Compressed Sensing Performance Analysis Via Replica Method Using Bayesian Framework

Solomon A. Tesfamicael

Dept. of Electronics and Telecommunications (IET)
(and Faculty of Teacher and Interpreter Education (FLT))
Norwegian University of Science and Technology (NTNU)
Email: tesfamic@ntnu.no

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

Keywords - Compressed Sensing; Performance Analysis; Replica Method; Bayesian Framework

I. INTRODUCTION

Recently questions like, *why go to so much effort to acquire all the data when most of what we get will be thrown away?Can we just directly measure the part that will not end up being thrown away?* that were paused by Donoho and others [1]–[4] triggered a new way of sampling or sensing called compact (”compressed”) sensing (CS).

The CS paradigm in signal processing requires three important ingredients [5]. First, the desired signal should have a sparse representation in a known transform domain, i.e., it should be compressible. If the signal is sparse spatially, for example consider an image which is sparse in the pixels, then the transform domain can be the identity. Second, the aliasing artifacts due to undersampling should be uncoherent in the transform domain. This creates a noise-like structure. This measurement noise then can be modeled using white Gaussian noise. Third, a nonlinear reconstruction scheme should be used to enforce sparsity and consisitency with the data [6]. Recently, this recovery using CS has been shown to be mathematically exact [1]–[4]. As a signal processing scheme, CS follows a similar framework: encoding, transmission/storing, and decoding. A block diagram is given in Figure 1 focusing on the encoding and decoding of such a system for noisy measurement.

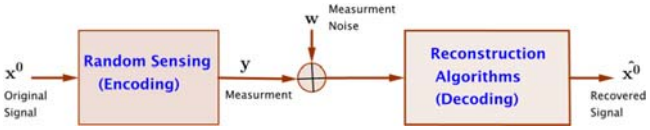

Fig. 1: Blockdiagram for CS-based reconstruction.

In CS the task is to estimate or recover a sparse or compressible vector $\mathbf{x}^0 \in R^N$ from a measurement vector $\mathbf{y} \in R^M$. These are related through the linear transform $\mathbf{y} = \mathbf{A}\mathbf{x}^0$. Here, $\mathbf{x}^0$ is a sparse vector and $M \ll N$. In the seminal papers [1], [3], [4], $\mathbf{x}^0$ is estimated from $\mathbf{y}$, by solving a convex optimization problem [7], [8]. Others have used greedy algorithms, like subspace pursuit (SP) [9], orthogonal matching pursuit (OMP) [10] to solve the problem. In this paper the focus is rather on the convex optimization methods. We consider the noisy measurement system and the linear relation becomes

$$\mathbf{y} = \mathbf{A}\mathbf{x}^0 + \sigma_0\mathbf{w}, \qquad \text{(I.1)}$$

where $\mathbf{y}$ and $\mathbf{x}^0$ are as above where as the noise term, $\mathbf{w} \sim N(0,\mathbf{I})$. There exists a large body of work on how to efficiently obtain an estimate for $\mathbf{x}^0$. The performance of such estimators are measured using metrics like Restricted Isometric Property (RIP) [2], Mutual Coherence (MC) [11], yet there is apparently no consensus on the bound determining how many measurements M are needed to approximate the sparse signal with lengeth N and sparsity k by using such metrics. The tool used in this paper gives performance bounds of large size CS systems [12] using these system parameters. Generally the linear model (I.1) is used to describe a multitude of linear systems like code division multiple access (CDMA) and multiple antenna systems like MIMO, to mention just a few. Tools from

statistical mechanics have been employed to analyze large CDMA [13] and MIMO systems [14]. In addition, this paper applies the same wisdom to analyze the performance of estimators used in CS. Guo et al. in [12] used a Bayesian framework for statistical inference with noisy measurements and characterize the posterior distribution of individual elements of the sparse signal by describing the mean square error (MSE) exactly. To do so, they consider (I.1) in a large system and applied the decoupling principle using tools from statistical mechanics.

One can also find work that has used the tools from statistical mechanics to analyze CS system performance. To mention some, in [12] as stated above, Guo et al. used the tools to describe the minimum mean square error (MMSE) estimator, in [15] Rangan et al. used the maximum a posterior(MAP) estimator of CS systems. These are referred as Replica MMSE claim and Replica MAP claim in [15]. In [16]–[18] and [19] authors have used Belief propagation and message passing algorithms for probabilistic reconstruction in CS using replica methods including RS. Especially, in [20] one finds excellent work about phase diagrams in CS systems while [21] generalizes replica analysis using free random matrices. Recently, authors in [22] have proposed a turbo compressed sensing algorithm with partial discrete Fourier transform (DFT) sensing matrices. They claim that algorithm outperforms the well-known approximate message passing (AMP) algorithm when a partial DFT sensing matrix is involved. Kabashima et. al in [23], Ganguli and Sompolinsky in [24] and Takeda and Kabashima [25]–[27] have shown statistical mechanical analysis of the CS by considering the noiseless recovery problem and they indicated that RSB analysis is needed in the phase regimes where the RS solution is not stable. In this paper the performance of these CS estimators, considered as MAP estimator, is shown for the noisy problem by using the replica method including RS and RSB as in [28], [29], where the RSB ansatz gives a better solution when the replica symmetry (RS) solution is unstable. This work is a kind of extension of [29], from MIMO systems to the CS systems.

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

II. BAYESIAN FRAMEWORK FOR SPARSE ESTIMATION

Beginning with a given vector of measurements $\mathbf{y} \in \mathrm{R}^M$ and measurement matrix $\mathbf{A} \in \mathrm{R}^{M\times N}$, we assume noisy measurement with $\mathbf{w} \in \mathrm{R}^M$ being i.i.d. Gaussian random variables with zero mean and covariance matrix $\mathbf{I}$, estimating the sparse vector $\mathbf{x}^0 \in \mathrm{R}^N$ is the problem that we are considering where these variables are related by the linear model (I.1).

A. Sparse Signal Estimation

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

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

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

$$\hat{\mathbf{x}^0} = \min_{\mathbf{x}\in\mathbb{R}^N}\left[\frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + uf(\mathbf{x})\right], \tag{II.2}$$

where $uf(\mathbf{x})$ is a regularizing term, for some non-negative u. By taking $f(\mathbf{x}) = \|\mathbf{x}\|_p$, emphasis is made on a solution with LP norm, and $\|\mathbf{x}\|_p$ is defined as a penalizing norm.

