1}\log\left[P^n\left(\boldsymbol{y}|\boldsymbol{\Phi}\right)\right]_{\boldsymbol{y},\boldsymbol{\Phi}}. \tag{4.12}$$

In particular, under the replica symmetric ansatz, where the dominant saddle-point is assumed to be of the form

$$q_{ab} = q_{ba} = \begin{cases} \rho & (a = b = 0) \\ m & (a = 1, 2, \ldots, n;\ b = 0) \\ Q & (a = b = 1, 2, \ldots, n) \\ q & (a \neq b = 1, 2, \ldots, n) \end{cases}, \tag{4.13}$$

when the distribution of nonzero entries in (4.1) is given as the standard Gaussian $\tilde{P}(x) = \exp(-x^2/2)/\sqrt{2\pi}$, the above procedure expresses the average free energy density as

$$\begin{aligned} \bar{f} \;=\; & -\operatorname*{extr}_{\omega}\left\{\int \mathrm{d}x^0 P\left(x^0\right)\int \mathrm{D}z\phi\left(\sqrt{\hat{q}}z + \hat{m}x^0; \hat{Q}\right) + \frac{1}{2}Q\hat{Q} + \frac{1}{2}q\hat{q} - m\hat{m}\right. \\ & \left. +2\alpha\int \mathrm{D}t\mathcal{Q}\left(\frac{m}{\sqrt{\rho q - m^2}}t\right)\log\mathcal{Q}\left(\sqrt{\frac{q}{Q-q}}t\right)\right\}. \end{aligned} \tag{4.14}$$

Here, $\alpha = M/N$, $\mathcal{Q}(x) = \int_x^{+\infty}\mathrm{D}z$, $\mathrm{D}z = \mathrm{d}z\exp -z^2/2/\sqrt{2\pi}$ is a Gaussian measure, $\operatorname{extr}_X\{g(X)\}$ denotes the extremization of a function $g(X)$ with respect to X, $\omega = \{Q, q, m, \hat{Q}, \hat{q}, \hat{m}\}$, and

$$\begin{aligned} & \phi\left(\sqrt{\hat{q}}z + \hat{m}x^0; \hat{Q}\right) \\ = \;& \log\left\{\int \mathrm{d}x P\left(x\right)\exp\left(-\frac{\hat{Q}+\hat{q}}{2}x^2 + (\sqrt{\hat{q}}z + \hat{m}x^0)x\right)\right\}. \end{aligned} \tag{4.15}$$

The derivation of (4.14) is provided in F.

In evaluating the right-hand side of (4.12), $P(\boldsymbol{y}|\boldsymbol{\Phi})$ not only gives the marginal likelihood (the partition function), but also the conditional density of $\boldsymbol{y}$ for taking the configurational average. This accordance between the partition function and the distribution of the quenched random variables is generally known as the Nishimori condition in spin glass theory [39], and yields the identity $\left[P^n\left(\boldsymbol{y}|\boldsymbol{\Phi}\right)\right]_{\boldsymbol{y},\boldsymbol{\Phi}} = \int \mathrm{d}\boldsymbol{\Phi}P(\boldsymbol{\Phi})\left(\sum_{\boldsymbol{y}}P^{n+1}(\boldsymbol{y}|\boldsymbol{\Phi})\right)$, which indicates that the true signal, $\boldsymbol{x}^0$, can be handled on an equal footing with the other n replicated signals $\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^n$ in the replica computation. As $n \to 0$, this higher replica symmetry among the $n + 1$ replicated variables allows us to further simplify the replica symmetric ansatz (4.13)

by imposing four extra constraints: $Q = \rho$, $q = m$, $\hat{Q} = 0$, and $\hat{q} = \hat{m}$. As a consequence, the extremization condition of (4.14) is summarized by the non-linear equations

$$m = \int \mathrm{D}t \frac{\left(\int \mathrm{d}x x e^{-\frac{\hat{m}}{2}x^2+\sqrt{\hat{m}}tx} P(x)\right)^2}{\int \mathrm{d}x e^{-\frac{\hat{m}}{2}x^2+\sqrt{\hat{m}}tx} P(x)} \tag{4.16}$$

$$\hat{m} = \frac{\alpha}{\pi\sqrt{2\pi}\,(\rho - m)} \int \mathrm{d}t \frac{\exp\left\{-\frac{\rho+m}{2(\rho-m)}t^2\right\}}{\mathcal{Q}\left(\sqrt{\frac{m}{\rho-m}}t\right)}. \tag{4.17}$$

In physical terms, the value of m determined by these equations is the typical overlap $N^{-1}\left[\boldsymbol{x}^0 \cdot \langle \boldsymbol{x} \rangle_{|\boldsymbol{y},\boldsymbol{\Phi}}\right]_{\boldsymbol{y},\boldsymbol{\Phi}}$ between the original signal $\boldsymbol{x}^0$ and the posterior mean $\langle \boldsymbol{x} \rangle_{|\boldsymbol{y},\boldsymbol{\Phi}}$. The law of large numbers and the self-averaging property guarantee that both $N^{-1}|\boldsymbol{x}|^2$ and $N^{-1}|\boldsymbol{x}^0|^2$ converge to ρ with a probability of unity for typical samples. This indicates that the typical value of the direction cosine between $\boldsymbol{x}^0$ and $\hat{\boldsymbol{x}}^{\mathrm{Bayes}}(\boldsymbol{y})$ can be evaluated as

$$\begin{aligned} & \left[(\boldsymbol{x}^0 \cdot \hat{\boldsymbol{x}}^{\mathrm{Bayes}}(\boldsymbol{y}))/(|\boldsymbol{x}^0||\hat{\boldsymbol{x}}^{\mathrm{Bayes}}(\boldsymbol{y})|)\right]_{\boldsymbol{y},\boldsymbol{\Phi}} \\ \simeq\ & \left[(\boldsymbol{x}^0 \cdot \langle \boldsymbol{x} \rangle_{|\boldsymbol{y},\boldsymbol{\Phi}})\right]_{\boldsymbol{y},\boldsymbol{\Phi}} / \left(\left[\left|\boldsymbol{x}^0\right|\right]_{\boldsymbol{x}^0} \left[\left|\langle \boldsymbol{x} \rangle_{|\boldsymbol{y},\boldsymbol{\Phi}}\right|\right]_{\boldsymbol{y},\boldsymbol{\Phi}}\right) \\ =\ & Nm/(\sqrt{N\rho}\sqrt{Nm}) \\ =\ & \sqrt{m/\rho}. \end{aligned} \tag{4.18}$$

Therefore, the MSE in (4.4) can be expressed using m and ρ as

$$\mathrm{MSE} = 2\left(1 - \sqrt{\frac{m}{\rho}}\right). \tag{4.19}$$

The symmetry between $\boldsymbol{x}^0$ and the other replicated variables $\boldsymbol{x}^a$ $(a = 1, 2 \ldots, n)$ provides $\bar{f}$ with further information-theoretic meanings. Inserting $P(\boldsymbol{y}, \boldsymbol{\Phi}) = P(\boldsymbol{y}|\boldsymbol{\Phi})P(\boldsymbol{\Phi})$ into the definition of $\bar{f}$ gives

$$\bar{f} = N^{-1} \int \mathrm{d}\boldsymbol{\Phi} P(\boldsymbol{\Phi}) \left(-\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{\Phi}) \log P(\boldsymbol{y}|\boldsymbol{\Phi})\right), \tag{4.20}$$

which indicates that $\bar{f}$ accords with the entropy density of $\boldsymbol{y}$ for typical measurement matrices $\boldsymbol{\Phi}$. The expression

$$P(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\Phi}) = \prod_{\mu=1}^{M} \Theta\left(y_\mu (\boldsymbol{\Phi}\boldsymbol{x})_\mu\right) \in \{0, 1\} \tag{4.21}$$

guarantees that the conditional entropy of $\boldsymbol{y}$ given $\boldsymbol{x}$ and $\boldsymbol{\Phi}$, $-\sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\Phi}) \log P(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\Phi})$, always vanishes. These indicate that $\bar{f}$ also implies a mutual information density between $\boldsymbol{y}$ and $\boldsymbol{x}$. This physically quantifies the optimal information gain (per entry) of $\boldsymbol{x}$ that can be extracted from the 1-bit measurement $\boldsymbol{y}$ for typical $\boldsymbol{\Phi}$.

4.3 Bayesian optimal signal reconstruction by GAMP

Equation (4.19) represents the potential performance of the Bayesian optimal signal reconstruction of 1-bit CS. However, in practice, exploiting this performance is a non-trivial task, because performing the exact Bayesian reconstruction (4.8) is computationally difficult. To resolve this difficulty, we now develop an approximate reconstruction algorithm following the framework of belief propagation (BP). Actually, BP has been successfully employed for standard CS problems with linear measurements, showing excellent performance in terms of both reconstruction accuracy and computational efficiency [12]. To incorporate the non-linearity of the 1-bit measurement, we employ a variant of BP known as generalized approximate message passing (GAMP) [41], which can also be regarded as an approximate Bayesian inference algorithm for perceptron-type networks [27].

In general, the canonical BP equations for the probability measure $P(\boldsymbol{x}|\boldsymbol{\Phi},\boldsymbol{y})$ are expressed in terms of $2MN$ messages, $m_{i\to\mu}(x_i)$ and $m_{\mu\to i}(x_i)$ $(i=1,2,\cdots,N;\mu=1,2,\cdots,M)$, which represent probability distribution functions that carry posterior information and output measurement information, respectively. They can be written as

$$m_{\mu\to i}(x_i) = \frac{1}{Z_{\mu\to i}}\int\prod_{j\neq i}\mathrm{d}x_j P(y_\mu|u_\mu)\prod_{j\neq i}m_{j\to\mu}(x_j) \tag{4.22}$$

$$m_{i\to\mu}(x_i) = \frac{1}{Z_{i\to\mu}}P(x_i)\prod_{\gamma\neq\mu}m_{\gamma\to i}(x_i) \tag{4.23}$$

Here, $Z_{\mu\to i}$ and $Z_{i\to\mu}$ are normalization factors ensuring that

$$\int\mathrm{d}x_i m_{\mu\to i}(x_i)=\int\mathrm{d}x_i m_{i\to\mu}(x_i)=1, \tag{4.24}$$

and we also define $u_\mu\equiv(\boldsymbol{\Phi}\boldsymbol{x})_\mu$. Using (4.22), the approximation of marginal distributions $P(x_i|\boldsymbol{\Phi},\boldsymbol{y})=\int\prod_{j\neq i}\mathrm{d}x_j P(\boldsymbol{x}|\boldsymbol{\Phi},\boldsymbol{y})$, which are often termed beliefs, are evaluated as

$$m_i(x_i)=\frac{1}{Z_i}P(x_i)\prod_{\mu=1}^{M}m_{\mu\to i}(x_i), \tag{4.25}$$

where Z_i is a normalization factor for $\int\mathrm{d}x_i m_i(x_i)=1$. To simplify the notation, we hereafter convert all measurement results to $+1$ by multiplying each row of the measurement matrix $\boldsymbol{\Phi}=(\Phi_{\mu i})$ by y_μ $(\mu=1,2,\ldots,N)$, giving $(\Phi_{\mu i})\to(y_\mu\Phi_{\mu i})$, and denote the resultant matrix as $\boldsymbol{\Phi}=(\Phi_{\mu i})$. In the new notation, $P(y_\mu|u_\mu)=\Theta(u_\mu)$.

Next, we introduce means and variances of x_i in the posterior information message distributions as

$$a_{i\to\mu} \equiv \int\mathrm{d}x_i x_i m_{i\to\mu}(x_i) \tag{4.26}$$

$$\nu_{i\to\mu} \equiv \int\mathrm{d}x_i x_i^2 m_{i\to\mu}(x_i)-a_{i\to\mu}^2. \tag{4.27}$$

We also define $\omega_\mu \equiv \sum_i \Phi_{\mu i} a_{i\to\mu}$ and $V_\mu \equiv \sum_i \Phi_{\mu i}^2 \nu_{i\to\mu}$ for notational convenience. Similarly, the means and variances of the beliefs, a_i and ν_i, are introduced as $a_i \equiv \int \mathrm{d}x_i x_i m_i(x_i)$ and $\nu_i \equiv \int \mathrm{d}x_i x_i^2 m_i(x_i) - a_i^2$. Note that $\boldsymbol{a} = (a_i)$ represents the approximation of the posterior mean $\langle \boldsymbol{x} \rangle_{|\boldsymbol{y},\boldsymbol{\Phi}}$. This, in conjunction with a consequence of the law of large numbers $\langle \boldsymbol{x}/|\boldsymbol{x}| \rangle_{|\boldsymbol{y},\boldsymbol{\Phi}} \simeq \langle \boldsymbol{x} \rangle_{|\boldsymbol{y},\boldsymbol{\Phi}} / \sqrt{N\rho}$, indicates that the Bayesian optimal reconstruction is approximately performed as $\hat{\boldsymbol{x}}^{\mathrm{Bayes}}(\boldsymbol{y}) \simeq \sqrt{N\rho}\boldsymbol{a}/|\boldsymbol{a}|$.

To enhance the computational tractability, let us rewrite the functional equations of (4.22) and (4.23) into algebraic equations using sets of $a_{i\to\mu}$ and $\nu_{i\to\mu}$. To do this, we insert the identity

$$\begin{aligned} 1 &= \int \mathrm{d}u_\mu \delta\left(u_\mu - \sum_{i=1}^N \Phi_{\mu i} x_i\right) \\ &= \int \mathrm{d}u_\mu \frac{1}{2\pi} \int \mathrm{d}\hat{u}_\mu \exp\left\{-i\hat{u}_\mu \left(u_\mu - \sum_{i=1}^N \Phi_{\mu i} x_i\right)\right\} \end{aligned} \tag{4.28}$$

into (4.22), which yields

$$\begin{aligned} m_{\mu\to i}(x_i) &= \frac{1}{2\pi Z_{\mu\to i}} \int \mathrm{d}u_\mu P(y_\mu|u_\mu) \int \mathrm{d}\hat{u}_\mu \exp\left\{-i\hat{u}_\mu (u_\mu - \Phi_{\mu i} x_i)\right\} \\ &\quad \times \prod_{j\neq i} \left\{ \int \mathrm{d}x_j m_{j\to\mu}(x_j) \exp\left\{i\hat{u}_\mu \Phi_{\mu j} x_j\right\}\right\}. \end{aligned} \tag{4.29}$$

The smallness of $\Phi_{\mu i}$ allows us to truncate the Taylor series of the last exponential in equation (4.29) up to the second order of $i\hat{u}_\mu \Phi_{\mu j} x_j$. Integrating $\int \mathrm{d}x_j m_{j\to\mu}(x_j)(\ldots)$ for $j \neq i$, we obtain the expression

$$\begin{aligned} m_{\mu\to i}(x_i) &= \frac{1}{2\pi Z_{\mu\to i}} \int \mathrm{d}u_\mu P(y_\mu|u_\mu) \int \mathrm{d}\hat{u}_\mu \exp\left\{-i\hat{u}_\mu (u_\mu - \Phi_{\mu i} x_i)\right\} \\ &\quad \times \exp\left\{ i\hat{u}_\mu(\omega_\mu - \Phi_{\mu i} a_{i\to\mu}) - \frac{\hat{u}_\mu^2}{2}(V_\mu - \Phi_{\mu i}^2 \nu_{i\to\mu})\right\}, \end{aligned} \tag{4.30}$$

and carrying out the resulting Gaussian intergral of $\hat{u}_\mu$, we obtain

$$\begin{aligned} m_{\mu\to i}(x_i) &= \frac{1}{Z_{\mu\to i}\sqrt{2\pi(V_\mu - \Phi_{\mu i}^2 \nu_{i\to\mu})}} \int \mathrm{d}u_\mu P(y_\mu|u_\mu) \\ &\quad \times \exp\left\{ -\frac{(u_\mu - \omega_\mu - \Phi_{\mu i}(x_i - a_{i\to\mu}))^2}{2(V_\mu - \Phi_{\mu i}^2 \nu_{i\to\mu})}\right\}. \end{aligned} \tag{4.31}$$