When $p = 2$, we get

$$\hat{\mathbf{x}^0} = \min_{\mathbf{x}\in\mathbb{R}^N}\left[\frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + u\|\mathbf{x}\|_2\right], \tag{II.3}$$

This is penalizing the least square error by the L2 norm and this performs badly as well, since it does not introduce sparsity into the problem. When $p = 0$, we get the L0 norm, which is defined as

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

the number of the non-zero entries of $\mathbf{x}$, which actually is a partial norm since it does not satisfy the triangle inequality property, but can be treated as norm by defining it as in [15], and get the L0 norm regularizing estimator

$$\hat{\mathbf{x}^0} = \min_{\mathbf{x}\in\mathbb{R}^N}\left[\frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + u\|\mathbf{x}\|_0\right], \tag{II.4}$$

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

approximate it using the L1 penalizing norm to get the estimator

$$\hat{\mathbf{x}}^0 = \min_{\mathbf{x}\in\mathbb{R}^N}\left[\frac{1}{2}\parallel \mathbf{y}-\mathbf{A}\mathbf{x}\parallel_2^2 + u\parallel \mathbf{x}\parallel_1\right], \tag{II.5}$$

which is a convex approximation to the L0 penalizing solution II.4. The best solution for estimating the sparse vector **x** is given by the Zero-Norm regularized estimator which is a hard combinatorial problem. These estimators, (II.3) - (II.5), can equivalently be presented as solutions to the constrained optimization problem [1], [3], [4]. This constrained optimization version of (II.5) is known as the L1 penalized L2 minimization called LASSO (Least Absolute Shrinkage and Selection Operator) or BPDN(Basis Pursuit Denosing), which can be set as Quadratic Programing (QP) and Quadratic Constrained Linear Programing (QCPL) optimization problems. [1] In the following subsection the above estimators are presented as a MAP estimator in a Bayesian framework.

B. Bayesian Framework for Sparse Signals

Equivalently, the estimator of $\mathbf{x}^0$ in (II.2) can generally be presented as MAP estimator under the Bayesian framework. Assume a prior probability distribution for **x** to be

$$p_u(\mathbf{x}) = \frac{e^{-uf(\mathbf{x})}}{\int_{\mathbf{x}\in\chi^N} e^{-uf(\mathbf{x})}d\mathbf{x}}, \tag{II.6}$$

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

$$f(\mathbf{x}) = \sum_{i=1}^{N} f(x_i). \tag{II.7}$$

such that for sufficiently large u, $\int_{x\in\chi}\exp(-uf(x))dx$ is finite as in [15]. Let the assumed variance of the noise be given by

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

where γ is system parameter which can be taken as $\gamma = \sigma_u^2 u$ where σ_u^2 is the assumed variance for each component of **w**. Note that we incorporate the sparsity in the prior pdf via $f(\mathbf{x})$. By (I.1) the probability density function of **y** given **x** is given by

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}\mid\mathbf{x};\mathbf{A}) = \frac{1}{(2\pi\sigma_u^2)^{N/2}}e^{-\frac{1}{2\sigma_u^2}\parallel \mathbf{y}-\mathbf{A}\mathbf{x}\parallel_2^2}, \tag{II.8}$$

and prior distribution of **x** by (II.6), the posterior distribution for the measurement channel (I.1) according to Bayes' law is

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}\mid\mathbf{y};\mathbf{A}) = \frac{e^{-u(\frac{1}{2\gamma}\parallel \mathbf{y}-\mathbf{A}\mathbf{x}\parallel_2^2 + f(\mathbf{x}))}}{\int_{\mathbf{x}\in\chi^N} e^{-u(\frac{1}{2\gamma}\parallel \mathbf{y}-\mathbf{A}\mathbf{x}\parallel_2^2 + f(\mathbf{x}))}d\mathbf{x}}. \tag{II.9}$$

Then the MAP estimator can be shown to be

$$\hat{\mathbf{x}}^{MAP} = \arg\min_{\mathbf{x}\in\chi^N}\left[\frac{1}{2\gamma}\parallel \mathbf{y}-\mathbf{A}\mathbf{x}\parallel_2^2 + f(\mathbf{x})\right]. \tag{II.10}$$

Now, as we choose a different penalizing function in (II.10) we get the different estimators defined above in Equations (II.3), (II.4), and (II.5) but this time under the framework of Bayesian framework as a MAP estimator [15].

1) Linear Estimators : when $f(x) = \|x\|_2^2$ (II.10) reduces to

$$\hat{\mathbf{x}}_{Linear}^{MAP} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T + \gamma\mathbf{I})^{-1}\mathbf{y}, \tag{II.11}$$

which is the LMMSE estimator.

2) LASSO Estimator: when $f(x) = \|x\|_1$ we get the LASSO estimator and (II.10) becomes

$$\hat{\mathbf{x}}_{Lasso}^{MAP} = \arg\min_{\mathbf{x}\in\chi^N}\left[\frac{1}{2\gamma}\parallel \mathbf{y}-\mathbf{A}\mathbf{x}\parallel_2^2 + \parallel \mathbf{x}\parallel_1\right] \tag{II.12}$$

3) Zero-Norm regularization estimator: when $f(x) = \|x\|_0$, we get the Zero-Norm regularization estimator and

(II.10) becomes

$$\hat{\mathbf{x}}_{Zero}^{MAP} = \arg\min_{\mathbf{x}\in\chi^N}\left[\frac{1}{2\gamma}\parallel \mathbf{y}-\mathbf{A}\mathbf{x}\parallel_2^2 + \parallel \mathbf{x}\parallel_0\right] \tag{II.13}$$

Whether these minimization problems are solvable or not, the replica analysis results can provide the asymptotic performances of all the above estimators via the replica method as given in [15], [23]–[25], [30]. We apply RS ansatz as used by Muller et al. in [30] and RSB ansatz as used by Zaidel et al.¨ [29] on vector precoding for MIMO. Actually, this work is an extension of the RSB analysis to MIMO systems done in [29] to the CS system.

III. Statistical Analysis

The performance of the Bayesian estimators like MMSE, and MAP can be done by determining the pdf of the error vector. The error is random and it should be centered about zero for the estimator to perform well. Kay showed in that way (see Section 11.6 in [31]) the performance analysis of MMSE estimator. We believe in general that inference for the asymptotic performance of MAP estimators is best done with statistical mechanical tools including RSB assumption. The outline of the replica analysis is done in this section. The reader is referred to [30], [32] and [29], for deeper

understanding of replica method which is a central idea in this section.

We begin our analysis from the posterior distribution (II.9), which is sufficient statistics to estimate $\mathbf{x}^0$ [30] and the denominator is called the normalizing factor or evidence in Bayesian inference according to [33] and Partition function in statistical mechanics [32]. Actually, it is this connection which provides the ground to apply the tools which are used in statistical mechanics. So the task of evaluating the above estimators for the sparse vector $\mathbf{x}^0$ can be translated to that of statistical physics task. Let us justify first how the analysis using statistical mechanical tool is able to do it.

The Boltzmann-Gibbs distribution is defined as

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

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

$$\mathcal{Z} = \int_{\chi^N} e^{-\beta \mathcal{H}(\mathbf{x})} d\mathbf{x}. \tag{III.2}$$

Often the Hamiltonian can be given by a quadratic form like

$$\mathrm{H}(\mathbf{x}) = \mathbf{x}^T \mathbf{J} \mathbf{x}, \tag{III.3}$$

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

$$\mathcal{E} = \int_{\mathbf{x} \in \chi^N} p_{\mathbf{x}}(\mathbf{x}) \mathcal{H}(\mathbf{x}) d\mathbf{x} \tag{III.4}$$

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

$$\mathcal{Z} = \int_{\chi^N} e^{-\beta \mathcal{H}(\mathbf{x})} d\mathbf{x}. \tag{III.5}$$

The free energy can be calculated using

$$\mathcal{F} \equiv \mathcal{E} - \frac{\mathcal{S}}{\beta}. \tag{III.6}$$

At thermal equilibrium, the energy of the system being preserved, according to the second law of thermodynamics the entropy of the system is maximized when the free energy is minimized, where β, the inverse temperature, is the Lagrange multiplier in the maximization of (III.5), subject to the mean energy constraint. Therefore, at equilibrium, the free energy can be expressed as

$$\mathcal{F} = -\frac{1}{\beta} \log \mathcal{Z} \tag{III.7}$$

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

$$\mathcal{E} = \frac{1}{N} \min_{\mathbf{x} \in \chi^N} \mathcal{H}(\mathbf{x}) \tag{III.8}$$

For our system that we address, which is given by (II.10) or equivalently by (II.2), the Hamiltonian becomes

$$\mathcal{H}(\mathbf{x}) = \frac{1}{2\sigma_u^2} (\mathbf{y} - \mathbf{A}\mathbf{x})^T (\mathbf{y} - \mathbf{A}\mathbf{x}) + u f(\mathbf{x}). \tag{III.9}$$

Compared to (III.3), the Hamiltonian in (III.9) has regularizing term in addition to the quadratic form in which the regularizing term $f(\mathbf{x})$ contributes to the additional terms involved in CS. After inserting (III.2) and (III.9) in (III.1) this gives information about the solution to (II.10) or to (II.2) in general, since they are equivalent problems. Therefore, one can use tools from statistical mechanics, which helps to analyze the performance of such estimators. For example in this paper we infer about the minimum energy of each component of $\mathbf{x}$ using these tools.

Let $\mathbf{x}^0$ and $\mathbf{x}$ being drawn from the same set, then the partition function of the posterior distribution given in (III.1) becomes

$$\mathcal{Z} = \int_{\mathbf{x} \in \chi^N} e^{-\beta \left[\frac{1}{2\sigma_u^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + u f(\mathbf{x}) \right]} d\mathbf{x}, \tag{III.10}$$

by using (III.2) and (III.9). The posterior distribution (II.9) depends on the predetermined random variables $\mathbf{y}$ and $\mathbf{A}$ called quenched states in physics literature [26], [27]. That is, we use fixed states $\mathbf{y} = \mathbf{A}\mathbf{x}^0 + \mathbf{w}$ instead of $\mathbf{y}$ for the large system limit, as $N, M \to \infty$, while maintaining N/M fixed. We then calculate the n^{th} moment of the partition function Z with respect to the predetermined variables, n replicas; hence this is where the name replica method came from. The replicated partition function is then given by

$$\mathcal{Z}^n = \int_{\{\mathbf{x}^a\}} e^{-\beta \left[\frac{1}{2\sigma_u^2} \sum_{a=1}^{n} \left(\|\mathbf{y} - \mathbf{A}\mathbf{x}^a\|_2^2 \right) + \frac{\gamma}{\sigma_u^2} \sum_{a=1}^{n} f(\mathbf{x}^a) \right]} \prod_{a=1}^{n} d\mathbf{x}^a \tag{III.11}$$

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

$$\mathcal{Z}^n = \int_{\{\mathbf{x}^a\}} e^{-\beta \left[\frac{1}{2\sigma_u^2} \sum_{a=1}^{n} \left(\|\mathbf{A}(\mathbf{x}^0 - \mathbf{x}^a) + \mathbf{w}\|_2^2 \right) + \frac{\gamma}{\sigma_u^2} \sum_{a=1}^{n} f(\mathbf{x}^a) \right]} \prod_{a=1}^{n} d\mathbf{x}^a \tag{III.12}$$

Averaging (III.12) over the noise $\mathbf{w}$, we get

$$\int_{\mathbb{R}^M} \frac{d\mathbf{w}}{\pi^M} e^{-\frac{1}{2\sigma_0} (\mathbf{w}^T \mathbf{w})} \mathcal{Z}^n$$

$$= \alpha^{N/2} \int_{\{\mathbf{x}^a\}} e^{-\beta \left[\frac{1}{2} Tr \mathbf{J} \mathbf{L}(n) + \frac{\gamma}{\sigma_u^2} \sum_{a=1}^{n} f(\mathbf{x}^a) \right]} \prod_{a=1}^{n} d\mathbf{x}^a, \tag{III.13}$$

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

$$\mathbf{J} = \mathbf{O}\mathbf{D}\mathbf{O}^{-1}, \tag{III.14}$$

where **D** is a diagonal matrix while **O** is $N \times N$ orthogonal matrix assumed to be drawn randomly from the uniform distribution defined by the Haar measure on the orthogonal group. For more clarity on this one can refer to in [26], [34]. $\mathbf{L}(n)$ is given by

$$\mathbf{L}(n) = -\frac{1}{\sigma_u^2}\sum_{a=1}^{n}(\mathbf{x}^0 - \mathbf{x}^a)(\mathbf{x}^0 - \mathbf{x}^a)^T + \frac{\sigma_0^4}{\sigma_u^2(\sigma_u^2 + n\sigma_0^2)}\left(\sum_{a=1}^{n}(\mathbf{x}^0 - \mathbf{x}^a)\right)\left(\sum_{b=1}^{n}(\mathbf{x}^0 - \mathbf{x}^b)\right)^T. \quad \text{(III.15)}$$

Further averaging what we get on the right hand side of (III.13) over the cross correlation matrix **J**, by assuming the eigenvalue spectrum of **J** to be self-averaging, we get

$$\underset{w,J}{E}\{\mathcal{Z}^n\} = \underset{J}{E}\left(\alpha^{N/2}\int_{\{\mathbf{x}^a\}} e^{-\beta\psi_1(\mathbf{x}^a,\mathbf{J},\mathbf{L}(n))}\prod_{a=1}^{n}d\mathbf{x}^a\right) = \alpha^{N/2}\int_{\{\mathbf{x}^a\}} e^{\frac{-\beta\gamma}{\sigma_u^2}\sum_{a=1}^{n}f(\mathbf{x}^a)}\psi_2(\mathbf{J},\mathbf{L}(n))\prod_{a=1}^{n}d\mathbf{x}^a, \quad \text{(III.16)}$$

The inner expectation in (III.16) is the Harish -Chandra Itzykoson-Zuber integral (again see in [28], [29] and the references therein). The plan here is to evaluate the fixed-rank matrices $\mathbf{L}(n)$ as $N \to \infty$. Further following the explanation in [29] (III.16) becomes

$$\underset{w,J}{E}\{\mathcal{Z}^n\} = \alpha^{N/2}\int_{\{\mathbf{x}^a\}}\psi_3(\mathbf{x}^a,\mathbf{J},\mathbf{L}(n))\prod_{a=1}^{n}d\mathbf{x}^a, \quad \text{(III.17)}$$

where

$$\psi_3(\mathbf{x}^a;R(-v)) = e^{\frac{-\beta\gamma}{\sigma_u^2}\sum_{a=1}^{n}f(\mathbf{x}^a)}e^{-N\sum_{a=1}^{n}\int_0^{\lambda a}R(-v)dv+o(N)} \quad \text{(III.18)}$$

where $R(v)$ is the R-transform of the limiting eigenvalue distribution of the matrix **J** (see, definition 1 in [28] of Rtransform or in [14] for better understanding of R-transform) and $\{\lambda_a\}$ denote the eigenvalues of $\mathbf{L}(n)$ as explained in [28], [29], [34].