Since $\Phi_{\mu i}^2$ vanishes as $O(N^{-1})$ while $\nu_{i\to\mu} \sim O(1)$, we can omit $\Phi_{\mu i}^2 \nu_{i\to\mu}$ in (4.31). In addition, we replace $\Phi_{\mu j}^2$ in $V_\mu = \sum_i \Phi_{\mu j}^2 \nu_{i\to\mu}$ with its expectation N^{-1}, utilizing the law of large numbers. This removes the dependence on the index μ, making all V_μ equal to their average

$$V \equiv \frac{1}{N} \sum_{i=1}^N \nu_i. \tag{4.32}$$

The smallness of $\Phi_{\mu i}(x_i - a_{i\to\mu})$ again allows us to truncate the Taylor series of the exponential in (4.31) up to the second order. Thus, we have a parameterized expression of $m_{\mu\to i}(x_i)$:

$$m_{\mu\to i}(x_i) \propto \exp\left\{-\frac{A_{\mu\to i}}{2}x_i^2 + B_{\mu\to i}x_i\right\}, \tag{4.33}$$

where the parameters $A_{\mu\to i}$ and $B_{\mu\to i}$ are evaluated as

$$A_{\mu\to i} = (g'_{\text{out}})_\mu \Phi_{\mu i}^2 \tag{4.34}$$
$$B_{\mu\to i} = (g_{\text{out}})_\mu \Phi_{\mu i} + (g'_{\text{out}})_\mu \Phi_{\mu i}^2 a_{i\to\mu} \tag{4.35}$$

using

$$(g_{\text{out}})_\mu \equiv \frac{\partial}{\partial\omega_\mu}\log\left(\int \mathrm{d}u_\mu P(y_\mu|u_\mu)\exp\left(-\frac{(u_\mu-\omega_\mu)^2}{2V}\right)\right) \tag{4.36}$$
$$(g'_{\text{out}})_\mu \equiv -\frac{\partial^2}{\partial\omega_\mu^2}\log\left(\int \mathrm{d}u_\mu P(y_\mu|u_\mu)\exp\left(-\frac{(u_\mu-\omega_\mu)^2}{2V}\right)\right). \tag{4.37}$$

The derivation of these is given in G. Equations (4.34) and (4.35) act as the algebraic expression of (4.22). In the sign output channel, inserting $P(y_\mu|u_\mu) = \Theta(u_\mu)$ into (4.36) gives $(g_{\text{out}})_\mu$ and $(g'_{\text{out}})_\mu$ for 1-bit CS as

$$(g_{\text{out}})_\mu = \frac{\exp\left(-\frac{\omega_\mu^2}{2V}\right)}{\sqrt{2\pi V}\mathcal{Q}\left(-\frac{\omega_\mu}{\sqrt{V}}\right)} \tag{4.38}$$
$$(g'_{\text{out}})_\mu = (g_{\text{out}})_\mu^2 + \frac{\omega_\mu}{V}(g_{\text{out}})_\mu. \tag{4.39}$$

To obtain a similar expression for (4.23), we substitute the last expression of (4.33) into (4.23), which leads to

$$m_{i\to\mu}(x_i) = \frac{1}{\tilde{Z}_{i\to\mu}}\left[(1-\rho)\delta(x_i) + \rho\tilde{P}(x_i)\right] e^{-(x_i^2/2)\sum\limits_{\gamma\neq\mu} A_{\gamma\to i} + x_i \sum\limits_{\gamma\neq\mu} B_{\gamma\to i}} \tag{4.40}$$

This indicates that $\prod_{\gamma\neq\mu} m_{\gamma\to i}(x_i)$ in (4.23) can be expressed as a Gaussian distribution with mean $(\sum_{\gamma\neq\mu} B_{\gamma\to i})/(\sum_{\gamma\neq\mu} A_{\gamma\to i})$ and variance $(\sum_{\gamma\neq\mu} A_{\gamma\to i})^{-1}$. Inserting these into (4.26) and (4.27) provides the algebraic expression of (4.23) as

$$a_{i\to\mu} = f_a\left(\frac{1}{\sum_{\gamma\neq\mu} A_{\gamma\to i}}, \frac{\sum_{\gamma\neq\mu} B_{\gamma\to i}}{\sum_{\gamma\neq\mu} A_{\gamma\to i}}\right), \tag{4.41}$$
$$\nu_{i\to\mu} = f_c\left(\frac{1}{\sum_{\gamma\neq\mu} A_{\gamma\to i}}, \frac{\sum_{\gamma\neq\mu} B_{\gamma\to i}}{\sum_{\gamma\neq\mu} A_{\gamma\to i}}\right), \tag{4.42}$$

where we define

$$f_a(\Sigma^2, R) \equiv \frac{\bar{x}\Sigma^2 + R\sigma^2}{\frac{(1-\rho)(\sigma^2+\Sigma^2)^{3/2}}{\rho\Sigma}\exp\left\{-\frac{R^2}{2\Sigma^2}+\frac{(R-\bar{x})^2}{2(\sigma^2+\Sigma^2)}\right\}+(\sigma^2+\Sigma^2)}, \tag{4.43}$$

$$\begin{aligned} f_c(\Sigma^2, R) \equiv & \left\{\rho(1-\rho)\Sigma\left(\sigma^2+\Sigma^2\right)^{-5/2}\left[\sigma^2\Sigma^2\left(\sigma^2+\Sigma^2\right)+\left(\bar{x}\Sigma^2+R\sigma^2\right)^2\right]\right. \\ & \times\exp\left\{-\frac{R^2}{2\Sigma^2}-\frac{(R-\bar{x})^2}{2(\sigma^2+\Sigma^2)}\right\}+\rho^2\exp\left\{-\frac{(R-\bar{x})^2}{\sigma^2+\Sigma^2}\right\}\frac{\sigma^2\Sigma^4}{(\sigma^2+\Sigma^2)^2}\Bigg\} \\ & \times\left\{(1-\rho)\exp\left\{-\frac{R^2}{2\Sigma^2}\right\}+\rho\frac{\Sigma}{\sqrt{\sigma^2+\Sigma^2}}\exp\left\{-\frac{(R-\bar{x})^2}{2(\sigma^2+\Sigma^2)}\right\}\right\}^{-2} \end{aligned} \tag{4.44}$$

$\bar{x}$ and σ^2 represent the average and variance of $\tilde{P}(x_i)$. In our case, we set $\bar{x}=0$ and $\sigma^2=1$. Note that the form of f_a and f_c depend only on the prior distribution $\tilde{P}(x)$.

For the signal reconstruction, we need to evaluate the moments of $m_i(x_i)$. This can be performed by simply adding back the μ dependent part to (4.41) and (4.42) as

$$a_i = f_a(\Sigma_i^2, R_i), \tag{4.45}$$

$$\nu_i = f_c(\Sigma_i^2, R_i), \tag{4.46}$$

where $\Sigma_i^2 = \left(\sum_\mu A_{\mu\to i}\right)^{-1}$, $R_i = \frac{\sum_\mu B_{\mu\to i}}{\sum_\mu A_{\mu\to i}}$. For large N, Σ_i^2 typically converges to a constant, independent of the index, as Σ^2. This, in conjunction with (4.34) and (4.35), yields

$$\Sigma^2 = \left(\frac{1}{N}\sum_\mu (g'_{\rm out})_\mu\right)^{-1}, \tag{4.47}$$

$$R_i = \left(\sum_\mu (g_{\rm out})_\mu \Phi_{\mu i}\right)\Sigma^2 + a_i. \tag{4.48}$$

BP updates $2MN$ messages using (4.34), (4.35), (4.41), and (4.42) ($i = 1, 2, \cdots N, \mu = 1, 2, \cdots M$) in each iteration. This requires a computational cost of $O(M^2\times N + M\times N^2)$ per iteration, which may limit the practical utility of BP to systems of relatively small size. To enhance the practical utility, let us rewrite the BP equations into those of $M+N$ messages for large N, which will result in a significant reduction of computational complexity to $O(M\times N)$ per iteration. To do this, we express $a_{i\to\mu}$ by applying Taylor's expansion to (4.41) around R_i as

$$\begin{aligned} a_{i\to\mu} &= f_a\left(\frac{1}{\sum_\gamma A_{\gamma\to i}-A_{\mu\to i}}, \frac{\sum_\gamma B_{\gamma\to i}-B_{\mu\to i}}{\sum_\gamma A_{\gamma\to i}-A_{\mu\to i}}\right) \\ &\simeq a_i + \frac{\partial f_a(\Sigma^2, R_i)}{\partial R_i}(-B_{\mu\to i}\Sigma^2) + O(N^{-1}), \end{aligned} \tag{4.49}$$

where $B_{\mu\to i}\sim O(N^{-1/2})$ and $\sum_\gamma A_{\gamma\to i} - A_{\mu\to i}$ is approximated as $\sum_\gamma A_{\gamma\to i} = \Sigma^{-2}$, because of the smallness of $A_{\mu\to i}\propto \Phi_{\mu i}^2 \sim O(N^{-1})$. Multiplying this

by $\Phi_{\mu i}$ and summing the resultant expressions over i yields

$$\omega_\mu = \sum_i \Phi_{\mu i} a_i - (g_{\rm out})_\mu V, \tag{4.50}$$

where we have used $\nu_i = f_c = \Sigma^2 \frac{\partial f_a}{\partial R_i}$, which can be confirmed by (4.43) and (4.44).

Let us assume that $\{(a_i, \nu_i)\}$ and $\{((g_{\rm out})_\mu, (g'_{\rm out})_\mu)\}$ are initially set to certain values. Inserting these into (4.32) and (4.50) gives V and $\{\omega_\mu\}$. Substituting these into equations (4.38) and (4.39) yields a set of $\{((g_{\rm out})_\mu, (g'_{\rm out})_\mu)\}$, which, in conjunction with $\{a_i\}$, offers Σ^2 and $\{R_i\}$ through (4.47) and (4.48). Inserting these into (4.45) and (4.46) offers a new set of $\{(a_i, \nu_i)\}$. In this way, the iteration of (4.32), (4.50) $\rightarrow$ (4.38), (4.39) $\rightarrow$ (4.47), (4.48) $\rightarrow$ (4.45), (4.46) $\rightarrow$ (4.32), (4.50) $\rightarrow \ldots$ constitutes a closed set of equations to update the sets of $\{(a_i, \nu_i)\}$ and $\{((g_{\rm out})_\mu, (g'_{\rm out})_\mu)\}$. This is the generic GAMP algorithm given a likelihood function $P(y|u)$ and a prior distribution $P(x)$ [41].

We term the entire procedure the Approximate Message Passing for 1-bit Compressed Sensing (1bitAMP) algorithm. The pseudocode of this algorithm is summarized in Figure 4.1. Three issues are noteworthy. First, for relatively large systems, e.g., $N = 1024$, the iterative procedure converges easily in most cases. Nevertheless, since it relies on the law of large numbers, some divergent behavior appears as N becomes smaller. Even for such cases, however, employing an appropriate damping factor in conjunction with a normalization of $|\boldsymbol{a}|$ at each update considerably improves the convergence property. Second, the most time-consuming parts of this iteration are the matrix-vector multiplications $\sum_\mu (g_{\rm out})_\mu \Phi_{\mu i}$ in (4.48) and $\sum_i \Phi_{\mu i} a_i$ in (4.50). This indicates that the computational complexity is $O(NM)$ per iteration. Finally, a_i in equation (4.48) and $(g_{\rm out})_\mu V$ in equation (4.50) correspond to what is known as the *Onsager reaction term* in the spin glass literature [50, 46]. These terms stabilize the convergence of 1bitAMP, effectively canceling the self-feedback effects.

4.4 Results

To examine the utility of 1bitAMP, we carried out numerical experiments for $N = 1024$ systems. We set initial conditions of $\boldsymbol{a} = 0\mathbf{1}, \boldsymbol{\nu} = \rho\mathbf{1}$, and $\boldsymbol{\omega} = \mathbf{1}$, where $\mathbf{1}$ is the N-dimensional vector whose entries are all unity, and stopped the algorithm after 20 iterations (Figure 4.3). The MSE results for various sets of α and ρ are shown as crosses in Figures 4.2 (a)–(d). Each cross denotes an experimental estimate obtained from 1000 experiments. The standard deviations are omitted, as they are smaller than the size of the symbols. The convergence time is short, which verifies the significant computational efficiency of 1bitAMP. For example, in a MATLAB® environment, for $\alpha = 3, \rho = 0.0625$, one experiment takes around 0.2 s.

To test the consistency of 1bitAMP with respect to replica theory, we solved the saddle-point equations (4.16) and (4.17) for each set of α and ρ. The blue curves in Figures 4.2 (a)–(d) show the theoretical MSE evaluated by (4.19) against α for $\rho = 0.03125, 0.0625, 0.125$, and 0.25. The excellent

Algorithm 1: APPROXIMATE MESSAGE PASSING FOR 1-BIT CS($\mathbf{a}^*, \nu^*, \omega^*$)

1) **Initialization** :
   a seed : $\mathbf{a}_0 \leftarrow \mathbf{a}^*$
   ν seed : $\nu_0 \leftarrow \nu^*$
   ω seed : $\omega_0 \leftarrow \omega^*$
   Counter : $k \leftarrow 0$
2) **Counter increase** :
   $k \leftarrow k + 1$
3) **Mean of variances of posterior information message distributions** :
   $\mathbf{V}_k \leftarrow \mathrm{N}^{-1}(\mathrm{sum}(\nu_{k-1}))\mathbf{1}$
4) **Self-feedback cancellation** :
   $\omega_k \leftarrow \mathbf{\Phi}\mathbf{a}_{k-1} - \mathbf{V}_k g_{\mathrm{out}}(\omega_{k-1}, \mathbf{V}_k)$
5) **Variances of output information message distributions** :
   $\Sigma_k^2 \leftarrow \mathrm{N}(\mathrm{sum}(g'_{\mathrm{out}}(\omega_k, \mathbf{V}_k)))^{-1}$
6) **Average of output information message distributions** :
   $(\mathbf{R})_k \leftarrow \mathbf{a}_{k-1} + (g_{\mathrm{out}}(\omega_k, \mathbf{V}_k)\mathbf{\Phi})\Sigma_k^2$
7) **Posterior mean** :
   $\mathbf{a}_k \leftarrow f_a(\Sigma_k^2\mathbf{1}, \mathbf{R}_k)$
8) **Posterior variance** :
   $\nu_k \leftarrow f_c(\Sigma_k^2\mathbf{1}, \mathbf{R}_k)$
9) **Iteration** : Repeat from step 2 until convergence.

FIGURE 4.1: Pseudocode for 1-bitAMP. $\mathbf{a}^*$, ν^*, and ω^* are the convergent vectors of $\mathbf{a}_k$, ν_k, and ω_k obtained in the previous loop. $\mathbf{1}$ is the N-dimensional vector whose entries are all unity. This figure is cited from [57]

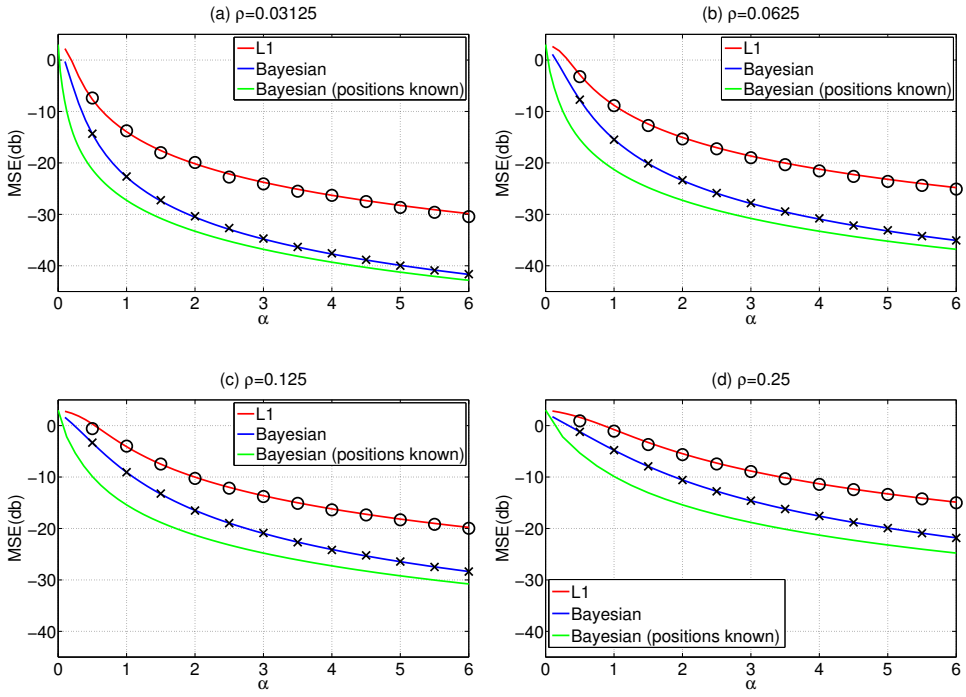

FIGURE 4.2: MSE (in decibels) versus measurement bit ratio α for 1-bit CS. (a), (b), (c), and (d) correspond to $\rho = 0.03125, 0.0625, 0.125$, and 0.25, respectively. Red curves represent the theoretical prediction of l_1-norm minimization [56]; blue curves represent the theoretical prediction of the Bayesian optimal approach; green curves represent the theoretical prediction of the Bayesian optimal approach when the positions of all nonzero components in the signal are known, which is obtained by setting $\alpha \to \alpha/\rho$ and $\rho \to 1$ in (4.16) and (4.17). Crosses represent the average of 1000 experimental results by the 1bitAMP algorithm in Figure 4.1 for a system size of $N = 1024$. Circles show the average of 1000 experimental results by an l_1-based algorithm RFPI proposed in [49] for 1-bit CS in the system size of $N = 128$. Although the replica symmetric prediction for the l_1-based approach is thermodynamically unstable, the experimental results of RFPI are numerically consistent with it very well.This figure is cited from [57]®IOP Publishing Ltd.