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

$$\beta\bar{\mathcal{F}} = -\lim_{N\to\infty}\frac{1}{N}\underset{w,J}{E}\{\log\ \mathcal{Z}\} = -\lim_{N\to\infty}\frac{1}{N}\lim_{n\to 0}\frac{\partial}{\partial n}\log\ \underset{w,J}{E}\{(\mathcal{Z})^n\}. \quad \text{(III.19)}$$

and then we calculate the cost function of the L_p-norm reconstruction using the quenched average of the free energy [25], [35] as

$$\bar{\mathcal{E}} = \lim_{\beta\to\infty}\frac{1}{\beta}\bar{\mathcal{F}} = -\lim_{\beta\to\infty}\frac{1}{\beta}\lim_{N\to\infty}\tfrac{1}{N}\underset{w,J}{E}\{\log\ \mathcal{Z}\} = -\lim_{\beta\to\infty}\frac{1}{\beta}\lim_{n\to 0}\frac{\partial}{\partial n}\underbrace{\lim_{N\to\infty}\tfrac{1}{N}\log\ \underset{w,J}{E}\{(\mathcal{Z})^n\}}_{\Xi_n}. \quad \text{(III.20)}$$

where we get (III.20) by using one of the assumptions used in replica calculations, after interchanging the order of the limits we assumed we get the same result. Further, for Ξ_n we have

$$\Xi_n = -\lim_{N\to\infty}\tfrac{1}{N}\log\left(\alpha^{N/2}\int_{\{\mathbf{x}^a\}}\psi_4(\mathbf{x}^a;R(-v))\prod_{a=1}^{n}d\mathbf{x}^a\right). \quad \text{(III.21)}$$

where

$$\psi_4(\mathbf{x}^a;R(-v)) = e^{\frac{-\beta\gamma}{\sigma_u^2}\sum_{a=1}^{n}f(\mathbf{x}^a)}e^{\sum_{a=1}^{n}\int_0^{\lambda a}R(-v)dv}. \quad \text{(III.22)}$$

Since the additive exponential terms of order $o(N)$ have no effect on the results when taking saddle point integration in the limiting regime as $N \to \infty$ due to the factor $\frac{1}{N}$ outside the logarithm in (III.21) any such terms are dropped further for notational simplicity as in [29].

In order to find the summation in (III.21) we employed the procedure in [29] and the nN dimensional space spanned by the replicas is split into subshells, defined through $n\times n$ matrix **Q**

$$S(\mathbf{Q}) = \{\mathbf{x}^1,\ldots,\mathbf{x}^n \mid (\mathbf{x}^0 - \mathbf{x}^a)^T(\mathbf{x}^0 - \mathbf{x}^b) = NQ_{ab}\}. \quad \text{(III.23)}$$

The limit $N \to \infty$ enables us to use saddle point integration. Hence we can have the following general result as similar to [29] but extended in this work with the term, which pertains to CS.

Proposition 1. *The energy* E *from* (III.8)*, for any inverse temperature β, any structure of* **Q** *consistent with* (III.23)*, and*

R-transform $R(\cdot)$*, such that* $R(\mathbf{Q})$ *is well-defined, is given by*

$$\bar{\mathcal{E}} = -\lim_{n\to 0}\frac{1}{n}Tr[\mathbf{Q}R(-\beta\mathbf{Q})], \quad \text{(III.24)}$$

where **Q** *is the solution to the saddle point equation*

$$\mathbf{Q} = \int\frac{\int_{\{\tilde{\mathbf{x}}\in\chi^n\}}\psi_2(\tilde{\mathbf{x}})e^{\psi_1(\tilde{\mathbf{x}})+\frac{1}{2}\ln\alpha-\frac{\beta\gamma}{\sigma_u^2}\tilde{\mathbf{x}}}d\tilde{\mathbf{x}}}{\int_{\{\tilde{\mathbf{x}}\in\chi^n\}}e^{\psi_1(\tilde{\mathbf{x}})+\frac{1}{2}\ln\alpha-\frac{\beta\gamma}{\sigma_u^2}\tilde{\mathbf{x}}}d\tilde{\mathbf{x}}}dF_{X^0}(x^0), \quad \text{(III.25)}$$

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

Proof: $R(\mathbf{Q})$ is well-defined means that the R-transform of the matrix $\mathbf{Q}$ is defined. For the sake of space we have omitted the proof in this paper. Interested readers can see Appendix B given at http://arxiv.org/abs/1409.2303. ■

Further, to simplify the result in (III.24), we assume a simple structure on to the $n \times n$ cross correlation matrix $\mathbf{Q}$ at the Saddel point. So we assume two different assumptions for the entries of $\mathbf{Q}$ called ansatz: Replica Symmetry(RS) and one step Replica Symmetric Breaking (1RSB) ansatz. For example given the convexity of the energy function (II.12), the replica symmetric ansatz for the saddle point is reasonable. Whereas the energy function in (II.13) is not convex and assuming 1RSB assumption can suffice. Since the CS recovery estimators applied in this paper are redefined as MAP estimators and generally it is believed that using 1RSB ansatz is enough to analyse such estimators if the local stability of the RS saddle point is lost against the perturbations [29], [36].

Showing the derivation of (III.24) analytically and further simplifying it using additional assumptions given under, is the central purpose of this paper as it was done in [28] and [29] for MIMO systems, but for a different purposes.

As in [28] and [29] we assume that the ansatz are given by:

1) replica symmetry ansatz :

$$\mathbf{Q} = q_0 \mathbf{1}_{n\times n} + \frac{b_0}{\beta}\mathbf{I}_{n\times n} \tag{III.26}$$

2) one replica symmetry breaking ansatz:

$$\mathbf{Q} = q_1 \mathbf{1}_{n\times n} + p_1 \mathbf{I}_{\frac{n\beta}{\mu_1}\times\frac{n\beta}{\mu_1}} \otimes \mathbf{1}_{\frac{\mu_1}{\beta}\times\frac{\mu_1}{\beta}} + \frac{b_1}{\beta}\mathbf{I}_{n\times n} \tag{III.27}$$

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

A. LASSO Estimator with RS Ansatz

Consider the LASSO estimator given in (II.12), which is equivalent to the solution of the main unconstrained optimization problem (II.2) in l_1 penalized sense. Its performance is calculated via energy per component, $\overline{\mathcal{E}}$, using two macroscopic variables q_0 and b_0 given by

$$q_0 = \int_{\mathbb{R}}\int_{\mathbb{C}} \left| x^0 - \Psi_1(x) \right|^2 Dz dF_{X^0}(x^0), \tag{III.28}$$

$$b_0 = \frac{1}{f_0}\int_{\mathbb{R}}\int_{\mathbb{C}} \Re\left\{ x^0 - \Psi_1(x) z^* \right\} Dz dF_{X^0}(x^0), \tag{III.29}$$

where

$$\Psi_1(x) = \arg\min_{x\in\chi}\left[\left| -z f_0 + 2e_0(x^0 - x) - \frac{\gamma}{\sigma_u^2} \right| \right], \tag{III.30}$$

$$e_0 = \frac{1}{\sigma_u^2} R\left(\frac{-b_0}{\sigma_u^2}\right), \tag{III.31}$$

$$f_0 = \sqrt{2\frac{q_0}{\sigma_u^4} R'\left(\frac{-b_0}{\sigma_u^2}\right)}, \tag{III.32}$$