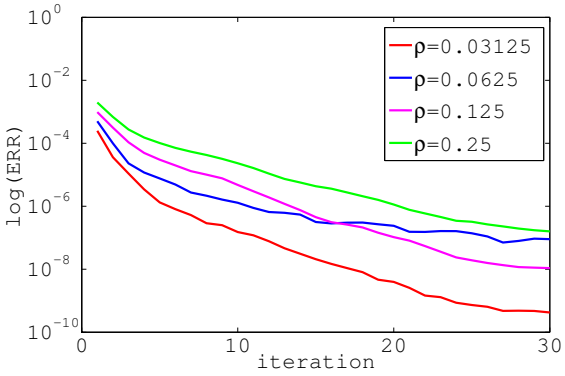

FIGURE 4.3: Mean square differences between estimated signals on two successive iterative update of 1bitAMP for a signal size of $N = 1024$ and $\alpha = 6$, which are evaluated from 100 experiments. Red, blue, magenta, and green represent $\rho = 0.03125, 0.0625, 0.125$, and 0.25, respectively. The cross between the blue and the magenta lines are considered as a result of the small number of experiment. This figure is cited from [57]®IOP Publishing Ltd.

agreement between the numerical experiments and the theoretical prediction indicates that 1bitAMP nearly saturates the potentially achievable MSE of the signal recovery scheme based on the Bayesian optimal approach.

For comparison, Figures 4.2 (a)–(d) also plot the replica symmetric prediction of MSEs for the l_1-norm minimization approach (red curves), which was examined in an earlier study [56]. Although the replica symmetric prediction is thermodynamically unstable, it is numerically consistent with the experimental results (circles) given by the algorithm proposed in [49]. Therefore, the prediction at least serves as a good approximation.

We also plot the MSEs of the Bayesian optimal approach when the positions of the non-zero components of $\boldsymbol{x}$ are known (green curves). These act as lower bounds for the MSEs of the Bayesian optimal approach. When the positions of non-zero components of $\boldsymbol{x}$ are known, we need not consider the part containing zero components. Therefore, the problem can be seen as that defined when a ρN-dimensional signal $\boldsymbol{x}$ is measured by an $\alpha N \times \rho N$-dimensional matrix. In such situations, performance can be evaluated by setting $\rho = 1$ and replacing α with α/ρ in (4.16) and (4.17), as the dimensionality of $\boldsymbol{x}$ is reduced from N to $N\rho$. Solving (4.16) and (4.17) for $\alpha \gg 1$ shows that the MSEs of the Bayesian optimal approach can be asymptotically expressed as

$$\mathrm{MSE}^{\mathrm{Bayes}} \simeq \frac{1.9258\rho^2}{\alpha^2} = 1.9258 \times \left(\frac{N\rho}{M}\right)^2 \tag{4.51}$$

for $\alpha \gg 1$, which accords exactly with the asymptotic form of the green curves (Figure 4.4: left panel, see H). On the other hand, the asymptotic form of the MSE for the l_1-norm approach is evaluated as

$$\mathrm{MSE}^{l_1} \simeq \frac{\pi^2\left[2(1-\rho)H\left(1/\sqrt{\hat{q}_{l_1}^{\infty}(\rho)}\right)+\rho\right]^2}{\alpha^2}, \tag{4.52}$$

where $\hat{q}_{l_1}^{\infty}(\rho)$ is the value of $\hat{q}$ for the l_1-norm approach obtained for $\alpha \to \infty$ (see I).

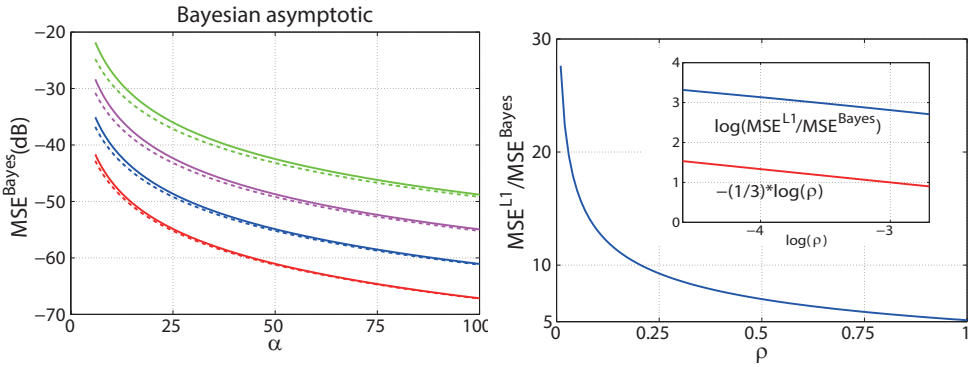

FIGURE 4.4: Left: MSE (in decibels) versus measurement bit ratio α for Bayesian optimal signal reconstruction of 1-bit CS. Red, blue, magenta, and green correspond to $\rho = 0.03125, 0.0625, 0.125$, and 0.25, respectively. The solid curves represent the theoretical prediction obtained by (4.16) and (4.17); dashed curves show the performance when the positions of non-zero entries are known, and dotted curves denote the asymptotic forms (4.51), which are indistinguishable from the dashed curves because they closely overlap. Right: Ratio of MSE between l_1-norm and Bayesian approaches when $\alpha \gg 1$ versus sparsity ρ of the signal. The inset shows a log-log plot for $0 < \rho < 0.1$. The least-squares fit implies that the ratio diverges as $O(\rho^{-0.33})$ as $\rho \to 0$. This figure is cited from [57]®IOP Publishing Ltd.

Equation (4.51) means that, at least in terms of MSEs, correct prior knowledge of the sparsity asymptotically becomes as informative as the knowledge of the exact positions of the non-zero components. In most statistical models, the accuracy of asymptotic inference is expressed as a function of the ratio $\alpha = M/N$ between the number of data M and the dimensionality of the variables to be inferred N [44, 54]. Equation (4.51) indicates that, in the current problem, the dimensionality N is replaced with the actual degree of the non-zero components $N\rho$, which originates from the singularity of the prior distribution (4.1). This implies that caution is necessary in testing the validity of statistical models when sparse priors are employed, since conventional information criteria such as Akaike's information criterion [1] and the minimum description length [42] mostly handle objective statistical models that are free of singularities, so that the model complexity is naively incorporated as the number of parameters N [53].

Equation (4.52) indicates that, even if prior knowledge of the sparsity is not available, optimal convergence can be achieved in terms of the "exponent" as $\alpha \to \infty$ using the l_1-norm approach. However, the performance can differ considerably in terms of the "pre-factor." The right panel of Figure 4.4 plots the ratio $\mathrm{MSE}^{l_1}/\mathrm{MSE}^{\mathrm{Bayes}}$, which diverges as $O(\rho^{-0.33})$ as $\rho \to 0$. This indicates that prior knowledge of the sparsity of the objective signal is more beneficial as ρ becomes smaller.

4.5 Summary

In this chapter, we have examined the typical performance of the Bayesian optimal signal recovery for 1-bit CS using methods from statistical mechanics. Using the replica method to compare the performance of the Bayesian optimal approach to the l_1-norm minimization, we have shown that the utility of correct prior knowledge on the objective signal, which is incorporated in the Bayesian optimal scheme, becomes more significant as the density of non-zero entries ρ in the signal decreases. In addition, we have clarified that the MSE performance asymptotically saturates that obtained when the exact positions of non-zero entries are exactly known as the number of 1-bit measurements increases. We have also developed a practically feasible approximate algorithm for Bayesian signal recovery, which can be regarded as a special case of the GAMP algorithm. The algorithm has a computational cost of the square of the system size per update, exhibiting a fairly good convergence property as the system size becomes larger. The experimental results show excellent agreement with the predictions made by the replica method. These indicate that almost-optimal reconstruction performance can be attained with a computational complexity of the square of the signal length per update, which is highly beneficial in practice.

Obtaining the correct prior distribution of the sparse signal may be an obstacle to applying the current approach in practical problems. One possible solution is to estimate hyper-parameters that characterize the prior distribution in the reconstruction stage, as has been proposed for nonquantized CS [33]. It was reported that orthogonal measurement matrices, rather than those of statistically independent entries, enhance the signal reconstruction performance for several problems related to CS [47, 29, 52, 28, 40, 55]. Such devices may also be effective for 1-bit CS.

Chapter 5

Conclusion

5.1 Summary of this thesis

In this thesis, we have analyzied typical performance of several differnt schemes of 1-bit compressed sensing using statistical mechanics methods. For each scheme, we also have developed efficient algorithms using statistical techniques. For simplicity, we consider the case that the measuring matrix has i.i.d entries, and the measurements are noiseless.

Main results of this thesis are summarized as follows:

- Analyzing the typical performance of an l_1-norm based signal recovery scheme for 1-bit CS for i.i.d Gaussian sparse signals (chapter 2).
- Developing an approximate recovery algorithm inspired by the cavity method for l_1-norm based 1-bit CS and numerically testing the reconstruction performance (chapter 3).
- Suggesting a strategy that introduce a threshold parameter to the quantization process in order to captures scale information for l_1-norm based signal recovery for 1-bit CS and analyzing the typical performance of it (chapter 3).
- Develop a heuristic that adaptively tunes the threshold parameter based on measurement results for l_1-norm based signal recovery for 1-bit CS and numerically checking the behaviour (chapter 3).
- Clarifying the statistical lower bound of the recovery performance of 1-bit CS by analyzing Bayesian inference approach (chapter 4).
- Showing that the Bayesian approach enables better reconstruction than the l_1-norm minimization approach, asymptotically saturating the performance obtained when the non-zero entries positions of the signal are known (chapter 4).
- Developing a message passing algorithm for signal reconstruction on the basis of belief propagation for Bayesian approach of 1-bit CS, and testing numerical experiments are consistent with those of the theoretical analysis (chapter 4).

We believe that our work has contributed to provide a deeper understanding and to the practical development to this field.

5.2 Future directions

Inspired by the study for non-quantized compressed sensing, we suggest some future research directions on 1-bit compressed sensing as follows:

- Improving the adaptive thresholding algorithm as well as the application of the algorithm to practical problems.
- Estimating hyper-parameters that characterize the prior distribution in the reconstruction stage when we do not know the exact prior distribution of the signal.
- Using orthogonal measurement matrices instead of those of statistically independent entries.

Of course it is not necessary to only focus on the extreem 1-bit case, it is also important to study the behaviour of multi-bits compressed sensing and clearify the relation between bits number and reconstruction performance.

Appendix A

Derivation of (2.8)

A.1 Assessment of $\left[Z^n\left(\beta;\Phi,\boldsymbol{x}^0\right)\right]_{\Phi,\boldsymbol{x}^0}$ for $n\in\mathbb{N}$

Averaging (2.5) with respect to $\boldsymbol{\Phi}$ and $\boldsymbol{x}^0$ offers the following expression of the n-th moment of the partition function:

$$\left[Z^n\left(\beta;\Phi,\boldsymbol{x}^0\right)\right]_{\Phi,\boldsymbol{x}^0}=\int\prod_{a=1}^{n}\left(d\boldsymbol{x}^a\delta\left(|\boldsymbol{x}^a|^2-N\right)\times e^{-\beta||\boldsymbol{x}^a||_1}\right)\times\left[\prod_{a=1}^{n}\prod_{\mu=1}^{M}\Theta\left((\boldsymbol{\Phi}\boldsymbol{x}^0)_\mu(\boldsymbol{\Phi}\boldsymbol{x}^a)_\mu\right)\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0}. \tag{A.1}$$

We insert $n(n+1)/2$ trivial identities

$$1=N\int dq_{ab}\delta\left(\boldsymbol{x}^a\cdot\boldsymbol{x}^b-Nq_{ab}\right), \tag{A.2}$$

where $a>b=0,1,2,\ldots,n$, into (A.1). Furthermore, we define a joint distribution of $n+1$ vectors $\{\boldsymbol{x}^a\}=\{\boldsymbol{x}^0,\boldsymbol{x}^1,\boldsymbol{x}^2,\ldots,\boldsymbol{x}^n\}$ as

$$P\left(\{\boldsymbol{x}^a\}|\boldsymbol{Q}\right)=\frac{1}{V\left(\boldsymbol{Q}\right)}P(\boldsymbol{x}^0)\times\prod_{a=1}^{n}\left(\delta\left(|\boldsymbol{x}^a|^2-N\right)\times e^{-\beta||\boldsymbol{x}^a||_1}\right)\times\prod_{a>b}\delta\left(\boldsymbol{x}^a\cdot\boldsymbol{x}^b-Nq_{ab}\right), \tag{A.3}$$

where $\boldsymbol{Q}=(q_{ab})$ is an $(n+1)\times(n+1)$ symmetric matrix whose 00 and the other diagonal entries are fixed as ρ and 1, respectively.

$$P(\boldsymbol{x}^0)=\prod_{i=1}^{N}\left((1-\rho)\delta(x_i^0)+\rho\tilde{P}(x_i^0)\right) \tag{A.4}$$

denotes the distribution of the original signal $\boldsymbol{x}^0$, and $V\left(\boldsymbol{Q}\right)$ is the normalization constant that makes $\int\prod_{a=0}^{n}d\boldsymbol{x}^aP\left(\{\boldsymbol{x}^a\}|\boldsymbol{Q}\right)=1$ hold. These indicate that (A.1) can also be expressed as

$$\left[Z^n\left(\beta;\Phi,\boldsymbol{x}^0\right)\right]_{\Phi,\boldsymbol{x}^0}=\int d\boldsymbol{Q}\left(V\left(\boldsymbol{Q}\right)\times\Xi\left(\boldsymbol{Q}\right)\right), \tag{A.5}$$

where $d\boldsymbol{Q} \equiv \prod_{a>b} dq_{ab}$ and

$$\Xi(\boldsymbol{Q}) = \int \prod_{a=0}^{n} d\boldsymbol{x}^a P(\{\boldsymbol{x}^a\}|\boldsymbol{Q}) \left[\prod_{a=1}^{n} \prod_{\mu=1}^{M} \Theta\left((\boldsymbol{\Phi}\boldsymbol{x}^0)_\mu (\boldsymbol{\Phi}\boldsymbol{x}^a)_\mu \right) \right]_{\boldsymbol{\Phi}}. \tag{A.6}$$

Equation (A.6) can be regarded as the average of $\prod_{a=1}^{n} \prod_{\mu=1}^{M} \Theta\left((\boldsymbol{\Phi}\boldsymbol{x}^0)_\mu (\boldsymbol{\Phi}\boldsymbol{x}^a)_\mu \right)$ with respect to $\{\boldsymbol{x}^a\}$ and $\boldsymbol{\Phi}$ over distributions of $P(\{\boldsymbol{x}^a\})$ and $P(\boldsymbol{\Phi}) \equiv \left(\sqrt{2\pi/N}\right)^{-MN} \exp\left(-(N/2)\sum_{\mu,i} \Phi_{\mu i}^2\right)$. In computing this, it is noteworthy that the central limit theorem guarantees that $u_\mu^a \equiv (\boldsymbol{\Phi}\boldsymbol{x}^a)_\mu = \sum_{i=1}^{N} \Phi_{\mu i} x_i^a$ can be handled as zero-mean multivariate Gaussian random numbers whose variance and covariance are provided by

$$\left[u_\mu^a u_\nu^b \right]_{\boldsymbol{\Phi}, \{\boldsymbol{x}^a\}} = \delta_{\mu\nu} q_{ab}, \tag{A.7}$$

when $\boldsymbol{\Phi}$ and $\{\boldsymbol{x}^a\}$ are generated independently from $P(\boldsymbol{\Phi})$ and $P(\{\boldsymbol{x}^a\})$, respectively. This means that (A.6) can be evaluated as

$$\begin{aligned} \Xi(\boldsymbol{Q}) &= \left(\frac{\int d\boldsymbol{u} \exp\left(-\frac{1}{2}\boldsymbol{u}^{\mathrm{T}} \boldsymbol{Q}^{-1} \boldsymbol{u}\right) \prod_{a=1}^{n} \Theta\left(u^0 u^a\right)}{(2\pi)^{(n+1)/2} (\det \boldsymbol{Q})^{1/2}} \right)^M \\ &= \left(2 \int \frac{d\boldsymbol{u} \exp\left(-\frac{1}{2}\boldsymbol{u}^{\mathrm{T}} \boldsymbol{Q}^{-1} \boldsymbol{u}\right) \Theta\left(u^0\right) \prod_{a=1}^{n} \Theta\left(u^a\right)}{(2\pi)^{(n+1)/2} (\det \boldsymbol{Q})^{1/2}} \right)^M. \end{aligned} \tag{A.8}$$

On the other hand, expressions

$$\delta\left(|\boldsymbol{x}^a|^2 - N\right) = \frac{1}{4\pi} \int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty} d\hat{q}_{aa} \exp\left(-\frac{1}{2} \hat{q}_{aa} \left(|\boldsymbol{x}^a|^2 - N \right) \right) \tag{A.9}$$

and

$$\delta\left(\boldsymbol{x}^a \cdot \boldsymbol{x}^b - N q_{ab}\right) = \frac{1}{2\pi} \int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty} d\hat{q}_{ab} \exp\left(\hat{q}_{ab} \left(\boldsymbol{x}^a \cdot \boldsymbol{x}^b - N q_{ab} \right) \right), \tag{A.10}$$

and use of the saddle point method offer

$$\begin{aligned} \frac{1}{N} \ln V(\boldsymbol{Q}) = \operatorname*{extr}_{\hat{\boldsymbol{Q}}} \Bigg\{ & -\frac{1}{2} \mathrm{Tr} \hat{\boldsymbol{Q}} \boldsymbol{Q} \\ & + \ln \left(\int d\boldsymbol{x} P(x^0) \exp\left(\frac{1}{2} \boldsymbol{x}^{\mathrm{T}} \hat{\boldsymbol{Q}} \boldsymbol{x} - \beta \sum_{a=1}^{n} \beta |x^a| \right) \right) \Bigg\} \end{aligned} \tag{A.11}$$

Here $\boldsymbol{x} = (x^0, x^1, \ldots, x^n)^{\mathrm{T}}$ and $\hat{\boldsymbol{Q}}$ is an $(n+1) \times (n+1)$ symmetric matrix whose 00 and other diagonal components are given as 0 and $-\hat{q}_{aa}$, respectively, while off-diagonal entries are offered as $\hat{q}_{ab}$. Equations (A.8) and (A.11) indicate that $N^{-1} \ln \left[Z^n(\beta; \boldsymbol{\Phi}, \boldsymbol{x}^0) \right]_{\boldsymbol{\Phi}, \boldsymbol{x}^0}$ is correctly evaluated by using the saddle point method with respect to $\boldsymbol{Q}$ in the assessment of the right-hand side of (A.5) when N and M tend to infinity keeping $\alpha = M/N$ finite.