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

Proposition 2. *Given the LASSO estimator in* (II.12) *and the macroscopic variables* q_0 *and* b_0, *in addition given the conditions in proposition 1, the typical value (per element) of the minimized cost function* (III.24) *simplifies to*

$$\bar{\mathcal{E}}_{rs}^{lasso} = \frac{q_0}{\sigma_u^2} R\left(\frac{-b_0}{\sigma_u^2}\right) - \frac{b_0 q_0}{\sigma_u^4} R'\left(\frac{-b_0}{\sigma_u^2}\right) \tag{III.33}$$

Proof: See Appendix C in the article given at http://arxiv.org/abs/1409.2303. ■

B. LASSO Estimator with 1RSB Ansatz

Moving further to the RSB ansatz instead of assuming RS ansatz, we get more macroscopic parameters involved in the calculation of the quantity given by (III.24): b_1, p_1, q_1, and μ_1. These are given by the following fixed point equations as $n \to 0$ and $\beta \to \infty$, and using the compact notation as in [29]. Let

$$\Delta(y,z) \equiv e^{-\mu_1 \min_{x\in\chi} \psi_5(x,y,z)}, \tag{III.34}$$

where

$$\psi_5(x,y,z) = -2\Re\{(x^0 - x)(f_1 z^* + g_1 y^*)\} + e_1(x^0 - x)^2 - \frac{\gamma}{\sigma_u^2}|x| \quad \text{for } (y,z) \in \Re^2 \tag{III.35}$$

and its normalized version

$$\tilde{\Delta}(y,z) = \frac{\Delta(y,z)}{\int_{\mathbb{C}} \Delta(\tilde{y},z) d\tilde{y}} \tag{III.36}$$

$$b_1 + p_1\mu_1 = \frac{1}{f_1}\int_{\mathbb{R}}\int_{\mathbb{C}^2} \Re\Big\{\Big(x^0 - \Psi_2\Big)z^*\Big\} DyDzdF_{X^0}(x^0) \quad \text{(III.37)}$$

$$b_1 + (q_1 + p_1)\mu_1 = \frac{1}{g_1}\int_{\mathbb{R}}\int_{\mathbb{C}^2} \Re\Big\{\Big(x^0 - \Psi_2\Big)y^*\Big\} DyDzdF_{X^0}(x^0) \quad \text{(III.38)}$$

$$q_1 + p_1 = \frac{1}{g_1}\int_{\mathbb{R}}\int_{\mathbb{C}^2} |\Psi_2|^2 DyDzdF_{X^0}(x^0) \quad \text{(III.39)}$$

and

where

$$\int_{\frac{b_1}{\sigma_u^2}}^{\frac{b_1+\mu_1 p_1}{\sigma_u^2}} R(-v)dv = -R(-\frac{b_1+\mu_1 p_1}{\sigma_u^2}) \quad \text{(III.40)}$$

$$-\mu_1^2\Big((q_1+p_1)g_1^2 + p_1 f_1^2\Big) + \int_{\mathbb{R}}\int_{\mathbb{C}} \log\Big(\int_{\mathbb{C}} \Delta(y,z)Dy\Big) DzdF_{X^0}(x^0), \quad \text{(III.41)}$$

where

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

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

$$e_1 = \frac{1}{\sigma_u^2}R(\frac{-b_1}{\sigma_u^2}), \quad \text{(III.42)}$$

$$g_1 = \sqrt{\frac{1}{\mu_1\sigma_u^2}\Big[R(\frac{-b_1}{\sigma_u^2}) - R(\frac{-b_1-\mu_1 p_1}{\sigma_u^2})\Big]}, \quad \text{(III.43)}$$

$$f_1 \overset{n\to 0}{\longrightarrow} \frac{1}{\sigma_u^2}\sqrt{q_1 R'(\frac{-b_1-\mu_1 p_1}{\sigma_u^2})} \quad \text{(III.44)}$$

and where $dF_{X^0}(x^0)$ is a probability measure of x^0 where as *Dy* and *Dz* refers to integration over Gaussian measure. All these equations are shown in Appendix D of the article given at http://arxiv.org/abs/1409.2303. The use of RSB ansatz for a convex optimization problem like LASSO can be unnecessary if the RS ansatz provide the global minimum solution. However, the use of the RSB ansatz becomes crucial for the CS problem since the best solution is provided by the Zero-Norm regularizing estimator, which is a non-convex problem. The next two propositions are provided as an extension of the proposition in [29] to CS problems.

Proposition 3. *Given the LASSO estimator in* (II.12) *and suppose the random matrix* **J** *satisfies the decomposability property* (III.14). *Then under some technical assumptions, including one-step replica symmetry breaking, and the macroscopic variables given by the above fixed point equations, the effective typical value of the minimized cost function per component converges in probability as N, M → ∞, N/M < ∞ , to*

$$\bar{\varepsilon}_{1rsb}^{LASSO} = \frac{1}{\sigma_u^2}(q_1 + p_1 + \frac{b_1}{\mu_1})R(\frac{-b_1-\mu_1 p_1}{\sigma_u^2}) - \frac{b_1}{\mu_1\sigma_u^2}R(-\frac{b_1}{\sigma_u^2}) - q_1(\frac{b_1+\mu_1 p_1}{\sigma_u^4})R'(\frac{-b_1-\mu_1 p_1}{\sigma_u^2}) \quad \text{(III.45)}$$

Proof: See Appendix D in the article given at http://arxiv.org/abs/1409.2303. ■

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

$$\varepsilon_{1rsb}^{LASSO} \to \varepsilon_{RS}^{LASSO} \quad \text{(III.46)}$$

C. Zero-Norm Regularizing Estimator with 1RSB Ansatz

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

Proposition 4. *Given the Zero-Norm regularizing estimator in* (II.13) *and suppose the random matrix* **J** *satisfies the decomposability property* (III.14). *Then under some technical assumptions, including one-step replica symmetry breaking, the effective energy penalty per component converges in probability as N, M → ∞, N/M < ∞ , to*

$$\bar{\varepsilon}_{1rsb}^{zero\text{-}norm} = \frac{1}{\sigma_u^2}(q_1 + p_1 + \frac{b_1}{\mu_1})R(\frac{-b_1-\mu_1 p_1}{\sigma_u^2}) - \frac{b_1}{\mu_1\sigma_u^2}R(-\frac{b_1}{\sigma_u^2}) - q_1(\frac{b_1+\mu_1 p_1}{\sigma_u^4})R'(\frac{-b_1-\mu_1 p_1}{\sigma_u^2})$$

(III.47)

Proof: See Appendix D in the article given at http://arxiv.org/abs/1409.2303.

■

Note that even though (III.45) and (III.47) have the same expression they are different since the macroscopic variables are calculated from different prior distributions for **x** and $\mathbf{x}^0$.