A.2 Treatment under the replica symmetric ansatz

Let us assume that the relevant saddle point in assessing (A.5) is of the form of (2.7) and, accordingly,

$$\hat{q}_{ab} = \hat{q}_{ba} = \begin{cases} 0, & (a = b = 0) \\ \hat{m}, & (a = 1, 2, \ldots, n; b = 0) \\ \hat{Q}, & (a = b = 1, 2, \ldots, n) \\ \hat{q}, & (a \neq b = 1, 2, \ldots, n) \end{cases} . \tag{A.12}$$

$n+1$ dimensional Gaussian random variables $u^0, u^1, \ldots u^n$ whose variance and covariance are provided as (2.7) can be expressed as

$$u^0 = \sqrt{\rho - \frac{m^2}{q}} s^0 + \frac{m}{\sqrt{q}} z, \tag{A.13}$$

$$u^a = \sqrt{1-q} s^a + \sqrt{q} z, \ (a = 1, 2, \ldots, n) \tag{A.14}$$

utilizing $n+2$ independent standard Gaussian random variables z and $s^0, s^1, \ldots, s^n$. This indicates that (A.8) is evaluated as

$$\Xi(\boldsymbol{Q}) = \left(2 \int \mathrm{D}z \mathcal{Q}\left(\frac{m}{\sqrt{\rho q - m^2}} z \right) \mathcal{Q}^n \left(\sqrt{\frac{q}{1-q}} z \right) \right)^M . \tag{A.15}$$

On the other hand, substituting (A.12) into (A.11), in conjunction with the identity

$$\exp\left(\hat{q} \sum_{a>b(\geq 1)} x^a x^b \right) = \int \mathrm{D}z \exp\left(\sum_{a=1}^n \left(-\frac{\hat{q}}{2} (x^a)^2 + \sqrt{\hat{q}} z x^a \right) \right) \tag{A.16}$$

provides

$$\frac{1}{N} \ln V(\boldsymbol{Q}) = \underset{\hat{Q}, \hat{q}, \hat{m}}{\mathrm{extr}} \left\{ \frac{n}{2} \hat{Q} - \frac{n(n-1)}{2} \hat{q} q - \hat{m} m \right.$$
$$\left. + \ln \left[\left(\int dx \exp\left(-\frac{\hat{Q} + \hat{q}}{2} x^2 + \left(\sqrt{\hat{q}} z + \hat{m} x^0 \right) x - \beta |x| \right) \right)^n \right]_{x^0, z} \right\} \tag{A.17}$$

Although we have assumed that $n \in \mathbb{N}$, the expressions of (A.15) and (A.17) are likely to hold for $n \in \mathbb{R}$ as well. Therefore the average free energy $\overline{f}$ can be evaluated by substituting these expressions into the formula $\overline{f} = -\lim_{n \to 0} (\partial / \partial n) \left((\beta N)^{-1} \ln \left[Z^n(\beta; \boldsymbol{\Phi}, \boldsymbol{x}^0) \right]_{\boldsymbol{\Phi}, \boldsymbol{x}^0} \right)$.

In the limit of $\beta \to \infty$, a nontrivial saddle point is obtained only when $\chi \equiv \beta(1-q)$ is kept finite. Accordingly, we change the notations of the auxiliary variables as $\hat{Q} + \hat{q} \to \beta \hat{Q}$, $\hat{q} \to \beta^2 \hat{q}$, and $\hat{m} \to \beta \hat{m}$. Furthermore, we use the asymptotic forms

$$\lim_{\beta \to \infty} \frac{1}{\beta} \int \mathrm{D}z \mathcal{Q}\left(\frac{m}{\sqrt{\rho q - m^2}} z \right) \ln \mathcal{Q}\left(\sqrt{\frac{q}{1-q}} z \right)$$

$$
\begin{aligned}
&= \int \mathrm{D}z\mathcal{Q}\left(\frac{m}{\sqrt{\rho-m^2}}z\right)\left(-\frac{z^2}{2\chi}\Theta(z)\right) \\
&= -\frac{1}{4\pi\chi}\left(\arctan\left(\frac{\sqrt{\rho-m^2}}{m}\right)-\frac{m}{\rho}\sqrt{\rho-m^2}\right)
\end{aligned} \tag{A.18}
$$

and

$$
\begin{aligned}
&\lim_{\beta\to\infty}\frac{1}{\beta}\ln\left(\int dx\exp\left(\beta\left(-\frac{\hat{Q}}{2}x^2+\left(\sqrt{\hat{q}}z+\hat{m}x^0\right)x-|x|\right)\right)\right) \\
&= -\phi\left(\sqrt{\hat{q}}z+\hat{m}x^0;\hat{Q}\right).
\end{aligned} \tag{A.19}
$$

Using these in the resultant expression of $\overline{f}$ offers (2.8).

Appendix B

Stability of the RS solution for l_1 approach

The 1-step replica symmetry breaking (1RSB) ansatz means that, at the relevant saddle point, n replica indices $1, 2, \ldots, n$ are classified into n/p groups of an equal size p, and $q_{ab} = q_1$ holds if a and b belong to an identical group and $q_0 (\leq q_1)$, otherwise. This yields the following expression of the average free energy of finite temperature:

$$\begin{aligned}\overline{f} = \underset{\omega}{\mathrm{extr}} \Bigg\{ & -\frac{1}{\beta} \left[\ln \left(\int \mathrm{D}t \exp\left(-p\mathcal{Y}_0\right) \right) \right]_{x^0,z} \\ & -\frac{1}{2\beta}(\hat{Q} + \hat{q}_1) + \frac{\hat{q}_1}{2\beta}(1 - q_1) + \frac{p}{2\beta}(\hat{q}_1 q_1 - \hat{q}_0 q_0) + \frac{1}{\beta}\hat{m}m \\ & -\frac{2\alpha}{\beta p} \int \mathrm{D}z \mathcal{Q}\left(\frac{m}{\sqrt{\rho q - m^2}} z \right) \ln \left(\int \mathrm{D}t \exp\left(-p\mathcal{Y}_1\right) \right) \Bigg\}, \end{aligned} \tag{B.1}$$

where $\mathcal{Y}_0 \equiv -\ln\left(\int dx \exp\left(-(\hat{Q} + \hat{q}_1)x^2/2 + (\sqrt{\hat{q}_1 - \hat{q}_0}t + \sqrt{\hat{q}_0}z + \hat{m}x^0)x - \beta|x| \right) \right)$, $\mathcal{Y}_1 \equiv -\ln\left(\int \mathrm{D}x \Theta\left(-\left(\sqrt{1 - q_1}x + \sqrt{q_1 - q_0}t + \sqrt{q_0}z \right) \right) \right)$, $\omega = \{q_1, q_0, m, \hat{Q}, \hat{q}_1, \hat{q}_0, \hat{m}\}$, and $[\cdots]_{x^0,z} = \int dx^0 P(x^0) \int \mathrm{D}z \left(\cdots\right)$. The RS solution is regarded as a special case of the 1RSB solution for which $q_1 = q_0$ holds. Therefore one can check the thermodynamical validity of the RS solution by examining the stability of the solution of $q_1 = q_0$ under the 1RSB ansatz.

The extremization condition of (B.1) indicates that

$$\begin{aligned} q_1 - q_0 = & \left[\frac{\int \mathrm{D}t e^{-p\mathcal{Y}_0} \left(\partial \mathcal{Y}_0 / \partial(\sqrt{\hat{q}_0}z) \right)^2}{\int \mathrm{D}t e^{-p\mathcal{Y}_0}} \right. \\ & \left. - \left(\frac{\int \mathrm{D}t e^{-p\mathcal{Y}_0} \left(\partial \mathcal{Y}_0 / \partial(\sqrt{\hat{q}_0}z) \right)}{\int \mathrm{D}t e^{-p\mathcal{Y}_0}} \right)^2 \right]_{x^0,z} \\ \simeq & \left[\left(\frac{\partial^2 \mathcal{Y}_0^{\mathrm{RS}}}{\partial(\sqrt{\hat{q}_0}z)^2} \right)^2 \left(\frac{\int \mathrm{D}t e^{-p\mathcal{Y}_0} t^2}{\int \mathrm{D}t e^{-p\mathcal{Y}_0}} - \left(\frac{\int \mathrm{D}t e^{-p\mathcal{Y}_0} t}{\int \mathrm{D}t e^{-p\mathcal{Y}_0}} \right)^2 \right) \right]_{x^0,z} (\hat{q}_1 - \hat{q}_0) \\ \simeq & \left[\left(\frac{\partial^2 \mathcal{Y}_0^{\mathrm{RS}}}{\partial(\sqrt{\hat{q}_0}z)^2} \right)^2 \right]_{x^0,z} (\hat{q}_1 - \hat{q}_0) \end{aligned} \tag{B.2}$$

and

$$\hat{q}_1 - \hat{q}_0 = 2\alpha \int \mathrm{D}z \mathcal{Q}\left(\frac{m}{\sqrt{\rho q - m^2}} z \right) \left(\frac{\int \mathrm{D}t e^{-p\mathcal{Y}_1} \left(\partial \mathcal{Y}_1 / \partial(\sqrt{q_0}z) \right)^2}{\int \mathrm{D}t e^{-p\mathcal{Y}_1}} \right.$$

$$
\begin{aligned}
&\quad - \left(\frac{\int \mathrm{D}t e^{-p\mathcal{Y}_1} \left(\partial \mathcal{Y}_1 / \partial (\sqrt{q_0} z) \right)}{\int \mathrm{D}t e^{-p\mathcal{Y}_1}} \right)^2 \Bigg) \\
&\simeq 2\alpha \int \mathrm{D}z \mathcal{Q} \left(\frac{m}{\sqrt{\rho q - m^2}} z \right) \left(\frac{\partial^2 \mathcal{Y}_1^{\mathrm{RS}}}{\partial (\sqrt{q_0} z)^2} \right)^2 \\
&\qquad \times \left(\frac{\int \mathrm{D}t e^{-p\mathcal{Y}_1} t^2}{\int \mathrm{D}t e^{-p\mathcal{Y}_1}} - \left(\frac{\int \mathrm{D}t e^{-p\mathcal{Y}_1} t}{\int \mathrm{D}t e^{-p\mathcal{Y}_1}} \right)^2 \right) (q_1 - q_0) \\
&\simeq 2\alpha \int \mathrm{D}z \mathcal{Q} \left(\frac{m}{\sqrt{\rho q - m^2}} z \right) \left(\frac{\partial^2 \mathcal{Y}_1^{\mathrm{RS}}}{\partial (\sqrt{q_0} z)^2} \right)^2 (q_1 - q_0) \qquad \text{(B.3)}
\end{aligned}
$$

hold for $|q_1 - q_0| \ll 1$ and $|\hat{q}_1 - \hat{q}_0| \ll 1$ irrespectively of the value of p. Here $\mathcal{Y}_0^{\mathrm{RS}}$ and $\mathcal{Y}_1^{\mathrm{RS}}$ represent assessments of $\mathcal{Y}_0$ and $\mathcal{Y}_1$ under the assumptions of $\hat{q}_1 = \hat{q}_0$ and $q_1 = q_0$, respectively. In (B.2) and (B.3) we used the Taylor expansion expressions

$$
\partial \mathcal{Y}_0 / \partial (\sqrt{\hat{q}_0} z) \sim \partial \mathcal{Y}_0^{\mathrm{RS}} / \partial (\sqrt{\hat{q}_0} z) + \partial^2 \mathcal{Y}_0^{\mathrm{RS}} / \partial (\sqrt{\hat{q}_0} z)^2 \sqrt{\hat{q}_1 - \hat{q}_0} t \qquad \text{(B.4)}
$$

and

$$
\partial \mathcal{Y}_1 / \partial (\sqrt{q_0} z) \sim \partial \mathcal{Y}_1^{\mathrm{RS}} / \partial (\sqrt{q_0} z) + \partial^2 \mathcal{Y}_1^{\mathrm{RS}} / \partial (\sqrt{q_0} z)^2 \sqrt{q_1 - q_0} t, \qquad \text{(B.5)}
$$

and the fact that the variances of t for the measures $\mathrm{D}t e^{-p\mathcal{Y}_0} / \int \mathrm{D}t e^{-p\mathcal{Y}_0}$ and $\mathrm{D}t e^{-p\mathcal{Y}_1} / \int \mathrm{D}t e^{-p\mathcal{Y}_1}$ become unity as $\hat{q}_1 - \hat{q}_0$ and $q_1 - q_0$ vanish, irrespectively of the value of p.

To examine the stability of the RS solution in the limit of $\beta \to \infty$, let us change the variable notations as $\chi = \beta(1-q)$, $\hat{Q} + \hat{q}_1 \to \beta \hat{Q}$, $\hat{q}_1 \to \beta^2 \hat{q}_1$, $\hat{q}_0 \to \beta^2 \hat{q}_0$, and $\hat{m} \to \beta \hat{m}$ and set $q_0 = q$ and $\hat{q}_0 = \hat{q}$. This yields expressions of $\mathcal{Y}_0^{\mathrm{RS}} \simeq \beta \phi(\sqrt{\hat{q}} z + \hat{m} x^0; \hat{Q}) = -\beta g(\sqrt{\hat{q}} z + \hat{m} x^0) / \hat{Q}$ and $\mathcal{Y}_1^{\mathrm{RS}} \simeq (\beta / \chi) f(-\sqrt{q} z)$ for $\beta \gg 1$. Substituting these into (B.2) and (B.3) leads to

$$
\Delta \simeq \frac{1}{\hat{Q}^2} \left[\left(g''(\sqrt{\hat{q}} z + \hat{m} x^0) \right)^2 \right]_{x^0, z} \hat{\Delta} \qquad \text{(B.6)}
$$

and

$$
\hat{\Delta} \simeq \frac{2\alpha}{\chi^2} \int \mathrm{D}z \mathcal{Q} \left(\frac{m}{\sqrt{\rho - m^2}} z \right) \left(f''(-z) \right)^2 \Delta, \qquad \text{(B.7)}
$$

where we set $\Delta = q_1 - q$ and $\hat{\Delta} = \hat{q}_1 - \hat{q}$, and used $q \to 1$. The condition that (B.6) and (B.7) allow a solution of $(\Delta, \hat{\Delta}) \neq (0, 0)$ offers (2.16).

Appendix C

Derivation of the cavity equations

We refer to the system in which x_i and (a_μ, z_μ) are kept out as the i-cavity and μ-cavity systems, respectively. In addition, we denote $\mathcal{L}_{i\to\mu}(x_i)$, $A_{i\to\mu}$ and $H_{i\to\mu}$ as the single-body cost function for the μ-cavity system and its parameters, respectively, and similarly for $\mathcal{L}_{\mu\to i}(a_\mu, z_\mu)$, $B_{\mu\to i}$ and $K_{\mu\to i}$. Self-consistent equations are derived from the following arguments.