IV. PARTICULAR EXAMPLE: BERNOULLI-GAUSSIAN MIXTURE DISTRIBUTION

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

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

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

A. Replica Symmetric Analysis

Considering the macroscopic variables given by (III.28) and (III.29) and inserting the assumed distributions above and simplfying, the fixed point equations become

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

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

(IV.3)

where e_0 and f_0 are given by (III.31) and (III.32), respectively, also

$$\Psi_3(z) = \left|\frac{z f_0 + \frac{\gamma}{\sigma_u^2}}{2e_0}\right|^2 e^{-|z|^2}, \quad \text{(IV.4)}$$

$$\Psi_4(z,x^0) = \left|\frac{z f_0 + \frac{\gamma}{\sigma_u^2}}{2e_0}\right|^2 e^{-\left(\frac{(x^0)^2}{2}+|z|^2\right)}, \quad \text{(IV.5)}$$

$$\Psi_5(z) = \Re\left\{\left(\frac{z f_0 + \frac{\gamma}{\sigma_u^2}}{2e_0}\right)z^*\right\}e^{-|z|^2}, \text{and} \quad \text{(IV.6)}$$

$$\Psi_6(z,x^0) = \Re\left\{x^0(1-z^*) + \left(\frac{z f_0 + \frac{\gamma}{\sigma_u^2}}{2e_0}\right)z^*\right\}e^{-\left(\frac{(x^0)^2}{2}+|z|^2\right)}$$

.(IV.7)

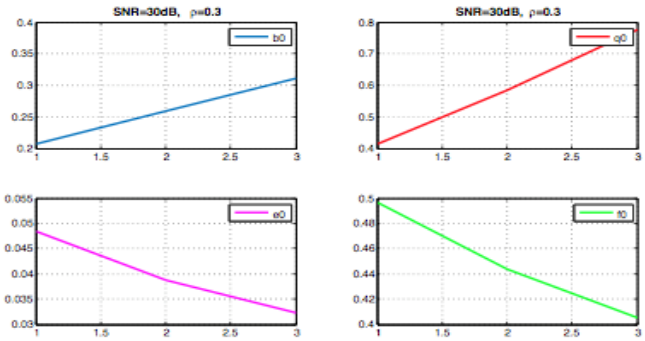

Fig. 2: The macroscopic variables of the RS ansatz for LASSO versus the measurement ratio M/N.

With the defined values for the parameters given in this subsection above we first plot these macroscopic variables and this is shown in Figure 2. Using these macroscopic variables, we plot the minimized cost function per component versus measurement ratio M/N in Figure 3 which is given under proposition 2 for different sparsity ratios (ρ = 0.1, ρ = 0.3).

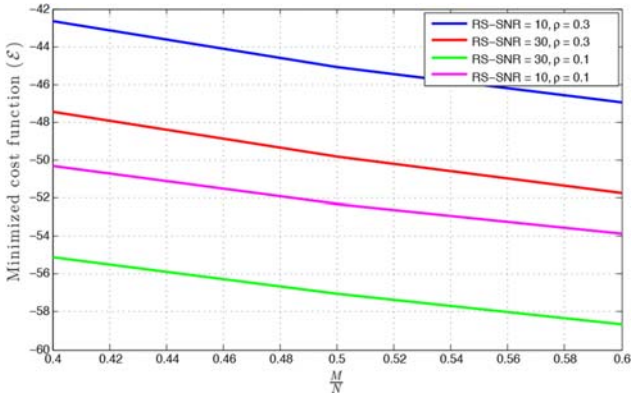

Fig. 3: The effective typical value of the minimized cost function per component against measurement ratio M/N for different sparsity.

B. Replica Symmetric Breaking Analysis

Considering the same Bernoulli-Gaussian mixture distribution (IV.1) assumed in this section we consider the macroscopic variables, which arises from 1RSB ansatz. Then the effective typical value of the minimized cost function per component as $M \to \infty$, $N \to \infty$, while M/N is finite, which are given by (III.45) and (III.47) are dependent up on four macroscopic variables given by (III.37)-(III.40).

It is possible to simplify these equations further and give numerical results. But this is deferred for further work. We expect that the free energy from the RSB ansatz to be greater than the free energy from the RS ansatz for the Zero-Norm regularizing, which can be seen from the analytical terms which have more parameters in (III.47). However, for LASSO these free energies, hence the typical value of the minimized cost function, will be quite similar since for convex minimization problems there is one global minimum and RS ansatz is sufficient enough to produce the solution.

V. CONCLUSION

In this paper we have used the replica method to analyze the performance of the estimators used in compressed sensing in which we generalized them as MAP estimators. The performance of MAP estimators can be shown using the replica method. 1RSB ansatz can be enough to analyze such estimators [29]. We have only shown here one particular example for the CS problem, i.e. for Bernoulli-Gaussian distribution. One may be interested to verify it using different examples. In addition we have only compared the performance of the estimators based on the free energy, but one can also use other metrics such as comparing the input/output distribution using replica analysis, as it is done in [29]. The main result of this paper is analytical analysis

for the performance of the estimators used in CS. This issues and others, like doing the numerical analysis for the 1RSB, are left for future work.

ACKNOWLEGMENT

We are grateful to Ralf R. Muller, Lars Lundheim, the late Birhutesfa Godana, Faraz Barzideh, Rodrigo Vicente de Miguel and Benjamin M. Zaidel for interesting discussions and suggestions.

APPENDIX

A. Green's function

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

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

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

B. R-transform

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

$$R(z) \equiv \sum_{n=1}^{\infty} C_n z^{n-1}, \tag{A.2}$$

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

$$G(R(z)+\frac{1}{z}) = z \quad \text{or} \quad R(G(z)) + \frac{1}{G(z)} = z. \tag{A.3}$$

The spectral density of the matrix $\mathbf{J} = \mathbf{A}^T\mathbf{A}$ converges almost surely to the Marchenko-Pastur law as $M = \alpha N \to \infty$ [28]. And the R-transform of this matrix is given by

$$R(z) = \frac{1}{1-\alpha z} \tag{A.4}$$

and its derivative with respect to z becomes

$$R'(z) = \frac{\alpha}{(1-\alpha z)^2}, \tag{A.5}$$

where $\alpha = N/M$ is system load. (One can see also about macroscopic versus microscopic properties and Self-averaging in the appendix of http://arxiv.org/abs/1409.2303).

C. Partition Function

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

$$P_\beta(\mathbf{x}|\mathbf{y}) = \frac{1}{\mathcal{Z}}e^{-\beta\mathcal{H}(\mathbf{x})}, \tag{A.6}$$

as in (III.1) for the CS system under consideration, where $Z(\beta;\mathbf{y})$ is the *partition function*,

$$\mathcal{Z} = \int_{\mathbf{x}\in\chi^N} e^{-\beta\left[\frac{1}{2\sigma_0^2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \mathrm{u}f(\mathbf{x})\right]}d\mathbf{x}, \tag{A.7}$$

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

D. Free Energy

One of the important macroscopic quantities in statistical mechanics which is also a self averaging quantity is *free energy* [37], [38], [39], [14], [12] and [29], defined as

$$\mathcal{F} = \mathcal{E} - \beta\,\mathcal{P} \tag{A.8}$$

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

$$\mathcal{F} = -\beta^{-1}\log \mathrm{Z} \tag{A.9}$$

$$\mathcal{E} = \partial(\mathcal{F}\beta)/\partial\beta \quad \text{and} \quad \mathcal{P} = -\partial\mathcal{F}/\partial(\frac{1}{\beta}). \tag{A.10}$$

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

$$\mathcal{F} = -\beta^{-1}\mathbb{E}(\log \mathcal{Z}). \tag{A.11}$$

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

E. Replica Method

Lets as assume that we are interested in evaluating $\mathbb{E}(\log z)$ in relation to the average free energy as shown above in equation (A.11). In order to ease the difficulty the following identity is used:

$$\log(\mathcal{Z}) = \lim_{n\to 0}\frac{\partial}{\partial n}\log \mathcal{Z}^n. \tag{A.12}$$

Using this identity (A.12), $\mathbb{E}(\log z)$ can be expressed as

$$\mathbb{E}(\log \mathcal{Z}) = \lim_{n\to 0}\frac{\partial}{\partial n}\log \mathbb{E}(\mathcal{Z}^n). \tag{A.13}$$

At this stage a number of assumptions are made: It is assumed that the limit and the expectation can be interchanged giving the opportunity to to deal with $E(Z^n)$ whenever dealing with $\mathbb{E}(\log \mathcal{Z})$ is difficult task. If n is a natural number then first we calculate Z^n, i.e., *n replicas* of the random variable. Then in order to calculate the derivative and the limit, n has to be continues variable. This is another critical assumption! These assumptions together with (A.13) is called *replica trick*, and this is where the name replica and the method, *replica method*, has evolved from. It is only recently that some of the assumptions made here have been rigorously proved [40]. Therefore, replica method is now a standard technique to study the free energy of disordered systems. Applying the replica method the averaged free energy given in (A.11) becomes

$$\begin{aligned} \bar{\mathcal{F}} &= -\lim_{N\to\infty} \mathbb{E}\left(\frac{1}{N\beta}\log \mathcal{Z}\right) \\ &= -\lim_{N\to\infty}\frac{1}{N\beta}\lim_{n\to 0}\frac{\partial}{\partial n}\log \mathbb{E}(\mathcal{Z}^n) \qquad \text{(A.14)} \\ &= -\lim_{n\to 0}\frac{\partial}{\partial n}\lim_{N\to\infty}\frac{1}{N\beta}\log \mathbb{E}(\mathcal{Z}^n). \qquad \text{(A.15)} \end{aligned}$$

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

$$\mathbb{E}(\mathcal{Z}^n(\beta;\mathbf{y})) = \mathbb{E}\Big(\prod_{a=1}^{n} Z_a(\beta;\mathbf{y})\Big). \qquad \text{(A.16)}$$

F. Replica Symmetry

The replica ansatz are arbitrary and often assumed to be independent random variables. This makes hard to calculate the expectation of the right hand side of (A.16). However, it is possible to assume symmetry among the replica ansatz, called *replica symmetry* (RS) ansatz here. Since the RS assumption on the ansatz has been shown to be sufficient if the optimization problem under consideration is a convex one [38], [40] [39], [14], [12], [29]. And the CS minimization problems, given by (II.11), and (II.12) for p = 2 and p = 1, is a convex optimization problem. For any fixed realization of the quenched disordered $\mathbf{A}$, the replicated sparse vector $\mathbf{x}^a$ are independent, and marginalizing (integrating) over the disorder introduces attractive interactions between replicas [41], and the interactions between replicas depends only on the overlap matrix $\mathbf{Q}$, with entries given by $\mathbf{Q}_{ab} = \frac{1}{N}\mathbf{x}^a \cdot \mathbf{x}^b$ where $a, b = 0,1,2,\cdots,n$. The landscape described by this overlap matrix $\mathbf{Q}$ is very complicated and simplification is needed, i.e., by assuming RS ansatz.

G. Replica Symmetry Breaking

The assumption of replica symmetry ansatz is a very simplified model for some problems. If the local stability of the RS saddle point is lost against the perturbations that break it then the solution fails to realize. The local instability of the RS solution is verified by the de Almeida-Thouless (AT) condition [40]. Therefore, replica symmetry breaking ansats has to be used. This corresponds to the physical picture in which there are many free energy landscapes with many valleys. A good figurative explanation is provided in [41] (see figure 4). For example if we consider the CS minimization problem, given by (II.13) for p = 0, it is a non-convex optimization problem, and the expression for the minimized l_p-norm per element, C_p, can not be calculated using the replica symmetry ansatz. In such cases the symmetry breaks. Therefore one has to use the replica symmetry breaking ansatz.

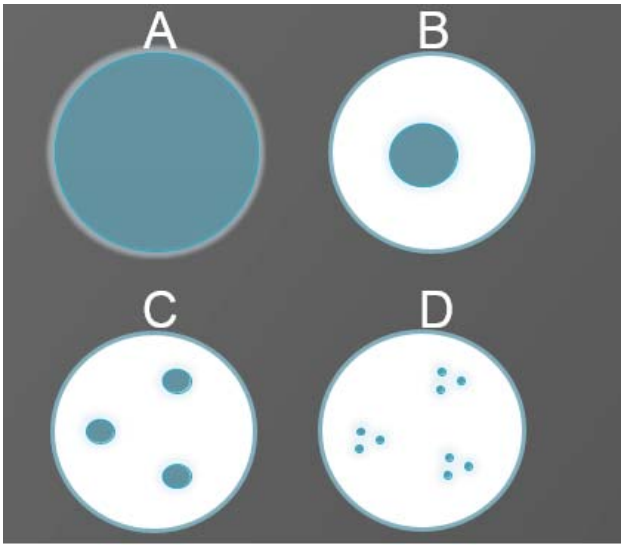

Fig. 4: Probability lumps in free energy valleys. The larger circles represent the space of all possible sparse vectors. The shaded regions represent the space of configurations (free energy valleys) with non-negligible probability under the Boltzman-Gibbs distribution. (A) At high temperature all configurations are explored by this distribution. (B) The replica symmetric ansatz for a low temperature phase. The interaction between replicas freeze into small set of configurations. (C) One possible ansatz for replica symmetry breaking (RSB) in which the replica overlap matrix Q is characterized by two order parameters. (D) There exists a series of k-step RSB schemes describing scenarios in which the distribution decomposes into a nested hierarchy of lumps of depth k, Here k=2. (Provided by Advani and et al [41]).

REFERENCES

[1] D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289 – 1306, Apr. 2006.

[2] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[3] E. Candes, R. J., and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,"

Information Theory, IEEE Transactions on, no. 2, pp. 489 – 509, Feb. 2006.

[4] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, no. 12, pp. 5406–5425, 2006.

[5] S. Qaisar, R. Bilal, W. Iqbal, M. Naureen, and S. Lee, "Compressive sensing: From theory to applications, a survey," *Communications and Networks, Journal of*, vol. 15, no. 5, pp. 443–456, Oct 2013.