Vertical step:
Let us suppose that x_i is put into the i-cavity system, which yields an approximation of the cost function of (2.17) as

$$(\Lambda/2)x_i^2 + |x_i| + \sum_{\nu=1}^{M} \left(\mathcal{L}_{\nu\to i}(a_\nu, z_\nu) + \Phi_{\nu i} a_\nu x_i\right). \tag{C.1}$$

From this function we remove all terms that are related to (a_μ, z_μ) of a certain index $\mu \in \{1, 2, \ldots, M\}$, which leads to an approximate cost function of the μ-cavity system. $\mathcal{L}_{i\to\mu}(x_i)$ must be obtained by partially optimizing the resulting μ-cavity cost function with respect to (a_ν, z_ν) of the remaining indices $\forall \nu \in \{1, 2, \ldots, M\}\backslash\mu$, where $S\backslash a$ generally denotes the set provided by removing an element a from a set S. This offers the relation

$$\mathcal{L}_{i\to\mu}(x_i) = \frac{\Lambda}{2}x_i^2 + |x_i| + \sum_{\nu\neq\mu} \left\{ \min_{z_\nu > 0} \max_{a_\nu} \left\{\mathcal{L}_{\nu\to i}(a_\nu, z_\nu) + \Phi_{\nu i} a_\nu x_i\right\} \right\}. \tag{C.2}$$

This relation and the fact that $\Phi_{\mu i}$ is a negligibly small independent sample from an identical Gaussian distribution with zero mean and variance N^{-1} yield the following equations evaluating $A_{i\to\mu}$ and $H_{i\to\mu}$ from a set of $\{B_{\nu\to i}\}$ and $\{K_{\nu\to i}\}$:

$$A_{i\to\mu} = \Lambda + \sum_{\nu\neq\mu} \frac{\Phi_{\nu i}^2}{B_{\nu\to i}} f''(K_{\nu\to i}), \tag{C.3}$$

$$H_{i\to\mu} = -\sum_{\nu\neq\mu} \frac{\Phi_{\nu i}}{B_{\nu\to i}} f'(K_{\nu\to i}). \tag{C.4}$$

Horizontal step:
Similarly, putting (a_μ, z_μ) into the μ-cavity system and removing x_i yields

another relation,

$$\mathcal{L}_{\mu\to i}(a_\mu, z_\mu) = -z_\mu a_\mu + \sum_{j\neq i}\left\{\min_{x_j}\left\{\mathcal{L}_{j\to\mu}(x_j) + \Phi_{\mu j}a_\mu x_j\right\}\right\}, \tag{C.5}$$

which offers

$$B_{\mu\to i} = \sum_{j\neq i}\frac{\Phi_{\mu j}^2}{A_{j\to\mu}}g''(H_{j\to\mu}), \tag{C.6}$$

$$K_{\mu\to i} = \sum_{j\neq i}\frac{\Phi_{\mu j}}{A_{j\to\mu}}g'(H_{j\to\mu}). \tag{C.7}$$

Recovery step:
A_i and H_i are evaluated from (C.6) and (C.7) as

$$A_i = \Lambda + \sum_{\mu=1}^{M}\frac{\Phi_{\mu i}^2}{B_{\mu\to i}}f''(K_{\mu\to i}), \tag{C.8}$$

$$H_i = -\sum_{\mu=1}^{M}\frac{\Phi_{\mu i}}{B_{\mu\to i}}f'(K_{\mu\to i}). \tag{C.9}$$

This means that the recovered signal is provided as

$$\hat{x}_i = \frac{1}{A_i}g'(H_i), \tag{C.10}$$

where Λ is determined in such a way that $\sum_{i=1}^{N}\hat{x}_i^2 = N$ holds. Similarly,

$$B_\mu = \sum_{i=1}^{N}\frac{\Phi_{\mu i}^2}{A_{i\to\mu}}g''(H_{i\to\mu}), \tag{C.11}$$

$$K_\mu = \sum_{i=1}^{N}\frac{\Phi_{\mu i}}{A_{i\to\mu}}g'(H_{i\to\mu}), \tag{C.12}$$

are obtained from (C.3) and (C.4). These offer the (approximate) optimal value of the Lagrange multiplier a_μ as

$$\hat{a}_\mu = -\frac{1}{B_\mu}f'(K_\mu). \tag{C.13}$$

Equations (C.4) and (C.9) indicate the difference between
$H_{i\to\mu}$ and H_i
is vanishingly small for $N\to\infty$ as $\Phi_{\mu i}$ scales as $O\left(N^{-1/2}\right)$, and similarly for
$K_{\mu\to i}$ and K_μ.
This also allows us to handle A_i and $A_{i\to\mu}$ as a single site-independent

parameter A, and we similarly deal with B_μ and $B_{\mu\to i}$ as B. These considerations, in conjunction with (C.8) and (C.11), offer

$$A = \Lambda + \frac{1}{NB}\sum_{\mu=1}^{M} f''(K_\mu), \tag{C.14}$$

$$B = \frac{1}{NA}\sum_{i=1}^{N} g''(H_i), \tag{C.15}$$

where we replaced $\Phi_{\mu i}^2$ in (C.8) and (C.11) with its expectation N^{-1} by utilizing the law of large numbers. Furthermore, inserting $f'(K_{\mu\to i}) \simeq f'(K_\mu - \Phi_{\mu i}\hat{x}_i) \simeq f'(K_\mu) - \Phi_{\mu i} f''(K_\mu)\hat{x}_i$ and $g'(H_{i\to\mu}) \simeq g'(H_i + \Phi_{\mu i}\hat{a}_\mu) \simeq g'(H_i) + \Phi_{\mu i} g''(H_i)\hat{a}_\mu$ into (C.9) and (C.12), respectively, yields

$$\begin{aligned} H_i &\simeq \sum_{\mu=1}^{M}\Phi_{\mu i}\hat{a}_\mu + \left(\frac{1}{B}\sum_{\mu=1}^{M}\Phi_{\mu i}^2 f''(K_\mu)\right)\hat{x}_i \\ &\simeq \sum_{\mu=1}^{M}\Phi_{\mu i}\hat{a}_\mu + \left(\frac{1}{NB}\sum_{\mu=1}^{M} f''(K_\mu)\right)\hat{x}_i \\ &= \sum_{\mu=1}^{M}\Phi_{\mu i}\hat{a}_\mu + \Gamma\hat{x}_i \end{aligned} \tag{C.16}$$

and

$$\begin{aligned} K_\mu &\simeq \sum_{i=1}^{N}\Phi_{\mu i}\hat{x}_i - \left(\frac{1}{A}\sum_{i=1}^{N}\Phi_{\mu i}^2 g''(H_i)\right)\hat{a}_\mu \\ &\simeq \sum_{i=1}^{N}\Phi_{\mu i}\hat{x}_i - \left(\frac{1}{NA}\sum_{i=1}^{N} g''(H_i)\right)\hat{a}_\mu \\ &= \sum_{i=1}^{N}\Phi_{\mu i}\hat{x}_i - B\hat{a}_\mu, \end{aligned} \tag{C.17}$$

where we set $\Gamma = (NB)^{-1}\sum_{\mu=1}^{M} f''(K_\mu)$. Equations (C.10), (C.13), and (C.14)–(C.17) lead to (2.20)–(2.23).

Appendix D

Derivation of (3.11)

D.0.1 Assessment of $\left[Z^n\left(\beta;\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}\right)\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}}$ for $n\in\mathbb{N}$

Averaging (3.9) with respect to $\boldsymbol{\Phi}$ and $\boldsymbol{x}^0$ offers the following expression of the n-th moment of the partition function:

$$\left[Z^n\left(\beta;\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}\right)\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}}=\int\prod_{a=1}^{n}\left(d\boldsymbol{x}^a e^{-\beta||\boldsymbol{x}^a||_1}\right)\times\left[\prod_{a=1}^{n}\prod_{\mu=1}^{M}\Theta\left((\boldsymbol{\Phi}\boldsymbol{x}^0+\boldsymbol{\lambda})_\mu(\boldsymbol{\Phi}\boldsymbol{x}^a+\boldsymbol{\lambda})_\mu\right)\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}}. \tag{D.1}$$

We insert $n(n+1)/2$ trivial identities

$$1=N\int dq_{ab}\delta\left(\boldsymbol{x}^a\cdot\boldsymbol{x}^b-Nq_{ab}\right), \tag{D.2}$$

where $a>b=0,1,2,\ldots,n$, into (D.1). Furthermore, we define a joint distribution of $n+1$ vectors $\{\boldsymbol{x}^a\}=\{\boldsymbol{x}^0,\boldsymbol{x}^1,\boldsymbol{x}^2,\ldots,\boldsymbol{x}^n\}$ as

$$\begin{aligned}P\left(\{\boldsymbol{x}^a\}|\boldsymbol{Q}\right) &= \frac{1}{V\left(\boldsymbol{Q}\right)}P(\boldsymbol{x}^0)\times\prod_{a=1}^{n}\left(e^{-\beta||\boldsymbol{x}^a||_1}\right)\\ &\quad\times\prod_{a>b}\delta\left(\boldsymbol{x}^a\cdot\boldsymbol{x}^b-Nq_{ab}\right),\end{aligned} \tag{D.3}$$

where $\boldsymbol{Q}=(q_{ab})$ is an $(n+1)\times(n+1)$ symmetric matrix whose 00 and the other diagonal entries are fixed as ρ and q_{aa}, respectively. $P(\boldsymbol{x}^0)=\prod_{i=1}^{N}\left((1-\rho)\delta(x_i^0)+\rho\tilde{P}(x_i^0)\right)$ denotes the distribution of the original signal $\boldsymbol{x}^0$, and $V\left(\boldsymbol{Q}\right)$ is the normalization constant that makes

$$\int\prod_{a=0}^{n}d\boldsymbol{x}^a P\left(\{\boldsymbol{x}^a\}|\boldsymbol{Q}\right)=1 \tag{D.4}$$

hold. These indicate that (D.1) can also be expressed as

$$\left[Z^n\left(\beta;\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}\right)\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}}=\int d\boldsymbol{Q}\left(V\left(\boldsymbol{Q}\right)\times\left[\Xi\left(\boldsymbol{Q}\right)\right]_{\boldsymbol{\lambda}}\right), \tag{D.5}$$

where $d\boldsymbol{Q}\equiv\prod_{a>b}dq_{ab}$ and

$$\Xi\left(\boldsymbol{Q}\right)=\int\prod_{a=0}^{n}d\boldsymbol{x}^a P\left(\{\boldsymbol{x}^a\}|\boldsymbol{Q}\right)$$

$$\times \left[\prod_{a=1}^{n} \prod_{\mu=1}^{M} \Theta\left((\boldsymbol{\Phi}\boldsymbol{x}^0 + \boldsymbol{\lambda})_\mu (\boldsymbol{\Phi}\boldsymbol{x}^a + \boldsymbol{\lambda})_\mu \right) \right]_{\boldsymbol{\Phi}}. \tag{D.6}$$

Equation (D.6) can be regarded as the average of $\prod_{a=1}^{n}\prod_{\mu=1}^{M}\Theta\left((\boldsymbol{\Phi}\boldsymbol{x}^0+\boldsymbol{\lambda})_\mu(\boldsymbol{\Phi}\boldsymbol{x}^a+\boldsymbol{\lambda})_\mu\right)$ with respect to $\{\boldsymbol{x}^a\}$ and $\boldsymbol{\Phi}$ over distributions of $P\left(\{\boldsymbol{x}^a\}\right)$ and $P(\boldsymbol{\Phi}) \equiv \left(\sqrt{2\pi/N}\right)^{-MN} \exp\left(-(N/2)\sum_{\mu,i}\Phi_{\mu i}^2\right)$. In computing this, it is noteworthy that when N and M tend to infinity while keeping $\alpha = \frac{M}{N}$ finite, the Central Limit Theorem guarantees that $u_\mu^a \equiv (\boldsymbol{\Phi}\boldsymbol{x}^a)_\mu = \sum_{i=1}^{N}\Phi_{\mu i}x_i^a$ can be handled as zero-mean multivariate Gaussian random numbers whose variance and covariance are provided by

$$\left[u_\mu^a u_\nu^b\right]_{\boldsymbol{\Phi},\{\boldsymbol{x}^a\}} = \delta_{\mu\nu}q_{ab}, \tag{D.7}$$

when $\boldsymbol{\Phi}$ and $\{\boldsymbol{x}^a\}$ are generated independently from $P(\boldsymbol{\Phi})$ and $P\left(\{\boldsymbol{x}^a\}\right)$, respectively. This means that (D.6) can be evaluated as

$$\Xi(\boldsymbol{Q}) = \left(\frac{\int d\boldsymbol{u} \exp\left(-\frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{u}\right) \prod_{a=1}^{n} \Theta\left((u^0+\lambda)(u^a+\lambda)\right)}{(2\pi)^{(n+1)/2}(\det \boldsymbol{Q})^{1/2}} \right)^M, \tag{D.8}$$

where u^0, u^a, and λ represent the typical elements of $\boldsymbol{u^0}$, $\boldsymbol{u^a}$ and $\boldsymbol{\lambda}$, respectively, since each μ is independently distributed.

On the other hand, expression

$$\delta\left(\boldsymbol{x}^a \cdot \boldsymbol{x}^b - N q_{ab}\right) = \frac{1}{2\pi}\int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty} d\hat{q}_{ab} e^{\hat{q}_{ab}\left(\boldsymbol{x}^a\cdot\boldsymbol{x}^b - N q_{ab}\right)}, \tag{D.9}$$

and use of the saddle point method offer

$$\begin{aligned} \frac{1}{N}\ln V(\boldsymbol{Q}) = \operatorname*{extr}_{\hat{\boldsymbol{Q}}} \Bigg\{ & -\frac{1}{2}\mathrm{Tr}\hat{\boldsymbol{Q}}\boldsymbol{Q} \\ & + \ln\left(\int d\boldsymbol{x} P(x^0) \exp\left(\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\hat{\boldsymbol{Q}}\boldsymbol{x} - \beta\sum_{a=1}^{n}\beta|x^a| \right) \right) \Bigg\}. \end{aligned} \tag{D.10}$$

Here, $\boldsymbol{x} = (x^0, x^1, \ldots, x^n)^{\mathrm{T}}$, and x^a represents the typical element of $\boldsymbol{x}^a$, since each x_i^a is independently distributed. $\hat{\boldsymbol{Q}}$ is an $(n+1)\times(n+1)$ symmetric matrix whose 00 and other diagonal components are given as 0 and $-\hat{q}_{aa}$, respectively, while off-diagonal entries are offered as $\hat{q}_{ab}$. Equations (D.8) and (D.10) indicate that $N^{-1}\ln\left[Z^n(\beta;\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda})\right]_{\boldsymbol{\Phi},\boldsymbol{x}^0,\boldsymbol{\lambda}}$ is correctly evaluated by using the saddle point method with respect to $\boldsymbol{Q}$ in the assessment of the right-hand side of (D.5), when N and M tend to infinity while keeping $\alpha = M/N$ finite.

D.0.2 Treatment under the replica symmetric ansatz

Let us assume that the relevant saddle point in assessing (D.5) is of the form of (3.10) and, accordingly,

$$\hat{q}_{ab} = \hat{q}_{ba} = \begin{cases} 0, & (a = b = 0) \\ \hat{m}, & (a = 1, 2, \ldots, n; b = 0) \\ \hat{Q}, & (a = b = 1, 2, \ldots, n) \\ \hat{q}, & (a \neq b = 1, 2, \ldots, n) \end{cases} . \tag{D.11}$$

The $n+1$-dimensional Gaussian random variables $u^0, u^1, \ldots, u^n$, whose variance and covariance are provided as (3.10), can be expressed as

$$u^0 = \sqrt{\rho\sigma_0^2 - \frac{m^2}{q}}s^0 + \frac{m}{\sqrt{q}}t, \tag{D.12}$$

$$u^a = \sqrt{Q-q}s^a + \sqrt{q}t, \ (a = 1, 2, \ldots, n) \tag{D.13}$$

by utilizing $n+2$ independent standard Gaussian random variables t and $s^0, s^1, \ldots, s^n$. This indicates that (D.8) is evaluated as

$$\Xi(\boldsymbol{Q}) = \left(\int \mathrm{D}t \mathcal{Q}\left(-\frac{\frac{m}{\sqrt{q}}t + \lambda}{\sqrt{\rho\sigma_0^2 - \frac{m^2}{q}}} \right) \mathcal{Q}^n \left(-\frac{\sqrt{q}t + \lambda}{\sqrt{Q-q}} \right) \right.$$
$$\left. + \mathcal{Q}\left(\frac{\frac{m}{\sqrt{q}}t + \lambda}{\sqrt{\rho\sigma_0^2 - \frac{m^2}{q}}} \right) \mathcal{Q}^n \left(\frac{\sqrt{q}t + \lambda}{\sqrt{Q-q}} \right) \right)^M . \tag{D.14}$$

On the other hand, substituting (D.11) into (D.10), in conjunction with the identity,

$$\exp\left(\hat{q} \sum_{a>b(\geq 1)} x^a x^b \right)$$
$$= \int \mathrm{D}z \exp\left(\sum_{a=1}^n \left(-\frac{\hat{q}}{2}(x^a)^2 + \sqrt{\hat{q}} z x^a \right) \right) \tag{D.15}$$

where z is a standard Gaussian random variable, yields

$$\frac{1}{N}\ln V(\boldsymbol{Q}) = \underset{\hat{Q},\hat{q},\hat{m}}{\mathrm{extr}} \left\{ \frac{n}{2}\hat{Q}Q - \frac{n(n-1)}{2}\hat{q}q - \hat{m}m\sigma_0^2 \right.$$
$$+ \ln \left[\left(\int dx \exp\left(-\frac{\hat{Q}+\hat{q}}{2}x^2 + \left(\sqrt{\hat{q}}z + \hat{m}x^0 \right) x \right. \right. \right.$$
$$\left. \left. \left. -\beta|x|) \right)^n \right]_{x^0,z} \right\}. \tag{D.16}$$

Although we have assumed that $n \in \mathbb{N}$, the expressions of (D.14) and (D.16) are likely to hold for $n \in \mathbb{R}$ as well. Therefore the average free energy $\overline{f}$ can be evaluated by substituting these expressions into the formula $\overline{f} = -\lim_{n\to 0}(\partial/\partial n)\left((\beta N)^{-1} \ln \left[Z^n(\beta; \boldsymbol{\Phi}, \boldsymbol{x}^0, \boldsymbol{\lambda}) \right]_{\boldsymbol{\Phi},\boldsymbol{x}^0}, \boldsymbol{\lambda} \right)$.