[6] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007. [Online]. Available: http://dx.doi.org/10.1002/mrm.21391

[7] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Measurements vs. bits: Compressed sensing meets information theory," in *in Proc. Allerton Conf. on Comm., Control, and Computing*, 2006.

[8] M. A. T. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 4, pp. 586–597, Dec 2007.

[9] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *Information Theory, IEEE Transactions on*, vol. 55, no. 5, pp. 2230–2249, May. 2009.

[10] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[11] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *Information Theory, IEEE Transactions*, vol. 47, no. 7, pp. 2845 – 2862, Nov. 2001.

[12] D. Guo and S. Verdu, "Randomly spread cdma: asymptotics via statistical physics," *Information Theory, IEEE Transactions on*, vol. 51, no. 6, pp. 1983–2010, June 2005.

[13] T. Tanaka, "A statistical-mechanics approach to large-system analysis of cdma multiuser detectors," *Information Theory, IEEE Transactions on*, vol. 48, no. 11, pp. 2888–2910, Nov 2002.

[14] R. R. Muller, G. Alfano, B. M. Zaidel, and R. de Miguel, "Applications¨ of large random matrices in communications engineering," *CoRR*, vol. abs/1310.5479, 2013.

[15] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of map estimation via the replica method and applications to compressed sensing," *CoRR*, vol. abs/0906.3234, 2009.

[16] D. Baron, S. Sarvotham, and R. Baraniuk, "Bayesian compressive sensing via belief propagation," *Signal Processing, IEEE Transactions on*, vol. 58, no. 1, pp. 269–280, Jan 2010.

[17] S. Rangan, "Estimation with random linear mixing, belief propagation and compressed sensing," *CoRR*, vol. abs/1001.2228, 2010.

[18] D. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012, pp. 1231–1235.

[19] A. Maleki, L. Anitori, Z. Yang, and R. Baraniuk, "Asymptotic analysis of complex lasso via complex approximate message passing (camp)," *Information Theory, IEEE Transactions on*, vol. 59, no. 7, pp. 4290– 4308, July 2013.

[20] F. Krzakala, M. Mzard, F. Sausset, Y. Sun, and L. Zdeborov, "Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices," Journal of Statistical Mechanics: Theory and Experiment, vol. 2012, no. 08, p. P08009, 2012. [Online]. Available: http://stacks.iop.org/1742-5468/2012/i=08/a= P08009.

[21] A. Tulino, G. Caire, S. Verdu, and S. Shamai, "Support recovery with sparsely sampled free random matrices," *Information Theory, IEEE Transactions on*, vol. 59, no. 7, pp. 4243–4271, July 2013.

[22] J. Ma, X. Yuan, and L. Ping, "Turbo compressed sensing with partial dft sensing matrix," *Signal Processing Letters, IEEE*, vol. 22, no. 2, pp. 158–161, Feb 2015.

[23] Y. Kabashima, T. Wadayama, and T. Tanaka, "A typical reconstruction limit for compressed sensing based on lp-norm minimization," *Journal of Statistical Mechanics: Theory and Experiment*, vol. Volume 2009, September 2009.

[24] S. Ganguli and H. Sompolinsky, "Statistical mechanics of compressed sensing," *Phys. Rev. Lett.*, vol. 104, p. 188701, May 2010. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.104.188701

[25] K. Takeda and Y. Kabashima, "Statistical mechanical analysis of compressed sensing utilizing correlated compression matrix," *CoRR*, vol. abs/1001.4361, 2010.

[26] S. U. K. Takeda and Y. Kabashima, "Analysis of cdma systems that are characterized by eigenvalue spectrum," *EPL (Europhysics Letters)*, vol. 76, p. 1193, 2010. [Online]. Available: http://iopscience.iop.org/ 0295-5075/76/6/1193

[27] K. Takeda, A. Hatabu, and Y. Kabashima, "Statistical mechanical analysis of the linear vector channel in digital communication," *CoRR*, vol. abs/0707.3336, 2007.

[28] R. Muller, D. Guo, and A. Moustakas, "Vector precoding for wireless¨ mimo systems and its replica analysis," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 3, pp. 530–540, April 2008.

[29] B. Zaidel, R. Muller, A. Moustakas, and R. De Miguel, "Vector precoding for gaussian mimo broadcast channels: Impact of replica symmetry breaking," *Information Theory, IEEE Transactions on*, vol. 58, no. 3, pp. 1413–1440, March 2012.

[30] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, Sept 2009, pp. 52–59.

[31] S. M. Kay, *Fundamentals of statistical signal processing : estimation theory.*, 1st ed. Prentice-Hall PTR, apr 2010. [Online]. Available: http://www.amazon.com/exec/obidos/redirect?tag= citeulike07-20&path=ASIN/0133457117

[32] R. Podgornik, "Book review: Statistical physics of spin glasses and information processing. hidetoshi nishimori, oxford science publications, oxford university press, 2001," *Journal of Statistical Physics*, vol. 109, no. 1-2, pp. 335–337, 2002. [Online]. Available: http://dx.doi.org/10.1023/A%3A1020091132740

[33] D. J. C. MacKay., *Information Theory, Inference and Learning Algorithms*, ser. Technical report (Stanford University. Department of Statistics). Stanford, Calif. : Division of Biostatistics, Stanford University, 2005, no. 2005-19B/233.

[34] K. Takeda, A. Hatabu, and Y. Kabashima, "Statistical mechanical analysis of the linear vector channel in digital communication," *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 47, p. 14085, 2007. [Online]. Available: http://stacks.iop.org/1751-8121/40/i=47/a= 004

[35] K. Takeda and Y. Kabashima, "Statistical mechanical assessment of a reconstruction limit of compressed sensing: Toward theoretical analysis of correlated signals," *Europhysics Letters Association*, vol. EPL 95 18006, 2011.

[36] T. Uezu, M. Yoshida, T. Tanaka, and M. Okada, "Statistical mechanical analysis of CDMA multiuser detectors: At stability and entropy of the RS solution, and 1RSB solution," *Progress of Theoretical Physics Supplement*, vol. 157, pp. 254–257, 2005. [Online]. Available: http://ptps.oxfordjournals.org/content/157/254.abstract

[37] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *Journal of Physics F: Metal Physics*, vol. 5, no. 5, p. 965, 1975. [Online]. Available: http://stacks.iop.org/0305-4608/5/i=5/a=017

[38] Seidenfeld, "R. a. fisher's fiducial argument and bayes' theorem," *Institute of Mathematical Statistics*, vol. 7, no. 3, pp. 358–368, Aug., 1992. [Online]. Available: http://www.jstor.org/stable/2246072

[39] T. Tanaka, "Moment problem in replica method," *Interdisciplinary information sciences*, vol. 13, pp. 17–23, 2007. [Online]. Available: http://hdl.handle.net/10097/17348

[40] M. Talagrand, "The parisi formula," *Annals of Math.*, vol. 163, pp. 221– 263, 2006. [Online]. Available: http://www.jstor.org/stable/20159953

[41] M. Advani, S. Lahiri, and S. Ganguli, "Statistical mechanics of complex neural systems and high dimensional data," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2013, no. 03, p. P03014, 2013. [Online]. Available: http://stacks.iop.org/1742-5468/2013/i=03/a=P03014