In the limit of $\beta \to \infty$, a nontrivial saddle point is obtained only when $\chi \equiv \beta(Q-q)$ is kept finite. Accordingly, we change the notations of the

auxiliary variables as $\hat{Q}+\hat{q}\to\beta\hat{Q}$, $\hat{q}\to\beta^2\hat{q}$, and $\hat{m}\to\beta\hat{m}$. Furthermore, we use the asymptotic forms

$$\begin{aligned}&\lim_{\beta\to\infty}\frac{1}{\beta}\int \mathrm{D}t\mathcal{Q}\left(\frac{\frac{m}{\sqrt{q}}t+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)\ln\mathcal{Q}\left(\frac{\sqrt{q}t+\lambda}{\sqrt{Q-q}}\right)\\&=\int \mathrm{D}t\mathcal{Q}\left(\frac{\frac{m}{\sqrt{q}}t+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)\left(-\frac{(\sqrt{q}t+\lambda)^2}{2\chi}\Theta(\sqrt{q}t+\lambda)\right)\end{aligned}\tag{D.17}$$

and

$$\begin{aligned}&\lim_{\beta\to\infty}\frac{1}{\beta}\ln\left(\int dx\exp\left(\beta\left(-\frac{\hat{Q}}{2}x^2+\left(\sqrt{\hat{q}}z+\hat{m}x^0\right)x-|x|\right)\right)\right)\\&=-\phi\left(\sqrt{\hat{q}}z+\hat{m}x^0;\hat{Q}\right).\end{aligned}\tag{D.18}$$

Using these in the resultant expression of $\overline{f}$ offers (3.11).

Appendix E

RS stability of thresholding 1-bit compressive sensing

To examine the validity of the RS ansatz, we also evaluated the local stability of the RS solutions against the disturbances that break the replica symmetry [2], which offers

$$\begin{aligned}
&\frac{\alpha}{\hat{Q}^2\chi^2}\left[\mathcal{Q}\left(-\frac{\frac{mt}{\sqrt{q}}+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)u''\left(-\sqrt{q}t-\lambda\right)\right.\\
&\left.+\mathcal{Q}\left(\frac{\frac{mt}{\sqrt{q}}+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)u''\left(\sqrt{q}t+\lambda\right)\right]_{t,\lambda}\\
&\times 2\left((1-\rho)\mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}}}\right)+\rho\mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}+\hat{m}^2\sigma_0^2}}\right)\right)\\
&-1<0,
\end{aligned} \tag{E.1}$$

as the stability condition. Unfortunately, the left handside of equation (E.1) is always zero, therefore it is not satisfied. This indicates that taking the replica symmetry breaking (RSB) into account is necessary for evaluating the exact performance of the signal recovery scheme defined by (3.2). Here we provid a brief sketch of the derivation of this condition by the 1-step replica symmetry breaking (1RSB) calculation.

1RSB ansatz means that, at the relevant saddle point, n replica indices $1, 2, \ldots, n$ are classified into n/p groups of an equal size p, and $q_{ab} = q_1$ holds if a and b belong to an identical group and $q_0 (\leq q_1)$, otherwise. This yields the following expression of the average free energy of finite temperature:

$$\begin{aligned}
\overline{f} = \underset{\omega}{\mathrm{extr}}\Bigg\{&-\frac{1}{\beta}\left[\ln\left(\int \mathrm{D}t\exp\left(-p\mathcal{Y}_0\right)\right)\right]_{x^0,z}\\
&-\frac{1}{2\beta}(\hat{Q}+\hat{q}_1)Q+\frac{\hat{q}_1}{2\beta}(Q-q_1)+\frac{p}{2\beta}(\hat{q}_1 q_1-\hat{q}_0 q_0)+\frac{1}{\beta}\hat{m}m\sigma_0^2\\
&-\frac{\alpha}{\beta p}\left[\int \mathrm{D}z\left(\mathcal{Q}\left(-\frac{\frac{m}{\sqrt{q}}z+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)\ln\left(\int \mathrm{D}z\exp\left(-p\mathcal{Y}_1\right)\right)\right.\right.\\
&\left.\left.+\mathcal{Q}\left(\frac{\frac{m}{\sqrt{q}}t+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)\ln\left(\int \mathrm{D}z\exp\left(-p\mathcal{Y}_2\right)\right)\right)\right]_{\lambda}\Bigg\},
\end{aligned} \tag{E.2}$$

where

$$\begin{aligned}
\mathcal{Y}_0 \equiv & -\ln\left(\int dx \exp\left(-(\hat{Q}+\hat{q}_1)x^2/2+(\sqrt{\hat{q}_1-\hat{q}_0}t+\right.\right.\\
& \left.\left.\sqrt{\hat{q}_0}z+\hat{m}x^0)x-\beta|x|\right)\right),\\
\mathcal{Y}_1 \equiv & -\ln\left(\int \mathrm{D}x\Theta\left(\sqrt{Q-q_1}x+\sqrt{q_1-q_0}t+\sqrt{q}_0z+\lambda\right)\right),\\
\mathcal{Y}_2 \equiv & -\ln\left(\int \mathrm{D}x\Theta\left(-\left(\sqrt{Q-q_1}x+\sqrt{q_1-q_0}t+\sqrt{q}_0z+\lambda\right)\right)\right),
\end{aligned}$$

$\omega=\{q_1,q_0,Q,m,\hat{Q},\hat{q}_1,\hat{q}_0,\hat{m}\}$, $[\cdots]_{x^0,z}=\int dx^0P(x^0)\int \mathrm{D}z\,(\cdots)$, and $[\cdots]_\lambda=\int d\lambda P(\lambda)\,(\cdots)$. The RS solution is regarded as a special case of the 1RSB solution for which $q_1=q_0$ holds. Therefore one can check the thermodynamical validity of the RS solution by examining the stability of the solution of $q_1=q_0$ under the 1RSB ansatz.

The extremization condition of (E.2) indicates that

$$\begin{aligned}
& q_1-q_0\\
&=\left[\left[\left(\partial\mathcal{Y}_0/\partial(\sqrt{\hat{q}_0}z)\right)^2\right]_{|\mathcal{Y}_0}-\left[\left(\partial\mathcal{Y}_0/\partial(\sqrt{\hat{q}_0}z)\right)\right]^2_{|\mathcal{Y}_0}\right]_{x^0,z}\\
&\simeq\left[\left(\frac{\partial^2\mathcal{Y}_0^{\mathrm{RS}}}{\partial(\sqrt{\hat{q}_0}z)^2}\right)^2\left(\left[t^2\right]_{|\mathcal{Y}_0}-[t]^2_{|\mathcal{Y}_0}\right)\right]_{x^0,z}(\hat{q}_1-\hat{q}_0)\\
&\simeq\left[\left(\frac{\partial^2\mathcal{Y}_0^{\mathrm{RS}}}{\partial(\sqrt{\hat{q}_0}z)^2}\right)^2\right]_{x^0,z}(\hat{q}_1-\hat{q}_0)
\end{aligned}\tag{E.3}$$

and

$$\begin{aligned}
& \hat{q}_1-\hat{q}_0\\
&=\alpha\int \mathrm{D}z\mathcal{Q}\left(-\frac{\frac{m}{\sqrt{q}}z+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)\\
&\quad\times\left(\left[\left(\frac{\partial\mathcal{Y}_1}{\partial(\sqrt{\hat{q}_0}z)}\right)^2\right]_{|\mathcal{Y}_1}-\left[\frac{\partial\mathcal{Y}_1}{\partial(\sqrt{\hat{q}_0}z)}\right]^2_{|\mathcal{Y}_1}\right)\\
&\quad+\alpha\int \mathrm{D}z\mathcal{Q}\left(\frac{\frac{m}{\sqrt{q}}z+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)\\
&\quad\times\left(\left[\left(\frac{\partial\mathcal{Y}_2}{\partial(\sqrt{\hat{q}_0}z)}\right)^2\right]_{|\mathcal{Y}_2}-\left[\frac{\partial\mathcal{Y}_2}{\partial(\sqrt{\hat{q}_0}z)}\right]^2_{|\mathcal{Y}_2}\right)\\
&\simeq\alpha\left\{\int \mathrm{D}z\mathcal{Q}\left(-\frac{\frac{m}{\sqrt{q}}z+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)\left(\frac{\partial^2\mathcal{Y}_1^{\mathrm{RS}}}{\partial(\sqrt{q_0}z)^2}\right)^2\right.\\
&\qquad\times\left(\left[t^2\right]_{|\mathcal{Y}_1}-[t]^2_{|\mathcal{Y}_1}\right)\\
&\quad+\int \mathrm{D}z\mathcal{Q}\left(\frac{\frac{m}{\sqrt{q}}z+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}}\right)\left(\frac{\partial^2\mathcal{Y}_2^{\mathrm{RS}}}{\partial(\sqrt{q_0}z)^2}\right)^2\\
&\qquad\left.\times\left(\left[t^2\right]_{|\mathcal{Y}_2}-[t]^2_{|\mathcal{Y}_2}\right)\right\}(q_1-q_0)
\end{aligned}$$

$$\simeq \alpha \left\{ \int \mathrm{D}z\mathcal{Q}\left(-\frac{\frac{m}{\sqrt{q}}z+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}} \right) \left(\frac{\partial^2 \mathcal{Y}_1^{\mathrm{RS}}}{\partial(\sqrt{q_0}z)^2} \right)^2 \right.$$
$$\left. + \int \mathrm{D}z\mathcal{Q}\left(\frac{\frac{m}{\sqrt{q}}z+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}} \right) \left(\frac{\partial^2 \mathcal{Y}_2^{\mathrm{RS}}}{\partial(\sqrt{q_0}z)^2} \right)^2 \right\}$$
$$\times (q_1 - q_0) \tag{E.4}$$

hold for $|q_1 - q_0| \ll 1$ and $|\hat{q}_1 - \hat{q}_0| \ll 1$ irrespectively of the value of p, where $[\cdots]_{|\mathcal{Y}} = \frac{\int \mathrm{D}te^{-p\mathcal{Y}}(\cdots)}{\int \mathrm{D}te^{-p\mathcal{Y}}}$. Here $\mathcal{Y}_0^{\mathrm{RS}}$ and $\mathcal{Y}_1^{\mathrm{RS}}$ represent assessments of $\mathcal{Y}_0$ and $\mathcal{Y}_1$ under the assumptions of $\hat{q}_1 = \hat{q}_0$ and $q_1 = q_0$, respectively. In (E.3) and (E.4) we used the Taylor expansion expressions $\partial\mathcal{Y}_0/\partial(\sqrt{\hat{q}_0}z) \sim \partial\mathcal{Y}_0^{\mathrm{RS}}/\partial(\sqrt{\hat{q}_0}z) + \partial^2\mathcal{Y}_0^{\mathrm{RS}}/\partial(\sqrt{\hat{q}_0}z)^2\sqrt{\hat{q}_1 - \hat{q}_0}t$ and $\partial\mathcal{Y}_1/\partial(\sqrt{q_0}z) \sim \partial\mathcal{Y}_1^{\mathrm{RS}}/\partial(\sqrt{q_0}z) + \partial^2\mathcal{Y}_1^{\mathrm{RS}}/\partial(\sqrt{q_0}z)^2\sqrt{q_1 - q_0}t$, and the fact that the variances of t for the measures $\mathrm{D}te^{-p\mathcal{Y}_0}/\int \mathrm{D}te^{-p\mathcal{Y}_0}$ and $\mathrm{D}te^{-p\mathcal{Y}_1}/\int \mathrm{D}te^{-p\mathcal{Y}_1}$ become unity as $\hat{q}_1 - \hat{q}_0$ and $q_1 - q_0$ vanish, irrespectively of the value of p.

To examine the stability of the RS solution in the limit of $\beta \to \infty$, let us change the variable notations as $\chi = \beta(Q - q)$, $\hat{Q} + \hat{q}_1 \to \beta\hat{Q}$, $\hat{q}_1 \to \beta^2\hat{q}_1$, $\hat{q}_0 \to \beta^2\hat{q}_0$, and $\hat{m} \to \beta\hat{m}$ and set $q_0 = q$ and $\hat{q}_0 = \hat{q}$. This yields expressions of $\mathcal{Y}_0^{\mathrm{RS}} \simeq \beta\phi(\sqrt{\hat{q}}z + \hat{m}x^0; \hat{Q}) = -\beta g(\sqrt{\hat{q}}z + \hat{m}x^0)/\hat{Q}$, $\mathcal{Y}_1^{\mathrm{RS}} \simeq (\beta/\chi)u(\sqrt{q}z + \lambda)$ and $\mathcal{Y}_2^{\mathrm{RS}} \simeq (\beta/\chi)u(-\sqrt{q}z - \lambda)$ for $\beta \gg 1$, where $g(x) = \frac{1}{2}(|x| - 1)^2\Theta(|x| - 1)$, $u(x) = x^2\Theta(x)$. Substituting these into (E.3) and (E.4) leads to

$$\Delta \simeq \frac{1}{\hat{Q}^2}\left[\left(g''(\sqrt{\hat{q}}z + \hat{m}x^0) \right)^2 \right]_{x^0,z} \hat{\Delta} \tag{E.5}$$

and

$$\hat{\Delta} \simeq \frac{\alpha}{\chi^2}\left\{ \int \mathrm{D}z\mathcal{Q}\left(-\frac{\frac{m}{\sqrt{q}}z+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}} \right) \left(u''(\sqrt{q}z + \lambda) \right)^2 \right.$$
$$\left. + \int \mathrm{D}z\mathcal{Q}\left(\frac{\frac{m}{\sqrt{q}}z+\lambda}{\sqrt{\rho\sigma_0^2-\frac{m^2}{q}}} \right) \left(u''(-\sqrt{q}z - \lambda) \right)^2 \right\} \Delta, \tag{E.6}$$

where we set $\Delta = q_1 - q$ and $\hat{\Delta} = \hat{q}_1 - \hat{q}$, and used $q \to Q$. The condition that (E.5) and (E.6) allow a solution of $(\Delta, \hat{\Delta}) \neq (0, 0)$ offers (E.1).

We nonetheless think that the RS analysis offers considerably accurate approximates of the exact performance in terms of MSE. Excellent consistency between the numerical experiments and the RS analysis suggests that even if (3.2) has many local optima, they are close to one another in terms of the l_2-norm yielding similar values of MSE.

Appendix F

Derivation of (4.14)

F.1 Assessment of $[P^n(\boldsymbol{y}|\boldsymbol{\Phi})]_{\boldsymbol{\Phi},\boldsymbol{y}}$ for $n \in \mathbb{N}$

Averaging (4.11) with respect to $\boldsymbol{\Phi}$ and $\boldsymbol{y}$ gives the following expression for the n-th moment of the partition function:

$$[P^n(\boldsymbol{y}|\boldsymbol{\Phi})]_{\boldsymbol{\Phi},\boldsymbol{y}} = \int \prod_{a=1}^{n} \left(d\boldsymbol{x}^a P(\boldsymbol{x}^a)\right) \times \left[\prod_{a=1}^{n}\prod_{\mu=1}^{M} \Theta\left((\boldsymbol{y})_\mu (\boldsymbol{\Phi}\boldsymbol{x}^a)_\mu\right)\right]_{\boldsymbol{\Phi},\boldsymbol{y}}. \quad \text{(F.1)}$$

We insert $n(n+1)/2$ trivial identities

$$1 = N \int dq_{ab}\delta\left(\boldsymbol{x}^a \cdot \boldsymbol{x}^b - Nq_{ab}\right), \quad \text{(F.2)}$$

where $a > b = 0, 1, 2, \ldots, n$, into (F.1). Furthermore, we define a joint distribution of $n+1$ vectors $\{\boldsymbol{x}^a\} = \{\boldsymbol{x}^0, \boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^n\}$ as

$$P(\{\boldsymbol{x}^a\}|\boldsymbol{Q}) = \frac{1}{V(\boldsymbol{Q})} P(\boldsymbol{x}^0) \times \prod_{a=1}^{n} \left(P(\boldsymbol{x}^a)\right) \times \prod_{a>b} \delta\left(\boldsymbol{x}^a \cdot \boldsymbol{x}^b - Nq_{ab}\right), \quad \text{(F.3)}$$

where $\boldsymbol{Q} = (q_{ab})$ is an $(n+1) \times (n+1)$ symmetric matrix whose 00 and other diagonal entries are fixed as ρ and Q, respectively.

$$P(\boldsymbol{x}^0) = \prod_{i=1}^{N} \left((1-\rho)\delta(x_i^0) + \rho\tilde{P}(x_i^0)\right) \quad \text{(F.4)}$$

denotes the distribution of the original signal $\boldsymbol{x}^0$, and $V(\boldsymbol{Q})$ is the normalization constant that ensures $\int \prod_{a=0}^{n} d\boldsymbol{x}^a P(\{\boldsymbol{x}^a\}|\boldsymbol{Q}) = 1$ holds. These indicate that (F.1) can also be expressed as

$$[P^n(\boldsymbol{y}|\boldsymbol{\Phi})]_{\boldsymbol{\Phi},\boldsymbol{y}} = \int d\boldsymbol{Q}\left(V(\boldsymbol{Q}) \times \Xi(\boldsymbol{Q})\right), \quad \text{(F.5)}$$

where $d\boldsymbol{Q} \equiv \prod_{a>b} dq_{ab}$ and

$$\Xi(\boldsymbol{Q}) = \int \prod_{a=0}^{n} d\boldsymbol{x}^a P(\{\boldsymbol{x}^a\}|\boldsymbol{Q}) \left[\prod_{a=1}^{n}\prod_{\mu=1}^{M} \Theta\left((\boldsymbol{\Phi}\boldsymbol{x}^0)_\mu (\boldsymbol{\Phi}\boldsymbol{x}^a)_\mu\right)\right]_{\boldsymbol{\Phi}}. \quad \text{(F.6)}$$

Equation (F.6) can be regarded as the average of $\prod_{a=1}^{n}\prod_{\mu=1}^{M} \Theta\left((\boldsymbol{y})_\mu (\boldsymbol{\Phi}\boldsymbol{x}^a)_\mu\right)$ with respect to $\{\boldsymbol{x}^a\}$ and $\boldsymbol{\Phi}$ over distributions of $P(\{\boldsymbol{x}^a\})$ and $P(\boldsymbol{\Phi}) \equiv$

$\left(\sqrt{2\pi/N}\right)^{-MN} \exp\left(-(N/2)\sum_{\mu,i}\Phi_{\mu i}^2\right)$. Notice that, here we replaced the average over $\boldsymbol{y}$ by the average over $\boldsymbol{x}^0$, since $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}^0$. In computing this, note that the central limit theorem guarantees that $u_\mu^a \equiv (\boldsymbol{\Phi}\boldsymbol{x}^a)_\mu = \sum_{i=1}^N \Phi_{\mu i} x_i^a$ can be handled as zero-mean multivariate Gaussian random numbers whose variance and covariance are given by

$$\left[u_\mu^a u_\nu^b\right]_{\boldsymbol{\Phi},\{\boldsymbol{x}^a\}} = \delta_{\mu\nu} q_{ab}, \tag{F.7}$$

when $\boldsymbol{\Phi}$ and $\{\boldsymbol{x}^a\}$ are generated independently from $P(\boldsymbol{\Phi})$ and $P(\{\boldsymbol{x}^a\})$, respectively. This means that (F.6) can be evaluated as

$$\begin{aligned}\Xi(\boldsymbol{Q}) &= \left(\frac{\int d\boldsymbol{u} \exp\left(-\frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{u}\right)\prod_{a=1}^n \Theta\left(u^0 u^a\right)}{(2\pi)^{(n+1)/2}(\det \boldsymbol{Q})^{1/2}}\right)^M \\ &= \left(2\int \frac{d\boldsymbol{u} \exp\left(-\frac{1}{2}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{u}\right)\Theta\left(u^0\right)\prod_{a=1}^n \Theta\left(u^a\right)}{(2\pi)^{(n+1)/2}(\det \boldsymbol{Q})^{1/2}}\right)^M . \end{aligned} \tag{F.8}$$

On the other hand, expressions

$$\delta\left(|\boldsymbol{x}^a|^2 - NQ\right) = \frac{1}{4\pi}\int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty} d\hat{q}_{aa} \exp\left(-\frac{1}{2}\hat{q}_{aa}\left(|\boldsymbol{x}^a|^2 - NQ\right)\right) \tag{F.9}$$

and

$$\delta\left(\boldsymbol{x}^a \cdot \boldsymbol{x}^b - N q_{ab}\right) = \frac{1}{2\pi}\int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty} d\hat{q}_{ab} \exp\left(\hat{q}_{ab}\left(\boldsymbol{x}^a \cdot \boldsymbol{x}^b - N q_{ab}\right)\right), \tag{F.10}$$

and use of the saddle-point method, offer

$$\begin{aligned}\frac{1}{N}\log V(\boldsymbol{Q}) = \underset{\hat{\boldsymbol{Q}}}{\mathrm{extr}} \Bigg\{ &-\frac{1}{2}\mathrm{Tr}\hat{\boldsymbol{Q}}\boldsymbol{Q} \\ &+ \log\left(\int d\boldsymbol{x} P(x^0)\prod_{a=1}^n P(x^a)\exp\left(\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\hat{\boldsymbol{Q}}\boldsymbol{x}\right)\right)\Bigg\}. \end{aligned} \tag{F.11}$$

Here, $\boldsymbol{x} = (x^0, x^1, \ldots, x^n)^{\mathrm{T}}$ and $\hat{\boldsymbol{Q}}$ is an $(n+1)\times(n+1)$ symmetric matrix whose 00 and other diagonal components are given as 0 and $-\hat{q}_{aa}$, respectively. The off-diagonal entries are $\hat{q}_{ab}$. Equations (F.8) and (F.11) indicate that $N^{-1}\log\left[P^n(\boldsymbol{y}|\boldsymbol{\Phi})\right]_{\boldsymbol{\Phi},\boldsymbol{y}}$ is correctly evaluated by the saddle-point method with respect to $\boldsymbol{Q}$ in the assessment of the right-hand side of (F.5), when N and M tend to infinity and $\alpha = M/N$ remains finite.

F.2 Treatment under the replica symmetric ansatz

Let us assume that the relevant saddle-point for assessing (F.5) is of the form (4.13) and, accordingly,

$$\hat{q}_{ab} = \hat{q}_{ba} = \begin{cases} 0, & (a = b = 0) \\ \hat{m}, & (a = 1, 2, \ldots, n; b = 0) \\ \hat{Q}, & (a = b = 1, 2, \ldots, n) \\ \hat{q}, & (a \neq b = 1, 2, \ldots, n) \end{cases} . \tag{F.12}$$

The $n+1$-dimensional Gaussian random variables $u^0, u^1, \dots u^n$ whose variance and covariance are given by (4.13) can be expressed as

$$u^0 = \sqrt{\rho - \frac{m^2}{q}} s^0 + \frac{m}{\sqrt{q}} z, \tag{F.13}$$

$$u^a = \sqrt{Q-q} s^a + \sqrt{q} z, \ (a = 1, 2, \dots, n) \tag{F.14}$$

utilizing $n+2$ independent standard Gaussian random variables z and $s^0, s^1, \dots, s^n$. This indicates that (F.8) is evaluated as

$$\Xi(\boldsymbol{Q}) = \left(2 \int \mathrm{D}z \mathcal{Q} \left(\frac{m}{\sqrt{\rho q - m^2}} z \right) \mathcal{Q}^n \left(\sqrt{\frac{q}{Q-q}} z \right) \right)^M . \tag{F.15}$$

On the other hand, substituting (F.12) into (F.11), in conjunction with the identity

$$\exp\left(\hat{q} \sum_{a>b(\geq 1)} x^a x^b \right) = \int \mathrm{D}z \exp\left(\sum_{a=1}^{n} \left(-\frac{\hat{q}}{2} (x^a)^2 + \sqrt{\hat{q}} z x^a \right) \right), \tag{F.16}$$

provides

$$\begin{aligned} \frac{1}{N} \log V(\boldsymbol{Q}) = \underset{\hat{Q},\hat{q},\hat{m}}{\mathrm{extr}} \Bigg\{ & \frac{n}{2} \hat{Q} Q - \frac{n(n-1)}{2} \hat{q} q - \hat{m} m \\ & + \log \left[\left(\int dx P(x) \exp\left(-\frac{\hat{Q}+\hat{q}}{2} x^2 + \left(\sqrt{\hat{q}} z + \hat{m} x^0 \right) x \right) \right)^n \right]_{x^0, z} \Bigg\} \end{aligned} \tag{F.17}$$

Although we have assumed that $n \in \mathbb{N}$, the expressions of (F.15) and (F.17) are likely to hold for $n \in \mathbb{R}$ as well. Therefore, the average free energy $\overline{f}$ can be evaluated by substituting these expressions into the formula $\overline{f} = \lim_{n\to 0} (\partial/\partial n) \left((N)^{-1} \log \left[P^n \left(\boldsymbol{y} | \boldsymbol{\Phi} \right) \right]_{\boldsymbol{\Phi}, \boldsymbol{y}} \right)$.

Furthermore, employing the approximate expressions

$$\lim_{n\to 0} \mathcal{Q}^n(x) = \lim_{n\to 0} \exp\left(n \log \mathcal{Q}(x) \right) \approx 1 + n \log \mathcal{Q}(x), \tag{F.18}$$

$$\lim_{n\to 0} \log\left(1 + nC(\cdot) \right) \approx nC(\cdot), \tag{F.19}$$

where $C(\cdot)$ is an arbitrary function, we obtain the form

$$\lim_{n\to 0} \frac{\partial}{\partial n} \frac{1}{N} \log \Xi(\boldsymbol{Q}) = 2\alpha \int \mathrm{D}z \mathcal{Q} \left(\frac{m}{\sqrt{\rho q - m^2}} z \right) \mathcal{Q} \left(\sqrt{\frac{q}{Q-q}} z \right). \tag{F.20}$$

And we have

$$\begin{aligned} \lim_{n\to 0} \frac{\partial}{\partial n} \frac{1}{N} \log V(\boldsymbol{Q}) \quad = \quad & \underset{\hat{Q},\hat{q},\hat{m}}{\mathrm{extr}} \Bigg\{ \int \mathrm{d}x^0 P\left(x^0 \right) \int \mathrm{D}z \phi \left(\sqrt{\hat{q}} z + \hat{m} x^0; \hat{Q} \right) \\ & + \frac{1}{2} Q \hat{Q} + \frac{1}{2} q \hat{q} - m \hat{m} \Bigg\}. \end{aligned} \tag{F.21}$$

Using these in the resultant expression of $\overline{f}$ gives (4.14).

Appendix G

Derivation of (4.33)–(4.37)

Expanding the exponential in (4.31) up to the second order of $\Phi_{\mu i}(x_i - a_{i\to\mu})$ and performing the integration with respect to u_μ gives

$$\begin{aligned} m_{\mu\to i}(x_i) &\simeq c_0 + c_1\Phi_{\mu i}(x_i - a_{i\to\mu}) + \frac{1}{2}c_2\Phi_{\mu i}^2(x_i - a_{i\to\mu})^2 \\ &\simeq \exp\left\{\ln c_0 + \frac{c_1}{c_0}\Phi_{\mu i}(x_i - a_{i\to\mu}) + \frac{c_0c_2 - c_1^2}{2c_0^2}\Phi_{\mu i}^2(x_i - a_{i\to\mu})^2\right\} \\ &\propto \exp\left\{-\frac{A_{\mu\to i}}{2}x_i^2 + B_{\mu\to i}x_i\right\}, \end{aligned} \tag{G.1}$$

where

$$c_0 \equiv \int \mathrm{d}u_\mu P(y_\mu|u_\mu)\exp\left(-\frac{(u_\mu - \omega_\mu)^2}{2V}\right), \tag{G.2}$$

$$c_1 \equiv \int \mathrm{d}u_\mu P(y_\mu|u_\mu)\left(\frac{u_\mu - \omega_\mu}{V}\right)\exp\left(-\frac{(u_\mu - \omega_\mu)^2}{2V}\right), \tag{G.3}$$

$$c_2 \equiv \int \mathrm{d}u_\mu P(y_\mu|u_\mu)\left(\left(\frac{u_\mu - \omega_\mu}{V}\right)^2 - \frac{1}{V}\right)\exp\left(-\frac{(u_\mu - \omega_\mu)^2}{2V}\right), \tag{G.4}$$

and

$$A_{\mu\to i} = \frac{c_1^2 - c_0c_2}{c_0^2}\Phi_{\mu i}^2, \tag{G.5}$$

$$B_{\mu\to i} = \frac{c_1}{c_0}\Phi_{\mu i} + \frac{c_1^2 - c_0c_2}{c_0^2}\Phi_{\mu i}^2 a_{i\to\mu}. \tag{G.6}$$

Equations (G.3) and (G.4) imply that c_1 and c_2 can be expressed as $c_1 = \partial c_0/\partial\omega_\mu$ and $c_2 = \partial^2 c_0/\partial\omega_\mu^2$, respectively. Inserting this into (G.5) and (G.6), we obtain (4.33)–(4.37).

Appendix H

Asymptotic form of $\mathrm{MSE}^{\mathrm{Bayes}}$

The behavior as $m \to \rho$ and $\hat{m} \to \infty$ is obtained as $\alpha \to \infty$. This implies that equations (4.16) and (4.17) can be evaluated as

$$\begin{aligned} m &= \int \mathrm{D}t \frac{\rho^2(1+\hat{m})^{-1} e^{\frac{\hat{m}}{1+\hat{m}}t^2} \frac{\hat{m}}{(1+\hat{m})^2} t^2}{1-\rho+\rho(1+\hat{m})^{-1/2} e^{\frac{\hat{m}}{2(1+\hat{m})}t^2}} \\ &= \frac{\rho^2 \hat{m}}{(1+\hat{m})} \int \mathrm{D}z z^2 \left[(1-\rho)(1+\hat{m})^{1/2} e^{-\frac{\hat{m}}{2}z^2} + \rho\right]^{-1} \\ &\simeq \rho(1-\hat{m}^{-1}) \end{aligned} \tag{H.1}$$

and

$$\begin{aligned} \hat{m} &= \frac{2\alpha}{\rho-m} \int \mathrm{D}t \frac{e^{-\frac{m}{\rho-m}t^2}/(2\pi)}{\mathcal{Q}\left(\sqrt{\frac{m}{\rho-m}}t\right)} = \frac{2\alpha}{\sqrt{m(\rho-m)}} \int \frac{\mathrm{d}z}{(2\pi)^{3/2}} \frac{e^{-\frac{\rho+m}{2m}z^2}}{\mathcal{Q}(z)} \\ &\simeq \frac{2C\alpha}{\sqrt{m(\rho-m)}}, \end{aligned} \tag{H.2}$$

respectively. Here, the integration variables have been changed to $(1+\hat{m})^{-1/2}t = z$ and $\sqrt{m/(\rho-m)}t = z$ in (H.1) and (H.2), respectively, and we set $C \equiv \int \mathrm{d}z (2\pi)^{-3/2} e^{-z^2}/\mathcal{Q}(z) = 0.3603\ldots$. Equations (H.1) and (H.2) yield an asymptotic expression for m:

$$m \simeq \rho\left(1 - \left(\frac{\rho}{2C\alpha}\right)^2\right). \tag{H.3}$$

Inserting this into (4.19) gives (4.51).

The performance when the positions of non-zero entries are known can be evaluated by setting $\rho = 1$ and replacing α with α/ρ in (4.16) and (4.17) as the dimensionality of $\boldsymbol{x}$ is reduced from N to $N\rho$. This reproduces (4.51) in the asymptotic region of $\alpha \gg 1$.

Appendix I

Asymptotic form of MSE^{l_1}

The saddle-point equations of the l_1-norm minimization approach under a normalization constraint of $|\boldsymbol{x}|^2 = N$ are as follows [56]:

$$\hat{q} = \frac{\alpha}{\pi\chi^2}\left(\arctan\left(\frac{\sqrt{\rho-m^2}}{m}\right)-\frac{m}{\rho}\sqrt{\rho-m^2}\right), \tag{I.1}$$

$$\hat{m} = \frac{\alpha}{\pi\chi\rho}\sqrt{\rho-m^2}, \tag{I.2}$$

$$\begin{aligned}\hat{Q}^2 = 2\Bigg\{&(1-\rho)\left[(\hat{q}+1)\,\mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}}}\right)-\sqrt{\frac{\hat{q}}{2\pi}}e^{-\frac{1}{2\hat{q}}}\right]\\&+\rho\left[\left(\hat{q}+\hat{m}^2+1\right)\mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}+\hat{m}^2}}\right)\right.\\&\left.-\sqrt{\frac{\hat{q}+\hat{m}^2}{2\pi}}e^{-\frac{1}{2(\hat{q}+\hat{m}^2)}}\right]\Bigg\},\end{aligned} \tag{I.3}$$

$$\chi = \frac{2}{\hat{Q}}\left[(1-\rho)\,\mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}}}\right)+\rho\mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}+\hat{m}^2}}\right)\right], \tag{I.4}$$

$$m = \frac{2\rho\hat{m}}{\hat{Q}}\mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}+\hat{m}^2}}\right). \tag{I.5}$$

The behavior as $m \to \sqrt{\rho}$ and $\hat{m} \to \infty$ is obtained as $\alpha \to \infty$. This implies that (I.3) can be evaluated as

$$\begin{aligned}\hat{Q} &\simeq \left(\rho\hat{m}^2-\frac{4\hat{m}}{\sqrt{2\pi}}+B(\hat{q},\rho)\right)^{1/2}\\&\simeq \sqrt{\rho}\hat{m}\left[1-\frac{2}{\sqrt{2\pi}\hat{m}}+\left(\frac{B(\hat{q},\rho)}{2\rho}-\frac{3}{\pi}\right)\right],\end{aligned} \tag{I.6}$$

where $B(\hat{q},\rho) \equiv \rho\,(\hat{q}+1)+2\,(1-\rho)\left[(\hat{q}+1)\,\mathcal{Q}\left(\frac{1}{\sqrt{\hat{q}}}\right)-\sqrt{\frac{\hat{q}}{2\pi}}e^{-\frac{1}{2\hat{q}}}\right]$. Inserting (I.6) into (I.5), we obtain

$$m \simeq \sqrt{\rho}\,(1-\delta), \tag{I.7}$$

where

$$\delta \equiv \left(\frac{B(\hat{q},\rho)}{2\rho}-\frac{1}{\pi}\right)/\hat{m}^2 = \pi^2\left[2(1-\rho)\mathcal{Q}\left(1/\sqrt{\hat{q}}\right)+\rho\right]^2/(2\alpha^2). \tag{I.8}$$

Inserting (I.2), (I.6), (I.7), and $\chi \simeq \left[2(1-\rho)\mathcal{Q}(1/\sqrt{\hat{q}})\right]/\hat{Q}$ into (I.1) yields a closed equation with respect to $\hat{q}$:

$$\hat{q} \simeq \frac{2}{3}\left(B(\hat{q},\rho) - \frac{2\rho}{\pi}\right)\left[2(1-\rho)\mathcal{Q}(1/\sqrt{\hat{q}}) + \rho\right]^{-1}. \tag{I.9}$$

This determines the value of $\hat{q}$ for $\alpha \to \infty$, $\hat{q}_{l_1}^{\infty}(\rho)$. Combining (I.8) and

$$\mathrm{MSE}^{l_1} = 2\left(1 - \frac{m}{\sqrt{\rho}}\right) \simeq 2\delta \tag{I.10}$$

gives (4.52) in the asymptotic region of $\alpha \gg 1$.

Bibliography

[1] H Akaike. "A new look at the statistical model identification". In: *IEEE Trans. on AC* 19 (1974), p. 716. URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1100705&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D1100705.

[2] J R L de Almeida and D J Thouless. "Stability of the Sherrington-Kirkpatrick solution of a spin glass model". In: *J. Phys. A* 11 (1978), p. 983. URL: http://iopscience.iop.org/article/10.1088/0305-4470/11/5/028/meta.

[3] F Bach. "Learning with Submodular Functions: A Convex Optimization Perspective". In: *arXiv* 1111.6453 (). URL: http://arxiv.org/abs/1111.6453.

[4] M Bayati and A Montanari. "The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing". In: *IEEE Trans. on Inform. Theory* 57 (2011), pp. 764–785. URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5695122&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D5695122.

[5] B Boser and B Wooley. "The design of sigma-delta modulation analog-to-digital converters". In: *Solid-State Circuits, IEEE Journal* 23 (1988), pp. 1298–1308. URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=90025&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D90025.

[6] E J Candès, J Romberg, and T Tao. "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information". In: *IEEE Trans. Inform. Theory* 52 (2006), pp. 489 –509. URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1580791&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D1580791.

[7] E J Candès and M B Wakin. "An Introduction To Compressive Sampling". In: *IEEE Signal Processing Magazine* (2008), pp. 21–30. URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=4472240&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D4472240.

[8] J Christensen. "A Brief History of Linear Algebra". In: *University of Utah, Grant Gustafson, Final Project Math 2270* (). URL: http://www.math.utah.edu/~gustafso/s2012/2270/web-projects/christensen-HistoryLinearAlgebra.pdf.

[9] "Compressed sensing hardware". In: (). URL: https://sites.google.com/site/igorcarron2/compressedsensinghardware.

[10] D L Donoho. "Compressed sensing". In: *IEEE Trans. Inform. Theory* 52 (2006), pp. 1289 –1306. URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1614066&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D1614066.

[11] D L Donoho. "High-Dimensional Centrally Symmetric Polytopes with Neighborliness Proportional to Dimension". In: *Discrete Computational Geometry* 35 (2006), pp. 617–652. URL: http://link.springer.com/article/10.1007%2Fs00454-005-1220-0.

[12] D L Donoho, A Maleki, and A Montanari. "Message-passing algorithms for compressed sensing". In: *Proc. Nat. Acad. Sci.* 106.18914–18919 (2009). URL: http://www.pnas.org/content/106/45/18914.full.

[13] D L Donoho and J Tanner. "Counting faces of randomly projected polytopes when the projection radically lowers dimension". In: *J. Amer. Math. Soc.* 22 (2009), pp. 1–53. URL: http://www.ams.org/journals/jams/2009-22-01/S0894-0347-08-00600-0/home.html.

[14] V S Dotsenko. "Introduction to the Replica Theory of Disordered Statistical Systems". In: *Cambridge: Cambridge University Press* (2001). URL: http://ebooks.cambridge.org/ebook.jsf?bid=CBO9780511524592.

[15] M Duarte et al. "Single-pixel imaging via compressive sampling". In: *EEE Signal Process. Mag.* 25.2 (2008), pp. 83–91. URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=4472247&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D4472247.

[16] E.Candes. "Compressive sampling". In: *presented at the Int. Congr. Math., Madrid, Spain* (2006). URL: http://www.signallake.com/innovation/CompressiveSampling06.pdf.

[17] M Elad. "Sparse and Redundant Repersentations–from theory to applications in signal and image processing". In: *New York: Springer* (2010).

[18] A.K. Fletcher, S Rangan, and V.K. Goyal. "On the rate-distortion performance of compressed sensing". In: *Acoustics, Speech and Signal Processing,ICASSP 2007, IEEE International Conference* 3 (2007). URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=4217852&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D4217852.

[19] S Ganguli and H Sompolinsky. "Statistical mechanics of compressed sensing". In: *Phys. Rev. Lett.* 104.188701 (2010). URL: http://elsc.huji.ac.il/sites/default/files/physrevlett.104.188701.pdf.

[20] M Grant, S Boyd, and Y Ye. "Matlab Software for Disciplined Convex Programming". In: Version 2.1 (2015). URL: http://cvxr.com/cvx/.

[21] K. Hayashi, M. Nagahara, and T. Tanaka. "A user's guide to compressed sensing for communications systems". In: *IEICE Trans. Commun.* E96-B.3 (). URL: http://repository.kulib.kyoto-u.ac.jp/dspace/bitstream/2433/170851/1/ieice-e96-b3_pp685.pdf.

[22] "Initiative for High-Dimensional Data-Driven Science through Deepening of Sparse Modeling, 2013-2017". In: (). URL: http://sparse-modeling.jp/program/A02-3.html.

[23] L Jacques et al. "Robust 1-Bit Compressive Sensing via Binary Stable Embeddings of Sparse Vectors". In: *Information Theory, IEEE Transactions on* (2011), pp. 2082–2102. URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6418031&url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F18%2F4667673%2F06418031.pdf%3Farnumber%3D6418031.

[24] Y Kabashima. "A CDMA multiuser detection algorithm on the basis of belief propagation". In: *J. Phys. A* 36.11111 (2003). URL: http://www.sp.dis.titech.ac.jp/~kaba/papers/cdma.pdf.

[25] Y Kabashima. "Overview of compressed sensing". In: *Tokyo Institute of Technology, Lecture of Advanced Topics in Mathematical Information Sciences I* (). URL: http://www.sp.dis.titech.ac.jp/syllabus/OverviewCS_Kabashima.pdf.

[26] Y Kabashima and D Saad. "Belief propagation vs. TAP for decoding corrupted messages". In: *Europhys. Lett.* 44 (1998), p. 668. URL: http://iopscience.iop.org/article/10.1209/epl/i1998-00524-7/fulltext/.

[27] Y Kabashima and S Uda. "A BP-based algorithm for performing Bayesian inference in large perceptron-type networks". In: *ALT 2004; Lecture Notes in AI, Springer* 3244 (2004), pp. 479–493. URL: http://link.springer.com/chapter/10.1007%2F978-3-540-30215-5_36.

[28] Y Kabashima and M Vehkaperä. "Signal recovery using expectation consistent approximation for linear observations". In: *Information Theory (ISIT), 2014 IEEE International Symposium on, Honolulu, HI* (2014), pp. 226–230. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6874828&punumber%3D6867217%26filter%3DAND%28p_IS_Number%3A6874773%29%26pageNumber%3D3.

[29] Y Kabashima, M Vehkaperä, and S Chatterjee. "Typical l1-recovery limit of sparse vectors represented by concatenations of random orthogonal matrices". In: *J. Stat. Mech.* P12003 (2012). URL: http://iopscience.iop.org/article/10.1088/1742-5468/2012/12/P12003/meta.

[30] Y Kabashima, T Wadayama, and T Tanaka. "A typical reconstruction limit of compressed sensing based on Lp-norm minimization". In: *J. Stat. Mech.* L09003; E07001 (2009,2012). URL: http://iopscience.iop.org/article/10.1088/1742-5468/2009/09/L09003.

[31] Y Kakiuchi. In: *Kyoto Sangyo Uni.* (). URL: http://www.cc.kyoto-su.ac.jp/~kano/pdf/study/student/2014KakiuchiPresen.pdf.

[32] U.S. Kamilov et al. "One-bit Measurements with Adaptive Thresholds". In: *IEEE Signal Process. Letters* 19.10 (2012), pp. 607–610. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=6244855.

[33] F Krzakala et al. "Statistical-Physics-Based Reconstruction in Compressed Sensing". In: *Phys. Rev. X* 2.021005 (2012). URL: http://journals.aps.org/prx/abstract/10.1103/PhysRevX.2.021005.

[34] D Lee et al. "Spectrum Sensing for Networked System Using 1-Bit Compressed Sensing with Partial Random Circulant Measurement Matrices". In: *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th* (2012). URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6240259&sortType%3Dasc_p_Sequence%26filter%3DAND%28p_IS_Number%3A6239848%29%26pageNumber%3D19.

[35] D J C MacKay. "Good error-correcting codes based on very sparse matrices". In: *IEEE Trans. Inform. Theory* 45 (1999), p. 399. URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=748992&type=ref&url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel4%2F18%2F16170%2F00748992.pdf%3Ftp%3D%26isnumber%3D16170%26arnumber%3D748992%26type%3Dref.

[36] D J C MacKay and R M Neal. "Near Shannon limit performance of low density parity check codes". In: *Elect. Lett.* 33 (1997), p. 457. URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=585036&url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel1%2F2220%2F12683%2F00585036.

[37] M Mézard and M Montanari. "Information, Physics, and Computation". In: *Oxford University Press* (2009). URL: http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198570837.001.0001/acprof-9780198570837.

[38] M Mézard, G Parisi, and M A Virasoro. "Spin Glass Theory and Beyond–An Introduction to the Replica Method and Its Applications". In: *Singapore: World Scientific Lecture Notes in Physics* 9 (1987). URL: http://www.worldscientific.com/worldscibooks/10.1142/0271.

[39] H Nishimori. "Statistical Physics of Spin Glasses and Information Processing". In: *Oxford: Oxford University Press* (2001). URL: http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198509417.001.0001/acprof-9780198509417.

[40] S Oymak and B Hassibi. "A Case for Orthogonal Measurements in Linear Inverse Problems". In: *Information Theory (ISIT), 2014 IEEE International Symposium on,Honolulu, HI* (2014), pp. 3175–3179. URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6875420&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D6875420.

[41] S Rangan. "Generalized approximate message passing for estimation with random linear mixing". In: *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on* (2010), pp. 2168 –2172. URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=

&arnumber=6033942&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D6033942.

[42] J Rissanen. "Modeling by shortest data description". In: *Automatica* 14 (1978), 465–471. URL: http://www.sciencedirect.com/science/article/pii/0005109878900055.

[43] S Sarvotham, D Baron, and R Baranuik. "Measurements vs. bits: Compressed sensing meets information theory". In: *Proceedings of 44th Allerton Conf. Comm., Ctrl., Computing.* (2006). URL: http://dsp.rice.edu/sites/dsp.rice.edu/files/cs/allerton2006SBB.pdf.

[44] H S Seung, H Sompolinsky, and N Tishby. "Statistical mechanics of learning from examples". In: *Phys. Rev. A* 45.6056 (1992). URL: http://journals.aps.org/pra/abstract/10.1103/PhysRevA.45.6056.

[45] C.E. Shannon. "Communication in the presence of noise". In: *Proceedings of the IEEE* 86.2 (). URL: http://web.stanford.edu/class/ee104/shannonpaper.pdf.

[46] M Shiino and T Fukai. "Study of self-inhibited analogue neural networks using the self-consistent signal-to-noise analysis". In: *J. Phys. A* 25.L375 (1992). URL: http://iopscience.iop.org/article/10.1088/0305-4470/25/18/014/meta.

[47] T Shinzato and Y Kabashima. "Learning from correlated patterns by simple perceptrons". In: *J. Phys. A* 42.015005 (2009). URL: http://iopscience.iop.org/article/10.1088/1751-8113/42/1/015005/meta.

[48] J-L Starck, F Murtagh, and J M Fadili. "Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity". In: *New York: Cambridge University Press* (2010).

[49] Boufounos P T and Baraniuk R G. In: *Proc. Conf. Inform. Science and Systems (CISS), Princeton, NJ* 16-21 (2008). URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4558487.

[50] D J Thouless, P W Anderson, and R G Palmer. "Solution of 'Solvable model of a spin glass'". In: *Phil. Mag.* 35 (1977), pp. 593–601. URL: http://www.tandfonline.com/doi/abs/10.1080/14786437708235992.

[51] J Tropp et al. "Beyond Nyquist: Efficient sampling of sparse, bandlimited signals". In: *IEEE Trans. Inf. Theory* 56 (2010), pp. 520–544. URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5361485&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D5361485.

[52] M Vehkaperä, Y Kabashima, and S Chatterjee. "Analysis of Regularized LS Reconstruction and Random Matrix Ensembles in Compressed Sensing". In: *Information Theory (ISIT), 2014 IEEE International Symposium on,Honolulu, HI* (2014), pp. 3185–3189. URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6875422&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D6875422.

[53] S Watanabe. "Algebraic Geometry and Statistical Learning Theory". In: *Cambridge University Press* (2009). URL: http://www.cambridge.org/fr/academic/subjects/computer-science/pattern-recognition-and-machine-learning/algebraic-geometry-and-statistical-learning-theory?format=HB.

[54] T L H Watkin, A Rau, and M Biehl. "The statistical mechanics of learning a rule". In: *Rev. Mod. Phys.* 65 (1993), p. 499. URL: http://journals.aps.org/rmp/abstract/10.1103/RevModPhys.65.499.

[55] C-K Wen and K-K Wong. "Analysis of Compressed Sensing with Spatially-Coupled Orthogonal Matrices". In: (2014). URL: http://arxiv.org/abs/1402.3215.

[56] Y Xu and Y Kabashima. "Statistical mechanics approach to 1-bit compressed sensing". In: *J. Stat. Mech.* P02041 (2013). URL: http://iopscience.iop.org/article/10.1088/1742-5468/2013/02/P02041/meta.

[57] Y Xu, Y Kabashima, and L Zdeborova. "Bayesian signal reconstruction for 1-bit compressed sensing". In: *J. Stat. Mech.* P11015 (2014). URL: http://iopscience.iop.org/article/10.1088/1742-5468/2014/11/P11015.